P5-29. CHARACTERIZATION OF REPETITIVE DNA IN THE SMALL GENOME OF FLATFISH *Cynoglossus semilaevis*

Francisca Robles, Selene García, Alexander García-Zea, Rafael Navajas-Pérez, Roberto de la Herrán, Carmelo Ruiz Rejón

Department of Genetics, University of Granada, Spain,
Corresponding author: Carmelo Ruiz Rejón (carmelo@ugr.es)

Despite their size, a main feature of eukaryotic genomes is the presence of repetitive sequences. Flatfish (order Pleuronectifoms) have smaller genomes than other teleost (381 to 1076 Mb), with C-values ranging from 0.39 to 1.10 pg (1).We used a NGS-based analysis to detect and characterize repetitive DNA in the first sequenced flatfish genome, *Cynoglossus semilaevis* (2). Genomic sequences were retrieved from the NCBI database (#PRJNA73987). The dataset was explored to estimate the nature and quantity of SSRs, mobile elements, and satellite DNAs. Assembled chromosomes were separately analyzed with WindowMasker to define low complexity sequences and short repeats. SSR analysis showed an average density of 20,559.78 b per Mb. The average number of loci per Mb was 1,134.09. All chromosomesshowed a similar number of SSR loci except W chromosome, with roughly an eight-foldincrease (8,724).

To detect interspersed repeats and tandemly-arrayed sequences, 500K paired reads were randomly selected and clustered with Galaxy Repeat Explorer, masked and annotated with RepeatMasker. Repeated DNA correspond to almost 22% of the genome.Also, up to 20 different mobile elements (10.50% of the genome) were detected: nine Ty3 elements (5.11%), seven LINEs (1.97%) and four MITEs (3.37%).

Clusters annotated as mobile elements and main clusters of tandemly-arrayed sequences (selected according to repeat unit size and abundance) were mapped *in silico*. The most represented clusters of satellite DNA (Cyse 1, Cyse 2, Cyse 8, Cyse 15,Cyse 29B, Cyse 33, Cyse 73, and Cyse 101B) accounted for a 6.84% of the genome. Cyse1 (the most abundant, ~3%) mapped at chromosomes 8 and W. Cyse29B and Cyse101B are restricted to chromosome 1 and Cyse15 is found in all chromosomes.

1 Robles et al. (2017) *Journal of Heredity* 108, 217–222

2 Chen et al. (2014) *Nat Genet* 46, 253–260

This work was supported by Ministerio de Ciencias, Innovación y Universidades of Spain (RTI2018-096847-B-C22)

CHARACTERIZATION OF REPETITIVE DNA IN THE SMALL GENOME OF FLATFISH *Cynoglossus semilaevis*

Francisca Robles, Selene García, Alexander García-Zea, Rafael Navajas-Pérez, Roberto de la Herrán, Carmelo Ruiz Rejón

Department of Genetics, University of Granada, Spain,

Corresponding author: Carmelo Ruiz Rejón (carmelo@ugr.es)

Facultad de Ciencias

UNIVERSIDAD DE GRANADA

BACKGROUND

Flatfish (order Pleuronectifoms) are represented by more than 700 species with worldwide distribution and some species as turbot, Japanese flounder, and tongue sole are used in aquaculture for industrial large-scale production. They show interesting characteristics for evolutionary studies such as its anatomy, with external asymmetry, and smaller genomes than other teleost (381 to 1076 Mb), with C-values ranging from 0.39 to 1.10 pg (1). A main feature of eukaryotic genomes is the presence of repetitive sequences. In this sense, in this study, we used a NGS-based analysis to detect and characterized repetitive DNA in the small genome of the flatfish *Cynoglossus semilaevis*, and its distribution, abundance, evolution, and origin are discussed.

METHODS

Genomic sequences were retrieved from the NCBI database (#PRJNA73987) (2). The dataset was explored to estimate the nature and quantity of SSRs, mobile elements, and satellite DNAs. Thus, assembled chromosomes were separately analyzed with WindowMasker to define low complexity sequences and short repeats. Additionally, also, SSRs also were found using Phobos (as implemented in Geneious v6.1.8). According to criteria used in fugu (3) and tetraodon (4), only dinucleotides to hexanucleotides with three perfect repeats or SSRs with a minimum of 12 nucleotides, with no more than two successive N's and a minimum score of six points were selected. To detect interspersed repeats and tandemly-arrayed sequences, 500K paired reads were randomly selected and clustered with Galaxy Repeat Explorer, masked and annotated with RepeatMasker. Finally, main clusters of tandemly-arrayed sequences (selected according to repeat unit size and abundance) were mapped *in silico*.

This work was supported by Ministerio de Ciencia e Innovación of Spain. Project RTI2018-096847-B-C22: Desarrollo del mapa de ligamiento diploide del lenguado y caracterización de secuencias repetidas y genes con expresión diferencial entre sexos.

RESULTS

SSR analysis showed an average density of 20,559.78 b per Mb. The average number of loci per Mb was 1,134.09. All chromosomes showed a similar microsatellite density, except W chromosome, with roughly the half of SSRs (Figures 1). In general, dinucleotides were the most abundant SSRs in the genome (35.62%), followed by tetranucleotides (20.59%), trinucleotides (18.83%), mononucleotides (16.20%), pentanucleotides (5.57%), and hexanucleotides (3.18%). These percentages of bases were similar to the percentages of loci: 33.58%, 20.65%, 19.21%, 19.22%, 4.96% and 2.37%, respectively. Taken the chromosomes separately, highest percentage of SSRs corresponds to dinucleotides, except for chromosome 17 and W, where tetranucleotides were predominant. As for the percentage of loci per chromosome in all cases the highest number corresponded to dinucleotides, except for the W chromosome, where tetranucleotides were more abundant. On the Z chromosome both percentages (dinucleotides and tetranucleotides) were similar.

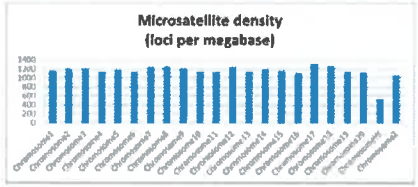

Figure 1A and B. Distribution of microsatellites by chromosomes

REFERENCES:

1. Robles F. et al. (2017) Journal of Heredity 108, 217–222
2. Chen S. et al. (2014) Nat. Genet. 46, 253–260
3. Edwards et al. 1998 Journal of Mol. Biol. 278(4):843-54
4.. Crollius H.R. et al (2000). Nat. Genet. 25, 235.

Clustering analysis showed that 193,172 reads (22.5% of genome) grouped in 23,467 clusters. Eight clusters accounting for a 6.84% of the genome showed a star-like or circular graph topology (Figures 2) and were confirmed as satellite-DNA families by dot-plot analysis (data no shown). We named them as Cyse1, Cyse2, Cyse8, Cyse15, Cyse29B, Cyse33, Cyse73, and Cyse101B, respectively. All of them revealed no similarities to other known sequences in GenBank. Additionally, up to 20 different mobile elements (10.50% of the genome) were detected and annotated as: nine Ty3 elements (5.11%), seven LINEs (1.97%) and four MITEs (3.37%).

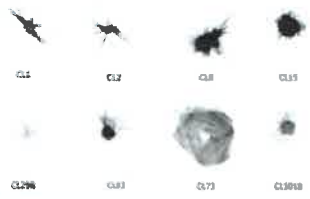

Figure 2. Topology of satellite-DNA clusters

After in silico mapping, using consensus monomeric repeats against the *C. semilaevis* chromosome sequences, we detected two groups of satellite DNAs attending to their location; Group I, integrated by Cyse1, Cyse29B and Cyse101B, includes satellite DNAs with small number of loci, while Group II, integrated by Cyse15, Cyse2, Cyse33 and Cyse73, includes satellite DNA with multiple loci. Cyse8 could not be mapped. Cyse1 (the most abundant, ~3%) mapped in only one locus at chromosomes 8 and W. Cyse29B (in one locus) and Cyse101B (in two loci) were located only in chromosome 1. Cyse15 was found in all chromosomes. Cyse2 and Cyse33 were present in all chromosomes except for chromosome 20 and chromosome 7, respectively. Cyse73 was present in half of the chromosomes (Table 1). Cyse15, Cyse2, Cyse33, and Cyse73 were found arrayed in different number of loci. The mean identities for each cluster on all chromosomes was 92.3% for Cyse1, 87.01% for Cyse2, 89.01% for Cyse15, 86.8% for Cyse29B, 74.84% for Cyse33 86.75 for Cyse73, and 96% for Cyse101B.

DISCUSSION

SSR analysis showed a percentage of 2,06% of microsatellite in *C. semilaevis* genome. Despite of its small size, this data is very similar to obtained in tetraodon with 3,21% (4) and fugu 2,12% (3). None of the monomeric sequences analyzed in the 8 families presented homology with elements of genomic databases, being, therefore, the first time that these satellite DNA families have been characterized. The location of the different families can help in the evolutionary studies of this group of fish. Thus, the presence of the same family of satellite DNA in chromosome 8 and W chromosome could indicate a possible origin of sex chromosomes from this pair of autosomes.

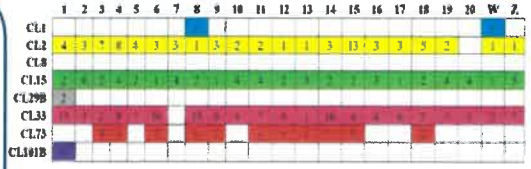

Table 1. Summary of chromosomal location for eight satellite-DNA families analyzed. The numbers indicate the loci found in each chromosome (all of the with at less two units in tandem)