N *unalighned FASTA sequences*
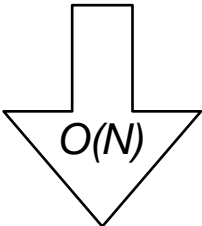
Reference Sequence, e.g. **HXB2 pr_rt**
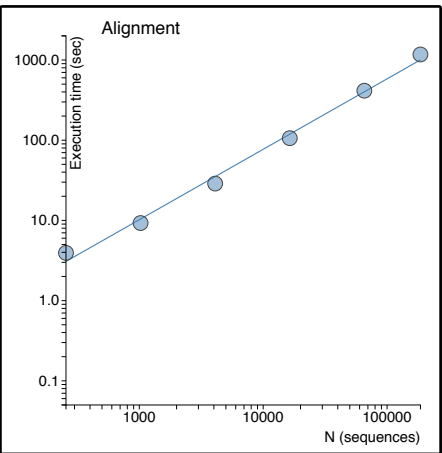
```
CCTCAGATCACTCTTTGGCAACGACCMC...
CAAATCACTCTTTGGCAACGACCTCTTG...
ATCACTCTTTGGCAGCGACCACTCGTCC...
TCACTCTTTGGCAACGACCCCTTGTCAC...
```

```
CCTCAGGTCACTCTTTGGCAACGACCCCTC...
```

Map each read to reference
with `bealign`

$O(N)$

Alignment — Execution time (sec) vs N (sequences)

Aligned FASTA sequences

```
CCTCAGATCACTCTTTGGCAACGACCMCTCGTCACAATAAAGATAGGGGGGCAATTAAAG
---CAAATCACTCTTTGGCAACGACCTCTTGTCACAGTAAAAATAGCAGGACAGCTAAAA
------ATCACTCTTTGGCAGCGACCACTCGTCCCAATAAGGATAGGGGGGCAACTAAAG
-------TCACTCTTTGGCAACGACCCCTTGTCACAGTAAAAATAGGAGGACAGATGATA
```

Compute pairwise distances
with `tn93,` *reporting those*
under 0.02

$O(N^2)$

Distances — Execution time (sec) vs N (sequences)
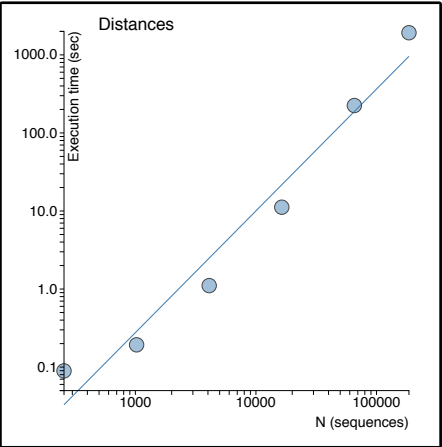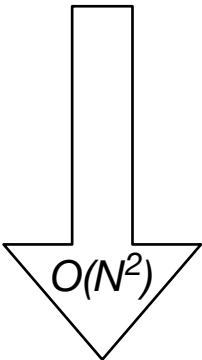
Distances between pairs of sequences, up to a threshold

ID1	ID2	Distance
B_US_KT167882_1985	B_US_GQ209649_2002	0.0152805
07_BC_CN_KT379202_2011	07_BC_CN_KR187686_2011	0.0100482
07_BC_CN_KT379202_2011	07_BC_CN_KR187866_2012	0.0110194
01_AE_CN_KX583284_2014	01_AE_CN_KU050634_2007	0.0186584
07_BC_CN_KR187686_2011	07_BC_CN_KR187866_2012	0.0124912

Construct and describe the
network using
`hivnetworkcsv`

$\sim O(N)$

Network inference — Execution time (sec) vs N (sequences)
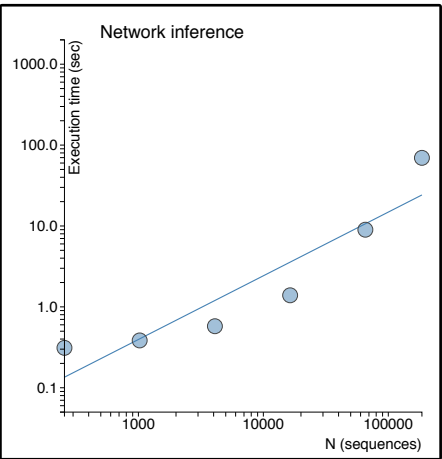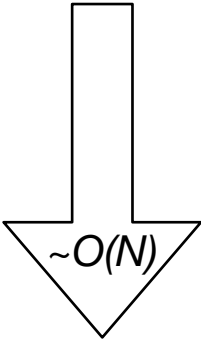
Network file (JSON), network visualization

```
…
Nodes": [
    {
        "attributes": [],
        "baseline": null,
        "cluster": 1,
        "edi": null,
        "id": "B_JP_AB868660_2010"
    },
…
```

Network Statistics

Clusters	1396
Edges	22138
Nodes	5466
Sequences used to make links	5466
Links/node	
Mean	8.1
Median	1.0
Range	1 - 247
Interquartile range	1 - 4
Cluster sizes	
Mean	3.9
Median	2.0
Range	2 - 361
Interquartile range	2 - 3

Genetic distances among linked nodes.