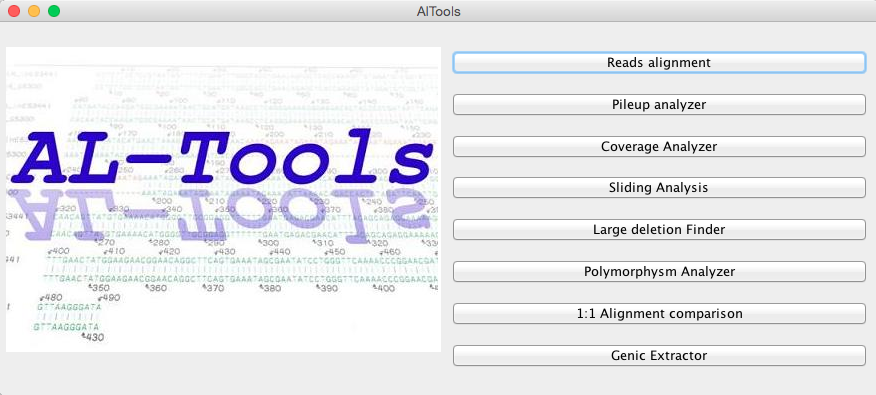
**Altools 0.1 manual**

**What is Altools**

Altools is a suite of software that will help you in calling, examining, visualising and confer biological meaning to your NGS data. Atools features a graphical user interface (GUI, figure 1) that allows it to be used even with no (o very small) previous experience with the unix terminal. It is based on c++ written algorithms and a Java based GUI and for this reason also the software installation is kept as hassle free as possible. Just a few system requirements and you are ready to start…….



**Figure 1:** Altools main window.

**System requirement**

Altools will generate several plots by using R scripts and for this reason R must be installed on the machine running the software. For installing R on your computer please refer to the page <http://cran.ms.unimelb.edu.au> .

Altools uses an R package called dnacopy to investigate the copy number variations and for this reason such a software must be installed on your machine as well. In order to install rnacopy open a terminal on your machine and type the following (in succession):

>R

>source("http://bioconductor.org/biocLite.R")

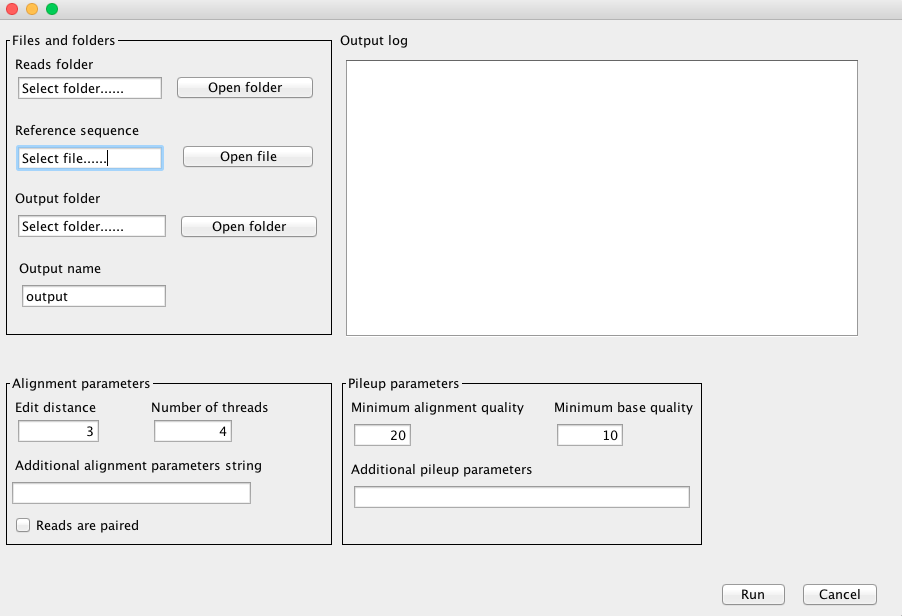
>biocLite("DNAcopy")

……. That’s it!

Altools uses also the softwares bwa, samtools and Varscan. Such tools are already included in the main Altools folder. You can replace the with the latest versions, however care must be taken in name them exactly as they are at the moment of download.

**Alignement and pileup**

In this section your reads will be aligned by using the bwa program after your reference genome has been indexed (figure 2). The resulting .sam file will be converted to bam format, sorted and elaborated by the samtools program that will generate a pileup file that will represent the starting point of the following elaborations.



**Figure 2:** Reads alignment GUI

In the reads alignment GUI you should select the folder containing the fastq formatted reads, the fasta formatted reference file, the output folder and a valid output file name. Just a few words about the reads folder. Files in such folder can only end with “\_1.fastq” or “\_2.fastq”. Altools will recognize potential paired end files by their name (e.g. pairedEnd\_1.fastq and pairedEnd\_2.fastq). File ending with “\_1.fastq” and without the corresponding “\_2.fastq” mate will be treated as single end.

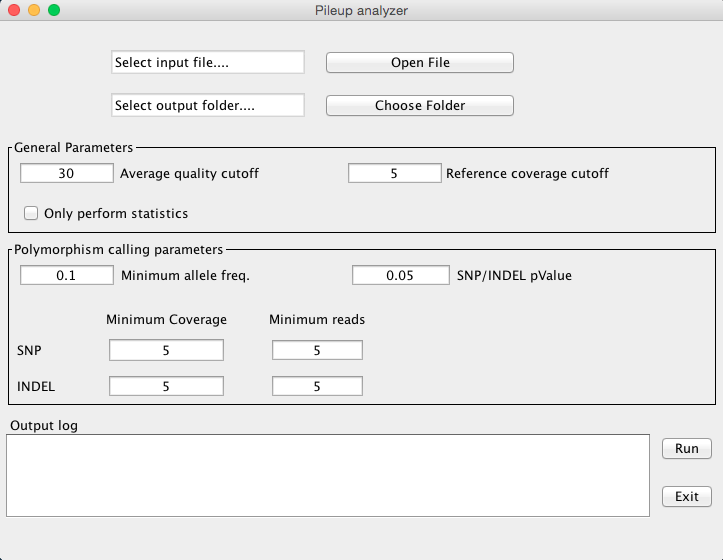
You can change 3 parameter for the alignment process: (a) the edit distance, (b) the number of threads and (c) any additional parameters for the bwa software (please refer to the bwa manual for more details). Additional parameters should be inserted by using the corresponding flags and values separated by spaces (e.g. “–i 4 –o 2”, without quotes, etc.). Finally you can decide to tick the “Reads are paired” checkbox. Doing so all the paired ends reads will be aligned accordingly, otherwise all the files in the reads in the reads folder will be treated as single end reads.

You can also tune the samtools pileup generation process by choosing (a) the minimum alignment quality, (b) the minimum base quality and (c) any additional samtools parameters, again by using the corresponding flags and values separated by spaces.

In the output folder 4 subdirectories will be generated: (a) Sam where all the sam files will be stored, (b) Bam where all the bam files will be stored, (c) indexedReference where the reference genome and the indexing files will be stored and (d) Pileup where the pileup file will be stored. In the latest two files will be placed , namely a \_pileup file and a \_pileupCorr file (an elaborated version of the \_pileup files). The latest should be used for the following elaborations with Altools.

**Pileup analyser**

This program (figure 3) will generate a pileup folder where coverage, SNPs and INDELs info will be sorted in a simple and lightweight fashion. Most of the remaining Alttols software will work on this folder. This tool takes in input a pileup file that has been generated during the alignment and pilep phase and more precisely the \_pileupCorr file.

 **Figure 3**

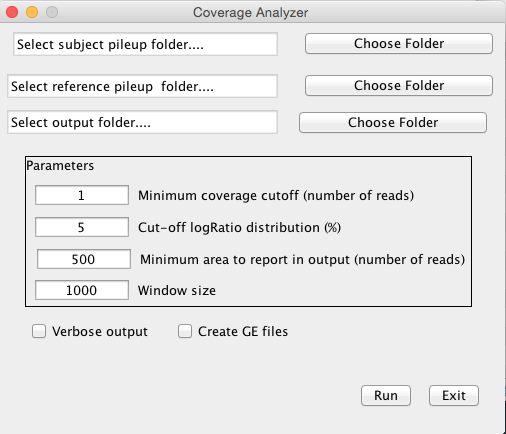
You can choose the output folder where the pileup information and the main statistics will be stored. Finally please feel free to play around with the remaining parameters that are tunable in the self-explained textboxes.

Pileup analyser uses Varscan to call snps and indel and will generate (a) a Varscan formatted pileup file (varscaPileup), (b) Statistics on the alignment in a tabular format with several information about the coverage and the called polymorphisms divided by chromosome and (c) the pileup folder that will be used in the remaining Altools programs (see below).

**Coverage analyser**

Copy number variation (CNV) and presence/absence variation (PAV) can be calculated with this tool (Figure 4). Since by definition loss or gain can be distinguished only by comparing two genomes this tool take in input the pileup folder that has been previously generated on your resequencing experiments and a second pileup folder that has been generated on the reference genome (therefore you should grab some reads on the reference too). Alternatively you can use this tool to compare two different resequencing projects relative to the same species.

Coverage analyser uses the R package dnacopy to segment the entire genome by considering the ratio between the coverage of your species and the coverage on the reference (i.e. the two pileup folders). A distribution of ratio values will be generated and you can set the percentage cutoff of this distribution to be used to call potential losses and gains. Since the coverage ratio will be calculated on adjacent segments of DNA (e.g. by using an adjacent window approach) you can choose the segment size (Window size) and also a minimum length for the found polymorphisms to be output. In the output folder several subdirectories will be generated corresponding to the different chromosomes. In each of these folder informations about losses, gains and zero coverage areas will be reported.

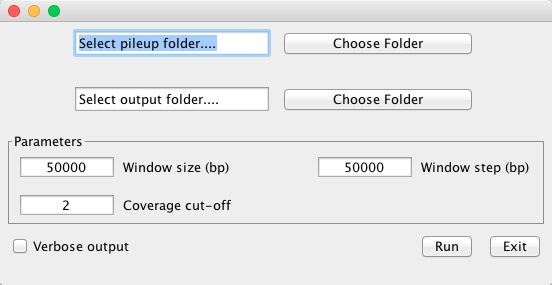


**Figure 4**

Finally you can check the “Create GE files” . Doing so several \_GE files will be generated that can be used with the Genic extractor tool (see below) to investigate whether the found structural variations includes annotated genes.

**Sliding analysis**

Want to just perform a sliding analysis of your chromosomes to estimate the variation of coverage, snps presence and indel presence? Just point to the pileup folder, window size and steps in the Sliding analysis GUI (figure 5) and you are done



**Figure 5**

**Deletion finder**

This tool (figure 6) will allow you to find potential large deletions that can not be found be conventional short reads alignment approach. Please note that in order to call large delations you need to providea sam files that is relative to the alignment of paired ends on the reference. Just provide the folder path where all your sam files are stored (e.g. the Sam folder in the output folder you created when run the “Reads alignment” tool), output file name and the range of variability of the insert size in order for a large deletion to be called.

You can also perform a coverage analysis on the found delations. This is particularly useful in understanding whether the delation is homozygous (i.e. you are expected to find zero or near zero coverage) or heterozygous.

Finally a minimum number of reads pairs that confirm the delation can be chosen in order to output the polymorphism

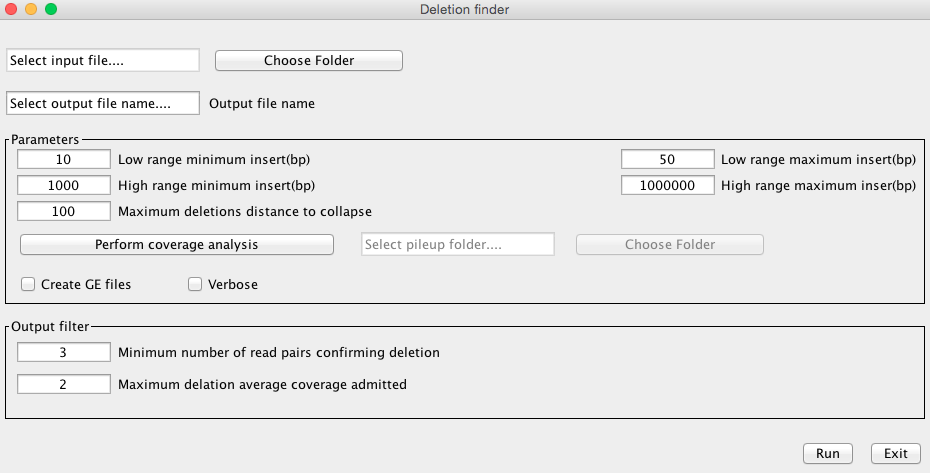


Figure 6

**Polymorphism analyser**

With this tool (figure 7) you can retrieve all snps and indels that are present in the genic regions. A file with several statistics such as the frame, the amino acid change, the position relative to the beginning of the sequence will be reported in output.

In order to perform such analysis you need to provide the pileup folder, an annotation file (gff format reporting the lines relative to “gene”, “mRNA”, “CDS”, and “UTR”), and a folder with 3 files (named cds, 5utr and 3utr) with the fasta formatted sequences of the coding sequences and the two UTR portions. Names in the annotation file and the sequences files should of course match.

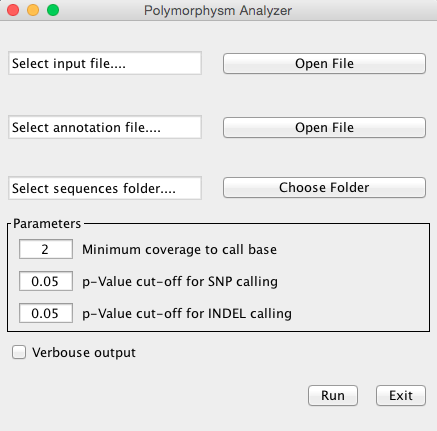


Figure 7

**1:1 Alignment comparison**

This tool (the gui is still work in progress and therefore you will not be able to use it at the moment, but it should be ready very soon) will allow you to insert two pileup folders relative to two resequencing project and receive back information about the peculiar polymorphisms of the two analysed species, plus the common polymorphisms.

**Genic extractor**

This tool (figure 8) is useful when you want to investigate whether the structural variations (found with the other Altools software) contains genic regions. Just provide the \_GE file and the annotation gff file and press the “run” button. A file reporting a flag (1=the gene is contained, 0 = the gene is only partially contained) for each gene containing structural variations will be output.

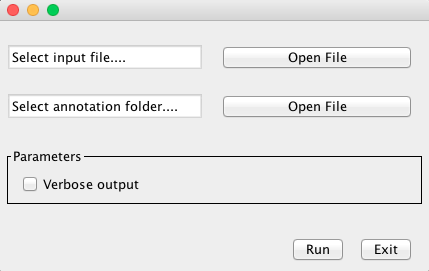


Figure 8