Curating reserve level species lists in an era of diverse and dynamic data sources

Elizabeth Wenk[a,*], Thomas Mesaglio[a,b,*,†], David Keith[c,d], Will Cornwell[a]

[a]Evolution & Ecology Research Centre; School of Biological, Earth and Environmental Sciences; UNSW Sydney; Sydney, NSW, Australia

[b]CSIRO Atlas of Living Australia, Sydney, NSW, Australia

[c]Centre for Ecosystem Science; School of Biological, Earth and Environmental Sciences; UNSW Sydney; Sydney, NSW, Australia

[d]IUCN Commission on Ecosystem Management, Gland, Switzerland

*equal lead authorship

[†]corresponding author: thomasmesaglio@gmail.com

Keywords

Biodiversity inventory, curation, iNaturalist, species list, species occurrence, voucher

Abstract

Dynamic yet accurate reserve-level species lists are essential for conservation and biodiversity research. Even when such lists exist, changing taxonomy, ongoing species migrations and invasions, and new discoveries of historically overlooked species mean static lists can become rapidly outdated. Biodiversity databases such as the Global Biodiversity Information Facility, and citizen science platforms such as iNaturalist, offer rapidly accessible, georeferenced data, but their accuracy is rarely tested. Here we compare species lists generated for two of the world's oldest, more famous natural reserves – Yosemite National Park in California, United States and Royal National Park in New South Wales, Australia – using both automated data extraction techniques and extensive manual curation steps. Here we show that automated list creation without manual curation offers inflated measures of species diversity at a reserve level. Lists generated from herbarium vouchers required far more curation than lists generated from iNaturalist, with both incorrect coordinates attached to vouchers and long-outdated names inflating voucher-based species lists. In comparison, iNaturalist data had relatively few errors, in part due to continual curation by a large community, including many botanical experts, and the frequent and automatic implementation of taxonomic updates. As such, iNaturalist will become an increasingly accurate source of automated biodiversity lists over time, but currently offers poor coverage of graminoid species and introduced species relative to showier, native taxa. At this point, researchers must manually curate lists extracted from herbarium vouchers or static reserve lists, and integrate these data with records from iNaturalist, to produce the most robust and taxonomically up-to-date reserve-level species lists.

1. Introduction

Change is an ever-present force in ecology, with equilibria being transient rather than permanent. Human activities, whether on a small or large scale, are continuously altering the composition of species in local ecosystems globally, affecting both disturbed and undisturbed environments. These alterations are partially systematic, often manifesting as local species extinctions in the warmest or driest parts of their ranges and expansions in the cooler areas (Slaton 2015, Rumpf et al. 2019). However, other changes are more unpredictable or have underlying causes that are challenging to identify. The concept of species turnover gained significant attention through MacArthur and Wilson's groundbreaking work on island biogeography (MacArthur and Wilson 1963). Their theory, which posits that all habitats are essentially islands – with conservation reserves being particularly island-like for many species – gained prominence due to its suggestion that isolated habitats face unique ecological challenges. Ongoing changes to the climate are forecast to drive these gains and losses, and there is a large literature forecasting these changes at different scales (Lenoir and Svenning 2015, Lawlor et al. 2024, Westoby et al. 2024).

A key element in applying this theory is species lists: these lists are relevant for global databasing and analyses (Weigelt et al. 2020). However, the quality and up-to-dateness of species lists are highly variable, and the ongoing increase in photo-vouchered specimens via citizen science platforms offers a new and valuable data stream to integrate into these lists. One of the largest and most successful such platforms is iNaturalist (www.inaturalist.org), with ~3 million observers from around the world having contributed more than 192 million observations across almost 477,000 species. iNaturalist data are extensively used in research and conservation, and there have already been many notable applications related to biodiversity monitoring, including documenting the decline of butterflies in western North America (Forister et al. 2021), detecting rare and cryptic species that are often missed by more structured surveys (Roberts et al. 2022), and documenting new country, state or regional records (e.g., Jones et al. 2019, Daniels et al. 2022, Orr et al. 2023). However, this new data stream also has potential new challenges: there are collection biases and types of errors to consider, which are partially shared with traditionally collected datasets (Binley and Bennett 2023).

Data integration necessarily involves curation: before species lists generated by citizen science initiatives can be merged with those generated from vouchers or with static park lists, one must assess both the accuracy and taxonomy of all three data sources and their constituent data providers. It is only in the past few years that citizen science data have become sufficiently comprehensive in some regions to compare their accuracy and coverage with that offered by voucher-based lists. Given research documenting species turnover across time often relies in large part on automatically generated species occurrence lists from biodiversity and citizen science portals, we assess what level of manual curation must supplement automated processes as part of developing a workflow for generating dynamic species lists.

Measuring species dynamics can be done at a variety of scales. Here we focus on the reserve scale both because of its management relevance and because of the synergy with theory from MacArthur and Wilson (1963). Plot- or transect-based efforts offer very repeatable methods but often miss rare taxa (Gaston 1999, Newmaster et al. 2005, Goslee 2006). For this reason, national park species lists are typically predominantly voucher based, with each positive observation and identification linked to a known specimen, which allows for ongoing curation as taxonomy and biogeography change. Notably, park lists are a *curated* subset of known collections.

Our goals were to recognize the dynamic nature of biodiversity; to build bridges between traditional and new data sources, foundational ecological theory, and conservation needs at the reserve scale; and, to capture the differences between species lists with and without detailed curation. We aim to create (1) a topology for syncing collections-based and citizen science-based observations in the context of a moving target and (2) a reserve-based prioritization algorithm for limited curation attention.

2. Methods

We present two case studies using vascular plants in two of the five oldest national parks in the world: Yosemite National Park in the United States and Royal National Park in Australia (Figure 1). Both are large, diverse, and well-sampled national parks that have largely been protected from development for over a century, but they have radically different habitats and native floras. Both are botanically diverse – more than 1000 native vascular plant species – and data rich, especially from citizen scientists. We found the use of two case studies informative to separate idiosyncrasies of individual national parks from possible general patterns.

2.1 Yosemite National Park

Indigenous Miwok and Paiute people lived in and around Yosemite Valley for up to 7,000 years following the latest ice age, but traditional knowledge of the flora was largely (but not entirely) lost during the colonization process, which included forced evictions by the federal government in 1851, 1906, 1929, and 1969 (Anderson 2005). A re-discovery of the flora in the European framework began with work by Albert Kellogg. The currently oldest known specimens were collected by Bolander and Brewer in the 1860s which began a long period of intense botanical effort. Parts of the current National Park were protected from development by Abraham Lincoln in 1864, with a larger area declared a National Park in 1890. Yosemite National Park currently spans ~300,000 ha.

Since 1860, more than 26,000 herbarium specimens have been collected within Yosemite National Park. Collections between 1950 and 2000 were relatively sparse due to strict collecting regulations, while a series of targeted collections from rare habitats – the Sky Island and Unusual Lakes Flora projects (see Colwell 2011) – bolstered herbarium specimens from 2005–2012. The organization of citizen science photos as observational species records began in the late 2000s but really began to take off in the park in 2015. As of May 2024 there were over 29,000 Research Grade iNaturalist observations of vascular plants in the reserve, already exceeding 160+ years of collecting.

2.2 Royal National Park

Royal National Park is located on the lands of the Dharawal people (King and Linnean Society of New South Wales 2022), who shared their name with the plant now known as the Cabbage Tree Palm (*Livistona australis*; Organ and Speechley 1997). Only a small fraction of the traditional knowledge of the local plants survived colonization (Organ and Speechley 1997). The oldest accessible collections were made by F.W.L. Leichhardt in August 1840, a few years before his disappearance on an expedition to central Australia. Royal National Park was declared as the second National Park in the world in 1879, following Yellowstone. Royal National Park currently spans ~15,000 ha.

Since Leichhardt, collections grew through time, peaking in the 1970s to early 1990s corresponding to a series of systematic flora surveys, with a subsequent decline in the number of new collections to the present day. In 2019 photographic vouchers taken by citizen science began to gain popularity. In the few years since then, the quantity of citizen science observations uploaded to iNaturalist has already almost doubled the herbarium collections from 180+ years of collecting effort: as of May 2024, there were almost 11,000 Research Grade observations in iNaturalist of vascular plants in the reserve, compared to just over 6,000 vouchered specimens.

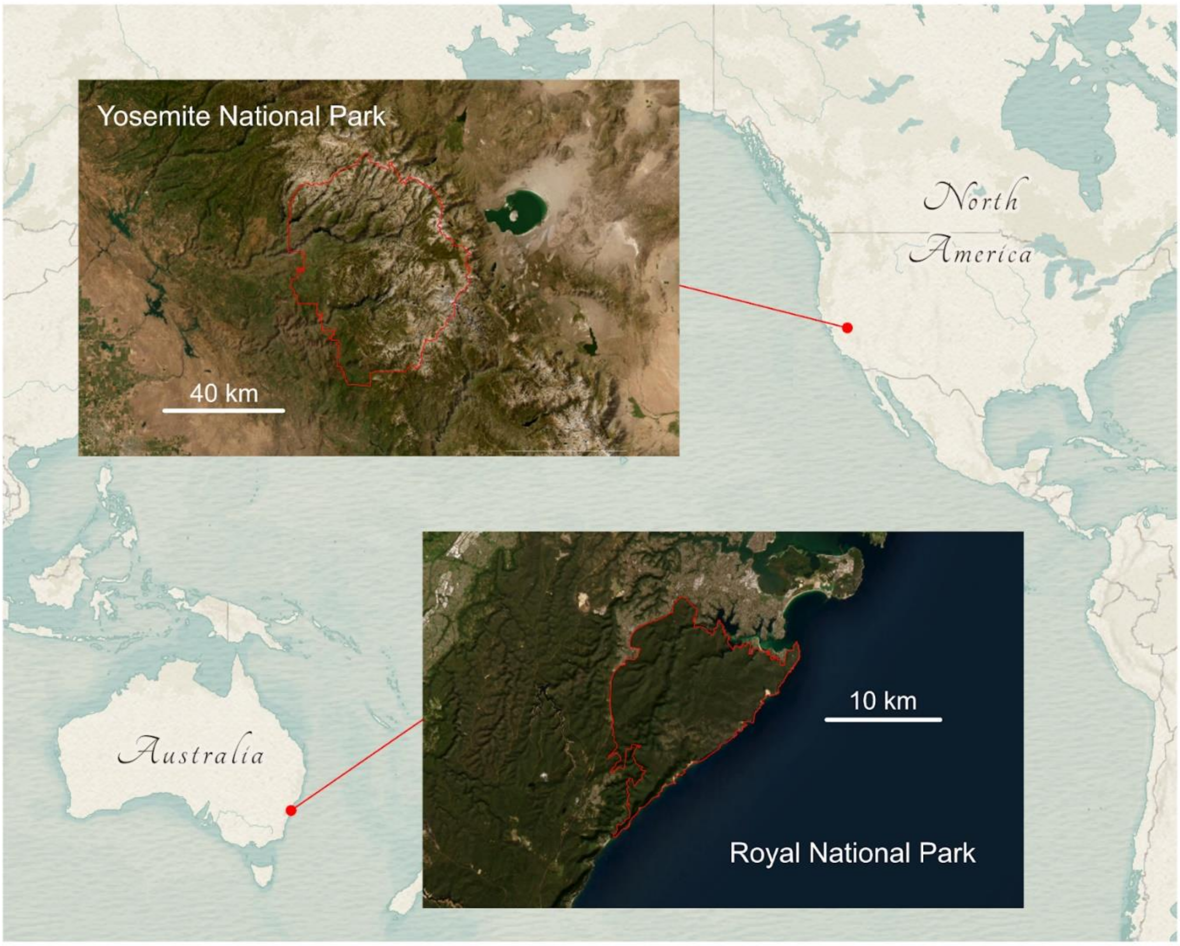

Figure 1. Locations of Yosemite National Park and Royal National Park. Map generated using Digital Atlas of Australia (https://digital.atlas.gov.au).

2.3 Data extraction and taxonomic alignment

For each reserve, we downloaded data from three different streams: vouchered specimens held at herbaria, Research Grade observations from the biodiversity citizen science platform iNaturalist, and an expert-curated park list (Figure 2).

2.3.1 Vouchers

We downloaded vascular plant occurrence data for Royal National Park (GBIF 2024a) and Yosemite National Park (GBIF 2024b) from the Global Biodiversity Information Facility (GBIF; www.gbif.org) using an approximate bounding box around each reserve. We then filtered each dataset to the exact park boundaries and only retained records associated with a preserved specimen to compile a list of vouchered taxa for each reserve. At the time of download, vouchers from some major herbaria in California were not in GBIF for technical reasons, so we also downloaded California Consortium of Herbarium 2 (CCH; https://www.cch2.org/portal/) data for Yosemite.

2.3.2 iNaturalist

Research Grade observations from iNaturalist are exported to GBIF approximately once per week. However, because this export only includes observations with a copyright licence of CC0, CC BY or CC BY-NC (Mesaglio 2024), we instead downloaded all Research Grade observations for Yosemite (7 May 2024) and Royal (27 May 2024) directly from iNaturalist to retain all observations regardless of licence.

2.3.3 Park lists

For each reserve we also obtained a curated park list. For Yosemite, this list was compiled by National Park Service botanists and is publicly available through the National Park Service's NPSpecies website (https://irma.nps.gov/NPSpecies/Search/SpeciesList/YOSE). The Royal list was also compiled by several expert botanists, and was most recently published in the second edition of the print publication *Field Guide to Royal National Park, New South Wales* (Wilson and Keith 2022). Both lists had been extensively curated by expert botanists combining their local knowledge of the park's flora with historical checklists, systematic vegetation survey datasets held by their respective National Parks Services, and a review of available vouchers. The Yosemite list was most recently updated approximately twenty years ago, whilst the Royal list was predominantly compiled approximately ten years ago and updated two years ago.

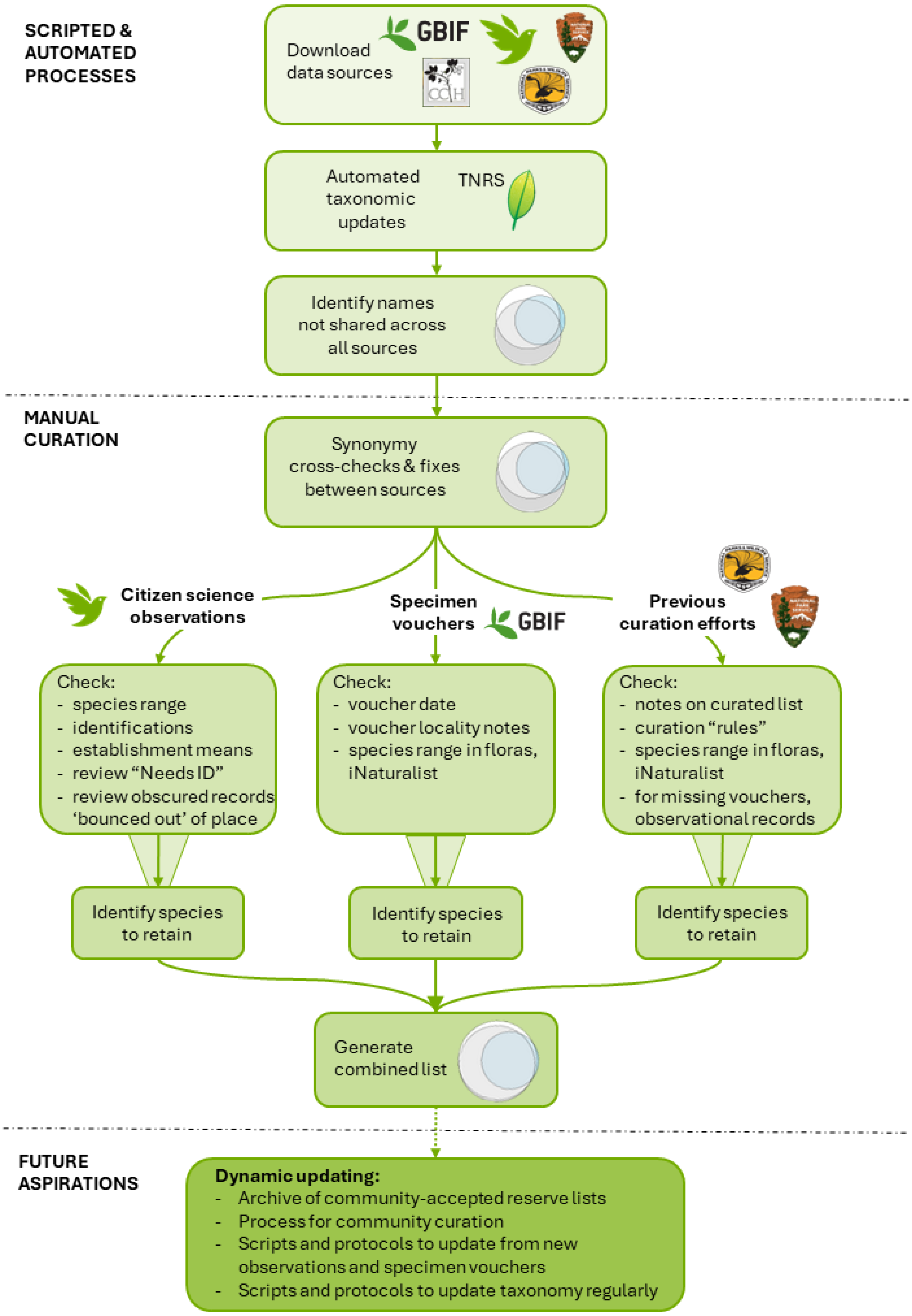

Figure 2. Generalised workflow for generating reserve-level species lists.

2.4 Taxonomy

Because multiple different taxonomies had initially been applied across our individual species lists, we decided to follow the current taxonomy used in iNaturalist. Given iNaturalist largely follows Kew's Plants of the World Online (POWO; https://powo.science.kew.org) as its taxonomic authority, which uses the World Checklist of Vascular Plants (WCVP; Govaerts et al. 2021) as its taxonomic backbone, this allowed us to apply a single consistent taxonomy to both Yosemite and Royal.

We initially investigated automated, systematic updating of names across all taxa, but found that this was often not feasible. For example, we encountered taxonomic uncertainty for some of the oldest records (i.e., from the nineteenth century), with vouchers assigned a name of a far-out-of-range species; these likely indicate taxa that have gone through many revisions (including splits) between the collection date and the current day, and unless names are constantly updated by herbaria following each split, it becomes difficult to accurately trace all nomenclatural changes. Overall, the oldest vouchers, whilst important records, require careful, manual attention at the original herbarium. Similarly, the nomenclature used in the park lists would have been current at the time they were last reviewed, but because taxonomy often changes rapidly, some names were already outdated.

Because the park list for Yosemite initially had quite outdated taxonomy, we used the Taxonomic Name Resolution Service (TNRS) to update all names to match WCVP taxonomy as a first step for only this list. However, manual alignment of names to iNaturalist taxonomy was still required for the voucher and park lists for both reserves.

We initially retained only species-level names, as many vouchers and most iNaturalist observations are not identified to infra-taxa. This step means that the species lists generated here will be somewhat shorter than the complete lists one could download from any of the three data streams. For instance, the park list for Yosemite has 109 species where two or more infra-taxa are documented as occurring within the reserve.

2.5 Manual curation

Each of the three species lists for each reserve required manual curation to verify that the listed species actually occur (or occurred) within the reserve. For each reserve, records for all species which did not occur on all three lists were manually checked. However, after noting some species erroneously attributed to the iNaturalist list for a reserve by a single misidentified Research Grade observation, we decided that there must exist more than five Research Grade observations for a species to avoid a manual check, even if it occurred on all three lists. For each iNaturalist observation we checked: 1) the coordinates, with careful attention to whether the habitat displayed in the photograph matched the purported location; 2) the identification, where we were able to make an informed decision; and, 3) who made the identifications, with additional scrutiny applied to both observer identifications made by users on a one-off visit to the reserve, and confirming identifications made by users who are not regular identifiers of the specific taxonomic group of plants for the reserve in question.

For each reserve, all species that had at least one voucher and were not also present in both the park and iNaturalist lists were manually verified. As we are not expert taxonomists and did not have access to the physical specimens themselves, with many herbarium sheets for both reserves yet to be digitized, we generally did not attempt to determine if an identification was botanically incorrect. Instead, we made judgements based on whether the metadata suggested an error, checking: 1) the number of vouchers in existence for a species; 2) the collection dates for each voucher; 3) the location of the voucher in the reserve, including habitat; 4) the verbatim locality and other collection notes for each voucher, including common place names and landmarks shared between inside and outside the reserve; and, 5) the collector name, especially in conjunction with collection date and location to resolve where specific collectors were likely to have been on particular dates. In particular, we scrutinized species for which very few vouchers were present, species for which all vouchers were only collected before 1900, and species for which vouchers only exist in small herbaria or herbaria in regions far from each reserve. For species where voucher metadata suggested a possible error, we also considered the species' range beyond the reserve in question to decide its validity. For Yosemite, for species where all vouchers were distrusted, we also referenced a regional flora compiled by Dr. Dean Taylor (Taylor 2010); this resource indicates vouchers he has examined and often indicates names that have previously been misapplied to a species, and whether a given species likely existed within the reserve boundaries. For all species, we always checked the oldest voucher and the most recent voucher, regardless of how many vouchers existed for that species and even if the species occurred on all three lists for that reserve.

If a species appeared only on the park list for a reserve, with no vouchers or Research Grade iNaturalist observations, we checked other sources of data that may have informed these listings, including scientific papers, observational data from flora surveys (especially for Royal, with these data readily accessible via the Atlas of Living Australia; https://www.ala.org.au), Research Grade iNaturalist observations from outside the reserve boundary, and historical checklists.

For species that did not appear in all three lists for each reserve, special attention was given to recent taxonomic changes. Taxon names were often looked up in POWO (which documents all synonyms, but not splits), iNaturalist (which documents recent splits and synonyms), as well as GBIF, CCH (for Yosemite), the online California flora (for Yosemite, Jepson eFlora; https://ucjeps.berkeley.edu/eflora/) and the taxonomic literature. We also checked the Australian Plant Name Index (APNI; https://biodiversity.org.au/nsl/services/search/taxonomy) for Royal, which provides extensive information on names and splits for Australian taxa.

Vouchers should represent known collections of species in a reserve across the duration of historic collecting campaigns. Although herbarium vouchers are considered the gold standard for biodiversity data (Cullen 2013, Funk et al. 2017), not all vouchers are accurately identified, are linked to the currently accepted name for a specific taxon collected in a specific location, or are free of errors introduced during electronic databasing efforts. Similarly, whilst Research Grade iNaturalist observations are generally reliably geolocated and identified (see Mesaglio 2024), there are a range of errors which can lower data quality and result in false positive records for a given reserve. We therefore compiled a list of likely error sources prior to our manual checks, and used these to guide our verification process as follows:

2.6 iNaturalist observations

1. Geographic errors

Whilst photographs taken with mobile devices and uploaded to iNaturalist are generally automatically and accurately geocoordinated, some observers manually enter coordinates or a place name to geolocate their observations, especially when taking photographs with a DSLR camera. In some cases, manual geolocation can position an occurrence in an implausible location or the wrong geographic jurisdiction. For example, Bridalveil Fall, a key

Yosemite Valley landmark, “attracts” observations made at other Bridalveil Falls worldwide when users do not double check the location they have entered.

There are also observations in iNaturalist that are uploaded retrospectively, sometimes years after the photographs were originally taken, and observers may not accurately remember where they were at the time. Finally, when manually typing location names in iNaturalist during the upload process, Google Maps retrieves an explicit point location (often with a large coordinate accuracy value) connected to a geographic place name such as “Yosemite National Park”. If observers manually enter such a place name, their observation is automatically linked to a location that might be quite distant to their actual observation location. For this project, as the point locations for “Yosemite National Park” and “Royal National Park” both fall within their respective boundaries, this does not introduce errors at the reserve list-level. However, the point location for Ansel Adams Wilderness, a large Wilderness Area just south of Yosemite, actually lies on the Yosemite side of a remote peak just within the National Park. Therefore, species that only occur south of Yosemite could ostensibly appear on the species list despite being observed outside of the reserve.

2. Identification errors

For an iNaturalist observation to reach Research Grade at a species level or finer, the observation must be identified by at least two users, with more than two thirds of these identifiers in agreement. This provides an important check, especially if the identifiers are truly independent. However, there may be instances where the second identification for an observer’s record is made by a friend or colleague who “accepts” their identification without question. This can occur for observations made by less knowledgeable and expert observers alike. Moreover, experts are not infallible and can, on occasion, make erroneous identifications, and when an expert has offered an identification, observers often uncritically agree with their identification without careful scrutiny (see e.g., Burgman 2016).

3. Databasing discrepancies

Whilst all Research Grade iNaturalist observations should be eligible to be exported to GBIF, only a subset of these records are actually piped to GBIF; observers can set copyright licence permissions such that their observations will not be exported. In particular, observers may set their observation licence to All Rights Reserved, perhaps mistakenly thinking they are setting this licence for their photographs (a separate setting) or not realizing the implications of this choice. In some cases, 100% of the observations for a given species within a reserve may fall under a non-GBIF compatible licence, especially for species with few records that have all been uploaded by the same observer, and thus be entirely excluded from GBIF at that location. Although these discrepancies did not influence our workflow given we downloaded observations directly from iNaturalist, they are an important consideration for data users accessing iNaturalist records directly through GBIF.

2.7 Vouchers

1. Geographic errors

For recently collected vouchers (since c. 2000), precise coordinates are usually explicitly provided by the collector in addition to a verbatim locality. For many older vouchers, however, sometimes nebulous, imprecise, or even deprecated place names must be converted into a best-guess set of coordinates during digitization. This is accomplished both through the manual addition of place names and automated processes such as the GEOLocate Batch Processing Tool (https://www.geo-locate.org/web/). In addition to human misinterpretation of place names, these processes are subject to transcription errors (manual coordinate entry) and database errors (automated processes).

2. Identification errors

A presumption among researchers is that the species names attached to herbarium vouchers are accurate, and whilst this is usually the case, identification errors are also present within herbarium collections, sometimes at high rates for certain taxa (Utjés et al. 2022, Cardoso et al. 2024, Coca-de-la-Iglesia et al. 2024). Misidentification of a voucher is more likely for small herbaria and herbaria far from the collection source, where the determiner may lack expertise with the regional flora of the voucher’s original location.

3. Databasing discrepancies

Researchers may assume that GBIF offers a comprehensive portal for biodiversity occurrence data and that, for any analysis for which they need species occurrence data, they can download GBIF records and be confident that collection records from the world’s major herbaria will be retrieved. This is not necessarily true as we discovered when downloading data, and it was only due to our in-depth knowledge of the herbaria that *should* have been in GBIF that we learned of a technical error that temporarily excluded all records from many of California’s largest herbaria. This is an error we would not have caught had we been downloading data from a region with which we were unfamiliar. Many institutions are also affected by specimen processing, accessioning, digitization, and data transfer delays as a consequence of both lack of resourcing and funding, and technical errors.

4. Taxonomic change/uncertainty

The earliest botanical collections for a given location are particularly susceptible to taxonomic errors, as nascent taxon concepts for newly “discovered” areas may still have been fuzzy, with names for similar-looking taxa from other locations applied until the taxon concepts were subsequently resolved. If herbaria did not frequently update the name attached to a voucher, the link between the name written on the voucher and the name that would be applied today to this taxon concept could be lost and the misapplied name retained (Goodwin et al. 2015). Taxon splits, especially where the original name is retained, are difficult to trace unless all splits and the dates of these splits are explicitly databased. In Australia, the Australian Plant Census (APC; https://biodiversity.org.au/nsl/services/search/taxonomy) documents all known splits, whereas California lacks such a resource. Making judgements on which names have been misapplied to a given voucher can be difficult for experts to make, and often impossible for a casual data user.

3. Results and Discussion

Building a data framework for dynamic, reserve-wide species lists is a difficult botanical and data science challenge, but one with clear potential for research and conservation impacts. Our case study of two well-sampled reserves, Yosemite National Park (Figure 3) and Royal National Park (Figure 4), offers a clear message for assessing species diversity using biodiversity portals: diverse and independent data sources are complementary. Expert-curated park lists are of high quality and an invaluable resource, but can go out of date if not regularly updated, whilst iNaturalist offers a recent, accurate, but incomplete picture of biodiversity at a reserve level. Broadly, extensive manual curation, including accounting for identification and geographic data errors, updating names, and aligning differing taxonomies is required to generate a comprehensive, reliable biodiversity list when integrating multiple different data streams (Supplementary Table 1, Supplementary Table 2). Targeted curation at different levels of the data flow system can make this process much more seamless into the future.

3.1 Taxonomic alignment

For Yosemite, after the use of TNRS, 98 of the names originally on the park list were synonyms and 6 names were misapplied, whilst 10 names originally on the voucher list were synonyms and 25 names were misapplied, with decisions on these 25 misapplied names based on notes in either Jepson (https://ucjeps.berkeley.edu/eflora/) or Taylor (2010); 5 of these were also on the park list. There were also 46 names from the voucher list rejected due to unspecified taxonomic mismatches, untraced synonyms, or other changes that we were not able to trace. Of these, only 3 were on the park list, insinuating park botanists and managers scoured voucher collections in the past, making near-identical decisions about species that were very unlikely to occur within the park. Most of these 46 were represented by very few vouchers that had only been collected before 1900, and were lodged at small herbaria.

For Royal, 105 of the names originally on the park list were synonyms and 7 names were misapplied, whilst 78 names originally on the voucher list were synonyms and 5 names were misapplied. Two species were also removed from the Royal voucher list due to unusual name matching errors; for example, a known issue in the ALA related to specimens from the National Herbarium of New South Wales resulted in vouchers originally identified as various *Eucalyptus* intergrades mapping to the name *Eucalyptus pauciflora* subsp. *debeuzevillei*, a taxon only found hundreds of kilometres to the southwest.

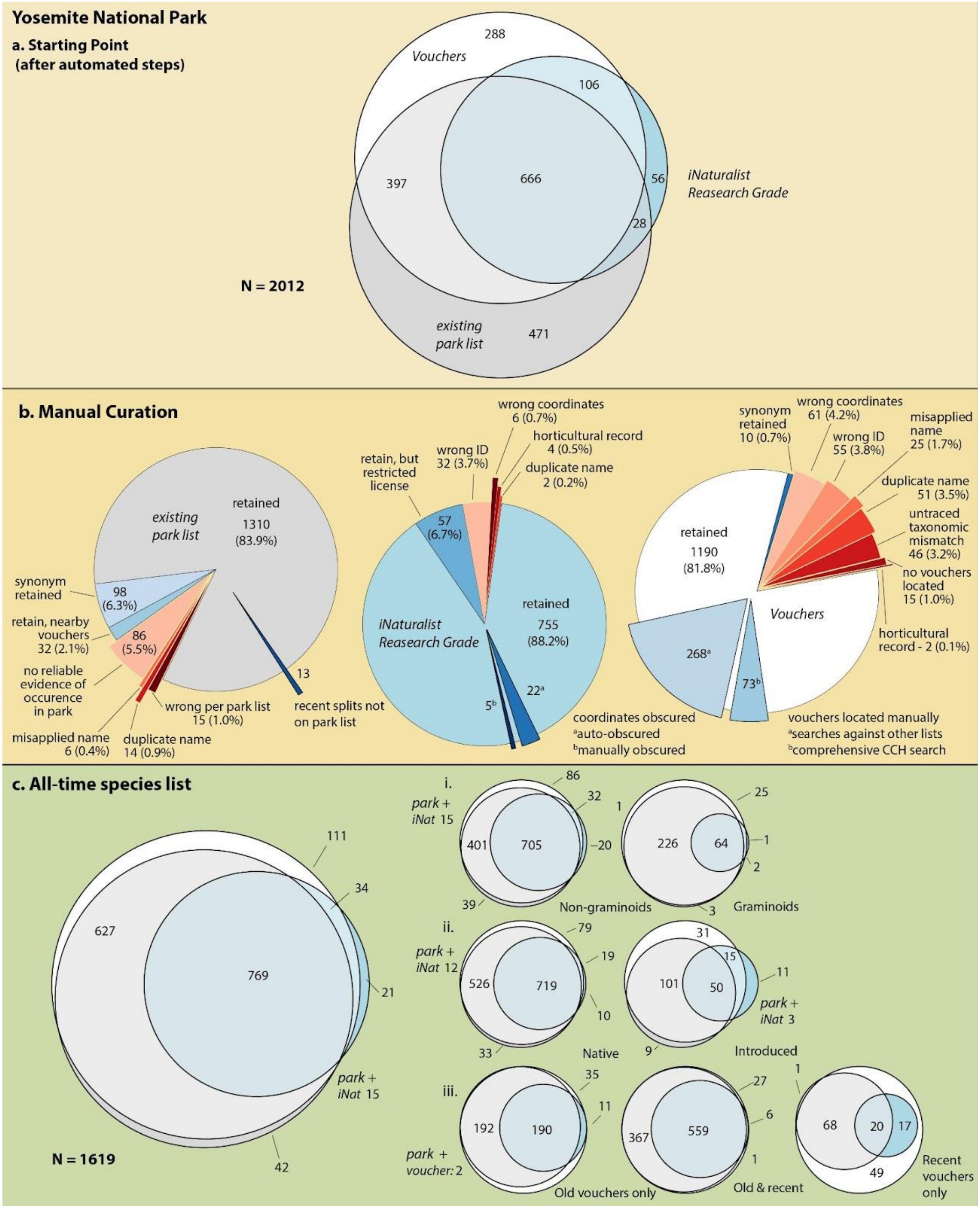

Figure 3. Species list curation for Yosemite National Park. We used 1990 as the cut-off between old and recent vouchers. ‘Graminoid’ taxa were defined as those in the families Cyperaceae, Juncaceae, Juncaginaceae, Poaceae, Potamogetonaceae, and Typhaceae. In panel b, red slices indicate species that were removed, and blue slices indicate species that were added.

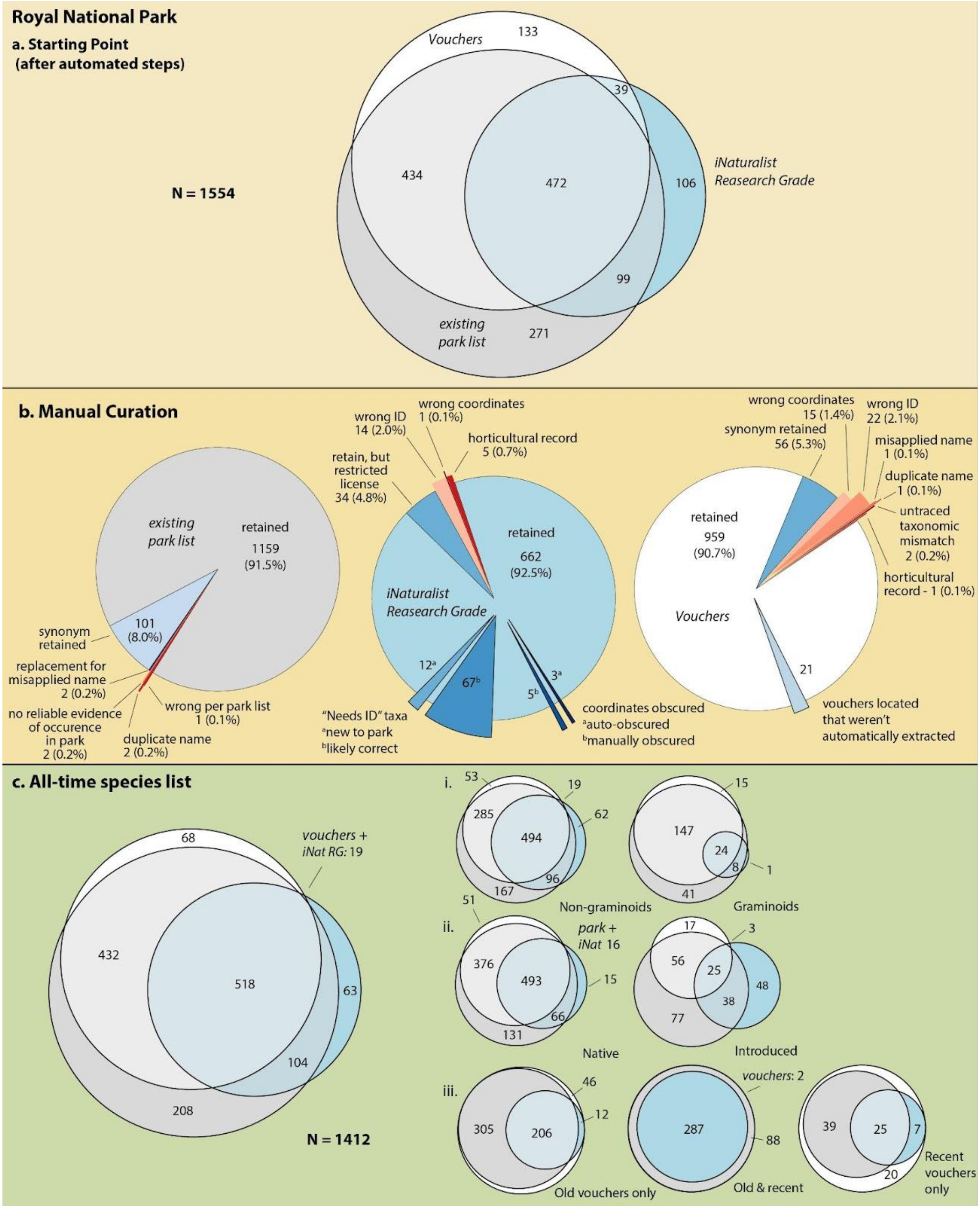

Figure 4. Species list curation for Royal National Park. We used 1990 as the cut-off between old and recent vouchers. 'Graminoid' taxa were defined as those in the families Centrolepidaceae, Cyperaceae, Juncaceae, Juncaginaceae, Poaceae, Posidoniaceae, Potamogetonaceae, Restionaceae and Zosteraceae. In panel b, red slices indicate species that were removed, and blue slices indicate species that were added.

3.2 Voucher errors

For Yosemite, relatively few scientific names (55) are explicitly designated as being misidentified, as that would require looking at individual herbarium vouchers and making taxonomic decisions we do not feel qualified to make; instead, this designation is largely reserved for instances where a species was represented on the list by a single voucher assigned the name of a taxon only known to occur far outside Yosemite's boundary; such vouchers were almost always in small collections. The 61 coordinate errors include a few suspected mislabelling errors on herbarium sheets, but predominantly represent errors made during the manual or automated translation of a place name into coordinates. For instance, a single recurring mistake generated by the GEOLocate Batch Processing Tool added 27 desert species to the list; the coordinates assigned to the locality "Big Maria Mountains" in California's Riverside County are erroneously positioned in southwestern Yosemite. Other mistakes added individual species and were quickly detected by a place name description that did not match the coordinates. There were three ambiguous place names in particular that could reference a location inside or outside the park; Snow Creek (there are two in Mariposa County); Eagle Meadows (one lies in Tuolumne County north of the reserve); and Red Peak (one lies in Tuolumne County north of the reserve).

An additional problem is that vouchers labelled, for instance, "Yosemite region", may include locations outside the park itself; moreover, the Yosemite park boundaries were partially redrawn in 1906, extending further west, but ceding a region southeast of the current park boundary (Ritter Range) to mining interests (Davis and Wenk 2020). Similarly, large changes were made to Royal's boundaries c. 1900, with some smaller additions in the 1980s. Geographic ambiguity or inaccuracy on the voucher label can also be a problem with the oldest vouchers, as colloquial place names have often changed. An unexpected source of voucher error was 341 species from Yosemite for which vouchers had not been downloaded automatically, and were only located with manual searches. 268 of these were initially found through manual searches due to their inclusion on either the park list or in iNaturalist. An additional 73 species were located when all vouchers in Tuolumne, Mariposa, and Madera counties were downloaded and vouchers attributed as having been collected in Yosemite based on the locality, independent of coordinates. For 55 of the 341 species, all vouchers lack coordinates, and a few have coordinates erroneously placed outside the park. Most of the other species were "missed" because CCH downloads do not update the taxonomy and therefore the download included names that were not recognized as "accepted" by our initial scripts; revised scripts and synonym-matching lookup tables were later developed to capture them.

Twenty-two species were removed from the Royal voucher list based on our assessment of them as misidentified. Like Yosemite, almost all of these were species with known ranges far outside of Royal, some of them by thousands of kilometres (e.g., *Passiflora kuranda*, which is found only around the Cairns region in tropical Queensland). One species from Royal, *Pinus nigra*, was represented by a single voucher of a cultivated specimen, and was also removed. The 15 species removed for coordinate errors featured a variety of different issues, including a voucher collected at Flat Rock in Vanuatu that had been geolocated at Flat Rock Creek in Royal; clear mismatches between the verbatim locality and the coordinates (e.g., a voucher collected at "Heathcote on western side of Princes Hwy" that had been placed into Royal on the eastern side of the highway); and, coordinate transcription errors that were only detected by inspecting the digitized herbarium sheet.

Vouchers for 21 species that were unambiguously collected in Royal based on the locality notes had to be manually retrieved, including 4 species which were not on the Research Grade iNaturalist or park lists. Most of these had been geolocated into the ocean to the east of Royal. One voucher indicated it was collected “c. 1/4 way from Lilyvale to B.P. [Burning Palms]”, a location unambiguously inside Royal, but the coordinates had been placed to the west of Lilyvale (instead of eastwards) just outside the park. There were also two species — *Drosera capensis* and *Utricularia subulata*, both carnivorous plants non-native to Royal — that have clearly been collected from inside Royal, with vouchers deposited at the National Herbarium of New South Wales explicitly cited in scientific papers documenting their discovery (Jobson and Conn 2012a, Jobson and Conn 2012b), but for which the vouchers have not been digitized, despite being collected more than ten years ago.

3.3 iNaturalist observation errors

Relatively few errors were detected in Research Grade iNaturalist data for either Yosemite or Royal. The data extraction for Yosemite included 6 species with incorrect coordinates, in each instance because the observer had manually added coordinates when posting their photos. For example, a photo from Bridal Veil Falls in Oregon had been geolocated at Yosemite’s Bridalveil Fall, and an observer visiting multiple locations in the park had erroneously added the coordinates of a high elevation lake to a photograph taken outside the park along the Merced River. In addition, there were 32 species where all Research Grade observations were misidentified; to avoid these observations being incorporated into iNaturalist’s evolving Computer Vision model, most were corrected as soon as they were detected. Promoting continued, and increased, expert identifier engagement with iNaturalist will be crucial for keeping the number of misidentifications at such low levels (Callaghan et al. 2022). There were a number of species for which iNaturalist observations existed that were excluded from GBIF, with 70 species for which all Research Grade observations have a non-GBIF compatible copyright licence (55 of which also had vouchers). Of these 70, 57 were retained as valid observations.

Only a single species was removed from the Research Grade iNaturalist list for Royal due to incorrect coordinates: an observation of *Grevillea acanthifolia* uploaded more than six years after it was observed was mistakenly geolocated into the park, with that user’s other observations from the same day and same time correctly geolocated in the Blue Mountains northwest of Royal. Just 14 species were represented only by misidentified Research Grade observations, with almost all of these cases being a single misidentified record of a similar species. There were also 5 species for which the only observation from Royal was of a planted specimen that had not been marked as such. There were 34 species for which all Research Grade observations are under a non-GBIF compatible copyright licence (22 of which also had valid vouchers).

Across both reserves, some species were represented exclusively by obscured Research Grade observations, with important implications for the discoverability of these records. In Yosemite, there were 27 species for which all Research Grade observations were obscured; 22 of these are rare or designated vulnerable species that have their locations automatically obscured by iNaturalist, and 5 species were singleton records obscured manually by the observer. There were far fewer such species in Royal, just 3 and 2 respectively. In iNaturalist, Royal National Park is a ‘community curated’ place, one of two main place types. The second type is ‘standard’ places; any iNaturalist observations that are originally geolocated inside their boundaries will always be indexed within that place, and appear in searches, regardless of whether they are ‘bounced out’ due to obscuration. Conversely, obscured observations can be bounced out of community curated places (see Mesaglio 2024) and thus not appear in searches for that taxon within that place. For this reason, observations for the 5 aforementioned Royal species were not automatically retrieved; instead, we had to manually locate them and confirm with the observers that the true coordinates did fall within Royal. More importantly, however, if any of these species were searched for directly in GBIF they would not show as occurring in Royal National Park. Although Yosemite, like all US National Parks, is designated as a standard place on iNaturalist, this is a nominal status only without holding typical standard place functionality. However, because Yosemite is so large, almost all of the species represented only by obscured observations within the park still appear in iNaturalist searches for the reserve, as the randomised coordinates and obscuration bounding box for each record fail to be bounced out. Nonetheless, obscured observations made near the edges of Yosemite can still be bounced out and lost, and require manual retrieval as for Royal.

3.4 Park list errors

The Yosemite park list has been curated repeatedly by park botanists, omitting nearly all of the same species that have vouchers but that we considered to not actually occur in the reserve. However, the park list curators seem to have made an explicit decision to also include “potentially present” species known close to the park boundaries. These are predominantly species that occur at lower, warmer elevations west of the reserve, but also include species whose ranges are outside the reserve boundaries to the north, east, and south. Per our definition of “occurring in Yosemite”, species from the park list where the closest voucher or iNaturalist observation was more than 5 km from the park boundary were all removed, while a subset of species occurring within this radius were retained and flagged as potentially occurring in the park. There were also 15 species on the parks list with an explicit note that they were added in error; as these species were indeed also absent from other sources, they were removed.

The Royal park list was also well-curated, with almost 100% of the originally listed species retained after our manual curation and accounting for the alignment of names to iNaturalist taxonomy. Similar to Yosemite, the Royal park list included a small number of species which have not actually been recorded in Royal; they have thus far only been observed in Heathcote National Park (west of Royal) and Garawarra State Conservation Area (south), but are predicted to be potentially present in Royal given the proximity and habitat similarity between the reserves. There were only three species from the park list for which we could find no reasonable evidence of their presence in Royal, and thus removed them from the park list. Two of these – *Allocasuarina diminuta* and *Veronica notabilis* – were aforementioned “potentially present” taxa, whilst the third – *Prostanthera marifolia* – was already explicitly noted on the park list as needing removal.

3.5 Curated species lists

3.5.1 Yosemite

The final curated species list indicates 1619 vascular plant species are known to occur in Yosemite (Figure 3, Supplementary Table 3). This includes 32 species on the park list that are likely to occur in the reserve, based on anecdotal observations and the proximity of observations or voucher collections to the park boundary, and 7 species with “needs ID” status in iNaturalist that we believe are correctly identified (the latter excluded from calculations in Figure 3). 51.9% of species known to occur within Yosemite have at least one Research Grade iNaturalist observation, including 21 species for which there are no vouchers and which are not included in the park list (Figure 5). Of the ten native species for which the first Yosemite occurrence data comes from iNaturalist, five are lower elevation or Coast Range species that have moved up in elevation, having been previously recorded only in the lower elevation Sierra Nevada foothills west of Yosemite. Just a few new Yosemite occurrences, such as *Hemitomes congestum* and *Erythranthe rubella*, represent “missed” species that have probably long occurred in the park but were missed by all specimen collectors. The other 11 species are new weed records for the park. Other species with iNaturalist observations that lacked vouchers had been included in the park list, such as *Leptosiphon nuttallii* and *Ivesia shockleyi*; it is unknown if these were on the park list due to expert knowledge of their existence in the park or because they had previously been documented near the park boundary. In comparison to vouchered species, iNaturalist observations are biased to ‘showy’ species (versus graminoid) and natives (versus introduced species). Species for which there are only recent voucher collections (post-1990) have the lowest proportion of iNaturalist observations (23.9%) than do species with only old vouchers (46.7%) or both old and new vouchers (58.9%).

3.5.2 Royal

The final curated species list indicates 1424 vascular plant species are known to occur in Royal (Figure 4, Supplementary Table 4), including 12 species represented only by needs ID iNaturalist observations that we believe are correctly identified (the latter excluded from most calculations in Figure 4). 49.4% of taxa known to occur within Royal have at least one Research Grade iNaturalist observation in the park, including 63 species for which there are no vouchers and which are not included in the park list (Figure 5). Of the 15 native species for which the first Royal occurrence data comes from iNaturalist, five are orchids, including the critically endangered *Thelymitra atronitida* and the highly localised and rare *Thelymitra improcera*. The other 48 taxa are new weed records for the park. As for Yosemite, iNaturalist observations in Royal are biased to ‘showy’ species (versus graminoid) and natives (versus introduced species) in comparison to vouchered taxa. Species for which there are only recent vouchers (post-1990) have the lowest proportion of iNaturalist observations (35.2%) compared to species with only old vouchers (38.3%) or both old and new vouchers (76.1%).

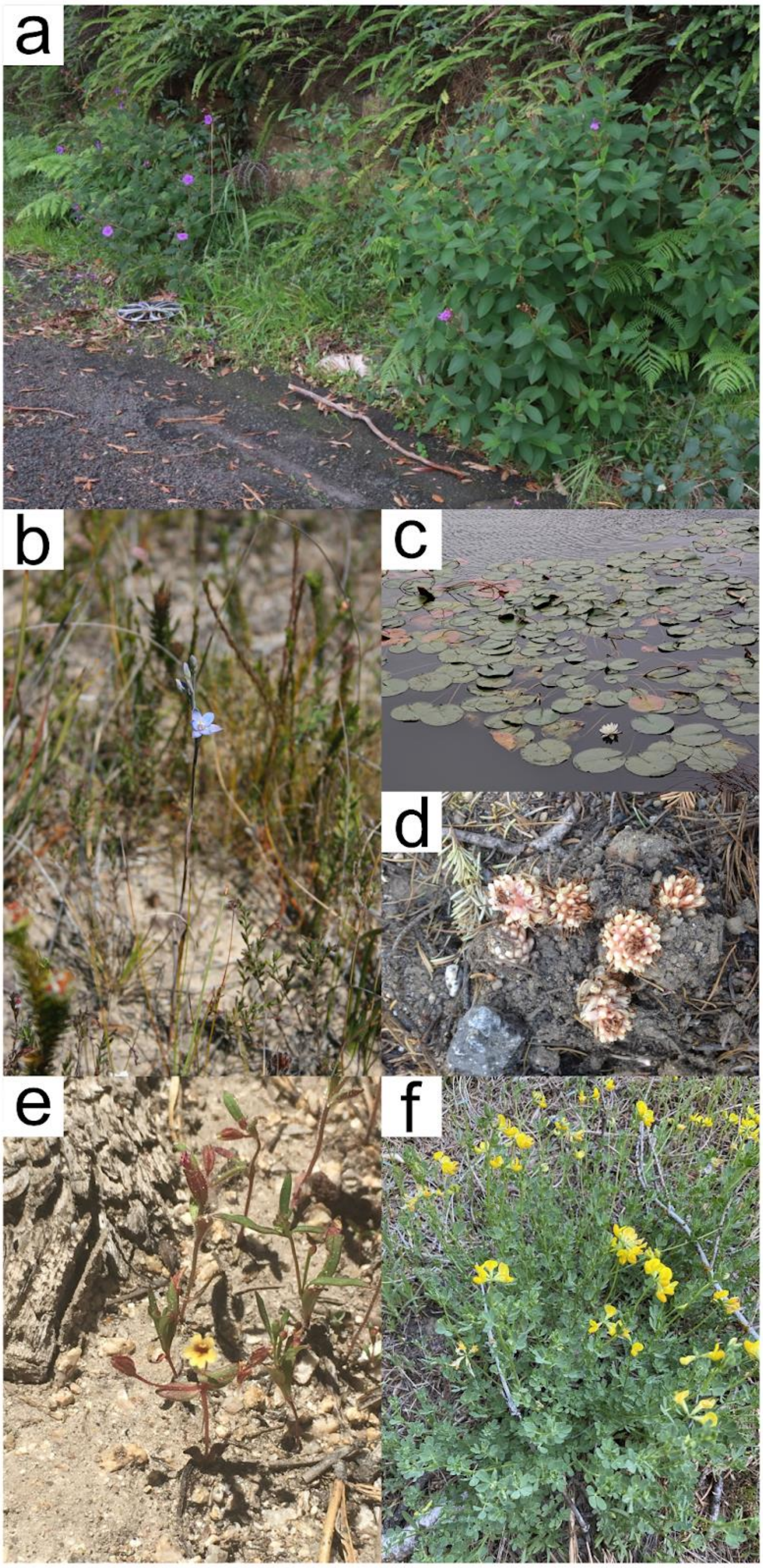

Figure 5. Plant species newly recorded for Royal National Park (a-c) and Yosemite National Park (d-f) from Research Grade iNaturalist observations. a) *Pleroma urvilleanum* (Melastomataceae), introduced, Thomas Mesaglio, CC BY; b) *Thelymitra improcera* (Orchidaceae), native, Robert Humphries, CC BY-NC; c) *Nymphaea alba* (Nymphaeaceae),

introduced, Russell Barrett, CC BY-NC; d) *Hemitomes congestum* (Ericaceae), native, Adam J. Searcy, CC BY; e) *Erythranthe rubella* (Phrymaceae), native, Brett Bell, CC BY-NC; f) *Lotus corniculatus* (Fabaceae), introduced, yhirama, CC BY-NC.

3.6 Lessons learned

Our two case studies demonstrate eight key lessons for the compilation of plant species lists at a reserve level:

1. Expert-curated park lists are an invaluable resource

It is clear that high quality species lists curated by botanical experts provide the most robust documentation of a reserve's flora. Experts are able to compile records from different data streams, including data not readily publicly available, and make informed judgements on the reliability of each record, and whether a species does or does not occur at a given location. This is especially the case for experts that have extensively surveyed and collected in the reserve in focus, as was the case for both Yosemite and Royal. Because these park lists generally utilize data from a range of sources, including vouchers and floristic surveys, they will almost always represent the most comprehensive account of a reserve's flora compared to voucher-only or iNaturalist-only lists. After our own manual curation, the Yosemite and Royal park lists contained 89.8% and 88.6% of all species recorded from their respective reserves, providing the most accurate picture of each reserve's total known flora.

If such a resource does not exist for a reserve, does not have up-to-date taxonomy, or is not publicly available, each new research team generating a reserve list from vouchers must invest effort to manually curate the list, and inevitably will make some different decisions to previous compilers, leading to disparate park lists and an unnecessary waste of precious expert curation time and resources.

2. Expert-curated park lists require ongoing updates

Given the intensive curatorial investment needed to maintain park lists, these resources often require taxonomic updates, first programmatically (e.g., with APCalign; Wenk et al. 2024) and second manually to account for splits. Given the dynamic nature of plant taxonomy, a delay between updates of even a few years can result in the build-up of numerous taxonomic changes that need to be addressed.

There are three broad categories of taxonomic change. The first is a large collection of globally unambiguous changes, for example, a species that has moved to a new genus. These "look-up" taxonomic updates can be implemented programmatically with tools like TNRS and APCalign, and are mostly already being handled well by data organisations like ALA and GBIF. The second category includes global splits that are ambiguous without geography but become unambiguous "look-up" taxonomic updates within the specific geographic context of a particular reserve. For these, it makes sense that botanists or managers for each reserve maintain their own taxonomic update lookup external to the global one so that the curation process is faster each time. The third and most difficult set of changes are ambiguous ones which need to be re-examined in light of new species concepts, and generally require direct examination of vouchers to assess whether a change needs to be made. In our case, given the examination of vouchers was largely beyond the scope of this project (except where vouchers had already been digitized), we were mostly unable to apply this third category to the Yosemite and Royal lists.

In addition to taxonomic updates, curation attention at the reserve level needs to also focus on physical and photographic vouchers that affect park lists. For example, park lists must be updated to include newly discovered species. In Yosemite, just 57.4% of species represented only by a recent voucher were included in the park list, compared to 96.6% of species with both old and new vouchers and 89.3% of species with only old vouchers. Similarly, for Royal, just 70.3% of species represented only by a recent voucher were included in the park list, compared to 99.5% of species with both old and new vouchers and 89.8% of species with only old vouchers.

3. Manual curation of vouchers is necessary

Vouchers are an integral component of species lists, constituting the fundamental verifiable evidence for species occurrences. However, like any other form of biodiversity data, vouchers can be misidentified or be associated with metadata errors such as incorrect coordinates. Only through manual curation did certain repeat voucher errors emerge that had not been expected to incorrectly contribute so many taxa to the list for each reserve. For example, the error with the GEOLocate Batch Processing Tool for Yosemite was only detected by individually looking up taxa for which there were vouchers in GBIF, but which were not on the park list and not in iNaturalist; only after repeatedly noting erroneous coordinates attached to Riverside County specimens was it apparent this was an error with the tool, not a one-off input error. There were also numerous species with vouchers that had been geolocated into the ocean to the east of Royal which were similarly only retrieved through manual searching. Across Yosemite and Royal, older vouchers were significantly more likely to have geospatial errors, most of which were introduced during the digitization process. These records should be scrutinized more closely during curation.

While data from both herbarium vouchers and citizen scientists are broadly accurate, rare occurrences of both should be carefully analyzed before making inferences about species occurrences, invasions or extinctions. Putative high value observations can be identified and subjected to additional scrutiny. These are taxa with a single (or small number) of herbarium vouchers or citizen scientist observations that suggest a range expansion. Such observations, whether recent or older, should be individually investigated for taxonomic (including naming and identification) and geographic accuracy before their putative occurrence is accepted.

4. The curation process should be streamlined with protocols and tools

Expert curation time is both valuable and limited, and thus should be focused on the most important vouchers (both physical and photographic) and taxonomic issues related to park lists and reserve management. Concurrently, many (not all) important errors can be corrected by non-experts. The democratisation of major parts of the curation process optimises the number of errors that can be corrected in a timely manner, and affords experts more time to focus on issues that actually require their botanical expertise.

iNaturalist provides a strong model system to demonstrate the efficiency of this division of curatory labour at both a taxon and an individual record scale. “Look-up” taxonomic updates can be implemented by site curators so that changes such as geographic-based splits or transferring a species to a new genus are automatically and instantaneously applied to all relevant observations, obviating the need for any manual updates to names for potentially hundreds or thousands of records. This allows expert identifiers to spend their time more productively on identifying records of taxa in their area of expertise, and reviewing records affected by the third category of taxonomic changes, such as splits resulting in numerous new species occurring in sympatry for which records cannot be automatically and programmatically reidentified based on geography. Similarly, data errors such as incorrect coordinates can easily be reviewed and annotated in the Data Quality Assessment section of each iNaturalist observation (see Mesaglio 2024) by any user at anytime from anywhere; for example, it does not require an expert to determine that a terrestrial plant record geolocated into the ocean is in error. This community-driven approach to data curation is clearly the most efficient system for allocating curatorial resources.

5. iNaturalist is an important complementary data stream

With almost 200 million observations submitted globally, iNaturalist is now one of the largest sources of biodiversity data in the world. Observations continue to accumulate at an exponential rate, with fewer than ten years' worth of records on the platform already exceeding close to two hundred years' worth of collecting efforts for Yosemite and Royal. Even for such heavily botanised locations with thousands of vouchers collected by hundreds of botanists over time, the Research Grade iNaturalist dataset contributed 21 new species for Yosemite and 63 new species for Royal which had previously never been recorded for these reserves. Lists derived from iNaturalist observations are largely accurate, and almost devoid of 'false positives' if only taxa with two or more Research Grade observations are retained. And indeed, for both Yosemite and Royal, a greater proportion of taxa with Research Grade iNaturalist observations were considered to actually be present in their respective reserves than taxa with vouchers (Figure 3, 4), with the latter dataset containing more errors resulting in the removal of species from the voucher lists.

However, we emphasize that iNaturalist cannot replace vouchers or expert-curated park lists. Even for two observation-dense locations, the Research Grade iNaturalist dataset still only includes observations for approximately 50% of the total species diversity in both reserves, with this percentage plummeting to just 15.0% and 13.9% for graminoids in Yosemite and Royal respectively. The value of iNaturalist is therefore as a complementary resource to the other data streams, with especially high value for detecting new invasive species (e.g., Werenkraut et al. 2020, Dimson et al. 2023, Gervazoni et al. 2023).

6. All three data streams contribute 'unique' species

Although the park list for each reserve contained ~90% of the total flora, significant contributions were still made by the voucher and Research Grade iNaturalist lists: 166 species from Yosemite were represented only from the voucher and/or iNaturalist list, being absent from the park list, whilst 150 species were added for Royal. It is thus clear that the combination of all three data streams provides the most complete picture of a reserve's plant diversity.

7. Data flow issues impede efficient data synthesis

Biodiversity occurrence data needs to flow from the original observer or collector through one intermediary – either a herbarium or iNaturalist – to the global aggregator, GBIF.

Unfortunately, numerous data flow issues can impede the efficient merging of data. In many cases there are other curation steps involved, most notably data entry and the geo-referencing of older specimens. At the scale of this project (tens of thousands of records), errors were detected that arose at all points in the data flow process. In our case study, especially for Yosemite data sources, disparate and quite outdated taxonomies can require significant effort to be invested to merge multiple scientific names that reference the same taxon concept across data streams. For some reserves, many vouchers lack coordinates, necessitating alternative searches by jurisdiction or locality names to locate vouchers; in Yosemite, for example, 43% of vouchers lacked coordinates and there were 55 species for which none of their vouchers had coordinates. Whilst in our case study more data flow errors became apparent working with voucher data, data were also "lost" from iNaturalist downloads, and in particular through the iNaturalist to GBIF pipeline. Observations of sensitive species with auto-obscured coordinates and observations with manually obscured coordinates may not be positioned within a reserve polygon, and observations with restrictive licensing (anything other than CC0, CC BY or CC BY-NC) are excluded from GBIF. All of these elements must be taken into consideration when synthesizing data across these resources to ensure the most robust final species list can be compiled.

The digitization of herbarium vouchers, the merging of multiple collections into biodiversity portals, the development of automated taxonomy-updating tools, and the emergence of iNaturalist herald an era of easy data access and manipulation. The Darwin Core Standard (Wieczorek et al. 2012) has been a great advance and is near-universally used for occurrence data, allowing columns to be readily merged across resources, but further improvements in the curation process in light of the data flow system would facilitate the assembly of reserve-level lists.

8. Reserves are idiosyncratic

Although there are many generalisable findings from our study that are transferable to any reserve in the world and to the process of building species lists broadly, and there are common challenges inherent to all lists, the list curation and assemblage process differs from place to place due to idiosyncratic differences between reserves. Each location has its own eccentricities, and these will influence the degree and type of curation needed. The mobilization of local experts intimately familiar with the flora of a given location, especially for regions where expert engagement may currently be variable or low, is therefore crucial for maximizing the reliability of reserve species lists.

4. Conclusions

The existence of pre-existing reserve-level species lists is required to document new species arrivals and local species extinctions. Botanists have long pursued this task, amalgamating their expert knowledge of local floras with information from vouchered specimens. Much like our manual curation steps, these efforts have required creating a list of all species for which vouchers are detected, then individually determining that some represent misidentifications or an outdated or misapplied scientific name. Huge amounts of time have been invested in these static lists, but due to constantly changing taxonomy (renaming and splits), new discoveries, new invasions, and changing range boundaries due to climate change, these lists rapidly become outdated, reducing their utility. Ideally, these lists are configured to be dynamic, with periodic expert curation combined with semi-automatic taxonomic updates. However, these lists are only useful if researchers and citizen scientists are aware of their existence and trust them: they must be made publicly available and the methods and decisions used to compile these lists must be documented and shared as list metadata.

Acknowledgements

We thank Karen Wilson, Peter Wilson, Tim Hammer, Frank Hemmings, Lachlan Copeland, and John Arney for their insights into Royal records; Alison Colwell and Peggy Moore for their insights about Yosemite voucher collecting, voucher curation, and the assembly of the Yosemite park list; and, Jane Widness for assistance with Royal shapefiles.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Anderson, K., 2005. Tending the wild: Native American knowledge and the management of California's natural resources. University of California Press.

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. and Lowry, S., 2013. The taxonomic name resolution service: an online tool for automated standardization of plant names. BMC Bioinformatics, 14, pp.1-15.

Burgman, M.A., 2016. Trusting judgements: how to get the best out of experts. Cambridge University Press.

Cardoso, P.H., Schindler, B., Figueira, M., O'Leary, N. and Santos-Silva, F., 2024. Taxonomic, nomenclatural, and distributional updates in *Casselia* (Verbenaceae): Are the species rare or are their specimens just in the wrong herbarium cabinet?. Acta Botanica Brasilica, 38, p.e20230151.

Coca-de-la-Iglesia, M., Gallego-Narbón, A., Alonso, A. and Valcárcel, V., 2024. High rate of species misidentification reduces the taxonomic certainty of European biodiversity databases of ivies (*Hedera* L.). Scientific Reports, 14(1), p.4876.

Colwell, A., 2011. *Yosemite's Sky Islands*. https://yosemite.org/wp-content/uploads/d7/useruploads/Yosemite-Sky-Islands-Colwell-FINAL-511.pdf

Culley, T.M., 2013. Why vouchers matter in botanical research. Applications in Plant Sciences, 1(11), p.1300076.

Daniels, S.R., Barnes, A., Peer, N., Egan, V.T., Taylor, R., Taylor, R.W. and van der Colff, D., 2022. iNaturalist is useful at enhancing biodiversity studies as evident from southern African freshwater crabs (Decapoda: Brachyura: Potamonautidae). Journal of Crustacean Biology, 42(3), p.ruac042.

Davis, T. and Wenk, E., 2020. "John Muir Trail. HALS-CA-144." Washington, DC: U.S. Department of Agriculture. National Park Service. Historic American Landscapes Survey.

Dimson, M., Berio Fortini, L., Tingley, M.W. and Gillespie, T.W., 2023. Citizen science can complement professional invasive plant surveys and improve estimates of suitable habitat. Diversity and Distributions, 29(9), pp.1141-1156.

Funk, V.A., Edwards, R. and Keeley, S., 2018. The problem with (out) vouchers. Taxon, 67(1), pp.3-5.

Gaston, K.J., 1999. Implications of interspecific and intraspecific abundance-occupancy relationships. Oikos, 86(2), pp.195-207.

Gervazoni, P., Minuti, G., Fuentes-Rodriguez, D., Coetzee, J., Sosa, A., Sabater, L. and Franceschini, C., 2023. Citizen Science Improves the Known and Potential Distribution of a Strong Wetland Invader: Implications for Niche Modeling and Invasion Management. Environmental Management, 71(6), pp.1176-1187.

[dataset] GBIF.org (22 January 2024a) GBIF Occurrence Download https://doi.org/10.15468/dl.ccdf6t

[dataset] GBIF.org (05 February 2024b) GBIF Occurrence Download https://doi.org/10.15468/dl.6d7fmm

Goodwin, Z.A., Harris, D.J., Filer, D., Wood, J.R. and Scotland, R.W., 2015. Widespread mistaken identity in tropical plant collections. Current Biology, 25(22), pp.R1066-R1067.

Goslee, S.C., 2006. Behavior of vegetation sampling methods in the presence of spatial autocorrelation. Plant Ecology, 187(2), pp.203-212.

Govaerts, R., Nic Lughadha, E., Black, N., Turner, R. and Paton, A., 2021. The World Checklist of Vascular Plants, a continuously updated resource for exploring global plant diversity. Scientific Data, 8(1), p.215.

Jobson, R.W. and Conn, B., 2012a. *Drosera capensis* (Droseraceae), a new naturalised record for Australia. Telopea, 14, pp.89-92.

Jobson, R.W. and Conn, B., 2012b. *Utricularia subulata* L.(Lentibulariaceae): a new weed record of an Australian native for New South Wales, Australia. Telopea, 14, pp.83-88.

Jones, C.D., Glon, M.G., Cedar, K., Paiero, S.M., Pratt, P.D. and Preney, T.J., 2019. First record of Paintedhand Mudbug (*Lacunicambarus polychromatus*) in Ontario and Canada and the significance of iNaturalist in making new discoveries. The Canadian Field-Naturalist, 133(2), pp.160-166.

King, R.J. and Linnean Society of New South Wales, 2022. Field Guide to Royal National Park, New South Wales (Ed. Robert J. King). Second edition. Linnean Society of New South Wales, Sydney.

Lawlor, J.A., Comte, L., Grenouillet, G., Lenoir, J., Baecher, J.A., Bandara, R.M.W.J., Bertrand, R., Chen, I.C., Diamond, S.E., Lancaster, L.T. and Moore, N., 2024. Mechanisms, detection and impacts of species redistributions under climate change. Nature Reviews Earth & Environment, 5, pp.351-368.

Lenoir, J. and Svenning, J.C., 2015. Climate-related range shifts–a global multidimensional synthesis and new research directions. Ecography, 38(1), pp.15-28.

MacArthur, R.H. and Wilson, E.O., 1963. An equilibrium theory of insular zoogeography. Evolution, 17(4), pp.373-387.

Mesaglio, T., 2024. A Guide to iNaturalist: An Australian perspective, Atlas of Living Australia, Publication Series No. 12, Canberra, Australia, pp. 81. https://doi.org/10.54102/ala.33279

Newmaster, S.G., Belland, R.J., Arsenault, A., Vitt, D.H. and Stephens, T.R., 2005. The ones we left behind: comparing plot sampling and floristic habitat sampling for estimating bryophyte diversity. Diversity and Distributions, 11(1), pp.57-72.

Organ, M.K. and Speechley, C., 1997. Illawarra Aborigines, in Hagan, J.S. and Wells, A. (eds), A History of Wollongong, University of Wollongong Press, pp.7-22

Orr, M.C., Hung, K.L.J., Wilson-Rankin, E.E., Simpson, P.M., Yanega, D., Kim, A.Y. and Ascher, J.S., 2023. Scientific note: First mainland records of an unusual island bee (*Anthophora urbana clementina*) highlight the value of community science for adventive species detection and monitoring. Apidologie, 54(5), p.46.

Roberts, C.J., Vergés, A., Callaghan, C.T. and Poore, A.G., 2022. Many cameras make light work: opportunistic photographs of rare species in iNaturalist complement structured surveys of reef fish to better understand species richness. Biodiversity and Conservation, 31(4), pp.1407-1425.

Rumpf, S.B., Hülber, K., Zimmermann, N.E. and Dullinger, S., 2019. Elevational rear edges shifted at least as much as leading edges over the last century. Global Ecology and Biogeography, 28(4), pp.533-543.

Slaton, M.R., 2015. The roles of disturbance, topography and climate in determining the leading and rear edges of population range limits. Journal of Biogeography, 42(2), pp.255-266.

Taylor, D.W., 2010. 'Flora of the Yosemite Sierra.' (La Vergne, TN).

Utjés, M., Nualart, N. and López-Pujol, J., 2022. Misidentifications of herbarium specimens are a great problem for distribution studies: *Kalanchoe* as a case study. https://digital.csic.es/bitstream/10261/284609/1/SPNHC_final.pdf

Weigelt, P., König, C. and Kreft, H., 2020. GIFT–A global inventory of floras and traits for macroecology and biogeography. Journal of Biogeography, 47(1), pp.16-43.

Wenk, E.H., Cornwell, W., Fuchs, A., Kar, F., Monro, A., Sauquet, H., Stephens, R.E. and Falster, D.S., 2024. APCalign: an R package workflow and app for aligning and updating flora names to the Australian Plant Census. Australian Journal of Botany, 72, BT24014. doi:10.1071/BT24014

Werenkraut, V., Baudino, F. and Roy, H.E., 2020. Citizen science reveals the distribution of the invasive harlequin ladybird (*Harmonia axyridis* Pallas) in Argentina. Biological Invasions, 22(10), pp.2915-2921.

Westoby, M., Andrew, S.C., Gallagher, R.V. and Schrader, J., 2024. Species gain and loss per degree Celsius. Oikos, 2024(4), p.e10556.

Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R., Robertson, T. and Vieglais, D., 2012. Darwin Core: an evolving community-developed biodiversity data standard. PloS One, 7(1), p.e29715.

Wilson, K.L. and Keith, D. A., 2022. Plants of Royal National Park. In 'Field Guide to the Royal National Park, New South Wales (Ed. Robert J. King), pp. 91-106. Second edition. Linnean Society of New South Wales, Sydney.