In-depth genomic characterization of a unique collection of rainbow trout isogenic lines

Delphine Lallias, Mekki Boussaha, Maria Bernard, Antoine Peigné, Céline Ciobotaru, Nicolas Dechamp, Mathilde Dupont-Nivet, Edwige Quillet

IN-DEPTH GENOMIC CHARACTERIZATION OF A UNIQUE COLLECTION OF RAINBOW TROUT ISOGENIC LINES

D. Lallias*, M. Boussaha, M. Bernard, A. Peigné, C. Ciobotaru, N. Dechamp, M. Dupont-Nivet and E. Quillet

GABI, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas (France)
E-mail: delphine.lallias@inra.fr

Introduction

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). Genomic characterization of the lines is pivotal to realize the whole benefit of this resource in integrative approaches aimed at dissecting complex traits and understanding the "genome to phenome" mechanisms. 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 $\geq$ 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.

Material and methods

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).

Results

Search for small genomic variations with GATK Haplotype Caller software resulted in the identification of 15 064 416 SNPs and 3 173 673 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. Analysis of chromosomal distribution was also performed (Figure 1). 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.

Discussion

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 add to the overall knowledge on rainbow trout genomic structure and polymorphisms and provide a description of broad interest of the structural and functional organization of its genome.

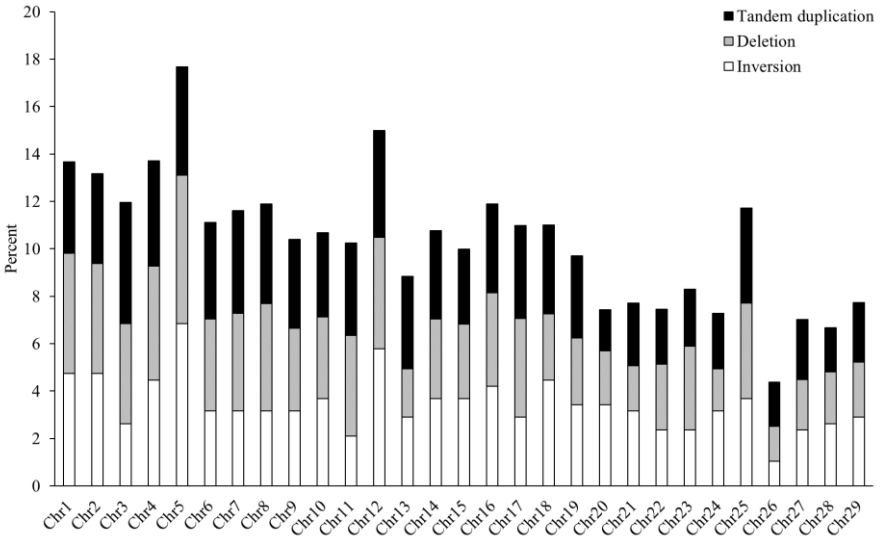

Fig. 1 Chromosomal distribution of SVs within the trout chromosomes.