

ÉPIGÉNÉTIQUE TRANSGÉNÉRATIONNELLE CHEZ LA CAILLE : ANALYSE DE LA METHYLATION DE L'ADN À L'ÉCHELLE DU GÉNOME

Chloé Cerutti¹, Sophie Leroux¹, Paul Terzian², Joanna Lledo³, David Gourichon⁴, Jean-Noël Hubert¹, Yann Labrune¹, Rémy-Félix Serre³, Céline Vandecasteele³, Christophe Klopp², Carole Iampietro³, Christine Gaspin², Denis Milan¹, Cécile Donnadiou³, Vincent Coustham^{5,6}, Frédérique Pitel¹ et Guillaume Devailly¹

¹GenPhySE, Université de Toulouse, INRAE, ENVT, 31326, CASTANET-TOLOSAN

²MIAT, UR875 INRAE, 31326, CASTANET-TOLOSAN

³INRAE, US1426, GeT-PlaGe, Genotoul, France Génomique, Université de Toulouse, 31326 CASTANET-TOLOSAN

⁴PEAT, UE1295 - INRAE Centre Val de Loire, 37380, NOUZILLY

⁵INRAE, Université de Tours, BOA, 37380, Nouzilly, France

⁶INRAE, Université de Pau & Pays de l'Adour, NuMeA, E2S UPPA, Aquapôle, 64310, SAINT-PEE-SUR-NIVELLE, France

frederique.pitel@inrae.fr

RÉSUMÉ

L'influence de l'environnement prénatal sur le développement du phénotype adulte est partiellement médiée par des phénomènes épigénétiques. Pour étudier une transmission "non génétique" de cette influence environnementale entre générations, l'objectif de ce travail était d'analyser une marque épigénétique, la méthylation des cytosines de l'ADN, chez des cailles issues d'une étude antérieure : des œufs fécondés ont été divisés en deux groupes, un groupe avec injection dans l'œuf d'un perturbateur endocrinien, la génistéine, et un groupe témoin sans injection. Deux "épilignées" ont été générées, respectivement issues des contrôles et des traités. Après trois générations sans aucune autre injection, plusieurs phénotypes ont été affectés par le traitement des ancêtres, tels que la reproduction et des caractéristiques comportementales. Des données de méthylation de l'ADN ont été générées par deux technologies de séquençage (Whole Genome Bisulfite Sequencing – WGBS, et Oxford Nanopore Technology - ONT) à partir d'échantillons de sang de la troisième génération. Nous avons détecté des milliers de cytosines différenciellement méthylées (DMC) entre épilignées, quelle que soit la technologie (18772 en ONT, 11818 en WGBS). Malgré la cause génétique probable d'une bonne part des différences observées entre les deux groupes, ces résultats démontrent un effet transgénérationnel de l'injection de génistéine dans l'œuf 3 générations plus tôt. Ces données nous ont également permis d'effectuer une analyse comparative entre deux technologies de séquençage pour l'analyse de la méthylation de l'ADN.

ABSTRACT

Transgenerational epigenetics in quail: genome-wide analysis of DNA methylation

The influence of the prenatal environment on the development of the adult phenotype is partially mediated by epigenetic phenomena. To study a "non-genetic" transmission of this environmental influence between generations, the aim of this work was to analyze an epigenetic mark, DNA methylation, in quails derived from a previous study: fertilized eggs were divided into two groups, one with an egg injection of an endocrine disruptor, genistein, and a control group without injection. Two "epilines" were generated, from the control and treated groups. After three generations without any further injections, several phenotypes were affected by the treatment of the ancestors, such as reproduction and behavioral traits. DNA methylation data were generated by two sequencing technologies (Whole Genome Bisulfite Sequencing - WGBS, and Oxford Nanopore Technology - ONT) from third-generation blood samples. We detected thousands of differentially methylated cytosines (DMCs) between epilines, from both technologies (18772 in ONT, 11818 in WGBS). Despite the probable genetic cause of many of the differences observed between the two groups, these results demonstrate a transgenerational effect of genistein injection into the egg, 3 generations earlier. These data also enabled us to carry out a comparative analysis between two sequencing technologies for DNA methylation analysis.

INTRODUCTION

L'influence de l'environnement prénatal sur le développement du phénotype adulte est partiellement médiée par des phénomènes épigénétiques (Feil et Fraga, 2012). Un nombre croissant d'études ont mis en évidence la transmission des marques épigénétiques entre les générations à la suite d'une exposition environnementale. Cependant, la question de leur transmission au-delà des individus exposés fait l'objet de nombreux débats. Des études récentes ont révélé que le phénomène de transmission non génétique de l'information entre générations était probablement présent chez les espèces aviaires (Vitorino Carvalho et al 2023). Dans l'une d'entre elles (Leroux et al., 2017), des œufs fécondés ont été divisés en deux groupes : un groupe auquel a été injecté un perturbateur endocrinien, la génistéine, et un groupe témoin non injecté. Des accouplements ont été réalisés au sein de chaque groupe pendant 3 générations pour constituer des "épilignées". Afin de minimiser la variabilité génétique entre ces épilignées, des croisements "en miroir" ont été réalisés, en maintenant des contributions d'ancêtres similaires à travers les générations dans chaque épilignée. Après trois générations sans aucune autre injection, plusieurs caractères ont révélé une influence du traitement des ancêtres, notamment des caractères de développement, de reproduction et de comportement. Dans cette étude, pour mieux comprendre la transmission trans-générationnelle de ces effets environnementaux, potentiellement non-génétique, nous avons étudié la méthylation de l'ADN (au niveau des sites CpG) à partir d'échantillons de sang de la troisième génération, en utilisant deux technologies différentes : le WGBS (Whole Genome Bisulfite Sequencing) et le séquençage de l'ADN natif ONT (Oxford Nanopore Technologies). Cette dernière technologie, récente, permet de discerner les cytosines méthylées des cytosines non méthylées, et génère des lectures longues, plus informatives que les lectures courtes. Cette étude nous permettait ainsi de comparer une technologie éprouvée (WGBS) avec une technologie utilisée pour la première fois chez la caille (ONT).

1. MATERIELS ET METHODES

1.1. Dispositif animal

Les animaux utilisés sont issus de la lignée expérimentale S+ (Mills et Faure, 1991). Ils ont été élevés en conditions standard à l'unité expérimentale INRAE PEAT (Nouzilly). Les œufs fécondés ont été divisés en deux groupes : un groupe témoin non injecté et un groupe dont les œufs ont été injectés avec 500 µg de génistéine avant incubation, formant la génération G0. Trois générations ont ensuite été produites sans

autre injection (Figure 1). Des échantillons de sang de 24 caillies ont été prélevés à l'âge de 3 semaines.

1.2. Séquençage pour les analyses de méthylation

L'ADN génomique de haut poids moléculaire a été extrait du sang de 23 échantillons, selon un protocole d'extraction décrit par Roussot et al, 2003, contrôlé sur Femto Pulse System (Agilent) et quantifié au Qubit (ThermoFisher). La préparation des librairies et le séquençage (2 x 150 pb) pour les analyses WGBS ont été réalisés par Novogene (UK) sur un séquenceur Illumina NovaSeq 6000, en "paired-ends" de 150 pb. Les librairies ONT ont été préparées et séquencées sur l'instrument PromethION (flow cells FLO-PRO002 R9.4.1), à 20 fmol, sur la plateforme GeT-PlaGe.

1.3. Analyses des données

Après séquençage, les lectures doivent être analysées : alignement sur le génome de la caille, filtres sur la qualité et la profondeur de lecture, calcul du taux de méthylation à chaque CpG observé (qui servira aux analyses de méthylation différentielle). L'alignement des lectures obtenues a été effectué sur le génome de référence Ensembl de la caille japonaise, qui contient 8 974 565 sites CpG (https://www.ebi.ac.uk/ena/data/view/GCA_001577835.1). Les analyses, réalisées sur la plateforme bioinformatique Genotoul, comprenaient des pipelines public (nf-core/methylseq, option Bismark, pour le WGBS) ou développé par le fournisseur (Megalodon pour l'ONT). Un pipeline R développé au laboratoire (<https://forgemia.inra.fr/chloe.cerutti/bseqmethdiffanalysis>) a été utilisé pour détecter les cytosines différenciellement méthylées entre les épilignées. Le pipeline effectue d'abord la fusion des informations de méthylation et de couverture des deux brins pour chaque site CpG, puis filtre les données : seuls les sites CpG présentant une couverture minimale de 5 et maximale de 1000 lectures dans au moins 80 % des échantillons ont été utilisés pour l'analyse ; les positions des CpG correspondant à des SNP, susceptibles d'être identifiés à tort comme des DMC, ont été éliminés. Ces SNP (n = 9 142 915) ont été détectés avec le logiciel Longshot v0.4.1 (Edge et al., 2019) à partir des données ONT, sur ADN natif. La détection des cytosines différenciellement méthylées (DMC) a ensuite été réalisée à l'aide du package edgeR (Chen et al., 2017), en prenant en compte les effets sexe et traitement. Les SNP détectés à partir des données ONT ont permis de comparer les fréquences alléliques à ces positions entre les épilignées, à l'aide de plink v1.9 (Purcell et al., 2007) et la fonction association.test du package R gaston (Perdry et al., 2022). L'annotation des DMC a été réalisée à l'aide du package R GenomeFeatures v1.3 (<https://forgemia.inra.fr/aurelien.brionne/GenomeFeatures>) en utilisant le fichier d'annotation (v103) d'Ensembl, en définissant le promoteur entre 3 Kpb en amont et 500 bp en aval du site de début de transcription - TSS). Une analyse de

surreprésentation des termes de la Gene Ontology (GO) a été réalisée à l'aide du package R ViSEAGO v1.4.0 (Brionne et al., 2019) et un test d'enrichissement a été réalisé à l'aide du test exact de Fisher (fonction runTest du package topGO v2.42.0).

2. RESULTATS ET DISCUSSION

Dans un premier temps nous avons comparé les données obtenues par chacune des méthodes, décrites dans le tableau 1. On voit que les deux technologies ont donné des résultats comparables. Le nombre de CpG analysables est légèrement supérieur en ONT, avec une couverture moyenne plus grande qu'en WGBS. La corrélation moyenne du taux de méthylation entre les données WGBS et ONT, avec 6 107 897 sites CpG communs aux deux types de données, est très satisfaisante ($0,89 \pm 0,02$) et permet de conclure que la technologie ONT est utilisable chez la caille au même titre que le WGBS pour analyser la méthylation de l'ADN. Notons que le séquençage ONT ne nécessite pas de traitement de l'ADN au bisulfite de sodium et permet une détection plus fiable des SNP, même à faible profondeur comme récemment montré chez le bovin (González-Recio et al, 2023).

Dans un second temps nous avons effectué l'analyse des données pour répondre à notre question scientifique : y a-t-il un effet de l'environnement de développement des ancêtres sur le profil de méthylation des individus ? L'analyse a permis d'identifier 14 019 (WGBS) et 23 967 DMC (ONT) associées au traitement, dont 7 497 communes aux deux technologies. La répartition des DMC est globalement similaire à celle des CpG, avec un nombre plus élevé dans les introns et les régions intergéniques. Les DMC ont été trouvées dans 3 142 gènes différents. Ces gènes sont enrichis en termes GO associés notamment aux processus de développement, à la reproduction et au développement du cerveau. En outre, même si ce n'était pas l'objet principal de l'étude, nous avons détecté plus de 100 000 DMC (109 216 en WGBS, 120 466 en ONT), et environ 2000 DMR (1779 en WGBS, 2223 en ONT) associées au sexe, principalement sur le chromosome Z.

On sait que la méthylation de l'ADN peut être sous contrôle génétique (Bell et al., 2011), donc non gouvernée par les variations d'environnement. L'étape suivante était donc d'estimer la part des DMC observées dues à une transmission non-génétique entre générations. Une analyse génétique préliminaire des données suggère que la plupart de ces positions différenciellement méthylées sont liées à la variabilité génétique entre les deux épilignées, bien qu'un croisement miroir ait été mis en place pour réduire les différences génétiques entre les deux lignées. Ces différences génétiques peuvent s'expliquer par un échantillonnage aléatoirement déséquilibré dans les fondateurs et / ou par une dérive génétique au cours des générations, et / ou par des modifications génétiques générées par le traitement. Des analyses complémentaires sont nécessaires pour approfondir ces hypothèses.

Malgré tout, des CpG dont le taux de méthylation différentiel entre les deux épilignées ne semble pas dû à une variabilité génétique ont été mises en évidence. Nous avons notamment détecté une position sur le chromosome 10 (chr10:9453125) qui ne semble pas être associée à une variation génétique. Cette DMC (Figure 2) est associée au promoteur du gène SPATA5L1 (spermatogenesis associated 5-like 1), impliqué dans la biogenèse des ribosomes.

CONCLUSION

Ces données nous ont permis de démontrer que les deux technologies de séquençage pour l'analyse de la méthylation de l'ADN, WGBS et ONT, donnent des résultats comparables. Malgré la cause génétique probable d'une bonne part des différences observées entre les deux groupes, ces résultats montrent un effet transgénérationnel de l'injection de génistéine dans l'œuf 3 générations plus tôt sur le niveau de méthylation de certaines régions du génome. Ces résultats très prometteurs (la modification de l'environnement des ascendants pourrait modifier le phénotype des générations futures, chez les oiseaux) méritent d'être confirmés et affinés par des études approfondies, à plus grande échelle.

REFERENCES BIBLIOGRAPHIQUES

Bell et al., 2011. Genome Biol, (12), R10
Brionne et al., 2019. BioData mining, 12(1), 1-13.
Chen et al., 2017. F1000Research, 6.
Feil et Fraga, 2012. Nat Rev Genet 13, 97–109.
González-Recio et al. J Anim Sci Biotechnol. 2023 Jul 12;14(1):98.
Leroux et al., 2017. Genet Sel Evol 49, 14.
Mills et Faure, 1991. J. Comp. Psy. 105, 25–38.
Perdry et al., 2022. R package version 1.5.9. <https://CRAN.R-project.org/package=gaston>
Purcell et al., 2007. Am J Hum Genet. 2007 Sep;81(3):559-75
Rousot et al., 2003. Genet Sel Evol.2003;35:559–72.
Vitorino Carvalho et al., 2023. J Anim Sci Biotechnol. 14(1):124

REMERCIEMENTS

Les auteurs remercient l'ensemble de l'équipe de PEAT – Financements : ANR EpiBird - SeqOccIn FEDER-FSE Midi-Pyrénées et Garonne 2014-2020 – Région Occitanie – Département de Génétique Animale INRAE

Tableau 1. Données obtenues pour chaque technologie

	WGBS	ONT
Nombre de lectures	284,8 ± 47,6 M	5,5 ± 1,4 M
CpG observés avant filtres	8 315 634 ± 72 776	8 788 404 ± 10 082
CpG analysés après filtres	6 142 783 ± 9 564	7 955 339 ± 547
Couverture moyenne des CpG analysés	20,5 ± 20,8	39 ± 17,5

Figure 1. Dispositif expérimental. Des échantillons de sang ont été prélevés sur 23 cailles (mâles et femelles) de 2 groupes : un groupe contrôle non-injecté (epi-) et un groupe injecté avec de la génistéine. Les cailles étudiées provenaient de la lignée expérimentale S+. (Cerutti, 2023)

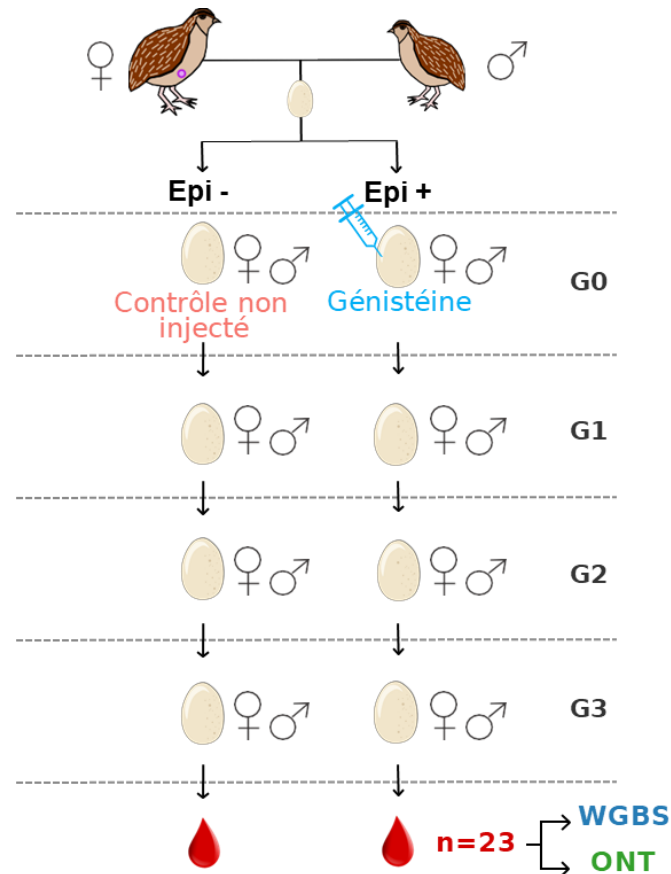


Figure 2. Taux de méthylation observé pour chaque technologie à une DMC localisée sur le chromosome 10, au niveau du promoteur du gène SPATA5L1 (Cerutti, 2023)

