|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | | |  |
| [RepeatMasker v4](https://github.com/shenwei356/seqkit/releases/tag/v2.4.0).1.5 | | |  |
| **应用移植指南** | | |
| **文档版本** | **01** | |
| **发布日期** | **2023-03-24** | |
| 华为网格系统---方案4-032.png | | | | |
|  | 华为技术有限公司 | | 附件1-16K |  |

|  |
| --- |
| 版权所有 © 华为技术有限公司2022。 保留一切权利。  非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。  商标声明  附件3-版权声明页图 和其他华为商标均为华为技术有限公司的商标。  本文档提及的其他所有商标或注册商标，由各自的所有人拥有。  注意  您购买的产品、服务或特性等应受华为公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为公司对本文档内容不做任何明示或默示的声明或保证。  由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。 |

|  |  |
| --- | --- |
| 华为技术有限公司 | |
| 地址： | 深圳市龙岗区坂田华为总部办公楼 邮编：518129 |
| 网址： | <https://www.huawei.com> |
| 客户服务邮箱： | [support@huawei.com](mailto:support@huawei.com) |
| 客户服务电话： | 4008302118 |

目 录

[目 录 ii](#_Toc23438)

[1 介绍 1](#_Toc23187)

[2 环境要求 2](#_Toc8120)

[3 移植规划数据 4](#_Toc4949)

[4 配置编译环境 5](#_Toc4682)

[4.1 Perl 6](#_Toc24376)

[4.2安装RMBlast 6](#_Toc21885)

[4.3安装TRF 7](#_Toc27704)

[4.1安装RepeatMasker 7](#_Toc18773)

[5 获取源码 9](#_Toc25663)

[6 运行和验证 10](#_Toc31280)

[7 修订记录 11](#_Toc19458)

# 介绍

RepeatMasker是重复序列检测的常用工具，通过与参考数据库的相似性比对来准确识别或屏蔽基因组中的重复序列，属于同源预测注释的方式

关于RepeatMasker的更多信息请访问<http://www.repeatmasker.org/>。

一句话描述：重复序列检测

建议的版本

建议使用版本为“RepeatMasker-4.1.5.tar.gz”。

# 环境要求

硬件要求

硬件要求如表2-1所示。

硬件要求

| 名称 | 说明 |
| --- | --- |
| 架构 | ARM |

软件要求

软件要求如表2-2所示。

注意

* 不同HPC应用的依赖软件不同，建议按照如下步骤判断其依赖软件：
* 1、查看其上游社区是否提供安装指导文档；
* 2、搜索网络上是否已经有社区提供相关安装文档或博客；
* 3、尝试安装该软件，根据报错情况，决定安装哪些依赖软件；
* 4、咨询华为工程师是否有相关经验。

软件要求

| 项目 | 版本 | 下载地址 |
| --- | --- | --- |
| RMBlast | V2.13.0 | <ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/2.13.0/ncbi-blast-2.13.0+-src.tar.gz> |
| TRF | V4.10.0 | <https://github.com/Benson-Genomics-Lab/TRF/archive/refs/tags/v4.09.1.tar.gz> |
| RepeatMasker | V4.1.4 | <http://www.repeatmasker.org/RepeatMasker/RepeatMasker-4.1.5.tar.gz> |

操作系统要求

操作系统要求如表2-3所示。

操作系统要求

| 项目 | 版本 | 下载地址 |
| --- | --- | --- |
| Kernel | 4.18.0 | Linux 4.18.0-147.el8.aarch64 |

# 移植规划数据

本章节给出BUSCO软件在移植过程中涉及到的相关软件安装规划路径的用途及详细说明。

移植规划数据

| 序号 | 软件安装规划路径 | 用途 | 说明 |
| --- | --- | --- | --- |
| 1 | /usr/local/RMBlast/ | RMBlast的安装规划路径。 | 这里的安装规划路径只是一个举例说明，建议部署在共享路径中。需要根据实际情况调整，**后续章节凡是遇到安装路径的命令，都以现网实际规划的安装路径为准进行替换，不再单独说明。** |
| 2 | /usr/local/TRF/ | TRF的安装规划路径。 |
| 3 | /usr/local/REPEATMASKER/ | RepeatMasker的安装规划路径。 |

# 配置编译环境

前提条件

使用SFTP工具将各安装包上传至服务器对应目录下。

配置流程

配置流程

| 序号 | 配置项 | 说明 |
| --- | --- | --- |
| 1 | 安装RMBlast | 参考4.2 **安装RMBlast软件**。 |
| 2 | 安装TRF | 参考4.3 **安装TRF软件**。 |
| 3 | 安装RepeatMasker | 参考4.4 **安装RepeatMasker软件**。 |

## **4.1 Perl**

操作步骤

使用SSH工具，登录服务器。

一般Linux会自带perl，查看Perl版本，推荐使用v5.28.0

**perl -version**

## **4.2安装RMBlast**

操作步骤

使用SSH工具，登录服务器。

下载并解压RMBlast安装包。

**mkdir -p /usr/local/RMBlast/**

**cd /usr/local/RMBlast/**

**wget ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/2.13.0/ncbi-blast-2.13.0+-src.tar.gz**

**wget https://www.repeatmasker.org/rmblast/isb-2.13.0+-rmblast.patch.gz**

**tar zxvf ncbi-blast-2.13.0+-src.tar.gz**

**gunzip isb-2.13.0+-rmblast.patch.gz**

*注：请以压缩包实际名称进行解压*

打补丁，打进ncbi。

**cd ncbi-blast-2.13.0+-src**

**patch -p1 < ../isb-2.13.0+-rmblast.patch**

编译安装。

**cd c++**

**./configure --prefix= /usr/local/RMBlast/ncbi-blast-2.13.0+-src CC=clang CXX=clang++ FC=flang**

**make & make install**

添加环境变量。

**export PATH= /usr/local/RMBlast/ncbi-blast-2.13.0+-src/c++/ReleaseMT/bin:$PATH**

**----结束**

## **4.3安装TRF**

操作步骤

使用SSH工具，登录服务器。

下载并解压TRF安装包。

**mkdir -p /usr/local/TRF/**

**cd /usr/local/TRF/**

**wget https://github.com/Benson-Genomics-Lab/TRF/archive/refs/tags/v4.09.1.tar.gz**

**tar -xzvf v4.09.1.tar.gz**

*注：请以压缩包实际名称进行解压*

编译安装。

**../configure --prefix=/usr/local/TRF/TRF-4.09.1/build/ CC=clang CXX=clang++ FC=flang**

**make & make install**

添加环境变量。

**export PATH=/usr/local/TRF/TRF-4.09.1/build/src:$PATH**

**----结束**

## **4.1安装RepeatMasker**

安装步骤详见http://www.repeatmasker.org/RepeatMasker/

操作步骤

使用SSH工具，登录服务器。

下载并解压RepeatMasker安装包。

**mkdir -p /usr/local/REPEATMASKER/**

**cd /usr/local/REPEATMASKER/**

**wget http://www.repeatmasker.org/RepeatMasker/RepeatMasker-4.1.4.tar.gz**

**tar -xzvf RepeatMasker-4.1.4.tar.gz**

*注：请以压缩包实际名称进行解压*

从www.girinst.org下载RepBaseRepeatMaskerEdition-20181026.tar.gz上传，但是需要注册并认证。

**cp RepBaseRepeatMaskerEdition-20181026.tar.gz /usr/local/REPEATMASKER/RepeatMasker-4.1.4**

**cd /usr/local/REPEATMASKER/RepeatMasker-4.1.4**

**gunzip RepBaseRepeatMaskerEdition-20181026.tar.gz**

**tar xvf RepBaseRepeatMaskerEdition-20181026.tar**

编译。

**perl ./configure CC=clang CXX=clang++ FC=flang**

**----结束**

添加环境变量。

**export PATH=/usr/local/REPEATMASKER/RepeatMasker-4.1.4:$PATH**

# 获取源码

操作步骤

下载RepeatMasker安装包“RepeatMasker-4.1.5.tar.gz”。

下载地址：<http://www.repeatmasker.org/RepeatMasker/RepeatMasker-4.1.5.tar.gz>

----结束

# 运行和验证

操作步骤

使用SSH工具，登录服务器。

下载并上传测试文件。

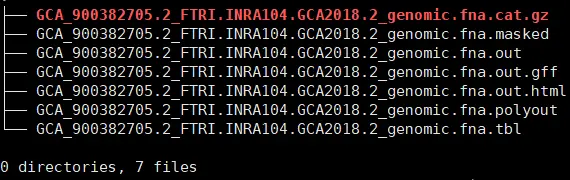
**cd /usr/local/REPEATMASKER/**

**wget https://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900/382/705/GCA\_900382705.2\_FTRI.INRA104.GCA2018.2/GCA\_900382705.2\_FTRI.INRA104.GCA2018.2\_genomic.fna.gz**

**gunzip GCA\_900382705.2\_FTRI.INRA104.GCA2018.2\_genomic.fna.gz**

运行

**repeatMasker -pa 4 -species Fusarium tricinctum -poly -a -dir /usr/local/REPEATMASKER/output/ GCA\_900382705.2\_FTRI.INRA104.GCA2018.2\_genomic.fna**



# 修订记录

| 发布日期 | 修订记录 |
| --- | --- |
| 2023-03-24 | 第一次正式发布。 |