

The GTrack ecosystem

- expressive file formats for analysis of genomic track data



ELIXIR All Hands 2017, 21-23 March, Rome, Italy

Sveinung Gundersen, Matúš Kalaš, Boris Simovski, Brynjar Rongved, Henrik Skjfield, Sivert Kronen Hatteberg, Abdulrahman Azab, Osman Abul, Arnoldo Frigessi, Geir Kjetil Sandve, and Eivind Hovig

Complementary file formats - GTrack, BTrack and GSuite are file formats designed to handle genomic track data of heterogeneous types. The file formats are designed to complement each other and work jointly as a complete ecosystem for representation and analysis of most types of data that can be located along a reference genome.

GTrack

GTrack¹ is a tabular format that was developed as part of the Genomic HyperBrowser^{2,3} system to provide a uniform representation of most types of genomic datasets. GTrack is able to replace common formats such as WIG, GFF, BED, FASTA, in addition to represent chromatin capture datasets, such as Hi-C and ChIA-PET.

BTrack

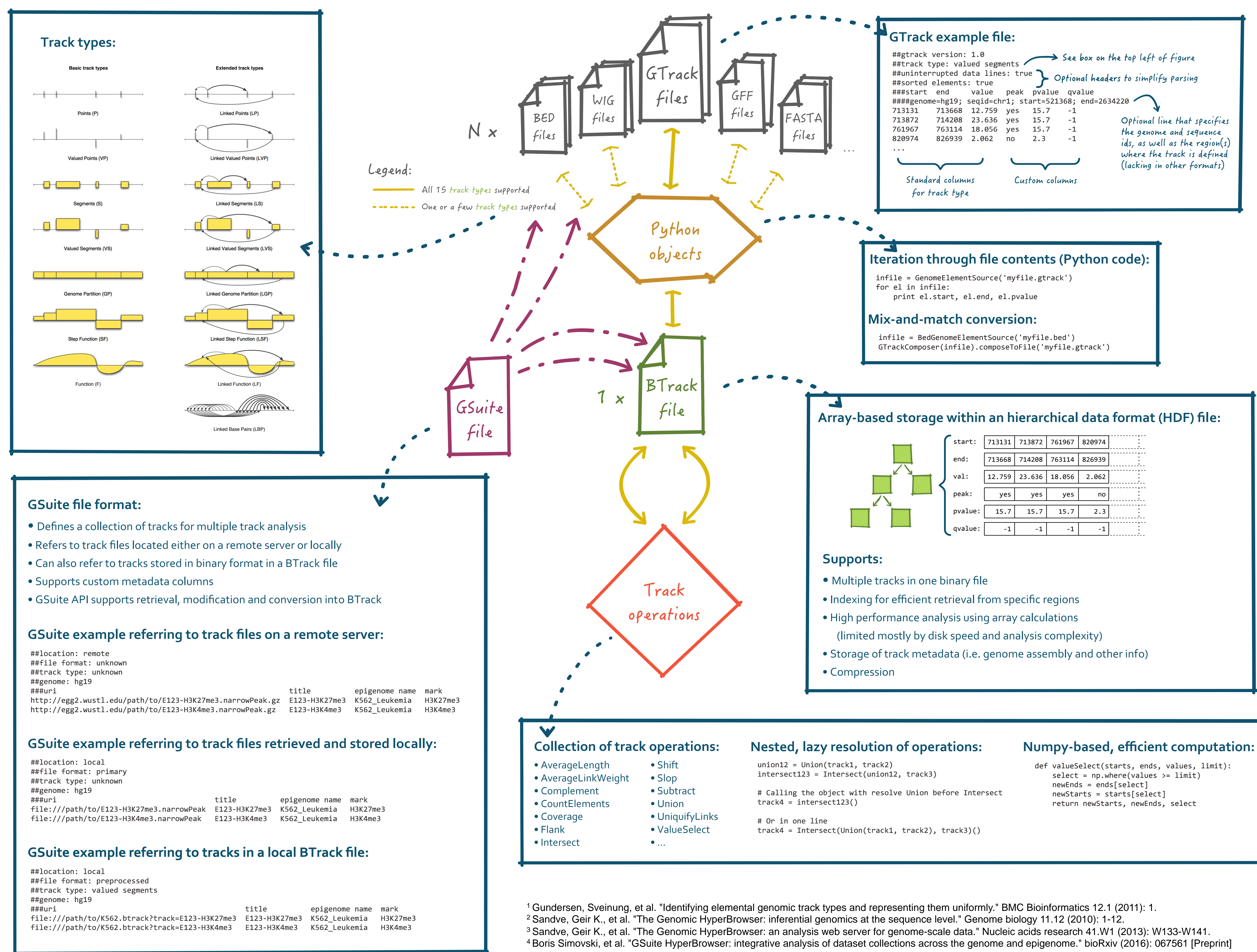
The BTrack format supports the same variety of informational content as GTrack, but in a binary form. BTrack is unique in supporting a collection of multiple tracks stored together in one HDF5-based binary file, while still supporting a high level of efficiency.

GSuite

The GSuite⁴ format is a unique tabular format that binds together the whole chain of multi-track analysis, from search and retrieval of genomic tracks, through intermediate processing, to analysis.

Python package

A Python package supporting parsing, conversion and operations is available from <https://github.com/elixir-no-nels/gtrackcore>. It is currently limited (excl. HDF, GSuite and track operations) but will be expanded to support a fully functional API with command line access.



¹ Gundersen, Sveinung, et al. "Identifying elemental genomic track types and representing them uniformly." BMC Bioinformatics 12.1 (2011): 1.

² Sandve, Geir K., et al. "The Genomic HyperBrowser: inferential genomics at the sequence level." Genome biology 11.12 (2010): 1-12.

³ Sandve, Geir K., et al. "The Genomic HyperBrowser: an analysis web server for genome-scale data." Nucleic acids research 41.W1 (2013): W133-W141.

⁴ Boris Simovski, et al. "GSuite HyperBrowser: integrative analysis of dataset collections across the genome and epigenome." bioRxiv (2016): 067561 [Preprint]

Contact:

Sveinung Gundersen
sveinung.gundersen@medisin.uio.no
Elixir Norway, Oslo Node, Univ. of Oslo,
Pb 1080 Blindern, 0316 Oslo, Norway



@ELIXIREurope



/company/elixir-europe



UiO : University of Oslo

