

植物标本标签文档类 nanicolle^{*}

杨宇昌[✉] (NEGIMORIYA)

2019 年 9 月 27 日 ver. 2.01

植物标本是经处理能够长期保存的植物材料。它们是植物居群的样本，通过标本可以了解植物居群。由于植物材料本身并不能完整地反映居群的重要信息，标本采集者需要作出补充并记录成**采集标签**，与标本一同保存。植物分类学者经研究可以判定标本对应居群所属的物种，其意见则以**鉴定标签**的形式，附在标本上。

本文介绍的 L^AT_EX 文档类 nanicolle 可用于植物标本中文采集标签和鉴定标签的排版。两种标签分别用 `\Collect` 命令和 `\Identify` 名称产生。排好的标签可直接打印。

nanicolle 依赖 ctex 宏集和 graphicx、geometry、multicol、calc、listofitems、color、xstring 等宏包。目前，nanicolle 文档类仅支持 X_YL^AT_EX 编译方式。

nanicolle 按 The L^AT_EX Project Public License (LPPL) 1.3c 协议¹授权。

目录

1	nanicolle 类文档的结构	2
2	用 <code>\Collect</code> 命令生成采集标签	2
3	用 <code>\Identify</code> 命令生成鉴定标签	4
4	其他问题	4
4.1	用工作表软件储存原始数据	4
4.2	PDF 文件的打印设置	5
5	完整实例	5
6	更新日志	7

^{*} 本项目的 Github 仓库地址为 <https://github.com/Mikumikunisiteageru/nanicolle>。

[✉] 电子邮箱为 yang.yc.allium@gmail.com。

¹ LPPL 1.3c 协议的详细内容见 <http://www.latex-project.org/lppl.txt>。

1 nanicolle 类文档的结构

一份使用 nanicolle 类的文档应是包含以下 5 部分、以后缀 .tex 结尾的纯文本：

1. 载入 nanicolle 文档类，即 `\documentclass[⟨选项⟩]{nanicolle}`。其中，⟨选项⟩控制文档类的行为，例如 `nomap` 可抑制采集标签中的地图，`autoduplicate` 则能按标本份数重复采集标签；选项超过一个时，其间要用 `,` 分隔。不需要加载选项时，可写 `\documentclass{nanicolle}`。
2. 在导言区使用 `\heading` 和 `\subheading` 命令定义个性化内容。要在采集标签上方添加统一的标题，可写 `\heading{⟨标题⟩}`；副标题如需要则可由 `\subheading{⟨副标题⟩}` 命令产生。若跳过此部分，则采集标签上方无标题。
3. 正文开始，即 `\begin{document}`。
4. 在正文中使用 `\Collect` 和 `\Identify` 两种命令分别产生采集标签和鉴定标签，每个命令占一行。两种命令的格式参见第 2 和第 3 节。
5. 正文结束，即 `\end{document}`。

2 用 `\Collect` 命令生成采集标签

`\Collect` 命令的格式为

```
\Collect→⟨记录号⟩→⟨采集人⟩→⟨采集号⟩→⟨采集日期⟩→⟨科名⟩
      →⟨中文名⟩→⟨学名⟩→⟨照片号⟩→⟨标本份数⟩→⟨产地⟩→⟨经度⟩→⟨纬度⟩
      →⟨海拔⟩→⟨生境⟩→⟨生活型⟩→⟨体高⟩→⟨胸径⟩→⟨附注⟩
```

当中 `→` 表示水平制表符 (U+0009, 可按键盘 Tab 键输入)。每行最多能有一个 `\Collect` 命令；每个 `\Collect` 命令须出现在一行开始处，其所有参数必须出现在同一行中。某些参数可以为空，但 `→` 都须保留。以下就诸参数分别说明。

1. `⟨记录号⟩`：仅为数据组织方便而设，不在采集标签上出现。
2. `⟨采集人⟩`：标本的采集者的姓名。如有多个采集者，应尽量列举姓名，其间可以用顿号分隔；不建议填写不包含人名的采集团队名称。此参数不得为空。
3. `⟨采集号⟩`：属于 `⟨采集人⟩` 中第一个采集者的记录标本采集的序列号码。建议为简单的整数，从 1 开始递增。
4. `⟨采集日期⟩`：采集植物材料的日期，可使用 `⟨年⟩.⟨月⟩.⟨日⟩` 的阿拉伯数字表示。此参数不得为空。
5. `⟨科名⟩`：临时鉴定的科的名称，建议使用中文表示。
6. `⟨中文名⟩`：临时鉴定的植物中文名。

7. **〈学名〉**: 临时鉴定的植物学名, 建议不带作者引证。当非空时, 可用通式 **〈属级部分〉〈种级部分〉〈种下等级部分〉** 表示。其中, **〈属级部分〉** 有 2 类结构:

i. **〈属名〉**; ii. **×〈属名〉**。

〈种级部分〉 有 9 类结构:

i. **␣sp.**; ii. **␣sp.␣nov.**; iii. **␣〈种加词〉**; iv. **␣×〈种加词〉**;
v. **␣aff.␣〈种加词〉**; vi. **␣aff.␣×〈种加词〉**; vii. **␣cf.␣〈种加词〉**;
viii. **␣cf.␣×〈种加词〉**; ix. **␣'〈栽培种名〉'**。

当中的 **␣** 表示一个空格 (U+0020)。**〈种下等级部分〉** 仅当 **〈种级部分〉** 为结构 iii 或结构 iv 时可以非空, 此时它有 4 类结构:

i. **␣subsp.␣〈亚种加词〉**; ii. **␣var.␣〈变种加词〉**; iii. **␣f.␣〈变型加词〉**;
iv. **␣'〈栽培品种名〉'**。

8. **〈照片号〉**: 仅为数据组织方便而设, 不在采集标签上出现。
9. **〈标本份数〉**: 标本的同号复份数目, 用正整数表示。不在采集标签上出现。当文档类的 **〈选项〉** 包含 **autoduplicate** 时, 每个 **\Collect** 命令将输出 **〈标本份数〉** 张重复的采集标签。
10. **〈产地〉**: 植物材料的采集地的自然语言表述, 要包含完整的省级、县级、乡级行政单元序列和小地点, 小地点可参照地标表示, 以便考证与回访。此参数不得为空。
11. **〈经度〉**: 植物材料采集地的经度坐标, 用十进制的浮点数表示, 以角度为单位, 不带单位符号, 正数为东经, 负数为西经。不接受度分秒表示。
12. **〈纬度〉**: 植物材料采集地的纬度坐标, 用十进制的浮点数表示, 以角度为单位, 不带单位符号, 正数为北纬, 负数为南纬。不接受度分秒表示。
13. **〈海拔〉**: 植物材料采集地的海拔高度, 用浮点数表示, 以米为单位, 不带单位符号, 可以为负数。
14. **〈生境〉**: 植物居群的野外生活环境, 如 **山地**、**林缘**、**溪边** 等, 但对于栽培的植物, 应填写 **栽培** 而非具体的生活环境。
15. **〈生活型〉**: 植物居群的生活型, 如 **乔木**、**灌木**、**藤本** 等。
16. **〈体高〉**: 植物居群中个体的典型高度, 用浮点数表示, 以米为单位, 不带单位符号。
17. **〈胸径〉**: 植物居群中个体的典型胸径, 用浮点数表示, 以厘米为单位, 不带单位符号。
18. **〈附注〉**: 包含其他在植物标本上无法观察的有价值的信息, 如各部的颜色、气味, 树皮的纹理, 传粉者, 植物在当地的俗名、用途、常见程度等。

nanicolle 文档类默认在经纬度均非空且经度在东经 73–136 度间、纬度在北纬 17–54 度间时在采集标签下方输出地图, 以标明该坐标的位置。要抑制这个功能, 可以在载入文档类时使用 **nomap** 选项, 见第 1 节。

3 用 `\Identify` 命令生成鉴定标签

`\Identify` 命令的格式为

```
\Identify→⟨记录号⟩→⟨学名⟩→⟨中文名⟩→⟨鉴定人⟩→⟨鉴定人标准形式⟩
→⟨鉴定日期⟩→⟨批注⟩
```

和 `\Collect` 一样，每个 `\Identify` 命令及其所有参数需独占一行。以下就诸参数分别说明。

1. `⟨记录号⟩`：仅为数据组织方便而设，不在采集标签上出现。
2. `⟨学名⟩`：正式鉴定的植物学名，须带有作者引证，可用通式 `⟨属级部分⟩⟨种级部分⟩⟨种下等级部分⟩⟨作者引证⟩` 表示。此参数不得为空。
3. `⟨中文名⟩`：与 `⟨学名⟩` 相关联的中文种名。
4. `⟨鉴定人⟩`：标本鉴定者的姓名。
5. `⟨鉴定人标准形式⟩`：标本鉴定者的姓名在分类学上约定的标准形式。此参数与 `⟨鉴定人⟩` 不得同时为空。
6. `⟨鉴定日期⟩`：鉴定标本的日期，格式同 `⟨采集日期⟩`。此参数不得为空。
7. `⟨批注⟩`：需另外表明的分类学意见。

不同于 `\Collect` 命令，`\Identify` 命令不带 `⟨标本份数⟩` 参数，故每个 `\Identify` 命令固定地输出一张鉴定标签。要输出重复的鉴定标签，只能手动复制命令所在的行。

4 其他问题

4.1 用工作表软件储存原始数据

`\Collect` 与 `\Identify` 之所以不遵照 L^AT_EX 惯例而用制表符 `→` 作为参数的定界符，其本意是使得原始数据可以储存工作表（Spreadsheet）软件²中。当纯文本粘贴到工作表中时，换行符分隔各行，制表符分隔行内各列。当工作表的内容粘贴为纯文本时，各行由换行符分隔，行内各列由制表符分隔（形成 TSV 格式）。事实上，这种机制允许储存于工作表中的原始数据可以直接为 X_QL^AT_EX 引擎所读取。具体来说，可以建立一个形如表 1 的工作表来储存原始数据。这样，工作表中的若干行被选中并复制到纯文本环境下后便直接转换成了 `\Collect` 或 `\Identify` 命令要求的格式。而工作表中的原始数据库在这些命令所要求的参数外还能储存更丰富的信息，这些额外信息会被 `\Collect` 或 `\Identify` 命令忽略，不影响标签的内容。

²常见的工作表软件如 Microsoft Office Excel。

命令	记录号	采集人	采集号	采集日期	科名	中文名	学名	...
\Collect	1
\Collect	2
\Collect	3

表 1: 储存采集标签原始数据的工作表示例

4.2 PDF 文件的打印设置

编译得到的 PDF 文件最终要用打印机打印并裁剪，成为可实际使用的标签。打印机的可打印范围一般比纸面小，为保留文件完整性，打印程序通常默认将文件缩小并布局在纸面中心。这样，按印出来的线条裁切，便会出现两侧的标签宽、中间的标签窄的现象。为避免这种问题，打印 PDF 文件时应选中“使用原始页面大小”选项。

5 完整实例

以下是一份符合 nanicolle 规范的文档实例，保存为 `nanicolle-example.tex`。

```

1 \documentclass[autoduplicate]{nanicolle}
2 \begin{document}
3 \Collect→2912→杨宇昌→6159→2018.9.24→莎草科→褐穗莎草
   →Cyperus_fuscus→7038→1
   →北京市怀柔区琉璃庙镇双文铺村白河湾湿地生态公园→116.678772
   →40.655164→214→河边荒草地→披散草本→0.2→
   →须根甚浅，不过2--3cm；花的鳞片两侧紫褐色，中间绿色。
4 \Collect→1997→杨宇昌→5731→2018.5.8→忍冬科→苦糖果
   →Lonicera_fragrantissima_subsp._standishii→7609→2
   →河北省邯郸市武安市管陶乡董家门村西至洞垆途中→→→1356.0
   →山坡灌草丛→灌木→3→→果实橙红色，味甜微苦。
5 \Identify→392→Allium_atrosanguineum_var._tibeticum_(Regel)_
   G._H._Zhu_&_Turland→藏葱→杨宇昌→→→2018.10.7→
6 \Identify→176→Acer_davidii_subsp._grosseri_(Pax)_P._C._de_
   Jong→葛萝槭→杨宇昌→→→2018.4.19→
7 \Identify→230→Erysimum_xcheiri_(L.)_Crantz→桂竹香→杨宇昌→
   →2018.5.17→
8 \Identify→590→Koenigia_alpina_(All.)_T._M._Schust._&_Reveal
   →高山神血宁→杨宇昌→→→2019.4.13→
9 \end{document}

```

注意，每个 `\Collect` 或 `\Identify` 命令及其所有参数实际上均位于同一行中，换行只是便于展示。将此文档放置在 `nanicolle.cls` 所在的目录下，命令行中运行 `xelatex nanicolle-example`，即可在同一目录中得到 `nanicolle-example.pdf`，其效果见图 1。

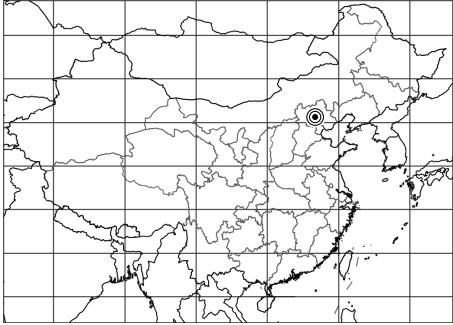
<div>采集人 杨宇昌</div> <div>采集号 6159 采集日期 2018.9.24</div> <div>产地 北京市怀柔区琉璃庙镇双文铺村白河湾湿地生态公园</div> <div>经度 116.678772°E 纬度 40.655164°N</div> <div>海拔 214 m 生境 河边荒草地</div> <div>生活型 披散草本 体高 0.2 m 胸径</div> <div>科名 莎草科 中文名 褐穗莎草</div> <div>学名 <i>Cyperus fuscus</i></div> <div>附注 须根甚浅，不过 2–3 cm；花的鳞片两侧紫褐色，中间绿色。</div>	<div>采集人 杨宇昌</div> <div>采集号 5731 采集日期 2018.5.8</div> <div>产地 河北省邯郸市武安市管陶乡董家门村西至洞埡途中</div> <div>经度 纬度</div> <div>海拔 1356.0 m 生境 山坡灌草丛</div> <div>生活型 灌木 体高 3 m 胸径</div> <div>科名 忍冬科 中文名 苦糖果</div> <div>学名 <i>Lonicera fragrantissima</i> subsp. <i>standishii</i></div> <div>附注 果实橙红色，味甜微苦。</div>
	<div><i>Allium atrosanguineum</i> var. <i>tibeticum</i> (Regel) G. H. Zhu et Turland</div> <div>藏葱</div> <div>杨宇昌 2018.10.7</div>
<div>采集人 杨宇昌</div> <div>采集号 5731 采集日期 2018.5.8</div> <div>产地 河北省邯郸市武安市管陶乡董家门村西至洞埡途中</div> <div>经度 纬度</div> <div>海拔 1356.0 m 生境 山坡灌草丛</div> <div>生活型 灌木 体高 3 m 胸径</div> <div>科名 忍冬科 中文名 苦糖果</div> <div>学名 <i>Lonicera fragrantissima</i> subsp. <i>standishii</i></div> <div>附注 果实橙红色，味甜微苦。</div>	<div><i>Acer davidii</i> subsp. <i>grosseri</i> (Pax) P. C. de Jong</div> <div>葛萝槭</div> <div>杨宇昌 2018.4.19</div>
	<div><i>Erysimum</i> × <i>cheiri</i> (L.) Crantz</div> <div>桂竹香</div> <div>杨宇昌 2018.5.17</div>
	<div><i>Koenigia alpina</i> (All.) T. M. Schust. et Reveal</div> <div>高山神血宁</div> <div>杨宇昌 2019.4.13</div>

图 1: 样例 PDF 文件 nanicolle-example.pdf 的前 2 栏（左半页）

6 更新日志

ver. 2.01 (2019.9.27)

改进采集标签中下划线表的实现。
支持在附注等项目中直接用 \LaTeX 命令如 `\textit` 等控制字体样式。
合并解析拉丁学名的两处代码，允许在采集标签的学名后引证作者。

ver. 2.00 (2019.4.28)

全部代码重构，打包为文档类，重新命名为 `nanicolle`。
新增地图自定义与自动选择功能。
暂时删去英文采集标签排版功能。
暂时删去标本条码标签排版功能。
暂时删去植物所建筑地图。
开始发布于 CTAN 和 Github 上。

ver. 1.14 (2018.12.2)

新增 `\ifinternal` 开关，以快速控制条码、照片号、标题等。

ver. 1.13 (2018.9.15)

改进采集标签中下划线表的实现。
修正地图中的国界。
采集标签中增加植物所建筑地图。
缩减采集标签中的悬挂缩进量。
代码模块化。

ver. 1.12 (2018.5.15)

优化拉丁学名断行位置。
调整标本条码标签的段落样式。

ver. 1.11 (2018.3.28)

调整原始数据结构。
鉴定标签中增加命名标准缩写。

ver. 1.10 (2018.2.9)

采集标签中增加东莞市镇街地图。

ver. 1.09 (2018.1.28)

实现标本条码标签排版。

ver. 1.08 (2017.10.31)

采集标签中增加地理范围为经度 $73^{\circ}\text{--}136^{\circ}\text{E}$ 、纬度 $17^{\circ}\text{--}54^{\circ}\text{N}$ 的国界与中国省界地图。

ver. 1.07 (2017.10.22)

采集标签与鉴定标签中增加条码。
实现带有尼泊尔分区地图的英文采集标签排版，用于 2017 年中国—尼泊尔联合植物考察。

ver. 1.06 (2017.10.21)

采集标签中增加标题。

ver. 1.05 (2017.7.4)

调整页面布局，使标签尺寸缩小为原来的 $\sqrt{2}/2$ ，与 A4 台纸相配。
改进水平切割线的样式。

ver. 1.03 (2017.7.2)

参数的表示由 \LaTeX 格式改成 TSV 格式，方便用电子表格储存。
此版本曾发布于 Github 上。

ver. 1.02 (2016.8.6)

新造“菰”字。

ver. 1.01 (2016.8.3)

实现采集标签与鉴定标签排版。