Minimizer VS. Open Syncmer sketching

Minimizer: k=5, w=4

A T C G T G A T

A T C G T

T C G T G

C G T G A

G T G A T
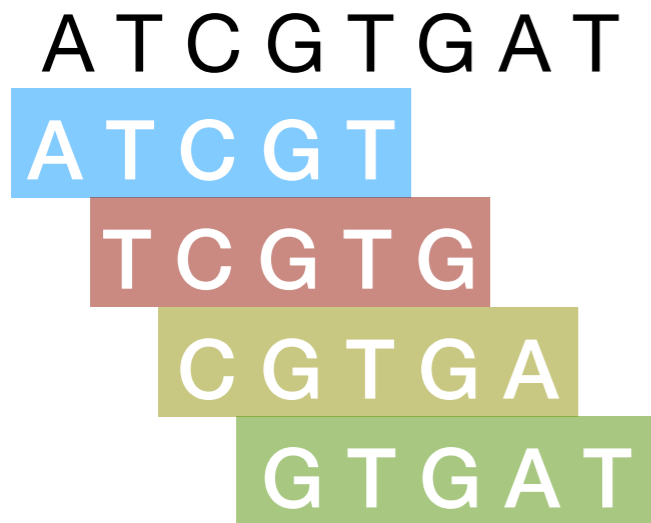
Label kmer by some ordering
(e.g., random hashing)

⟶

3 1 4 2

⟶

Select smallest kmer in the ordering

T C G T G

Open Syncmer: k=5, s=3, t=2

For each kmer above:

A T C G T

A T C

T C G

C G T

Label s-kmer by some ordering
(e.g., random hashing)

⟶

3 2 1

⟶

If t =3, then this kmer
will be selected;

If t≠3, this kmer will
not be selected

$$OS(A^k) = \{x \in A^k \ \ s.t. \ \ \arg\min_{m\in[1,k-s+1]} h(x[m:m+s-1]) = t\}$$

$$2 \leq s < k; 1 \leq t \leq s$$

- Open Syncmer is more sensitive to distantly related sequences