Gene 644 (2018) 4–12

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Research paper

Cytogenomic analysis of several repetitive DNA elements in turbot (*Scophthalmus maximus*)

Xoana Taboada[a], Magalí Rey[a], Carmen Bouza[b], Ana Viñas[a,*]

[a] Departamento de Zoología, Genética y Antropología Física, Facultad de Biología, CIBUS, Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Spain
[b] Departamento de Zoología, Genética y Antropología Física, Facultad de Veterinaria, Universidade de Santiago de Compostela, 27002 Lugo, Spain

ARTICLE INFO

Keywords:
Turbot
Histone H1
Rex
Telomeric sequence
5S rRNA
Gene family
Cytogenomics

ABSTRACT

Repetitive DNA plays a fundamental role in the organization, size and evolution of eukaryotic genomes. The sequencing of the turbot revealed a small and compact genome, as in all flatfish studied to date. The assembly of repetitive regions is still incomplete because it is difficult to correctly identify their position, number and array. The combination of classical cytogenetic techniques along with high quality sequencing is essential to increase the knowledge of the structure and composition of these sequences and, thus, of the structure and function of the whole genome. In this work, the *in silico* analysis of H1 histone, 5S rDNA, telomeric and Rex repetitive sequences, was compared to their chromosomal mapping by fluorescent *in situ* hybridization (FISH), providing a more comprehensive picture of these elements in the turbot genome. FISH assays confirmed the location of H1 in LG8; 5S rDNA in LG4 and LG6; telomeric sequences at the end of all chromosomes whereas Rex elements were dispersed along most chromosomes. The discrepancies found between both approaches could be related to the sequencing methodology applied in this species and also to the resolution limitations of the FISH technique. Turbot cytogenomic analyses have proven to add new chromosomal landmarks in the karyotype of this species, representing a powerful tool to investigate targeted genomic sequences or regions in the genetic and physical maps of this species.

1. Introduction

Turbot (*Scophthalmus maximus*, family Scophthalmidae, Order Pleuronectiformes) is the most important commercialized flatfish worldwide (PR China with 65,000 t is the first producer in the world followed by Europe with 11,849 t, FAO, 2014). In this species, important genetic and genomic resources have been developed *i.e.* high-density genetic maps (Hermida et al., 2013; Wang et al., 2015), and transcriptomic data related to sex, growth and disease resistance (Robledo et al., 2014, 2016; Hu et al., 2016; Ma et al., 2016; Ribas et al., 2016; Ronza et al., 2016). Furthermore the genome of this species has been sequenced (Figueras et al., 2016), being the total length of the assembly 544 Mb (approximately 96% of the estimated size). In this species, classical cytogenetic studies have been carried out using both mitotic and meiotic chromosomes (Bouza et al., 1994; Pardo et al., 2001; Cuñado et al., 2001). Worth mentioning is the study that integrated the genetic and cytogenetic maps of turbot using BAC-FISH (Bacterial Artificial Chromosome-Fluorescent *In Situ* Hybridization), which allowed the assignment of 24 linkage groups (LGs) into 22 chromosome pairs according to the haploid karyotype of the species (LGs 8 + 18 and 21 + 24; Taboada et al., 2014b). This BAC panel tool also represented an important support for identifying homologous chromosomes and guiding genome assembly.

Despite all the genomic resources available and previous cytogenetic work, information concerning turbot repetitive DNA is still scarce. Repetitive DNA sequences (also called repeatome) constitute a highly variable and important portion of eukaryotic genomes. Initially considered useless genomic elements, “junk”, parasitic or selfish DNA (Ohno, 1972; Doolittle and Sapienza, 1980; Nowak, 1994), the repeatome has important structural and functional roles in the eukaryotic genomes. Repetitive sequences are very different in structure, origin and function. According to an organizational criterion, they can be classified in tandem (head to tail orientation) and dispersed repetitive DNA. Among tandem-repetitive DNA there are rDNA and protein-coding gene families, as histones, or telomeric repeats. On the other hand, DNA transposable elements (TE) represent dispersed repetitive elements (López-Flores and Garrido-Ramos, 2012; Biscotti et al., 2015). In the turbot genome 8.5% of repetitive sequences were reported,

Abbreviations: TE, transposable element; BAC, bacterial artificial chromosome; FISH, fluorescent *in situ* hybridization; LG, linkage group; NTS, non-transcribed spacer; LTR, long terminal repeats; PE, paired-end; MP, mate pair; BLAST, Basic Local Alignment Search Tool

* Corresponding author.
E-mail address: anamaria.vinas@usc.es (A. Viñas).

https://doi.org/10.1016/j.gene.2017.12.013
Received 16 May 2017; Received in revised form 23 November 2017; Accepted 11 December 2017
Available online 12 December 2017

mostly analyzing the types and fraction of TE (5%), which was similar to that of other sequenced flatfish (Figueras et al., 2016). Previous cytogenetic analyses in turbot were restricted to heterocromatin banding and major rDNA class using 28S rDNA probes (Pardo et al., 2001; Taboada et al., 2014b). Although the repetitive sequences analysis in other fish are also scarce, a wider range than in turbot of these sequences have been explored, including coding and no-coding elements such as histone H1, 5S rDNA, telomeric, Rex1 and Rex3 (Cioffi and Bertollo, 2012).

H1 linker histones are architectural proteins which facilitate the formation of higher order chromatin structures interacting with the linker DNA which joins adjacent nucleosomes (Pan and Fan, 2016). H1 gene family represents the most heterogeneous group of histone proteins (Cole, 1987; Brown, 2001) and is less evolutionarily conserved than the core histone genes (Harshman et al., 2013). In the last years, its function has moved from a static chromatin component stabilizing higher-order structures to a protein involved in the regulation of different DNA processes (Hergeth and Schneider, 2015).

The multigene family of 5S rDNA is the smallest ribosomal RNA component, known as the minor rDNA, which comprises tandem repetitions of one transcriptional unit of about 120 base pairs separated by a non-transcribed spacer (NTS) (Long and Dawid, 1980). The coding sequence is highly conserved, even among unrelated species, whereas the NTS is highly variable both in sequence and length, even in closely related species.

Telomere structures are composed by DNA-protein complexes involving highly conserved tandem-repeated DNA sequences. In vertebrates the repeated motif is the 6 bp minisatellite TTAGGG (Meyne et al., 1989), and the total size of the telomeres ranges from 2 to 25 kb in fish (Okalewicz, 2013). These motifs are usually located at the end of the chromosomes, but can also be found in interstitial and pericentromeric regions indicating chromosome fusion events, thus being very useful in karyotype evolution analysis (Okalewicz, 2013).

Mobile DNA or transposable elements (TE) are sequences with the ability to move from one location in the genome to another. These elements can make up an important fraction of eukaryotic genomes (from 5% in the smallest fish genomes, such as turbot, Figueras et al., 2016; to more than half in the human genome, Lander et al., 2001). TE sequences have a huge potential to generate gene and chromosome structural mutations, and also important functions in the regulation and reparation of several genes (Shapiro, 2005). Although most fish genomes are compact, TEs are very diverse, containing all types of mobile elements described (Volff et al., 2003). These elements are classified in two categories depending on their method for “jumping” from one genomic position to another; class I or retroelements move through an RNA intermediate, while class II or DNA transposons move by cutting and pasting into other locations. Rex1 and Rex3 are non-LTR (long terminal repeats) retrotransposons (Volff et al., 2001) which are widely distributed in fish genomes and are associated with heterochromatin organization and chromosome rearrangements (Volff et al., 2001; Ozouf-Costaz et al., 2004; Valente et al., 2010; Ferreira et al., 2011; Voltolin et al., 2013).

The integration of cytogenetic and genomic maps increases our knowledge about the architecture and function of the genomes. This cytogenomic approach is particularly important to characterize repetitive DNA, which has been associated with technical challenges for sequence alignment and assembly even in the extensively studied human genome (Miga, 2015). Frequently, bioinformatic tools do not allow the correct assembly of large segments of repeated DNA, whereas the *in situ* hybridization technique may allow its visualization over chromosomes. On the contrary, sequences with single or low copy number cannot be detected by cytogenetic approaches, but may be localized with *in silico* genomic studies. Data related to the structure and organization of the repetitive sequences in many fish genomes are still scarce. Most molecular studies analyzed the distribution of repetitive sequence classes over chromosomes using cytogenetic approaches, but the availability of genomic data in several fish species offers the opportunity to perform a more integrative analysis. To our knowledge only two cytogenomic studies have been published, both in cichlids (Mazzuchelli et al., 2012; Valente et al., 2016). In this work, for the first time in turbot, we undertake a cytogenomic analysis of different classes of uncharacterized repetitive DNA, including coding (histone H1 and 5S rDNA) and non-coding sequences (telomeric, Rex1 and Rex3), previously studied in related fish species (Cioffi and Bertollo, 2012; Ferreira et al., 2011; Okalewicz, 2013; Costa et al., 2014; Jiang et al., 2014), with the aim of improving the knowledge about the structure and composition of the turbot genome using an integrative approach including physical, genetic and chromosomal mapping. This strategy allowed the evaluation of the robustness of the turbot genome assembly by the comparison of FISH results with the localization of the sequences analyzed on specific linkage groups. In addition; this study identified a set of genomic scaffolds bearing repetitive sequence markers useful for tracing the identity of homologous chromosomes in the turbot karyotype, and further supporting genome assembly refinements.

2. Materials and methods

2.1. Genome analysis and amplification of repetitive sequences

The turbot genome assembly spans 544 Mb (~96% of estimated genome content) with a scaffold N50 of 4.3 Mb. The 95% of the genome assembly is found in 287 scaffolds ≥ 166 kb (longest scaffold: 19 Mb) and 80% (156 scaffolds) was anchored to the turbot genetic map. Furthermore, some non-anchored scaffolds were predictively mapped to turbot LGs by comparative mapping with other model fish species.

Genomic DNA was extracted from muscle tissue of five turbot fish using standard phenol–chloroform procedures (Sambrook et al., 1989). Primers to amplify the repetitive sequences of 5S rDNA, H1, Rex1 and Rex3 (Table 1) were designed based on the turbot genome using Primer3 (http://primer3.ut.ee/). PCRs were carried out in a MyCycler Thermal cycler (Bio-Rad) on a volume of 50 μL including 75 ng of genomic DNA, 20 pmol of each primer, 0.2 mM of each dNTP, 1 × PCR buffer and 2.5 U of GreenTaq DNA polymerase (GenScript). The

Table 1
Repetitive sequences amplified on the turbot genome. Primers, genome regions where primers were designed, GenBank accession number and conditions for PCR amplification.

	Primers	Genome region Scaffold: start-end (bp)	Amplicon size	Accession number	Tm
Histone 1	F: GCCAAGAAGAAGACGACACC R: GCCTTCTTAGGGCTCTTGGT	Sm5_s00067: 1,195,717–1,196,168	452 bp	MG200027	58 °C
5S rDNA	F: GCCATACCAGCCTGAATACG R: AAAATGCTTACAGCACCTGGTA	Sm5_s06938: 214–329	116 bp	MG200028	60 °C
REX1	F: TTCTCCAGTGCCTTCAACACC R: CCCTCAGCAGAAAGAGTCTGCTC	Sm5_s00053: 726,051–726,622	574 bp	MG200029	58 °C
REX3	F: AATCCGACTTACTGCCGAGA R: AAAGGGGGACCAGAGAGTGT	Sm5_s00129: 928,982–929,367	386 bp	MG200030	58 °C