Molecular Biology, Vol. 35, No. 3, 2001, pp. 331–335. Translated from Molekulyarnaya Biologiya, Vol. 35, No. 3, 2001, pp. 391–396.

GENOMICS. PROTEOMICS. BIOINFORMATICS

UDC 577.113:598.8

Structural and Functional Characterization of a Centromeric Highly Repetitive DNA Sequence of *Fringilla coelebs* L. (Aves: Passeriformes)

A. G. Malykh[1], V. G. Zhurov[2], A. F. Saifitdinova[2], S. E. Deryusheva[2], and E. R. Gaginskaya[2]

[1] *Fidelity Systems, Inc., Gaithersburg, MD 20879, United States*

[2] *Biological Institute, St. Petersburg State University, Staryi Peterhof, 198504 Russia;*
E-mail: chromas@paloma.spbu.ru

Received November 10, 2000

Abstract—A highly repetitive centromeric *Fringilla coelebs Pst*I (FCP) element was cloned and sequenced. The FCP tandem repeats with a unit length 505 or 506 accounted for about 0.9% of the entire genome and had 57% GC. Direct genomic sequencing with FCP-specific primers and ThermoFidelase 2A revealed the consensus sequence and the five most common single-nucleotide polymorphisms (SNPs) for the FCP unit. FCP may be transcribed and may play a role in the spatial arrangement of the genome.

Key words: *Fringilla coelebs* L., pericentric satellite DNA, direct genome sequencing, ThermoFidelase 2A

INTRODUCTION

Repetitive sequences varying in nucleotide composition, copy number, and genome distribution are characteristic of the eukaryotic genome. Although such sequences have been revealed in higher organisms more than 30 years ago, their origin, evolution, and especially function are far from fully understood.

Blocks of multicopied tandem repeats mark pericentric chromosome regions in higher eukaryotes. Satellite DNA of pericentric heterochromatin has been well studied in mammals (including humans, mice, and cattle) but only fragmentarily studied in birds [1, 2]. Here we analyze a new family of pericentric repeats found in the genome of chaffinch *Fringilla coelebs* L. (Fringillidae, Passeriformes).

EXPERIMENTAL

Genomic DNA was isolated from whole blood by the standard technique. Blood (100–150 μl) was obtained from five females, which were trapped for ringing at the Ornithological Station, Zoological Institute, Russian Academy of Sciences, on the Kurish Spit during their seasonal migration in 1997. Blood was stabilized with 0.2% sodium citrate, 0.3% glucose.

Isolated DNA was digested with *Bam*HI, *Pvu*II, *Hin*dIII, *Pst*I, *Sau*3AI, *Msp*I, and *Hpa*II. The products were electrophoretically separated in 1% agarose gel. The major fraction of the *F. coelebs Pst*I fragments (FCP) sized about 500 bp was eluted from gel with a QIAGEN QIAquick kit and cloned in pTZ19U. Clones pFCP1 and pFCP2 were chosen for further experiments.

Genomic proportion of FCP was estimated by dot blotting with the cloned FCP labeled with [^{32}P]dATP (Institute of Applied Chemistry, St. Petersburg) [3]. After hybridization, filters were exposed with an X-ray film. Comparative densitometry of hybridization signals was carried out with the PhotoM 1.0 program (A. Chernigovskii).

Southern hybridization of the radiolabeled cloned FCP fragments to genomic DNA digested with the above restriction enzymes, electrophoretically separated, and immobilized on a filter followed the standard protocol [3].

Sequencing of pFCP1 and pFCP2 with a BigDye kit was done according to a Perkin-Elmer protocol with modification. The reaction mixture (5 μl) contained 50 ng of plasmid DNA and 3.2 pmol of the standard primers for pTZ19U.

Divergence of FCP repetitive units was analyzed by direct genomic sequencing with enzyme complex ThermoFidelase 2A (Fidelity Systems, United States). Based on the established sequence of the cloned FCP, we synthesized forward primers ZF1 (AACACTGGCACTGAGCGTGC) and ZF2 (CTTCTCCGGCAAGGAAGCCT) and reverse primers ZR1 (GCACGCTCAGTGCCAGTGTT) and ZR2 (AGGCTTCCTTGCCGGAGAAG).

Direct genomic sequencing was performed in 5 μl of the reaction mixture containing 300 ng of total genomic DNA, 10 pmol of a primer, 0.1 μl of Thermo-

0026-8933/01/3503-0331$25.00

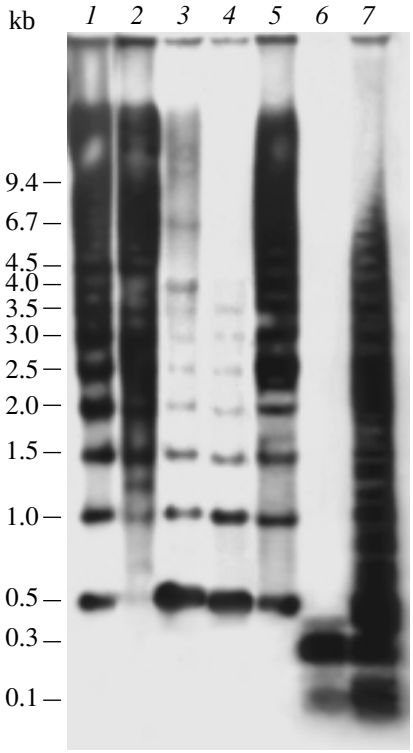

Fig. 1. Southern blotting of ^{32}P-labeled FCP to *F. coelebs* genomic DNA digested with (*1*) *Bam*HI, (*2*) *Pvu*II, (*3*) *Pst*I, (4) *Hin*dIII, (*5*) *Sau*3AI, (*6*) *Msp*I, and (*7*) *Hpa*II.

Fidelase 2A, and 2 μl of the standard BigDye reagent. Amplification included 100 cycles of 5 s at 95°C, 30 s at 55°C, and 4 min at 60°C, with first denaturation for 2 min. The products were analyzed in an ABI377 DNA sequencer.

The FCP nucleotide sequence was analyzed with the BLAST [4], TRANSFAC [5], and ChrClass [6] programs, using the Internet resources.

RESULTS AND DISCUSSION

Each of the enzymes used yielded a distinct restriction pattern of *F. coelebs* genomic DNA. Fragments about 500 bp were the major digestion products with *Bam*HI, *Hin*dIII, *Pst*I, and *Sau*3AI and were not obtained with *Pvu*II, *Msp*I, and *Hpa*II. A band corresponding to fragments about 200 bp was seen on the *Pvu*II and *Pst*I patterns. Several fractions of DNA repeats less than 500 bp were obtained with *Msp*I and *Hpa*II. A fraction of highly repetitive *Pst*I fragments about 500 bp were selected for further analysis, cloned in pTZ19U, and termed FCP. Two independent clones, pFCP1 and pFCP2, were sequenced and used as hybridization probes.

Southern hybridization of genomic DNA with radiolabeled pFCP1 and pFCP2 yielded a regular "ladder" in the cases of *Bam*HI, *Pst*I, *Hin*dIII, and *Sau*3AI (Fig. 1), suggesting that FCP is a tandem repeat of the cloned fragment. Most monomers contained restriction sites for *Bam*HI, *Hin*dIII, *Pst*I, *Sau*3AI, and *Msp*I and lacked the *Pvu*II sites. The *Msp*I fragments less than 500 bp originated from FCP, whereas the *Pst*I and *Pvu*II fragments about 200 bp were not homologous to FCP. Hence, in addition to FCP, the *F. coelebs* genome contains another highly repetitive sequence, which can be obtained with *Pst*I and *Pvu*II.

The relative proportion of FCP in the genome was estimated by dot blotting of the cloned FCP fragment with DNA of pFCP2 and with genomic DNA (Fig. 2). The results of two independent experiments were similar and showed that FCP accounts for about 0.9% of the entire genome. Weighting 2.24 ng [7], the *F. coelebs* diploid genome contains $4 \cdot 10^4$ copies of FCP.

The 506- and 505-nt inserts of pFCP1 and pFCP2 were sequenced, and the sequence was deposited in GenBank (accession no. AF160980). The divergence between FCP1 and FCP2 was 0.6% and was due to nucleotide substitutions. Since FCP accounts for an appreciable proportion of the *F. coelebs* genome, it was necessary to analyze many independent clones in order to establish the monomer consensus sequence and the positions of the most common substitutions. To avoid this labor-consuming procedure, we used direct genomic sequencing with specific primers and ThermoFidelase 2A. The latter contains several enzymes of thermophilic microorganisms and has earlier been employed with success in direct sequencing of bacterial genomes. ThermoFidelase 2A facilitates specific primer annealing and extension and protects DNA against heat damage during denaturation and synthesis, thereby allowing far more amplification cycles.

With direct genomic sequencing, we established the consensus sequence and the five most common single-nucleotide polymorphisms (SNPs) for the FCP unit. The results also confirmed the head-to-tail arrangement of repeated units (Fig. 3). We found a site for *Bam*HI, *Hin*dIII, *Pst*I, and *Sau*3AI and seven sites for *Hpa*II/*Msp*I in the repeated unit, which was consistent with the Southern hybridization data. The *Bam*HI and *Sau*3AI sites each contained SNP and, consequently, occurred in only a fraction of monomers. This explained more intense hybridization of FCP oligomers in the cases of *Bam*HI and *Sau*3AI compared with *Hin*dIII and *Pst*I (Fig. 1).

Satellite DNA mostly occurs in pericentric heterochromatin in higher eukaryotes. On evidence of fluorescent *in situ* hybridization (FISH), FCP occupies the pericentric regions of all *F. coelebs* chromosomes [8]. The data on the nucleotide sequence and chromosomal location of FCP are consistent with negative staining of the pericentric regions with AT-specific

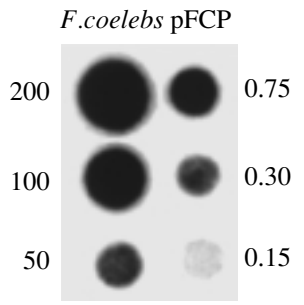

Fig. 2. Dot hybridization of *F. coelebs* genomic DNA and pFCP1 with ^{32}P-labeled insert isolated from pFCP1. The amounts (ng) of genomic DNA applied on a filter and of the cloned FCP fragment, which was computed from the insert and pTZ19U sizes and the amount of pFCP1 DNA applied on a filter, are indicated on the left and on the right, respectively.

fluorochrome DAPI [8]. On average, FCP has 57% GC, testifying again to the enrichment in GC of pericentric heterochromatin.

Nucleotide sequence comparisons with sequences from the NCBI databases assigned FCP to a new family of highly repetitive elements. FCP is the first satellite isolated from the *F. coelebs* genome and lacks a homology to the known pericentric sequences of other bird genomes. It should be noted that, notwithstanding the cloning and analysis of many genomic sequences of various birds, only two families of pericentric repeats are known. One is a highly polymorphic family of repeats about 190 nt occurring in the genome of virtually all bird species studied [1], including Passeriformes. We have isolated such a sequence (GenBank accession no. AF160981) from the sparrow genome. Possibly, the highly repeated 200-bp *Pvu*II and *Pst*I fragments of *F. coelebs* also belong to this family.

Another pericentric repeat, PR1, has been isolated from the genome of two pigeon species [2]. This avian nucleotide sequence contains a region homologous to the CENP-B box and seems to be functionally similar to mammal alphoid satellite DNA. CENP-B is a conserved DNA-binding protein which is involved in formation of the kinetochore plate and has been found in many eukaryotes, including birds [9]. We did not find a putative CENP-B box in the 190-nt repeats and in FCP. Possibly, these repeats interact with other centromere-associated proteins with unknown binding sites. On the other hand, the kinetochore domain is only one out of three functional subdomains recognized in the centromeric region of the eukaryotic chromosome [10].

Until recently, highly repetitive eukaryotic sequences were considered as junk DNA playing no role in the genome. However, pericentric heterochromatin, which contains satellite DNA, is essential for chromosome pairing and segregation during cell division and for cohesion of sister chromatids [10, 11]. Metaphase chromosomes are arranged in a rosette, the centromeres forming a central ring [12]. This implicates the pericentric regions in association of nonhomologous chromosomes in the metaphase rosette. In FISH of the cloned FCP to *F. coelebs* metaphase chromosomes, signals have been detected not only in pericentric heterochromatin, but also in intercentromeric connectives between nonhomologous chromosomes [8]. Such connectives are artifacts of chromosome spreading on the slide surface. Resulting from disruption of the native chromosome structure during cell preparation as they are, these DNA–protein threads do suggest tight association of chromosomes in the cell cycle. Since the threads contain FCP, it is possible to assume that FCP-containing pericentric domains of nonhomologous chromosomes are close together in intact cells, and that their interaction is tight enough for the centromeres to remain connected even in chromosome spreads. Thus, the FCP repeats are possibly involved in spatial arrangement of chromosomes during cell division.

Interchromosomal interactions have been attributed to formation of DNA triplexes [13]. We did not

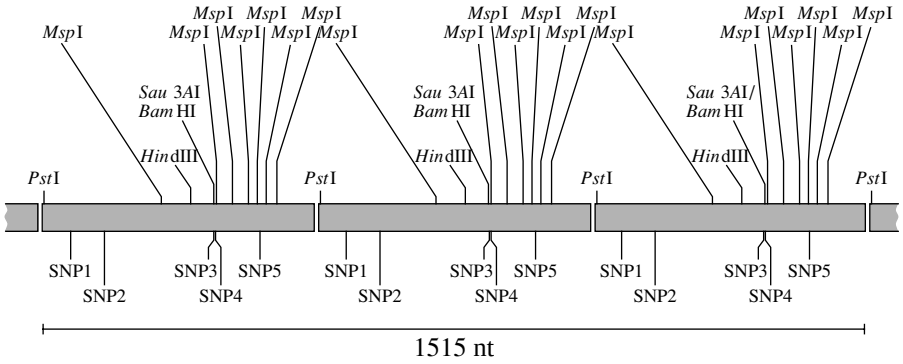

Fig. 3. Scheme of the FCP tandem repeat. Restriction sites and SNPs T/G (SNP1), G/C (SNP2), G/C (SNP3), C/G (SNP4), and A/C (SNP5) are indicated.

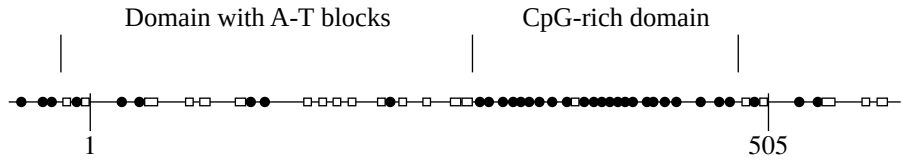

Fig. 4. Distribution of (●) CpG dinucleotides and (□) A-T blocks affecting DNA conformation in FCP.

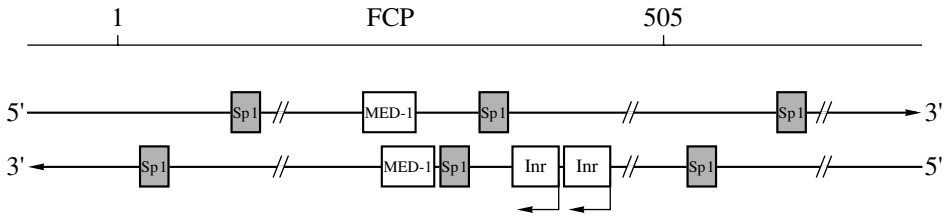

Fig. 5. Arrangement of sequences homologous to regulatory elements MED-1 [17], Inr [18], and the Sp1-binding site [5] in the FCP repeated unit.

detect sequences able to form triplexes and hairpins in FCP. However, the FCP unit contains several interspersed short direct repeats and A-T blocks, which are known to cause DNA bending. These elements proved to be nonuniformly distributed throughout the FCP unit, A-T blocks being clustered in one half of the unit and CpG in the other (Fig. 4).

In vertebrates, cytosines may be methylated in up to 70% of all CpG. Methylation plays an important role in regulating gene expression, as methylated promoters are inactive [14]. With isoschizomers *Msp*I and *Hpa*II differing in methylation sensitivity, we revealed high CpG methylation in the FCP repeats (Fig. 1, lanes *6*, *7*). This suggests inactivation of FCP-containing chromatin in *F. coelebs* blood cells. It is known, however, that demethylation and *de novo* methylation occur in ontogenesis. DNA is demethylated in germline cells and is partly methylated during their differentiation [14]. Hence DNA methylation in developing oocytes is possibly lower than in somatic cells. As observed with diplotene lampbrush chromosomes, noncoding repetitive sequences are transcribed during oogenesis [2, 15, 16]. Thus, transcription of the FCP repeats is also possible. This is supported by the fact that, though lacking the TATA box characteristic of promoters, FCP contains hexanucleotide GCTCCS at positions 377 and 381 in the plus and minus strands, respectively. This hexanucleotide is a conserved multiple start site element downstream (MED-1), which is typical of TATA-less promoters [17]. Two potential transcription initiation regions (Inr) [18] are 95 and 112 nt upstream of MED-1 in the minus strand (Fig. 5). It is known that TATA box and Inr are functionally similar and interchangeable. In addition, an important role in the function of TATA-less promoters is ascribed to transcriptional activator Sp1 [18], and FCP contains Sp1-binding sites at positions 23 and 385 on the minus strand and 90 and 463 on the plus strand (Fig. 5).

The clustering of CpG dinucleotides in the vicinity of Inr and MED-1 also testifies to our assumption, because CpG islands mostly occur in promoter regions and near transcription start sites in the vertebrate genome [14]. If FCP were methylated in all cells throughout the ontogenesis, CpG would be replaced with TpG as a result of spontaneous deamination of 5-methylcytosine during evolution. Since such a tendency was not observed, the GC-rich domain of FCP is possibly nonmethylated and potentially active at least in germline cells.

As already mentioned, CpG dinucleotides are important regulatory elements. Mammals and birds possess multifunctional proteins (MeCP1, MeCP2, ARBP) that specifically bind to mCpG and regulate gene expression, DNA replication, formation of chromatin loops via interaction with scaffold- and matrix-associated regions (SARs/MARs), and specific arrangement of pericentric heterochromatin [14, 19].

Sequence analysis with the ChrClass program [6] did not reveal domains typical for SARs/MARs in the FCP repeats. However, an FCP dimer is similar to SARs/MARs in distribution of motifs associated with DNA bending and, therefore, is potentially able to assume a similar secondary structure. Along with specific nucleotide sequences, specific secondary structure elements are necessary for DNA recognition by various proteins [20]. Possibly, the FCP repeats are new SARs/MARs that are located in pericentric chromosome regions and tend to associate with each other at various cell-cycle stages. The existence of such SARs/MARs has been proposed by R.M. Donev and L.P. Dzhondzhurov [21].

Thus, on the strength of its high conservation, the presence in the pericentric regions of all chromosomes and in fibers connecting the centromeres of nonhomologous chromosomes, and potential capability of transcription and interaction with transcription regulators and with the nuclear matrix, we think that the FCP satellite is functionally significant.

ACKNOWLEDGMENTS

We are grateful to G.V. Glazko for sequence analysis with the ChrClass program and to T.F. Andreeva and V.N. Kulikov for help and advise in manipulations with recombinant DNAs. This work was supported by the Russian Foundation for Basic Research (project nos. 99-04-49592, 00-04-55013); the program Russian Universities—Basic Research (project no. 015.07.01.78) of the RF Ministry of Education; and by the Unique Instrument Support grant (COMUS-1, state registration no. 06-03) from the RF Ministry of Industry, Science, and Technology.

REFERENCES

1. Madsen, C.S., Brooks, J.E., de Kloet, E., and de Kloet, S.R., *Genome*, 1994, vol. 37, pp. 351–355.
2. Solovei, I.V., Joffe, B.I., Gaginskaya, E.R., and Macgregor, H.C., *Chromosome Res.*, 1996, vol. 4, pp. 588–603.
3. Maniatis, T., Fritsch, E.F., and Sambrook, J., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, N.Y.: Cold Spring Harbor Lab. Press, 1982.
4. Zhang, J. and Madden, T.L., *Genome Res.*, 1997, vol. 7, pp. 649–656.
5. Wingender, E., Chen, X., Hehl, R., *et al.*, *Nucleic Acids Res.*, 2000, vol. 28, pp. 316–319.
6. Glazko, G.V., Rogozin, I.B., and Glazkov, M.V., *Mol. Biol.*, 2000, vol. 34, pp. 5–10.
7. Monaghan, P. and Metcalfe, N.B., *Trends Genet.*, 2000, vol. 16, pp. 331–332.
8. Saifitdinova, A.F., Timofejeva, L.P., Zhurov, V.G., and Gaginskaya, E.R., *Tsitologiya*, 2000, vol. 42, pp. 588–593.
9. Saffery, R., Earle, E., Irvine, D.V., *et al.*, *Chromosome Res.*, 1999, vol. 7, pp. 261–265.
10. Craig, J.M., Earnshaw, W.C., and Vagnarelli, P., *Exp. Cell Res.*, 1999, vol. 246, pp. 249–262.
11. Hennig, W., *Chromosoma*, 1999, vol. 108, pp. 1–9.
12. Nagele, R., Freeman, T., McMorrow, L., and Lee, H., *Science*, 1995, vol. 270, pp. 1831–1834.
13. Ikemura, T., Ohno, M., Uemura, T., *et al.*, *Cytogenet. Cell Genet.*, 1998, vol. 81, pp. 101–102.
14. Goto, T. and Monk, M., *Mol. Cell. Biol.*, 1998, vol. 62, pp. 362–378.
15. Diaz, M.O., Barsacchi-Pilone, G., Mahon, K.A., and Gall, J.G., *Cell*, 1981, vol. 24, pp. 649–659.
16. Solovei, I., Macgregor, H.C., and Gaginskaya, E.R., *Kew Chromosome Conf. IV*, Brandham, P.E. and Bennett, M.D., Eds., Kew: Royal Botanical Gardens, 1995, pp. 323–330.
17. Ince, T.A. and Scotto, K.W., *J. Biol. Chem.*, 1995, vol. 270, pp. 30 249–30 252.
18. Kollmar, R., Sukow, K.A., Sponagle, S.K., and Farnham, P.J., *J. Biol. Chem.*, 1994, vol. 269, pp. 2252–2257.
19. Weitzel, J.M., Buhrmester, H., and Stratling, W.H., *Mol. Cell. Biol.*, 1997, vol. 17, pp. 5656–5666.
20. Mäueler, W., Bassili, G., Epplen, C., *et al.*, *Chromosome Res.*, 1999, vol. 7, pp. 163–166.
21. Donev, R.M. and Dzhondzhurov, L.P., *Mol. Biol.*, 2000, vol. 34, pp. 137–142.