ER-flow Application Description Template

Application Name:
Simulation of low coverage sequence experiments

Application domain:
Bioinformatics, DNA sequence analysis

Brief description of application
Randomly samples alignments from a sequence alignment file. It is used to simulate low coverage sequence experiments.

Inputs:
PREFIX – Helper file for the workflow. It is a text file containing the word "BAM". It is used to deal with internal file names in WS-PGRADE/gUse
VALUES – Text file containing the path(s) to sequence alignment file(s). The sequence alignment files are in bam format
PERCENTAGE – text file containing number(s). These are the fractions that you would like to sample from the alignment file(s). E.g. 0.1 (10%), 0.5 (50%)

Outputs:
BAM – Text file which contains the paths to the output sequence alignment file(s), i.e. the sampled fraction of the original file(s)

data:
input data format: bam input data value range
output data format: bam output data value range
sample data: (link) to be determined
application (link) Picard DownsampleSam
documentation (link) http://picard.sourceforge.net/
publication (link)

Execution environment
DCI: (computing, data, VO, etc)
middleware: glite, vlemed workflow system: guse

Execution characteristics
data size (per unit, typical number of units):
input 8-50GB temporary range: 10-100% of input data output
processing time (per unit): depends on input size and amount of sequences that are sampled from this file
memory usage: tbd disk usage: input plus a fraction of the input

Target users
Community, projects: (link) http://www.bioinformaticslaboratory.nl/
number of users: 1, but can be of interest for other studies

user type: developer+ end-user

Usage scenario for workflow in the ER-FLOW (how workflow will be reused, metaworkflow, how expected to contribute to project indicators, etc.).

Contact information (author)
name: Barbera van Schaik
e-mail b.d.vanschaik@amc.uva.nl, support@ebioscience.amc.nl
Bioinformatics Laboratory, KEBB, Academic Medical Center, Amsterdam, NL