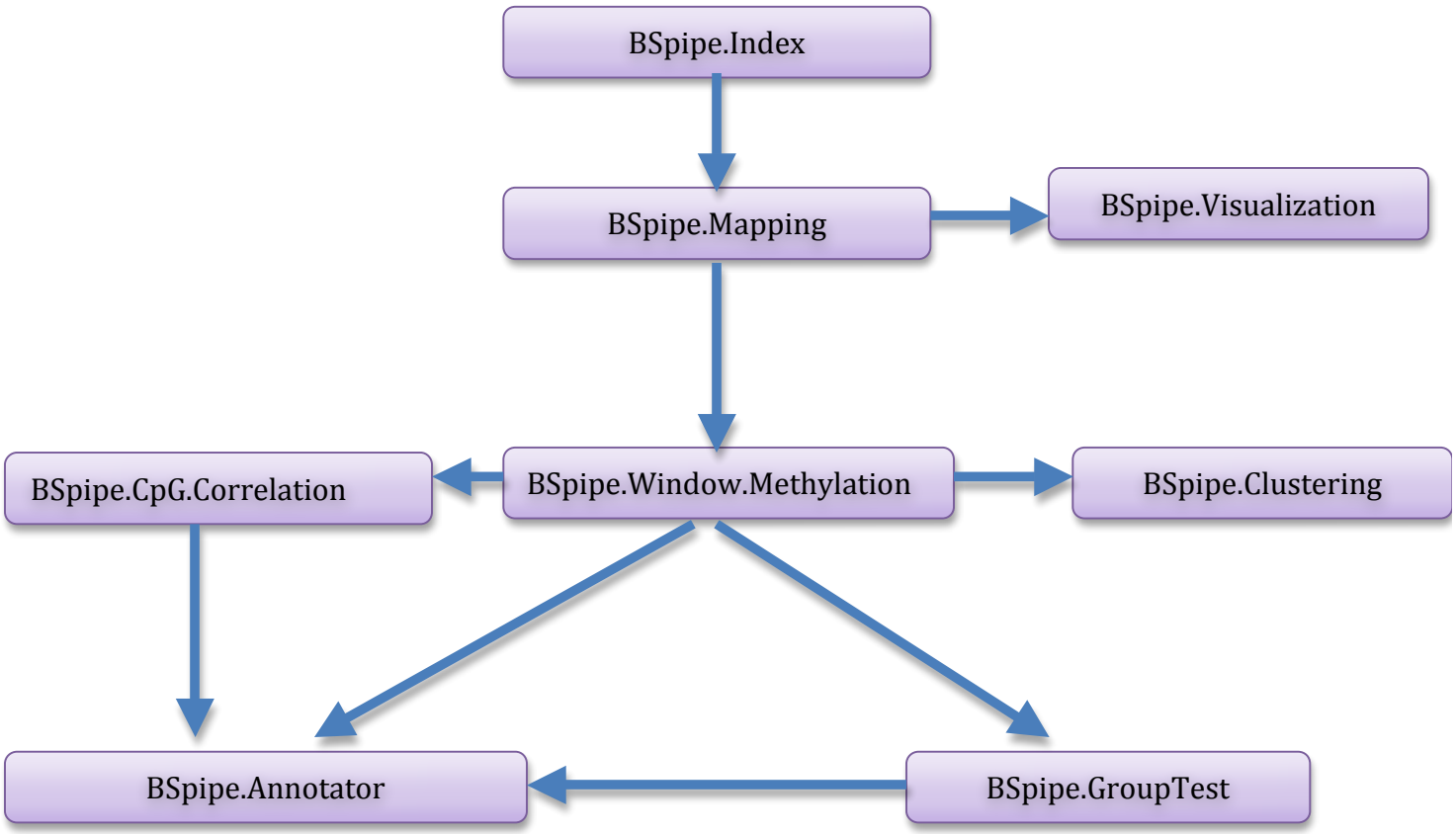
BSpipe.Index Documentation

Description: Creates index files for BSpipe mapping module
Author: GHSU COMICS
BSpipe Version: 1.0

Bisulfite sequencing is a powerful technique to study DNA cytosine methylation. Bisulfite treatment followed by PCR amplification specifically converts unmethylated cytosines to thymine. Coupled with next generation sequencing technology, it is able to detect the methylation status of every cytosine in the genome. BSpipe is an efficient bisulfite mapping software. It comes with a couple of utilities that includes mapping, annotation, calculation of CpG correlation etc.

The following is the flowchart of use of utilities in BSpipe package.



BSpipe.Index is specific utility under BSpipe software for creating index files based on the reference sequence file provided. The module also creates a reference configuration file, which is used by subsequent modules in BSpipe.

Parameters:

Name	**Type**	**Description**
Input.fasta.file	file (.fa / .fasta)	Input reference fasta file whose index files need to be created
Mapping.configuration	tab delimited txt (.conf)	Refer next section
RRBS.name	text	RRBS name in the reference configuration file. [default-'mspi']
Recognition.sequence	text	Recognition sequence denoted by '-' for cutting site. For example, use C-CGG for MspI. [default – 'C-CGG']

Configuration files:

- **Mapping Configuration file**

  This file saves information and parameters for the mapping program to be used (either BWA or BOWTIE or SOAP2). The rows in the file are as follows (a sample bowtie configuration)

  # bowtie configuration

  PROGRAM=bowtie
  WATSON=*parametes in quotes*
  CRICK=*parameters in quotes*
  INDEX=*parameters in quotes*
  PHRED64= --phred64-quals
  PHRED33=

  The above format makes sure different parameters be set for WATSON strand and CRICK strand. A snapshot of the mapping configuration file for bowtie is shown below.

```
  GNU nano 1.3.12

#bowtie configuration

PROGRAM=bowtie
WATSON='--norc -n 2 -e 150 -k 10 --chunkmbs 250 --maxbts 800 --best -t -S REF SEQ'
CRICK='--nofw -n 2 -e 150 -k 10 --chunkmbs 250 --maxbts 800 --best -t -S REF SEQ'
INDEX='bowtie-build -o 3 REF SEQ'
PHRED64=--phred64-quals
PHRED33=
```

  N.B: For convenience, conf files for all 3 mapping programs are available with the uploaded scripts in genepattern. To edit it, users can download the file, make changes and upload it back.

- **Running jobs on SGE**

  Users can submit mapping jobs to the Sun Grid Engine (SGE), by including the SGE configuration file (optional). This option is not set by default. Administrators can edit the wrapper shell script to include the sge.conf file using the prefix '-c' in the command line.

  N.B: For convenience, sge.conf file is available with the uploaded scripts in genepattern, which can be edited by the administrators.

Output files:

The program creates all the files required by BSpipe.Mapping program. This includes the fa.length file, index files based on the mapping program and fasta files for 'c' and 'g' sites. It also creates a reference configuration file (.conf) that can be used in all subsequent BSpipe modules.