

FIGURES

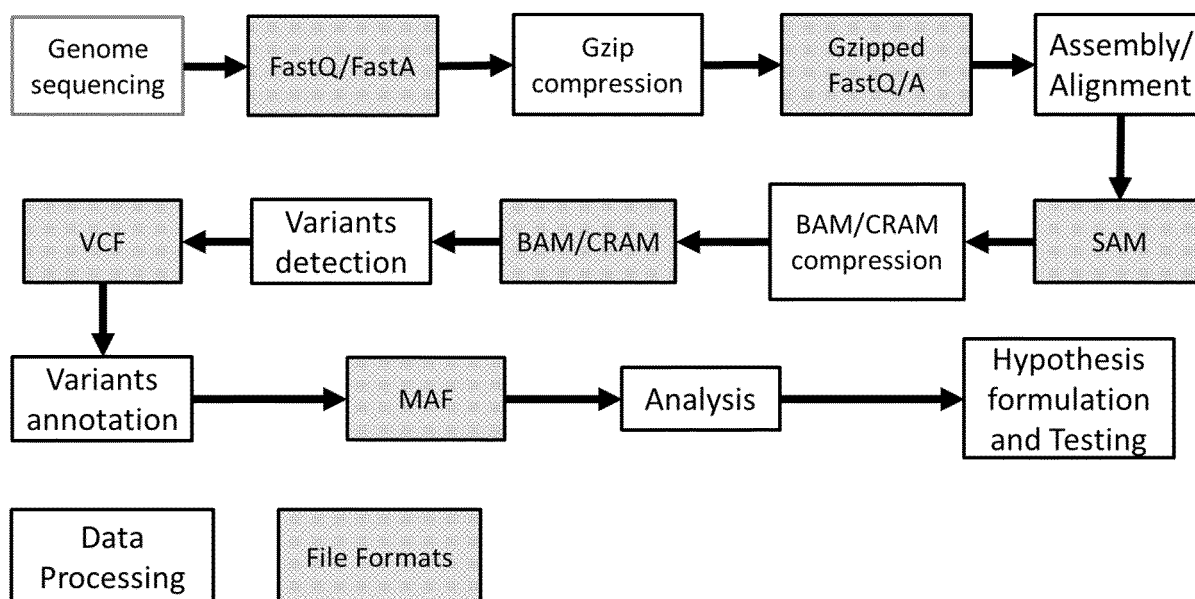


Figure 1- Genomic data life cycle from sequencing to analysis and hypothesis formulation.

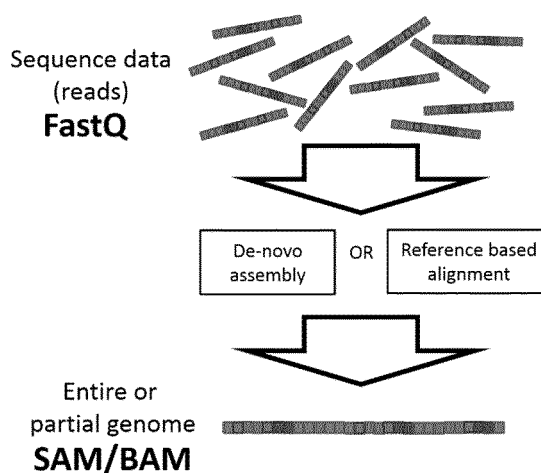


Figure 2 - Genome sequence reads alignment.

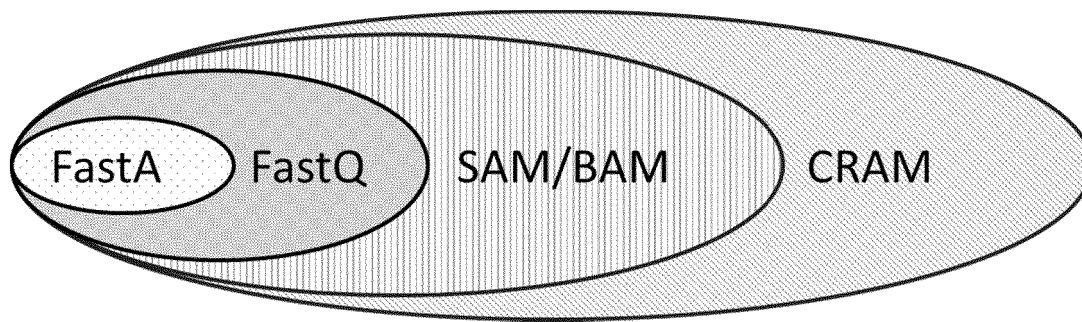


Figure 3 - Genomic file formats and their relationships.

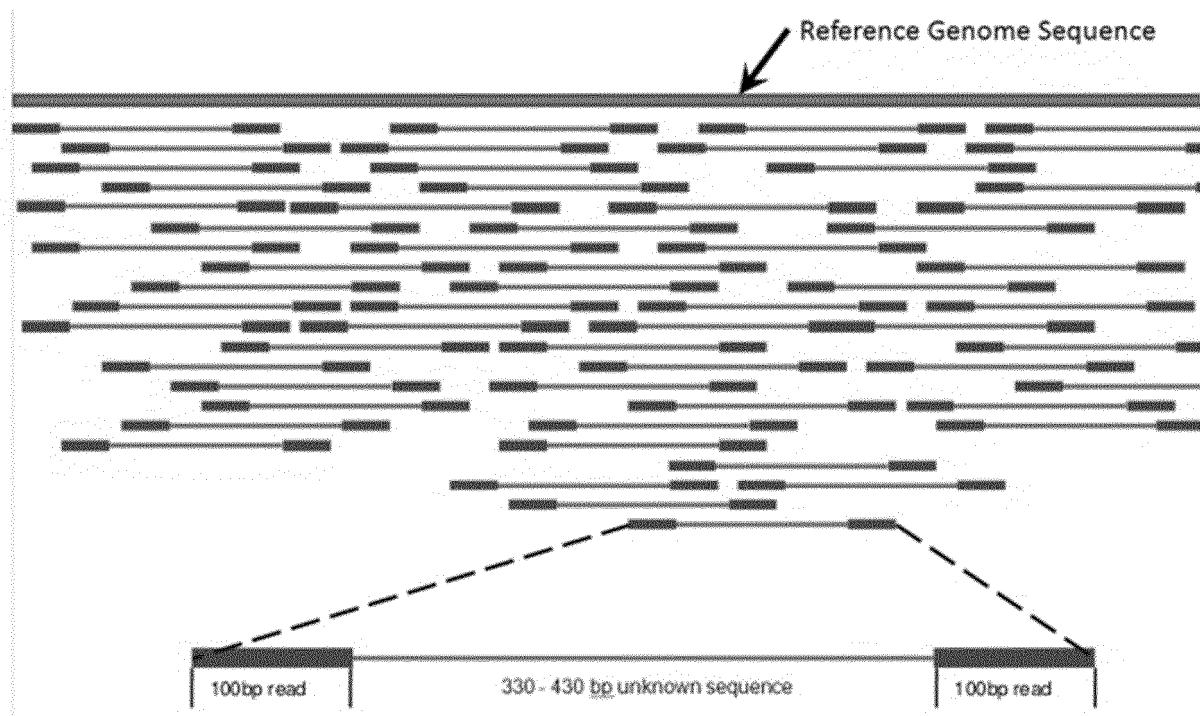


Figure 4 - Reads pairs mapped to a reference sequence (Genome).

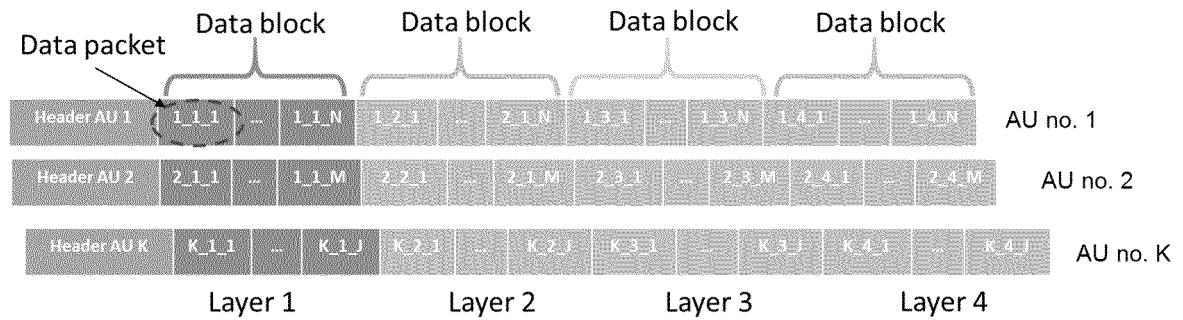


Figure 5 – Access Units are composed by Data Blocks. Each data block contains packets of descriptors belonging to the same type of data.

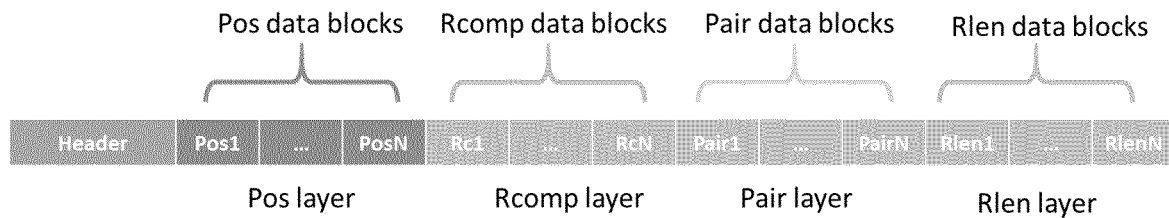


Figure 6 - Example of Access Unit including a header and layers composed by data blocks.

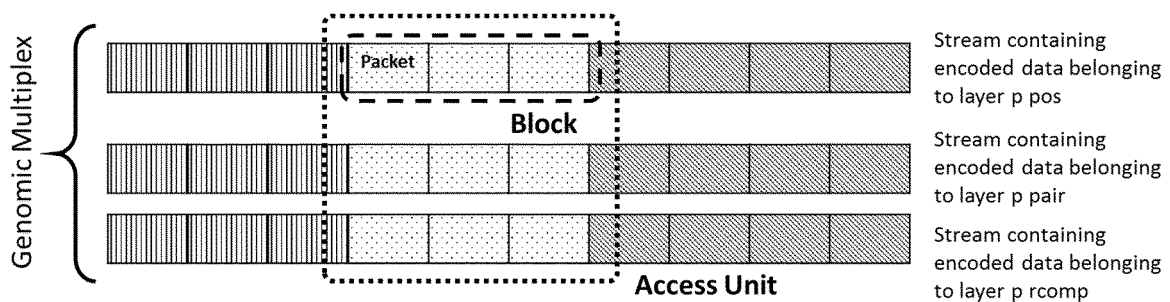


Figure 7 - Relation among genomic "Data Packets", "Blocks", Access Units, layers and Streams Reads Classes.

Master Index Table

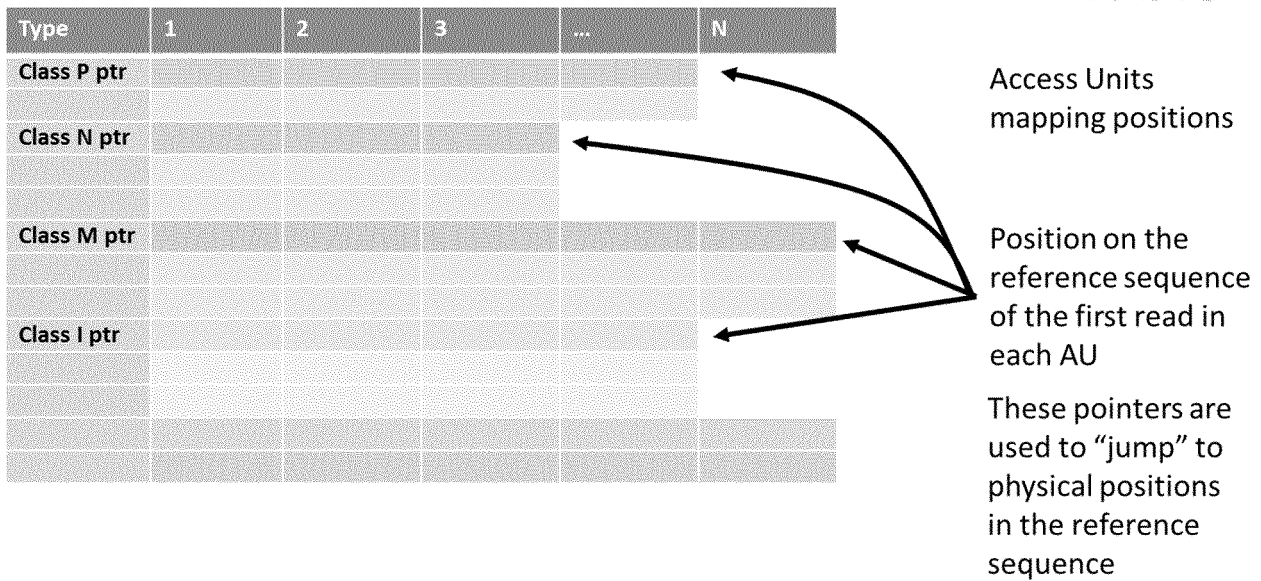


Figure 8 - The first type of data contained in the MIT are vectors of mapping loci of the first read contained by each Access Unit.

Unique ID	Version	Header Size	Ref count	Reads Len.	N. AUs	Ref Ids	M.I.T.
-----------	---------	-------------	-----------	------------	--------	---------	--------

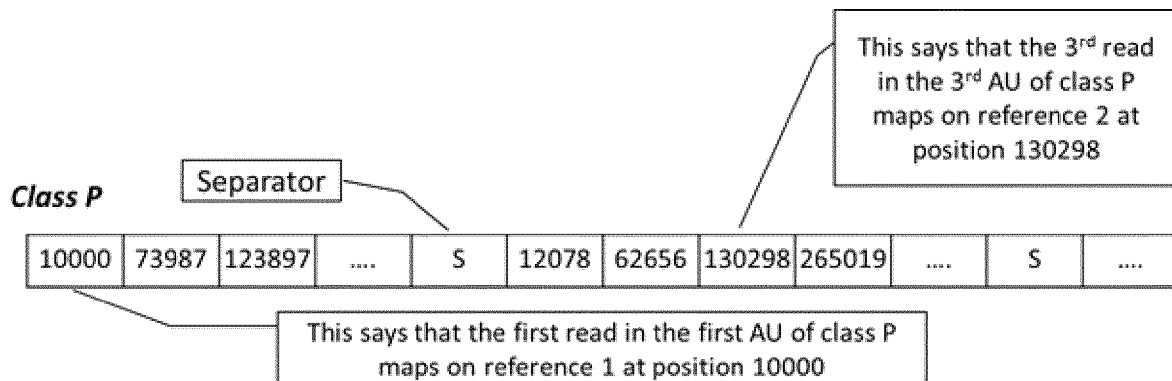


Figure 9 - The generic structure of the Main Header and a partial representation of MIT showing the mapping positions of the first read in each pos AU of class P.

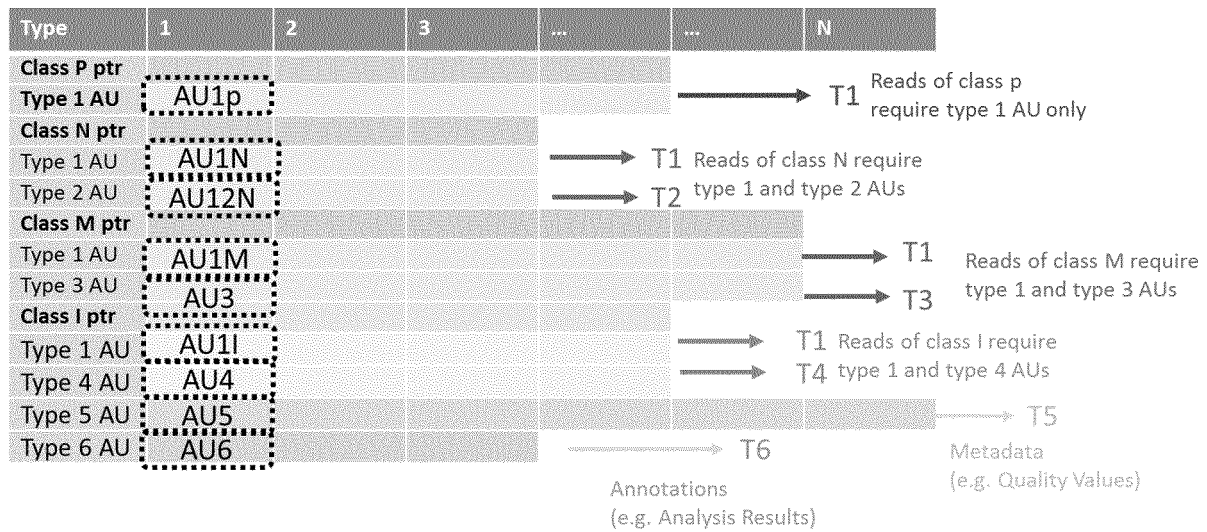


Figure 10 – The second type of data contained by the MIT are vectors of pointers to the physical location of each Access Unit. These vectors are referred to as Local Index Tables.

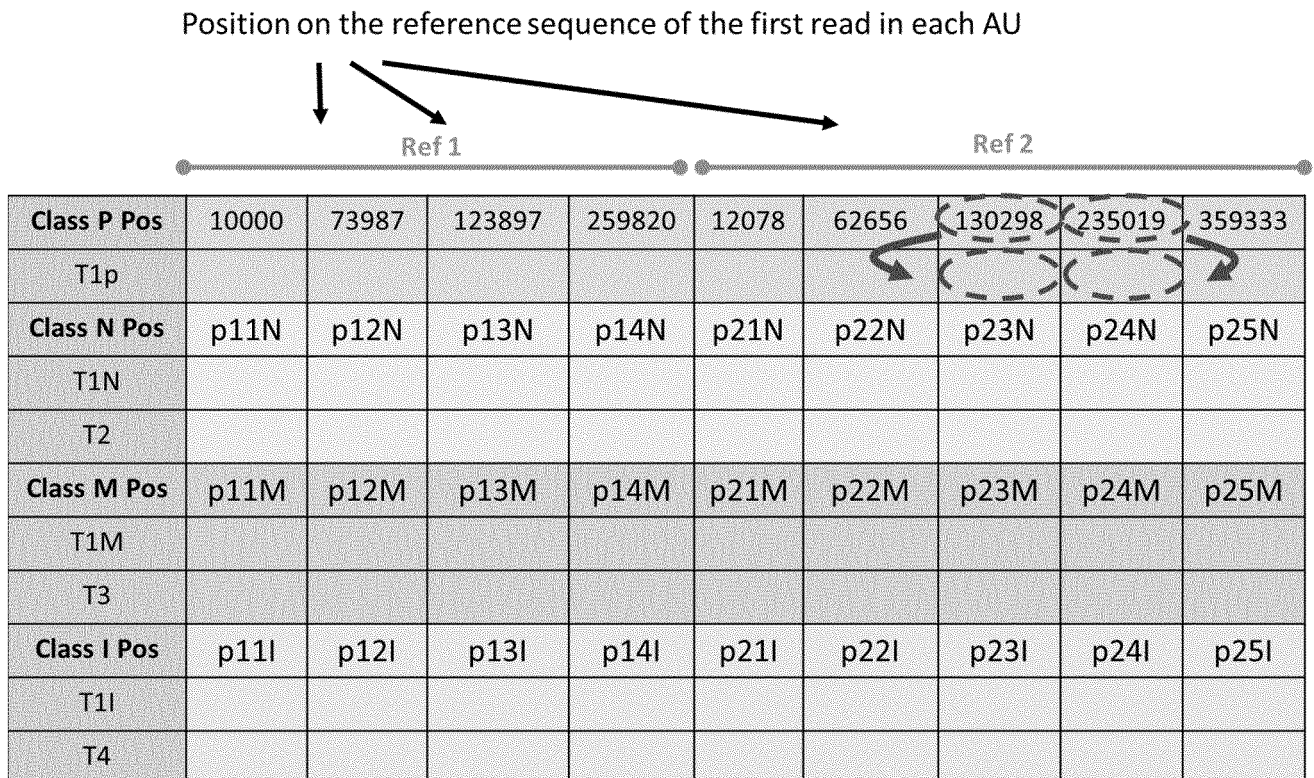


Figure 11 - The Access Units containing reads of class P mapped on reference sequence no. 2 between position 150,000 and 250,000 are accessed using the values contained in the T1p vector

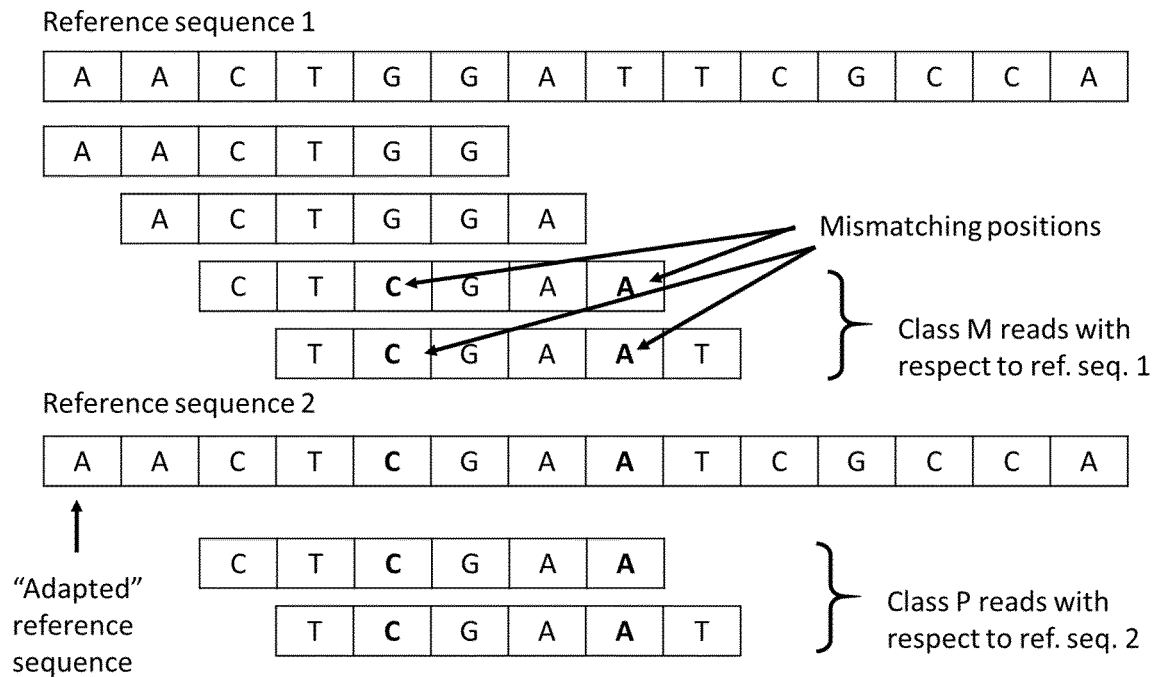


Figure 12 - A modification in the reference sequence can transform M reads in P reads.

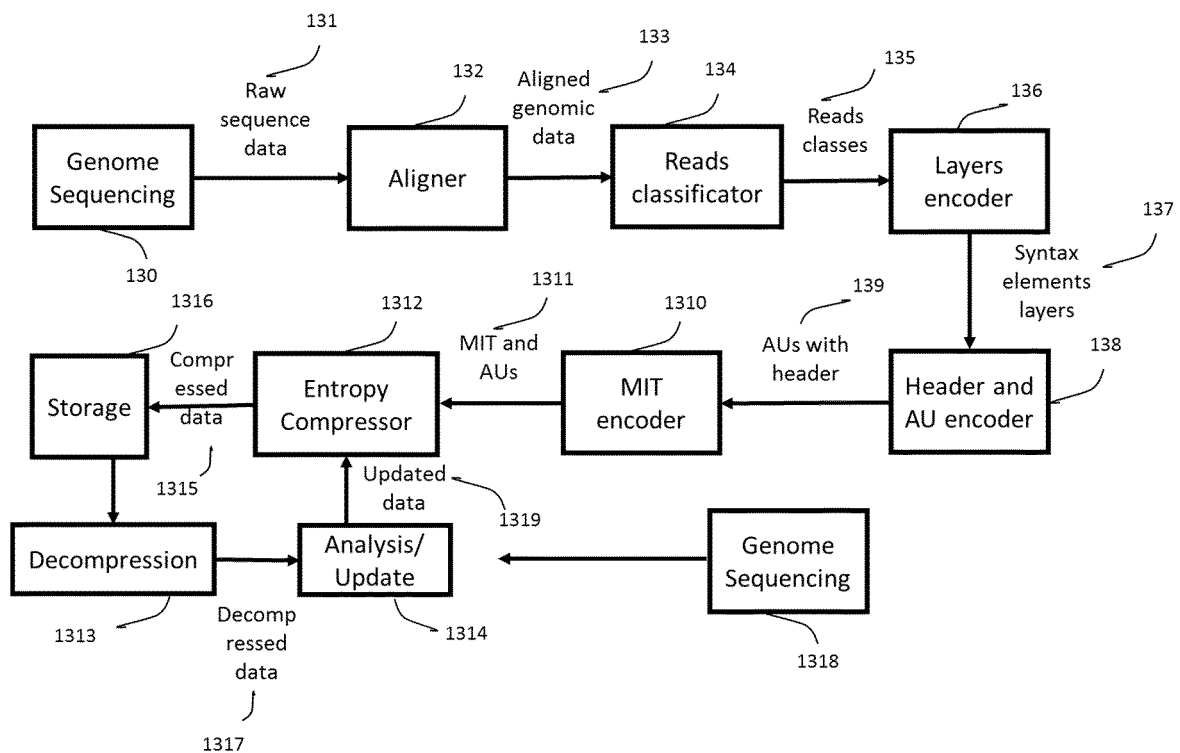


Figure 13 - Overall Architecture.

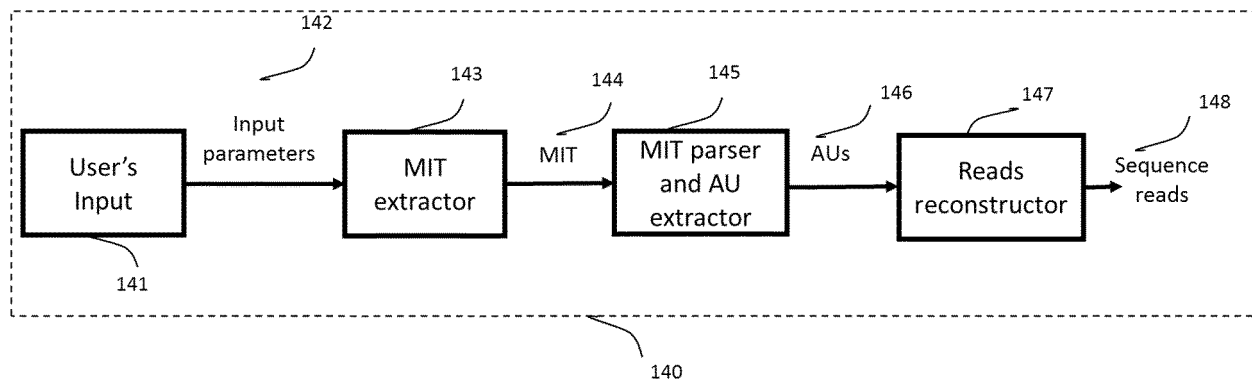


Figure 14 – Sequence reads extractor

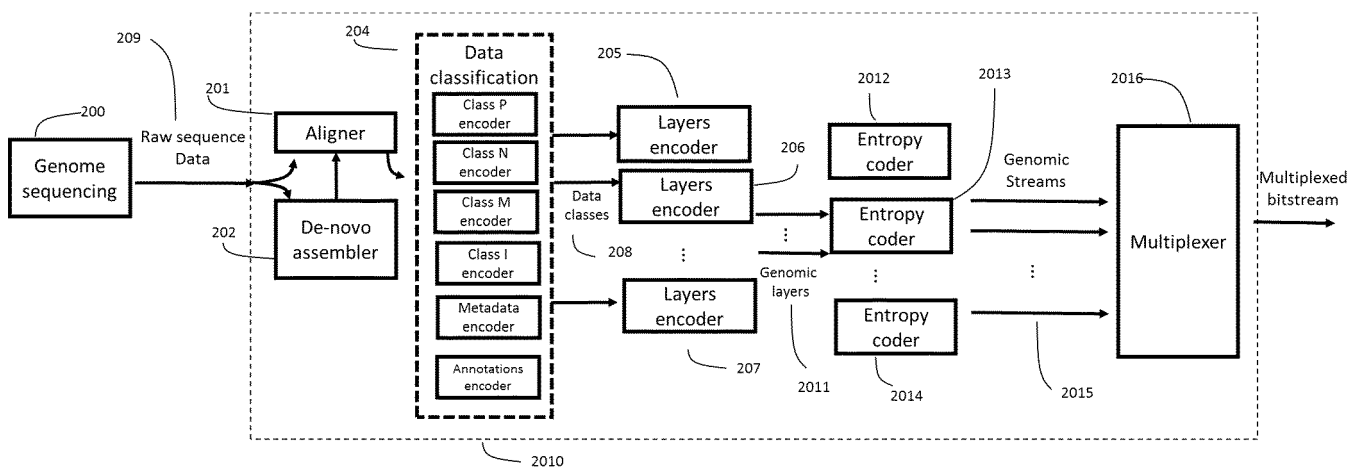


Figure 15 - Genomic encoder

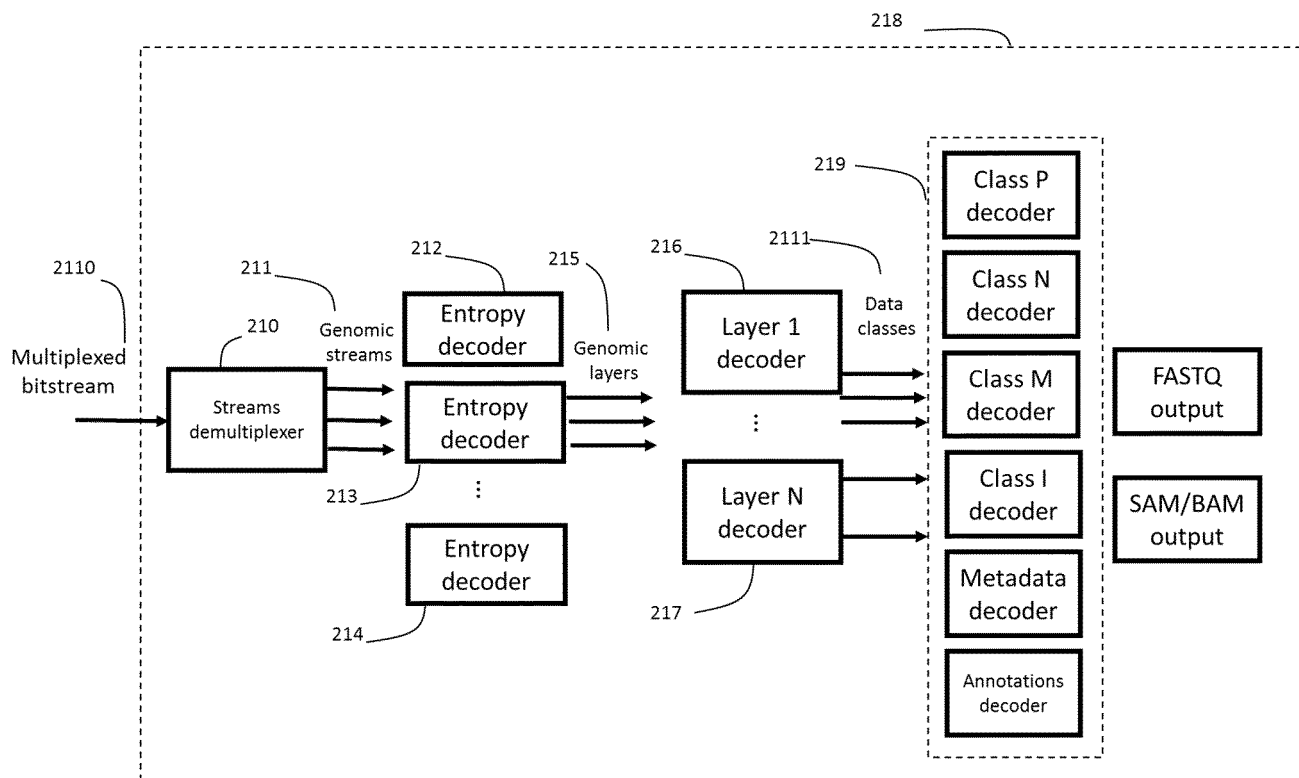


Figure 16 – Genomic Decoder