

•数据论文•

植物标本标签设计的原则及R程序包herblabel

张金龙* 朱慧玲 刘金刚 Gunter A. Fischer

(嘉道理农场暨植物园植物保育部, 香港)

摘要: 植物标本是分类学、生态学和分子生物学最重要的凭证之一。标本的采集和鉴定信息需清晰、准确、美观地展示和保存于标本标签中, 不能有歧义以及拼写错误。在标签的制作过程中, 数据输入的方式要简单、直接, 标签文件生成过程中最好能自动分析错误, 且在打印之前要便于修改和调整。本文探讨了打印植物标本标签的若干原则以及注意事项, 并介绍了用R语言编写的herblabel程序包生成植物标本标签以及鉴定标签的过程。herblabel程序包基于Darwin Core和CVH5.0数据交换标准, 可快速批量生成几种样式的RTF标签, 且标签简洁、美观, 易于编辑。herblabel程序包具有检查地点完整程度, 学名拼写和接受状态, 科、属在APG等新系统下的对应关系等功能, 可有效减少数据录入过程中产生的错误。此外, 本程序包在打印标签时使用的是基于Darwin Core标准保存的标本数据库, 不仅方便统计和管理, 也可以直接用于全球生物多样性信息网络(GBIF)数据共享或者数字植物标本馆的建设。该程序包可显著提高植物标本馆标本制作、管理和信息录入的工作效率, 减轻工作人员的负担, 并在植物生物多样性编目中发挥重要作用。

关键词: 标本馆; 植物标本; 标签; R语言程序包; 拼写; APG III; Darwin Core

Principles behind designing herbarium specimen labels and the R package ‘herblabel’

Jinlong Zhang*, Huiling Zhu, Jingang Liu, Gunter A. Fischer

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

Abstract: Voucher specimens are essential for plant taxonomy, ecology, and plant molecular biology. Collection and identification information should be clearly, accurately and formally printed on herbarium labels, without confusion or spelling mistakes. During the preparation of the labels, data entry should be straight-forward and easy to manage. Ideally, the herbarium label printing software should conduct a thorough evaluation to avoid potential errors, and should issue warnings accordingly. By taking these principles into account, we developed the herblabel R package for generating herbarium labels based on Darwin Core and CVH5.0 data exchange standards. During the implementation of herblabel, the spelling and status of scientific names and the family-genera relationship are compared with the embedded databases to determine potential mistakes or incongruence against APG or other new classification systems. As the template is based on Darwin Core exchange standards, the database could not only be used to generate summaries and reports but also could be shared via GBIF or virtual herbaria. This package will increase efficiency and reduce workload for herbarium specimen preparation, management, and data entry, and should play an important role in cataloging plant diversity.

Key words: herbarium; specimens; label; R package; spelling; APG III; Darwin Core

植物标本是系统分类学、生态学、中药学、林学、园林学等很多领域的研究基础, 它包含了物种形态特征、地理分布、生境和物候期等大量信息。

植物标本除了用于分类学研究、编撰植物志和植物名录外, 其采集信息在生物多样性研究中也具有极其重要的价值。植物标本是植被调查的重要凭证,

收稿日期: 2016-08-23; 接受日期: 2016-12-14

* 通讯作者 Author for correspondence. E-mail: jinlongzhang01@gmail.com

优势种一般需采集标本;在野外未能鉴定出的种往往也需要采集标本进一步鉴定和分析。植物标本还是DNA条形码以及群体遗传学研究的重要凭证,分子生物学家可以从标本上取样,提取DNA并测序,从而获得各物种的DNA序列,以推断各类群的系统发育关系(CBOL Plant Working Group et al, 2009)。通过标本采集记录,人们还可以了解物种分布区大小(Rivers et al, 2011),分析或预测物种潜在分布区对气候变化或人为干扰的响应(McGraw, 2001),探讨物种适应性进化的格局与过程(Evans et al, 2009; Smith & Donoghue, 2010),评估生物入侵的风险(Weber, 1998; Delisle et al, 2003; 朱丽和马克平, 2010)等。标本数据更是全球生物多样性信息网络(Global Biodiversity Information Facility, GBIF, <http://www.gbif.org/>)最基础的数据之一。

采集和整理标本是植物标本馆最基本的工作(Bridson & Forman, 1998)。据纽约植物园世界标本馆索引(Herbarium Index, <http://sweetgum.nybg.org/science/ih/>)统计,截止到2016年12月,全世界约有3,000家标本馆,收藏了大约3.5亿份植物标本。中国在世界标本馆索引登记过的标本馆约350家,馆藏最为丰富的是中国科学院植物研究所标本馆(PE, 260万份),其次是中国科学院昆明植物研究所标本馆(KUN, 110万份)和中国科学院华南植物园(SCIB, 100万份)。即使是小型的标本馆,如香港植物标本室(HK)也保存约4万份,嘉道理农场暨植物园植物标本室(KFBG)的标本数量也超过1万份。标本的采集、鉴定信息等一般都包括在标本标签中。面对数量庞大的标本,植物考察队以及标本馆工作人员每天要耗费大量时间和精力为成百上千的标本准备标签,并校对标签上的采集和鉴定信息,以确保准确无误。

我国标本馆的标签通常分为采集标签(collecting notes)和鉴定标签(determination/annotation slips)两种。采集标签是植物工作者在野外采集标本的记录,其信息包括:采集人、采集号、采集日期、采集地点、植株高度、花果特征、生境以及伴生种等,过去多为手写。采集标签一般贴在台纸的左上角。鉴定标签或备注标签主要是植物学家或者标本馆工作人员对该标本的鉴定或者其他备注(如DNA采样、异名、科属变化等)。记录的信息一般包括:完整学名、鉴定人、鉴定日期等。如果该份标本为模

式标本,还需要标注模式标本的类型(如Holotype、Isotype)、参考文献以及与近缘种的区别等。鉴定标签一般较小,贴于标本的右下角或者左下角。

随着电脑和激光打印机的普及,国际上多数标本馆都已将标本采集和鉴定信息更加简洁、美观地整合到一张标本标签中。为了方便查阅,标本标签一般位于台纸的右下角。如果标本有新的鉴定,则增加新的鉴定标签,贴在原标本标签上方或台纸的左下角。

打印植物标本标签常通过数据库生成报表完成。国内外用于打印标本标签的软件已经有很多,如中国科学院植物研究所开发的EasyLabels(<http://pe.ibcas.ac.cn/download/download.html/>),英国牛津大学植物科学系编写的BRAHMS(<http://herbaria.plants.ox.ac.uk/bol/brahms/>),美国佛罗里达大学植物标本馆Kent Perkins编写的PLabel(<http://www.flmnh.ufl.edu/herbarium/pl/>),康涅狄格大学Robert K. Colwell编写的Biota(<http://vicero.yeeb.uconn.edu/biota/biota/>),堪萨斯大学生物多样性研究所编写的Specify(<http://specifyx.specifysoftware.org/>)等。也有人通过Microsoft Word的邮件合并(mail merge)功能生成标本标签(<https://cmlr.uq.edu.au/guide-creating-herbarium-specimen-labels-using-mail-merge/>, <https://www.idigbio.org/sites/default/files/workshop-presentations/small-herbarium2013/Using%20Label%20Merge.pdf/>)。

虽然生成标本标签的软件众多,但是不同的软件运行的操作系统可能不同,如EasyLabels、BRAHMS、PLabel只能在Windows上运行,且不同软件数据录入的格式和字段多有不同,往往导致数据传输上的不便。为了更方便地进行标本数据传输,方便数据的集中处理和分析,人们提出了Darwin Core数据标准(Wieczorek et al, 2012, <http://rs.tdwg.org/dwc/>)。Darwin Core规定了标本数据传输的标准字段,从而简化了数据共享等阶段需要进行的数据标准化过程,减少了工作量。Darwin Core在2009年成为全球生物多样性信息网络的标准数据格式。然而使用Darwin Core格式保存国内标本馆的标本记录时,某些字段需要扩展,某些字段需要简化,这就形成了中国数字植物标本馆CVH5.0标准(<http://www.cvh.org.cn/news/3/>)。

基于简化过的Darwin Core数据标准和CVH5.0

标准, 我们用R语言(R Core Team, 2016)编写了herblabel程序包, 它可以在Windows、MacOS、Linux等多种操作系统上运行。考虑到易用性, 标本采集记录以人们熟悉的Microsoft Excel格式输入, 继而读取到R中。herblabel生成的标签文件为富文本格式(rich text format, RTF), 可在Microsoft Word等办公软件中方便地编辑。最重要的是, herblabel在生成标本标签过程中加入了纠错机制, 可检查所输入采集信息的完整性以及拼写错误, 并将日期转换为完整格式, 避免歧义(图1)。为了更好地在标签中展示采集、鉴定信息, 本文总结了植物标本标签设计的原则, 介绍了herblabel程序包及其功能, 并通过实例介绍了如何通过herblabel生成标本标签和鉴定标签。

1 植物标本标签设计的原则

为了提高标本标签的可读性与科研价值, 同时减少标签内容检查的工作量, 建议在打印标本标签

时参考以下原则: 信息完整、准确、简明清晰、统一、美观和可编辑、自动纠错。

1.1 信息完整

(1)基本采集信息的完整性。标本标签需要提供采集人、采集号、采集日期、采集地点、经纬度、海拔, 以及所采标本的植株高度、颜色、气味、多度、伴生种等信息。同时, 标本标签上应该有初步鉴定结果。这几项中, 采集人、采集号、采集地点、采集日期是4项最基本的信息。如果这些信息不完整, 标本的价值就会大大降低。其中采集人和采集号相当于标本的身份证, 不能缺失或有歧义。

(2)地名等级的完整性。从国家、省份、县到乡镇、村寨等各地名等级, 都必须完整且规范, 符合当前的地名标准。例如, 中国广东省肇庆市鼎湖山鸡笼山, 英文地点仍然按照由大到小的顺序, 即CHINA, Guangdong, Zhaoqing, Dinghushan, Jilongshan。也可以在标签标题中注明是某一国家或

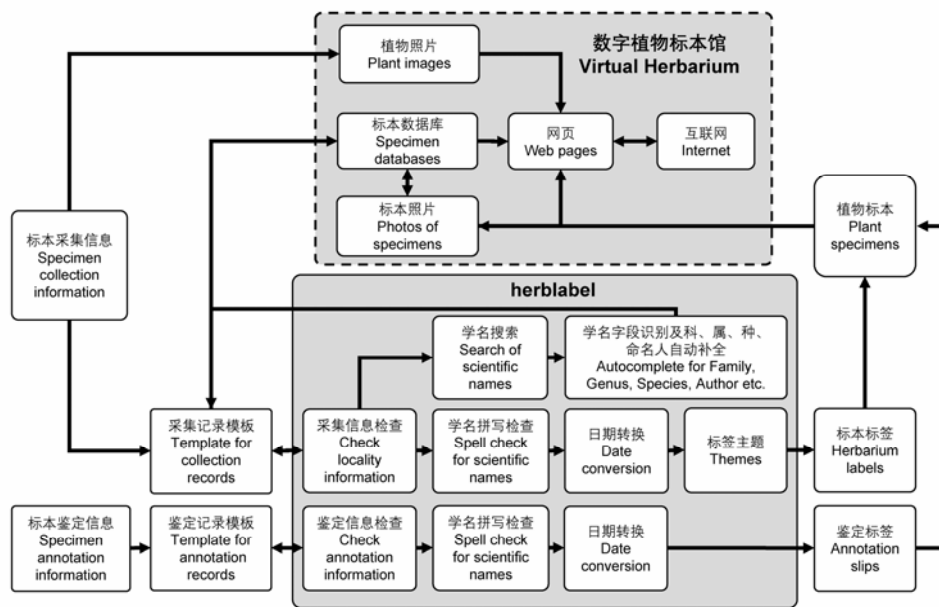


图1 R程序包herblabel的工作流程及其在标本馆中的作用。herblabel要求标本采集数据以简化的Darwin Core数据交换标准的数据框输入, 可自动进行采集信息完整性、学名拼写以及科属对应关系检查, 用户只需要提供完整学名或《中国植物志》、Flora of China的中文名, herblabel便可在内置的数据库中查找对应的学名, 并对科、属、种和命名人信息自动补全, 生成富文本格式(RTF)的标本标签或者鉴定标签。贴好采集或鉴定标签后, 标本的照片可用于数字植物标本馆中。同时, 经过herblabel检查过的标本采集数据, 可以直接用于数字植物标本馆的数据库建设中。

Fig. 1 The workflow of the herblabel R package and its role in a herbarium. herblabel uses data.frame based on the simplified Darwin Core data exchange template, and can check spelling of scientific names, find out the family in new classification systems as well as examine completeness of the collection records. Users need to supply complete scientific names for a species or Chinese names either from Flora Reipublicae Popularis Sinicae or Flora of China. The package will automatically search and find out the corresponding scientific names based on embedded databases, and will fill in the family, genus, specific epithet, authority etc. for the records. Herbarium specimen labels and annotation labels will be in Rich Text Format (RTF), suitable for being viewed and edited in Microsoft Word. The photos of the herbarium specimen, and the examined collection records could then be used in a virtual herbarium.

地区的植物,地名由省份开始。

(3)学名的完整性。目前有些标本馆标本标签上仅有中文名,或者仅有不带作者的学名,这种情况需要避免。标签上的学名必须包含命名人和鉴定日期,同时命名人的缩写要规范。

1.2 准确

标签中的信息不能有歧义。如采集日期09-07-13, 09/07/13, 09/July/13等在不同国家或地区含义不同,这样的格式是不能采用的。我们建议采用英文全称,如09 July 2013,或者中文全称的格式,如2013年7月9日。

1.3 简明清晰

充分利用标签的空间,不显示无信息的项目。例如,无经纬度的标本就不要留出经纬度的位置,无花果特征的描述则不要显示相应的项目。采集信息应该按照条理组织,建议以如下的顺序打印:(1)标本馆名称;(2)科名;(3)学名;(4)采集地点;(5)经度;(6)纬度;(7)海拔;(8)花果特征;(9)生境;(10)采集人;(11)采集号;(12)采集日期;(13)鉴定人;(14)鉴定日期。标本采集信息按照顺序排列,会提高标签的可读性。

1.4 统一

包括格式和度量单位统一。命名人的拼写要遵循学界接受的标准;学名的拼写以专家意见或者植物志的接受名为准。经纬度统一采用WGS1984坐标系,并遵循度、分、秒格式,标明东经、北纬,如22°23'16" N, 114°59'59" E。海拔则统一用米。

1.5 美观

标签的字体、布局应美观易读,采用中英文对照或纯英文打印,以方便国际交流。字体大小适中;段落之间的行距合适;拉丁文统一用斜体,以便和英文相区别;标签的边距适当,文字不能太靠近标签边缘,以免裁切时损坏;标签的高度应随文字多少而自动调整;标签需要有裁切线作为指引,保证边缘整齐。建议英文使用Times New Roman字体、9磅,中文使用宋体、小五号;行距以1.2–1.5倍为佳。

1.6 可编辑

生成的标签文件要能灵活地编辑,对于发现的问题,无论是在标本采集数据本身还是在生成的标签中,都可以快速修改。

1.7 具备纠错机制

生成标签的程序最好具备纠错机制,能够自动

检查的信息包括:(1)学名的拼写,如是否为接受名或者有无拼写错误等;(2)科属的拼写;(3)科属的对应关系等。在检查过程中如果发现错误,需要给出相应的提示。

我们在设计herblabel程序包时充分考虑了以上原则,并努力将标本数据的输入标准化,以使生成的标签准确、清晰、易读。

2 herblabel程序包及其功能

herblabel是用R语言开发的,用于植物标本标签和鉴定标签打印的程序包。源代码保存在Rforge (Theußl & Zeileis, 2008, https://r-forge.r-project.org/R/?group_id=2053/)以及github (<https://github.com/helixcn/herblabel/>)上,可以自由下载和使用(参见附录1)。该程序包可以在Windows、MacOS以及Linux等多种操作系统上运行,可直接通过Rforge安装,安装命令为install.packages("herbalbel", repos="http://R-Forge.R-project.org")。生成的标签为富文本格式(RTF)的文件,可通过Microsoft Word等办公软件进行进一步编辑修改。一般情况下,herblabel自动生成的标签已经足够美观,无需进一步编辑。

在当前版本(0.5.5)中,herblabel的函数包括:herbarium_label、annotation_label、fill_dwc和parse_taxa等,下面分别介绍。

2.1 herbarium_label: 生成植物标本标签

调用格式为herbarium_label(dat = NULL, spellcheck = TRUE, theme = c("KFBG", "PE", "KUN", "HU"), outfile = "herblabel.rtf"),该函数能够基于简化的Darwin Core格式的数据框(data.frame,其字段参见附录3)生成RTF格式的标本标签(图2),各标签之间设3个点组成的裁切线。其中:dat为基于简化的Darwin Core数据标准的标本采集记录,必须是数据框格式。spellcheck为逻辑选项,为TRUE或者FALSE,默认为TRUE,即进行拼写检查,此时函数会检查学名的接受状态(参见4.1),科名的拼写以及科属的对应关系,如果与内置数据库出现冲突,会在生成的RTF标签文件中提示正确的科属关系(图3)。theme为标签的样式,这里称为主题(见附录1),目前只能为"KFBG", "PE", "KUN", "HU"中的一个,分别为嘉道理农场暨植物园植物标本室、中国科学院植物研究所标本馆、中国科学院昆明植物研究所标本馆、哈佛大学植物标本馆的缩写。这

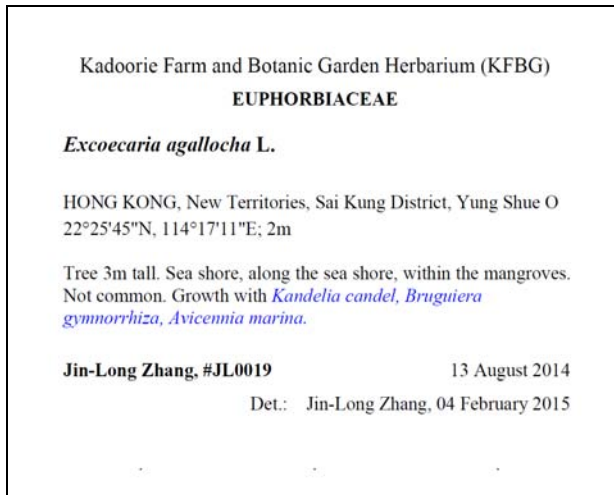


图2 herblabel生成的RTF标本标签。拉丁名自动设定为斜体, 请注意底部由三个点组成的裁切线。

Fig. 2 An RTF herbarium label generated by herblabel, note the scientific names have been automatically identified and changed to italic. Please note the cutting lines comprised by three dots at the bottom.

些标签的样式均参考中国数字植物标本馆CVH新近收录的标本设计, 样式上与以上标本馆目前所用的标签类似, 但不一定完全相同。outfile字符串表示生成的RTF标签文件的文件名。

2.2 annotation_label: 生成鉴定标签

调用格式为 `annotation_label(dat = NULL, spellcheck = TRUE, outfile = "Annotation_Labels.RTF")`, 可基于标本鉴定记录生成RTF的鉴定标签。其中dat为标本鉴定记录模板(参见附录4), 格式为数据框; `spellcheck = TRUE`, 表示进行学名的拼写检查。在设定进行学名检查后, 如果标签中的学名并非接受名(参见4.1), RTF标签中就会给出提示。outfile为字符串, 表示生成的RTF标签文件的文件名。

2.3 fill_dwc: 查找学名并补齐科、属、种、命名人等信息

将一个完整的学名按照属、种、命名人、种下等级、种下等级加词、种下等级命名人逐项填写相当费时费力。fill_dwc函数能根据《中国植物志》或Flora of China收录的中文标准名或完整的学名, 自动将以上信息填写完整。调用格式为 `fill_dwc(dat, namedb = c("spfrps", "spfoc"))`, 其中dat是简化的Darwin Core格式数据框, namedb只能是"spfrps"或"spfoc", 表示按照中文名将模板数据中的学名自动



图3 标本标签中拼写以及科属关系的检查示例。图中大戟科EUPHORBIACEAE因为输入时不慎加入数字, 导致herblabel提示其科名未获接受。Excoecaria在APG III系统属于大戟科EUPHORBIACEAE, 但是因为科名的拼写错误造成不匹配, 所以这里提示“按照The Plant List网站, 应为大戟科”。

Fig. 3 Checking of spelling and the family-genus relationship. Note that a digit was appended to EUPHORBIACEAE, so the family name could not be found among the embedded family name database. Excoecaria should be in EUPHORBIACEAE, however it could not be matched due to the spelling error and a warning was issued.

补充完整时, 分别依据的是《中国植物志》或Flora of China。该函数同时能为高等植物各类群的学名匹配科名: 其中苔藓植物分科完全遵循The Plant List网站数据, 蕨类植物(Christenhusz et al, 2011b)、裸子植物(Christenhusz et al, 2011a)、被子植物主要遵循The Plant List网站, 按照APG III分类系统(The Angiosperm Phylogeny Group, 2009), 部分属的科名参考刘冰等(2015)关于APG III系统的论文。

fill_dwc以及herbarium_label函数配合时使用的R代码如下:

```
## 导入openxlsx程序包, 以读取xlsx文件
library(openxlsx)
## 读取程序包内置模板文件
dat <- read.xlsx(system.file("extdata",
"DARWIN_CORE_HERBARIUM_RECORDS.xlsx", package
= "herblabel"))
## 自动填写科属种以及命名人信息
dwc_filled <- fill_dwc(dat)
```

2.4 parse_taxa: 将学名分成属、种、命名人等字段
格式为parse_taxa(taxa), 用于将学名划分成属、种加词、命名人、种下等级(如subsp., var., f.)、种下等级加词、命名人等字段。

3 用herblabel生成标本标签和采集标签

3.1 打印标本标签

假设有名为herbarium_records.xlsx的标本记录数据保存在D:/herbarium/, 则在R控制台(RGui)中输入以下命令, 即可生成RTF标签:

```
## 设定工作路径, 注意斜杠要向右
setwd("D:/herbarium/")
library(herblabel)
library(openxlsx)
dat <- read.xlsx("herbarium_records.xlsx")
#### 检查拼写等
herbarium_label(dat, outfile = "herbarium_labels.RTF")
#### 不进行拼写检查
herbarium_label(dat, spellcheck = FALSE, outfile = "herbarium_labels.RTF")
```

默认情况下, spellcheck = TRUE, 此时herbarium_label函数将进行以下检查: (1)学名有效性, 即是否是The Plant List网站上的接受名; (2)属名的拼写; (3)科的拼写; (4)科属对应关系(图1)。这些信息会在生成的RTF文件中给出提示。

3.2 打印鉴定标签

将该annotation_template.xlsx拷贝到D:/herbarium/, 填好相应的鉴定信息后, 在R控制台中运行如下命令:

```
setwd("D:/herbarium")
library(herblabel)
library(openxlsx)
dat <- read.xlsx("annotation_template.xlsx")
annotation_label(dat)
```

由herbarium_label和annotation_label函数生成的RTF文件(图2-4), 可方便地用Microsoft Word进一步编辑。

为了更方便地使用herblabel打印标本标签及鉴定标签, 并做好数据备份, 作者编写了run_herblabel R脚本(下载地址: <https://github.com/helixcn/>

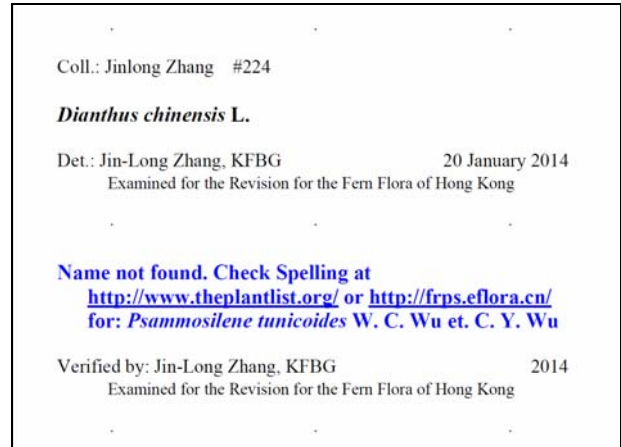


图4 标本鉴定标签举例。根据输入的文字, 鉴定标签的大小可以非常灵活。鉴定标签也可以检查拼写以及学名是否为接受名。请注意底部由三个点组成的裁切线。

Fig. 4 Example of Annotation labels. These labels can be quite flexible in size depending on the text provided. Checking of the spelling and validity could also be implemented. Please note the cutting lines comprised by three dots at the bottom.

run_herblabel/ 或 <http://blog.sciencenet.cn/home.php?mod=attachment&id=92319/>, 文件夹说明参见附录2)。用户在采集信息模板输入数据后(附录3), 通过双击相同文件夹下的Window批处理(.bat)文件, 就可以直接生成相应的标签, 同时按照时间顺序做好备份。程序会自动检查出现的错误或者问题, 同时会在Microsoft Excel模板文件中给出提示。由于数据以简化的Darwin Core标准保存, 可以直接用于GBIF数据共享或者数字植物标本馆(纪力强, 2000; 许哲平等, 2010)的建设中。

4 标签上学名、科属的可能错误及其改正

4.1 学名或者命名人拼写错误或者为异名

在准备标签时直接输入学名往往容易出错, 这些错误通常包括: (1)属名或者种加词的拼写错误; (2)大小写错误; (3)命名人的缩写不符合规范等。命名人的拼写方式不统一严格说来虽然不能称为错误, 但也可能会影响标签的阅读, 这些问题将影响标本的整理。herbarium_label函数可以将Microsoft Excel中输入的学名与内置的数据库进行严格比对, 内置的数据库基于The Plant List网站(<http://www.theplantlist.org/>)所有的接受名以及《中国植物志》(<http://frps.eflora.cn/>)和Flora of China(http://www.efloras.org/flora_page.aspx?flora_id=2/)

的接受名。若输入的学名与数据库中的学名拼写不符, 生成的RTF文件中就会给出提示, 以使用户及时发现和处理。

4.2 科属拼写错误以及对应关系的检查

APG分类系统是目前为止最接近被子植物真实系统发育关系的分类系统(The Angiosperm Phylogeny Group, 2016), 很多标本馆(包括英国的邱园, Royal Botanical Gardens, Kew)正在转向使用APG分类系统, 而该系统的科属划分和《中国植物志》以及地方志有很多不一致的地方。物种所在的APG III等现代分类系统的科名以及接受状态可在The Plant List网站查询, 但如果每一条记录都在该网站查询会十分费时。为此herbarium_label函数在生成标本标签时会进行科名和属名的拼写以及对应关系检查。若该条标本存在科、属拼写错误, 或者标本的科属对应关系与APG III等新分类系统分科不同时给出提示(图3)。由于分子以及解剖学等多方面的证据不断更新, APG等新系统也在不断变化中, 最新的APG IV (The Angiosperm Phylogeny Group, 2016)已经发表, 但是涉及的科属变化还有待进一步整理。高等植物其他类群的新系统也在不断变化中。我们会适时将herblabel的内置数据更新到APG IV等新的分类系统。

4.3 伴生种学名的斜体

生境描述中如出现伴生种的学名, herbarium_label函数可基于内置的拉丁文数据库自动匹配拉丁文, 并将其转换为斜体, 用蓝色标记。因此, 如果伴生种的学名出现拼写错误则不会被识别, 这样在查看标签时容易发现并改正。

4.4 处理有歧义日期

herblabel生成的标签中, 日期尽量采用英文全称, 输出格式为“21 July 2016”, 其中月份为英文全称, 年份用4位数字, 置于最后。但在模板中输入日期时, 不必按照英文全称填写, 只需按照yyyy-mm-dd的ISO 8601格式(Wolf & Wicksteed, 1998)填写即可, 即年、月、日用“-”隔开, 例如“2015-12-8”表示2015年12月8日。如果用户使用自己的日期格式导致软件无法识别和自动转换, herblabel将保留作者提供的日期格式。由于输入数据用的是Microsoft Excel模板, 而Excel倾向于将数据转换为一系列整数, herblabel考虑到这种情况, 在0.5.5版本后能够读取并识别Excel转换为整数之后

的日期。

基于简化的Darwin Core和CVH5.0的数据交换标准利用R语言编写的程序包herblabel, 实现了标本标签的快速批量生成; 其自动纠错功能能够对学名的拼写以及科属对应关系进行检查, 有效减少了数据录入过程中产生的错误。本程序包在打印标签时使用的是以Darwin Core标准保存的标本数据库, 不仅方便统计和管理, 也可以直接用于GBIF数据共享或者数字植物标本馆的建设。该程序包将显著提高植物标本馆在标本制作、管理和信息录入时的工作质量和效率, 并在植物生物多样性编目相关的研究中发挥重要作用。

致谢: 感谢Stephan Gale博士、Pankaj Kumar博士、黄世芳女士、冯嘉恩女士对植物标本标签的设计提出宝贵意见。衷心感谢三位审稿专家以及编委提出的建设性意见和建议。

参考文献

- Bridson D, Forman L (1998) The Herbarium Handbook, 3rd edn. Royal Botanic Gardens: Kew, Richmond.
- CBOL Plant Working Group, Hollingsworth PM, Forrest LL, Spouge JL, Hajibabaei M, Ratnasingham S, van der Bank M, Chase MW, Cowan RS, Erickson DL, Fazekas AJ, Graham SW, James KE, Kim KJ, Kress WJ, Schneider H, van AlphenStahl J, Barrett SCH, van den Berg C, Bogarin D, Burgess KS, Cameron KM, Carine M, Chacón J, Clark A, Clarkson JJ, Conrad F, Devey DS, Ford CS, Hedderson TAJ, Hollingsworth ML, Husband BC, Kelly LJ, Kesanakurti PR, Kim JS, Kim YD, Lahaye R, Lee HL, Long DG, Madriñán S, Maurin O, Meusnier I, Newmaster SG, Park CW, Percy DM, Petersen G, Richardson JE, Salazar GA, Savolainen V, Seberg O, Wilkinson MJ, Yi DK, Little DP (2009) A DNA barcode for land plants. Proceedings of the National Academy of Sciences, USA, 106, 12794–12797.
- Christenhusz M, Reveal J, Farjon A, Gardner MF, Mill RR, Chase MW (2011a) A new classification and linear sequence of extant gymnosperms. Phytotaxa, 19, 55–70.
- Christenhusz M, Zhang XC, Schneider H (2011b) A linear sequence of extant families and genera of lycophytes and ferns. Phytotaxa, 19, 7–54.
- Delisle F, Lavoie C, Jean M, Lachance D (2003) Reconstructing the spread of invasive plants: taking into account biases associated with herbarium specimens. Journal of Biogeography, 30, 1033–1042.
- Evans MEK, Smith SA, Flynn RS, Donoghue MJ (2009) Climate, niche evolution, and diversification of the “Bird-Cage” evening primroses (*Oenothera*, sections *Anogra* and *Kleinia*). The American Naturalist, 173,

- 225–240.
- Ji LQ (2000) The construction status of biodiversity information system and a brief introduction of CBIS. *Chinese Biodiversity*, 8, 41–49. (in Chinese with English abstract) [纪力强 (2000) 生物多样性信息系统建设的现状及CBIS简介. *生物多样性*, 8, 41–49.]
- Liu B, Ye JF, Liu S, Wang Y, Yang Y, Lai YJ, Zeng G, Lin QW (2015) Families and genera of Chinese angiosperms: a synoptic classification based on APG III. *Biodiversity Science*, 23, 225–231. (in Chinese with English abstract) [刘冰, 叶建飞, 刘夙, 汪远, 杨永, 赖阳均, 曾刚, 林秦文 (2015) 中国被子植物科属概览: 依据APG III系统. *生物多样性*, 23, 225–231.]
- McGraw JB (2001) Evidence for decline in stature of American ginseng plants from herbarium specimens. *Biological Conservation*, 98, 25–32.
- Rivers MC, Taylor L, Brummitt NA, Meagher TR, Roberts DL, Lughadha EN (2011) How many herbarium specimens are needed to detect threatened species? *Biological Conservation*, 144, 2541–2547.
- R Core Team (2016) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>. (accessed on 2016-8-18)
- Smith SA, Donoghue MJ (2010) Combining historical biogeography with niche modeling in the *Caprifolium* clade of *Lonicera* (Caprifoliaceae, Dipsacales). *Systematic Biology*, 59, 322–341.
- The Angiosperm Phylogeny Group (2016) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society*, 181, 1–20.
- Theußl S, Zeileis A (2008) Collaborative software development using R-Forge. <https://r-forge.r-project.org/>. (accessed on 2016-8-18)
- Weber E (1998) The dynamics of plant invasions: a case study of three exotic goldenrod species (*Solidago* L.) in Europe. *Journal of Biogeography*, 25, 147–154.
- Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, Giovanni R, Robertson T, Vieglais D (2012) Darwin Core: an evolving community-developed biodiversity data standard. *PLoS ONE*, 7, e29715.
- Wolf M, Wicksteed C (1998) Date and time formats. <http://www.w3.org/TR/NOTE-datetime/>. (accessed on 2016-8-18)
- Xu ZP, Cui JZ, Qin HN, Ma KP (2010) On the architecture of biodiversity e-Science infrastructure in China. *Biodiversity Science*, 18, 480–488. (in Chinese with English abstract) [许哲平, 崔金钟, 覃海宁, 马克平 (2010) 中国生物多样性 e-Science平台建设构想. *生物多样性*, 18, 480–488.]
- Zhu L, Ma KP (2010) On the niche stasis of intercontinental invasive plants. *Biodiversity Science*, 18, 547–558. (in Chinese with English abstract) [朱丽, 马克平 (2010) 洲际入侵植物生态位稳定性研究进展. *生物多样性*, 18, 547–558.]

(责任编辑: 严岳鸿 责任编辑: 黄祥忠)

附录 Supplementary Material

附录1 herblabel程序包的安装、配置和使用

Appendix 1 Installation, configuration and using of herblabel
<http://www.biodiversity-science.net/fileup/PDF/2016230-1.pdf>

附录2 对run_herblabel文件夹下各文件的详细说明

Appendix 2 Description for the files in folder "run_herblabel"
<http://www.biodiversity-science.net/fileup/PDF/2016230-2.pdf>

附录3 标本标签模板各列说明

Appendix 3 Description for the columns of the herbarium label template
<http://www.biodiversity-science.net/fileup/PDF/2016230-3.pdf>

附录4 鉴定标签模板各列说明

Appendix 4 Description for the columns of the annotation label template
<http://www.biodiversity-science.net/fileup/PDF/2016230-4.pdf>