



# 11 扩增子软件安装

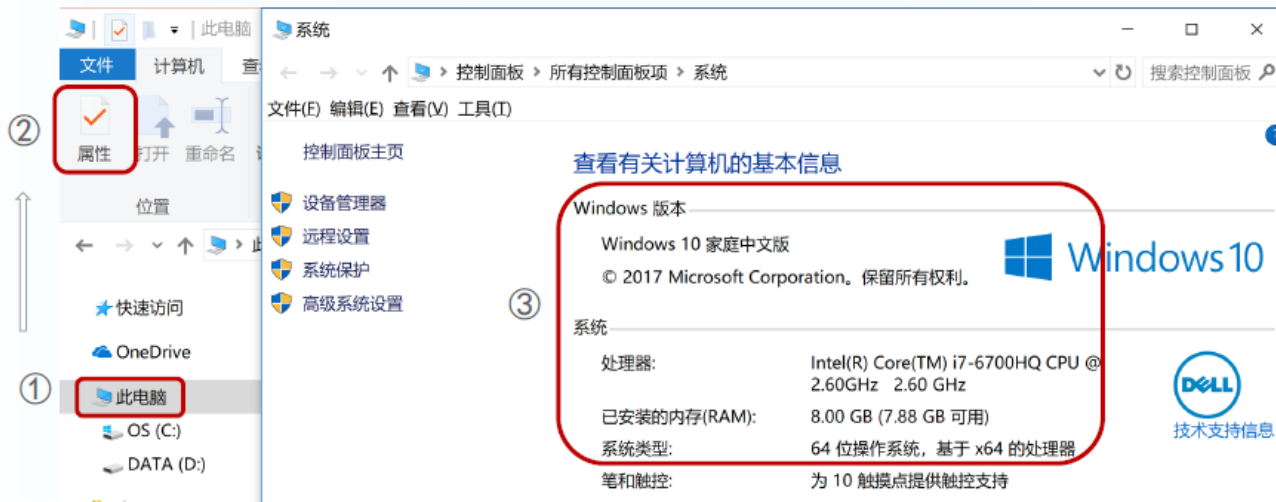
易生信  
2023年2月3日



# 上课自带笔记本要求



- **操作系统：**推荐Windows10+ 64位版(主流兼容性好)，不推荐32位系统、Mac
- **最低配置：**CPU双核、内存4G、硬盘10GB空间
- **推荐配置：**CPU四核+、内存8G+、硬盘30GB+空间



Windows 10用户点击“资源管理器”中的“此电脑”，再点击--“属性”即可显示电脑基本信息



# 常用软件列表



1. 数据分析环境Shell + R + IDE: GitForWindows、R+RStudio+R包
2. 扩增子分析流程: EasyAmplicon, EasyMicrobiome,  
EasyAmplicon: 测试数据、分析流程、示例结果(正对照)  
EasyMicrobiome: 分析依赖软件、数据库等, 如序列工具seqkit、表格工具csvtk
3. 差异分析和可视化: STAMP
4. 网络分析及可视化: Cytoscape、Gephi
5. 图片排版: Adobe Illustrator
6. Win子系统+QIIME 2



# 快速下载易扩增子(EasyAmplicon)



## ○ 安装依赖软件

Windows用户 (占PC中87.5%) 为例, 快速下载[GitForWindows](#), [R](#), [RStudio](#), [STAMP v2.1.3](#) 或 4合1压缩包[zip](#).

## ○ 安装易扩增子 (EasyAmplicon)

EasyAmplicon流程(正对照) <https://github.com/YongxinLiu/EasyAmplicon>

EasyMicrobiome软件和数据库 <https://github.com/YongxinLiu/EasyMicrobiome>

方法 1. 访问上方链接, 点击 Code -- Download

方法2. 采用国内镜像链接下载: [EasyAmplicon](#) [EasyMicrobiome](#)



# 1.1 分析环境：GitForWindows(仅Win用户安装)



- 提供Windows运行Shell命令的环境，可在RStudio的Terminal中使用
- 官网：<http://gitforwindows.org/>，点击Download下载最新版
- 安装包Git-2.39.0.2-64-bit.exe右键以管理员身份运行，按默认参数安装。  
(若不能调用，此电脑-属性-高级-环境变量-系统变量Path中添加C:\Program Files\Git\usr\bin)
- 具体使用见：[Windows轻松实现linux shell环境：gitforwindows](#)



## 1.2 数据分析环境：R语言



- R语言是目前生物学、经济学等领域最流行的统计分析语言
- 官网：<https://www.r-project.org/> 下载最新版：Download CRAN - China Tsinghua - Download R for Windows(Mac) —— base —— Download R 4.x.x
- Win文件夹中有R-4.2.2-win.exe安装程序
- 双击安装程序
- Mac文件夹中有Mac系统安装包R-4.2.2.pkg

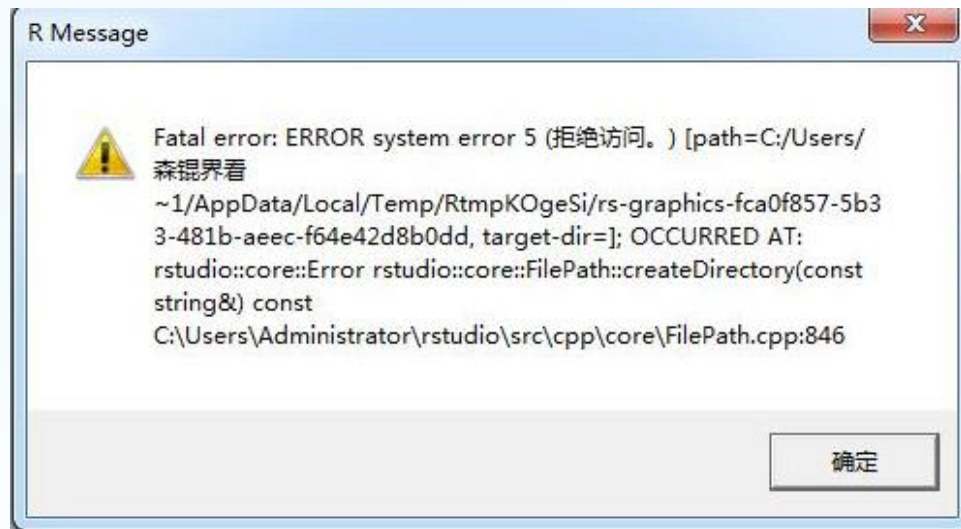


# 常见问题：中文用户名导致乱码及无法使用



- 如果您碰到如下错误，是因为用户名中存在中文，导致乱码不能识别，请新建一个用户，名字为纯英文，重新安装以上工具。

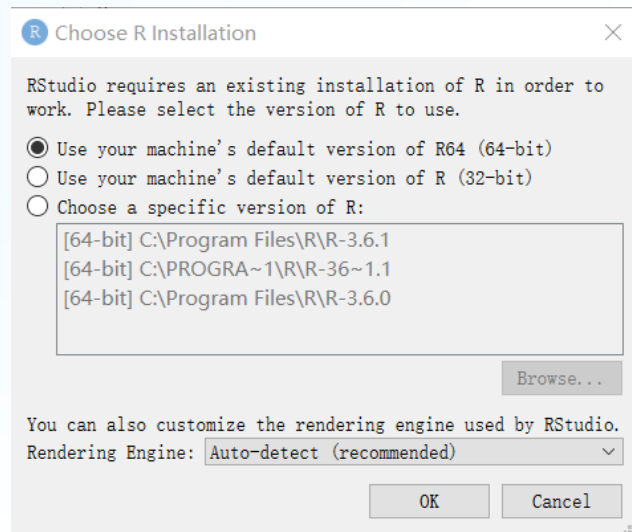
Win10下新建用户操作方法：  
Win10开始 — 设置 — 帐号 —  
家庭和其它人员 — 我没有... —  
添加一个... — 输入用户名和密  
码 — 下一步 — 按提示操作至  
完成



# 1.3 R/Shell编程环境——RStudio

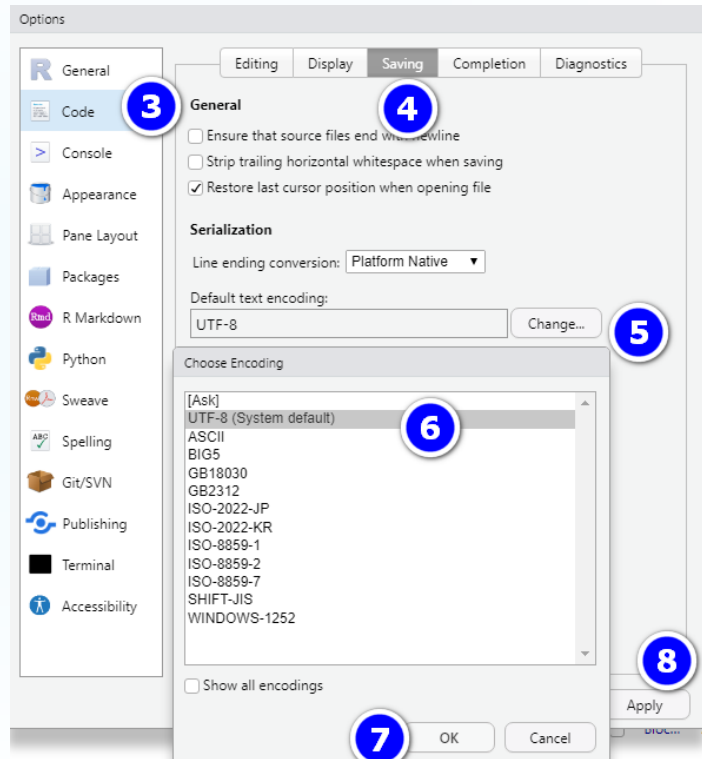
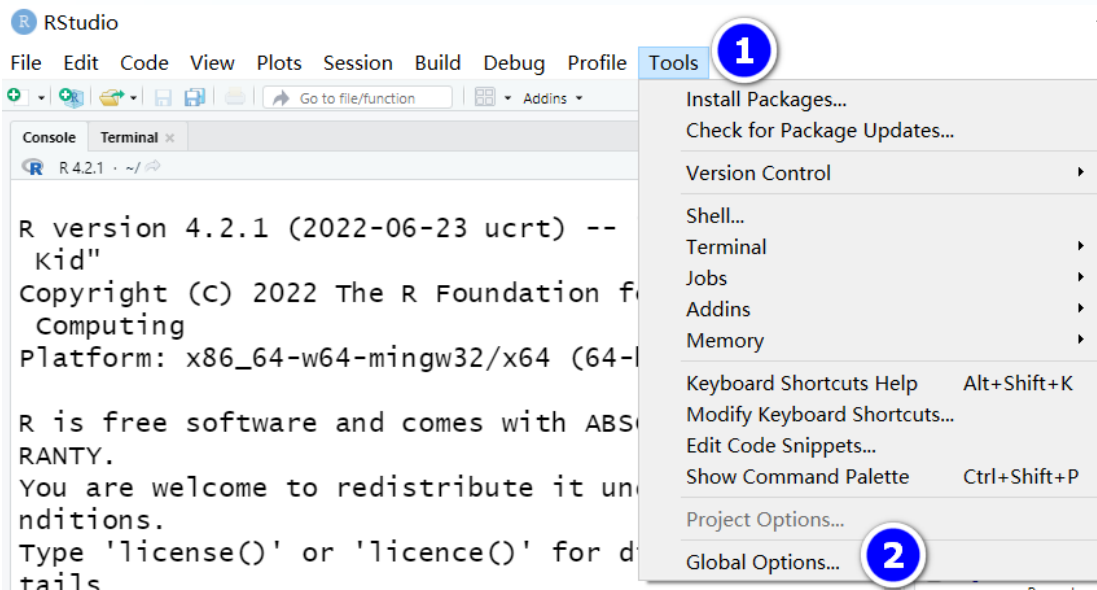


- 下载页面: <https://posit.co/download/rstudio-desktop/>
- 选择适合自己系统的版本(Windows 10+或macOS 10.14+), 下载安装程序的最新版, 或在文件夹使用预下载安装程序Win版RStudio-RStudio-2022.12.0-353.exe或Mac版RStudio-2022.12.0-353.dmg
- **右键使用管理员身份安装**
- 完成后打开时, 如存在R多版本会提示选择  
选第一项 - “使用系统默认R64位版本” - 点击OK  
即使用安装的最新64位版R语言环境





# 常见问题：RStudio中字符乱码处理



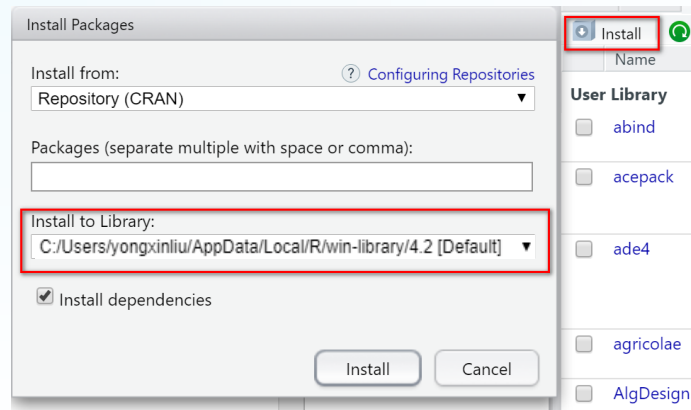
按数字标号顺序切换编码，关闭文件，关闭程序，再打开RStudio，查看文章即可恢复正常。

# 1.4 安装R包——批量解压4.2.zip



- 4.2.zip包含了几百个常用R包，解压即可使用，省去了下载安装过程。  
详见易汉博[R学习教程](#) R包下载链接：[Windows版](#)或[Mac版](#)
- 首先启动RStudio，菜单Tools-Install Packages查看默认R包安装目录
- 将压缩包4.2.zip复制到win-library目录，选中4.2.zip，右键选择解压缩至当前文件夹，如提示文件替换，可选择全部选是。（注意不要在4.2文件夹里再出现4.2文件夹，导致系统找不到包的问题）

Mac用户解压Mac文件夹中的压缩包于  
下载(Downloads)目录，运行如下命令  
`cp -r ~/Downloads/library/*  
/Library/Frameworks/R.framework/Ve  
rsions/4.2/Resources/library/`



# 安装R 4.2包——测试和常用问题解决



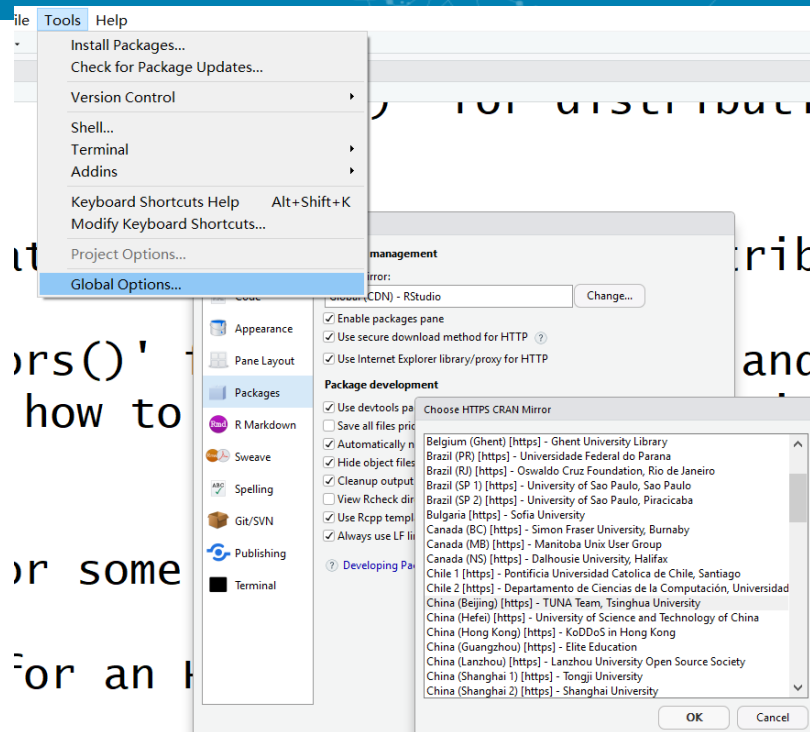
- 替换之后，重新启动RStudio，在左下角“>”后面输入 `library(ggplot2)` 测试，如下图则是安装成功。

```
[Workspace loaded from /var/www/html/ehbio_doc/train/.RData]
> library(ggplot2)
>
```

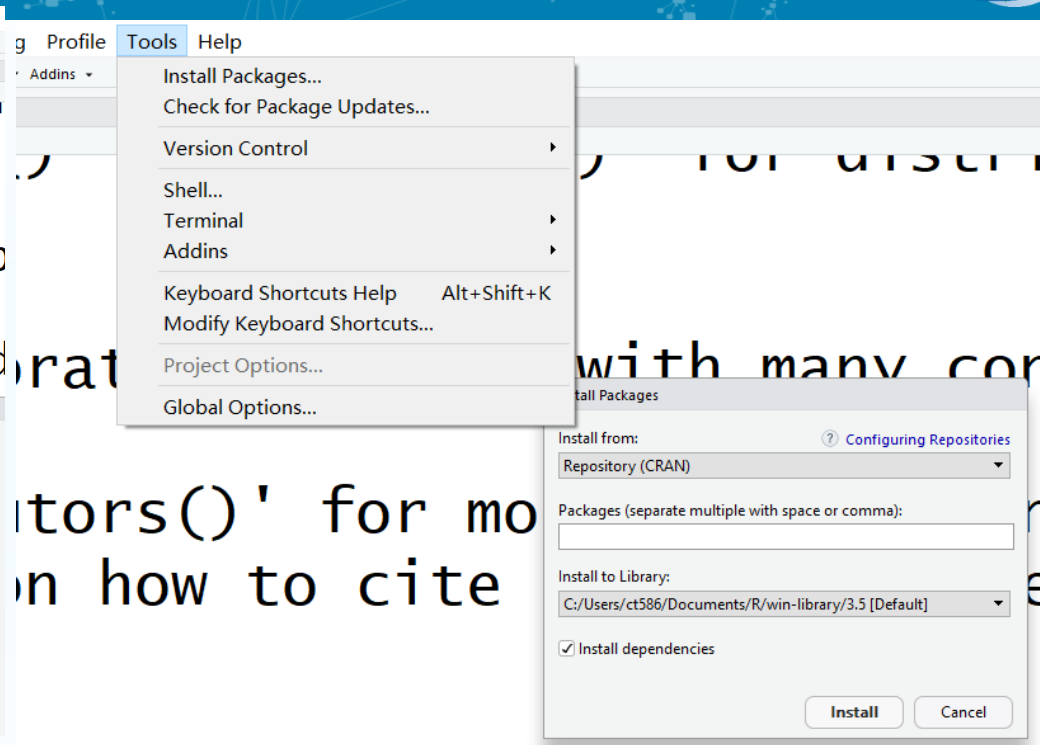
- 如果出现 **Error in library("ggplot2") : 不存在叫 'ggplot2'这个名字的程辑包**，则说明包未安装成功。可能是前面操作解压没有成功，或位置不正确，请仔细核对并重复上面解压安装操作。
- 如果是自己需要用的新包，不存在于易生信提供的压缩包中，则需按后面的方式重头安装。



# 重头菜单安装R包 ---- CRAN官方包



配置距离你最近的CRAN镜像  
方便快速下载(可选)



输入待安装包的名字  
安装CRAN的包



# 重头代码方式安装R包 ---- CRAN / Bioconductor / Github



- # 安装CRAN包 (*斜体部分为包的名字, 可替换*。注: devtools包用于安装github来源R包)
- `install.packages("devtools")`
- # 安装Bioconductor包的新方法 (R版本  $\geq 3.5$ 。注: edgeR是最流行的测序数据差异比较R包 )
- `if (!requireNamespace("BiocManager", quietly = TRUE))`
- `install.packages("BiocManager")`
- `BiocManager::install("edgeR")`
- # 安装Github的R包 (*斜体部分为包的名字, 需要替换*。注: github包名由用户名和包名两部分组成)
- `library(devtools)`
- `devtools::install_github("microbiota/amplicon")`



# 1.5 RStudio中使用Shell命令



- 在RStudio的Tools菜单 — Global Options——Terminal中选择Git Bash, 点OK
- 用 Rstudio 打开 12Linux 目录中 shell.sh文件, 光标处理代码行或选中要执行的代码, 点运行按钮 (Run)/按Ctrl+Enter执行下面代码

# 进入C盘

cd /c

# 创建一个目录

mkdir -p amplicon

# 查看文件列表

ls

```
1 cd /c/2amplicon
2 ls
3 cat <<END >design.txt
4 samp1 ko
5 samp2 ko
6 samp3 wt
7 samp4 wt
8 END
9 cut -f 1 design.txt
```

1. 确认文档类型为Shell  
2. 定位到需要运行的代码, 点 Run  
3. 代码运行结果在左下角输出  
4. 左下角直接输入命令再回车也可执行对应命令

```
ehbio@LAPTOP-FMB2A343 /c/2amplicon
$ ls
doc pipeline_reference_linux.sh script temp
gg pipeline_reference_win.sh seq

ehbio@LAPTOP-FMB2A343 /c/2amplicon
$ cat <<END >design.txt
> samp1 ko
> samp2 ko
> samp3 wt
> samp4 wt
> END

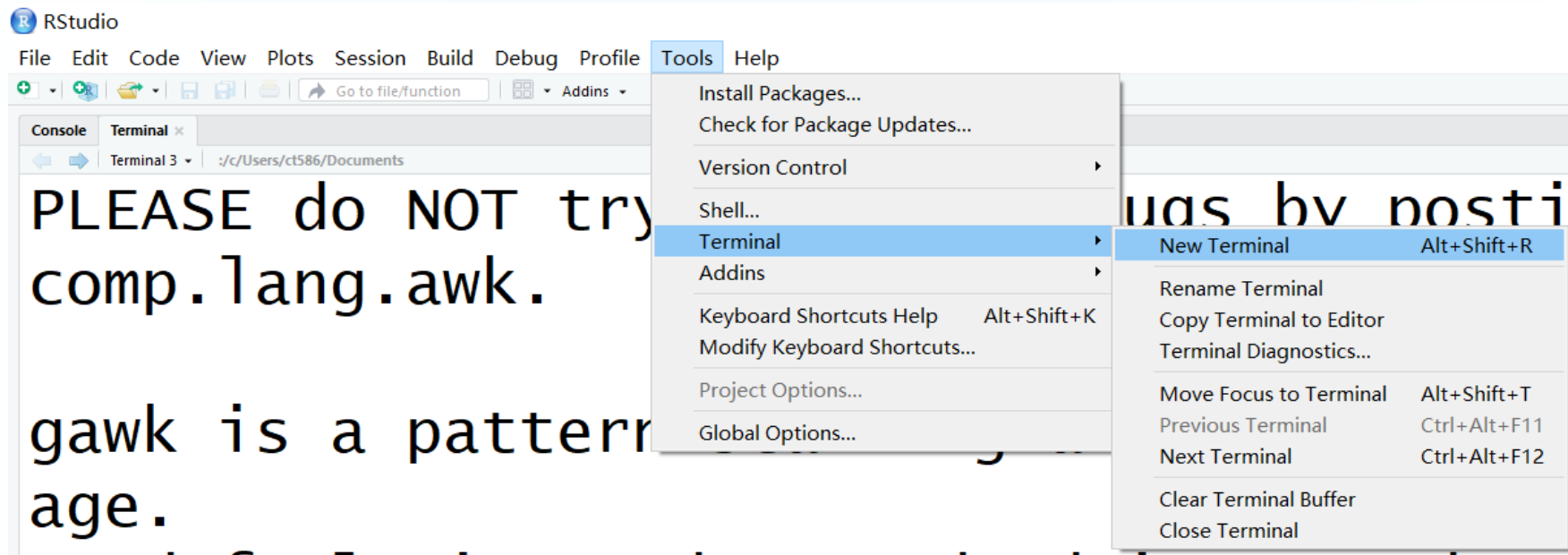
ehbio@LAPTOP-FMB2A343 /c/2amplicon
$ cut -f 1 design.txt
samp1 ko
samp2 ko
samp3 wt
samp4 wt
```



# 常见问题：RStudio中调出Terminal



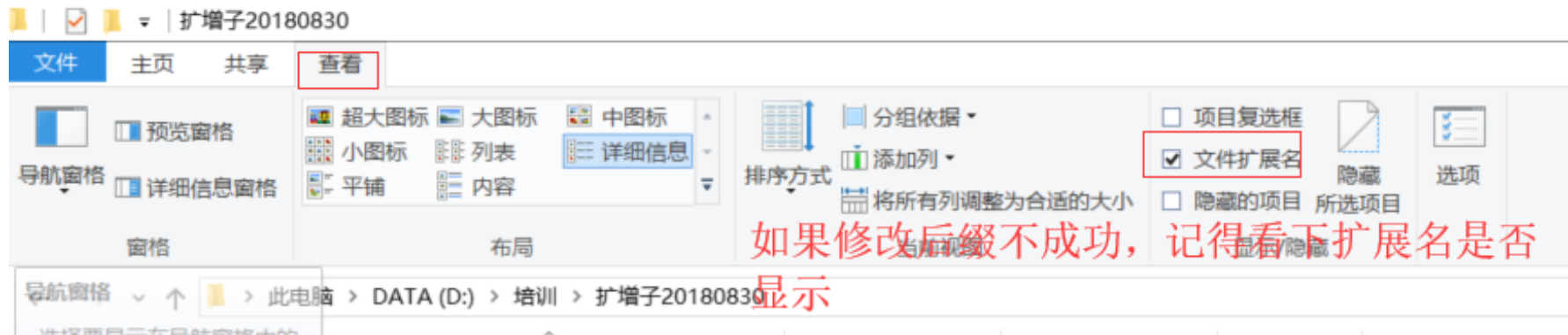
- 若未看到Terminal或不小心关掉了Terminal，可按下图所示打开。
- Tools —— Terminal —— New Terminal (快捷键Alt+Shift+R)



# 常见问题：Win下不显示文件扩展名问题



- 人们常用文件扩展名决定文件类型，如程序一般为.exe
- Linux Shell脚本为.sh，R语言的脚本为.R，R Markdown为.Rmd
- 只有扩展名正确，RStudio才能正确选择合适的环境运行
- Windows中**查看** - 勾选**文件扩展名**，方便修改正确识别代码文件





# 2.1 扩增子分析流程USEARCH / VSEARCH下载 (整合至EasyMicrobiome , 可跳过)



## ○ USEARCH下载

访问 <http://www.drive5.com/usearch/download.html> , 选择USEARCH 10版本下载:

Windows: usearch10.0.240\_win32.exe.gz

Mac OS: usearch10.0.240\_i86osx32.gz

下载后解压, 并改名为usearch.exe。如windows用户保存于C:/EasyMicrobiome/win/目录

## ○ VSEARCH下载

<https://github.com/torognes/vsearch> 主页中右侧release找最新下载链接, 如 vsearch-2.15.2-win-x86\_64.zip

下载后解压其中的 vsearch.exe 至C:/EasyMicrobiome/win/目录



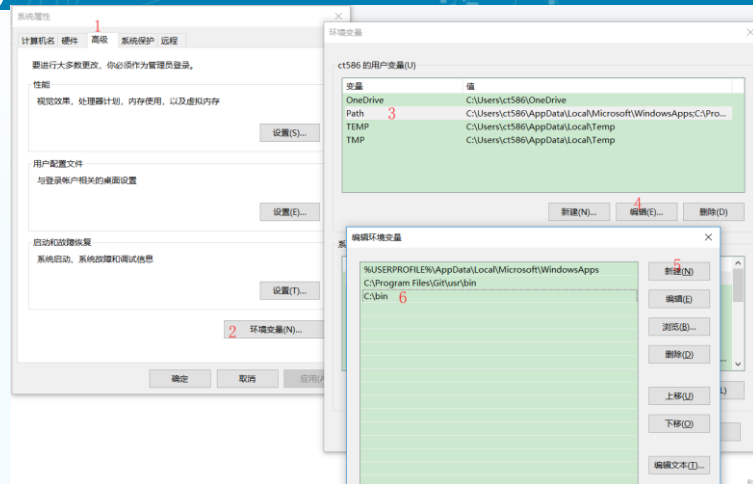
# 扩增子分析流程USEARCH / VSEARCH安装(整合至EasyMicrobiome, 可跳过)



- 将U盘/下载amplicon目录复制到C盘。
- 移动public目标至C盘，添加程序位置至环境变量：我的电脑 - 右键属性 - 按右侧截图操作，新建C:\public\win。
- 测试是否安装成功：

在RStudio的Terminal下输入 usearch 回车，如有，则表明安装成功。

若不成功，可检查环境变量配置，或双击 usearch.exe 查看是否有报错信息。



```
$ usearch
usearch v10.0.240_win32, 2.0Gb RAM (8.5Gb total), 12 cores
(C) Copyright 2013-17 Robert C. Edgar, all rights reserved.
http://drive5.com/usearch
```

License: personal use, non-transferrable

```
$ vsearch
vsearch v2.15.2_win_x86_64, 7.9GB RAM, 12 cores
https://github.com/torognes/vsearch
```

For help, please enter: C:\public\win\vsearch.exe --help | less  
For further details, please consult the manual by entering: man vsearch



## 2.2 扩增子分析流程QIIME 2安装



- Windows 10系统需要安装应用商店中的Ubuntu 20.04程序，请参阅 [Windows10安装Linux子系统Ubuntu 20.04LTS](#)，轻松使用生信软件，效率秒杀虚拟机
- 本质上QIIME不支持Windows，我们在Windows中安装了Ubuntu系统中运行QIIME 2，可直接在Ubuntu程序界面中使用，也可以RStudio的Terminal中选择Bash (Windows Subsystem for Linux)来使用
- 然后在Ubuntu/Mac环境中安装Conda，下载QIIME 2最新版软件列表，并用conda安装QIIME 2，可参考 [QIIME 2简介和安装](#)
- 如果在线安装失败，还可选下载安装包本地安装，代码见附录12。
- 详细使用方案另可参考：[MPB：使用QIIME 2分析微生物组16S rRNA基因扩增子测序数据\(视频\)](#)



## 2.2 扩增子分析流程QIIME 2本地安装



- 需上文配置好Windows10的Linux子系统

- 下载、安装和启动conda并安装conda

```
wget -c https://repo.anaconda.com/miniconda/Miniconda3-latest-Linux-x86_64.sh
```

```
bash Miniconda3-latest-Linux-x86_64.sh -b -f
```

```
~/miniconda3/condabin/conda init
```

- 将安装包 public 存放于 C 盘 (/mnt/c/EasyMicrobiome/linux/qiime2-2022.11.tar.gz)

```
mkdir -p ~/miniconda3/envs/qiime2-2022.11
```

```
tar -xzf /mnt/c/db/conda/qiime2-2022.11.tar.gz -C ~/miniconda3/envs/qiime2-2022.11
```

```
conda activate qiime2-2022.11
```

# 3.1 辅助跨平台工具(整合至EasyMicrobiome)



- Windows下命令行下载工具 wget

<https://eternallybored.org/misc/wget/>

- 沈伟 <https://github.com/shenwei356>

支持Windows / Linux / MacOS的32/64位系统

SeqKit 0.15.0: 序列处理

- <https://github.com/shenwei356/seqkit>

csvtk v0.22.0: 表格处理

- <https://github.com/shenwei356/csvtk>

TaxonKit v0.7.2: NCBI物种信息查询和整理

- <https://github.com/shenwei356/taxonkit>

Rush v0.4.2: 任务并行管理软件

- <https://github.com/shenwei356/rush>



## 3.2 EditPlus纯文本编辑器



- <https://www.editplus.com/>
- 体积小巧，启动快，功能齐全
- 可远程打开、保存
- 支持正则表达式匹配和替换
- 支持显示空格、制表符
- 官网下载5.0最新版，或安装public/win/epp510\_1828\_64bit.exe
- 注册码：Vovan      3AG46-JJ48E-CEACC-8E6EW-ECUAW



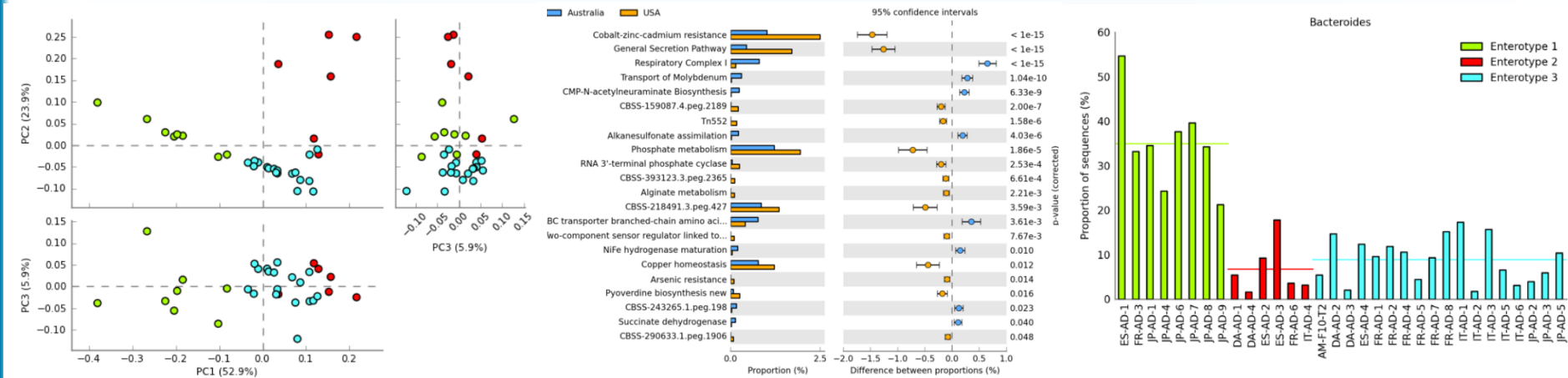
## 4. 微生物组差异比较STAMP



- STAMP是一款分析微生物分类和功能谱的软件，最新版本2.1.3，Downloads部分可下载适合自己Windows/Linux/MacOS版本的软件。
- Examples处提供了示例分析结果，以及演示数据实例。
- STAMP可以现实不同平台下兼容，实现Beta多样性散点图、物种丰度柱状图、箱线图，以及Post-hoc图展示差异菌。还可以绘制带误差线柱状图、误差线和柱分离组合图、相关散点图、密度柱状图、p值柱状图等分析和绘图。
- <http://kiwi.cs.dal.ca/Software/STAMP> 下载系统对应版本
- public\win目录中有windows版安装包 STAMP\_2\_1\_3.exe



# STAMP结果展示



微生物组间差异分析神器-STAMP [简明教程](#) [中文帮助文档](#)



## 5. 网络分析和可视化Cytoscape



- Cytoscape是一款图形化显示网络软件，生物学中常用于分析转录因子与基因或蛋白与蛋白之间互作关系、GO和KEGG富集分析
- 软件下载地址：<http://www.cytoscape.org>
- 没安装过Java运行环境的用户，先安装jdk-11.0.7\_windows-x64\_bin.exe
- 在文件夹下或点击链接下载安装程序  
Cytoscape\_3\_8\_2\_windows\_64bit.exe
- 按默认参数完成安装即可
- Cytoscape使用视频教程：<https://ke.qq.com/course/261290>



## 6. 图片美化和排版Adobe Illustrator



- 简称"AI", 是一款非常好的矢量图形处理工具、图片排版工具
- 是Adobe公司开发的一款收费软件, 在此为大家提供一个破解免安装版的包仅供学习使用
- 在文件夹下或点击链接下载安装包  
Adobe\_Illustrator\_CC\_2018\_v22.1.0.314\_x64\_zh\_CN\_Portable.7z ,  
解压后运行Illustrator.exe程序即可使用(打不开尝试管理员运行或  
App\Illustrator\Support Files\Contents\Windows\Illustrator.exe)
- 使用视频教程: <https://ke.qq.com/course/261607>
- 安装前可体验下网页版: [https://mp.weixin.qq.com/s/0IfurkWazDj\\_gXVltSly6w](https://mp.weixin.qq.com/s/0IfurkWazDj_gXVltSly6w)



# 常见问题：Adobe Illustrator无法打开



- 当双击IllustratorPortable.exe无法打开或者遇到报错情况如下图时  
删掉软件包中的Data文件夹，再双击IllustratorPortable.exe即可打开  
(Data文件夹是每次运行AI时自动产生的文件，每次打开前删除即可)



App	2018/8/5 9:34	文件夹
Data	2018/8/7 9:47	文件夹
IllustratorPortable.exe	2015/7/27 4:14	应用程序
IllustratorPortable.ini	2015/7/28 10:08	配置设置
说明.txt	2015/7/27 3:42	文本文档
未来软件园_百度搜索	2014/10/30 11:31	Internet 快捷方式
未来软件园-首页	2006/7/20 22:26	Internet 快捷方式
下载安装说明.txt	2015/3/18 16:58	文本文档





- R包编译环境Rtools
- (可选)浏览器：谷歌Chrome/微软Edge
- 登录服务器：XShell / PuTTY;
- 上传下载文件：Filezilla / WinSCP



# R包源码安装编译软件Rtools



- 访问 <https://mirrors.tuna.tsinghua.edu.cn/CRAN/> -  
Download R for Windows – [Rtools](#) - [RTools 4.2](#) -  
[Rtools42 installer](#)
- 下载安装包约460 MB，双击完成安装，即可支持源码安装R包



## 7. 谷歌Chrome浏览器(兼容性好/可选Win10 Edge)



- 谷歌 Chrome 浏览器——远程访问 RStudio server或其它网页工具兼容性最好的浏览器
- 网址 <https://www.google.cn/chrome/> 在线安装适合你操作系统的最新版(Google网站访问可能需要科学上网，**可选使用360或电脑管家快速安装最新版**)
- 按默认参数完成安装即可
- Windows 10系统最新版的Microsoft Edge也采用Chrome内核，兼容性较接近Chrome



## 8. 服务器通讯——Xshell 和 FileZilla



- Xshell是终端，用于登录服务器。填写姓名、邮箱获取免费版下载链接 <https://www.netsarang.com/zh/free-for-home-school/>，Xshell无法正常运行用户备选PuTTY <http://www.putty.be/> putty.exe
- FileZilla可以断点续传进行上传、下载，上传数据至NCBI。下载链接 <https://filezilla-project.org/download.php?type=client> 无法正常运行此软件备选WinSCP <https://winscp.net/eng/download.php> 或 Xftp
- 按默认参数完成安装即可，Xshell 需要按图示选择许可类型才可免费使用(界面可能为英文)。

请选择一个许可类型。如果你是注册用户或评估Xftp5，请选择商业。

☐ 商业（包括评估）

☒ 免费为家庭/学校



# 8.1 Xshell安装和使用



- Xshell是Windows下功能强大的终端。Linux和Mac可以直接使用ssh命令访问服务器
- 最新版下载地址: <https://www.netsarang.com/zh/free-for-home-school/>
- 安装: 注意选择Home/School类型可免费使用
- 运行: 首次需要配置服务器信息, 文件 - 新建 - 名称YSX - 主机IP - 确定 - 输入用户名并勾选记住 - 输入密码并勾选记录 - 确定

- RStudio中的Terminal也可以访问服务器
- ssh [username@serverIP](#)

如 ssh amplicon@192.168.1.130

```
bailab - Xshell 6 (Free for Home/School)

1 bailab x +
https://microk8s.io/ has doc

0 packages can be updated.
0 updates are security updates.

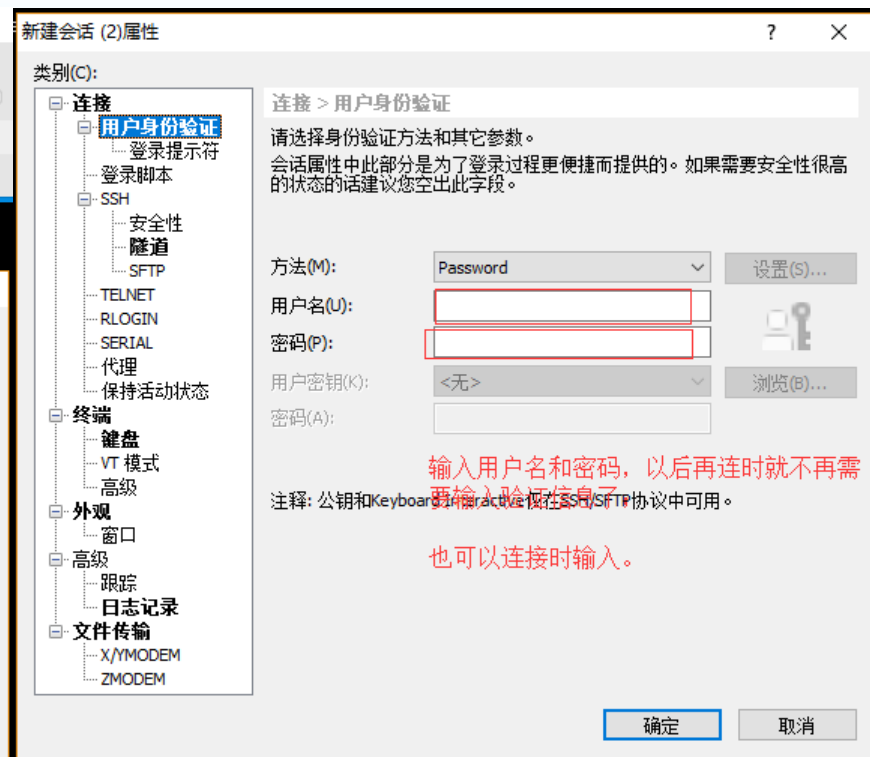
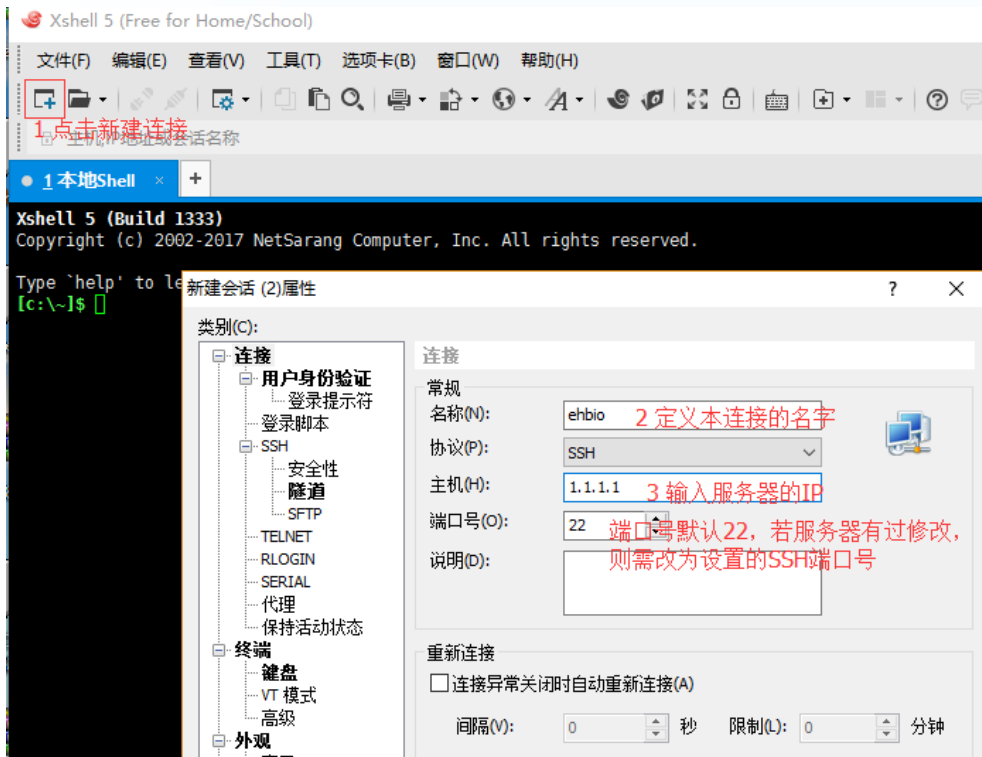
New release '18.04.5 LTS' avail
Run 'do-release-upgrade' to upg

*** System restart required ***
Last login: Wed Aug 26 16:55:12
(base) [yongxin@biocloud:~]$
```

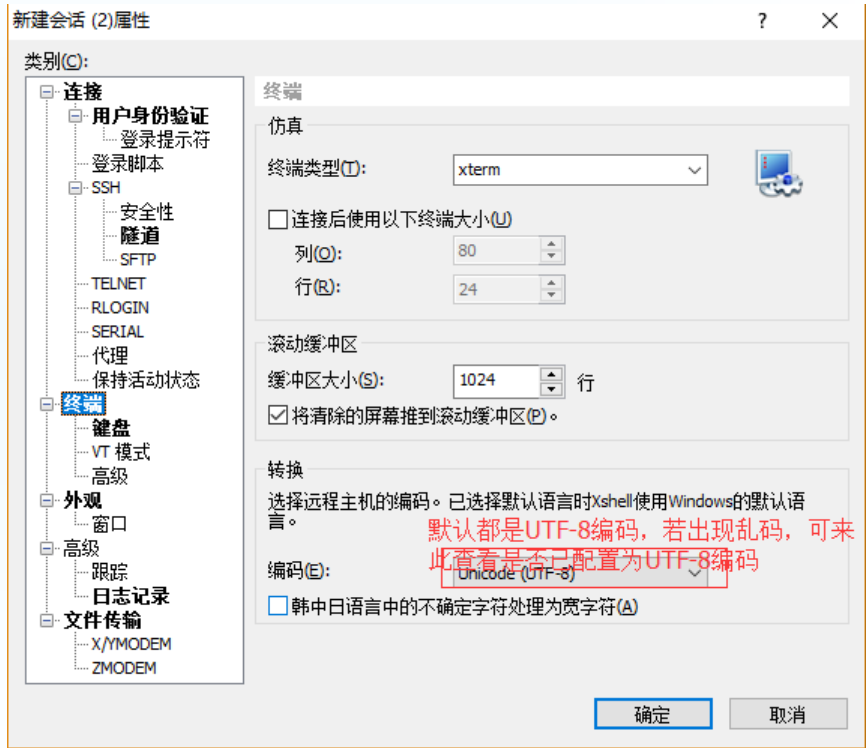
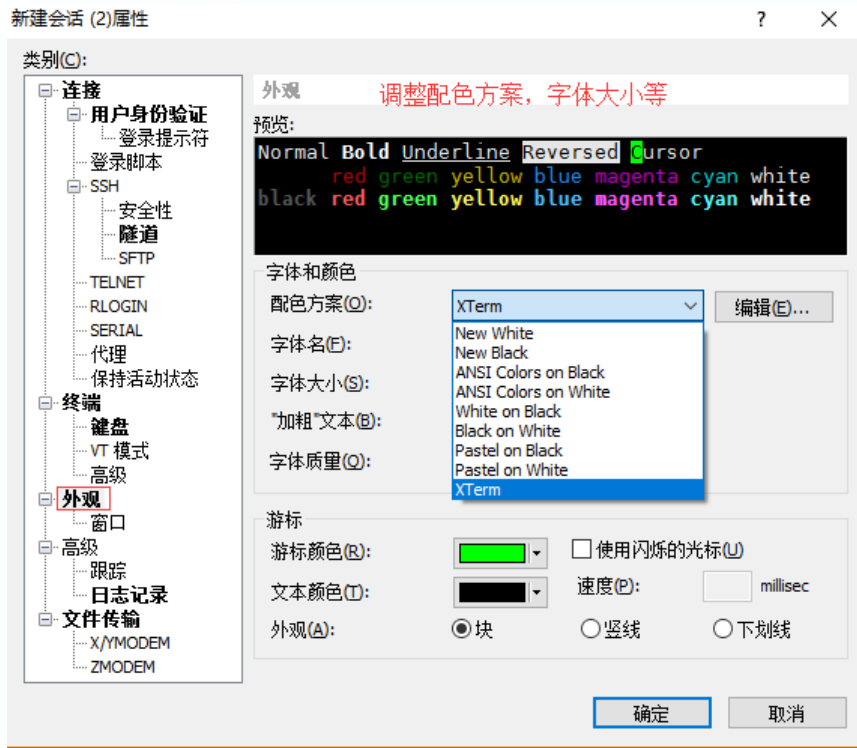




# Xshell使用



# Xshell使用



## 8.2 FileZilla的安装和使用



- 官网: <https://filezilla-project.org/> Download FileZilla Client 下载适合自己系统的最新版安装包

- 配置 Configure

打开软件-左上角图标 ‘打开站点管理器’ - 新站点 - 命名为 YSX - 主机 填写 192.168.1.130 (IP地址, 填自己服务器或上课临时IP), 端口 22; 协议 SFTP, 登陆类型 普通, 用户名为姓名全拼, 密码默认为ysx\_amplicon。首次连接选择总是信任, 确定。

# FileZilla上传下载使用



主机(H):  用户名(U):  密码(W):  端口(P):

上传下载前的登录设置，注意sftp是必须的，端口号根据需要修改

本地站点: D:\

- 桌面
- 文档
- 此电脑
- C: (Windows)
- D: (DATA)
- E: (RECOVERY)

文件名	文件大小
..	
11940958FOCA	4
VitaminD switches BAF complexes to protect b...	6,827,
业务宣传片.wmv	19,866
基因组分析测序.docx	166,33
FSCapture.rar	4,264,
poster.pdf	14,618
EHBIO	

设置 隶属于“编辑”菜单栏

选择页面(P): 请选择目标文件存在时的默认操作。

文件存在时的默认操作

下载(D):

上传(U):

要是选择“如果较新就覆盖”，您的系统时间需要与服务器同步。若时间不同 (例如不同的时区)，请在站点管理器中选择一个时区。

☐ 允许续传 ASCII 文件(L)  
如果服务器使用不同格式的行结尾，继续传输 ASCII 格式文件可能导致问题。

- 连接
  - FTP
    - 主动模式
    - 被动模式
    - FTP 代理服务器
  - SFTP
  - 通用代理
- 传输
  - 文件类型
  - 对已存在文件的操作
- 界面
  - 主题
  - 日期/时间格式
  - 文件大小格式
  - 文件列表
- 语言
- 文件编辑

## 原始数据极速上传NCBI SRA教程



扫码关注生信宝典，学习更多生信知识



扫码关注宏基因组，获取专业学习资料

# 易生信，没有难学的生信知识



易汉博基因科技(北京)有限公司  
EHBIO Gene Technology (Beijing) co., LTD