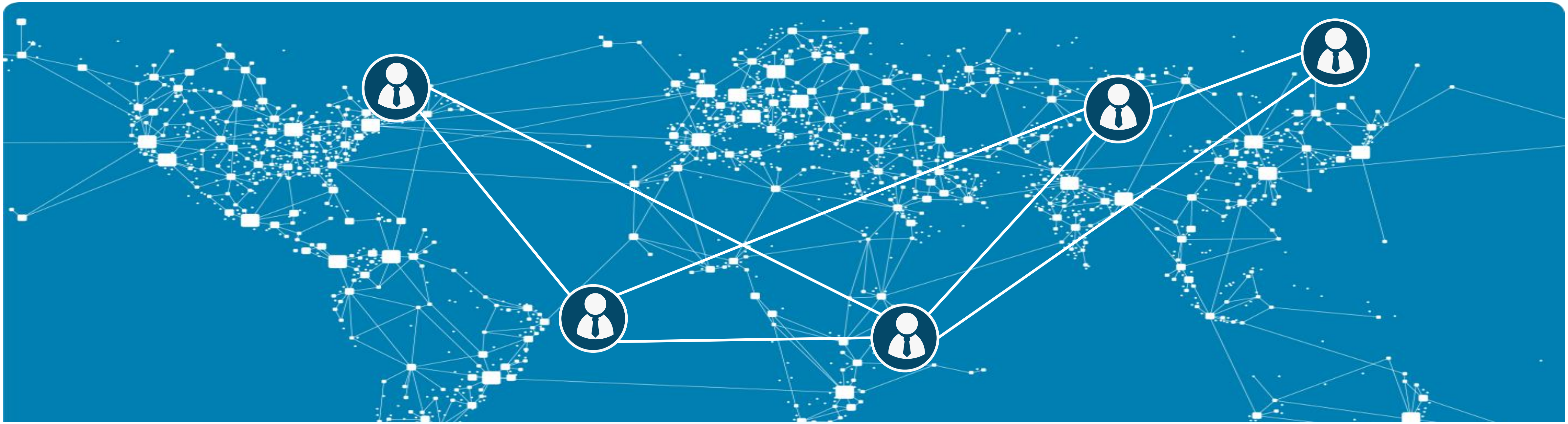
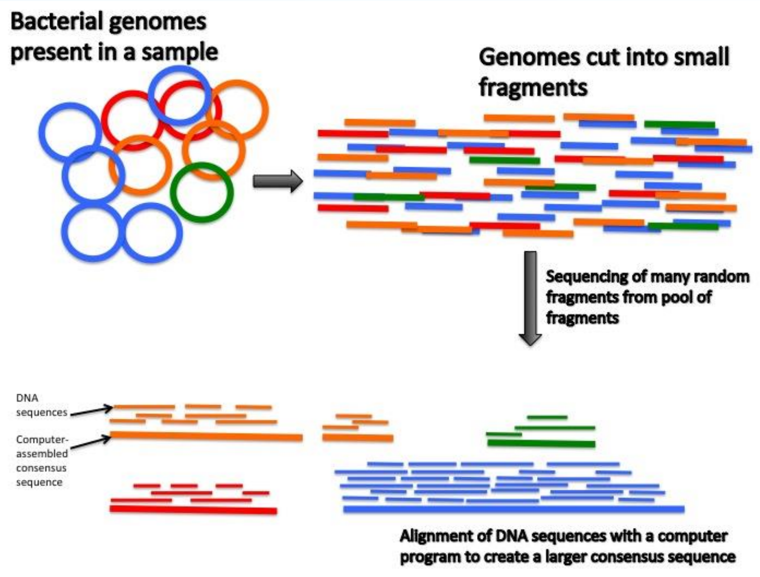


32 拼接/组装Assembling

易生信
2019年6月23日



Bacterial genomes present in a sample

Genomes cut into small fragments

Sequencing of many random fragments from pool of fragments

DNA sequences

Computer-assembled consensus sequence

Alignment of DNA sequences with a computer program to create a larger consensus sequence

目录

一. KneadData质控

二. Kraken2物种分类

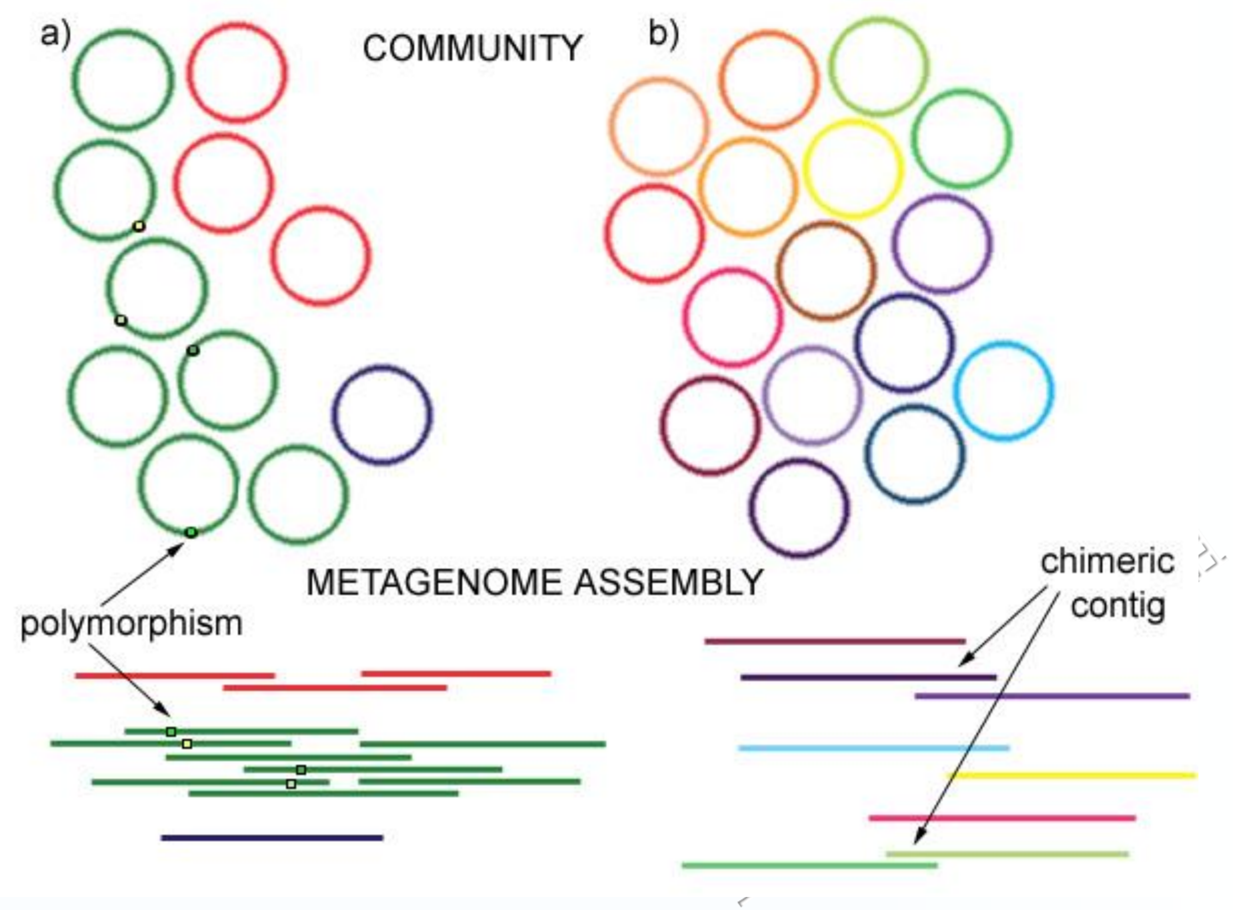
三. **Megahit拼接**

四. Prokka基因注释

五. Cd-hit非冗余基因集(可选)

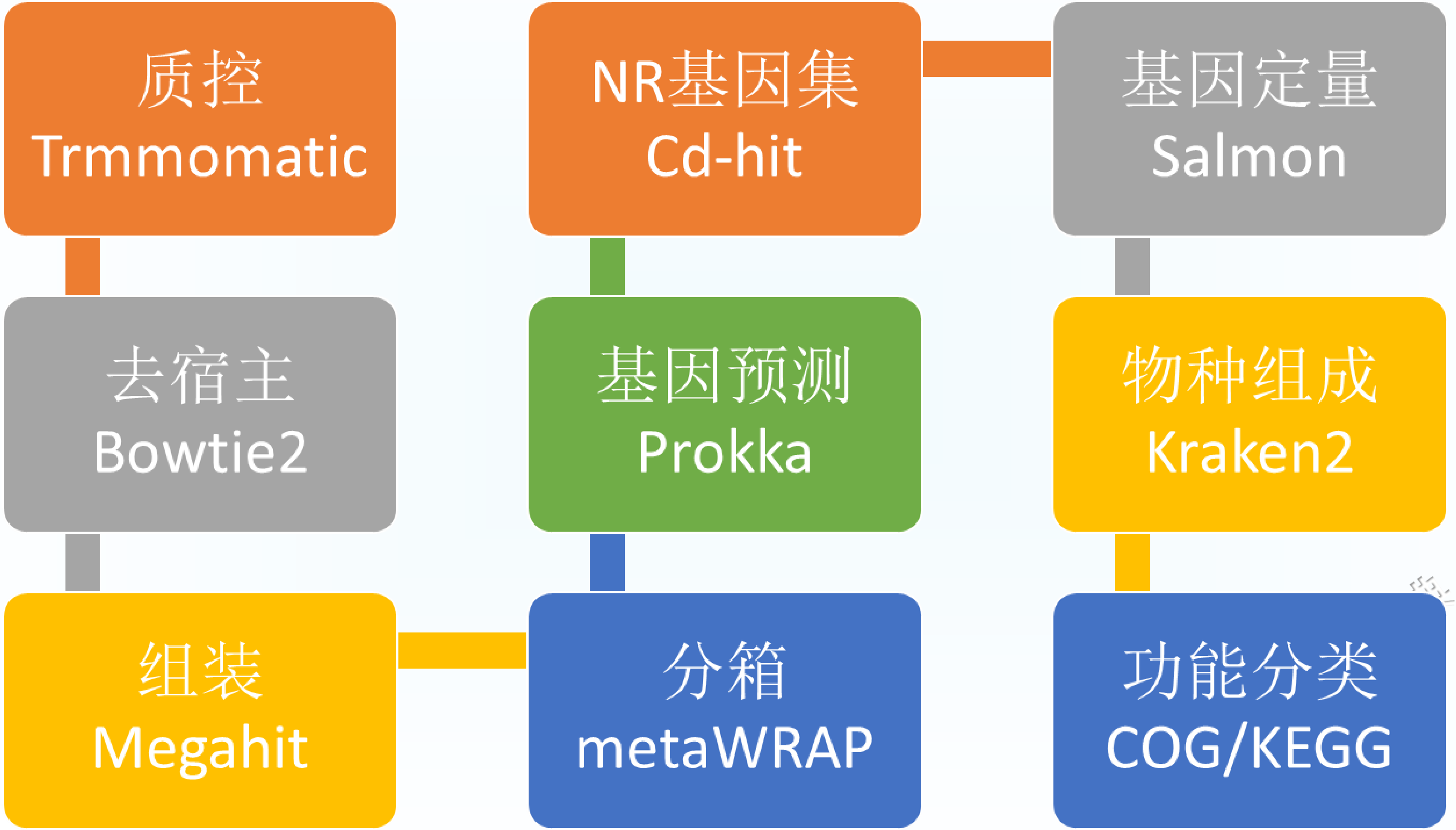
六. Salmon基因定量

七. 基因功能注释



http://wiki.biomine.skelleftea.se/wiki/index.php/Metagenomics

宏基因组无参(*De novo*)分析主要软件和数据库



质控
Trmmomatic

去宿主
Bowtie2

组装
Megahit

分箱
metaWRAP

基因预测
Prokka

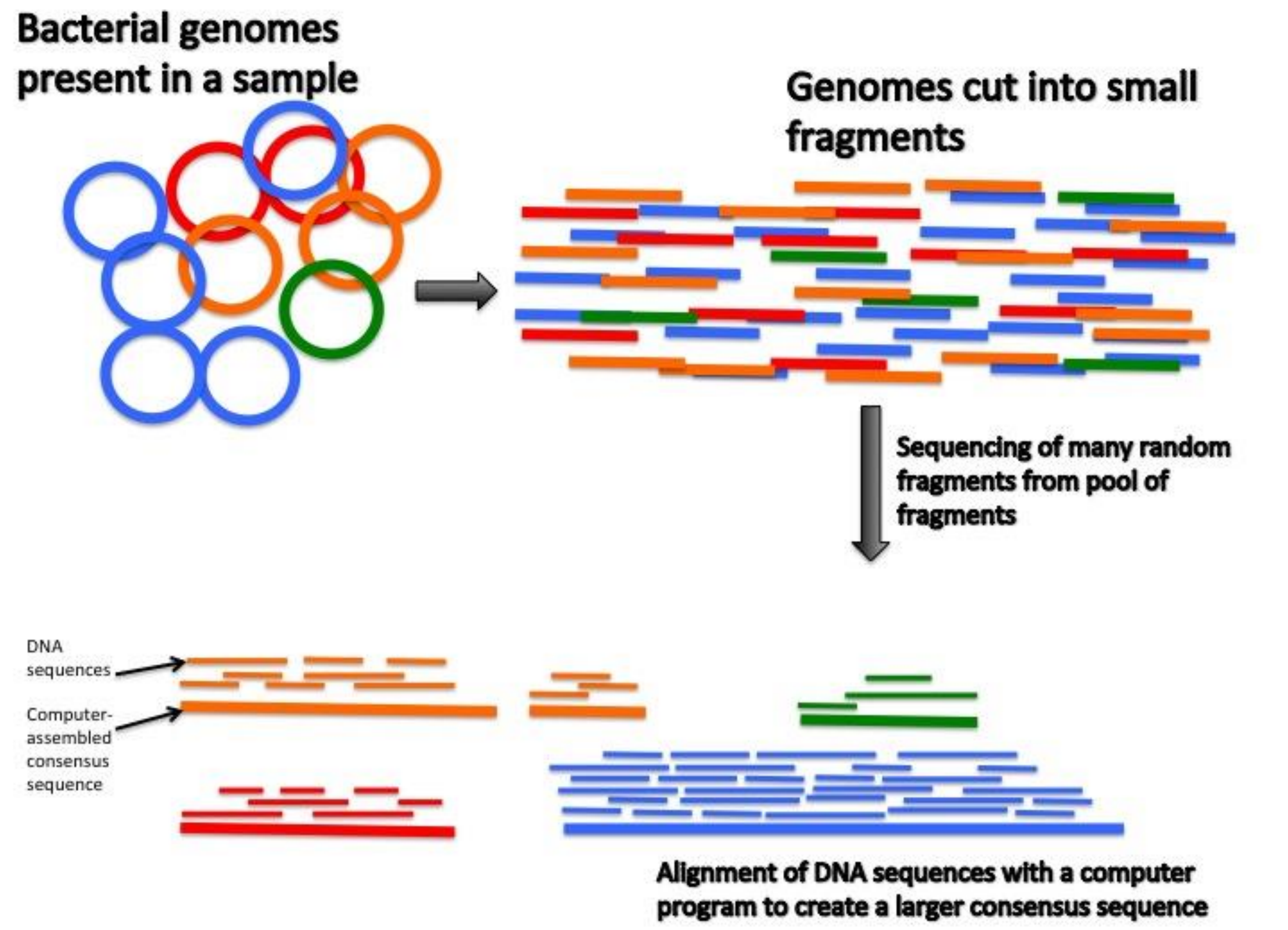
NR基因集
Cd-hit

基因定量
Salmon

物种组成
Kraken2

功能分类
COG/KEGG

拼接/组装(Assemble)的基本原理


Bacterial genomes present in a sample
Genomes cut into small fragments
Sequencing of many random fragments from pool of fragments
DNA sequences
Computer-assembled consensus sequence
Alignment of DNA sequences with a computer program to create a larger consensus sequence

https://teachthemicrobiome.weebly.com/sequencing-the-microbiome.html

拼接中常见名词

- Read: 高通量测序平台产生的序列
- Contig：拼接软件基于reads之间的overlap区，拼接获得的序列
- Scaffold：双端测序时，同一条序列两端的reads分布于不同的contigs上，可确定contigs的位置距离时中间用N连接
- N50: 将Contig或Scaffold按长度由大到小排列 ，累加总长度50%时，所在序列长度，用于表示拼接质量的重要参数
- Depth：测序深度，即测序总碱基与基因组大小的比值，如人类30X，即90G数据，宏基因组中要求较完整获得相对丰度1%的细菌基因组，测序量为：5 MB * 30X / 1% = 15GB
- 覆盖度Coverage：测序获得的序列占整个基因组的比例，如97%即3%没测到。

https://wenku.baidu.com/view/e40659fe9e3143323968937a.html

组装软件评估

- 组装结果中存在大量错误
- 高复杂度的宏基因组推荐使用MetaSPAdes

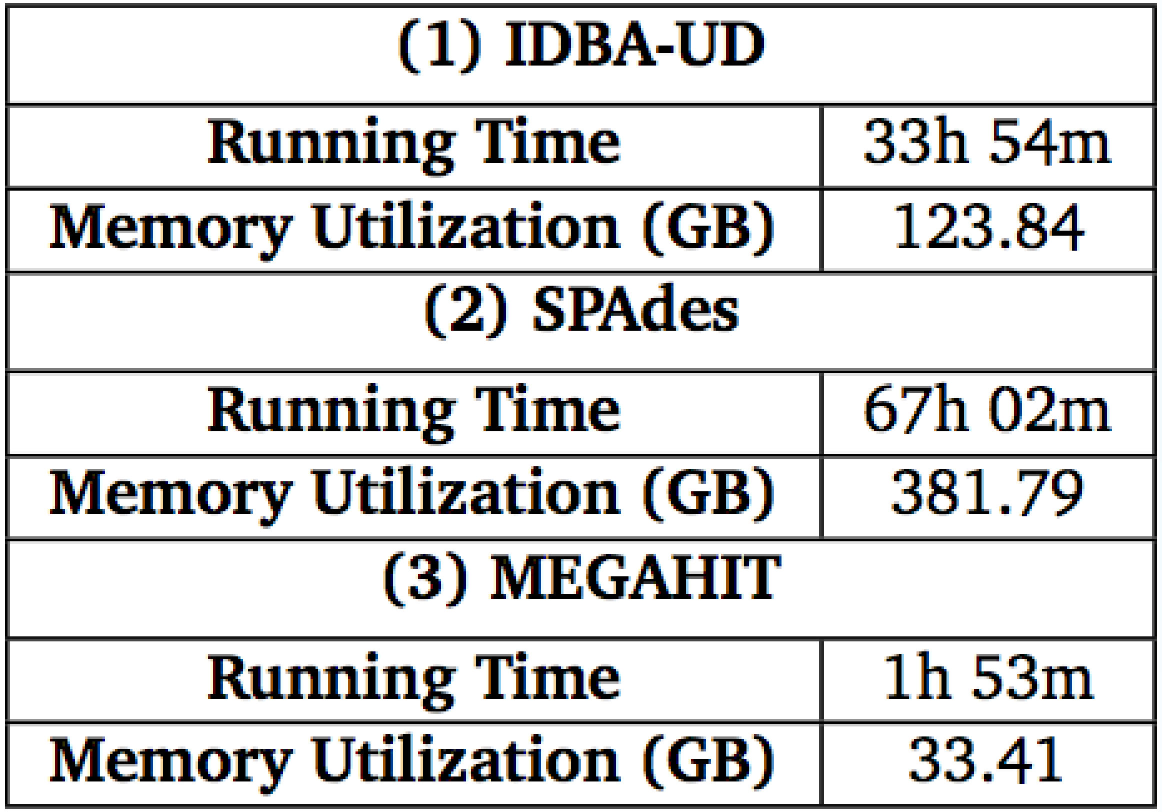
  Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome Research* **27**, 824-834, doi:10.1101/gr.213959.116 (2017).
- 低复杂度的宏基因组推荐使用MaSuRCA

  Zimin, A. V. *et al.* The MaSuRCA genome assembler. *Bioinformatics* **29**, 2669-2677, doi:10.1093/bioinformatics/btt476 (2013).
- **Megahit是最保守的组装软件，拥有最小的N50和错误率**

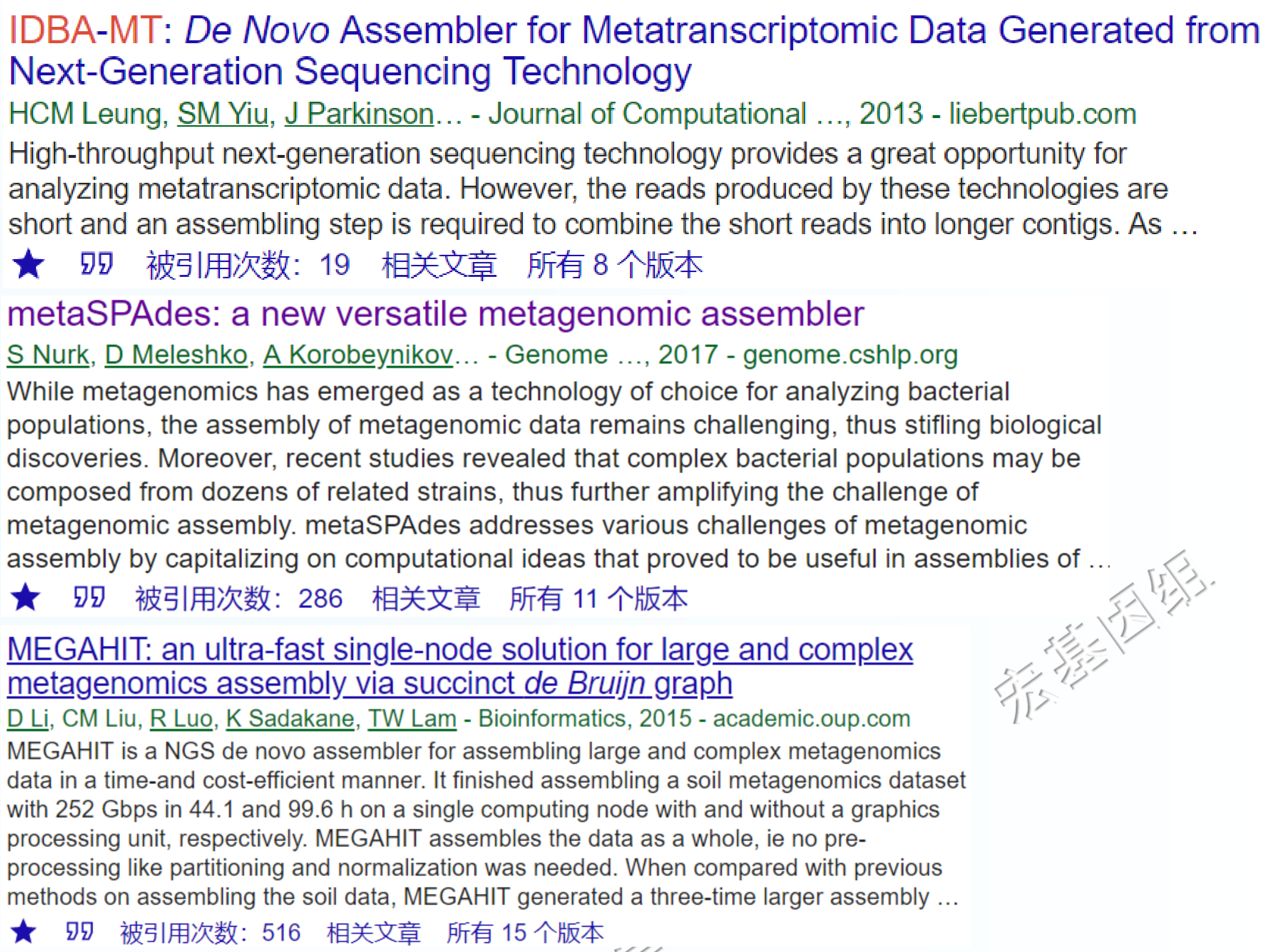
  Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674-1676, doi:10.1093/bioinformatics/btv033 (2015).

Forouzan, E., Shariati, P., Mousavi Maleki, M. S., Karkhane, A. A. & Yakhchali, B. Practical evaluation of 11 de novo assemblers in metagenome assembly. *Journal of Microbiological Methods* **151**, 99-105, doi:https://doi.org/10.1016/j.mimet.2018.06.007 (2018).

主流拼接软件：运行时间和内存需求比较



(1) IDBA-UD	
Running Time	33h 54m
Memory Utilization (GB)	123.84
(2) SPAdes	
Running Time	67h 02m
Memory Utilization (GB)	381.79
(3) MEGAHIT	
Running Time	1h 53m
Memory Utilization (GB)	33.41



IDBA-MT: *De Novo* Assembler for Metatranscriptomic Data Generated from Next-Generation Sequencing Technology

HCM Leung, SM Yiu, J Parkinson… - Journal of Computational …, 2013 - liebertpub.com

High-throughput next-generation sequencing technology provides a great opportunity for analyzing metatranscriptomic data. However, the reads produced by these technologies are short and an assembling step is required to combine the short reads into longer contigs. As …

★ ❞ 被引用次数：19 相关文章 所有 8 个版本

metaSPAdes: a new versatile metagenomic assembler

S Nurk, D Meleshko, A Korobeynikov… - Genome …, 2017 - genome.cshlp.org

While metagenomics has emerged as a technology of choice for analyzing bacterial populations, the assembly of metagenomic data remains challenging, thus stifling biological discoveries. Moreover, recent studies revealed that complex bacterial populations may be composed from dozens of related strains, thus further amplifying the challenge of metagenomic assembly. metaSPAdes addresses various challenges of metagenomic assembly by capitalizing on computational ideas that proved to be useful in assemblies of …

★ ❞ 被引用次数：286 相关文章 所有 11 个版本

MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct *de Bruijn* graph

D Li, CM Liu, R Luo, K Sadakane, TW Lam - Bioinformatics, 2015 - academic.oup.com

MEGAHIT is a NGS de novo assembler for assembling large and complex metagenomics data in a time-and cost-efficient manner. It finished assembling a soil metagenomics dataset with 252 Gbps in 44.1 and 99.6 h on a single computing node with and without a graphics processing unit, respectively. MEGAHIT assembles the data as a whole, ie no pre-processing like partitioning and normalization was needed. When compared with previous methods on assembling the soil data, MEGAHIT generated a three-time larger assembly …

★ ❞ 被引用次数：516 相关文章 所有 15 个版本

Leung H C M, Yiu S M, Parkinson J, et al. IDBA-MT: de novo assembler for metatranscriptomic data generated from next-generation sequencing technology[J]. Journal of Computational Biology, 2013, 20(7): 540-550.

Nurk S, Meleshko D, Korobeynikov A, et al. metaSPAdes: a new versatile metagenomic assembler[J]. Genome research, 2017, 27(5): 824-834.

Li D, Liu C M, Luo R, et al. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph[J]. Bioinformatics, 2015, 31(10): 1674-1676.

MEGAHIT——多快好省的组装神器

- 最快，最省内存，且在宏基因组拼接中质量可接受的软件
- -h显示参数详细
- -1/2左或右端文件，支持多文件；--12双端交替(interleave)的单文件；-r单端
- -t设置线程数，默认全用
- --use-gpu 支持GPU运算
- --continue 支持中断继续运行
- --k-min 27 --k-max 191 --k-step 20 # 手动设置kmer

Li, Dinghua, et al. "MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph." *Bioinformatics* 31.10 (2015): 1674-1676.
Li, Dinghua, et al. "MEGAHIT v1. 0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices." *Methods* 102 (2016): 3-11.

3.1.1 Megahit拼接

```
rm -rf temp/megahit
time megahit -t 9 \
  -1 `ls temp/qc/*_1_kneaddata_paired_1.fastq|tr '\n' ','|sed 's/,$//'` \
  -2 `ls temp/qc/*_1_kneaddata_paired_2.fastq|tr '\n' ','|sed 's/,$//'` \
  -o temp/megahit # --k-min 27 --k-max 191 --k-step 20
# `ls...`用于自动获得文件列表，无需手动添写
# 默认使用所有线程，如本机192线程，real 43s, user 96m, sys 1m
# 设定9线程，real 1m, user 8m, sys 1m；虽才快了17s，但浪费了10倍资源
# 查看拼接结果
head temp/megahit/final.contigs.fa
```

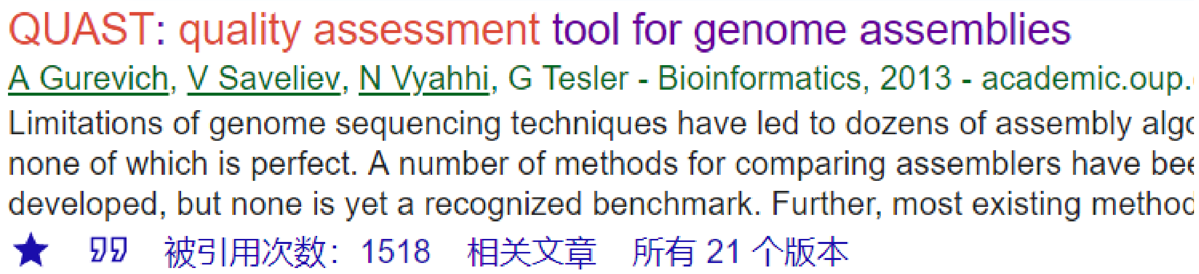
3.1.2 metaSPAdes拼接(可选)

- conda install spades # 安装软件
- metaspades.py -h # 查看帮助
- metaspades.py --test # 运行测试数据

```
time metaspades.py -t 24 -m 500 \
  `ls temp/qc/*_1_kneaddata_paired_1.fastq|sed 's/^/-1 /'| tr '\n' ' '` \
  `ls temp/qc/*_1_kneaddata_paired_2.fastq|sed 's/^/-2 /'| tr '\n' ' '` \
  -o metaspades
# 24线程 17m，内存500G
# 此外 --iontorrent 支持PGM数据，甚至支持--pacbio和--nanopore
```

Nurk, Sergey, et al. "metaSPAdes: a new versatile metagenomic assembler." *Genome research* (2017): gr-213959.
主页：http://cab.spbu.ru/software/spades/ Meta帮助: http://cab.spbu.ru/files/release3.12.0/manual.html#meta

3.1.3 QUAST评估

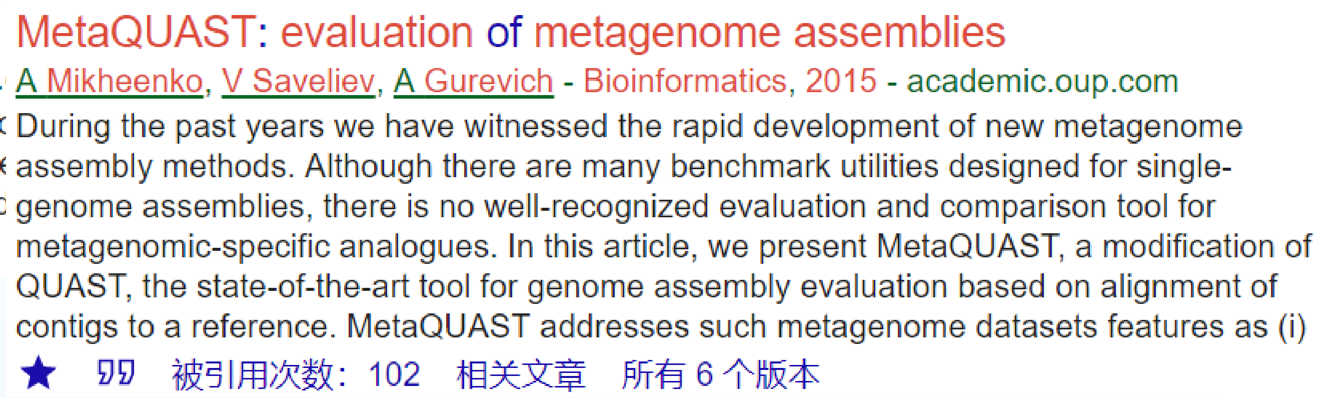


QUAST: quality assessment tool for genome assemblies

A Gurevich, V Saveliev, N Vyahhi, G Tesler - Bioinformatics, 2013 - academic.oup.

Limitations of genome sequencing techniques have led to dozens of assembly algo none of which is perfect. A number of methods for comparing assemblers have bee developed, but none is yet a recognized benchmark. Further, most existing method

★ ꟷ 被引用次数：1518 相关文章 所有 21 个版本



MetaQUAST: evaluation of metagenome assemblies

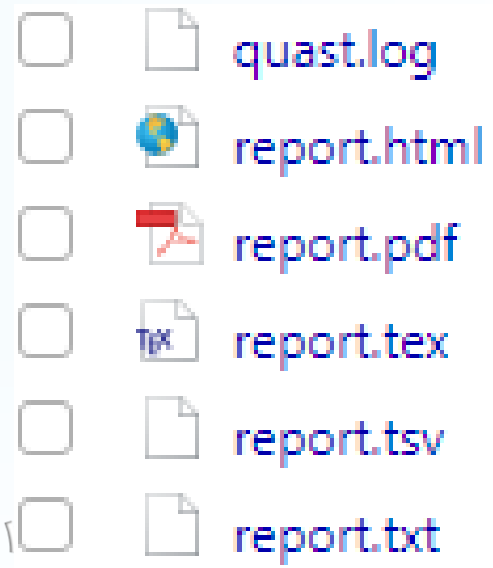
A Mikheenko, V Saveliev, A Gurevich - Bioinformatics, 2015 - academic.oup.com

During the past years we have witnessed the rapid development of new metagenome assembly methods. Although there are many benchmark utilities designed for single-genome assemblies, there is no well-recognized evaluation and comparison tool for metagenomic-specific analogues. In this article, we present MetaQUAST, a modification of QUAST, the state-of-the-art tool for genome assembly evaluation based on alignment of contigs to a reference. MetaQUAST addresses such metagenome datasets features as (i)

★ ꟷ 被引用次数：102 相关文章 所有 6 个版本

```
quast.py -h # 显示帮助
mkdir -p result/megahit/
ln temp/megahit/final.contigs.fa result/megahit/
quast.py result/megahit/final.contigs.fa -o result/megahit/
# 生成report文本tsv/txt、网页html、PDF等格式报告

# 依赖数据库更全面评估
metaquast.py result/megahit/final.contigs.fa -o
result/megahit/
```



□ quast.log

□ report.html

□ report.pdf

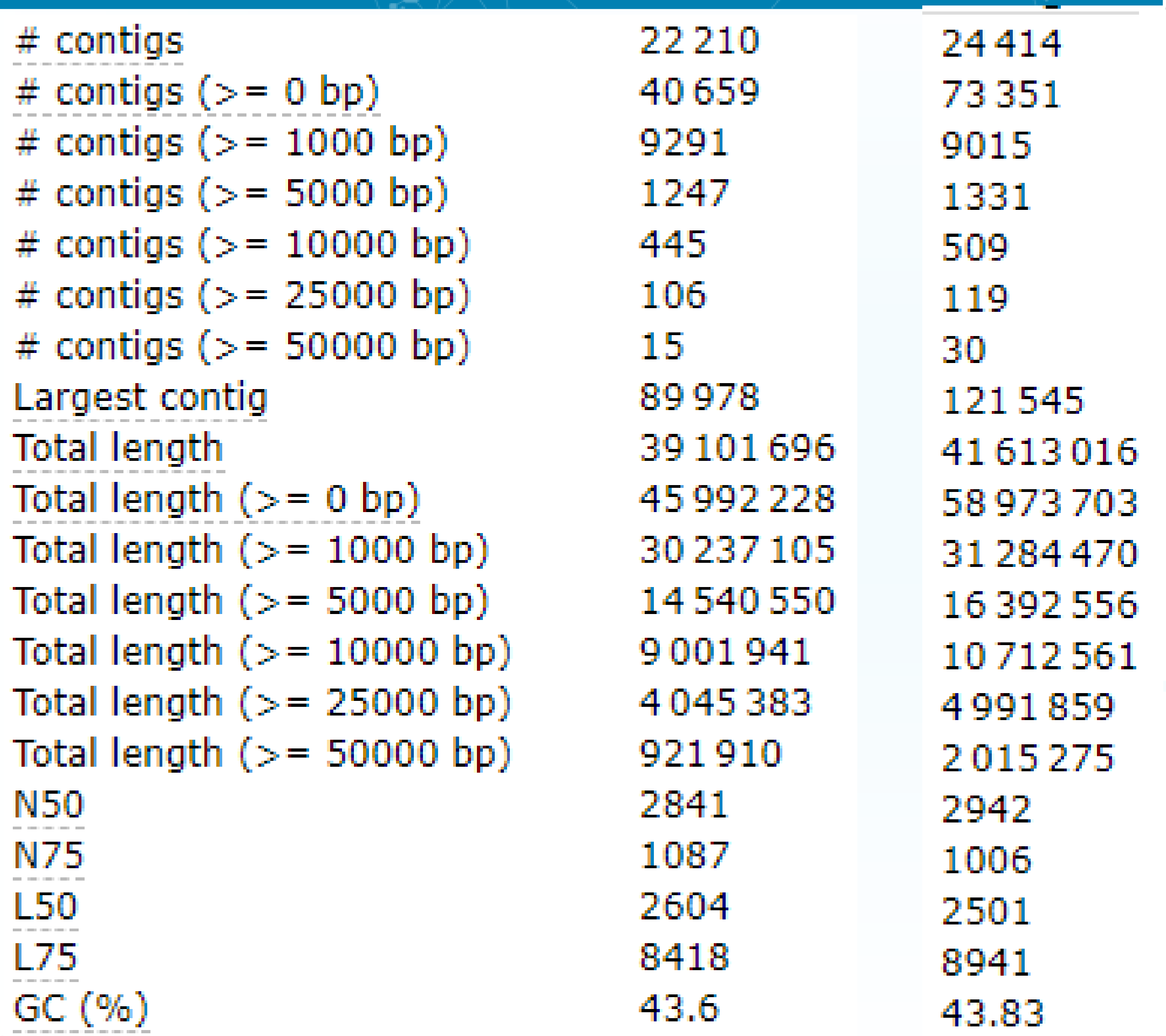
□ report.tex

□ report.tsv

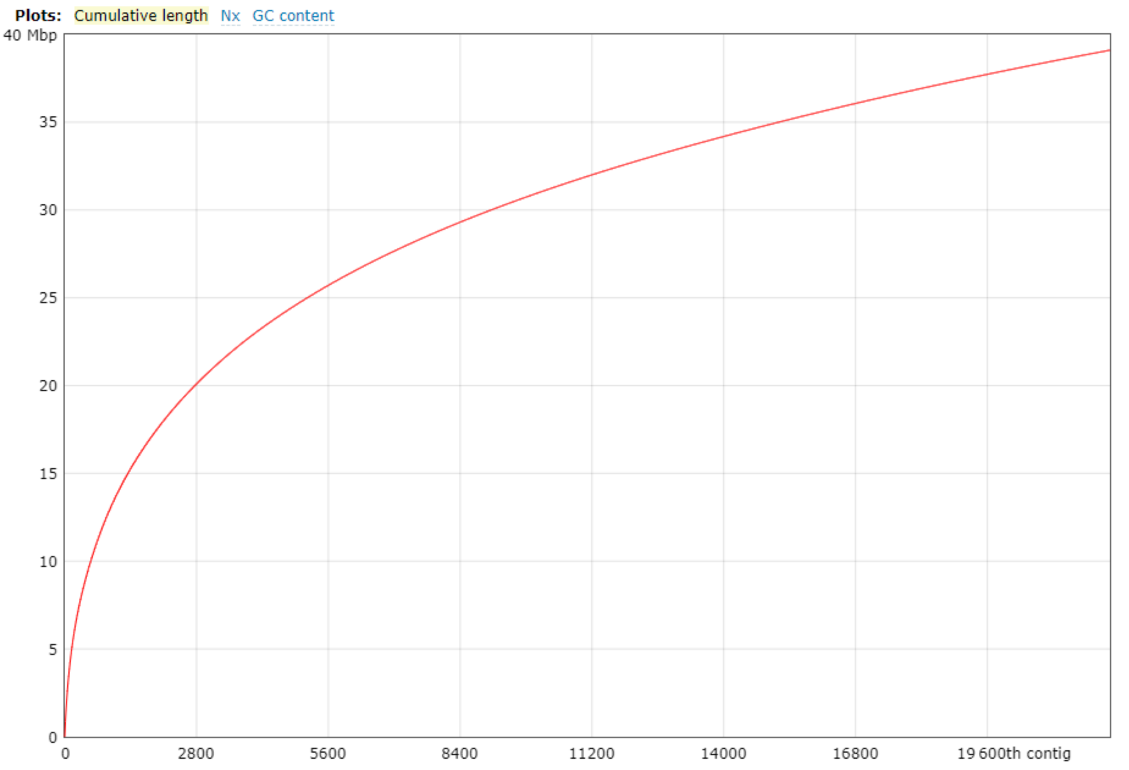
□ report.txt

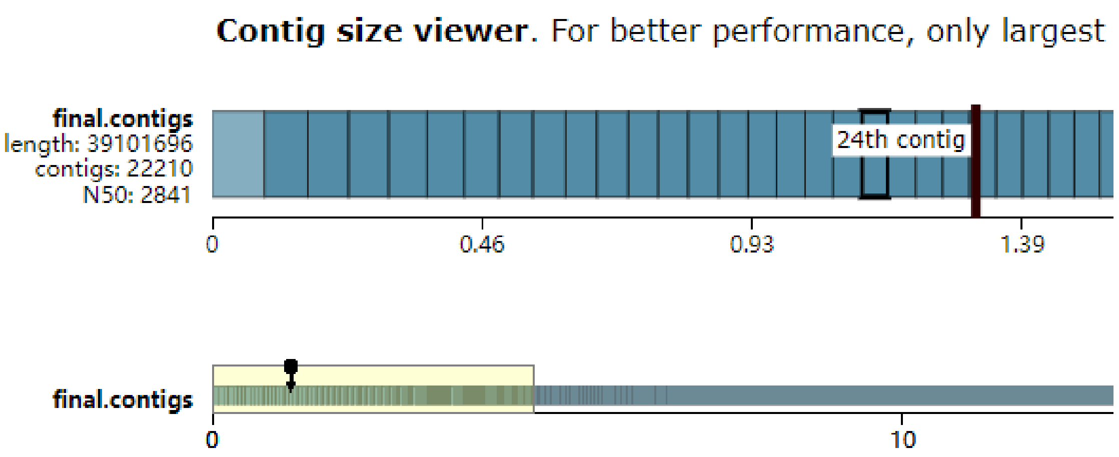
软件安装和使用指南帮助 http://quast.bioinf.spbau.ru/manual.html

评估结果：megahit vs metaspades



# contigs	22 210	24 414
# contigs (>= 0 bp)	40 659	73 351
# contigs (>= 1000 bp)	9291	9015
# contigs (>= 5000 bp)	1247	1331
# contigs (>= 10000 bp)	445	509
# contigs (>= 25000 bp)	106	119
# contigs (>= 50000 bp)	15	30
Largest contig	89 978	121 545
Total length	39 101 696	41 613 016
Total length (>= 0 bp)	45 992 228	58 973 703
Total length (>= 1000 bp)	30 237 105	31 284 470
Total length (>= 5000 bp)	14 540 550	16 392 556
Total length (>= 10000 bp)	9 001 941	10 712 561
Total length (>= 25000 bp)	4 045 383	4 991 859
Total length (>= 50000 bp)	921 910	2 015 275
N50	2841	2942
N75	1087	1006
L50	2604	2501
L75	8418	8941
GC (%)	43.6	43.83





Contig size viewer. For better performance, only largest

目录

一. KneadData质控(已完成)

二. Kraken物种分类

三. Megahit拼接

四. **Prokka基因注释**

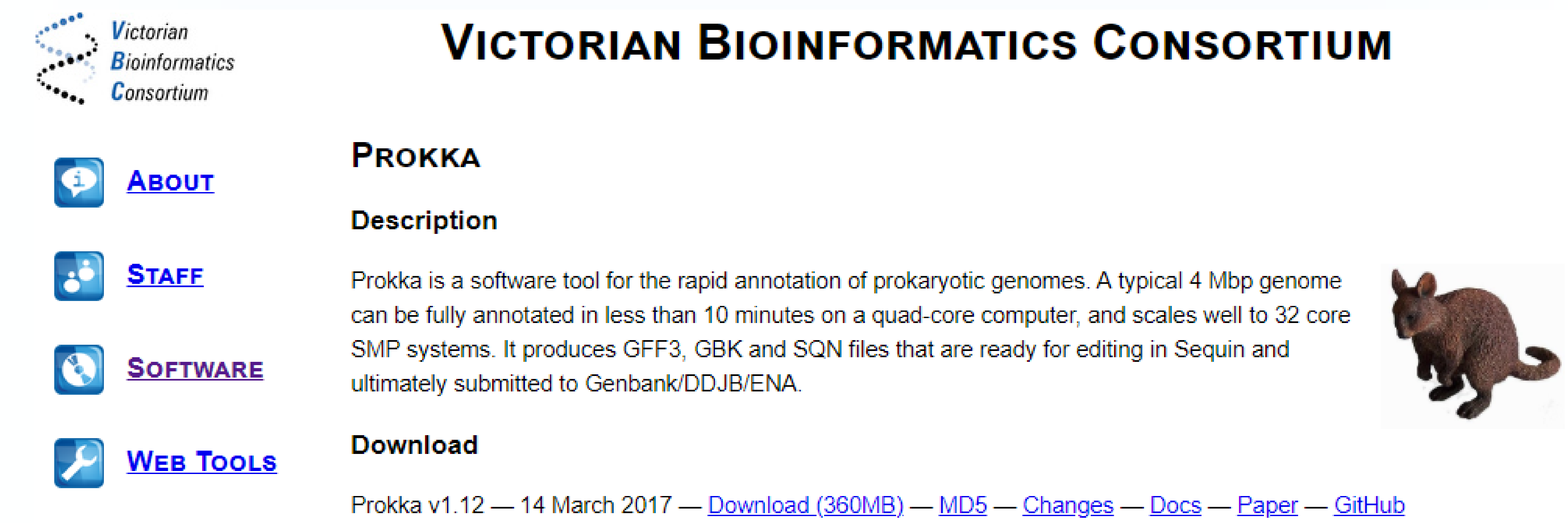
五. Cd-hit非冗余基因集(可选)

六. Salmon基因定量

七. 基因功能注释

http://wiki.biomine.skelleftea.se/wiki/index.php/Metagenomics

Prokka基因注释



Victorian Bioinformatics Consortium

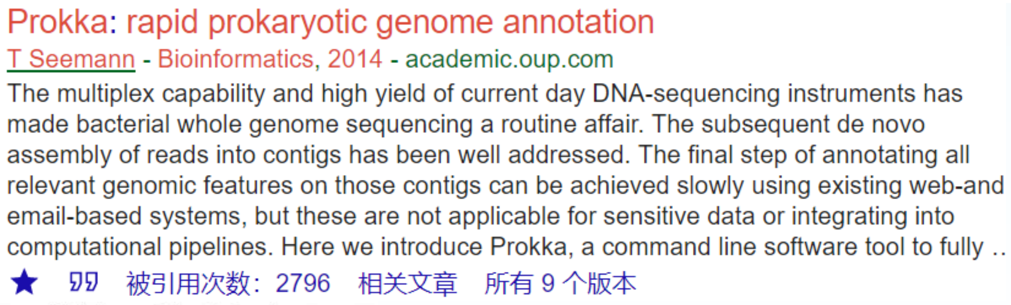
Prokka

Description

Prokka is a software tool for the rapid annotation of prokaryotic genomes. A typical 4 Mbp genome can be fully annotated in less than 10 minutes on a quad-core computer, and scales well to 32 core SMP systems. It produces GFF3, GBK and SQN files that are ready for editing in Sequin and ultimately submitted to Genbank/DDJB/ENA.

Download

Prokka v1.12 — 14 March 2017 — Download (360MB) — MD5 — Changes — Docs — Paper — GitHub



Prokka: rapid prokaryotic genome annotation

T Seemann - Bioinformatics, 2014 - academic.oup.com

The multiplex capability and high yield of current day DNA-sequencing instruments has made bacterial whole genome sequencing a routine affair. The subsequent de novo assembly of reads into contigs has been well addressed. The final step of annotating all relevant genomic features on those contigs can be achieved slowly using existing web-and email-based systems, but these are not applicable for sensitive data or integrating into computational pipelines. Here we introduce Prokka, a command line software tool to fully ..

截止2019年6月3日，Google Scholar统计引用2796次，神奇的软件只有一个作者。

http://www.vicbioinformatics.com/software.prokka.shtml

基因注释Prokka

- Prokka: rapid prokaryotic genome annotation
- Prokka是一个命令行软件工具，可以在一台典型台式机上在约10分钟内充分注释一个细菌基因组草图。它产生标准兼容的输出文件以进行进一步分析或者在基因组浏览器中查看。
- 2014年发表于Bioinformatics，最新版本1.12于2017年3月14日更新，大小360MB。因为它是一个复杂的分析流程，依赖关系众多。
- 安装：conda install prokka

Seemann, Torsten. "Prokka: rapid prokaryotic genome annotation." *Bioinformatics* 30.14 (2014): 2068-2069.
http://www.vicbioinformatics.com/software.prokka.shtml

3.2.1 prokka基因注释

```
# 查看文件大小，预估时间
ll temp/megahit/final.contigs.fa # 31 Mb
time prokka temp/megahit/final.contigs.fa --outdir temp/prokka \
  --prefix mg --metagenome --kingdom Archaea,Bacteria,Mitochondria,Viruses \
  --force --cpus 8
# 以mg开头，注释宏基因组，多界，强制覆盖输出
# 8线程，耗时40s, 1m56s, 29s
```

prokka结果说明

.gff: 基因注释文件，包括gff和序列，可用igv直接查看

.gbk: Genebank格式，来自gff

.fna: 输入contig核酸文件

.faa: 翻译CDS的AA序列

.ffn: 所有转录本核酸序列

.sqn: 用于提交的序列

.fsa: 输入序列，但有sqn的描述，用于tbl2asn生成sqn文件

.tbl: 特征表，用于tbl2asn生成sqn文件

.err: 错误报告

.log: 日志

.txt: 统计结果

.tsv: 所有注释基因特征表格

其它基因注释软件

- MetaGeneAnnotator　　conda install metagene_annotator

  http://metagene.nig.ac.jp/ 2006 NAR，2008 DNA Res，引用416和377次

- FragGeneScan　　conda install fraggenescan

  https://sourceforge.net/projects/fraggenescan/ 2010年发表于NAR，引用460次

- MetaGeneMark　　末被conda收录，有在线工具

  http://exon.gatech.edu/GeneMark/ 2010年发表于NAR，引用537次

Noguchi, Hideki, Jungho Park, and Toshihisa Takagi. "MetaGene: prokaryotic gene finding from environmental genome shotgun sequences." *Nucleic acids research* 34.19 (2006): 5623-5630.

Noguchi, Hideki, Takeaki Taniguchi, and Takehiko Itoh. "MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes." DNA research 15.6 (2008): 387-396.

Rho, Mina, Haixu Tang, and Yuzhen Ye. "FragGeneScan: predicting genes in short and error-prone reads." *Nucleic acids research* 38.20 (2010): e191-e191.

Zhu, Wenhan, Alexandre Lomsadze, and Mark Borodovsky. "Ab initio gene identification in metagenomic sequences." *Nucleic acids research* 38.12 (2010): e132-e132.

目录

一. KneadData质控(已完成)

二. Kraken物种分类

三. Megahit拼接

四. Prokka基因注释

五. **Cd-hit非冗余基因集(可选)**

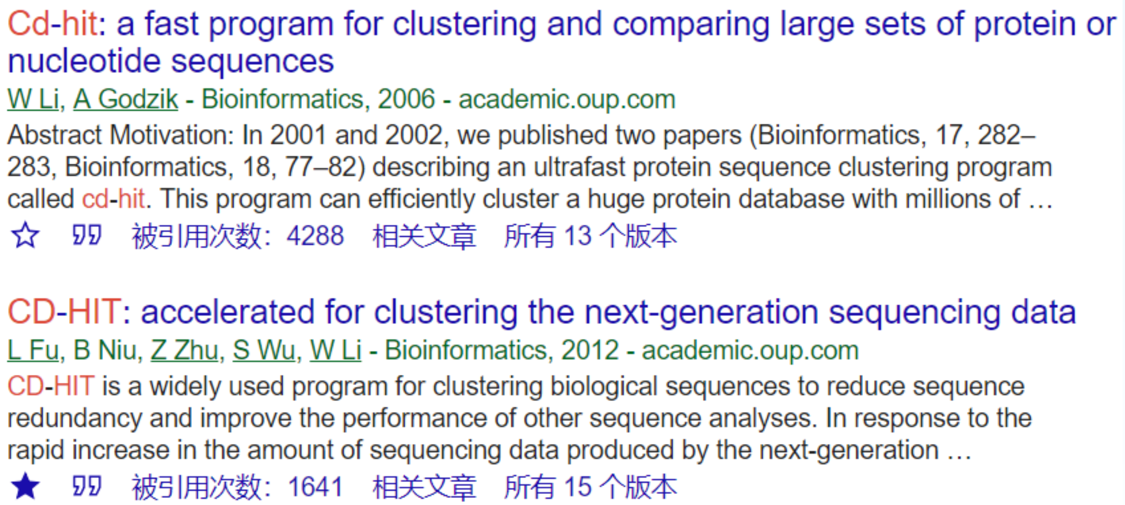
六. Salmon基因定量

七. 基因功能注释

http://wiki.biomine.skelleftea.se/wiki/index.php/Metagenomics

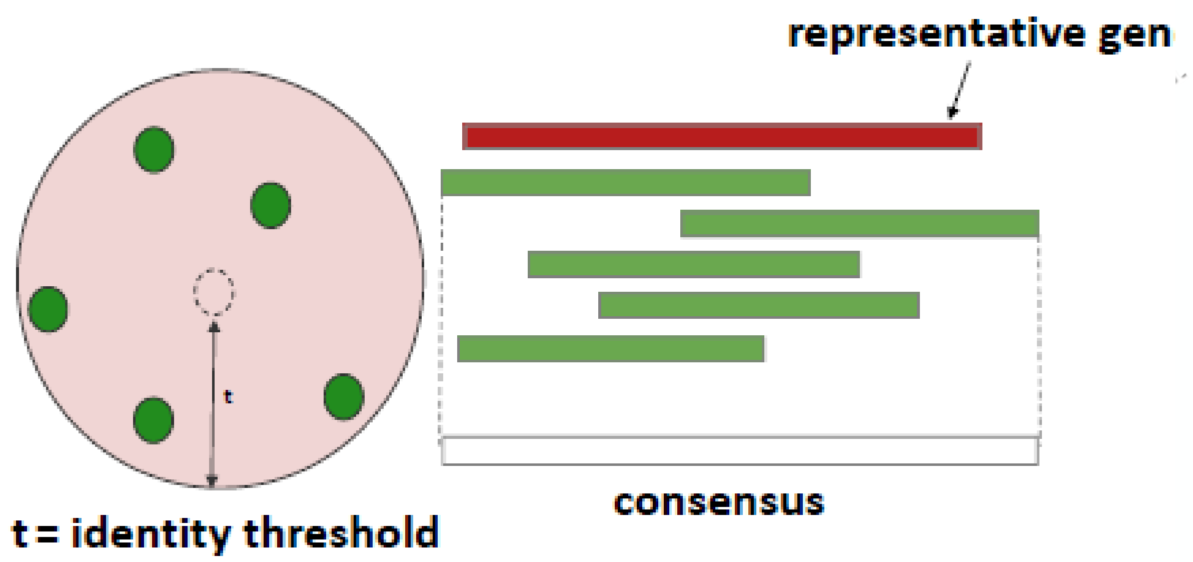
构建非冗余基因集(可选)

- 去除冗余基因、降低基因数量级
- 多样本、批量合并为一致参考序列Reference
- 通过CD-HIT将所有样本的基因序列根据序列相似性进行聚类，去除冗余序列(coverage > 90%, identity > 95%）。



Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences
W Li, A Godzik - Bioinformatics, 2006 - academic.oup.com
Abstract Motivation: In 2001 and 2002, we published two papers (Bioinformatics, 17, 282–283, Bioinformatics, 18, 77–82) describing an ultrafast protein sequence clustering program called cd-hit. This program can efficiently cluster a huge protein database with millions of …
☆ 99 被引用次数：4288 相关文章 所有 13 个版本

CD-HIT: accelerated for clustering the next-generation sequencing data
L Fu, B Niu, Z Zhu, S Wu, W Li - Bioinformatics, 2012 - academic.oup.com
CD-HIT is a widely used program for clustering biological sequences to reduce sequence redundancy and improve the performance of other sequence analyses. In response to the rapid increase in the amount of sequencing data produced by the next-generation …
★ 99 被引用次数：1641 相关文章 所有 15 个版本



主页：http://weizhongli-lab.org/cd-hit/ 软件下载、在线服务器、帮助

CD-HIT家族

小工具	功能	应用
cd-hit	按指定相似度聚类蛋白质序列	非冗余蛋白集构建，如UniRef
cd-hit-est	按指定相似度聚类核酸序列	非冗余基因集构建、重复序列家族分析、聚类OTUs
cd-hit(-est)-2d	两个数据库比对	多批量、来源宏基因组构建非冗余基因集
cd-hit-dup	从Illumina单双端序列中鉴定冗余	对高通量测序数据的双端去冗余
cd-hit-otu	16S序列聚类	早期OTU鉴定方法
cd-hit-lap	鉴定重叠序列	我没用过

3.2.2 cd-hit构建非冗余基因集

```
# 输入文件可由多个样本、组、批次序列合并文件，方便整合分析
# aS覆盖度，c相似度，G局部比对，M内存0不限制，T多线程，g最优解
time cd-hit-est -i temp/prokka/mg.ffn -o temp/NR/mg.ffn.nr \
  -aS 0.9 -c 0.95 -G 0 -M 0 -T 9 -g 1
ln temp/NR/mg.ffn.nr result/NR/nucleotide.fa
# 翻译核酸为对应蛋白序列
transeq -sequence result/NR/nucleotide.fa -outseq result/NR/protein.fa
# 序列名自动添加了_1，为与核酸对应要去除
sed -i 's/_1 / /' result/NR/protein.fa
```

cd-hit对注释的基因进行聚类

目录

一. KneadData质控(已完成)

二. Kraken物种分类

三. Megahit拼接

四. Prokka基因注释

五. Cd-hit非冗余基因集(可选)

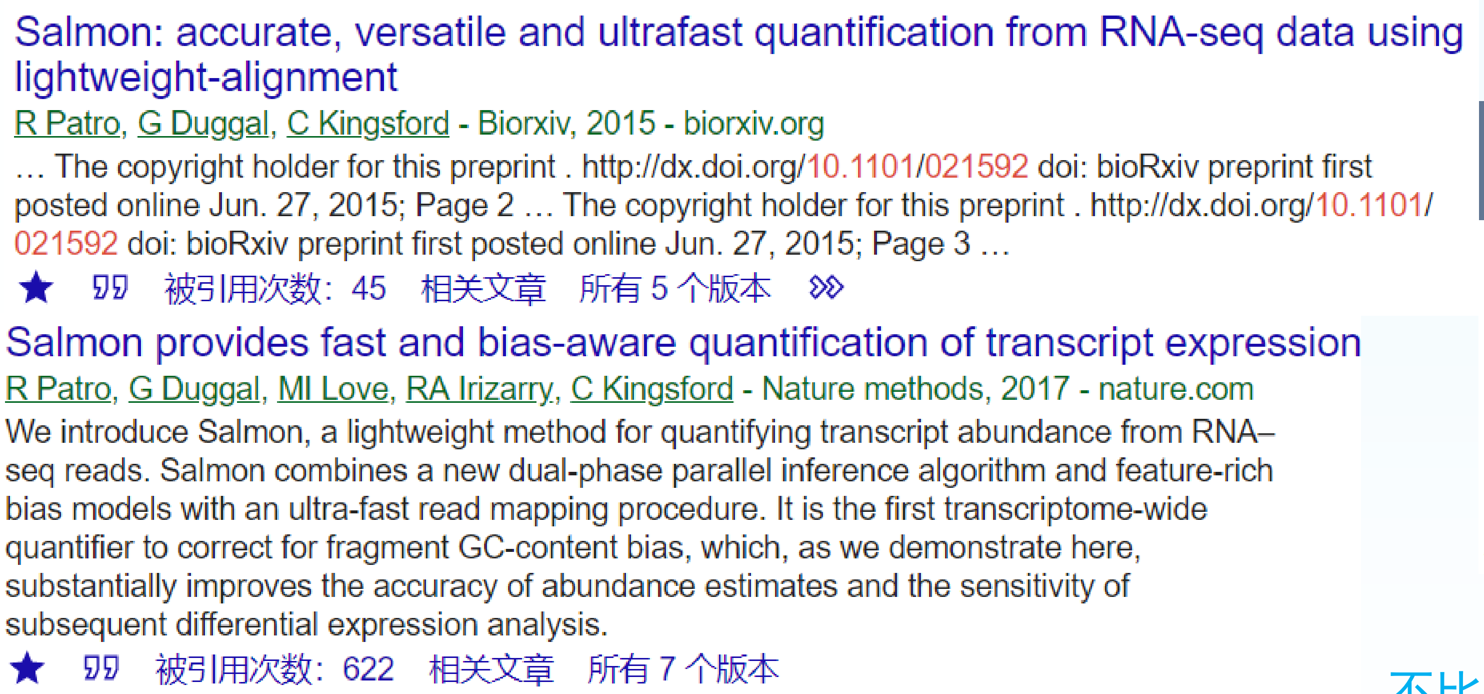
六. **Salmon基因定量**

七. 基因功能注释

http://wiki.biomine.skelleftea.se/wiki/index.php/Metagenomics

Salmon非比对定量

- Salmon(三文鱼)是一款新的、极快的转录组计数软件。它与Kallisto(熊神星)和Sailfish(旗鱼)类似，可以不通过mapping而获得基因的counts值。Salmon的结果可由edgeR / DESeq2等进行counts值的下游分析。



Salmon: accurate, versatile and ultrafast quantification from RNA-seq data using lightweight-alignment

R Patro, G Duggal, C Kingsford - Biorxiv, 2015 - biorxiv.org

… The copyright holder for this preprint . http://dx.doi.org/10.1101/021592 doi: bioRxiv preprint first posted online Jun. 27, 2015; Page 2 … The copyright holder for this preprint . http://dx.doi.org/10.1101/021592 doi: bioRxiv preprint first posted online Jun. 27, 2015; Page 3 …

★ ❞ 被引用次数：45 相关文章 所有 5 个版本 »

Salmon provides fast and bias-aware quantification of transcript expression

R Patro, G Duggal, MI Love, RA Irizarry, C Kingsford - Nature methods, 2017 - nature.com

We introduce Salmon, a lightweight method for quantifying transcript abundance from RNA–seq reads. Salmon combines a new dual-phase parallel inference algorithm and feature-rich bias models with an ultra-fast read mapping procedure. It is the first transcriptome-wide quantifier to correct for fragment GC-content bias, which, as we demonstrate here, substantially improves the accuracy of abundance estimates and the sensitivity of subsequent differential expression analysis.

★ ❞ 被引用次数：622 相关文章 所有 7 个版本

nature methods

Brief Communication | Published: 06 March 2017

Salmon provides fast and bias-aware quantification of transcript expression

Rob Patro ✉, Geet Duggal, Michael I Love, Rafael A Irizarry & Carl Kingsford ✉

Nature Methods **14**, 417–419 (2017) | Download Citation ⤓

不比对快速估计基因丰度Salmon
https://www.nature.com/articles/nmeth.4197

Salmon安装和基本功能

- conda install salmon # 安装
- samlon -h # 查看帮助
- salmon v0.11.2，主要提供以下5类功能

  index Create a salmon index # 建索引

  quant Quantify a sample # 样本定量

  alevin single cell analysis # 单细胞分析

  swim Perform super-secret operation

  quantmerge Merge multiple quantifications into a single file # 合并样本结果

3.2.3 基因定量——建索引

```
# 创建索引目录
mkdir -p temp/salmon

# 建索引,-t转录本，--type类型fmd/quasi
# -k kmer长度默认31, -i 索引
salmon index -t result/NR/nucleotide.fa \
    -p 9 --type quasi -k 31 \
    -i temp/salmon/index
```

3.2.3 基因定量——定量

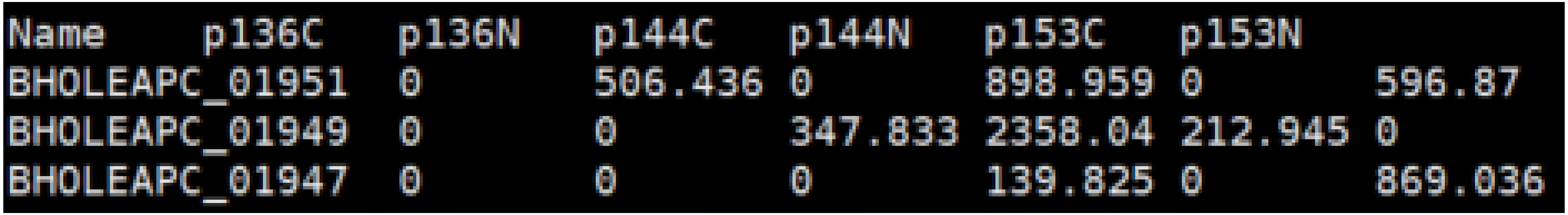
```
# 定量，l文库类型自动选择，p线程，--meta宏基因组模式
time parallel -j 3 \
  'salmon quant -i temp/salmon/index -l A -p 3 --meta \
  -1 temp/qc/{1}_1_kneaddata_paired_1.fastq -2 temp/qc/{1}_1_kneaddata_paired_2.fastq \
  -o temp/salmon/{1}.quant' \
  ::: `tail -n+2 result/design.txt | cut -f 1`
```

3.2.3 基因定量——合并为丰度矩阵

```
mkdir -p result/salmon
# 合并百万分比TPM
salmon quantmerge --quants temp/salmon/*.quant \
  -o result/salmon/gene.TPM
# 合并原始reads count值
salmon quantmerge --quants temp/salmon/*.quant \
  --column NumReads -o result/salmon/gene.count
sed -i '1 s/.quant//g' result/salmon/gene.*
```

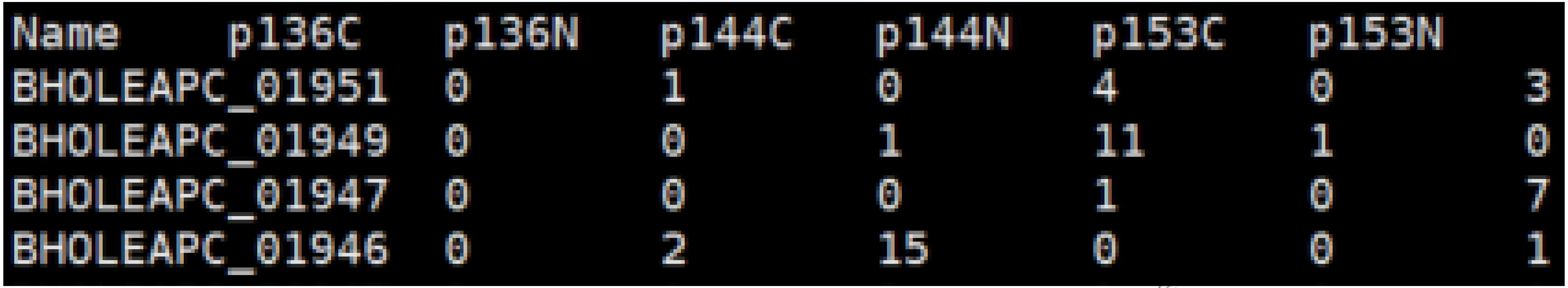
基因定量完成

- 基因丰度矩阵(TPM)，可下游STAMP、LEfSe、limma、t.test等统计



```
Name    p136C   p136N   p144C   p144N   p153C   p153N
BHOLEAPC_01951  0       506.436 0       898.959 0       596.87
BHOLEAPC_01949  0       0       347.833 2358.04 212.945 0
BHOLEAPC_01947  0       0       0       139.825 0       869.036
```

- 基因定量原始reads counts，下游接edgeR、DESeq2统计差异



```
Name    p136C   p136N   p144C   p144N   p153C   p153N
BHOLEAPC_01951  0       1       0       4       0       3
BHOLEAPC_01949  0       0       1       11      1       0
BHOLEAPC_01947  0       0       0       1       0       7
BHOLEAPC_01946  0       2       15      0       0       1
```

总结

- Megahit快速组装，节省计算资源；metaSPAdes精细拼接，但内存和时间消耗大；拼接长度和错误率也成正比，N50提高也伴随时嵌合体升高风险；
- Quast快速评估常用组装指标，提供txt / html / pdf格式报告；metaquast基于参考数据库进行更细致的评估；
- Prokka提供了基因鉴定、注释一站式解决方案，依赖关系多也容易报错；
- Cd-hit用于多基因集合并，建立非冗余基因集，方便开展多样品定量、比较；
- Salmon基于k-mer的非比对定量方法：快速，准确，节约空间；主要分为建索引，定量和合并3步。非比对方法没有序列比对中间文件。
- 软件更新快，使用出问题时，需要正对照。如果软件没问题，可能是版本不对，查看帮助，修改为最新格式和参数即可。新版本有问题也可安装教程中指定版本。

参考资源

- 宏基因组公众号文章目录
- 生信宝典公众号文章目录
- 加拿大生信网 https://bioinformatics.ca/
- 加拿大生信网宏基因组课程中文版——挖掘微生物组生物标记
- 美国高通量开源课程 https://github.com/ngs-docs
- The Huttenhower Lab http://huttenhower.sph.harvard.edu/
- 微生物组助手 https://github.com/LangilleLab/microbiome_helper
- Susan Holmes http://statweb.stanford.edu/~susan/

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