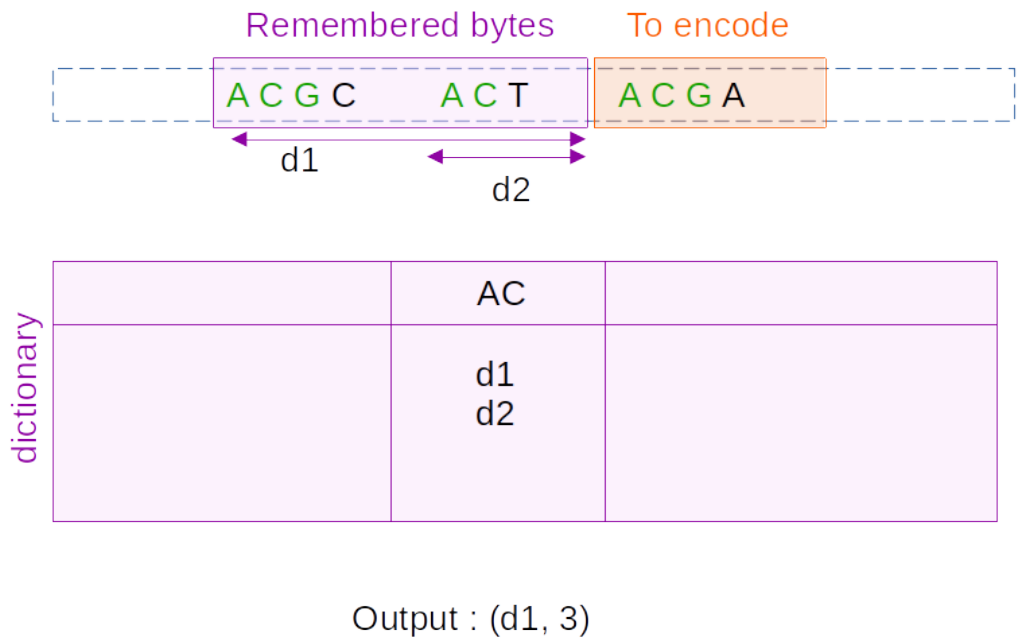
Influence of preset parameter on XZ compression

Marie Picard
–
Supervisors : Karel Břinda, Leo Ackermann

October 23, 2025

XZ in short

1. sliding window algorithm to detect matches

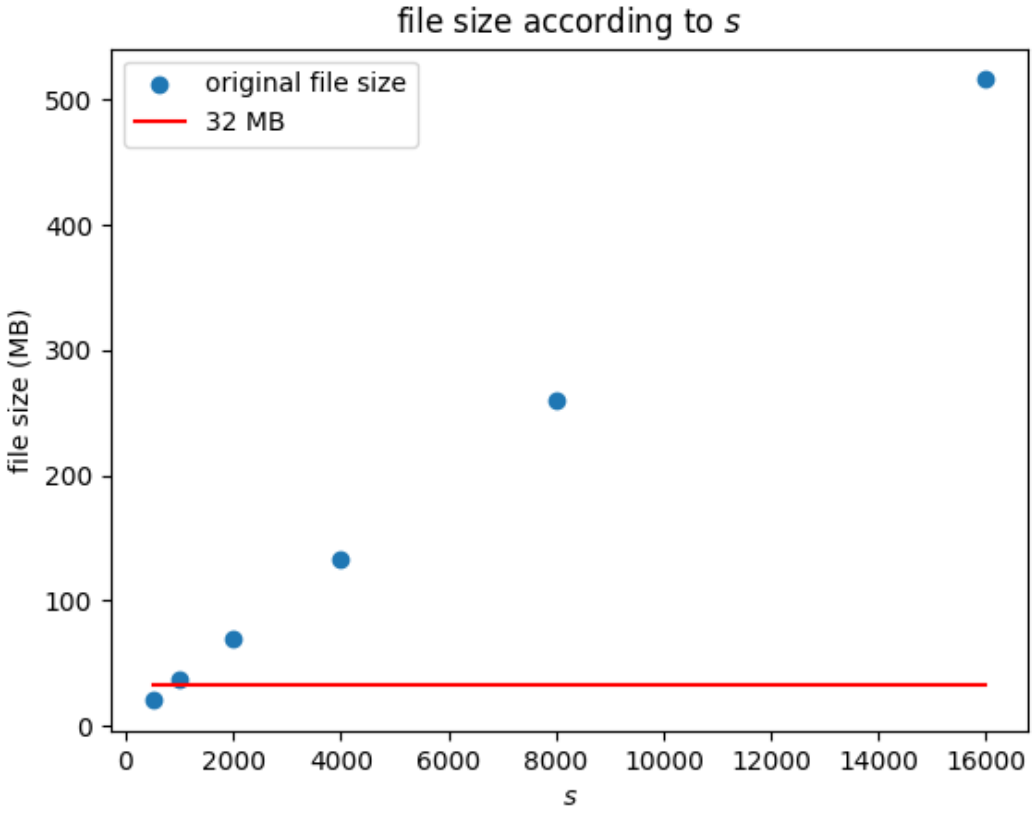


2. range encoding of the output

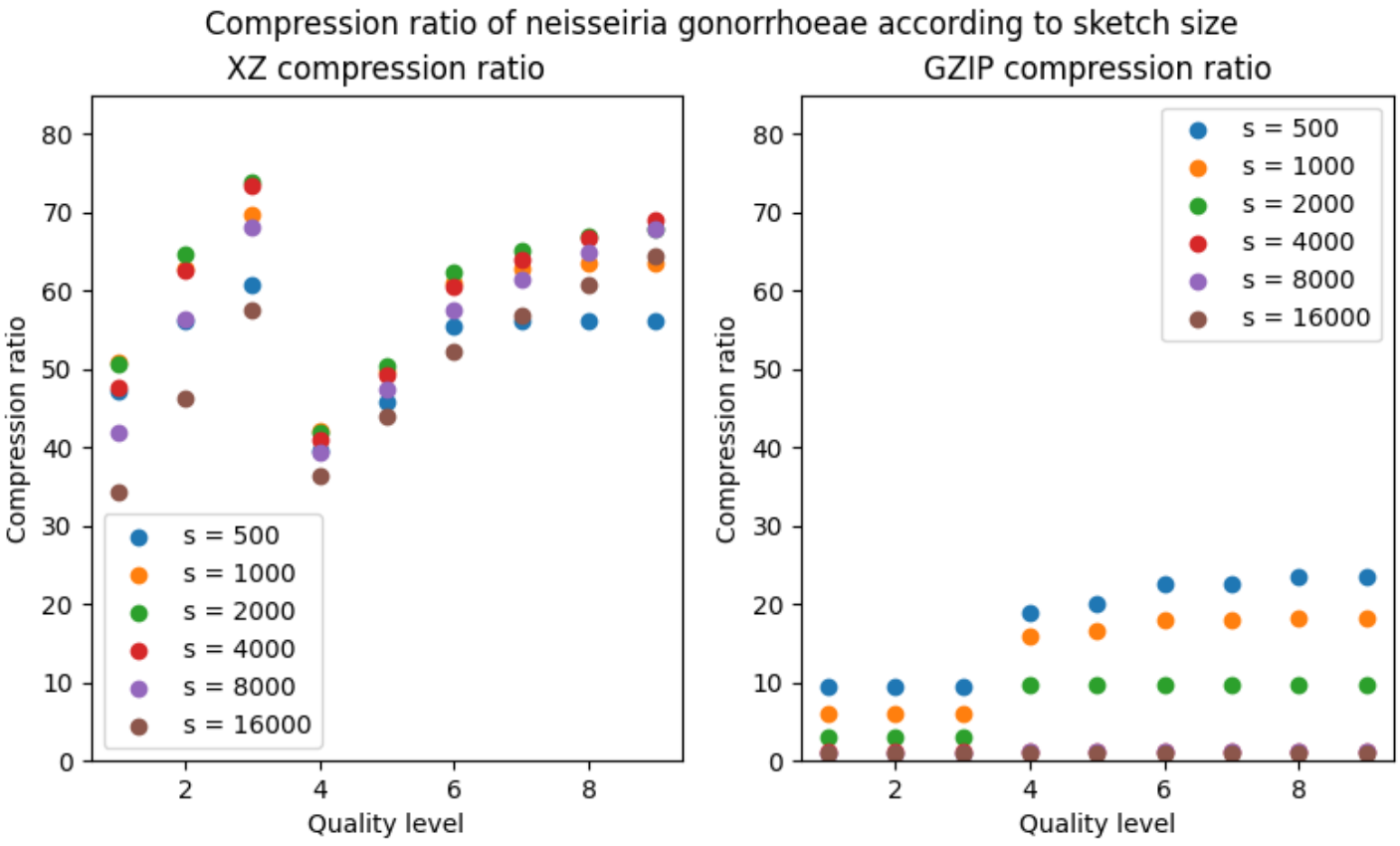
XZ presets memory constraints

Preset	DictSize (MiB)	min file size (MiB)
1	1	0.5
2	2	1
3	4	2
4	4	2
5	8	4
6	8	4
7	16	8
8	32	16
9	64	32

File size according to s



Compression ratio of XZ/GZIP over an archive



Analysis

Observations :

- low compression rate for high presets
- plateau for high presets for smaller values of s

Possible explanation :

- Too high redundancy because of mash format and phylogenetic closeness

Todo :

- look into a reformatting of .msh files (change compression order or smaller representation)
- try using specific algorithms for small values/increasing order