

DOI: 10.11931/guihaia.gxzw202209066

刘慧圆, 覃海宁, 包伯坚, 等, 2022. 中国高等植物数字化标本分析 [J]. 广西植物, 42(增刊 1): 29–45.

LIU HY, QIN HN, BAO BJ, et al., 2022. Analysis of digitized specimens of higher plants in China [J]. *Guihaia*, 42(Suppl. 1): 29–45.



中国高等植物数字化标本分析

刘慧圆^{1,2}, 覃海宁², 包伯坚², 陈天翔², 韩国霞², 刘全儒^{1*}

(1. 北京师范大学 生命科学学院, 北京 100875; 2. 中国科学院植物研究所 系统与进化植物学国家重点实验室, 北京 100093)

摘要: 传统上馆藏标本, 主要用于植物分类学、植物资源学的研究。数字标本的出现将标本的使用拓展到从研究生物多样性时间空间分布到生态学和进化学理论、生物多样性保护、农业和人类健康等广泛领域。截至目前, 从互联网上获取的采自中国的植物标本数量已有1 200多万份。该文通过整理和分析这些数据以了解中国植物标本的数字化精度、采集时间和采集地区规律以及采集空缺等状况。结果表明: 中国标本采集形成了4个高峰, 即20世纪30年代、60年代、80年代和21世纪初, 中国植物标本采集和研究工作主要在20世纪50年代后由中国学者完成。标本采集地区覆盖度在省级较好, 县级标本采集则很不平衡; 标本采集类群在科属层面覆盖率高, 但近五分之一的物种采集不足; 标本的采集量既与植物分布幅度相关, 也与采集地区的知名度、所获科研项目及采集者偏好有关。未来中国植物标本数字化方向应该在继续挖掘馆藏标本的同时, 一方面开展对现有数字化标本信息再审核及补充, 并加强与欧美大馆的信息共享以获取早期历史标本信息; 另一方面应用数字化标本信息分析结果, 指导境内标本的精准采集, 包括采集薄弱/空白地区、采集薄弱/空白属种的采集, 以进一步增强实体标本馆能力, 提高数字化标本质量, 为进一步完善植物标本数字化和精准化采集提供依据, 更好地服务科学和社会的发展。

关键词: 植物标本, 标本数字化, 采集空缺, 精准采集, 高等植物

中图分类号: Q94 **文献标识码:** A **文章编号:** 1000-3142(2022)增刊 1-0029-17

Analysis of digitized specimens of higher plants in China

LIU Huiyuan^{1,2}, QIN Haining², BAO Bojian², CHEN Tianxiang²,
HAN Guoxia², LIU Quanru^{1*}

(1. *College of Life Sciences, Beijing Normal University, Beijing 100875, China*; 2. *State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China*)

Abstract: Traditionally, herbarium specimens are mainly used for plant taxonomy and resource science research. The emergence of digitized specimens has expanded the use of specimens to a wider range of fields, from the study of temporal and spatial distribution of biodiversity to theoretical ecology and evolutionary, biodiversity conservation,

收稿日期: 2022-09-06

基金项目: 国家自然科学基金(31770213); 国家植物标本资源库(E0117G1001)。

第一作者: 刘慧圆(1982-), 博士, 工程师, 主要从事植物地理、生物多样性保护和科学数据研究, (E-mail) liuhy@ibcas.ac.cn。

*通信作者: 刘全儒, 教授, 主要从事植物分类学和植物地理学研究, (E-mail) liuquanru@bnu.edu.cn。

agriculture and human health. Up to now, information of more than 12 million plant specimen sheets collected from China has been available on the internet. This paper compiles and analyzes these data to evaluate the digitization accuracy, the temporal and spatial pattern of herbarium collection, and the collection vacancy of Chinese plant specimens, so as to provide a basis for further improving the digitization and accurate collection of plant specimens. The results are as follows: The collection of specimens in China formed four peaks, namely, the 1930s, the 1960s, the 1980s and the early 21st century, and the collection and research of Chinese plant specimens have been mainly completed by Chinese scholars since the 1950s. The coverage of specimen collection is rather high on province level while greatly unbalanced on county level. The coverage of specimen collection by plant groups is high on family and genus level, but nearly one fifth of the species have not been collected sufficiently. The number of collected specimens is related to the distribution range of plants, as well as the popularity of the collection area, the scientific research projects, and the preferences of the collectors. To better digitize Chinese plant specimens in the future, we should continue to explore the undigitized herbarium specimens, review and supplement the existing digitized specimen information, and strengthen the information sharing with European and American herbaria to obtain the information of early historical specimens. The results from our analysis can be applied as a guidance for accurate collection of specimens in China, showing the weak or blank areas and genera/species, which can further enhance the capacity of the herbaria and improve the quality of digitized specimens to better serve science and the society.

Key words: plant specimens, specimen digitization, collection vacancy, accurate collection, higher plants

各类标本馆和博物馆中保藏的生物标本包括植物标本,是物种在地球上存在的历史凭证,西方文献常常称其为自然历史标本(natural history collections)。传统上,馆藏标本主要用途是作为分类学、系统学及相关学科如植物资源学等的研究素材(Lane 1996; 王利松等 2010; 杨永 2012)。但是,随着大量数字化标本的出现和全球范围内生物多样性保护的需求,生物标本数据广泛应用于系统学、生物地理学、生态学、物候学和基因组学等众多学科领域,在农业发展、人类健康、生物多样性保护政策制定等社会应用领域中发挥着重要作用(Lane 1996; Ponder et al., 2001; O'Connell et al., 2004; Johnson et al., 2011; Robbirt et al., 2011; Wen 2015; Page et al., 2015)。标本数字化工作始于 20 世纪 70 年代末 80 年代初欧美,至 90 年代奠定了基础。直到 20 世纪初,全球性的生物多样性保护热潮让人们认识到各国各类博物馆、标本馆中保藏的标本是生物多样性监测及保护策略制定所参考的重要且易于获取的材料(Ponder et al., 2001; Suarez & Tsutsui 2004; Pyke & Ehrlich 2010; Page et al., 2015; Nualart et al., 2017; James et al., 2018; Hedrick, 2020)。此后,借助蓬勃发展的计算机技术及网络通信技术,欧美发达国家及南非(<https://www.sanbi.ac.za/>)、巴西([\[taPublicoHVUC.do\]\(http://taPublicoHVUC.do\)\)等许多国家纷纷加速推进本国标本数字化进程,并支持建立 GBIF \(<http://www.gbif.org/>\) 和 JSTOR \(<http://plants.jstor.org/>\) 等全球性标本信息共享网站。](http://reflora.jbrj.gov.br/reflora/herbarioVirtual/ConsultaPublicoHVUC/Consul</p></div><div data-bbox=)

依据《国际植物标本馆索引》(*Index Herbariorum*),全球共有 176 个国家 3 522 个标本馆,馆藏标本总量达 3.9 亿份,其中馆藏量前五位的国家为美国(7 846 万)、法国(2 404 万)、英国(2 365 万)、德国(2 212 万)和中国(2 037 万)(Thiers, 2021)。中国是唯一馆藏量主要来自本国的国家。早在 20 世纪 80 年代初期,我国科技部、中国科学院等部门及生物科学工作者就已了解到国际生物标本馆研究领域的数字化趋势,着手开展与国际最新动向同步的生物物种和标本信息数字化的研究。其中,实施力度最大、影响最远的是科技部国家科技基础条件平台项目——原国家标本资源共享平台。国家标本资源共享平台(缩写为 NSII, <http://www.nsii.org.cn>),由中国科学院植物研究所牵头组建,自 2004 年启动以来,通过旗下植物子平台(共享网站为“中国数字植物标本馆”,缩写为 CVH, <http://www.cvh.ac.cn>)等相关子平台,共完成全国 100 多家植物标本馆共 1 000 万份标本的规范化整理及数字化表达,并全部实现网络共享(马克平等 2010; 许哲平等 2010)。该项工作(网站)成为国内外用户查询中

国植物标本及相关植物学信息的重要门户网站,在世界同行中享有广泛的影响。

至今,已有多篇论文分析中国数字化植物标本所覆盖的植物类群、时空情况等。阳文静(2013)对 900 万条(包括 650 万份标本记录和 250 万条文献记录)物种分布数据进行了去重,得到 431 万条县级植物分布记录,并借此对中国植物采集的地理格局和成因进行了详细分析。姜承勇等(2018)对 216 万份标本记录进行了采集时间进程和省份采集情况分析,并根据当前采集信息,预测新疆、西藏地区植物标本采集空间较大。桂略宁等(2018)对 1 053 万份标本记录进行了省份采集情况分析,认为不同地区的采集密度呈现非均质性。尹朝露等(2018)和王凯莉等(2018)则分别分析兰科和蔷薇科数字标本信息,揭示相关科的标本采集分布特点。张玉雪等(2018)利用 14.9 万份杜鹃花科标本信息分析了中国杜鹃花科植物物种丰富度的分布格局及影响其分布的气候因子之间的关系,结果显示杜鹃花科植物标本采集较为完整的县级行政区集中分布于长江以南地区以及西南地区,东北、西北及东部地区相对较少。Qian 等(2018)对从 GBIF 和 NSII 获取的共 1 100 万份中国标本信息与物种编目的完整性进行比较,认为标本数据库的完整程度对物种编目的完整性影响非常大。这些工作从多个层面分析和揭示了中国植物数字化标本所蕴藏的信息和规律,但仍然缺乏从标本数据的层面对标本采集的时空和类群进行全面深入分析与研究。

本文基于国内外共享平台获取的中国植物数字标本的海量信息,对这些标本信息在时间、空间、类群等方面积累程度进行分析,以了解中国植物标本在采集时间、空间、类群等方面的特点及采集空缺,为未来数字化方向及开展标本精准化采集提供方向性建议及策略。

1 数字标本来源与分析方法

1.1 数字标本来源

数字标本数据来源于 CVH、NSII 两个标本共享平台,其中 NSII 只选取了教学子平台和保护区子平台的数据。同时补充了 JSTOR 和 GBIF 标本数据网络共享平台上采自中国的标本,共提取了 12 464 469 条采自中国境内的高等植物标本信息

(表 1)。所收集标本采集时间跨度为 1700—2020 年。

表 1 中国数字化植物标本数据来源

Table 1 Sources of digitized Chinese plant specimen data

数据来源 Source	网址 Website	记录数 Number of records	标准化记录数 Number of standardized records
CVH	http://www.cvh.ac.cn	7 701 464	6 166 124
NSII	http://www.nsii.org.cn	2 449 694	1 976 144
GBIF & JSTOR	http://plants.jstor.org , http://www.gbif.org	2 313 311	744 843
总计 Total		12 464 469	9 906 121

1.2 数据的核准及清理

由于数据来源不一,故在进行分析之前,首先对标本数据进行了核准和整理工作:(1)删除标本重复记录;(2)规范各个字段内容格式;(3)通过 TNRS 网站(<https://tnrs.biendata.org/>),更正与核查提取的拉丁学名等。

1.3 标本数据库植物名称及县级采集地名处理

使用《中国生物物种名录(光盘版 2017)》(中国科学院生物多样性委员会,2017)(简称《生物物种名录》)作为标准库对标本数据的植物拉丁学名、属名及科名进行比对,包括接受名和异名的匹配转化和标准化。《生物物种名录》中各类群使用的分类系统如下:被子植物为 APG III 分类系统(Chase & Reveal, 2009)、裸子植物为克氏系统(Christenhusz et al., 2011)、蕨类植物根据 *Flora of China*(Wu et al., 2013a)的系统及类群处理意见、苔藓植物根据 Frey (2009)主编的 *Syllabus of Plant Families* (Part 3 Bryophytes and seedless Vascular Plants)。匹配处理得到的物种名称包括 462 科 3 929 属 32 815 种,以及种下等级 5 048 种。

对于未符合目前行政区划的旧县级名称,通过搜集获得原中国科学院中国植物编委会编辑的各省区地名参考资料(油印本),《中国地名录》(国家测绘局地名研究所,1983),《云南植物采集史略》(包士英等,1995),《中国行政区划沿革手册》(陈潮,2007)等公开或未公开发表文献,并参考了中华人民共和国民政部(<http://www.mca.gov.cn/article/sj/xzqh/>)的各年行政区划沿革信息来进行甄别,并对采集地名地标到现用县级单位。

共标准化和校正 9.8 万条县名,涉及省名 5.1 万笔,植物名称 21.76 万笔。其中学名归属:属名 1.1 万笔,科名 0.15 万笔。最终得到有效标本 9 906 121 份(占收集标本量 84.3%),成为本研究使用的中国植物标本数据库(简称标本数据库)。标本数据库内容包括常规字段,如标本采集人、采集时间、采集号、省份、市县、鉴定名称、接受名、属、科、生境、海拔、经纬度等信息。标本数据库覆盖中国包括港澳台在内的 34 个省(区、市)以及 2 217 个县,占全国行政区划 2 377 个县的 93%(国家基础地理信息系统 http://nfgis.nsd.gov.cn/nfgis/chinese/c_xz.htm)。

1.4 物种采集强度为薄弱的界定

本文把标本量少于 5 份(标本份数 ≤ 4) 的物种定义为采集薄弱物种。

1.5 县级物种采集完整程度分析

利用物种累积曲线(species accumulation curve, SAC)的弯曲程度来表示不同地理单元的物种采集完整程度,(Gotelli & Colwell, 2001; 阳文静, 2013)。其计算方法为对每个地理单元中的样本进行重复抽样,抽取 N 个个体,计算这 N 个个体所包含的平均物种数量。通过计算,抽取标本量达到 90% 物种的标本比例开始计算随标本数增加的物种累计曲线的平均斜率,即每增加一份标本,平均增加的物种数,来表示该地理单元的取样完整程度。斜率接近 0 时,表明该地理单元的取样趋于完整;斜率接近 1 时,表明该地理单元取样

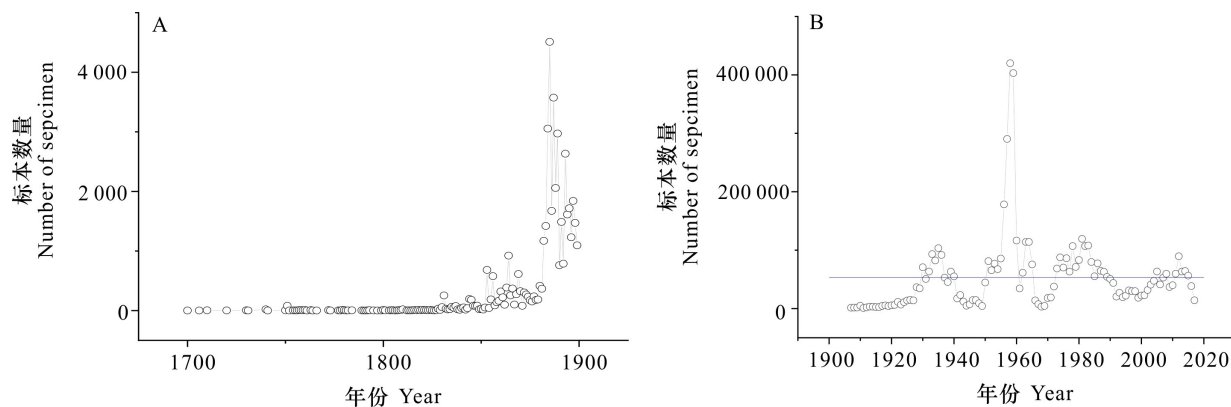
很不完整。根据 Yang 等(2014),对斜率定义阈值为 0.05,当平均斜率值大于 0.05,认为该地理单元的采集程度较为完整,反之则采集程度不完整。

所有统计分析都在 OriginPro 9.0 软件中完成,地图制图在 ArcGIS 10.4.1 软件中完成。

2 结果与分析

2.1 采集时间分析

中国植物标本最早采集始于 18 世纪(王印政等, 2004),据标本数据库在 19 世纪末出现第一个采集高峰,其中 1885 年的采集数量最多,达 4 510 份。19 世纪末 20 世纪初在中国采集的植物标本主要由欧美俄等国牧师和外交官进行,标本主要存放于中国境外。20 世纪以后采集的标本主要形成 4 个高峰,分别是 20 世纪 30 年代、60 年代、80 年代和 21 世纪初(图 1)。第一个采集高峰在 20 世纪 30 年代,其中 1935 年采集量最高,标本量达 123 267 份,这一采集高峰代表近代中国植物学研究的兴起,是中国植物分类学众多开拓者们艰苦卓绝的标本采集活动的成果,同时还有不少外国采集者的参与和贡献;第二个采集高峰在 1958 年前后,其中 1958 年采集量最高,标本量达 523 807 份;第三个采集高峰在 20 世纪 80 年代,其中 1983 年采集量最高,标本量达 181 122 份;在 2013 年前后形成了第四个采集小高峰,其中 2013 年采集量最高,标本量达 87 127 份。



A. 20 世纪之前的标本采集数量; B. 20 世纪之后的标本采集数量。

A. Number of specimens collected before 20th century; B. Number of specimens collected after 20th century.

图 1 不同采集年份所采集标本数量情况

Fig. 1 Number of specimens collected in different collection years

新中国成立后,国家为调查国内自然资源本底,组织开展了多次大型综合及单项科考活动,采集大量标本,形成 50 至 90 年代的第二个和第三个采集高峰。例如,黄河中游水土保持考察队(1954—1957)采集标本 3 万余号,全国野生经济植物资源普查(1959—1960)在全国共采集标本约 20 万号,青藏综合考察队(1973—1976)共采集标本 1.5 万余号,横断山地区综合科学考察队(1981—1983)共采集标本约 4 万号等(王印政等,2004)。进入 21 世纪后,国家中医药管理局从 2011 年起组织实施了第四次全国中药资源普查,启动 10 个省 205 个县的中草药资源普查试点(黄璐琦等,2012),再次迎来了这个时期标本采集的小高峰,从而形成了第四个采集高峰。

中国目前有活力的 220 余家标本馆共计馆藏 2 100 余万份标本(高等植物、藻类及菌物标本)(覃海宁等,2019),其中大部分为中国学者在 20

世纪 50 年代后独立搜集。戴迈凡等(2018)对《中国植物志》学名发表年代统计结果表明,中国植物描述发表出现两个高峰,20 世纪初至 40 年代和 80、90 年代;后一高峰延续时间短而描述新物种数量多。这说明大部分中国植物分类研究是在 20 世纪 50 年代之后由中国学者完成的。

2.2 采集地分析

2.2.1 省级和县级采集情况分析 对标本数据库统计的结果显示,标本采集覆盖了中国包括港澳台在内的 34 个省(区、市)(图 2)。其中,采集量达 100 万份的只有云南和四川两省,分别为 111 万和 147 万(图 2,表 2)。云南和四川地处中国西南地区,是国内公认的植物物种最为丰富的省份,也是历来吸引最多植物采集人(队)前往调查研究的省份。据最新统计结果,2000—2019 年间,云南、广西和四川位列发表新物种和新记录物种数量最多的 3 个省(区),仅云南就发表了 1 164 个植

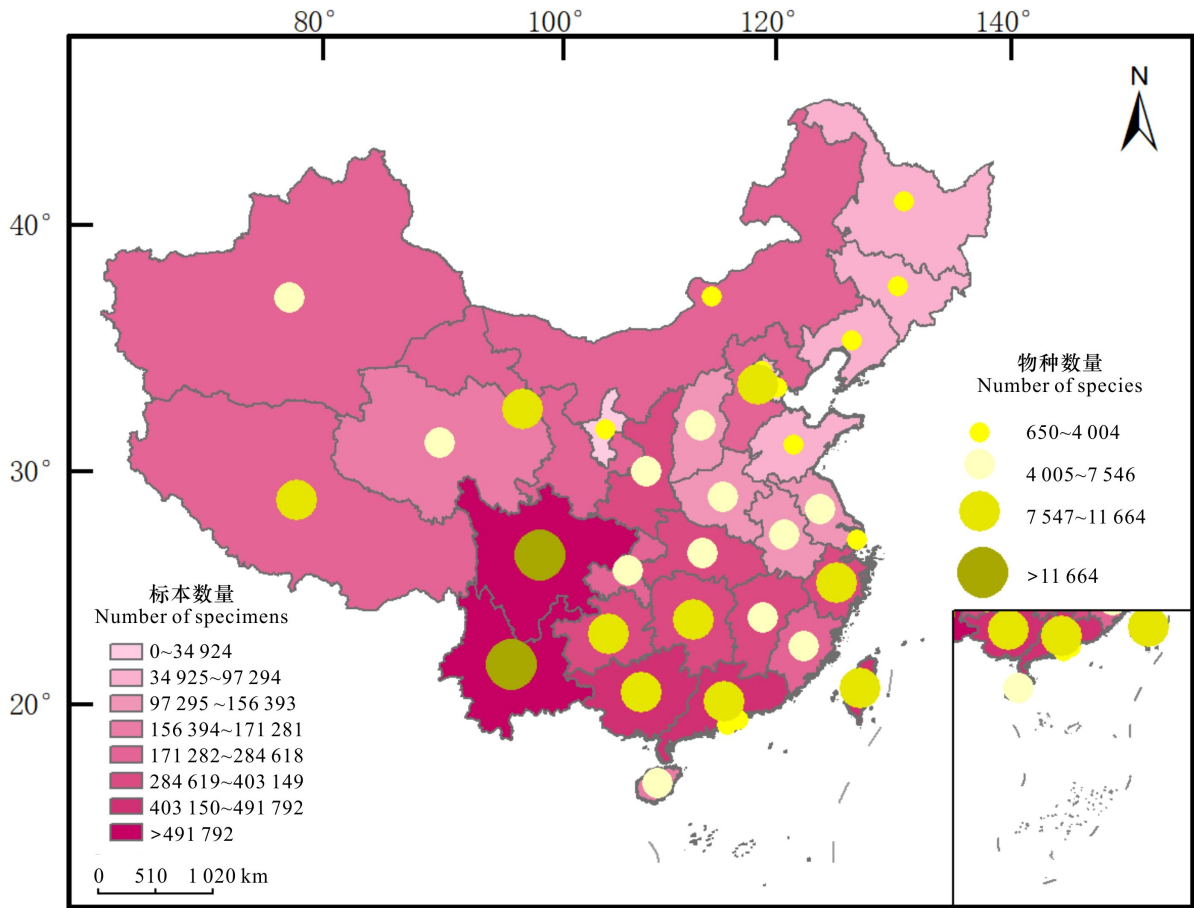


图 2 中国省级标本采集强度和物种数

Fig. 2 Specimen collection intensity and number of species collected on China's province-level

表 2 中国各省(区、市)标本采集强度和采集物种数

Table 2 Specimen collection intensity and number of species collected in each province (region, municipality) of China

省(区、市) Province (region, municipality)	标本数量 Number of specimens	采集物种数量 Number of species with specimens	面积 Area (km ²)	省(区、市) Province (region, municipality)	标本数量 Number of specimens	采集物种数量 Number of species with specimens	面积 Area (km ²)
云南 Yunnan	1 118 500	21 269	390 000	江苏 Jiangsu	131 488	4 889	100 000
四川 Sichuan	1 477 807	17 041	490 000	河南 Henan	142 125	4 859	170 000
广西 Guangxi	449 462	11 664	240 000	安徽 Anhui	129 404	4 732	140 000
西藏 Xizang	265 812	10 557	1 230 000	河北 Hebei	156 393	4 469	190 000
贵州 Guizhou	321 079	10 381	180 000	青海 Qinghai	163 518	4 439	720 000
广东 Guangdong	443 455	9 489	180 000	山西 Shanxi	129 928	4 407	160 000
甘肃 Gansu	265 690	8 977	430 000	内蒙古 Inner Mongolia	190 943	4 004	120 000
湖南 Hunan	336 352	8 471	210 000	北京 Beijing	97 294	3 441	17 000
台湾 Taiwan	491 792	8 348	36 000	吉林 Jilin	86 920	3 158	90 000
湖北 Hubei	314 153	7 546	190 000	黑龙江 Heilongjiang	91 244	3 019	460 000
江西 Jiangxi	320 027	7 490	170 000	辽宁 Liaoning	78 923	3 014	140 000
重庆 Chongqing	192 558	7 159	82 000	山东 Shandong	72 744	2 982	160 000
陕西 Shaanxi	292 550	7 138	210 000	上海 Shanghai	34 924	2 779	6 340
福建 Fujian	205 302	6 818	120 000	香港 Hong Kong	15 344	2 366	1 113.8
浙江 Zhejiang	368 225	6 581	100 000	宁夏 Ningxia	25 155	2 082	66 000
海南 Hainan	171 281	6 035	34 000	天津 Tianjin	7 085	820	12 000
新疆 Xinjiang	284 618	5 472	1 664 900	澳门 Macao	1 277	650	32.9

物新物种和 230 个新记录物种,分别占到同期全国总量的四分之一和三分之一(Du et al., 2020)。

标本采集量在 10 万份以下的省份中,面积较大的省份有山东、黑龙江、吉林和辽宁。王印政等(2004)研究发现,新中国成立后山东省的标本采集主要存放于山东师范大学生命科学学院植物标本室(SDNU)及山东大学生命科学院植物标本室(JSPC)。然而,SDNU 尚未开展植物标本数字化,JSPC 馆藏量为 1.3 万份(覃海宁等,2019),已数字化了三分之一。标本数据库中三省数字化标本共有 25.6 万份,包括来自中国科学院沈阳应用生态研究所东北生物标本馆(IFP)的 18 万份。覃海宁等(2019)研究发现,IFP 馆藏标本达 60 万份。显然,由于东三省植物标本数字化率偏低的缘故,因此本库中数据与实际馆藏量存在一定差距。建议东三省标本馆成为今后开展数字化工作的重点地区之一。

通过对标本数据库的 2 164 个县的统计,得到平均每个县的标本量为 2 678 份,物种数为 482 种。标本采集量前 20 县共来自 7 个省(区、市),

其中 9 个县位于云南,5 个县位于四川(表 3)。标本采集量最多的前三个县(市、区)分别是四川省的峨眉山市、重庆市的南川区 and 四川省的康定市,标本量分别为 17.8 万份、16.6 万份和 6.3 万份;含物种数最高的前三个县是重庆市的南川区、四川省的峨眉山市和云南省的贡山独龙族怒族自治县,含物种数分别为 5 022 种、4 460 种和 3 993 种(表 3)。标本库中有 9%($n = 216$)的县没有标本记录,如山东的高青县、夏津县等;有 22%($n = 527$)的县标本量不到 100 份,有 31%($n = 731$)的县物种数不到 100 种,如吉林图们市的标本量为 63 份 31 种、江苏泰州市的标本量为 25 份 23 种(图 3)。由进一步分析可知,这些采集量位列前茅的县,大多为著名的生物多样性高的地区,同时是宗教圣地、旅游名胜所在地,或大专院校长期定点实习采集点,或某采集人/队长期采集地。如横断山地区综合科学考察队(1981—1983)对云南四川境内的几个县采集了相当数量的标本,此后这些县还分别开展了综考,采集了大量标本;针对佛教圣地峨眉山和金佛山的采集主

表 3 标本采集强度前 20 位的县名单
Table 3 Top 20 counties for specimen collection

省(区、市) Province (region, municipality)	县(市、区) County (city, region)	标本数量 Number of specimens	采集物种数量 Number of species with specimens
四川 Sichuan	峨眉山市 Emeishan City	178 022	4 460
重庆 Chongqing	南川区 Nanchuan District	166 449	5 022
四川 Sichuan	康定市 Kangding City	63 720	3 421
湖北 Hubei	神农架林区 Shennongjia Forest	59 311	3 129
四川 Sichuan	天全县 Tianquan County	47 394	2 891
浙江 Zhejiang	杭州市 Hangzhou City	47 029	3 097
云南 Yunnan	贡山独龙族怒族自治县 Gongshan Derungzu Nuzu Autonomous County	43 907	3 993
云南 Yunnan	昆明市 Kunming City	43 203	3 557
四川 Sichuan	木里藏族自治县 Muli Zangzu Autonomous County	41 645	3 323
四川 Sichuan	雷波县 Leibo County	37 186	2 883
云南 Yunnan	香格里拉市 Shangri-La City	36 051	3 654
云南 Yunnan	德钦县 Deqen County	31 646	3 075
云南 Yunnan	屏边苗族自治县 Pingbian Miaozi Autonomous County	29 977	3 383
云南 Yunnan	景洪市 Jinghong City	27 501	3 339
云南 Yunnan	勐腊县 Mengla County	27 297	3 842
西藏 Xizang	墨脱县 Meetog County	25 222	2 854
云南 Yunnan	丽江市 Lijiang City	22 205	2 954
云南 Yunnan	景东彝族自治县 Jingdong Yizu Autonomous County	21 088	3 095
广西 Guangxi	兴安县 Xing'an County	19 005	2 902
广西 Guangxi	那坡县 Napo County	18 155	3 151

要由四川大学完成,如 1951—1957 年间,在方文培教授(其个人在峨眉山采集标本超过 1 万份)的主持下派出的金佛山植物资源考察队(1986)以及在康定、二郎山等地采集的川西队(1963)等采集活动。根据标本数据库信息统计,标本采集在 500 份以上的采集人或采集队,在峨眉山地区采集的有 38 人(队),金佛山地区有 17 人(队)。

我们将各个省(区、市)和县级的物种数量和标本数量进行了 Pearson 相关分析,相关系数分别为 0.95 和 0.75 ($P < 0.001$),说明物种的数量和标本的采集量成正相关。

2.2.2 县级物种采集完整性评估 对 2 217 个县(占全国 2 377 个县的 93%)的物种采集完整性进行评估, Slope 平均值为 0.225。其中, Slope 小于 0.05 的县仅有 4% ($n = 93$),表明这些县的物种采集较为全面,如四川省的峨眉山市 (Slope = 0.007)、重庆市的南川区 (Slope = 0.01) 等(图 4)。Slope 小于 0.05 的县与阳文静(2013)的相比少了 100 多个,推测是其根据文献补充植物分布记录后带来的结果差别,在一定程度上补充了数字化标本的数据未覆盖的信

息。96% 的县 Slope 大于 0.05,即 96% 的县属于采集不完整;斜率较大的县主要集中在河北、山东、安徽北部;有 160 个县(7%,分布在 21 个省份)没有标本记录。将采集完整性指数在县级地图上进行可视化(图 4),可以明显看出,在县级水平上,各地区的物种采集完整程度极度不平衡。

2.3 采集类群分析

2.3.1 类群覆盖程度 标本数据库中科、属和种的标本覆盖情况见表 4。被子植物物种数量是高等植物各类群中的“大户”,其标本数量也是最大的,并且类群覆盖度上,其科、属和种的覆盖度均在 90% 以上。裸子植物是 4 个类群中覆盖率最高的类群,科、属和种的覆盖率均在 96% 以上;蕨类植物的科、属和种的覆盖率均在 90% 以上;苔藓植物的科和属覆盖率在 90% 以上,但种仅覆盖了 81%。

2.3.2 不同分类水平标本采集情况 对高等植物 4 个类群的科、属和种的标本量前 10 名进行统计(表 5)。

在科级水平上,平均标本量为 21 365 份。标本数据库中标本数量超过 21 365 份的科有 98 个,占

表 4 标本库与中国生物物种名录在类群水平的匹配情况
Table 4 Matching of the specimen database and the Col-China at taxa level

项目 Item	类群 Group	被子植物 Angiosperm	裸子植物 Gymnosperm	蕨类植物 Pteridophyte	苔藓植物 Bryophyte	合计 Total
中国生物物种名录 Col-China	科 Family	264	10	40	151	465
	属 Genus	3 185	45	178	595	4 003
	种 Species	30 724	263	2 217	3 045	36 249
标本库 Specimen database	科 Family	263 (8 867 625)	10 (158 092)	40 (641 407)	149 (203 600)	462 (9 870 724)
	属 genus	3 151 (8 441 225)	45 (154 207)	176 (611 823)	557 (192 354)	3 929 (9 399 609)
	种 species	28 112 (7 156 453)	252 (137 015)	1 988 (526 497)	2 463 (114 101)	32 815 (7 934 066)

注: 括号内数字表示标本数量。

Note: Numbers in brackets indicate the number of specimens.

总科数的 21%, 标本总量达 8 805 603 份, 占标本数据库总量的 89%。标本量在 10 万份以上的科有 24 个, 标本总量达 5 142 754 份, 占标本库总量的 52%, 除蕨类植物的水龙骨科 (Polypodiaceae) 和鳞毛蕨科 (Dryopteridaceae) 外, 其余 22 科均为被子植物 (表 5)。被子植物中科的标本数量排名前三的分别是蔷薇科 (Rosaceae)、菊科 (Asteraceae) 和禾本科 (Poaceae), 标本量均在 50 万份以上; 裸子植物中科的标本数量排名前三的分别是松科 (Pinaceae)、柏科 (Cupressaceae) 和红豆杉科 (Taxaceae), 标本量均在 2 万份以上; 苔藓植物中科的标本数量排名前三的分别是丛藓科 (Pottiaceae)、羽藓科 (Thuidiaceae) 和青藓科 (Brachytheciaceae), 标本量均在 1 万份以上。采集量排名靠前的科同时也是各大类群中的物种数量大的科, 如裸子植物的松科和柏科, 被子植物的菊科、禾本科、蔷薇科和唇形科 (Lamiaceae)。

在属级水平上, 平均标本量为 2 392 份。标本数据库中标本数量超过 2 392 份的属有 840 个, 标本总量达 7 944 795 份, 占标本库总量的 80%。标本采集量在 1 万份以上的属有 211 个, 占总属数 5%, 标本总量达 5 114 353 份, 占标本库总量的 51%, 其中被子植物有 190 个属, 蕨类植物 18 个属, 裸子植物 3 个属。含 5 万份标本以上的属有 21 个, 全部为被子植物, 多为大属或含较多广布种的属, 其中双子叶植物 18 个属, 蔷薇科 4 属为最多, 单子叶植物仅有藎草属 (*Carex*) 和菝葜属

(*Smilax*) 两个属。排名前五的属分别是杜鹃花科的杜鹃花属 (*Rhododendron*)、蓼科的蓼属 (*Polygonum*)、五福花科的荚蒾属 (*Viburnum*)、蔷薇科的悬钩子属 (*Rubus*) 和李属 (*Prunus*), 标本采集量均在 8 万份以上。裸子植物排名前五的属分别是松科松属 (*Pinus*)、柏科刺柏属 (*Juniperus*)、松科云杉属 (*Picea*)、红豆杉科三尖杉属 (*Cephalotaxus*) 和红豆杉属 (*Taxus*), 标本采集量均在 8 000 份以上。采集量排名靠前的属同样多为物种数量多的大属 (如杜鹃花属) 和常见属 (如蓼属和松属), 以及观赏花卉 (如荚蒾属)。

在种级水平上, 平均标本数量为 241 份。标本数大于 241 份的有 7 129 种, 占总种数的 22%, 隶属于 254 科, 占总科数的 55%, 这些物种所涉及的标本份数共为 6 468 959 份, 占整个标本库的 84%。种的标本采集量排前 10 名见表 5, 其中前三分别是蔷薇科的梅 (*Prunus mume*, 26 422 份)、禾本科的稻 (*Oryza sativa*, 20 506 份)、豆科的大豆 (*Glycine max*, 15 985 份)。标本采集量较多的物种多为观赏植物、栽培植物或同时为广布植物。

2.3.3 类群采集薄弱分析

(1) 缺乏标本记录的植物类群

标本采集在科级的覆盖度较好, 仅有 3 个科缺乏标本记录, 占全部科的 0.6%, 包含疑似灭绝植物白玉簪 (*Corsiopsis chinensis*) 所在的白玉簪科 (Corsiaceae) (覃海宁等, 2017)。在属级水平上, 无标本采集的属为 74 个属 (表 4), 占全部属的

表 5 各类群标本采集量排名前 10 的科属种

Table 5 The top 10 families, genera, species in terms of the number of specimens at taxa level

类别 Type	被子 植物 Angiosperm	标本 数量 NOC	物种 数量 NOS	裸子 植物 Gymnosperm	标本 数量 NOC	物种 数量 NOS	蕨类 植物 Pteridophyte	标本 数量 NOC	物种 数量 NOS	苔藓 植物 Bryophyte	标本 数量 NOC	物种 数量 NOS
科 Family	蔷薇科 Rosaceae	642 212	1 721	松科 Pinaceae	63 634	158	鳞毛蕨科 Dryopteridaceae	110 211	505	丛藓科 Pottiaceae	13 958	197
	菊科 Asteraceae	552 607	2 700	柏科 Cupressaceae	49 940	81	水龙骨科 Polypodiaceae	103 207	354	羽藓科 Thuidiaceae	11 569	40
	禾本科 Poaceae	515 871	2 310	红豆杉科 Taxaceae	22 322	34	凤尾蕨科 Pteridaceae	80 264	266	青藓科 Brachytheciaceae	11 416	120
	豆科 Fabaceae	498 119	2 349	麻黄科 Ephedraceae	6 783	18	蹄盖蕨科 Athyraceae	63 918	323	蔓藓科 Metesoriaceae	11 180	74
	唇形科 Lamiaceae	282 745	1 336	罗汉松科 Podocarpaceae	5 342	13	金星蕨科 Thelypteridaceae	41 724	209	提灯藓科 Mniaceae	10 371	90
	毛茛科 Ranunculaceae	255 259	1 409	苏铁科 Cycadaceae	3 935	28	铁角蕨科 Aspleniaceae	33 922	111	灰藓科 Hypnaceae	9 587	111
	杜鹃花科 Ericaceae	213 649	1 023	银杏科 Ginkgoaceae	3 028	1	卷柏科 Selaginellaceae	29 600	73	金发藓科 Polytrichaceae	9 007	56
	樟科 Lauraceae	211 077	528	买麻藤科 Gnetaceae	2 522	10	碗蕨科 Dennstaedtiaceae	26 216	60	平藓科 Neckeraceae	7 458	85
	茜草科 Rubiaceae	177 631	857	南洋杉科 Araucariaceae	488	4	石松科 Lycopodiaceae	20 235	70	真藓科 Bryaceae	7 212	81
	蓼科 Polygonaceae	168 157	289	金松科 Sciadopityaceae	98	1	膜蕨科 Hymenophyllaceae	14 004	53	绢藓科 Entodontaceae	5 951	38
属 Genus	杜鹃花属 <i>Rhododendron</i>	131 684	713	松属 <i>Pinus</i>	22 265	56	鳞毛蕨属 <i>Dryopteris</i>	44 790	176	羽藓属 <i>Thuidium</i>	7 090	14
	蓼属 <i>Polygonum</i>	110 509	145	刺柏属 <i>Juniperus</i>	18 873	39	耳蕨属 <i>Polystichum</i>	31 269	209	青藓属 <i>Brachythecium</i>	6 656	50
	荚蒾属 <i>Viburnum</i>	88 634	107	云杉属 <i>Picea</i>	13 889	33	铁角蕨属 <i>Asplenium</i>	30 101	93	小金发藓属 <i>Pogonatum</i>	5 838	22
	悬钩子属 <i>Rubus</i>	83 300	364	三尖杉属 <i>Cephalotaxus</i>	9 830	11	凤尾蕨属 <i>Pteris</i>	29 973	97	绢藓属 <i>Entodon</i>	5 217	34
	李属 <i>Prunus</i>	82 786	149	红豆杉属 <i>Taxus</i>	8 442	10	卷柏属 <i>Selaginella</i>	28 753	73	真藓属 <i>Bryum</i>	5 138	54
	铁线莲属 <i>Clematis</i>	74 359	276	冷杉属 <i>Abies</i>	7 947	32	蹄盖蕨属 <i>Athyrium</i>	23 112	137	灰藓属 <i>Hypnum</i>	4 752	23
	槭属 <i>Acer</i>	74 311	129	麻黄属 <i>Ephedra</i>	6 739	18	瓦韦属 <i>Lepisorus</i>	21 216	51	匐灯藓属 <i>Plagiomnium</i>	4 698	29
	柳属 <i>Salix</i>	72 719	365	落叶松属 <i>Larix</i>	6 020	16	双盖蕨属 <i>Diplazium</i>	20 547	98	凤尾藓属 <i>Fissidens</i>	4 664	69
	蔷薇属 <i>Rosa</i>	71 781	166	铁杉属 <i>Tsuga</i>	5 907	6	石韦属 <i>Pyrrisia</i>	19 264	32	光萼苔属 <i>Porella</i>	4 290	62
	栎属 <i>Quercus</i>	71 567	121	柏木属 <i>Cupressus</i>	5 431	8	铁线蕨属 <i>Adiantum</i>	14 245	38	羽苔属 <i>Plagiochila</i>	3 665	88
种 Species	梅 <i>Prunus mume</i>	26 465		须弥红豆杉 <i>Taxus wallichiana</i>	7 672		野雉尾金粉蕨 <i>Onychium japonicum</i>	5 389		大羽藓 <i>Thuidium cymbifolium</i>	1 496	
	稻 <i>Oryza sativa</i>	20 521		三尖杉 <i>Cephalotaxus fortunei</i>	5 620		海金沙 <i>Lygodium japonicum</i>	4 941		东亚小金发藓 <i>Pogonatum inflexum</i>	1 320	
	大豆 <i>Glycine max</i>	15 324		杉木 <i>Cunninghamia lanceolata</i>	4 965		节节草 <i>Equisetum ramosissimum</i>	4 881		匐灯藓 <i>Plagiomnium cuspidatum</i>	811	

续表5

类别 Type	被子 植物 Angiosperm	标本 数量 NOC	物种 数量 NOS	裸子 植物 Gymnosperm	标本 数量 NOC	物种 数量 NOS	蕨类 植物 Pteridophyte	标本 数量 NOC	物种 数量 NOS	苔藓 植物 Bryophyte	标本 数量 NOC	物种 数量 NOS
	唐古碎米荠 <i>Cardamine tangutorum</i>	14	432	侧柏 <i>Platycladus orientalis</i>	4	759	贯众 <i>Cyrtomium fortunei</i>	4	793	羽枝青藓 <i>Brachythecium plumosum</i>		759
	青榨槭 <i>Acer davidii</i>	10	845	马尾松 <i>Pinus massoniana</i>	4	191	蜈蚣草 <i>Pteris vittata</i>	4	115	绿羽藓 <i>Thuidium assimile</i>		755
	白檀 <i>Symplocos paniculata</i>	10	157	刺柏 <i>Juniperus formosana</i>	3	814	边缘鳞盖蕨 <i>Microlepia marginata</i>	3	993	刀叶树平藓 <i>Homaliodendron scalpellifolium</i>		749
	珍珠花 <i>Lyonia ovalifolia</i>	9	637	柏木 <i>Cupressus funebris</i>	3	793	乌蕨 <i>Odontosoria chinensis</i>	3	835	桧叶白发藓 <i>Leucobryum juniperoideum</i>		687
	桦叶荚蒾 <i>Viburnum betulifolium</i>	9	624	日本柳杉 <i>Cryptomeria japonica</i>	3	455	渐尖毛蕨 <i>Cyclosorus acuminatus</i>	3	548	扭叶藓 <i>Trachypus bicolor</i>		686
	山鸡椒 <i>Litsea cubeba</i>	9	489	铁杉 <i>Tsuga chinensis</i>	3	215	欧洲凤尾蕨 <i>Pteris cretica</i>	3	536	侧枝匍灯藓 <i>Plagiomnium maximoviczii</i>		685
	龙芽草 <i>Agrimonia pilosa</i>	9	486	圆柏 <i>Juniperus chinensis</i>	3	048	狗脊 <i>Woodwardia japonica</i>	3	506	真藓 <i>Bryum argenteum</i>		664

注: NOC. 标本数量; NOS. Col 中物种数量。

Note: NOC. Number of specimens; NOS. Number of species in Col.

1.8%。其中被子植物有 34 个属无标本采集,无标本属最多的科是兰科,有 7 个属缺乏标本记录,大都为近些年发表的新属,如 *Singchia* (Liu & Chen, 2009)、*Shizhenia* 和 *Gennaria* (金伟涛等, 2015); 十字花科有 6 个属,有的该属为单种属较难采集,如 *Atelanthera*; 菊科有 4 个属,即 *Schischkinia*、*Pseudohandelia*、*Microcephala* 和 *Frolovia*, 据《中国植物志》(1987) 记载,大都分布在新疆,并且植株矮小,较难采集。无标本的属大多集中在采集量在 10 万份以上的物种数量大的科,说明这些大科仍旧需要在分类学上开展更多的研究,同时各馆仍需加强模式标本的数字化工作。

在种级水平上,无标本采集的种为 3 250 种,占全部物种的 9%,其中被子植物 2 445 种、裸子植物 11 种、蕨类植物 212 种、苔藓植物 582 种。有 623 个无标本物种是在 2007 年之后发表的。无标本物种数较多的科,仍集中在物种数量和标本数量较大的科,如禾本科、兰科、豆科和菊科(表 6),其中禾本科无标本物种数最多,有 225 种,涉及 66 个属。无标本的物种所在的科或者是野外分类较难的类群,如禾本科和莎草科;或者是近些年发表的新种等狭域分布物种,如 *Phyllostachys*

purpureociliata G. H. Lai (2013)、*Senecio changii* C. Ren & Q. E. Yang (2016) 等。

依据“中国高等植物受威胁物种名录”(覃海宁等, 2017),无标本物种中有 16 种为疑似灭绝植物; 331 种为受威胁物种,占全部受威胁物种及无标本物种的 10%,其中有 197 个物种属于狭域分布物种,仅分布在 1~3 个县范围内。当用 *Flora of China* (Wu et al., 2013b) 认可的 197 个特有属(含 200 余种)的物种名录与标本数据库物种名称进行匹配时发现,有一半以上物种缺乏标本,除学名归并及拼写错误之外,缺乏标本记录的比例仍然不低。

(2) 标本数小于 5 份的植物类群

根据标本数据库,有 4 698 个物种标本数量过少,标本份数在 5 份以下(标本份数 ≤ 4),涉及 275 科 1 325 属,约十分之一(12.9%)的物种是采集薄弱的,其中有 404 个物种的标本仅为模式标本。包括被子植物 3 835 种、裸子植物 7 种、蕨类植物 194 种、苔藓植物 662 种。仅有 1 份标本的物种数为 1 507 个,涉及 195 科 690 属,占标本库物种数的 4.1%,其中被子植物 1 202 种、裸子植物 3 种、蕨类植物 59 种、苔藓植物 243 种。

23 个科的物种中一半以上标本份数在 5 份以

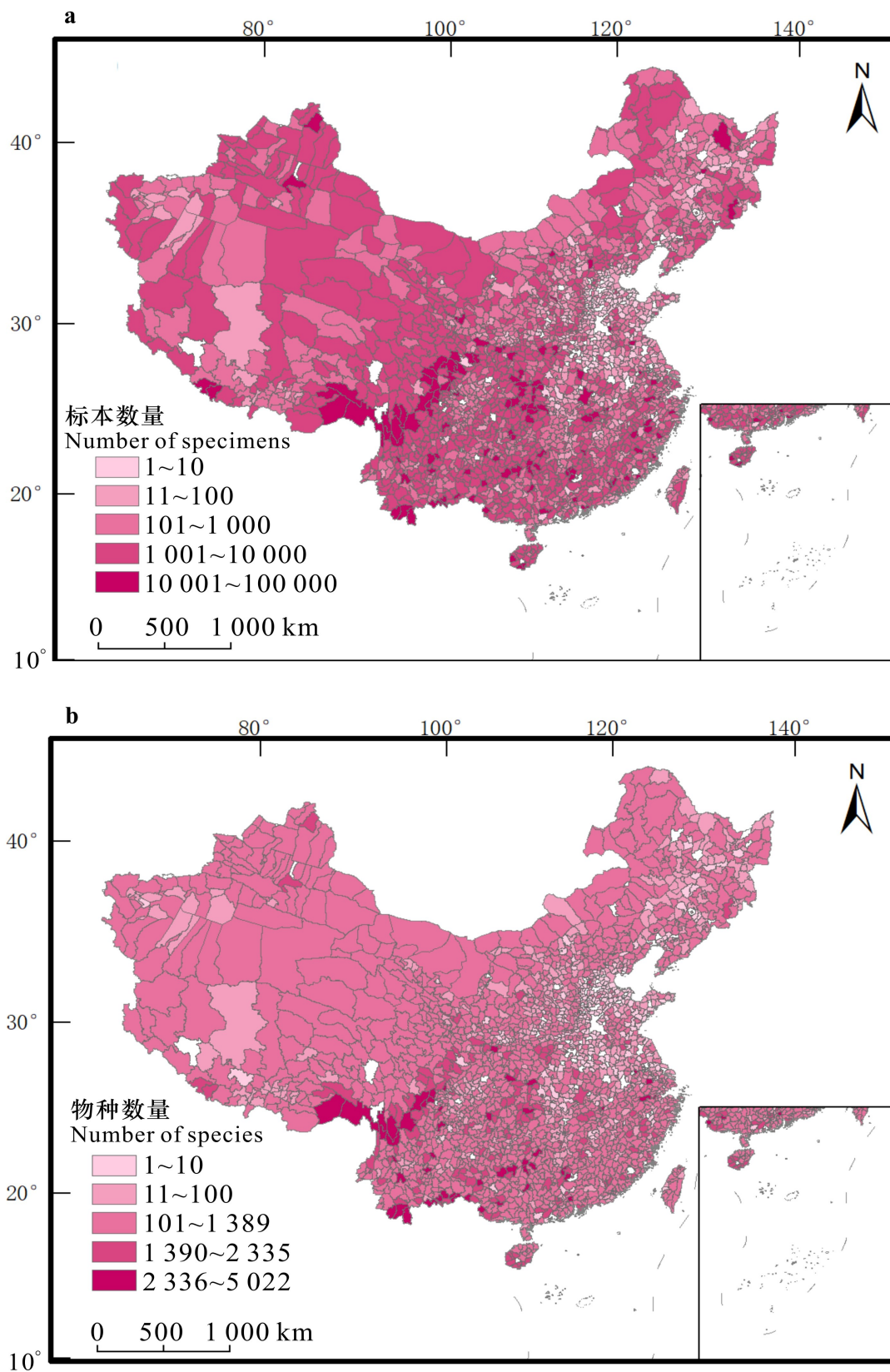
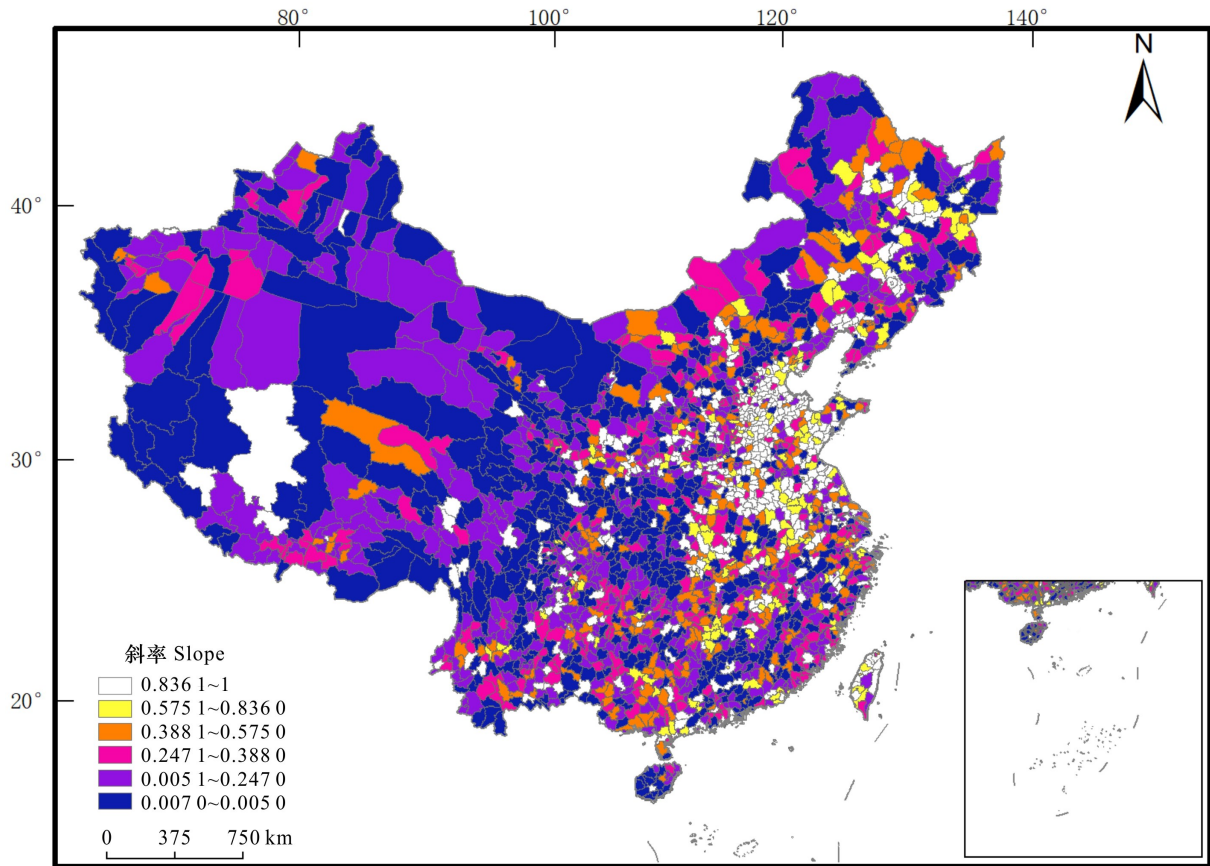


图 3 中国县级标本采集强度 (a) 和物种数 (b)

Fig. 3 Specimen collection intensity (a) and number of species (b) collected on China's county-level



深色(指数小)表示物种采集完整程度较高,浅色(指数大)表示物种采集完整程度较低。

Dark color (small index) indicates that the species collection is complete, and light color (large index) indicates poor collection of species.

图 4 中国2 377个县的物种采集完整程度

Fig. 4 Completeness of species collection in 2 377 counties in China

下(表7),除单型科以及物种数量较少的科外,还包括一些60%左右的物种都是狭域分布的科如秋海棠科、苦苣苔科等。标本份数在5份以下的物种中,有565个物种是受威胁物种(覃海宁等,2017),其中分布地点仅为1~3个的狭域物种有425个,约占采集薄弱的物种数的十分之一。

3 讨论与结论

3.1 中国植物标本采集和研究工作主要在20世纪50年代后由中国作者完成,中国植物标本采集有力地支撑着中国植物分类学的发展

阳文静(2013)通过对650万条标本数据的统计认为我国有3个采集高峰,这三个采集高峰与

本文的前三个采集高峰一致,但其所用数据库时间范围仅到2009年,而根据本文的标本数据库第4个采集小高峰是在2013年前后形成。我国植物标本的4个采集高峰真实地反映了我国植物学发展以及植物标本采集的几个重要阶段,并与政府在国家发展的不同阶段所执行的方针政策相呼应。第2和第3个高峰形成于20世纪50至90年代,同时大部分中国植物分类研究也是在20世纪50年代之后完成的,表明中国植物标本采集和研究工作主要在20世纪50年代后由中国学者完成。各国标本采集积累量,与一个国家植物资源的丰富程度和植物学的研究水平成正比(王文采,2021)。中国植物标本的采集在短短一百年间,就积累了2 000多万份,平均不到30年就形成了一个

表 6 无标本物种数在 50 种以上的科
Table 6 Families with more than 50 species without specimen

科 Family	物种数 Number of species	无标本物种数 Number of species without specimen	无标本物种 涉及的属数 Genus number of species without specimen	种的标本缺乏率 Specimen deficiency rate of species (%)
禾本科 Poaceae	2 310	225	66	10%
兰科 Orchidaceae	1 535	212	84	14%
豆科 Fabaceae	2 349	179	42	8%
菊科 Asteraceae	2 700	174	56	7%
莎草科 Cyperaceae	1 067	91	11	9%
杜鹃花科 Ericaceae	1 023	91	6	9%
苦苣苔科 Gesneriaceae	689	89	17	13%
蔷薇科 Rosaceae	1 721	75	16	4%
天门冬科 Asparagaceae	342	74	15	23%
天南星科 Araceae	237	56	14	24%
茜草科 Rubiaceae	857	54	21	7%
毛茛科 Ranunculaceae	1 409	50	8	4%

采集高峰,并且在 2000 年之后还有采集高峰,说明中国的标本采集与资源调查工作一直在持续进行,有力地支撑中国植物分类学的发展,给植物分类学研究提供了丰厚的养分。我国的植物标本数字化工作在后植物志时代,继续引领国际潮流,对世界植物分类学的发展持续发挥重要作用。

3.2 我国标本采集空间覆盖度在省级较好,县级标本采集强度和不完整程度极度不平衡

东北平原地区、华北平原地区、长江中下游平原及四川盆地等低海拔地区的标本物种采集完整性严重不足。而山区如横断山区、云贵高原、秦岭、武夷山等生物多样性热点地区(黄继红等, 2014)均有相对较高的标本采集量和较好的物种采集完整性。这个结果和阳文静(2013)的研究结果基本是一致的。造成县级标本采集不平衡现象的主要原因有两个:一方面可能受到采集者偏好(Daru et al., 2018)的影响,如平原地区相对来说多为农耕地,生境异质性较低,物种多样性较低,而我国的标本采集活动大多为专业采集,植物学家相对不重视物种多样性较低地区的采集;山区由于海拔和不同的地理环境差异导致的生境异质性较高,物种的多样性非常高,植物学家更倾向于去山区进行植物采集。另一方面可能与采集地区的知名度相关,如我国一些宗教圣地所处的名山大川,经常吸引植物学家进行采集;此外,还与国

家方针政策,植物学家所获科研项目直接相关,如国家部署的资源调查项目等。

3.3 我国标本采集类群覆盖率高,但近五分之一的物种采集不足

我国标本采集类群覆盖率高,但苔藓植物相对于维管植物的采集强度较低。一方面由于研究苔藓植物的专家较少,并且苔藓植物本身矮小使其不易被发现以及高的生境质量要求使其不易被采集,而且鉴定苔藓标本需要花费更多的时间,导致苔藓植物的采集强度及鉴定数量低于其他类群。物种数较多的科,通常标本采集量也较多,并且标本的采集量与植物类群的常见程度紧密相关。物种中采集标本量较多的种,大多是栽培、广布和经济植物,这与该物种的经济价值、采集的便利程度及物种分布幅度有明显的关联,上述结论与阳文静(2013)的研究结果基本一致。另一方面,在类群数字化空缺上,尽管科级采集覆盖度较好,但在科级水平上仍有如白玉簪科等类群存在采集空白的情况。近十分之一的物种标本还没有被数字化,这些物种大多集中在被子植物和苔藓植物中,其中苔藓植物的种级缺乏率最高。近五分之一的物种标本是采集不足的。

3.4 基于标本数字化建立的数据库对物种多样性的完整性分析仍有一定的局限性

我们在运用千万数量级的数据进行时空及类

表 7 采集薄弱物种在该科有标本物种数中占比大于 50% 的科

Table 7 Family with the proportion of species with weak collection in the number of species with specimens more than 50%

科 Family	物种数量 Number of species	有标本的 物种数 Number of species with specimen	标本采集薄弱 的物种数 Number of species with weak collection	采集薄弱物种在该科 有标本物种数中比例 Proportion of species with weak collection in the number of species with specimens (%)
节蒴木科 Borthwickiaceae	1	1	1	100
波喜荡科 Posidoniaceae	1	1	1	100
大花草科 Rafflesiaceae	1	1	1	100
翼盖蕨科 Didymochlaenaceae	1	1	1	100
角胡麻科 Martyniaceae	1	1	1	100
冰沼草科 Scheuchzeriaceae	1	1	1	100
花柱草科 Styliidiaceae	2	2	2	100
无叶莲科 Petrosaviaceae	2	2	2	100
木犀草科 Resedaceae	4	3	3	100
川苔草科 Podostemaceae	6	4	4	100
蛇菰科 Balanophoraceae	13	13	10	77
闭鞘姜科 Costaceae	5	4	3	75
熏倒牛科 Biebersteiniaceae	3	3	2	67
金鱼藻科 Ceratophyllaceae	3	3	2	67
水玉簪科 Burmanniaceae	15	12	8	67
沟繁缕科 Elatinaceae	6	6	4	67
鹤望兰科 Strelitziaceae	4	3	2	67
秋海棠科 Begoniaceae	195	160	103	64
苦苣苔科 Gesneriaceae	583	494	312	63
芭蕉科 Musaceae	20	18	10	56
竹芋科 Marantaceae	12	11	6	55
兰科 Orchidaceae	1 487	1 269	695	55
凤仙花科 Balsaminaceae	271	239	123	51

群的现状分析时发现,尽管使用海量的标本数据对我国植物采集情况进行分析,所得到的结论仍可能与真实的情况有所差异,这主要取决于数据库的完整程度,因为其完整程度对物种编目的完整性影响非常大(Qian et al., 2018),而采集地理偏差也会导致分析得出的结论不一定真实(Yang et al., 2014; Daru et al., 2018)。之所以造成这种差异,可能主要存在如下因素:(1)类群方面的处理。一方面,数字化标本的数据绝大部分由人工录入,因此可能存在各种问题导致名称的录入错误,从而在我们对标本名称进行校对时,可能会有

一些物种因为录入错误而被剔除,从而导致一些类群形成采集空缺;另一方面,由于馆藏标本的鉴定通常滞后于分类学的研究,很多类群在最新的名录系统中已经被处理,然而标本上却未及时更新,以致没有被统计到。尽管我们考虑到这个情况,使用的是 2017 年的生物物种名录,但是难免还会出现类似的问题。另外,在 2017 年之后发表的新种的标本采集情况,这里没有进行类群统计。(2)我国标本数字化程度的不均衡。我国标本数字化的程度与参与数字化项目的标本馆数量、运行状况及参与程度直接相关。尽管本文利用了数

千万条标本数据进行分析,然而覃海宁等(2019)研究发现,这些标本数据仅占我国馆藏标本的约50%,各馆仍有大量的馆藏标本需要进行数字化,以填补采集空白和薄弱的类群和地区。(3)标本的采集在空间和类群上的分析,没有结合时间分布来进行充分讨论。未来可以考虑分析在一些时间节点之后,哪些物种和地区没有新的标本,据此来布局标本采集的重点。

总体来说,目前中国植物标本数字化工作卓有成效,在类群和时空的覆盖度上都达到了一定的规模。今后的标本数字化工作在继续挖掘馆藏标本的同时,可以考虑:一方面开展对现有数字化标本信息再审核及补充工作,继续面向全国标本馆进行馆藏标本数字化,尤其是加强模式标本及采集薄弱地区,如东北地区馆藏标本数字化,以提高标本数据所承载的时间、空间和类群信息的覆盖度,同时籍此提高馆藏标本的鉴定比例,促进植物分类学的发展;另一方面需要加强与欧美大馆的信息共享以获取更多早期采自中国的历史标本信息。

本文通过对我国目前为止标本收集最全的标本数据库在时间和空间分布以及类群采集方面的分析,了解了我国植物标本在时空、类群方面的采集特点及空缺,所得到的标本采集的未来参考方向,对于国家植物标本资源库生物标本精准采集工作具有重要的参考意义。对采集空白以及采集薄弱的物种进行精准采集,有助于人们了解该物种在目前的野外生存状况,进而有效开展相应的生物多样性评估和采取保护措施;对采集空白及薄弱的地区进行精准采集,可以增加人们对这些地区的植物资源情况的了解且有助于发现新类群。因此,在开展生物标本精准采集时,需要结合分类修订的最新工作,针对标本资源库中空白类群和地区进行精准采集;之后在空缺和薄弱的名单中,按照物种系统学及分类学上的重要意义,以及保护上的重要意义,也可结合其经济上的重要价值进行筛选,如对根据特有属、单种属或寡种属中尤其是木本植物、国家发布的最新《国家重点保护野生植物名录》、新近发表的以及公众关心的一些物种进行精准采集。通过精准采集的科学布局、合理的资源配置,未来国家植物标本资源库将进一步增强实体标本馆能力建设,提高数字化标本质量,以期更好地服务科学及社会。

参考文献:

- BAO SY, MAO PY, YUAN S, 1995. A brief history of plant collection in Yunnan: 1919–1950 [M]. Beijing: Science and Technology of China Press. [包士英, 毛品一, 苑淑秀, 1995. 云南植物采集史略: 1919–1950 [M]. 北京: 中国科学技术出版社.]
- CHASE MW, REVEAL JL, 2009. A phylogenetic classification of the land plants to accompany APGIII [J]. *Bot J Linn Soc*, 161: 122–127.
- CHEN C, 2002. China administrative divisions history handbook [M]. Beijing: Cartographic Publishing Press: 1–440. [陈潮, 2000. 中国行政区划沿革手册 [M]. 北京: 地图出版社: 1–440.]
- CHRISTENHUSZ MJ, REVEAL JL, FARJON A, et al., 2011. A new classification and linear sequence of extant gymnosperms [J]. *Phytotaxa*, 19 (1): 55–70.
- DAI MF, XIANG JP, LIU JA, et al., 2018. Composition Chinese plants based on analysis on specimens and the possibility of discovering new species [J]. *Front Data Comput*, 9(5): 27–35. [戴迈凡, 向杰平, 刘吉安, 等, 2018. 基于标本探究中国植物的组成变化趋势和新物种发现的可能性 [J]. *科研信息化技术与应用*, 9(5): 27–35.]
- DARU BH, PARK DS, PRIMACK RB, et al., 2018. Widespread sampling biases in herbaria revealed from large-scale digitization [J]. *Newphytol*, 217(2): 939–955.
- DU C, LIAO S, BOUFFORD DE, et al., 2020. Twenty years of Chinese vascular plant novelties, 2000 through 2019 [J]. *Plant Divers*, 42(5): 393–398.
- FREY W, STECH M, FISCHER E, 2009. Bryophytes and seedless vascular plants [M]. Stuttgart: Gebr. Brontraeger Verlagsbuchhandlung: 1–263.
- GOTELLI NJ, COLWELL RK, 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness [J]. *Ecol Lett*, 4(4): 379–391.
- GUI LN, LI CY, WANG KN, et al., 2018. Assessment of collection densities of hard-to-collect plant taxa through specimen data and the shortcomings of conducting related studies [J]. *Front Data Comput*, 49(5): 41–46. [桂略宁, 李辰昱, 王康浓, 等, 2018. 通过标本数据评估难以采集的植物类群采集密度及进行相关研究的缺陷 [J]. *科研信息化技术与应用*, 49(5): 41–46.]
- HEDRICK BP, HEBERLING JM, MEINEKE EK, et al., 2020. Digitization and the future of natural history collections [J]. *BioScience*, 70(3): 245–251.
- HUANG JH, MA KP, CHEN B, 2014. Diversity and geographical distribution of endemic seed plants in China

- [M]. Beijing: Higher Education Press: 65-78. [黄继红, 马克平, 陈彬, 2014. 中国特有种子植物的多样性及其地理分布 [M]. 北京: 高等教育出版社: 65-78.]
- HUANG LQ, ZHAO RH, CHEN SL, et al., 2012. Progress of the Fourth National Census of Chinese Medicine Resources Preparation and Pilot Work [J]. *Mod Chin Med*, 14(1): 13-15. [黄璐琦, 赵润怀, 陈士林, 等, 2012. 第四次全国中药资源普查筹备与试点工作进展 [J]. *中国现代中药*, 14(1): 13-15.]
- IGNNB SM, 1983. *China Gazetteer* [M]. Beijing: Cartographic Publishing Press: 1-313. [国家测绘局地名研究所, 1983. *中国地名录* [M]. 北京: 中国地图出版社: 1-313.]
- JAMES SA, SOLTIS PS, BELBIN L, et al., 2018. Herbarium data: Global biodiversity and societal botanical needs for novel research [J]. *Appl Plant Sci*, 6(6325): e01024.
- JIANG CY, YU WX, YANG T, et al., 2018. Analysis of the current status of national botanical specimen collection and prediction of collection trends based on specimen data in Chinese collections [J]. *Front Data Comput*, 49(5): 94-101. [姜承勇, 余卫星, 杨婷等, 2018. 基于中国馆藏标本数据分析全国植物标本采集现状及采集趋势预测 [J]. *科研信息化技术与应用*, 49(5): 94-101.]
- JIN WT, XIANG XG, JIN XH, 2015. Generic delimitation of Orchidaceae from China: current situation and perspective [J]. *Biodivers Sci*, 23(2): 237-242. [金伟涛, 向小果, 金效华, 2015. 中国兰科植物属的界定: 现状与展望 [J]. *生物多样性*, 23(2): 237-242.]
- JOHNSON KG, BROOKS SJ, FENBERG PB, et al., 2011. Climate change and biosphere response: Unlocking the collections vault [J]. *BioScience*, 61(2): 147-153.
- LAI GH, 2013. New species of the genus *Phyllostachys* (Gramineae: Bambusoideae) from Anhui Province, China [J]. *Plant Divers*, 35(2): 131-140.
- LANE MA, 1996. Roles of natural history collections [J]. *Ann Mo Bot Gard*, 83(4): 536-545.
- LIU ZJ, CHEN LJ, 2009. *Singhia* and *Gunnaria*, two new genera of Orchidaceae [J]. *J Syst Evol*, 47(6): 599-604.
- MA KP, LOU ZP, SU RH, 2010. Review and prospects of biodiversity research in Chinese Academy of Sciences [J]. *Bull Chin Acad Sci*, 25(6): 634-644. [马克平, 娄治平, 苏荣辉, 2010. 中国科学院生物多样性研究回顾与展望 [J]. *中国科学院院刊*, 25(6): 634-644.]
- NUALART N, NEUS I, SORIANO I, et al., 2017. Assessing the relevance of herbarium collections as tools for conservation biology [J]. *Bot Rev*, 83(3): 1-23.
- O'CONNELL AF, GOLBERT AT, HATFIELD JS, 2004. Contribution of natural history collection data to biodiversity assessment in national parks [J]. *Conserv Biol*, 18(5): 1254-1261.
- PAGE LM, MACFADDEN BJ, FORTES JA, et al., 2015. Digitization of biodiversity collections reveals biggest data on biodiversity [J]. *BioScience*, 65(9): 841-842.
- PONDER W, CARTER G, FLEMONS P, et al., 2001. Evaluation of museum collection data for use in biodiversity assessment [J]. *Conserv Biol*, 5(18): 1523-1739.
- PYKE GH, EHRLICH PR, 2010. Biological collections and ecological/environmental research: a review, some observations and a look to the future [J]. *Biol Rev*, 85(2): 247-266.
- QIAN H, DENG T, BECK J, et al., 2018. Incomplete species lists derived from global and regional specimen-record databases affect macroecological analyses: A case study on the vascular plants of China [J]. *J Biogeogr*, 45(12): 2718-2729.
- QIN HN, LIU HY, HE Q, et al., 2019. *Index Herbariorum Sinicorum* [M]. 2nd ed. Beijing: Science Press: 1-352. [覃海宁, 刘慧圆, 何强, 等, 2019. *中国植物标本馆索引* [M]. 2版. 北京: 科学出版社: 1-352.]
- QIN HN, YANG Y, DONG SY, et al., 2017. Threatened species list of China's higher plants [J]. *Biodivers Sci*, 25(7): 696-744. [覃海宁, 杨永, 董仕勇, 等, 2017. 中国高等植物受威胁物种名录 [J]. *生物多样性*, 25(7): 696-744.]
- REN C, TONG TJ, HONG Y, et al., 2016. *Seneciochangii* (Asteraceae: Senecioneae), a new species from Sichuan, China [J]. *PLoS ONE*, 11(4): e0151423.
- ROBBIRT KM, DAVY AJ, HUTCHINGS MJ, et al., 2011. Validation of biological collections as a source of phenological data for use in climate change studies: A case study with the orchid *Ophrys sphegodes* [J]. *J Ecol*, 99(1): 235-241.
- SUAREZ AV, TSUTSUI ND, 2004. The value of museum collections for research and society [J]. *BioScience*, 54(1): 66-74.
- The Biodiversity Committee of Chinese Academy of Sciences, 2017. *Catalogue of life China: 2017 annual checklist* [CD]. Beijing: Sciences Press. [中国科学院生物多样性委员会, 2017. *中国生物物种名录 2017 版* [CD]. 北京: 科学出版社.]
- THIERS BM, 2021. *The World's Herbaria 2021: A Summary Report Based on Data from Index Herbariorum (Issue 6.0)* [R]. <http://sweetgum.nybg.org/science/ih/>.
- WANG KL, CHEN JX, FAN X, 2018. Analysis of geographical deviations in the collection of specimen records of Rosaceae on the national specimen resource sharing platform [J]. *Front Data Comput*, 49(5): 54-63. [王凯莉, 陈嘉欣, 范雪, 2018. 国家标本资源共享平台蔷薇科植物标本记录采集地理偏差分析 [J]. *科研信息化技术与应用*, 49(5):

- 54-63.]
- WANG LS, CHEN B, JI LQ, et al., 2010. Progress in biodiversity informatics [J]. *Biodivers Sci*, 18(5): 429-443. [王利松, 陈彬, 纪力强, 等, 2010. 生物多样性信息学研究进展 [J]. *生物多样性*, 18(5): 429-443.]
- WANG WC, 2021. *Global plants introduction* [M]. Beijing: Beijing Publishing Group Co., Ltd, Beijing Press: 23. [王文采, 2021. *世界植物简志* [M]. 北京: 北京出版集团, 北京出版社: 23.]
- WANG YZ, QIN HN, FU DZ, 2004. A brief history of Chinese plant collection [M]// WU ZY, CHEN XQ. *Flora Reipublicae Popularis Sinicae: Tomus 1*. Beijing: Sciences Press: 659-703. [王印政, 覃海宁, 傅德志, 2004. 中国植物采集简史 [M]//吴征镒, 陈心启. *中国植物志: 第一卷*. 北京: 科学出版社: 659-703.]
- WEN J, ICKERT-BOND SM, APPELHANS MS, et al., 2015. Collections-based systematics: Opportunities and outlook for 2050 [J]. *J Syst Evol*, 53(6): 477-488.
- WU CY, RAVEN PH, HONG DY, 2013a. *Flora of China: Vol. 1* [M]. Beijing: Science Press; St. Louis: Missouri Botanical Garden Press: 1-254.
- WU CY, RAVEN PH, HONG DY, 2013b. *Flora of China: Vol. 2-3* [M]. Beijing: Science Press; St. Louis: Missouri Botanical Garden Press.
- XU ZP, CUI JZ, QIN HN, 2012. The concept of building China's biodiversity e-Science platform [J]. *Biodivers Sci*, 18(5): 480-488. [许哲平, 崔金钟, 覃海宁, 2010. 中国生物多样性 e-Science 平台建设构想 [J]. *生物多样性*, 18(5): 480-488.]
- XU ZP, QIN HN, MA KP, et al., 2012. Research on the management, sharing and application of natural science and technology resources: The case of China Digital Herbarium [J]. *Chin Sci Technol Resour Rev*, 44(1): 27-33. [许哲平, 覃海宁, 马克平, 等, 2012. 自然科技资源的管理、共享和应用研究——以中国数字植物标本馆为例 [J]. *中国科技资源导刊*, 44(1): 27-33.]
- YANG WJ, MA KP, KREFT H, 2014. Environmental and socio-economic factors shaping the geography of floristic collections in China [J]. *Glob Ecol Biogeogr*, 23(11): 1284-1292.
- YANG WJ, 2013. Geographical bias in plant collections in China and its impact on the analysis of biodiversity patterns [D]. Beijing: Institute of Botany, Chinese Academy of Sciences: 1-57. [阳文静, 2013. 中国植物采集的地理偏差及其对生物多样性格局分析的影响 [D]. 北京: 中国科学院植物研究所: 1-57.]
- YANG Y, 2012. Holdings of type specimens of plants in herbaria of China [J]. *Biodivers Sci*, 20(4): 512-516. [杨永, 2012. 我国植物模式标本的馆藏量 [J]. *生物多样性*, 20(4): 512-516.]
- YIN ZL, LIU YQ, XIAO C, 2018. Geographical deviation of orchid specimen record collection and its environmental factor interpretation in the national specimen resource sharing platform [J]. *Front Data Comput*, 49(5): 64-71. [尹朝露, 刘雨晴, 肖翠, 2018. 国家标本资源共享平台兰科植物标本记录采集地理偏差及其环境因子解释 [J]. *科研信息化技术与应用*, 49(5): 64-71.]
- ZHANG YX, REN M, XIAO C, 2018. Analysis of geographic deviation in the collection of Chinese Rhododendronaceae specimens based on data from the National Specimen Resource Sharing Platform [J]. *Front Data Comput*, 49(5): 72-83. [张玉雪, 任敏, 肖翠, 2018. 基于国家标本资源共享平台数据的中国杜鹃花科植物标本采集地理偏差分析 [J]. *科研信息化技术与应用*, 49(5): 72-83.]

(责任编辑 蒋巧媛 邓斯丽)