bíogo/hts: high throughput sequence handling for the Go language

R Daniel Kortschak[1]**, Brent S Pedersen**[2]**, and David L Adelson**[1]

1 School of Biological Sciences, The University of Adelaide **2** Department of Human Genetics, University of Utah

Summary

bíogo/hts provides a Go native implementation of the SAM specification (Group 2016) for SAM and BAM alignment formats (H. et al. 2012) commonly used for representation of high throughput genomic data, the BAI, CSI and tabix indexing formats, and the BGZF blocked compression format. The bíogo/hts packages perform parallelized read and write operations and are able to cache recent reads according to user-specified caching methods. The parallelisation approach used by the bíogo/hts package is influenced by the approach of the D implementation, sambamba by Tarazov et al. (T. A. et al. 2015). The bíogo/hts APIs have been constructed to provide a consistent interface to sequence alignment data and the underlying compression system in order to aid ease of use and tool development.

References

A., Tarasov, Vilella A., Cuppen E., Nijman I. J., and Prins P. 2015. "Sambamba: Fast Processing of Ngs Alignment Formats." *Bioinformatics* 31 (12): 2032–4. doi:10.1093/bioinformatics/btv098.

Group, The SAM/BAM Format Specification Working. 2016. "Sequence Alignment/Map Format Specification." https://samtools.github.io/hts-specs/SAMv1.pdf.

H., Li, Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R., and 1000 Genome Project Data Processing Subgroup. 2012. "The Sequence Alignment/Map Format and Samtools." *Bioinformatics* 25 (16): 2078–9. doi:10.1093/bioinformatics/btp352.