



•技术与方法•

获取生物物种名录信息的R程序包SP2000

丁刘勇^{1,2}, 李昊³, 陶捐^{1,2}, 张金龙⁴, 黄敏睿^{1,2}, 杨科^{1,2}, 王军^{1,2}, 丁城志^{1,2*}, 何大明^{1,2*}

1. 云南大学国际河流与生态安全研究院, 昆明 650504; 2. 云南大学云南省国际河流与跨境生态安全重点实验室, 昆明 650504; 3. 云南大学国家示范性软件学院, 昆明 650504; 4. 香港嘉道理农场暨植物园植物保育部, 香港 999077

摘要: 物种名录为衡量区域和全球生物多样性提供了数据基础。随着互联网的兴起与发展, 人们将地球上已知的动物、植物、微生物等类群的物种名录信息存储到公共数据平台中, 并对物种名录进行快速及时地更新, 这极大地促进了分类学、保护生物学和宏观生态学等学科的发展, 成为政府或国际组织开展物种保育现状评估、红色名录编撰和生物多样性保护的重要依据。物种2000中国节点(<http://www.sp2000.org.cn>)和Catalogue of Life网站(<http://www.catalogueoflife.org>)分别是中国和全球最大的生物物种名录数据平台, 截至2020年6月4日, 其收录的物种数分别为122,280种和1,829,672种。然而这些数据平台仅提供物种查询、检索、下载等基本功能, 难以满足使用者准确、快速地批量获取所需生物物种名录数据信息的需求, 制约了这些大数据平台在生物多样性研究和保护中的作用。因此, 我们选取R语言开发了程序包SP2000, 旨在帮助用户批量获取中国或全球生物物种名录信息。该程序包具有跨Windows、MacOS、Linux等多个系统运行、操作便捷、代码开源等特点。为了方便用户使用, 本文详细介绍了SP2000的基本原理、特点及使用指南, 包括程序包的下载、安装、运行和参数设置等。

关键词: 物种名录; 红色名录; 中国生物多样性; R语言程序包

丁刘勇, 李昊, 陶捐, 张金龙, 黄敏睿, 杨科, 王军, 丁城志, 何大明 (2021) 获取生物物种名录信息的R程序包SP2000. 生物多样性, 29, 118–122. doi: 10.17520/biods.2020235.

Ding LY, Li H, Tao J, Zhang JL, Huang MR, Yang K, Wang J, Ding CZ, He DM (2021) SP2000: An open-sourced R package for querying the Catalogue of Life. Biodiversity Science, 29, 118–122. doi: 10.17520/biods.2020235.

SP2000: An open-sourced R package for querying the Catalogue of Life

Liuyong Ding^{1,2}, Hao Li³, Juan Tao^{1,2}, Jinlong Zhang⁴, Minrui Huang^{1,2}, Ke Yang^{1,2,4}, Jun Wang^{1,2}, Chengzhi Ding^{1,2*}, Daming He^{1,2*}

1 Institute of International Rivers and Eco-security, Yunnan University, Kunming 650504

2 Yunnan Key Laboratory of International Rivers and Transboundary Eco-security, Yunnan University, Kunming 650504

3 National Pilot School of Software, Yunnan University, Kunming 650504

4 Flora Conservation Department, Kadoorie Farm and Botanic Garden, Hong Kong 999077

ABSTRACT

Aims: The Catalogue of Life provides the basis for understanding both regional and global biodiversity. With the invention and development of the internet, the up-to-date species checklists stored in the public databases has greatly promoted the development of taxonomy, conservation biology, and macroecology. Public species checklists play an indispensable role in biodiversity conservation and aid in the assessment of species' conservation status. The Species 2000 China Node (<http://www.sp2000.org.cn>) and the Catalogue of Life (<http://www.catalogueoflife.org>) are among the leading online databases in cataloguing biodiversity, contain 122,280 and 1,829,672 taxa respectively (including infraspecific taxa). Although searching the content of the websites may be relatively straightforward, downloading the data and transferring it into a statistical environment for further analysis can present challenges.

Method: To address this issue, we developed the package SP2000 using the R programming language.

收稿日期: 2020-06-11; 接受日期: 2020-08-12

基金项目: 生态环境部生物多样性调查评估项目(2019HJ2096001006)、国家重点研发计划(2016YFA0601600)和云南大学研究生科研创新基金(2019227)

* 共同通讯作者 Co-authors for correspondence. E-mail: chzhding@ynu.edu.cn; dmhe@ynu.edu.cn

Application: SP2000 is an open-source, cross-platform, and user-friendly package which aims to help users query and download the checklist of organisms (including animals, plants, fungi, and microbes) from within and outside China. Here we introduce and describe the usage of SP2000 including installation, and configuration of parameters.

Key words: species checklist; redlist; China's biodiversity; R package

物种名录是人们评估生物多样性状况、开展生物多样性保护和管理的基礎(Reichhardt, 1999; Banki et al, 2019; Ower & Roskov, 2019)。随着互联网的发展,人们将地球上已知的动物、植物和微生物的物种名录信息存储到公共数据平台中,并对物种名录进行快速及时(如年度、季度和月度)的更新,这极大地促进了分类学、保护生物学和宏观生态学等学科的发展(蒋志刚等, 2015)。目前,这些数据库已经成为开展物种保育现状评估、红色名录编撰和生物多样性保护的重要依据。

物种2000中国节点(Species 2000 China Node, <http://www.sp2000.org.cn/>)是由中国科学院生物多样性委员会(The Biodiversity Committee of Chinese Academy of Sciences, BC-CAS)于2006年与其合作单位(中国科学院植物研究所、动物研究所、微生物研究所、海洋研究所)共同建设的中国生物多样性大数据平台。自2008年以来,该平台每年更新中国生物物种名录信息,并通过网络和光盘向社会公布,是世界上唯一连续多年发布年度生物物种名录的国家生物多样性数据平台(蒋志刚等, 2015; 马克平等, 2018)。2020年5月22日发布的《中国生物物种名录2020版》共收录物种及种下单元122,280个(物种110,231个,种下单元12,049个),其中包括动物54,359种,植物37,793种,真菌12,506种(The Biodiversity Committee of Chinese Academy of Sciences, 2020)。Catalogue of Life (<http://www.catalogueoflife.org/>)是全球最大的生物物种名录数据平台,截至2020年6月4日,该平台收录了全球已知的动物、植物和微生物等类群现存物种1,829,672种、灭绝物种38,145种。《全球生物物种名录》与《中国生物物种名录》都是生物物种数据的汇编,前者是全球生物物种名录研究的集合,后者则是国别研究的子集,两者研究的空間尺度不同。《全球生物物种名录》为大尺度生物多样性研究提供了大数据;国别生物物种名录确定了各国物种名录信息,填补了前者的知识空缺,两份生物物种名录互为补充。与《全球生物物种名录》相比,《中国生物物种名录》

除了包含每个物种的科学名、同物异名、别名、参考文献、分类系统、分布区等数据外,还提供了中文名和中文名的汉语拼音等内容,目前通过光盘和网站数据库的方式为用户提供信息服务。然而上述两个数据平台仅提供物种查询、检索、下载等基本功能,很难满足使用者准确、快速地批量获取所需生物物种名录数据的需求,制约了生物物种名录数据的广泛使用和高效利用。

针对以上问题,我们选取R编程语言(R Core Team, 2020),开发了SP2000程序包(<https://cran.r-project.org/package=SP2000/>),旨在帮助用户从物种2000中国节点和Catalogue of Life网站准确、快速查找和批量获取所需生物物种名录数据。R语言具有开源、在线资源丰富、支持多操作系统、易掌握等特点(Tippmann, 2014; 张金龙等, 2016; Lai et al, 2019)。R程序包SP2000的基本原理、特点和使用指南见下。

1 SP2000的基本原理

R程序包SP2000的编写基于物种2000中国节点网站提供的API服务(<http://sp2000.org.cn/api/document/>, version 2)和Catalogue of Life数据平台提供的web服务(<http://webservice.catalogueoflife.org/col/webservice/>, version 1.9)。

2 SP2000的特点

相比其他生物物种名录获取工具,如浏览器插件(Catalogue of Life Search Plugin, <http://www.catalogueoflife.org/content/web-browser-page-plugin>)、《中国生物物种名录》光盘(<http://sp2000.org.cn/download/>)和JavaScript语言开发的portal-components检索工具(<https://github.com/CatalogueOfLife/portal-components/>), R程序包SP2000具有以下特点: (1)可在Windows、MacOS、Linux等多个操作系统上运行; (2)操作便捷; (3)代码开源,程序包可及时更新; (4)查询返回data.frame数据类型,易于数据的读取、储存和挖掘,如输出的

物种名录数据可通过R程序包spocc (Chamberlain, 2020)跨生物多样性数据平台进一步挖掘物种的更多信息。

3 SP2000的使用指南

3.1 版本与安装

R语言编写的程序包SP2000已被R语言官网接收(<https://cran.r-project.org/package=SP2000/>), 源代码详见<https://github.com/Otoliths/SP2000/>。用户可直接通过R命令install.packages("SP2000")安装, 也可以通过devtools::install_github("Otoliths/SP2000")安装。当前发行的SP2000程序包版本为0.1.0, 主要包含8个函数: set_search_key、search_family_id、search_taxon_id、search_checklist、get_redlist_china、get_col_global、find_synonyms和get_col_taiwan。其详细使用指南见附录1。

3.2 set_search_key设置Species 2000密钥

函数set_search_key只有1个参数key。在运行函数search_*之前需先申请“Species 2000”密钥, 注册物种2000中国节点网站, 点击“用户信息”便可获取一个自己独享的API密钥服务, 然后调用一次函数set_search_key("your key"), 便可运行所有search_*函数(即search_family_id、search_taxon_id和search_checklist)。

3.3 search_family_id查询中国生物物种名录的科id

科是生物学分类中最常用的分类等级, 通过科可更方便查询到种或种下的分类单位。物种2000中国节点数据平台对分类单位科、种(种下)都编写了id, 具有唯一性, 保证了数据查询的准确率。函数search_family_id提供了查询科id集合的功能, 它有4个参数: query、start、limit和mc.cores。(1)参数query输入中国生物的科名或科名的一部分、或科拉丁名和中文名, 支持单个或多个变量查询; (2)参数start设置单次查询开始的页数, 默认值为1; (3)参数limit设置单次查询返回的记录, 默认值为20; (4)参数mc.cores设置计算机的内核数量, 缩短查询时间, 默认值为2。以“鳗鲡科(Anguillidae)”为例, 调用方式为search_family_id(query = "Anguillidae", mc.cores = 2)。

3.4 search_taxon_id查询中国生物物种名录的种id

通过科id可以直接获取种或种下的id(唯一性)列表, 然后使用函数search_checklist获取物种名录

的详细信息。函数search_taxon_id支持科id、物种学名和中文名查询种或种的id, 它有5个参数: query、name、start、limit和mc.cores。(1)参数query输入函数search_family_id查询返回的科名id、或中国生物物种的学名和中文名, 支持单个或多个变量查询; (2)参数name设置查询方式, 与参数query配合使用, 可选择参数有“familyID”, “scientificName”和“commonName”, 默认值为“scientificName”; (3)参数start设置单次查询开始的页数, 默认值为1; (4)参数limit设置单次查询返回的记录, 默认值为20; (5)参数mc.cores设置同3.3。函数search_taxon_id可以与search_family_id配合使用, 以“鳗鲡科(Anguillidae)”为例, R代码如下:

```
## 加载包
library("SP2000")
## 设置“Species 2000”密钥
set_search_key <- "your apikey"
## 根据科名查询, 返回科id集合
familyid <- search_family_id(query = "Anguillidae")
## 根据科id返回种或种下的id列表
query <- familyid$Anguillidae$data$record_id
taxonid <- search_taxon_id(query = query, name = "familyID")
```

3.5 search_checklist查询中国生物物种名录信息

函数search_checklist的主要功能是通过种或种下id(唯一性)查询返回物种的详细信息, 包括科学名、同物异名、别名、文献、分类系统、分布区等数据, 以及中文名和中文名的汉语拼音等内容。该函数需要与函数search_family_id和search_taxon_id配合使用, 有2个参数: query和mc.cores。(1)参数query输入函数search_taxon_id查询返回的种或种下id集合, 支持单个或多个变量查询; (2)参数mc.cores同3.3。

以3.4查询结果taxonid为例, 调用方式为:

```
query <- taxonid[["3851c5311bed46c19529cb155d37aa9b"]][["data"]][["namecode"]]
```

```
search_checklist(query = query)
```

3.6 get_redlist_china查询中国物种红色名录信息

函数get_redlist_china有4个参数: query、option、group和viewDT。(1)参数query输入科学名或中文名;

(2)参数option设置查询方式,与参数query配合使用,可选择参数有“Chinese Names”和“Scientific Names”,默认值为“Scientific Names”;(3)参数group设置查询类群,可选择参数包括“Amphibians”,“Birds”,“Mammals”,“Inland Fishes”,“Reptiles”,“Plants”和“Fungi”,分别是两栖类、鸟类、哺乳类、内陆鱼类、爬行类、植物和真菌;(4)参数viewDT是逻辑值,与参数group配合使用,如果viewDT = TRUE,查询结果显示一个交互性页面,以“内陆鱼类(Inland Fishes)”为例,调用方式为get_redlist_china (... , group = "Inland Fishes", viewDT = TRUE)。

以“鳗鲡属(Anguilla)”查询为例,R代码如下:

```
## 加载包
```

```
library("SP2000")
```

```
## 查询鳗鲡属的中国红色名录信息
```

```
get_redlist_china (query = "Anguilla", option = "Scientific Names")
```

```
## 查询“内陆鱼类”中国红色名录信息,显示一个可查询、下载的交互性页面
```

```
get_redlist_china (group = "Inland Fishes", viewDT = TRUE)
```

3.7 get_col_global查询全球生物物种名录信息

函数get_col_global运行不受“Species 2000”密钥限制,可独立使用,它包含6个参数: query、option、response、start、limit和mc.cores。(1)参数query输入种或种下id、物种名;(2)参数option设置查询方式,与参数query配合使用,可选择参数有“id”和“name”,默认值为“name”;(3)参数response设置查询返回结果,可选择“full”或“terse”,“full”返回完整查询结果,“terse”返回简短查询结果,默认值为“terse”;(4)参数start设置查询返回的第一个记录,默认值为0,与参数response配合使用;(5)参数limit设置单次查询返回的记录,默认值为500,单次简短查询返回结果的最大数量为500,单次完整查询返回结果的最大数量为50;(6)参数mc.cores设置同3.3。

以批量获取全球“鳗鲡属(Anguilla)”名录信息为例,R代码如下:

```
## 加载包
```

```
library("SP2000")
```

```
x <- get_col_global (query = "Anguilla", response = "full")
```

```
## 总的查询结果为208
```

```
x[["Anguilla"]][["meta"]][["total_number_of_results"]] [1]
```

3.8 find_synonyms查询全球生物物种同物异名

函数find_synonyms有2个参数: query和mc.cores。(1)参数query输入物种名,(2)参数mc.cores设置计算机的内核数量,缩短查询时间,默认值为2。以“欧洲鳗鲡(Anguilla anguilla)”为例,调用方式为find_synonyms ("Anguilla anguilla")。

3.9 get_col_taiwan查询台湾生物物种名录信息


函数get_col_taiwan有4个参数: query、level、option、include_synonyms。(1)参数query查询分类名称信息;(2)参数level设置查询方式,与参数query配合使用,可选择“kingdom”,“phylum”,“class”,“order”,“family”,“genus”,“species”中的一个,分别表示界、门、纲、目、科、属和种,默认值为“species”;(3)参数option设置查询方式,可选择“contain”,“equal”和“beginning”中的一个,分别表示“包含”,“等于”,“开始于”,默认值为“equal”;(4)参数include_synonyms是逻辑值,查询结果是否包含同物异名信息,默认值为TRUE。以“鳗鲡科(Anguillidae)”为例,调用方式为get_col_taiwan(query = "Anguillidae", level = "family")。


为了更好地满足生物多样性信息时代的用户需求,在R程序包SP2000基础上,我们也使用Python语言(Perkel, 2015; Python Software Foundation, 2020)编写了程序包SP2000(<https://pypi.org/project/SP2000/>; <https://github.com/ynulihao/SP2000/>)。用户可直接通过命令pip3 install SP2000或python3 -m pip install SP2000安装,其参数设置和函数调用方式基本与R程序包SP2000一致,其使用指南详见附录2。


致谢:衷心感谢两位审稿专家以及编委对R程序包SP2000提出的宝贵意见和建议。

ORCID

丁刘勇  <https://orcid.org/0000-0002-5490-182X>

张金龙  <https://orcid.org/0000-0002-1161-5460>

王军  <https://orcid.org/0000-0003-2481-1409>

丁城志  <https://orcid.org/0000-0001-5215-7374>

参考文献

- Banki O, Hobern D, Döring M, Remsen D (2019) Catalogue of Life Plus: A collaborative project to complete the checklist of the world's species. *Biodiversity Information Science and Standards*, 3, e37652.
- Chamberlain S (2020) spocc: Interface to Species Occurrence Data Sources. R package version 1.0.8. <https://CRAN.R-project.org/package=spocc/>. (accessed on 2020-06-01)
- Jiang ZG, Qin HN, Liu YN, Ji LQ, Ma KP (2015) Protecting biodiversity and promoting sustainable development: In memory of the releasing of Catalogue of Life China 2015 and China Biodiversity Red List on the International Day for Biological Diversity 2015. *Biodiversity Science*, 23, 433–434. (in Chinese) [蒋志刚, 覃海宁, 刘忆南, 纪力强, 马克平 (2015) 保护生物多样性, 促进可持续发展——纪念《中国生物物种名录》和《中国生物多样性红色名录》发布. *生物多样性*, 23, 433–434.]
- Lai JS, Lortie CJ, Muenchen RA, Yang J, Ma KP (2019) Evaluating the popularity of R in ecology. *Ecosphere*, 10, e02567.
- Ma KP, Zhu M, Ji LQ, Ma JC, Guo QH, Ouyang ZY, Zhu L (2018) Establishing China Infrastructure for Big Biodiversity Data. *Bulletin of the Chinese Academy of Sciences*, 33(8), 80–87. (in Chinese with English abstract) [马克平, 朱敏, 纪力强, 马俊才, 郭庆华, 欧阳志云, 朱丽 (2018) 中国生物多样性大数据平台建设. *中国科学院院刊*, 33(8), 80–87.]
- Ower G, Roskov Y (2019) The Catalogue of Life: Assembling data into a global taxonomic checklist. *Biodiversity Information Science and Standards*, 3, e37221.
- Perkel JM (2015) Programming: Pick up Python. *Nature*, 518, 125–126.
- Python Software Foundation (2020) Python Language Reference, version 3.7. <https://www.python.org/>. (accessed on 2020-06-01)
- R Core Team (2020) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. <https://www.R-project.org/>. (accessed on 2020-06-01)
- Reichhardt T (1999) Catalogue of life could become reality. *Nature*, 399, 519–519.
- The Biodiversity Committee of Chinese Academy of Sciences (2020) Catalogue of Life China: 2020 Annual Checklist, Beijing. (in Chinese and in English) [中国科学院生物多样性委员会 (2020) 中国生物物种名录: 2020年版. 北京] <http://www.sp2000.org.cn/CoLChina>. (accessed on 2020-05-22)
- Tippmann S (2014) Programming tools: Adventures with R. *Nature*, 517, 109–110.
- Zhang JL, Zhu HL, Liu JG, Fischer GA (2016) Principles behind designing herbarium specimen labels and the R package 'herblabel'. *Biodiversity Science*, 24, 1345–1352. (in Chinese with English abstract) [张金龙, 朱慧玲, 刘金刚, Fischer GA (2016) 植物标本标签设计的原则及R程序包herblabel. *生物多样性*, 24, 1345–1352.]

(责任编委: 张健 责任编辑: 黄祥忠)

附录 Supplementary Material

附录1 R程序包SP2000的安装、配置和使用指南

Appendix 1 The installation, configuration and usage for R package SP2000
<http://www.biodiversity-science.net/fileup/PDF/2020235-1.pdf>

附录2 Python程序包SP2000的简介、安装、配置和使用指南

Appendix 2 The introduction, Installation, configuration and usage for Python package SP2000
<http://www.biodiversity-science.net/fileup/PDF/2020235-2.pdf>

附录3 本文英文版

Appendix 3 Full-text in English
<http://www.biodiversity-science.net/fileup/PDF/2020235-3.pdf>

附录1 R程序包SP2000的安装、配置和使用指南

Appendix 1 The installation, configuration and usage for R package SP2000

1. 安装 SP2000	1
2. 使用指南	2
2.1 加载 SP2000	2
2.2 set_search_key 设置 Species 2000 密钥.....	2
2.3 search_family_id 查询中国生物物种名录的科 id	2
2.4 search_taxon_id 查询中国生物物种名录的种 id.....	3
2.5 search_checklist 查询中国生物物种名录信息	4
2.6 get_col_global 查询全球生物物种名录信息	5
2.7 get_redlist_china 查询中国物种红色名录信息.....	5
2.8 find_synonyms 查询全球生物物种同物异名.....	5
2.9 get_col_taiwan 查询台湾生物物种名录信息.....	6
2.10 测试环境.....	6
参考文献.....	7

附录 1 文档由 R 程序包 *rmarkdown* (Allaire et al, 2020) 自动生成。使用 R 程序包 SP2000 之前, 用户需要安装 R 和 RStudio 软件, 该程序包可在 Windows、MacOS、Linux 等多个操作系统上运行。R 软件下载的地址为: <http://cran.r-project.org/bin/windows/base/>, 用户根据自己的计算机操作系统, 选取相应 R 软件, 建议用户下载最新版的 R, 因为涉及到字符编码转换, R 版本不低于 3.0.0; RStudio 软件下载的地址为: <https://rstudio.com/products/rstudio/download/>, 用户根据安装的 R 软件选取相应的 RStudio 软件, 因为旧版本不能运行 Python 代码, 建议用户下载最新版的 RStudio。下文重点介绍 R 程序包 SP2000 的安装、配置和使用指南。

1. 安装 SP2000

#Current official release

```
install.packages("SP2000", repos = "https://cran.r-project.org")
```

#Current beta / GitHub release:

```
# if (!requireNamespace("devtools", quietly = TRUE))  
#   install.packages("devtools")
```

```
devtools::install_github("Otoliths/SP2000")
```

#or

```
devtools::install_git("git://github.com/Otoliths/SP2000.git")
```

#or

```
devtool::install_gitlab("Otoliths/SP2000")
```

#Check if reticulate is installed

丁刘勇, 李昊, 陶捐, 张金龙, 黄敏睿, 杨科, 王军, 丁城志, 何大明 (2021) 获取生物物种名录信息的 R 程序包 SP2000. 生物多样性, 29 (1): 118–122. <http://www.biodiversity-science.net/CN/10.17520/biods.2020235>

```
if (!requireNamespace("reticulate", quietly = TRUE))  
  install.packages("reticulate")
```

2. 使用指南

R 程序包 SP2000 主要包含 8 个函数: `set_search_key`、`search_family_id`、`search_taxon_id`、`search_checklist`、`get_redlist_china`、`get_col_global`、`find_synonyms` 和 `get_col_taiwan`, 下面以“鳗鲡科(Anguillidae)”为例, 详细介绍以上函数的使用和测试代码的操作系统, 并且通过 R 包 `reticulate` 包提供 `repl_python()` 和 `exit` 命令, 实现 Python 代码的进入和退出, 可以使用 Python 调用 R 运行的结果, 调用方式 `r.*`。

2.1 加载 SP2000

```
#Load the SP2000 package  
library("SP2000")  
  
# Welcome to R Package SP2000 0.1.0 !!!  
  
# To start with the SP2000, please digit:  
  
# https://cran.r-project.org/package=SP2000\(R\)  
  
# https://pypi.org/project/SP2000\(Python\)  
  
# Load the reticulate package  
library("reticulate")  
  
# install IPython package  
reticulate::py_install (packages = "IPython", pip = TRUE)
```

2.2 `set_search_key` 设置 Species 2000 密钥

注意: 在运行函数 `search_family_id`、`search_taxon_id`、`search_checklist` 之前需先申请“Species 2000”密钥, 注册物种 2000 中国节点 <http://www.sp2000.org.cn> 网站, 点击“用户信息”便可获取一个自己独享的 API 密钥服务, 然后调用一次函数 `set_search_key`(“your key”), 便可运行所有 `search_*` 函数。

#Note: You need to apply for the apiKey <<http://www.sp2000.org.cn>>

#to run search_ functions of this package.*

```
set_search_key("06319834*****")
```

2.3 `search_family_id` 查询中国生物物种名录的科 id

```
#Search family IDs via family name  
#Family name, or part of family name, supports Latin and Chinese names.  
str (search_family_id (query = "鳗鲡科"))  
  
# Request returned successfully!!!  
# Download date: 2020-08-04  
# |=====| 100%, Elapsed 00:00  
# List of 1  
# $ 鳗鲡科:List of 2  
# ..$ meta:List of 5  
# ...$ code :int 200  
# ...$ limit :int 20  
# ...$ count :int 1  
# ...$ page :int 1
```

丁刘勇, 李昊, 陶捐, 张金龙, 黄敏睿, 杨科, 王军, 丁城志, 何大明 (2021) 获取生物物种名录信息的 R 程序包 SP2000. 生物多样性, 29 (1): 118–122. <http://www.biodiversity-science.net/CN/10.17520/biods.2020235>

```
# ...$ message: chr "success"
# ..$ data: tibble [1 × 14] (S3: tbl_df/tbl/data.frame)
# ...$ family_c : chr "鳗鲡科"
# ...$ phylum_c : chr "脊索动物门"
# ...$ superfamily : logi NA
# ...$ kingdom : chr "Animalia"
# ...$ record_id : chr "3851c5311bed46c19529cb155d37aa9b"
# ...$ phylum : chr "Chordata"
# ...$ kingdom_c : chr "动物界"
# ...$ family : chr "Anguillidae"
# ...$ class : chr "Actinopterygii"
# ...$ class_c : chr "辐鳍鱼纲"
# ...$ order_c : chr "鳗鲡目"
# ...$ order : chr "Anguilliformes"
# ...$ superfamily_c: logi NA
# ...$ download_date: Date[1:1], format: "2020-08-03"

familyid <- search_family_id (query = "Anguillidae")

# Request returned successfully!!!
# Download date: 2020-08-04
# |=====| 100%, Elapsed 00:00

#Calling R from Python
repl_python ()

from pprint import pprint

pprint (r.familyid)

# {'Anguillidae': {'data': family_c phylum_c superfamily ... order superfamily_c download_date
# 0 鳗鲡科 脊索动物门 True ... Anguilliformes True 2020-08-03
#
# [1 rows x 14 columns],
# 'meta': {'code': 200,
# 'count': 1,
# 'limit': 20,
# 'message': 'success',
# 'page': 1}}
exit
```

2.4 search_taxon_id 查询中国生物物种名录的种 id

```
#Search taxon IDs via familyID
query1 = familyid$Anguillidae$data$record_id

taxonid1 <- search_taxon_id (query = query1, name = "familyID")

# Request returned successfully!!!
# Download date: 2020-08-04
# |=====| 100%, Elapsed 00:00
# Research type: familyID

str(taxonid1[["3851c5311bed46c19529cb155d37aa9b"]] [["meta"]])

# List of 5
# $ code : int 200
# $ limit : int 20
# $ count : int 5
```


丁刘勇, 李昊, 陶捐, 张金龙, 黄敏睿, 杨科, 王军, 丁城志, 何大明 (2021) 获取生物物种名录信息的 R 程序包 SP2000. 生物多样性, 29 (1): 118–122. <http://www.biodiversity-science.net/CN/10.17520/biods.2020235>

```
# $ page : int 1
# $ message: chr "success"

#Search taxon IDs via scientificName
queries <- c ("Anguilla marmorata", "Anguilla japonica", "Anguilla bicolor", "Anguilla nebulosa", "Anguilla luzonensis")

taxonid2 <- search_taxon_id (query = queries)

# Request returned successfully!!!
# Download date: 2020-08-04
# |=====| 100%, Elapsed 00:02
# Research type: scientificName
```

2.5 search_checklist 查询中国生物物种名录信息

```
#Download detailed lists via species or infraspecies ID
query2 <- taxonid1[["3851c5311bed46c19529cb155d37aa9b"]] [["data"]][["namecode"]]

x1 <- search_checklist (query = query2)

# Request returned successfully!!!
# Download date: 2020-08-04
# |=====| 100%, Elapsed 00:01
# Records - Found: 5

str(x1[["Anguilla bicolor"]])

# List of 2
# $ meta:List of 2
# ..$ code : int 200
# ..$ message: chr "success"
# $ data:List of 12
# ..$ searchCodeStatus: chr "accepted name"
# ..$ Synonyms : 'data.frame': 1 obs. of 2 variables:
# ...$ synonym: chr "Anguilla foochowensis"
# ...$ refs :List of 1
# ...$ : list()
# ..$ namecode : chr "1bcb107bcbf74c6eb81554e398beb840"
# ..$ scientificName : chr "Anguilla bicolor"
# ..$ author : chr "McClelland, 1844"
# ..$ Refs : chr [1:2] "McClelland J. (1844) Apodal fishes of Bengal. Calcutta Journal of Natural History, 5 (18):151-226." ""
# ..$ taxonTree : tibble [1 × 8] (S3: tbl_df/tbl/data.frame)
# ...$ phylum : chr "Chordata"
# ...$ genus : chr "Anguilla"
# ...$ species : chr "bicolor"
# ...$ infraspecies: chr ""
# ...$ family : chr "Anguillidae"
# ...$ kingdom : chr "Animalia"
# ...$ class : chr "Actinopterygii"
# ...$ order : chr "Anguilliformes"
# ..$ chineseName : chr "双色鳗鲡"
# ..$ searchCode : chr "1bcb107bcbf74c6eb81554e398beb840"
# ..$ CommonNames : chr "福州鳗"
# ..$ SpecialistInfo : 'data.frame': 3 obs. of 4 variables:
# ...$ E-Mail : chr [1:3] "zhangcg@ioz.ac.cn" "zoskt@gate.sinica.edu.tw" ""
# ...$ Address : chr [1:3] "1 Beichen West Road, Chaoyang District, Beijing 100101, P.R.China(北京市朝阳区北辰西路 1 号院 5 号 中国科学院动物研究所)" "()" "No.999, Huchenghuan Rd , Nanhui New City, Shanghai, P.R. China(上海市浦东新区沪城环路 999 号)"
# ...$ name : chr [1:3] "Zhang Chunguang(张春光)" "Shao, Kwang-Tsao(邵广昭)" "Wu Hanlin(伍汉霖)"
```

丁刘勇, 李昊, 陶捐, 张金龙, 黄敏睿, 杨科, 王军, 丁城志, 何大明 (2021) 获取生物物种名录信息的 R 程序包 SP2000. 生物多样性, 29 (1): 118–122. <http://www.biodiversity-science.net/CN/10.17520/biods.2020235>

```
# ...$ Institution: chr [1:3] "Institute of Zoology, Chinese Academy of Sciences(中国科学院动物研究所)" "(中央研究院生物多样性研究中心)" "College of Life Science & Technology, Shanghai Ocean University(上海海洋大学生命科学与技术学院)"
# ..$ download_date : Date[1:1], format: "2020-08-03"
```

2.6 get_col_global 查询全球生物物种名录信息

```
#Get Catalogue of Life Global checklist via species name and id
x2 <- get_col_global (query = queries, option = "name")

# Download date: 2020-08-04
# Research type: name
# |=====| 100%, Elapsed 00:03

str(x2[["Anguilla bicolor"]][["meta"]])

# List of 8
# $ id : chr ""
# $ name : chr "Anguilla bicolor"
# $ total_number_of_results : int 3
# $ number_of_results_returned: int 3
# $ start : int 0
# $ error_mexage : NULL
# $ version : chr "1.9 rev 2126ab0"
# $ rank : chr ""
```

2.7 get_redlist_china 查询中国物种红色名录信息

```
#Query Redlist of Chinese Biodiversity
get_redlist_china (query = "Anguilla", option = "Scientific Names")

## Download date: 2020-08-04

## # A tibble: 4 x 11
## `Chinese Family... Family` `Chinese Names` ScientificNames Status
## <chr> <chr> <chr> <chr> <chr>
## 1 鳗鲡科 Angui... 日本鳗鲡 Anguilla japon... EN
## 2 鳗鲡科 Angui... 花鳗鲡 Anguilla marmo... EN
## 3 鳗鲡科 Angui... 双色鳗鲡 Anguilla bicol... NT
## 4 鳗鲡科 Angui... 云纹鳗鲡 Anguilla nebul... NT
## # ... with 6 more variables: `Assessment Criteria` <chr>, Endemic <chr>,
## # Taxon <chr>, `Chinese Taxon` <chr>, Group <chr>, `Chinese Group` <chr>
```

2.8 find_synonyms 查询全球生物物种同物异名

```
#Find synonyms via species name from Catalogue of Life Global
find_synonyms (queries)

# Download date: 2020-08-04
# |=====| 100%, Elapsed 00:03

## Find 8 results of synonyms for Anguilla marmorata are as follows:
## Find 6 results of synonyms for Anguilla japonica are as follows:
## Find 23 results of synonyms for Anguilla bicolor are as follows:
## Find 4 results of synonyms for Anguilla nebulosa are as follows:
## Find 1 results of synonyms for Anguilla luzonensis are as follows:
```

丁刘勇, 李昊, 陶捐, 张金龙, 黄敏睿, 杨科, 王军, 丁城志, 何大明 (2021) 获取生物物种名录信息的 R 程序包 SP2000. 生物多样性, 29 (1): 118–122. <http://www.biodiversity-science.net/CN/10.17520/biods.2020235>

```
## $`Anguilla marmorata`
## [1] "Anguilla fidiensis" "Anguilla hildebrandti" "Anguilla johannae"
## [4] "Anguilla labrosa" "Anguilla marmorata" "Anguilla mauritiana"
## [7] "Muraena manillensis" "Muraena mossambica"
##
## $`Anguilla japonica`
## [1] "Anguilla angustidens" "Anguilla breviceps" "Anguilla manabei"
## [4] "Anguilla nigricans" "Anguilla remifera" "Muraena pekinensis"
##
## $`Anguilla bicolor`
## [1] "Anguilla amblyodon" "Anguilla australis"
## [3] "Anguilla bicolor bicolor" "Anguilla bicolor pacifica"
## [5] "Anguilla bicolor" "Anguilla bicolor bicolor"
## [7] "Anguilla bleekeri" "Anguilla cantori"
## [9] "Anguilla dussumieri" "Anguilla malabarica"
## [11] "Anguilla marmorata" "Anguilla mauritiana"
## [13] "Anguilla moa" "Anguilla mowat"
## [15] "Anguilla pacifica" "Anguilla sidat"
## [17] "Anguilla spengeli" "Anguilla virescens"
## [19] "Muraena halmaherensis" "Muraena macrocephala"
## [21] "Muraena moa" "Muraena mossambica"
## [23] "Muraena virescens"
##
## $`Anguilla nebulosa`
## [1] "Anguilla bengalensis" "Anguilla elphinstonei"
## [3] "Anguilla nebulosa nebulosa" "Muraena maculata"
##
## $`Anguilla luzonensis`
## [1] "Anguilla huangi"
```

2.9 get_col_taiwan 查询台湾生物物种名录信息

#Search Catalogue of Life Taiwan checklist

get_col_taiwan (query="Anguillidae", level="family")

Download date: 2020-08-04

A tibble: 5 x 23

name_code kingdom kingdom_c phylum phylum_c class class_c order order_c family

<chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr>

1 380710 Animal... 動物界 Chord... 脊索動物門... Acti... 條鰭魚綱... Angu... 鰻形目 Angui...

2 395489 Animal... 動物界 Chord... 脊索動物門... Acti... 條鰭魚綱... Angu... 鰻形目 Angui...

3 380711 Animal... 動物界 Chord... 脊索動物門... Acti... 條鰭魚綱... Angu... 鰻形目 Angui...

4 395491 Animal... 動物界 Chord... 脊索動物門... Acti... 條鰭魚綱... Angu... 鰻形目 Angui...

5 380712 Animal... 動物界 Chord... 脊索動物門... Acti... 條鰭魚綱... Angu... 鰻形目 Angui...

... with 13 more variables: family_c <chr>, genus <chr>, genus_c <chr>,

species <chr>, infraspecies_marker <chr>, infraspecies <chr>,

infraspecies2_marker <chr>, infraspecies2 <chr>, author <chr>,

author2 <chr>, common_name_c <chr>, endemic <chr>, dataprovider <chr>

2.10 测试环境

#Collect Information About the Current R Session

sessionInfo ()

R version 4.0.0 (2020-04-24)

Platform: x86_64-apple-darwin17.0 (64-bit)

Running under: macOS Catalina 10.15.5

#

丁刘勇, 李昊, 陶捐, 张金龙, 黄敏睿, 杨科, 王军, 丁城志, 何大明 (2021) 获取生物物种名录信息的 R 程序包 SP2000. 生物多样性, 29 (1): 118–122. <http://www.biodiversity-science.net/CN/10.17520/biods.2020235>

```
# Matrix products: default
# BLAS: /System/Library/Frameworks/Accelerate.framework/Versions/A/Frameworks/vecLib.framework/Versions/A/libBLAS.dylib
# LAPACK: /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRlapack.dylib
#
# locale:
# [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
#
# attached base packages:
# [1] stats    graphics  grDevices  utils      datasets  methods   base
#
# other attached packages:
# [1] SP2000_0.1.0  reticulate_1.16
#
# loaded via a namespace (and not attached):
# [1] Rcpp_1.0.5    rstudioapi_0.11 xml2_1.3.2    magrittr_1.5  lattice_0.20-41
# [6] rlang_0.4.7   fansi_0.4.1    tools_4.0.0   parallel_4.0.0 DT_0.14
# [11] grid_4.0.0    data.table_1.12.8 png_0.1-7     utf8_1.1.4    cli_2.0.2
# [16] htmltools_0.5.0 ellipsis_0.3.1 assertthat_0.2.1 yaml_2.2.1    digest_0.6.25
# [21] tibble_3.0.3  lifecycle_0.2.0 crayon_1.3.4  Matrix_1.2-18 pbmcapply_1.5.0
# [26] purrr_0.3.4   vctrs_0.3.1    htmlwidgets_1.5.1 curl_4.3      rlist_0.4.6.1
# [31] glue_1.4.1    compiler_4.0.0 pillar_1.4.6  XML_3.99-0.5  jsonlite_1.7.0
# [36] pkgconfig_2.0.3
```

参考文献

Allaire J, Xie YH, McPherson J, Luraschi J, Ushey K, Atkins A, Wickham H, Cheng J, Chang W, Iannone R (2020) rmarkdown: Dynamic Documents for R. R package version 2.1. <https://rmarkdown.rstudio.com>. (accessed on 2020-06-08)

丁刘勇, 李昊, 陶捐, 张金龙, 黄敏睿, 杨科, 王军, 丁城志, 何大明 (2021) 获取生物物种名录信息的 R 程序包 SP2000. 生物多样性, 29 (1): 118–122. <http://www.biodiversity-science.net/CN/10.17520/biods.2020235>

附录2 Python程序包SP2000的简介、安装、配置和使用指南

Appendix 2 The introduction, Installation, configuration and usage for Python package SP2000

1. 简介	1
2. 安装 SP2000	2
3. 使用指南	3
3.1 加载 SP2000	3
3.2 set_search_key 设置 Species 2000 密钥	3
3.3 search_family_id 查询中国生物物种名录的科 id	4
3.4 search_taxon_id 查询中国生物物种名录的种 id	4
3.5 search_checklist 查询中国生物物种名录信息	5
3.6 get_col_global 查询全球生物物种名录信息	6
3.7 get_redlist_china 查询中国物种红色名录信息	7
3.8 find_synonyms 查询全球生物物种同物异名	7
3.9 get_col_taiwan 查询台湾生物物种名录信息	8
3.10 测试环境	10
参考文献	10

1. 简介

Python 程序包 SP2000 基于 Python3 编写的, 可以在 Windows、MacOS 以及 Linux 等多种操作系统上运行, 已上传到 Python 官网(<https://pypi.org/project/SP2000>, version 0.6)。其程序包含 8 个函数: set_search_key、search_family_id、search_taxon_id、search_checklist、get_redlist_china、get_col_global、find_synonyms 和 get_col_taiwan。

附录 2 文档由 R 程序包 *rmarkdown* (Allaire et al, 2020) 自动生成。Python 程序包 SP2000 测试基于 RStudio 软件, 用户首先需要先安装 R 软件, 再安装 RStudio 软件, R 软件下载的地址为: <http://cran.r-project.org/bin/windows/base/>, 用户根据自己的计算机操作系统, 选取相应 R 软件, 建议用户下载最新版的 R, 因为涉及到字符编码转换, R 版本不低于 3.0.0; RStudio 软件下载的地址为: <https://rstudio.com/products/rstudio/download/>, 用户根据安装的 R 软件选取相应的 RStudio 软件, 因为旧版本不能运行 Python 代码, 建议用户下载最新版的 RStudio。下面以“鳗鲡科(Anguillidae)”为例, 详细介绍 Python 程序包 SP2000 函数的使用, 并且通过 R 包 *reticulate* 包完成 R 调用 Python 结果, 调用方式 `py$*`。下文重点介绍 Python 程序包 SP2000 的安装、配置和使用指南。

丁刘勇, 李昊, 陶捐, 张金龙, 黄敏睿, 杨科, 王军, 丁城志, 何大明 (2021) 获取生物物种名录信息的 R 程序包 SP2000. 生物多样性, 29 (1): 118–122. <http://www.biodiversity-science.net/CN/10.17520/biods.2020235>

2. 安装 SP2000

```
#Check if reticulate is installed
if (!requireNamespace ("reticulate", quietly = TRUE))
  install.packages ("reticulate")

## trying URL 'https://mirrors.tongji.edu.cn/CRAN/bin/macosx/contrib/4.0/reticulate_1.16.tgz'
## Content type 'application/octet-stream' length 1864511 bytes (1.8 MB)
## =====
##
## downloaded 1.8 MB
## The downloaded binary packages are in
## /var/folders/1x/5mjvdx0m754gnfv6_dyxqm0000gn/T/Rtmp9PBvfl/downloaded_packages

# Load the reticulate package
library ("reticulate")

# install SP2000 package
reticulate::py_install ("SP2000", pip = TRUE)

# Collecting package metadata (current_repodata.json): ...working... done
# Solving environment: ...working... done
#
# # All requested packages already installed.
#
# Collecting SP2000
# Using cached SP2000-0.6-py3-none-any.whl (12 kB)
# Requirement already satisfied, skipping upgrade: requests in /Users/yong/Library/r-miniconda/envs/r-reticulate/lib/python3.6/site-packages (from SP2000) (2.23.0)
# Requirement already satisfied, skipping upgrade: pandas in /Users/yong/Library/r-miniconda/envs/r-reticulate/lib/python3.6/site-packages (from SP2000) (1.0.4)
# Requirement already satisfied, skipping upgrade: certifi>=2017.4.17 in /Users/yong/Library/r-miniconda/envs/r-reticulate/lib/python3.6/site-packages (from requests->SP2000) (2020.6.20)
# Requirement already satisfied, skipping upgrade: urllib3!=1.25.0,!<1.25.1,<1.26,>=1.21.1 in /Users/yong/Library/r-miniconda/envs/r-reticulate/lib/python3.6/site-packages (from requests->SP2000) (1.25.9)
# Requirement already satisfied, skipping upgrade: idna<3,>=2.5 in /Users/yong/Library/r-miniconda/envs/r-reticulate/lib/python3.6/site-packages (from requests->SP2000) (2.9)
# Requirement already satisfied, skipping upgrade: chardet<4,>=3.0.2 in /Users/yong/Library/r-miniconda/envs/r-reticulate/lib/python3.6/site-packages (from requests->SP2000) (3.0.4)
# Requirement already satisfied, skipping upgrade: pytz>=2017.2 in /Users/yong/Library/r-miniconda/envs/r-reticulate/lib/python3.6/site-packages (from pandas->SP2000) (2020.1)
# Requirement already satisfied, skipping upgrade: numpy>=1.13.3 in /Users/yong/Library/r-miniconda/envs/r-reticulate/lib/python3.6/site-packages (from pandas->SP2000) (1.18.5)
# Requirement already satisfied, skipping upgrade: python-dateutil>=2.6.1 in /Users/yong/Library/r-miniconda/envs/r-reticulate/lib/python3.6/site-packages (from pandas->SP2000) (2.8.1)
# Requirement already satisfied, skipping upgrade: six>=1.5 in /Users/yong/Library/r-miniconda/envs/r-reticulate/lib/python3.6/site-packages (from python-dateutil>=2.6.1->pandas->SP2000) (1.15.0)
# Installing collected packages: SP2000
# Successfully installed SP2000-0.6

# install IPython package
reticulate::py_install (packages = "IPython", pip = TRUE)

# Collecting package metadata (current_repodata.json): ...working... done
# Solving environment: ...working... done
#
# # All requested packages already installed.
#
# Collecting IPython
# Using cached ipython-7.16.1-py3-none-any.whl (785 kB)
```

丁刘勇, 李昊, 陶捐, 张金龙, 黄敏睿, 杨科, 王军, 丁城志, 何大明 (2021) 获取生物物种名录信息的 R 程序包 SP2000. 生物多样性, 29 (1): 118–122. <http://www.biodiversity-science.net/CN/10.17520/biods.2020235>

```
# Requirement already satisfied, skipping upgrade: backcall in /Users/yong/Library/r-miniconda/envs/r-reticulate/lib/python3.6/site-packages (from IPython) (0.1.0)
# Requirement already satisfied, skipping upgrade: setuptools>=18.5 in /Users/yong/Library/r-miniconda/envs/r-reticulate/lib/python3.6/site-packages (from IPython) (47.1.1.post20200529)
# Requirement already satisfied, skipping upgrade: pexpect; sys_platform != "win32" in /Users/yong/Library/r-miniconda/envs/r-reticulate/lib/python3.6/site-packages (from IPython) (4.8.0)
# Requirement already satisfied, skipping upgrade: prompt-toolkit!=3.0.0,!<3.0.1,<3.1.0,>=2.0.0 in /Users/yong/Library/r-miniconda/envs/r-reticulate/lib/python3.6/site-packages (from IPython) (3.0.5)
# Requirement already satisfied, skipping upgrade: appnope; sys_platform == "darwin" in /Users/yong/Library/r-miniconda/envs/r-reticulate/lib/python3.6/site-packages (from IPython) (0.1.0)
# Requirement already satisfied, skipping upgrade: pickleshare in /Users/yong/Library/r-miniconda/envs/r-reticulate/lib/python3.6/site-packages (from IPython) (0.7.5)
# Requirement already satisfied, skipping upgrade: traitlets>=4.2 in /Users/yong/Library/r-miniconda/envs/r-reticulate/lib/python3.6/site-packages (from IPython) (4.3.3)
# Requirement already satisfied, skipping upgrade: pygments in /Users/yong/Library/r-miniconda/envs/r-reticulate/lib/python3.6/site-packages (from IPython) (2.6.1)
# Requirement already satisfied, skipping upgrade: jedi>=0.10 in /Users/yong/Library/r-miniconda/envs/r-reticulate/lib/python3.6/site-packages (from IPython) (0.17.0)
# Requirement already satisfied, skipping upgrade: decorator in /Users/yong/Library/r-miniconda/envs/r-reticulate/lib/python3.6/site-packages (from IPython) (4.4.2)
# Requirement already satisfied, skipping upgrade: ptyprocess>=0.5 in /Users/yong/Library/r-miniconda/envs/r-reticulate/lib/python3.6/site-packages (from pexpect; sys_platform != "win32"->IPython) (0.6.0)
# Requirement already satisfied, skipping upgrade: wcwidth in /Users/yong/Library/r-miniconda/envs/r-reticulate/lib/python3.6/site-packages (from prompt-toolkit!=3.0.0,!<3.0.1,<3.1.0,>=2.0.0->IPython) (0.2.3)
# Requirement already satisfied, skipping upgrade: ipython-genutils in /Users/yong/Library/r-miniconda/envs/r-reticulate/lib/python3.6/site-packages (from traitlets>=4.2->IPython) (0.2.0)
# Requirement already satisfied, skipping upgrade: six in /Users/yong/Library/r-miniconda/envs/r-reticulate/lib/python3.6/site-packages (from traitlets>=4.2->IPython) (1.15.0)
# Requirement already satisfied, skipping upgrade: parso>=0.7.0 in /Users/yong/Library/r-miniconda/envs/r-reticulate/lib/python3.6/site-packages (from jedi>=0.10->IPython) (0.7.0)
# Installing collected packages: IPython
# Successfully installed IPython-7.16.1
```

3. 使用指南

3.1 加载 SP2000

```
#Import the SP2000 package
from SP2000.sp2000 import *

#Import the pprint package
from pprint import pprint
```

3.2 set_search_key 设置 Species 2000 密钥

注意: 在运行函数 `search_family_id`、`search_taxon_id`、`search_checklist` 之前需先申请“Species 2000”密钥, 注册物种 2000 中国节点 <http://www.sp2000.org.cn> 网站, 然后点击“用户信息”便可获取一个自己独享的 API 密钥服务, 最后调用一次函数 `set_search_key`(“your key”), 便可运行以上 `search_*` 函数。

```
#Note: You need to apply for the apiKey <http://www.sp2000.org.cn>

#to run search_* functions of this package.

set_search_key("06319834*****")
```

丁刘勇, 李昊, 陶捐, 张金龙, 黄敏睿, 杨科, 王军, 丁城志, 何大明 (2021) 获取生物物种名录信息的 R 程序包 SP2000. 生物多样性, 29 (1): 118–122. <http://www.biodiversity-science.net/CN/10.17520/biods.2020235>

3.3 search_family_id 查询中国生物物种名录的科 id

```
#Search family IDs via family name
#Family name, or part of family name, supports Latin and Chinese names.
search_family_id ("鳗鲡科")

## {'鳗鲡科': ['3851c5311bed46c19529cb155d37aa9b']}

familyid = search_family_id ("Anguillidae")

pprint (familyid)

## {'Anguillidae': ['3851c5311bed46c19529cb155d37aa9b']}

#Calling Python from R
py$familyid

## $Anguillidae
## [1] "3851c5311bed46c19529cb155d37aa9b"
```

3.4 search_taxon_id 查询中国生物物种名录的种 id

```
#Search taxon IDs via familyID
taxonid1 = search_taxon_id (" ".join (familyid["Anguillidae"]),
                               name = "family_id")

pprint (taxonid1)

## {'3851c5311bed46c19529cb155d37aa9b': ['1bcb107bcbf74c6eb81554e398beb840',
##
##                               '9b9b328f6fa045089021ba38f912a0e8',
##                               'cbf03e5022f94c3daad91843b9f0b1e7',
##                               'e192fbc15df24049bcd0fd01d307affa',
##                               'f542929f776246efa44e559c389139d8']}

#Calling Python from R
py$taxonid1

## $`3851c5311bed46c19529cb155d37aa9b`
## [1] "1bcb107bcbf74c6eb81554e398beb840" "9b9b328f6fa045089021ba38f912a0e8"
## [3] "cbf03e5022f94c3daad91843b9f0b1e7" "e192fbc15df24049bcd0fd01d307affa"
## [5] "f542929f776246efa44e559c389139d8"

#Search taxon IDs via scientificName
taxonid2 = search_taxon_id ("Anguilla marmorata",
                              "Anguilla japonica",
                              "Anguilla bicolor",
                              "Anguilla nebulosa",
                              "Anguilla luzonensis",
                              name = "scientific_name")

pprint (taxonid2)

## {'Anguilla bicolor': ['1bcb107bcbf74c6eb81554e398beb840'],
## 'Anguilla japonica': ['f542929f776246efa44e559c389139d8'],
## 'Anguilla luzonensis': ['cbf03e5022f94c3daad91843b9f0b1e7'],
## 'Anguilla marmorata': ['e192fbc15df24049bcd0fd01d307affa'],
## 'Anguilla nebulosa': ['9b9b328f6fa045089021ba38f912a0e8']}
```

丁刘勇, 李昊, 陶捐, 张金龙, 黄敏睿, 杨科, 王军, 丁城志, 何大明 (2021) 获取生物物种名录信息的 R 程序包 SP2000. 生物多样性, 29 (1): 118–122. <http://www.biodiversity-science.net/CN/10.17520/biods.2020235>

```
#Calling Python from R
py$taxonid2

## $`Anguilla marmorata`
## [1] "e192fbc15df24049bcd0fd01d307affa"
##
## $`Anguilla japonica`
## [1] "f542929f776246efa44e559c389139d8"
##
## $`Anguilla bicolor`
## [1] "1bcb107bcbf74c6eb81554e398beb840"
##
## $`Anguilla nebulosa`
## [1] "9b9b328f6fa045089021ba38f912a0e8"
##
## $`Anguilla luzonensis`
## [1] "cbf03e5022f94c3daad91843b9f0b1e7"
```

3.5 search_checklist 查询中国生物物种名录信息

```
#Download detailed lists via species or infraspecies ID
x1 = search_checklist ('e192fbc15df24049bcd0fd01d307affa',
                      'f542929f776246efa44e559c389139d8',
                      '1bcb107bcbf74c6eb81554e398beb840',
                      '9b9b328f6fa045089021ba38f912a0e8',
                      'cbf03e5022f94c3daad91843b9f0b1e7')

pprint (x1['e192fbc15df24049bcd0fd01d307affa'])

# {'CommonNames': [],
#  'Distribution': 'Zhejiang(浙江)',
#  'Refs': [{'1': ''}, {'2': ''}],
#  'SpecialistInfo': [{'Address': '1 Beichen West Road, Chaoyang District, '
#                                'Beijing 100101, P.R.China(北京市朝阳区北辰西路 1 号院 5 号 '
#                                '中国科学院动物研究所)',
#  'E-Mail': 'zhangcg@ioz.ac.cn',
#  'Institution': 'Institute of Zoology, Chinese Academy of '
#  'Sciences(中国科学院动物研究所)',
#  'name': 'Zhang Chunguang(张春光)'},
#  {'Address': '()',
#  'E-Mail': 'zoskt@gate.sinica.edu.tw',
#  'Institution': '(中央研究院生物多样性研究中心)',
#  'name': 'Shao, Kwang-Tsao(邵广昭)'},
#  {'Address': 'No.999, Huchenghuan Rd , Nanhui New City, '
#  'Shanghai, P.R. China(上海市浦东新区沪城环路 999 号)',
#  'E-Mail': '',
#  'Institution': 'College of Life Science & Technology, '
#  'Shanghai Ocean '
#  'University(上海海洋大学生命科学与技术学院)',
#  'name': 'Wu Hanlin(伍汉霖)}],
#  'author': 'Quoy et Gaimard, 1824',
#  'chineseName': '花鳗鲡',
#  'namecode': 'e192fbc15df24049bcd0fd01d307affa',
#  'scientificName': 'Anguilla marmorata',
#  'searchCode': 'e192fbc15df24049bcd0fd01d307affa',
#  'searchCodeStatus': 'accepted name',
#  'taxonTree': {'class': 'Actinopterygii',
#  'family': 'Anguillidae',
#  'genus': 'Anguilla',
```

丁刘勇, 李昊, 陶捐, 张金龙, 黄敏睿, 杨科, 王军, 丁城志, 何大明 (2021) 获取生物物种名录信息的 R 程序包 SP2000. 生物多样性, 29 (1): 118–122. <http://www.biodiversity-science.net/CN/10.17520/biods.2020235>

```
# 'infraspecies': '',
# 'kingdom': 'Animalia',
# 'order': 'Anguilliformes',
# 'phylum': 'Chordata',
# 'species': 'marmorata'}}
```

3.6 get_col_global 查询全球生物物种名录信息

```
#Get Catalogue of Life Global checklist via species name and id
x2 = get_col_global ("Anguilla marmorata",
                    "Anguilla japonica",
                    "Anguilla bicolor",
                    "Anguilla nebulosa",
                    "Anguilla luzonensis",
                    option="name")

#
pprint (x2['Anguilla marmorata'])

# [{'bibliographic_citation': 'Froese R. & Pauly D. (eds). (2020). FishBase '
#                               '(version Feb 2018). In: Species 2000 & ITIS '
#                               'Catalogue of Life, 2020-06-04 Beta (Roskov Y.; '
#                               'Ower G.; Orrell T.; Nicolson D.; Bailly N.; Kirk '
#                               'P.M.; Bourgoïn T.; DeWalt R.E.; Decock W.; '
#                               'Nieukerken E. van; Penev L.; eds.). Digital '
#                               'resource at www.catalogueoflife.org/col. Species '
#                               '2000: Naturalis, Leiden, the Netherlands. ISSN '
#                               '2405-8858.',
# 'id': '433e0a4fe332e565c1679fa149543d83',
# 'is_extinct': 'false',
# 'name': 'Anguilla marmorata',
# 'name_html': '<i>Anguilla marmorata</i> Quoy & Gaimard, 1824',
# 'name_status': 'accepted name',
# 'online_resource': 'http://www.fishbase.org/Summary/SpeciesSummary.php?ID=1275',
# 'rank': 'Species',
# 'record_scrutiny_date': [],
# 'source_database': 'FishBase',
# 'source_database_url': 'http://www.fishbase.org',
# 'url': 'http://www.catalogueoflife.org/col/details/species/id/433e0a4fe332e565c1679fa149543d83'},
# {'accepted_name': {'bibliographic_citation': 'Froese R. & Pauly D. (eds). '
#                                               '(2020). FishBase (version Feb '
#                                               '2018). In: Species 2000 & ITIS '
#                                               'Catalogue of Life, 2020-06-04 '
#                                               'Beta (Roskov Y.; Ower G.; '
#                                               'Orrell T.; Nicolson D.; Bailly '
#                                               'N.; Kirk P.M.; Bourgoïn T.; '
#                                               'DeWalt R.E.; Decock W.; '
#                                               'Nieukerken E. van; Penev L.; '
#                                               'eds.). Digital resource at '
#                                               'www.catalogueoflife.org/col. '
#                                               'Species 2000: Naturalis, '
#                                               'Leiden, the Netherlands. ISSN '
#                                               '2405-8858.',
# 'id': '94f902df44fd76bed84cdea361b24fd6',
# 'is_extinct': 'false',
# 'name': 'Anguilla bengalensis',
# 'name_html': '<i>Anguilla bengalensis</i> (Gray, 1831)',
# 'name_status': 'accepted name',
# 'online_resource': 'http://www.fishbase.org/Summary/SpeciesSummary.php?ID=1272',
# 'rank': 'Species',
# 'record_scrutiny_date': [],
# 'source_database': 'FishBase',
```


丁刘勇, 李昊, 陶捐, 张金龙, 黄敏睿, 杨科, 王军, 丁城志, 何大明 (2021) 获取生物物种名录信息的 R 程序包 SP2000. 生物多样性, 29 (1): 118–122. <http://www.biodiversity-science.net/CN/10.17520/biods.2020235>

```
# 'source_database_url': 'http://www.fishbase.org',
# 'url': 'http://www.catalogueoflife.org/col/details/species/id/94f902df44fd76bed84cdea361b24fd6'},
# 'bibliographic_citation': 'Froese R. & Pauly D. (eds). (2020). FishBase '
# '(version Feb 2018). In: Species 2000 & ITIS '
# 'Catalogue of Life, 2020-06-04 Beta (Roskov Y.; '
# 'Ower G.; Orrell T.; Nicolson D.; Bailly N.; Kirk '
# 'P.M.; Bourgoin T.; DeWalt R.E.; Decock W.; '
# 'Nieukerken E. van; Penev L.; eds.). Digital '
# 'resource at www.catalogueoflife.org/col. Species '
# '2000: Naturalis, Leiden, the Netherlands. ISSN '
# '2405-8858.',
# 'id': '1a44a32b532e87f048b6bfa8fed07098',
# 'name': 'Anguilla marmorata',
# 'name_html': '<i>Anguilla marmorata</i> (non Quoy & Gaimard, 1824)',
# 'name_status': 'misapplied name',
# 'online_resource': '',
# 'rank': 'Species',
# 'source_database': 'FishBase',
# 'source_database_url': 'http://www.fishbase.org',
# 'url': 'http://www.catalogueoflife.org/col/details/species/id/94f902df44fd76bed84cdea361b24fd6/synonym/1a44a32b532e87f048b6bfa8fed07098']}]
```

3.7 get_redlist_china 查询中国物种红色名录信息

#Query Redlist of Chinese Biodiversity

```
x3 = get_redlist_china (group = 'Inland Fishes')
```

```
pprint (x3) # pandas DataFrame
```

#	family_c	family	species_c ...	group_c	kingdom	kingdom_c
# 10436	七鳃鳗科	Petromyzontidae	东北七鳃鳗 ...	内陆鱼类	Vertebrate	脊椎动物卷
# 10437	七鳃鳗科	Petromyzontidae	雷氏七鳃鳗 ...	内陆鱼类	Vertebrate	脊椎动物卷
# 10438	七鳃鳗科	Petromyzontidae	北极七鳃鳗 ...	内陆鱼类	Vertebrate	脊椎动物卷
# 10439	丝足鲈科	Osphronemidae	线足鲈 ...	内陆鱼类	Vertebrate	脊椎动物卷
# 10440	亚口鱼科	Catostomidae	胭脂鱼 ...	内陆鱼类	Vertebrate	脊椎动物卷
#
# 11874	鳢科	Channidae	沃氏鳢 ...	内陆鱼类	Vertebrate	脊椎动物卷
# 11875	鳢科	Channidae	黑月鳢 ...	内陆鱼类	Vertebrate	脊椎动物卷
# 11876	鲢科	Pangasiidae	长丝鲢 ...	内陆鱼类	Vertebrate	脊椎动物卷
# 11877	鲢科	Pangasiidae	贾巴鲢 ...	内陆鱼类	Vertebrate	脊椎动物卷
# 11878	鲢科	Pangasiidae	巨无齿鲢 ...	内陆鱼类	Vertebrate	脊椎动物卷

```
#
# [1443 rows x 11 columns]
```

3.8 find_synonyms 查询全球生物物种同物异名

#Find synonyms via species name from Catalogue of Life Global

```
x4 = find_synonyms ("Anguilla marmorata",
                    "Anguilla japonica")
```

```
pprint (x4)
```

```
# {'Anguilla japonica': {'Anguilla angustidens',
# 'Anguilla breviceps',
# 'Anguilla manabei',
# 'Anguilla nigricans',
# 'Anguilla remifera',
# 'Muraena pkinensis'}},
```

丁刘勇, 李昊, 陶捐, 张金龙, 黄敏睿, 杨科, 王军, 丁城志, 何大明 (2021) 获取生物物种名录信息的 R 程序包 SP2000. 生物多样性, 29 (1): 118–122. <http://www.biodiversity-science.net/CN/10.17520/biods.2020235>

```
# 'Anguilla marmorata': {'Anguilla fidjiensis',  
#                        'Anguilla hildebrandti',  
#                        'Anguilla johannae',  
#                        'Anguilla labrosa',  
#                        'Anguilla marmolata',  
#                        'Anguilla mauritiana',  
#                        'Muraena manillensis',  
#                        'Muraena mossambica'}}}
```

3.9 get_col_taiwan 查询台湾生物物种名录信息

```
#Search Catalogue of Life Taiwan checklist
```

```
x5 = get_col_taiwan ("Anguillidae", level = "family")
```

```
pprint (x5)
```

```
# {'Anguillidae': [{ 'author': 'Schmidt, 1928',  
#                   'author2': None,  
#                   'class': 'Actinopterygii',  
#                   'class_c': '條鰭魚綱',  
#                   'common_name_c': '太平洋雙色鰻鱺;短鰭鰻;二色鰻',  
#                   'dataprovder': None,  
#                   'endemic': None,  
#                   'family': 'Anguillidae',  
#                   'family_c': '鰻鱺科',  
#                   'genus': 'Anguilla',  
#                   'genus_c': '鰻鱺屬',  
#                   'infraspecies': 'pacifica',  
#                   'infraspecies2': None,  
#                   'infraspecies2_marker': None,  
#                   'infraspecies_marker': None,  
#                   'kingdom': 'Animalia',  
#                   'kingdom_c': '動物界',  
#                   'name_code': '380710',  
#                   'order': 'Anguilliformes',  
#                   'order_c': '鰻形目',  
#                   'phylum': 'Chordata',  
#                   'phylum_c': '脊索動物門',  
#                   'species': 'bicolor'},  
#                   { 'author': 'Kaup, 1856',  
#                     'author2': None,  
#                     'class': 'Actinopterygii',  
#                     'class_c': '條鰭魚綱',  
#                     'common_name_c': '西里伯斯鰻鱺;西里伯斯鰻;鰻;黑鰻',  
#                     'dataprovder': None,  
#                     'endemic': None,  
#                     'family': 'Anguillidae',  
#                     'family_c': '鰻鱺科',  
#                     'genus': 'Anguilla',  
#                     'genus_c': '鰻鱺屬',  
#                     'infraspecies': None,  
#                     'infraspecies2': None,  
#                     'infraspecies2_marker': None,  
#                     'infraspecies_marker': None,  
#                     'kingdom': 'Animalia',  
#                     'kingdom_c': '動物界',  
#                     'name_code': '395489',  
#                     'order': 'Anguilliformes',  
#                     'order_c': '鰻形目',
```

丁刘勇, 李昊, 陶捐, 张金龙, 黄敏睿, 杨科, 王军, 丁城志, 何大明 (2021) 获取生物物种名录信息的 R 程序包 SP2000. 生物多样性, 29 (1): 118–122. <http://www.biodiversity-science.net/CN/10.17520/biods.2020235>

```
#      'phylum': 'Chordata',
#      'phylum_c': '脊索動物門',
#      'species': 'celebesensis'},
#      {'author': 'Temminck & Schlegel, 1846',
#      'author2': None,
#      'class': 'Actinopterygii',
#      'class_c': '條鰭魚綱',
#      'common_name_c': '日本鰻鱺;白鰻;日本鰻;正鰻;白鰻;鰻鱺;土鰻;淡水鰻',
#      'dataprovider': None,
#      'endemic': None,
#      'family': 'Anguillidae',
#      'family_c': '鰻鱺科',
#      'genus': 'Anguilla',
#      'genus_c': '鰻鱺屬',
#      'infraspecies': None,
#      'infraspecies2': None,
#      'infraspecies2_marker': None,
#      'infraspecies_marker': None,
#      'kingdom': 'Animalia',
#      'kingdom_c': '動物界',
#      'name_code': '380711',
#      'order': 'Anguilliformes',
#      'order_c': '鰻形目',
#      'phylum': 'Chordata',
#      'phylum_c': '脊索動物門',
#      'species': 'japonica'},
#      {'author': 'Watanabe, Aoyama & Tsukamoto, 2009',
#      'author2': None,
#      'class': 'Actinopterygii',
#      'class_c': '條鰭魚綱',
#      'common_name_c': '呂宋鰻鱺;呂宋鰻;黃氏鱸鰻',
#      'dataprovider': None,
#      'endemic': None,
#      'family': 'Anguillidae',
#      'family_c': '鰻鱺科',
#      'genus': 'Anguilla',
#      'genus_c': '鰻鱺屬',
#      'infraspecies': None,
#      'infraspecies2': None,
#      'infraspecies2_marker': None,
#      'infraspecies_marker': None,
#      'kingdom': 'Animalia',
#      'kingdom_c': '動物界',
#      'name_code': '395491',
#      'order': 'Anguilliformes',
#      'order_c': '鰻形目',
#      'phylum': 'Chordata',
#      'phylum_c': '脊索動物門',
#      'species': 'luzonensis'},
#      {'author': 'Quoy & Gaimard, 1824',
#      'author2': None,
#      'class': 'Actinopterygii',
#      'class_c': '條鰭魚綱',
#      'common_name_c': '花鰻鱺;鱸鰻;花鰻;烏耳鰻;土龍;黑鰻',
#      'dataprovider': None,
#      'endemic': None,
#      'family': 'Anguillidae',
#      'family_c': '鰻鱺科',
#      'genus': 'Anguilla',
#      'genus_c': '鰻鱺屬',
```

丁刘勇, 李昊, 陶捐, 张金龙, 黄敏睿, 杨科, 王军, 丁城志, 何大明 (2021) 获取生物物种名录信息的 R 程序包 SP2000. 生物多样性, 29 (1): 118–122. <http://www.biodiversity-science.net/CN/10.17520/biods.2020235>

```
#      'infraspecies': None,
#      'infraspecies2': None,
#      'infraspecies2_marker': None,
#      'infraspecies_marker': None,
#      'kingdom': 'Animalia',
#      'kingdom_c': '動物界',
#      'name_code': '380712',
#      'order': 'Anguilliformes',
#      'order_c': '鰻形目',
#      'phylum': 'Chordata',
#      'phylum_c': '脊索動物門',
#      'species': 'marmorata'}}
```

3.10 测试环境

```
#Collect Information About the Current Python Session
import IPython
print (IPython.sys_info())

## {'commit_hash': '2486838d9',
##  'commit_source': 'installation',
##  'default_encoding': 'UTF-8',
##  'ipython_path': '/Users/yong/Library/r-miniconda/envs/r-reticulate/lib/python3.6/site-packages/IPython',
##  'ipython_version': '7.16.1',
##  'os_name': 'posix',
##  'platform': 'Darwin-19.5.0-x86_64-i386-64bit',
##  'sys_executable': '/Users/yong/Library/r-miniconda/envs/r-reticulate/bin/python',
##  'sys_platform': 'darwin',
##  'sys_version': '3.6.10 | packaged by conda-forge | (default, Apr 24 2020, '
##               '16:29:39) \n'
##               '[GCC Clang 9.0.1 ]'}
```

参考文献

Allaire J, Xie YH, McPherson J, Luraschi J, Ushey K, Atkins A, Wickham H, Cheng J, Chang W, Iannone R (2020) rmarkdown: Dynamic Documents for R. R package version 2.1. <https://rmarkdown.rstudio.com>. (accessed on 2020-06-08)

SP2000: An open-sourced R package for querying the Catalogue of Life

Liuyong Ding^{1,2}, Hao Li³, Juan Tao^{1,2}, Jinlong Zhang⁴, Minrui Huang^{1,2}, Ke Yang^{1,2,4}, Jun Wang^{1,2}, Chengzhi Ding^{1,2*}, Daming He^{1,2*}

1 *Institute of International Rivers and Eco-security, Yunnan University, Kunming 650504*

2 *Yunnan Key Laboratory of International Rivers and Transboundary Eco-security, Yunnan University, Kunming 650504*

3 *National Pilot School of Software, Yunnan University, Kunming 650504*

4 *Flora Conservation Department, Kadoorie Farm and Botanic Garden, Hong Kong 999077*

ABSTRACT

Aims: The Catalogue of Life provides the basis for understanding both regional and global biodiversity. With the invention and development of the internet, the up-to-date species checklists stored in the public databases has greatly promoted the development of taxonomy, conservation biology, and macroecology. Public species checklists play an indispensable role in biodiversity conservation and aid in the assessment of species' conservation status. The Species 2000 China Node (<http://www.sp2000.org.cn>) and the Catalogue of Life (<http://www.catalogueoflife.org>) are among the leading online databases in cataloguing biodiversity, contain 122,280 and 1,829,672 taxa respectively (including infraspecific taxa). Although searching the content of the websites may be relatively straightforward, downloading the data and transferring it into a statistical environment for further analysis can present challenges.

Method: To address this issue, we developed the package SP2000 using the R programming language.

Application: SP2000 is an open-source, cross-platform, and user-friendly package which aims to help users query and download the checklist of organisms (including animals, plants, fungi, and microbes) from within and outside China. Here we introduce and describe the usage of SP2000 including installation, and configuration of parameters.

Key words: species checklist; redlist; China's biodiversity; R package

1 Introduction

The Catalogue of Life provides the basis for understanding both regional and global biodiversity (Reichhardt, 1999; Banki et al, 2019; Ower & Roskov, 2019). With the invention and development of the internet, Annual or monthly editions of species checklists have been numerous stored in the public databases (e.g. the Species 2000 China Node and Catalogue of Life), which have greatly promoted the development of taxonomy, conservation biology and macroecology (Jiang et al, 2015). At present, these databases have been widely used in the assessment of species status, red list compilation and biodiversity conservation for governments or international organizations.

What are the similarities and differences between the Species 2000 China Node and Catalogue of Life? Their goals are to provide a validated checklist of the known species to all users in the world. As of June 4, 2020, the Species 2000 China Node (<http://www.sp2000.org.cn>, the Biodiversity Committee of Chinese Academy of Sciences, 2020) and the Catalogue of Life (<http://www.catalogueoflife.org>) record 122,280 and 1,829,672 taxa (including infraspecific taxa), respectively. The latter is a

collection of more than 130 global species databases, while the former is a subset of national studies. Therefore, the spatial scale of the both checklists is different and complementary. The Species 2000 China Node was established in 2006 by the Biodiversity Committee of the Chinese Academy of Sciences (BC-CAS), which is contributed by the Institute of Botany, Institute of Microbiology, Institute of Oceanology and Institute of Zoology, CAS. The first annual checklist of Catalogue of Life China was released in 2008 and has been updated annually, which is an important data source for the Catalogue of Life annual checklist (Jiang et al, 2015; Ma et al, 2018). Compared with global Catalogue of Life, Catalogue of Life China also provides the Chinese name (i.e., characters and pinyin) in addition to containing the scientific name of each species, synonyms, alias, references, classification system, distribution area and other information. Although all information in two websites will be available to all users in the world freely, downloading and getting required data into a statistical environment for further analysis are not straight-forward, which has become the main obstacle to restrict the widespread use of these checklists.

To address the above problems, we developed the R package SP2000 using R programming language (due to its features of open-sourced, cross-platformed,

丁刘勇, 李昊, 陶捐, 张金龙, 黄敏睿, 杨科, 王军, 丁城志, 何大明 (2021) 获取生物物种名录信息的 R 程序包 SP2000. 生物多样性, 29 (1): 118 - 122. <http://www.biodiversity-science.net/CN/10.17520/biods.2020235>

etc.; Tippmann, 2014; Zhang et al, 2016; Lai et al, 2019; R Core Team, 2020), which aims to help users accurately, quickly to query and download the required species checklists from the Species 2000 China node and Catalogue of Life website.

2 Methods

The R package *SP2000*, a programmatic interface to <http://sp2000.org.cn>, re-written based on an accompanying 'Species 2000' API (<http://sp2000.org.cn/api/document>, version 2), and access tables describing catalogue of the Chinese known species of animals, plants, fungi, micro-organisms, and more. This package also supports access to Catalogue of Life (<http://web-service.catalogueoflife.org/col/web-service>, version 1.9).

Compared to other tools for acquiring species checklists such as Catalogue of Life Search Plugin (<http://www.catalogueoflife.org/content/web-browser-page-plugin>), CD-ROM of Catalogue of Life China (<http://sp2000.org.cn/download>) and portal-components developed via JavaScript language (<https://github.com/CatalogueOfLife/portal-components>), the SP2000 is an open-sourced, cross-platformed, and user-friendly package which aims to help users to query and download the checklist of animals, plants, fungi and micro-organisms both in and outside China. The downloaded information goes directly into R statistical environment for further analysis, for example, mining more information of biodiversity through R package *spocc* (Chamberlain, 2020).

3 Usage

3.1 Version and installation

The package *SP2000* written in R language has been submitted to the CRAN (<https://cran.r-project.org/package=SP2000>, version 0.1.0), users can easily install R packages SP2000 using R commands `install.packages("SP2000", repos = "https://cran.r-project.org")`. It mainly consists of eight functions covering `set_search_key`, `search_family_id`, `search_taxon_id`, `search_checklist`, `get_redlist_China`, `get_col_global`, `find_synonyms` and `get_col_taiwan`, and the configuration of the parameters for these functions are as follows.

3.2 set_search_key

This function `set_search_key` allows users to set the key variable used all `search_*` functions (e.g. `search_family_id`, `search_taxon_id` and `search_checklist`). Users can obtain a key by registering

at <http://sp2000.org.cn/api/document>, clicking on the user information. It is worth noting that the upper limit of daily API visits for ordinary users is 2000, and users can apply for increasing the daily API request limit, through filling in the application form <http://col.especies.cn/doc/API.docx> and send an email to SP2000CN@ibcas.ac.cn entitled "Application for increasing API Request Times". Set the API key by running the R code:

```
set_search_key <- "your apikey"
```

3.3 search_family_id

The family is the most commonly used classification rank in biological classification, through which the taxonomic unit of species or subspecies can be more easily inquired. The Species 2000 China Node defines unique identity (id) for family and species (subspecies) to ensure the accuracy of data query. The function `search_family_id` provides the function querying the collection of family's ids. There are four arguments including `query`, `start`, `limit` and `mc.cores`. (1) The parameter '`query`' supports one or more queries for family name, or part of family name, (2) the parameter '`start`' sets the number of record to start at, the default value of 1, (3) the parameter '`limit`' sets the number of records to return, the default value is 20, and (4) the parameter '`mc.cores`' can set the number of cores to use, the default value is 2. Search family ids by running:

```
search_family_id(query = "Anguillidae")
```

3.4 search_taxon_id

The family ids can be used to directly obtain the list of ids for the species or subspecies, and then the details of the species list can be obtained using the function `search_checklist`. The `search_taxon_id` supports multiple types of queries for family's ids, scientific name and common name (including Chinese name). There are five arguments including `query`, `name`, `start`, `limit` and `mc.cores`. (1) The parameter `query` supports one or more queries, (2) the parameter `name` sets the query mode, in conjunction with the parameter `query`, parameters to be selected are "`familyID`", "`scientificName`" and "`commonName`", the default value is "`scientificName`", (3) the parameter `start` sets the number of record to start at, the default value of 1, (4) the parameter `limit` sets the number of records to return, the default value is 20, and (5) the parameter `mc.cores` setting is the same as 3.3. Take "`Anguillidae`" as an example, the R code is as follows:

```
## loading package
library("SP2000")
## Set your Species 2000 API key
set_search_key <- "your apikey"
## Search family ids via family name
```

丁刘勇, 李昊, 陶捐, 张金龙, 黄敏睿, 杨科, 王军, 丁城志, 何大明 (2021) 获取生物物种名录信息的 R 程序包 SP2000. 生物多样性, 29 (1): 118 - 122. <http://www.biodiversity-science.net/CN/10.17520/biods.2020235>

```
familyid <- search_family_id (query =  
"Anguillidae")  
## Search taxon ids via family's ids  
query <- familyid$Anguillidae$data$record_id  
taxonid <- search_taxon_id (query = query, name  
= "familyID")
```

3.5 search_checklist

The function `search_checklist` gets detailed information of species through species ids, including scientific name, synonym, alias, literature, classification system, distribution region and other data, as well as Chinese name and Chinese name pinyin and other contents. This function needs to be used in combination with the functions `search_family_id` and `search_taxon_id`. There are two arguments including `query` and `mc.cores`. (1) The parameter `query` supports one or more queries and (2) The parameters `mc.cores` is the same as 3.3.

Take the query result of 3.4 as an example:

```
query <- taxonid[["3851c5311bed46c19529cb1  
55d37aa9b"]][["data"]][["namecode"]]  
search_checklist (query = query)
```

3.6 get_redlist_china

The function `get_redlist_china` has four parameters: `query`, `option`, `group`, and `viewDT`. (1) The parameter 'query' supports one or more queries for scientific name or Chinese name, (2) the parameter 'option' sets the query mode, which is used in conjunction with the parameter `query`. The parameters `option` include "Chinese Names" and "Scientific Names", and the default value is "Scientific Names", (3) the optional parameters 'group' includes "Amphibians," "Birds", "Mammals", "Inland Fishes", "Reptiles," "Plants" and "Fungi", and (4) the parameter 'viewDT' is the logical value, which is used together with the parameter `group`. If `viewDT = TRUE`, the query result will display an interactive page. Taking Inland Fishes as an example, it is called `get_redlist_China (... , group = "Inland Fishes", viewDT = TRUE)`.

Take the *Anguilla* query as an example, the R code is as follows:

```
## Get Chinese Red List of the genus Anguilla  
get_redlist_china (query = "Anguilla", option =  
"Scientific Names")  
## Query "Inland Fish" China Red List  
information displaying searchable, downloadable and  
interactive page  
get_redlist_china (group = "Inland Fishes",  
viewDT = TRUE)
```

3.7 get_col_global

The function `get_col_global` is unrestricted by the Species 2000 key and can be used independently. It contains six parameters: `query`, `option`, `response`, `start`, `limit` and `mc.cores`. (1) The parameter 'query' inputs

one or more ids or the species name, (2) the parameter 'option' sets the query mode, which is used in conjunction with the parameter `query`. The optional parameters have "ID" and "name", and the default value is "name", (3) the parameter 'response' sets the query to return the result, which can be selected as "Full", one of "terse". "Full" returns the full query result, and "terse" returns the short query result. The default value is "terse"; (4) The parameter 'start' sets the first record returned by the query. The default value is 0, which is used in conjunction with the parameter 'response', (5) the parameter 'limit' sets the record returned by a single query, the default value is 500, the maximum number of results returned by a single short query is 500, and the maximum number of results returned by a single complete query is 50, and (6) the parameter 'mc.cores' setting is the same as 3.3.

Take the *Anguilla* query as an example, the R code is as follows:

```
x <- get_col_global (query = "Anguilla", response  
= "full")  
## The total query result is 208  
x[["Anguilla"]][["meta"]][["total_number_of_res  
ults"]] [1]
```

3.8 find_synonyms

The function `find_synonyms` has two arguments `query` and `mc.cores`. (1) The argument 'query' enters one or more species name, and (2) the argument 'mc.cores' is the same as 3.3. Take "*Anguilla Anguilla*" as an example and call it `find_synonyms ("Anguilla Anguilla")`.

3.9 get_col_taiwan

The function `get_col_taiwan` has four parameters: `query`, `level`, `option` and `include_synonyms`. (1) The parameter 'query' supports one or more queries, (2) the parameter 'level', which can be used in combination with the parameter `query` to select one of "kingdom", "phylum", "class", "order", "family", "genus" and "species", (3) the parameter 'option' includes "Contain", "Equal" and "beginning", the default is "equal", and (4) the parameter 'include_synonyms' is the logical value, and the query result contains synonym information, with the default value of TRUE. Take *Anguillidae* as an example, the call method is

```
get_col_taiwan (query = "Anguillidae", level =  
"family").
```

4 Conclusion

In this paper, we presented a new tool (R package *SP2000*) of re-written via Web API, which provides an interface for application program to download the data of the species checklists. Its detailed usage makes

丁刘勇, 李昊, 陶捐, 张金龙, 黄敏睿, 杨科, 王军, 丁城志, 何大明 (2021) 获取生物物种名录信息的 R 程序包 SP2000. 生物多样性, 29 (1): 118 - 122. <http://www.biodiversity-science.net/CN/10.17520/biods.2020235>

SP2000 a useful tool for biodiversity researchers and taxonomists.

In addition to R package *SP2000*, we also developed Python package *SP2000* (<https://pypi.org/project/SP2000>) through Python programming language (Perkel, 2015; Python Software Foundation, 2020) to better meet the needs of users in the era of big biodiversity data (Bisby, 2000). Users can easily install python packages SP2000 using commands “install pip3 install SP2000” or “python3 -m pip install SP2000”. The configuration of the parameters and query are basically the same as that of R package *SP2000*. Future work on *SP2000* includes enhancements such as search for more taxonomic information of insects and invertebrates distributed in China. We will also add China Animal Scientific Database (<http://zoology.especies.cn>) to the package *SP2000* via Web API (<http://zoology.especies.cn/database/api>).





Acknowledgements

The R package *SP2000* was made possible by leveraging integral R packages including *jsonlite* (Ooms, 2014), *tibble* (Müller & Wickham, 2020), *rlist* (Ren, 2016) and many others. We thank two anonymous reviewers whose helpful feedback helped improve the package and clarify this manuscript.

Availability

The R and Python package SP2000 are freely available under the permissive MIT license at <https://cran.r-project.org/package=SP2000> and <https://pypi.org/project/SP2000>, respectively.

ORCID

Liuyong Ding  <https://orcid.org/0000-0002-5490-182X>
Jinlong Zhang  <https://orcid.org/0000-0002-1161-5460>
Jun Wang  <https://orcid.org/0000-0003-2481-1409>
Chengzhi Ding  <https://orcid.org/0000-0001-5215-7374>

References

Banki O, Hobern D, Döring M, Remsen D (2019) Catalogue of Life Plus: A collaborative project to complete the checklist of the world's species. *Biodiversity Information Science and Standards*, 3, e37652.
Bisby FA (2000) The quiet revolution: Biodiversity informatics and the internet. *Science*, 289, 2309–2312.
Chamberlain S (2020) spocc: Interface to Species Occurrence

Data Sources. R package version 1.0.8. <https://CRAN.R-project.org/package=spocc/>. (accessed on 2020-06-01)
Jiang ZG, Qin HN, Liu YN, Ji LQ, Ma KP (2015) Protecting biodiversity and promoting sustainable development: In memory of the releasing of Catalogue of Life China 2015 and China Biodiversity Red List on the International Day for Biological Diversity 2015. *Biodiversity Science*, 23, 433–434. (in Chinese)
Lai JS, Lortie CJ, Muenchen RA, Yang J, Ma KP (2019) Evaluating the popularity of R in ecology. *Ecosphere*, 10, e02567.
Ma KP, Zhu M, Ji LQ, Ma JC, Guo QH, Ouyang ZY, Zhu L (2018) Establishing China Infrastructure for Big Biodiversity Data. *Bulletin of the Chinese Academy of Sciences*, 33(8), 80–87. (in Chinese with English abstract)
Müller K, Wickham H (2020). *tibble*: Simple Data Frames. R package version 3.0.3. <https://CRAN.R-project.org/package=tibble>. (accessed on 2020-08-01)
Ooms J (2014) The jsonlite Package: A Practical and Consistent Mapping Between JSON Data and R Objects. *arXiv*, 1403, 2805.
Ower G, Roskov Y (2019) The Catalogue of Life: Assembling data into a global taxonomic checklist. *Biodiversity Information Science and Standards*, 3, e37221.
Perkel JM (2015) Programming: Pick up Python. *Nature*, 518, 125–126.
Python Software Foundation (2020) Python Language Reference, version 3.7. <https://www.python.org/>. (accessed on 2020-06-01)
R Core Team (2020) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>. (accessed on 2020-06-01)
Reichhardt T (1999) Catalogue of life could become reality. *Nature*, 399, 519–519.
Ren K (2016) rlist: A Toolbox for Non-Tabular Data Manipulation. R package version 0.4.6.1. <https://CRAN.R-project.org/package=rlist>. (accessed on 2020-08-01)
The Biodiversity Committee of Chinese Academy of Sciences (2020) Catalogue of Life China: 2020 Annual Checklist, Beijing, China (in Chinese and in English). <http://www.sp2000.org.cn/CoLChina>. (accessed on 2020-05-22)
Tippmann S (2014) Programming tools: Adventures with R*Nature*, 517(7532), 109–110.
Zhang JL, Zhu HL, Liu JG, Fischer GA (2016) Principles behind designing herbarium specimen labels and the R package ‘herblabel’. *Biodiversity Science*, 24, 1345–1352. (in Chinese with English abstract)