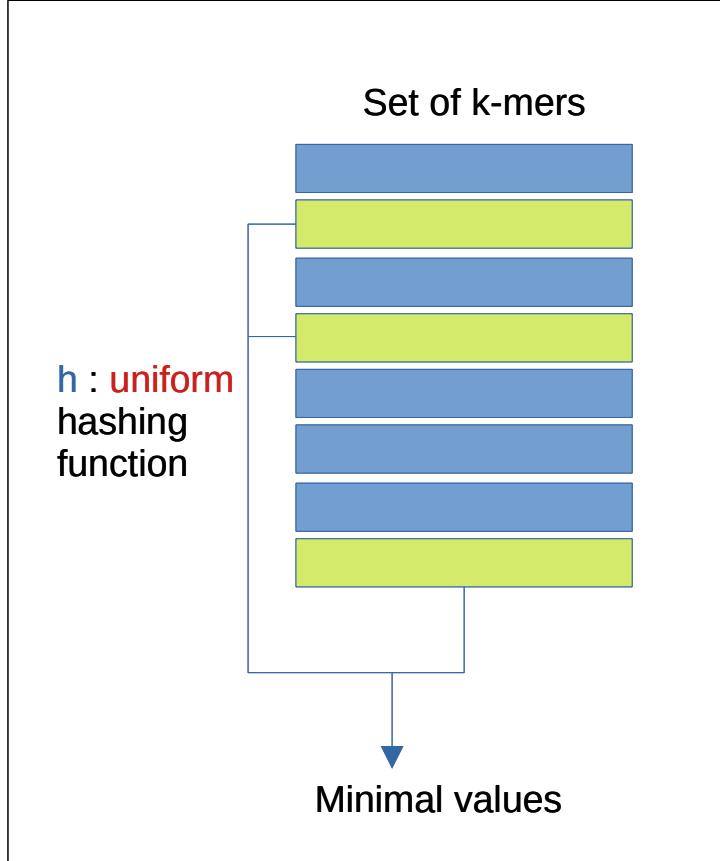
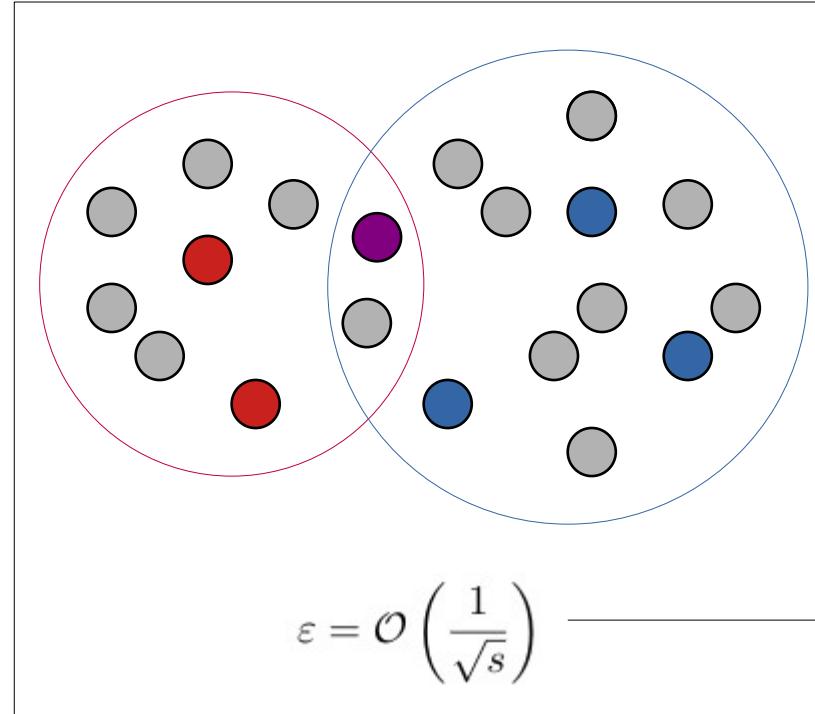


MASH



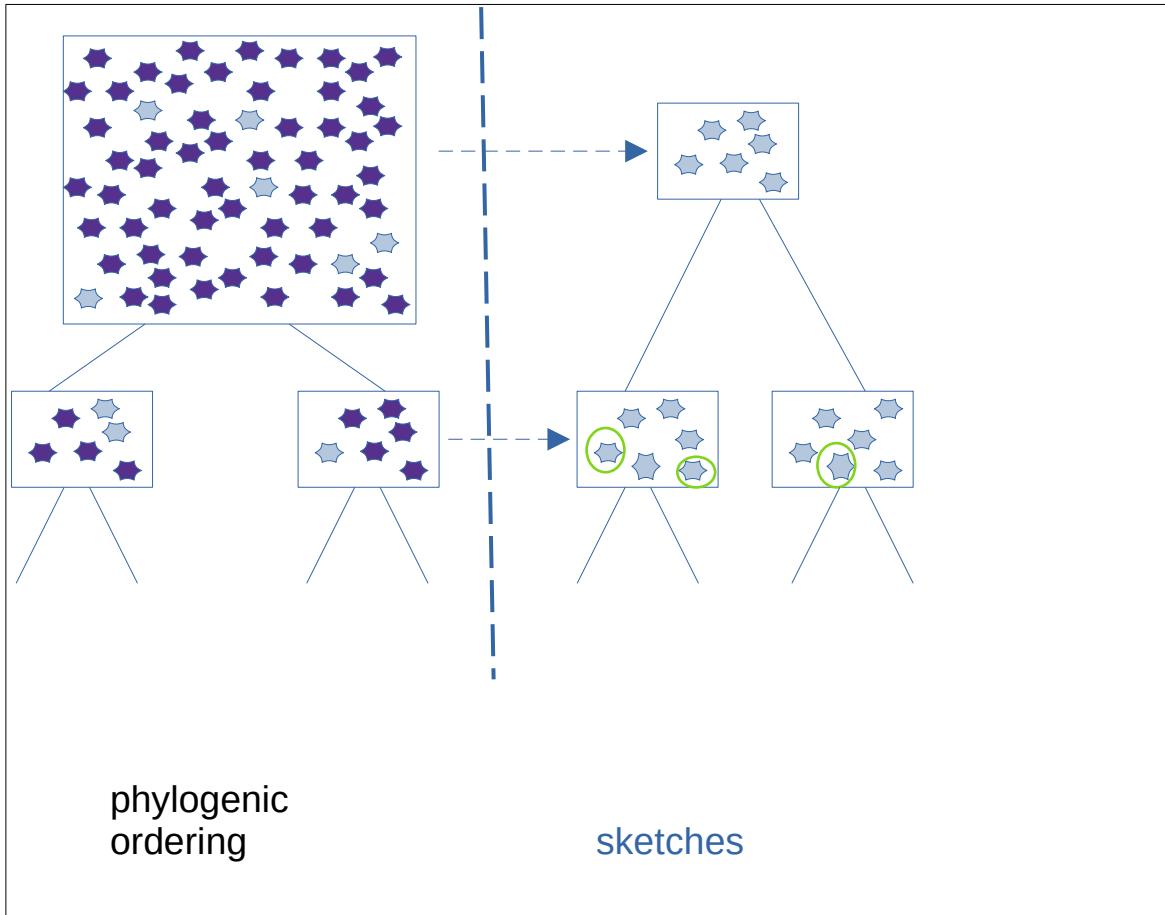
sketch



Estimation of Jaccard index : j

Hypotheses :
- random DNA
- uniform h

Bottom-up compression



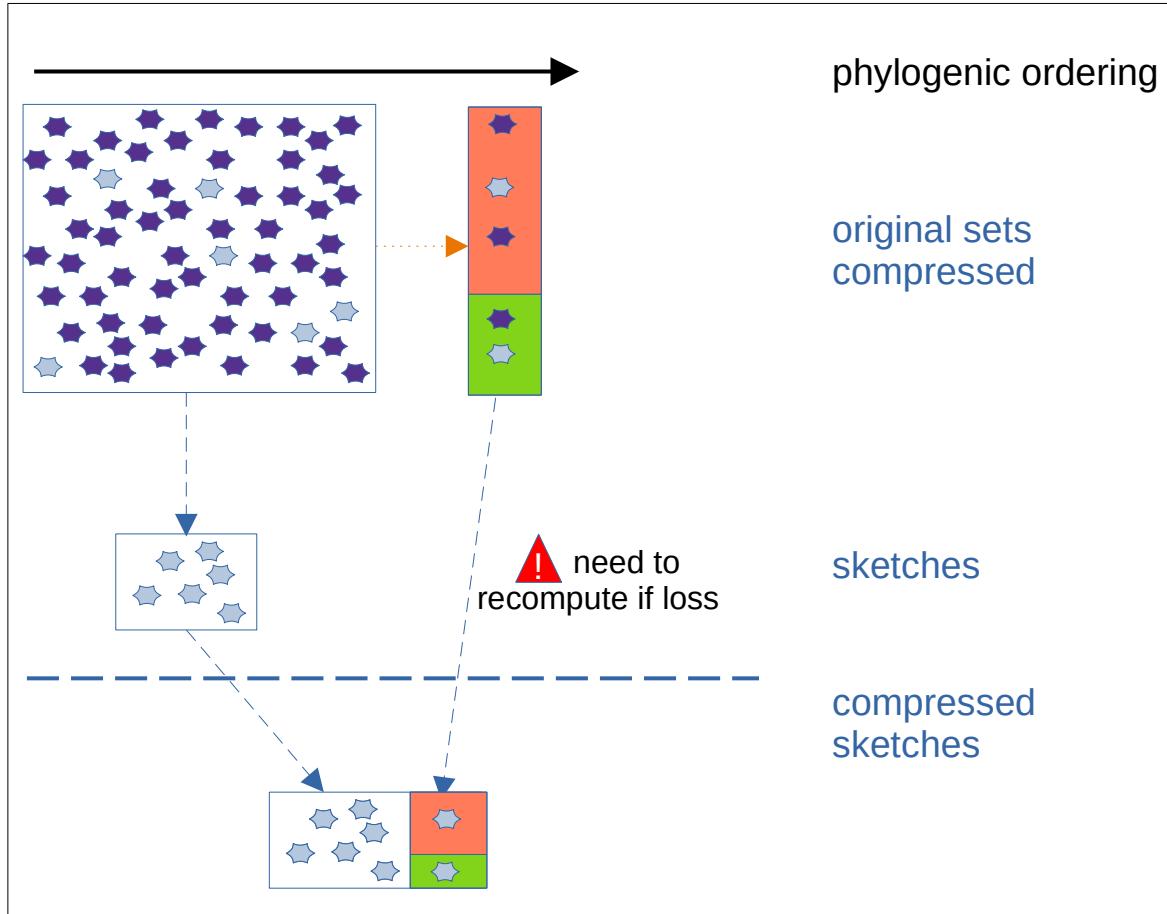
PLUS :

- Better distance estimation
- Reduced memory usage
- Compatible with the algorithm

MINUS

- Extra pre-processing

Left-to-right compression



- Might need to recompute (can be somewhat fixed with $s' > s$)