

DMR Analysis Package

Contents

1 DMR installation and usage.....	1
2 Configuring DMR.....	3
2.1 workflow.properties.....	3
2.2 mysql.properties.....	4
2.3 DMR_config.properties.....	4
2.4 reference.xml.....	6
3 Contact and Support.....	8

1 DMR installation and usage

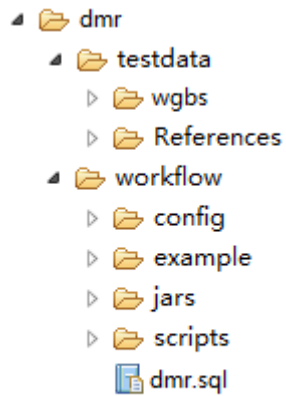
1. Install the required software and configure the environment. (see the prerequisites document)

2. Download the dmr_pipeline.tar.gz from WBSA website (<http://bsap.big.ac.cn/download/download.jsp>) put it under a given directory such as /home/test

Note: In the following paragraph we will use “/home/test” to make examples.

```
cd /home/test  
  
tar xzvf dmr_pipeline.tar.gz
```

After uncompressed, you will see this directory structures like the following figure:



Note: The files in the “dmr/workflow/example” use “/home/test” as relative directory.

Folder name	Description
testdata/wgbs	required ,a test data for test this package
testdata/References	required ,a test reference for test the package which need to be configured in the reference.xml
workflow/config	contains configuration template for the package
workflow/example	required ,contains all the configuration files to run the package
workflow/jars	required , contains all the jars to run the package
workflow/scripts	required , contains all the scripts to run the package
dmr.sql	required ,a database schema need to load into MySQL database

3. Import the MySQL script.

You need to login into Mysql database as root, by default the password is empty, so just click enter.

```
mysql -uroot -p
```

```
mysql> create database dmr;

mysql>source /home/test/dmr/workflow/dmr.sql;

mysql> grant all on dmr.* to 'pipeline'@'localhost' identified by 'pipeline'
```

4. Run the DMR module

- 1) Prepare the references;
- 2) Fulfill the configuration files.
- 3) Run the command

```
java -jar /home/test/dmr/workflow/jars/workflow.jar \
/home/test/dmr/workflow/jars/example/DMR_config.xml \
/home/test/dmr/workflow/ example /config
```

2 Configuring DMR

The package uses a number of configuration files to configure a process, the configuration files in the “example” directory.

- workflow.properties
- mysql.properties
- dmr_config.xml
- reference.xml

2.1 workflow.properties

Note: the parameters are identified by bold font you need to change to your real directory. This is a PBS package, so you need to configure the cluster.submit.queueName parameter otherwise the whole package will not run correctly.

```
path.result=/home/test/dmr/workflow/ example/result
path.perl=/home/test/dmr/workflow/ scripts
path.reference=/home/test/dmr/workflow/ example/config/reference.xml
path.config=/home/test/dmr/workflow/ example/config

database.use=1
cluster.name=PBS
cluster.submit.queueName=workq
```

Table 1: workflow.properties

Parameter	Description	Example
path.result	Absolute path of the result directory	result

path.perl	Absolute path of the Perl script directory	scripts
path.reference	Absolute path of the reference.xml file	reference.xml
path.config	Absolute path of directory of the configuration files	config
cluster.name	Type of the cluster, we use PBS in this workflow.	PBS
cluster.submit.queue name	Work queue name in PBS used to submit jobs.	workq

2.2 mysql.properties

Note: If the Mysql database which you have installed is not located in your current machine, you need change "localhost" to the IP of that machine.

```
jdbc.driverClassName=com.mysql.jdbc.Driver

jdbc.url=jdbc:mysql://localhost/dmr?user=pipeline&password=pipeline&useUnicode=true&characterEncoding=UTF-8&autoReconnect=true
```

2.3 DMR_config.properties

Note: The parameters sampleOne ,sampleTwo refer to the WGBS's finished jobid, and they must locate in path.result/wgbs directory(path.result from 2.1). We use these two jobid's result to do DMR analysis.

```
pipeline.type=dmr
process=5
refId=d1
sampleOne= 3333331
sampleTwo= 3333332
cFdrPvalue= 0.01
cFdr=n
cc=0.8
cf=0.2
cNumber= 70
cNumberMove= 5
cPvalueForWilcox= 0.01
cNumberm=0.8
cNumberf=0.2
cgFdrPvalue= 0.01
cgFdr=n
cgC=0.8
```

```

cgf=0.2
cgNumber=70
cgNumberMove=5
cgPvalueForWilcox=0.01
cgNumberm=0.8
cgNumberf=0.2
chWinLen= 5000
chWinNumber=10
chFdrPvalue=0.01
chFdr=n
chFoldValue=2
chc=0.8
jar.package=/home/test/dmr/workflow/jars/dmr.jar
jar.html=/home/test/dmr /workflow/jars/pipelinehtml.jar

```

The following table gives all the configurable parameters.

Parameter	Description	Example
pipeline.type	pipeline type, must be dmr	dmr
process	Number of processes used during the processing	5
SampleOne	Job ID of sample 1.	123456
SmampleTwo	Job ID of sample 2.	123457
refId	Reference ID for the data, which is included in the file reference.xml.	1
cFdrPvalue	for C static window method,p-value	0.01
cFdr	for C static window method, FDR correction, y:yes, n:no	n y
cc	for C static window method, the minimum sequence coverage rate	0.8
cf	for C static window method, the minimum delta methylation level	0.2
cNumber	for C dynamic window method, C base number in one window	70
cNumberMove	for C dynamic window method ,C number in a step size	5
cPvalueForWilcox	for C dynamic window method ,p-value	0.01
cNumberm	for C dynamic window method , the minimum percent of C base coverage in one test window	0.8
cNumberf	for C dynamic window method , the minimum delta methylation level	0.2
cgFdrPvalue	for CG static window method ,p-value	0.01

cgFdr	for CG static window method ,FDR correction, y:yes, n:no	n y
cgc	for CG static window method ,the minimum sequence coverage rate	0.8
cgf	for CG static window method , the minimum delta methylation level	0.2
cgNumber	for CG dynamic window method, CG number in one window	70
cgNumberMove	for CG dynamic window method, CG number in a step	5
cgPvalueForWilcox	for CG dynamic window method,p-value	0.01
cgNumbererm	for CG dynamic window method, the minimum percent of CG coverage in one test window	0.8
cgNumberf	for CG dynamic window method, the minimum delta methylation level	0.2
chWinLen	for CH static window method ,base number in the window	5000
chWinNumber	for CH static window method ,considered window number	10
chFdrPvalue	for CH static window method ,p-value	0.01
chFdr	for CH static window method ,FDR correction, y:yes, n:no	y n
chFoldValue	for CH static window method ,fold change of the methylation level between samples	0.2
chc	for CH static window method , the minimum sequence coverage rate	0.8
jar.package	for DMR pipeline, must be dmr.jar	
jar.html	to generate html page in the result directory, must be pipelinehtml.jar	

2.4 reference.xml

Note: The parameter which identified by bold font you need to change to satisfy your real demands, for the left parameters and the whole XML format you should not change it in order to run the package correctly.

If you have multiple references, just add the <reference> element as the example file.

```
<?xml version="1.0" encoding="UTF-8"?>
```

```

<references>
<reference id="d1" name="Rice" >
  <params>
    <param name="refDir">/home/test/dmr/testdata/References/rice/ref</param>
    <param
name="refC2TG2AFile">/home/test/dmr/testdata/References/rice/Ref_C-T_G-A/ref_all_C-T_G
-A.fa</param>
    <param
name="genelistChromDir">/home/test/dmr/testdata/References/rice/genes</param>
    <param name="repeatDir">/home/test/dmr/testdata/References/rice/repeats</param>
    <param name="goNumberFile">/home/test/dmr
/testdata/References/rice/GO/rice_gene_GO.txt</param>
    <param
name="geneOntologyFile">/home/test/dmr/testdata/References/rice/GO/gene_ontology_ext
.obo</param>
    <param name="refCGIDir">/home/test/dmr/testdata/References/rice/cpgislands</param>
    <param
name="ideogramDir">/home/test/dmr/testdata/References/rice/ideogram</param>
  </params>
</reference>
</references>

```

The following table gives all the configurable parameters.

Param	Description	Example
refDir	Absolute path of directory of raw reference files that must be .fa format and named chr*.fa	/home/test/dmr/testdata/References/rice/ref
refC2TG2AFile	Absolute path of reference file that has merged all the raw reference files and converted C to T and G to A. The file must have been indexed by BWA before used.	/home/test/dmr/testdata/References/rice/Ref_C-T_G-A/ref_all_C-T_G-A.fa
genelistChromDir	Absolute path of gene file directory. The gene file is downloaded from UCSC. The genes on forward strand of each chromosome are stored in a file named chr*_C-T.gene. The genes on reverse strand of each chromosome are stored in a file named chr*_G-A.gene. #geneid #geneid #chromosome #strand #start #end #start #end #exon number #first exon start pos #second exon start	/home/test/dmr/testdata/References/rice/genes

	<p>pos #first exon end pos #second exon end pos</p> <p>All the start positions are count from 0 and the end positions are count from 1.</p> <p>NOTE: In condition on DMR analysis, all the genes on each chromosome are merged into one file named all_chr_genes.</p>	
repeatDir	Absolute path of repeat file directory. It is downloaded from UCSC and in the standard format.	/home/test/dmr/te stdata /References/rice/re peats
goNumberFile	File with GO information. Each line begins with a gene id followed by the GO number.	/home/test/dmr/te stdata /References/rice/G O/rice_gene_GO.txt
geneOntologyFile	Full ontology file, including cross-products, inter-ontology links, and has part relationships. It could be downloaded at http://www.geneontology.org	/home/test/dmr/te stdata /References/rice/G O/gene_ontology_e xt.obo
refCGIDir	Absolute path of CpG islands file directory. It is downloaded from UCSC and in the standard format.	/home/test/dmr/te stdata /References/rice/cp gislands
ideogramDir	Absolute path of ideogram files of the species including: chromosomes.png, chromosomes.map.	/home/test/dmr/te stdata /References/rice/id eogram

3 Contact and Support

DMR Analysis package is developed and maintained by [Beijing Institute of Genomics\(BIG\)](#), Chinese Academy of Sciences. If you have feedback or questions, please feel free to contact us at bsap@big.ac.cn.