

PanGu manual

V 1.2

create by Qianshu Zhu

2019-9-26

System release: CentOS Linux release 7.4.1708

CPU: Intel(R) Xeon(R) Gold 6148 CPU @ 2.40GHz -- total 160 processors

Disk: /dev/sdb(home)--46TB, /dev/mapper/mpatha(home1)--59TB

MemTotal: 1056 GB

--

pangu系统中的所有软件请尽量交由管理员安装，包括python和R的packages等

本文档不定期完善更新

因PATH中加入的路径越来越多，某些不常用的工具将可能会出现执行文件重名的情况（目前没有发现），请一定主要用type或which查看调用的是不是预期的软件，并在/etc/profile.d/PATH中查看PATH的顺序

已安装的软件：

软件	调用命令	版本号	备注
bamtools	bamtools	2.5.1	依赖cmake
bat	bat	0.11	可替代cat，以语法格式高亮显示文件
bcftools	bcftools	1.9	隶属htslib
bedtools	bedtools	2.28.0	toolset for genome arithmetic
bsmap	bsmap	2.90	已停止更新
bsmapz	bsmapz	1.1.0	bsmap适应新版samtools的分支版本
BEDOPS	bedops; hotspot	2.4.36	因bedops依赖的库较多，安装了一个miniconda来装bedops;使用前先使用conda-init激活环境；其存放路径：/opt/miniconda2/bin/bedops
blast	blastn	2.7.1	NCBI blast tools
BLAT	blat	36x2	fast sequence search command line tool
bismark	bismark..	0.22.1	index:/home/share/bismark_index
bowtie	bowtie	1.2.3	已安装多个版本，index已配置在全局变量BOWTIE_INDEXES=/home/share/bowtie_index
bowtie2	bowtie2	2.3.5.1	已安装多个版本，index已配置在全局变量BOWTIE2_INDEXES=/home/share/bowtie2_index
bwa	bwa	0.7.17	index: /home/share/bwa_index
cufflinks	cufflinks; cuffdiff..	2.2.1	软件下包含多个工具，可通过which查看
cutadapt	cutadapt	2.5	python3版本，可多线程使用；pip3安装
deepTools	deeptools..	3.3.0	多工具集；pip安装

datamash	datamash	1.4	shell下的表格处理脚本
GeneTrack	python run.py	1.0.3	/usr/local/software/GeneTrack-1.0.3，使用方法较复杂
GATK	gatk	4.1.3.0	多工具集；注释：/home/share/gatk_resource
GSEA	gsea	3.0	java图形界面运行
htslib	htslib	1.9	隶属htslib
HiC-Pro	HiC-Pro	2.11.1	HiC比对pipeline，注释文件：\$ConfigHP
HiCExplorer	hicexplorer	2.2.1.1	python3版本；pip3安装
hisat2	hisat2	2.1.0	推荐使用的index：/home/share/hisat2_index/new_index
homer	homer	--	一套多软件工具集
IGV	igv	2.6.3	高dpi版：igv_hidpi；包含IGVtools：igvtools
linuxnd	linuxnd	--	下载诺禾数据的工具
picard	picard	2.20.7	多种工具集
R	R; Rscript	3.6.0	默认使用的R，yum安装；3.5.3已安装
rstudio-server	pangu:8787	1.2.1335	通过浏览器打开10.11.41.109:8787使用；默认使用R-3.6.0，若需要其他版本可另开启一个端口；
STAR	STAR	2.7.2b	--
MACS	macs14	1.4.3	python2版本，pip安装
MACS2	macs2	2.1.2	python2版本，pip安装
moabs	moabs	1.3.3	--
meme	meme	5.0.5	motif相关工具
ChromHMM	ChromHMM	1.19	java -jar ChromHMM.jar
stringtie	stringtie	2.0.3	--
fastp	fastp	0.12.0	cutadapt和fastq的多线程改进版本
3DChromatin_ReplicateQC	--	--	四种HiC的QC工具包；包括：HiCRep,HiC-Spector,QuASAR-Rep,genomedisco
Domaincaller	--	--	matlab、perl脚本，需要分多步骤执行，/usr/local/software/Domaincaller
samtools	samtools	1.9	隶属htslib
sambamba	sambamba	0.7.0	依赖libc，高效的sam bam处理工具
python2	python	2.7.16	默认使用的python，调用命令python或python2，附带pip
python3	python3	3.7.4	调用命令python3，附pip3
subread	featureCounts..	2.0.0	多工具集，常用featureCounts
matlab	matlab	R2019a	因版权问题，只能一个用户使用
Juicebox	Juicebox	1.11.08	类似igv的HiC浏览器
trash-cli	trash..	0.17.1.14	回收站机制，可替代rm防止误删除；以trash开头的多个命令；pip安装
fithic	fithic	2.0.7	python3版本；pip3安装

translation	trans; etc; cte..	0.9.6.11	一个调用网络词典的翻译脚本；i.e. etc:英译中；cte:中译英
motd	motd	1.0	motto of the day; 一个随机获得随机motto的脚本，login时自动调用
wechat	wechat; wait-m; wait-n	1.0	一个给微信发提醒消息的脚本；需要先加入本实验室的企业微信；i.e. 在后台运行程序后运行wait-m，可在程序运行结束后及时收到提醒
xsel	xsel	1.2.0	shell下的剪切板功能；配合多种脚本使用；i.e. cwd;

安装路径：/usr/local/software
 部分单一的程序脚本在/usr/local/bin中
 部分软件链接在/usr/local/LINKS，其他添加在PATH中
 软件安装根据大家需求不定期更新
 默认使用最新版本，需要老版本软件的可通过绝对路径访问
 indexes、codes和annotation都放在/home/share中
 /home/share/hisat2_index是原有的index；/home/share/hisat2_index/new_index是自己制作的index

Babraham projects:

(Babraham所有的工具包，路径/usr/local/software/Babraham-Projects，某些不常用的自行查看文档使用)

软件	调用命令	版本号	备注
ASAP	--	0.1.2	--
bareback	--	1.0	--
bismark	bismark	0.22.1	--
fastq_screen	fastq_screen	0.14.0	--
hicup	hicup	0.7.2	--
seqmonk	seqmonk	1.45.4	--
sierra	sierra	0.3	--
SNPsplit	SNPsplit	0.3.2	--
fastqc	fastqc	0.11.8	--
Sherman	Sherman	0.1.8	--
redotable	redotable	1.1	--
chipmonk	--	1.2.3	<i>java -classpath . uk.ac.bbsrc.babraham.ChipMonk.ChipMonkApplication</i>
focalpoint	--	0.6	--
fretsaw	--	1.0	--
mzViewer	--	0.4	--
sparkspotter	--	1.1	--
stackmeasure	--	1.0	--

UCSC-tools

http://hgdownload.cse.ucsc.edu/admin/exe/linux.x86_64/
 路径：/usr/local/software/UCSC-tools，已加到环境变量PATH中
 工具较多，主要有：bigWigToBedGraph, faToFastq, fetchChromSizes等等

已安装的R library

BiocManager
devtools

CRAN packages:

```
("magrittr","ggplot2","data.table","reshape2","pheatmap","stringr","RColorBrewer")  
("ggrepel","gplots","stringi","stringr","Rcpp","rlang","tidyr","dplyr","ggsignif","ggforce")
```

bioconductor:

```
("DESeq2","clusterProfiler","GSEABase","edgeR","fgsea","biomaRt","Seurat","GEOquery","ggtree","GSEABase","limma","enrichplot")  
hicrep, FithiC  
BiocManager::install("multiHiCcompare")
```

github:

```
devtools::install_github("kassambara/easyGgplot2")  
devtools::install_github("3DGenomes/binless",subdir="binless") #依赖gsl-devel
```

已安装的python packages（包括pip和源码安装）：

- bx-python #for HiC-Pro
- trash-cli
- deepTools
- MACS2
- MACS
- scipy
- biopython

python3下安装：

- fithic
- cutadapt
- HiCExplorer
- multiqc
-

yum安装的软件

- vim-X11
- xsel
- jq
- R-3.6.0
- rstudio-server-rhel-1.2.1335
- iptables-service
- ccache
- docker
- gsl-devel

某些重要文件的存放路径

- 所有软件、packages和library的安装，以及系统配置记录：/root/git/LinuxSet
- 软件安装目录：/usr/local/software
- 软件软连接目录：/usr/local/LINKS

- 环境路径PATH存放位置: /etc/profile.d/PATH
- bowtie2 index: /home/share/bowtie2_index
- bowtie index: /home/share/bowtie_index
- bismark index: /home/share/bismark_index
- baw index: /home/share/bwa_index
- hisat2 index: /home/share/hisat2_index/new_index (上一级上hisat2自带的index)
- STAR index: /home/share/STAR_index
- 各课题原始数据: /home/share/DATA/clean_reads
- gene annotation: /home/share/ann
- 从UCSC上下载的annotation: /home/share/ann/useful_annotation/download_FTP
- 共享的脚本代码: /home/share/codes
- 从Illumina下载的annotation等: /home/share/Illumina_iGenome
- 课程以及其他共享资源: /home/share/resource

其他资源和来源情况请查看/home/share下的数据结构和log、README文件