



Biological Resource Centers
for Domestic Animals

3rd International Seminar of CRB-Anim Infrastructure

Domestic Animals, Biobanks and Biodiversity

November 26th, 2019

ANALYSE GÉNOME ENTIER DES LIGNÉES ISOGENIQUES DE TRUITE ARC-EN-CIEL

D. Lallias, M. Boussaha, M. Bernard, A. Peigné, C. Ciobotaru, N. Dechamp, M. Dupont-Nivet, E. Quillet, INRA, UMR GABI, Jouy-en-Josas

Une collection unique de 19 lignées isogéniques homozygotes de truite arc-en-ciel a été établie et maintenue dans les installations expérimentales piscicoles de l'INRA. Ces lignées ont été caractérisées pour un grand nombre de caractères et se sont avérées être une ressource pertinente pour étudier les bases moléculaires de caractères complexes à différentes échelles biologiques (cellule, tissu, organisme). Avoir accès à la variabilité génomique entre les lignées est essentielle pour les études fonctionnelles et de recherche de QTL afin d'identifier le(s) polymorphisme(s) responsable(s) des phénotypes d'intérêt et de leur variation. L'objectif de cette étude était de mener une caractérisation génomique approfondie des lignées isogéniques, en étudiant à la fois les petites variations génomiques (SNPs et petits InDels) et les variants structuraux (VS). Les VS sont définis comme des altérations génomiques qui affectent de larges fragments d'ADN ≥ 50 nucléotides, causant ainsi des modifications soit dans la quantité d'ADN (insertions, délétions et duplications) soit dans la structure d'ADN (inversions). Même si les VS ont reçu un intérêt croissant chez de nombreuses espèces et ont été montrés associés à plusieurs maladies et phénotypes, ils ne sont que mal décrits chez les poissons.

Toutes les lignées isogéniques (un ou deux individus par lignée) ont été séquencées à une profondeur variant de 10X à 32X, sur une plateforme Illumina HiSeq X-Ten dans une configuration pairée 2x150 pb. L'analyse des petites variations génomiques a été effectuée en suivant le guide des bonnes pratiques de GATK. L'identification et la caractérisation des VS a été faite avec 3 outils différents correspondant à deux approches distinctes mais complémentaires : i) Pindel et Delly (approche split-read); ii) BreakDancer (approche pairée).

La recherche de petites variations génomiques avec GATK Haplotype Caller a résulté en l'identification de 14 439 713 SNPs et 3 183 663 InDels. Un total de 17 037 VS putatifs ont été identifiés, correspondant à 13 271 délétions, 3 386 duplications en tandem et 380 inversions. Une analyse de distribution des tailles a révélé que la plupart des délétions (98.3%) étaient inférieures à 10Kb (27.8% entre 51bp and 1Kb), alors que la majorité des duplications en tandem (95.1%) étaient supérieures à 1Kb. L'annotation des VS a montré que 8 326 régions contenaient soit des gènes entiers soit des portions de gènes, parmi elles 5 967 délétions, 2 091 duplications en tandem et 268 inversions.

La caractérisation fine des lignées isogéniques homozygotes de truite arc-en-ciel permettra non seulement la production des informations nécessaires à la pleine exploitation des études tirant profit des phénotypes contrastés et/ou des caractéristiques génomiques originales des lignées, mais également de fournir une description d'intérêt général de l'organisation structurale et fonctionnelle du génome de la truite.





Biological Resource Centers
for Domestic Animals

3rd International Seminar of CRB-Anim Infrastructure

Domestic Animals, Biobanks and Biodiversity

November 26th, 2019

WHOLE GENOME ANALYSIS OF RAINBOW TROUT ISOGENIC LINES

D. Lallias, M. Bousaha, M. Bernard, A. Peigné, C. Ciobotaru, N. Dechamp, M. Dupont-Nivet, E. Quillet, INRA, UMR GABI, Jouy-en-Josas

A unique collection of 19 isogenic homozygous rainbow trout lines has been established and maintained in INRA fish experimental facilities. These lines have been characterized for a variety of traits and proved to be a relevant resource to investigate the molecular bases of complex traits at different biological scales (cell, tissue, organism). Having access to the genomic variability among lines is essential in functional or QTL studies in order to identify polymorphism(s) responsible for the phenotypes of interest and their variation. The objective of this study was to carry out in-depth genomic characterization of the trout isogenic lines, by investigating both small genomic variations (SNPs and InDels) and structural variants (SVs). SVs are defined as genomic alterations that affect large DNA segments ≥ 50 nucleotides, thereby causing modifications in either DNA quantity (insertions, deletions and duplications) or DNA structure (inversions). Although SVs have received increasing interest in many species and were shown to be associated with several diseases and phenotypes, they are poorly documented in fish.

All isogenic lines (one or two individuals per line) have been resequenced at a depth of coverage ranging from 10X to 32X, on an Illumina HiSeq X-Ten platform, in paired-end 2x150 bp configuration. Analysis of small genomic variants was performed according to the GATK Best Practices. The identification and characterization of SVs was done by using 3 different tools corresponding to two distinct but complementary approaches: i) Pindel and Delly (split-read approach); ii) BreakDancer (paired-end approach).

Search for small genomic variations with GATK Haplotype Caller software resulted in the identification of 14 439 713 SNPs and 3 183 663 InDels. A total of 17 037 putative SVs were identified, corresponding to 13 271 deletions, 3 386 tandem duplications and 380 inversions. Analysis of the length distribution revealed that most deletions (98.3%) were less than 10Kb (27.8% between 51bp and 1Kb), whereas the vast majority of tandem duplications (95.1%) were larger than 1Kb. Annotation of the SVs revealed a total of 8 326 regions which contain either entire genes or parts of genes, among them 5 967 deletions, 2 091 tandem duplications and 268 inversions.

The fine characterization of rainbow trout homozygous isogenic lines will allow not only production of the information necessary for a full exploitation of the ongoing and future studies taking advantage of the contrasted phenotypes and/or of the original genomic features of the lines, but also to provide a description of broad interest of the structural and functional organization of its genome.

