

# Introdução à metagenômica: curso prático

Montagem de metagenomas

Em um mundo ideal...



# Na realidade...



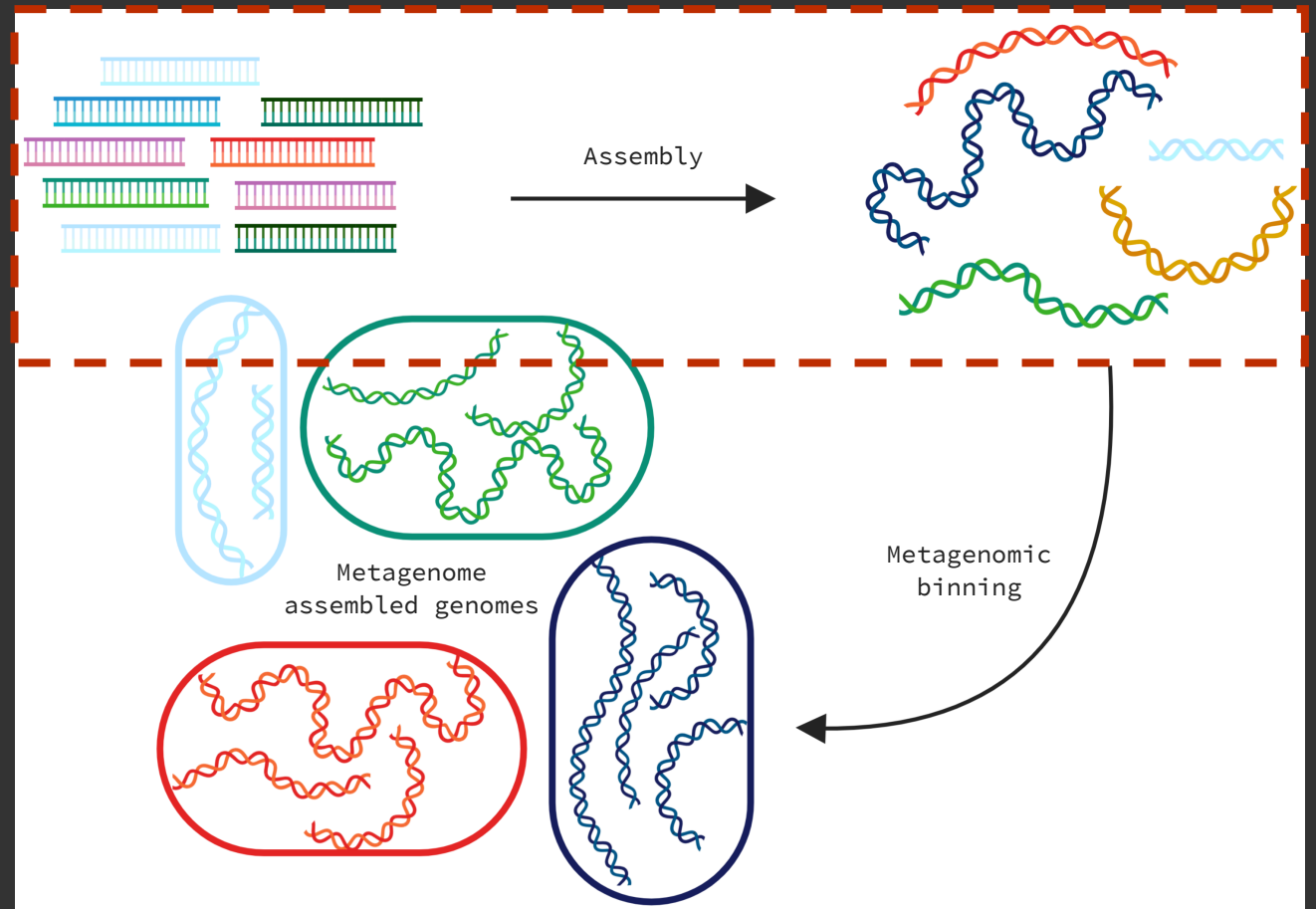
**Fragmentação**

# De sequências à contigs: montagem de novo

Obtenção de sequências  
contíguas (*contigs*)

Montagem de  
genomas vs. metagenomas

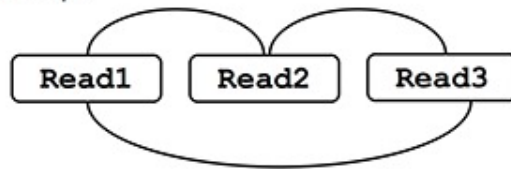
- # Comunidades complexas e pouco uniformes
- # Ausência de genomas de referência



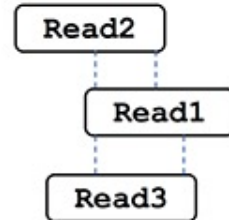
# Estratégias de montagem

(a) Overlap, Layout, Consensus assembly

(i) Find overlaps



(ii) Layout reads



(iii) Build consensus

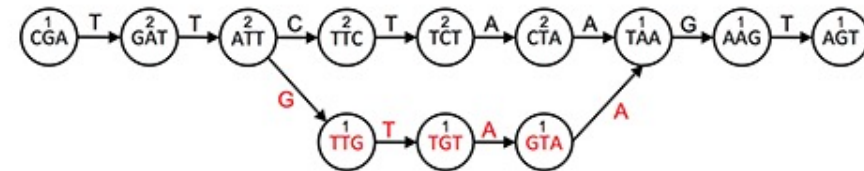
```
CGATTCTA
  TTCTAAGT
  GATTGTAA
  -----
CGATTCTAAGT
```

(b) De Bruijn graph assembly

(i) Make kmers

|                        |                        |                        |
|------------------------|------------------------|------------------------|
| <b>Read1: TTCTAAGT</b> | <b>Read2: CGATTCTA</b> | <b>Read3: GATTGTAA</b> |
| Kmers: TTC             | Kmers: CGA             | Kmers: GAT             |
| TCT                    | GAT                    | ATT                    |
| CTA                    | ATT                    | TTG                    |
| TAA                    | TTC                    | TGT                    |
| AAG                    | TCT                    | GTA                    |
| AGT                    | CTA                    | TAA                    |

(ii) Build graph



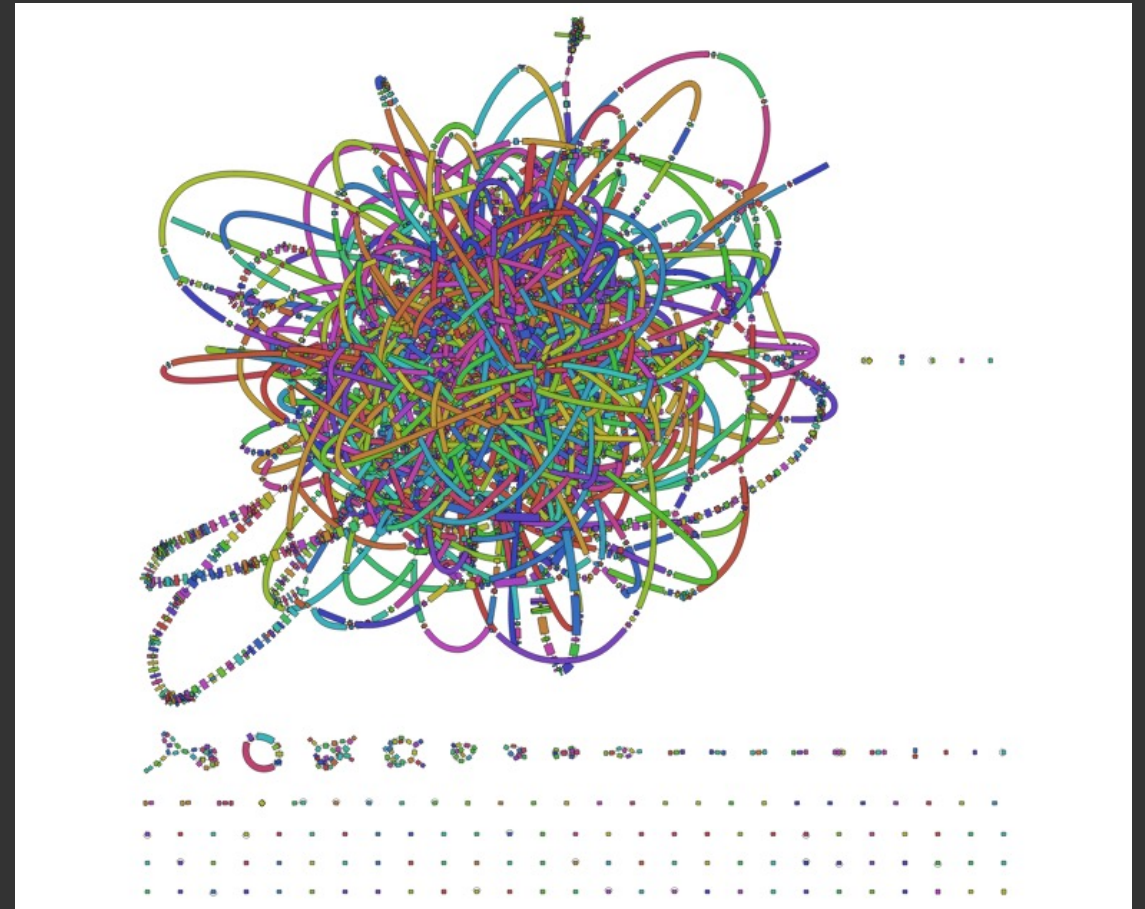
(iii) Walk graph and output contigs



*Ayling et al. (2020)*

# Qual valor de kmer utilizar?

k = 51



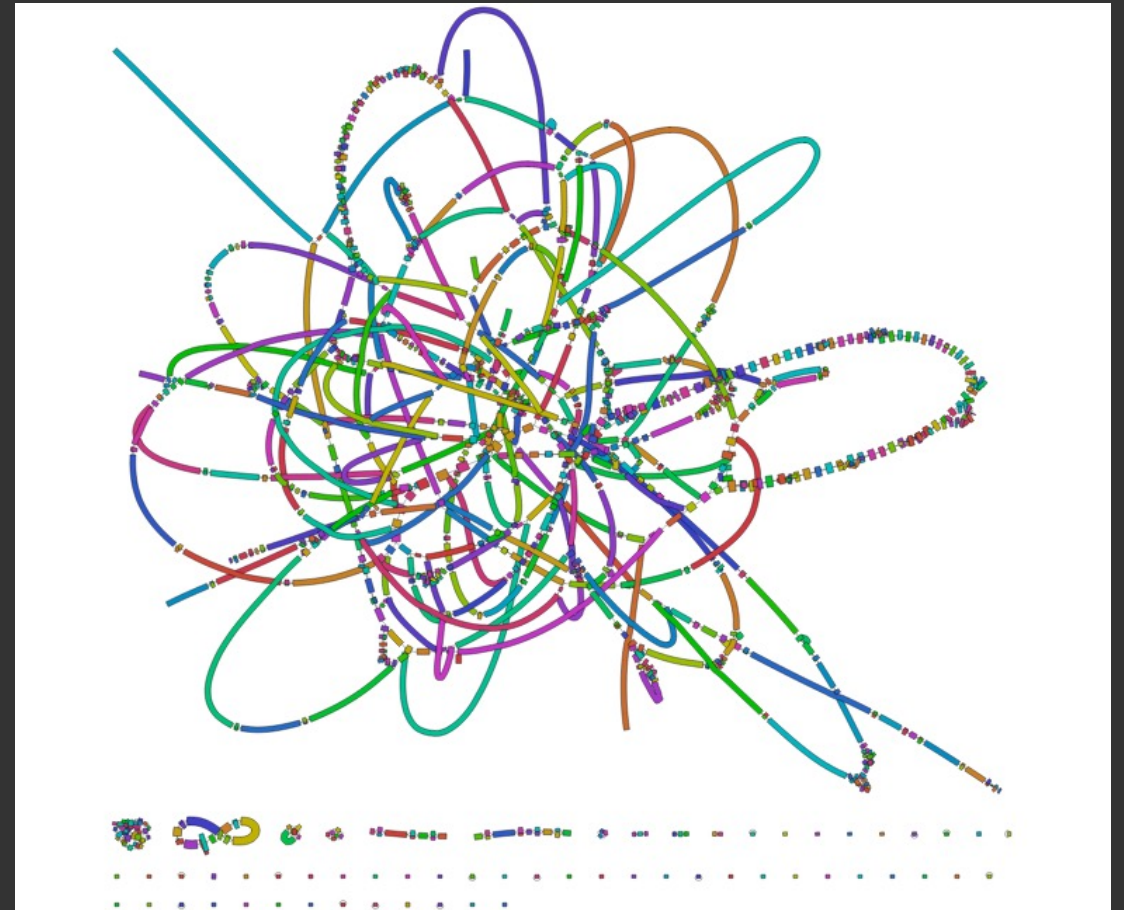
[github.com/rrwick/Bandage/wiki/Effect-of-kmer-size](https://github.com/rrwick/Bandage/wiki/Effect-of-kmer-size)



# Qual valor de kmer utilizar?

k = 51

k = 61



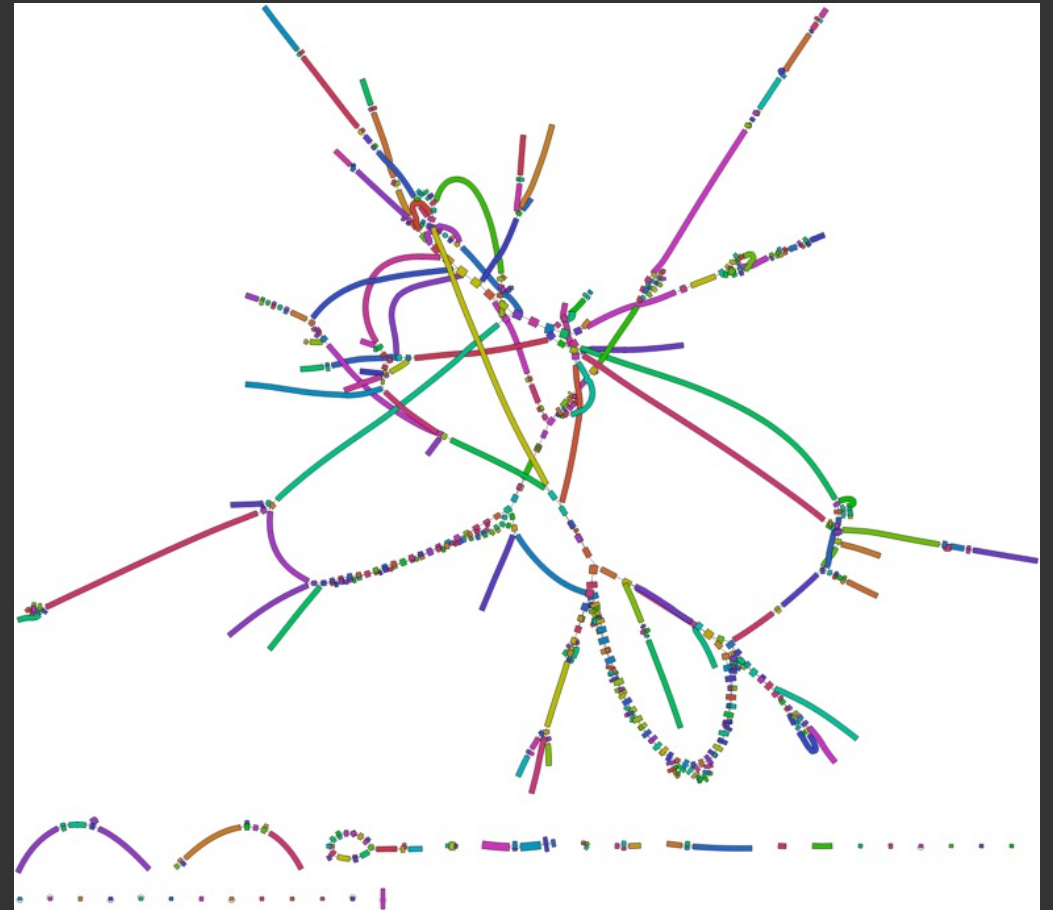
[github.com/rrwick/Bandage/wiki/Effect-of-kmer-size](https://github.com/rrwick/Bandage/wiki/Effect-of-kmer-size)

# Qual valor de kmer utilizar?

k = 51

k = 61

k = 71



[github.com/rrwick/Bandage/wiki/Effect-of-kmer-size](https://github.com/rrwick/Bandage/wiki/Effect-of-kmer-size)



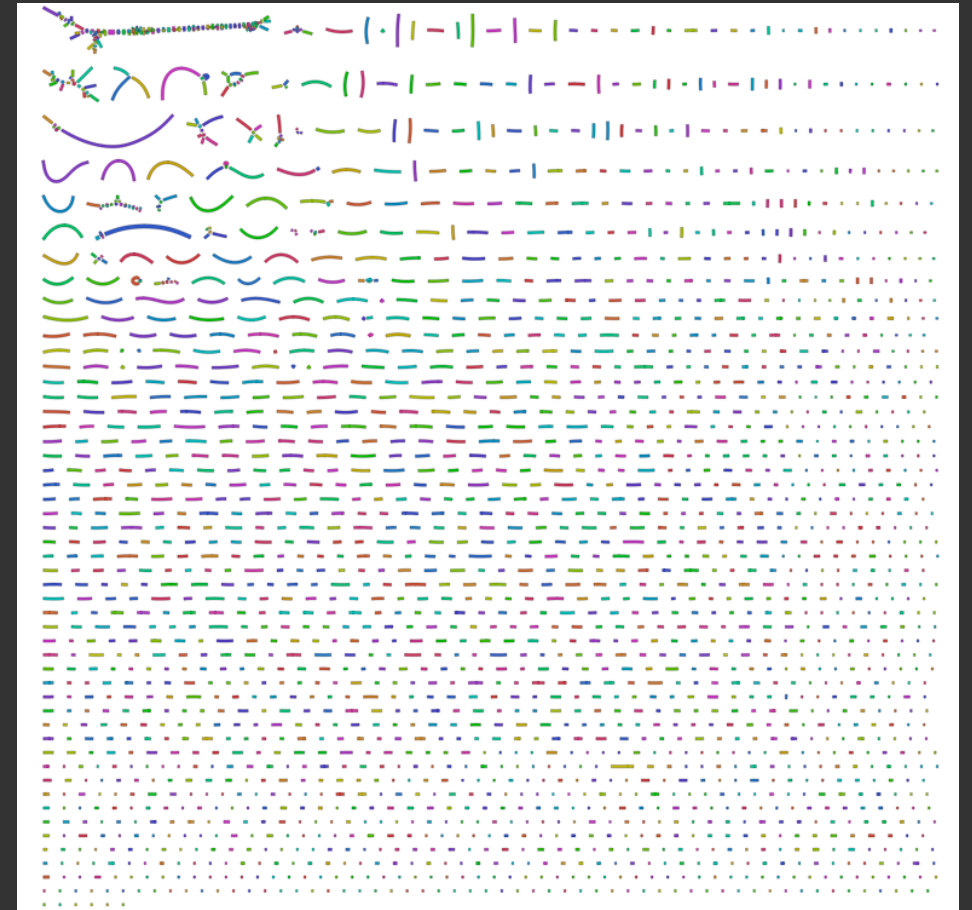
# Qual valor de kmer utilizar?

k = 51

k = 61

k = 71

k = 91



[github.com/rrwick/Bandage/wiki/Effect-of-kmer-size](https://github.com/rrwick/Bandage/wiki/Effect-of-kmer-size)

# Qual valor de kmer utilizar?

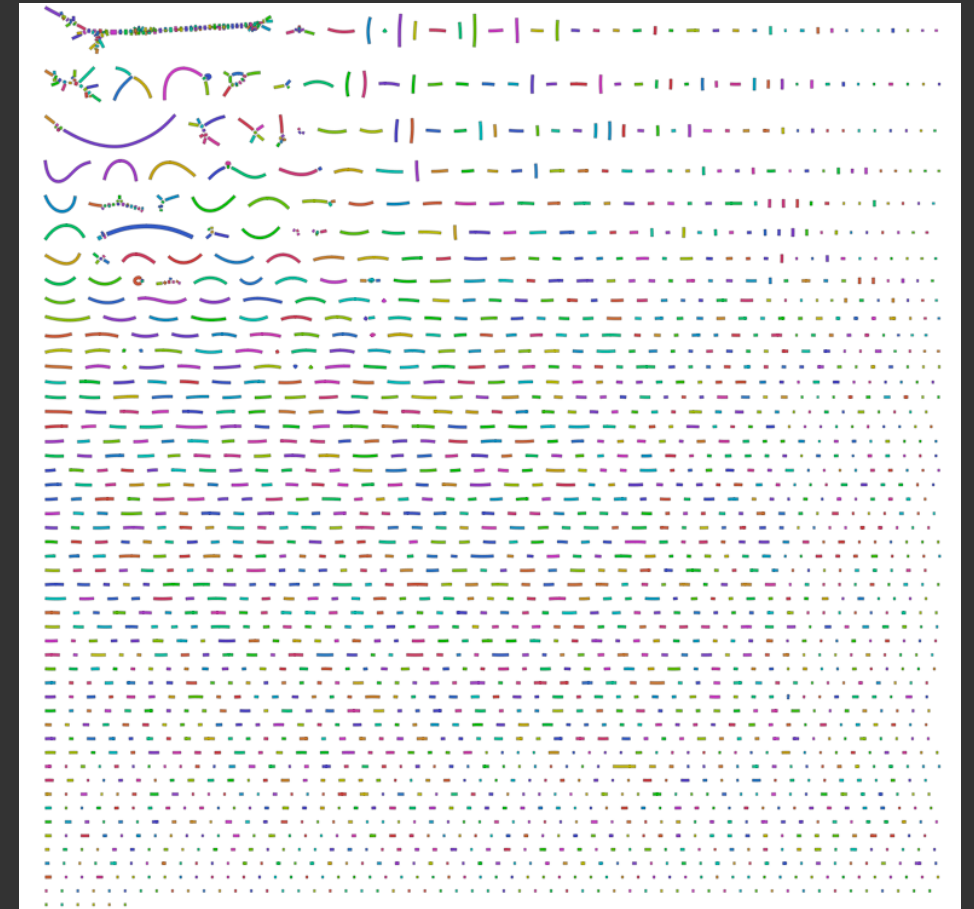
## Dicotomia

- # kmers pequenos → contigs curtos com muitas conexões
- # kmers grandes → contigs longos com menos conexões

## Fatores que afetam o tamanho de kmer ideal

- # Tamanho das sequências
- # Volume de sequenciamento
- # Complexidade das sequências

Que tal o uso de múltiplos kmer?



[github.com/rrwick/Bandage/wiki/Effect-of-kmer-size](https://github.com/rrwick/Bandage/wiki/Effect-of-kmer-size)

# Diferentes programas para montagem de sequências curtas

**Table 1** Assembly statistics and computational requirements for assembly of the Tara Oceans metagenome. Time required is given in seconds, minutes and hours for illustrative purposes and memory in GB of RAM required

|  | Tara Ocean |             |             |             |            |            |           |             |            |
|--|------------|-------------|-------------|-------------|------------|------------|-----------|-------------|------------|
|  | CLC        | IDBA-UD     | MEGAHIT     | metaSPAdes  | MetaVelvet | Omega      | Ray Meta  | SPAdes      | Velvet     |
| Number of contigs ( $\geq 500$ bp)         | 50,716     | 163,815     | 216,938     | 185,419     | 67,161     | 15,982     | 6128      | 220,178     | 57,816     |
| Total length                               | 46,069,409 | 179,686,756 | 210,621,485 | 202,770,058 | 55,972,515 | 34,861,819 | 7,277,214 | 275,920,632 | 45,425,460 |
| No. of long contigs ( $\geq 1$ kbp)        | 10,720     | 50,498      | 56,243      | 48,640      | 12,590     | 13,305     | 2179      | 70,711      | 8802       |
| No. of ultra-long contigs ( $\geq 50$ kbp) | 0          | 2           | 1           | 37          | 0          | 9          | 0         | 54          | 0          |
| Largest contig                             | 39,748     | 101,400     | 62,649      | 141,519     | 30,177     | 102,255    | 41,443    | 197,381     | 21,980     |
| <i>N50</i>                                 | 880        | 1166        | 982         | 1124        | 805        | 2691       | 1329      | 1415        | 749        |
| <i>L50</i>                                 | 14,113     | 38,236      | 58,246      | 39,033      | 21,544     | 2737       | 1345      | 39,617      | 19,631     |
| Mapping rate (%)                           | 38.98      | 52.24       | 55.92       | 64.03       | 4117       | 13.64      | 8.25      | 64.46       | 48.19      |
| Time (seconds)                             | 3527       | 69,782      | 10,455      | 125,862     | 2527       | 168,213    | 16,419    | 80,039      | 2342       |
| Time (minutes)                             | 58.78      | 1163.03     | 174.25      | 2097.70     | 42.12      | 2803.55    | 273.65    | 1333.98     | 39.03      |
| Time (hours)                               | 0.98       | 19.38       | 2.90        | 34.96       | 0.70       | 46.73      | 4.56      | 22.23       | 0.65       |
| Memory required (GB)                       | 16.23      | 42.84       | 10.58       | 66.53       | 109.37     | 30.7       | 42        | 157.75      | 109.37     |

van der Walt et al. (2017)

# Questões abertas a respeito de montagem de metagenomas

Tamanho de kmer

Co-montagem ou não?

Falta de referências: como saber se uma montagem é boa ou não

## # Desafios

- # Regiões com repetições

- # Transferência horizontal

- # Elementos extra-cromossomais