

Instruction

ABACAS用于根据参考基因组快速快速连接(align, order, orientate), 查看并设计引物用于contigs之间gaps的填充。根据参考基因组, 使用mummer查询比对位置并识别共线性。输出结果使用N代表contigs之间的重叠和gaps, contigs之间的重叠重由于contigs末端的低质量以及低复杂度区域。同时, ABACAS能自动提取gaps, 并设计引物用于gaps填充, 设计引物的特异性可通过mummer比对查看。

Usage

```
abacas -r <reference> -q <contigs> -p <nucmer|promer>
```

ABACAS将会输出多个不同结果文件, 使用ACT进行比较。

-c, 参考序列为环状

-i, 最小一致性比对, 默认40%, 值越小, 比对contigs更多

-v, 最小contigs覆盖度, 默认40%

-V, 最小contigs覆盖度差异, 0表示contigs多重比对时, 随机放置多重contigs

-t, 针对未使用的contigs(.bin文件中), 采用tblastx比对; 使用-b生成该文件

-N, 不使用N对contigs间的重叠以及gaps进行填充

使用默认参数将会输出以下文件:

经过排序和排列方向的序列文件(reference_query.fasta or prefix.fasta), 如果发现contigs比对到了反向链, 那么.fasta文件中的contigs为反向互补排列(输出中所有对应的定向为“-”都经过了个反向互补输出), 然而ACT中将显示contigs初始方向。

特征文件(reference_query.tab or prefix.tab)

Bin文件, 包含没有用于排序的contigs(reference_query.bin or prefix.bin)

比较文件(reference_query.crunch or prefix.crunch)

Gap信息(reference_query.gaps or prefix.gaps)

关于有比对信息但是没有用于排序的contigs信息(unused_contigs.out)

-m, 输出经过排序和重定向的多重fasta格式文件

-b, 输出包含未比对的contigs的多重fasta文件

-g, 输出reference上的多重fasta文件对应gaps

Output

```
1 100 100 1 99526 NODE_22_length_99526_cov_55.320033 21078 120603 unknown NONE
2 100 100 104507 191398 NODE_24_length_86892_cov_55.495465 125584 212475 unknown NONE
3 100 100 196379 201120 NODE_61_length_4742_cov_51.740628 217456 222197 unknown NONE
4 100 100 201403 236394 NODE_36_length_34992_cov_54.860376 222480 257471 unknown NONE
5 100 100 238786 239907 NODE_90_length_1122_cov_390.503518 259863 260984 unknown NONE
6 100 100 241483 370991 NODE_16_length_129509_cov_54.359339 262560 392068 unknown NONE
7 100 100 372184 564452 NODE_13_length_192269_cov_53.771742 393261 585529 unknown NONE
8 72 100 564552 567426 NODE_69_length_2875_cov_45.612445 583817 586691 unknown NONE
```

Comparison file(.crunch file): 7列对应, 覆盖度, 一致性, pseudomolecule起点, 终点, contig ID, reference起点, 终点

| | | | | | | | |
|---|-----|------|--------|--------|--------|--------|-----------------|
| 1 | Gap | 4980 | 99527 | 104506 | 120604 | 125583 | NON-Overlapping |
| 2 | Gap | 4980 | 191399 | 196378 | 212476 | 217455 | NON-Overlapping |
| 3 | Gap | 282 | 201121 | 201402 | 222198 | 222479 | NON-Overlapping |
| 4 | Gap | 2391 | 236395 | 238785 | 257472 | 259862 | NON-Overlapping |
| 5 | Gap | 1575 | 239908 | 241482 | 260985 | 262559 | NON-Overlapping |
| 6 | Gap | 1192 | 370992 | 372183 | 392069 | 393260 | NON-Overlapping |
| 7 | Gap | 99 | 564453 | 564551 | 585530 | 583816 | Overlapping |
| 8 | Gap | 939 | 567427 | 568365 | 586692 | 587630 | NON-Overlapping |

Gap file(.gaps): 2-6列表, gap大小, pseudomolecule起点, 终点, reference起点, 终点

```
1 NODE_65_length_3137_cov_102.733555
2 NODE_66_length_3035_cov_47.143054
3 NODE_77_length_1723_cov_41.990602
4 NODE_142_length_515_cov_256.404639
5 NODE_51_length_10187_cov_563.764513
6 NODE_42_length_25237_cov_48.201235
7 NODE_104_length_813_cov_1463.790087
8 NODE_45_length_20420_cov_98.846745
```

Bin file(.bin): 未能用于生成pseudomolecule的contigs名称

```
1 >ordered_NC_016845.1
2 GCGGTGGTCCCACCTGACCCCATGCCGAAGTGAACGCCGTAGCGCCGATGGT
3 AGTGTGGGGTCTCCCCATGTGAGAGTAGGGAAGTCCAGGCATCAAACAAGTGAAGAGCC
4 CATCCTGTGCGATGGGCTTTTTGCGTTTCTGTCGTCGCGCCAGCGTGCCGGCATGTCCCC
5 CCGGTGGCGCTGCGCTTGCCGGGGCTACGGATCCGACACGGGGCTACGTCCCTGTGGCCG
6 GACAAGGTGCGCAGCACCGCCTCCGGGAAAATCCCTCCTATATCACCGTCGATAAATGAC
7 TATTGCCTCTCTAACCAACTCACAATAAATGTCGCGATCTGTGGTACATCATTTAAATCC
8 AGCAGCGGAAGTGAGGTCGCGATCGATACATCGCTGGCCACGGCGATGGTATGCTCATCA
```

经过排序和重定向后contigs生成的pseudomolecule文件, 重叠contigs使用100个N进行分割, 同时gaps也使用N表示

```
abacas -v 5 -c -N -m -g ref_query_gaps -b -r ASM24018v2_chromosomal_seq.fna -q
39401_scaffolds_filtered_500_30.fasta -p nucmer -o ref_query_5
```

To view your results in ACT

```
Sequence file 1: NC_016845.1.fna
Comparison file 1: NC_016845.1._39401.crunch
Sequence file 2: NC_016845.1._39401.fasta
```

ACT feature file is: NC_016845.1._39401.tab

Contigs bin file is: NC_016845.1._39401.bin

Gaps in pseudomolecule are in: NC_016845.1._39401.gaps