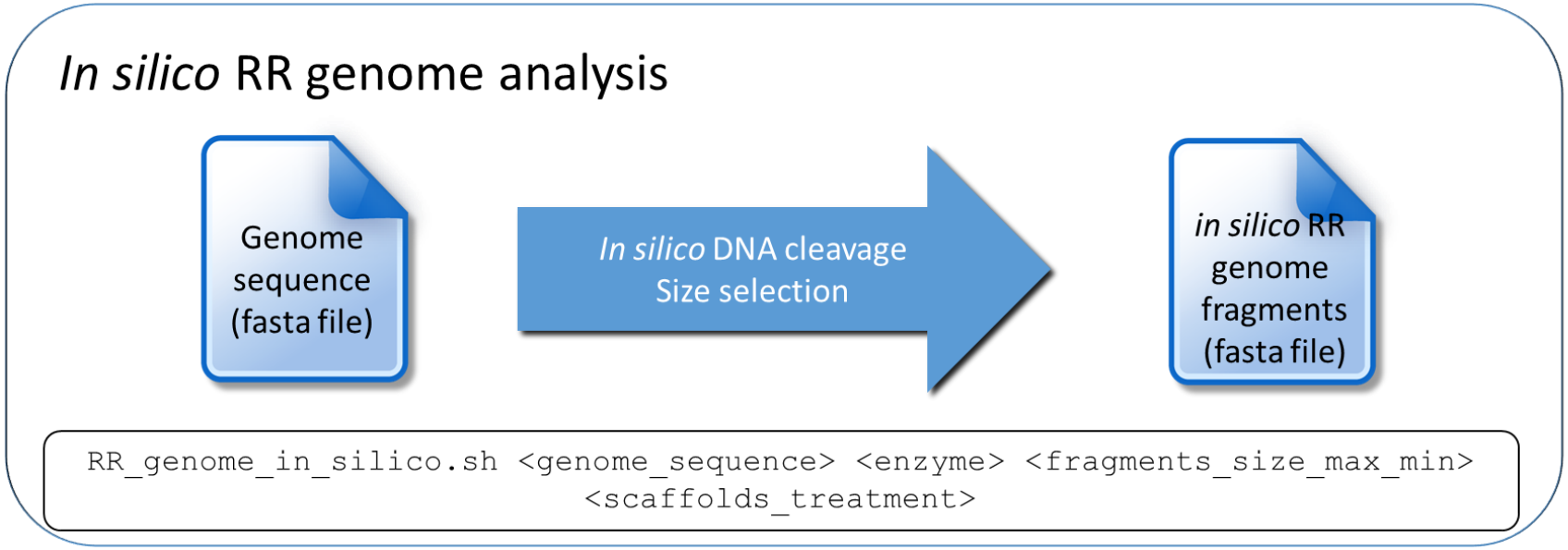
In silico RR genome analysis : Principle

This process allows to optimize the experimental design according to the number of covered CpGs and their genomic annotation.

Genomic DNA is *in silico* digested with restriction enzymes into fragments. These collection of fragments or “RR genome” can be selected according to their size.

An output report file displaying the features of this RR genome is also performed. Be careful, this report is only valid if Msp1, XmaC1 or BssS1 have been used to *in silico* digest the genome.



In silico RR genome analysis : Input / Output

INPUT :

The script takes as input a genomic fasta file along with some parameters :

Enzyme recognition sequence or recognition site (i.e. “CCGG” for MspI)
Fragment size selection :
if selection : enter the fragment minimal size followed by the fragment maximal size
if no selection : enter “0” following by “-1”
Keep scaffolds
if yes, enter “1”. If no, enter “0” (the default value is “0”)

OUTPUT:

- A fragment or “RR genome” fasta file.
- A log file reminding parameter selection and some features such as
- A text report file displaying the following results:
  - Name of the sample
  - RR genome size
  - Percentage of whole genome
  - Number of fragments
  - Number of CpG sites (RR genome)
  - Percentage of total genomic CpG sites

In silico RR genome analysis : Example #1

Genomic DNA “Sus_scrofa.Sscrofa10.2.chr18.fa” digested by MspI (“CCGG” recognition sequence) with fragment selection between 40 and 220, scaffold sequences are discarded :

- RRBS_HOME/RR_genome/RR_genome_in_silico.sh A.fasta CCGG 40 220

Output :
- Sus_scrofa.Sscrofa10.2.chr18_frag_in_silico_CCGG_40_220.fasta: fasta file with fragments

- Sus_scrofa.Sscrofa10.2.chr18_frag_in_silico_CCGG_40_220.fasta_results.txt:

Sample	RR genome size	% of whole genome	number of fragments	number of CpG sites (RR genome)	% of total genomic CpG sites
Sus_scrofa.Sscrofa10.2.chr18_frag_in_silico_CCGG_40_220.fasta	2080380	3.4	18583	104402	12.4

- RR_genome_in_silico.log:
Input file : Sus_scrofa.Sscrofa10.2.chr18.fa
Restriction site : CCGG
Size selection : [40;220]
Treat scaffold : No
Output file : /work/fpiumi/Test/Sus_scrofa.Sscrofa10.2.chr18_frag_in_silico_CCGG_40_220.fasta
Fragments total number = 18583
Empty fragments total number = 279
Sorting initialization...
fasta RR dig ok
RR parameters ok

In silico RR genome analysis : Example #2

Genomic DNA “Sus_scrofa.Sscrofa10.2.chr18.fa” digested by MspI (“CCGG” recognition sequence) with no fragment selection and with kept scaffolds :

- RRBS_HOME/RR_genome**/**RR_genome_in_silico.sh Sus_scrofa.Sscrofa10.2.chr18.fa CCGG 0 -1 1

Output :
- Sus_scrofa.Sscrofa10.2.chr18_frag_in_silico_CCGG.fasta : fasta file with fragments

- Sus_scrofa.Sscrofa10.2.chr18_frag_in_silico_CCGG.fasta_results.txt :

Sample	RR genome size	% of whole genome	number of fragments	number of CpG sites (RR genome)	% of total genomic CpG sites
Sus_scrofa.Sscrofa10.2.chr18_frag_in_silico_CCGG.fasta	61220071	100.0	64644	842519	100.0

- RR_genome_in_silico.log:
Input file : Sus_scrofa.Sscrofa10.2.chr18.fa
Restriction site : CCGG
Size selection : No selection
Treat scaffold : Yes
Output file : /work/fpiumi/Test/Sus_scrofa.Sscrofa10.2.chr18_frag_in_silico_CCGG.fasta
Fragments total number = 64644
Empty fragments total number = 279
Sorting initialization...
fasta RR dig ok
RR parameters ok