Big hitting collectors make massive and disproportionate contribution to the discovery of plant species

Daniel P. Bebber1, Mark A. Carine2, Gerrit Davidse3,
David J. Harris4, Elspeth M. Haston4, Malcolm G. Penn2,
Steve Cafferty2, John R. I. Wood5 and Robert W. Scotland5,*

1Earthwatch Institute, 256 Banbury Road, Oxford OX2 7DE, UK
2Department of Botany, Natural History Museum, Cromwell Road, London SW7 5BD, UK
3Missouri Botanical Garden, PO Box 299, St Louis, MO 63166-0299, USA
4Royal Botanic Garden Edinburgh, 20A Inverleith Row, Edinburgh EH3 3LR, UK
5Department of Plant Sciences, University of Oxford, South Parks Road, Oxford OX1 3RB, UK

Discovering biological diversity is a fundamental goal—made urgent by the alarmingly high rate of extinction. We have compiled information from more than 100 000 type specimens to quantify the role of collectors in the discovery of plant diversity. Our results show that more than half of all type specimens were collected by less than 2 per cent of collectors. This highly skewed pattern has persisted through time. We demonstrate that a number of attributes are associated with prolific plant collectors: a long career with increasing productivity and experience in several countries and plant families. These results imply that funding a small number of expert plant collectors in the right geographical locations should be an important element in any effective strategy to find undiscovered plant species and complete the inventory of the world flora.

Keywords: type specimen; plant collectors; discovery process

1. INTRODUCTION
Enhanced understanding of how the natural world is discovered and monitored has recently become possible because of the increased number of electronic initiatives that are providing extensive amounts of specimen data from the world’s natural history collections, which have hitherto been largely inaccessible [1–6]. The extent and magnitude of these data provide an opportunity to ask novel questions that have not been possible until now.

Plants, especially, are pivotal organisms for monitoring and measuring global biodiversity because they comprise a species-rich component of almost all habitats on the earth [7] and are relatively well known. Nevertheless, 15–30 per cent of flowering plants remain to be discovered [8–10]. A key question for policy-makers and taxonomic institutions is how the remaining species can be efficiently discovered? To answer this question, sound knowledge and a good scientific understanding of the processes that lead to the discovery of new species is required.

The process of species discovery comprises many distinct elements, but includes three pivotal stages: collection of specimens, recognition of species and publication of species [8]. Plant collecting is one aspect of the discovery process that has been much written about, especially the role of the great botanical explorers such as Sir Joseph Banks [11] or Alexander von Humboldt [12]. These pioneers form an integral part of the story of the discovery, colonization and exploitation of the world by Europeans in their search for food, beverages, spices and new garden plants as well as knowledge for its own intrinsic value [13]. The portrayal of these so-called ‘great plant hunters’ and their role in discovering new and economically important exotic plants is relatively well documented and forms part of an extensive literature related to the history of colonial expansion and botanical discovery [11,13–18]. Although often discussed in the context of regional and country flora projects [19], the dynamics of plant collecting in relation to our knowledge of how new species are discovered has not been quantified from a global perspective. This has recently become possible owing to funding from the Andrew W. Mellon Foundation, and subsequent institutional commitment to database type specimens and deposit these data in a central repository such as JSTOR Plant Science [2]. Here, we focus on the quantitative contribution that individual collectors have made for the discovery of plant diversity. We use the number of type specimens associated with each collector as a proxy for the overall contribution of that collector to the discovery process.

2. MATERIAL AND METHODS

(a) Data
We assembled four datasets of plant collectors, and the type specimens of flowering plants associated with each of those collectors that are housed at four major taxonomic institutions: Missouri Botanical Garden, Natural History Museum London (NHM), Royal Botanic Garden Edinburgh, and the Royal Botanic Gardens Melbourne. For each...
Institutional database, the largest task was to standardize how each collector was recorded, as many collectors were recorded in various formats, which had to be identified and made consistent. We excluded: all types of infraspecific names; any with collector = Anon; any with type status as follows: type? Chironomus, Chirotonus, Co-type, Epitype, Neotype, Original material, Paralectotype, Paratype; we also excluded 4000 Linnean types housed at NHM. For each institution, we included only one type specimen per collector associated with each binomial, to reduce redundancy that would be caused by including all isotypes and syntypes. For situations where there were types of the same name in different institutions, we counted only one. For the analyses that needed dates, we estimated missing dates for some collectors using the median date from the range of dates from the specimens with dates of that collector. The final datasets for the four institutions comprised 50,398, 25,116, 16,700 and 10,741 type specimens from NHM, Missouri, Edinburgh and Melbourne, respectively, with a combined total of 102,955. The cleaned up data reported in this paper are tabulated in the electronic supplementary material, available on request from the four institutions and archived in their original format at JSTOR Plant Science.

(b) Analyses
To analyse change in the frequency distribution of types per collector over time, the log of the number of types per collector was regressed against the median date per collector. The regression for geographical specialization (fraction of types collected in the most-collected country) excluded collectors with one type. Data for plant diversity in different countries were taken from Pitman & Jørgensen [20]. Genera were matched to families using data from Mabberley [21]. Kernel density was calculated using the density function for R v. 2.13.0 [22].

3. RESULTS AND DISCUSSION
The history of plant discovery is dominated by a small number of highly productive collectors. In all four institutions, the distribution of types per collector (T) was highly skewed, with around 2 per cent of collectors responsible for half the type specimens, whereas around half the collectors contributed only to a single type (figure 1). We term these top approximately 2 per cent of collectors the ‘Big Hitters’. The frequency distribution was remarkably similar across all datasets (figure 1) and has been more or less conserved through time (figure 2). However, closer analysis shows that the mean T has declined slightly but significantly over time (electronic supplementary material, table S1), i.e. the relative importance of Big Hitters, in recent times, is less than in the past. Depending on the institution, the reduced importance of Big Hitters can be due to both a reduction in T and to an increase in the number of collectors contributing a small number of types (figure 2). As a whole, however, T is markedly skewed.

A number of traits are associated with the prolific collection of types. T increases with duration of collection, as measured by the interval between the earliest and latest type per collector (electronic supplementary material, table S1). Median T for collectors active for a decade or more ranged from 9 to 18, compared with one for those active for less than a decade (electronic supplementary material, table S1). However, this increase in species discovery with experience is nonlinear (figure 3), such that the rate of species discovery per year of collection activity increases with T (electronic supplementary material, figure S1), but decreases with the duration of collection (figure 4). Big Hitters, therefore, collect more efficiently, but this efficiency cannot be explained solely by time spent in the field. The importance of experience is illustrated by the pattern of collection within the careers of individual collectors. Despite great variability, there is a significant increase in the skewness of the collection—year distribution (i.e. more specimens are collected towards the end of the career; see electronic supplementary material, table S1) with T. Big Hitters, therefore, collect for longer—collect more rapidly—and get better at collecting new species as they go. This increased efficiency is probably related to increased knowledge of the flora in the areas they are collecting as well as increased taxonomic knowledge.

Geography plays some part in explaining the success of Big Hitters. There was no evidence that Big Hitters were more likely to visit more botanically diverse countries than
Small Hitters, as there was no significant correlation between $T$ and the average plant diversity of countries from which types were collected, except for the NHM collection (electronic supplementary material, table S1). However, $T$ was significantly correlated with total number of countries from which those types originated (electronic supplementary material, table S1). While Big Hitters tend to visit more countries, the vast majority of their types come from a single country (electronic supplementary material, tables S1 and S2). Big Hitters, while not focusing on high-diversity countries any more than Small Hitters, collect from more countries but specialize on a particular country.

With few exceptions, Big Hitters were taxonomically cosmopolitan, the number of plant families increasing with $T$ (electronic supplementary material, table S1). Some collectors, such as Carl Luer who has contributed hundreds of new species of orchid to the Missouri collection, are highly specialized, but this is unusual. As with geographical specialization, although Big Hitters collected from several families, the most-collected family per Big Hitter comprised a large fraction of types (electronic supplementary material, table S1).

To summarize, Big Hitters are distinguished by five attributes: they collect over many years, they collect more types per year, they collect from several different countries although specializing in one particular country, they collect from a wide range of plant families (although again, often specializing in a particular family) and they collect more types towards the end of their careers. Conversely, they are no more likely to collect from countries with high plant diversity than Small Hitters.

4. CONCLUSIONS
These observations point to the critical role of expertise gained from many years spent in the field as a determinant of a productive plant-hunting career. Our results demonstrate that prolific plant collectors, as measured by the number of associated type specimens, are not restricted in time to the eighteenth and nineteenth centuries but are equally a phenomenon of contemporary botany, for example, Peter Davis, a big hitter from Edinburgh, collected his first type specimen from Turkey, which was even then considered relatively well known, in 1938. Over the next 38 years, he collected a further 348 types—the vast majority from Turkey. Davis built up a wealth of experience in both the geography and plants of Turkey, which enabled him to identify and collect many new species. Today, it is often much harder to built up this kind of expertise, partly because funding agencies provide grants only for a limited period and often turn down applications from experienced workers to favour newer applicants. Experienced collectors may not collect as many numbers as a novice collector but they will use their experience and accumulated botanical competence to be selective so increasing the number of new species found. This explains why they will be more successful in finding novelties towards the end of their career. Our results also demonstrate, for the first time,
that many productive plant collectors are not necessarily those who first botanize a particular region. Although this ‘first to collect in an area/country’ effect explains the large number of types associated with, for example, Robert Brown in Australia, the majority of big hitters collected types over many years. Since 2003, several authors on this paper have been part of capacity-building projects involving plant collecting in Bolivia, funded by the UK Darwin initiative. Our experience during that project suggests that there are relatively few people who develop the key skills and interests for really productive plant collecting and our strategy has been to identify these individuals, encourage them and offer modest funding to help them work. Overall, our message is that plant collecting is a specific part of the discovery process that demands experience and skill that seems to be developed by certain individuals over a considerable length of time. We consider that efforts should be made to identify, train and fund these individuals throughout their career once they have demonstrated their motivation and capacity as they can make a substantial contribution to discovering new species. The rapid loss of global biodiversity [23] and the perceived ‘taxonomic bottleneck’ owing to declining numbers of professional taxonomists [3,8,24] has led to calls for a massive increase in the utilization of non-professionals in the plant discovery process [25]. These non-professionals comprise several distinct classes, including parataxonomists (local resident, field-based, biodiversity inventory specialists working with professional taxonomists), citizen scientists (interested members of the public) and students. While parataxonomists tend to spend many years in the field and work closely with professionals, thereby developing expertise in the local biota [26,27] students are likely to be involved in collecting for much shorter periods and have limited training [28]. Citizen scientists are mostly found in developed countries and generally lack time, resources or permits to collect from geographically remote areas where unknown plants are likely to be found in any number. Our results imply that over reliance on students, parataxonomists and citizen scientists [25] to the neglect of experienced plant collectors may not entirely yield the hoped for results. Recent research has demonstrated that half of the undiscovered plant species are already in herbaria [8], and also that more undiscovered plant species are to be collected from biodiversity hotspots [29]. These findings, alongside the results presented here, provide important insights into our understanding of the discovery process that can inform policy-makers and funding agencies in any future initiatives to design the optimal strategy to complete the inventory of the world flora. We consider that the optimal strategy should specifically examine the role of specialization and expertise. Much can be said for urging collaborative teams to work together, with collectors, data workers, morphologists and molecular geneticists each contributing important knowledge of

Figure 3. Types by duration of collecting activity. (a) Edinburgh. (b) Melbourne. (c) Missouri. (d) NHM. Bold line and broken lines show best-fit standard error for power regressions. Narrow line shows predicted value where exponent = 1. Exponents for the data were significantly less than one (electronic supplementary material, table S1).
plants (and animals) in proportion to their interests, aptitudes, passions and inclinations.

We acknowledge Alison Vaughan and the Royal Botanic Gardens Melbourne MELISR database and Teri Bilsborrow, Jeany Davidse and Jay Paige at the Missouri Botanical Garden. Thanks to Kevin Gaston and two anonymous reviewers for comments on an earlier draft of this manuscript. D.P.B. is funded by HSBC Climate Partnership, and R.W.S. and J.R.I.W. acknowledge funding from the UK Darwin initiative. The data reported in this paper were made publicly available at JSTOR Plant Sciences by funding from the Andrew W. Mellon Foundation.

REFERENCES


22 R Development Core Team. 2011 R: a language and environment for statistical computing. Vienna, Austria.


