Extensive length variation in the ribosomal DNA intergenic spacer of yellow perch, (*Perca flavescens*)

Journal:	<i>Genome</i>
Manuscript ID	gen-2015-0114.R2
Manuscript Type:	Article
Date Submitted by the Author:	18-Nov-2015
Complete List of Authors:	Kakou, Bidenam; Universite du Quebec a Trois-Rivieres, Departement des sciences de l'environnement Angers, Bernard; Universite de Montreal, Departement des sciences biologiques Glemet, Helene; Universite du Quebec a Trois-Rivieres, Departement des sciences de l'environnement
Keyword:	IGS, spacer length variants, rRNA genes, repetitive elements, <i>Perca flavescens</i>

SCHOLARONETM
Manuscripts

Extensive length variation in the ribosomal DNA intergenic spacer of yellow perch, *Perca flavescens*

Bidénam Kakou^{*}, Bernard Angers^{§†} and Hélène Glémet^{*†}

^{*} Département des sciences de l'environnement, Université du Québec à Trois-Rivières, Trois-Rivières, Québec, Canada, G9A 5H7

Bidenam.Kakou@uqtr.ca; Helene.GleMET@uqtr.ca

[§] Department of biological sciences, Université de Montréal, Montréal, Québec, Canada, H3C 3J7
Bernard.Angers@umontreal.ca

[†] GRIL - Groupe de recherche interuniversitaire en limnologie et en environnement aquatique

Running title IGS length classes in yellow perch

Corresponding author

Hélène Glémet

Département des sciences de l'environnement

Université du Québec à Trois-Rivières

C.P. 500

Trois-Rivières, Québec, Canada, G9A 5H7

Helene.GleMET@uqtr.ca

(819) 376-5011 ext. 3364

Draft

ABSTRACT

The intergenic spacer (IGS) is located between ribosomal RNA (rRNA) gene copies. Within the IGS, regulatory elements for rRNA gene transcription are found, as well as a varying number of other repetitive elements that are at the root of IGS length heterogeneity. This heterogeneity has been shown to have functional significance through its effect on growth rate. Here, we present the structural organization of yellow perch (*Perca flavescens*) IGS based on its entire sequence, as well as the IGS length variation within a natural population. Yellow perch IGS structure has four discrete regions containing tandem repeat elements. For three of these regions, no specific length class was detected as allele size was seemingly normally distributed. However, for one repeat region, PCR amplification uncovered the presence of two distinctive IGS variants representing a length difference of 1116 bp. This repeat region was also devoid of any CpG sites despite a high GC content. Balanced selection may be holding the alleles in the population and would account for the high diversity of length variants observed for adjacent regions. Our study is an important precursor for further work aiming to assess the role of IGS length variation in influencing growth rate in fish.

L'espaceur intergénique (IGS) est situé entre chaque copie des gènes de l'ADN ribosomique (ADNr). Le nombre d'éléments régulateurs de la transcription des gènes ADNr ainsi qu'un nombre variable d'autres éléments répétés sont à la base de la variation en taille de l'IGS. Il a été démontré que cette hétérogénéité a un rôle fonctionnel de par son effet sur le taux de croissance. Dans cette étude, nous présentons l'organisation de l'IGS de la perchaude (*Perca flavescens*) à partir de la séquence complète ainsi que la variation en taille de l'IGS dans une population naturelle. L'IGS de la perchaude contient quatre régions différentes contenant des éléments

répétés en tandem. Pour trois de ces régions, il n'y a pas de catégories distinctives et la taille des allèles est normalement distribuée. Cependant, pour la région restante, les amplifications par PCR révèlent la présence de deux allèles distinctifs, présentant une différence de 1116 pb. Cette région répétée est également dépourvue de CpG malgré un fort contenu en GC. La sélection balancée pourrait maintenir ces deux allèles depuis longtemps dans la population et expliquer la diversité élevée d'allèles de tailles différentes observée dans les régions adjacentes. Cette étude est un important précurseur aux futurs travaux portant sur le rôle de la variation en taille des IGS et de son influence sur la croissance des poissons.

Keywords IGS, spacer length variants, rRNA genes, repetitive elements, *Perca flavescens*

INTRODUCTION

The multicopy genes encoding ribosomal RNA (rRNA) are arranged as tandem head-to-tail repeats in ribosomal DNA (rDNA). Each copy is composed of three coding genes (18S, 5.8S and 28S) separated by non-coding internal transcribed spacers (Goldman *et al.* 1983). Copies are separated from each other by the intergenic spacer (IGS) containing terminators of transcription of the previous gene unit in the 5' end and transcription initiation sites of the next gene unit in the 3' end. IGS also includes tandemly-arranged repetitive elements (subrepeats) shown to enhance transcription of rRNA genes from the adjacent promoter (Pikaard *et al.* 1990; Jacob 1995; Banditt *et al.* 1999; Caudy and Pikaard 2002).

Variation in the number of subrepeats results in IGS length heterogeneity (Weider *et al.* 2005) among populations, individuals, and even within individuals (Chou *et al.* 1999). IGS length variation has functional significance through its impact on the evolutionary ecology of organisms, mainly by its effects on growth rate via RNA production (Weider *et al.* 2005). A study conducted on clonal lineages of *Daphnia pulex* subjected to divergent selection, showed that in lineages for high juvenile growth rate, a predominance of long IGS variants was observed over those with lower growth rates (Gorokhova *et al.* 2002). The observed variations in IGS length were related to the number of subrepeat units carrying a promotor sequence in the repetitive region. Furthermore, in two lineages of *Drosophila melanogaster* selected for differential development rates, individuals of the fast developing line possessed longer IGS unlike those in the slow developing line; a higher number of repeats were found within the longer IGS (Cluster *et al.* 1987). Moreover, many studies on invertebrate species have found that the

longer the IGS, the more repeats it contains, notably enhancers which increase rRNA gene transcription rates allowing the organism to synthesize sufficient ribosomes needed to support protein requirements for rapid growth (Cluster *et al.* 1987; Grimaldi and Di Nocera 1988; Gorokhova *et al.* 2002; Fahy *et al.* 2005).

Extensive IGS size variation has been observed in numerous taxa, including fishes. Although fishes display an astonishing variation in growth capacity (Belk *et al.* 2005; McDermid *et al.* 2010), relationships between IGS length and growth rate as seen for invertebrates, have yet to be established. Previous studies on yellow perch, *Perca flavescens* from Lake St. Pierre (LSP) revealed important growth differences for both larvae and adults (Tardif *et al.* 2005; Glémet and Rodríguez 2007). LSP is a highly heterogeneous shallow fluvial lake presenting contrasting environmental conditions (Frenette *et al.* 2006) likely to influence growth rate.

This study aims at determining IGS length variation in yellow perch, *Perca flavescens*. We performed the sequence analysis of the complete rDNA IGS region in order to highlight the location and reveal the organization of regions containing repeats. We investigate each repeat region in individuals from a natural population to assess rDNA length variation arising from the IGS sequence. Although we do not attempt to relate observed IGS variation to growth performance in fish, this study represents a first step for future studies wishing to test this relationship in vertebrates.

MATERIALS AND METHODS

IGS sequence, assembly and characterization

One individual yellow perch was used for IGS sequence characterization. The individual originated from a previous study (Glémet and Rodríguez 2007) where fish were collected in Lake St. Pierre (lat: 46°12; long: 72°50; Quebec, Canada). Genomic DNA was extracted from the liver tissue using the DNeasy Tissue kit (Qiagen, Mississauga, Canada). The entire intergenic spacer (IGS) was amplified by PCR using the primers designed in the 3' end of 28S (IGSF) and the 5' end of 18S (IGSR) (Table 1), using 28S and 18S conserved regions among nine fish species (Genbank accession numbers U34341; BX537263; X02995; AF308735; AF133089; X98839; AF518195; AF021880; X98841). The conserved regions in the 28S and 18S genes were identified by sequence comparison with the CLUSTALW algorithm (Larkin *et al.* 2007). The PCR reaction was performed in a 50 µl final volume containing 4 ng/ul template DNA, 400 nM of each primer, 10 mM of each dNTP, 1X Buffer for AccuTaq LATM DNA Polymerase, 2% DMSO, 0.05 units/ul AccuTaq LATM DNA Polymerase Mix (Sigma-Aldrich Inc., Oakville, USA). PCR conditions were as follows: initial denaturation at 96°C for 30 s, 30 cycles at 94°C for 25 s, 63°C for 15 s, 10 min. at 72°C and 7 min. of final extension at 72°C.

To overcome the problem caused by the presence of repetitive elements frequently detected in IGS (Weider *et al.* 2005), the sequencing strategy adopted consisted of a first step of IGS cloning followed by the creation of nested-deleted plasmids with the Erase-a-base system (Promega, Madison, USA) (Butler and Doers 1995). The amplified fragment was cloned in a plasmid using the TOPO XL PCR cloning Kit (Invitrogen Inc., Burlington, Canada) following manufacturer's

instructions and a library was constructed from the recombinant plasmid. From this library a recombinant plasmid was purified and subjected to an unidirectional Exonuclease III digestion. As a result, a collection of plasmids carrying gradually deleted inserts were obtained and used to create a new library from which plasmids were randomly selected and purified. The last step was the sequencing of each deleted insert using the same M13 Reverse priming site present in all plasmids. Sequencing was conducted at either Université Laval (*Centre de Bio-informatique et de biologie Computationnelle*, Québec, Canada) or Qiagen (Qiagen sequencing services, Hilden, Germany) using either a standard or a hairpin run sequencing method. The latter is a special sequencing reaction that enables accurate sequencing of difficult templates including hairpin structures that technically contain two inverted repeats separated by at least three nucleotides (Kieleczawa 2006).

Sequences were assembled into contigs using the bl2seq program (National Center for Biotechnology Information website) based on the algorithm developed by Altschul *et al.* (1997), and the Seqmerge program (Accelrys, Inc., 2005). Following the amplified fragment sequence reconstitution, the IGS boundaries were positioned by aligning the reconstituted sequence with the ends of 28S and 18S genes from sea bass, *Dicentrarchus labrax* (Genbank accession number DQ660358) and rainbow trout, *Oncorhynchus mykiss* (Genbank accession number AF308735), respectively. Further sequence analyses were required since an initial BLAST search resulted in no sequence similarities to any other known sequences, nor was there recognition of regulatory elements from available published databases. IGS regions containing repeats (repetitive regions) were determined with the Tandem Repeats Finder program (Benson 1999). Several Open Source programs were used for IGS sequence analysis: The Emboss programs for GC content (Isochore) and CpG identification (Cpgplot and Newcpgreport) (Rice *et al.* 2000); Poly (A) Signal Miner to

predict sequence Poly (A) signals (Liu *et al.* 2003) and Proscan: version 1.7 to predict promotor regions based on scoring homologies with putative eukaryotic Pol II promoter sequences (Prestridge 1995).

IGS size variation

To investigate IGS size variation in natural population of yellow perch, 40 individuals from Lake St. Pierre (Leung *et al.* 2011) were analyzed. Total DNA used was previously extracted from individuals by proteinase K digestion followed by phenol-chloroform purification and ethanol precipitation (Leung *et al.* 2011). PCR reactions were carried out to individually amplify regions identified to contain tandem repeats (R1, R2, R3 and R4, Table 2) using pairs of primers reported in Table 1. For R1 region, PCR reaction was performed in a total volume of 12.5 µl containing: 1.25 µl 10X Taq reaction buffer, 0.2mM of each dNTP, 0.4 µM of each primer, 0.5 U of Taq DNA polymerase (Feldan®) and 2µl of DNA template. For R2, R3, and R4 regions, 0.5 ul of Band Sharpener (Feldan®) was used in addition to all reagents used for R1 region. Amplification conditions were as follows for R1: an initial denaturation of 92°C for 30s, 35 cycles of 92°C for 30s, 50°C for 15s and 68°C for 120s followed by one final extension step of 68°C for 5 min. For R2, R3 and R4 regions, amplification conditions were similar except for the annealing step where temperature and duration differed slightly, 62°C for 30 s. For all PCR products, electrophoresis was performed on a denaturing 5% polyacrylamide (19:1 acrylamide:bis-acrylamide) gel. Silver nitrate staining was used to visualize bands (Bassam et al. 1991). A molecular size standard was used on each gel in order to estimate fragment size. The resulting banding patterns for each region were visually analyzed. Allele size was assessed according to the number of repeats contained. For the R2 region, to ascertain that length variants represented true repeat units and

were not a result from INDEL formation, an additional PCR was performed using two restriction enzymes matched to the internal bp frame of the repeat regions. Results were as expected where clear and distinct «cut fragments» were observed for each variant (data not shown) thus confirming the presence of true repeats.

Draft

RESULTS AND DISCUSSION

The fragment amplified by PCR and sequenced is 4792 bp (Tab. 1); the 28S 3'end position is found at nucleotide (nt) 466 and the 18S 5'end starts at nt 4456. Therefore, the putative length of yellow perch IGS is 3990 bp ranging from nt 467 to nt 4455. The sequence has been submitted to GenBank and the accession number is EU325541. Compared to the IGS of some fish, yellow perch IGS length is similar to that of the zebra fish, *Danio rerio* which ranges from 4kb to 7kb (Pelan 2004) but shorter than the IGSs of lake trout, *Salvelinus namaycush* and brown trout, *Salmo trutta* whose lengths are typically 9kb-18kb and 8-16kb, respectively (Castro *et al.* 1997; Reed and Phillips 2000). Among invertebrates, IGS lengths are more similar to yellow perch IGS, where in fact, *Daphnia* species and tadpole shrimp IGS ranges from 4kb to 5kb (Crease 1993; Luchetti *et al.* 2006).

Within the IGS are included four successive regions containing tandemly-arranged repeats (Figure 1). From the 28S 3'end to the 18S 5'end, these repeat regions, R1, R2, R3, and R4 are separated by non-repetitive sequence regions, NR1, NR2, NR3, NR4, and NR5. Such IGS structure is found in most eukaryotes species in which the general IGS organization is constituted by a central region composed of repetitions and non-repetitive regions next to the 28S 3'end and the 18S 5'end (Reeder 1990; Baldrige and Fallon 1992; Crease 1993; Ryu *et al.* 1999). The length of the repeat elements is 106, 62, 91, and 105 bp for the R1, R2, R3, and R4 regions respectively (Table 2). A BLAST search of repeat elements in yellow perch resulted in no sequence similarities to other known sequences; as well, a search of available published databases did not allow the identification of precise regulatory elements. Many studies have established that IGS repetitive elements are regulatory elements for rRNA gene transcription,

notably enhancers whose role is to increase rRNA transcription rate (Wu and Fallon 1998; Morales-Hojas *et al.* 2002). In fact, the enhancement of transcription has been shown for repetitive elements of 60/81 pb, 140 pb, 240 pb/330 pb found respectively in *Xenopus*, mouse and *Drosophila melanogaster* IGS (Busby and Reeder 1983; Labhart and Reeder 1984; Reeder 1984; Grimaldi and Di Nocera 1988; Grimaldi *et al.* 1990; Kuhn *et al.* 1990; Pikaard *et al.* 1990). It is probable that in yellow perch, IGS repetitive regions (regions R1, R2, R3, and R4) include regulatory elements such as enhancers, spacer promoters and terminators that have diverged from known regulatory sequences of other species.

The IGS sequence of yellow perch contains other elements that may have an implication in the transcriptional regulation of rRNA genes. Two predictions of polyadenylation poly (A) signals (PAS), AATAAA were detected slightly upstream and at the beginning of the R1 repeat region. During transcription, RNA polymerase is known to pass through one or more AATAAA sequences beyond the 3' end of the coding region, which communicate the termination of transcription (Proudfoot 2011). Likely these PASs in yellow perch IGS are putative structures of a termination transcription signal for rDNA transcription as they are located in the IGS downstream of the 28S gene. Among other elements are also poly (dA) tracts > 10 bases located mainly within the R1 region and between R2 and R3 regions. The presence of such poly (dA:dT) stretches in intergenic sequences of others eukaryotes has been related to the formation of a specific curved DNA that allows transcription factors to access DNA (Koch and Thiele 1999; Anderson and Widom 2001). Thus, these tracts are considered enhancers elements for gene transcription in general, and constitute functional promoter elements in yeast (Struhl 1985; Winter and Varshavsky 1989; Suter *et al.* 2000). It is probable that poly (dA) tracts present in the

IGS sequence of yellow perch also play an important role in the transcriptional regulation of rRNA genes, as they do for other eukaryotes.

Additionally, although no precise regulatory elements were identified, two candidate promoter regions were predicted within yellow perch IGS based on scoring homologies with putative eukaryotic mammalian polymerase II (PolII) promoter sequences (Prestridge 1995). A first promoter region located between nt 835-1085, corresponding to the end of R1 and beginning of NR2 IGS regions, contained two significant signals for transcriptional elements, E1IF and E2F, with scoring weight of > 50 indicating that the signal is found only in promoter sequences and never in non-promoter sequences. A total of 167 primate Pol II promoter sequences, obtained from the Eukaryotic Promoter Database, and 999 primate non-promoter sequences, obtained from the GenBank sequence databank, are used for establishing scoring weights (Prestridge, 1995). A second region between 2329-2579 nt, corresponding to the end of NR4 and beginning of R4 IGS regions contained a TATA box at nt 2564 nt and the estimated transcription start site at nt 2595 in R4. Although the RNA Polymerase I (PolI) rather than PolII is generally associated with the transcription of rDNA in mammalian cells, there are substantial similarities in their transcription machinery (Russel and Zomerdijsk 2006). Predicted candidate promoter regions based on PolII are likely also implicated in RNA transcriptional regulation by PolI however further investigation to confirm these predictions and to test specific spacer promoter functionality in yellow perch IGS is warranted.

The IGS sequence of yellow perch contains two CpG islands, the first spanning regions R3 to R4, and the second downstream of R4, essentially in the NR5 region (Figure 2). CpG islands have been found in, or near, approximately 40% of promoters of mammalian genes (Fatemi *et al.*

2005) and about half of all CpG islands contain transcription start sites (Deaton and Bird 2011). Typically in mammalian genomes, CpG islands are 300-3,000 base pairs in length, with regions greater than 500 bp considered more likely to be «true» CpG islands if they have a GC content greater than 55%, and an observed-to-expected (obs/exp) CpG ratio of 65%. The CpG islands in yellow perch IGS were identified upstream of the 5' 18S gene, in the vicinity of the predicted promoter and transcription start site (see above). Yellow perch CpG island characteristics seemingly concord with mammalian criteria; CpG1 has 55.9% GC, an obs/exp CpG ratio of 66.8% and is 809bp in length while CpG2 has 62% GC, an obs/exp CpG ratio of 95% and is 903bp in length. These facts suggest CpG islands found within yellow perch IGS subrepeats likely play a role in the regulation of rRNA transcription however in order to determine their exact functional role, extensive analyses still need to be carried out in the future.

CpG islands are conventionally involved in the epigenetic regulation of genes via their methylation status (Gronbaek *et al.* 2007). In vertebrates, contrary to other genes in the genome, housekeeping genes, which include rRNA genes, are closely associated with hypomethylated CpG islands which typically occur at or near the transcription start site of the genes (Gardiner-Garden and Frommer 1987). Here, the unmethylated state of promoters with CpG islands allows genes to be expressed and is contrary to the highly methylated CpG sites found in the coding regions. In fact, hypermethylated rDNA regions are not involved in the transcription regulation of mouse and human rDNA (Mason *et al.* 1997; Shiraishi *et al.* 1999; Grozdanov and Karagyozev 2002). In the human genome, the frequency of the CG nucleotides is 1 %, which is less than 1/4 of the expected frequency. The consensus explanation for CpG deficiency is a spontaneous deamination of methylated cytosines into thymines, effectively decreasing CpG frequency (Duret

and Galtier 2000). In the regions that escape methylation (e.g., the promoter regions of housekeeping genes), CpG dinucleotides are less suppressed. Curiously, the R2 region of yellow perch IGS was completely devoid of any CpG despite a relatively high GC percentage of 57 % (Tab. 2). The functional significance of such a marked CpG deficiency for the R2 region of yellow perch is not evident and not likely related to high mutation of C to T in methylated CpGs given the considerable C content. The R2 repetitive spacer sequence is clearly CpG atypical and its length may serve as a buffer effectively isolating the transcription end site of the 28 S from the transcription start site of the 18S genes, or still yet, it may have other non-recognized transcriptional elements important for rRNA regulation.

Regions R1 to R4 were PCR-amplified separately to assess IGS length variation in a natural population (Figure 3). The number of repeats contained within alleles ranged from two to eight for R1, two to 10 for R3, and from two to 11 for R4. This represents an allele size variation of 636, 728 and 945 bp for each of those regions respectively. No specific IGS length class was representative of the R1, R3, and R4 regions since allele abundance for each of these regions was seemingly normally distributed for the yellow perch population. In contrast, for region R2, the number of repeats ranged from four to 22 representing a length difference of 1116 bp. Allele distribution revealed a clear predominance of both the four- and 22-repeat alleles in the population (Figure 3). In addition, no difference was detected among size variants of R1, R3, and R4 regions for individuals typed as homozygotes 4-4 vs 22-22 of the R2 region (Figure 4). The R2 region is thus unique amongst the other IGS regions, as it reveals the most contrasting IGS alleles. In crustacean and fish, some studies have demonstrated that IGS length heterogeneity is confined to specific regions. For example, the IGS length variation in *Daphnia pulex* is due to a

central region containing a 330 pb/326 bp element (Gorokhova *et al.* 2002); for lake trout, *Salvelinus namaycush* the region adjacent to the 28S is responsible for IGS length variation (Zhuo *et al.* 1995) whereas in *Salmo trutta*, this variation is due to a region upstream of the 3'ETS (Castro *et al.* 1997). Here, the R2 region revealed two contrasting variants implying that in yellow perch, distinct IGS length classes are mainly due to a 62 bp element, repeated either four or 22 times.

The observed diversity of length variants for regions R1, R3, and R4, and not for region R2 in individuals typed as homozygotes 4-4 vs 22-22 of the R2 region (figure 4) is intriguing and indicates that many mutations have persisted in these latter regions while they are absent from region R2. This is seemingly not what is expected according to neutral molecular evolution. A comparison of a two-gene region in *Drosophila melanogaster* revealed that a higher than expected level of silent polymorphisms in the coding region of *Adh* was the cause of departure from selective neutrality and was attributed to a single balanced polymorphism in the *Adh* locus, at the site encoding the *Adh* fast and slow alleles (Kreitman and Hudson 1991). The authors postulate that because a balanced polymorphism maintained by natural selection has existed over a long period of time, any tightly-linked neutral mutations arising thereafter, are inclined to persist longer than would be expected under drift alone. It follows that, if selection has been maintaining the balanced polymorphism observed in the present study over a long time period, it could also have favored the persistence of later mutations appearing in regions surrounding region R2. This could have effectively given rise to the diversity of length variants observed for regions R1, R3, and R4 in individuals typed as homozygotes 4-4 vs 22-22 of the R2 region. That the R2 region be under balanced selection and holding the 4 and 22-repeat alleles in the population, is a plausible hypothesis that warrants further investigation, especially given the

putative functional role of IGS on growth performance.

The observed IGS length variants for R2 region in the present study, where two contrasting allele lengths are present among yellow perch individuals of Lake St. Pierre presents an exciting opportunity for future test of a possible link between IGS length and growth rate in natural vertebrate populations. Furthermore, if the R2 region is shown to influence growth performance, it could be subject to selective pressure through its influence on growth rate. As a genetic marker, the R2 region may be a better indicator of adaptive variation than neutral markers (Beebee and Rowe 2004), and thus more likely to reflect population differentiation due to selection in natural systems.

In conclusion, our study showed that the structural organization of yellow perch IGS is similar to that of many other eukaryotes. While yellow perch IGS has many repetitions confined to four clusters each separated by non-repetitive regions, they have diverged from known regulatory sequences of other species. Still, several elements seemingly implicated in the transcriptional regulation of rRNA genes were identified including, poly A signals, poly (dA) tracts, putative promoter regions, and CpG islands. The repeat R2 region appears to be unique as CpGs are completely absent despite a relatively high GC percentage. In addition, the most contrasting IGS length variants were found for the R2 region among yellow perch individuals. Our study paves the way to future investigations by providing both a genetic marker for population differentiation studies as well as an important tool for unraveling the mechanisms connecting IGS length variation in vertebrates with a major life history trait, notably growth rate in fish.

ACKNOWLEDGEMENTS

We thank M. Léveillé and G. Trottier, for assistance with fish collection and for their laboratory work. Financial support was provided by an individual grant to H.G. from the Natural Sciences and Engineering Research Council of Canada.

Draft

REFERENCES

- Altschul, S. F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z. *et al.* 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**: 3389-3402.
- Anderson, J.D., and Widom, J. 2001. Poly(dA-dT) promoter elements increase the equilibrium accessibility of nucleosomal DNA target sites. *Mol. Cell Biol.* **21**: 3830-3839.
- Baldrige, G. D., and Fallon, A.M. 1992. Primary structure of the ribosomal DNA intergenic spacer from the mosquito, *Aedes albopictus*. *DNA Cell Biol.* **11**: 51-59.
- Banditt, M., Koller, T., and Sogo, J.M. 1999. Transcriptional activity and chromatin structure of enhancer-deleted rRNA genes in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* **19**: 4953-4960.
- Bassam, B.J., Caetano-Anollés, G., and Gresshoff, P.M. 1991. Fast and sensitive silver staining of DNA in polyacrylamide gels. *Anal Biochem.* **196**: 80-83.
- Beebee, T.J.C., and Rowe, G. 2004. Molecular and adaptive variation *In* An introduction to molecular ecology, Oxford University Press, New York. pp. 183-208.
- Belk, M.E., Johnson, J.B., Wilson, K.W., Smith, M.E., and Houston, D.D. 2005. Variation in intrinsic individual growth rate among populations of leatherside chub (*Snyderichthys copei* Jordan & Gilbert): adaptation to temperature or length of growing season? *Ecol. Freshw. Fish*, **14**: 177-184.
- Benson, G., 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**: 573-580.

- Busby, S.J., and Reeder, R.H. 1983. Spacer sequences regulate transcription of ribosomal gene plasmids injected into *Xenopus* embryos. *Cell*, **34**: 989-996.
- Butler, B., and Doers, M. 1995. Generation of Nested Deletions using the Erase-A-Base® System. *Promega Notes*, **53**: 36-38.
- Castro, J., De Lucchini, S., Nardi, I., Sanchez, L., and Martinez, P. 1997. Molecular analysis of a NOR site polymorphism in brown trout (*Salmo trutta*): organization of rDNA intergenic spacers. *Genome*, **40**: 916-922.
- Caudy, A.A., and Pikaard, C.S. 2002. *Xenopus* ribosomal RNA gene intergenic spacer elements conferring transcriptional enhancement and nucleolar dominance-like competition in oocytes. *J. Biol. Chem.* **277**: 31577-31584.
- Chou, C.H., Chlang, Y.C., and Chiang, T.Y. 1999. Within- and between-individual length heterogeneity of the rDNA-IGS in *Miscanthus sinensis* var. glaber (Poaceae): phylogenetic analyses. *Genome*, **42**: 1088-1093.
- Cluster, P.D., Marinkovic, D., Allard, R.W., and Ayala, F.J. 1987. Correlations between development rates, enzyme activities, ribosomal DNA spacer-length phenotypes, and adaptation in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA*. **84**: 610-614.
- Crease, T.J., 1993. Sequence of the intergenic spacer between the 28S and 18S rRNA-encoding genes of the crustacean, *Daphnia pulex*. *Gene*, **134**: 245-249.
- Deaton, A.M., and Bird, A. 2011. [CpG islands and the regulation of transcription](#). *Genes Dev.* **25**: 1010-1022.
- Duret, L., and Galtier, N. 2000. The covariation between TpA deficiency, CpG deficiency, and G+C content of human isochores is due to a mathematical artifact. *Mol. Biol. Evol.* **17**: 1620-1625.
- Fahy, D., Conconi, A., and Smerdon, M. J. 2005. Rapid changes in transcription and chromatin

- structure of ribosomal genes in yeast during growth phase transitions. *Exp. Cell. Res.* **305**: 365-373.
- Fatemi, M., Pao, M.M., Jeong, S., Gal-Yam, E.N., Egger, G., Weisenberger, D.J., *et al.*, 2005. Footprinting of mammalian promoters: use of a CpG DNA methyltransferase revealing nucleosome positions at a single molecule level. *Nucleic Acids Res.* **33**: e176.
- Frenette, J.J., Arts, M.T., Morin, J., Gratton, D., and Martin, C. 2006. Hydrodynamic control of the underwater light climate of the fluvial Lac Saint-Pierre. *Limnol. Oceanogr.* **51**: 2632-2645.
- Gardiner-Garden, M., and Frommer, M. 1987. CpG islands in vertebrate genomes. *J. Mol. Biol.* **196**: 261-282.
- Glémet, H., and Rodríguez, M.A. 2007. Short-term growth (RNA/DNA ratio) of yellow perch in relation to environmental influences and spatio-temporal variation in a shallow fluvial lake. *Can. J. Fish. Aquat. Sci.* **64**: 1646-1655.
- Goldman, W.E., Goldberg, G., Bowman, L.H., Steinmetz, D., and Schlessinger, D. 1983. Mouse rDNA: sequences and evolutionary analysis of spacer and mature RNA regions. *Mol. Cell Biol.* **3**: 1488-1500.
- Gorokhova, E., Dowling, T.E., Weider, L.J., Crease, T.J., and Elser, J.J. 2002. Functional and ecological significance of rDNA intergenic spacer variation in a clonal organism under divergent selection for production rate. *Proc. Biol. Sci.* **269**: 2373-2379.
- Grimaldi, G., and Di Nocera, P.P. 1988. Multiple repeated units in *Drosophila melanogaster* ribosomal DNA spacer stimulate rRNA precursor transcription. *Proc. Natl. Acad. Sci. USA.* **85**: 5502-5506.

- Grimaldi, G., Fiorentini, P., and Di Nocera, P.P. 1990. Spacer promoters are orientation-dependent activators of pre-rRNA transcription in *Drosophila melanogaster*. *Mol. Cell Biol.* **10**: 4667-4677.
- Gronbaek, K., Hother, C., and Jones, P.A. 2007. Epigenetic changes in cancer. *APMIS.* **115**: 1039-1059.
- Grozdanov, P.N., and Karagyozov, L. 2002. The mouse ribosomal DNA amplification promoting sequence 1 is highly methylated and repeated in the genome. *Z. Naturforsch.* **57**: 897-901.
- Jacob, S.T. 1995. Regulation of ribosomal gene transcription. *Biochem. J.* **306**: 617-626.
- Kieleczawa, J. 2006. Fundamentals of sequencing difficult templates- An overview. *J. Bio. Tech.* **17**: 207-217.
- Koch, K.A., and Thiele, D.J. 1999. Functional analysis of a homopolymeric (dA-dT) element that provides nucleosomal access to yeast and mammalian transcription factors. *J. Biol. Chem.* **274**: 23752-23760.
- Kreitman, M., and Hudson, R.R. 1991. Inferring the evolutionary histories of the *Adh* and *Adh-dup* loci in *Drosophila melanogaster* from patterns of polymorphism and divergence. *Genetics*, **127**: 565-582.
- Kuhn, A., Deppert, U., and Grummt, I. 1990. A 140-base-pair repetitive sequence element in the mouse rRNA gene spacer enhances transcription by RNA polymerase I in a cell-free system. *Proc. Natl. Acad. Sci. USA.* **87**: 7527-7531.
- Labhart, P., and Reeder, R.H. 1984. Enhancer-like properties of the 60/81 bp elements in the ribosomal gene spacer of *Xenopus laevis*. *Cell*, **37**: 285-289.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., *et al.* 2007. Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**: 2947-2948.

- Leung, C., Magnan, P., Angers, B. 2011, Genetic evidence for sympatric populations of yellow perch (*Perca flavescens*) in lake Saint-Pierre (Canada): the crucial first step in developing a fishery management plan. J. Aquac. Res. Development, Doi :10.4172/2155-9546.S6-001.
- Liu, H., Han, H., Li, J., and Wong, L. 2003. An *in-silico* method for prediction of polyadenylation signals in human sequences. Genome Inform. **14**: 84-93
- Luchetti, A., Scanabissi, F., and Mantovani, B. 2006. Molecular characterization of ribosomal intergenic spacer in the tadpole shrimp *Triops cancriformis* (Crustacea, Branchiopoda, Notostraca). Genome, **49**: 888-893.
- Mason, S.W., Sander, E.E., and Grummt, I. 1997. Identification of a transcript release activity acting on ternary transcription complexes containing murine RNA polymerase I. EMBO J. **16**: 163-172.
- McDermid, J.L., Sloan, W.N., Wilson, C.C., and Shuter, B.J. 2010. Early life history variation among hatchery and wild-origin lake trout reared in a hatchery environment. Trans. Amer. Fish. Soc. **139**: 21-28.
- Morales-Hojas, R., Post, R.J., Cheke, R.A., and Wilson, M.D. 2002. Assessment of rDNA IGS as a molecular marker in the *Simulium damnosum* complex. Med. Vet. Entomol. **16**: 395-403.
- Pelan, S. 2004. Zebrafish DNA sequence from clone DKEY-208G1 in linkage group 14, complete sequence. Available from <http://www.ncbi.nlm.nih.gov>
- Pikaard, C.S., Pape, L.K., Henderson, S.L., Ryan, K., Paalman, M.H. *et al.* 1990. Enhancers for RNA polymerase I in mouse ribosomal DNA. Mol. Cell Biol. **10**: 4816-4825.
- Prestridge, D.L. 1995. Predicting Pol II Promoter Sequences using Transcription Factor Binding Sites. J. Mol. Biol. **249**: 923-932.
- Proudfoot, N.J. 2011. Ending the message: poly(A) signals then and now. Gene. Dev. **25**: 1770-

1782.

- Reed, K.M., and Phillips, R.B. 2000. Structure and organization of the rDNA intergenic spacer in lake trout (*Salvelinus namaycush*). *Chromosome Res.* **8**: 5-16.
- Reeder, R.H. 1984. Enhancers and ribosomal gene spacers. *Cell*, **38**: 349-351.
- Reeder, R.H. 1990. rRNA synthesis in the nucleolus. *Trends Genet.* **6**: 390-395.
- Rice, P., Longden, I., and Bleasby, A. 2000. EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet.* **16**: 276-277.
- Russell, J., and Zomerdijk, J.C.B.M. 2006. The RNA Polymerase I transcription machinery. *Biochem. Soc. Symp.* **73**: 203-216.
- Ryu, S.H., Do, Y.K., Hwang, U.W., Choe, C.P., and Kim, W. 1999. Ribosomal DNA intergenic spacer of the swimming crab, *Charybdis japonica*. *J. Mol. Evol.* **49**: 806-809.
- Shiraishi, M., Sekiguchi, A., Chuu, Y.H., and Sekiya, T. 1999. Alteration of mosaic methylation of the repeat unit of the human ribosomal RNA genes in lung cancer. *Biol. Chem.* **380**: 81-84.
- Struhl, K. 1985. Naturally occurring poly(dA-dT) sequences are upstream promoter elements for constitutive transcription in yeast. *Proc. Natl. Acad. Sci. USA.* **82**: 8419-8423.
- Suter, B., Schnappauf, G., and Thoma, F. 2000. Poly(dA.dT) sequences exist as rigid DNA structures in nucleosome-free yeast promoters *in vivo*. *Nucleic Acids Res.* **28**: 4083-4089.
- Tardif, D., Glémet, H., Brodeur, P., and Mingelbier, M. 2005. RNA/DNA ratio and total length of yellow perch (*Perca flavescens*) in managed and natural wetlands of a large fluvial lake. *Can. J. Fish. Aquat. Sci.* **62**: 2211-2218.

- Weider, L. J., Elser, J. J., Crease, T. J., Mateos, M., Cotner, J. B. *et al.* 2005. The functional significance of ribosomal (r) DNA variation: Impacts on the evolutionary ecology of organisms. *Annu. Rev. Ecol. Syst.* **36**: 219-242.
- Winter, E., and Varshavsky, A. 1989. A DNA binding protein that recognizes oligo(dA).oligo(dT) tracts. *EMBO J.* **8**: 1867-1877.
- Wu, C.C., and Fallon, A.M. 1998. Analysis of a ribosomal DNA intergenic spacer region from the yellow fever mosquito, *Aedes aegypti*. *Insect Mol. Biol.* **7**: 19-29.
- Zhuo, L., Reed, K.M., and Phillips, R.B. 1995. Hypervariability of ribosomal DNA at multiple chromosomal sites in lake trout (*Salvelinus namaycush*). *Genome*, **38**: 487-496.

Draft

Table 1. PCR primers design for amplification of sequences containing IGS and associated repeat regions, R1, R2, R3, and R4 of yellow perch, *Perca flavescens*. F and R designate forward and reverse primer respectively. W = A or T.

Primer name	Sequence (5'-3')	Target sequence size (bp)
IGS-F	CTAATAGGGAACGTGAGCTGGGTTTAGACCGTC	4792 ^a
IGS-R	GAAAGTTGATAGGGCAGACATTCGAAWGAGWCG	
R1-F	GTAGAGCCGTTCTGTGACA	974
R1-R	CCTAACCCCTAACCCCTCCA	
R2-F	TGGAGGGTTAGGGTTAGG	491
R2-R	CACTTTCAACATGCTCCA	
R3-F	TGGAGCATGTTGAAAGTG	725
R3-R	TCTCCGGTCAAACACTTAG	
R4-F	CTAAGTGTTTGACCGGAGA	1001
R4-R	GGTAACCCGACTTTCAAC	

^a available from GenBank (accession number is EU325541)

Table 2. Sequence characteristics of IGS repeat, (R1, R2, R3, and R4) and non-repeat, (NR1, NR2, NR3, NR4, and NR5) regions in a single yellow perch individual¹.

Region	Relative repeat position (nt)	Total size (bp)	G content	C content	%GC	CpG ratio (Obs/Exp)	Individual repeats (Number and size, bp ²)	
NR1	1-300	300	81	77	52.3	0.77	–	–
R1	301-952	652	165	75	36.8	1.26	6	106 (23)
NR2	953-1181	229	105	37	62.0	0.53	–	–
R2	1182-1461	280	87	71	56.6	0.05	4	62 (32)
NR3	1462-1562	101	38	23	59.8	0.35	–	–
R3' ²	1563-1610 ³	48	2	0	4.2	0.00	2	18 (14)
NR3' ²	1611-1639	29	9	4	44.8	0.81	–	–
R3	1640-2125	486	126	120	50.6	1.06	5	91 (32)
NR4	2126-2351	226	61	53	50.44	0.84	–	–
R4	2352-2943	592	219	123	57.8	0.59	5	105 (67)
NR5	2944-3990	1047	291	343	60.55	0.93	–	–
Total IGS	1-3990	3990	1184	926	52.9	0.78	–	–

¹based on completely sequenced IGS

²number in parentheses indicates truncated repeat size

³minor repeat (R3') and non-repeat (NR3') region included for completeness

Figure captions

Figure 1. General intergenic spacer (IGS) structure for yellow perch, *Perca flavescens*. The IGS is the non-transcribed region found between 18S, 5.8S, and 28S clusters of ribosomal RNA genes. Repeat regions, R1, R2, R3, and R4 within in the IGS are indicated by shaded boxes while non-repeat regions, NR1, NR2, NR3, NR4, and NR5 are indicated by clear boxes. ITS, internal transcribed spacers 1 and 2, numbered from 18S 3' end. Repeat region primers were located upstream and downstream in non-repetitive regions.

Figure 2. GC content in the intergenic spacer (IGS) of yellow perch, *Perca flavescens*. Relative positions of repeat regions R1, R2, R3, and R4 are represented by solid lines. Those of CpG islands, CpG1 (55.9% GC, obs/exp CpG=0.668, 809bp) and CpG2 (62% GC, obs/exp CpG=0.950, 903bp) are represented dashed lines. All regions are numbered from the 5' end.

Figure 3. Allele distribution histograms for intergenic spacer (IGS) sequence containing repeat regions R1, R2, R3, and R4 of yellow perch, *Perca flavescens*. N=40 individuals from Lake St-Pierre. Abundance refers to the number of specific repeat alleles present in the surveyed population.

Shaded bar indicates individual for which entire IGS sequence was obtained.

Figure 4. Allele distribution histograms of R1, R3 and R4 regions of yellow perch, *Perca flavescens* individuals typed as homozygotes 4-4 vs 22-22 of the R2 region. For homozygote 4-4,

N=10 and for homozygote 22-22, N=12 individuals from Lake St-Pierre. Abundance refers to the number of specific repeat alleles present in the surveyed population.

Draft

Figure 1

Figure 2

Figure 3

Figure 4

R1 (0.32%)

```

1 AAATAGGTTGGTATACCAGGGGCGCGGTTAAACGGCGGGAGTAACATTGACTCTCTTAAGGTATAGTGAAAAAGTTAAAGTATTGAAACACTTAGAAAAAAAAA-
2 .....G.....A
3 .....-
4 .....A
5 .....--
6 .....--

```

R2 (2.15%)

```

1 TTAACCCAGACATAGAGCCACGTACTGGGGCTTATGGTGGCCCATGGGGGTCATGGAGCATG
2 .....TC.T.....TC.....C.....A.....
3 .....C.T.....TC.....C.....A.....
4 .....TC.T.....TC.....C.....A.....

```

R3 (0.88%)

```

1 GGAGGCGGTACGAGGTCAACCGCTTCTGCATTCCCCCTTCCCCGGTCGAAAACTTAGAAAAATGTCAAAAAGATTTTGACCGGAGAAGG
2 .....
3 .....G.....
4 .....
5 .....C.....-----

```

R4 (0.83%)

```

1 CCGGATAAAGGGAGGGAGAGGTGGCAGACCTCTACCCGAGTATGGGGCACCTTCCCTGGGGTAAATTGAAGGGTGTTTCGGTGGGGTTTTCTCCACGGTTTTTCGG
2 .....
3 .....C.....
4 .....
5 .....C.....
6 .....C.....-----

```

Figure 1.

Figure Legend

Figure 1. An alignment of the repeats from each of the regions, R1, R2, R3 and R4 in the sequenced IGS of yellow perch, *Perca flavescens*. Estimates of pairwise sequence divergence between repeats in each region are indicated in parentheses.

Draft