Title: Workflow-BS an integrative workflow for RRBS and WGBS data. From the BS-seq to the DMR

Authors:
Gaelle Lefort[1], Marjorie Mersch[2], Sylvain Foissac[2], Frédérique Pitel[2], Céline Noirot[1].
[1]*INRA - UR875 MIAT – 31326 CASTANET-TOLOSAN*
[2]*INRA – UMR1388 GenPhySE – 31326 CASTANET-TOLOSAN*

Abstract (300 words max):
DNA methylation is an epigenetic mark that has suspected regulatory roles in a broad range of biological processes and diseases. The technology is now available for genome-wide methylation studies, at a high resolution and with possibly a large number of samples. Many specific aligners for BS-seq data exist, such as BSMAP and Bismark. Also, R packages (methylKit and DSS) were designed to detect differentially methylated cytosines (DMC) and differentially methylated regions (DMR). Methy-Pipe (Peiyong Jiang *et al.* 2014. *PLOS one*) fill the gap between those analyses by combining a complete pipeline from raw data to statistical outputs but it requires a specific cluster environment (SGE software). Here, we propose a workflow which deals with fastq files from BS-seq (WGBS and RRBS) and goes through all steps to provide bed files of DMC and DMR. It can support most distributed resource management systems (Condor, SGE, …).

Our pipeline uses standard software to i) clean data ii) align WGBS or RRBS reads to a reference genome iii) extract methylation and iv) identify DMC and DMR. Raw data are cleaned with Trim_galore and aligned with Bismark. The base-resolution methylation level is extracted by context and sample with methylKit. If a SNP file is provided, its polymorphic positions are removed from the analysis. Several tests to detect DMC and DMR are then performed, according to the experimental design supplied by the user. Statistics and graphics are also provided. As the pipeline is based on Jflow (Mariette *et al.* 2015. *Bioinformatics)*, it can be used on command line or through a web server. Adding a new aligner or a new component has been made as simple as possible for future evolution of the tool.

We will present results obtained by using this pipeline on chicken and plant genomes.

Keywords:
DNA methylation
pipeline
web interface
WGBS and RRBS alignements
DMC and DMR identification