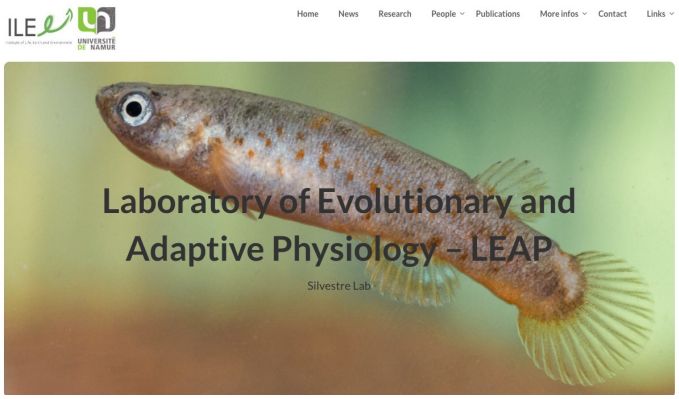
Bisufite sequencing analysis
Bioinformatic workshop

Frédéric Silvestre

Laboratory of Evolutionary and Adaptive Physiology (LEAP) – URBE

University of Namur



www.evolution-physiology.be

LEAP
Laboratory of Evolutionary
and Adaptive Physiology

https://github.com/fredsilvestre/bioinfoworkshop_epigenetics

- R Markdown scripts to be fulfilled parts 1 to 3 (copy/paste in Rstudio)
- Dataset for the exercise
- Context: « README.md »
- Studentspack
- Introduction slides

Goals of the workshop:

- Learn how to use Bioconductor for bioinformatic analyses
- Learn how to use R Markdown
- Learn how to work with DNA sequences in RStudio and with genomic objects
- Learn how to analyse a methylome from bisulfite sequencing
- Work on real data

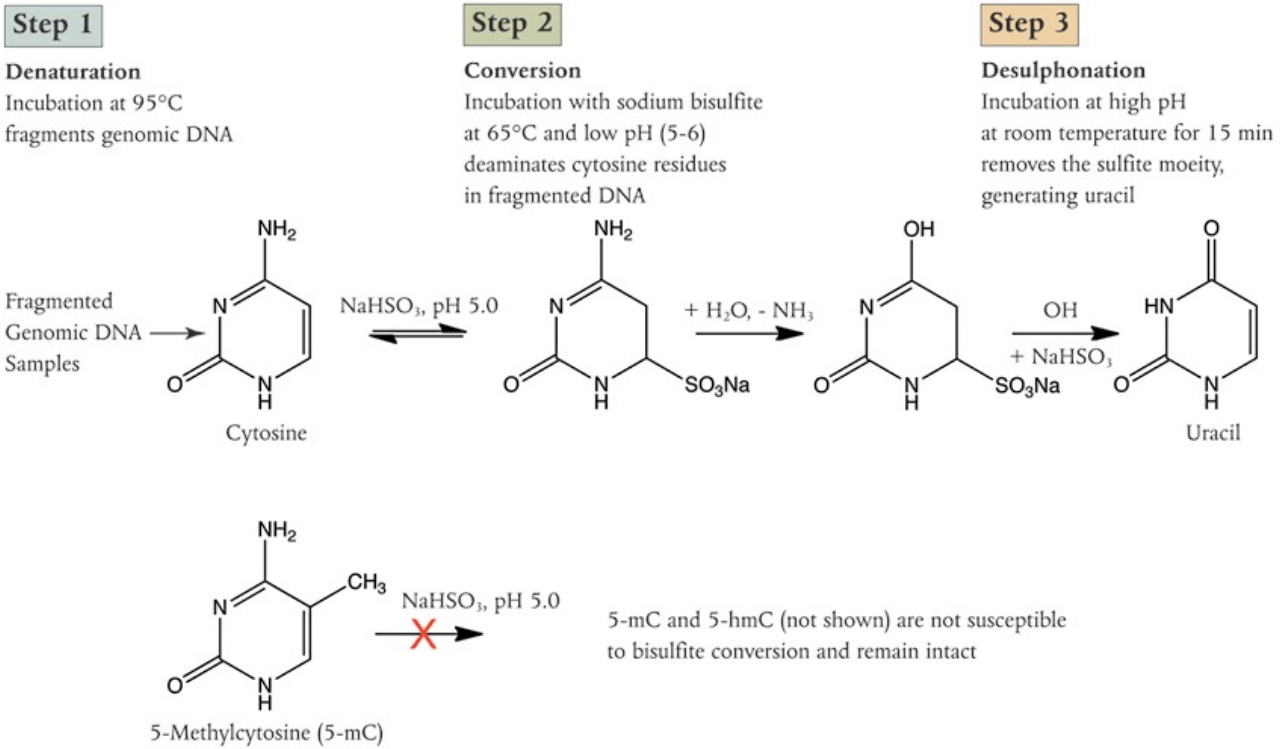
Before the workshop:

- Install R and Rstudio (the most recent versions)
- Run Rstudio and try to get used with the environment ; try to upload some packages on CRAN
- Download the files in the studentspack file on github

Few online references:

To get familiar with R: https://www.statmethods.net/index.html
Discussion forum about bioinformatics: https://stackoverflow.com/
Reference book: http://compgenomr.github.io/book/ (mostly Ch10)

Bisulfite conversion



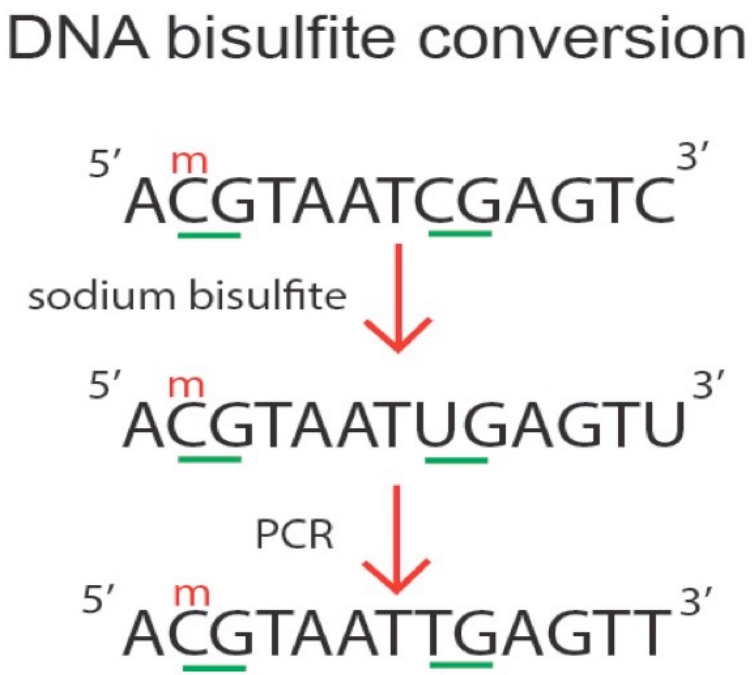


Table 1. Sequences resulting from bisulfite conversion and PCR

	Original sequence	After bisulfite treatment	After PCR amplification
Unmethylated DNA	A-C-G-T-C-G-T-C-A	A-U-G-T-U-G-T-U-A	A-T-G-T-T-G-T-T-A
Methylated DNA	A-C-G-T-C-G-T-C-A	A-C-G-T-C-G-T-U-A	A-C-G-T-C-G-T-T-A

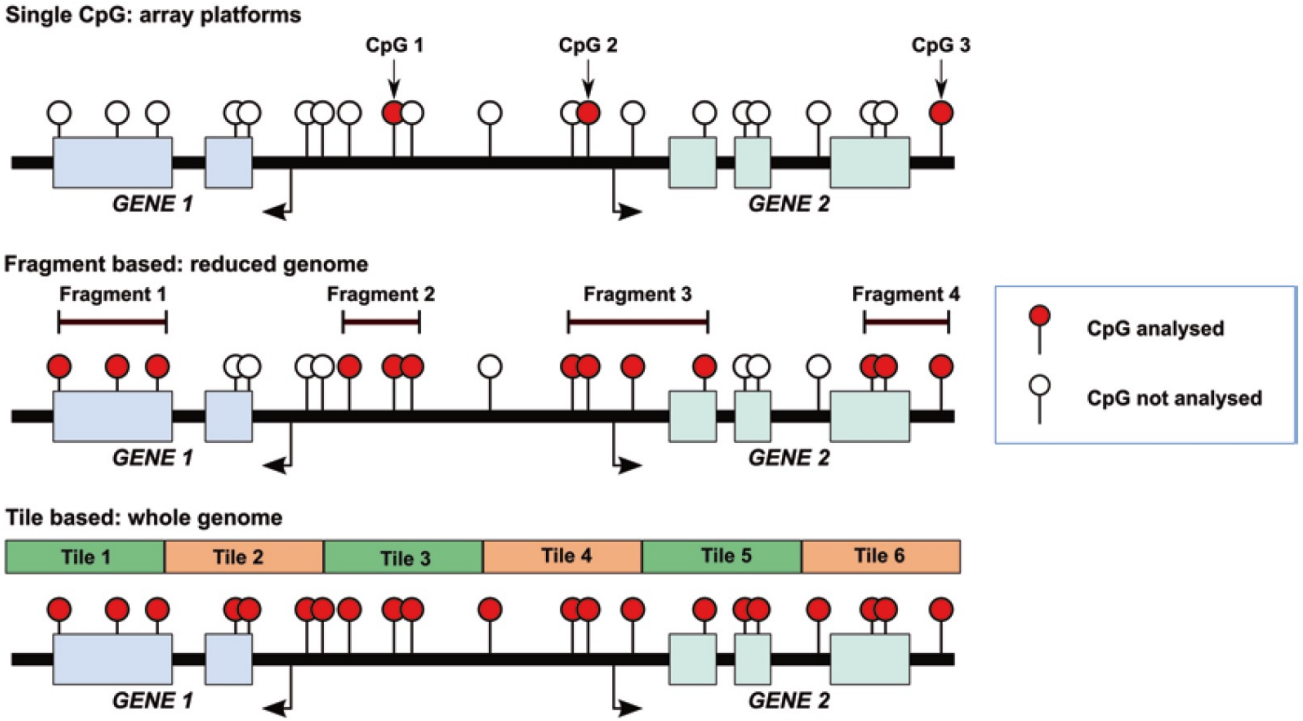
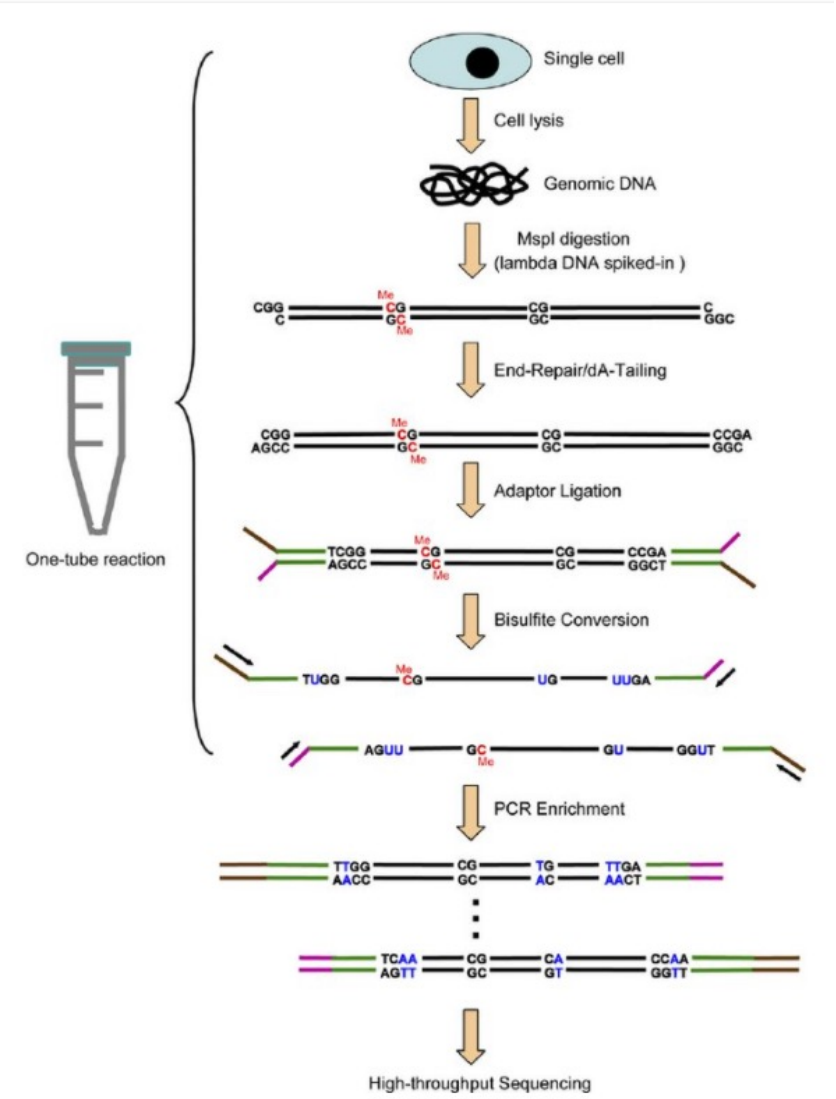


Fig. 1 Major analysis approaches for genome-wide DNA methylation analysis. There are several approaches for analyzing differential methylation between different groups and conditions. These approaches differ based on the unit of analysis: (1) the single CpG site approach independently analyzes each CpG site in investigated samples; (2) for RRBS, MspI-digested fragments can be used as the unit of analysis (implemented in DMAP [32] package); (3) a common approach for large bisulfite sequencing data is to investigate regions with fixed size genomic windows. It is possible to use sliding windows based on user-specified criteria

RRBS workflow



The workflow of bioinformatic analysis:

10.2 Analyzing DNA methylation data

For the remainder of this chapter, we will explain how to do DNA methylation analysis using R. The analysis process is somewhat similar to the analysis patterns observed in other sequencing data analyses. The process can be chunked to four main parts with further sub-chunks:

1. Processing raw data

- Quality check
- Alignment and post-alignment processing
- Methylation calling
- Filtering bases

2. Exploratory analysis

- Clustering
- PCA

3. Finding interesting regions

- Differential methylation
- Methylation segmentation

4. Annotating interesting regions

- Nearest genes
- Annotation with other genomic features
- Integration with other quantitative genomics data

Scientific context for part 3: mangrove rivulus populations

The mangrove rivulus: surviving in the mangroves

Kryptolebias marmoratus

The mangrove rivulus: surviving in the mangroves

Rivulus microhabitat and sampling

The mangrove rivulus: surviving in the mangroves

Simultaneous Hermaphrodites

Males (1-25%)
(Primary and secondary males)

K. marmoratus = Only known self-fertilizing vertebrate

= androdioecy
(≠ parthenogenesis)

The mangrove rivulus: surviving in the mangroves

Simultaneous Hermaphrodites

Outcrossing

Males (1-25%)
(Primary and secondary males)

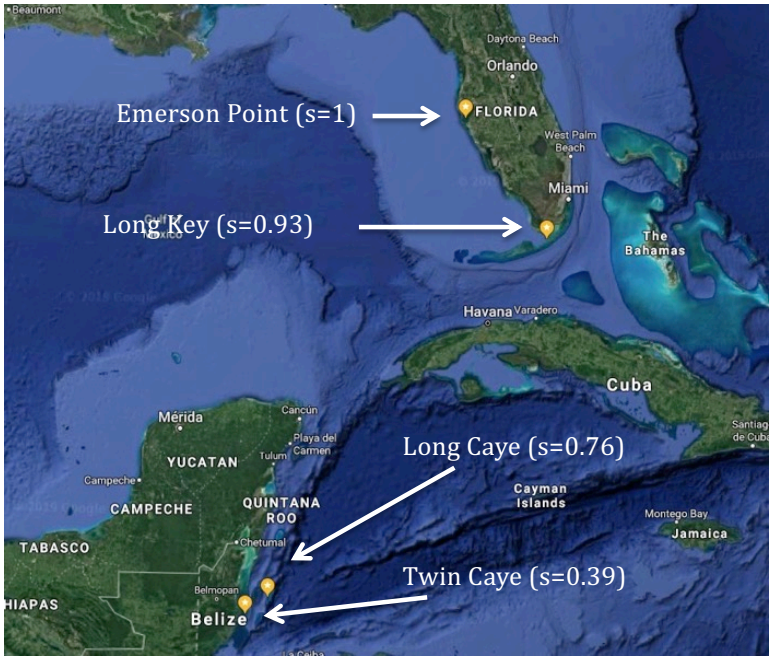
Variable OC (low OC in Florida ; higher in Belize)

Self-fertilization

The mangrove rivulus: surviving in the mangroves

Scientific field missions 2019: Belize and Florida

Valentine Chapelle thesis (FRIA since 10/2018): "*The epigenetic origin of behavioral traits variability in a self-fertilizing fish : the mangrove rivulus*"



The mangrove rivulus: surviving in the mangroves

The mangrove rivulus: surviving in the mangroves

Population	N total fish	N males	N herma	N juveniles	% of male
Twin Caye (PG)	177	74	103	0	**41.8**
Long Caye (LC)	31	3	26	2	10.3
Emerson Point (EPP)	540	2	538	0	0.4
Long Key (LK)	44	1	35	8	2.8