

**CARACTERISATION D'UNE DUPLICATION DE 71 KB AU LOCUS ASIP
IMPLIQUEE DANS LE PHENOTYPE BEIGE/FAWN
DU PLUMAGE DE LA CAILLE JAPONAISE**

**Robic Annie¹, Leroux Sophie¹, Gourichon David², Trussardi Damien¹, Zerjal Tatiana³,
Esquerre Diane⁴, Plisson-Petit Florence¹, Fillon Valérie¹, Vignal Alain¹,
Minvielle Francis³, Pitel Frédérique¹, Morisson Mireille¹**

¹ GenPhySE, Université de Toulouse INRA INPT ENVT 31326 Castanet-Tolosan,

² INRA PEAT 37380 Tours, ³ GABI UMR INRA/AgroParisTech 78352 Jouy-en-Josas,

⁴ INRA GeT-PlaGe Genotoul 31326 Castanet Tolosan,

annie.robic@inra.fr

RÉSUMÉ

La récente mise à disposition du génome de référence de la caille japonaise (mars 2016) a rendu possible la caractérisation moléculaire par séquençage haut débit de mutations causales de phénotypes dont le locus impliqué est connu, comme le phénotype *Beige* (Fawn) impliquant le gène ASIP. Les lectures appairées obtenues par séquençage de deux cailles de phénotype *Beige* ont été alignées sur le génome de référence de la caille. Cet alignement a été examiné attentivement dans la région du gène ASIP (Chromosome 20 de la caille) et une duplication d'une région de près de 71 kb a été mise en évidence. Dans cette région dupliquée, il y a une copie du gène ASIP qui est dans une situation normale alors que la seconde copie est incluse dans un gène de fusion ITCH-ASIP. Après avoir caractérisé les différents transcrits issus de la région, nous avons caractérisé le transcrit du gène de fusion : ce transcrit contient l'exon 1 (non codant) d'ITCH et les trois exons codants d'ASIP. La mutation était déjà connue pour entraîner un excès de transcrits ASIP, évalués sur la partie codante, aussi nous avons réalisé des PCR quantitatives pour évaluer l'importance des différents transcrits d'ASIP et du gène de fusion. Ces dosages montrent clairement que c'est la transcription excessive du gène ITCH-ASIP en un transcrit de fusion qui aboutit à cet excès de transcrits codants pour une protéine ASIP.

ABSTRACT

Characterization of a 71 kb duplication at ASIP locus involved in the Fawn phenotype of the plumage of Japanese quail.

The recent availability of the reference genome of the Japanese quail (March 2016) allows the molecular characterization by high-throughput sequencing of causal mutations of phenotypes for which the locus responsible is known, as for the Fawn phenotype involving the ASIP gene. The reads obtained by paired-end sequencing of two Fawn quails were aligned on the quail reference genome. This alignment was carefully examined in the ASIP gene region (quail chromosome 20) and we found a tandem duplication of a region of about 71 kb. In this new region, there was a copy of the ASIP gene that was in a normal situation, while the second copy was included in an ITCH-ASIP fusion gene. After characterization of different transcripts from the region, we characterized the transcript of the fusion gene: this transcript contained exon 1 (non-coding) of ITCH and the three coding exons of ASIP. The mutation is already known to cause an excess of ASIP transcripts, measured on the coding region. In addition, quantitative PCR was performed to evaluate the importance of different transcripts issued from this region. These assays clearly show that it is the excessive transcription of the ITCH-ASIP gene into a fusion transcript that leads to the excess of transcripts coding for the ASIP protein.

INTRODUCTION

La récente mise à disposition du génome de référence de la caille japonaise (mars 2016) a rendu possible la caractérisation moléculaire par séquençage haut débit de mutations causales de phénotypes dont le locus impliqué est connu. Quand cette mutation relève d'une modification de la structure du génome de ce locus (et pas uniquement une modification ponctuelle, SNP), le séquençage d'un ou deux individus peut suffire.

Nous nous sommes intéressés au phénotype *Beige* (*Fawn*) qui justement était susceptible d'avoir ces deux caractéristiques :

- Nous savions déjà que la mutation causale impliquait le locus du gène ASIP (chromosome 20 de la caille japonaise) (Hiragaki et al. 2008).
- Plusieurs mutations du gène ASIP conduisant à des phanères plus claires que chez l'animal de phénotype sauvage sont connues (*Lethal Yellow* chez la souris ; *Viable Yellow* chez la souris, le mouton *Merinos* ; Alpaga blanc ; et *Lethal Yellow* chez la caille) et elles sont toutes explicables par une modification structurale de leur génome, au locus ASIP.

Etant donné que nous souhaitons rechercher prioritairement une modification structurale, il nous « suffisait » de réaliser le séquençage « appairé » d'un ou deux individus porteurs de cette mutation et de comparer ces données avec le génome de référence.

Le phénotype *Fawn* est causé par une mutation mise en évidence par plusieurs auteurs et est appelée *Fawn*, *Fawn-2* ou *Beige* (Tsudzuki 2008). C'est un phénotype Mendélien à transmission autosomale dominante. A la naissance, le plumage de l'homozygote se caractérise par un plumage jaune clair avec trois bandes sombres sur le dos (mais pas sur la tête). A l'état adulte, les homozygotes sont plus clairs et avec un dimorphisme sexuel plus marqué que pour les individus de phénotype sauvage (Tsudzuki et al. 1996). D'autre part, une étude avait précédemment permis de montrer que le gène ASIP bénéficiait d'une transcription plus active au niveau de la peau chez les individus *Fawn* (Hiragaki et al. 2008).

Après avoir séquençé deux individus de phénotype *Beige* et caractérisé un segment dupliqué chez ceux-ci, nous avons étudié les conséquences de cette duplication, sur la transcription des gènes de la région. La présence excessive de transcrits codants pour une protéine ASIP chez les individus beige a trouvé une explication.

1. MATERIELS ET METHODES

1.1. Le séquençage « appairé » d'un génome

Nous avons choisi de séquençer deux animaux issus de deux populations séparées depuis plus de 25 ans, maintenues à l'UE PEAT, présentant le phénotype

Beige/Fawn à l'état homozygote (fixé). L'ADN de ces animaux a été extrait à partir de sang. Les séquençages ont été réalisés sur la plateforme GeT-PlaGe de la Génomole de Toulouse.

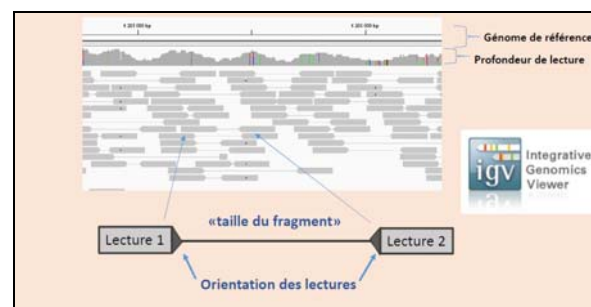
L'ADN de l'individu choisi est extrait puis fragmenté afin d'en retenir une banque de fragments de 350 pb environ. Le séquençage de cette banque a été réalisé avec la technologie illumina Hiseq3000 afin de proposer des lectures de tous ces fragments par chacune des deux extrémités. Il s'agit de lectures de 150 pb, associées par paire (les lectures sont dites appairées) qui sont comparables aux séquençages d'un fragment inséré dans un vecteur, obtenu par clonage. Ici la taille des fragments attendus est donc de 350 pb.

1.2. Alignement du génome séquençé sur le génome de référence

Le génome de référence de la caille est un ensemble de 2012 fichiers de séquences, attribués ou non à des chromosomes.

Le résultat du séquençage de l'individu qui nous intéresse, est un ensemble de lectures associées par paire. Nous utilisons un programme d'alignement, *BWA-MEM* spécialement adapté à l'alignement de ces lectures de 150 pb sur le génome de référence. Ces alignements sont ensuite nettoyés des artéfacts possibles de séquençage redondant (*Rmdup*). Tous ces traitements bioinformatiques ont été réalisés sur la plateforme Galaxy de la Génomole de Toulouse (<http://bioinfo.genotoul.fr/index.php?id=124>).

Figure 1 : Visualisation des alignements avec IGV



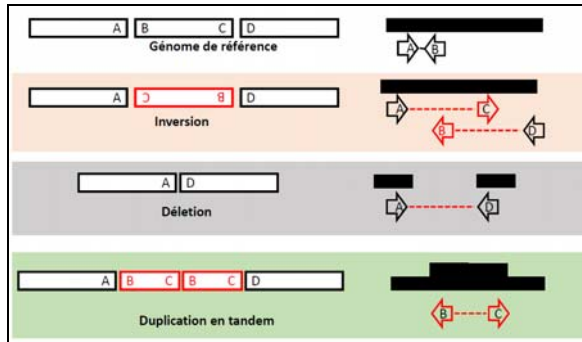
Les paires de lectures sont alignées sur le génome de référence. Chacune d'entre-elles est symbolisée par deux grosses flèches, orientées tête-tête, séparées par un trait mince (partie du fragment non séquençé).

Pour visualiser l'alignement de ces lectures sur le génome de référence, nous avons utilisé le programme *Integrative Genome Viewer*, IGV. Classiquement les paires de lectures s'alignent tête à tête sur le locus considéré du génome de référence (Figure 1) avec une taille de fragment compatible avec la sélection de taille réalisée avant le séquençage. De plus ce logiciel propose une évaluation de la quantité de séquences alignées à chaque locus (profondeur de lecture, sur la Figure 1).

1.3. Détection visuelle des anomalies de structure

Les principaux types de modifications structurales reconnaissables par cette méthode visuelle sont l'inversion d'un segment de séquence, la délétion ou perte d'un segment de séquence et la duplication en tandem d'un segment (Figure 2).

Figure 2 : Anomalies de structure du génome et détection visuelle dans IGV



Les anomalies de l'alignement des lectures dans la région de la modification de la structure s'observent sur l'orientation de certaines extrémités de lecture, sur la taille apparente du fragment, et éventuellement une abondance différente de lectures dans la région de la modification.

1.4. Techniques biologiques

Les ARN d'embryons ont été extraits à partir de peau (avec le duvet) : nous avons utilisé des embryons de 15 jours, issus d'une population beige et des embryons d'une population au plumage sauvage. Pour pouvoir rechercher et doser les différents transcrits, les ADNc ont été synthétisés par une transcription inverse des ARN avec un amorçage au hasard (dn9).

Des amplifications PCR ont été réalisées en choisissant deux amorces aux extrémités de la région à amplifier, à partir des séquences connues. Les PCR quantitative (qPCR) ont été effectuées sur ABI7900HT avec du Sybr green et deux transcrits de référence (RPS13 et GAPDH). Dans tous les cas, les deux amorces ont été choisies sur deux exons différents.

Des séquençages de fragments PCR ont été réalisés (ABI3700).

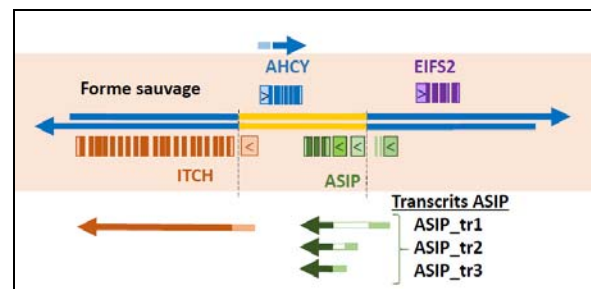
2. RESULTATS ET DISCUSSION

2.1. Présentation du contenu de la région ASIP

L'annotation du génome de la caille a été faite parallèlement à l'assemblage du génome de référence et nous disposons donc d'un génome de référence annoté. Dans la région, on notera que le gène ASIP se trouve sur le même brin que le gène ITCH, alors qu'AHCY et EIFS2 sont sur l'autre brin (Figure 3).

Le gène ASIP est correctement annoté pour sa partie codante (trois exons) mais nous savons par les connaissances acquises chez la poule qu'il possède vraisemblablement cinq exons non codants (dits 5'NC) dont le n°1, n°4 et le n°5 qui pourraient être départ de transcription. Le gène ASIP a donc probablement trois promoteurs et sa transcription pourrait aboutir à trois types de transcrits (Figure 3). La séquence codante débute dans l'exon n°6. Les gènes ITCH et AHCY sont plus simples puisqu'ils n'ont qu'un exon 5'NC et donc un seul départ possible pour leur transcription.

Figure 3 : La région du gène ASIP chez la caille



Chaque gène est représenté par une couleur : les exons codants sont schématisés par des rectangles foncés alors que les exons 5'NC le sont par des rectangles de même couleur mais plus claire. Les exons 5'NC pouvant être le site de démarrage de la transcription sont symbolisés par un rectangle avec une flèche. Nous attendons trois types de transcrits codants pour la même protéine ASIP, alors que ITCH et AHCY n'ont qu'un transcrit.

2.2. Mise en évidence d'une duplication en tandem

Dans cette étude nous avons obtenu 100 millions de lectures de 150 pb par individu, soit 15 Gb séquencées par individu (séquences brutes). Après alignement sur le génome de référence, près de 80 % de ces lectures peuvent être alignées correctement sur le génome de référence de la caille japonaise.

Figure 4 : Caractérisation d'une duplication de près de 71 Kb



Grace à IGV, nous visualisons des paires de lectures orientées queue-queue et éloignées de près de 71 kb. De plus, une présence excessive de lectures dans l'ensemble de cette région permet de soupçonner l'existence d'une duplication en tandem de 71 kb.

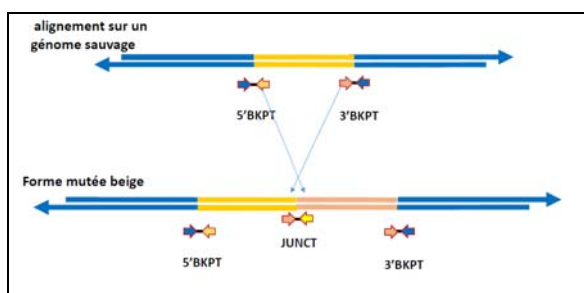
Ces alignements ont été examinés attentivement dans la région du gène ASIP (Chromosome 20 de la caille). En utilisant IGV, nous avons trouvé les trois éléments caractérisant une duplication en tandem (Figure 2)

dans la région d'ASIP (Figure 4). Cette duplication de 71 kb est localisable à quelques paires de bases près, grâce à ce logiciel : il suffit de zoomer sur les deux limites que nous appellerons points de cassures 5' et 3' (5'BKPT et 3'BKPT).

2.3. Caractérisation de la duplication et lien avec le phénotype *Beige/Fawn*

Pour séquencer les deux points de cassure, il suffit de définir deux couples d'amorces de part et d'autres des points de cassure et de séquencer les deux fragments PCR correspondants. La Figure 5 permet de comprendre qu'en cas de duplication (et seulement dans ce cas de figure), il y a un troisième couple qui génèrera un fragment PCR.

Figure 5 : Caractérisation de la duplication par PCR



Les deux points de cassure (5'BKPT et 3'BKPT) peuvent être amplifiés par deux couples d'amorces. Si on prend la première amorce de la PCR-3'BKPT avec la seconde amorce de la PCR-5'BKPT, on amplifie chez les animaux ayant la duplication un fragment de jonction (JUNCT).

Les deux points de cassure ont été séquencés chez un individu beige et chez un autre de phénotype sauvage. Le point de jonction a été amplifié et séquencé sur un individu beige. Les séquences sont conformes à ce que nous attendions et permettent de caractériser une apparente délétion de 70895 pb. Aux trois points stratégiques, une micro-homologie de 3 pb est observée. L'absence d'une séquence répétée importante exclut a priori la possibilité de trouver des caillies avec 3 ou 4 copies de ce segment.

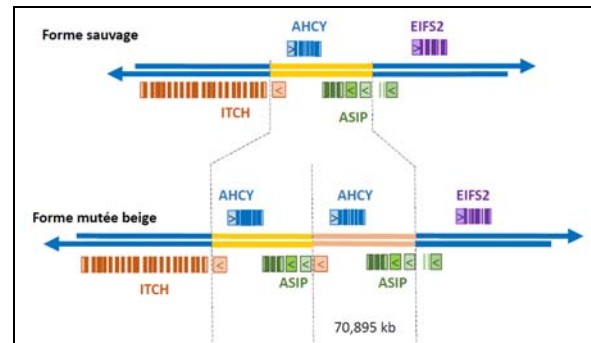
Les fragments 3'BKPT et JUNCT ont été utilisés pour montrer que la duplication existait chez toutes les caillies présentant le phénotype *Beige* et seulement chez celles-ci.

2.4. Conséquences de la duplication sur le locus ASIP

Les points de cassure de cette duplication en tandem sont localisés en aval de l'exon n°1 d'ITCH et en amont de l'exon n°4 d'ASIP (5'NC). La région dupliquée de 70895 pb inclut sur un brin le gène AHCY dans son ensemble et sur l'autre brin, une partie des gènes ASIP et ITCH (Figure 6). Aucun gène ne paraît détruit dans ce remaniement, mais la

nouvelle copie d'ASIP (privée de ses trois premiers exons) se trouve placée juste après le premier exon d'ITCH, point de démarrage possible d'une transcription. De plus, ce gène de fusion ITCH-ASIP possède encore les deux exons ASIP (n°4 et n°5) permettant éventuellement un démarrage de la transcription.

Figure 6 : Région ASIP chez les caillies de phénotypes sauvage et beige



Chez la caille beige, la duplication en tandem d'un segment conduit sur un brin à la présence de l'exon 1 dupliqué d'ITCH suivi d'une nouvelle copie tronquée du gène ASIP, et à une seconde copie du gène AHCY sur l'autre brin.

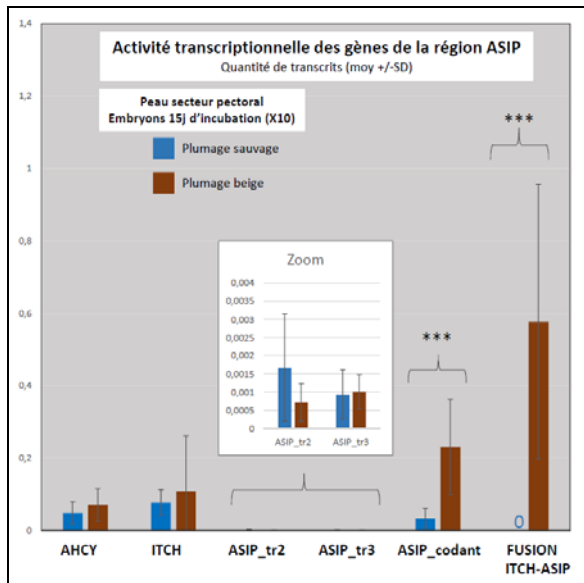
Après avoir séquencé dans des prélèvements de peau d'embryon de 15 jours, les trois transcrits d'ASIP et le transcrit d'ITCH chez la caille, nous avons caractérisé le transcrit du gène de fusion : ce transcrit contient l'exon 1 (NC) d'ITCH et les 3 exons codants d'ASIP. Les exons non codants n°4 et n°5 d'ASIP ne sont pas incorporés dans le transcrit de fusion. Le gène de fusion peut aussi proposer des transcrits ASIP_tr2 et ASIP_tr3, mais nous n'avons aucun moyen de les différencier de ceux issus de la première copie du gène. Après avoir testé plusieurs prélèvements de peau d'embryons, le prélèvement de peau du secteur pectoral, a été retenu pour la suite.

2.5. Conséquences sur la transcription des gènes de la région

Nous avons réalisé des PCR quantitatives pour évaluer l'importance des différents transcrits d'ASIP et du gène de fusion dans la responsabilité moléculaire de cette surexpression.

Comme l'avaient déjà montré Hiragaki et coll. (2008) nous observons une différence statistiquement significative des transcrits codants pour ASIP, entre les caillies de phénotype sauvages et les caillies beiges (Figure 7). Le transcrit de fusion ITCH-ASIP n'est retrouvé que chez les caillies beiges (il ne peut d'ailleurs pas exister chez les autres caillies). Ces dosages montrent clairement que c'est la transcription excessive de la copie fusionnée, ITCH-ASIP, en un transcrit de fusion qui aboutit à cet excès de transcrits codants pour une protéine ASIP.

Figure 7 : Evaluation de l'expression des gènes



Les transcrits ASIP_tr2 et ASIP_tr3 sont trouvés en quantités très faibles (zoom) alors que le transcrit de fusion n'est trouvé que chez les animaux beiges. Cette étude a été faite dans la peau d'embryons de 15 jours.

Néanmoins nous observons plus de transcrits de fusion que de transcrits contenant les trois exons codants d'ASIP (ASIP_codant). Ceci montre qu'il existe probablement une partie de ces transcrits issus du gène de fusion qui se terminent prématurément et qui ne contiennent probablement pas l'ensemble des trois exons codants pour ASIP. La technologie de qPCR ne permet pas de doser le transcrit de fusion autrement qu'en dosant le transcrit contenant l'exon-1 d'ITCH et l'exon-6 d'ASIP (1^{er} exon codant d'ASIP). Les différences de quantités de transcrits ASIP_tr2, ASIP_tr3, d'AHCY, et d'ITCH ne sont pas significativement différentes entre animaux de phénotypes sauvages et beiges. Il est à noter que le transcrit de fusion est trouvé en plus grande quantité que le transcrit ITCH, alors que ces deux transcrits sont transcrits à partir d'un promoteur ITCH.

2.6. Ressemblances avec les moutons *Merinos*

La région du gène ASIP est particulièrement bien conservée chez les oiseaux mais aussi chez les mammifères (ordre des gènes et orientation). Chez les

moutons *Merinos*, une duplication similaire a déjà été caractérisée (Norris and Whan 2008) : les points de cassure de cette duplication en tandem sont localisés en aval de l'exon-1 d'ITCH et en amont des exons codants d'ASIP et un transcrit de fusion ITCH-ASIP est aussi mis en évidence. Toutefois, le gène ASIP du segment original paraît être inactivé chez les moutons *Merinos* (blancs). Dans le cas de la caille, nous n'avons recueilli aucun élément allant dans ce sens. De plus chez le mouton, on trouve une séquence répétée, de type SINE, aux pieds de chaque répétition et il est possible de trouver des animaux présentant plusieurs répétitions de ce segment génomique.

CONCLUSION

Une duplication de près de 71 kb dans la région d'ASIP a été mise en évidence chez les cailles beiges et seulement chez celles-ci. La région issue de cette duplication contient deux copies du gène AHCY et surtout un gène issu de la fusion entre le 1^{er} exon d'ITCH (5'NC) et une partie du gène ASIP qui a conservé ses trois exons codants. La transcription de ce gène fusion abouti, en particulier, à un transcrit ITCH-ASIP codant pour une protéine ASIP. Nous avons montré que, chez les animaux beiges, ce transcrit de fusion était le principal transcrit codant pour une protéine ASIP. La surproduction de transcrits codants pour une protéine ASIP normale dans la peau à la base des plumes est très certainement responsable du phénotype de plumage beige.

Le locus ASIP est assez particulier puisque l'on trouve une nouvelle fois le même type de mutations chez un oiseau (caille beige) que chez un mammifère (mouton *Merinos*). En effet on connaissait déjà la délétion mise en évidence chez la souris *Lethal Yellow* qui est très similaire à celle repérée chez la caille *Lethal Yellow* (Nadeau et al. 2008).

D'un point de vue méthodologique, l'insertion de séquences issues d'un autre locus (comme chez l'Alpaga blanc, (Chandramohan et al. 2013)) aurait été très difficile à mettre en évidence par cette méthode, c'est pourquoi des chercheurs travaillent à la mise au point de techniques de détection par voie bioinformatique.

REFERENCES BIBLIOGRAPHIQUES

- Chandramohan, B. et al. 2013. *Gene* (521) 303-310.
- Hiragaki, T. et al.. 2008. *Genetics* (178) 771-775.
- Nadeau, N. J. et al. 2008. *Genetics* (178)777-786.
- Norris, B. J. and V. A. Whan. 2008. *Genome Research* (18) 1282-1293.
- Tsudzuki, M. 2008. *The Journal of Poultry Science* (45) 159-179.
- Tsudzuki, M. et al. 1996. *The Journal of Heredity* (87) 248-2

