**Overview and development of the GBS-MeDIP protocol:**

High- throughput DNA genetic and epigenetic profiling can help to understand genome-wide variations such as SNPs, CNVs, and differentiated methylated regions (DMR) to ramp up the identification of markers of susceptibility to diseases such as cancer. Although the cost of whole genome interrogation across many individuals is still a limiting factor, reduced representation libraries allow to reduce sequencing costs in this scenario30.

We integrated MeDIP with GBS to simultaneously investigate genetic and methylomic variability across reduced individual genomes. The idea was to take advantage of the fact that samples used as the input of the GBS are barcoded before the creation of a pooled library1,2. GBS is a genotyping approach that uses RE to reduce the genome complexity followed by the addition of barcodes to individual samples. In the current protocol, we used a previously optimized3 *PstI* digestion of the genomic DNA to generate fragments with overhangs. The use of RE allows to access regions of the genome that are not easily reached by sequence capture methods1. This happens because lower copy number regions (common in species with high genetic variations1) are enriched with much higher efficiency when using a RE compared with random shearing.

*PstI* is a RE that recognizes the 5’CTGCAG 3’ motif and cuts the genome in a methylation insensitive manner, as this motif does not contain CpG sites3. Initially, we evaluated different REs both *in silico* and *in vitro*, including *PstI*, *MseI*, *ApeKI, SbfI, MspI*, and *EcoRI*. Among these enzymes tested, the most efficient in enriching regions of the human genome in a suitable length size (ranging from 200 to 500 bps) for sequencing was the *PstI*. Previously, in a study in chickens, we observed a broad distribution of the *PstI* generated fragments across the chicken genome3. Subsequently, *in silico* digestions of the genomes of six different species, namely*Danio rerio, Gallus gallus, Mus musculus,* *Homo sapiens*, *Sus scrofa* and *Canis familiaris* were performed using the *PstI* restriction site (Figure 1, “a” to “f”). The selection of these species was based on the animal models we investigate3-10 and species widely used in genomic studies11,12. The fragment pattern obtained by these *in silico* digestions indicates that *PstI* cleavage of these genomes (including the human) produces high quantities of fragments of the desired length (ranging 200-500bp) for Illumina sequencing platforms (Figure 1). Importantly, the pattern of *in silico* *PstI* cleavage of the human genome (Figures 1b and d)is similar to that of the chicken, in which GBS has been recently optimized3.

Additionally, we performed *in silico* digestions of the same five model specie genomes considering only fragments with at least one CpG in their composition. This analysis simulates what would happen if an entire genome would be methylated to the fullest, then cleaved by *PstI*, and subsequently subjected to MeDIP. With the human genome the *PstI* *in silico* digestion generated 178,002 fragments ranging 200-500 bp, which corresponds to 14% of all the fragments obtained. Interesting, more than 90% of the fragments ranging 200-500bp have at least one CpG dinucleotide in their composition (Figure 1g). This shows an enrichment of CpG fragments after *PstI* cleavage as recently reported4. In comparison, *PstI* digestion of the chicken genome generated 171,959 fragments with at least one CpG, which corresponds to ~20% of all the fragments generated. Importantly, fragments smaller than 200 bp produced by the *Pst1* cleavage are less prone to be captured by the MeDIP (Figure 1). Thus, the MeDIP step pre-selects for larger fragments that are suitable for Illumina sequencing.

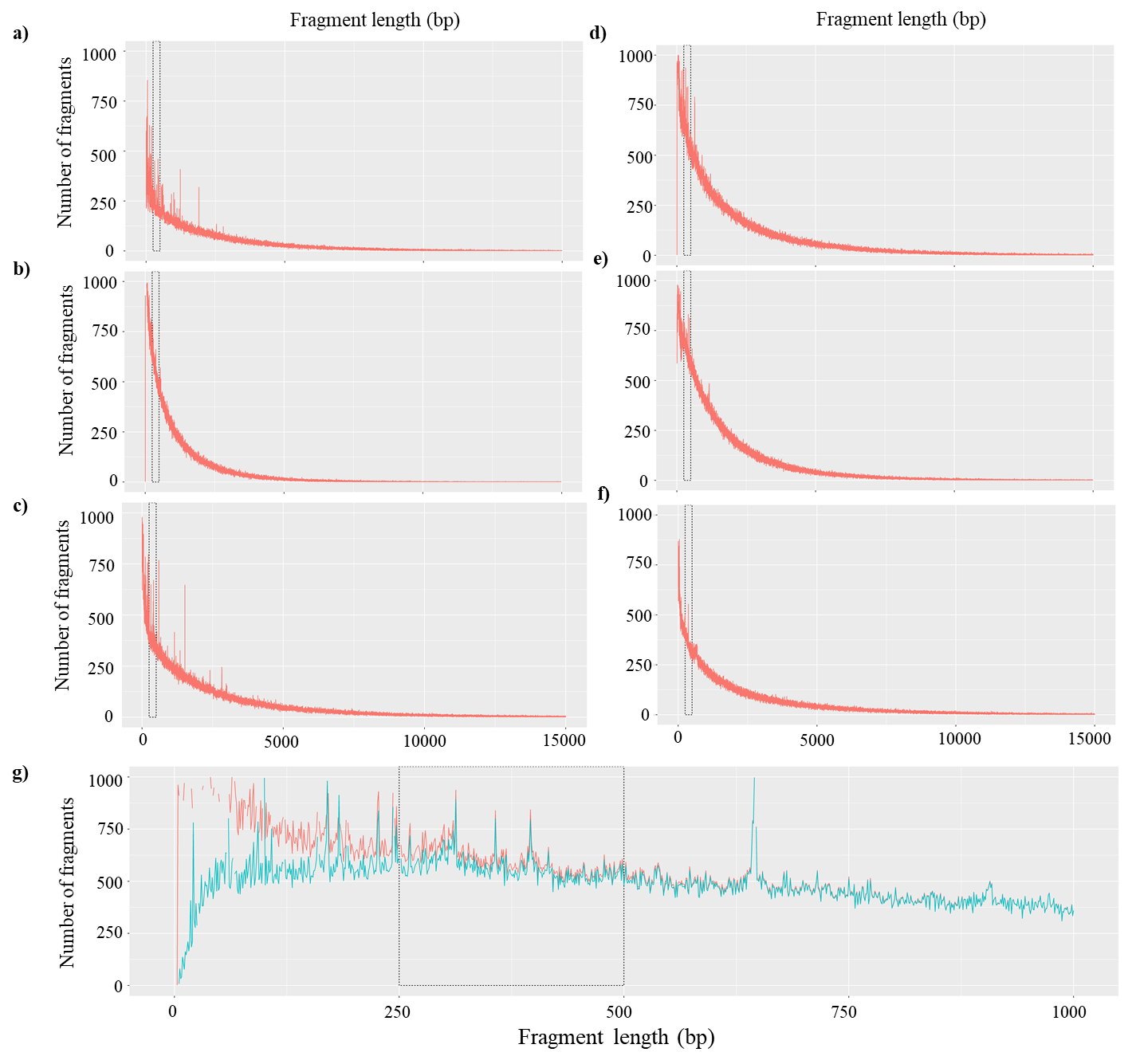


Figure 1: *In silico* *Pst1* digestion prediction of six different reference genomes : (a) *D.rerio*; (b) *G. gallus*; (c) *M. musculus*; (d) *H. sapiens*; (e) *S. scrofa* and (f) *C. familiaris*. (g) represents the *H. sapiens Pst1 in silico* digestion of the reference genome (red) and a subset with fragments containing at least one CpG in their composition (blue). The region of interest (200-500bp) is indicated by the box with the black dotted line.

After the DNA is fully digested by the enzyme, two Illumina adapters are added at the ends of each digested DNA fragment. This is a crucial step for the clustering of the Illumina probe within the sequencing flowcell. One of the adapters is called the ‘common adapter’, which is the same for all the samples. The other is called the ‘barcode adapter’, which allows for the later interindividual identification of the sequences through bioinformatic analyses 2. After the ligation of the adapters, the samples are pooled. Subsequently, the methylated fraction of the whole pool is captured by MeDIP. This captured genomic fraction enriched for DNA methylation is then PCR amplified. However, only DNA regions containing both adapters are able to be amplified and subsequently sequenced2.

**References:**

1 Elshire, R. J. *et al.* A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS one* **6**, e19379, doi:10.1371/journal.pone.0019379 (2011).

2 Poland, J. A. & Rife, T. W. Genotyping-by-Sequencing for Plant Breeding and Genetics. *The Plant Genome* **5**, 92–102, doi:10.3835/plantgenome2012.05.0005 (2012).

3 Pertille, F. *et al.* High-throughput and Cost-effective Chicken Genotyping Using Next-Generation Sequencing. *Scientific Reports* **6**, doi:10.1038/srep26929 (2016).

4 Pértille, F. *et al.* Putative Epigenetic Biomarkers of Stress in Red Blood Cells of Chickens Reared Across Different Biomes. *Frontiers in Genetics* **11**, 1-44, doi:10.3389/fgene.2020.508809 (2020).

5 Guerrero-Bosagna, C. *et al.* DNA methylation variation in the brain of laying hens in relation to differential behavioral patterns. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*, 100700-100700, doi:10.1016/j.cbd.2020.100700 (2020).

6 Nunes, J. R. S. *et al.* Genome-wide association study reveals genes associated with the absence of intermuscular bones in tambaqui (Colossoma macropomum). *Animal Genetics*, doi:10.1111/age.13001 (2020).

7 Pertille, F. *et al.* Mutation dynamics of CpG dinucleotides during a recent event of vertebrate diversification. *Epigenetics* **14**, 685-707, doi:10.1080/15592294.2019.1609868

10.1080/15592294.2019.1609868. Epub 2019 May 9. (2019).

8 Sundman, A.-s. *et al.* DNA methylation in canine brains is related to domestication and dog-breed formation. *PLOS ONE* **15**, e0240787-e0240787, doi:10.1371/journal.pone.0240787 (2020).

9 Pertille, F. *et al.* Genome-wide association study for performance traits in chickens using genotype by sequencing approach. *Sci Rep* **7**, 41748-41748, doi:10.1038/srep41748

10.1038/srep41748. (2017).

10 Pértille, F. *et al.* Sperm Methylome Profiling Can Discern Fertility Levels in the Porcine Biomedical Model. *International Journal of Molecular Sciences* **22**, 2679 (2021).

11 Demars, J. *et al.* High resolution physical map of porcine chromosome 7 QTL region and comparative mapping of this region among vertebrate genomes. *BMC genomics* **7**, 13, doi:10.1186/1471-2164-7-13 (2006).

12 Ouedraogo, M. *et al.* The duplicated genes database: identification and functional annotation of co-localised duplicated genes across genomes. *PloS one* **7**, e50653, doi:10.1371/journal.pone.0050653 (2012).