

Documentation to BinningTestData

- Obligatories:
 - all of the C++ files must be kept in the same directory
- Librarys used:
 - R:
 - Rsamtools
 - C++:
 - iostream
 - vector
 - list
- Functions:
 - binningTestData:
 - arguments:
 - filenames_csv
 - the name of a file containing either the names of the bam-files and the according fasta-file or the fasta-file, the file/s containing the reads and a name for the bowtie2-prebuild
 - example of version nr.1
 - headerForFasta, headerForBams
fastaNr1, BamNr1
fastaNr2, bamNr2
...
 - example of version nr.2:
 - headerForFasta, headerForFatsqNr1, headerForFastqNr2, prebuild-name
FatsaNr1, FastqNr1, FastqNr2, Buildname
FastaNr2, FastqNr3,, BuildnameNr2
...
 - coverage
 - an integer specifieng the requested coverage over the sequence, from which the number of reads to be used will be determined
 - prebuild
 - a boolean that determins wether the bowtie2-prebuild function is going to be used or not

- readAsBams
 - a boolean that determines whether filenames_csv will be read as version nr.1 or nr.2. If it is True it will be read as version Nr. 1.
 - humanReadable
 - a boolean that determines whether there will be graphical output in a pdf file or not
 - repeatable
 - a boolean that determines whether there will be a seed set, so that with the same data and the repeatable value as true the function will always produce the same result
- output:
- the output is a list. In this list the first element is an IRanges object, the second a vector containing the mean coverage values to each of the ranges/contigs, and the third contains the N50 value to the assembly. From there on this pattern repeats itself for all sequences given as arguments, in the same order they are listed in the argument file. If there is more than one sequence in a given fasta file the output is listed in the same order as the sequences occur in the fasta file.
 - example:
 - list(IRangesObjectNr1, meanCoverageVectorNr1, N50Nr1, IRangesObjectNr2, meanCoverageVectorNr2, N50Nr2 ...)
 - as mentioned above there is a way to also get graphical output, the output will be a graph of the coverage to the position on the sequence and a histogram of the contig lengths for every sequence