東邦大学審査学位論文（博士）

THESIS

Molecular cytogenetic analysis of highly repetitive sequences in a Japanese hagfish, *Eptatretus burgeri*: The implication of complicated elimination system

Kohei Nagao

Supervised by Professor Souichirou Kubota

2023

Contents

Abstract

In some organisms, chromosome number and DNA amounts can differ significantly between germ and somatic cells due to the removal of specific chromosomal fragments or entire chromosomes during early embryogenesis. This process, known as programmed genome rearrangement (PGR), has been reported in a variety of organisms, including ciliates, insects, crustaceans, and vertebrates. In the Japanese hagfish, *Eptatretus burgeri*, 16 chromosomes (eliminated [E]-chromosomes) are eliminated from presumptive somatic cells (2n=36), which is equivalent to 20.9% of the nuclear DNA in germ cells (2n=52). To date, these PGR events have been observed across eight hagfish species, and 16 repetitive DNA families have been identified as the eliminated sequences from these species. At least 11 of the 16 eliminated repetitive DNA families were selectively amplified in the germline genome in *E. burgeri*. Fluorescence *in situ* hybridization (FISH) assay clearly showed that one of these eliminated DNA families, Eliminated Element of *E. burgeri* 1 (EEEb1), is exclusively located on all E-chromosomes. NGS analysis has recently facilitated the assembly of two distinct draft genomes of *E. burgeri*, derived from the testis and liver. This advancement allows for the prediction of not only nonrepetitive eliminated sequences but also over 100 repetitive and eliminated sequences, accomplished through K-mer-based analysis.

The NGS analysis has also revealed that ten repetitive DNA sequences (EEEb1–10) constitute more than 90% of the eliminated genome, thereby enabling the inference of the fundamental mechanisms and patterns of E-chromosome evolution through their utilization in cytogenetical analysis. Therefore, I first investigated their chromosomal localization using two-color FISH analysis in spermatocytes. The results demonstrated the similar localization patterns of these repetitive DNA families among all E-chromosomes, suggesting that the eight pairs of E-chromosomes are derived from a single pair of ancestral chromosomes through the multiple duplication events driven by meiotic drive over a prolonged evolutionary period.

Next, three-color FISH analysis using confocal microscopy was performed to elucidate the evolution of E-chromosomes more precisely with enhanced resolution and relative position of the ten eliminated repetitive DNA sequences. Based on the signal distribution, the E-chromosomes could be divided into six groups. To delve deeper into the local distribution of EEEb1–10 across extended chromatin fibers obtained from testicular cells, fiber-FISH experiments were subsequently conducted. The localization of EEEb1 on the extended chromatin fibers did not exhibit a continuous pattern, encompassing most of the E-chromosomes as observed in meiotic metaphases. Instead, it displayed interspersed distribution along the extended chromatin fibers. Likewise, the remaining nine eliminated DNA families were also dispersed across the EEEb1-positive chromatin fibers, intertwined with EEEb1.

Finally, the chromosomal localization of the repetitive DNA families identified in other species, namely EEEo1 and EEEo2, was explored in *E. burgeri*. Intriguingly, their FISH signals were not detected on the E-chromosomes in the meiotic metaphase of spermatocytes. Alternatively, pinpoint signals were observed within the interstitial regions of the EEEb1-negative chromosomes (non-E-chromosomes). Furthermore, in the mitotically somatic cells, neither EEEo1 nor EEEo2 signals were observed on any of the chromosomes. This thesis presents the first report providing evidence that, in addition to chromosomal elimination, internal deletion and rejoining of repetitive DNA families on non-E-chromosomes, i.e., undergoes genome rearrangement, occurred in this species.

Introduction

In most multicellular organisms, the genetic information (genome) of all the cells, originating from a fertilized egg and comprising the individual, is qualitatively and quantitatively identical throughout the process of development and differentiation. However, in some species, the genomic composition (i.e., chromosome number and DNA content) significantly varies between germ and somatic cells as a consequence of the programmed loss of the specific chromosome segments and/or the entire chromosomes during early embryogenesis. This intricate phenomenon is known as a programmed genome rearrangement (PGR). The first case of PGR was reported in the nematode, *Parascaris equorum* by Boveri (1887). In this species, the fertilized egg and blastomeres of the germline lineage have a pair of large chromosomes (2n=2), whereas these large chromosomes are fragmented and certain fragments of them are eliminated through the non-disjunction at the M phase in the presumptive somatic lineages. This phenomenon, referred to as chromatin diminution or chromosome elimination, has been reported in a variety of organisms, including ciliates, insects, crustaceans, and vertebrates (Wang and Davis 2014; Dedukh and Krasikova 2022). The eliminated chromosomes (E-chromosomes) usually contain a large number of highly repetitive sequences and exhibit the characteristics of the constitutive heterochromatin, while it has been reported that protein-coding genes that could be involved in the early stages of embryogenesis or gonadogenesis are contained on the eliminated chromosomes in songbird and sea lamprey (Biederman *et al.* 2018; Smith *et al.* 2018; Kinsella *et al.* 2019; Yasmin *et al.* 2022).

PGR in vertebrates was first discovered in the Japanese hagfish, *Eptatretus burgeri.* In spermatogonial cells, a total of 36 C-band negative and 16 C-band positive chromosomes (2n=52) were observed. However, only the 36 C-band negative chromosomes were retained in somatic cells, suggesting the selective elimination of the C-band positive chromosomes (Kohno *et al.* 1986). A quantitative analysis of nucleic DNA content in germline and somatic cells has revealed that

approximately 20.9% of germline genomic DNA is lost in somatic cells. Subsequent studies have confirmed that PGR occurs in seven hagfish species, including *E. stoutii*, *E. okinoseanus*, *E. cirrhatus*, *Paramyxine atami*, *P. sheni*, *Myxine glutinosa*, and *M. garmani*. Interestingly, the number of chromosomes to be eliminated in these species varied from two to 62, corresponding to 20.9–74.5% of the germline genomic DNA (Kohno *et al.* 1998). Moreover, a total of 16 highly repetitive DNA families have been identified as sequences that undergo selective amplification in the germline cells but are eliminated from presumptive somatic cells through molecular genetic analyses (Kubota *et al.* 1993; 1997; 2001; Goto *et al.* 1998; Kojima *et al.* 2010; Nagao *et al.* 2022). Four eliminated repetitive DNA families, EEEo1 (Eliminated Element of *E. okinoseanus* 1), EEEo2, EEPa1, and EEEb2, are conserved in the germline genomes of the studied hagfish species (Kubota *et al.* 1993; 1997; 2001; Goto *et al.* 1998; Nabeyama *et al.* 2000; Kojima *et al.* 2010).

In *E. burgeri*, out of the 16 eliminated DNA families identified in eight hagfish species, 11 families (EEEb1, EEEb2, EEEb3, EEEb4, EEEb5, EEEb6, EEEo1, EEEo2, EEPa1, EEPs1, and EEPs4) were selectively amplified in the germline genome. These families accounted for about 43.3% of the eliminated DNA, as demonstrated by Southern- and slot-blot assays (Nagao *et al.* 2022). One of the eliminated DNA families, i.e., EEEb1, is exclusively localized on all E-chromosomes (Kubota *et al.* 2001). To elucidate all the detail of an eliminated genome (E-genome) including the non-repetitive regions and protein-coding genes, the germline genome (G-genome) and somatic genome (S-genome) of *E. burgeri* have been assembled using the next-generation sequencing (NGS). Based on the comprehensive analysis using the G-genome and S-genome, the amount of E-genome is estimated to be about 320 MB. This is roughly equivalent to what it estimated previously by the comparative quantification of nuclear DNA (Kohno *et al.* 1986). Unfortunately, the assembled G-genome failed to incorporate the aforementioned repetitive DNA families due to their extensive copy number and tandem array. In contrast, an additional k-mer-based analysis using NGS reads data found over 130 novel repetitive DNA families, which are amplified more than 100-fold in the germline

genome compared to the somatic genome, in addition to the above 11 eliminated DNA families. The total number of bases of these DNA families in the G-genome is estimated to be over 300 MB, suggesting that most of the E-genome is occupied by these DNA families (Tanaka *et al.* 2023).

Since E-genomes are mostly composed of repetitive DNA sequences in most cases (Kubota *et al.* 1993; 1997; 2001; Staiber *et al.* 1997; Goto *et al.* 1998; Degtyarev *et al.* 2004; Itoh *et al.* 2009; Smith *et al.* 2009; Kojima *et al.* 2010), the repetitive sequences seem to play an important role in PGR. However, the most eliminated repetitive DNA families are divergent between species, indicating that the elimination system, specifically specific chromosome targeting mechanisms, evolved in a lineage- and species-specific manner. Hagfish are primitive jawless fish (Agnathan), the earliest-branching vertebrate group. The elucidation of the PGR mechanism of *E. burgeri* provides insight into how to evolve the primitive vertebrate genome. Hence, in this thesis, I attempted to analyze the eliminated DNA families to elucidate the evolutionary process of *E. burgeri* E-chromosomes. Additionally, I carried out the chromosomal localization analysis of eliminated repetitive sequences of *E. burgeri* to gain insight into the PGR mechanism. For this aim, I initially confirmed whether the novel repetitive families inferred by the above computer analysis are indeed present in the G-genome of *E. burgeri*. Based on the size of the repeat unit and the copy number, a total of four repetitive families were selected and were amplified by PCR using germline DNA as a template for subsequent cloning and sequencing. As expected, all of them were found to be selectively amplified in the G-genome and were designated as EEEb7 to 10. FISH analysis using EEEb1–6 (identified before) and EEEb7–10 also revealed that they are exclusively localized on the eight pairs of the E-chromosomes. From these results, I proposed the hypothesis that the eight pairs of E-chromosomes originated from a single pair of ancestral chromosomes, which have been duplicated several times by meiotic drive over a long evolutionary period. Subsequently, three-color FISH analysis with confocal microscopy was conducted to investigate the relative position of the ten DNA families among E-chromosomes more precisely. The results revealed that the eight pairs of E-

chromosomes could be classified into six groups based on the signal distribution of each DNA family. The fiber-FISH experiment revealed that the EEEb2–10 families were dispersed on the EEEb1-positive extended chromatin fiber. From these data, the chromosomal organization of eight pairs of E-chromosomes was discussed.

Apart from these, which focused on the species-specific eliminated DNA families, the final section of this thesis clarified the chromosomal localization of EEEo1 and EEEo2, which are conserved among multiple hagfish species as an eliminated DNA family, in the meiotic metaphases of the germ cells and somatic cells in *E. burgeri*. Unexpectedly, EEEo1 and EEEo2 were not detected on any of the E-chromosomes, whereas the pinpoint signals were detected on the interstitial region of the identical bivalent EEEb1-negative chromosome (non-E-chromosome) in the spermatocytes. Since neither EEEo1 nor EEEo2 signals were observed in somatic cells, this is the first report showing the internal deletion of eliminated repetitive DNA families from non-E-chromosomes in this species.

Materials and methods

Ethical statement

All animal experiments in this study were approved (Protocol#21-52-446) and conducted according to the guidelines of the Institutional Animal Care and Use Committee of Toho University.

Animals

Japanese hagfish, *Eptatretus burgeri*, were collected from Sagami Bay in Kanagawa, Japan. The animals were injected intraperitoneally with colchicine (0.375 mg/kg body weight) 2 hours before sacrifice to enrich mitotic cells. All animals were euthanized with a high dose of ethyl m-aminobenzoate methanesulfonate (MS-222) (Nacalai Tesque).

PCR amplification and molecular cloning

For PCR amplification of nine DNA families (EEEb2, EEEb3, EEEb4, EEEb5, EEEb6, EEEb7, EEEb8, EEEb9, EEEb10, EEEo1, and EEEo2), genomic DNA from germline was extracted from testes by a standard protocol using proteinase K, phenol/chloroform extraction, and ethanol precipitation as described previously (Kubota *et al.* 1993).

Primer pairs for the amplification of nine repetitive DNA families were designed by Primer3plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) with the consensus sequences of each repetitive DNA family and are summarized in Table 1. Each DNA family was amplified with 2×GoTaq® Hot Start Green Master Mix (Promega) using 0.8–2.2 ng of germline DNA and the primer pair according to the manufacturer's instructions. Amplification was performed on a T100TM thermal cycler (Bio-Rad) under the conditions listed in Table 2.

The PCR products were separated on 2.0% agarose gel for EEEb3 and EEEb7 and 3.0% agarose gel for the other families with a molecular size marker (Gene Ladder wide1; Nippon Gene).

The PCR products were purified and ligated into pMD-20 vector or pANT vector using Mighty TA-cloning Kit (Takara) or TA-Enhancer Cloning Kit (Nippon Gene) according to the protocols recommended by the respective supplier. After transformation into DH5α Competent Cells (Toyobo) and subsequent blue/white selection on ampicillin plates, plasmid DNAs were prepared from positive clones by NucleoSpin® Plasmid EasyPure (Takara). The nucleotide sequences of the inserted DNA were confirmed by Sanger sequencing (Eurofins Genomics).

Sequence analysis

All nucleotide sequences were aligned by GENETYX-MAC ver. 19.0.2, which was manually modified when necessary. Gap sites were not included in the calculation of intraspecific sequence diversity.

Two-color FISH

Chromosome slides from testes and gills were prepared according to the method described by Goto *et al.* (2016) with a slight modification of the duration of hypotonic treatment from 10 to 30 min. These slides were treated with 100 μg/mL RNase A (type I-AS; Merck) in 1×SSC for 30 min at 37°C, followed by dehydration and drying through 70% and 100% ethanol series. The plasmid DNAs, namely EEEb2_No.1, EEEb3_No.1, EEEb4_No.3, EEEb5_No.8, EEEb6_No.1, EEEb7_No.7, EEEb8_No.2, EEEb9_No.20, EEEb10_No.5, EEEo1_No.3, and EEEo2_No.5 were labeled with biotin-16-dUTP (Promo Kine), while plasmid DNA containing EEEb1 previously cloned in a study by Kubota *et al.* (2001), was labeled with digoxigenin-11-dUTP (ENZO) by nick translation as described by Green and Sambrook (2012).

After ethanol precipitation with 25 μg of yeast tRNA (Invitrogen), labeled probe DNA was thoroughly resuspended in 20 μL of formamide, and denaturation was then performed at 75°C for 10 min. The denaturation of chromosomal DNA, hybridization, washing, and detection were performed

as described by Kubota *et al.* (1993) with slight modifications. Chromosomal DNA was denatured with 70% formamide/2×SSC at 70°C for 2 min and then immediately dipped in ice-cold 70% and 100% ethanol for 5 min, respectively. Approximately 500 ng of probe DNA was applied per slide in 20 μL of hybridization mixture (2×SSC, 2 mg/mL bovine serum albumin [BSA], 10% dextran sulfate, 50% formamide). After overnight hybridization at 37°C in a dark humid chamber, the slides were extensively washed in 2×SSC/0.05% Tween 20 for 10 min, 50% formamide/0.5×SSC for 20 min, 2×SSC/0.05% Tween 20 for 20 min at 42°C, and Tris-NaCl-Tween 20 buffer (TNT) for 5 min at room temperature.

After pretreatment with TNT buffer containing 0.5% blocking solution (Merck) for 30 min at 37°C, the slides were incubated with 4 μg/mL of anti-digoxigenin fluorescence Fab fragments (Merck) and 1/1000-diluted streptavidin conjugated with DyLight™ 549 fluorescent dye (Vector Laboratories, Burlingame, CA) Tris-buffered saline containing 0.05% Tween 20 buffer (TBST) for 1 h at 37°C in a dark humid chamber. After three washes with TBST and counterstaining with 0.4 μg/mL Hoechst 33342 (Thermo Fisher Scientific) in TBST, the slides were mounted with Fluoro-Keeper Antifade Reagent (Nacalai Tesque, Kyoto, Japan). Immunofluorescence images and DNA FISH images were obtained using a Microscope Axio Imager.A2 (Carl Zeiss, Jena, Germany) with a CCD camera (Carl Zeiss) and the software program AxioVision (Carl Zeiss).

Three-color FISH

For three-color FISH, plasmid DNA containing EEEb1 was labeled with cyanine-5-dUTP by Nick Translation Mix (Roche), while plasmid DNA harboring one of the other nine DNA families (EEEb2–EEEb10) was labeled with digoxigenin-11-dUTP using DIG-Nick Translation Mix (Roche) or with biotin-16-dUTP (Promo Kine) using Nick Translation Mix (Roche). Denaturation of chromosomal DNA and hybridization were performed as described above. After hybridization, the slides were washed in 50% formamide/2×SSC for 15 min at 37°C, 2×SSC for 30 min, and TNT for

5 min at room temperature. After pretreatment with 1/2-diluted Blocking One Histo (Nacalai Tesque) in double-distilled water for 10 min at room temperature, the slides were incubated with 4 μg/ml of anti-digoxigenin Fab fragments conjugated with FITC (Merck), and 1/2,000-diluted streptavidin conjugated with alexa405 (Vector Laboratories) or 1/2,000-diluted streptavidin conjugated with cyanine-3 (Vector Laboratories) in TNT for 1 hour at room temperature in a dark humid chamber. After three times washes with TNT and counterstaining with 0.05 μg/mL propidium iodide (Fujifilm Wako) or 0.4 μg/ml Hoechst 33342 in TNT, the slides were mounted with Fluoro-KEEPER Antifade Reagent (Nacalai Tesque). Immunofluorescence images and DNA FISH images were obtained by Nikon A1R confocal microscopy controlled by NIS-Elements (Nikon) with Plan Apo ×100/1.4 NA oil immersion objective and lasers with excitation lines 405, 488, 594, and 633 nm. The distribution of the fluorescence signal was analyzed using ImageJ2 (Fiji) ver. 2.3.0/1.53q.

Construction of karyogram

The signal distribution in the XY-section was used for linescan analysis. The signal localizations of each combination were analyzed using over 30 metaphase chromosomes from at least three different samples. The E-chromosome marker, EEEb1 signal localization patterns exhibited a dichotomy. In other EEEb families, the signal distributions exhibited up to four patterns. Then, I clustered the signal distributions on individual E-chromosomes by comparing to EEEb1 distributions and to one another.

Fiber-FISH

The testis was homogenized in a microcentrifuge tube containing hagfish ringer's solution (0.5 M NaCl, 8 mM KCl, 2 mM $CaCl_2$, 2 mM $NaHCO_3$, 4 mM $MgCl_2$) on the ice. Subsequently, the tube was briefly centrifuged at 1500 rpm for a few seconds at room temperature, and the supernatant was aspirated. The cell pellet was then resuspended in a 2/3-diluted hagfish ringer's solution and

subject to centrifugation under the same conditions. This process was repeated several times, and the suspensions were subsequently stored at 4°C until the chromatin preparation. Thereafter, 15 μL of the cell suspension was spread on glass slides in a circular pattern and hemi-dried at room temperature before crystallization of the droplet edge. The slides were vertically immersed in lysis solution (25 mM Tris-HCl [pH 8.0], 1% TritonX-100, 2 M NaCl) for 5 seconds at room temperature and slowly removed from the lysis solution. The preparation was immediately fixed by immersion in 70% ethanol for 30 min at room temperature. The slides were then immersed in 100% ethanol for 5 min and air-dried. The slides were stored at –20°C until the following experiments. Probe labeling, RNase treatment, denaturation of chromosomal DNA, hybridization, washing, detection, image capture, and image analysis were carried out as described above for three-color FISH analysis.

Results

PCR amplification and molecular cloning

To determine the chromosomal localization of repetitive DNA families (EEEb1–10) that are selectively amplified in the germ cells identified in *E. burgeri* by FISH analysis, I first cloned these sequences for probe DNA preparation. In general, it is more efficient to use plasmids in which the target sequences are multiply and tandemly inserted when used for probe preparation. Although EEEb3 and 7 have repeat unit sizes greater than 300 bp, the remaining EEEb1, 2, 4, 5, 6, 8, 9, 10, EEEo1, and 2 have small repeat unit sizes (≤ 120 bp), then plasmid clones with multi-copy insertion of these were newly cloned for preparation of the probe DNA, respectively. EEEb1 had already been cloned in a previous study (Kubota *et al.* 2001) and cloned plasmid DNA having multi copies of EEEb1 was available for probe preparation. For DNA cloning, EEEb2–10 were amplified by PCR using germline DNA with each primer pair summarized in Table 1, respectively. As shown in Figure 1, the amplified DNA fragments were successfully detected at the size corresponding to the repeating units and their multimers (EEEb2: 57 bp/unit, EEEb3: 404–446 bp/unit, EEEb4: 67 bp/unit, EEEb5: 58 bp/unit, EEEb6: 56 bp/unit, EEEb7: 360 bp/unit, EEEb8: 47 bp/unit, EEEb9: 84 bp/unit, EEEb10: 120 bp/unit, EEEo1: 177 bp/unit, and EEEo2: 84 bp/unit).

After agarose gel separation and purification of the DNA band corresponding to dimer to tetramer of each repeat, all PCR products were independently cloned into cloning vectors, and selected the positive clones by insert PCR. EEEb2 was identified as eliminated repetitive DNA families by restriction enzyme analysis and Southern-blot hybridization (Kubota *et al.* 2001). EEEb2 clones were sequenced from 14 repeats from five clones, designated EEEb2_No.1, 3, 4, 19, and 22 (Figure 2a). One pair of three inverted repeats was detected in EEEb2. GC content of the consensus sequence was 26.3%. The average intraspecific divergence was 9.0%. EEEb3 was first identified as rDNA-related sequences which were selectively amplified in the germline genome (Takano 1998).

Eight sequences of EEEb3 were obtained from five positive clones (designated EEEb3_No.1, 2, 3, 4, and 5) (Figure 2b). GC content of the consensus sequence was 46.0%. EEEb3 consists of a fragment of 5S rDNA, a SINE2 family homologous region, and microsatellite-like GTA repeats. The average intraspecific divergence was 12.4%. EEEb4 to 6 were initially isolated as sequences that were preferentially transcribed and eliminated from G-genome (Otsuzumi 2009; Chinone 2010). Clones EEEb4 were sequenced from six repeats from two clones, designated EEEb4_No.3 and 5 (Figure 2c). One pair of direct and inverted repeats was detected in EEEb4. GC content of the consensus sequence was 55.2%. The average intraspecific divergence was 17.9 %. EEEb5 clone was sequenced from three repeats from one clone, designated EEEb5_No.8 (Figure 2d). One pair of sub-repeats was detected in EEEb5. GC content of the consensus sequence was 43.1%. The average intraspecific divergence was 10.4%. EEEb6 clones were sequenced from 13 repeats from five clones, designated EEEb6_No.1, 6, 14, 19, and 22 (Figure 2e). Two pairs of sub-repeats were detected in EEEb6. GC content of the consensus sequence was 37.5%. The average intraspecific divergence was 18.3%. The nucleotide sequences from EEEb2 to EEEb6 clones were identical to their consensus sequences described by Kubota *et al.* (2001) and Nagao *et al.* (2022).

EEEb7–10 were identified by comparative genome analysis between the germline and somatic genomes as mentioned above (Tanaka *et al.* 2023). EEEb7 clones were sequenced from 16 repeats from eight clones, designated EEEb7_No.1, 5, 6, 7, 10, 11, 12, and 14 (Figure 2f). Four pairs of direct repeats were detected in the EEEb7. GC content of the consensus sequence was 40.8%. The average intraspecific divergence was 4.5%. EEEb8 clones were sequenced from 40 repeats from four clones, designated EEEb8_No.2, 5, 6, and 7 (Figure 2g). GC content of the consensus sequence was 29.8%. The average intraspecific divergence was 7.8%. EEEb9 clones were sequenced from 34 repeats from six clones, designated EEEb9_No.7, 9, 12, 16, 20, and 21 (Figure 2h). Two pairs of direct repeats were detected in the EEEb9. GC content of the consensus sequence was 39.3%. The average intraspecific divergence was 18.6%. EEEb10 clones were sequenced from 12 repeats from

five clones, designated EEEb10_No.1, 5, 6, 9, and 17 (Figure 2i). Two pairs of direct repeats were detected in the EEEb10. GC content of the consensus sequence was 39.2%. The average intraspecific divergence was 2.2%.

EEEo1 and 2 were initially isolated as eliminated sequences from other hagfish species, *E. okinoseanus* (Kubota *et al.* 1997). EEEo1 clones were sequenced from ten repeats from six clones, designated EEEo1_No.3, 4, 5, 7, 8, and 10 (Figure 2j). GC content of the consensus sequence was 40.3%. The average intraspecific divergence was 7.6%. EEEo2 clones were sequenced from 16 repeats from three clones, designated EEEo2_No.1, 4, and 5 (Figure 2k). One pair of three sub-repeats was detected in EEEo2. GC content of the consensus sequence was 29.8%. The average intraspecific divergence was 12.5%. The nucleotide sequences from EEEo1 and EEEo2 clones were identical to their consensus sequences described by Kubota *et al.* (1997).

Two-color FISH analysis indicates the origin and evolution of eliminated chromosomes

Two-color FISH analysis of EEEb family

To examine the chromosomal localization of EEEb2–10, probe DNAs were prepared from the cloned plasmid DNAs mentioned above. Using EEEb1, which has been previously shown to localize to all E-chromosomes, as a control, EEEb1 and other sequences (EEEb2–10) were detected with different fluorochromes for comparative analysis of the localization of each sequence. The sets of labeled probes were then subjected to hybridization with metaphase chromosomal DNA derived from the testis and liver of *E. burgeri*. Testis-derived chromosome samples contain mitotic spermatogonia and meiotic spermatocytes, as well as mitotic somatic cells constituting the testis. Therefore, images of at least 50 metaphases of meiotic spermatocytes were used to analyze the chromosomal localization of each DNA family in each combination. On the other hand, the metaphases of spermatogonia and somatic cells were rarely observed. Hence, the FISH analysis was performed insufficiently on the spermatogonia and somatic cells (3–5 mitotic metaphases were

observed in each combination).

As shown in Figures 3 and 4, EEEb1-positive chromosomes (E-chromosomes) were usually clustered with each other in the meiotic spermatocytes, but not in the mitotic spermatogonia, as observed in previous studies (Kohno *et al.* 1986; Kubota *et al.* 2001). In EEEb2, fluorescent signals were clear on all EEEb1-positive but not EEEb1-negative chromosomes in the metaphases of the *E. burgeri* spermatocytes (Figure 3a). The colocalization analysis revealed that the EEEb2 signal was mostly located on the vicinity of EEEb1 signals or EEEb1-negative regions. Similar localization patterns were observed in the mitotic spermatogonia metaphases; the signals of EEEb2 rarely overlapped with those of EEEb1 (Figure 4a). EEEb3 also appeared to be clustered on all EEEb1-positive chromosomes in the first meiotic metaphase, and EEEb1-negative chromosomes had no EEEb3 signals (Figure 3b). The signals of EEEb3 rarely overlapped with those of EEEb1. The same results were observed in the mitotic spermatogonia metaphases; they rarely overlapped with each other on the EEEb1-positive chromosomes (Figure 4b). In the case of EEEb4, major signals were detected in the regions adjacent to the EEEb1 signals on the EEEb1-positive chromosomes in the spermatocytes, although several minor signals seemed to be colocalized with those of EEEb1. None of EEEb4 signals were located on EEEb1-negative chromosomes (Figure 3c). The signals of EEEb5 were constantly detected on all EEEb1-positive chromosomes in the spermatocytes and spermatogonia. However, the major signals were infrequently colocalized with EEEb1 signals on the E-chromosomes in both cells (Figures 3d and 4c). On the other hand, no signals of EEEb5 were observed on EEEb1-negative chromosomes (Figure 3j). The EEEb6 signals were also detected on all EEEb1-positive chromosomes in spermatocytes and spermatogonia as dense signals (Figures 3e and 4e). Approximately one-third of EEEb6 signals overlapped with EEEb1 signals in both cells. The signals of EEEb6 were absent on EEEb1-negative chromosomes.

In EEEb7, fluorescent signals were detected on all EEEb1-positive but not EEEb1-negative chromosomes in the metaphases of spermatocytes (Figure 3f). By image analysis, the majority of

EEEb7 signals colocalized with those of EEEb1, although some EEEb7 signals were partly detected in the periphery of EEEb1 signals. Similarly, the most of EEEb7 signals were observed on EEEb1-positive regions on EEEb1-positive chromosomes in spermatogonial cells (Figure 4f). In any cell lines, EEEb7 signals were never observed on EEEb1-negative chromosomes (Figure 3j). EEEb8 also appeared to cluster on all EEEb1-positive chromosomes in the first meiotic metaphase, while EEEb1-negative chromosomes exhibited no discernible EEEb8 signals (Figure 3g). The distribution and size of EEEb8 signals were entirely consistent with those of EEEb1. Furthermore, the EEEb8 signals showed complete colocalization with EEEb1 signals in spermatogonia (Figure 4g). Among the EEEb family, EEEb8 revealed the highest degree of colocalization with the EEEb1. On the other hand, no signals of EEEb8 were observed on EEEb1-negative chromosomes (Figure 3j). In the case of EEEb9, the large signals were detected in the regions adjacent to the EEEb1 signals, whereas small or faint signals colocalized with those of EEEb1 on EEEb1-positive chromosomes in spermatocytes (Figure 3h). In the spermatogonial cells, the EEEb9 signals were observed as a dot-like signal on the EEEb1-positive chromosomes (Figure 4h). No EEEb9 signals were ever observed on EEEb1-negative chromosomes. As for EEEb10, weak signals were detected on all chromosomes in both spermatocytes and spermatogonia, and the signals detected on EEEb1-positive chromosomes tended to be intense (Figures 3i and 4i). Weak signals were similarly detected on all chromosomes in somatic cells (Figure 3j).

Fine-scale chromosomal mapping of eliminated repetitive DNA families reveals chromosomal substructure of eliminated chromosomes

1. Three-color FISH analysis with confocal microscopy

To enhance the resolution of the signals, confocal laser scanning microscopy was adapted to FISH assay. I analyzed the resolution of the three-color FISH signals in the optimized pinhole state (65.13 μm) and the open state (90.68 μm) (Figure 5). The results revealed that imaging with the

optimal pinhole in confocal laser scanning microscopy resulted in a clear boundary of the signal localization. Furthermore, Z-stack imaging, acquiring continuous images by shifting the focal plane in the Z-axis, were performed to examine the signal distribution in the XZ- and YZ-section (Figure 5). As a result, the three-dimensional structure in the Z-axis was not well preserved in the Carnoy's fixed chromosomal samples and the overlap of the signals in the Z-axis plane did not achieve a higher resolution compared to the XY-section. Therefore, I analyzed the signal distribution in the XY-section using confocal laser scanning microscopy with the combination of three probes. The signal localizations of each combination were analyzed using over 30 metaphase chromosomes from at least three different samples, respectively.

As shown in Figure 6, the results were consistent with present study (Figures 3 and 4), wherein fluorescent signals for each DNA family were detected on all EEEb1-positive chromosomes in the metaphase of spermatocytes. Therefore, I performed linescan analysis to compare the distribution of EEEb2–10 on each EEEb1-positive chromosome in the spermatocytes (Figure 6; bottom). I clustered signal localization of EEEb1–10 on each E-chromosomes. The E-chromosome marker, EEEb1 signal localization patterns exhibited a dichotomy: (*i*) two clusters were symmetrically located on the terminal regions of seven pairs of E-chromosomes, and (*ii*) a single cluster was located on the terminal region of one pair of E-chromosomes. In other EEEb families, the signal distributions exhibited up to four patterns: (*i*) two clusters were symmetrically located on the terminal region of E-chromosomes, (*ii*) a single cluster was located on the middle region of E-chromosomes, (*iii*) a single cluster was located on the terminal region of E-chromosomes, and (*iv*) two clusters were located on the middle region and terminal region of E-chromosomes. As a result, I divided the signal distributions of the nine DNA families (EEEb2–10) on E-chromosomes into four groups by comparing them with the distributions of EEEb1, despite the similarity in their individual E-chromosome distributions. Specifically, (*i*) the signal clusters of EEEb8 were fully aligned with EEEb1 clusters; (*ii*) EEEb2, EEEb4, and EEEb7 largely overlapped with EEEb1 clusters; (*iii*) half of

the signals of EEEb6 and EEEb9 were overlapped with EEEb1 clusters; and (*iv*) those of EEEb3, EEEb5, and EEEb10 were predominantly located on EEEb1-negative regions and the peripheries of EEEb1-positive regions.

Based on this categorization, additional FISH experiments were conducted using combinations of each DNA family classified into the same signal distribution group with the EEEb1 probe (Figure 7). As shown in Figures 7a–7c, the signal localizations of EEEb2, EEEb4, and EEEb7 were slightly different, although all were classified into the same signal distribution group in above experiment (Figure 6). Half of the signals overlapped with each other on the EEEb1-positive regions of all E-chromosomes, while the remaining signals were exclusively detected on the EEEb1-negative regions of certain E-chromosomes. On the other hand, the signals of EEEb6 predominantly corresponded to those of EEEb9, encompassing both the EEEb1-negative and -positive regions of all E-chromosomes (Figure 7d). In the cases of EEEb3 and EEEb5, their signals exhibited scarce concordance on both the EEEb1-positive and -negative regions of some of the E-chromosomes (Figure 7e). Notably, some of the EEEb5 signals overlapped with the intense EEEb10 signals on both the EEEb1-positive and -negative regions of the E-chromosomes (Figure 7f). The EEEb3 signals tended to coincide with the intense EEEb10 signals on both the EEEb1-positive and -negative regions of E-chromosomes (Figure 7g). Further clustering analysis revealed that the signal distributions were consistent with the present study (Figure 6).

Using these hybridization signals, I clustered the signal distributions on individual E-chromosomes by comparing to EEEb1 distributions and to one another with the signal distributions sorted above. Then, I constructed a karyogram encompassing eight E-chromosomes, along with a map illustrating the locations of EEEb1–10 DNA families on individual E-chromosomes (Figure 8). These signal distribution patterns were observed over 84% metaphase I spermatocytes (Table 3). Notably, two pairs of E-chromosomes exhibited identical signal distribution patterns (patterns 1 and 2). In most cases, the signal localization of each DNA family exhibited an almost symmetrical

distribution across all E-chromosomes except for pattern 6 E-chromosomes. This remarkable symmetry observed in the karyogram strongly suggests an isochromosomal nature for patterns 1–5 E-chromosomes in this species.

2. Fiber-FISH analysis with confocal microscopy

To gain a more detailed understanding of the distribution of EEEb1–10 within E-chromosomes, fiber-FISH analysis was performed on elongated chromatin fibers obtained from testicular interphase cells, using the same probe combination as in Figure 6. In the fiber-FISH method, chromatin fibers in the interphase nuclei are elongated and spread on a glass slide, followed by the hybridization of probe DNA. Consequently, it enables visualization of the localization of each DNA with higher resolution than FISH using metaphase chromosomes. On the other hand, since it is difficult to identify whether the chromatin is from the E-chromosome or not, the EEEb1 probe was again used as the E-chromosome identification marker for all FISH here. The signal localizations of each combination were analyzed using over 70 chromatin fibers at least four different samples, respectively. First, the signals of EEEb1, which was used as an E-chromosome marker, were observed as dispersed small clusters in the extended chromatins of interphase nuclei (Figures 9a–9e), although EEEb1 localization was detected as two distinct large blocks in the metaphase chromosomes (Figures 6a–6e). These results indicate that the cluster size of EEEb1 was much smaller than expected before and was widely dispersed over the E-chromosomes, suggesting that two large blocks of EEEb1 signals on the metaphase E-chromosomes were likely a result of chromatin aggregation. In EEEb2–10, except for EEEb8, whose localization was almost identical to that of EEEb1 on the metaphase chromosomes, the signal distribution of each sequence on chromatin was not as simple as those observed in the metaphase chromosomes. Similar to EEEb1, small signal clusters of these DNA families were dispersed throughout the observed chromatins. The signals of each DNA family were observed in EEEb1-positive regions, EEEb1-negative regions, or their boundary regions on EEEb1-

positive chromatin (Figures 9a–9e). The distribution patterns were highly diverse, making it difficult to cluster them.

Based on the signal distribution of each DNA family on metaphase chromosomes, additional fiber-FISH experiments were conducted using the same probe combination as in Figure 7 (Figure 10). Consistent with the findings from the multicolor FISH on metaphase chromosomes, the signals of the nine eliminated DNA families (EEEb2 to 10) were interspersed along EEEb1-positive chromatin fibers, entwined with EEEb1 signals and with one another. In each probe combination, the specific patterns of signal distribution were not identified due to the complicated dispersion of small clusters on the extended chromatins, although they showed quite similar distributions on the metaphase chromosomes. As observed in the above experiment (Figure 9), the signals of each DNA family were observed in EEEb1-positive regions, EEEb1-negative regions, or their boundary regions on EEEb1-positive chromatin (Figures 10a–10e). These signal distribution patterns were observed on over 92% chromatin fibers in each probe combination. On the other hand, I could not carry out fiber-FISH analysis with BAC clones to estimate the resolution of the extended chromatin fiber, because despite the hagfish BAC clone library was constructed in the previous study (Pascual-Anaya *et al.* 2018), the application of this is not practical.

Two-color FISH analysis indicates internal deletion of highly repetitive DNA families from the non-eliminated chromosomes

Two-color FISH analysis of EEEo family

Chromosomal localization of the *E. burgeri*-specific repetitive DNA families was investigated in the above subsections. In this subsection, I used two evolutionary conserved repetitive DNA sequences, EEEo1 and EEEo2. EEEo1, which was preserved in the germline genome but not in the somatic genome in two Eptatretidae hagfish species (*E. okinoseanus* and *E. burgeri*), was dominantly distributed on all E-chromosomes in the metaphase I spermatocytes of *E. okinoseanus*

(Kubota *et al.* 1993). EEEo2 was conserved as an eliminated sequence in the germline genomes of five species (*E. okinoseanus*, *E. cirrhatus*, *E. burgeri*, *P. sheni*, and *P. atami*), although their chromosomal localization and amount varied. Germline selective amplification of EEEo1 and EEEo2 were previously reported in the germline genome of *E. burgeri* by Southern-blot analysis, but chromosomal localizations of those have been never determined (Nabeyama *et al.* 2000). Therefore, the chromosomal localization of EEEo1 and EEEo2 was investigated by multicolor FISH analyses with EEEb1 probes as a marker of E-chromosomes. As shown in Figures 11a to 11f, a pair of pinpoint EEEo1 or EEEo2 signals appear at the interstitial region of an EEEb1-negative bivalent chromosome in the metaphase I spermatocytes, respectively. On the other hand, no signals of EEEo1 or EEEo2 were detected in the somatic cells (Figures 11g and 11h). In the combination of EEEo1 and 2 probes, both signals were localized on the identical non-E-chromosome in 209 of 241 (86.7%) metaphase I spermatocytes (Figures 11i–11k).

Discussion

In this thesis, I clarified three major aspects regarding the localization of 10 eliminated DNA families in the entire E-chromosome of *E. burgeri* through various combination of FISH analyses. Firstly, it was found that EEEb1, which was previously thought to be a major component of the E-genome and widely distributed throughout the E-chromosome, is indeed distributed as smaller tandem clusters divided by remaining EEEb families and other sequences, rather than being uniformly distributed. Secondly, detailed localization analysis of EEEb1 to EEEb10 revealed three distinct groups: sequences that closely matched the localization of EEEb1 at the chromosome level (EEEb8), sequences mainly located in proximity to the edges of the EEEb1 cluster (EEEb2, EEEb4, EEEb6, EEEb7, and EEEb9), and sequences primarily located in regions where the localization of EEEb1 was sparse (EEEb3, EEEb5, and EEEb10). Based on these distribution patterns, the eight E-chromosomes could be classified into six patterns. Finally, I demonstrated that EEEo1 and EEEo2, which are repetitive sequences conserved in multiple species of the Eptatretidae family, exist within the somatically retained chromosomes and are eliminated through PGR involving internal deletions and chromosomal rejoining.

The evolution of E-chromosomes in *E. burgeri*

Many studies have attempted to elucidate the evolutionary process of PGR, the origin and the differentiation of E-chromosomes in animals undergoing PGR, including hagfish. Since E-chromosomes are mostly heterochromatic and composed of many repetitive DNA families (Kubota *et al.* 1993; 1997; 2001; Staiber *et al.* 1997; Goto *et al.* 1998; Degtyarev *et al.* 2004; Itoh *et al.* 2009; Smith *et al.* 2009; Kojima *et al.* 2010), it remains unclear. In previous studies, genome sequencing using NGS analysis has revealed that a significant portion of E-genome in the several species consists of various tandem repetitive sequences. However, due to the vast diversity of those repetitive

sequences, their chromosomal localization analysis has been largely overlooked (Wang *et al.* 2017; Dockendorff *et al.* 2022; Timoshevskaya *et al.* 2023). In my laboratory, genome assembly of both germ cells and somatic cells in *E. burgeri*, as well as estimation of the E-genome were performed. Based on the comparative analysis of both genomes, the 10 eliminated DNA families used in this study (EEEb1 to EEEb10) account for over 90% of the E-genome. Therefore, this study is considered the first comprehensive investigation to systematically detect the chromosomal localization of eliminated DNA sequences. The results of this study using FISH analysis showed that EEEb1 to EEEb9, but not EEEb10, were selectively detected on all E-chromosomes in the spermatocytes of *E. burgeri*, and their distribution patterns were relatively similar across all E-chromosomes (Figures 3 and 4). These results suggest that the eight pairs of E-chromosomes originated from a single pair of ancestral chromosomes and diverged through multiple duplication events caused by meiotic drive during a long evolutionary period. In support of this hypothesis, in the genus *Eptatretus*, the number of E-chromosomes varies among examined species (2n=14–62), and many of the eliminated DNA families identified previously exhibit the species-specific manner. Furthermore, supernumerary chromosomes, also known as B chromosomes (which are additional dispensable chromosomes that occur frequently among multicellular organisms) has been found in the E-chromosome in some species (Kohno *et al.* 1998). Regarding B chromosomes, it is known that they also originate from fragmented or degenerated autosomal chromosome complement in phylogenetically divergent species (Utsunomia *et al.* 2016; Silva *et al.* 2017; Hanlon *et al.* 2018; Makunin *et al.* 2018; Milani *et al.* 2018; Ruiz-Ruano *et al.* 2018; Jehangir *et al.* 2019; Dalla Benetta *et al.* 2020). PGR is thought to have been acquired in the ancestral species of Cyclostomata before the Petromyzontiformes-Myxiniformes divergence (Yan *et al.* 2016; Timoshevskiy *et al.* 2017; Smith *et al.* 2018), at least 400 million years ago (Smith *et al.* 2010). Bachmann-Waldmann *et al*. (2004) hypothesized that PGR accelerated the nucleotide divergence of the eliminated sequences. In accordance with these findings, in *E. burgeri*, the E-chromosomes seem to have diverged rapidly over a period of 400 million years

following their origination from somatic chromosomes. The original ancestral chromosome of 16 E-chromosomes thus appears to have degenerated or been replaced by other elements, since most of the eliminated DNA families, EEEb1 to EEEb9, were never detected on non-E-chromosomes. Supporting this hypothesis, a similar result was reported from one of the other hagfish species, *E. cirrhatus*; all of three *E. cirrhatus*-specific eliminated DNA families, EEEc1–3, were detected on their E-chromosomes but not on somatically retained chromosomes by FISH analysis (Goto *et al.* 1998).

In the case of chironomids and songbirds, which eliminate germline-restricted chromosomes (GRCs) during early embryogenesis (Bauer and Beermann 1952; Pigozzi and Solari 1998), microdissected GRC-specific probes were clearly detected not only on the GRC itself but also on a pair of autosomal chromosomes that are retained in the somatic cells. This finding strongly indicates that the GRCs originated from the intraspecific complement of autosomes (Itoh *et al.* 2009; Staiber 2004; Torgasheva *et al.* 2019). On the other hand, in the case of the gnat, which also undergoes PGR during early embryogenesis, different results have been reported from comparative genome analysis between related lineages. In the Sciaridae gnat, *Bradysia coprophila*, most genes found on the GRCs exhibited significant divergence from their paralogs on the autosomes, while these GRC genes showed sequence similarity to the orthologs present in the Cecidomyiidae gnat lineage. These results suggest that the GRCs of the Sciaridae gnat likely arose through introgression from the Cecidomyiidae lineage via a hybridization event during the early divergence between Sciaridae and Cecidomyiidae (Hodson *et al.* 2022; Metz 1938). In contrast to the Sciaridae gnat, the E-chromosome of *E. burgeri*, particularly EEEb1, contains components that have not been detected in other hagfish species (Kubota *et al.* 2001). Additionally, EEEb10 is shared among all E-chromosomes and the remaining chromosomes retained in somatic cells (Figures 3i and 4i). These observations suggest that the E-chromosomes of hagfish species, at least in the case of *E. burgeri*, likely originated from intraspecific autosomes (somatically retained chromosomes). This origin involves the loss and

replacement of sequences from one or more ancestral autosomes by eliminated repetitive DNA families, such as EEEb1 and EEEb7, rather than chromosomes originating from different species.

In *E. burgeri*, although the chromosomes are morphologically similar, the E-chromosomes could be divided into six groups based on the distribution patterns of the ten eliminated DNA families during meiotic metaphase-I (Figure 8). The presence of two pairs of chromosomes with pattern 1 and pattern 2 indicates that these chromosome groups have remained conserved without significant chromosomal rearrangements within each group following the chromosome duplication event. Among the examined DNA families, EEEb1 and EEEb8 in particular showed symmetric distributions in patterns 1–5 E-chromosomes, but pattern 6 did not (Figure 8). However, the degree of conservation of symmetric distribution varied among different eliminated DNA families and between groups. In Characidae fish species, B chromosomes are known to be present with intraspecific variants (Mestriner *et al.* 2000; Silva *et al.* 2014; 2016; 2017). Some of these B chromosome variants exhibit isochromosome features, characterized by the symmetrical distribution of repetitive DNA families that are found on intraspecific autosomes. Additionally, meiotic self-pairing between the two arms of the chromosome is observed. These findings indicate that the B chromosome variants with isochromosome characteristics originated from intraspecific autosomes through the incorrect division of the acrocentric chromosomes. Similarly, in *E. burgeri*, although the chromosomal arms are morphologically indistinguishable (Shichiri *et al.* 2001), it is possible that the E-chromosomes originated as isochromosomes through chromatid nondisjunction and chromosomal fusion events involving ancestral chromosomes, similar to what is observed in Characidae species.

The contradiction of genomic novelty and sequence homogenization within the E-chromosomes could be explained by birth-and-death and concerted evolution of repetitive DNA families (Nei and Rooney 2005). The birth-and-death model assumes that new sequences are generated by repeated duplication, and some duplicated sequences remain in the genome for a long evolutionary period, while others are deleted or inactivated through deleterious mutations. On the

other hand, in the concerted evolution model, repetitive sequences are assumed to evolve in a concerted manner with mutations through the entire member by repeated unequal crossover, leading to intraspecific sequence homogeneity. Further investigations of both repetitive sequences and protein-coding sequences will help us understand the evolution of E-chromosomes and PGR.

Present FISH analyses revealed that all of the eliminated DNA families were exclusively detected on either E-chromosomes (EEEb1–9) or somatically retained chromosomes (EEEb10, EEEo1 and 2) in *E. burgeri*, with the exception of EEEb10, which was detected on both germline-restricted and somatically retained chromosomes (Figures 3, 4, 6, and 7). Although the genomic contents were conserved among E-chromosomes, indicating the presence of an autonomous mechanism such as recombination-mediated DNA repair system to maintain the repetitive sequence array, these findings imply limited homogenization of the genomic contents between the germline-restricted and somatically retained chromosomes in this species. In bird species, their karyotypes usually have large numbers of diploid chromosomes, typically around 80, and consist of morphologically distinguishable macrochromosomes and microchromosomes. Centromeric repetitive sequences are preferentially amplified in either several macrochromosomes or all microchromosomes (Chen *et al.* 1989; Yamada *et al.* 2002; Shang *et al.* 2010; Ishishita *et al.* 2014). These findings indicate that centromeric repetitive sequences are homogenized within macrochromosomes or microchromosomes, with the homogenization process being more pronounced among microchromosomes. In chicken interphase nuclei, macrochromosomes and microchromosomes show different spatial organizations: the former cluster in the nuclear center while the latter are distributed at the nuclear periphery (Habermann *et al.* 2001; Tanabe *et al.* 2002). Considering these findings, it is evident that DNA transposition (homogenization) through unequal crossing over frequently takes place in the centromeric regions between physically close, nonhomologous macro- or microchromosomes. Similarly, in the interphase nuclei of *E. burgeri* testicular cells, EEEb1-positive chromatins were observed as several large blocks, implying the

clustering in the certain chromatin territory (Chisuwa 2017). Moreover, the bivalent E-chromosomes tended to assemble each other in the mitotic metaphase nuclei (Kohno *et al.* 1986). This may explain the differential conservation of genomic contents between E-chromosomes and somatically retained chromosomes in this species.

The genomic organization of EEEb family at the chromatin level

Fiber-FISH analysis of extended chromatin fibers from germ cells of *E. burgeri* revealed that the signal localizations of the ten eliminated repetitive DNA families were intermingled with each other, indicating the absence of clear boundaries between clusters of different DNA families at the chromatin level (Figures 9 and 10). To the best of our knowledge, fiber-FISH analysis on animals exhibiting PGR has not been previously conducted. However, the intermingled arrangement of nonhomologous repetitive DNA families has been observed in diverse taxa (Alkhimova *et al.* 2004; Kuhn *et al.* 2008; 2009; de Barros *et al.* 2011; Paco *et al.* 2015; Milani *et al.* 2017; Viana *et al.* 2017; Souza *et al.* 2021). For instance, in the grasshopper species *Abracris flavolineata*, FISH analysis showed that two distinct DNA families, namely AflaSAT-1 and AflaSAT-2, were colocalized on the large pericentromeric region of all autosomal chromosomes and the small centromeric region of B chromosomes. Subsequent fiber-FISH analysis also demonstrated that these two DNA families were intermingled with each other on the extended chromatin fibers (Milani *et al.* 2017). Although fiber-FISH analysis with BAC clones to estimate the resolution of extended chromatin fiber were not performed due to the unavailability of hagfish BAC clones in this study, other previous studies showed that the DNA fiber can be extended approximately 2.5–3.5 kb/μm on slides (Jackson *et al.* 1998; Fransz *et al.* 1996; Jackson *et al.* 1999; Cheng *et al.* 2002; Wang *et al.* 2013). This resolution seems to be influenced by the physical status of chromosomes or chromatin. For example, euchromatin could be extended greater than heterochromatin since the former is more physically relaxed than the later. These findings indicate that, although tandemly repetitive DNA families appear

as a single visible FISH signal cluster on metaphase chromosomes, the clusters of different DNA families are intermixed and nested within each other at the chromatin level, rather than a single DNA family being exclusively repeated in a tandem array. However, the specific boundary sequences between DNA families remain unclear. Hence, further investigation of both the genomic contents and chromosomal organization will provide insights into the evolution of E-chromosomes and PGR.

The complicated elimination system in *E. burgeri*

FISH analysis with EEEo1 and EEEo2 probes clearly demonstrated the internal deletion of eliminated repetitive DNA families from non-E-chromosomes in *E. burgeri* (Figure 11). About mechanism of PGR, the absence of centromeric histone or centromere protein has been observed in internal chromosomal regions or entire chromosomes that are eliminated prior to PGR in other animals. In the parasitic nematode *Ascaris suum*, a pair of large holocentric chromosomes (2n=2) is retained in the fertilized egg and in the presumptive germ cells. During germline mitoses, centromeric histone CENP-A is distributed along the entire length of these chromosomes. Prior to PGR in the 4-cell embryo, CENP-A is significantly diminished in the middle regions of the chromosomes to be eliminated. This leads to the absence of kinetochores and attachment sites necessary for chromosome segregation, resulting in the elimination of internal chromosomal regions during mitosis (Kang *et al.* 2016). In contrast, in passerine birds, the chromosomes are present as monocentric chromosomes, and the elimination of entire chromosomes was observed during early spermatogenesis (Pigozzi and Solari 1998; Malinovskaya *et al.* 2020). One of these passerine birds, *Riparia diluta*, showed that centromere protein was located not only on non-eliminated chromosomes as centromere-specific signal but also on GRC as non-centromere-specific diffused signals in spermatocytes prior to PGR. This indicates that the absence of centromeric protein also leads to the elimination of entire chromosomes (Malinovskaya *et al.* 2020). An ultrastructural analysis of mitotic and meiotic chromosomes has been performed on *E. burgeri* using electron microscopy. Both types of

chromosomes have diffused kinetochores over 12–50% of the chromosome length without any constriction (Shichiri *et al.* 2001). These findings raise the possibility that such epigenetic changes lead to elimination of both internal chromosomal regions and entire chromosomes via PGR in *E. burgeri*, as observed in nematodes and passerine bird.

Other possibilities for elimination mechanisms have been suggested by several previous studies. Terminal chromosomal deletion followed by *de novo* telomere addition has been observed in certain animals (Kirk and Blackburn 1995; Kojima *et al.* 2010; Dockendorff *et al.* 2022). In the free-living nematode *Oscheius tipulae*, which eliminates all chromosomal ends (2n=10) by excision of specific sequence motifs in presumptive somatic cells during the 8- to 16-cell embryo stage, END-seq analysis that identifies DNA double strand breaks (DSB) and the DSB end resection revealed that not only canonical excised sites that define the somatic telomere ends but also additional and alternative excised sites within eliminated regions in the wild-type and excised sequence mutant. This indicates that the additional and alternative excised sites provide a fail-safe mechanism for PGR. Moreover, this analysis also revealed that the excised sites differ among 23 geographically isolated wild populations (Dockendorff *et al.* 2022). Changing telomere structure followed by terminal chromosomal deletion has also been observed in the ciliate *Tetrahymena*, unicellular protozoa organisms. Within a single cell, two functionally distinct nuclei exist: the inactive, diploid micronucleus (MIC; germline nucleus) and the active, polyploid macronucleus (MAC; somatic nucleus). During sexual reproduction, the MAC is derived from the MIC through PGR (Gorovsky 1973). During MAC development, piggyBac transposon-derived genes are involved in the recognition of the position of internal deletion and rejoining (Noto and Mochizuki 2017). While the MAC telomeres are composed of 5′-TTGGGG-3′ (T2G4) repeats, the MIC telomeres contain inner 5′-TTTGGGG-3′ (T3G4) repeats adjacent to the distal (T2G4) repeats (Kirk and Blackburn 1995). Genome sequencing of MIC and MAC revealed that the MIC telomeres are removed during MAC development through PGR and replaced by the MAC telomeres through *de novo* telomere addition

(Hamilton *et al.* 2016). Although no domesticated transposases or DNA motifs have been identified in *E. burgeri*, a similar mechanism may be involved in internal chromosomal elimination. Additionally, these observations in nematode and ciliate suggest that excised sequences may differ among populations and that telomere structures may differ between germline and somatic cells in *E. burgeri*. Further investigation provides new insights to understand the mechanisms involved in the PGR.

The evolution of EEEo family in Eptatretidae hagfish species

As mentioned before, EEEo1, which was preserved in the germline genome but not in the somatic genome in two Eptatretidae hagfish species (*E. okinoseanus* and *E. burgeri*), was dominantly distributed on all E-chromosomes in the metaphase I spermatocytes of *E. okinoseanus* (Kubota *et al.* 1993; Kohno *et al.* 1998). EEEo2 was also conserved as an eliminated sequence in the germline genomes of five species (*E. okinoseanus*, *E. cirrhatus*, *E. burgeri*, *P. sheni*, and *P. atami*; (Nabeyama *et al.* 2000), although their chromosomal localization and amount varied, as shown in Figure 12. In three of the five species (*E. okinoseanus*, *E. cirrhatus*, and *P. atami*), EEEo2 was exclusively distributed on all or some E-chromosomes. In contrast, it was detected only at both peripheral regions of the 34 non-E-chromosomes of the mitotic spermatocytes in *P. sheni* and was completely absent in somatic cells, suggesting the excision of the terminal regions of the non-E-chromosomes containing EEEo2. Terminal excision of the non-E-chromosome was also observed in *E. okinoseanus* type B (Figure 12) and *E. stoutii* (Kohno *et al.* 1998). Therefore, our results strongly demonstrated that this was the first case of the elimination of internal chromosomal regions followed by rejoining of the boundaries on non-E-chromosomes during chromosome elimination in hagfish. The variation of the distribution and their copies of EEEo1 and EEEo2 between *E. burgeri* and other Eptatretidae hagfish species are well explained by the library hypothesis (Ugarkovic and Plohl 2002). It predicts that the related species share a common ancestral set of conserved repetitive DNA families, each of which

may be differentially amplified in each species. In previous analysis, EEEo2 family was rarely amplified in euchromatic eliminated region in *P. sheni* and *E. burgeri*, whereas it was highly amplified in heterochromatic E-chromosomes in *E. okinoseanus*, *E. cirrhatus*, and *P. atami* by a concerted evolution (Nabeyama *et al.* 2000). In some species, chromosomal position and copy number of EEEo1 appeared to coordinate with those of EEEo2, but not others (Kubota *et al.* 1993). Thus, the repetitive families previously found in eight hagfish may be shared as an original set in ancestral hagfish, and the EEEo1 and EEEo2 families appeared to be sympatrically amplified on a pair of non-E-chromosome in *E. burgeri*. In future studies, it will be necessary to elucidate how to precisely eliminate internal chromosomal regions during chromosome elimination.

Table 1. PCR primers used in this study

Primer name	Primer sequence (5′-3′)	Amplified sequence
EEEb2_Primer_F	AAAGTTTGAGTGGTGTAAGAAG	EEEb2 (57 bp/unit)
EEEb2_Primer_R	AAAATAAACATCTCACCTCAAAA	
EEEb3_Primer_F	GAGACTGCCTGGGAATACC	EEEb3 (428 bp/unit)
EEEb3_Primer_R	GTCTCCCATCAAGTACTAACC	
EEEb4_Primer_F	CAACATGATTGGTCCA	EEEb4 (67 bp/unit)
EEEb4_Primer_R	GGGTGATTTACCTCGA	
EEEb5_Primer_F	GGCTCATAGAAGGAAGAAGC	EEEb5 (58 bp/unit)
EEEb5_Primer_R	CAACATCGCCACCTACATG	
EEEb6_Primer_F	CTCAGTAAGAAACGTCG	EEEb6 (56 bp/unit)
EEEb6_Primer_R	GTTAAACCACGCCCAC	
EEEb7_Primer_F	CCGCCAAATGTGTTAGGAATG	EEEb7 (360 bp/unit)
EEEb7_Primer_R	CCTGTTGAAAGAGGCCAAAAAC	
EEEb8_Primer_F	AAAAAAAGGGATTATTGTATATTTTG	EEEb8 (47 bp/unit)
EEEb8_Primer_R	CGGTTCGGAATTTTCCAC	
EEEb9_Primer_F	AAACCCTACAATTGTTCTG	EEEb9 (84 bp/unit)
EEEb9_Primer_R	CTCTCCGGTGTGTATTC	
EEEb10_Primer_F	GTTCTAAATATGGGGCCTACCTT	EEEb10 (120 bp/unit)
EEEb10_Primer_R	CATCTTGGCAGAACCCTTTTC	
EEEo1_Primer_F	GTKKWCGGATTCKGA	EEEo1 (177 bp/unit)
EEEo1_Primer_R	AACTTTTTGAGATATCHKG	
EEEo2_Primer_F	CTTTTTTGGTGACCAGAAGTG	EEEo2 (84 bp/unit)
EEEo2_Primer_R	AGGAAAAATCACTTCCAGAAAAAG	

Table 2. Condition of the PCR amplification

Amplified sequence	Preheat	35 Cycles			Final extension
		Denaturation	Annealing	Extension	
EEEb2	95°C, 2 min	95°C, 30 sec	43°C, 30 sec	72°C, 15 sec	72°C, 5 min
EEEb3	95°C, 2 min	95°C, 30 sec	60°C, 30 sec	72°C, 1 min	72°C, 5 min
EEEb4	95°C, 2 min	95°C, 30 sec	50°C, 30 sec	72°C, 15 sec	72°C, 5 min
EEEb5	95°C, 2 min	95°C, 30 sec	66°C, 30 sec	72°C, 15 sec	72°C, 5 min
EEEb6	95°C, 2 min	95°C, 30 sec	50°C, 30 sec	72°C, 15 sec	72°C, 5 min
EEEb7	95°C, 2 min	95°C, 30 sec	53°C, 30 sec	72°C, 1 min	72°C, 5 min
EEEb8	95°C, 2 min	95°C, 30 sec	57°C, 30 sec	72°C, 30 sec	72°C, 5 min
EEEb9	95°C, 2 min	95°C, 30 sec	55°C, 30 sec	72°C, 15 sec	72°C, 5 min
EEEb10	95°C, 2 min	95°C, 30 sec	57°C, 30 sec	72°C, 30 sec	72°C, 5 min
EEEo1	95°C, 2 min	95°C, 30 sec	53°C, 30 sec	72°C, 1 min	72°C, 5 min
EEEo2	95°C, 2 min	95°C, 30 sec	58°C, 30 sec	72°C, 1 min	72°C, 5 min

Table 3. Percentages of pettern-fitted nucleus

Probe combination	Pattern-fitted nucleus percentages (pattern-fitted nucleus/analyzed nucleus)
EEEb1+EEEb2+EEEb3	86.7% (26/30)
EEEb1+EEEb4+EEEb5	85.0% (34/40)
EEEb1+EEEb6+EEEb7	85.7% (30/35)
EEEb1+EEEb8+EEEb9	85.3% (29/34)
EEEb1+EEEb2+EEEb10	87.5% (35/40)
EEEb1+EEEb2+EEEb4	87.2% (35/40)
EEEb1+EEEb4+EEEb7	93.5% (29/31)
EEEb1+EEEb2+EEEb7	87.5% (28/32)
EEEb1+EEEb6+EEEb9	93.9% (31/33)
EEEb1+EEEb3+EEEb5	93.9% (31/33)
EEEb1+EEEb5+EEEb10	93.5% (29/31)
EEEb1+EEEb3+EEEb10	84.3% (27/32)

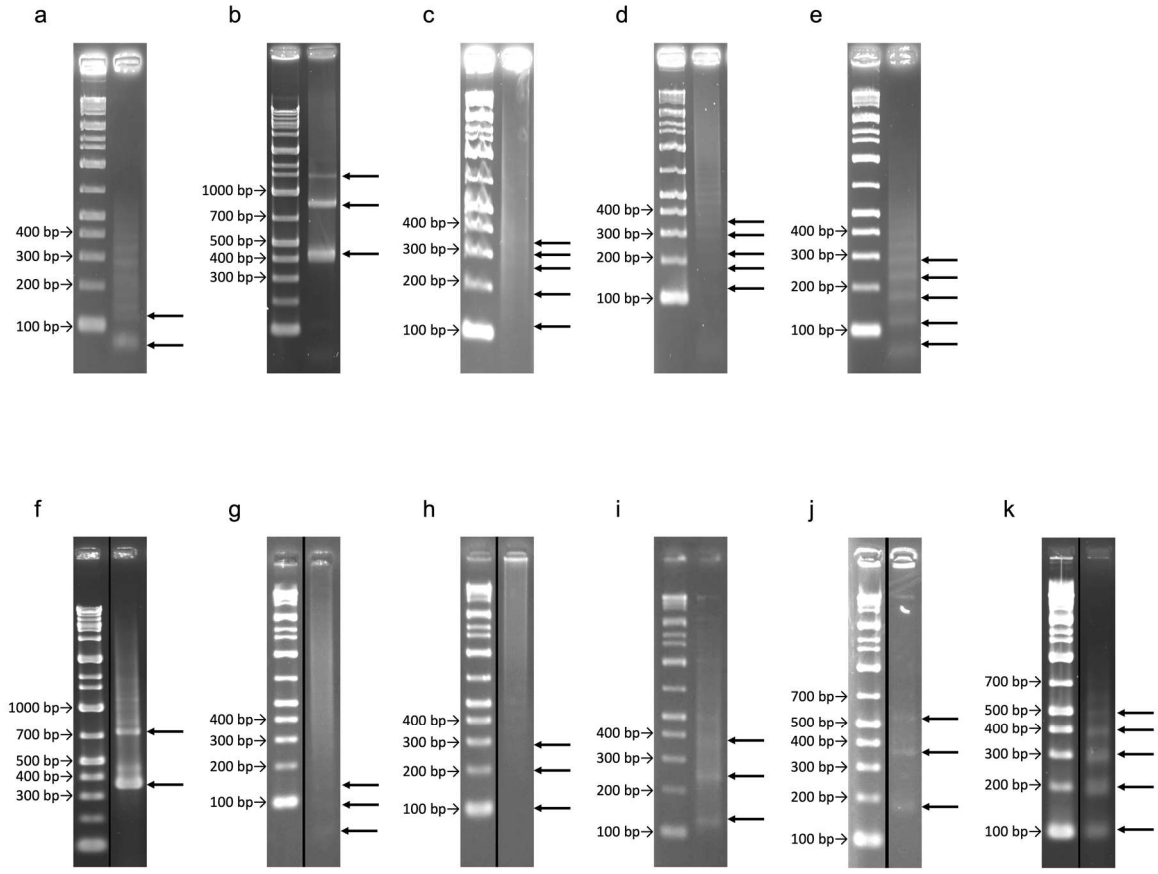
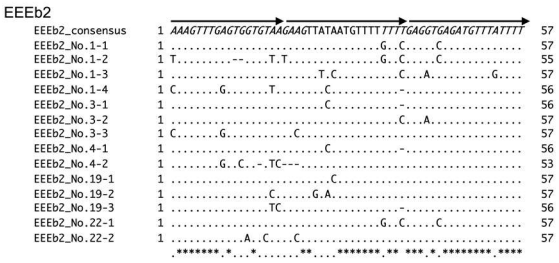


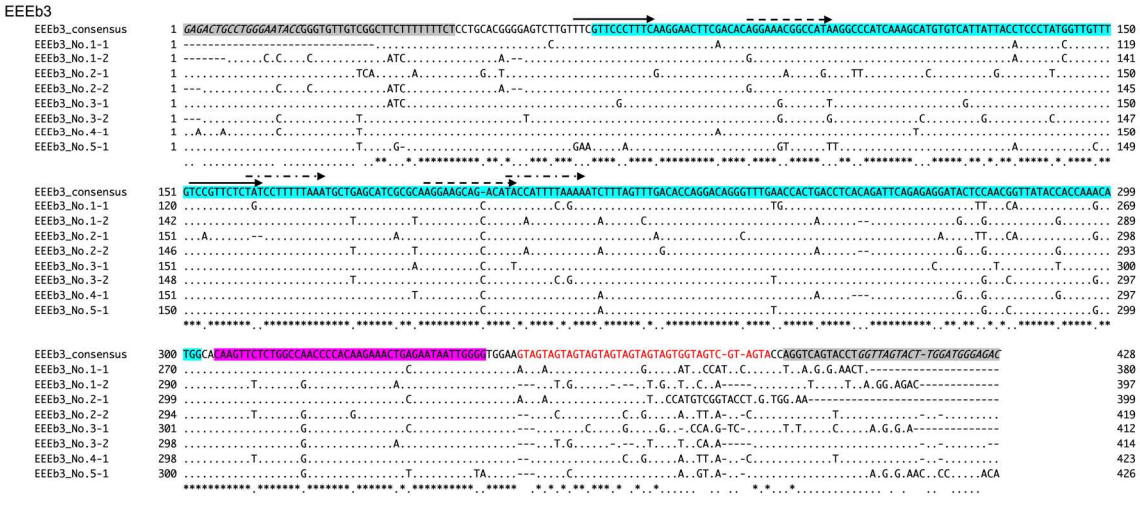
Figure 1. Amplification of 11 repetitive DNA families.

PCR amplification was carried out using germline of *E. burgeri* as templates. The left lane in the photograph contains DNA molecular size marker (Gene Ladder wide1). The arrows indicate the corresponding size of monomeric and multimeric EEEb2 (a), EEEb3 (b), EEEb4 (c), EEEb5 (d), EEEb6 (e), EEEb7 (f), EEEb8 (g), EEEb9 (h), EEEb10 (i), EEEo1 (j), and EEEo2 (k).

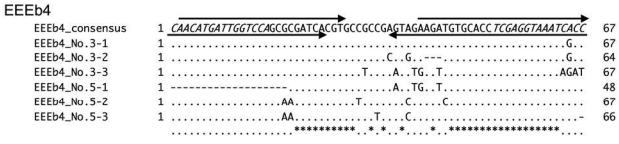
a


EEEb2
EEEb2_consensus
EEEb2_No.1-1
EEEb2_No.1-2
EEEb2_No.1-3
EEEb2_No.1-4
EEEb2_No.3-1
EEEb2_No.3-2
EEEb2_No.3-3
EEEb2_No.4-1
EEEb2_No.4-2
EEEb2_No.19-1
EEEb2_No.19-2
EEEb2_No.19-3
EEEb2_No.22-1
EEEb2_No.22-2

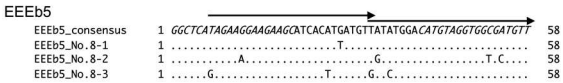
b


EEEb3
EEEb3_consensus
EEEb3_No.1-1
EEEb3_No.1-2
EEEb3_No.2-1
EEEb3_No.2-2
EEEb3_No.3-1
EEEb3_No.3-2
EEEb3_No.4-1
EEEb3_No.5-1

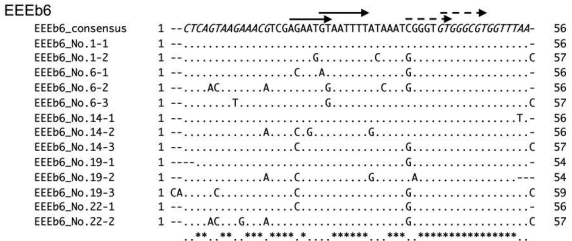
c


EEEb4
EEEb4_consensus
EEEb4_No.3-1
EEEb4_No.3-2
EEEb4_No.3-3
EEEb4_No.5-1
EEEb4_No.5-2
EEEb4_No.5-3

d


EEEb5
EEEb5_consensus
EEEb5_No.8-1
EEEb5_No.8-2
EEEb5_No.8-3

e


EEEb6
EEEb6_consensus
EEEb6_No.1-1
EEEb6_No.1-2
EEEb6_No.6-1
EEEb6_No.6-2
EEEb6_No.6-3
EEEb6_No.14-1
EEEb6_No.14-2
EEEb6_No.14-3
EEEb6_No.19-1
EEEb6_No.19-2
EEEb6_No.19-3
EEEb6_No.22-1
EEEb6_No.22-2

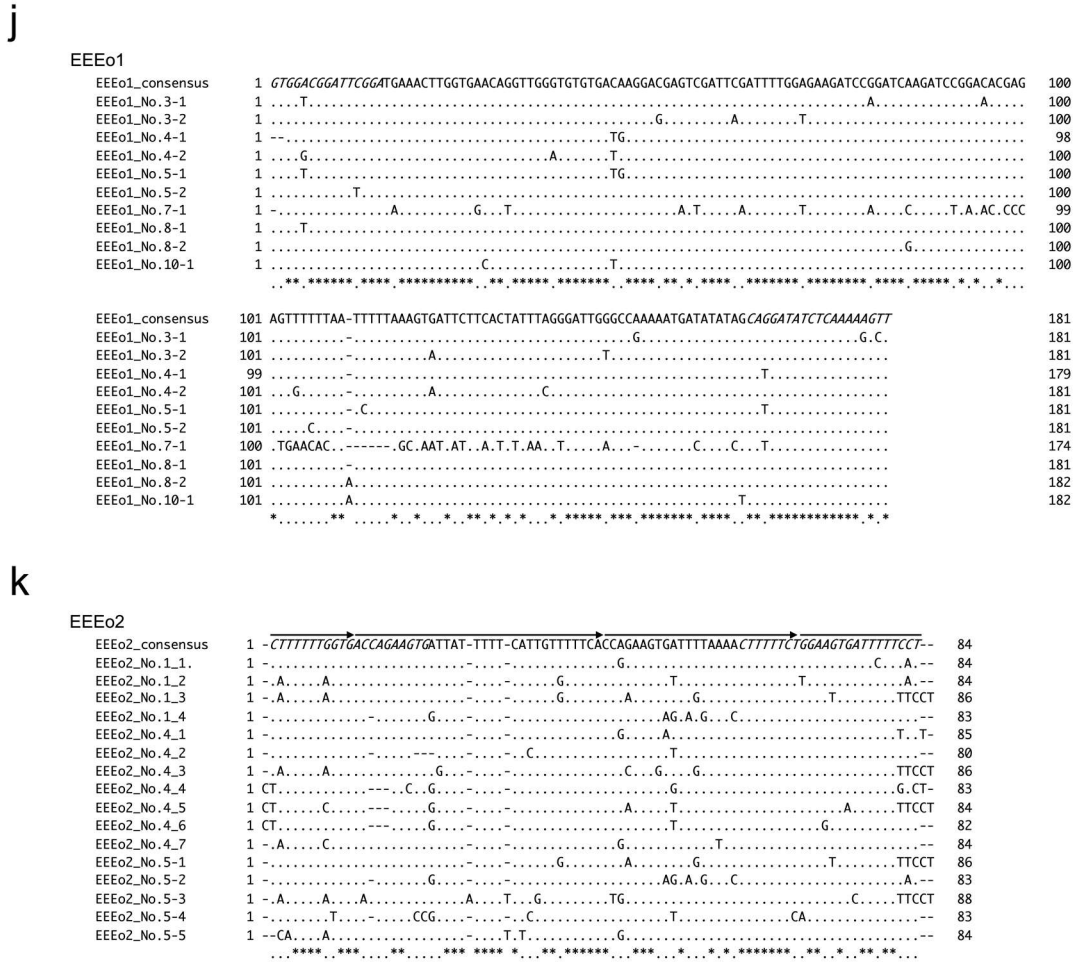


Figure 2. Cloned sequences of nine repetitive DNA families.

Each consensus nucleotide sequences (top) are aligned with the inserts of EEEb2 (a), EEEb3 (b), EEEb4 (c), EEEb5 (d), EEEb6 (e), EEEb7 (f), EEEb8 (g), EEEb9 (h), EEEb10 (i), EEEo1 (j), and EEEEo2 (k) (bottom). Dots (.) represent nucleotides identical to those in the consensus sequence at the top. Base substitutions are indicated by the respective bases. Dashes (-) represent alignment gaps. Primer regions were marked with italics letters. The upper arrows indicate the direct repeats, while the lower arrows indicate the inverted repeats. In (b), gray shading: 5S rDNA, light-blue: SINE2-5_EBu, and purple: SINE2-6_EBu homologous regions. The GTA repeats detected are shown in red letters.

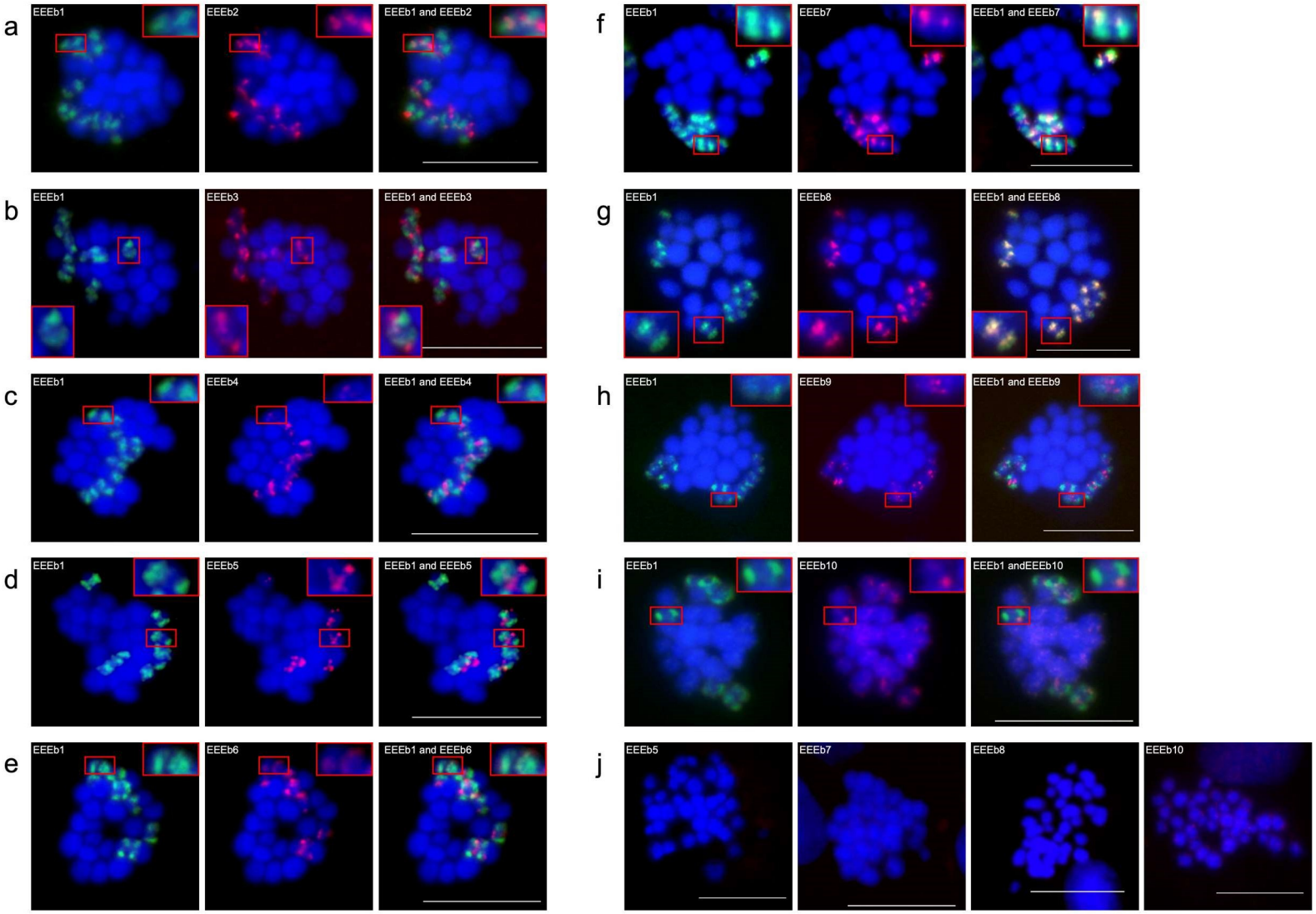


Figure 3. Chromosomal mapping of ten eliminated DNA families in *E. burgeri*.

Metaphase chromosomes from spermatocytes were hybridized using a digoxigenin-labeled EEEb1 probe (green) with biotin-labeled EEEb2 (a), EEEb3 (b), EEEb4 (c), EEEb5 (d), EEEb6 (e), EEEb7(f), EEEb8 (g), EEEb9 (h), and EEEb10 (i) (red). The chromosomal localization of EEEb5, EEEb7, EEEb8, and EEEb10 in somatic cells were shown at (j). Chromosomes were counterstained with Hoechst 33342 (blue). Insets show magnified images of E-chromosomes with each signal. Scale bar = 5 μm.

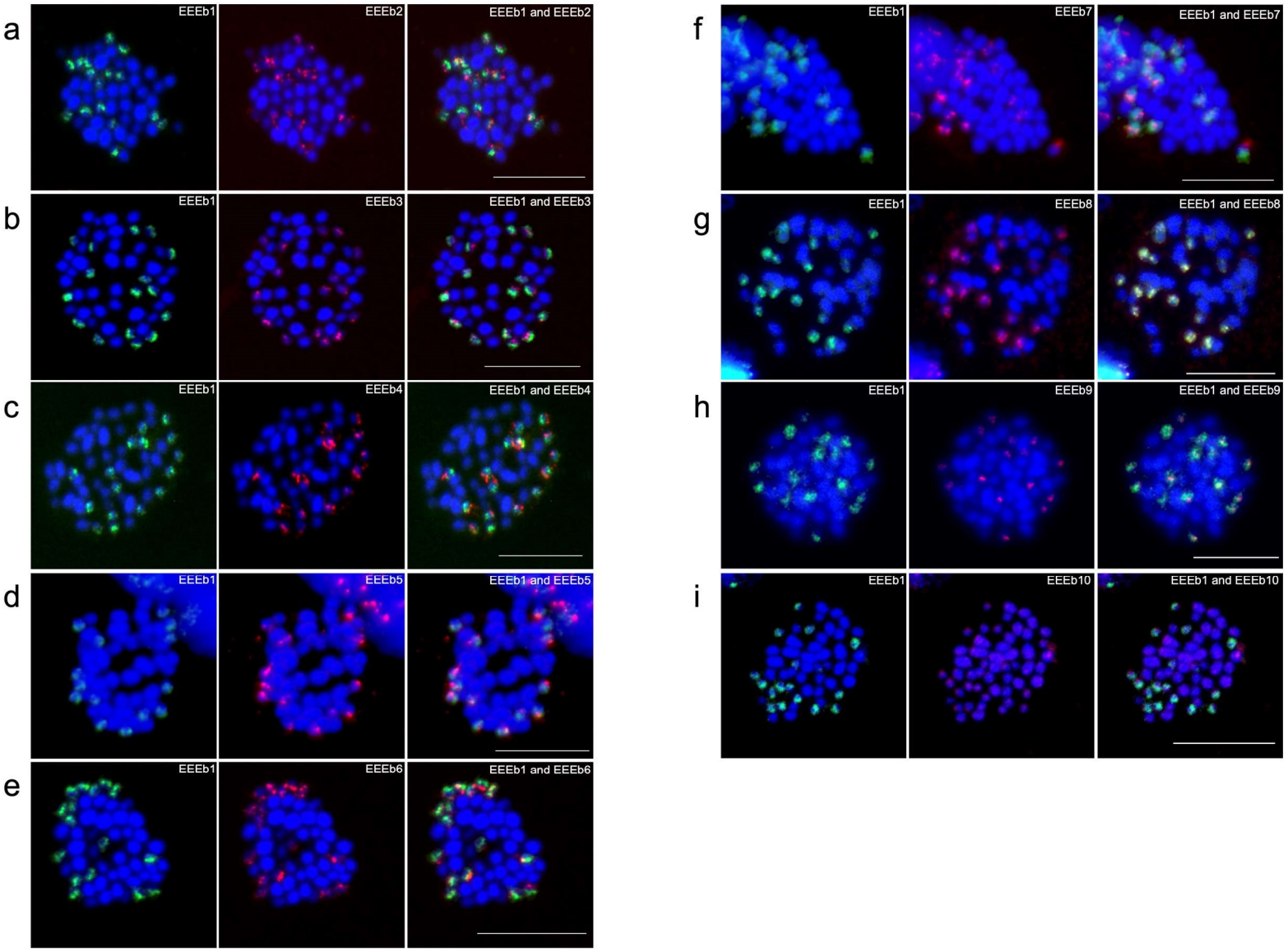


Figure 4. Colocalization analysis of ten eliminated DNA families in *E. burgeri*.

Metaphase chromosomes from spermatogonia were hybridized using a digoxigenin-labeled EEEb1 probe (green) with biotin-labeled EEEb2 (a), EEEb3 (b), EEEb4 (c), EEEb5 (d), EEEb6 (e), EEEb7(f), EEEb8 (g), EEEb9 (h), and EEEb10 (i) (red). Chromosomes were counterstained with Hoechst 33342 (blue). Insets show magnified images of E-chromosomes with each signal. Scale bar = 5 μm.

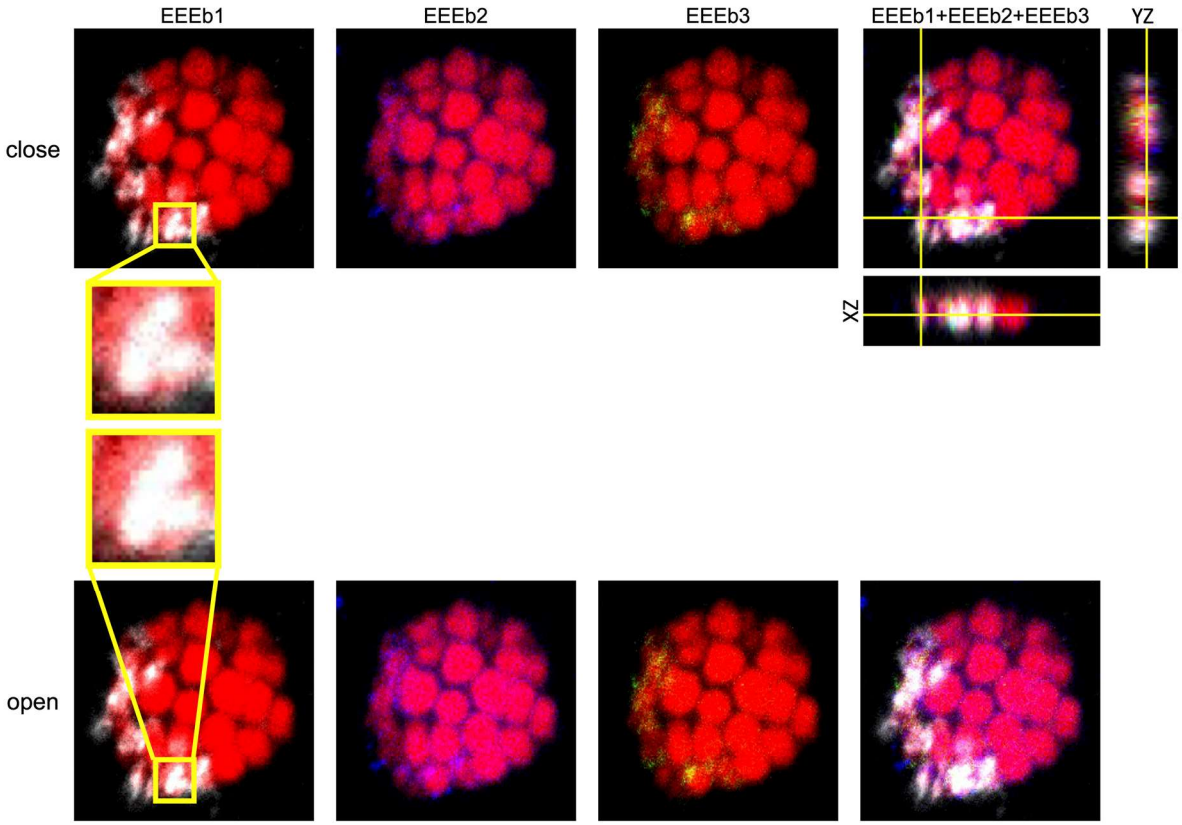


Figure 5. Chromosomal mapping of eliminated DNA families with confocal microscopy.

Metaphase chromosomes in spermatocytes were hybridized using cyanine-5-labeled EEEb1 probes (gray) along with biotin-labelled EEEb2 (blue) and digoxigenin-labelled EEEb3. The upper and lower panels were captured by close (65.13 μm) and open (90.68 μm) pinhole size, respectively. Magnified images of E-chromosomes are shown in the middle. Fluorescence images derived from XZ and YZ sections were displayed only in the small pinhole-size image. Chromosomes were counterstained with propidium iodide (red).

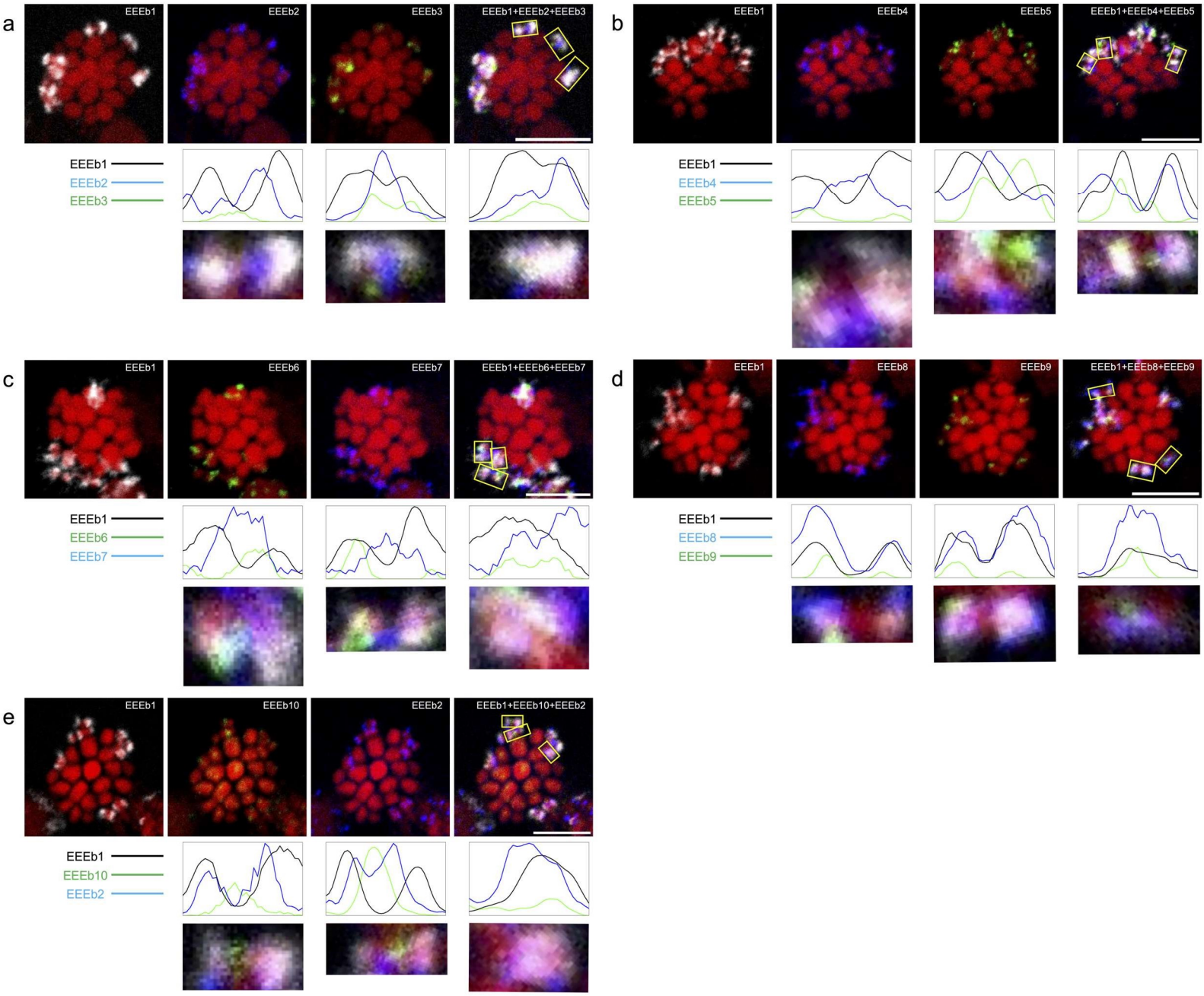


Figure 6. Chromosomal mapping of ten eliminated DNA families in *E. burgeri*.

Metaphase chromosomes in spermatocytes were hybridized using cyanine-5-labeled EEEb1 probes (gray) along with biotin-labelled EEEb2 (a and e), EEEb4 (b), EEEb7 (c), and EEEb8 (d) (blue), and digoxigenin-labelled EEEb3 (a), EEEb5 (b), EEEb6 (c), EEEb9 (d) and EEEb10 (e) (green) (top). Fluorescence intensity histograms, derived from XY-sections, were displayed below representative individual E-chromosomes (bottom). Chromosomes were counterstained with propidium iodide (red). Scale bar = 5 μm.

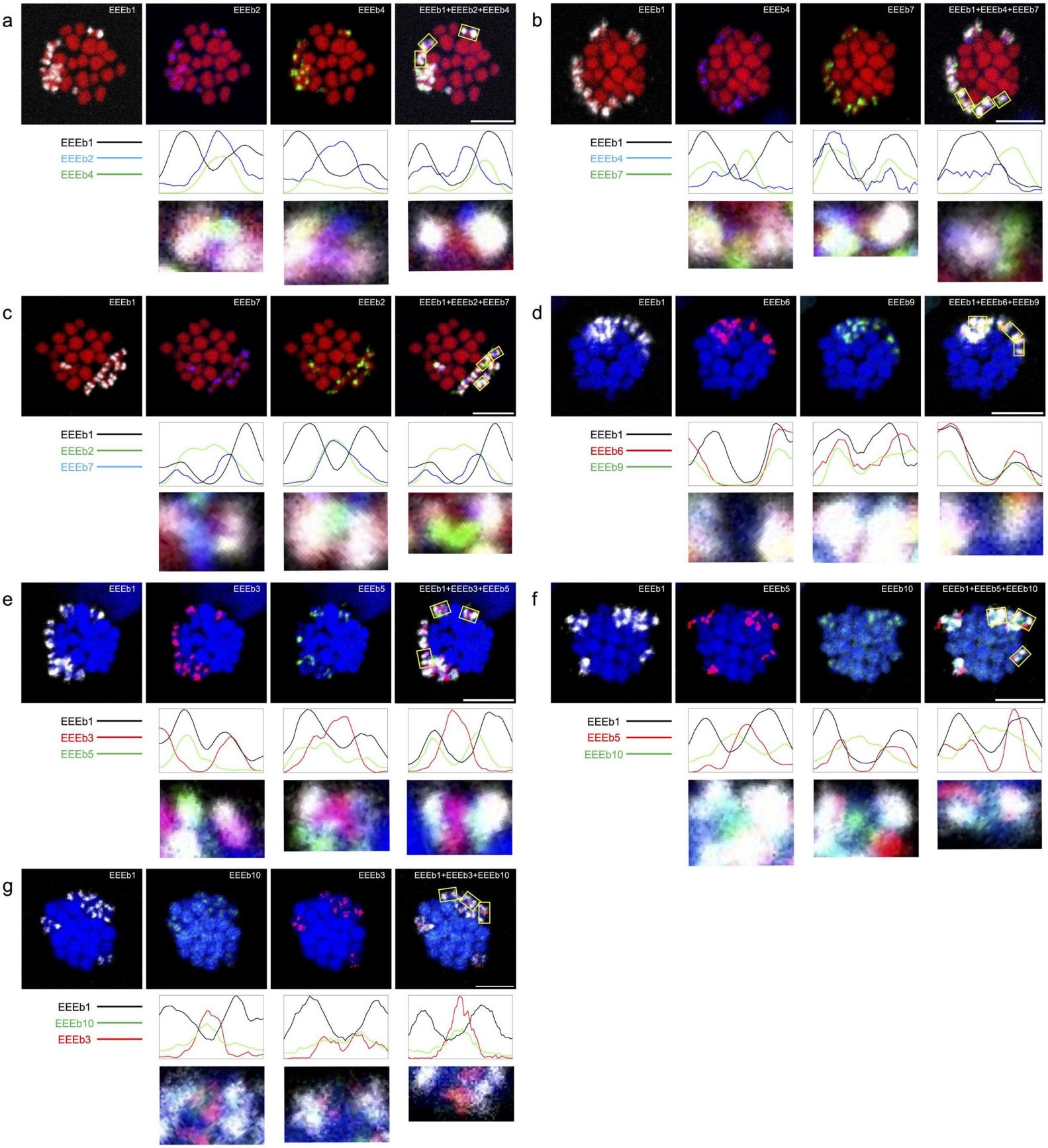
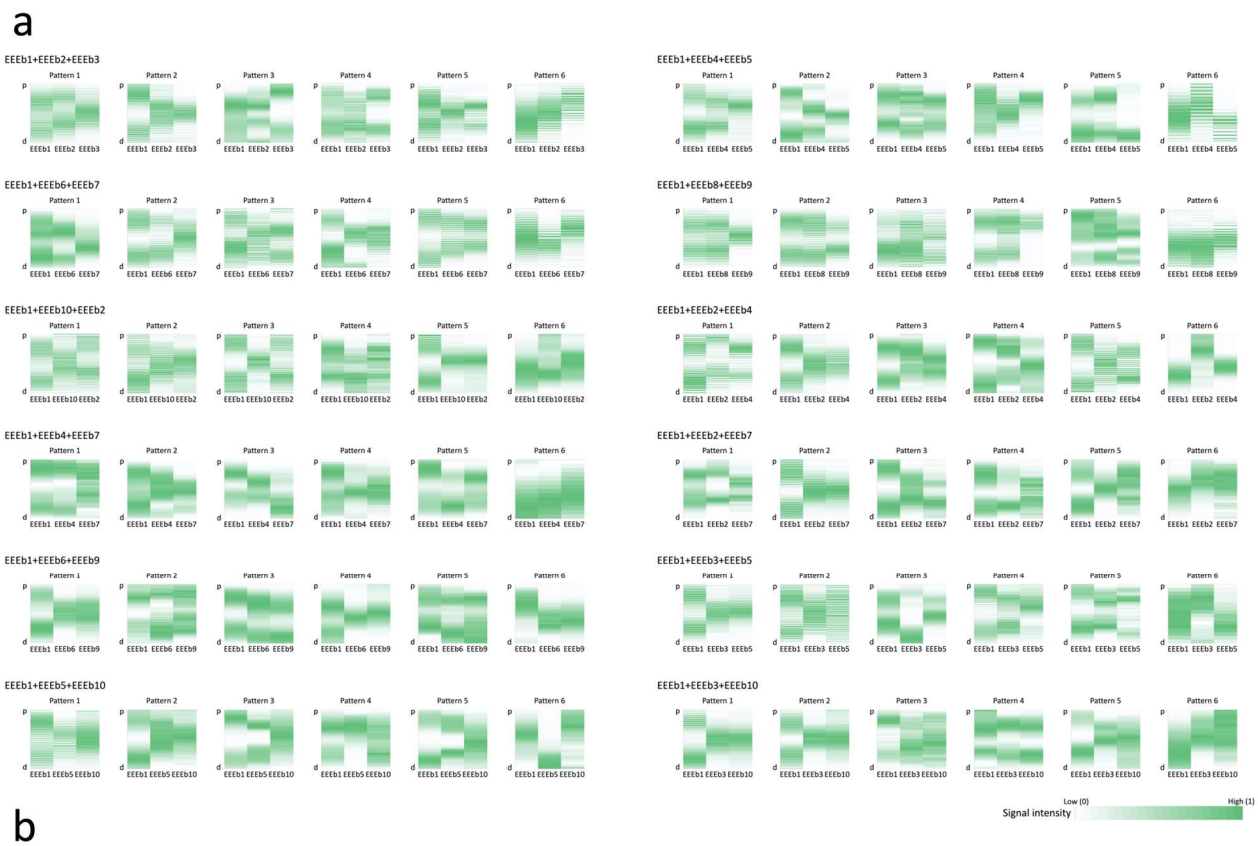

a
EEEb1
EEEb2
EEEb4
EEEb1+EEEb2+EEEb4
EEEb1
EEEb2
EEEb4
b
EEEb1
EEEb4
EEEb7
EEEb1+EEEb4+EEEb7
EEEb1
EEEb4
EEEb7
c
EEEb1
EEEb7
EEEb2
EEEb1+EEEb2+EEEb7
EEEb1
EEEb2
EEEb7
d
EEEb1
EEEb6
EEEb9
EEEb1+EEEb6+EEEb9
EEEb1
EEEb6
EEEb9
e
EEEb1
EEEb3
EEEb5
EEEb1+EEEb3+EEEb5
EEEb1
EEEb3
EEEb5
f
EEEb1
EEEb5
EEEb10
EEEb1+EEEb5+EEEb10
EEEb1
EEEb5
EEEb10
g
EEEb1
EEEb10
EEEb3
EEEb1+EEEb3+EEEb10
EEEb1
EEEb10
EEEb3

Figure 7. Chromosomal mapping of similarly distributed DNA families in *E. burgeri*.

Metaphase chromosomes from spermatocytes were hybridized using cyanine-5-labeled EEEb1 probes (gray) with biotin-labelled EEEb2 (blue) and digoxigenin-labelled EEEb4 (green) (a), digoxigenin-labelled EEEb4 (green) and biotin-labelled EEEb7 (blue) (b), biotin-labelled EEEb7 (blue) and digoxigenin-labelled EEEb2 (green) (c), biotin-labelled EEEb6 (red) and digoxigenin-labelled EEEb9 (green) (d), biotin-labelled EEEb3 (red) and digoxigenin-labelled EEEb5 (green) (e), biotin-labelled EEEb5 (red) and digoxigenin-labelled EEEb10 (green) (e), or digoxigenin-labelled EEEb10 (green) and biotin-labelled EEEb3 (red) (g). Chromosomes were counterstained with propidium iodide (a–c) (red) or Hoechst33342 (d–g) (blue), respectively. The other notions correspond to those to Figure 6.


a
EEEb1+EEEb2+EEEb3
EEEb1+EEEb4+EEEb5
EEEb1+EEEb6+EEEb7
EEEb1+EEEb8+EEEb9
EEEb1+EEEb10+EEEb2
EEEb1+EEEb2+EEEb4
EEEb1+EEEb4+EEEb7
EEEb1+EEEb2+EEEb7
EEEb1+EEEb6+EEEb9
EEEb1+EEEb3+EEEb5
EEEb1+EEEb5+EEEb10
EEEb1+EEEb3+EEEb10
Pattern 1
Pattern 2
Pattern 3
Pattern 4
Pattern 5
Pattern 6
Low (0)
High (1)
Signal intensity

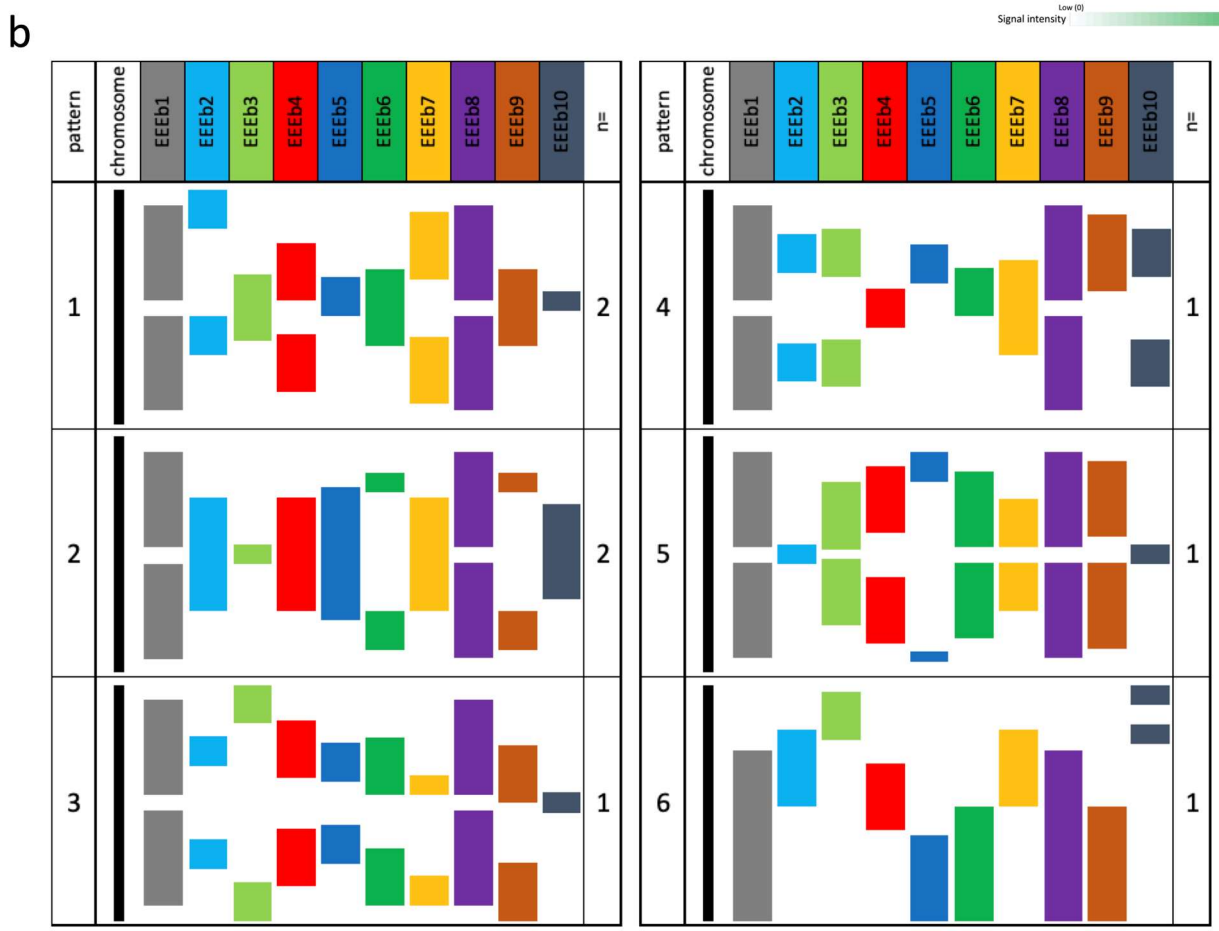

b
pattern
chromosome
EEEb1
EEEb2
EEEb3
EEEb4
EEEb5
EEEb6
EEEb7
EEEb8
EEEb9
EEEb10
n=
1
2
3
4
5
6

Figure 8. Schematic karyogram of the distribution patterns of the ten eliminated DNA families in *E. burgeri*.

This heatmap and karyogram are based on the linescaned images of FISH using EEEb1 to EEEb10 presented in this study (Figures 2 and 3). (a) The heatmap is based on the results of FISH using EEEb1 to EEEb10 presented in this study (Figures 2 and 3). The relative signal intensities on each E-chromosomes were investigated by the heatmap analysis. The eight pairs of E-chromosomes were divided into six patterns based on the signal distribution patterns of EEEb1–EEEb10 in metaphase chromosomes of spermatocytes (patterns 1 to 6). "p" and "d" at the horizontal axis indicate the hypothetically defined proximal region and distal region of each chromosome. (b) The schematic karyogram is based on the results of the above heatmap analysis. Although the signals of EEEb10 were detected on the entire region of E-chromosomes (and somatically retained chromosomes), only the intense signal regions are depicted in this karyogram. E-chromosomes in pattern1–5 showed symmetric distribution patterns at least seven eliminated DNA families, while pattern 6 E-chromosomes did not show symmetric signal distribution of any DNA families. "n=" indicates the chromosomal pair(s) of the E-chromosomes categorized into each pattern.

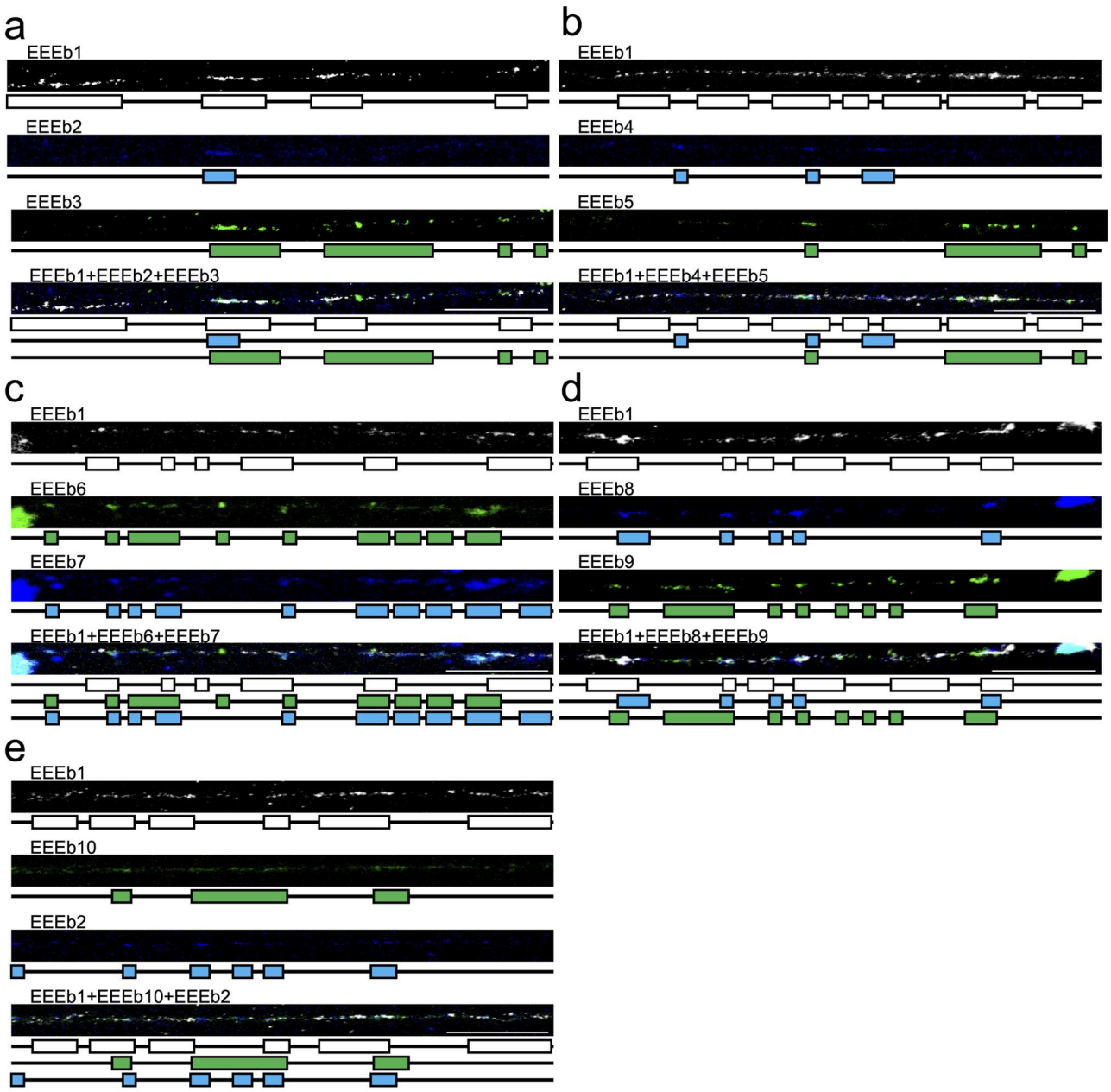


Figure 9. Fiber-FISH analysis of ten eliminated DNA families in *E. burgeri*.

Extended chromatin fibers from testicular cells were hybridized using cyanine-5-labeled EEEb1 probes (gray) with biotin-labelled EEEb2 (a and e), EEEb4 (b), EEEb7 (c), and EEEb8 (d) (blue) and digoxigenin-labelled EEEb3 (a), EEEb5 (b), EEEb6 (c), EEEb9 (d), and EEEb10 (e) (green). Schematic diagrams are shown at below of each signal image. Scale bar = 20 μm.

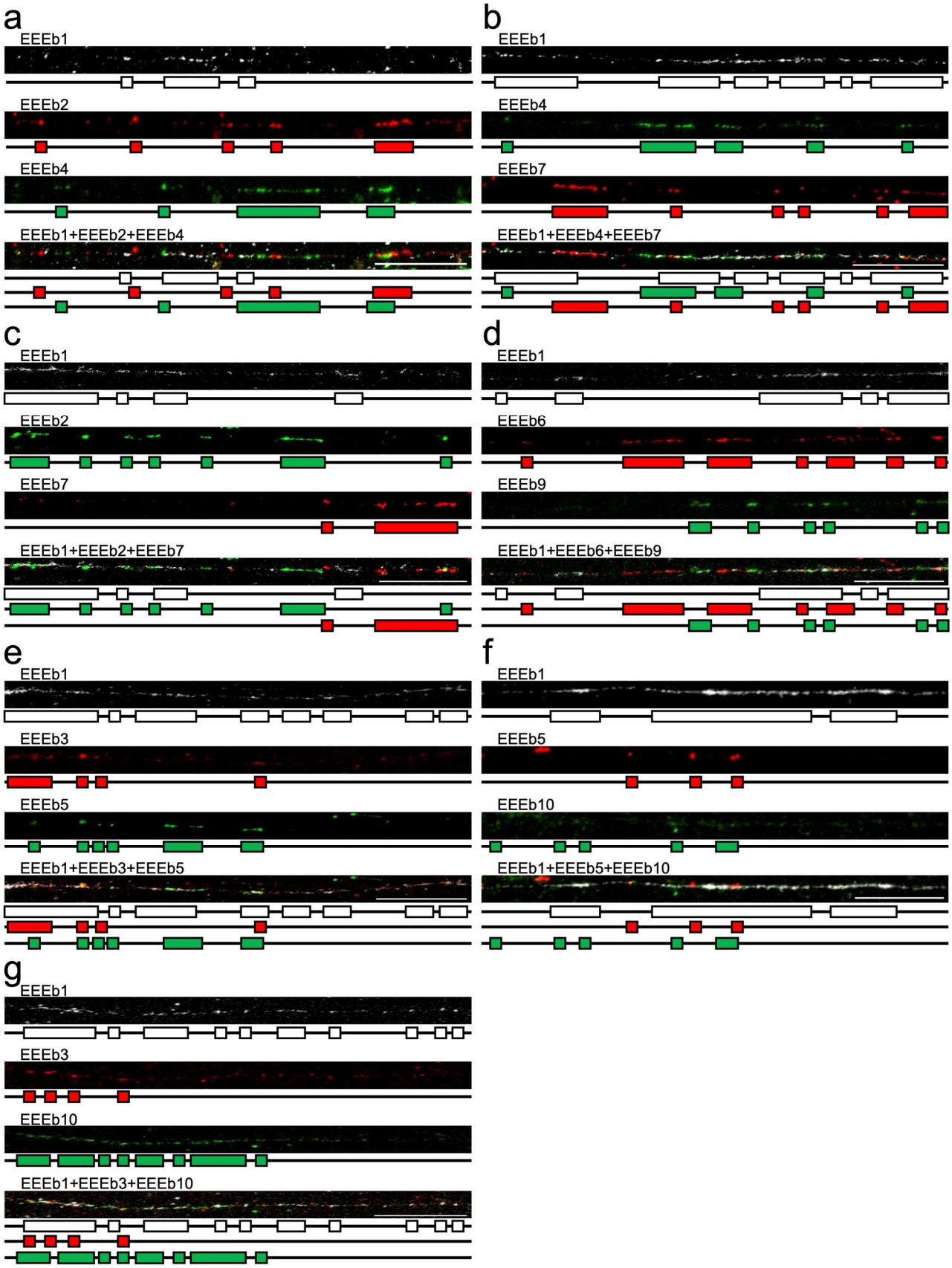

a
EEEb1
EEEb2
EEEb4
EEEb1+EEEb2+EEEb4
b
EEEb1
EEEb4
EEEb7
EEEb1+EEEb4+EEEb7
c
EEEb1
EEEb2
EEEb7
EEEb1+EEEb2+EEEb7
d
EEEb1
EEEb6
EEEb9
EEEb1+EEEb6+EEEb9
e
EEEb1
EEEb3
EEEb5
EEEb1+EEEb3+EEEb5
f
EEEb1
EEEb5
EEEb10
EEEb1+EEEb5+EEEb10
g
EEEb1
EEEb3
EEEb10
EEEb1+EEEb3+EEEb10

Figure 10. Fiber-FISH analysis of similarly distributed DNA families in *E. burgeri*.

Extended chromatin fibers from testicular cells were hybridized using cyanine-5-labeled EEEb1 probes (gray) with biotin-labelled EEEb2 (red) and digoxigenin-labelled EEEb4 (green) (a), digoxigenin-labelled EEEb4 (green) and biotin-labelled EEEb7 (red) (b), digoxigenin-labelled EEEb2 (green) and biotin-labelled EEEb7 (red) (c), biotin-labelled EEEb6 (red) and digoxigenin-labelled EEEb9 (green) (d), biotin-labelled EEEb3 (red) and digoxigenin-labelled EEEb5 (green) (e), biotin-labelled EEEb5 (red) and digoxigenin-labelled EEEb10 (green) (e), or biotin-labelled EEEb3 (red) and digoxigenin-labelled EEEb10 (green) (g). The other notions correspond to those to Figure 9.

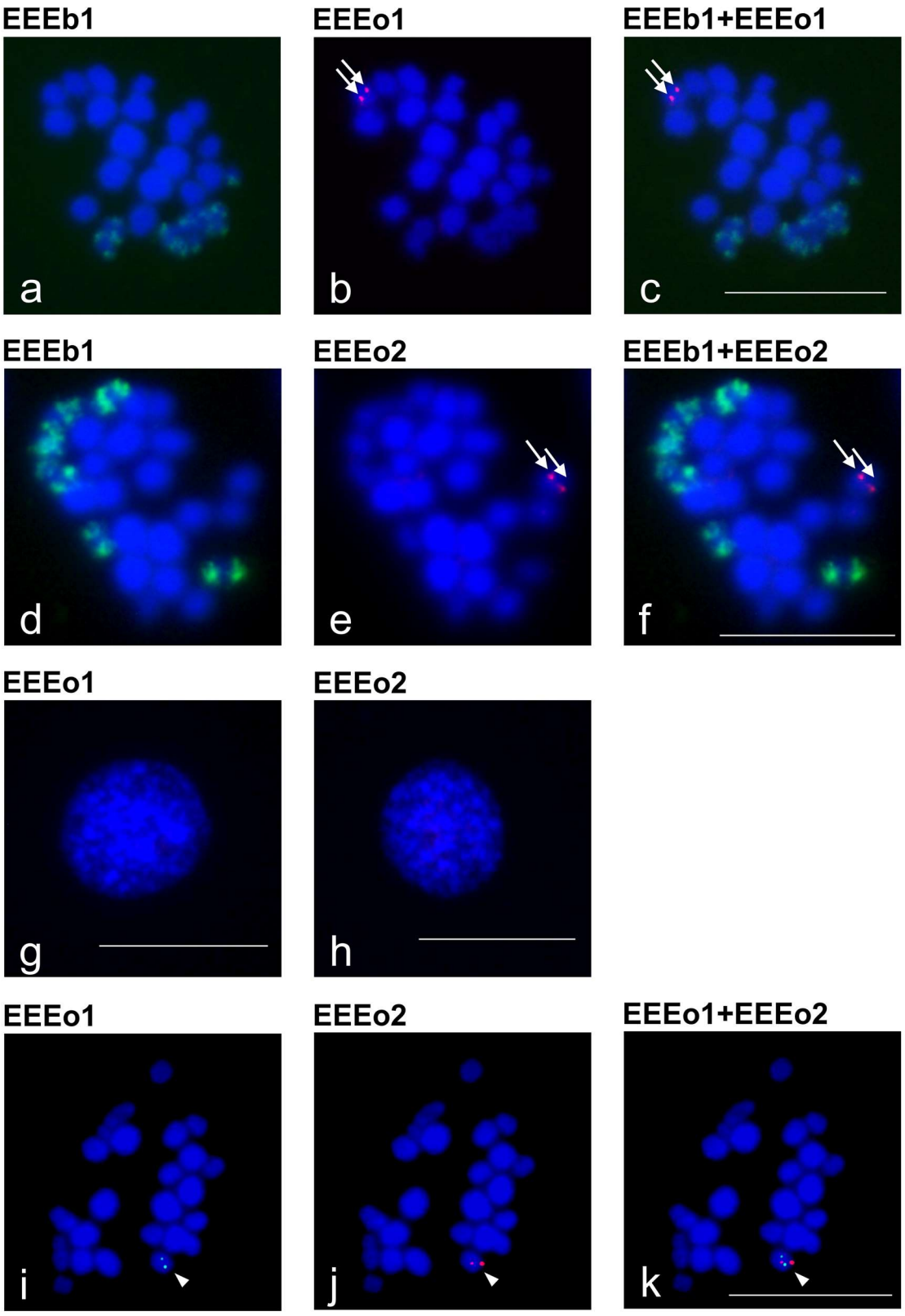

EEEb1
EEEo1
EEEb1+EEEo1
a
b
c
EEEb1
EEEo2
EEEb1+EEEo2
d
e
f
EEEo1
EEEo2
g
h
EEEo1
EEEo2
EEEo1+EEEo2
i
j
k

Figure 11. Chromosomal mapping of EEEo1 and EEEo2 families in *E. burgeri*.

EEEb1 (green), EEEo1 (red), and merged image in meiosis I metaphases (a to c). EEEb1 (green), EEEo2 (red), and merged image (d to f). Neither EEEo1 (red) nor EEEo2 (red) signals appear in the somatic cells (g and h). EEEo1 (green), EEEo2 (red), and merged image in meiosis I metaphases (i to k). Arrows shown in b, c, e, and f indicate EEEo1 and EEEo2 signals. Arrowheads denote the chromosome with hybridization signals of EEEo1 and/or EEEo2 probes (i–k). Scale bar = 5 μm.

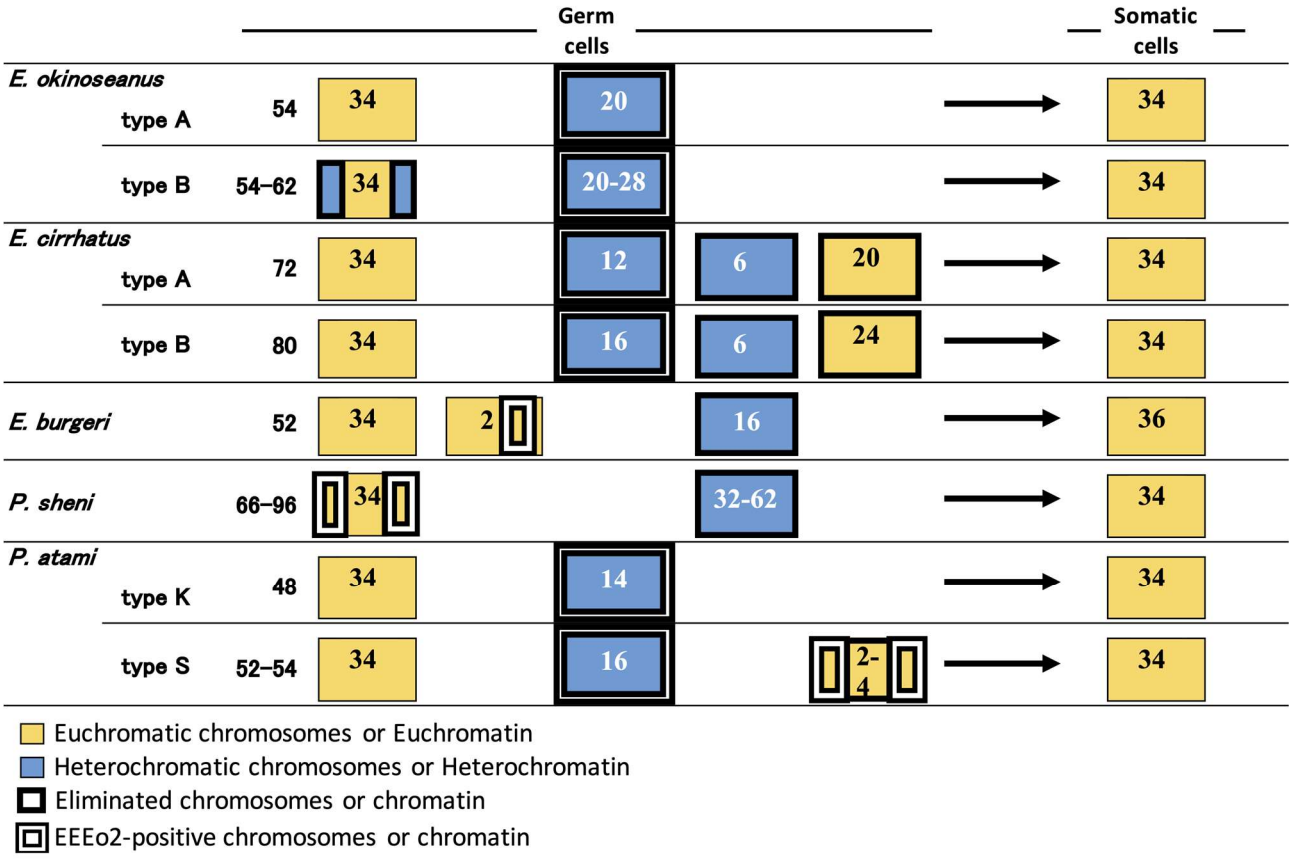


Figure 12. Schematic diagrams of elimination patterns of EEEo2 and E-chromosome in five hagfish.

Chromosomes having the same features are grouped, and their numbers are listed in the squares. The total number of chromosomes in germ cells are shown at the left. The other notations are shown at bottom.

Acknowledgements

I would like to express my sincere appreciation to Associate Professor Yuji Goto, Professor Souichirou Kubota, and Emeritus Professor Sei-ichi Kohno (Toho University), for their invaluable guidance and critical feedback on the manuscript. I am grateful to Mr. Yoshiki Tanaka, Assistant Professor Rei Kajitani, and Professor Takehiko Itoh (Tokyo Institute of Technology), as well as to Mr. Tomoyuki Aizu, Ms. Hinako Ishizaki, and Specifically Appointed Professor Atushi Toyoda (National Institute of Genomics), for their technical support of DNA sequencing. Additionally, I would like to thank Mr. Susumu Hatanaka (Shinshoumaru, Fishermen's Cooperative of Enosima-Katase), for supplying me with experimental materials.

References

Alkhimova OG, Mazurok NA, Potapova TA, Zakian SM, Heslop-Harrison JS, Vershinin AV (2004) Diverse patterns of the tandem repeats organization in rye chromosomes. *Chromosoma* 113(1):42-52

Bachmann-Waldmann C, Jentsch S, Tobler H, Muller F (2004) Chromatin diminution leads to rapid evolutionary changes in the organization of the germ line genomes of the parasitic nematodes *A. suum* and *P. univalens*. *Mol Biochem Parasitol* 134(1):53-64

Bauer H, Beermann W (1952) Der Chromosomencyclus der Orthocladiinen (*Nematocera*, *Diptera*). *Zeitschrift für Naturforschung B* 7(9-10):557-563

Biederman MK, Nelson MM, Asalone KC, Pedersen AL, Saldanha CJ, Bracht JR (2018) Discovery of the First Germline-Restricted Gene by Subtractive Transcriptomic Analysis in the Zebra Finch, *Taeniopygia guttata*. *Curr Biol* 28(10):1620-1627 e1625

Boveri T (1887) Über Differenzierung der Zellkerne während der Furchung des Eies von *Ascaris megalocephala*. *Anat Anz* 2(688-693

Chen ZQ, Lin CC, Hodgetts RB (1989) Cloning and characterization of a tandemly repeated DNA sequence in the crane family (Gruidae). *Genome* 32(4):646-654

Cheng Z, Buell CR, Wing RA, Jiang J (2002) Resolution of fluorescence *in-situ* hybridization mapping on rice mitotic prometaphase chromosomes, meiotic pachytene chromosomes and extended DNA fibers. *Chromosome Res* 10(5):379-387

Chinone H. (2010) Analysis of repetitive sequences exclusively amplified in germline genome of *Eptatretus burgeri*. Master Dissertation. Department of Biology, Faculty of Science, Toho University, Japan.

Chisuwa R. (2017) The cytogenetic analysis of the histone modifications in *Eptatretus burgeri*. Master Dissertation. Department of Biology, Faculty of Science, Toho University, Japan.

Dalla Benetta E, Antoshechkin I, Yang T, Nguyen HQM, Ferree PM, Akbari OS (2020) Genome elimination mediated by gene expression from a selfish chromosome. *Sci Adv* 6(14):eaaz9808

de Barros AV, Sczepanski TS, Cabrero J, Camacho JPM, Vicari MR, Artoni RF (2011) Fiber FISH reveals different patterns of high-resolution physical mapping for repetitive DNA in fish. *Aquaculture* 322-323(47-50

Dedukh D, Krasikova A (2022) Delete and survive: strategies of programmed genetic material elimination in eukaryotes. *Biol Rev Camb Philos Soc* 97(1):195-216

Degtyarev S, Boykova T, Grishanin A, Belyakin S, Rubtsov N, Karamysheva T, Makarevich G, Akifyev A, Zhimulev I (2004) The molecular structure of the DNA fragments eliminated during chromatin diminution in *Cyclops kolensis*. *Genome Res* 14(11):2287-2294

Dockendorff TC, Estrem B, Reed J, Simmons JR, Zadegan SB, Zagoskin MV, Terta V, Villalobos E, Seaberry EM, Wang J (2022) The nematode *Oscheius tipulae* as a genetic model for programmed DNA elimination. *Curr Biol* 32(23):5083-5098 e5086

Fransz PF, Alonso-Blanco C, Liharska TB, Peeters AJ, Zabel P, de Jong JH (1996) High-resolution physical mapping in *Arabidopsis thaliana* and tomato by fluorescence in situ hybridization to extended DNA fibres. *Plant J* 9(3):421-430

Gorovsky MA (1973) Macro- and micronuclei of *Tetrahymena pyriformis*: a model system for studying the structure and function of eukaryotic nuclei. *J Protozool* 20(1):19-25

Goto Y, Kubota S, Kohno S (1998) Highly repetitive DNA sequences that are restricted to the germ line in the hagfish *Eptatretus cirrhatus*: a mosaic of eliminated elements. *Chromosoma* 107(1):17-32

Goto Y, Osawa D, Kubota S (2016) Preferential demethylation of DNA cytosine on the chromosomes restricted to germ cells in the spermatocytes but not the spermatogonia in the inshore hagfish, *Eptatretus burgeri*. *Chromosome Sci* 19(1-4):11-18

Green MR, Sambrook J (2012) *Molecular Cloning: A Laboratory Manual (Fourth Edition)*. Cold

Spring Harbor Laboratory Press.

Habermann FA, Cremer M, Walter J, Kreth G, von Hase J, Bauer K, Wienberg J, Cremer C, Cremer T, Solovei I (2001) Arrangements of macro- and microchromosomes in chicken cells. *Chromosome Res* 9(7):569-584

Hamilton EP, Kapusta A, Huvos PE, Bidwell SL, Zafar N, Tang H, Hadjithomas M, Krishnakumar V, Badger JH, Caler EV et al. (2016) Structure of the germline genome of *Tetrahymena thermophila* and relationship to the massively rearranged somatic genome. *Elife* 5(

Hanlon SL, Miller DE, Eche S, Hawley RS (2018) Origin, Composition, and Structure of the Supernumerary B Chromosome of *Drosophila melanogaster*. *Genetics* 210(4):1197-1212

Hodson CN, Jaron KS, Gerbi S, Ross L (2022) Gene-rich germline-restricted chromosomes in black-winged fungus gnats evolved through hybridization. *PLoS Biol* 20(2):e3001559

Ishishita S, Tsuruta Y, Uno Y, Nakamura A, Nishida C, Griffin DK, Tsudzuki M, Ono T, Matsuda Y (2014) Chromosome size-correlated and chromosome size-uncorrelated homogenization of centromeric repetitive sequences in New World quails. *Chromosome Res* 22(1):15-34

Itoh Y, Kampf K, Pigozzi MI, Arnold AP (2009) Molecular cloning and characterization of the germline-restricted chromosome sequence in the zebra finch. *Chromosoma* 118(4):527-536

Jackson SA, Dong F, Jiang J (1999) Digital mapping of bacterial artificial chromosomes by fluorescence *in situ* hybridization. *Plant J* 17(5):581-587

Jackson SA, Wang ML, Goodman HM, Jiang J (1998) Application of fiber-FISH in physical mapping of *Arabidopsis thaliana*. *Genome* 41(4):566-572

Jehangir M, Ahmad SF, Cardoso AL, Ramos E, Valente GT, Martins C (2019) De novo genome assembly of the cichlid fish *Astatotilapia latifasciata* reveals a higher level of genomic polymorphism and genes related to B chromosomes. *Chromosoma* 128(2):81-96

Kang Y, Wang J, Neff A, Kratzer S, Kimura H, Davis RE (2016) Differential Chromosomal Localization of Centromeric Histone CENP-A Contributes to Nematode Programmed DNA

Elimination. *Cell Rep* 16(9):2308-2316

Kinsella CM, Ruiz-Ruano FJ, Dion-Cote AM, Charles AJ, Gossmann TI, Cabrero J, Kappei D, Hemmings N, Simons MJP, Camacho JPM et al. (2019) Programmed DNA elimination of germline development genes in songbirds. *Nat Commun* 10(1):5468

Kirk KE, Blackburn EH (1995) An unusual sequence arrangement in the telomeres of the germ-line micronucleus in *Tetrahymena thermophila*. *Genes Dev* 9(1):59-71

Kohno S, Kubota S, Nakai Y (1998) Chromatin Diminution and Chromosome Elimination in Hagfishes. In *The Biology of Hagfishes* (ed. JM Jørgensen, et al.), pp. 81-100. Springer Netherlands, Dordrecht.

Kohno S, Nakai Y, Satoh S, Yoshida M, Kobayashi H (1986) Chromosome elimination in the Japanese hagfish, *Eptatretus burgeri* (Agnatha, Cyclostomata). *Cytogenet Cell Genet* 41(4):209-214

Kojima NF, Kojima KK, Kobayakawa S, Higashide N, Hamanaka C, Nitta A, Koeda I, Yamaguchi T, Shichiri M, Kohno S et al. (2010) Whole chromosome elimination and chromosome terminus elimination both contribute to somatic differentiation in Taiwanese hagfish *Paramyxine sheni*. *Chromosome Res* 18(3):383-400

Kubota S, Ishibashi T, Kohno S (1997) A germline restricted, highly repetitive DNA sequence in *Paramyxine atami*: an interspecifically conserved, but somatically eliminated, element. *Mol Gen Genet* 256(3):252-256

Kubota S, Kuro-o M, Mizuno S, Kohno S (1993) Germ line-restricted, highly repeated DNA sequences and their chromosomal localization in a Japanese hagfish (*Eptatretus okinoseanus*). *Chromosoma* 102(3):163-173

Kubota S, Takano J, Tsuneishi R, Kobayakawa S, Fujikawa N, Nabeyama M, Kohno S (2001) Highly repetitive DNA families restricted to germ cells in a Japanese hagfish (*Eptatretus burgeri*): a hierarchical and mosaic structure in eliminated chromosomes. *Genetica* 111(1-3):319-328

Kuhn GC, Sene FM, Moreira-Filho O, Schwarzacher T, Heslop-Harrison JS (2008) Sequence analysis, chromosomal distribution and long-range organization show that rapid turnover of new and old pBuM satellite DNA repeats leads to different patterns of variation in seven species of the *Drosophila buzzatii* cluster. *Chromosome Res* 16(2):307-324

Kuhn GC, Teo CH, Schwarzacher T, Heslop-Harrison JS (2009) Evolutionary dynamics and sites of illegitimate recombination revealed in the interspersion and sequence junctions of two nonhomologous satellite DNAs in cactophilic *Drosophila* species. *Heredity (Edinb)* 102(5):453-464

Makunin AI, Romanenko SA, Beklemisheva VR, Perelman PL, Druzhkova AS, Petrova KO, Prokopov DY, Chernyaeva EN, Johnson JL, Kukekova AV et al. (2018) Sequencing of Supernumerary Chromosomes of Red Fox and Raccoon Dog Confirms a Non-Random Gene Acquisition by B Chromosomes. *Genes (Basel)* 9(8)

Malinovskaya LP, Zadesenets KS, Karamysheva TV, Akberdina EA, Kizilova EA, Romanenko MV, Shnaider EP, Scherbakova MM, Korobitsyn IG, Rubtsov NB et al. (2020) Germline-restricted chromosome (GRC) in the sand martin and the pale martin (Hirundinidae, Aves): synapsis, recombination and copy number variation. *Sci Rep* 10(1):1058

Mestriner CA, Galetti PM, Jr., Valentini SR, Ruiz IR, Abel LD, Moreira-Filho O, Camacho JP (2000) Structural and functional evidence that a B chromosome in the characid fish *Astyanax scabripinnis* is an isochromosome. *Heredity (Edinb)* 85 (Pt 1)(1):1-9

Metz CW (1938) Chromosome Behavior, Inheritance and Sex Determination in Sciara. *The American Naturalist* 72(743):485-520

Milani D, Bardella VB, Ferretti A, Palacios-Gimenez OM, Melo AS, Moura RC, Loreto V, Song H, Cabral-de-Mello DC (2018) Satellite DNAs Unveil Clues about the Ancestry and Composition of B Chromosomes in Three Grasshopper Species. *Genes (Basel)* 9(11)

Milani D, Ramos E, Loreto V, Marti DA, Cardoso AL, de Moraes KCM, Martins C, Cabral-de-Mello

DC (2017) The satellite DNA AflaSAT-1 in the A and B chromosomes of the grasshopper *Abracris flavolineata*. *BMC Genet* 18(1):81

Nabeyama M, Kubota S, Kohno S (2000) Concerted evolution of a highly repetitive DNA family in eptatretidae (Cyclostomata, agnatha) implies specifically differential homogenization and amplification events in their germ cells. *J Mol Evol* 50(2):154-169

Nagao K, Otsuzumi T, Chinone H, Sasaki T, Yoshimoto J, Matsuda M, Kubota S, Goto Y (2022) Novel selectively amplified DNA sequences in the germline genome of the Japanese hagfish, *Eptatretus burgeri*. *Sci Rep* 12(1):21373

Nei M, Rooney AP (2005) Concerted and birth-and-death evolution of multigene families. *Annu Rev Genet* 39(121-152

Noto T, Mochizuki K (2017) Whats, hows and whys of programmed DNA elimination in *Tetrahymena*. *Open Biol* 7(10)

Otsuzumi T. (2009) Genomic study of subtracted DNAs that are specifically expressed in germline genome of *Eptatretus burgeri*. Master Dissertation. Department of Biology, Faculty of Science, Toho University, Japan.

Paco A, Adega F, Mestrovic N, Plohl M, Chaves R (2015) The puzzling character of repetitive DNA in *Phodopus* genomes (Cricetidae, Rodentia). *Chromosome Res* 23(3):427-440

Pascual-Anaya J, Sato I, Sugahara F, Higuchi S, Paps J, Ren Y, Takagi W, Ruiz-Villalba A, Ota KG, Wang W et al. (2018) Hagfish and lamprey *Hox* genes reveal conservation of temporal colinearity in vertebrates. *Nat Ecol Evol* 2(5):859-866

Pigozzi MI, Solari AJ (1998) Germ cell restriction and regular transmission of an accessory chromosome that mimics a sex body in the zebra finch, *Taeniopygia guttata*. *Chromosome Res* 6(2):105-113

Ruiz-Ruano FJ, Cabrero J, Lopez-Leon MD, Sanchez A, Camacho JPM (2018) Quantitative sequence characterization for repetitive DNA content in the supernumerary chromosome of the

migratory locust. *Chromosoma* 127(1):45-57

Shang WH, Hori T, Toyoda A, Kato J, Popendorf K, Sakakibara Y, Fujiyama A, Fukagawa T (2010) Chickens possess centromeres with both extended tandem repeats and short non-tandem-repetitive sequences. *Genome Res* 20(9):1219-1228

Shichiri M, Kubota S, Miyaji K, Kohno S (2001) Ultrastructure of the kinetochores in Japanese hagfish, *Eptatretus burgeri*. *Chromosome Sci* 5(3):139-144

Silva DMZA, Daniel SN, Camacho JP, Utsunomia R, Ruiz-Ruano FJ, Penitente M, Pansonato-Alves JC, Hashimoto DT, Oliveira C, Porto-Foresti F et al. (2016) Origin of B chromosomes in the genus *Astyanax* (Characiformes, Characidae) and the limits of chromosome painting. *Mol Genet Genomics* 291(3):1407-1418

Silva DMZA, Pansonato-Alves JC, Utsunomia R, Araya-Jaime C, Ruiz-Ruano FJ, Daniel SN, Hashimoto DT, Oliveira C, Camacho JP, Porto-Foresti F et al. (2014) Delimiting the origin of a B chromosome by FISH mapping, chromosome painting and DNA sequence analysis in *Astyanax paranae* (Teleostei, Characiformes). *PLoS One* 9(4):e94896

Silva DMZA, Utsunomia R, Ruiz-Ruano FJ, Daniel SN, Porto-Foresti F, Hashimoto DT, Oliveira C, Camacho JPM, Foresti F (2017) High-throughput analysis unveils a highly shared satellite DNA library among three species of fish genus *Astyanax*. *Sci Rep* 7(1):12726

Smith JJ, Antonacci F, Eichler EE, Amemiya CT (2009) Programmed loss of millions of base pairs from a vertebrate genome. *Proc Natl Acad Sci U S A* 106(27):11212-11217

Smith JJ, Stuart AB, Sauka-Spengler T, Clifton SW, Amemiya CT (2010) Development and analysis of a germline BAC resource for the sea lamprey, a vertebrate that undergoes substantial chromatin diminution. *Chromosoma* 119(4):381-389

Smith JJ, Timoshevskaya N, Ye C, Holt C, Keinath MC, Parker HJ, Cook ME, Hess JE, Narum SR, Lamanna F et al. (2018) The sea lamprey germline genome provides insights into programmed genome rearrangement and vertebrate evolution. *Nat Genet* 50(2):270-277

Souza J, Guimaraes E, Pinheiro-Figliuolo V, Cioffi MB, Bertollo LAC, Feldberg E (2021) Chromosomal Analysis of *Ctenolucius hujeta* Valenciennes, 1850 (Characiformes): A New Piece in the Chromosomal Evolution of the Ctenoluciidae. *Cytogenet Genome Res* 161(3-4):195-202

Staiber W (2004) Molecular evolution of homologous gene sequences in germline-limited and somatic chromosomes of *Acricotopus*. *Genome* 47(4):732-741

Staiber W, Wech I, Preiss A (1997) Isolation and chromosomal localization of a germ line-specific highly repetitive DNA family in *Acricotopus lucidus* (Diptera, Chironomidae). *Chromosoma* 106(5):267-275

Takano J. (1998) Analyses of germline-restricted, highly repetitive DNA families in a Japanese hagfish (*Eptatretus burgeri*). Master Dissertation. Department of Biology, Faculty of Science, Toho University, Japan.

Tanabe H, Habermann FA, Solovei I, Cremer M, Cremer T (2002) Non-random radial arrangements of interphase chromosome territories: evolutionary considerations and functional implications. *Mutat Res* 504(1-2):37-45

Tanaka Y, Kajitani R, Itoh T (2023) cycle_finder: *de novo* analysis of tandem and interspersed repeats based on cycle-finding. *bioRxiv* doi:10.1101/2023.07.17.549334:2023.2007.2017.549334

Timoshevskaya N, Eskut KI, Timoshevskiy VA, Robb SMC, Holt C, Hess JE, Parker HJ, Baker CF, Miller AK, Saraceno C et al. (2023) An improved germline genome assembly for the sea lamprey *Petromyzon marinus* illuminates the evolution of germline-specific chromosomes. *Cell Rep* 42(3):112263

Timoshevskiy VA, Lampman RT, Hess JE, Porter LL, Smith JJ (2017) Deep ancestry of programmed genome rearrangement in lampreys. *Dev Biol* 429(1):31-34

Torgasheva AA, Malinovskaya LP, Zadesenets KS, Karamysheva TV, Kizilova EA, Akberdina EA, Pristyazhnyuk IE, Shnaider EP, Volodkina VA, Saifitdinova AF et al. (2019) Germline-

restricted chromosome (GRC) is widespread among songbirds. *Proc Natl Acad Sci U S A* 116(24):11845-11850

Ugarkovic D, Plohl M (2002) Variation in satellite DNA profiles--causes and effects. *EMBO J* 21(22):5955-5959

Utsunomia R, Silva DM, Ruiz-Ruano FJ, Araya-Jaime C, Pansonato-Alves JC, Scacchetti PC, Hashimoto DT, Oliveira C, Trifonov VA, Porto-Foresti F et al. (2016) Uncovering the Ancestry of B Chromosomes in *Moenkhausia sanctaefilomenae* (Teleostei, Characidae). *PLoS One* 11(3):e0150573

Viana PF, Ezaz T, Marajo L, Ferreira M, Zuanon J, Cioffi MB, Bertollo LAC, Gross MC, Feldberg E (2017) Genomic Organization of Repetitive DNAs and Differentiation of an XX/XY Sex Chromosome System in the Amazonian Puffer Fish, *Colomesus asellus* (Tetraodontiformes). *Cytogenet Genome Res* 153(2):96-104

Wang J, Davis RE (2014) Programmed DNA elimination in multicellular organisms. *Curr Opin Genet Dev* 27(26-34

Wang J, Gao S, Mostovoy Y, Kang Y, Zagoskin M, Sun Y, Zhang B, White LK, Easton A, Nutman TB et al. (2017) Comparative genome analysis of programmed DNA elimination in nematodes. *Genome Res* 27(12):2001-2014

Wang K, Zhang W, Jiang Y, Zhang T (2013) Systematic application of DNA fiber-FISH technique in cotton. *PLoS One* 8(9):e75674

Yamada K, Nishida-Umehara C, Matsuda Y (2002) Characterization and chromosomal distribution of novel satellite DNA sequences of the lesser rhea (*Pterocnemia pennata*) and the greater rhea (*Rhea americana*). *Chromosome Res* 10(6):513-523

Yan X, Meng W, Wu F, Xu A, Chen S, Huang S (2016) The Nuclear DNA Content and Genetic Diversity of *Lampetra morii*. *PLoS One* 11(7):e0157494

Yasmin T, Grayson P, Docker MF, Good SV (2022) Pervasive male-biased expression throughout the

germline-specific regions of the sea lamprey genome supports key roles in sex differentiation and spermatogenesis. *Commun Biol* 5(1):434