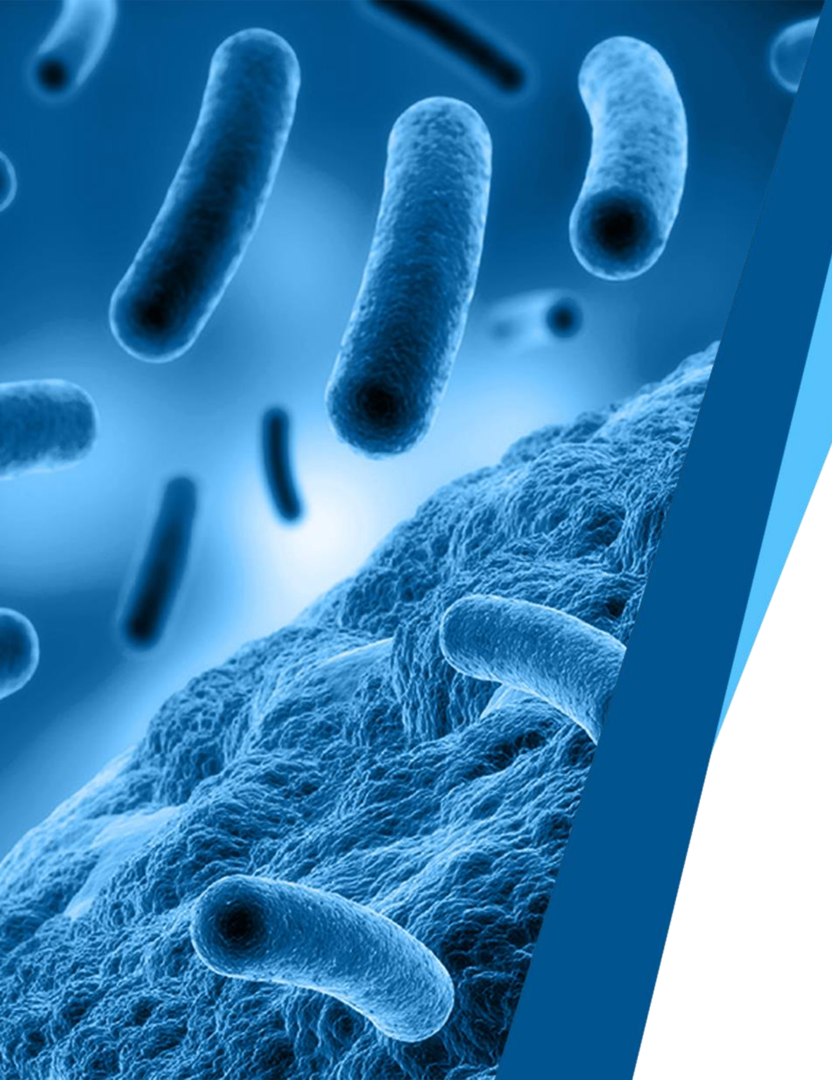
国家微生物科学数据中心
National Microbiology Data Center

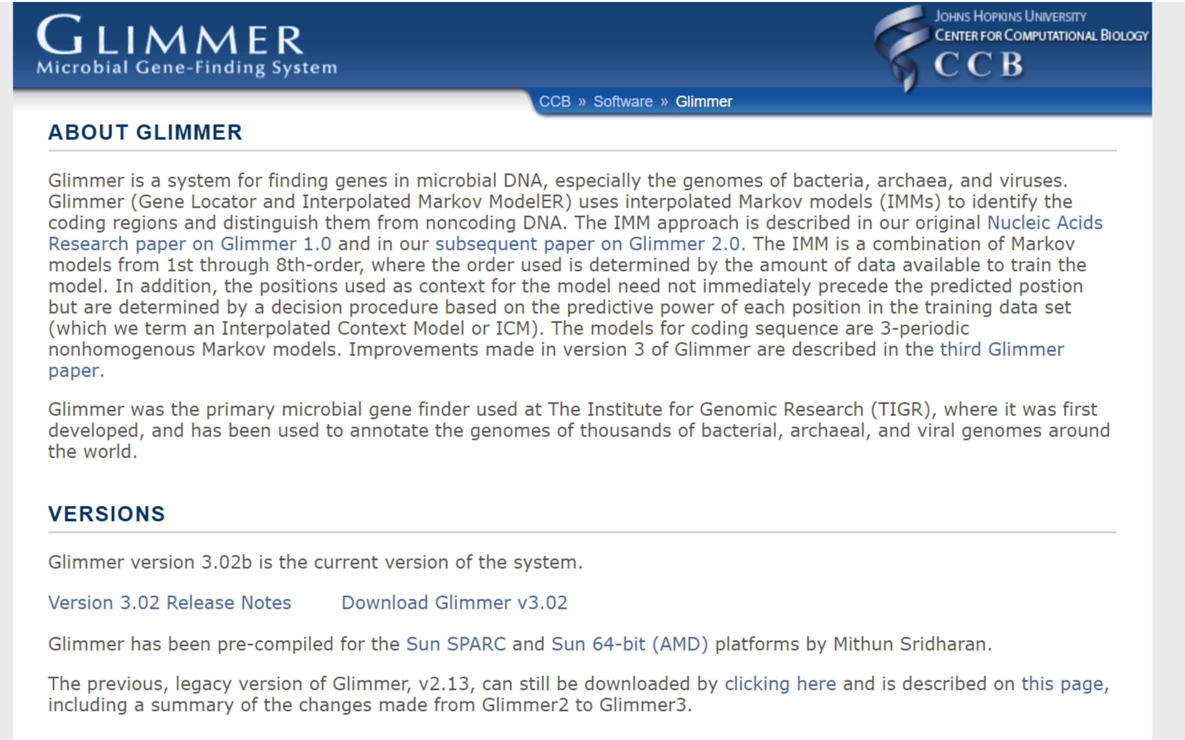


基因预测——Glimmer

Glimmer软件基本介绍

Glimmer是一个用于在微生物DNA中查找基因的软件，尤其是细菌、古细菌和病毒的基因组。

Glimmer使用插值马尔可夫模型（IMM）来识别编码区，并将其与非编码DNA区分开。



GLIMMER
Microbial Gene-Finding System

JOHNS HOPKINS UNIVERSITY
CENTER FOR COMPUTATIONAL BIOLOGY
CCB

CCB » Software » Glimmer

ABOUT GLIMMER

Glimmer is a system for finding genes in microbial DNA, especially the genomes of bacteria, archaea, and viruses. Glimmer (Gene Locator and Interpolated Markov ModelER) uses interpolated Markov models (IMMs) to identify the coding regions and distinguish them from noncoding DNA. The IMM approach is described in our original Nucleic Acids Research paper on Glimmer 1.0 and in our subsequent paper on Glimmer 2.0. The IMM is a combination of Markov models from 1st through 8th-order, where the order used is determined by the amount of data available to train the model. In addition, the positions used as context for the model need not immediately precede the predicted postion but are determined by a decision procedure based on the predictive power of each position in the training data set (which we term an Interpolated Context Model or ICM). The models for coding sequence are 3-periodic nonhomogenous Markov models. Improvements made in version 3 of Glimmer are described in the third Glimmer paper.

Glimmer was the primary microbial gene finder used at The Institute for Genomic Research (TIGR), where it was first developed, and has been used to annotate the genomes of thousands of bacterial, archaeal, and viral genomes around the world.

VERSIONS

Glimmer version 3.02b is the current version of the system.

Version 3.02 Release Notes　　Download Glimmer v3.02

Glimmer has been pre-compiled for the Sun SPARC and Sun 64-bit (AMD) platforms by Mithun Sridharan.

The previous, legacy version of Glimmer, v2.13, can still be downloaded by clicking here and is described on this page, including a summary of the changes made from Glimmer2 to Glimmer3.

http://ccb.jhu.edu/software/glimmer/index.shtml

Glimmer分析步骤

1. 产生长orf数据：long-orfs -n -t 1.15 genom.seq run1.longorfs
   - -n 输出文件去除首行，只包含orf
   - -t 熵距离得分阈值，小于阈值才被保留
   - 输入genom.seq 输出run1.longorfs
2. 提取数据集：extract -t genom.seq run1.longorfs > run1.train
3. 生成预测模型：build-icm -r run1.icm < run1.train
4. 基因预测：glimmer3 -o50 -g110 -t30 genom.seq run1.icm run1
   - -o 最大重叠片段长度阈值，小于阈值保留
   - -g 基因片段长度阈值，大于阈值保留
   - -t orf得分阈值，大于阈值保留
   - 生成两个文件run1. detail和run1.predict
5. 根据预测结果提取序列：extract -t genomseq run1.predict > predict.fasta

国家微生物科学数据中心

http://www.nmdc.cn

国家微生物科学数据中心

当前位置：首页>分析工具

Glimmer

搜索示例：SOAPdenovo2 / CANU / PfamScan / LEfSe /

Blast分析工具	🔥(5)
宏基因组分析流程	(3)
基因组拼接工具	(25)
基因组结构分析	(11)
基因组注释分析	(4)
元基因组分析	(20)
比较基因组分析	(13)

Glimmer

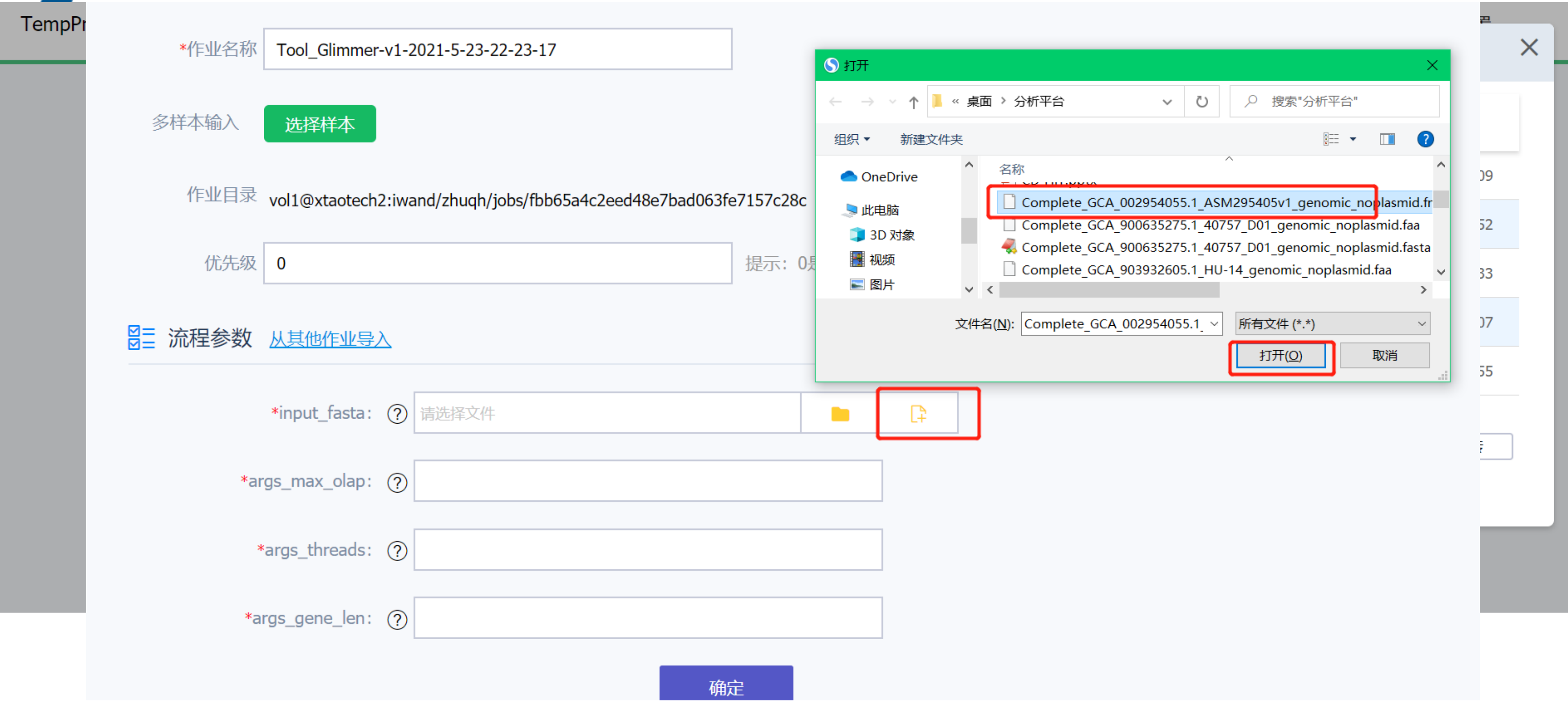
原始链接：http://ccb.jhu.edu/so...

工具介绍：Glimmer系统用以寻找微生物DNA中的基因，尤...

关 键 字：gene prediction

文　　章：Delcher, A.L., Harmon, D., Kasif, S., White, O....

国家微生物科学数据中心



*作业名称 Tool_Glimmer-v1-2021-5-23-22-23-17

多样本输入 选择样本

作业目录 vol1@xtaotech2:iwand/zhuqh/jobs/fbb65a4c2eed48e7bad063fe7157c28c

优先级 0 提示：0

流程参数 从其他作业导入

*input_fasta： 请选择文件

*args_max_olap：

*args_threads：

*args_gene_len：

确定

打开

桌面 › 分析平台

搜索"分析平台"

名称

Complete_GCA_002954055.1_ASM295405v1_genomic_noplasmid.fr

Complete_GCA_900635275.1_40757_D01_genomic_noplasmid.faa

Complete_GCA_900635275.1_40757_D01_genomic_noplasmid.fasta

Complete_GCA_903932605.1_HU-14_genomic_noplasmid.faa

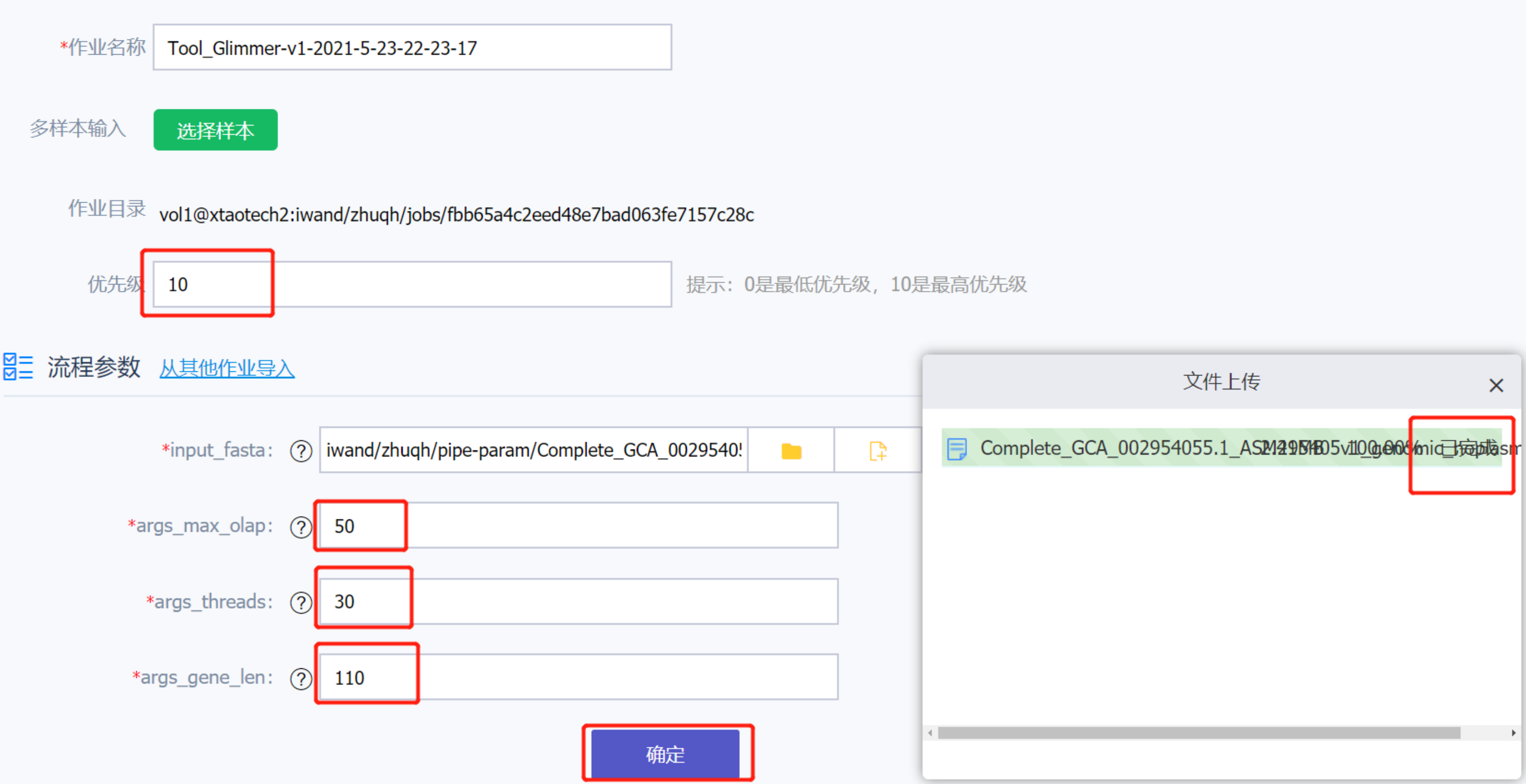
文件名(N): Complete_GCA_002954055.1_

所有文件 (*.*)

国家微生物科学数据中心

```
>NZ_CP021252.1 Corynebacterium striatum strain KC-Na-01 chromosome, complete genome
CTGCCGTGGATGCTTCATCAGAAAGCACACGCATCGCATCGCCGACGGCGCGGAAGGATGGGTCCGCGTTACGGGCGGCA
TCAATCGCCCTTTCGTATTCATGGCGCGCAGATTCCAGATCAGCAGCTGCCTTTTGTGCGGCCTCGCCCTGGTCATTCCA
CCAGTGCCCATCTGTGTTTTCGAGCTGGGAAATCAGCTGCTGGTACTCGGCACCACCATCGGCGATTATCTTTCCCAAGT
CATCCCATGTTTTACCCTGCTTTTCCAGGGTGTCCCCGGTCAAAGTGGCAGCATCGGCCATCGTCTCCATTTTGCCGACA
ACGGATTCGGTGCTGTAGCCAAGCAAGGAAGAGACATCGACAGCTCCTAAAGCAGCGCCACGCCAACCTTCCATTGTCTC
TGCACCCGCGCGCATAGAAGCAGTAAATCCCGCTGCTACCGTAGCGGCGTCTTCCATCTGCTCTTTACTTAGCGGAGTCG
TAGTGCCGGCAAGGGCTACATTAAGCCTTTGCTGCGCTTCAGCCGCATCATCCGTGGTGCGTCGGTAGTGCTCTTGAGCT
TGGGCAGCGCGCTGGTTGAAAGCGATAAGGTCGCCGATAACTGCAGTGGCTCCGGCAAGAGCAATGCCCCACGGACCACC
GAGCATATTCATTAGGCCACTGCCTGCGCCCTTGAGCTTGTCCATTCCTGCGGCGGCGACGCCGCCCTGGCCACCAAGTA
CTCGCATCGCGGCGCCGGTGCGACTCATTTCTGGGTTGGCCTGACGCATATAGCCGTAGGCCTCGGTCAGAGAAGAGCCG
AATACCTTGATTCCAGCGCGAGAAGACGTAAAGGCATCCGTGAGCTTAGTGACCATTGCGGCACCACGTTCAGTCTGCAA
AGCCTTCAGCGCGAACTGGACAGTGCCAAGAGCGATCGCAGCTTCCTTCACCGGCTGCGGAAGCTGACCCCATACTTCAA
CCGACTTTCCGATTGCCTGGGTAGCTGTTCCGAAAGTAGAAGCAGTTAACCCGGCGAATGTCTCGATTGCTGGTGACGCT
TTATCAACCATCGATACCGCGGCATTACCAGTATCGACCAGCTCATCTTTCACGGCATCGAATACACCAAGTCGGACGTC
TTCCATGGTGTTCTGGAAGCGCTCCCACACACCCGGTAGGCCCTGAGTCTGCGCCGCTGCGACCTCCGCGGCCTGGCCCT
GCCGTGTGACCATTTCGCGCAGCTTCGCGAAGCCTTCTGAACCTTGCTCTGCTGCAATACCTGCCAAGCGCATTGCGTCA
GAGCCGAAAAGTGTAGCCGTTGCGGCTTGATACTGCTCGTCGGTCATTCGCTTGGAAGCGTCCTGGAGTTGTCCCATCAG
GGACTCCATGCCTACAAACTTTCCTTGCGCGTCATAAACCGTCAGGCCCAACTCTTCGATTGCCTGCTGCGCTGGCTTTC
CCTGATCCGTCAAAGCAAGCATTGCAGATTTCAGCAGCGTACCTGCGTCGCTTCCCTGGATACCAGCATTCGCCATCATC
GCCAGCGAGGTCGCCGTGTCATCAATCGAGACACCAAAGGCGTTTGCAACTGTGCCCGCCTGCTGCAGACCTTGAGAGAT
ACCGGTCATCTCCGCAGAAGACGCATTCGCTGCACCTGCCAAGATGTCAGAAACGCGGGCAGCATCGCCTGCGCCTAAGG
AAAAAGCTTGAAGTGCCTGCGACTGGATGGTTGCGGCCTCGGCAGCATCGACCTGTGCAGCAGCAGCCAGCTGCAGGGTG
CCACGTGCAGCAGTCATAGACTGCTCCACTGAGAAACCACCCTTGGCAAGCTCTGTCATAGCGGCAGCAGCATCAGCGGC
AGACGTTCCCGTCAGCTGGGTATCCGTGCCCAAATCACGAGCTTTAGCACGGACAGCGTCGAGCTGCGCGCCTGTGGCTT
GAGAAACTGCCGCCATGGAGTTAATTTGAGACTCAAACTCCATGCCGATCTGTCCAACTGCGCGCGCCATCTCTGCGCCA
CCGAGTGCAAGGCCGAGCGAAGCGCCAAGCTTGGACGCAACGCCCATTGCCGGGCCCAACCCAGACTCAAGTTTGGAACT
AAAGTTCTTGACGTCTGGCTCGACGAGAATATCAATCTTGCCGCCTGCCATTTTGCCTCCTTATTCGCCGCGCGCAGCCA
AGAGCTGGCGAATAGTGCCTTTCTTACGCTGGCTCTTATCCGCTCCTGACGACATCGGTGTATCTCGTGGTTTTTCTTTC
TCGGGATGGAAGGAAATAAATTCCTGGTGGGCGTCGAGCAGCAGCTCCATATGGATTCGTTCAGGCCTTTTCGCCACAGG
GAACACAATCGGATGCGGCGGAGCCTTATACCCGGCGCGCTTTCGCTCTGCGCGGGCGCGCTTTATCTCGGGGTCTTCCG
GGTCTGTGGTCCATTGGGTGTATTCGGTGCCTAGCCAGTAGTCCAGGCGGTCGAGTATGCGGGCTACGTTTTCTTCTAGG
TCTGTCCATTCGTCGAGTTTTTGGACGAGGATGGCTAGGTCTAGGTGGCACATAGTCTCGCAGGCTTGCCGCCAGTCCAG
GCTGTATTGGCGTCGGAATCCTGTGAGGGCTTTGGCGTAGCCCTCGTGGGTGGTTAGCGCTGCTGTCCAGCTAAAAAATT
TCCGTCGTTATCGCGGAGTCCTGCTACCTTGCCCAGTTCGATGCATACTTTCGATGCGACCTTGATGTCGGCGGCTAGGA
GCCATTCGACGAGGTTCTTGGTGTCTTCCTTCGAAAGGTCGGTGAGGAAGTCGAGTTGGCGCTTGATTTGTTCCTCTGGG
GACGGCACGGTGTCGGGTGTCCACTTGGAGGTGTTGAAGAACTCGATGTAGTTGGCGACTTCTGCGCCGGTGAAGTTGCG
GCGGATGTTGAAGTCGATGCCGTTTAGCGTTGCTGGTTGTGGTTCTCCGCCGTTGGTTGCAAATGCGTTTTCGAAGATGT
CAATTCGGTTGGTTTCGGCAGTTTTCTTCGAGTTCTTTTGGTTAGCCATGGTTATTTCCTTTTCCGTCTCGGTCGGTTGG
```

国家微生物科学数据中心



*作业名称 Tool_Glimmer-v1-2021-5-23-22-23-17

多样本输入 选择样本

作业目录 vol1@xtaotech2:iwand/zhuqh/jobs/fbb65a4c2eed48e7bad063fe7157c28c

优先级 10 提示：0是最低优先级，10是最高优先级

流程参数 从其他作业导入

*input_fasta： iwand/zhuqh/pipe-param/Complete_GCA_0029540!

*args_max_olap： 50

*args_threads： 30

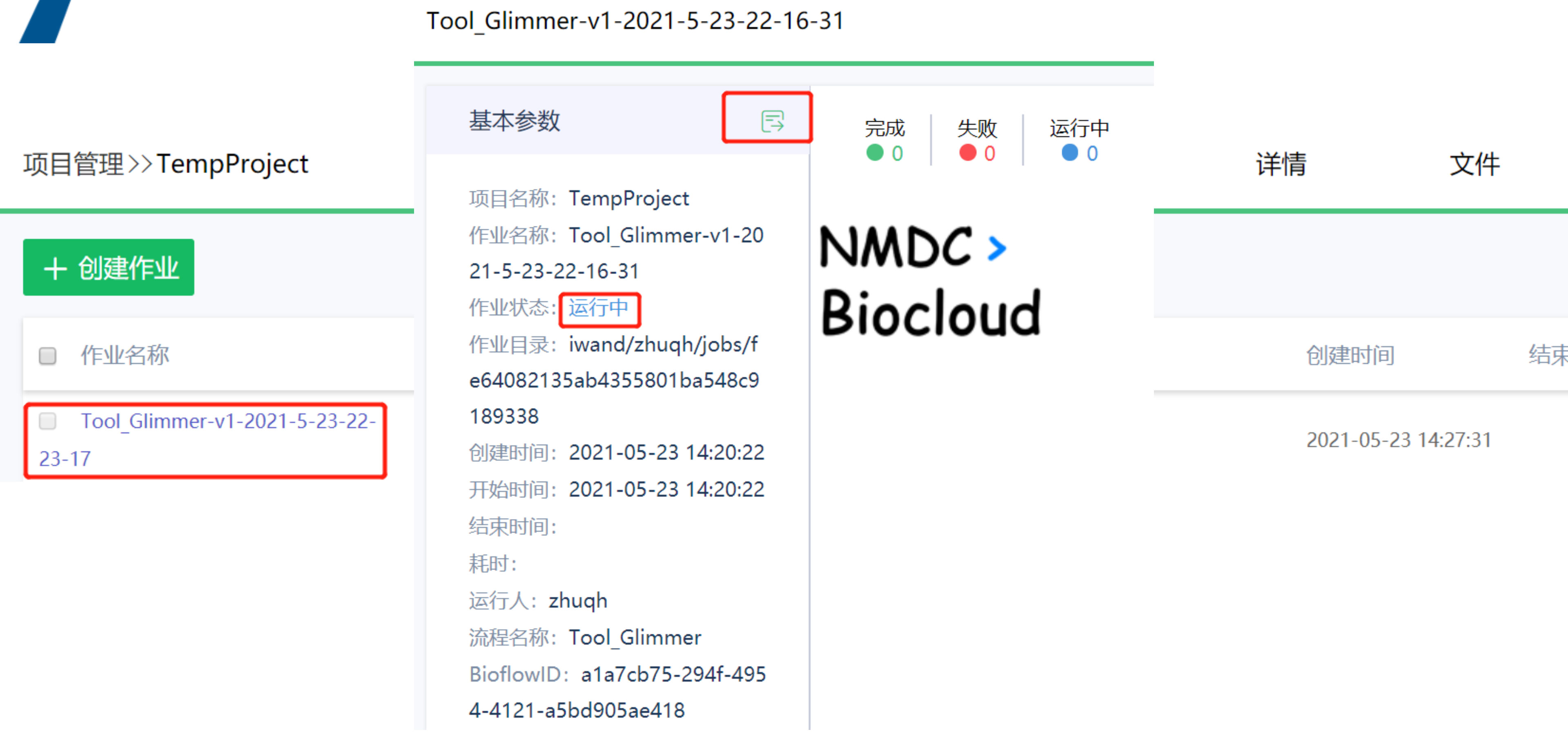
*args_gene_len： 110

确定

文件上传

Complete_GCA_002954055.1_ASM295405v1_genomic_plasm 2.19MB 100.00% 已完成

国家微生物科学数据中心



Tool_Glimmer-v1-2021-5-23-22-16-31

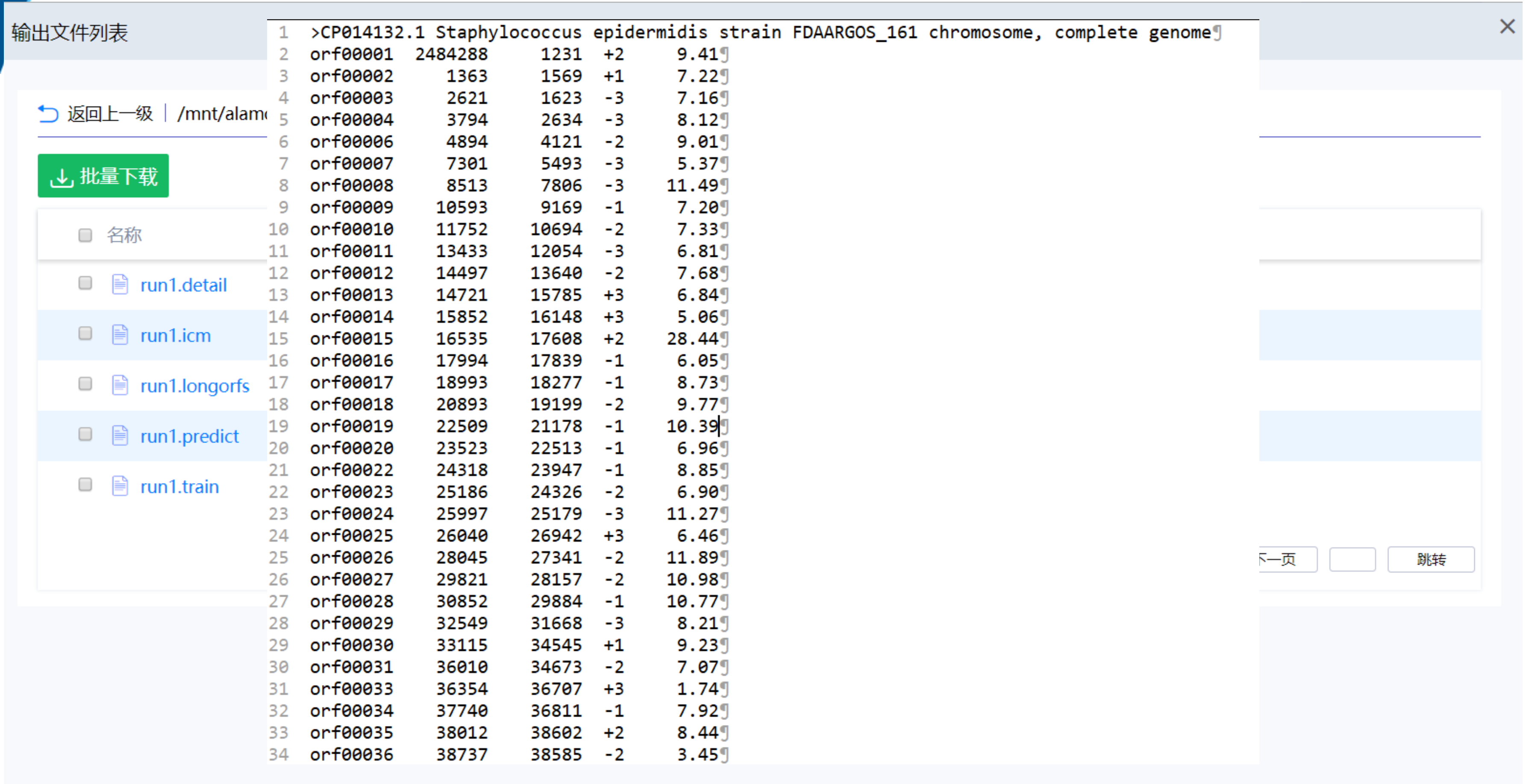
基本参数

项目名称：TempProject
作业名称：Tool_Glimmer-v1-2021-5-23-22-16-31
作业状态：运行中
作业目录：iwand/zhuqh/jobs/fe64082135ab4355801ba548c9189338
创建时间：2021-05-23 14:20:22
开始时间：2021-05-23 14:20:22
结束时间：
耗时：
运行人：zhuqh
流程名称：Tool_Glimmer
BioflowID：a1a7cb75-294f-4954-4121-a5bd905ae418

☐ 作业名称	创建时间	结束
☐ Tool_Glimmer-v1-2021-5-23-22-23-17	2021-05-23 14:27:31	

NMDC > Biocloud

国家微生物科学数据中心



```
>CP014132.1 Staphylococcus epidermidis strain FDAARGOS_161 chromosome, complete genome
orf00001  2484288     1231  +2     9.41
orf00002     1363     1569  +1     7.22
orf00003     2621     1623  -3     7.16
orf00004     3794     2634  -3     8.12
orf00006     4894     4121  -2     9.01
orf00007     7301     5493  -3     5.37
orf00008     8513     7806  -3    11.49
orf00009    10593     9169  -1     7.20
orf00010    11752    10694  -2     7.33
orf00011    13433    12054  -3     6.81
orf00012    14497    13640  -2     7.68
orf00013    14721    15785  +3     6.84
orf00014    15852    16148  +3     5.06
orf00015    16535    17608  +2    28.44
orf00016    17994    17839  -1     6.05
orf00017    18993    18277  -1     8.73
orf00018    20893    19199  -2     9.77
orf00019    22509    21178  -1    10.39
orf00020    23523    22513  -1     6.96
orf00022    24318    23947  -1     8.85
orf00023    25186    24326  -2     6.90
orf00024    25997    25179  -3    11.27
orf00025    26040    26942  +3     6.46
orf00026    28045    27341  -2    11.89
orf00027    29821    28157  -2    10.98
orf00028    30852    29884  -1    10.77
orf00029    32549    31668  -3     8.21
orf00030    33115    34545  +1     9.23
orf00031    36010    34673  -2     7.07
orf00033    36354    36707  +3     1.74
orf00034    37740    36811  -1     7.92
orf00035    38012    38602  +2     8.44
orf00036    38737    38585  -2     3.45
```

谢谢大家