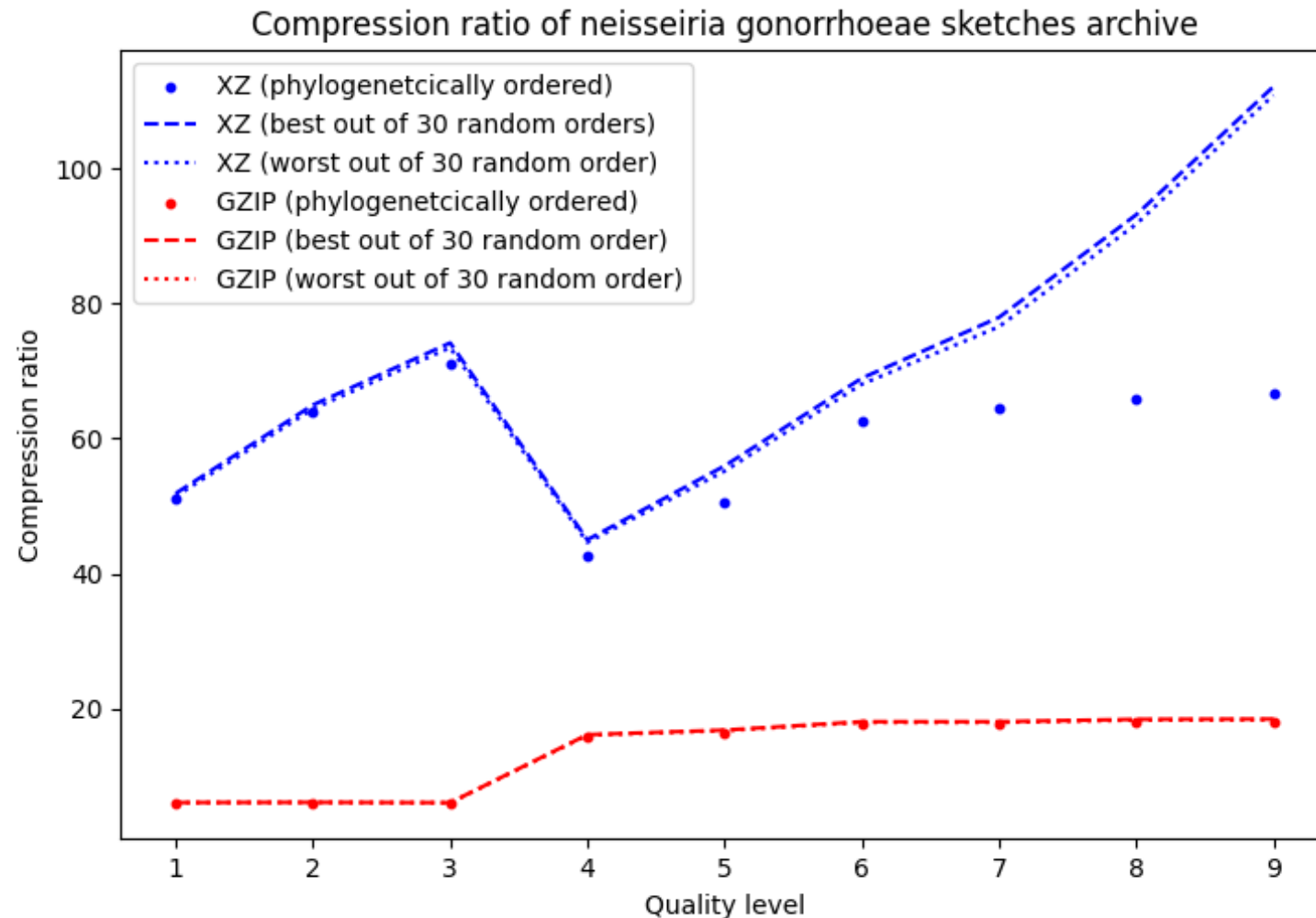


Benchmarking sketch compression using Gzip and XZ

16/10/2025

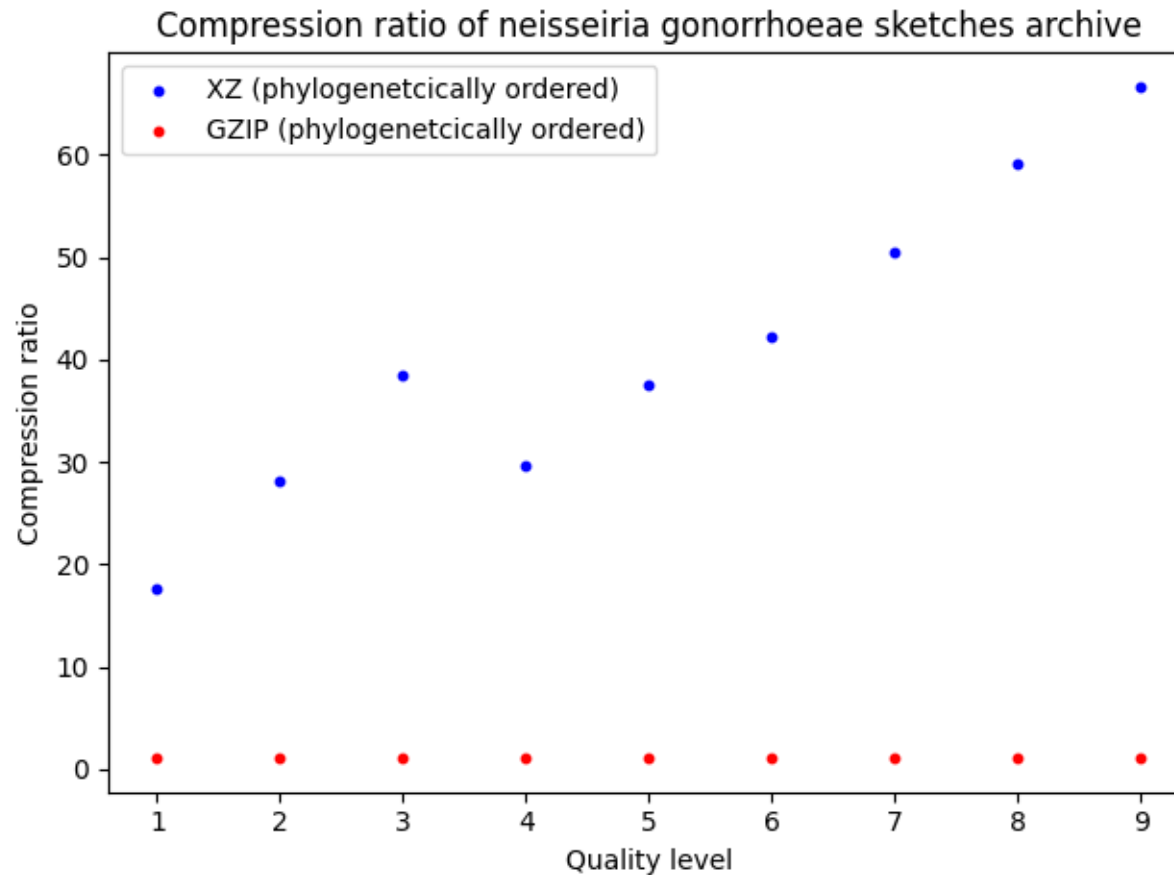
Quick reminder of last week's results



Hypotheses to test

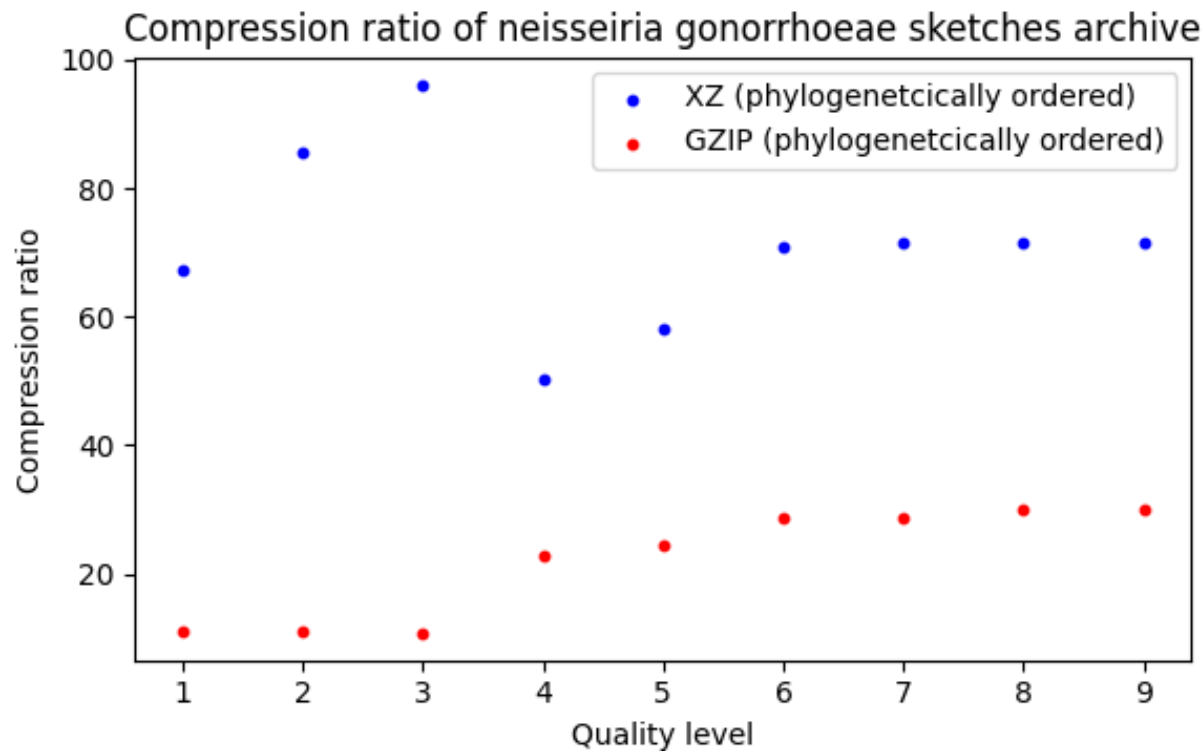
- 1) Headers impede compression rate
- 2) Higher compression rate between phylogenetically-close organisms

1st hypothesis : headers impede compression rate



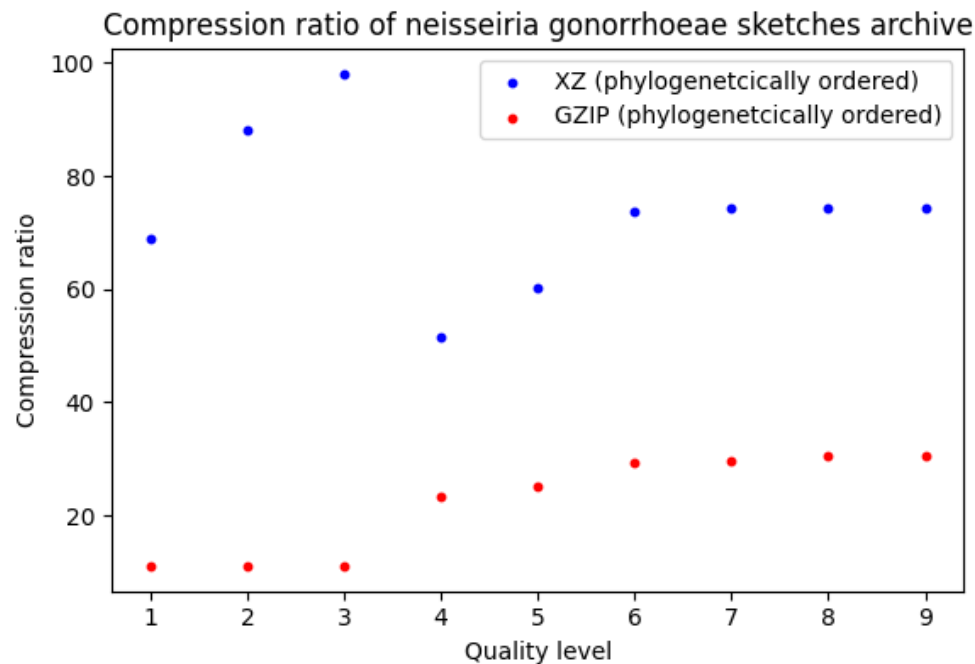
S = 50000

1st hypothesis : headers impede compression rate

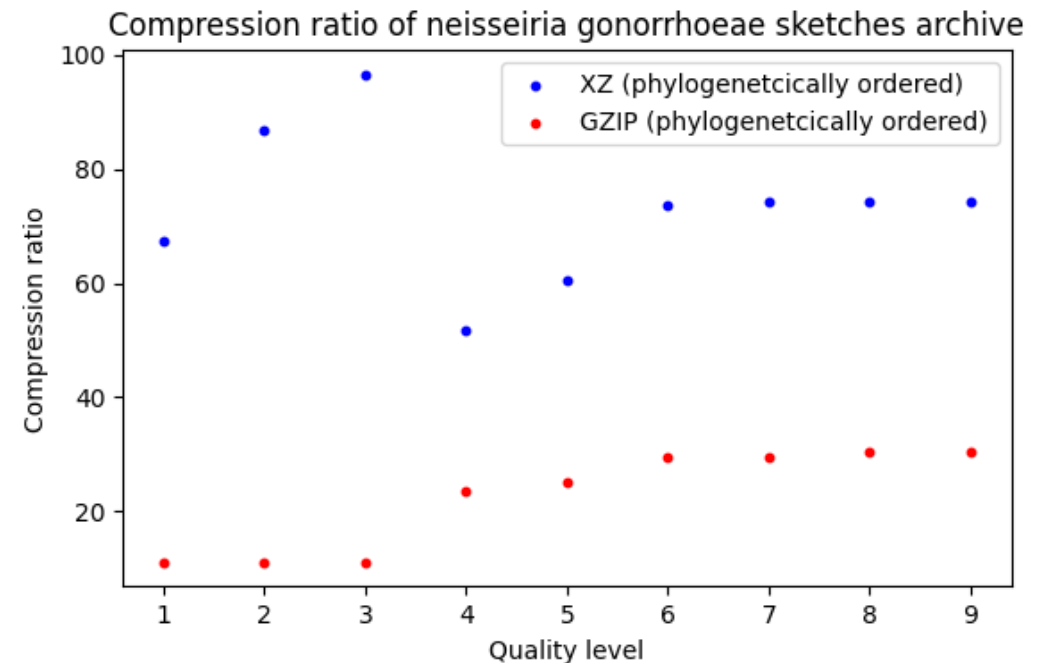


S = 500, no headers

1st hypothesis : headers impede compression rate



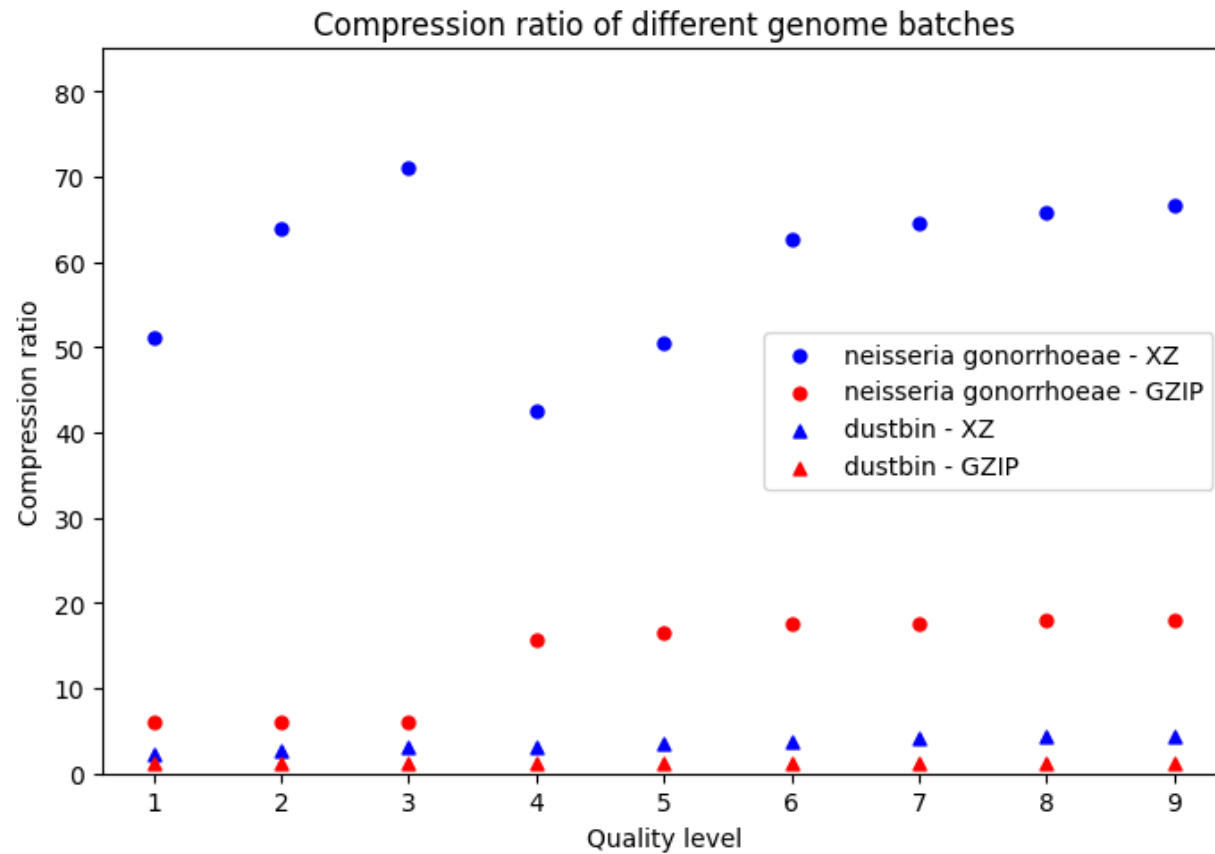
lexicographic order



shuffled

S = 500, no headers

2nd hypothesis : better compression between phylogenetically-close organisms



S = 1000

Conclusion

Conclusion :

- 1st hypothesis : mixed results
- 2nd hypothesis : seems confirmed

To do :

- Improve random tests
- Try other compression algorithms