



中国数字植物标本馆 (CVH) 大数据分析

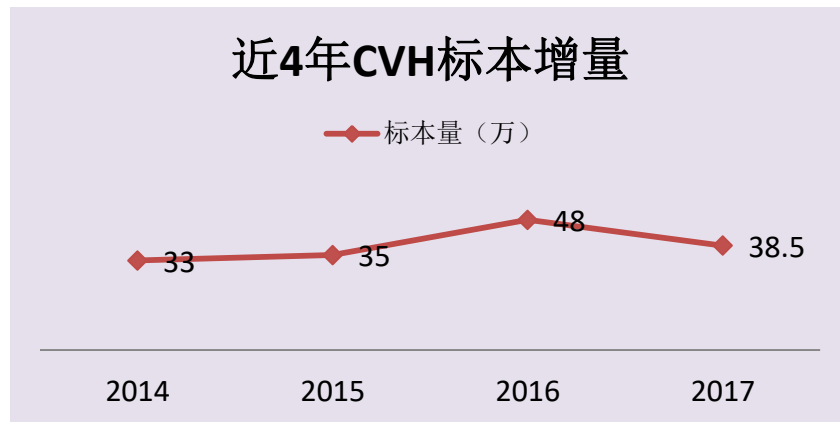
刘慧圆 赵莉娜 包伯坚
中国科学院植物研究所
2018-08-01

内容

- CVH标本总体情况
- CVH标本采集信息分析
- 标本采集空缺分析情况
- 科研支撑

CVH标本总体情况

- 总量
 - 693万份标本（2018年8月）
- 年增量
 - 38万（近4年）
- 数据来源
 - 86家标本馆
- 范围
 - 覆盖34个省市自治区



CVH标本总体情况

CVH全部标本主要字段的非空和有效数据统计

| | 非空数据 | 比率 | 有效数据 | 比率 | 比率差值 |
|------------|---------|--------|---------|--------|---------|
| 国家 | 6304396 | 96.70% | 5589589 | 85.73% | 10.96% |
| 中国 | 6015711 | 95.42% | | | |
| 国外 | 288685 | 4.58% | | | |
| 省名 | 5942517 | 91.15% | 4720350 | 72.40% | 18.746% |
| 县名 | 4023931 | 61.72% | 3879000 | 59.50% | 2.223% |
| 地名 | 4601443 | 70.58% | 4601302 | 70.57% | 0.002% |
| 科名 | 5672864 | 87.01% | 5140121 | 78.84% | 8.171% |
| 属名 | 5398532 | 82.80% | 5398489 | 82.80% | 0.001% |
| 种名 | 5133637 | 78.74% | 5120068 | 78.53% | 0.208% |
| 模式类型 | 4798494 | 73.60% | 4793274 | 73.52% | 0.080% |
| 采集人 | 5418902 | 83.11% | 5141128 | 78.85% | 4.260% |
| 鉴定人 | 4349295 | 66.71% | 4332824 | 66.46% | 0.253% |
| 海拔 | 5314668 | 81.52% | 3698359 | 56.73% | 24.791% |
| 经纬度 | 1869085 | 28.67% | 1550553 | 23.78% | 4.886% |
| 采集日期 | 5336950 | 81.86% | 5242666 | 80.41% | 1.446% |

本次分析

- 总记录数**651万**
- 截至**2017年12月**
- 未包含**2018年1月**的**43万份数据**



CVH标本总体情况



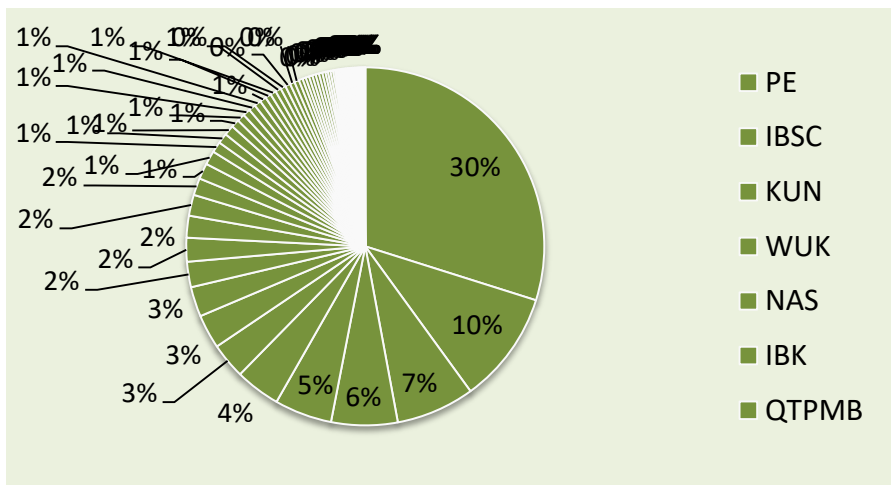
• 数据质量

- 有效数据比例平均达70% 省不规范化的情况（“云南省”案例）
- 各字段普遍存在问题
 - 无效字符
 - 录错
 - 格式不统一

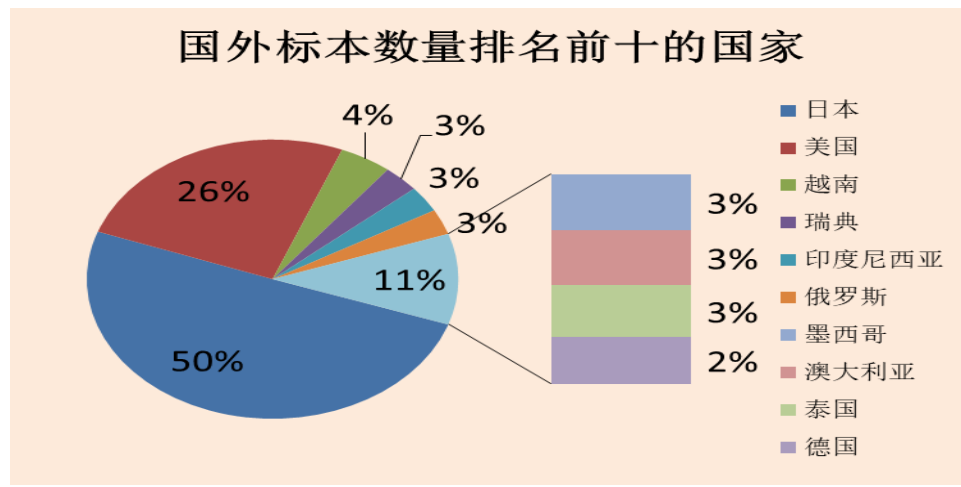
省级字段分组统计
得到2944组

| 省 | 数字 | 省 | 数字 |
|----------------------|--------|----------------------|--------|
| 云南省 | 899050 | 云南 | 4 |
| 云南 | 10974 | 中国云南西双版纳景洪眉公山庄 | 3 |
| 云南； YUNNAN | 2159 | 云南松 | 3 |
| 云南省 | 179 | 中国云南西双版纳州勐仑镇植物园白果园 | 1 |
| 云南 | 111 | 中国云南西双版纳州勐腊县回都河山坡 | 1 |
| 云 南 | 29 | 中国云南西双版纳勐仑镇植物园藤本区 | 1 |
| 云南 | 16 | 中国云南西双版纳勐腊县观榄里 | 1 |
| 云南 ; YUNNAN | 13 | 中国云南思茅市大寨 | 1 |
| 云南、西藏东南； Yunnan-Sout | 10 | 中国云南澜沧江思茅港榆绿木 | 1 |
| 云南县 | 9 | 浙江云南省 | 1 |
| 云南、四川； YUNNAN, SZECH | 9 | 云南省或湖南省 | 1 |
| 云南； YUNAN | 5 | 云南` | 1 |
| 云南、西藏； YUNNAN, TIBET | 5 | 云南， 四川； Yunnan, Szec | 1 |
| 中国云南西双版纳景洪热带作物研究所 | 4 | 云南 | 1 |
| 云南； Northwestern Yun | 4 | 共计 | 912600 |

CVH标本总体情况



CVH各馆标本数量比例

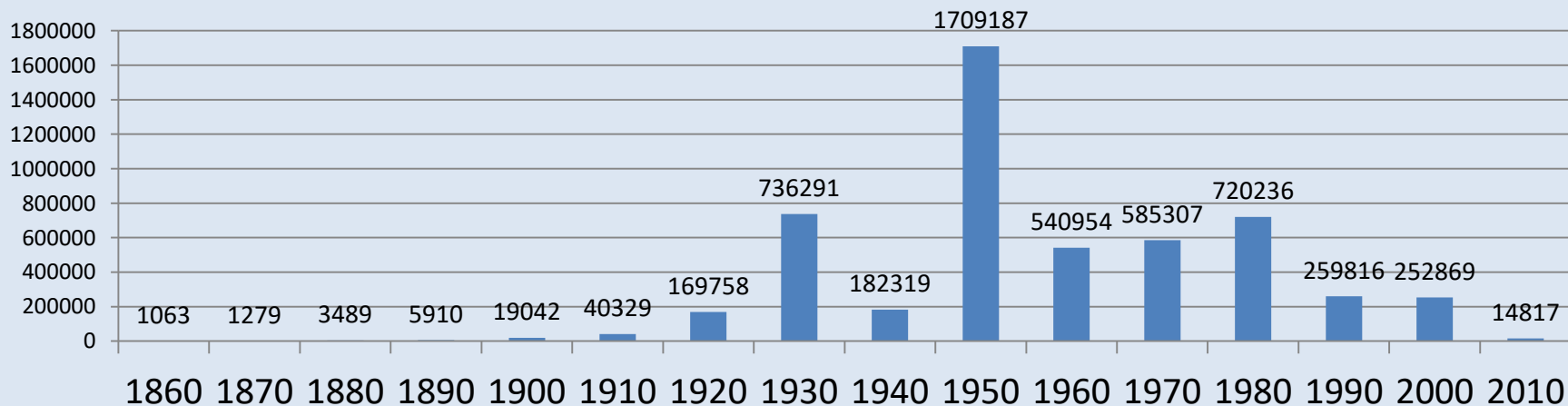


国外标本数量排名前十的国家



CVH标本分析

- 植物标本采集在时间上的特点

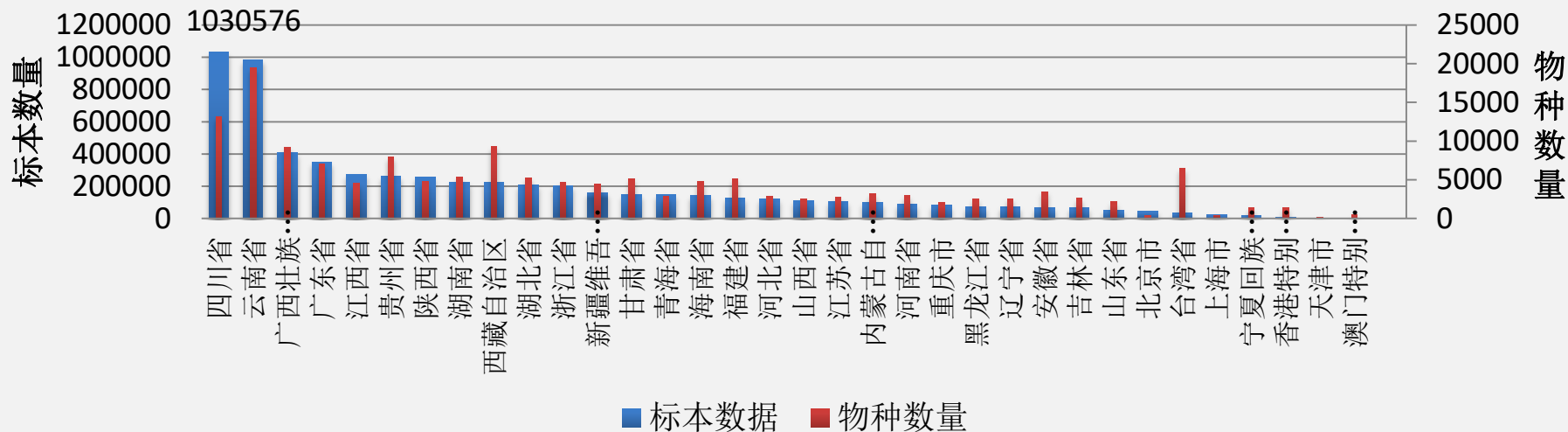


各年代标本采集量统计

CVH标本分析

- 植物标本采集在空间上的特点

各省标本数量和物种数量对比图



CVH标本分析

• CVH标本类群初步统计

| | 全 部 | | | 国 产 | | |
|------|-----|-------|-------|-----|------|-------|
| | 科数 | 属数 | 种数 | 科数 | 属数 | 种数 |
| 苔藓植物 | 153 | 902 | 4552 | 118 | 662 | 2975 |
| 蕨类植物 | 62 | 663 | 7126 | 62 | 513 | 4718 |
| 裸子植物 | 13 | 116 | 1004 | 13 | 95 | 691 |
| 被子植物 | 279 | 10873 | 46279 | 267 | 6975 | 32016 |
| 合计 | 507 | 12554 | 58961 | 460 | 8245 | 40400 |



CVH标本分析

CVH标本各级别覆盖比率

| 科级 | 类群\来源 | Col-China | CVH | 覆盖率 |
|----|-----------|------------|------------|------------|
| | 被子 | 264 | 263 | 99.6% |
| | 裸子 | 10 | 10 | 100% |
| | 蕨类 | 40 | 39 | 98% |
| | 苔藓 | 151 | 148 | 98% |
| | 合计 | 465 | 460 | 99% |

| 属级 | 类群\来源 | Col-China | CVH | 覆盖率 |
|----|-----------|-------------|-------------|------------|
| | 被子 | 3185 | 3093 | 97% |
| | 裸子 | 45 | 45 | 100% |
| | 蕨类 | 178 | 172 | 97% |
| | 苔藓 | 595 | 543 | 91% |
| | 合计 | 4003 | 3853 | 96% |

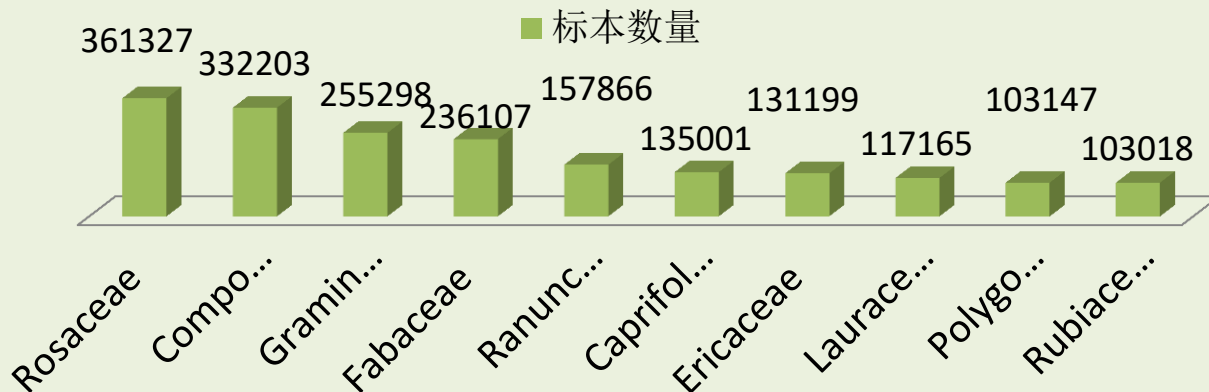
| 种级 | 类群\来源 | Col-China | CVH | 覆盖率 |
|----|-----------|--------------|--------------|------------|
| | 被子 | 31098 | 30779 | 99% |
| | 裸子 | 264 | 311 | 118% |
| | 蕨类 | 2252 | 1994 | 89% |
| | 苔藓 | 3031 | 2572 | 85% |
| | 合计 | 36645 | 35656 | 97% |

注：Col-China为中国生物物种名录2018年光盘版

CVH标本分析

- CVH平均每种标本量为209份
- 标本采集量超过10万份的科共10个科

标本数量在10万以上的科



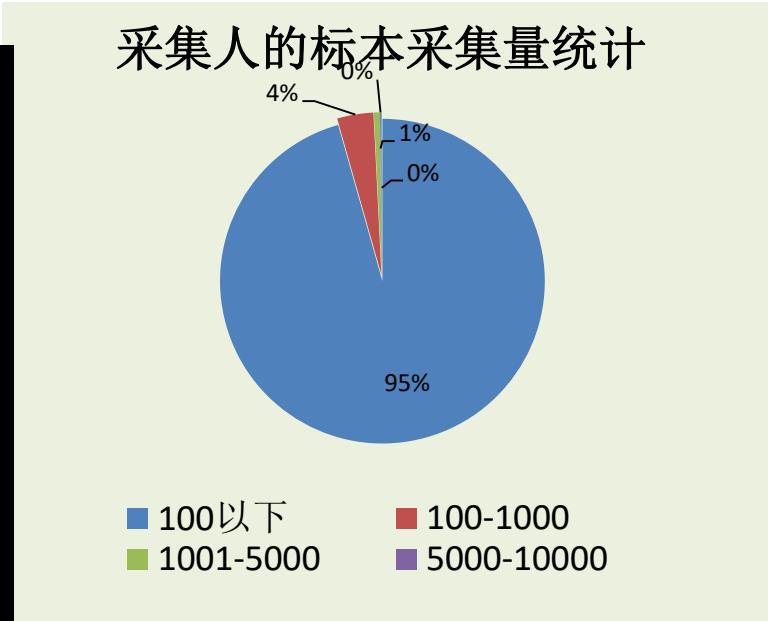
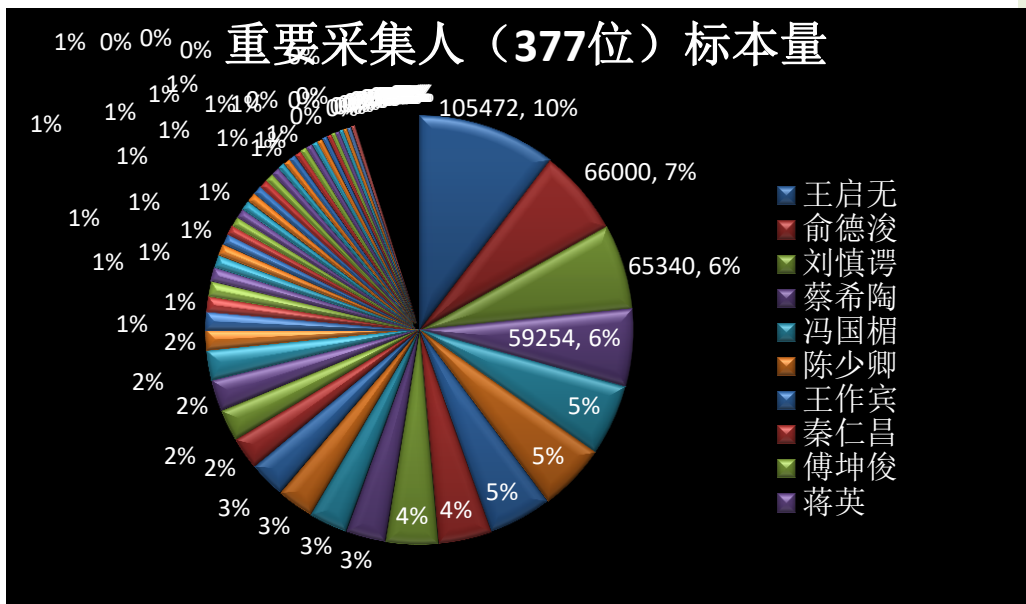
CVH标本分析

- 所有物种均有标本收录的科----103科

所有物种均有标本收录的前10科

| Family | 科名 | 该科总物种数 | 有标本的物种数 | 有标本的物种比例 (%) |
|-----------------|------|--------|---------|--------------|
| Combretaceae | 使君子科 | 31 | 31 | 100 |
| Rhizophoraceae | 红树科 | 19 | 19 | 100 |
| Linaceae | 亚麻科 | 18 | 18 | 100 |
| Altingiaceae | 阿丁枫科 | 16 | 16 | 100 |
| Balanophoraceae | 蛇菰科 | 15 | 15 | 100 |
| Burseraceae | 橄榄科 | 15 | 15 | 100 |
| Connaraceae | 牛栓藤科 | 15 | 15 | 100 |
| Verbenaceae | 马鞭草科 | 13 | 13 | 100 |
| Opiliaceae | 山柚子科 | 11 | 11 | 100 |
| Gnetaceae | 买麻藤科 | 11 | 11 | 100 |

标本采集人分析





重要采集人的统计情况

| 标本采集量 | 中文名 | 主要采集年代 | 标本馆藏 | 主要采集省 |
|--------|-----|---------------------------------|---|--------|
| 105472 | 王启无 | 1936; 1935 | PE; KUN; IBSC | 云南 |
| 70182 | 俞德浚 | 1935-36; 1937-38 | PE; KUN | 云南; 四川 |
| 61744 | 刘慎谔 | 1945; 1929-32; 1938-41; 1955-56 | PE; IFP; IBSC; WUK; KUN | 云南 |
| 59682 | 蔡希陶 | 1932-34 | KUN; PE; IBSC; LBG; SZ | 云南 |
| 55249 | 冯国楣 | 1939-40; 1959; 1947; 1962 | KUN; PE; IBSC | 云南 |
| 49700 | 王作宾 | 1933-39 | WUK; PE | 四川; 陕西 |
| 41366 | 秦仁昌 | 1956 | PE; KUN | 安徽; 浙江 |
| 40877 | 傅坤俊 | 1937-39 | UK; PE; KUN | 陕西; 甘肃 |
| 36749 | 李洪钧 | 1956-58 | IBSC; WUK; PE; HIB | 湖北; 湖南 |
| 36583 | 张志英 | 1958-1959; 1963; | LBG; WUK | 陕西; 甘肃 |
| 32956 | 贺贤育 | 1957-58; 1934-35 | WUK; IBSC; HHBG; NAS; PE | 浙江; 四川 |
| 31257 | 戴天伦 | 1958-59 | NAS; IBK; SZ; PE; IBSC; KUN; WUK; CDBI; QTPMB | 四川; 重庆 |
| 30516 | 蒋英 | 1929-30; 1932-33 | IBSC; NAS; PE | 广东 |

| 鉴定人 | 鉴定数量 |
|----------------|--------|
| 陈少卿 | 104502 |
| 吴征镒 | 105196 |
| 王文采 | 72246 |
| 陈艺林 | 49675 |
| 陈焕镛 | 44441 |
| 关克俭 | 42190 |
| 李安仁 | 42165 |
| 杨金祥 | 37057 |
| 李锡文 | 35278 |
| Miyoshi Furuse | 32918 |
| E.D.Merrill | 31460 |
| 秦仁昌 | 29970 |
| 傅书遐 | 28900 |
| 方文培 | 27999 |
| 李冀云 | 27945 |
| 郎楷永 | 27330 |
| 姚淦 | 26672 |
| 陈彦生 | 24965 |
| 邓懋彬 | 22932 |
| 李恒 | 22749 |

标本份数在3万份以上的采集人主要采集情况

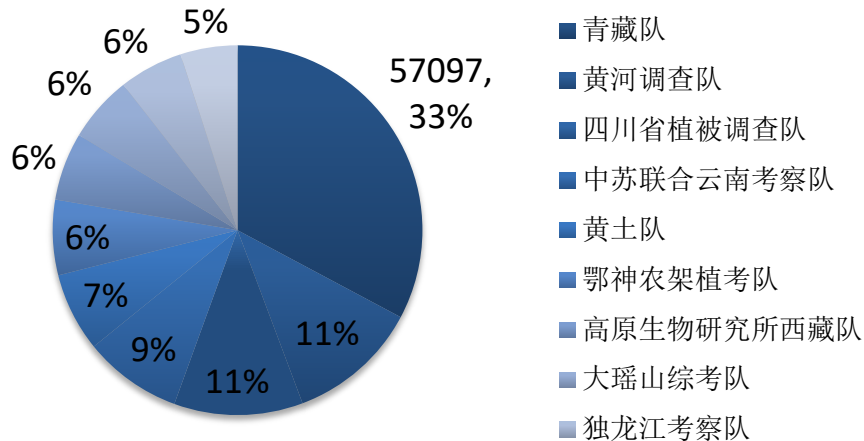
鉴定标本量位于前20的鉴定人



标本采集队分析

- 采集人数在2人以上的为采集队
- 采集量在300份以上的采集队共479支

主要采集队标本量 (TOP10)



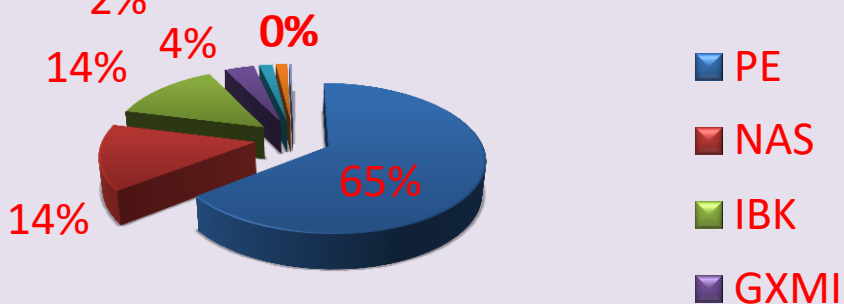


CVH模式标本

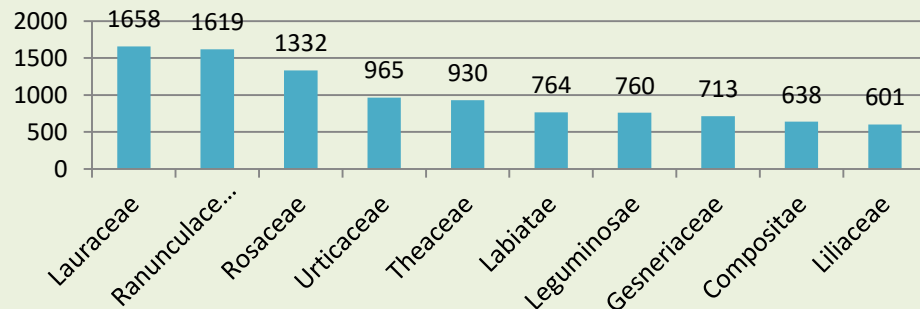
CVH模式标本数量

| | 科 | 属 | 种 | 合计 |
|--------|-----|------|-------|-------|
| 模式标本数量 | 239 | 1497 | 10973 | 27635 |

CVH模式标本各馆比例



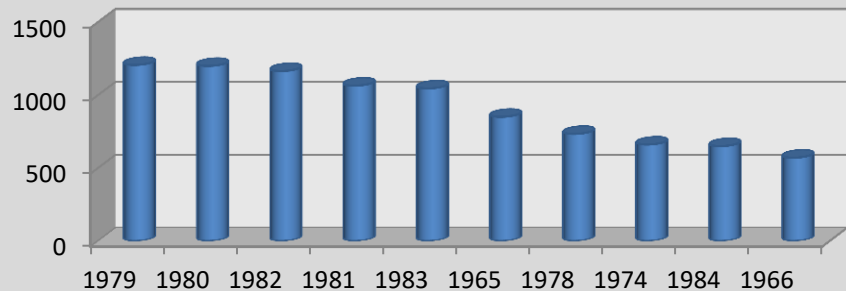
前十科模式标本数量



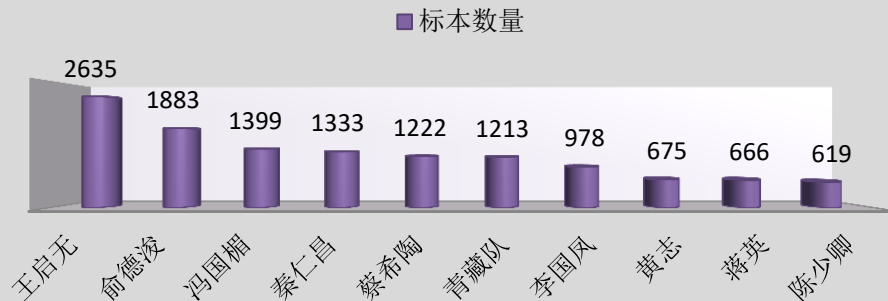


CVH模式标本

发表物种最多的年份(Top 10)



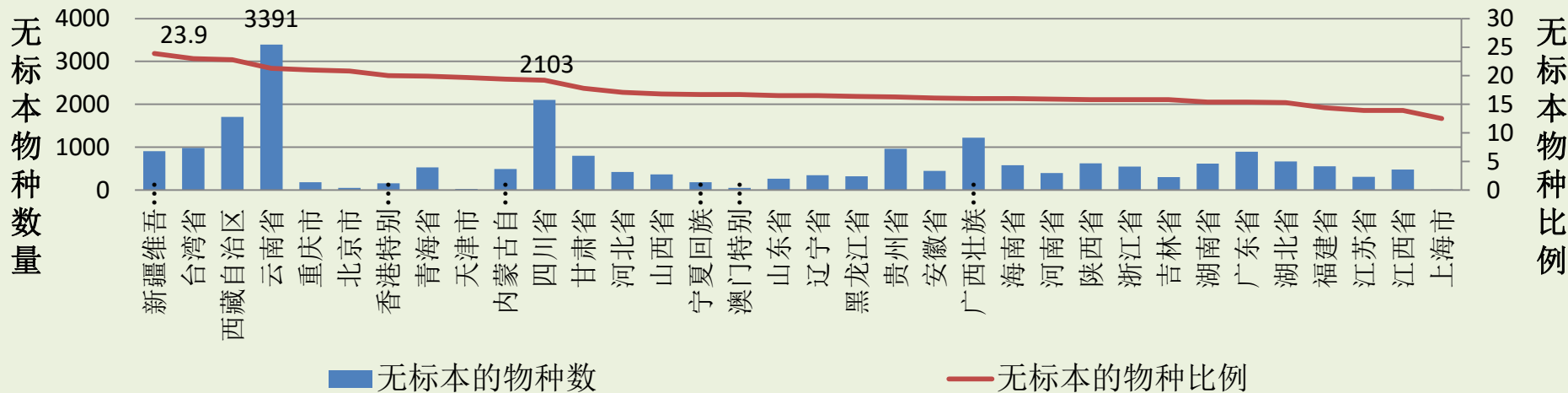
标本采集最多的采集人 (Top10)



标本采集空缺分析

采集薄弱省份

各省中无标本的物种情况



标本采集薄弱情况

- 采集薄弱类群

- 无标本物种数为7671,297科

种标本缺乏率最多的前10科

| Fam | 科名 | 种数 | 种 (无标本) | 属 (无标本) | 种 (标本缺乏率) |
|--------------|------|------|---------|---------|-----------|
| Gesneriaceae | 苦苣苔科 | 689 | 415 | 27 | 60.2 |
| Musaceae | 芭蕉科 | 37 | 21 | 3 | 56.8 |
| Magnoliaceae | 木兰科 | 142 | 64 | 8 | 45.1 |
| Begoniaceae | 秋海棠科 | 223 | 94 | 1 | 42.2 |
| Orchidaceae | 兰科 | 1534 | 637 | 131 | 41.5 |
| Araceae | 天南星科 | 236 | 91 | 18 | 38.6 |
| Poaceae | 禾本科 | 2292 | 834 | 132 | 36.4 |
| Juglandaceae | 胡桃科 | 30 | 10 | 3 | 33.3 |
| Fabaceae | 豆科 | 2349 | 763 | 88 | 32.5 |
| Asparagaceae | 天门冬科 | 339 | 110 | 15 | 32.4 |



标本采集薄弱情况

- 每种标本份数在1-4份的物种共5558种，15%
标本量低于4份的前十科

| Fam | 科 | 种数 | 种 (有标本) | 种数 (1-4份物种) | 种数 (1-4份物种) 占本科有标本物种比例 | 种数 (1-4份物种) 占本科物种比例 |
|------------------|------|------|---------|-------------|------------------------|---------------------|
| Musaceae | 芭蕉科 | 37 | 16 | 7 | 43.8 | 18.9 |
| Hydrocharitaceae | 水鳖科 | 46 | 36 | 15 | 41.7 | 32.6 |
| Thymelaeaceae | 瑞香科 | 132 | 102 | 34 | 33.3 | 25.8 |
| Begoniaceae | 秋海棠科 | 223 | 129 | 43 | 33.3 | 19.3 |
| Orchidaceae | 兰科 | 1534 | 897 | 277 | 30.9 | 18.1 |
| Eriocaulaceae | 谷精草科 | 38 | 26 | 8 | 30.8 | 21.1 |
| Zingiberaceae | 姜科 | 271 | 187 | 57 | 30.5 | 21 |
| Acanthaceae | 爵床科 | 328 | 240 | 73 | 30.4 | 22.3 |
| Urticaceae | 荨麻科 | 630 | 432 | 131 | 30.3 | 20.8 |
| Papaveraceae | 罂粟科 | 516 | 352 | 106 | 30.1 | 20.5 |



标本采集薄弱情况

- 无标本特有种数为6139种，占全部特有种数的30%

无标本特有种数量前十科

| Fam | 科 | 种数 | 无标本种 (all) | 无标本种 (特有) | 无标本的特有种比例 |
|------------------|------|-----|------------|-----------|-----------|
| PHRYMACEAE | 透骨草科 | 44 | 7 | 7 | 100 |
| ACTINIDIACEAE | 猕猴桃科 | 93 | 7 | 7 | 100 |
| HYDRANGEACEAE | 绣球花科 | 162 | 10 | 10 | 100 |
| CLUSIACEAE | 藤黄科 | 22 | 1 | 1 | 100 |
| ARISTOLOCHIACEAE | 马兜铃科 | 100 | 21 | 21 | 100 |
| SAPOTACEAE | 山榄科 | 28 | 1 | 1 | 100 |
| BERBERIDACEAE | 小檗科 | 356 | 67 | 63 | 94.03 |
| THEACEAE | 山茶科 | 206 | 28 | 26 | 92.86 |
| GESNERIACEAE | 苦苣苔科 | 689 | 415 | 376 | 90.6 |
| PENTAPHYLACACEAE | 五列木科 | 153 | 10 | 9 | 90 |
| THYMELAEACEAE | 瑞香科 | 132 | 30 | 27 | 90 |

标本采集薄弱情况

- 每种标本为1-4份的特有种为3548，占全部特有种的18%

特有物种标本份数低于4份的前十科

| 科 | 种数 | 种 (有标本) | 种数 (1-4份物种) | 种数 (1-4份特有种) | 种数 (1-4份特有种) 占本科有标本特有种比例 | 种数 (1-4份物种) 占本科物种比例 |
|------|-----|---------|-------------|--------------|--------------------------|---------------------|
| 猕猴桃科 | 93 | 86 | 9 | 9 | 100 | 10.5 |
| 绣球花科 | 162 | 152 | 21 | 20 | 95.2 | 13.2 |
| 小檗科 | 356 | 289 | 63 | 59 | 93.7 | 20.4 |
| 冬青科 | 270 | 217 | 45 | 42 | 93.3 | 19.4 |
| 马兜铃科 | 100 | 79 | 20 | 18 | 90 | 22.8 |
| 瑞香科 | 132 | 102 | 34 | 30 | 88.2 | 29.4 |
| 凤仙花科 | 283 | 206 | 49 | 43 | 87.8 | 20.9 |
| 天门冬科 | 339 | 229 | 48 | 42 | 87.5 | 18.3 |
| 苦苣苔科 | 689 | 274 | 79 | 69 | 87.3 | 25.2 |
| 灯心草科 | 128 | 92 | 15 | 13 | 86.7 | 14.1 |



标本采集薄弱情况

- 珍稀濒危物种（CR,EN,VU,NT）无标本共688种，占全部受威胁物种的18.3%。

无标本的珍稀濒危物种前十科

| 科 | 种数 | 无标本种 (CrEnVuNt) | 种 (CrEnVuNt) | 受威胁物种的无 标本率 |
|------|------|--------------------|--------------|----------------|
| 兰科 | 1534 | 243 | 791 | 30.7 |
| 杜鹃花科 | 1021 | 23 | 227 | 10.1 |
| 报春花科 | 750 | 37 | 226 | 16.4 |
| 豆科 | 2349 | 28 | 189 | 14.8 |
| 樟科 | 528 | 15 | 179 | 8.4 |
| 蔷薇科 | 1708 | 27 | 145 | 18.6 |
| 毛茛科 | 1376 | 23 | 143 | 16.1 |
| 莎草科 | 1054 | 34 | 132 | 25.8 |
| 唇形科 | 1327 | 14 | 118 | 11.9 |
| 禾本科 | 2292 | 44 | 115 | 38.3 |

标本采集薄弱情况

- 珍稀濒危物种每种标本份 ≤ 4 的物种为1171种，占所有受威胁物种的31%

珍稀濒危物种每种标本量低于4份的前十科

| 科 | 种数 | 种 (有标本) | 种数 (1-4份物种) | 种数 (1-4份珍稀濒危数) | 受威胁1-4份比例 | 受威胁1-4份比例2 |
|------|------|---------|-------------|----------------|-----------|------------|
| 兰科 | 1534 | 897 | 277 | 178 | 64.3 | 19.8 |
| 阿丁枫科 | 13 | 13 | 2 | 2 | 100 | 15.4 |
| 谷精草科 | 38 | 26 | 8 | 4 | 50 | 15.4 |
| 马兜铃科 | 100 | 79 | 20 | 12 | 60 | 15.2 |
| 姜科 | 271 | 187 | 57 | 26 | 45.6 | 13.9 |
| 木兰科 | 142 | 78 | 14 | 10 | 71.4 | 12.8 |
| 苏铁科 | 28 | 25 | 6 | 3 | 50 | 12 |
| 牛栓藤科 | 9 | 9 | 2 | 1 | 50 | 11.1 |
| 天南星科 | 236 | 145 | 43 | 15 | 34.9 | 10.3 |
| 山茶科 | 206 | 178 | 32 | 18 | 56.2 | 10.1 |



小结

- 标本采集与时代发展紧密相关（经济、学科、政治）
- 标本采集多集中在生物多样性较高的地区
- 标本采集仍不完整（类群、省级、县级）
 - 继续保持数字化增量（薄弱地区）
 - 加强采集薄弱地区和薄弱类群的采集
- 标本数据质量有待提高

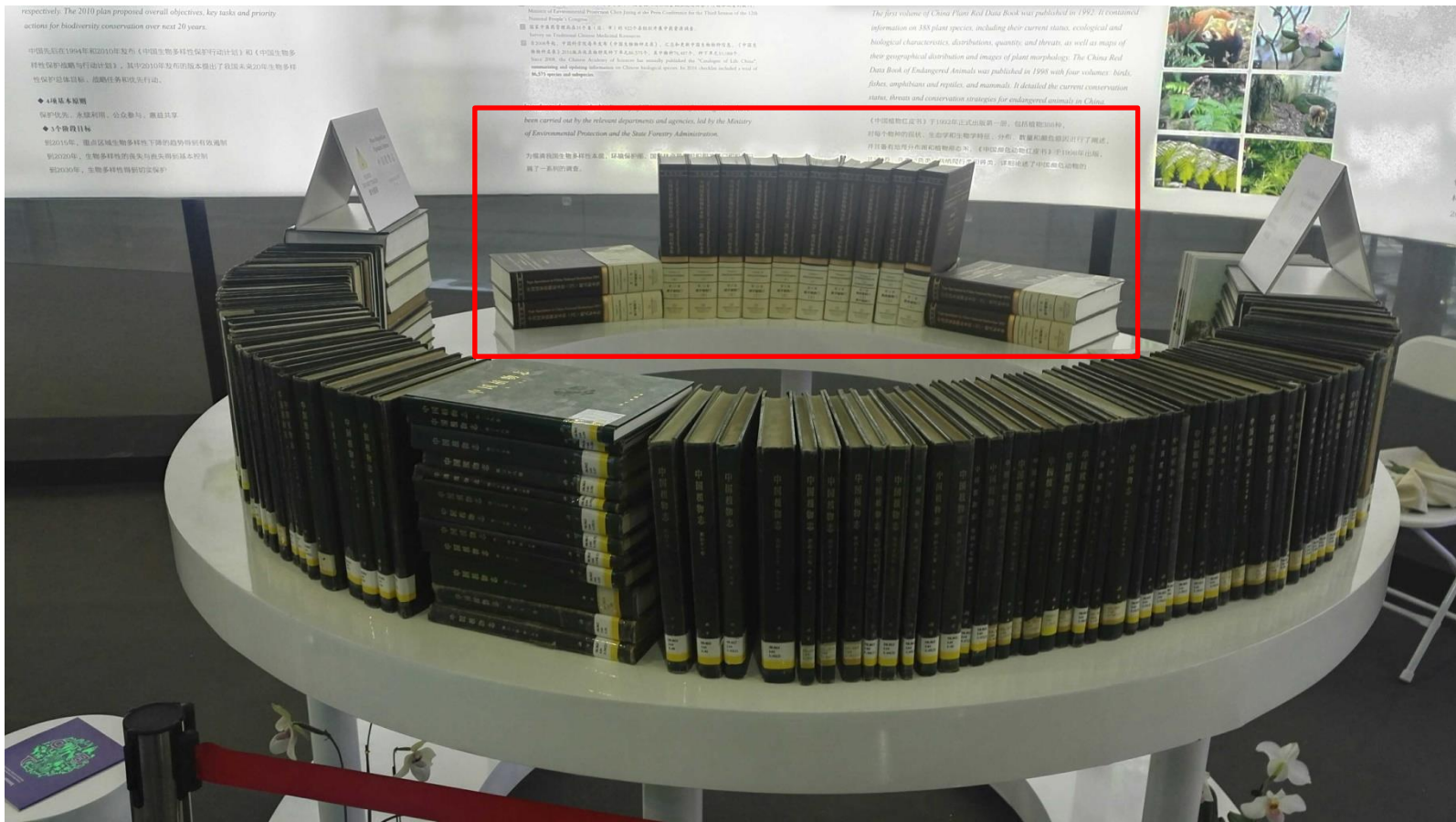


科研支撑

- 2007年-2017年
- 670篇科研论文=601(国内)+69(SCI)
- 涉及领域-自然科学（生态，保护，外来入侵等分布预测的文章比例最高），中医药，教学。
- 支撑项目-973项目4个；国家自然科学基金委国际重大合作项目1个；国家自然科学基金16个；中科院项目6个；其他60余个
- 论著30余本（中国国家植物标本馆（PE）模式标本集15卷）



PE模式集亮相第十九届国际植物学大会





中国数字植物标本馆
Chinese Virtual Herbarium

谢谢！



中国数字植物标本馆
Chinese Virtual Herbarium