

RESEARCH ARTICLE

Open Access



Higher-order organisation of extremely amplified, potentially functional and massively methylated 5S rDNA in European pikes (*Esox* sp.)

Radka Symonová^{1*}, Konrad Ocalewicz^{2,3}, Lech Kirtiklis⁴, Giovanni Battista Delmastro⁵, Šárka Pelikánová⁶, Sonia Garcia⁷ and Aleš Kovařík^{8*}

Abstract

Background: Pikes represent an important genus (*Esox*) harbouring a pre-duplication karyotype ($2n = 2x = 50$) of economically important salmonid pseudopolyploids. Here, we have characterized the 5S ribosomal RNA genes (rDNA) in *Esox lucius* and its closely related *E. cisalpinus* using cytogenetic, molecular and genomic approaches. Intragenomic homogeneity and copy number estimation was carried out using Illumina reads. The higher-order structure of rDNA arrays was investigated by the analysis of long PacBio reads. Position of loci on chromosomes was determined by FISH. DNA methylation was analysed by methylation-sensitive restriction enzymes.

Results: The 5S rDNA loci occupy exclusively (peri)centromeric regions on 30–38 acrocentric chromosomes in both *E. lucius* and *E. cisalpinus*. The large number of loci is accompanied by extreme amplification of genes (>20,000 copies), which is to the best of our knowledge one of the highest copy number of rRNA genes in animals ever reported. Conserved secondary structures of predicted 5S rRNAs indicate that most of the amplified genes are potentially functional. Only few SNPs were found in genic regions indicating their high homogeneity while intergenic spacers were more heterogeneous and several families were identified. Analysis of 10–30 kb-long molecules sequenced by the PacBio technology (containing about 40% of total 5S rDNA) revealed that the vast majority (96%) of genes are organised in large several kilobase-long blocks. Dispersed genes or short tandems were less common (4%). The adjacent 5S blocks were directly linked, separated by intervening DNA and even inverted. The 5S units differing in the intergenic spacers formed both homogeneous and heterogeneous (mixed) blocks indicating variable degree of homogenisation between the loci. Both *E. lucius* and *E. cisalpinus* 5S rDNA was heavily methylated at CG dinucleotides.

Conclusions: Extreme amplification of 5S rRNA genes in the *Esox* genome occurred in the absence of significant pseudogenisation suggesting its recent origin and/or intensive homogenisation processes. The dense methylation of units indicates that powerful epigenetic mechanisms have evolved in this group of fish to silence amplified genes. We discuss how the higher-order repeat structures impact on homogenisation of 5S rDNA in the genome.

Keywords: rDNA, Evolution, Chromosome, Fish, *Esox*, Single cell PacBio sequencing

* Correspondence: Radka.Symonova@uibk.ac.at; kovarik@ibp.cz

¹Research Institute for Limnology, University of Innsbruck, Mondseeerstrasse 9, A-5310 Mondsee, Austria

⁸Laboratory of Molecular Epigenetics, Institute of Biophysics, Czech Academy of Science, Kralovopolska 135, 61265 Brno, Czech Republic

Full list of author information is available at the end of the article



Background

Esox, the only genus in the family Esocidae (Esociformes) contains seven pike species in two monophyletic subgenera (*Esox* and *Kenoza*) with a circumpolar distribution [1]. *Esox americanus*, *E. masquinongy* and *E. niger* live naturally in North America, *E. reicherti* is the only Euroasian esocid endemic to the Amur River basin (Russia and China) while the *E. cisalpinus* (*E. flaviae*) and *E. aquitanicus* are native to Europe [2]. The Northern pike (*E. lucius*) occurs in North America and Eurasia. Its wide distribution and easy access make the Northern pike the most studied esocid species in terms of behaviour, ecology, genetics and evolution. The Northern pike inhabits lakes, rivers and brackish waters. It is an important commercial and recreational species. In the recent years, overexploitation of the natural stocks and climate change has resulted in the dramatic decline of some pike populations [3].

In the Northern pike, nuclear and mitochondrial DNAs exhibit relatively low genetic variability while there is a considerable genetic differentiation among populations [4, 5]. This may be explained by its ecological status—the top predator population size is related to the prey density and/or the bottlenecks that accompanied post-glacial expansion of the pike. Phylogenetic studies confirmed esocids as the closest relatives of the autotetraploid ancestor of salmonid fishes (trout, charrs, salmon, ciscoes and grayling) [1, 6]. Both genome size and the number of chromosome arms are about doubled in salmonids when compared to the pike [7, 8]. The Northern pike genome sequence has been recently published [9] and its linkage groups were successfully mapped on the salmonid reference genomes revealing the importance of *Esox* as the pre-duplication outgroup of salmonids [10]. The genus *Esox* possesses 50 acrocentric chromosomes and the number of chromosomal arms (FN) equalled also 50 [11]. The *Esox* karyotype is thus similar to the presumed karyotype of the diploid common ancestor of salmonids. Hypothetically, the salmonid ancestral karyotype after the salmonid-specific whole-genome duplication (WGD) was composed of 100 uni-armed chromosomes (FN = 100). Subsequent diploidization process included vast chromosome rearrangements that have resulted in formation of different karyotypes in the extant salmonids [7, 12]. Centric fusions retaining the number of chromosome arms and decreasing number of chromosome resulted in the formation of Coregoninae and Salmoninae karyotypes. The Thymallinae karyotypes experienced mostly inversions that increased the number of chromosome arms and retained chromosome number close to the presumed ancestral tetraploid karyotype [12]. Therefore, a deep knowledge concerning the *Esox* karyotype and genome organisation is of much importance to reconstruct and understand the complex evolution of salmonid lineages. However, although

salmonids are one of the best cytogenetically and molecularly studied fish lineages [13–15], published data on *Esox* sp. chromosomes are limited to basic information concerning chromosome number, their morphology and location of the Nucleolus Organizer Regions (NORs) in *E. lucius*, *E. niger*, *E. masquinongy* and *E. americanus* [7, 11].

The 5S rDNA unit consists of a conserved 120 bp genic region and a more variable intergenic spacer (IGS), also known as Non-Transcribed Spacer (NTS). The genic region contains a tripartite RNA Polymerase III promoter composed of three motifs, Box-A, Internal Element (IE) and Box-C [16], which appear to be conserved throughout the tree of life [17, 18]. Thus the 5S rRNA gene appears to be under extreme selection constraints that maintain structural rRNA features essential for ribosome function and conserved transcription binding sites. The presence of RNA Polymerase III promoter within the short genic region makes the 5S gene a relatively autonomous element prone to change its location on chromosomes. Based on the available data, fish harbour an extraordinary capacity to amplify and spread their rDNA sites across chromosomes - according to the currently assembled animal rDNA database (comprising >900 species of vertebrate, arthropods and molluscs), there are eight fish genera among the top ten species with the largest number of 5S rDNA loci (<http://www.animalrDNAdatabase.com>). In fish, up to 30–40 sites were evidenced in diverged lineages – Cypriniformes, Siluriformes, Characiformes, Salmoniformes, and Esociformes [14, 19–22].

Chromosomal mapping and sequence analysis of 5S rDNA give a powerful tool in genome evolution studies. Two basic models have been proposed for the evolution of rRNA genes (reviewed in [23, 24]). The first model, “concerted evolution”, is based on the observation that rDNA units are uniform within the genome while they differ across the genomes [25, 26]. Under this model, a mutation is rapidly spread across the arrays (genome) or is lost and as a result intragenomic homogeneity of genic and non-coding regions is similar. An alternative, the “birth and death” model has been proposed for multigenic families [27]. According to this model, new genes originate by successive duplications, and are either maintained for a long time or are lost, or else degenerate into pseudogenes. Sequence diversity of coding and spacer regions have been taken as criteria to distinguish between both models and a mixed model of these two has also been proposed [28]. Intragenomic heterogeneity of rDNA paralogs has been correlated with the number of loci in the genome [29]. The known high rDNA mobility across chromosomes [30] has been usually ascribed to various translocation [31] and transposition events [19, 32] as well as to polyploidy and interspecies hybridisation [33]. The latter seems to be the most plausible explanation of the heterogeneity because parental genomes may donate divergent paralogs to their

hybrid nucleus. In fish hybrids, the intra- and intergenomic variation was exploited in numerous phylogenetic studies [28, 34].

Despite the wealth of knowledge about 5S rDNA sequence and its chromosome position little is known about the higher-order repeat organisation, in which a block of multiple basic repeat units forms a larger repeat unit of repeats. Here, we exploited PacBio genomic resources now available [9] to determine the higher-order organisation of 5S rRNA genes that to our best knowledge has not been examined to date in any system. We addressed the question of structure, sequence homogeneity and evolution of 5S rRNA genes in two related species *E. lucius* and *E. cisalpinus*. We applied an integrative approach involving classical cytogenetics, molecular biology and *in silico* genomics methods.

Methods

Species and sample collections

Following specimens of the Northern pike (*Esox lucius*) were cytogenetically studied: ten young unsexed specimens purchased in a fish farm in Libechov, Czech Republic and 22 8-month old specimens (12 males, nine females and one unsexed) from a fish farm in Olsztyn, Poland. Of the only recently described Southern Pike (*Esox cisalpinus*) we analysed cytogenetically: fourteen young unsexed specimens from the Provincial Fish hatchery of Carmagnola, Italy (young specimens obtained from wild pike parents collected in the Turin Province, determined by G. B. Delmastro and E. Sala), a single specimen from the Po river, Carmagnola, locality Gerbasso, determined by G. B. Delmastro and finally, two specimens determined by G. B. Delmastro that are deposited in the ichthyologic collection of the Muséum National d'Histoire Naturelle (Paris, France).

Molecular cytogenetics

Molecular cytogenetic methods were carried out independently in laboratories in Poland and Czech Republic. The karyotypes were assessed following standard procedures [14, 35]. Briefly, fish were injected with 0.3% colchicine (Sigma-Aldrich, St. Louis, MO, USA) solution (0.25 ml/100 g body weight) 60 min. before being sacrificed. Portions of cephalic kidneys were removed, placed in 5 ml of a hypotonic solution (0.075 M KCl) and homogenised using glass homogenizers. Cell suspensions were then transferred to the 10 ml glass tubes, hypotonised for 40 min at RT and fixed with freshly prepared ice cold fixative (methanol: acetic acid, 3: 1, v/v). The fixative was changed three times before splashing on microscopic slides. Somatic metaphase plates were prepared by conventional air-drying technique with some modifications [36]. For visualization and description of chromosome morphology, metaphase spreads were

stained with 4', 6-diamidino-2-phenylindole DAPI (Vector, Burlingame, CA, USA).

Based on the number and the quality of the metaphase spreads, we selected two (*E. cisalpinus*) and five (*E. lucius*) individuals for the Fluorescence in situ Hybridisation (FISH) carried out according to [15, 35]. The 5S rDNA probe was obtained by amplification of *Esox* genomic DNA using the forward primer 5S-1: 5'-TACGCCCG ATCT CGTCCGATC-3' and the reverse primer 5S-2: 5'-CAGGCTGGTATGGCCGTAAGC-3' [37]. For the 28S rDNA probe, following primers were used: 5'-AAAC TCTGGTGGAGGTCCGT-3' and 5'-CTTACCAAAGT GGCCCACTA-3'. PCR products were purified using the GeneElute PCR Clean-Up Kit (Sigma, USA) and labeled by incorporation of Biotin-16-dUTP (28S) and Digoxigenin-11-dUTP (5S) by nick-translation method (Roche, Switzerland). FISH with 150 ng of rDNA probe per slide was performed with RNase-pretreated and formamide-denatured chromosome slides. Post-hybridisation wash was performed at 37 °C for 20 min. Detection of FISH signals was performed using Avidin-FITC and anti-Digoxigenin-Rhodamine, Fab fragments, respectively (Roche, Switzerland). Only high quality metaphase cells were examined under Zeiss Axio Imager A1 (Zeiss, Germany) and Nikon 90i (Nikon, Japan) microscopes equipped with epi-fluorescence and digital (Applied Spectral Imaging, Galilee, Israel) and monochromatic ProgRes MFcool (Jenoptik, Germany) cameras, respectively. Images were captured and the electronic processing of the images was performed using the Band View/FISH View software (Applied Spectral Imaging) and Lucia software ver. 2.0 (Laboratory Imaging, Czech Republic). Post-processing elaboration of the pictures was made using CorelDRAW Graphics Suite 11 (Corel Corporation, Canada).

Cloning and Sanger sequencing

DNA was isolated from blood cells, muscles and fins of adult individuals from several *E. lucius* and *E. cisalpinus* individuals using the classical phenol-chloroform extraction method. The crude DNA fraction was re-purified by DNeasy Blood and Tissue kit (Qiagen, Germany). PCR was used to amplify 5S units from genomic DNA of *E. lucius* and *E. cisalpinus* using PCR as described above. After agarose gel electrophoresis, two fragments of about 220 and 450 bp were visualised in each species. Both fragments were purified and cloned into the pGEM-T vector (Promega, USA). Two clones from each species were sequenced: the short inserts contained monomeric units carrying the 120 bp genic region and an intergenic spacer; the long inserts contained dimers with two copies of the genic region and a spacer. In addition, one 5S-carrying clone was obtained from the *E. lucius* genomic library prepared by digestion of DNA

with the *AseI* restriction enzyme. The sequences were submitted to the GenBank under the accession numbers (KX950799, KX965715-KX965718).

Southern and slot blot hybridisation

The procedure followed the protocol described by [38]. The 5S rDNA probe was a 243 bp insert of the 5S_Eci_a clone (GenBank KX965716) from *E. cisalpinus*. The plasmid insert was amplified and labelled with the 32P-dCTP (DekaPrime kit, Fermentas, Lithuania). The probe was hybridised at high stringency conditions (washing 2x SSC, 0.1% SDS followed by 0.1xSSC, 0.1% SDS at 65 °C). The hybridisation signals were visualised by Phosphor imaging (Typhoon 9410, GE Healthcare, PA, USA) and signals were quantified using ImageQuant software (GE Healthcare, PA, USA). The copy number of 5S genes was estimated using slot blot hybridisation. Briefly, the DNA concentration was estimated spectrophotometrically at OD_{260nm} using Nanodrop 3300 Fluorospectrometer (Thermo Fisher Scientific, USA). Concentrations were verified by the electrophoresis in agarose gels using dilutions of lambda DNA as standards. The three dilutions of genomic DNA (100, 50 and 25 ng), together with a serial dilutions of unlabelled plasmid inserts corresponding to the 5S monomers (GenBank KX965715-6), were denatured in 0.4 M NaOH and blotted onto a positively charged Nylon membrane (Hybond XC, GE Healthcare, USA) using a vacuum slot blotter (Schleicher-Schuell, Germany). The probe and the hybridisation conditions and visualisation of signals were the same as described above.

Intragenomic variation and rDNA copy number determined from NGS reads

The source data was the SRR1197512 archive containing Illumina HiSeq 2000 reads from the whole genome sequencing project of *Esox lucius* (SRX494131, University of Victoria, isolate CL-BC-CA-002) [9]. Sequence downloads and basic read manipulations of the genomic reads

were done with the aid of the Galaxy Server [39]. The starting read pool of NGS reads consisted of more than 900 million unpaired reads. Before mapping all reads with Ns, reads less than 90 nt in length or reads below quality Phred scores 30 were removed using 'QC and Manipulations' tools at the Galaxy server. The data in FASTQ formats were imported into the CLC Genomics Workbench 6.5.1 (Qiagen, Germany) (CLC). The number of reads was then reduced (Table 1) to decrease computing time using a Sample Reads" command. The high quality reads were mapped to the reference sequences: the 220 bp fragment of 5S rDNA from *E. lucius* (GenBank KX965716) and the 1581 bp fragment of 18S sequence from the pike icefish (*Champscephalus esox*) (AF518187) using the 'Map Read Reference' tool (CLC). The parameter settings were as follows: mismatch cost value: 2, insertion cost value: 3, deletion cost value: 3, with both the length fraction value and the similarity fraction value set at 0.5 and 0.8, respectively. Files with the mapped reads were saved and used for the downstream copy number and SNP analyses. Transcriptomic reads obtained from SRR1228710-12, SRR1228725 and SRR1228729 archives were mapped to 5S genic region (GenBank KX965716) as described above. Relatively few cDNA reads were mapped probably due to the fact that RNA was polyA-filtrated prior to library construction.

The genome proportion (GP, in percentages) of rDNA was calculated as the number of mapped reads versus total number of reads. The genome space (GS, in megabases) was calculated from the formula: genome size x GP of rDNA units. The copy number was calculated from GS values (in bp) divided by the size of the 5S monomer (220 bp) or part (1581 bp) of the 18S gene. The mapped rDNA units were relatively evenly covered by NGS (each 5S gene nucleotide was covered >3000 x) allowing reliable calculation of genome proportions (not shown).

Variants were called via the 'Probabilistic Variant Detection' function tool in CLC using default settings. SNPs were filtered as follows: minimum read coverage=300, count (the number of countable reads supporting

Table 1 Copy number of rRNA genes in *E. lucius* determined from NGS reads and Southern blot hybridisation

rDNA	Platform ^a /method	Read Archive accession	Total reads	Mapped reads ^b /BLAST hits	GP ^c (%)	GS ^d rDNA (Mb)	Copies ^e (1C)
5S	<i>IL</i>	SRR1197513	166,395,976	366,676	0.22	2.42	~20,200
	<i>PB</i>	SRR1930096	555,762	29,151	0.16	1.81	~18,000
	<i>S.blot^f</i>				5	53	~200,000
18S	<i>IL</i>	SRR1197513	166,395,976	199,694	0.12	1.32	~820
	<i>PB</i>	SRR1930096	555,762	1,436	0.08	0.91	~570

^aSequencing platform: *IL* – Illumina; *PB* – PacBio

^bThe number refers to the number of Illumina reads mapped to 5S reference or 5S hits with the PacBio data base

^cGP – genome proportion

^dGS – 5S genome space was calculated from GP: genome size (Mb) * GP (%), considering 1100 Mb/1C the *E. lucius* genome size

^eThe copies were calculated as follows: GS (bp) divided by size of the 5S unit monomer (220 bp)

^fExperimental evaluation of copies by the slot blot hybridisation (Additional file 2: Figure S2)

the allele)–50, frequency (the ratio of “the number of ‘countable’ reads supporting the allele” to “the number of ‘countable’ reads covering the position of the variant”): $\geq 5\%$ (high frequency SNPs).

Analysis of long 5S rDNA arrays within the PacBio reads

We created a BLAST library from the sequence archive SRR1930096 comprising >500,000 PacBio reads. The library was BLASTed against the NGS consensus (the 5S genic region plus the spacer) built from the *E. lucius* Illumina mapped reads. These PacBio reads (2640) were then extracted (primary 5S archive). Because the higher-order organization of units was the primary goal we filtered the primary archive to obtain longer reads of ≥ 10 kb. This step yielded 286 reads with Phred quality scores $Q = 10$ –11. The average size of read was 12.5 kb; the longest read was 29,914 bp. To determine the number of 5S rRNA genes in each read we used MultiBlast search ($e = 1.0$) and queried the 286 sequences against the 120 bp 5S NGS consensus (genic region). MultiBlast outputs were exported in the csv format to MS Excel and further processed. Pairwise comparisons of “self to self” read or “read to the 5S rDNA unit NGS consensus” was carried out for each read using the YASS genomic similarity search tool [40]. The alignment parameters were as follows: Scoring matrix (match, transversion, transition, other): +5, -4, -3, -4; gap costs (opening, extension): -16, -4; E-value threshold 0.001. X-drop threshold: 30. In order to reveal degenerate genes, the E-value was increased to 1.0 in some cases. This less stringency search usually resulted in about 15% increase in the number of hits. Outputs were presented as dot-plots.

Spacer variants in long PacBio reads were analysed using the ‘Search Motif Tool’ function in CLC. The search query involved a 10 nucleotide intergenic spacer sequence containing four highly polymorphic positions in the middle. Stringent conditions were applied scoring alignments with more than 90% matches along the 10 nucleotide query. Detected motifs were annotated and counted.

Phylogenetic and secondary 5S structure analysis

Alignments were built using the 120-bp long genic sequences originating from: (i) clones isolated in this work, (ii) clones from the GenBank, (iii) sequenced rRNA and (iv) cDNA consensus sequences prepared from the mapped transcriptomic reads. ClustalW alignment was implemented within the BioEdit program [41]. A phylogeny NJ tree was constructed using the Seaview program [42]. All calculations were run at default conditions using the Jukes-Kantor model and 1000 replicates.

Phylogenetic NJ trees were constructed from Illumina reads using short 70 bp subregions derived from the 5S genic region (Elu_b clone, position 201–270, GenBank KX965717) and the intergenic spacer (the same clone,

position 108–177). The stand-alone BLAST library of SRR1197512 was searched using these subregions as queries. Hit reads were extracted, trimmed for unique lengths (70 bp), sampled to 500 and aligned (‘Multiple Alignment’ tool function of the CLC). Unrooted NJ trees were constructed from aligned reads employing the Jukes-Cantor model and visualised in radial projections. Haplotypic diversity was calculated from aligned reads using the DNAsp4 program [43].

Secondary structure modelling was carried out using an online tool at the RNAfold web server ([44], <http://rna.tbi.univie.ac.at/>). The secondary structures were based on minimum free energy (MFE) calculations using the Turner 2004 model. The program setting was as follows: isolated nucleotides were avoided; vote for dangling energies on both sides of a helix in any case.

DNA methylation analysis

The purified genomic DNA samples from *E. lucius* (2 individuals) and *E. cisalpinus* (3 individuals) were digested with methylation-sensitive *HpaII* (sensitive to CG methylation) and its methylation-insensitive *MspI* isoschizomere (both enzymes are cutting at CCGG). The restriction fragments were hybridised on blots with the alpha[P32]dCTP-labelled 5S probe (Eci_a clone, GenBank KX965716). Control of digestion efficiency was carried out by spiking the *Esox* genomic DNA with a non-methylated plasmid DNA (pBluescript, Stratagen) and subsequent hybridisation with a plasmid probe. Both *MspI* and *HpaII* enzymes yielded expected restriction fragments (not shown).

Results

Localisation of 5S and 45S rDNA loci and heterochromatin on *Esox* chromosomes

Both *E. lucius* and *E. cisalpinus* showed the same number of chromosomes ($2n = 50$) exhibiting strict acrocentric morphology ($FN = 50$). We used FISH to determine the number and position of rDNA loci in *Esox* chromosomes using the 5S and 28S rDNA probes (Fig. 1a and b). In *E. cisalpinus*, the 5S rDNA probe hybridised to 30–34 sites (Fig. 1a, quantitative data are summarized in Additional file 1: Figure S1), all in the (peri)centromeric regions. The 28S probe hybridised to (peri)centromeric sites corresponding to NORs on two homologs (single locus) that lacked the 5S rDNA signals (Fig. 1a). In *E. lucius*, the 5S rDNA probe hybridised to 30–38 sites in (peri)centromeric regions (Fig. 1b, quantitative data are summarised in Additional file 1: Figure S1). The 28S rDNA probe hybridised to two NORs that co-localised but not overlapped with 5S signals. The 28S rDNA probe hybridised more distally compared to the 5S rDNA probe. These different patterns of the chromosomal distribution of both major and minor rDNA

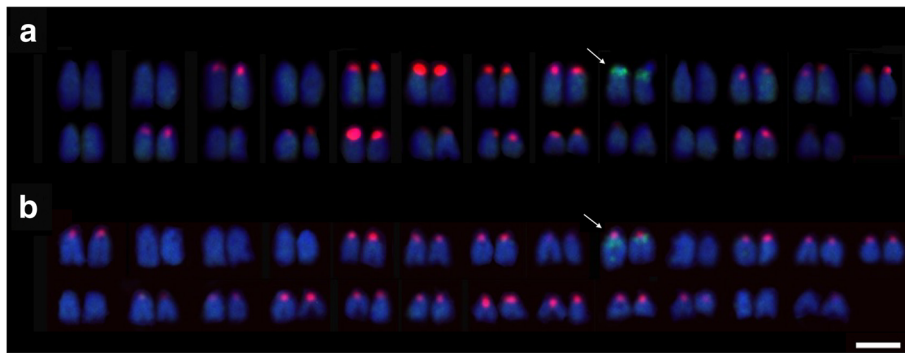


Fig. 1 Molecular cytogenetic FISH analysis with rDNA probes to *Esox cisalpinus* (a) and *E. lucius* (b) chromosomes shown on representative karyotypes. The probes were 5S rDNA in red (31 signals in *E. cisalpinus* and 37 signals in *E. lucius*) and 18S rDNA in green (two signals in both species marked by arrows). Chromosomes are arranged in pairs approximately in decreasing size. More quantitative data on counts of 5S rDNA signals are provided in Additional file 1: Figure S1. Scale equals to 5 μ m

sequences observed in *E. lucius* and *E. cisalpinus* enabled cytogenetic identification of these two species.

Massive amplification of 5S gene copies

FISH showed an extraordinary high number of strong 5S probe signals on numerous *Esox* chromosomes indicating a high copy repeat. This observation provoked a question about the 5S gene copy number in both *Esox* genomes. To determine the 5S copy number we first applied computation approach based on the proportion of mapped reads relative to the total reads (Table 1). The 5S rDNA copy number calculated from both data sets (coming from Illumina and PacBio platforms) were in agreement. The 18S rDNA copy number was at least 20 fold lower than that of the 5S rDNA. This is in line with differences in sites number (single 18S site and about thirty six 5S sites, Fig. 1).

We also determined the 5S rDNA copy number using the classical slot blot hybridisation (Additional file 2: Figure S2). Both *E. lucius* DNA isolates yielded more than 200 thousand copies confirming extreme gene amplification. However, the copy number determined by slot blot hybridisation was >10 fold higher than that one calculated from NGS reads.

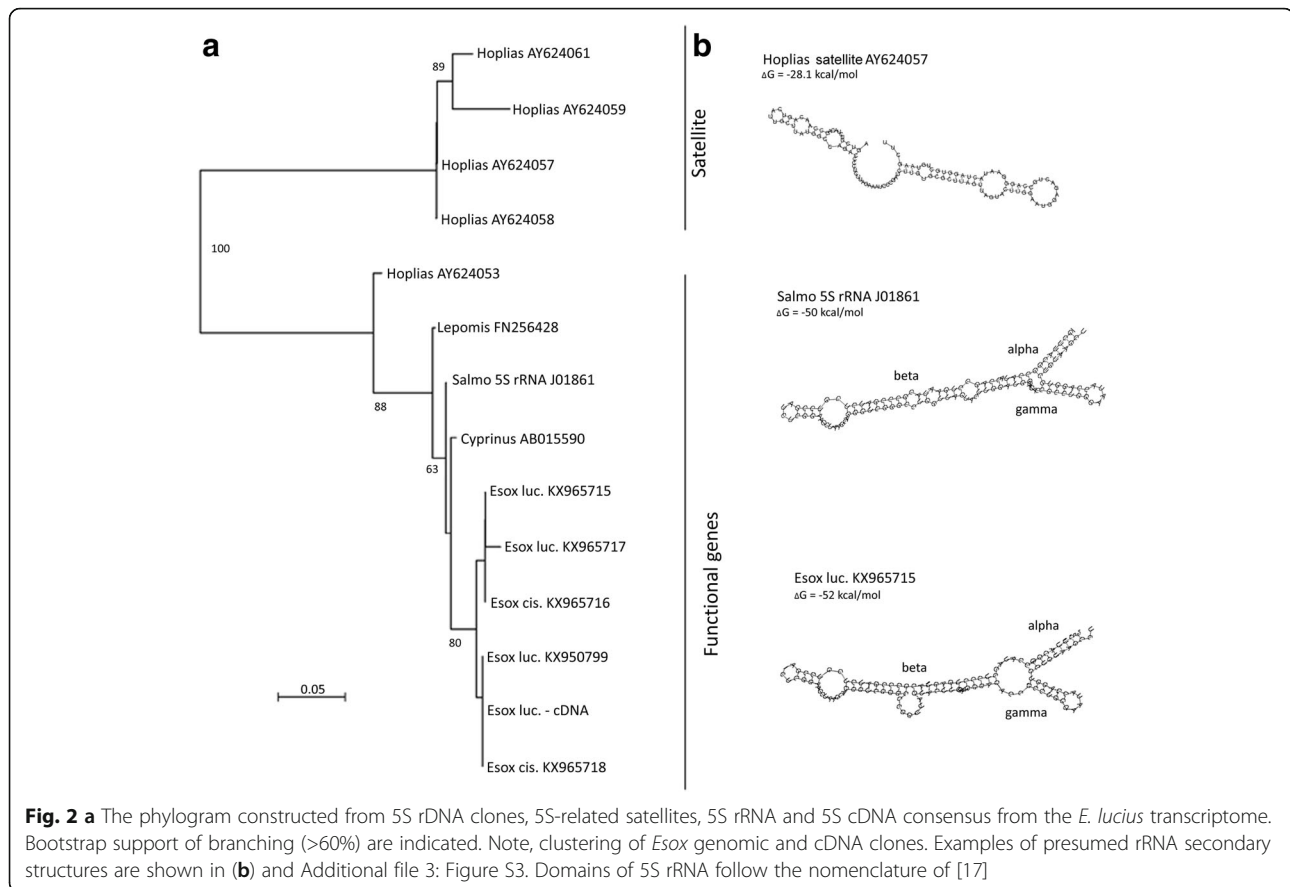
Cloning, sequencing and the phylogenetic analysis

A phylogenetic tree (Fig. 2a) was constructed from an alignment comprising several genomic clones from the GenBank, sequenced 5S rRNA genes [37] and a cDNA consensus sequence built from *E. lucius* transcriptomic reads. In addition, a 5S-derived satellite from *Hoplias malabaricus* was included. The sequences grouped into two main clades: (i) The upper clade contained the 5S-derived satellite clones from *H. malabaricus*. Relatively long branches indicate a considerable sequence divergence. (ii) The second clade was formed by a group of genomic clones from different species including those of *Esox*, sequenced 5S rRNA from rainbow trout (*Salmo*)

and the cDNA consensus from *E. lucius*. The sequences of clones from *E. lucius* and *E. cisalpinus* were similar both within and across species. Consequently, clones from *E. lucius* and *E. cisalpinus* could not be separated and formed a common branch with the 5S cDNA consensus. Within the functional genes, the *H. malabaricus* sequence (accession AY624053) was relatively well separated from other sequences isolated from *Lepomis*, *Salmo*, *Cyprinus* and *Esox*. This is understandable since *Hoplias* represents the most divergent genus out of the five fish genera analysed [45] indicating that the 5S tree roughly reflected the phylogeny.

Conserved secondary structures of *Esox* 5S rRNA

Thermodynamic stability of 5S rRNA secondary structure has been considered as an important criterion for gene functionality [46]. In general, the stable three domain structure is an attribute of functional genes. To investigate whether the amplified *Esox* 5S rDNA code for any functional molecules, we predicted the rRNA structures of several *Esox* molecules by computer modelling and results were compared with those of other genera. Indeed, the *E. lucius* clones produced comparable molecule shapes as the 5S rRNA of *Salmo* (rainbow trout) (Fig. 2b) and other potentially functional fish 5S rRNA genes (Additional file 3: Figure S3). The thermodynamic stability was comparable (high) between the species. The 5S-derived satellite from *Hoplias malabaricus* had significantly lower ($\Delta G \sim 28$ kcal/mol, in average) thermodynamic stability than the functional genes (typically $\Delta G \sim 50$ kcal/mol) explaining why the satellite structures were vastly different from functional genes of both *Salmo* and *E. lucius*. The typical Y-shape of the fish 5S rRNA was not so pronounced as in other species (Additional file 3: Figure S3) [17, 23]. The structural asymmetry of fish 5S rRNA molecules was apparently caused by a longer beta and shorter gamma domain, respectively.



Low intragenomic heterogeneity of the 5S genic region contrasts with higher diversity in the intergenic spacers

In order to determine the intragenomic homogeneity of 5S rDNA we explored the sequenced genome of *E. lucius*. The Illumina reads were mapped to the 5S rDNA reference clone ‘a’ from *E. lucius* (GenBank KX965715) and subjected to the analysis of variance. We considered only high confident SNPs occurring at $\geq 5\%$ frequency, i.e. at least 1000 genes that carry such a variant (considering there may be ~20,000 copies of 5S genes in the genome). Quantification of four kinds of sequence variation (indels and substitutions) along the genic region is shown in Fig. 3A and Table S1 in Additional file 4. Indels were rare and most polymorphisms could be attributed to single nucleotide substitutions. In the 120 bp genic region, only three SNPs, all substitutions, were found. The SNPs located outside of the regulatory motifs (Boxes A and C and IE) and did not significantly affect secondary structure (not shown). The site containing a T>C mutation at position +45 was also polymorphic among the cDNA reads (Additional file 4: Table S2) suggesting that both variants are expressed. The intragenomic SNPs were approximately five fold more abundant (12.8 SNPs per 100 bp) in the spacer than in the genic region (2.5 SNPs per 100 bp). The 28S

rRNA gene was slightly more homogeneous having only 1.2 SNPs per 100 bp of sequence.

In order to determine phylogenetic relationships between the 5S families we constructed haplotypic networks from 500 randomly selected Illumina reads mapping parts of the genic and IGS regions, respectively (Additional file 5: Figure S4A, B). It is evident that the NJ tree constructed from IGS was highly bifurcated compared to the one constructed from genic sequences. The overall diversity expressed as the number of substitutions/100 bp (Pi) was higher (about three-fold) in the spacer region (Pi = 0.0508) than in the genic region (Pi = 0.0145).

Higher-order organisation of 5S arrays

To date, the regularity of 5S tandems and distribution of variants of arrays in the genome have not been assessed due to technical difficulties related with sequencing of long and repetitive molecules. Only now, the single molecule sequencing technologies, such as the PacBio, appear to be suitable for such studies since they generate longer reads than other sequencing platforms. We took advantage of the availability of the PacBio-sequenced *E. lucius* genome (ENA archive SRR1930096) and addressed the question of distribution of 5S intergenic spacer variants (i) and higher-order organisation of 5S arrays (ii):

ii. The higher-order organisation of 5S repeats was investigated in 286 PacBio molecules extracted from the 5S rDNA BLAST search dataset and size-filtered for 10–30 kb. This subset representing about 43% of total *E. lucius* 5S rDNA was analysed for gene richness (Fig. 4). The 5S genes number varied markedly (1–110) between individual reads indicating differences in genomic organisation of repeats. To analyse the higher-order repeat organisation, each sequence was subjected to self to self comparison (Additional file 6: Figure S5). Based on the resulting dot-plot profiles, four types of higher-order organisation can be distinguished (Table 2 and Additional file 7: Tables S3): (i) Group I, molecules containing continuous blocks of 5S tandem repeats spanning the entire read length (Fig. 5a). (ii) Group II, molecules containing one or several blocks of 5S head to tail tandems plus variable portions of unrelated mostly unique sequences (Fig. 5b). In one read, the 5S block was attached to another block of 5S-unrelated tandems (Additional file 7: Table S3). (iii) Group III, molecules harbouring invertedly repeated blocks of 5S tandems that may or may not be separated by spacers (Fig. 5c and Additional file 8: Figure S6). (iv) Group IV, containing no longer blocks (Fig. 5d), but rather dispersed 5S copies. Despite relative abundance (37% reads) the 5S gene richness was low (~4%) in this group (Table 2). In contrast, gene representation in

reads bearing long blocks (Groups I–III) was high (96%). Pairwise comparisons of reads with the 5S reference (the genic region consensus built from Illumina reads) allowed us to determine variation in length of intergenic spacers. The three major spacer length variants were identified: (i) Short 95–116 bp variant occurring in 7035 (94%) units; (ii) Long 321–340 bp variant found in 357 (5%) units; (iii) Ultralong 1153–1209 bp variant found in 65 (1%) units (Additional file 7: Table S3). The sequences of short and long spacer variants were unrelated. Except for one read, spacer variants formed independent arrays.

The majority of 5S rRNA genes are highly methylated

The methylation status of 5S rDNA genes in *E. lucius* and *E. cisalpinus* was determined by enzymatic digestion of genomic DNA with methylation-sensitive *HpaII* and insensitive *MspI*. In all samples, the probe hybridised to high molecular weight bands after the digestion of DNA with methylation-sensitive *HpaII* (Fig. 6). In contrast, a major low molecular weight band of about 220 bp was visible after the digestion with *MspI*. The slow-migrating oligomeric *MspI* bands were faint confirming a high homogeneity of arrays. The near complete resistance of 5S rDNA to *HpaII* digestion indicated high level of methylation of most of the units. There were no differences in methylation profiles between the blood and fin

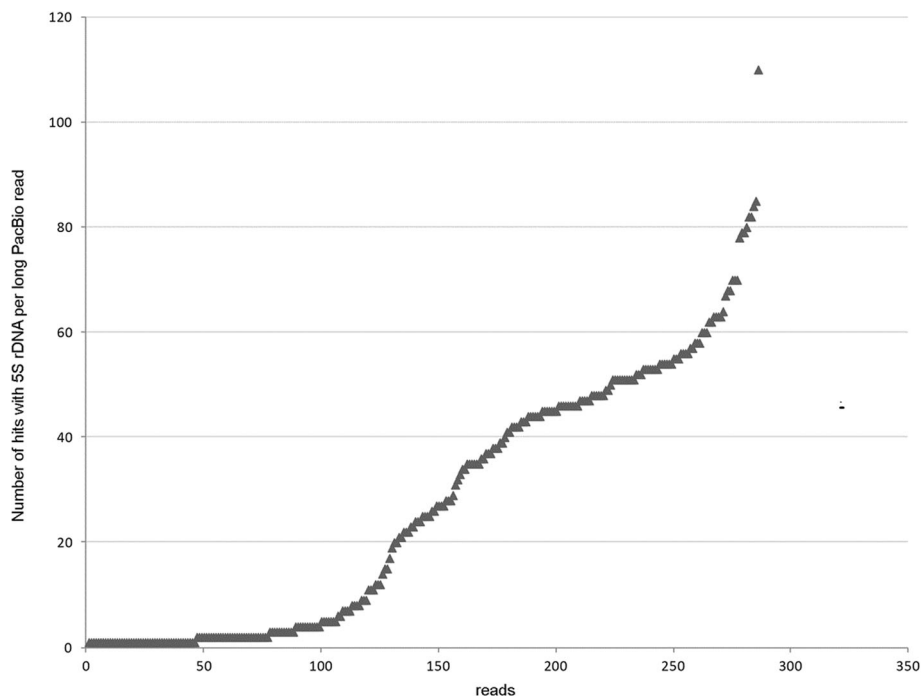


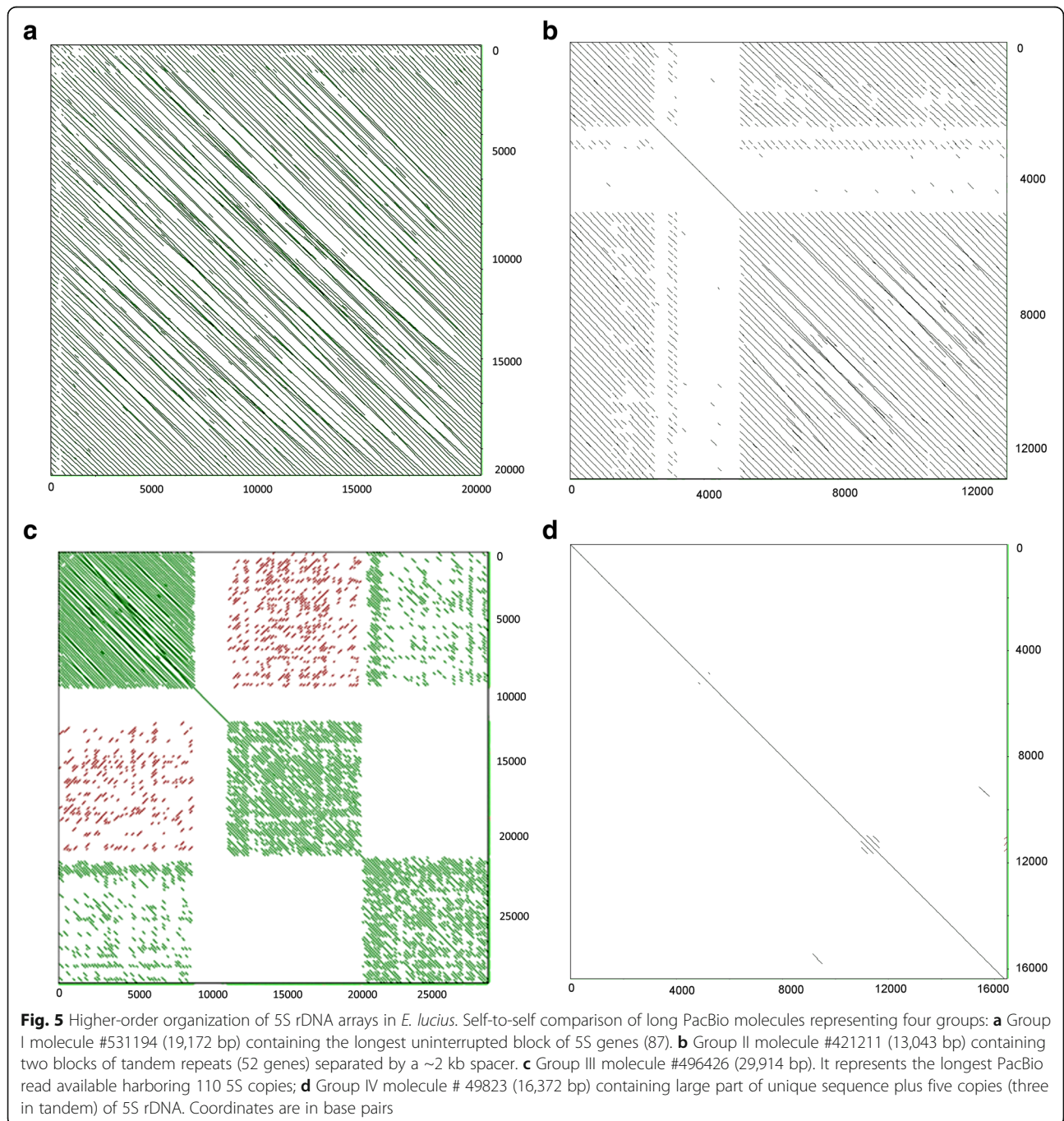
Fig. 4 5S rDNA content in long >10 kb PacBio reads of *E. lucius*. The total number of reads was 286. The sequences were queried with the 5S consensus (genic part). The MultiBlast search yielded 7768 alignments (about 43% of total 18,000 hits, Table 1) here considered as gene copies

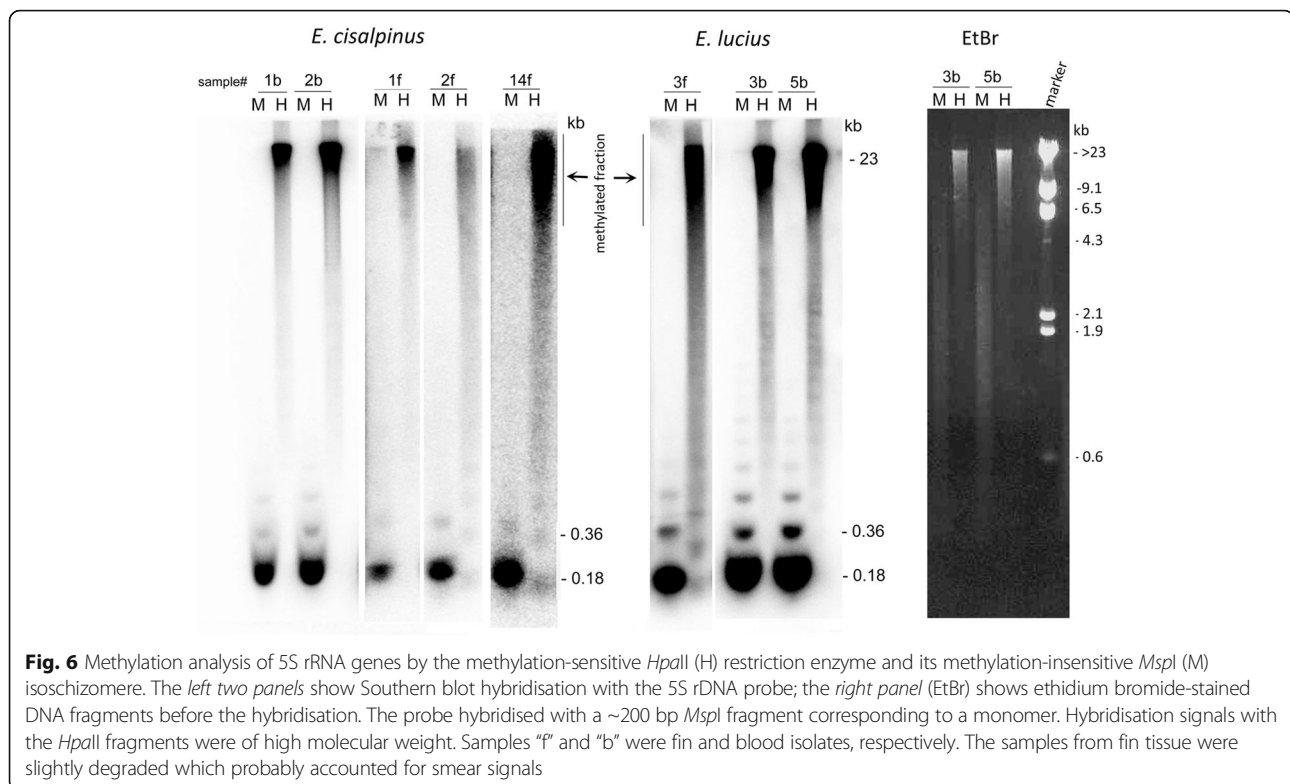
Table 2 Classification of PacBio reads according to 5S rDNA higher-order organisation

Group	5S arrangement in a read ^a	Number of reads (percentage)	Number of 5S genes (percentage) ^b
I	Tandem repeats, no unique DNA	95 (33)	4539 (58)
II	Tandem repeats + unique DNA	72 (25)	2352 (30)
III	Blocks of inverted repeats	11 (4)	639 (8)
IV	Dispersed or short (<5 units) tandems	108 (38)	274 (4)
Total		286 (100)	7769 (100)

^aThe data sets and read annotations are in Additional file 7: Table S3

^bThe number of genes in reads was determined based on MultiBlast search using the 5S genic region (NGS consensus) as a query





tissues. The globally high methylation level is evident from ethidium-bromide stained DNA fragments. Most *HpaII*-fragments migrated as high molecular weight relic while the DNA was relatively efficiently digested into shorter fragments with methylation-insensitive *MspI*.

Discussion

Massive amplification of 5S rRNA genes in *Esox lucius*

In animals, the number of rRNA genes typically reaches up to hundreds of copies [46–48]. It is therefore striking that these *Esox* species harbour tens of thousands of 5S rDNA copies. The actual copy number could be even higher since the experimental copy number estimates exceed 200 thousand copies forming about 5% of the *Esox* genomes. The apparent discrepancy (more than 10 fold) between NGS and slot blot hybridisation could be attributed to the fact that tandem repeats are typically underrepresented in the DNA sequencing libraries [49, 50]. The second explanation could be potential variation in 5S copy number in different populations of *E. lucius*. In our study, the copy number was calculated from the whole genome sequencing dataset of an American population while the experimental copy number estimation was performed in different individuals of European origin. Thus, we cannot exclude inter-population differences in copy numbers, already reported in mammals [51]. In any case, the number of 5S rRNA genes exceeding tens of thousands

of copies is far more than seen in most metazoans (<http://www.animalrDNAdatabase.com>).

The 5S rDNA copy number, sequence and position on chromosomes are similar in *E. lucius* and *E. cisalpinus*. The genomic spreading of 5S rDNA was not accompanied by any concomitant expansion of 45S rDNA whose copy number was limited to about 800 (Table 1). Independent amplification of both types of rDNA has also been reported in other fish genera [19, 21] suggesting that concerted amplification of 45S and 5S rDNA [51] may not be operating in fish or it is limited to particular groups. Related salmonid fishes harbour far more 45S than 5S loci [13–15, 36]. Thus, it is reasonable to suggest that amplification of 5S must have occurred after the divergence of a common *Esox* ancestor from the rest of salmonids.

Possible epigenetic regulation of amplified 5S rRNA genes by DNA methylation

The extremely high copy number of 5S genes in *Esox* resembles that of some amphibians [52]. However, in amphibians amplified repeats are attributed to be pseudogenes while most repeats we analysed here are probably capable to encode functional 5S rRNA. This was evidenced by their high intragenomic homogeneity, good matches between genomic and transcriptomic reads and conserved thermodynamically stable secondary structures. This raises an important question about

their transcription regulation. In mammals, only about one hundred of 45S rRNA genes are estimated to be transcribed at any one time (for review see [53]). In *Danio rerio* (zebrafish) only twelve 5S copies appear to be active in oocytes while a large number of genes in another locus are silenced and activated only at later developmental stages [54]. Perhaps, transcription activity of amplified genes in *Esox* might be regulated via epigenetic mechanisms as well. Indeed, 5S genes in both *E. lucius* and *E. cisalpinus* were heavily methylated at CG motifs suggesting that a powerful epigenetic system has evolved in this genus. There were no apparent differences in methylation levels between different tissues reported in other systems [55] suggesting that most genes are evenly methylated. However, our restriction enzyme-based methylation assay reveals global level of 5S rDNA methylation while its resolution is unable to detect changes at the single gene level. It is worth noting that sequence polymorphisms are located within the first half of the 5S gene both in *E. lucius* (Fig. 3) and *D. rerio* [54]. Therefore, developmental regulation of 5S rDNA expression cannot be excluded in *Esox*.

The 5S genes are organised in large blocks of variable sequence homogeneity

Tandem arrays of repeated sequences are generally considered as problematic regions often refractory to in depth genomic analysis. Hence, the organisation of repeat variants in the genome has been intensively discussed [47, 56, 57]. Based on conventional Southern hybridisation methods, repeat variants are believed to form separate arrays [58]. However, at the genomic scale evidence is missing due to technical difficulties related to sequencing of long and repetitive molecules. In *E. lucius*, variation in the 5S intergenic spacer was about 5-fold higher than in the genic region indicating relaxed selection constraints on intergenic spacers. We took the single cell PacBio sequencing approach to study distribution of spacer variants in tandem arrays in this species. Being aware that the randomly occurring sequencing errors are relatively high (~13% in DNA polymerase single pass) using this technology [59] making sequence polymorphisms difficult to interpret. However, sequencing errors cannot account for all the variation observed in our data sets. This is because SNPs (spacer variants) residing in the array were regularly phased (Fig. 3B) while sequencing errors are distributed randomly [59]. Secondly, the type and position of multinucleotide variation detected in long PacBio molecules were similar to those detected in high quality Illumina reads. Thus, the analysis of PacBio reads seems to be an adequate way to address the question of distribution of major variants. We found that 5S spacer variants form both separate and mixed arrays. Strict tandem arrangement of 5S units

in the heterogeneous arrays suggests that single or multiple nucleotide polymorphisms do not impair array regularity. In contrast, major spacer length variants tend to form independent arrays (Additional file 7: Table S3). Interestingly, in *Arabidopsis* where 5S loci are located on three chromosomes, a similar block-like structures composed of homo- and heterogenous arrays were detected in BAC libraries [60] and recent PacBio sequencing revealed substantial spacer polymorphisms in the 45S rDNA intergenic spacers located on two chromosomes [61]. These results suggest that the higher-order repeat structure may be evolutionary conserved. Despite the variation, evolution of 5S rRNA genes in *Esox* almost certainly fits the concept of concerted evolution considering that tens of thousands of copies are present in the genome and the SNPs being relatively infrequent. Relative heterogeneity of some arrays can be explained by reduced efficiency of interlocus recombination and/or less stringent selection constrains imposed on the spacers.

Mechanisms of arrays spreading

Accumulation of repeats at similar chromosomal positions raises questions on the mechanisms mediating spreading and homogenisation of 5S rDNA units. Several hypotheses can be drawn:

- i. Chromosome location may affect recombination rate and homogenisation of 5S arrays. In mouse, a huge copy number variability in 45S rRNA genes has been associated with their purely centromeric location in mostly acrocentric/telocentric chromosomes [62]. Thus, it has been proposed that DNA breaks may appear quite frequently in such located rDNA sequences leading to translocations of rDNA to other chromosomes. In *Esox*, 5S loci are uniformly located on short arms of nearly all acrocentric chromosomes. By analogy, in *Esox*, interlocus recombination of 5S genes could be driven by their (peri)centromeric position in acrocentric chromosomes.
- ii. 5S rRNA genes are able to multiply and integrate into other areas of the genome using a mechanism similar to retrotransposition among others. The 5S rRNA genes and SINEs harbor the same type of internal RNA polymerase III promoter [63]. On the top of that, some authors [64, 65] found a unique class of SINEs that have been formed by fusion of a 5S rRNA gene and a LINE, showing that 5S and retroelements may interact. In support, SINE elements with 5S features were reported in several fish [54, 64] and a retroelement was co-localised with 5S loci on *Erythrinus erythrinus* (red wolf fish) chromosomes [19] and recently among members of another fish genus, *Gymnotus* [66]. In the future it will be interesting to analyse inter-block spacers in

Group II reads (those containing one or few blocks of 5S genes head to tail, plus unrelated sequences) for the presence of transposable elements which may support their potential role in 5S rDNA mobilisation.

- iii. 5S genes spread through extrachromosomal replication and reintegration in new locations. Covalent extrachromosomal circles of rDNA have been identified in diverged biological taxa [67, 68]. In *Xenopus laevis* rDNA is known to replicate extrachromosomally during development [69]. In our datasets some loci (4% PacBio reads) contained invertedly repeated large blocks of 5S tandems. There is experimental evidence that extrachromosomal DNA can be generated by replication errors at the inverted repeats [70]. Therefore, one can hypothesize that the homogenisation of 5S loci across *Esox* chromosomes is mediated by the initial replication block at the inverted repeat, excision, extrachromosomal replication and reintegration by homologous or non-homologous recombination into a new genomic location. Large palindromes may thus transpose and seed 5S blocks to distal locations. Supporting this, a recent study of a human centromeric satellite (using long read PacBio sequencing) showed increased frequencies of inversions in acrocentric chromosomes compared to other chromosomes [71]. Perhaps, acrocentromeric positions of rDNA could be particularly vulnerable to such rearrangements and/or inversions. This model may also explain why interlocus homogenisation of rDNA often occurs without extensive chromosome rearrangements [30, 72].

Long term scenario of arrays evolution

- i. It can be envisaged that as long as 5S rRNA genes undergo interlocus homogenisation the number of loci would remain relatively constant. This is likely happening in both closely related species *E. lucius* and *E. cisalpinus* that show similar number of genes, sequence of the loci (including intergenic spacers) and their chromosome positions. Yet, at the cytogenetic level, small differences were noted: in contrast to *E. lucius*, *E. cisalpinus* lacked the co-localisation of 5S and 45S rDNA. Given that the 45S loci occur on homeologous chromosomes it follows that there was either 5S locus gain (*E. lucius*) or locus loss (*E. cisalpinus*) after the separation from the common ancestor. Perhaps, adjacent blocks to 45S and 5S could be represented by an unstable chromatin configuration leading to breaks and chromosomal translocation.
- ii. As methylated cytosines are more susceptible to mutations [73] we may expect gradual accumulation

of C > T and G > A mutations in the *Esox* 5S genes and their subsequent pseudogenisation. Indeed, in some plant species most rRNA genes were converted into pseudogenes [74]. However, we have no evidence for such significant erosion in *Esox* 5S rDNA despite its high methylation (methylation-induced mutations were not significantly enriched in *Esox* 5S rDNA, $p > 0.05$). Moreover, two major variants of a genic region seem to be expressed suggesting that not all mutations automatically lead to pseudogenisation. It is likely that *Esox* genomes seem to be in a dynamic phase of evolution, where most pseudogenes are being removed by genetic recombination.

- iii. Amplified 5S genes acquire centromeric function. Retrotransposons and RNA-polymerase-III-transcribed genes, including tRNA and 5S rRNA (the so called Pol III genes) have been found to be associated with centromeres in fission yeasts. Furthermore, results of some studies suggest a functional link between the centromeric localisation of the Pol III genes and chromosome condensation resulting in the proper assembly of mitotic chromosomes [75]. In the fish *H. malabaricus*, 5S genes gave rise to an independent satellite with apparently centromeric function [76]. Similar conversion might have happened in other species [77, 78] as well. Thus, if 5S rDNA is supportive to the centromeric function and plays a role in the assembly of mitotic chromosomes then its “invasions” to centromeric positions might be favored in evolution. Natural selection would select those variants that bind centromeric histones and adopt centromere-specific chromatin configuration. Relative heterogeneity of some blocks bearing degenerated units (Fig. 5c and Additional file 6: Figure S5) suggests that the process of satellite formation could have already been started. It is also possible that satellite repeats may arise from orphanised degenerated 5S insertions accounting for about 4% of *E. lucius* rDNA (Table 2).

Conclusions

In two European *Esox* species we have witnessed the extreme amplification of 5S rRNA genes reaching up to tens of thousands of copies and their distribution across more than half of the chromosomes. Such a high number is exceptional in animals, generally thought to contain moderate number of these genes. Most of the amplified genes appear to be functional and heavily epigenetically modified. Detailed analysis of long PacBio reads suggests a considerable variation in the phasing of unit variants and in the arrangement of large blocks of 5S tandem repeats. These higher-order structure polymorphisms may potentially influence the expression and homogenisation of these genes.

Additional files

Additional file 1: Figure S1. Summary on counts of FISH signals of 5S rDNA on chromosomes of two individuals of *E. cisalpinus* (Eci1 and Eci2) and five individuals of *E. lucius* two of which originated from the Czech Republic (EluCz3, EluCz5) and three from Poland (EluP8, EluP9, EluP15). (PDF 309 kb)

Additional file 2: Figure S2. Estimation of 5S rDNA copy number by slot blot hybridisation. The results are shown for two independent genomic DNA isolates and two 5S insert standards (*E. lucius* and *E. cisalpinus*). The 5S rDNAs (genic + spacer regions) account for about 5% of *E. lucius* genome equaling to about 250,000 copies (the data collected from two blot replicates and averaged). (PDF 267 kb)

Additional file 3: Figure S3. Secondary structure models of 5S rDNA molecule for fish (A) and non-fish species (B). (PDF 186 kb)

Additional file 4: Tables S1 and S2. Type, position, frequency and coverage of 5S rDNA polymorphisms. Table S1 – Analysis of genomic Illumina reads. Table S2 – Analysis of transcriptomic Illumina reads. SNPs along whole 5S units (genic and the intergenic spacer) were analyzed using genomic reads; for the transcriptome, only the genic region was considered. Note, high number of SNPs in spacer region compared to the genic region despite the shorter length (Table S1). SNP – single nucleotide polymorphism; MNP – multiple nucleotide polymorphism. (PDF 870 kb)

Additional file 5: Figure S4. Phylogenetic NJ trees constructed from 500 aligned Illumina reads derived from the 5S genic part (A) and an intergenic spacer (B). Well supported branches (bootstrap >60%) are indicated by blue arrows. (C) A sequence of the 5S clone b from *E. lucius* (GenBank KX965717.1) with highlighted subregions used in the phylogenetic analysis. (PDF 856 kb)

Additional file 6: Figure S5. Dot plot diagrams resulting from self to self comparison of long PacBio reads. (PDF 1371 kb)

Additional file 7: Table S3. Analysis of higher-order repeat structure of 5S rDNA using long (≥10 kb) PacBio reads. The selected sequences are ordered according to lengths (descending). Number of gene copies in reads was determined by MultiBlast. Arrangement was assessed by visual inspection of dot plot matrices (Additional file 6: Figure S5). Grouping followed the nomenclature in Table 2. Intergenic spacer variants: S–short (95–116 bp); L–long (321–340 bp) and UL–ultralong (1153–1209 bp). (PDF 471 kb)

Additional file 8: Figure S6. A group III molecule #499258 (19,900 bp) organised in two large immediately linked inverted blocks of tandem repeats. Green and red slanted lines indicate direct and inverted orientation of units, respectively. (B) The junction region alignment to 5S. Note, absence of any 5S-unrelated sequence between the inverted repeats. Note, a partial deletion of IGS in the third copy. (C) Nucleotide sequence of the junction region with annotated genic sequence (brown) and IGS (green). (PDF 291 kb)

Abbreviations

ΔG: Change in Gibbs free energy; FISH: Fluorescence In situ Hybridisation; FN: Fundamental number, number of chromosomal arms; GB: Gigabase; GP: Genome proportion; GS: Genome space; IE: Internal Element; IGS: Intergenic spacer; IL: Illumina sequencing; MB: Megabase; MNP: Multiple nucleotide polymorphism; NGS: Next Generation Sequencing; NJ: Neighbour Joining; NOR: Nucleolus Organizer Region; NTS: Non-Transcribed Spacer; PB: PacBio sequencing; rDNA: Ribosomal RNA genes; SNP: Single nucleotide polymorphism; WGS: Whole Genome Duplication

Acknowledgements

We thank to O. Cocco, G. Rocco and E. Sala (Associazione Pescatori Carmagnolesi) and to P. Lo Conte (Città Metropolitana di Torino, Servizio Caccia e Pesca) who provided young pike specimens from the Provincial Fish hatchery of Carmagnola.

Funding

This study was supported by a young researchers fellowship (NWF15/BIO-7) of the University of Innsbruck, Austria to RS (design of the study, sample collection, cytogenetic analysis, writing the manuscript); the Czech Science

Foundation projects P501/12/G090 to AK (design of the study, epigenetic analysis, bioinformatics procedures, writing the manuscript) and 14-02940S to RS and ŠP (sample collection, DNA isolation); a project financed by University of Warmia and Mazury in Olsztyn, Poland (No. 18.610.003-300) to KO (writing the manuscript, cytogenetic analysis) and a project from the government of Spain (CGL2016-75694-P) to SG (genomic data analysis, manuscript editing).

Availability of data and materials

All data generated or analysed during this study are included in this published article and its supplementary information files. Newly generated sequences have been submitted to GenBank under the following accession numbers: KX950799, KX965715-KX965718. Datasets used in the current study are available from the corresponding authors upon request. Chromosomal preparations and preparations after FISH experiments are deposited in laboratories of KO and ŠP.

Authors' contributions

RS and AK designed the study and performed laboratory experiments. KO and LK performed FISH with 5S rDNA and data analysis, GD provided *E. cisalpinus* samples, ŠP and RS collected biological material and prepared microscopic samples. AK, SG and RS carried out bioinformatic analysis. RS, AK, KO, GD co-drafted the manuscript, all authors read the manuscript. All authors approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval

In Poland, this study was carried out in strict accordance with the recommendations in the Polish ACT of 21 January 2005 of Animal Experiments (Dz. U. z 2005 r. Nr 33, poz. 289). The protocol was approved by the Local Ethical Committee for the Experiments on Animals of the University of Warmia and Mazury in Olsztyn, Poland (Permit Number: 11/2012). In the Czech Republic, all national guidelines for the care and use of animals were followed. The experimental procedures involving fish were approved by the Institutional Animal Care and Use Committee of the Institute of Animal Physiology and Genetics, CAS, v.v.i, according with directives from State Veterinary Administration of the Czech Republic, permit number 124/2009, and by the permit number CZ 00221 issued by Ministry of Agriculture of the Czech Republic.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Research Institute for Limnology, University of Innsbruck, Mondseestrasse 9, A-5310 Mondsee, Austria. ²Department of Marine Biology and Ecology, Faculty of Oceanography and Geography, University of Gdansk, Av. Marszałka Piłsudskiego 46, Gdynia 81-378, Poland. ³Department of Ichthyology, Faculty of Environmental Sciences, University of Warmia and Mazury, M. Oczapowskiego Str. 5, 10-718 Olsztyn, Poland. ⁴Department of Zoology, Faculty of Biology and Biotechnology, University of Warmia and Mazury, M. Oczapowskiego Str. 5, 10-718 Olsztyn, Poland. ⁵Museum of Natural History, Cascina Vigna, Via S. Francesco di Sales, 188, Carmagnola, Italy. ⁶Laboratory of Fish Genetics, Institute of Animal Physiology and Genetics, Czech Academy of Sciences, Rumburská 89, Liběchov 277 21, Czech Republic. ⁷Institut Botànic de Barcelona (IBB-CSIC-ICUB), Passeig del Migdia s/n, 08038 Barcelona, Catalonia, Spain. ⁸Laboratory of Molecular Epigenetics, Institute of Biophysics, Czech Academy of Science, Kralovopolska 135, 61265 Brno, Czech Republic.

Received: 4 November 2016 Accepted: 9 May 2017

Published online: 18 May 2017

References

- Nelson JS, Grande TC, Wilson MVH. *Fishes of the World*, 5th Ed. Hoboken, New Jersey, USA, Wiley; 2016.

2. FishBase, a global information on fishes. <http://www.fishbase.org>. Accessed 31 Mar 2017.
3. Lucentini L, Palomba A, Gigliarelli L, Sgaravizzi G, Lancioni H, Lanfaloni L, et al. Temporal changes and effective population size of an Italian isolated and supportive-breeding managed northern pike (*Esox lucius*) population. *Fish Res.* 2009;96:139–47.
4. Maes GE, VAN Houdt KJK, De Charleroy D, Volckaert FAM. Indications for a recent holarctic expansion of pike based on a preliminary study of mtDNA variation. *J Fish Biol.* 2003;63:254–9.
5. Jacobsen BH, Hansen MM, Loeschcke V. Microsatellite DNA analysis of northern pike (*Esox lucius* L.) populations: insights into the genetic structure and demographic history of a genetically depauperate species. *Biol J Linn Soc.* 2005;84:91–101.
6. Ishiguro NB, Miya M, Nishida M. Basal euteleostean relationships: a mitogenomic perspective on the phylogenetic reality of the “Protacanthopterygii”. *Mol Phylogenet Evol.* 2003;27:476–88.
7. Rab P, Flajshans M, Ludwig A, Lieckfeldt D, Ene C, Rabova M, et al. The second highest chromosome count among vertebrates is associated with extreme ploidy diversity in hybrid sturgeons. *Cytogenet Genome Res.* 2004;106:24.
8. Animal genome size database. <http://www.genomesize.com>. Accessed 4 Jan 2017.
9. Rondeau EB, Minkley DR, Leong JS, Messmer AM, Jantzen JR, von Schalburg KR, et al. The genome and linkage map of the northern pike (*Esox lucius*): conserved synteny revealed between the salmonid sister group and the Neoteleostei. *PLoS One.* 2014;9:e102089.
10. Sutherland B, Gosselin T, Normandeau E, Lamothe M, Isabel N, Audet C, et al. Novel method for comparing RADseq linkage maps reveals chromosome evolution in Salmonids. *Genome Biol Evol.* 2016;8:3600–17.
11. Rab P, Crossman EJ. Chromosomal NOR phenotypes in north-American pikes and pickerels, genus *Esox*, with notes on the Umbridae (Euteleostei, Esocae). *Can J Zool.* 1994;72:1951–6.
12. Phillips R, Rab P. Chromosome evolution in the Salmonidae (Pisces): an update. *Biol Rev Camb Philos Soc.* 2001;76:1–25.
13. Ocalewicz K, Woznicki P, Jankun M. Mapping of rRNA genes and telomeric sequences in Danube salmon (*Hucho hucho*) chromosomes using primed in situ labeling technique (PRINS). *Genetica.* 2008;134:199–203.
14. Symonova R, Majtanova Z, Sember A, Staaks GB, Bohlen J, Freyhof J, et al. Genome differentiation in a species pair of coregonine fishes: an extremely rapid speciation driven by stress-activated retrotransposons mediating extensive ribosomal DNA duplications. *BMC Evol Biol.* 2013;13:42.
15. Fujiwara A, Abe S, Yamaha E, Yamazaki F, Yoshida MC. Chromosomal localization and heterochromatin association of ribosomal RNA gene loci and silver-stained nucleolar organizer regions in salmonid fishes. *Chromosom Res.* 1998;6:463–71.
16. Pieler T, Hamm J, Roeder RG. The 5S gene internal control region is composed of three distinct sequence elements, organized as two functional domains with variable spacing. *Cell.* 1987;48:91–100.
17. Wicke S, Costa A, Munoz J, Quandt D. Restless 5S: the re-arrangement(s) and evolution of the nuclear ribosomal DNA in land plants. *Mol Phylogenet Evol.* 2011;61:321–32.
18. Vierna J, Gonzalez-Tizon AM, Martinez-Lage A. Long-term evolution of 5S ribosomal DNA seems to be driven by birth-and-death processes and selection in Ensis razor shells (Mollusca: Bivalvia). *Biochem Genet.* 2009;47:635–44.
19. Cioffi MB, Martins C, Bertollo LA. Chromosome spreading of associated transposable elements and ribosomal DNA in the fish *Erythrinus erythrinus*. Implications for genome change and karyoevolution in fish. *BMC Evol Biol.* 2010;10:271.
20. Symonova R, Matjanova Z, Korinkova T, Jankun M, Dion-Coté A-M, Bernatchez L, et al. Chromosomal characteristics of rDNA genes in salmonid fishes: trends in their patterns and evolution. In: 20th International colloquium on animal cytogenetics and gene mapping, Cordoba, Spain 2012. Cordoba: Chromosom Res; 2012. p. 810.
21. Sember A, Bohlen J, Slechtova V, Altmanova M, Symonova R, Rab P. Karyotype differentiation in 19 species of river loach fishes (Nemacheilidae, Teleostei): extensive variability associated with rDNA and heterochromatin distribution and its phylogenetic and ecological interpretation. *BMC Evol Biol.* 2015;15:251.
22. Maneechot N, Yano CF, Bertollo LA, Getlekha N, Molina WF, Ditcharoen S, et al. Genomic organization of repetitive DNAs highlights chromosomal evolution in the genus *Clarias* (Clariidae, Siluriformes). *Mol Cytogenet.* 2016;9:4.
23. Rebordinos L, Cross I, Merlo A. High evolutionary dynamism in 5S rDNA of fish: state of the art. *Cytogenet Genome Res.* 2013;141:103–13.
24. Nieto Feliner G, Rossello JA, et al. Concerted evolution of multigene families and homeologous recombination. In: *Plant Genome Diversity*, vol. 1. Wien: Springer; 2012. p. 171–94.
25. Dover GA. Molecular drive: a cohesive mode of species evolution. *Nature.* 1982;299:111–7.
26. Eickbush TH, Eickbush DG. Finely orchestrated movements: evolution of the ribosomal RNA genes. *Genetics.* 2007;175:477–85.
27. Nei M, Rooney AP. Concerted and birth-and-death evolution of multigene families. *Annu Rev Genet.* 2005;39:121–52.
28. Pinhal D, Yoshimura TS, Araki CS, Martins C. The 5S rDNA family evolves through concerted and birth-and-death evolution in fish genomes: an example from freshwater stingrays. *BMC Evol Biol.* 2011;11:151.
29. Matyasek R, Renny-Byfield S, Fulneck J, Macas J, Grandbastien MA, Nichols R, et al. Next generation sequencing analysis reveals a relationship between rDNA unit diversity and locus number in *Nicotiana diploids*. *BMC Genomics.* 2012;13:722.
30. Dubcovsky J, Dvorak J. Ribosomal RNA multigene loci—nomads of the Triticeae genomes. *Genetics.* 1995;140:1367–77.
31. Britton-Davidian J, Cazaux B, Catalan J. Chromosomal dynamics of nucleolar organizer regions (NORs) in the house mouse: micro-evolutionary insights. *Heredity.* 2012;108:68–74.
32. Raskina O, Belyayev A, Nevo E. Activity of the *Er/Spm*-like transposons in meiosis as a base for chromosome reprogramming in a small, isolated, peripheral population of *Aegilops speltoides* Tausch. *Chromosom Res.* 2004;12:153–61.
33. West C, James SA, Davey RP, Dicks J, Roberts IN. Ribosomal DNA sequence heterogeneity reflects intraspecific phylogenies and predicts genome structure in two contrasting yeast species. *Syst Biol.* 2014;63:543–54.
34. Wasko AP, Martins C, Wright JM, Galetti Jr PM. Molecular organization of 5S rDNA in fishes of the genus *Brycon*. *Genome.* 2001;44:893–902.
35. Kirtiklis L, Ocalewicz K, Wiechowska M, Boron A, Hliwa P. Molecular cytogenetic study of the European bitterling *Rhodeus amarus* (Teleostei: Cyprinidae: Acheilognathinae). *Genetica.* 2014;142:141–8.
36. Jankun M, Woznicki P, Dajnowicz G, Demska-Zakes K, Luczynski MJ, Luczynski M. Heterochromatin and NOR location in Northern Pike (*Esox lucius*). *Aquat Sci.* 1998;60:17–21.
37. Komiya H, Takemura S. Nucleotide sequence of 5S ribosomal RNA from rainbow trout (*Salmo gairdnerii*) liver. *J Biochem.* 1979;86:1067–80.
38. Koukalova B, Moraes AP, Renny-Byfield S, Matyasek R, Leitch AR, Kovarik A. Fall and rise of satellite repeats in allopolyploids of *Nicotiana* over c. 5 million years. *New Phytol.* 2010;186:148–60.
39. Goecks J, Nekrutenko A, Taylor J, Team G. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* 2010;11:R86.
40. Noe L, Kucherov G. YASS: enhancing the sensitivity of DNA similarity search. *Nucleic Acids Res.* 2005;33:W540–3.
41. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser.* 1999;41:95–8.
42. Guo X, Han F. Asymmetric epigenetic modification and elimination of rDNA sequences by polyploidization in wheat. *Plant Cell.* 2014;26:4311–27.
43. Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics.* 2003;19:2496–7.
44. Gruber AR, Lorenz R, Bernhart SH, Neubock R, Hofacker IL. The Vienna RNA websuite. *Nucleic Acids Res.* 2008;36:W70–4.
45. Miya M, Takeshima H, Endo H, Ishiguro NB, Inoue JG, Mukai T, et al. Major patterns of higher teleostean phylogenies: a new perspective based on 100 complete mitochondrial DNA sequences. *Mol Phylogenet Evol.* 2003;26:121–38.
46. Smirnov AV, Entelis NS, Krashennnikov IA, Martin R, Tarassov IA. Specific features of 5S rRNA structure—its interactions with macromolecules and possible functions. *Biochemistry (Mosc).* 2008;73:1418–37.
47. Vierna J, Wehner S, Honer zu Siederdisen C, Martinez-Lage A, Marz M. Systematic analysis and evolution of 5S ribosomal DNA in metazoans. *Heredity.* 2013;111:410–21.
48. Prokopowich CD, Gregory TR, Crease TJ. The correlation between rDNA copy number and genome size in eukaryotes. *Genome.* 2003;46:48–50.
49. Stankova H, Hastie AR, Chan S, Vrana J, Tulpova Z, Kubalaková M, et al. BioNano genome mapping of individual chromosomes supports physical mapping and sequence assembly in complex plant genomes. *Plant Biotechnol J.* 2016;14:1523–31.

50. Emadzade K, Jang TS, Macas J, Kovarik A, Novak P, Parker J, et al. Differential amplification of satellite PaB6 in chromosomally hypervariable *Prospero autumnale* complex (Hyacinthaceae). *Ann Bot.* 2014;114:1597–608.
51. Gibbons JG, Branco AT, Godinho SA, Yu S, Lemos B. Concerted copy number variation balances ribosomal DNA dosage in human and mouse genomes. *Proc Natl Acad Sci U S A.* 2015;112:2485–90.
52. Hilder VA, Dawson GA, Vlad MT. Ribosomal 5S genes in relation to C-value in amphibians. *Nucleic Acids Res.* 1983;11:2381–90.
53. Grummt I, Pikaard CS. Epigenetic silencing of RNA polymerase I transcription. *Nat Rev Mol Cell Biol.* 2003;4:641–9.
54. Locati MD, Pagano JF, Ensink WA, van Olst M, van Leeuwen S, Nehrlich U, et al. Linking maternal and somatic 5S rRNA types with different sequence-specific non-LTR retrotransposons. *RNA.* 2016;4:446–56.
55. Venney CJ, Johansson ML, Heath DD. Inbreeding effects on gene-specific DNA methylation among tissues of Chinook salmon. *Mol Ecol.* 2016;25:4521–33.
56. Plohl M, Mestrovic N, Mravinac B. Satellite DNA evolution. *Genome Dyn.* 2012;7:126–52.
57. Heslop-Harrison JS, Schwarzacher T. Organisation of the plant genome in chromosomes. *Plant J.* 2011;66:18–33.
58. Arnheim N, Treco D, Taylor B, Eicher EM. Distribution of ribosomal gene length variants among mouse chromosomes. *Proc Natl Acad Sci U S A.* 1982;79:4677–80.
59. Roberts RJ, Carneiro MO, Schatz MC. The advantages of SMRT sequencing. *Genome Biol.* 2013;14:405.
60. Cloix C, Tutois S, Mathieu O, Cuvillier C, Espagnol MC, Picard G, et al. Analysis of 5S rDNA arrays in *Arabidopsis thaliana*: physical mapping and chromosome-specific polymorphisms. *Genome Res.* 2000;10:679–90.
61. Havlova K, Dvorackova M, Peiro R, Abia D, Mozgova I, Vansacova L, et al. Variation of 45S rDNA intergenic spacers in *Arabidopsis thaliana*. *Plant Mol Biol.* 2016;92:457–71.
62. Cazaux B, Catalan J, Veyrunes F, Douzery EJ, Britton-Davidian J. Are ribosomal DNA clusters rearrangement hotspots?: a case study in the genus *Mus* (Rodentia, Muridae). *BMC Evol Biol.* 2011;11:124.
63. Paule MR, White RJ. Survey and summary: transcription by RNA polymerases I and III. *Nucleic Acids Res.* 2000;28:1283–98.
64. Kapitonov VV, Jurka J. A novel class of SINE elements derived from 5S rRNA. *Mol Biol Evol.* 2003;20:694–702.
65. Kalendar R, Tanskanen J, Chang W, Antonius K, Sela H, Peleg O, et al. Cassandra retrotransposons carry independently transcribed 5S RNA. *Proc Natl Acad Sci U S A.* 2008;105:5833–8.
66. da Silva M, Barbosa P, Artoni RF, Feldberg E. Evolutionary dynamics of 5S rDNA and recurrent association of transposable elements in electric fish of the family Gymnotidae (Gymnotiformes): The case of *Gymnotus mamiraua*. *Cytogenet Genome Res.* 2016;149:297–303.
67. Cohen S, Agmon N, Sobol O, Segal D. Extrachromosomal circles of satellite repeats and 5S ribosomal DNA in human cells. *Mob DNA.* 2010;1:11.
68. Navratilova A, Koblizkova A, Macas J. Survey of extrachromosomal circular DNA derived from plant satellite repeats. *BMC Plant Biol.* 2008;8:90.
69. Hourcade D, Dressler D, Wolfson J. The amplification of ribosomal RNA genes involves a rolling circle intermediate. *Proc Natl Acad Sci U S A.* 1973;70:2926–30.
70. Jack CV, Cruz C, Hull RM, Keller MA, Ralser M, Houseley J. Regulation of ribosomal DNA amplification by the TOR pathway. *Proc Natl Acad Sci U S A.* 2015;112:9674–9.
71. Sevim V, Bashir A, Chin CS, Miga KH. Alpha-CENTAURI: assessing novel centromeric repeat sequence variation with long read sequencing. *Bioinformatics.* 2016;32:1921–4.
72. Kovarik A, Matyasek R, Lim KY, Skalicka K, Koukalova B, Knapp S, et al. Concerted evolution of 18-5.8-26S rDNA repeats in *Nicotiana* allotetraploids. *Biol J Linn Soc.* 2004;82:615–25.
73. Vinson C, Chatterjee R. CG methylation. *Epigenomics.* 2012;4:655–63.
74. Wang W, Ma L, Becher H, Garcia S, Kovarikova A, Leitch IJ, et al. Astonishing 35S rDNA diversity in the gymnosperm species *Cycas revoluta* Thunb. *Chromosoma.* 2016;125:683–99.
75. Iwasaki O, Tanaka A, Tanizawa H, Grewal SJ, Noma K. Centromeric localization of dispersed Pol III genes in fission yeast. *Mol Biol Cell.* 2010;21:254–65.
76. Martins C, Ferreira IA, Oliveira C, Foresti F, Galetti Jr PM. A tandemly repetitive centromeric DNA sequence of the fish *Hoplias malabaricus* (Characiformes: Erythrinidae) is derived from 5S rDNA. *Genetica.* 2006;127:133–41.
77. Vittorazzi SE, Lourenco LB, Del-Grande ML, Recco-Pimentel SM. Satellite DNA Derived from 5S rDNA in *Physalaemus cuvieri* (Anura, Leiuperidae). *Cytogenet Genome Res.* 2011;134:101–7.
78. Kumke K, Macas J, Fuchs J, Altschmied L, Kour J, Dhar MK, et al. *Plantago lagopus* B Chromosome is enriched in 5S rDNA-derived satellite DNA. *Cytogenet Genome Res.* 2016;148:68–73.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

