

令和 6 年 6 月 7 日現在

機関番号：32661

研究種目：基盤研究(C) (一般)

研究期間：2019～2023

課題番号：19K06783

研究課題名(和文)ヌタウナギにおける染色体放出ではDNA鎖内部領域欠失・再結合はおこなわれているか

研究課題名(英文) A chromosome-level assembly of somatic and germline genomes in a Japanese hagfish, *Eptatretus burgeri*

研究代表者

久保田 宗一郎 (KUBOTA, Souichirou)

東邦大学・理学部・教授

研究者番号：30277347

交付決定額(研究期間全体)：(直接経費) 3,400,000円

研究成果の概要(和文)：無顎類のヌタウナギ目魚類は、染色体放出という特異なゲノム再編成(Programmed Genome Rearrangement)を行う。近年、研究代表者はNGS解析により得られたヌタウナギの体細胞と生殖細胞の二つのゲノムの全配列の比較解析により、この種の生殖細胞特異的すなわち放出ゲノムの中に「DNA鎖の内部領域の欠失と再結合」いわゆるInternal DNA deletion(ID)の領域が多数存在する可能性を見いだした。本研究ではHi-C解析により得られた両ゲノムの染色体レベルで再構築された18本のscaffold配列の比較から、染色体末端切除を含むID領域の十数カ所同定、その存在を証明した。

研究成果の学術的意義や社会的意義

ヌタウナギの染色体放出では、これまでは16本のヘテロクロマチンから成る「染色体を丸ごと失う」単純な様式しか想定されていなかったが、今回はじめて想定外の「DNA鎖内部領域の欠失・再結合」(ID)の領域を18本のユークロマチンからなる染色体からの検出に成功した。これはヌタウナギの染色体放出の抜本的な理解の進展に留まらず、すべてのPGRを行う真核生物は皆同様の機構を共有している可能性を強く示唆している。PGRは、一部の生物種にしか観察されないとはいえ、動物界に広く分布しており、また生殖細胞系列と体細胞系列の分化に直結する普遍的な現象であるため、今後この機構のより深い解析が望まれる。

研究成果の概要(英文)：It is known that chromosome elimination (referred to as Programmed Genome Rearrangement; PGR) occurs in hagfish species (Agnatha). In inshore hagfish, *Eptatretus burgeri*, 16 chromosomes have been lost in somatic cells ( $2n = 36$ ), which is equivalent to approx. 21% of the genomic DNA in germ cells ( $2n = 52$ ). A previous study using the next-generation sequencing technology showed a possibility that the eliminated genome (E genome) may contain many regions of Internal DNA deletion (ID) in this species. In this study, using the Hi-C sequencing analysis, whole DNA sequences of somatic and germline genomes were determined and also a chromosome-level assembly of all somatic chromosomes ( $2n = 36$ ) and homologous chromosomes in germ cells were constructed completely. By the comparative analysis of the 18 chromosome-level scaffolds deduced several ID regions including the terminal and internal deletion. The data will serve as a valuable genetic information of the biological significance of PGR.

研究分野：ゲノム科学

キーワード：ヌタウナギ ゲノム再編成 染色体放出 DNA鎖内部欠失 染色体末端切除 染色体レベルゲノム配列構築 Hi-C解析 NGSデータ

## 様式 C - 19、F - 19 - 1 (共通)

### 1. 研究開始当初の背景

高等動物は、体を構成するどの器官・組織由来の細胞でも、核内の染色体数 (DNA 量) は基本的に同じである。しかし、幾つかの生物群では体細胞と生殖細胞の間で細胞あたりの染色体数や DNA 量が大きく異なる現象が知られている。これは、発生初期に 2 つの細胞系列が分化する過程において、始原体細胞になる割球から染色体 (染色質) が失われることによる。この現象は、染色体放出 (染色質削減) と呼ばれ、1887 年 Boveri の線形動物ウマカイチュウ *Parascaris* における観察に始まり、その後、節足動物橈脚目や双翅目等で確認されている。また、原生動物繊毛虫類のテトラヒメナなどでは、接合後に小核から一連のゲノム再編成を受けて大核が分化する。このゲノム再編成も生物学的には染色体放出と同義の現象と考えられ、これらはまとめて PGR (プログラムされたゲノム再編成; Programed Genome Rearrangement) と呼ばれている。このように PGR は、一部の生物種に限られるが動物界に広く観察される現象である。しかし、この現象の起源やその役割は未だ十分に解明されていない。1986 年 Kohno らは、脊椎動物無顎類に分類されるヌタウナギ目のヌタウナギ (*Eptatretus burgeri*) において、後口動物では初めて染色体放出を観察・報告した。以来、研究代表者は本邦産 4 種と海外産 4 種のヌタウナギ目魚類において染色体解析を行い、8 種全ての種で最小 2 本、最大 62 本の (DNA 量では最小 21%; 最大 75%) 染色体放出を確認している。更に、この仲間は主に染色体をまるごと失い体細胞に分化するが、種によっては一部の染色体末端部分を切断して削減すること、体細胞が失う染色体は主にヘテロクロマチンで、その中身は多様な高頻度反復配列のモザイクであることを明らかにした (Kubota ら 1993; 2001; Kojima ら 2010; 他)。近年、研究代表者は次世代シーケンサー (NGS) を用いて、ヌタウナギの生殖細胞ゲノム (G ゲノム) と体細胞ゲノム (S ゲノム) の全配列をそれぞれ決定し、比較することで、生殖細胞にのみ存在するゲノムの中身を網羅的に把握する研究計画を立案し、科研費 { 基盤研究(C)(一般); 2013~15 年度 } 並びに新学術領域研究「ゲノム支援 (2014 年度)」の援助を得て遂行した。また、*k*-mer ベースによるデータも加えた結果この種のゲノムサイズは、生殖細胞で 2.09Gb、体細胞で 1.77Gb となり、これまでに想定されていたサイズ (それぞれ 2.17Gb、1.7Gb) に近似した。現在、両ゲノムの比較解析により生殖細胞特異的なゲノム領域 (E ゲノム; 約 320Mbp) の検出・解析を精力的に進めている。その解析の中で、G ゲノムの scaffold 配列には存在するが S ゲノムの相同な scaffold 配列では欠失している領域がおよそ 5.5Mbp あり、そのおよそ半分にあたる 2.6 Mbp は「DNA 鎖の内部領域の欠失と再結合」いわゆる Internal DNA deletion (ID) されていることが強く示唆された。この結果は、ヌタウナギのようにヒトと変わらない巨大ゲノムを持つ種における PGR では、全く想定されていなかった発見といえる。

### 2. 研究の目的

近年、研究代表者は本邦産のヌタウナギの体細胞 (染色体数  $2n=36$ ) と生殖細胞 ( $2n=52$ ) の 2 つの全ゲノムをそれぞれ決定した。これまではこの種では 16 本のヘテロクロマチンから成る「染色体を丸ごと失う」単純な様式の染色体放出を行うと考えられていたが、2 つのドラフトゲノムの比較解析により、この種の生殖細胞特異的、すなわち放出ゲノム (E ゲノム) の同定を進めた結果、全く想定外の ID の領域を 6800 カ所以上 (欠失の大きさは最大 11kbp、欠失領域のうち遺伝子配列の 80% 以上が含まれていると推定された領域は約 330、逆転写酵素や転位因子に関連する配列が 163 あまり検出されている) と多数検出した。上述の原生動物繊毛虫類に属するテトラヒメナでは、小核ゲノムの 15% あまりが取り除かれ、残りが 50 倍増幅されて大核ゲノムとなる。このゲノム再編成は、染色体が切断されて末端配列が削られた後にテロメアが付加される (Fragmentation & Telomere addition; F&T) 様式と、IES と呼ばれる配列が取り除かれて両側の DNA が繋がる (ID) 様式の 2 種類の反応によって進行する。IES は小核ゲノム中に 6 千から 1 万あまり検出され、長さは 0.5kbp から 20kbp 以上にわたる。テトラヒメナとヌタウナギとは進化的に遠く離れた生物種であるが、2 種における様々な関連事象が、2 種の ID 様式の共通性を強く示唆している。一方、同じ無顎類のヤツメウナギ目のウミヤツメでも 2009 年に染色体放出が報告され、近年体細胞 (Smith ら, 2013) 並びに生殖細胞 (Smith ら, 2018) のドラフトゲノムも公表された。しかし、この種における E ゲノム解析では ID 様式によるゲノム再編成は現在まで報告されていない。本研究では、塩基配列解析と染色体解析を織り交ぜ、ヒトと同程度のゲノムサイズを持つヌタウナギがはたして本当に ID を行っているのか詳細に追求し、その全容の解明を目的とした。また、本研究によって導かれる結果は、ヌタウナギの染色体放出の抜本的な理解に留まらず、真核生物ゲノム防御機構の進化を考えるうえでも貴重なデータとなる。

### 3. 研究の方法

科研費申請時には、まず検出されている ID 領域の大きさの違いで区切った塩基配列のデータベースをいくつか構築した後、以下の 3 つの作業目標を設けた; 1. 大きさに分けた幾つかの配

列について、その領域が本当にEゲノムに相当するのかを、S、G両ゲノムDNAを鋳型にしたPCRで検証し、またその領域の染色体上の位置をFISH解析で特定する、2.塩基配列を詳細に解析し、遺伝子等の分布や既知の転位因子との相同性等を解析する、3.個体によってID領域に差がないのかを、上記1の工程で確認し、必要なら関連領域についてのリシーケンシングをNGSによって行い、データの信憑性を検証する。

初年度は、まず作業目標1に沿って、SゲノムにおいてEゲノムの上下流が再結合したと考えられるS-(Sゲノム由来)scaffoldとそれに対応するGゲノムの配列をランダムに抽出し比較解析した結果、再結合の様式は概ね3通りに分類された。また同時に再結合が確認された複数のG-(Gゲノム由来)scaffold配列を基にEゲノムの上・下流を増幅する複数のPCR解析を行った結果、GゲノムからだけでなくSゲノムDNAからも増幅されるケースがあったことから、ID領域に細胞間、個体間での違いが存在する可能性が強く示唆された。また、実際に複数個体のゲノムDNAを用いた検証実験でもID領域の個体差が明確になった。そこで個々のID領域についてのPCRによる検証は時間と手間が掛かり多くを処理することが難しいこと、さらに解析に用いる領域がユニークでない場合、解析結果が安定しないこと等の問題点が明らかとなった。そこで上記の作業目標1の途中であったが、3に掲げたリシーケンシングを複数個体、複数組織で行うこととしたが、予算に限度もあり、初年度はまずは同一個体由来のSゲノムとGゲノムの2検体についてのみ受託解析にリシーケンスを依頼した。

一方、本研究課題について研究支援による研究解析体制の充実を目的に、2020年度の文部科学省科学研究費助成事業「先進ゲノム支援(詳細後述)」へ申請し、採択された。そのため、先進ゲノム支援チームの助言に従い、前年度のリシーケンス・データの情報解析を支援チームに依頼、また今後の研究遂行手順について協議を重ねた結果、新規に抽出し直した個体のSゲノムとGゲノムの1個体のLong-readによるドラフト・ゲノム再構築を行い、リシーケンス・データも同様に新規に抽出し直した別個体のDNA試料から行うこととした。新規採集した成熟個体の組織を国立遺伝学研究所の支援チームに送り、その後1個体分(血液、精巣)のPacBioを用いたLong-readデータ、2個体分のIlluminaを用いたshort-readデータが東京工業大学の支援チームへ送られ、新たなドラフト・ゲノム再構築とEゲノム推定並びにID領域の検出が行われた。*k*-mer頻度解析を踏まえたドラフト・ゲノム再構築によって2個体から得たゲノムサイズは、当初NGSデータからの概算値とほぼ等しいが、アセンブリの精度は著しく向上した。一方ID領域と末端切除が推定された領域は8800と400あまりとなり、当初の値6800カ所を越えた推定値となったが、候補領域を精査すれば、領域数は1/10程度になると見積もられた。しかしID領域数が1/10程度になるとしても数百と膨大な数になることから、新たにHi-Cによる解析を導入して染色体レベルでのゲノム配列の再構築が必要と考え、Hi-C解析を新規に1個体分(血液、精巣)実施して、この解析データを東京工業大学の支援チームから受け取った。

#### 4. 研究成果

- (1) ゲノムシーケンス結果は右並びに右下の表のような値となり、*k*-mer頻度解析を踏まえたドラフト・ゲノム再構築によって2個体から得たゲノムサイズは、Sゲノムはともに1.79Gbp、Gゲノムはそれぞれ、2.06Gbpと2.09Gbpとなり、Eゲノムは270~300Mbpと推定された。これらの値は当初NGSデータからの概算値(Sゲノムは1.77Gbp、Gゲノムは、2.09Gbp)とほぼ等しいが、推定されたEゲノムサイズは320Mから270~300Mと若干小さくなり、アセンブリの精度は、下表のように著しく向上した。また、Eゲノムの2/3(約200Mbp)は縦列型反復配列で占められるが残りの領域(約70Mbp)のうちの31Mbpは、2900(推定値)あまりの遺伝子配列と見積もられた。更に

PacBio ロングリード (CLR)

個体	組織	合計長 (Gbp)	平均長 (kbp)	最大長 (kbp)	N50長 (kbp)
Eb1	germ	299.6	19.6	356.2	30.6
Eb1	soma	328.9	21.5	338.6	33.3

N50: 加重中央値

Illumina paired-ends

個体	組織	合計長 (Gbp)	平均長 (bp)
Eb1	germ	198.3	137
	soma	222.7	140
Eb2	germ	89.3	137
	soma	91.8	140
TT_210517_Eb2	germ	32.8	131
	soma	31.3	133
TT_210517_Eb6	germ	39.5	131
	soma	36.6	132
Toho_190704	germ	68.9	142
	soma	56.8	142

#### アセンブリ統計値

Assembly	Total (bp)	Scaffold N50 (bp)	Contig N50 (bp)	Gap rate (%)	備考
今回のgermドラフトゲノム	2,072,412,291	2,170,495	2,142,373	0.00	
以前のgermドラフトゲノム	1,603,613,555	327,774	5,766	25.25	Illuminaリードのみで構築
今回のsomaドラフトゲノム	1,968,509,487	3,120,219	3,087,097	0.00	
以前のsomaドラフトゲノム	1,714,039,509	256,085	8,339	13.72	Illuminaリードのみで構築
Riken Eburgeri_3.2	2,608,303,295	2,692,996	7,983	44.97	理研発表の公開データ
Sea lamprey Pmar_germline_1.0	1,130,410,935	11,932,523	159,182	7.72	ヤツメウナギ公開データ (Smith et al. 2018)
Sea lamprey kPetMar1.pri	1,089,050,413	12,997,950	2,540,784	1.38	ヤツメウナギ公開データ (VGP)

### 眼珠の欠失割合量(bp)

Element	先行研究の推定値 (germ)	先行研究の推定値 (soma)	新germドラフトゲノム	新somaドラフトゲノム	旧germドラフトゲノム	旧somaドラフトゲノム
EEEb1	176,000,000	12,160	110,043	3,371	226	2,244
EEEb2	11,115,000	NA	601,798	14,963	1,243	609
EEEb3	20,972,000	14,338,000	343,280	983	6,328	764
EEEb4	32,495,000	2,144,000	397,651	21,882	12,528	17,550
EEEb5	2,581,000	17,980	198,525	793	504	745
EEEb6	28,000,000	1,764	711,099	292	620	40
EEEb7	NA	NA	385,054	562	345	560
EEEb8	NA	NA	570,508	164	1,320	387
EEEb9	NA	NA	344,419	4,229	13,177	4,339
EEEb10	NA	NA	711,288	688,874	26,851	80,613
EEEo1	8,319,000	3,982,500	12,372,540	11,936,927	11,365,151	11,641,157
EEEo2	714,000	546,000	61,313	69,959	10,147	20,003
EPPa1	14,940,000	2,905,000	313,452	377	849	208
5SrDNA_germ	NA	NA	26,148	6,267	871	3,202
5SrDNA_soma	NA	NA	27,951	6,948	1,540	3,871

機知の高頻度反復配列の両ゲノムにおけるアノテーション(コピー数の算出)も、上記の表のように大幅に改善された。一方 ID 領域と末端切除が推定された領域は 8800 と 400 あまりとなり、当初の値 6800 ヲ所を越えた推定値となった。しかし、候補領域を精査すれば、領域数は 1/10 程度になる可能性が示唆された。

- (2) 染色体レベルでのゲノム配列の再構築のため、Hi-C による解析を導入してライブラリの作成を試みた結果(右表)、最終年度である 2023 年度、S・G 両ゲノム 1 個体について極めて高い精度のドラフト・ゲノムが完成した。こ

### Hi-C (Omni-C) リードstats

サンプル	リード数	合計長	平均長
Eb1 germ (testis)	982,396,142	122,212,532,127	124
Eb1 soma (blood)	1,034,226,530	132,346,069,286	128

cutadaptによりトリム処理後の値を表示

### Chromosome-scale scaffolds (>10 Mb; n=18)

Tissue	Assembly	# scaffolds	Total (bp)	Scaffold max (bp)	Scaffold N50 (bp)	Scaffold L50 (#)	# contigs	Contig N50 (bp)	Contig L50 (#)	Gap rate (%)
Germ	YaHS (corrected)	18	1,938,529,808	170,147,799	115,442,621	7	1,507	2,129,054	276	0.04
	YaHS	18	1,912,937,206	166,262,877	113,575,749	7	1,432	2,147,478	268	0.02
Soma	YaHS (corrected)	18	1,941,446,575	168,007,596	115,781,728	7	1,075	2,997,688	206	0.03
	YaHS	21	1,913,339,421	162,718,173	115,609,856	7	1,032	3,070,454	200	0.01

\*correctd\*: マニュアル修正後

### Short scaffolds (≤ 10 Mb)

Tissue	Assembly	# scaffolds	Total (bp)	Scaffold max (bp)	Scaffold N50 (bp)	Scaffold L50 (#)	# contigs	Contig N50 (bp)	Contig L50 (#)	Gap rate (%)
Germ	YaHS (corrected)	567	134,593,300	2,764,804	384,442	77	661	291,848	143	0.03
	YaHS	617	159,754,989	2,762,704	402,883	93	720	315,266	161	0.01
Soma	YaHS (corrected)	130	27,541,944	674,207	296,154	32	131	296,000	32	0.00
	YaHS	145	55,373,466	7,482,352	600,633	19	160	519,670	28	0.01

のドラフト・ゲノムを用いた最新の解析結果によると、ゲノムサイズは、S ゲノムで 1.79Gbp、G ゲノムでは 2.07Gbp となり、E ゲノムは 280Mbp と推定された。また、E ゲノムの約 200Mbp は縦列型反復配列で占められ、残りの約 80Mbp が 1700 (推定値) あまりの遺伝子配列を含むユニーク配列からなることが示された。更に染色体レベルとしての scaffold が 18 本得られ(上記 2 つの表)、これは S ゲノムについては全配列が染色体にマッピングできたことを意味している。また S・G 両ゲノムの染色体レベルの 18 本の scaffold 配列の比較から、末端切除も含んだ ID 領域の数は 100 以下(この領域内にある遺伝子数は 56) と見積もられた。その後のより詳細な解析の結果、ID 領域の数は 2 番染色体に 4 から 5 力所と集中し、大きな ID 領域は 2 番、5 番、6 番、14 番染色体に見られ、5 番、6 番、14 番染色体の ID 領域は染色体末端であること、他に小さな ID 領域が 1 番、6 番、9 番に検出された。現在 2 つの ID 領域に注目し、ID 領域内や上・下流の近傍の塩基配列や、G ゲノムの ID 領域内の E ゲノム配列の検出等の詳細な解析を FISH 解析と織り交ぜ継続している。今後、得られたデータを基に、ヌタウナギの染色体放出における内部欠失型のゲノム再編成の実態や機序、意義などの解明に繋げ公表していく。

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計0件

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	後藤 友二  (GOTO Yuji)  (70362522)	東邦大学・理学部・教授    (32661)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関