Decolonise Herbaria and Specimen Data;

Quantifying the Contribution of Local Herbaria to Biodiversity Patterns

Jay Delves

September 2021

Thesis submitted in partial fulfilment for the MSc in the Biodiversity and Taxonomy of Plants

THE UNIVERSITY
of EDINBURGH

Royal
Botanic Garden
Edinburgh

Abstract

Botanical institutions are currently seeking to address their role in colonialism, during which the access granted to colonial land enabled the production and development of European botanical science, of herbaria, and of botanical expertise. There is a disparity between the size and spread of herbaria globally, where European herbaria have a disproportionate amount of staff, expert taxonomists visiting and working on their collections, and of specimen data collected from across the world, as well as a disproportionate amount of more secure funding. Meanwhile, local herbaria often hold small but regional collections, are greatly understaffed and underfunded and thus, overlooked by international researchers. Where resources are retained and agendas set in wealthier countries in the Global North, colonial legacies persist in the unequal distribution of globally accessible biodiversity data and data flowing back up north. In global scale biodiversity analyses, poorer counties are quickly deemed data deficient, yet global data repositories such as GBIF, omit the majority of global herbarium data. To examine these biases, we examined how local herbaria and their specimen data contribute to our knowledge of biodiversity patterns. We use taxonomically verified databases including local and international specimen data on the biodiverse genera *Begonia* and *Solanum* in Peru, to examine differences both in the uniqueness of local specimen data and how local data affects IUCN Red List threat assessments. We demonstrate that local herbaria have equally unique herbarium data, but without local herbaria we greatly overestimate the rarity of species, and species even become data deficient. Local herbaria are therefore integral to an accurate understanding of biodiversity patterns. The data void is smaller than we think, but filling it requires an increased focus on mobilising data from local herbaria, which must benefit local botanists and herbaria or risk perpetuating the legacies of colonialism.

Title page: Original drawing from botanical expedition of Ruiz and Pavon to Peru and Chile (1777-1816) held at Real Jardin Botanico (MA); (Left) *Begonia incarnata* by Jose Brunete. (Brunete, no date); (Right) Solanum grandiflorum by Isidro Galvez

Acknowledgements

I would like to thank my principal supervisor Peter Moonlight; whose knowledge and unyielding support was instrumental in the completion of this thesis. Thank you for your patience, openness and creativity, and essential mediation between science and the social sciencies. Many thanks to my supervisors Tiina Sarkinen and Sandra Knapp, for your energy and *Solanum* wisdom. Thank you Edeline Gagnon for coming to the R rescue, Antje Ahrends for your warm and attentive teaching and Martin Pullman for your Padme wizardry. Thank you, mentor, Mark Watson, for your kindness and guidance over the last two years. Thank you to the MSc cohort of 2019 who eased my ignorance and answered my incessant questions, and to fellow artist-scientist Sarah Roberts for reminding me that it’s okay to be an outsider. To my grandma Mary Oram, for whose passing granted me the funding for this course. To mum and dad and Merryn for the endless encouragement, whatever I do. And lastly to William Harper whose politics has shaped my every day, companionship grounds me and for reminding me of what is most important when things got too much.

Contents

1 Introduction: 1

1.1 Decolonise Herbaria: 1

1.1.1 Current Conversations on Decolonising Botanic Gardens, Collections and Scientific Practice 1

1.1.2 What is Decolonisation? Applications and Realities 2

1.2 Herbaria Present and Past: 4

1.2.1 Origins of Herbaria and Botanical Science in Colonialism 4

1.2.2 Contemporary Aims and Uses of Herbaria and Specimen Data 5

1.2.3 Size as a Comparison of the Distribution of Global Herbaria and Herbarium Specimens 7

1.3 How Colonialism has Affected Distribution of Plant knowledge, of Biodiversity patterns and Species distributions 12

1.4 Data Digitisation and Mobilisation 14

1.5 Aims and Objectives: 16

1.6 Study System: *Begonia* L. and *Solanum* L. in Peru: 17

1.6.1 Begonia and *Solanum L.* 17

1.6.2 Peru's Geography and Independence 18

1.7 What is Means to Decolonise Herbaria in this Study 19

2 Methods 20

2.1 Data sources, geo-referencing and cleaning 20

2.1.1 Dataset Summary 22

2.1.2 Duplicates 22

2.1.3 Defining Herbaria and the Data they Hold 23

2.2 Geographic Distribution of *Begonia* and *Solanum* Specimens 27

2.3 The Age and Distribution of Specimen Data 27

2.4 Temporal Geographic Distribution of Specimen Data 28

2.5 Modelling the Uniqueness of Herbaria 28

2.5.1 Uniqueness Indices 29

2.5.2 Multiple Linear Regression 30

2.6 Herbaria and the IUCN Status of *Begonia* and *Solanum* in Peru 31

3 Results 33

3.1 Differences between data in local and international herbaria 33

3.2 Exploring the Temporal and Geographic Distribution of Data in Herbaria.........................34

3.2.1 The Age of Specimens in Local and International Herbaria..34

3.2.2 The Distribution of Collections in Local and International Herbaria across Space and Time 36

3.3 Measuring the Uniqueness of Herbarium Specimen Data 38

3.3.1 The Size and Observed Uniqueness of Herbarium Collections...38

3.3.2 The Mean Uniqueness of Specimens for Local or International Herbaria....................................42

3.3.3 Results from Modelling Uniqueness Indices by in Peru, by Herbarium Accession, by Genera Accession...44

3.4 Contribution of Local Herbaria to Estimates of Range Size and Red List Threat Status 51

3.4.1 Contribution of Herbaria to Range Size Estimates 51

3.4.2 Contribution of Herbaria to Estimates of Range Size by Threat Category...................................52

3.4.3 Changes of Threat Status as Herbaria are Removed...61

4 Discussion...66

4.1 Differences between data in local and international herbaria 67

4.1.1 Can the Value of Herbaria be Quantified by the Uniqueness of the Specimens?..........................68

4.1.2 Discrepancy in the Temporal Geographic Distribution and Age Distribution of Specimens in Local and International Herbaria...73

4.1.3 The Relationship between Herbaria and Duplicates 74

4.1.4 Future Differences to Examine between Local and International Herbaria...................................75

4.2 Local Herbaria are Essential for Measuring Red List Threat Status...................................77

4.2.1 Mobilisation of Biodiversity Data..80

4.3 Future Integration of Decoloniality into Biodiversity Research Practice 82

5 Conclusion: 83

6 Bibliography 85

7 Appendix 94

1 Introduction:

1.1 Decolonise Herbaria:

1.1.1 Current Conversations on Decolonising Botanic Gardens, Collections and Scientific Practice

Natural history museums and botanic gardens are currently seeking to acknowledge and address both their contemporary and historic roles in colonialism, roles which have served to perpetuate structural racism within their institutions and to whitewash history (Das and Lowe, 2018). The term *decolonise* has been used to describe this process. Natural history institutions have proposed to decolonise their work by raising awareness of narrative contained within their biological collections, narratives that highlight the violent and oppressive relationships between people and plants that occurred in the expansion of European empire (National Museums Scotland, no date; Dundee, 2020; Museums Galleries Scotland, 2020; NatSCA, 2020). RBG Kew's director of science Alexandre Antonelli has publicised aims to 'decolonise botanical collections' (Antonelli, 2020). To this effect that Kew have outlined plans in their 2021 'Our Manifesto for Change' that aim to contextualise their position both now and in the past, protect plants for the well-being of all people and open a dialogue on the role they have played in Britain's colonial legacy (RBG Kew, 2021). They propose to 'promote diversity', 'additional perspectives', and 'disseminate their knowledge for the global benefit' (RBG Kew, 2021). RBG Edinburgh in approaching racial injustice and discrimination have similarly promised to decolonise their collections through storytelling, improving access to collections digitally and improving equity and opportunity across their institution (RBGE, 2021b). In 2017, Leiden University and Naturalis Biodiversity Center announced they were intending to decolonise their plant collections by making them accessible to the people who Holland colonised, from whom knowledge was taken for the advancement of western botanical science (van Andel, 2017). Science academics are also beginning to recognise the role that colonialism has both had and continues to play within each discipline, and are suggesting ways to decolonise current ecology, biogeography and geography practice (Legg, 2017; Radcliffe, 2017; Baker, Eichhorn and Griffiths, 2019; Trisos, Auerbach and Katti, 2021).

1.1.2 What is Decolonisation? Applications and Realities

Decolonisation or to decolonise is a broad term, that whilst describes a complex set of anti-colonial ideas or actions, holds different values and meanings unique to many people, as is the desired outcome (Sium, Desai and Ritskes, 2012). Decolonisation in its simplest form is the formal grant of constitutional independence to a subordinate state by a colonial power, most frequently referred to in the dismantling of European imperialism from the late 19th century which accelerated after World War II (Riches and Palmowski, 2021). A more comprehensive definition involves the act of a state gaining economic and cultural sovereignty and self-determination post colonisation (Coggins, 2018). The extent to which this broader term would suggest that complete decolonisation has yet to be achieved in the 20th Century (Coggins, 2018). Further revealing the longstanding structures and conditioning enforced by the dominion over a state and indigenous people. Decolonisation is then ‘a long-term process involving the bureaucratic, cultural, linguistic and psychological divesting of colonial power’ (Smith, 2010) which centres indigenous, marginalised life, community and epistemology (Sium, Desai and Ritskes, 2012). It seeks to undermine dominant Euro

American knowledge production which though formed in the imperial project and in Enlightenment-modernity's pursuit of knowledge, and were built on claims to be able to pronounce universal truths about the world (Radcliffe, 2017), still linger.

The 'colonial presence' (Stoler, 2016) and applications of western knowledge continue to be reproduced for example in academia, as evident in the principal residency of authors in Europe and North America that oversee projects beyond their national boundaries, that retain resources, funding and set research agendas overseas (Eichhorn, Baker and Griffiths, 2020). As a beginning, in confronting decolonial approaches to ecology, Trisos, Auerbach and Katti., (2021) outline five arguments for anti-oppressive practice: to decolonise your mind, know your histories, decolonise access and expertise, and integrate inclusive teams. Actionable ideas which are echoed in approaching decolonisation in similar academic fields (Baker, Eichhorn and Griffiths, 2019; Eichhorn, Baker and Griffiths, 2020). Decolonise in these contexts takes on an extended, potentially intangible meaning from the original definition, that is subject to criticism that decolonise as a verb and as a metaphor reproduces settler appropriation, disguised as reconciliation, by acting to do anything other than repatriation of land, power and privilege to indigenous people (Yang and Wayne, 2012). Instead, decolonisation is 'incommensurable' (having no common standard of measurement) and 'not equivocal to other anti-colonial struggles' (Yang and Wayne, 2012). Dispossession may be the 'truth' of colonialism particularly in the settler-colonialism in the Americas, yet it does not encompass the various forms of imperial debris and of formerly colonised people, historically and geographically uniquely situated (Gurminder, Bhambra and Nişancıoğlu, 2018). Whilst decolonial authors from the Global North may seek to challenge their knowledge systems born within their scientific fields, dismantling of colonial epistemologies and theoretical orthodoxies cannot necessarily be fully achieved within the very institution and infrastructures that founded and reproduce them (Jazeel, 2017; Noxolo, 2017). Discourse on Decolonising the University recognise these limitations but are committed to addressing the ties of coloniality and pedagogy from within the imperial centre (Gurminder, Bhambra and Nişancıoğlu, 2018). In search of better fitting word, Legg (2017) proposes *decolonialism* rather than decolonising, to denote the ongoing process of challenging the practices that made and sustain colonies, not the act of relinquishing colonies.

1.2 Herbaria Present and Past:

1.2.1 Origins of Herbaria and Botanical Science in Colonialism

Herbaria, the collections of preserved plant specimens, were formed within the imperial project alongside botanical gardens and the advancement of botany as scientific discipline. Botanic gardens such as Kew whilst initially set up for medicinal research quickly became a revenue-making business to fund the objectives and expansion of empire through the production of profitable plants (Brockway, 1979) . Most noticeably implicated in the illegal smuggling of *Cinchona* seeds (for quinine) from Peru and Bolivia which approved by Kew director William Hooker in late 19th century, enabled the treatment of the colonial Indian Army, suffering Malaria, to go into unexplored territories and successfully colonise parts of Africa previously unattainable (Baber, 2016). Networks of botanic gardens established at that time were set up strategically within and outside colonies, to assist the mobility of plants, knowledge, people, power and profit (Baber, 2016) and eventually globalisation. Botanists directly benefited from the access to colonial land (Trisos, Auerbach and Katti, 2021) in which they were able to record, collect and study the worlds flora to develop their discipline.

Moreover in advancement of their scientific discipline they aided the colonial conquest by providing knowledge on the cultivation of crops, species selection and suitable habitat for plantations in colonies (Brockway, 1979). Botanical collections in their amasses were fundamental to the development of a European rationalisation of knowledge, a framework in which to impose order on morphological chaos, with a system for naming all the flora and fauna of the world, taxonomy (Bowker and Star, 2000). The rise of Linnean binomial system simply overwrote indigenous names and plant knowledge, and often in honour of foreign sponsors or collectors (Gillman and Wright, 2020). Further, polymath Jose Antonio Alzate argued that the dominion of botanical classification, solely based on morphology, additionally decontextualises and erases local plant knowledge by ignoring the internal plant properties and plant uses (Bleichmar, 2012). This universal system devised the professionalisation of botany from the amateur overseas explorer into a formal discipline with botanists; those learned in the European taxonomy and nomenclature of plants (Baber, 2016) and herbaria, the ever expanding collections from which to record the diversity of the natural world. While the collection and knowledge of plants and their uses vastly pre-dates European cultures, the establishment of collections of preserved specimens (herbaria) was distinctly formed in Europe (Thiers, 2020a) with the development of western botanical science. Herbaria are hence rooted in European colonialism.

1.2.2 Contemporary Aims and Uses of Herbaria and Specimen Data

Herbaria whilst primarily served to train botanists in the formal taxonomy of plants, are currently being used in ways previously unimaginable, far from their original collected intention (Heberling and Isaac, 2017; Soltis, 2017; Marsico *et al.*, 2020). Currently being described as exaptation’s due to their co-option into other fields, specimens are also being used at a growing rate (Heberling and Isaac, 2017). Herbaria in addition to displaying the preserved plant morphology and material for DNA extraction, are repositories for a diversity of data available across vast temporal and spatial scales. For example, they are being used in biodiversity conservation to measure species distributions and extinction risk, can be used to assess phenological changes in response to climate change, to measure levels of industrial pollution in the soil and to decipher evolutionary relationships in phylogenetic studies (Lavoie, 2013; Soltis and Soltis, 2016; Soltis, 2017; Heberling, Prather and Tonsor, 2019; López and Sassone, 2019; Marsico *et al.*, 2020). Whilst central to taxonomic work, they are utilised in the scientific fields of population genetics (Wandeler, Hoeck and Keller, 2007), molecular systematics (Soltis and Soltis, 2016), environmental chemistry (Gritcan *et al.*, 2016), herbivory (Zangerl and Berenbaum, 2005), phytopathology (Hood *et al.*, 2010), to name a few. As we enter a new era of herbaria (Heberling, Prather and Tonsor, 2019), the importance of collections cannot be understated for contributing to novel scientific inquiries.

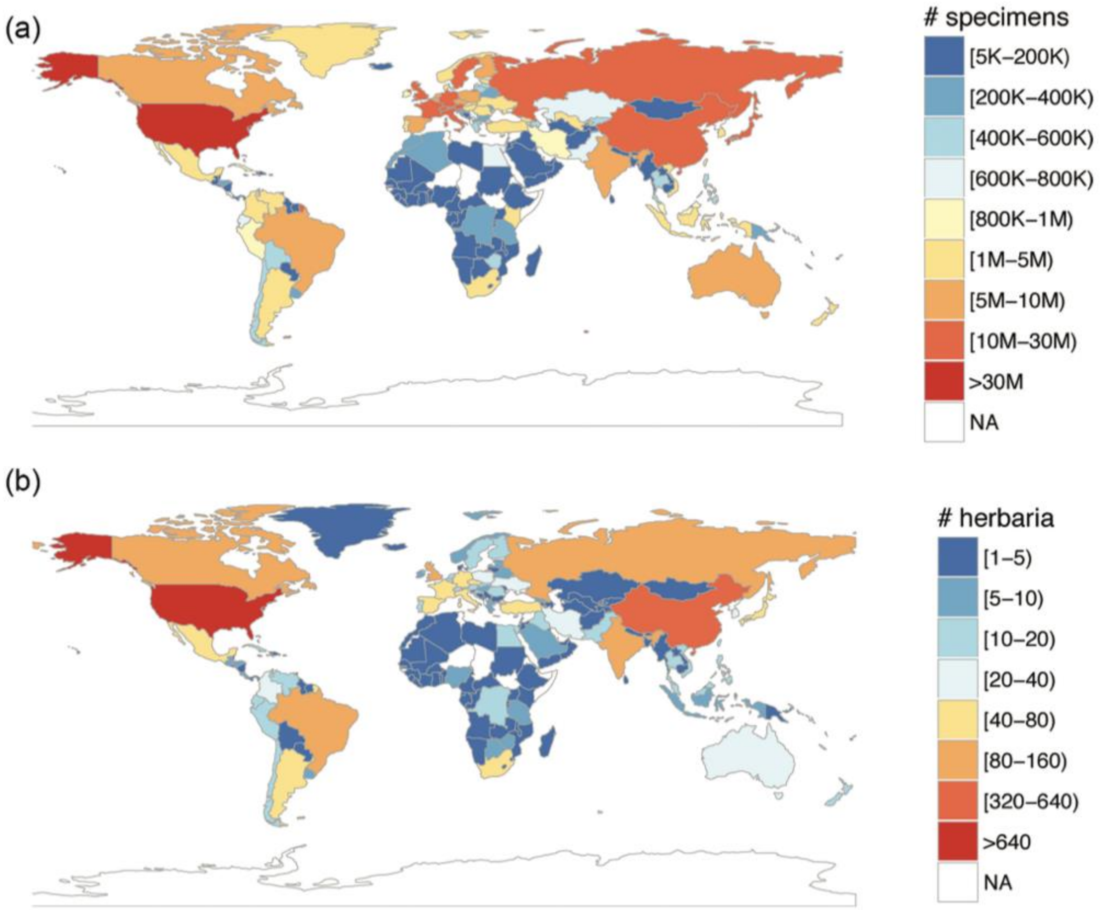


Fig.1 - Figure from Heberling et al. 2019: (a) the number of specimens; (b) the number of active herbaria as recorded in Index Herbariorum: The Worlds Herbaria 2018 Report (Thiers 2018)

1.2.3 Size as a Comparison of the Distribution of Global Herbaria and Herbarium Specimens

Herbaria today, are a vast archival source recording the worlds plant and fungal diversity, some collections dating back over five centuries (Heberling, Prather and Tonsor, 2019). There are currently 3,426 active and registered herbaria in the world today, holding a total of 396,204,891 specimens (Thiers, 2020b). Herbaria differ majorly across national and international borders, in their history and collections, in their structure, funding and consequently capacity, and focus. They can be associated and placed within botanical gardens, museums, universities or in private collections. Herbaria can differ greatly in the amount and size of the collection, where often the largest herbaria are in found within the Global North, dominant in Europe and N. America (Fig. 1) (Heberling, Prather and Tonsor, 2019). Very large herbaria with more than 5 million specimens and an early commence date (Table 1) are found predominantly in European countries with a colonial past (Lavoie, 2013). Large herbaria additionally have a geographic focus with their collections that is at a worldwide scale (Table 1) in comparison to local or smaller herbariums that do not, and are national in scope, or concentrated on a particular ecological area or reserve (Table 2). The "international vocation" of large herbaria has resulted not only in the frequent use by internal and external researchers in which they are consulted three to six times more than smaller herbaria, but interestingly they are consulted more than 90% for their international rather than local collections (Lavoie, 2013). Only possible through the access colonialism granted, large herbaria and the institutions that hold them have greatly benefited from established collections of world-wide flora, that have sustained the funding, careers and international scientific reputation of the countries they are placed within.

Considering the frequently used boundary where small herbaria have less than 100,000 specimens, 85% of herbaria in the world are small (Marsico *et al.*, 2020; Thiers, 2020b). Due to their size they are often underfunded, operate with minimal staff, curatorial roles are usually split with other responsibilities including teaching or research, and digitisation not possible (Harris and Marsico, 2017). Typically, herbaria with a smaller number of accessions have collections that are regionally (or locally) focused. They are however found to hold important ecological, taxonomic or geographically bias plant data, and hold local and indigenous knowledge (Monfils, A.K., Nelson, 2014; Monfils *et al.*, 2020). “Small” herbaria in the USA have been shown to have geographically and temporally unique herbaria and records not duplicated elsewhere (Marsico *et al.*, 2020). Similarly regional herbaria in Brazil, often overlooked, were found to significantly contribute to biodiversity analysis on national flora (Colombo, Kaehler and Calvente, 2016). Differences in size of herbaria can largely be accounted for from the unequal distribution of power, funding and knowledge over time. Yet there is an argument for the preservation, funding, development and digitisation across all herbaria equally, as size is not relative to the value of records they hold (Marsico *et al.*, 2020; Monfils *et al.*, 2020)

Largest Herbaria Globally

Institution	Herbarium Code	Number of Staff	Total Accession	Country	Locality	Geography of Herbaria	Date Commenced
Royal Botanic Gardens	K	69	8125000	U.K.	England	Worldwide, especially Africa, tropical Asia, and Australasia	1852
The New York Botanical Garden	NY	36	7921000	U.S.A.	New York	Worldwide with greatest strength in tropical America and North America	1891
Missouri Botanical Garden	MO	58	6850000	U.S.A.	Missouri	Worldwide; Central America (especially Costa Rica, Nicaragua, and Panama), tropical South America (especially Bolivia, Colombia, Ecuador, Paraguay, Peru, and Venezuela), Africa, and Madagascar; North America, Neotropics	1859
Komarov Botanical Institute of RAS	LE	35	6000000	Russia	Saint Petersburg	Worldwide; Russia, Asia, Europe, North America, Africa, Antarctic, arctic, Indian Ocean	1823
Conservatoire et Jardin botaniques de la Ville de Genève	G	30	6000000	Switzerland	Genève	Worldwide, especially Mediterranean, Middle East, South America, Africa, Madagascar, and regional	1824
Muséum National d'Histoire Naturelle	P	21	6000000	France	Paris	Worldwide, especially Africa (northern, western, and equatorial), south-eastern Asia, France, Europe, French Guiana, Madagascar, Mascarene and subantarctic islands, New Caledonia, Horne Islands, and Vanuatu	1635
Naturhistorisches Museum Wien	W	6	5500000	Austria	Wien	Worldwide	1807
The Natural History Museum	BM	32	5200000	U.K.	England	Worldwide, especially of British Isles, Europe, Africa, North America, West Indies, and Himalaya	1753
Smithsonian Institution	US	34	5100000	U.S.A.	District of Columbia	Worldwide with emphasis on neotropics, North America, Pacific Islands, Philippines, and Indian subcontinent	1848

Table 1: The top 10 largest herbaria in the world from (Thiers, 2016); Large accessioned herbaria tend to have a worldwide geographic focus, a greater amount of staff, and have an early commence date established during colonial rule

Herbaria with <5000 Accession							
Institution	Herbarium Code	Number of Staff	Total Accession	Country	Locality	Geography of Herbaria	Date Commenced
Palestine Museum of Natural History	PMNHH	2	755	Palestinian Territories	West Bank	Historic Palestine; the West Bank; the Middle East; Israel; Palestinian Territories	2018
National Institute of Technology and Evaluation	NBRC	3	900	Japan	Chiba	Japan	2002
Manitoba Conservation	MDNR	1	1525	Canada	Winnipeg	Manitoba.	1938
Universidad Politécnica De Cartagena	UPCT	5	1900	Spain	Murcia	SE Spain (Murcia and Albacete provinces, principally)	1999
University of Botswana	PSUB	4	2000	Botswana	Maun	Okavango Delta; northern Botswana	1996
Universidad Austral de Chile	VALD	3	2010	Chile	Valdivia	Central, Magallanic and Valdivian Region of Chile.	1965
Saasveld, Port Elizabeth Technikon	SAAS	2	2500	South Africa	George	Western Cape	1920
Mindanao State University, Tawi-Tawi College of Technology and Oceanography	MSU	3	2500	Philippines	Western Cape Province	Sulu Archipelago, Philippines	2019
Central Salt and Marine Chemicals Research Institute	BHAV	1	3000	India	Gujarat	India, especially from Gujarat and Maharashtra coast	1961
Botanica Royal Park of Palm Hills	BRPP	3	5000	Egypt	Cairo	Local region	2015

Table 2: A selection of 10 global herbaria with <5000 specimens (Thiers, 2016); Small accessioned herbaria tend to have a narrow geographic focus, often regionally or nationally, have a small amount of staff and have a recent commence date in the last century

1.3 How Colonialism has Affected Distribution of Plant knowledge, of Biodiversity Patterns and Species Distributions

The climate and biodiversity crisis has intensified efforts to quantify the scale and impact of biodiversity loss at a global scale (Marchese, 2015; Enquist *et al.*, 2019; Nic Lughadha *et al.*, 2020; Pimm, 2021). Understanding predicted changes in species distributions and habitat availability over time are essential for prioritising conservation actions and directing action to achieve globally agreed biodiversity goals, such as the Aitchi targets (Meyer *et al.*, 2015). Global mapping of biodiversity hotspots (Myers *et al.*, 2000), has largely revealed that the tropics are known to contain the vast majority of the world's terrestrial biodiversity (Feeley and Silman, 2011). These areas, subject to much global attention, have however been deemed as data voids (Kier *et al.*, 2005). Data voids consist of large gaps in knowledge, specifically on a habitat or group of organisms, and for plant data this is characterised by a paucity of herbarium specimens. In fact, most tropical plant species are known from only one specimen, redeeming them data deficient, while in South America 33% are known from ≥ 5 specimens (Feeley and Silman, 2011). Known as the Wallacean shortfall, the sustained lack of knowledge on the distributions of species is a historic as well as contemporary predicament (Hortal *et al.*, 2015). The predicted geographical spread of species however closely resembles survey effort, subject to human-mediated collecting bias (Hortal *et al.*, 2015; Daru *et al.*, 2018). These voids greatly affect the accuracy of ecological modelling, and identifying the bias that creates them is vital to improving biodiversity analysis (Meyer *et al.*, 2015)

Data deficiency disproportionately affects poorer areas of the world (Eichhorn, Baker and Griffiths, 2020) and the wealth of a country (GDP) has been found to equate to the lack of and mobilisation of data. Further socio-economic factors, such as, geographic location between country of collection and host databases, speaking English, and national security amplify the bias in available biodiversity data (Amano and Sutherland, 2013). Meyer et al., (2015) found that the prominent drivers of inventory incompleteness for animals consisted of the spatial distance from data-contributing institutions, national participation in the largest global data repository, the Global Biodiversity Informatic Facility (GBIF), and national research funding. While some countries are data deficient, others are resource rich and have been intensively studied (Kier *et al.*, 2005) resulting in a chronic inequality in the distribution of knowledge globally (Vorontsova *et al.*, 2021). For example, Vorontsova et al., (2021) found that geographic differences in taxonomic knowledge between grasses in Britain and Madagascar, greatly affect our understanding of species occurrences, to the extent that resource-poor areas have experienced unrecorded plant extinctions.

Yet data, albeit digital and desk-based, is still extractive (Baker, Eichhorn and Griffiths, 2019), evidenced in the flow of data out of poor and into rich countries that maintain the resources, funding and expertise (Eichhorn, Baker and Griffiths, 2020). A global assessment of taxonomic expertise of Amphibians found that a great amount of expertise resided in economically rich countries in North America and Europe, but was ‘exported’ to other parts of the world where researchers travelled for work (Rodrigues *et al.*, 2010). Further to this, an assessment of the residency of authors writing on global biodiversity, showed that 73% live in the Global North whereas 13% in the tropics (Eichhorn, Baker and Griffiths, 2020). Another study on Amazonian publishing found that while publications on the local geography increased, local authors did not (Malhado *et al.*, 2014). This disparity between country of origin and country of study highlights the stark reality that Global North authors still depend on the flow of biogeographical knowledge from global south, tracing the same scientific colonialism of knowledge making and dissemination (Eichhorn, Baker and

Griffiths, 2020). The consequence is that dominant knowledge paradigms persist, biological agendas are set abroad for “data deficient” countries and solidified in policy. This can often be to the detriment of local people, where local knowledge is continually written out of research processes (Toomey, 2016; Baker, Eichhorn and Griffiths, 2019). For example, the biogeographical composition of Amazonian forests has been greatly shaped by pre-Colombian human disturbances but previously not included in analyses on species composition (McMichael *et al.*, 2017). This demonstrating the skewed geographical knowledge and ignorance that can persist when research is taken from afar (Baker, Eichhorn and Griffiths, 2019) and further demystifies naive foreign concepts of an ‘exotic’, untouched, unrecorded biodiversity. These disproportionate differences in knowledge, geographic spread of data, data flow and expertise are evidence of how colonialism shaped and sustains our current knowledge of plant biodiversity (Eichhorn, Baker and Griffiths, 2020).

1.4 Data Digitisation and Mobilisation

Efforts to digitise and democratise specimen data have greatly accelerated in the last two decades (Canteiro *et al.*, 2019; Heberling *et al.*, 2021), aiming to improve research opportunities globally and enabling the study of global change (Soltis, Nelson and James, 2018; Sweeney *et al.*, 2018). Despite digitisation of biological collections in some institutions commencing in the late 1970’s (Sunderland, 2013), digitising collections is a time-consuming and expensive process (Sweeney *et al.*, 2018) and efforts have been unevenly spread across herbaria globally. The same biases that shape the unequal distribution of biodiversity data, knowledge and expertise, unsurprisingly are formed by the ability to digitise and mobilise specimen data globally. Participation even, is reliant on the national structure of funding in each country (Beck *et al.*, 2014) and moreover, to even utilise accessible data is dependent on reliable internet connection and computer access (Drew, Moreau and Stiassny, 2017; Vorontsova *et al.*, 2021). Digitisation, further, is driven variably between institutions, often by funding priorities, convenience or individual interests (Daru *et al.*, 2018; Hedrick *et al.*, 2020). RBGE for instance, has prioritised and completed digitisation for all type specimens, two geographical regions and four vascular plant families (RBGE, 2021a); likely pertaining to large and internationally unique collections and unique type specimens.

Whilst the largest global bioinformatic portal, GBIF, with almost 2 billion records, has the potential to diversify research access and expertise globally, evidenced in increased visitation from global south countries (India, Brazil, Columbia, Mexico) (Drew, Moreau and Stiassny, 2017), the data available is regionally biased (Meyer *et al.*, 2015; Daru *et al.*, 2018). If the database is predominantly formed of Global North institutions and authors (Heberling *et al.*, 2021) that have the long-term funding to digitise and disseminate specimen data, immobilised specimen data from other parts of the world would highly influence any biodiversity analysis outcome. Legacies of colonialism are further evident in the dominant authorship of GBIF data from European authors, and where research on global south topics is authored by Global North researchers (Heberling *et al.*, 2021). Therefore, a biased database can further enhance the careers of authors whose data is available and utilised in published manuscripts. Outside Europe and North America, several large digital repositories have been successful mobilising regional data, such as Mexico's Comision Nacional Para el Conocimiento y Uso de la Biodiversidad (CONABIO) and Brazil's Centro de Referencia em Informacao (CRIA), holding 6 millions and 9 million records, respectively, some of which are served by GBIF (Nelson and Ellis, 2019). The government of Brazil also distinctively launched Reflora programme in 2010, with the purpose of retrieving digitised specimens from overseas herbaria (Canteiro *et al.*, 2019; Nelson and Ellis, 2019). These databases are proven sufficient in reducing the number of data deficient species in Brazil (Sousa-Baena, Garcia and Townsend Peterson, 2014). Integration of data must therefore be focused towards global participation and mobilisation of non-western data, in order to optimise biodiversity analysis and equity in science practice (Meyer, Weigelt and Kreft, 2016; Heberling *et al.*, 2021).

1.5 Aims and Objectives:

Considering the recent studies highlighting the importance of regional and small herbaria in the USA and Brazil (Colombo, Kaehler and Calvente, 2016; Marsico *et al.*, 2020), we expect that exploring the specimen data of biodiverse genera from a unique locality, will contribute to quantifying the value of often overlooked herbaria in biodiverse regions. Therefore, the main aim of this study is to investigate whether data held in local herbaria contribute to our understanding of biodiversity patterns and investigate whether there are *spatial, temporal, or taxonomic* biases in the specimen data held in local and international herbaria. Further, we aim to reflect on how historic and contemporary colonialism has affected the size and distribution of biodiversity data and consequently the outcome of biodiversity analyses. Our objectives are to determine whether:

a. Local and international herbaria have different temporal or geographic distributions of specimens
b. Local and international herbaria differ in the age of their specimens
c. Local herbaria hold unique collections in comparison to International herbaria
d. Species are estimated to be less threatened when data from local herbaria are included (re-calculation of ICUN status using local herbaria)

1.6 Study System: *Begonia* L. and *Solanum* L. in Peru:

1.6.1 Begonia and *Solanum L.*

Our study system in which to compare biodiversity patterns between local and international herbaria is *Begonia* and *Solanum* found in Peru. Both genera are placed within the top 10 largest genera in the world (Frodin, 2004), currently with >2000 *Begonia* spp., and c.1300 Solanum spp., worldwide (Hughes *et al.*, 2015; Christenhusz, Fay and Chase, 2017). *Begonia* is one of two genera placed with family Begoniaceae, in order Cucurbitales (Mabberley, 2017). The most recent molecular phylogenetic study on the genus found strong support for monophyly between sections and clades, largely grouped by continental distribution (Moonlight *et al.*, 2018). Placed within family, Solanaceae, order Solanales, the genus *Solanum* contains the greatest number of species compared to other genera in the family (Mabberley, 2017). Major clades have strong monophyletic support at family level and between major clades in *Solanum* (Särkinen *et al.*, 2013). Recent treatments on the Flora of Peru for *Begoniaceae* are currently in prep., (Moonlight et al., 2021), revised from the last family treatment by Smith and Schubert nearly a century ago (Smith and Schubert, 1941). An annotated checklist for *Solanum* in Peru has been recently compiled by Särkinen et al., (2015), updated from a checklist of endemic Solanaceae spp., by Knapp, Spooner and Leon in 2006 (Knapp, Sánchez and León, 2006) and revised from Flora of Peru on Solanaceae published in 1962 by Macbride (Macbride, 1962).

Both genera are distributed worldwide, yet each genus has great species diversity within the tropical Americas and in particular within the Andes (Moonlight *et al.*, 2015; Särkinen *et al.*, 2015). In Peru, *Begonia* is represented by 76 spp., most of which are herbaceous and 38 that are endemic to Peru (León and Monsalve, 2006). Endemics are found within the humid montane, premontane forest regions and humid lowland Amazonian forests at an elevation between 400-1500m (León and Monsalve, 2006). Species of *Solanum* can be herbs, shrubs, trees or lianas, the genus containing many important cultivated crop species (Särkinen *et al.*, 2013). There are 276 spp., of *Solanum* in Peru, 253 spp., that are native, 23 that have been introduced or cultivated and 74 spp., of which are endemic to Peru (Särkinen *et al.*, 2015). Most diversity of *Solanum* in Peru is found at an elevation of 2500-3000m (Särkinen *et al.*, 2015).

1.6.2 Peru's Geography and Independence

Peru is South Americas third largest country, located in the west by the pacific ocean, bordered by Ecuador and Columbia in the north, Brazil and Bolivia in the east, and Chile to the south (Hill, 2017). Peru is divided into three geographic areas, the coastal plains and deserts in the west, the high Andean mountains from north to south and lowland amazon basin in the east (Weberbauer, 1936). Country-wide phytogeographic patterns are greatly influenced by the Andes and it is generally agreed, that ancient Andean uplift formed mountains that isolated the lowlands and limited dispersal, resulting in high endemism (Swenson *et al.*, 2012). The tropical Andes is the top biodiversity hotspot globally, recognised for the highest number of endemic plants and animals' species globally (Myers *et al.*, 2000). Prior to the independence in 1821, Peru as we know it today was a significant part Spanish colony and part of the Spanish Empire (*South America: Peru*, 2021). Spanish entry in 1521 and conquest in 1533 saw the decline of the largest pre-Columbian empire, the Incas (*South America: Peru*, 2021).

1.7 What is Means to Decolonise Herbaria in this Study

Considering the origins and production of herbaria in colonialism, and the of dominant paradigm of European botanical science that continue to shape an unequal distribution of biodiversity knowledge, expertise and resources, it is necessary to reflect on how these legacies can be challenged. In this study we follow the paper by Trisos, Auerbach, Katti., (2021) on anti-oppressive and decolonial practices as guidance to studying the differences in herbaria and specimen data. In particular we will focus on 'know your histories' to reflect on how historic colonialism has formed dominant knowledge systems, differences in size of resources and funding in biodiversity institutions. We will reflect on 'decolonise access', to recognise how access to herbaria and digital specimen data is asymmetrically spread, which affects both expertise, authorships and data flow globally. We will consider 'decolonise expertise' by endeavouring to demonstrate how local knowledge and local herbaria on a national scale are important to our understanding of global biodiversity. Recognising that local knowledge and local experts are best fit to study and map their local flora. However, in this study we only consider differences in knowledge on a global scale and under the paradigm on European botanical science and herbarium science, and in this study do not discuss alternative knowledge systems.

2 Methods

2.1 Data sources, geo-referencing and cleaning

Herbarium specimens of Peruvian *Begonia* and *Solanum* were extracted from expert verified specialist online taxonomic databases (Hughes *et al.*, 2015; PBI Solanum Project, 2021). The list of accepted species for the two study genera were taken from the recently published taxonomic checklists of *Solanum* in Peru (Särkinen *et al.*, 2015) and the Flora of Peru account for Begoniaceae (Moonlight et al., 2021), in which authors on both genera visited many herbariums in Peru to assemble their accounts. A minor difference between the two datasets is that all Peruvian *Begonia* specimens held in any given herbarium at the time of an expert visit were databased and incorporated into the Begonia Resource Centre, where any specimens incorporated after these dates are missing from the dataset. However, not all Peruvian *Solanum* specimens from Solanaceae Source were digitised on expert visits to herbaria in Peru, nor were all Peruvian herbaria visited, leaving more *Solanum* data in Peru herbaria to be digitised and mobilised in the future.

Georeferencing was completed for specimens of *Begonia* that had no coordinates. Georeferencing was done manually using locality descriptions and elevation available on specimen labels with Google Earth Pro v.12.2.2. Maps acquired from The Ministry of Transport and Communications (Ministerio de Transported y Comunicaciones Peru, 2016) were essential in finding small localities, roads or reserves undetectable from Google Earth Pro v.12.2.2. Search engine, Google and Google Translate were used for language translation and to interpret localities where names may have been incorrectly annotated, misspelled or changed over time. Once a locality was found specimens were geo-referenced to degrees, decimal minutes, prioritising the elevation specified in the label and searching along roads. No geo-reference was assigned if the specimen could not be determined to be collected within a 15 km radius of a locality. Georeferenced coordinates once allocated were added to the Begonia Resource Centre database (Hughes *et al.*, 2015), and to the Flora of Peru, Begoniaceae. No geo-referencing was required for the *Solanum* data.

Data of *Begonia* and *Solanum* specimens in Peru were downloaded to Microsoft Excel and sorted for analysis. For *Begonia* a total of 4,784 specimens were downloaded, and a total of 3,288 had been 'seen', signifying that duplicates has been distributed and processed at intended herbariums. A total of 2,952 specimens were georeferenced, and 10.22% could not be georeferenced and were excluded from any analysis. Further adjustments were made by updating synonyms and merging sub-herbariums to main on-site herbarium (G, G-BOIS, G-DC). Introduced species, *B. cucullata and B. heracleifolia,* were not excluded from this dataset due to the nature of the study question, whereby unnatural distributions of species will not affect our proposed study hypothesis on differences in herbaria. Infraspecific species were changed to species level to avoid errors from common misidentification at infraspecific level. Any species not identified to species level, such as only to genus or section within *Begonia*, were excluded from the dataset.

For *Solanum*, 23,047 specimens from Peru were downloaded. A total of 20,481 specimens were georeferenced, and 11.13% were excluded from the analysis because they lacked coordinate data. Sub-herbariums were merged in the dataset and cultivated, and invasive species were included as outlined above. The infraspecific taxon *Solanum stramoniifolium* var. *inerme* was treated as *S. stramoniifolium* for ease of analysis. Three outlier species specimens of *Solanum robustifrons* were removed removed after expert verification of our input data.

Data was also downloaded of all Peruvian *Begonia and Solanum* specimens from GBIF and the percentage calculated for specimens held in local versus international herbaria.

2.1.1 Dataset Summary

Our dataset comprised of 76 species of Peruvian *Begonia* and 282 species of *Solanum*. The *Begonia* dataset had a total of 3288 specimens including duplicates, of which 2952 were georeferenced and 3132 had a collection date. The mean number of specimens per species in *Begonia* is 43. The *Solanum* dataset had a total of 23,047 specimens, of which 20,481 were georeferenced and 22,443 had a collection date. The mean number of specimens per species in *Solanum* is 82.

2.1.2 Duplicates

All duplicates (i.e. individual specimens collected by the same author, on the same date, from the same location) held in different herbaria, were treated as independent collections due to the reality that when studying specimens, the location of the duplicates are often, if not always unknown. Using duplicates from these datasets, where collectors have recorded all duplicates they have found, therefore provides unique information in which to compare herbaria.

2.1.3 Defining Herbaria and the Data they Hold

In order to compare herbaria and the specimen data they hold, we will refer to local herbaria as any herbarium located within the national borders of Peru. Whereas international herbaria will refer to any herbaria and specimen data held outside Peru. In studies looking at small herbaria and their collections, the accession number (the number of sheets of all species held within an herbarium) has been used to categorise herbaria into small or large, whereby small has been counted as <100,000 accessions and large >100,000 accessions (<175,000 has also been used to categorise small herbaria) (Lavoie, 2013; Monfils, A.K., Nelson, 2014; Glon *et al.*, 2017; Marsico *et al.*, 2020; Thiers, 2020b). Size categories, whilst informative, are an arbitrary value to define herbaria by, where size of herbaria can vary widely across taxonomic discipline and terminology can undermine the impact potential of small herbaria (Monfils, A.K., Nelson, 2014; Monfils *et al.*, 2020). That being said, for this study size of accession will be referred to in order to be comparative to other studies using size, where local herbaria will also denote that the size of the herbarium accession is less than <100,000. This being the conservative boundary to which a small herbarium is defined. The only large herbarium in Peru USM, will be referred to as a large local herbarium. In our datasets, the *Begonia* specimens are stored at 33 herbaria globally, including 7 Peruvian herbaria and 26 international herbaria. *Solanum* specimens are held in 72 herbariums globally, including 10 Peruvian herbaria and 62 international herbaria (Table 3). All herbaria in Peru have a geographic focus of their collections either regionally or nationally (Table 3).

Herbarium	*Total Accession*	*Country*	*Locality*	*Number Staff*	*Commenced Date*	*Geographic Focus*
HAO	3000	Peru	Trujillo	3	1991	Northern Peru
HUSA	18500	Peru	Arequipa	4	2000	Peru.
CPUN	35000	Peru	Cajamarca	6	1966	Peru
MOL	38000	Peru	Lima	9	1948	Peru, especially central Peru and the Department of Lima
AMAZ	45000	Peru	Iquitos	2	1972	Amazonian Peru
CUZ	52000	Peru	Cusco	9	1936	Southern Peru
HUT	60000	Peru	Trujillo	11	1941	Northern Peru
HOXA	73000	Peru	Oxapampa	8	2003	Central forests of Peru
USM	800000	Peru	Lima	16	1918	Tropical Amazon Basin and Andes of Peru
CGG	14000	U.K.	Cambridge	2	1846	
PTIS	15000	U.S.A.	Sturgeon Bay	2		Southwestern U.S. to south-central Chile
GL	45000	U.K.	Glasgow	2	1780	Worldwide; British, especially of Glasgow
HB	85000	Brazil	Rio de Janeiro	9	1958	Brazil.
UT	127000	U.S.A.	Salt Lake City	7	1870	Especially of Utah, western U.S., and temperate regions
ALCB	130954	Brazil	Salvador	7	1950	Northeastern Brazil
PMA	148375	Panama	Panama	8	1968	Panama, some from Costa Rica, Colombia, and other regions
BHCB	190000	Brazil	Belo Horizonte	6	1968	Cerrados, campos rupestres (rocky grasslands), and inland semideciduous forests of Minas Gerais; Serra da Piedade and Serra do Caraca
QCA	205000	Ecuador	Quito	5	1971	Ecuador, including GalÃ¡pagos Islands and South and Central America
MG	209320	Brazil	Belem	21	1895	Amazonia, including regions outside of Brazil; Projeto Flora AmazÃ´nica, Amazonian restingas, Serra dos CarajÃ¡s, FLONA CaixuanÃ£, Volta Grande do Xingu, and Serra do Cachimbo.
CONN	225000	U.S.A.	Storrs	9	1898	Northeastern North America, especially Connecticut and New England; North America; New World tropics; northeastern U.S.
IBE	235000	U.S.A.	University	0	1962	Worldwide, emphasis on Alaska, Mississippi, and adjacent states; tropical America (especially Amazon region)
QCNE	250000	Ecuador	Quito	12	1979	Ecuador; some other countries
INPA	287000	Brazil	Manaus	10	1954	Amazonia
LPB	300000	Bolivia	La Paz	9	1984	Bolivia.
TCD	300000	Ireland	Dublin	3	1835	Worldwide, especially Europe, India, South Africa, Australia, North America, and Thailand
BKL	330000	U.S.A.	Bronx	0	1910	Long Island; North America
AK	350000	New Zealand	Auckland	6	1870	Worldwide, especially Pacific, Australia, New Zealand, northern New Zealand and its offshore islands
BREM	400000	Germany	Bremen	3	1865	Worldwide, especially northwestern Germany including North Sea Islands, western Mediterranean, and high mountains of Costa Rica
MBM	412000	Brazil	Curitiba	4	1965	Worldwide with emphasis on New World
CORD	500000	Argentina	Cordoba	29	1870	Central Argentina
OXF	500000	U.K.	Oxford	7	1621	Worldwide, especially British Isles, arctic, and South America.
COL	600000	Colombia	Bogota	17	1931	Worldwide, especially Colombia and adjacent countries
UBC	713000	Canada	Vancouver	11	1912	Worldwide; Pacific Rim countries; British Columbia, Washington, and Alaska
LIL	720000	Argentina	San Miguel de Tucuman	15	1931	Worldwide, especially Argentina and neighboring countries

SI	749999	Argentina	San Isidro	37	1911	Worldwide, Argentina and neighboring countries, North and South America, and Europe
GOET	750000	Germany	Gottingen	4	1832	Worldwide, with emphasis on Europe, southwestern Asia, South America
RB	800000	Brazil	Rio de Janeiro	49	1890	Worldwide, especially Brazil
DUKE	800000	U.S.A.	Durham	15	1932	Southeastern U.S., Mexico, Central America, and West Indies; worldwide
BH	845000	U.S.A.	Ithaca	13	1935	U.S. and tropics; New York; northeastern U.S.; Japan; worldwide
U	900000	Netherlands	Leiden	1	1816	Central and South America, especially the Guianas
EA	1000000	Kenya	Nairobi	13	1902	Mainly eastern Africa (Kenya, Tanzania, Uganda, Somalia) plus other African countries
WAG	1000000	Netherlands	Leiden	0	1896	Worldwide, especially tropical Africa, Mediterranean, and Europe
TEX	1006000	U.S.A.	Austin	26	1900	Texas; southwestern U.S.; Latin America, especially Mexico and northern Central America; worldwide, with emphasis on Latin America
WIS	1078000	U.S.A.	Madison	12	1849	Wisconsin; central and southeastern U.S.; Ozarks; Mexico; tropical America; U.S.S.R.; worldwide
CGE	1100000	U.K.	Cambridge	1	1761	Worldwide
MA	1158116	Spain	Madrid	17	1755	Worldwide, with greatest strength in Western Mediterranean, Central and South America, Africa, Australia and New Zealand
GB	1160000	Sweden	Goteborg	18	1926	Worldwide, especially South America, Hawaii, Middle East, Mediterranean, and Scandinavia
PH	1430000	U.S.A.	Philadelphia	10	1812	Asian-Malesian-Pacific and Australasian regions, especially peninsular Malaysia; southeastern Pennsylvania, southern New Jersey, northern Delaware, and northeastern Maryland
BRIT	1482000	U.S.A.	Fort Worth	22	1987	Worldwide, especially Texas and southeastern U.S.
Z	1500000	Switzerland	Zurich	15	1834	Worldwide; central Europe; southern Africa; New Caledonia
MEXU	1600000	Mexico	Mexico City	41	1888	New World, mostly Mexico and Central America
HBG	1800000	Germany	Hamburg	7	1879	Worldwide, especially Europe, South America, and South Africa
CAL	2086650	India	Howrah	40	1795	Worldwide, especially India, southern and southeastern Asia
UC	2100000	U.S.A.	Berkeley	15	1872	Worldwide, with emphasis on California, western North America, Mexico, Andean South America, Pacific Basin, and eastern Asia
CAS	2300000	U.S.A.	San Francisco	22	1853	Worldwide, especially western North America, California, and northern Latin America, Europe, Madagascar, China, Philippines and Galapagos Islands
F	2700000	U.S.A.	Chicago	17	1893	Worldwide with emphasis on tropical and North America, especially Mexico, Guatemala, Costa Rica, Colombia, Venezuela, Ecuador, and Peru; Central America, Andean South America, and Australasia; north temperate, South America, and south temperate
C	2900000	Denmark	Copenhagen	26	1759	Worldwide, especially Denmark, Greenland, Iceland, and Faeroe Islands; some arctic, Mediterranean, especially Greece, tropical eastern and northeastern Africa, especially Ethiopia, southeastern Asia, especially Thailand, and Central and South America
E	3000000	U.K.	Edinburgh	47	1839	Southwestern and southeastern Asia, Arabia, Turkey, Bhutan, Brazil, Britain, China, Himalayas, Mediterranean, Chile, Argentina, and southern Africa
M	3200000	Germany	Munchen	9	1813	Worldwide
H	3350501	Finland	Helsinki	39	1751	Worldwide, especially Fennoscandia (incl. NW Russia) and other areas of boreal and temperate Europe, Asia and North America; Japan, China, Australasia, New Guinea, Brazil, Argentina, Chile, and many other tropical areas

MPU	3500000	France	Montpellier	2	1809	Worldwide, especially Mediterranean Basin, Africa and Americas
B	3800000	Germany	Berlin	27	1815	Worldwide, especially central Europe, Mediterranean area, southwestern Asia, Africa, Cuba, and South America
BR	4000000	Belgium	Meise	25	1870	Worldwide, especially Belgium and central Africa
S	4570000	Sweden	Stockholm	12	1739	Worldwide
L	5000000	Netherlands	Leiden	30	1829	Wordwide, especially tropical Asia, tropical Africa, Central and South America
GH	5005000	U.S.A.	Cambridge	8		Worldwide; North America including West Indies and Mexico; eastern and southeastern Asia and Malesia; Philippines
US	5100000	U.S.A.	Washington	34	1848	Worldwide with emphasis on neotropics, North America, Pacific Islands, Philippines, and Indian subcontinent
BM	5200000	U.K.	London	32	1753	Worldwide, especially of British Isles, Europe, Africa, North America, West Indies, and Himalaya
W	5500000	Austria	Wien	6	1807	Worldwide
P	6000000	France	Paris	21	1635	Worldwide, especially Africa (northern, western, and equatorial), southeastern Asia, France, Europe, French Guiana, Madagascar, Mascarene and subantarctic islands, New Caledonia, Horne Islands, and Vanuatu
LE	6000000	Russia	Saint Petersburg	35	1823	Worldwide; Russia, Asia, Europe, North America, Africa, antarctic, arctic, Indian Ocean
G	6000000	Switzerland	Geneve	30	1824	Worldwide, especially Mediterranean, Middle East, South America, Africa, Madacascar, and regional
MO	6850000	U.S.A.	Saint Louis	58	1859	Worldwide; Central America (especially Costa Rica, Nicaragua, and Panama), tropical South America (especially Bolivia, Colombia, Ecuador, Paraguay, Peru, and Venezuela), Africa, and Madagascar; North America, Neotropics
NY	7921000	U.S.A.	Bronx	36	1891	Worldwide with greatest strength in tropical America and North America
K	8125000	U.K.	Kew	69	1852	Worldwide, especially Africa, tropical Asia, and Australasia

Table 3: List of all herbaria that have digitised collections of Peruvian *Begonia* and *Solanum* in Begonia Resource Centre and Solanaceae Source*;* sorted first by Peruvian herbaria (in red text) by accession size smallest-largest, and then by the international herbaria.

2.2 Geographic Distribution of *Begonia* and *Solanum* Specimens

To assess the geographic distributions of specimens from Peruvian and international herbaria (objective a), maps of the distribution of *Begonia* and *Solanum* specimens in Peru were created and plotted in R (R Core Team, 2020). Species occurrences were mapped against a base map adapted from Moonlight and Reynel (2018), showing elevation, major rivers, main roads, and political divisions within Peru. The distribution of occurrences for each species of both genera was initially plotted on the base map. The distribution of all occurrences of *Begonia* and *Solanum* were plotted by the individual herbarium they are held at. Species checklists were produced from the region in Peru they were collected from, 19 regions from the *Begonia* dataset and 24 regions from the *Solanum* dataset.

Each genus dataset was additionally split into Peruvian and International herbaria and plotted on the base map; a total of 1,205 *Begonia* georeferenced specimens are held within Peruvian herbaria and 1,722 georeferenced specimens are held in herbaria outside Peru. For *Solanum*, 9008 georeferenced specimens are held within Peru and 11,199 georeferenced specimens are held outside Peru. Additional R packages required were “maptools” (Bivand and Lewin-Koh, 2021), “maps” (Brownrigg, Minka and Deckmyn, 2018), “raster” (Hijmans, 2021), “rgdal” (Bivand, Keitt and Rowlinson, 2021), “stringr” (Wickham, 2019) and “png” (Urbanek, 2013). All R scripts are included in supplementary scripts (Supplementary R scripts).

2.3 The Age and Distribution of Specimen Data

To evaluate differences in the age ranges of specimen data among all herbaria, and between Peruvian and international herbaria (aim and objective b), boxplots were created in R (R Core Team, 2020) using dated specimens. There were 3,132 *Begonia* specimens with a collection date, and 125 with no date. A total of 22,443 *Solanum* specimens were dated, 583 specimens had no date and a further 7 specimens did not have an herbarium assigned. Additional herbaria holding dated specimen data were cleaned, including CIP herbarium in Peru where the herbarium accession number was gathered online (Vargas *et al.*, 2016) and multiple herbaria that needed updating to their main collection (LL to TEX; A to GH). Scatterplots of all herbaria in Peru and all international herbaria were also created in R (R Core Team, 2020) for both genera using R packages ‘ggplot2’ (Wickham, 2016).

2.4 Temporal Geographic Distribution of Specimen Data

To assess the temporal distributions of occurrences in Peru between Peruvian and international herbaria (aim and objective a), the data sets were combined, grouped into Peruvian or international herbaria and then split by collection date into 50-year time periods. Six density maps were produced each for Peruvian and international herbaria across time periods dating from 1750 to 2050. These were produced in R (R Core Team, 2020) with a raster cell size at 0.5 degree resolution (ca.50km at the equator) and mapped onto the base map (Moonlight and Reynel, 2018). Additional R packages required were ‘raster’ (Hijmans, 2021), ‘png’(Urbanek, 2013) and ‘viridis’ (Garnier *et al.*, 2021).

2.5 Modelling the Uniqueness of Herbaria

In considering the value of local herbaria beyond accession and size alone, several factors were evaluated to quantify the uniqueness of the specimen data between Peruvian and international herbaria (aim and objective c). Indices were designed to summarise variables relevant for biodiversity studies, including indices that quantify: the number of species only found in certain herbaria, the number of specimens of poorly collected species, the number of duplicated specimens, the number of type specimens, the number of rare species and whether an herbarium holds geographically exclusive specimen data. Further factors could be explored for different disciplines, i.e. flowering time for phenological studies. To estimate the statistical differences between local and international herbaria, the uniqueness indices were used to perform a linear regression model.

2.5.1 Uniqueness Indices

Mean uniqueness indices were calculated for 53 herbaria in our dataset from the total of specimen data per herbarium (Supplementary Table 1 and 2). Evaluating for each herbarium, the number of *Begonia* or *Solanum* specimens, the taxonomic uniqueness, geographic uniqueness, quantity of duplicates, quantity of type specimens, herbarium geographic uniqueness and the rarity of specimens. For each herbaria containing *Begonia* and *Solanum* in Peru, the following data was either retrieved from Index Herbariorum (Thiers, 2016) or calculated from our dataset and then presented as the mean of all specimens for each indices:

From *Indexherbariorum*:

i) *the total number of accessions per herbarium*
ii) *total number of staff per herbarium*
iii) *the date the herbarium commenced*

Uniqueness indices per herbarium in Peru:

iv) *number of Begonia specimens, number of Solanum specimens*
v) *taxonomic uniqueness of each species*
vi) *geographical uniqueness of each specimen*

vii) *duplicate uniqueness*

viii) *number of type specimens*

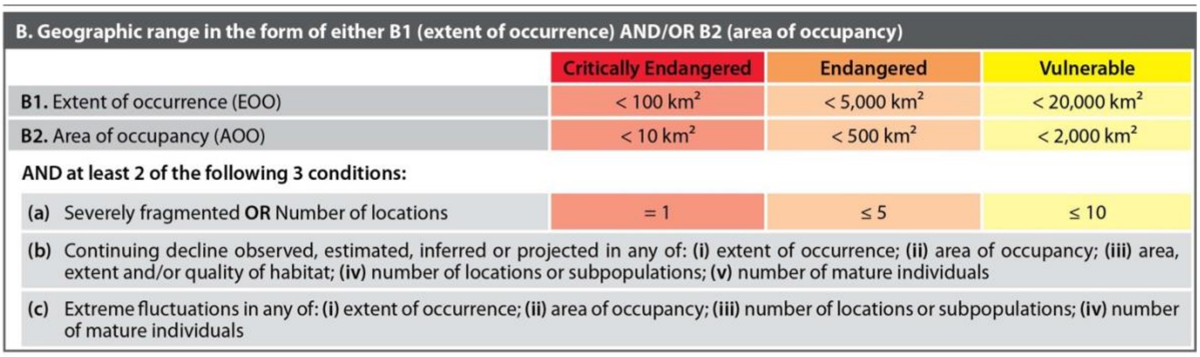
ix) *herbarium geographical uniqueness*

x) *rarity*

To clarify the calculations made: v) simply calculates the number of total specimens (including duplicates) of *Begonia* and *Solanum* collected in Peru at each herbarium; vi) calculates the number of specimens of each species (including duplicates), for each herbarium a high mean of the total of specimens per species indicates that the herbarium has a greater abundance of taxonomically unique species/specimens; vii) calculates the distance to the nearest 25 specimens (km), the greater the distance the more geographically unique the specimen is; viii) calculates if a specimen has unique coordinates to suggest whether it is duplicated, unduplicated or a repeated collection point, individual unduplicated specimens have a numeric value of 1; ix) calculates the total amount of types; x) calculates the distance to the nearest 25 specimens (km) excluding specimens from the same herbaria, in order to compare geographic uniqueness between herbaria; xi) calculates the total EOO of all specimens in each herbarium, where a lower EOO implies a greater total of range restricted species. The means of all specimens' indices were calculated by herbaria in preparation putting into a model.

2.5.2 Multiple Linear Regression

The linear regression model was performed in R (R Core Team, 2020) using multiple uniqueness indices to test whether local herbaria held more unique collections than from international herbaria. The analysis was run separately for each genus and asked whether each herbaria uniqueness variable can be explained by predictor variables - herbaria accession number, whether it is held in herbaria in Peru or the number of *Begonia* or *Solanum* specimens in each herbarium. The fit of the model was visually checked by plotting the residuals which were normally distributed.



B. Geographic range in the form of either B1 (extent of occurrence) AND/OR B2 (area of occupancy)

	Critically Endangered	Endangered	Vulnerable
B1. Extent of occurrence (EOO)	$< 100\ \text{km}^2$	$< 5{,}000\ \text{km}^2$	$< 20{,}000\ \text{km}^2$
B2. Area of occupancy (AOO)	$< 10\ \text{km}^2$	$< 500\ \text{km}^2$	$< 2{,}000\ \text{km}^2$
AND at least 2 of the following 3 conditions:			
(a) Severely fragmented **OR** Number of locations	= 1	≤ 5	≤ 10
(b) Continuing decline observed, estimated, inferred or projected in any of: **(i)** extent of occurrence; **(ii)** area of occupancy; **(iii)** area, extent and/or quality of habitat; **(iv)** number of locations or subpopulations; **(v)** number of mature individuals			
(c) Extreme fluctuations in any of: **(i)** extent of occurrence; **(ii)** area of occupancy; **(iii)** number of locations or subpopulations; **(iv)** number of mature individuals			

Fig 2: Criteria B from Red List guidelines (IUCN, 2019); parameters for calculating extinction risk using EOO and AOO

2.6 Herbaria and the IUCN Status of *Begonia* and *Solanum* in Peru

In order to quantify the contribution of local herbaria to our knowledge of biodiversity patterns (aim and objective d), we measured how data from local herbaria contributes to IUCN Red List threat assessments (IUCN, 2019). Based on criterion B (Fig. 2), the most widely used predictor which estimates the extinction risk status of a given species when unique georeferenced occurrences are available (Gaston and Fuller, 2009; Dauby *et al.*, 2017), the Extent of Occurrence (EOO) and the Area of Occupancy (AOO) was calculated for each species of *Begonia* and *Solanum* in Peru. This was performed using R package ‘ConR’ (Dauby *et al.*, 2017) that calculates the geographic range parameters EOO and AOO (Fig 2). IUCN Red List thresholds organise species into critically endangered, endangered, vulnerable, least concern and data deficient (IUCN, 2019). Out of 76 *Begonia* species in our dataset, it was possible to calculate the EOO for 51 species and AOO for all species. Out of a total of 282 *Solanum* species in our dataset, the EOO was calculated for 249 species and AOO for all species.

To examine changes in IUCN Red List assessments between herbaria, EOO and AOO was calculated for each study species removing the smallest accessioned herbaria first. Peruvian herbaria occur as the majority if not all of smallest herbaria in each dataset. The percentage of specimens removed at each step was also calculated to evaluate the impact of specimen abundance on EOO or AOO calculation.

IUCN Red list assessments were performed by assigning threat status with the EOO and AOO calculations of each species after the removal of herbaria. For each threat category, the EOO or AOO value were plotted removing each herbarium in the order of smallest to largest. Mean and median EOO and AOO values were calculated for species falling under each threat category. Mean and median values were log transformed in order to distribute data evenly for comparison between each result. All data handling was performed and plotted in R (R Core Team, 2020) using R package 'ggplot2' (Wickham, 2016); additional packages used were 'tidyverse' (Wickham *et al.*, 2019), 'ggrepel' (Slowikowski, 2021), 'ggtext'(Wilke, 2020), 'plyr' (Wickham, 2011). 'dplyr' (Wickham *et al.*, 2021), 'stringr' (Wickham, 2019), 'viridis' (Garnier *et al.*, 2021) and 'showtext'(Qui, 2021).

3 Results

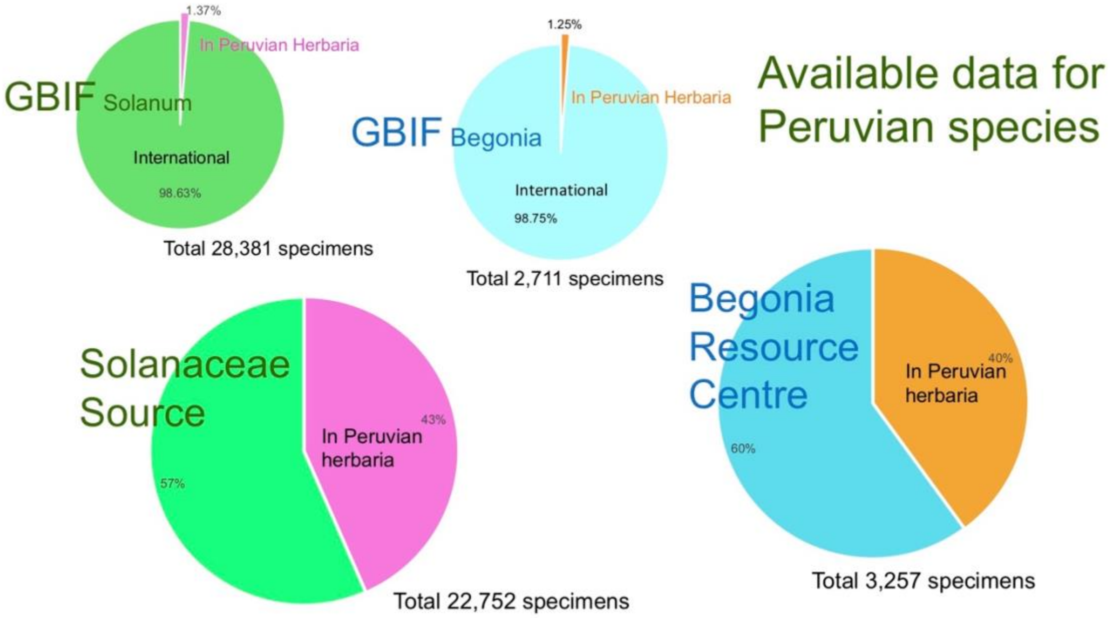


Fig 3: The proportion of *Begonia* and *Solanum* specimens from Peruvian herbaria downloaded and compiled from GBIF is 1%; in comparison to our dataset, Solanaceae Source and Begonia Resource Centre, that have between 40-43% of specimens from Peruvian herbaria

3.1 Differences between data in local and international herbaria

In a comparison between the proportion of Peruvian *Begonia* and *Solanum* specimen data held in local herbaria in Peru compared to international herbaria on GBIF, we found that only 1% are held in local herbaria, compared to our datasets that hold between 40-43% of specimen data in Peru (Fig 3).

3.2 Exploring the Temporal and Geographic Distribution of Data in Herbaria

3.2.1 The Age of Specimens in Local and International Herbaria

Our results show that there are clear differences in the age distributions of specimens held in individual herbaria for *Begonia* (Fig. 4a) and *Solanum (*Fig.4b), and also between the ages of the specimens held in local versus international herbaria (Fig. 4a/4b). Boxplots show that local herbaria have a recent age distribution of *Begonia* (Fig. 4a) and *Solanum* (Fig. 4b) specimens, whereas international herbaria have a greater spread of ages compared to local herbaria. Overall, specimens in local herbaria are younger than those in international herbaria (Fig. 4.a/4.b): the mean age of all local herbaria of *Begonia* is 1987, and 1970 international herbaria. In *Solanum*, the mean age of all local herbaria is 1984, and 1978 for international herbaria.

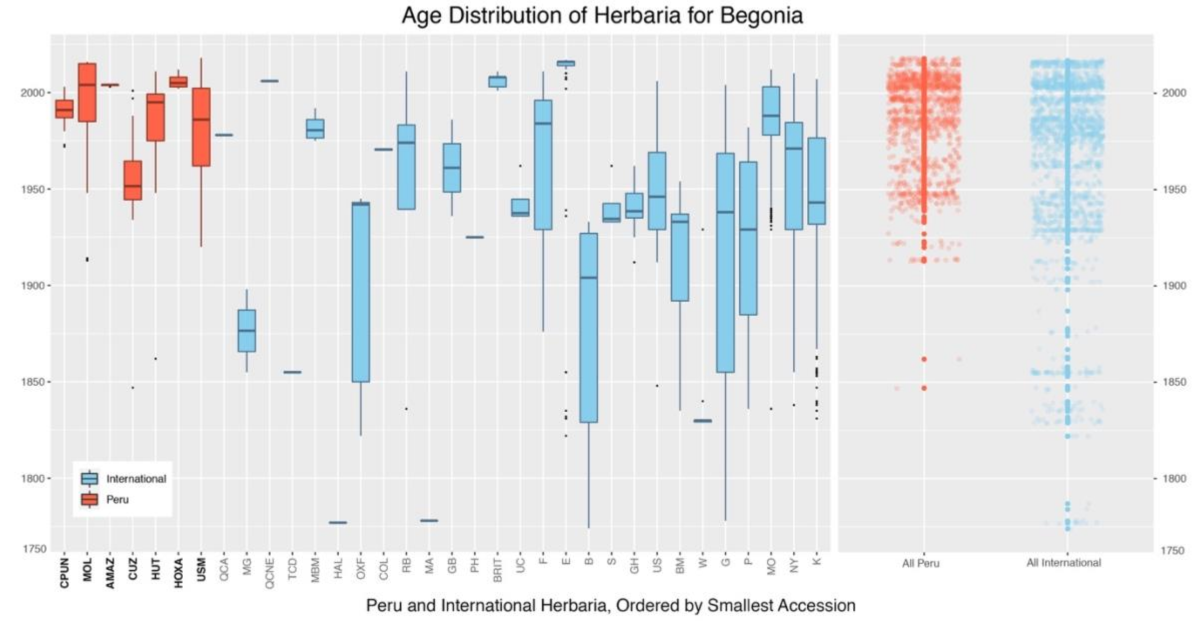


Fig 4a: Boxplots showing the age distribution of specimens of Peruvian *Begonia* in individual herbaria. Local Peruvian herbaria are shown in bold and red and are ordered by the total number of accessions they hold (smallest to largest, left to right). International herbaria are shown in blue and are ordered by the total number of accessions they hold (smallest to largest, left to right). Scatterplots on the right show the total age distribution of all specimens in local and international herbaria.

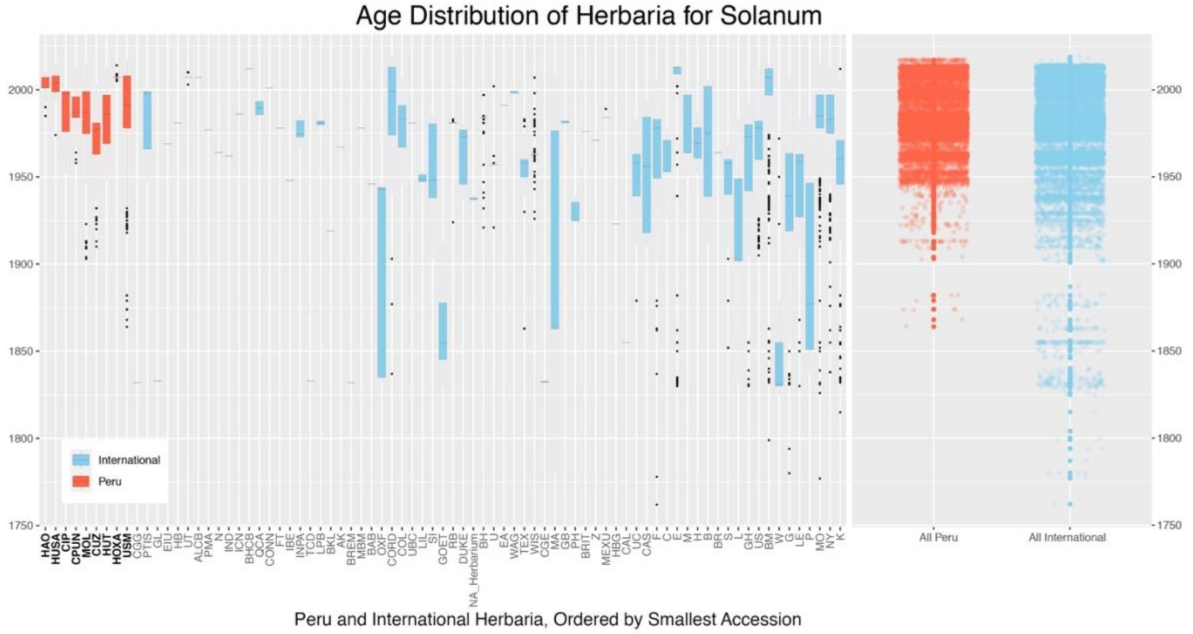
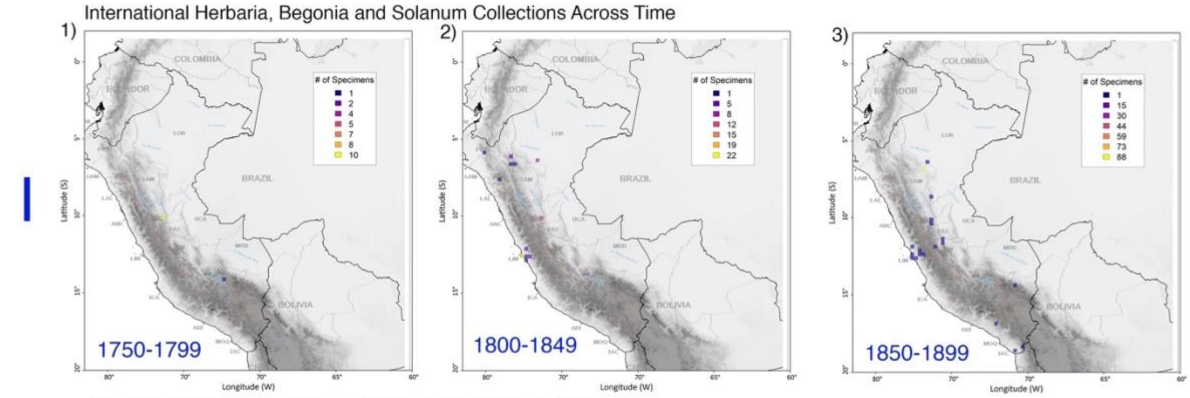


Fig 4b: Boxplots showing the age distribution specimens of Peruvian *Solanum* in individual herbaria. Local Peruvian herbaria are shown in bold and red and are ordered by the total number of accessions they hold (smallest to largest, left to right). International herbaria are shown in blue and are ordered by the total number of accessions they hold (smallest to largest, left to right). Scatterplots on the right show the total age distribution of all specimens in local and international herbaria.

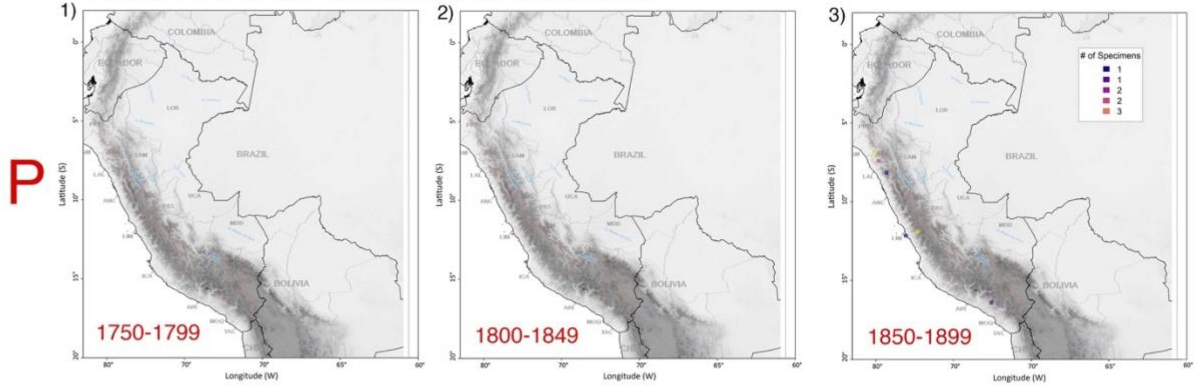
3.2.2 The Distribution of Collections in Local and International Herbaria across Space and Time

Our results demonstrate clear differences in the geographic distribution of herbaria across time between local (Fig. 5, P:1-3, P:4-6) and international herbaria (Fig. 5, I:1-3, I:4-6). The geographic distribution and abundance of occurrences in the earliest time periods 1750-1899 for international herbaria is very limited (Fig. 5, I:1-3) but begins to increase from 1850-1899 with 147 specimens (Fig. 5, I:3). Collections in international herbaria in the earliest time periods are associated major roads (Fig. 5, I:1-3). Local herbaria hold no collections from before 1849 (Fig. 5, P:1-2) and only 11 specimens collected between 1850-1899 (Fig. 5, P:3). Specimens in local herbaria increase both in number and geographic distribution from 1900 (Fig. 5, P:4). From 1950-1999 (Fig. 5, P:5) local herbaria have a broader distribution of occurrences than international herbaria with strong geographic clustering between herbaria. From 1950 onwards (Fig. 5, P:5-6), the number of specimens in local herbaria is similar to international herbaria.

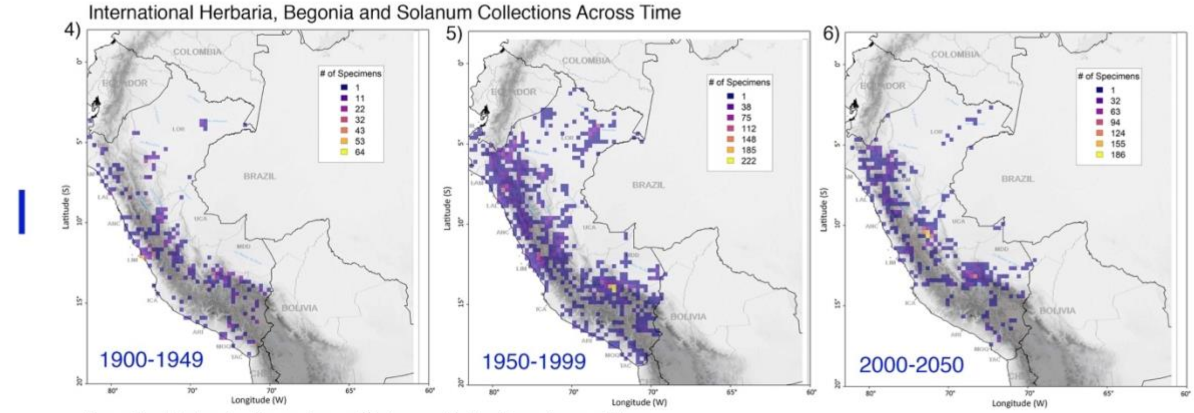
International Herbaria, Begonia and Solanum Collections Across Time



Peruvian Herbaria, Begonia and Solanum Collections Across Time



International Herbaria, Begonia and Solanum Collections Across Time



Peruvian Herbaria, Begonia and Solanum Collections Across Time

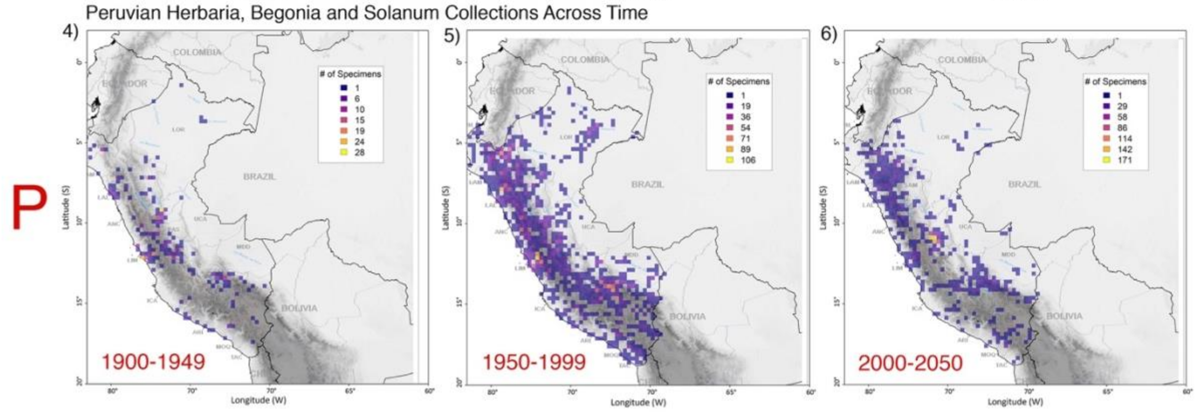


Fig. 5: Comparing the spatial and geographic distribution of local Peruvian and International herbaria of *Begonia* and *Solanum* collections from 1970 to today. Total georeferenced specimens per time period are: for International herbaria (I:1-3) from 1750=11, 1800=67, 1850=147; (I:4-6) from 1900=1261, 1950=7742, 2000=3513; for local Peruvian herbaria (P:1-3) no specimens are available for 1750 and 1800, 1850=11, (P:4-6) from 1900=501, 1950=6736, 2000=2843.

3.3 Measuring the Uniqueness of Herbarium Specimen Data

3.3.1 The Size and Observed Uniqueness of Herbarium Collections

We observe that local herbaria and the size of the accession has a great impact on the number of Peruvian specimens based on Supplementary Table 1 and 2. Local herbaria hold large numbers of *Begonia* and *Solanum* specimens, for example MOL in Lima has 135 *Begonia* specimens and 818 *Solanum* specimens, which is considerable considering their total herbarium accession is 38,000. A greater accession size can also influence the number of specimens in local herbaria, evident in the only large herbarium in Peru, USM, that has an herbarium accession of 800,000, holding the greatest national collections with 727 *Begonia* and 5,498 *Solanum* specimens. Large international herbaria additionally hold considerable number of specimens, most notably MO in Missouri USA that has an herbarium accession of 6,850,000 and holds 918 *Begonia* and 3592 *Solanum* specimens respectively. As an exception local herbaria AMAZ in Peru, has only four *Begonia* and no *Solanum* specimens, we suspect this is due to its regional geographic speciality that solely focus' on Amazonian Peru from which these genera are geographically scarce (Table 3). International herbaria with a large accession can also have a scarcity of *Begonia* and *Solanum* specimens, with 15 large international herbaria holding <10 specimens.

The taxonomic uniqueness of an herbarium's collections is not necessarily greater in local or international herbaria or for herbaria with a greater accession size (Supplementary Table 1 and 2). In fact, we observe that the larger the number of specimens an herbarium has, the average taxonomic uniqueness decreases. This implies that a greater amount of well collected species are held at herbaria with a larger number of specimens, to the effect of reducing the mean uniqueness. For example, the large international herbarium MO, despite having 918 *Begonia* specimens only have a taxonomic uniqueness of 17.09, in contrast to QCNE in Ecuador that holds only 1 Peruvian *Begonia* specimen, but it has a taxonomic uniqueness of 142.86. A similar pattern is observed for *Solanum* where large international herbaria MO, whilst has the second largest number of specimens (3592), only has a taxonomic uniqueness of 212.66. In comparison to large international herbaria, BRIT in USA, with 1 *Solanum* specimen but a taxonomic uniqueness of 115. Large Peruvian herbaria USM with the greatest *Solanum* specimens at 5298, only has a taxonomic uniqueness of 241.15.

The observed geographic uniqueness indices do not show a distinct pattern according to the herbarium accession size, local or international herbaria, or number of specimens (Supplementary Table 1 and 2). A local and small herbarium can have a similar geographic uniqueness compared to a local and small herbarium with a greater amount of specimens, for example, CPUN in Cajamarca Peru has 52 *Begonia* specimens and a geographic uniqueness of 13.86 km whereas HUT has 140 *Begonia* specimens and geographic uniqueness of 15.81 km. Distinctly, large international herbarium BRIT in Texas USA, has only 22 *Begonia* specimens but the largest geographic uniqueness of 27.66 km. PH herbarium in USA with only 6 *Solanum* specimens has the largest geographic uniqueness of 8.55 km.

Herbarium geographic uniqueness follows a similar pattern to geographic uniqueness for values between herbaria, however values are all incrementally higher for herbarium geographic uniqueness (Supplementary Table 1 and 2). This pattern reflects an overall increased uniqueness of the specimens at each herbarium. Whilst no distinct patterns were observed between local and international herbaria, by herbarium accession size or by the number of specimens, the most notable increases were between *Begonia* geographic uniqueness and herbarium geographic uniqueness. For example, the local herbarium HUT, and large local herbarium USM, demonstrate great increases compared to other herbaria: from 15.81 km to 21.18 km, and from 15.32 km to 20.89 km despite holding 140 and 727 *Begonia* specimens, respectively. Similarly, herbarium geographic uniqueness for *Begonia* is greater for international large herbaria MO, 12.72 km to 19.65 km, and BRIT, 27.66 km to 41.3 km, the former herbaria holds 918 *Begonia* specimens and the latter 22. Less distinct differences between the two herbarium indices were observed in *Solanum*, except for large local herbaria USM and large international herbaria MO that had the greatest increase in herbarium geographic uniqueness.

There are no observed robust patterns that show there is a greater duplicate uniqueness between local and international herbaria, whether there is a greater number of specimens in an herbarium or greater herbarium accession (Supplementary Table 1 and 2). However there seem to be an overall higher rate of unduplicated *Begonia* specimens in local herbaria in Peru, in comparison to the majority of international herbaria. With the exception that two large international herbaria hold the greatest unduplicated *Begonia* specimens, E in Edinburgh Scotland have 0.87 and BRIT have 0.72. Local herbaria in Peru additionally have an overall high number of unduplicated records in comparison to the majority of international herbaria, for example, CUZ in Cusco Peru valuing 0.47 and holding a great number, 2714, *Solanum* specimens. Observed duplicate uniqueness does not show a clear pattern with the amount of *Begonia* or *Solanum* specimens in an herbarium. For example, E holds 194 *Begonia* specimens whereas BRIT only holds 22. The highest unduplicated *Solanum* specimens at 0.5 are in large international herbarium P, in Paris France, which only has 22 specimens.

The observed type uniqueness shows that International and herbaria with a large accession have a greater number of types compared to local herbaria (Supplementary Table 1 and 2).. A great number of types however is not dependant on a great number of *Begonia* or *Solanum* specimens in an herbarium. This pattern is very clear for *Begonia* type uniqueness where values are significantly greater in international herbaria, despites holding few *Begonia* or *Solanum* specimens. For example, S in Stockholm Sweden is a large herbarium accession of 4,570,000, holds only 4 Peruvian *Begonia* specimens, but have mean type uniqueness of 100, the same patterns observed for UC California in USA and PH in Pennsylvania USA that have an herbarium accession over 2 million but hold less than 4 *Begonia* specimens and 100 type uniqueness. The large European International herbarium B in Berlin Germany, and W Austria both have an herbarium accession over 3.5 million, hold <32 specimens but have a great herbarium uniqueness of 90. The large local herbaria in Peru, USM, whilst has 727 *Begonia* specimens, has a type uniqueness of 3.71. A similar but weaker pattern is observed in type uniqueness in *Solanum* where four large international and European herbaria have a much greater type uniqueness than local herbaria. International herbaria MA in Madrid Spain, holds the greatest type uniqueness at 65.24 with 164 *Solanum* specimens, followed by large international European herbaria W, P, G in Geneva Switzerland. Local herbaria MOL has the greatest type uniqueness at 8.92 yet has *Solanum* 818 specimens.

The observed rarity uniqueness indices do not show clear patterns between Peru and international herbaria, nor between the herbarium accession size and number of specimens in an herbarium (Supplementary Table 1 and 2).. However, we observe that the number of specimens of *Begonia* and *Solanum*, has a great effect on the rarity uniqueness, where the calculations of rare species in an herbarium are influenced by common species with greater EOO. For example, the *Solanum* rarity uniqueness indices for local herbaria USM, whilst has the greatest number of specimens at 5498, the rarity indices are 333057, similar to large international herbarium OXF in Oxford UK, that holds only 13 specimens but has a rarity index of 329694.5. This suggests considering the great amount of *Solanum* specimens at USM that they hold many rare species. The herbarium containing the least rare species for *Begonia* is MBM in Parana Brazil has a rarity index of 602767.5, calculated from only 4 specimens.

3.3.2 The Mean Uniqueness of Specimens for Local or International Herbaria

A summary of uniqueness indices for all local and all international herbaria is shown in Table 4. Local herbaria have a greater mean uniqueness of *Begonia* and *Solanum* specimens than international herbaria, where the mean taxonomic uniqueness per local herbarium is 90 for *Begonia*, over double the mean value per international herbaria. The mean taxonomic uniqueness for *Solanum* per local herbaria is 1,387, significantly greater in comparison to the mean per international herbaria (323). The mean geographic uniqueness of *Begonia* per local herbarium is 26 greater in comparison to 15 per international herbaria. The mean geographic uniqueness of *Solanum* per local herbarium is 37, greater in comparison to 7 per international herbaria. The mean duplicate uniqueness of *Begonia* per local herbarium is 0.63, greater in comparison to 0.27 per international herbaria. The mean duplicate uniqueness of *Solanum* per local herbarium is 1.03, fivefold greater in comparison to 0.2 per international herbaria. The mean type uniqueness of *Begonia* per local herbarium is 91, over double when compared to per international herbaria at 37. The mean type uniqueness of *Solanum* per local herbarium is

105, over fivefold greater than per international herbaria at 20. The mean herbarium geographic uniqueness of *Begonia* per local herbarium is 31, greater in comparison to 18 per international herbaria. The mean herbarium geographic uniqueness of *Solanum* per local herbarium is 41, almost fivefold greater when compared to 8 per international herbaria. The mean rarity uniqueness of *Begonia* per local herbarium is 570117, greater in comparison to 273850 per international herbaria. The mean rarity uniqueness of *Solanum* per local herbarium is 1516903, almost fivefold greater when compared to 320181.

Uniqueness Indices:	*# of Begonia Specimens*	*Taxonomic*	*Geographic*	*Duplicate*	*Type*	*Herbarium Geographic*	*Rarity*
Mean Peruvian Herbaria	199	90	26	0.63	91	31	570117
Mean International Herbaria	113	37	15	0.27	38	18	273850

Table 4a: The mean uniqueness indices of *Begonia* specimens per Peruvian herbaria, and per International herbaria. Taxonomic, duplicate, and type uniqueness are unitless indices. Geographic and herbarium geographic uniqueness are measured in km, and rarity is measured in km2.

Uniqueness Indices:	*# of Solanum Specimens*	*Taxonomic*	*Geographic*	*Duplicate*	*Type*	*Herbarium Geographic*	*Rarity*
Mean Peruvian Herbaria	1418	1387	37	1.03	105	41	1516903
Mean International Herbaria	303	323	7	0.20	20	8	320181

Table 4b: The mean uniqueness indices of *Solanum* specimens per Peruvian herbaria, and per International herbaria. Taxonomic, duplicate, and type uniqueness are unitless indices. Geographic and herbarium geographic uniqueness are measured in km, and rarity is measured in km2.

3.3.3 Results from Modelling Uniqueness Indices by in Peru, by Herbarium Accession, by Genera Accession

Results from the regression model showed that specimens of *Begonia* and *Solanum* in local herbaria are not statistically more or less unique than those in international herbaria, based on our uniqueness indices, with no significant relationship between either herbarium accession and uniqueness, nor between the number of specimens in an herbarium and uniqueness (Table 5-11). However, a significant relationship was found between the number of Peruvian *Begonia* or *Solanum* specimens in herbaria and their accession size (Table 5a/5b).

The number of *Begonia* and *Solanum* specimens in Peruvian herbaria is statistically greater in local herbaria (both p values $<2 \times 10^{16}$) (Table 5a/5b). This shows that a greater the number of *Begonia and Solanum* specimens are held within local herbaria compared to international herbaria; the mean increase measured at 6.22×10^1 for *Begonia* and $^{4.71} \times 10^2$ for *Solanum*. There is a statistically significant relationship between the number of specimens of both *Begonia* and *Solanum* as measured and herbarium accession size (both p values $<2 \times 10^{16}$), with a mean increase of 1.61×10^5 in Begonia and 9.06×10^5 for *Solanum.*

The taxonomic uniqueness of *Begonia* and *Solanum* specimens has no significant relationship with herbarium accession, in Peru, or by the amount of *Begonia* or *Solanum* specimens per herbarium (Table 6a/6b). *Begonia* taxonomic uniqueness decreases as the herbarium accession increases (by -2.07×10^{-7}), whether it is in Peru (by -8.48×10^{00}) or whether the herbarium holds a greater amount of *Begonia* specimens (by -1.52×10^{-2}) (Table 6a). *Solanum* taxonomic uniqueness increases as the herbarium accession increases (by $2.26 \times 10^{-}$

^{5}), whether it is in Peru (by 6.36 x 10^{1}) or whether the herbarium holds a greater amount of *Solanum* specimens (by 4.59 x 10^{-2}) (Table 6b).

The geographic uniqueness of *Begonia* and *Solanum* specimens has no significant relationship with herbarium accession size, in Peru, or by the amount of *Begonia* or *Solanum* specimens per herbarium (Table 7a/7b). *Begonia* geographic uniqueness decreases as the herbarium accession increases (by -1.83 x 10^{-7}), whether it is in Peru (by -1.61 x 10^{00}) but increases whether the herbarium holds a greater amount of *Begonia* specimens (by 3.25 x 10^{-3}) (Table 7a). *Solanum* geographic uniqueness decreases as the herbarium accession increases (by -3.14 x 10^{-7}), whether it is in Peru (by -2.4 x 10^{00}) or whether the herbarium holds a greater amount of *Solanum* specimens (by -2.34 x 10^{-4}) (Table 7b).

The duplicate uniqueness of *Begonia* and *Solanum* specimens has no significant relationship with herbarium accession, in Peru, or by the amount of *Begonia* or *Solanum* specimens per herbarium (Table 8a/8b). *Begonia* duplicate uniqueness increases as the herbarium accession increases (by 6.69 x 10^{-9}), whether it is in Peru (by 1.47 x 10^{-1}), and whether the herbarium holds a greater amount of *Begonia* specimens (by 1.37 x 10^{-4}) (Table 8a). *Solanum* duplicate uniqueness increases as the herbarium accession increases (by 1.49 x 10^{-9}), whether it is in Peru (by 9.33 x 10^{-2}) or whether the herbarium holds a greater amount of *Solanum* specimens (by 1.97 x 10^{-5}) (Table 8b).

The type uniqueness of *Begonia* and *Solanum* specimens has no significant relationship with herbarium accession, in Peru, or by the amount of *Begonia* or *Solanum* specimens per herbarium (Table 9a/9b). *Begonia* type uniqueness increases as the herbarium accession increases (by 4.11 x 10^{-7}) or whether the herbarium holds a greater amount of *Begonia* specimens (by $4.01x\ 10^{-2}$) (Table 9a). *Begonia* type uniqueness decreases whether it is in Peru (by $-2.66 \text{ x } 10\ ^{1}$) (Table 9a). *Solanum* type uniqueness decreases as the herbarium accession increases (by $-6.16 \text{ x } 10^{-7}$), whether it is in Peru (by $-1.42 \text{ x } 10^{1}$) or whether the herbarium holds a greater amount of *Solanum* specimens (by $-3.75 \text{ x } 10^{-3}$) (Table 9b).

The herbarium geographic uniqueness of *Begonia* and *Solanum* specimens has no significant relationship with herbarium accession, in Peru, or by the amount of *Begonia* or *Solanum* specimens per herbarium (Table 10a/10b). *Begonia* herbarium geographic uniqueness decreases as the herbarium accession increases (by $-1.82 \text{ x } 10\ ^{-7}$), whether it is in Peru (by -1.48 x $10\ ^{00}$) but increases whether the herbarium holds a greater amount of *Begonia* specimens (by $9.02 \text{ x } 10^{-3}$) (Table 10a). *Solanum* herbarium geographic uniqueness decreases as the herbarium accession increases (by $-3.08 \text{ x } 10^{-7}$) and whether it is in Peru (by -2.11 x 10^{00}) (Table 10b). *Solanum* herbarium geographic uniqueness increases if the herbarium holds a greater amount of *Solanum* specimens (by $4.06 \text{ x } 10^{-4}$) (Table 10b).

The rarity uniqueness of *Begonia* and *Solanum* specimens has no significant relationship with herbarium accession, in Peru, or by the amount of *Begonia* or *Solanum* specimens per herbarium (Table 11a/11b). *Begonia* rarity uniqueness increases as the herbarium accession increases (by $5.67 \text{ x } 10\ ^{-4}$), whether it is in Peru (by $1.32\ 10^{3}$) and whether the herbarium holds a greater amount of *Begonia* specimens (by $1.63 \text{ x } 10^{2}$) (Table 11a). *Solanum* rarity uniqueness increases as the herbarium accession increases (by $9.95 \text{ x } 10^{-3}$) and whether the herbarium holds a greater amount of *Solanum* specimens (by $4.89 \text{ x } 10^{00}$) (Table 11b). *Solanum* rarity uniqueness decreases whether it is in Peru (by $-2.35 \text{ x } 10^{4}$) (Table 11b).

Number of Begonia Specimens

Coefficients:	Estimate	Standard Error	T Value	P Value
(Intercept)	-1.17 x 10^{00}	3.11 x 10^{1}	-3.76	<0.001
Herbarium Accession	1.61 x 19^{5}	6.83 x 10^{7}	23.52	< 2 x 10 16
In Peru	6.22 x 10^{1}	4.08 x 10^{00}	15.26	< 2 x 10 16

Table 5a: Coefficients of a regression model for the number of *Begonia* specimens, calculated from independent variables herbarium accession number and whether the herbarium is local in Peru. The residuals for this model are as follows: min = -96.64; median = 0.94; max= 808.65. The residual standard error is 18.62 based upon 3806 degrees of freedom. Significant values *** are in bold.

Number of Solanum Specimens

Coefficients:	Estimate	Standard Error	T Value	P Value
(Intercept)	-6.08 x 10^{00}	1.9 x 10^{00}	-3.19	0.0014
Herbarium Accession	9.06 x 10^{5}	4.18 x 10^{6}	21.66	< 2 x 10 16
In Peru	4.71 x 10^{2}	1.49 x 10^{1}	18.91	< 2 x 10 16

Table 5b: Coefficients of a regression model for number of *Solanum* specimens, calculated from independent variables herbarium accession number and whether the herbarium is local in Peru. The residuals for this model are as follows: min = -504.6; median = 4.8; max= 4960.4. The residual standard error is 113.9 based upon 3806 degrees of freedom. Significant values *** are in bold.

Begonia Taxonomic Uniqueness

Coefficients:	Estimate	Standard Error	T Value	P Value
(Intercept)	3.35 x 10^{1}	7.73 x 10^{00}	4.32	<0.001
Herbarium Accession	-2.07 x 10^{-7}	2.15 x 10^{-6}	-0.96	0.92
In Peru	-8.48 x 10^{00}	1.39 x 10^{1}	-0.61	0.54
Begonia Accession	-1.52 x 10^{-2}	2.7 x 10^{-2}	-0.43	0.67

Table 6a: Coefficients of a regression model for *Begonia* taxonomic uniqueness, calculated from independent variables herbarium accession number, Begonia accession number, and whether the herbarium is local in Peru. The residuals for this model are as follows: min = 26.27; median = -5.92; max= 109.48. The residual standard error is 27.2 based upon 32 degrees of freedom.

Solanum Taxonomic Uniqueness

Coefficients:	Estimate	Standard Error	T Value	P Value
(Intercept)	2.28 x 10^{2}	3.84 x 10^{1}	5.93	7.87 x 10 8
Herbarium Accession	2.26 x 10^{-5}	1.48 x 10^{-5}	1.52	0.13
In Peru	6.36 x 10^{1}	1.09 x 10^{2}	0.58	0.56
Solanum Accession	4.59 x 10^{-2}	3.99 x 10^{-2}	1.14	0.25

Table 6b: Coefficients of a regression model for *Solanum* taxonomic uniqueness, calculated from independent variables herbarium accession number, Solanum accession number, and whether the herbarium is local in Peru. The residuals for this model are as follows: min = -355.66; median = -87.78; max= 829.95. The residual standard error is 251.7 based upon 78 degrees of freedom. Significant values are in bold.

Begonia Geographic Uniqueness

Coefficients:	Estimate	Standard Error	T Value	P Value
(Intercept)	1.34 x 10^{1}	2.66 x 10^{00}	5.03	2.16 x 10-5
Herbarium Accession	-1.83 x 10^{-7}	7.38 x 10^{7}	-0.25	0.81
In Peru	-1.61 x 10^{00}	4.67 x 10^{00}	-0.35	0.73
Begonia Accession	3.25 x 10^{-3}	9.04 x 10^{-3}	0.36	0.72

Table 7a: Coefficients of a regression model for *Begonia* geographic uniqueness, calculated from independent variables herbarium accession number, Begonia accession number, and whether the herbarium is local in Peru. The residuals for this model are as follows: min =12.8; median = 0.38; max = 30.22. The residual standard error is 9.031 based upon 30 degrees of freedom. Significant values are in bold.

Solanum Geographic Uniqueness

Coefficients:	Estimate	Standard Error	T Value	P Value
(Intercept)	8.22 x 10^{00}	1.87 x 10^{00}	4.4	3.8 x 10 -5
Herbarium Accession	-3.14 x 10^{-7}	6.86 x 19^{-7}	-0.46	0.65
In Peru	-2.4 x 10^{00}	4.96 x 10^{00}	-0.48	0.63
Solanum Accession	-2.34 x 10^{-4}	1.79 x 10^{-3}	-0.13	0.9

Table 7b: Coefficients of a regression model for *Solanum* geographic uniqueness, calculated from independent variables herbarium accession number, Solanum accession number, and whether the herbarium is local in Peru. The residuals for this model are as follows: min = -8.21; median = -2.38; max= 53.74. The residual standard error is 11.3 based upon 71 degrees of freedom. Significant values are in bold.

Begonia Duplicate Uniqueness

Coefficients:	Estimate	Standard Error	T Value	P Value
(Intercept)	1.73 x 10^{-1}	8.16 x 10^{-2}	2.12	0.04
Herbarium Accession	6.69 x 10^{-9}	2.27 x 10^{-8}	0.3	0.77
In Peru	1.47 x 10^{-1}	1.43 x 10^{-1}	1.03	0.31
Begonia Accession	1.37 x 10^{-4}	2.78 x 10^{-4}	0.5	0.63

Table 8a: Coefficients of a regression model for *Begonia* duplicate uniqueness, calculated from independent variables herbarium accession number, Begonia accession number, and whether the herbarium is local in Peru. The residuals for this model are as follows: min = -0.32; median = -0.18; max= 0.83. The residual standard error is 0.28 based upon 30 degrees of freedom. Significant values are in bold.

Solanum Duplicate Uniqueness

Coefficients:	Estimate	Standard Error	T Value	P Value
(Intercept)	1.84 x. 10^{-1}	4.84 x 10^{-2}	3.78	<0.001
Herbarium Accession	1.49 x 10^{-9}	1.77 x 10^{-8}	0.08	0.93
In Peru	9.33 x 10^{-2}	1.29 x 10^{-1}	0.73	0.47
Solanum Accession	1.97 x 10^{-5}	4.65 x 10^{-5}	0.42	0.67

Table 8b: Coefficients of a regression model for *Solanum* duplicate uniqueness, calculated from independent variables herbarium accession number, Solanum accession number, and whether the herbarium is local in Peru. The residuals for this model are as follows: min = -0.28; median = -0.13; max= 0.82. The residual standard error is 0.29 based upon 71 degrees of freedom. Significant values are in bold.

Begonia Type Uniqueness

Coefficients:	Estimate	Standard Error	T Value	P Value
(Intercept)	3.79 x 10^{1}	1.03 x 10^{1}	3.7	<0.001
Herbarium Accession	4.11 x 10^{-7}	2.85 x 10^{-6}	0.15	0.89
In Peru	-2.66 x 10^{1}	1.84 x 10^{1}	-1.44	0.16
Begonia Accession	4.01 x 10^{-2}	3.58 x 10^{-2}	-1.12	0.28

Table 9a: Coefficients of a regression model for *Begonia* type uniqueness, calculated from independent variables herbarium accession number, Begonia accession number, and whether the herbarium is local in Peru. The residuals for this model are as follows: min = -39.88; median = -9.25; max= 62.26. The residual standard error is 36.06 based upon 32 degrees of freedom. Significant values are in bold.

Solanum Type Uniqueness

Coefficients:	Estimate	Standard Error	T Value	P Value
(Intercept)	2.12 x 10^{1}	4.98 x 10^{00}	4.26	5.6 x 10 -5
Herbarium Accession	-6.16 x 10^{-7}	1.92 x 10^{-6}	-0.31	0.75
In Peru	-1.42 x 10^{1}	1.41 x 10^{1}	-1	0.32
Solanum Accession	-3.75 x 10^{-3}	5.17 x 10^{-3}	-0.73	0.47

Table 9b: Coefficients of a regression model for *Solanum* type uniqueness, calculated from independent variables herbarium accession number, Solanum accession number, and whether the herbarium is local in Peru. The residuals for this model are as follows: min = -21.19; median = -14.35; max= 80.98. The residual standard error is 32.57 based upon 78 degrees of freedom. Significant values are in bold.

Begonia Herbarium Geographic Uniqueness

Coefficients:	Estimate	Standard Error	T Value	P Value
(Intercept)	1.52 x 10^{1}	3.12 x 10^{00}	4.88	3.28 x 10 -5
Herbarium Accession	-1.82 x 10^{-7}	8.65 x 10^{-7}	-0.22	0.83
In Peru	-1.48 x 10^{00}	5.48 x 10^{00}	-0.27	0.79
Begonia Accession	9.02 x 10^{-3}	1.06 x 10^{-2}	0.85	0.4

Table 10a: Coefficients of a regression model for *Begonia* herbarium geographic uniqueness, calculated from independent variables herbarium accession number, Begonia accession number, and whether the herbarium is local in Peru. The residuals for this model are as follows: min = -14.65; median = -0.2; max= 30.73. The residual standard error is 10.6 based upon 30 degrees of freedom. Significant values are in bold.

Solanum Herbarium Geographic Uniqueness

Coefficients:	Estimate	Standard Error	T Value	P Value
(Intercept)	8.85 x 10^{00}	1.99 x 10^{00}	4.43	3.39 x 10 5
Herbarium Accession	-3.08 x 10^{-7}	7.33 x 10^{-7}	-0.42	0.64
In Peru	-2.11 x 10^{00}	5.3 x 10^{00}	-0.4	0.69
Solanum Accession	4.06 x 10^{-4}	1.92 x 10^{-3}	0.21	0.83

Table 10b: Coefficients of a regression model for *Solanum* herbarium geographic uniqueness, calculated from independent variables herbarium accession number, Solanum accession number, and whether the herbarium is local in Peru. The residuals for this model are as follows: min = -8.84; median = -2.49; max= 57.04. The residual standard error is 12.08 based upon 71 degrees of freedom. Significant values are in bold.

Begonia Rarity Uniqueness

Coefficients:	Estimate	Standard Error	T Value	P Value
(Intercept)	2.19 x 10^{5}	3.89 x 10^{4}	5.53	4.74 x 10 6
Herbarium Accession	5.67 x 10^{-4}	1.06 x 10^{-2}	0.05	0.96
In Peru	1.32 x 10^{3}	6.81 x 10^{4}	0.01	0.99
Begonia Accession	1.63 x 10^{2}	1.31 x 10^{2}	1.25	0.22

Table 11a: Coefficients of a regression model for *Begonia* rarity uniqueness, calculated from independent variables herbarium accession number, Begonia accession number, and whether the herbarium is local in Peru. The residuals for this model are as follows: min = -198,156; median = 3879; max= 387,084. The residual standard error is 131,600 based upon 31 degrees of freedom. Significant values are in bold.

Solanum Rarity Uniqueness

Coefficients:	Estimate	Standard Error	T Value	P Value
(Intercept)	2.77 x 10^{5}	2.99 x 10^{4}	9.24	3.79 x 10 -14
Herbarium Accession	9.95 x 10^{-3}	1.16 x 10^{-2}	0.77	0.44
In Peru	-2.35 x 10^{4}	8.52 x 10^{4}	-0.28	0.78
Begonia Accession	4.89 x 10^{00}	3.12 x 10^{1}	0.16	0.88

Table 11b: Coefficients of a regression model for *Solanum* rarity uniqueness, calculated from independent variables herbarium accession number, Solanum accession number, and whether the herbarium is local in Peru. The residuals for this model are as follows: min = -273,245; median = -34,391; max= 957,106. The residual standard error is 196,400 based upon 78 degrees of freedom. Significant values are in bold.

3.4 Contribution of Local Herbaria to Estimates of Range Size and Red List Threat Status

3.4.1 Contribution of Herbaria to Range Size Estimates

Our results demonstrate that all herbaria, including the smallest herbaria, contribute to accurate estimates of species range size rarity. The overall pattern from the results show that as you remove herbaria, in order of smallest to largest, the EOO and AOO of all species decrease in both genera (Fig. 7-10). When all the data are included, the median EOO is 1227 km^2, yet when 10 and 20 herbaria are removed (out of a total of 33 herbaria) the EOO drops to 663 km^2 and to 400 km^2, respectively. The mean EOO for all *Begonia* species is 75,120 km^2, and when 10 and 20 herbaria are removed it drops to 71,348 km^2 and to 59,361 km^2, respectively. The median AOO for all the dataset of *Begonia* is 20 km^2, and when 10 and 20 herbaria are removed it drops to 14 km^2 and 12 km^2. The mean AOO km^2 calculated from all herbaria for *Begonia* is 70 km^2, and when 10 and 20 herbaria are removed it drops to 62 km^2 and 49 km^2, respectively. Similarly, for calculations of EOO for *Solanum*, when all herbaria are included the median EOO is 54,576 km^2, and when 30 and 60 herbaria are removed (out of a total of 72 herbaria) the EOO drops to 43,695 km^2 and 20,587 km^2, respectively. The mean EOO km^2 of all *Solanum* herbaria is 161,064 km^2, but in removing 30 and 60 herbaria the EOO decreases to 156,837 km^2 and 119,353 km^2. The median AOO when including all the dataset of *Solanum* is 54 km^2, but in removing 30 and 60 herbaria is calculated at 48 km^2 and 32 km^2, respectively. The mean AOO of all herbaria for *Solanum* is 114 km^2, yet when 30 and 60 herbaria are removed the AOO is 99 km^2 and 68 km^2, respectively.

3.4.2 Contribution of Herbaria to Estimates of Range Size by Threat Category

Local herbaria (<100,000 herbarium accession) contribute to ca. 22% of *Begonia* specimens (Fig. 7b/9b) and contribute to ca. 22% *Solanum* specimens (Fig. 8b/10b). The large local herbarium USM in Peru (800,000 herbarium accession), contributes an additional 20-25% to total of specimens of *Begonia* and *Solanum*.

Species of each genus were assigned a Red List threat category from calculations of EOO and AOO using data from all herbaria. As herbaria are removed in order of smallest to largest, our median calculations of log EOO and AOO decrease (Fig. 6a/6b). The mean EOO and AOO of critically endangered species of both genera greatly decreases when relatively few herbaria have been removed. For critically endangered species of *Begonia,* when local herbaria are removed the Log EOO becomes data deficient (Fig. 6a). For critically endangered species of *Solanum*, when USM are removed the Log EOO becomes data deficient (Fig. 6b). The median EOO for all species of *Begonia* shows that the removal of specific herbaria greatly influences the EOO; these are when removing local herbaria, USM and international herbarium E (Fig. 6a). The same pattern is shown in the median EOO for all species of *Solanum,* where the removal of local herbaria, USM and international herbarium BM in London UK, greatly decrease the EOO.

3.4.2.1 Contribution of Herbaria to EOO Calculations of Threatened Species

The EOO of critically endangered *Begonia* spp. is greatly affected when specimen data from local herbaria are removed. For example, when data from local herbaria are removed (Fig.7c), the EOO of *B. yuracyacuensis* and *B. rodriguezii* can no longer be calculated because the number of geographically unique specimens falls below three. The EOO of endangered *Begonia* species shows a similar pattern (Fig. 7d): when data from local herbaria are removed the EOO drops significantly and even reassigns several species as critically endangered (*B. joshii*, *B. altoperuviana*, *B. bifurcata*). When USM is removed, the EOO of many species decreases (Fig. 7d) and the EOO of *B.chemillenensis* and *B. psuedopleiopetala* fall below the amount of geographically unique species required to calculate an EOO. For vulnerable *Begonia* species, the EOO decreases when local Peruvian herbaria are removed, for example the EOO of *B. hirta* drops by almost half from 10,154 km² to 5125 km² when the data from local CUZ is removed (Fig. 7e). However vulnerable species are only estimated as a higher threat category, endangered, when the largest 11 international herbaria are removed (Fig. 7e). The EOO for several least concern *Begonia* species (Fig. 7f) decreases when local herbaria are removed, and again when USM is removed. For example, the EOO of *B. guaduensis* decreases significantly from 59,368 km² to 12,457 km² when the local herbarium CUZ are removed. Least concern *Begonia* species are greatly affected by the removal of large international herbaria where many species increase in threat category status.

The EOO of critically endangered *Solanum* spp. is greatly affected when herbaria are removed (Fig. 8c). The removal of local herbaria results in no possible calculation of EOO for species *S. trinitense and S. gracilifrons.* When herbarium USM is removed, three further species cannot either be calculated (*S. juglandifolium*, *S. contumazaense)* and the EOO drops significantly for *S. inelegans* from 50 km² to 7 km² (Fig. 8c). Out of 36 endangered *Solanum* spp., when local herbaria (MOL, CUZ) are removed the EOO of 8 species can no longer be calculated (Fig. 8d). These and other endangered species are assigned a higher threat category, critically endangered, when herbaria continue to be removed. Vulnerable *Solanum* spp. are also greatly affected by removing local herbaria, with many species significantly decreasing in EOO (Fig. 8e). Three species either move up into the endangered threat category (*S. raquialatum)* or no calculation is possible (*S. savanillense* and *S. acroglossum*) when local herbaria are removed (CUZ, HOXA). When USM are removed the EOO km² of *S. urubambaense* decreases from 5186 km² to 112 km², moving this species into a higher threat category along with an additional four spp. The EOO of least concern *Solanum* spp., decreased for many species when both local herbaria are removed, and USM is removed (Fig. 8f). For example, *S. chacoense* decreased from 20,353 km² to 128 km² when local herbarium CUZ has been removed. Many least concern *Solanum* spp., either change into a higher threat category, vulnerable, or become data deficient when small, local and international herbaria are removed.

3.4.2.2 Contribution of Herbaria to AOO Calculations of Threatened Species

In calculating the AOO for *Begonia* spp. the majority of all species (Fig. 9a) and species in a threat category decrease as local herbaria are removed (Fig. 9c-e). When local herbaria are removed, the AOO of 5/28 critically endangered species (Fig. 9c) either decreases by half, from 8-4 km² (*B. brevicordata, B .thyrsoidea, B. yuracyacuensis)* or are rendered data deficient due to not enough data remaining to calculate the AOO (*B. granpajatensis, B. heracleifolia).* The AOO of endangered *Begonia* species decreases incrementally as local herbaria are removed. (Fig. 9d) *B. rodriguezii* decreases from 12-8 km² when local herbarium HUT is removed, placing this species in critically endangered category. Other substantial decreases in AOO occur when the data from USM are removed, at which point *B. joshii* becomes data deficient. Additionally, in the removal of large international herbaria E and MO. Whilst vulnerable *Begonia* species decrease incrementally as local herbaria are removed, it is with the removal of USM that both vulnerable species are reassigned to a higher threat category, endangered (Fig. 9e).

Similarly, the AOO for *Solanum* spp. decreases with the removal of herbaria, but notably in the removal of local herbaria, USM, and international herbaria E and BM (Fig. 10a). The AOO of two critically endangered species decrease significantly in the removal of local herbaria, *S. tergosericeum* becomes data deficient when MOL is removed and *S. salasianum* drops from 8 km² to 4 km² when CUZ is removed (Fig. 10c). The AOO of the majority if not all endangered *Solanum* species decreases with the removal of local herbaria (Fig. 10d). When local herbaria CUZ are removed, *S. trinitense, S. scabrifolium, S. simplicissimum, S. rhomboideilanceolatum, S. limbaniense* and *S. augustii* increase in threat category to critically endangered or become data deficient as with *S. ayacuchense.* A further species *S.gracilifrons* is falls into the critically endangered category when local herbaria MOL is removed. A large drop of AOO is observed for *S. multiinterruptum* and *S. medians* that drop from 276 km² to 184 km², and 408 km² to 324 km² when CUZ specimen data is removed (Fig. 10d). In the removal of USM great decreases are evident in all endangered species. While all vulnerable *Solanum* species AOO decreases with the removal of local herbaria, when CUZ are removed *S. acaule* drops from 796 km² to 332 km² and would be moved into a greater threat category, endangered (Fig. 10e). A total of 8/9 vulnerable species drop into a high threat category when USM is removed (Fig. 10e). Removing large international herbaria BM also shows sizeable decreases in AOO. The only least concern species *S. candolleanum*, is greatly affected when local herbaria CUZ is removed, its AOO decrease almost by half from 2324 km² to 1292 km², moving it into vulnerable threat category (Fig. 10f).

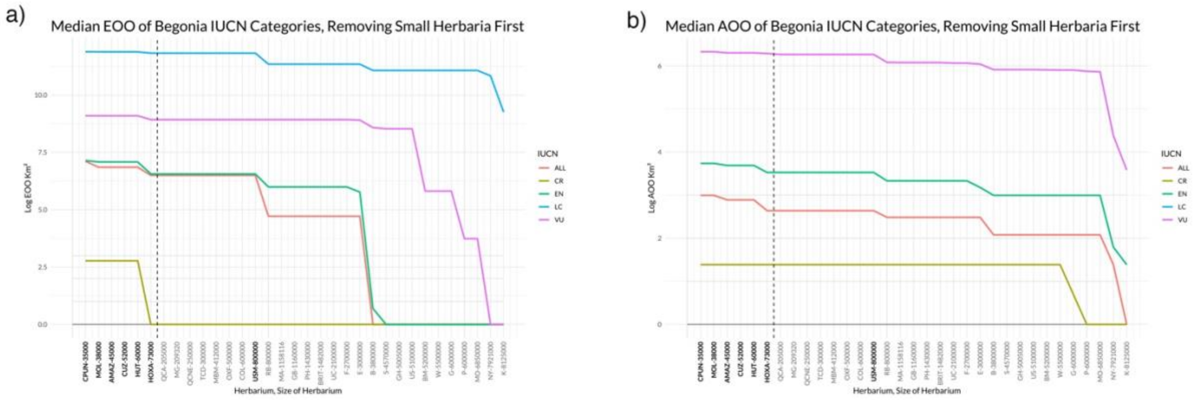


Fig. 6a: The median log EOO (a) and Log AOO (b) of *Begonia* species categorised in each Red List Threat category when removing herbaria in order of smallest - largest. Local herbaria are in bold; dashed vertical line marks the small herbaria boundary.

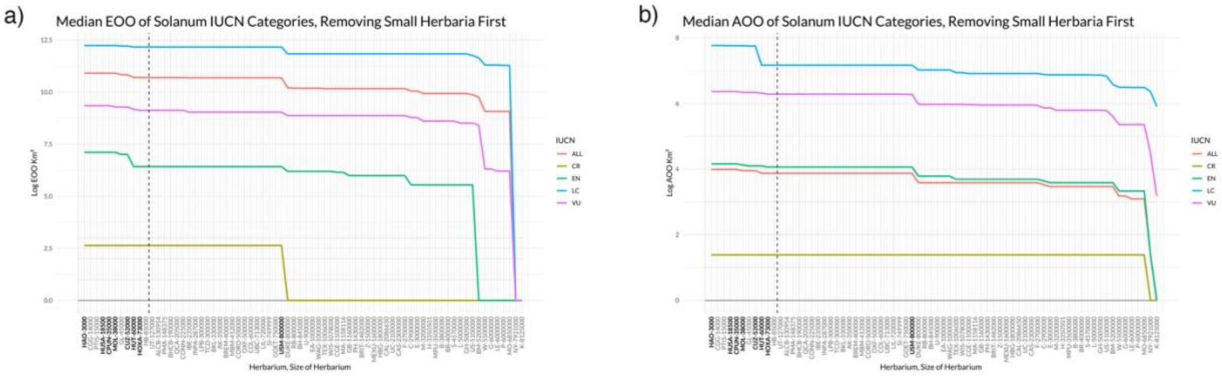


Fig. 6b: The median Log EOO (a) and Log AOO (b) of Solanum species categorised in each Red List Threat category when removing herbaria in order of smallest - largest. Local herbaria are in bold; dashed vertical line marks the small herbaria boundary.

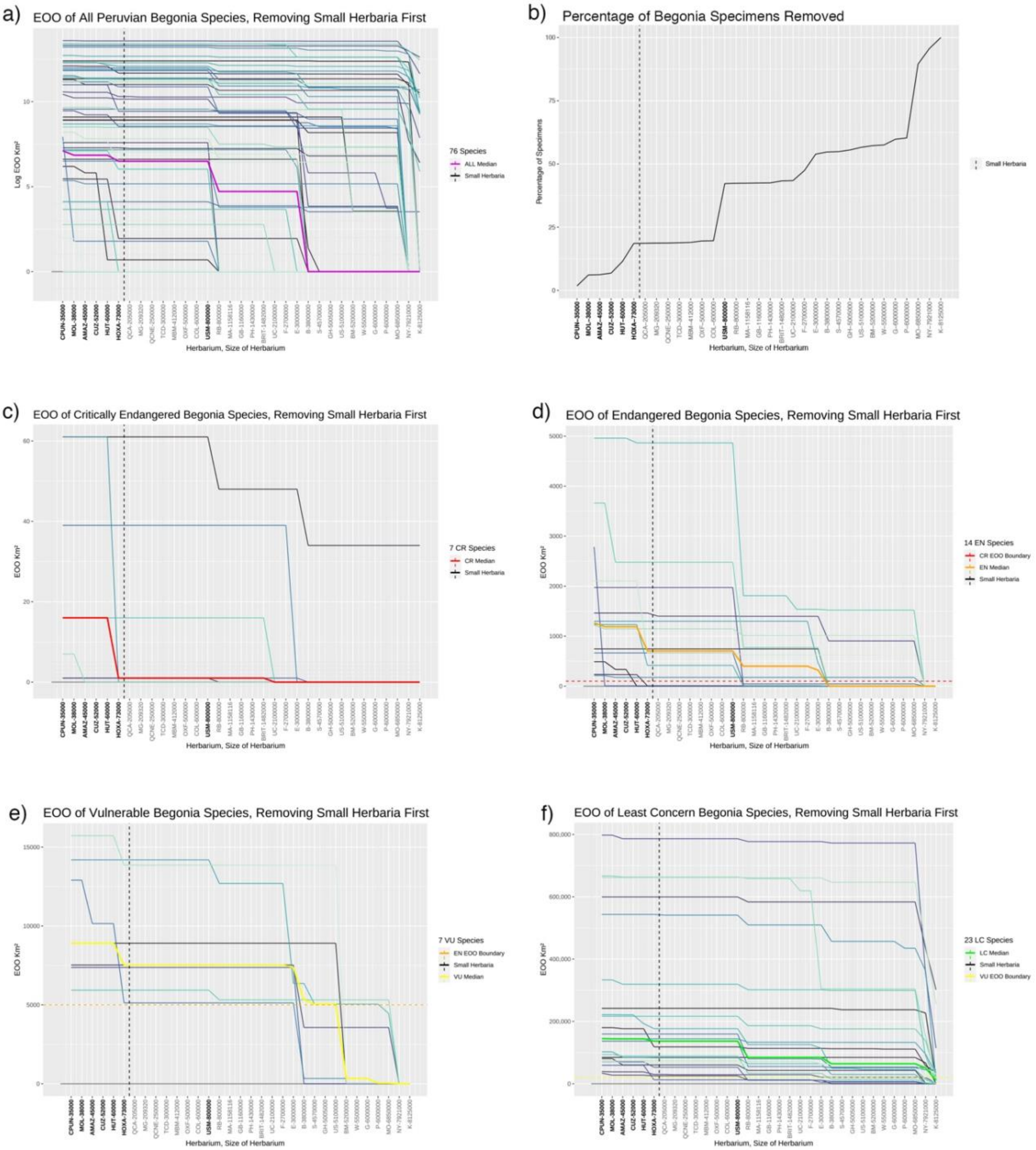


Fig 7: Measuring the changes in EOO km² for *Begonia* when removing herbaria in order of smallest - largest (Local herbaria are in bold); (a) Log EOO of all *Begonia* occurrences, (b) The percentage of *Begonia* specimens removed when removing herbaria smallest to largest, (c-d) The EOO km² of Critically Endangered (CR), Endangered (EN), Vulnerable (VU) and Least Concern (LC) *Begonia* species when removing herbaria smallest - largest.

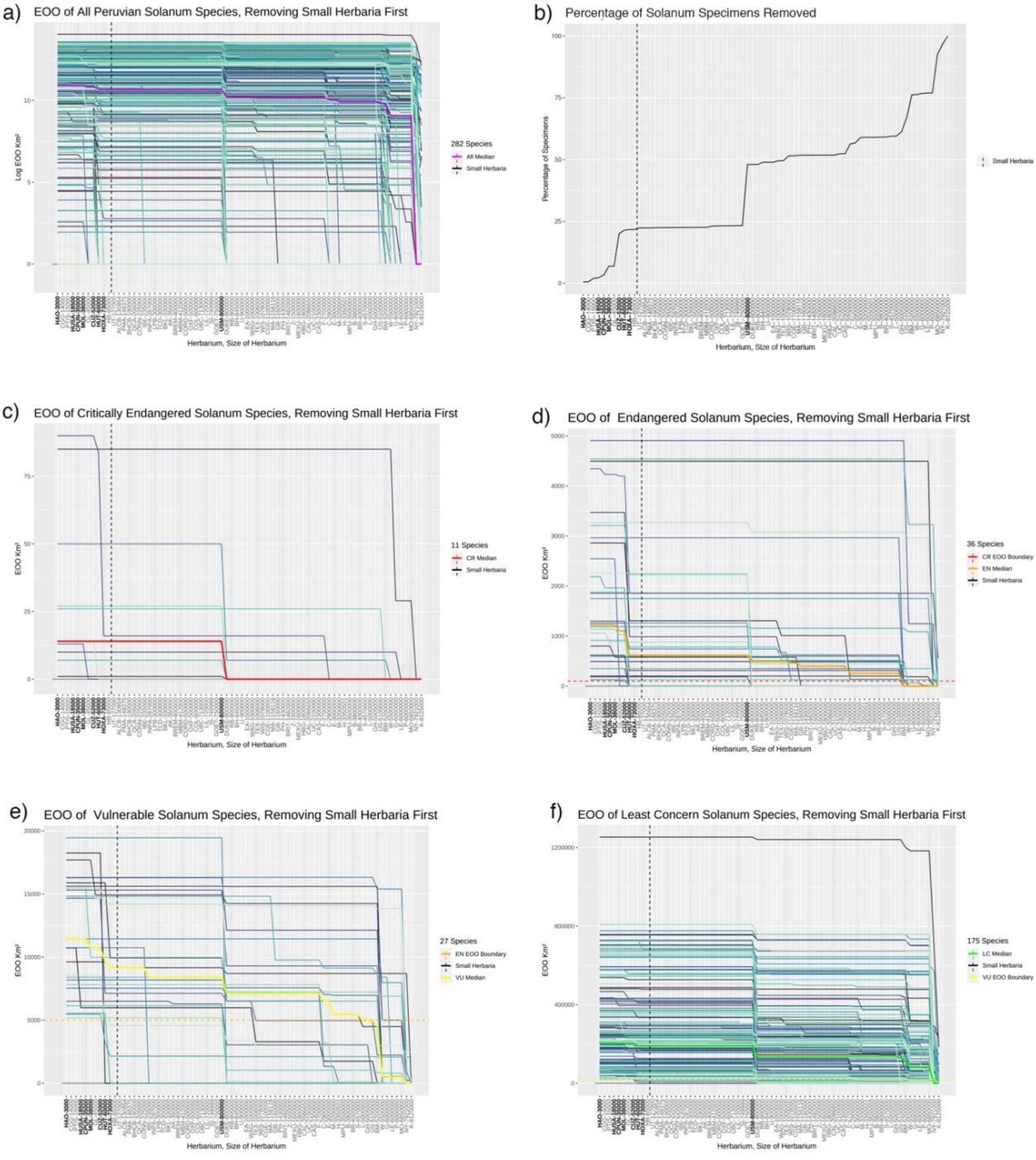
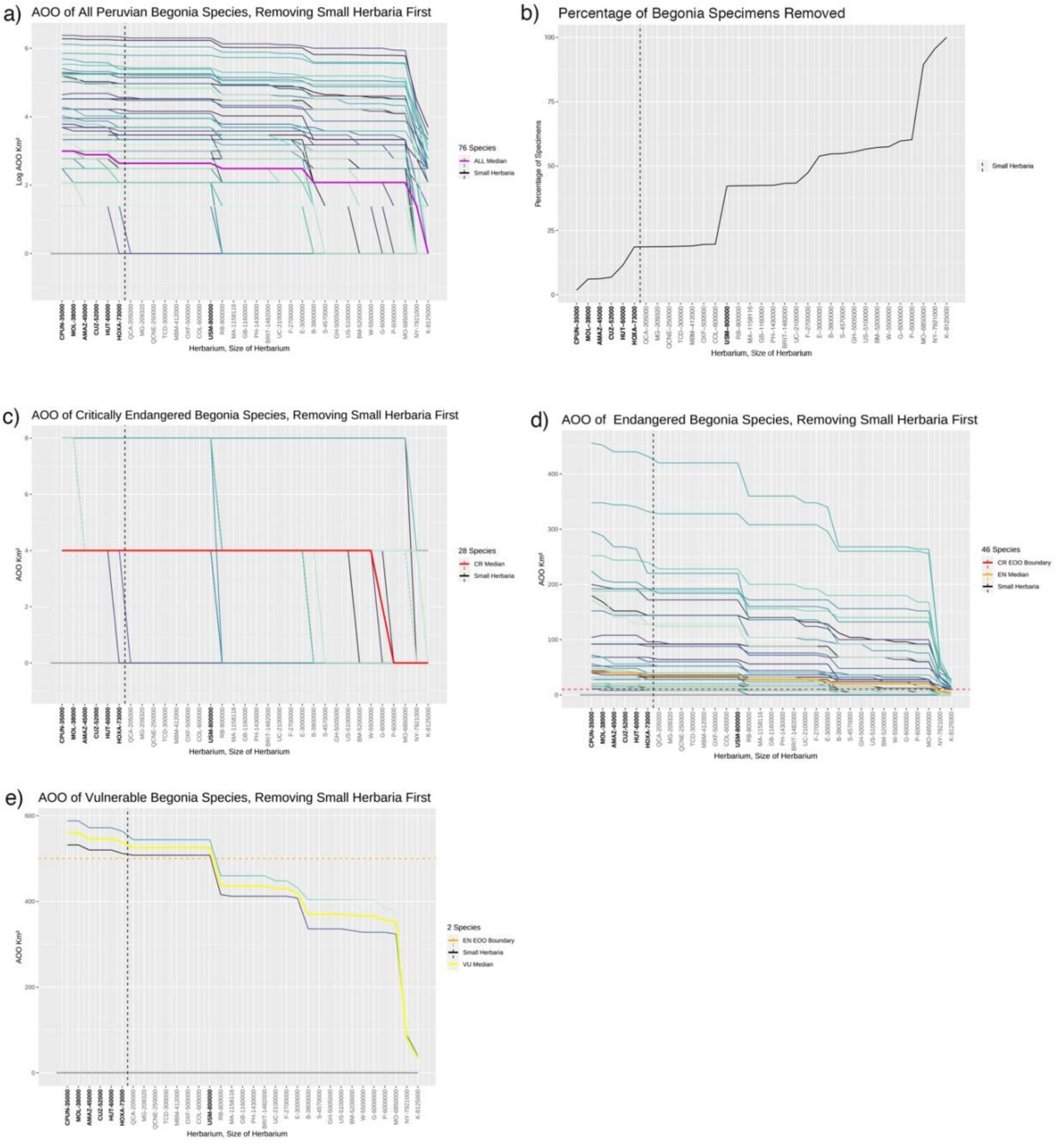


Fig 8: Measuring the changes in EOO km² for *Solanum* when removing herbaria in order of smallest - largest (Local herbaria are in bold); (a) Log EOO of all *Solanum* occurrences, (b) The percentage of *Solanum* specimens removed when removing herbaria smallest to largest, (c-d) The EOO km² of Critically Endangered (CR), Endangered (EN), Vulnerable (VU) and Least Concern (LC) *Solanum* species when removing herbaria smallest - largest.


a) AOO of All Peruvian Begonia Species, Removing Small Herbaria First

b) Percentage of Begonia Specimens Removed

c) AOO of Critically Endangered Begonia Species, Removing Small Herbaria First

d) AOO of Endangered Begonia Species, Removing Small Herbaria First

e) AOO of Vulnerable Begonia Species, Removing Small Herbaria First

Fig 9: Measuring the changes in AOO km² for *Begonia* when removing herbaria in order of smallest - largest (local herbaria are in bold); (a) Log AOO of all *Begonia* occurrences, (b) The percentage of *Begonia* specimens removed when removing herbaria smallest to largest, (c-d) The AOO km² of Critically Endangered (CR), Endangered (EN) and Vulnerable (VU) *Begonia* species when removing herbaria smallest - largest. (No species fit into Least Concern (LC) parameters).

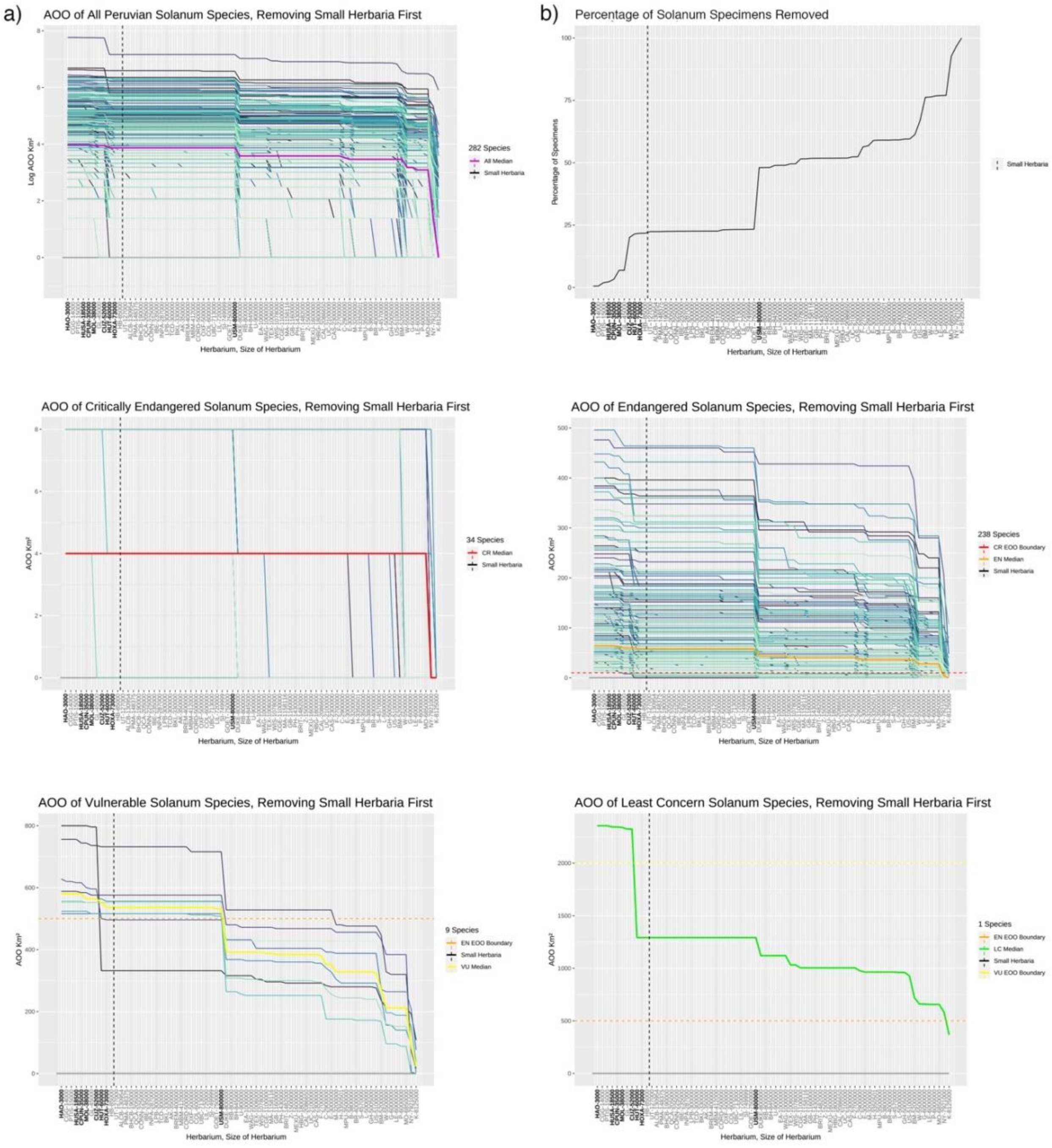


Fig 10: Measuring the changes in AOO km² for *Solanum* when removing herbaria in order of smallest - largest (local herbaria are in bold); (a) Log AOO of all *Solanum* occurrences, (b) The percentage of *Solanum* specimens removed when removing herbaria smallest to largest, (c-d) The AOO km² of Critically Endangered (CR), Endangered (EN) and Vulnerable (VU) *Solanum* species when removing herbaria smallest - largest.

3.4.3 Changes of Threat Status as Herbaria are Removed

The total number of species in each Red List threat category was calculated using EOO and AOO guidelines for the complete dataset and when removing herbaria at incrementally in order of smallest to largest herbarium accession (Tables 12-13). As the data from herbaria are removed in groups of five from EOO Red List assessments of *Begonia* species, there is an increase of data deficient (DD) species, as species either move threat category or not enough data is available to calculate a threat (Table 12a). The quantity of critically endangered and vulnerable *Begonia* species initially increases before decreasing as a greater number of herbaria are removed and species change threat category. The number of endangered species and least concern species gradually decrease as more herbaria are removed and species change threat category. As the data from herbaria are removed from AOO Red List assessments of *Begonia* species, the amount of critically endangered species increases, revealing that the decrease in specimen data increases the threat category (Table 12b). The number of endangered species decreases as herbaria are removed, and as species move into the critically endangered category. When the data from the 10 smallest herbaria are removed the only two vulnerable *Begonia* species move into a higher threat category.

The EOO Red List threat status of *Begonia* species was calculated removing the data from all Peruvian herbaria (Table 12c). When Peruvian herbaria are removed there is an increase in data deficient species from 25 to 32 in the complete dataset. This increase of data deficient species in reflected in the changes in assigned threat categories as Peruvian herbaria are removed, where only 44 species can be assigned a threat category when Peruvian herbaria are removed, down from 51 in a complete dataset (Table 12c). Removing only small herbaria from calculations of EOO threat assessments leaves 49 species with a threat category and 27 species data deficient however greater differences are exhibited when Peruvian data are removed (Table 12c). In summary there is an 11% increase in data deficient species when Peruvian herbaria are removed from calculations of EOO to assess Red List threat status (33% DD in complete dataset, 42% when Peru herbaria removed). When removing the data from all Peruvian herbaria from calculations of AOO Red List threat status of *Begonia* species, the number of endangered species increases to critically endangered, from 32 CR in the whole dataset to 34 critically endangered without the Peruvian herbarium data (Table 12d). Two vulnerable species also increase in threat status to endangered. Removing only small herbaria increases estimations of AOO threat status but not as significantly as all Peruvian herbaria (Table 12d).

As the data from herbaria are removed in groups of ten from EOO Red List assessments of *Solanum* species, there is an increase of data deficient (DD) species, as species either move threat category or not enough data is available to calculate a threat (Table 13a). The amount of critically endangered and least concern species decreases the more herbaria are removed, and species change threat category. Endangered and vulnerable species decrease as herbaria are removed, yet slightly increase at the point where 60 herbaria have been removed, before decreases again and a large increase in data deficient species is observed (Table 13a). As the data from herbaria are removed from AOO Red List assessments of *Solanum* species, the amount of critically endangered species increases, revealing that the decrease in specimen data greatly increases the threat category (Table 13b). This is evident in the reduction of endangered and vulnerable species as herbaria are removed and species move into the critically endangered category.

The EOO Red List threat status of *Solanum* species was calculated removing the data from all Peruvian herbaria (Table 13c). When Peruvian herbaria are removed there is an increase in data deficient species from 33 to 55 (Table 13c). The increase of data deficient species in reflected in the changes of species as they move threat. With the complete dataset 249 *Solanum* species have an assigned threat category, but when all Peruvian data are removed the total amount of species assigned a threat category decrease to 217. The removal of only small herbaria increases the amount of data deficient species and increases the threat category of species, not to as great an extent as all Peruvian herbaria (Table 13c). In total, there is an 8% increase in data deficient species when Peruvian herbaria are removed from calculations of EOO to assess Red List threat status of *Solanum* (12% DD in complete dataset, 20% when Peru herbaria removed). When removing the data from all Peruvian herbaria from calculations of AOO Red List threat status of *Solanum* species, the amount of critically endangered species increases by 22 spp., from 34 to 56 (Table 13d). As Peruvian herbaria are removed endangered, vulnerable and least concern species decrease as they move into higher threat categories.

12a.

EOO, # of Herbaria Removed		*0*	*5*	*10*	*15*	*20*	*25*	*30*	*32*
Red List Threat Status based on EOO # of spp.	DD	25	27	27	32	33	40	40	59
	CR	7	8	8	6	5	3	6	1
	EN	14	11	11	8	8	8	8	3
	VU	7	8	8	10	10	6	3	5
	LC	23	22	22	20	20	19	19	8

12b.

AOO, # of Herbaria Removed		*0*	*5*	*10*	*15*	*20*	*25*	*30*	*32*
Red List Threat Status based on AOO # of spp.	DD	0	0	0	0	0	0	0	0
	CR	28	30	30	34	35	41	42	59
	EN	46	44	44	42	41	35	34	17
	VU	2	2	2	0	0	0	0	0
	LC	0	0	0	0	0	0	0	0

Table 12 a/b: The number of *Begonia* species categorised into each Red List Threat category out of a total of 76, measured either by EOO or AOO as herbaria are removed. Demonstrating perceived increase in threat and increase of data deficient species as herbaria are removed. Species were treated as data deficient (DD) if it was not possible to calculate an EOO.

12c.

EOO, Herbaria removed		*All Herbaria*	*Removing Peruvian Herbaria*	*Removing Small Herbaria*
Red List Threat Status based on EOO # of spp.	DD	25	32	27
	CR	7	6	8
	EN	14	8	11
	VU	7	10	8
	LC	23	20	22

12d.

AOO, Herbaria Removed		*All Herbaria*	*Removing Peruvian Herbaria*	*Removing Small Herbaria*
Red List Threat Status based on AOO # of spp.	DD	0	0	0
	CR	28	34	30
	EN	46	42	44
	VU	2	0	2
	LC	0	0	0

Table 12 c/d: The number of *Begonia* species categorised into each Red List Threat category out of a total of 76 species, measured either by EOO or AOO as local Peruvian herbaria and small herbaria are removed. Demonstrating perceived increase in threat and increase of data deficient species as Peruvian herbaria and just small herbaria are removed. Species were treated as data deficient (DD) if it was not possible to calculate an EOO/AOO.

13a.

EOO, # of Herbaria Removed		*0*	*10*	*20*	*30*	*40*	*50*	*60*	*70*	*71*
IUCN Status	DD	33	45	45	45	64	64	67	202	249
# of spp.	CR	11	9	9	9	5	5	6	2	1
	EN	36	32	32	32	36	35	36	28	13
	VU	27	28	28	28	28	29	31	13	3
	LC	175	168	168	168	149	149	142	37	16

13b.

AOO, # of Herbaria Removed		*0*	*10*	*20*	*30*	*40*	*50*	*60*	*70*	*71*
Red List Status	DD	0	0	0	0	0	0	0	0	0
# of spp.	CR	34	46	47	47	63	64	67	203	250
	EN	238	228	227	227	217	216	214	78	32
	VU	9	8	8	8	2	2	1	1	0
	LC	1	0	0	0	0	0	0	0	0

Table 13 a/b: The number of *Solanum* species categorised into each Red List Threat category out of a total of 282, measured either by EOO or AOO as herbaria are removed. Demonstrating perceived increase in threat and increase of data deficient species as herbaria are removed. Species were treated as data deficient (DD) if it was not possible to calculate an EOO or AOO.

13c.

EOO, Herbaria removed		*All Herbaria*	*Removing Peruvian Herbaria*	*Removing Small Herbaria*
Red List Status	DD	33	55	41
# of spp.	CR	11	7	10
	EN	36	36	35
	VU	27	31	28
	LC	175	153	168

13d.

AOO, Herbaria removed		*All Herbaria*	*Removing Peruvian Herbaria*	*Removing Small Herbaria*
IUCN Status	DD	0	0	0
# of spp.	CR	34	56	43
	EN	238	224	230
	VU	9	2	9
	LC	1	0	0

Table 13 c/d: The number of *Solanum* species categorised into each Red List Threat category out of a total of 282 species, measured either by EOO or AOO as local Peruvian herbaria and small herbaria are removed. Demonstrating perceived increase in threat and increase of data deficient species as Peruvian herbaria and just small herbaria are removed. Species were treated as data deficient (DD) if it was not possible to calculate an EOO or AOO.

4 Discussion

Knowledge inequality, is reflected by the unequal distribution and abundance of herbaria, botanical expertise and data flow, established in colonialism and sustained in colonising countries. Known to scientists as bias, ignorance materialises into biodiversity data voids, characterised by a scarcity of specimens and knowledge on an area or country's species. To scientists concerned with global biodiversity, identifying data voids and bias in biodiversity data is essential for accelerating the progress of predicting the impact of rapid climate change and biodiversity break-down, and to enable the prioritisation of effective conservation planning (Hortal *et al.*, 2015). The Linnean and Wallacean shortfalls have the greatest influence on biodiversity data gaps and are essential for understanding the fundamentals of species relationships, distribution and conservation. Biodiversity researchers utilise data available from GBIF to model global scale data deficient areas (Feeley and Silman, 2011; Feeley, 2015). Unsurprisingly, the most significant gaps in global biodiversity knowledge are often found in biodiverse and tropical countries, where a low socio-economic status is an impediment to data availability (Meyer *et al.*, 2016; Vorontsova *et al.*, 2021). Despite attempts to democratise data, GBIF, the only biodiversity data repository that is global in scope and is globally accessibility, is predominantly formed of herbaria from wealthier countries that have the capacity to digitise and disseminate their collections. Without including data from all herbaria, how can we build an accurate and unbiased picture of global biodiversity patterns?

In Peru, a large percentage of specimens are held within local herbaria, yet the majority of local herbaria are digitally unavailable. Our datasets of two biodiverse genera in Peru are comprised of both local and international herbaria, from which we could evaluate differences between within and out-country herbaria. In this study we investigated whether the local herbaria contain specimen data with unique characteristics, in terms of their taxonomic uniqueness, geographic uniqueness, percentage of duplicates or the rarity of the species they represent. We found that local herbaria do not have more unique collections than international herbaria, nor are they less unique. However, our study reveals that local herbaria are key to understanding the distribution and particularly the conservation status of a species. Data held in local herbaria could significantly contribute to filling global biodiversity data voids, yet it is largely excluded from data repositories, analyses and scientific discourse.

4.1 Differences between data in local and international herbaria

Local herbaria are predominantly small in accession and young in age, often established in the late 20th or even 21st century. Local Peruvian herbaria hold 40-41% of specimens of *Begonia* and *Solanum* in our dataset (Fig. 3). Only 1% of the specimen data on GBIF for *Begonia* and *Solanum* are from Peruvian herbaria, with the remaining 99% from International herbaria (Fig. 3). This is expected to be representative for other groups of plants in Peru and for other poor biodiverse countries where local herbaria are digitally inaccessible. Considering our dataset only includes specimens collected from 9 out of 24 local herbaria in Peru, we predict that the patterns we have found would be more acute with an increase of local specimens. Correspondingly, our dataset does not include all specimens from international herbaria, and whilst digitisation is a priority in large international herbaria, it is a slow and variable process.

4.1.1 Can the Value of Herbaria be Quantified by the Uniqueness of the Specimens?

Monfils et al., (2020) recognised that regional or local herbaria could greatly contribute to fill gaps in our taxonomic, geographic and temporal understanding of global biodiversity. In order to quantify the value of local herbaria, we measured the uniqueness of the specimen data in each herbarium and tested their significance it in a regression model. Uniqueness indices were chosen based on their relevance to a broad range of biodiversity studies, including the taxonomy, geography and conservation of the species in question. Uniqueness was evaluated on the taxonomy, geography, quantity of duplicates, quantity of type specimens and rarity of species. Despite our calculations appearing to show that local Peruvian herbaria hold more mean unique specimen data in all indices compared to international herbaria (Table 4a/4b), our model was unable to detect any statistical significance to corroborate this finding. Local herbaria however are not less statistically unique, our results instead implying that each herbarium has a range an equally unique range of specimens according to the parameters we measured.

Large herbaria have incomparable numbers of specimens globally, despite the majority of global herbaria having a small accession (Thiers, 2016). Correspondingly, large herbaria also have a worldwide geographic focus (Thiers, 2016) or “international vocation” and are predominantly consulted for more than 90% for their international collections (Lavoie, 2013). We investigated whether local herbaria and larger herbaria had greater number *Begonia* and *Solanum* specimens. Our model detected that a greater number of specimens are found in local herbaria and in larger herbaria. This outcome was statistically significant and supports our estimations that local herbaria have a high abundance of regionally diverse specimens, irrespective of a smaller herbarium accession. It also supports our prediction that larger herbaria have a greater number of specimens, despite being geographically distant from the country of collection. This reveals the longstanding collection effort at large herbaria over time and moreover traces the continual presence of collectors from large herbaria overseas. It further implies that herbaria with a greater accession that have highly utilised collections, can justify greater funding for the up-keep and intensity of research at their institution.

Large herbaria have often been the priority for a researcher due to their high quality specimen data (Lavoie, 2013), greater resource management for loans (Casas-Marce *et al.*, 2012), and a greater abundance of species. Where a single visitation saves time and money, local herbaria have been overlooked, disregarding the value and range of their collections. We examined how taxonomically unique the specimen data held at local herbaria was compared to international herbaria and found that the overall observed mean of taxonomic uniqueness at local herbaria is much greater (Table 4a/4b). Our model suggests that the *Begonia* data from local herbaria, in larger herbaria or herbaria with more specimens decreases in taxonomic uniqueness, whereas increases in Solanum. Overall, these relationships are found to be statistically insignificant. In fact, the observed data shows a great impact on the uniqueness value the greater the number of specimens there are in the herbarium, suggesting a range of geographically wide and range restricted species dilutes the total uniqueness value. However, we can conclude that local and international herbaria are equally well represented in taxonomically unique data, and not only should be consulted in investigations but should be acknowledged for their potential for filling data voids. Collecting bias will strongly influence the taxonomic uniqueness of herbarium data, shaped by the clade interests of the staff and curation (Daru *et al.*, 2018) and this should be factored into consultation irrespective of herbarium size.

Specimens from poorly collected localities are crucial for furthering our understanding of biodiversity patterns and have a disproportionate impact on our understanding of the distributions and ecology of species, and the evolutionary biogeography of lineages. We investigated whether local herbaria have more geographically unique specimens and found that the observed mean geographic uniqueness for local herbaria is greater than for international herbaria (Table 4a/4b). Our model, however, demonstrates that in *Begonia*, local and larger herbaria have less geographically unique specimens but that herbaria with more *Begonia* specimens, conversely, have more geographically unique specimens. Similarly, local herbaria, larger herbaria, and herbaria with more *Solanum* specimens tend to have less geographically unique specimens. None of these effects are however significant. We find these relationships between geographic uniqueness and herbaria surprising, as geographical biases can affect international collectors who can have a limited knowledge of the local area and can be restricted to accessible areas, such as roads, lower elevations and localities near herbaria (Daru *et al.*, 2018). Our results instead may reflect resource imbalances and collecting priorities or agendas between local and international herbaria (Eichhorn, Baker and Griffiths, 2020). Staff from international herbaria may have access to larger budgets, allowing them to visit more remote areas than staff from local herbaria. Equally, the collecting priorities of local herbaria may include smaller geographic areas, which are better collected precisely because of the existence of those local herbaria. Correspondingly, local herbaria and local researchers may have different agendas to international herbaria and researchers, such a collecting for teaching (Marsico *et al.*, 2020), or for regional floras, in comparison to global biodiversity mapping.

Further to this, we considered whether herbaria hold data from a unique locality not found in other herbaria. Whilst we observed no distinct pattern of herbarium geographic uniqueness if in a local or larger herbarium, or if a herbarium held more *Begonia* or *Solanum* specimens, the mean herbarium geographic uniqueness of local herbaria was much greater than international herbaria (Table 4a/4b). Our model however found this relationship to be insignificant. Instead, interestingly this index generally increases incrementally in comparison to geographic uniqueness. More so for specific herbaria, such as large herbaria USM and MO for both genera, local herbarium HUT for *Begonia,* and large international herbaria BRIT *Solanum.* This implies that these herbaria hold collections with an overall greater unique locality, albeit insignificant, and could be prioritised for visitation by researchers.

Local herbaria hold a great amount of unduplicated specimens with a strong ecological, geographic and taxonomically unique data (Monfils, A.K., Nelson, 2014; Marsico *et al.*, 2020). We measured the duplicate uniqueness of *Begonia* and *Solanum* at local and international herbaria and gathered that the mean uniqueness is greater at local herbaria (Table 4a/4b). When we modelled the duplicate uniqueness, our results revealed a positive but insignificant relationship between duplicate uniqueness and the size of the accession, whether in local herbaria or measured by the number of specimens at an herbarium. We observed that duplicate uniqueness indices were greater in local herbaria and several large international herbaria which may influence our model outcome. Yet, our results may be representative of how the number of duplicates in local herbaria are highly influenced by foreign researchers depositing duplicates, and the implementation of duplication laws. Further, collecting and duplicate sharing has been subject to the bias of collectors personal preference, their social connections, or the proximity to herbaria (Marsico *et al.*, 2020). We consider that the utilisation of geographically relevant duplicated specimens may be greater in local herbaria compared to international herbaria (Marsico *et al.*, 2020). Unduplicated specimens however hold a unique record of a species and so increase the extrinsic value of that herbaria, incentivise herbarium visitation and digitisation of herbaria.

Type specimens, the reference specimens for published species, are invaluable for creating a strong taxonomy of a group of plants, and consequently the development of taxonomic expertise. Types are often held in old and large herbaria in colonising countries, where the majority of botanical experts has and still resides (Eichhorn, Baker and Griffiths, 2020). It is variable or unregulated whether the types of species new to taxonomy are held in herbaria of the country of collection or herbaria from which collector is from (Marsico *et al.*, 2020). We investigated whether there is a greater abundance of type specimens in local or international herbaria. Whilst our type uniqueness indices indicate that international herbaria with a large accession have a greater number of types irrespective of the amount of specimens of *Begonia* and *Solanum* they hold, our mean type uniqueness shows that local herbaria on average have a greater type uniqueness than international herbaria. Our model results were insignificant, however, they showed that the type uniqueness for *Begonia* increases if in a larger herbarium, or in herbaria with more *Begonia* specimens, conversely decreases if in a local herbarium, somewhat matching our observed indices. Then again, the type uniqueness of *Solanum* specimens decreases if in a larger herbarium, in local herbaria or in herbaria with more *Solanum* specimens. These results indicate that type specimens are spread unevenly across all herbaria and are greatly shaped by expertise and interests of collectors over time.

Specimens of rare species are important in increasing our understanding of species richness, ecosystem functioning and local extinction (Jain *et al.*, 2014). Without data on rare species, species are deemed data deficient and no conservation protection put in place (Parsons, 2016). We looked at whether local herbaria held a greater number of rare species compared to international herbaria. The mean indices for rarity uniqueness showed a greater amount of rarity in local herbaria in comparison to international herbaria (Table 4a/4b). Our model found that the rarity uniqueness for *Begonia* increased in local herbaria, larger herbaria or herbaria with a greater number of specimens. For *Solanum* rarity uniqueness increased in larger herbaria and herbaria with more *Solanum* specimens yet decreases in local herbaria. These results were however insignificant. We did not anticipate these results, as we would expect the knowledge of local botanists of local flora to increase the number of rare species collected. Collections of rare specimens are however greatly affected by sampling bias, where threatened species are either oversampled due to high scientific value or usually under sampled due to small or diminishing populations (Daru *et al.*, 2018). We suggest that model results could be influenced by number of specimens in an herbarium, where a great number of herbarium specimens can include common species, which can overshadow assessing where the rare species are held. Future calculations could only include the EOO of species with a restricted range, i.e., species that EOO is within Red List threat status. In general, local herbaria hold equally valuable specimen data on rare species compared to international herbaria and would contribute greatly to our understand species patterns under large-scale threat.

4.1.2 Discrepancy in the Temporal Geographic Distribution and Age Distribution of Specimens in Local and International Herbaria

There is a clear divergence in the age distribution and tempo-geographic distribution of *Begonia* and *Solanum* specimens between local and international herbaria. The presence of international herbaria in Peru for up to two centuries prior to local herbaria is indicative of the opportunities colonising countries were granted to far-reaching areas of the world.

Similarly, the early geographic distribution of international herbaria is clustered, with numerous collections at the same coordinates, indicative of colonial collectors and lack of historical infrastructure, access limited to prominent passes or popular localities (Daru *et al.*, 2018). By contrast, local herbaria in Peru have distinctly younger specimen data, whose geographic distribution of specimens rapidly increases from 1950 both in abundance and spread. This is reflected by the commence date of local herbaria, Peru's first and largest herbarium (USM) opening in 1918, with its youngest herbarium opening in 2003 (HOXA). While collecting on roads still is commonplace and causes ongoing, large biases in collecting patterns (Daru *et al.*, 2018), the geographic coverage of collecting greatly increases in local and international herbaria through time. The age distribution of all herbaria reveals how each herbarium has a unique collecting range in Peru, which undoubtably has been and is influenced by the agenda of each country geopolitically, by the research interests of herbaria and their staff (Eichhorn, Baker and Griffiths, 2020), and of temporal societal factors such as conflict and war, and migration of experts (Daru *et al.*, 2018; Proćków *et al.*, 2020; Vorontsova *et al.*, 2021).

4.1.3 The Relationship between Herbaria and Duplicates

Considering small herbaria in the USA were found to have unique specimens by locality, geographically and temporally (Marsico *et al.*, 2020), it was surprising that our model found no significance for local herbaria holding a greater uniqueness of specimen data than international herbaria. One explanation for this difference is due to the handling and relationship of duplicates in this study and between country herbaria. In this study we included all duplicates from all herbaria, whereas Marsico et al (2020) randomly selected one specimen from all duplicated collections. Duplicated specimens may mask significant patterns in our results between herbaria, for example, in taxonomic uniqueness where each specimen duplicated or not would contribute to the uniqueness indices in each herbarium. The duplicate index, however, does indicate the rate of duplicated specimens in an herbarium and was found to be insignificant across all herbaria. The extraction of biological and genetic resources outside the country of collection became illegal in many countries with the implementation of the international legal framework, the Nagoya Protocol, in 2014 (Rabeler *et al.*, 2019). This formulated that the benefits gained from being granted access for the collection of specimens and research, must be shared with the country of origin (Lendemer *et al.*, 2020). Thus, solidifying in law, the duplication specimens. In Peru, it being legally binding to leave duplicate specimens at local herbaria since 2001 (*Decreto Supremo Nº 014-2001-AG*, 2001; *Ley Forestal y de Fauna Silvestre LEY N° 29763*, 2011). Therefore, the increase of specimens observed in the temporal geographic distribution and age distribution of herbaria both in Peru and international herbaria from 2000 to-date (Fig. 4 and Fig. 5) could be influenced by duplicate specimens. The increase of collections at local herbaria in Peru from 1950-1999, comparable to international herbaria, is also striking and could signify the preference of international collectors leaving specimens at local herbaria and the increased opportunity to do so, as the number of local herbaria in Peru opening in this time period increased.

4.1.4 Future Differences to Examine between Local and International Herbaria

Begonia Resource Centre and Solanaceae Source are databases designed to facilitate biodiversity research, therefore the research design and questions we were able to ask were comparing local and international herbaria was limited to this field. Additional specimen details could be informative to study differences in herbarium data. For instance, qualitative or quantitative measures could be taken on the quality of label data, morphological trait data, flowering time, or if the specimen has available material for DNA extraction. Herbarium curation is another factor that could lead to differences the data in local and international herbaria, such as the time taken to mount specimens, time taken to send duplicates, quality of specimen paper, percentage of updated nomenclature of collections and time allocated to herbarium curation (Monfils, A.K., Nelson, 2014). The amount of funding for a herbarium could also be informative for assessing the capacity of the research and upkeep of an herbarium (Monfils, A.K., Nelson, 2014). Whilst the GDP of a country has been shown to increase the number of specimens and subsequently a barrier to building a robust picture of global conservation biodiversity (Amano and Sutherland, 2013), using the GDP of a country could be a beneficial factor to look at the differences between herbaria. Qualitative indices could be harder to measure but could have an effect in modelling the uniqueness of herbarium data.

Additional variables could be modelled to provide statistical evidence for relationships between local and international herbaria. The accession size could be modelled by the herbarium predictors, date commenced and number of staff, to elucidate if older herbaria, including those with colonial histories, are larger than younger herbaria (Lavoie, 2013) and maintaining and greater number of staff (also as indicated in our significant result between larger herbaria and number of *Begonia/Solanum* specimens). Further, the number of *Begonia* and *Solanum* specimens in an herbarium could be predicted in a model by the size of the accession and number of staff, indicating herbarium capacity for collecting. It could be assumed that older and larger herbaria have more type specimens, therefore types could be measured by modelling by the commence date and accession size. The relationship between the size of herbaria and the geography of the collections could be quantified by modelling the variable of herbaria geography as either national or worldwide, this could be predicted by accession size and show how large collections determine geographic scope and agendas of the herbarium. In order to study the relationship between colonised and colonising country, this factor could be assigned as a qualitative character and measured across the time of collections to see who was collecting and when. In effort to understand the bias in herbaria, patterns potentially covered by duplicates, it would be the interesting to assess the percentage of local collectors temporally in comparison to foreign researchers.

4.2 Local Herbaria are Essential for Measuring Red List Threat Status

Red List assessments have been internationally adopted by the IUCN since 2014 and provide a standardised method for assessing the conservation status of a species (Bland *et al.*, 2019). Species are evaluated and placed into five categories based on characteristics including their EOO and AOO: critically endangered, endangered, vulnerable, least concern and data deficient. These assessments are of vital importance as they are used to inform the conservation of global biodiversity ecosystems, and moreover inform legislation, land-use planning, protected areas management and ecosystem management (Parsons, 2016; Bland *et al.*, 2019; List, 2020). The EOO and AOO however, are subject to the quantity of spatial specimen data available for given species (Gaston and Fuller, 2009) and therefore subject to the bias from human-mediated collecting patterns (Daru *et al.*, 2018). Data deficient species do not have enough data to classify a threat status, and this could reflect a bias in collecting effort; low abundance, rarely sighted, or cryptic species from inaccessible localities are more likely to be data deficient (Parsons, 2016). Therefore, it should be considered whether these species are in fact ‘assume threatened’, as determining a species as data deficient has a great impact on the future conservation and handling of that species (Parsons, 2016; Vorontsova *et al.*, 2021). To the extent that there is little funding for data deficient species, nor incentive and time to study the species in question (Parsons, 2016). Our results show that international datasets can be missing specimen data that greatly modifies the threat status of a species, with heavy consequences for the conservation of that species.

In our study we quantified the contribution of local herbaria to calculations of Red List threat status by removing local herbaria and their specimen data from calculations using the complete dataset. We found, that without local herbaria, we greatly overestimate the quantity of rare species, and the rarity of individual species. Moreover, we lose the ability to even calculate the threat status for significant numbers of species, especially species considered rare or endangered when the whole dataset is used. In other words, high threat species lose vital specimen data when local herbaria are ignored, where specimen data was already scarce, so they appear rarer than they are or data deficient. Without local herbaria, the following species are lost to data deficiency, to name a few, *B. yuracyacuensis B. rodriguezii, S. trinitense and S. gracilifrons.* Additionally, 8 out of 36 *Solanum* spp., are reassigned to critically endangered, as are *B. joshii, B. altoperuviana* and *B. bifurcata.*
Without an accurate IUCN assessment, enigmatic, rare, and endangered species like *B. rodriguezii* won’t be receive the necessary conservation needed to protect it from extinction, data from local herbaria if therefore essential to the survival of endangered species.

The amount of *Begonia* and *Solanum* specimen data stored in local herbaria reflects these patterns, where local herbaria hold 22% of our total dataset. The large local herbarium USM also holds an additional 20-25% of our data and substantial decreases in EOO and AOO of species were demonstrated when this herbarium was removed. Correspondingly, specific international herbaria hold great amounts of specimen data for these species: for *Begonia,* E and MO, and for *Solanum* BM. The differences between the way threat category parameters EOO and AOO are calculated affects the outcome of threat status of a species. While EOO requires a minimum of 3 specimens to make a convex polygon, only 1 specimen is needed to calculate the grid-cell distribution of a species for AOO. The greatest distinction is in the number of data deficient species; where the EOO cannot be calculated, species become data deficient the less data is available. Considering parameter EOO is the most widely used parameter for calculating threat status in plants (Gaston and Fuller, 2009) and a more conservative measure, we consider this as the most useful measure of threat changes when local herbaria are removed. This part of our analysis could be further improved by comparing our results to a randomisation model, that would measure the EOO and AOO when removing specimen data randomly and demonstrate whether removing local and small herbaria is more significant than by chance. This would also corroborate patterns we have found in the contribution of local herbaria to biodiversity analysis, although it is apparent that herbariums with large quantities of data on these plant groups contribute greatly of our understanding of their range size.

Where local herbaria greatly contribute to calculating the threat status of a species, the measure of species that become data deficient represents the changes in threat category or the lack of remaining data available (Table 12 c-d/13 c-d). When local herbaria are removed from a total dataset of EOO, there is an 11% increase in data deficient *Begonia* species; similarly, there is an increase of 8% in data deficient *Solanum* spp.,. We predict our results could be representative for other biodiverse genera in biodiverse countries, where calculations of threat status must include local herbaria to decrease the amount of data deficient species. There are an estimated 19,147 species of vascular plants in Peru (Ulloa Ulloa *et al.*, 2017), and our results suggest that around 2,106 spp. would be deemed data deficient if the threat status of species were to be calculated only using international herbaria. We believe this finding could further be representative of all Andean countries and assessments of threat status, where around 11,271 out of 102,471 species (Ulloa Ulloa *et al.*, 2017) would not have enough data to calculate the threat status using EOO without local herbaria. Considering our *Solanum* dataset is not comprised of all local herbaria in Peru, we expect that additional local herbarium specimen data would decrease data deficiency for calculating to threat assessments. While in this study we only look at the how local herbaria contribute to measuring threat status, we predict that our results will be applicable to any study that uses specimen data to measure species geographic ranges', such as species distribution models (SDM), taxonomic checklists, and biogeographic studies (Meyer, Weigelt and Kreft, 2016). Areas of biodiversity research that use herbarium specimen data are all subject to the same sampling bias created from unequal knowledge systems, geographic spread and mobilisation of data.

4.2.1 Mobilisation of Biodiversity Data

In our study we highlight how data deficiency, can be an artifact of data mobilisation rather than a lack of data. Without local knowledge and local specimen data, international researchers, and currently biased datasets such as GBIF, are creating inaccurate biodiversity assessments. Unsurprisingly, colonial legacies persist not only in the unequal distribution of physical herbaria, but follow into the digital realm, where the ability to digitise and disseminate data, and then benefit from accessing data is asymmetrically experienced, further prolonging disparity between knowledge and expertise globally. Mobilisation of biodiversity data must therefore be prioritised and intensified across countries, where knowledge prevails but resources may be limited (Meyer *et al.*, 2015).

Where mobilisation and digitisation of biodiversity data is paramount for equity in global science, we can look to positive examples of collaborative mobilisation of specimen data internationally. Leading examples include Brazil’s government funded CRIA formed a virtual herbarium made up of the majority of Brazilian herbaria and their ReFlora programme repatriating digital specimen data from countries that hold their data. Examples of internationally cross-funded mobilisation of biodiversity databases include Mexico's CONABIO biodiversity database the National Biodiversity Information System (SNIB) and The Atlas of Living Costa Rica (CRBIO). However a large proportion of databases still remain in private collections or maintained by single individuals (Eichhorn, Baker and Griffiths, 2020), and the importance of making this information widely accessible must be conveyed to national and international funding bodies. Both databases used in this study, Begonia Resource Centre and Solanaceae Source, are free and digitally accessible, however due their taxonomic specialism, they are relatively unknown. Currently GBIF rarely accepts taxonomically specific databases. Plans to mobilise these databases are with their integration into The World Flora Online (WFO) (*Taxonomic Expert Networks (TENs): World Flora Online*, 2021), a taxonomically verified and freely accessible resource aimed at fulfilling Target 1 of Global Strategy for Plant Conservation (GSPC) with “An online flora of all known plants” (GSPC, 2012).

4.3 Future Integration of Decoloniality into Biodiversity Research Practice

To avoid reproducing oppressive and colonial research, botanical and biodiversity scientists must consider integrating inclusive and ethical practices within their research by considering both plants and people in the ecosystem for a sustainable future (Trisos, Auerbach and Katti, 2021). It is key for researchers to reflect on how scientific objectivity has obscured incentives to impart foreign knowledge onto different ecosystems and places globally (Baker, Eichhorn and Griffiths, 2019), sometimes to the detriment and displacement of local people (Trisos, Auerbach and Katti, 2021). With the opportunity to study global scale biology with access to "big data", researchers should realise that accurate scientific investigations are only achievable with a diversity of global authors whose local knowledge is essential. If global north scientists continue their work overseas, they must acknowledge their own knowledge biases and be open to other ways of looking at and working with biodiversity. Where scientists from colonised countries have had to adapt to dominant scientific frameworks and English language communication necessary for science careers, reciprocal efforts must be made by white western researchers (Trisos, Auerbach and Katti, 2021). Not only should research in countries of interest be collaborative with local partners, but project-leading and authoring should be prioritised for within-country researchers with whom have intrinsically valuable local knowledge (Baker, Eichhorn and Griffiths, 2019; Trisos, Auerbach and Katti, 2021). Where reading and referencing highly cited published papers enables the reproduction of dominantly Global North authors, a more ethical practice would seek local scientists voices those writing may be on free publishable or local platforms, in other languages and in other knowledge forms such as interviews. Similarly, good practice would involve making your research findings available and accessible to the local people it affects. Biodiversity science education would benefit from teaching ethical and critical practice, drawing from other disciplines such human geography and political ecology - reflecting on scientific objectivity and the perceived neutral voice, ultimately disconnects scientists from reflecting on the ethical impact of their work. Furthermore, botanical and biodiversity scientists, as other disciplines, should incorporate a positionality statement in their published works, which enables reflection and author transparency, from the viewpoint at which they write (Baker, Eichhorn and Griffiths, 2019; Eichhorn, Baker and Griffiths, 2020).

5 Conclusion:

In an attempt to decolonise herbaria and the specimen data they hold, we have shown how colonialism has produced and sustained an unequal distribution of botanical knowledge, expertise, and resources globally. We recognise how colonial legacies persist in the present, unequal distribution and mobilisation of biodiversity data, where Global North authors produce, access, and benefit from global biodiversity data. In this study we have demonstrated the significant value of local knowledge, local herbaria and the data they hold. Local herbaria hold unique biodiversity data which are imperative to understanding and analysing accurate biodiversity patterns. “Data deficiency” instead, is both declared and perceived by international researchers who lack local data and local knowledge. Where global biodiversity agendas are set in the Global North (Eichhorn, Baker and Griffiths, 2020), whether or not local herbaria filling “data voids” is in the interest of local people and botanists, should be discussed. Instead, incentives to mobilise local data must be in the interest of aiding the careers and expertise of local botanists and local herbaria, otherwise data mobilisation will continue to be colonising in the extraction of data that benefits the authors and institutions from the Global North.

Decoloniality and Positionality Statement:

I acknowledge the limitations of decolonising within this study. That whilst we use Peru and its flora as study system to understand how colonialism has shaped global biodiversity knowledge and patterns, that the practice itself of using and writing about data from a country I do not reside, georeferencing land I have never been, from a country whose language I do not speak, could in itself perpetuate colonising attitudes. For this study to truly benefit local botanists from Peru, local authors, local priorities, and concerns must be consulted. Furthermore, in this analysis we compare the differences between local within-country knowledge with international out-country knowledge, but do not consider Peruvian indigenous knowledge as an alternative to dominant modes of knowledge found in European botany, herbarium science or biodiversity science. While I advocate for the inclusion indigenous knowledge within botanical science, scientists must first be open to alternative and creative ways of studying and assessing plant biodiversity.

6 Bibliography

Amano, T. and Sutherland, W. J. (2013) ‘Four barriers to the global understanding of biodiversity conservation: Wealth, language, geographical location and security’, *Proceedings of the Royal Society B: Biological Sciences*. doi: 10.1098/rspb.2012.2649.

van Andel, T. (2017) *Decolonise the botanical treasure house*, *Universiteit Leiden*. Available at: https://www.universiteitleiden.nl/en/news/2017/01/decolonise-the-botanical-treasure-house.

Antonelli, A. (2020) *It’s time to decolonise botanical collections*. Available at: https://www.kew.org/read-and-watch/time-to-decolonise-botanical-collections (Accessed: 14 May 2021).

Baber, Z. (2016) ‘The Plants of Empire: Botanic Gardens, Colonial Power and Botanical Knowledge’, *Journal of Contemporary Asia*, pp. 659–679. doi: 10.1080/00472336.2016.1185796.

Baker, K., Eichhorn, M. P. and Griffiths, M. (2019) ‘Decolonizing field ecology’, *Biotropica*, 51(3), pp. 288–292. doi: 10.1111/btp.12663.

Beck, J. *et al.* (2014) ‘Spatial bias in the GBIF database and its effect on modeling species’ geographic distributions’, *Ecological Informatics*, 19, pp. 10–15. doi: 10.1016/j.ecoinf.2013.11.002.

Bivand, R., Keitt, T. and Rowlinson, B. (2021) ‘rgdal: Bindings for the “Geospatial” Data Abstraction Library’. Available at: https://cran.r-project.org/package=rgdal.

Bivand, R. and Lewin-Koh (2021) ‘maptools: Tools for Handling Spatial Objects’. Available at: https://cran.r-project.org/package=maptools.

Bland, L. M. *et al.* (2019) ‘Impacts of the IUCN Red List of Ecosystems on conservation policy and practice’, *Conservation Letters*, 12(5), pp. 1–8. doi: 10.1111/conl.12666.

Bleichmar, D. (2012) *Visible Empire: Botanical Expeditions and Visual Culture in the Hispanic Enlightenment*. Chicago University Press.

Bowker, G. and Star, S. (2000) *Sorting Things Out: Classification and its Consequences*. Cambridge: MA: MIT Press.

Brockway, L. H. (1979) ‘Science and Colonial Expansion : The Role of the British Royal Botanic Gardens’, *American Ethnologist*, 6(3), pp. 449–465.

Brownrigg, R., Minka, T. . and Deckmyn, A. (2018) ‘maps: Draw Geographical Maps’. Available at: https://cran.r-project.org/packages=maps.

Brunete, J. (no date) *Begonia [incarnata] by Brunete. Original drawing from Ruiz & Pavón’s Expedition (1777-1816)*, *JSTOR*. Available at: https://plants.jstor.org/stable/pdf/10.5555/al.ap.visual.ma-ajb04-d-1399.

Canteiro, C. *et al.* (2019) ‘Enhancement of conservation knowledge through increased access to botanical information’, *Conservation Biology,* 33(3), pp. 523–533. doi: 10.1111/cobi.13291.

Casas-Marce, M. *et al.* (2012) ‘The value of hidden scientific resources: Preserved animal specimens from private collections and small museums’, *BioScience*, 62(12), pp. 1077–1082. doi: 10.1525/bio.2012.62.12.9.

Christenhusz, J. ., Fay, M. and Chase, M. . (2017) *Plants of the World: An Illustrated Encyclopedia of Vascular Plants*. London: Kew Publishing.

Coggins, R. (2018) ‘decolonization. In A Concise Oxford Dictionary of Politics and International Relations.’, *Oxford University Press*. Available at: https://www.oxfordreference.com/view/10.1093/acref/9780199670840.001.0001/acref-9780199670840-e-322. (Accessed: 1 July 2021).

Colombo, B., Kaehler, M. and Calvente, A. (2016) ‘An inventory of the bignoniaceae from the Brazilian state of Rio Grande do Norte highlights the importance of small herbaria to biodiversity studies’, *Phytotaxa,* 278(1), pp. 19–28. doi: 10.11646/phytotaxa.278.1.2.

Daru, B. H. *et al.* (2018) ‘Widespread sampling biases in herbaria revealed from large-scale digitization’, *New Phytologist,* 217(2), pp. 939–955. doi: 10.1111/nph.14855.

Das, S. and Lowe, M. (2018) ‘Nature Read in Black and White: decolonial approaches to interpreting natural history collections’, *JoNSC,* 6, pp. 4–14. Available at: http://www.natsca.org/article/2509.

Dauby, G. *et al.* (2017) ‘ConR: An R package to assist large-scale multispecies preliminary conservation assessments using distribution data’, *Ecology and Evolution,* 7(24), pp. 11292–11303. doi: 10.1002/ece3.3704.

Decreto Supremo Nº 014-2001-AG (2001) *Aprueban el Reglamento de la Ley Forestal y de Fauna Silvestre*.

Drew, J. A., Moreau, C. S. and Stiassny, M. L. J. (2017) ‘Digitization of museum collections holds the potential to enhance researcher diversity’, *Nature Ecology and Evolution,* 1(12), pp. 1789–1790. doi: 10.1038/s41559-017-0401-6.

Dundee, V. (2020) *Decolonising our galleries: An introduction*. Available at: https://www.vam.ac.uk/dundee/articles/decolonising-our-galleries-an-introduction (Accessed: 18 June 2021).

Eichhorn, M. P., Baker, K. and Griffiths, M. (2020) ‘Steps towards decolonising biogeography’, *Frontiers of Biogeography,* 12(1), pp. 0–7. doi: 10.21425/F5FBG44795.

Enquist, B. J. *et al.* (2019) ‘The commonness of rarity: Global and future distribution of rarity across land plants’, *Science Advances,* 5(11), pp. 1–14. doi: 10.1126/sciadv.aaz0414.

Feeley, K. (2015) ‘Are we filling the data void? An assessment of the amount and extent of plant collection records and census data available for tropical South America’, *PLoS ONE,* 10(4). doi: 10.1371/journal.pone.0125629.

Feeley, K. and Silman, M. (2011) ‘The data void in modeling current and future distributions of tropical species’, *Global Change Biology*, 17(1), pp. 626–630. doi: 10.1111/j.1365-2486.2010.02239.x.

Frodin, D. G. (2004) ‘History and Concepts of Big Plant Genera Published by : International Association for Plant Taxonomy (IAPT) Stable URL : https://www.jstor.org/stable/4135449 REFERENCES Linked references are available on JSTOR for this article : History and concepts of’, *Taxon*, 53(3), pp. 753–776.

Garnier, S. *et al.* (2021) ‘viridis: Colorblind-Friendly Color Maps for R’.

Gaston, K. J. and Fuller, R. A. (2009) ‘The sizes of species’ geographic ranges’, *Journal of Applied Ecology*, 46(1), pp. 1–9. doi: 10.1111/j.1365-2664.2008.01596.x.

Gillman, L. N. and Wright, S. D. (2020) ‘Restoring indigenous names in taxonomy’, *Communications Biology*, 3(1), pp. 20–22. doi: 10.1038/s42003-020-01344-y.

Glon, H. E. *et al.* (2017) ‘The contribution of small collections to species distribution modelling: A case study from Fuireneae (Cyperaceae)’, *Ecological Informatics*, 42(September), pp. 67–78. doi: 10.1016/j.ecoinf.2017.09.009.

Gritcan, I. *et al.* (2016) ‘Leaf stable isotope and nutrient status of temperate mangroves as ecological indicators to assess anthropogenic activity and recovery from Eutrophication’, *Frontiers in Plant Science*, 7(DECEMBER2016), pp. 1–11. doi: 10.3389/fpls.2016.01922.

GSPC (2012) *Global Strategy for Plant Conservation*. Available at: https://www.bgci.org/wp/wp-content/uploads/2019/04/Guide_to_GSPC_english.pdf.

Gurminder, K., Bhambra, D. G. and Nişancıoğlu, K. (2018) ‘Introduction: Decolonising the University?’, in *Decolonising the University*. London: Pluto Press, pp. 70–115.

Harris, K. M. and Marsico, T. D. (2017) ‘Digitizing specimens in a small herbarium: A viable workflow for collections working with limited resources’, *Applications in Plant Sciences*, 5(4), p. 1600125. doi: 10.3732/apps.1600125.

Heberling, J. M. *et al.* (2021) ‘Data integration enables global biodiversity synthesis’, *Proceedings of the National Academy of Sciences of the United States of America*, 118(6), pp. 1–7. doi: 10.1073/pnas.2018093118.

Heberling, J. M. and Isaac, B. L. (2017) ‘Herbarium specimens as exaptations: New uses for old collections’, *American Journal of Botany*, 104(7), pp. 963–965. doi: 10.3732/ajb.1700125.

Heberling, J. M., Prather, L. A. and Tonsor, S. J. (2019) ‘The Changing Uses of Herbarium Data in an Era of Global Change: An Overview Using Automated Content Analysis’, *BioScience*, 69(10), pp. 812–822. doi: 10.1093/biosci/biz094.

Hedrick, B. P. *et al.* (2020) ‘Digitization and the Future of Natural History Collections’, *BioScience*, 70(3), pp. 243–251. doi: 10.1093/biosci/biz163.

Hijmans, R. . (2021) ‘raster: Geographic Data Analysis and Modeling’. doi: R package version 3.4-10.

Hill, M. . (2017) ‘Peru’, in *WORLDMARK ENCYCLOPEDIA OF THE NATIONS*. 14TH EDITI. Gale, A Cengage Company.

Hood, M. E. *et al.* (2010) ‘Distribution of the anther-smut pathogen Microbotryum on species of the Caryophyllaceae’, *New Phytologist*, 187(1), pp. 217–229. doi: 10.1111/j.1469-8137.2010.03268.x.

Hortal, J. *et al.* (2015) ‘Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity’, *Annual Review of Ecology, Evolution, and Systematics*, 46(1), pp. 523–549. doi: 10.1146/annurev-ecolsys-112414-054400.

Hughes, M. *et al.* (2015) *Begonia Resource Centre*. Available at: https://padme.rbge.org.uk/begonia/page/home.

IUCN (2019) *Guidelines for Using the IUCN Red List Categories and Criteria, Version 14*. Available at: http://www.iucnredlist.org/documents/RedListGuidelines.pdf.

Jain, M. *et al.* (2014) ‘The importance of rare species: A trait-based assessment of rare species contributions to functional diversity and possible ecosystem function in tall-grass prairies’, *Ecology and Evolution*, 4(1), pp. 104–112. doi: 10.1002/ece3.915.

Jazeel, T. (2017) ‘Mainstreaming geography’s decolonial imperative’, *Transactions of the Institute of British Geographers*, 42(3), pp. 334–337. doi: 10.1111/tran.12200.

Kier, G. *et al.* (2005) ‘Global patterns of plant diversity and floristic knowledge’, *Journal of Biogeography*, 32(7), pp. 1107–1116. doi: 10.1111/j.1365-2699.2005.01272.x.

Knapp, S., Sánchez, I. and León, B. (2006) ‘Solanaceae endémicas del Perú’, *Revista Peruana de Biologia*, 13(2), p. 473s.

Lavoie, C. (2013) ‘Biological collections in an ever changing world: Herbaria as tools for biogeographical and environmental studies’, *Perspectives in Plant Ecology, Evolution and Systematics*, 15(1), pp. 68–76. doi: 10.1016/j.ppees.2012.10.002.

Legg, S. (2017) ‘Decolonialism’, *Transactions of the Institute of British Geographers*, 42(3), pp. 345–348. doi: 10.1111/tran.12203.

Lendemer, J. *et al.* (2020) ‘The Extended Specimen Network: A Strategy to Enhance US Biodiversity Collections, Promote Research and Education’, *BioScience*, 70(1), pp. 23–30. doi: 10.1093/biosci/biz140.

León, B. and Monsalve, C. (2006) ‘Begoniaceae endémicas del Perú’, *Revista Peruana de Biología*, 13(2).

Ley Forestal y de Fauna Silvestre LEY N° 29763 (2011) *Diario Oficial el Peruano*. Available at: https://leyes.congreso.gob.pe/Documentos/Leyes/29763.pdf (Accessed: 30 August 2021).

List, R. (2020) *IUCN Red List 2017 – 2020 Report.* Available at: https://nc.iucnredlist.org/redlist/resources/files/1630480997-IUCN_RED_LIST_QUADRENNIAL_REPORT_2017-2020.pdf (Accessed: 7 August 2021).

López, A. and Sassone, A. B. (2019) 'The Uses of Herbaria in Botanical Research. A Review Based on Evidence From Argentina', *Frontiers in Plant Science*, 10(November), pp. 1–10. doi: 10.3389/fpls.2019.01363.

Mabberley, D. . (2017) *Mabberley's Plant Book*. 3rd edn. Cambridge University Press.

Macbride, J. . (1962) *Flora of Peru (5B)*. Chicago Field Museum of Natural History (Botany).

Malhado, A. C. M. *et al.* (2014) 'Geographic and Temporal Trends in Amazonian Knowledge Production', *Biotropica*, 46(1), pp. 6–13.

Marchese, C. (2015) 'Biodiversity hotspots: A shortcut for a more complicated concept', *Global Ecology and Conservation*, 3, pp. 297–309. doi: 10.1016/j.gecco.2014.12.008.

Marsico, T. D. *et al.* (2020) 'Small herbaria contribute unique biogeographic records to county, locality, and temporal scales', *American Journal of Botany*, 107(11), pp. 1577–1587. doi: 10.1002/ajb2.1563.

McMichael, C. N. H. *et al.* (2017) 'Ancient human disturbances may be skewing our understanding of Amazonian forests', *Proceedings of the National Academy of Sciences of the United States of America*, 114(3), pp. 522–527. doi: 10.1073/pnas.1614577114.

Meyer, C. *et al.* (2015) 'Global priorities for an effective information basis of biodiversity distributions', *Nature Communications*, 6, pp. 1–8. doi: 10.1038/ncomms9221.

Meyer, C. *et al.* (2016) 'Range geometry and socio-economics dominate species-level biases in occurrence information', *Global Ecology and Biogeography*, 25(10), pp. 1181–1193. doi: 10.1111/geb.12483.

Meyer, C., Weigelt, P. and Kreft, H. (2016) 'Multidimensional biases, gaps and uncertainties in global plant occurrence information', *Ecology letters*, 19(8), pp. 992–1006. doi: 10.1111/ele.12624.

Ministerio de Transported y Comunicaciones Peru (2016) *Mapas Viales*. Available at: http://portal.mtc.gob.pe/transportes/caminos/normas_carreteras/mapas_viales.html.

Monfils, A.K., Nelson, G. (2014) 'The Small Collections Network: Recruiting, Retaining, and Sustaining Small Collections in Biodiversity Digitization', *Society for the Preservation of Natural History Collections*, 28(1), pp. 32–33.

Monfils, A. K. *et al.* (2020) 'Regional collections are an essential component of biodiversity research infrastructure', *BioScience*, 70(12), pp. 1045–1047. doi: 10.1093/biosci/biaa102.

Moonlight et al. (2021) 'Flora of Peru, Begoniaceae.', *Edinburgh Journal of Botany*.

Moonlight, P. W. *et al.* (2015) 'Continental-scale diversification patterns in a megadiverse genus: The biogeography of Neotropical Begonia', *Journal of Biogeography*, 42(6), pp. 1137–1149. doi: 10.1111/jbi.12496.

Moonlight, P. W. *et al.* (2018) 'Dividing and conquering the fastest-growing genus: Towards a natural sectional classification of the mega-diverse genus begonia (Begoniaceae)', *Taxon*, 67(2), pp. 267–323. doi: 10.12705/672.3.

Moonlight, P. W. and Reynel, C. (2018) 'Two new species of begonia from Andean Peru', *Phytotaxa*, 381(1), pp. 116–126. doi: 10.11646/phytotaxa.381.1.15.

Museums Galleries Scotland (2020) *Decolonising your museum*. Available at: https://www.museumsgalleriesscotland.org.uk/advice/mgs-knowledge-exchange-resources/decolonising-your-museum/ (Accessed: 18 June 2021).

Myers, N. *et al.* (2000) 'Biodiversity hotspots for conservation priorities', *Nature*, 403. doi: 10.1080/21564574.1998.9650003.

National Museums Scotland (no date) *Colonial histories and legacies in our museums*. Available at: https://www.nms.ac.uk/collections-research/our-collections/colonial-histories-and-legacies-in-our-museums/ (Accessed: 18 June 2021).

NatSCA (2020) *Decolonising Natural Sciences Collections*. Available at: https://natsca.blog/2020/01/23/decolonising-natural-sciences-collections/ (Accessed: 18 June 2021).

Nelson, G. and Ellis, S. (2019) 'The history and impact of digitization and digital data mobilization on biodiversity research', *Philosophical Transactions of the Royal Society B: Biological Sciences*. doi: 10.1098/rstb.2017.0391.

Nic Lughadha, E. *et al.* (2020) 'Extinction risk and threats to plants and fungi', *PLANTS, PEOPLE, PLANET*, 2(5), pp. 389–408. doi: 10.1002/ppp3.10146.

Noxolo, P. (2017) 'Decolonial theory in a time of the re-colonisation of UK research', *Transactions of the Institute of British Geographers*, 42(3), pp. 342–344. doi: 10.1111/tran.12202.

Parsons, E. C. M. (2016) 'Why IUCN should replace "data deficient" conservation status with a precautionary "assume threatened" status-a cetacean case study', *Frontiers in Marine Science*, 3(OCT), pp. 2015–2017. doi: 10.3389/fmars.2016.00193.

PBI Solanum Project (2021) *Solanaceae Source*. Available at: http://www.solanaceaesource.org/.

Pimm, S. L. (2021) 'What we need to know to prevent a mass extinction of plant species', *Plants People Planet*, 3(1), pp. 7–15. doi: 10.1002/ppp3.10160.

Proćków, J. *et al.* (2020) 'How many type specimens can be stored in old lesser-known herbaria with turbulent histories? – A Juncus case study reveals their importance in taxonomy and biodiversity research', *PhytoKeys*, 153, pp. 85–110. doi: 10.3897/phytokeys.153.50735.

Qui, Y. (2021) 'showtext: Using Fonts More Easily in R Graphs'.

R Core Team (2020) 'R: A language and environment for computing.' Vienna, Austria: R Foundation for Statistical Computing.

Rabeler, R. K. *et al.* (2019) 'Herbarium Practices and Ethics, III', *Systematic Botany*, 44(1), pp. 7–13. doi: 10.1600/036364419X697840.

Radcliffe, S. A. (2017) 'Decolonising geographical knowledges', *Transactions of the Institute of British Geographers*, 42(3), pp. 329–333. doi: 10.1111/tran.12195.

RBG Kew (2021) *Our Manifesto for Change 2021-2030*. Available at: https://www.kew.org/about-us/press-media/manifesto-for-change-2021.

RBGE (2021a) *DIGITISATION*. Available at: https://www.rbge.org.uk/science-and-conservation/herbarium/accessing-the-collections/digitisation.

RBGE (2021b) *EQUITY AND ACCESS*. Available at: https://www.rbge.org.uk/about-us/equity-and-access/ (Accessed: 18 June 2021).

Riches, C. and Palmowski, J. (2021) 'decolonization'. Oxford University Press. doi: 10.1093/acref/9780191890949.013.0630.

Rodrigues, A. S. L. *et al.* (2010) 'A global assessment of amphibian taxonomic effort and expertise', *BioScience*, 60(10), pp. 798–806. doi: 10.1525/bio.2010.60.10.6.

Särkinen, T. *et al.* (2013) 'A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): A dated 1000-tip tree', *BMC Evolutionary Biology*, 13(1). doi: 10.1186/1471-2148-13-214.

Särkinen, T. *et al.* (2015) 'Listado anotado de Solanum L . (Solanaceae en el Perú; Annotated checklist of Solanum L . (Solanaceae) for Peru', *Revista Peruana de Biologia*, 22(1), pp. 3–62. doi: http://dx.doi.org/10.15381/rpb.v22i1.11121.

Sium, A., Desai, C. and Ritskes, E. (2012) 'Towards the "tangible unknown": Decolonization and the Indigenous future', *Decolonization: Indigeneity, Education & Society*, 1(1), pp. I–XIII. Available at: http://decolonization.org.

Slowikowski, K. (2021) 'ggrepel: Automatically Position Non-Overlapping Text Labels with "ggplot2"'.

Smith, L. . (2010) *Decolonizing methodologies*. London: Zed Books.

Smith, L. . and Schubert, B. . (1941) ‘Begoniaceae’, in *Flora of Peru*. Chicago Field Museum of Natural History (Botany), pp. 181–202.

Soltis, D. E. and Soltis, P. S. (2016) ‘Mobilizing and integrating big data in studies of spatial and phylogenetic patterns of biodiversity’, *Plant Diversity*, 38(6), pp. 264–270. doi: 10.1016/j.pld.2016.12.001.

Soltis, P. ., Nelson, G. and James, S. . (2018) ‘Green digitization: Online botanical collections data answering real-world questions: Online’, *Applications in Plant Sciences*, 6(2), pp. 4–7. doi: 10.1002/aps3.1028.

Soltis, P. S. (2017) ‘Digitization of herbaria enables novel research’, *American Journal of Botany*, 104(9), pp. 1281–1284. doi: 10.3732/ajb.1700281.

Sousa-Baena, M. S., Garcia, L. C. and Townsend Peterson, A. (2014) ‘Knowledge behind conservation status decisions: Data basis for “Data Deficient” Brazilian plant species’, *Biological Conservation*, 173, pp. 80–89. doi: 10.1016/j.biocon.2013.06.034.

South America: Peru (2021) *The World Factbook*. Available at: https://www.cia.gov/the-world-factbook/countries/peru/#introduction.

Stoler, A. . (2016) ‘I . Concept Work: FRAGILITIES AND FILIATIONS’, in *Duress: Imperial Durabilities in Our Times*. Durham, North Carolina: Duke University Press.

Sunderland, M. E. (2013) ‘Computerizing natural history collections’, *Endeavour*, 37(3), pp. 150–161. doi: 10.1016/j.endeavour.2013.04.001.

Sweeney, P. W. *et al.* (2018) ‘Large-scale digitization of herbarium specimens: Development and usage of an automated, high-throughput conveyor system’, *Taxon*, 67(1), pp. 165–178. doi: 10.12705/671.9.

Swenson, J. J. *et al.* (2012) ‘Plant and animal endemism in the eastern Andean slope: Challenges to conservation’, *BMC Ecology*, 12(January). doi: 10.1186/1472-6785-12-1.

Taxonomic Expert Networks (TENs): World Flora Online (2021) *World Flora Online*. Available at: http://about.worldfloraonline.org/tens.shtml.

Thiers, B. (2016) *Index Herbariorum: A Global Directory of Public Herbaria and Associated Staff. New York Botanical Garden’s Virtual Herbarium.* Available at: http://sweetgum.nybg.org/science/ih/.

Thiers, B. . (2020a) *Herbarium: The Quest to Preserve and Classify the World’s Plants*. Portland, Oregon: Timber Press.

Thiers, B. . (2020b) *The World’s Herbaria 2020: A Summary Report Based on Data from Index Herbariorum Issue*.

Toomey, A. H. (2016) ‘What happens at the gap between knowledge and practice? Spaces of encounter and misencounter between environmental scientists and local people’, *Ecology and*

Society, 21(2). doi: 10.5751/ES-08409-210228.

Trisos, C. H., Auerbach, J. and Katti, M. (2021) ‘Decoloniality and anti-oppressive practices for a more ethical ecology’, *Nature Ecology & Evolution*. doi: 10.1038/s41559-021-01460-w.

Ulloa Ulloa, C. *et al.* (2017) ‘An integrated assessment of the vascular plant species of the Americas’, *Science*, 358(6370), pp. 1614–1617. doi: 10.1126/science.aao0398.

Urbanek, S. (2013) ‘png: Read and Write PNG images’. Available at: https://cran.r-project.org/package=png.

Vargas, F. *et al.* (2016) *Collection and preservation of herbarium specimens for the CIP genebank.*

Vorontsova, M. S. *et al.* (2021) ‘Inequality in plant diversity knowledge and unrecorded plant extinctions: An example from the grasses of Madagascar’, *PLANTS, PEOPLE, PLANET*, 3(1), pp. 45–60. doi: 10.1002/ppp3.10123.

Wandeler, P., Hoeck, P. E. A. and Keller, L. F. (2007) ‘Back to the future: museum specimens in population genetics’, *Trends in Ecology and Evolution*, 22(12), pp. 634–642. doi: 10.1016/j.tree.2007.08.017.

Weberbauer, A. (1936) ‘Phytogeography of the Peruvian Andes’, in *Flora of Peru, Part 1*. Chiacgo, pp. 18–36.

Wickham, H. (2011) ‘The split-apply-combine strategy for data analysis’, *Journal of Statistical Software*, 40(1), pp. 1–29. doi: 10.18637/jss.v040.i01.

Wickham, H. (2016) ‘ggplot2: Elegant Graphics for Data Analysis’. New York: Springer-Verlag. Available at: https://ggplot2.tidyverse.org.

Wickham, H. (2019) ‘stringr: Simple, Consistent Wrappers for Common String Operations.’ Available at: https://cran.r-project.org/packag=stringr.

Wickham, H. *et al.* (2019) ‘Welcome to the Tidyverse’, *Journal of Open Source Software*, 4(43), p. 1686. doi: 10.21105/joss.01686.

Wickham, H. *et al.* (2021) ‘plyr: A Grammar of Data Manipulation’. doi: https://cran.r-project.org/web/packages/dplyr.

Wilke, C. O. (2020) ‘ggtext: Improved Text Rendering Support for “ggplot2”’.

Yang, E. and and Wayne, T. K. (2012) ‘Decolonization is not a metaphor’, *Decolonization: Indigeneity, Education and Society*, 1(1), pp. 1–40.

Zangerl, A. R. and Berenbaum, M. R. (2005) ‘Increase in toxicity of an invasive weed after reassociation with its coevolved herbivore’, *Proceedings of the National Academy of Sciences of the United States of America*, 102(43), pp. 15529–15532. doi: 10.1073/pnas.0507805102.

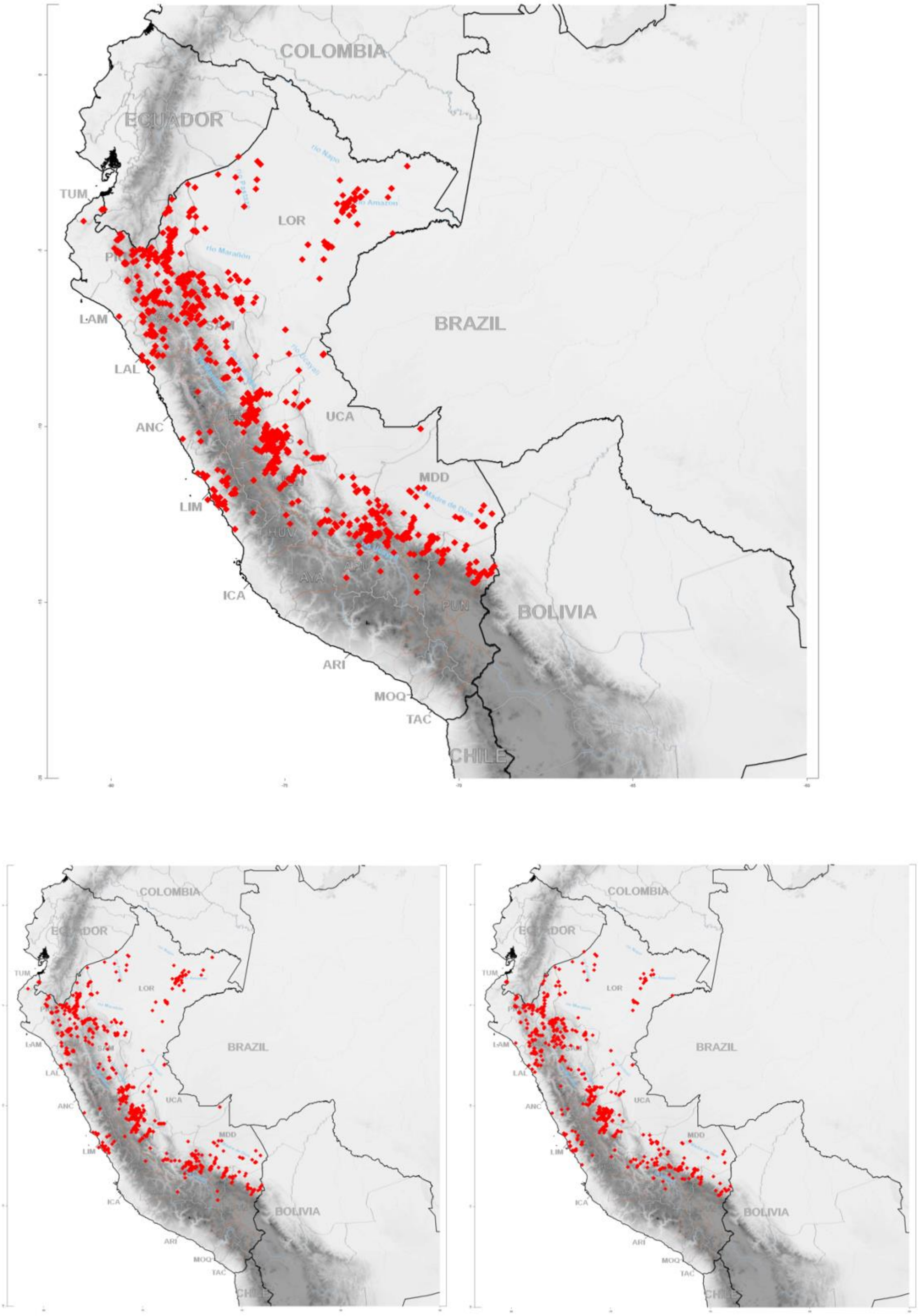
7 Appendix

code	number.staff	total.accessioned	physical.country	commenced.date	no.beg	b.tax.unq	b.geog.unq	b.dup.unq	b.type.per	b.herb.geog.unq	b.rarity	SL
CPUN	6	35000	Peru	1966	52	22.26115385	13.86090385	0.442307692	1.923076923	16.73948077	190290.1923	S
MOL	9	38000	Peru	1948	135	23.17911111	16.94692063	0.571428571	7.407407407	20.82731746	261360.6015	S
AMAZ	2	45000	Peru	1972	4	8.14	2.545	0	0	2.84375	165994	S
CUZ	9	52000	Peru	1936	33	38.00060606	17.97966667	0.444444444	3.03030303	19.86794444	215058.1613	S
HUT	11	60000	Peru	1941	140	31.2825	15.81464444	0.481481481	2.857142857	21.18291852	260824.8433	S
HOXA	8	73000	Peru	2003	210	11.84261905	3.801454545	0.14354067	8.571428571	5.315133971	311467.6827	S
USM	16	800000	Peru	1918	727	24.82833563	15.32104236	0.343419062	3.71389271	20.89709682	320751.2289	L
A	5	0	U.S.A.	1872	3	26.89333333	8.187333333	0	100	8.790333333	75064.33333	S
QCA	5	205000	Ecuador	1971	1	10.31	23.599	1	0	25.717	39132	L
MG	21	209320	Brazil	1895	2	12.295	43.5505	0	100	45.927	313977	L
QCNE	12	250000	Ecuador	1979	1	142.86	6.298	0	0	7.062	NA	L
TCD	3	300000	Ireland	1835	3	34.24	0.516	0	66.66666667	0.540666667	247550.6667	L
MBM	4	412000	Brazil	1965	4	7.555	13.69325	0	0	15.19	602767.5	L
HAL	7	450000	Germany	1812	3	12.44666667	NA	NA	0	NA	77570.33333	L
OXF	7	500000	U.K.	1621	28	16.28321429	4.053	0	10.71428571	4.388058824	281415.2593	L
COL	17	600000	Colombia	1931	2	18.525	20.352	0	0	21.8825	122009.5	L
RB	49	800000	Brazil	1890	4	7.6375	9.043	0.25	25	9.686	385878.75	L
MA	17	1158116	Spain	1755	8	23.88625	7.506	0	0	8.344	174306	L
GB	18	1160000	Sweden	1926	2	63.975	1.6085	0	50	1.7285	17625.5	L
PH	10	1430000	U.S.A.	1812	2	8.365	2.033	0	100	2.197	229325.5	L
BRIT	22	1482000	U.S.A.	1987	22	15.12636364	27.66477273	0.727272727	0	41.29477273	482795.4545	L
UC	15	2100000	U.S.A.	1872	4	69.8225	12.1185	0.25	100	13.22	29827	L
F	17	2700000	U.S.A.	1893	130	30.755	20.46460345	0.086206897	20.76923077	24.56959483	258530.2683	L
E	47	3000000	U.K.	1839	194	42.63427835	11.13426842	0.873684211	8.762886598	15.01767368	206114.8245	L
B	27	3800000	Germany	1815	32	69.8853125	16.02524	0.28	90.625	19.38744	121953.6296	L
S	12	4570000	Sweden	1739	4	37.8775	16.788	0	100	19.4325	427945.6667	L
L	30	5000000	Netherlands	1829	1	6.13	NA	NA	0	NA	221675	L
GH	8	5005000	U.S.A.		20	56.3055	9.573789474	0.105263158	45	10.827	153146.7647	L
US	34	5100000	U.S.A.	1848	32	53.033125	12.89854839	0.032258065	59.375	14.29258065	76147.27586	L
BM	32	5200000	U.K.	1753	25	19.5812	9.109842105	0.157894737	12	10.26442105	298191.9167	L
W	6	5500000	Austria		10	48.678	0.86	0.625	90	1.45	101326.3333	L
G	30	6000000	Switzerland	1824	105	23.78161905	17.65210606	0.090909091	22.85714286	20.30978788	250714.18	L
P	21	6000000	France		33	11.28151515	16.9894	0.333333333	27.27272727	19.3472	231382.75	L
MO	58	6850000	U.S.A.	1859	918	17.08948802	12.72091345	0.30994152	3.159041394	19.65333918	342272.2739	L
NY	36	7921000	U.S.A.	1891	198	17.8689899	14.66611667	0.211111111	11.11111111	17.95305556	330497.8093	L
K	69	8125000	U.K.	1852	165	24.54042424	13.27715873	0.142857143	11.51515152	15.06565873	282752.821	L

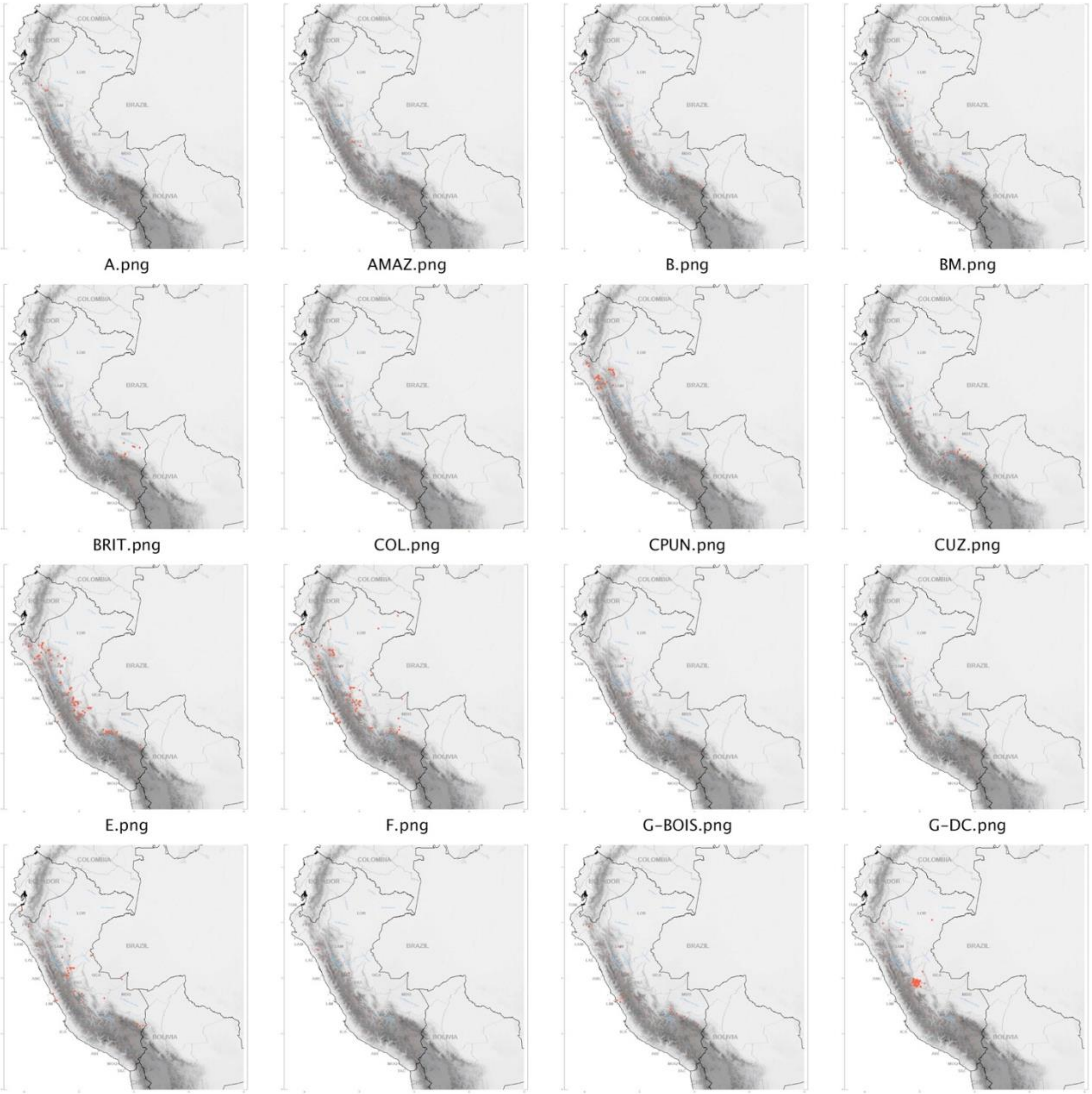
Supplementary Table 1: Herbarium indices displaying (left to right) the herbarium code, number of staff, herbarium accession, herbarium country, herbarium commenced date from IndexHerbariorum (Thiers, 2016); Mean herbarium uniqueness indices of *Begonia* specimens showing: the mean number of *Begonia* specimens, the mean taxonomic uniqueness, the mean geographic uniqueness, the mean number of duplicates, the mean number of type specimens, the mean herbarium geographic uniqueness, the mean rarity uniqueness and size S/L.

code	number.staff	total.accessioned	physical.country	commenced.date	SL	no.sol	s.tax.unq	s.geog.unq	s.dup.unq	s.type.per	s.herb.geog.unq	s.rarity
CPUN	6	35000	Peru	1966	S	270	303.685185	5.76521801	0.36492891	4.81481482	6.688971564	205739.639
MOL	9	38000	Peru	1948	S	818	411.089242	5.54008252	0.17342657	8.92420538	7.118018182	218765.56
CUZ	9	52000	Peru	1936	S	2714	1139.47605	5.56017504	0.47336378	2.65291083	9.712014079	256747.954
CAL	40	2086650	India	1795	L	3	175.333333	0	0	100	0	687297.667
HUT	11	60000	Peru	1941	S	340	160.064706	5.55967606	0.30633803	0.58823529	6.493852113	155133.581
HOXA	8	73000	Peru	2003	S	51	110.078431	1.16109804	0	0	1.362470588	315669.471
USM	16	800000	Peru	1918	L	5498	241.154238	6.69009875	0.38425739	0.5638414	10.14223998	333056.985
A	5	0	U.S.A.	1872	S	11	151.181818	8.221	0	9.09090909	8.978333333	334947.636
QCA	5	205000	Ecuador	1971	L	4	356	5.774	0	0	6.239	212437.25
TEX	26	1006000	U.S.A.	1900	L	7	220.857143	4.07271429	0.14285714	14.2857143	4.424857143	422620
TCD	3	300000	Ireland	1835	L	1	137	0	0	100	0	8568
MBM	4	412000	Brazil	1965	L	1	103	0	0	0	0	702472
OXF	7	500000	U.K.	1621	L	13	211.692308	3.028625	0	23.0769231	3.25725	329694.462
COL	17	600000	Colombia	1931	L	39	151.282051	2.76933333	0	2.56410256	3.112933333	334134.282
RB	49	800000	Brazil	1890	L	9	203.555556	8.458	0	0	9.039	246804.889
MA	17	1158116	Spain	1755	L	164	171.079268	5.93917073	0.41463415	65.2439024	6.958536585	304845.506
GB	18	1160000	Sweden	1926	L	3	298.333333	2.21366667	0	0	2.380333333	84036
Z	15	1500000	Switzerland	1834	L	1	409	0.048	0	0	0.054	175237
DUKE	15	800000	U.S.A.	1932	L	6	66.8333333	28.3915	0.5	0	29.9945	471788.5
LIL	15	720000	Argentina	1931	L	6	92.3333333	14.95	1	0	15.809	140022.667
EA	13	1000000	Kenya	1902	L	1	90	2.083	0	0	2.231	248080
PH	10	1430000	U.S.A.	1812	L	6	540.666667	8.557	0.33333333	0	9.722666667	268640.833
BRIT	22	1482000	U.S.A.	1987	L	1	115	2.118	0	0	2.383	70805
UBC	11	713000	Canada	1912	L	1	77	11.231	1	0	11.738	432318
UC	15	2100000	U.S.A.	1872	L	104	576.711539	5.01425253	0.09090909	4.80769231	5.49230303	229535.279
F	17	2700000	U.S.A.	1893	L	955	562.671204	5.66218561	0.13888889	3.2460733	6.502569444	252442.64
CONN	9	225000	U.S.A.	1898	L	9	15.6666667	5.428125	0.375	44.4444444	6.293	37642
N	9	150000	China	1915	L	1	44	NA	NA	100	NA	55806
HB	9	85000	Brazil	1958	S	1	101	1.234	0	0	1.309	242676
E	47	3000000	U.K.	1839	L	475	191.23579	3.54596279	0.01860465	2.10526316	4.750104651	315910.306
B	27	3800000	Germany	1815	L	21	180.571429	6.27752941	0.29411765	14.2857143	6.994058824	317797.85
IBE	0	235000	U.S.A.	1962	L	1	55	5.154	0	100	5.637	38449
S	12	4570000	Sweden	1739	L	30	281.9	3.92092593	0.11111111	6.66666667	4.242666667	296524.833
L	30	5000000	Netherlands	1829	L	7	417.571429	4.7946	0	14.2857143	5.169	367402.429
GH	8	5005000	U.S.A.		L	421	747.223278	5.5805989	0.07142857	7.36342043	6.225629121	246576.991
US	34	5100000	U.S.A.	1848	L	1313	500.315309	4.17561116	0.09123118	3.27494288	5.054171833	271177.034
ALCB	7	130954	Brazil	1950	L	1	78	14.554	0	0	15.477	230385
BM	32	5200000	U.K.	1753	L	1970	197.966498	4.46466053	0.16621549	1.06598985	5.854463454	318291.067
AK	6	350000	New Zealand	1870	L	1	15	7.813	1	0	8.402	212453
W	6	5500000	Austria		L	56	164.642857	3.73917391	0.17391304	35.7142857	4.958	558709.302
G	30	6000000	Switzerland	1824	L	126	228.857143	5.30928889	0.11111111	30.952381	5.830255556	297192.381
P	21	6000000	France	visit	L	33	212.121212	5.698125	0.5	39.3939394	6.372875	256876.03
MO	58	6850000	U.S.A.	1859	L	3592	211.659521	6.71360918	0.35798534	1.22494432	10.85617086	402179.492
GL	2	45000	U.K.	1780	S	1	137	0	0	100	0	8568
BREM	3	400000	Germany	1865	L	1	144	0	0	100	0.087	34316
NY	36	7921000	U.S.A.	1891	L	926	532.065875	4.59991862	0.11097411	1.07991361	5.189663379	301125.262
EIU	3	84000	U.S.A.	1906	S	1	371	NA	NA	0	NA	807586
LINN	3	33800	U.K.		S	1	155	NA	NA	100	NA	428144
MPU	2	3500000	France	1809	L	1	114	3.881	0	100	4.128	199381
K	69	8125000	U.K.	1852	L	740	780.554054	4.25586983	0.24808576	9.86486487	5.550350689	264817.715
CGG	2	14000	U.K.	1846	S	1	102	0.647	0	0	0.677	43108
IND	1	153605	U.S.A.	1885	L	1	413	NA	NA	0	NA	1253271
DS	1	0	U.S.A.		S	1	409	0.001	1	0	0.001	175237

Supplementary Table 2: Herbarium indices displaying (left to right) the herbarium code, number of staff, herbarium accession, herbarium country, herbarium commenced date from IndexHerbariorum (Thiers, 2016); Mean herbarium uniqueness indices of *Solanum* specimens showing: the mean number of *Solanum* specimens, the mean taxonomic uniqueness, the mean geographic uniqueness, the mean number of duplicates, the mean number of type specimens, the mean herbarium geographic uniqueness, the mean rarity uniqueness and size S/L.



(Top) All georeferences specimen data from Peruvian Begonia and Solanum from Begonia Resource Centre and Solanacea Source; (Bottom Left) All international herbarium specimen data from combined dataset*; (*Bottom Right*)* All local Peruvian specimen data from combined dataset*.*



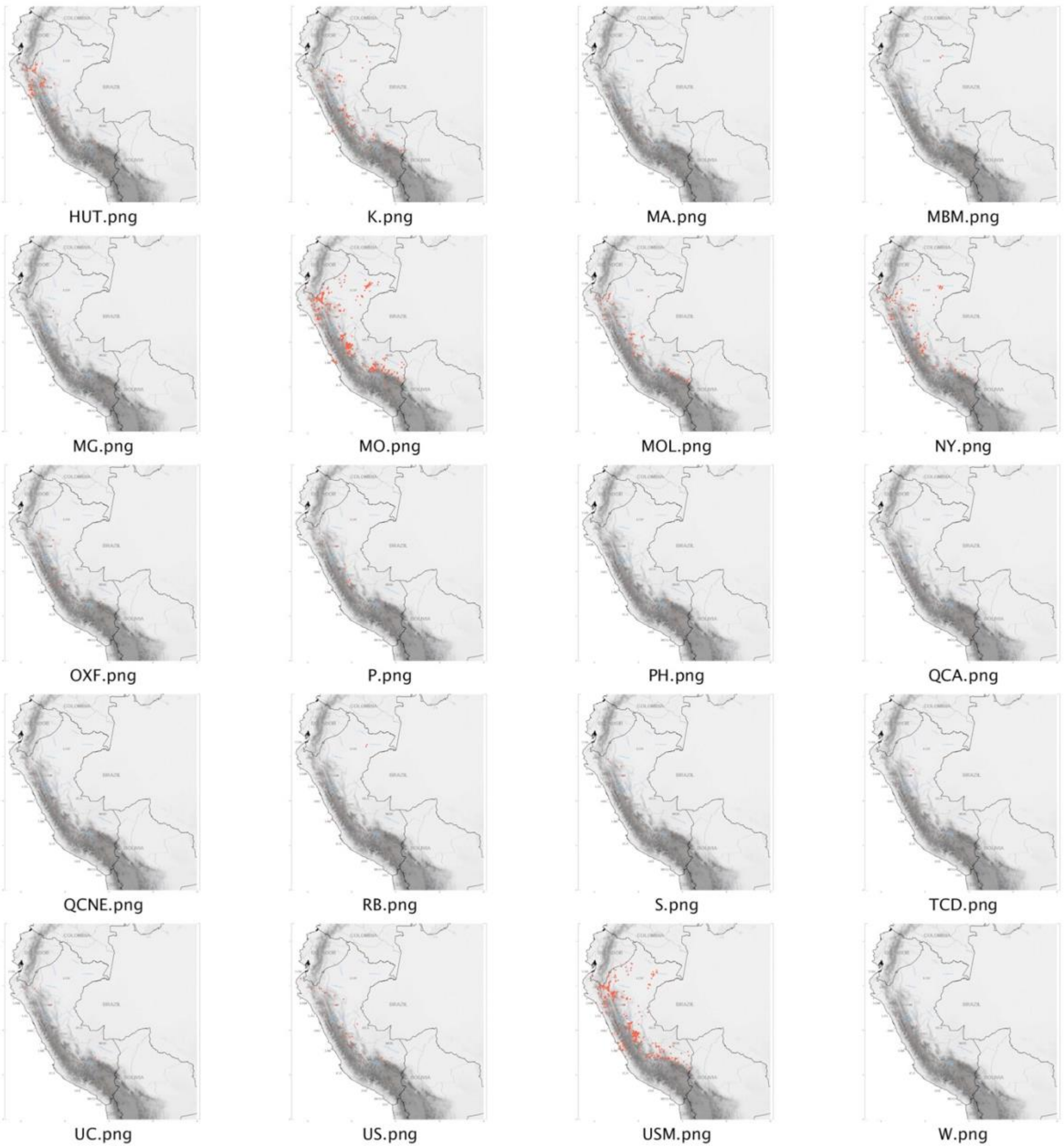
A.png AMAZ.png B.png BM.png

BRIT.png COL.png CPUN.png CUZ.png

E.png F.png G-BOIS.png G-DC.png

G.png GB.png GH.png HOXA.png

Georeferenced specimen data of Peruvian *Begonia* distributed by individual herbaria



HUT.png K.png MA.png MBM.png

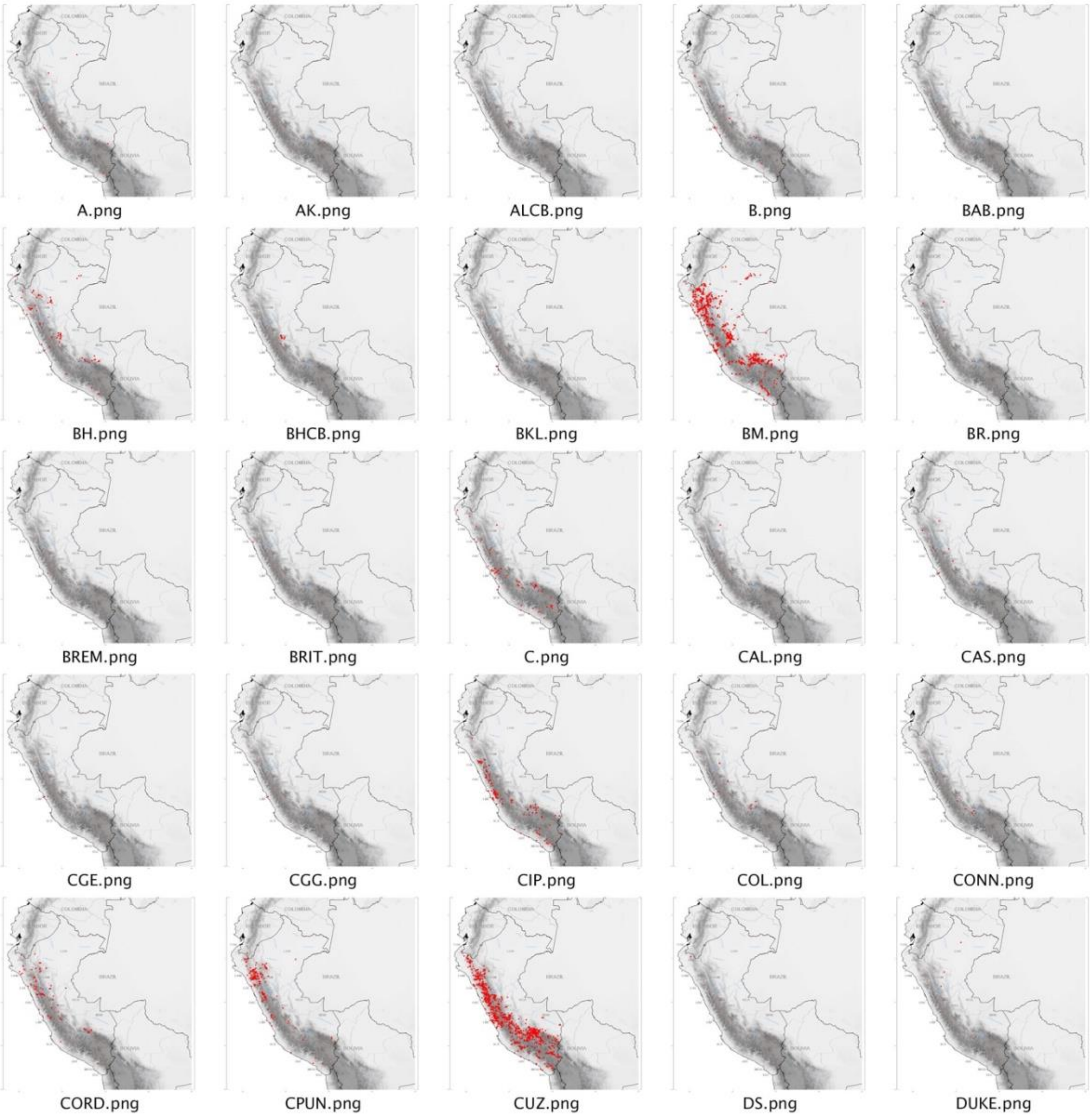
MG.png MO.png MOL.png NY.png

OXF.png P.png PH.png QCA.png

QCNE.png RB.png S.png TCD.png

UC.png US.png USM.png W.png

Georeferenced specimen data of Peruvian *Begonia* distributed by individual herbaria



A.png AK.png ALCB.png B.png BAB.png

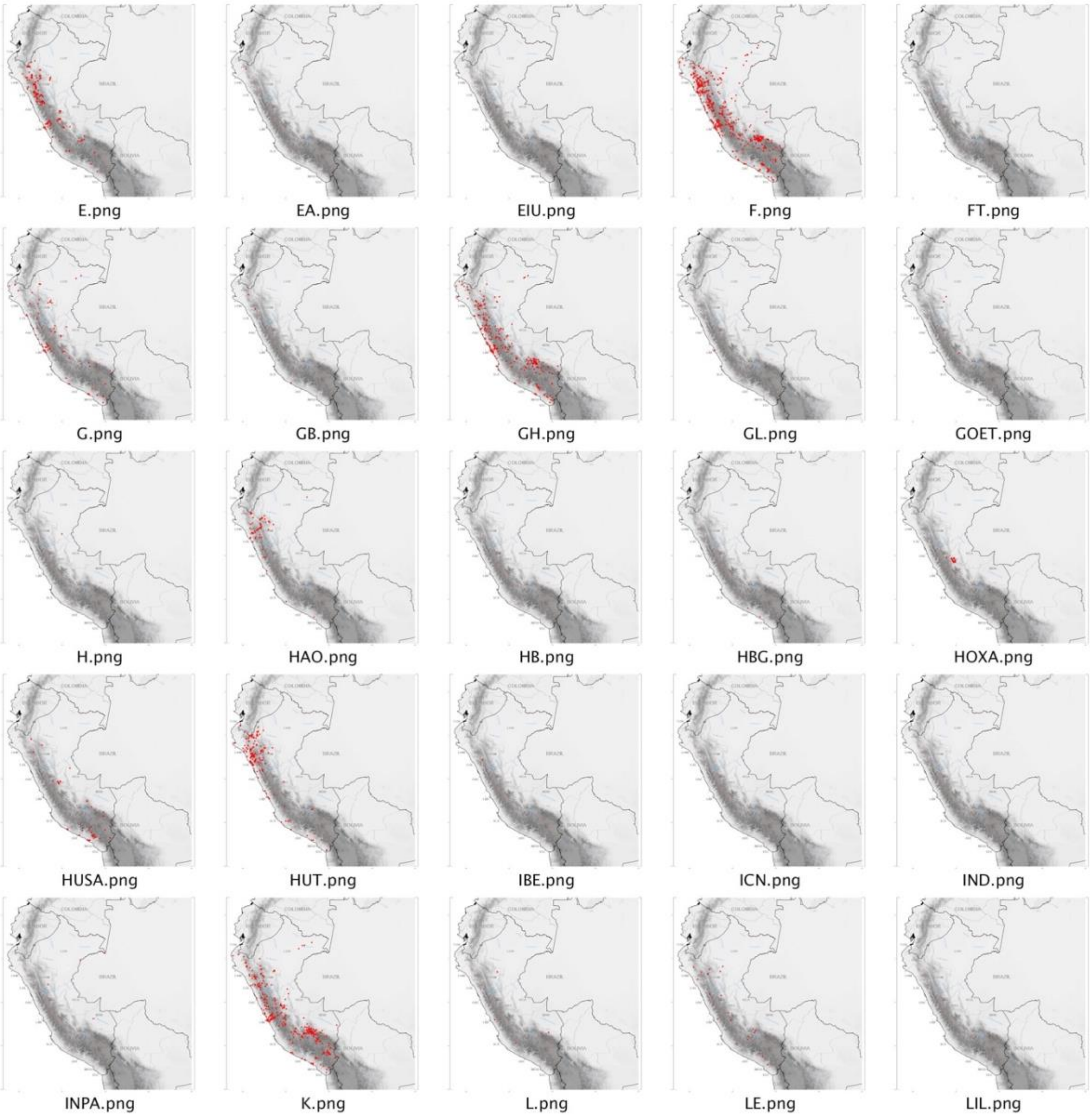
BH.png BHCB.png BKL.png BM.png BR.png

BREM.png BRIT.png C.png CAL.png CAS.png

CGE.png CGG.png CIP.png COL.png CONN.png

CORD.png CPUN.png CUZ.png DS.png DUKE.png

Georeferenced specimen data of Peruvian *Solanum* distributed by individual herbaria



E.png EA.png EIU.png F.png FT.png

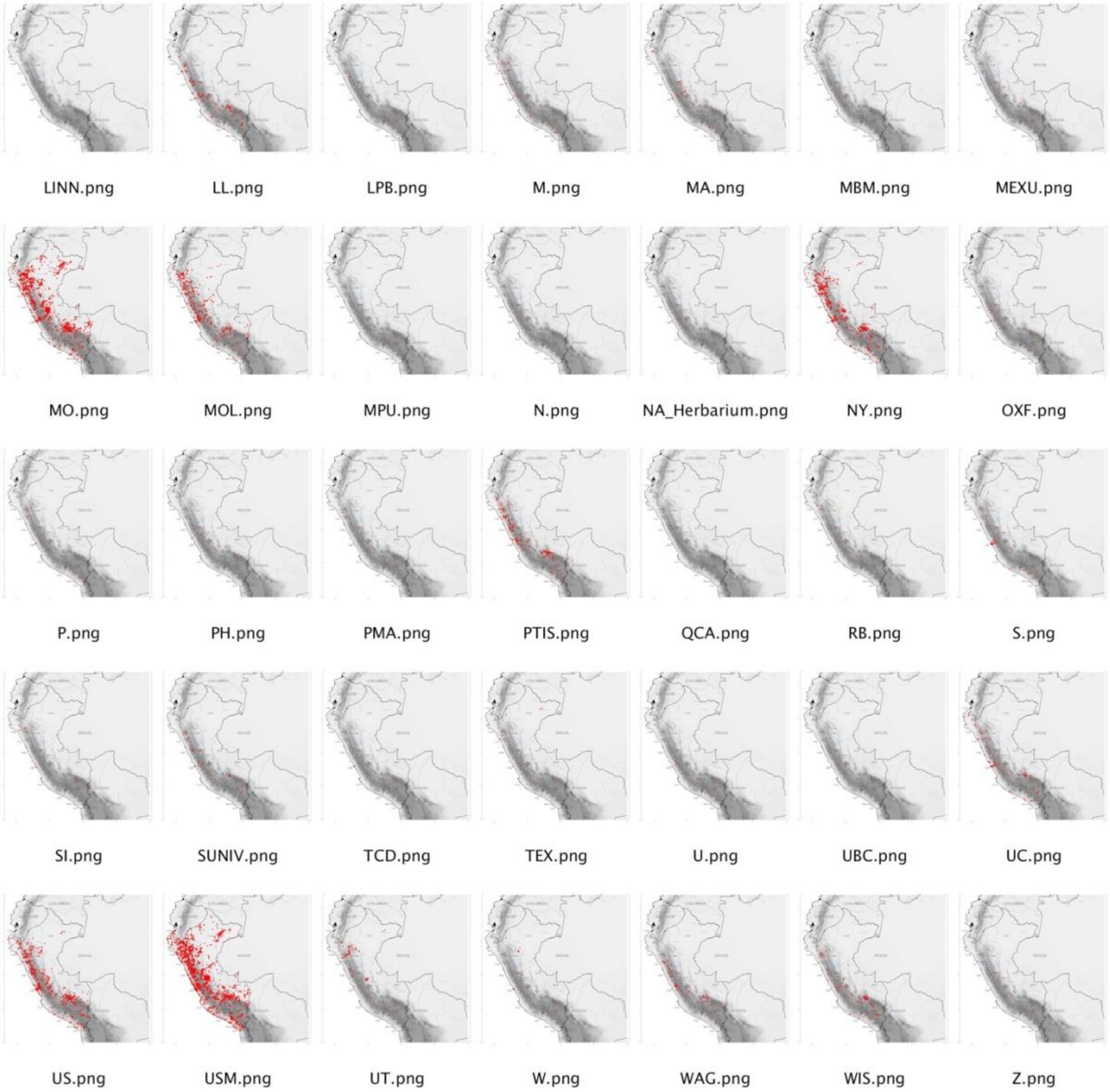
G.png GB.png GH.png GL.png GOET.png

H.png HAO.png HB.png HBG.png HOXA.png

HUSA.png HUT.png IBE.png ICN.png IND.png

INPA.png K.png L.png LE.png LIL.png

Georeferenced specimen data of Peruvian *Solanum* distributed by individual herbaria



LINN.png LL.png LPB.png M.png MA.png MBM.png MEXU.png

MO.png MOL.png MPU.png N.png NA_Herbarium.png NY.png OXF.png

P.png PH.png PMA.png PTIS.png QCA.png RB.png S.png

SI.png SUNIV.png TCD.png TEX.png U.png UBC.png UC.png

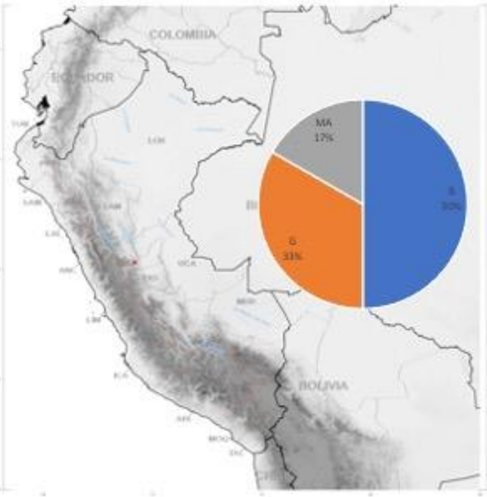
US.png USM.png UT.png W.png WAG.png WIS.png Z.png

Georeferenced specimen data of Peruvian *Solanum* distributed by individual herbaria

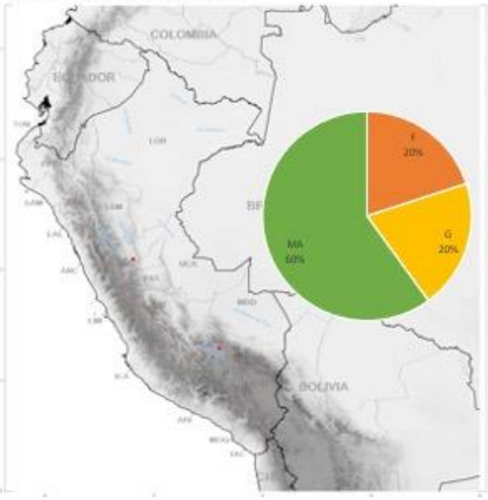
Maps of Collections in Peru across time

1750-1799

Begonia: – 6 specimens

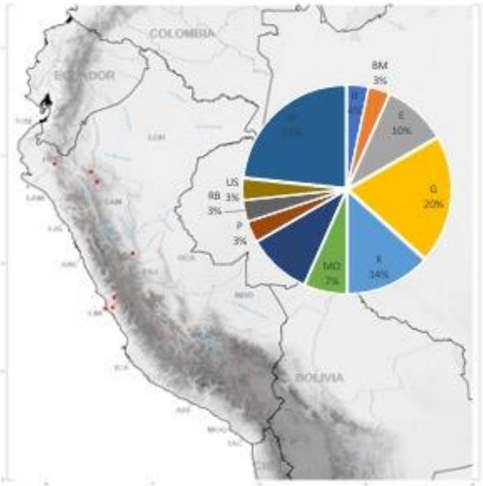


Solanum: - 11 specimens

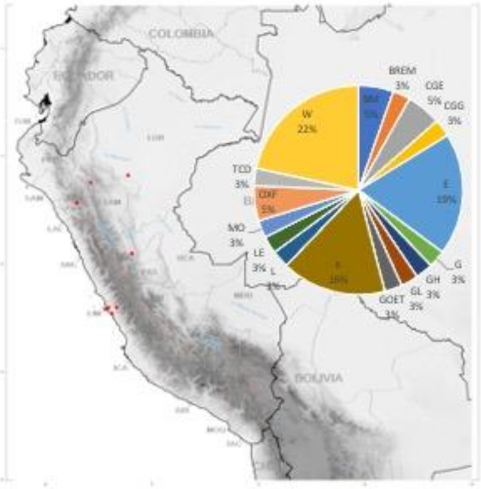


1800-1849

Begonia: - 30 specimens

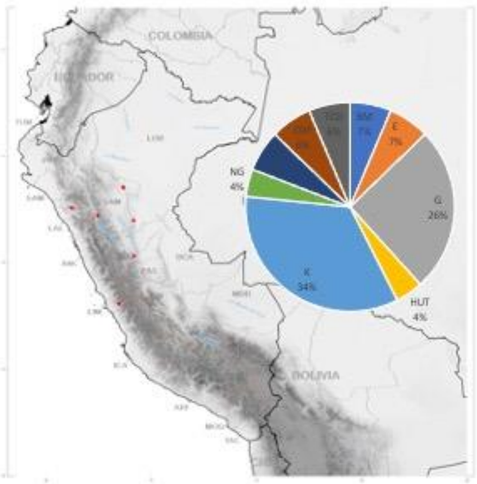


Solanum:- 116 specimens

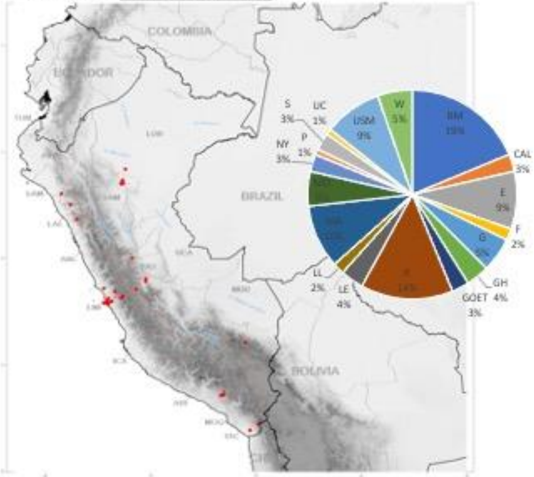


1850-1899

Begonia: – 47 specimens

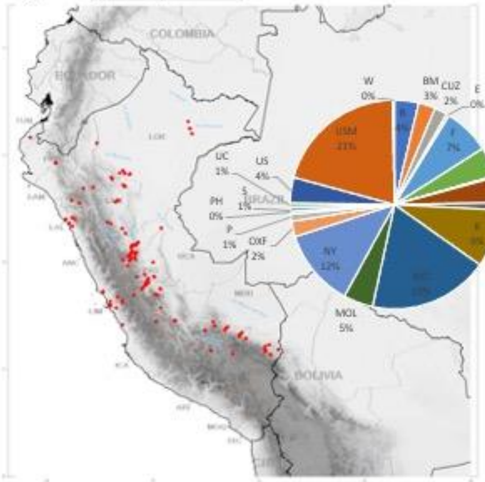


Solanum: – 169 specimens

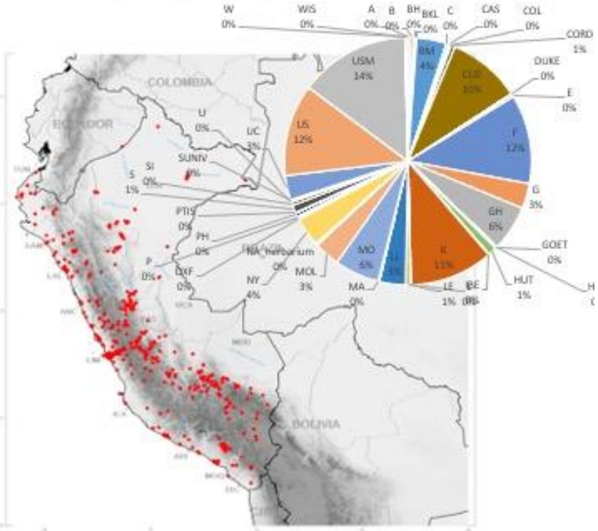


1900-1949

Begonia: – 457 specimens

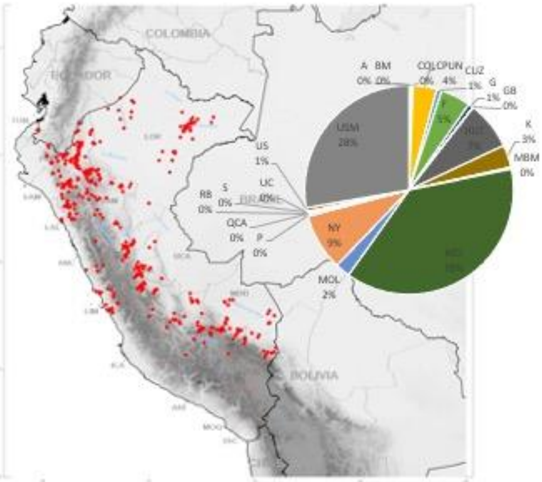


Solanum: – 1704 specimens

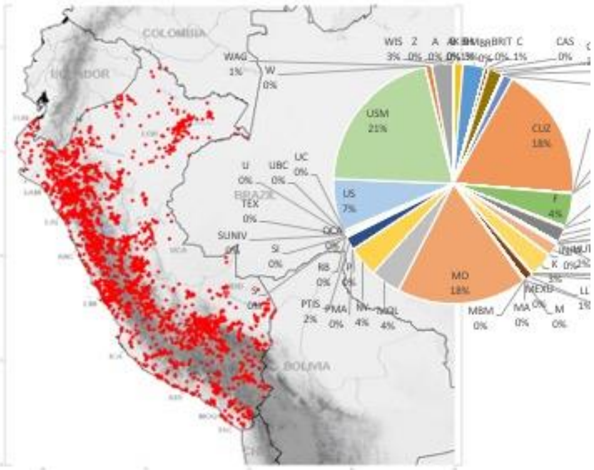


1950-1999

Begonia: – 1298 specimens

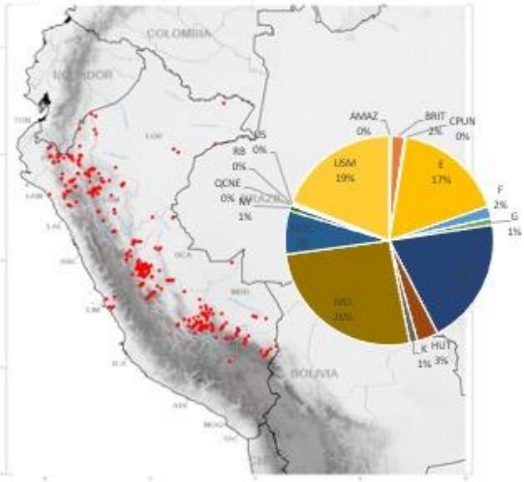


Solanum: - 1950-1999 – 15,038 specimens



2000- 2050

Begonia: – 1065 specimens



Solanum:– 5,424 specimens