

2024年3月6日

本件の取り扱いについては、下記の解禁時間以降でお願い申し上げます。

TV・ラジオ・WEB … 日本時間 2024年3月7日(木)午後9時

新聞 … 日本時間 2024年3月8日(金)朝刊

## サンマのゲノム情報を読み取り公開 ～ほぼ未開だったサンマの科学的研究の土台として～

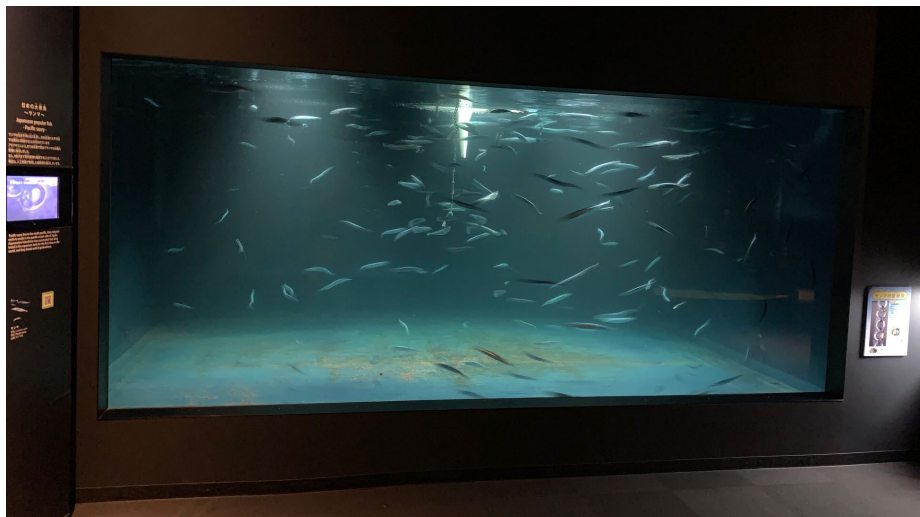
### ■ 概要

「秋の味覚」として日本人に古くから親しまれてきたサンマは、近年、漁獲量が減少し、価格が高騰したことで大きな話題となりました。これまで、サンマは漁獲量が比較的安定しており、魚価も安かったことから、養殖技術は確立されてきませんでした。このような背景から、身近な魚であるにもかかわらず、サンマの生物学的研究はあまりなされていませんでした。さらに、分子生物学研究の基本となる全ゲノム配列など DNA 情報が整備されていなかったことも、サンマについての研究が進んでいない大きな要因です。

情報・システム研究機構 国立遺伝学研究所分子生命史研究室の工樂樹洋教授(理化学研究所生命機能科学研究センター客員研究員)、北里大学の福田和也助教、理化学研究所生命機能科学研究センターの門田満隆技師、アクアマリンふくしまの山内信弥首席技師らからなる研究チームは、サンマ(図1)の全ゲノム情報を読み取り、多様な分子生物学研究のための情報基盤を整えました。

本研究では、世界で唯一サンマの継代飼育および展示を成功させた水族館であるアクアマリンふくしまにて陸上水槽内で人工ふ化し養成したサンマ成魚を用いました。まず、一分子リアルタイム(SMRT)塩基読み取り技術<sup>(1)</sup>による高精度な長鎖配列(ロングリード)解析により、このサンマの DNA 断片情報を取得しました。そして、染色体立体配座捕捉法(Hi-C)<sup>(2)</sup>により、数千万塩基にもなる染色体の DNA 配列を再構築しました。今後、他生物と配列を比較することで、サンマの分子生物学的特徴が明らかになると期待されます。

2022年5月に着手した本研究は、高い専門性をもつ技術者や研究者が、分野横断的に協力することで、短期間で成果を出すことができました。近年の漁獲量減少と価格高騰により大きな注目を浴びているサンマについて、その学術研究を速やかに推進するために、研究論文の発表に先立って、得られたゲノム情報はすでにインターネット上で広く公開しています。これに付随した論文は、日本で育まれてきた国際学術雑誌「DNA Research」に2024年3月7日(日本時間)に掲載されます。



## ■ 成果掲載誌

本研究成果は、国際科学雑誌「DNA Research」に 2024 年 3 月 7 日に掲載されます。

論文タイトル: Chromosomal DNA sequences of the Pacific saury genome: versatile resources for fishery science and comparative biology

(サンマゲノムの染色体規模の DNA 配列: 水産科学と比較生物学のための有用リソース)

著者: Mana Sato, Kazuya Fukuda, Mitsutaka Kadota, Hatsune Makino-Itou, Kaori Tatsumi, Shinya Yamauchi, Shigehiro Kuraku

(佐藤茉菜、福田和也、門田満隆、伊藤初音、辰見香織、山内信弥、工樂樹洋)

## ■ 研究の詳細

### ● 研究の背景

硬骨魚類<sup>(3)</sup>のダツ目<sup>(4)</sup>に属するサンマ (学名 *Cololabis saira*) は、日本の秋の食卓の象徴でありながら、近年の漁獲量の減少により価格の高騰が著しい水産物です。サンマは日本人にとって馴染み深い魚のひとつでありながら、サンマについての学術研究は盛んに行われてきたとは決していえません。そのことは、インターネット上の DNA 配列の国際公共データベースにおけるサンマの塩基配列登録数が、モデル生物であるメダカでは 133 万件に上るのに対して、100 件以下とごく少数である(2023 年 12 月現在)ことから明らかです。塩基配列登録数は、種内集団の地域差による遺伝的多様性に加えて、性別や体の異なる部位、そして発生段階による遺伝子発現プロファイリングといった多様な分子生物学的解析がどのくらい盛んに行われているかを表す、わかり易い指標なのです。この状況を打破する手立てとして、DNA 配列情報の 1 セットである全ゲノム配列を取得することが重要です。サンマのゲノム配列を得ることにより、外洋でのサンマの集団構造と遺伝的多様性の把握、そして、他の生物種との関係や集団の進化的な来歴が分かるだけでなく、種を特徴づける形質の成り立ちを調べるための基礎情報を得ることができます。

本研究に従事した研究者は、決して従来からサンマを研究してきた専門家ではありませんが、近年の漁獲量減少に鑑み、その学術研究を促進することが喫緊の課題であることを感じチームを結成しました。まず、世界で唯一サンマを長期的に継代飼育し展示してきた水族館アクアマリンふくしまの試料(図1)を用い、多様な生物種での実績がある sQuantGenome プロトコル<sup>(5)</sup>および iconHi-C (アイコニック) プロトコル<sup>(6)</sup>を活用することで、全ゲノム情報の効率的な取得が可能であると考え、研究に着手しました。

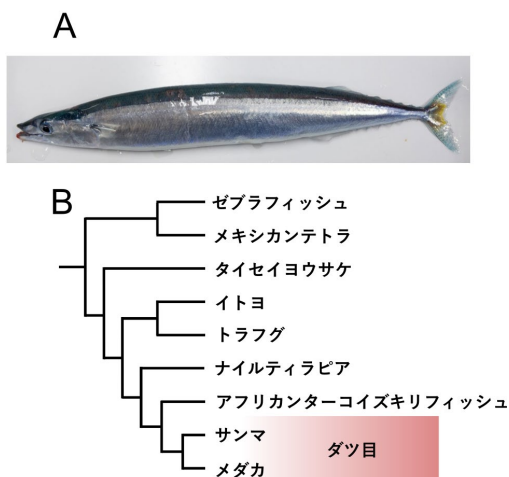


図 1: サンマ

A. 本研究で使用した飼育下のサンマ。アクアマリンふくしまにおいて人工ふ化し養成した成魚。

B. サンマと他の硬骨魚類<sup>(3)</sup>との系統関係。サンマはサヨリなどとともにダツ目<sup>(4)</sup>に属する。分子系統解析に基づいて、同じくダツ目に含まれるメダカとは約 7000 万年前に分岐したと推定されている。

## ● 本研究の成果

はじめに、全ゲノム解析の事前情報として、総塩基数を精査するためのゲノムサイズの測定を行いました。生きた細胞の核染色に基づく従来法ではなく、研究チームが独自に最適化した定量 PCR に基づく手法 (sQuantGenome プロトコル<sup>(5)</sup>) による測定の結果、サンマは約 11.7 億塩基 (1.17 ギガベース (Gb)) のゲノムサイズとわかり、メダカ (0.75 Gb) やゼブラフィッシュ (1.4 Gb) などの硬骨魚類として想定されるゲノムサイズであることがわかりました。次に、サンマの全ゲノム配列情報を新規に読み取るため、研究チームは、DNA 読み取り手法として 99% 以上とされる高い精度で約 1.5~2 万塩基の配列を安定して取得できる一分子リアルタイム (SMRT) 技術<sup>(1)</sup> を採用しました。取得した配列 (HiFi リード) はゲノムサイズの約 30 倍の塩基数に及ぶもので、いわゆるアセンブリと呼ばれる繋ぎ合わせなどのステップ (図2) により、冗長性が殆どなくゲノム全体を網羅した DNA 配列セットに整えることができました。この繋ぎ合わせのための計算は、「パソコン」と通常呼ばれるような廉価型の計算機 (CPU: Intel Core i9、メモリ: 128GB) において 1 日以内に済むような規模でした。このことは、ノウハウさえあれば、さまざまな機関や立場でゲノム解析が進められることを示しています。

すべての真核生物においてゲノムに含まれる DNA は染色体に納められています。今回、サンマについて、得られた上記の DNA 配列 (コンティグ) の多くは、染色体としては不十分な長さのものでした。そこで、それらの配列をコンピュータ上でさらに長くつなぎ合わせる Hi-C<sup>(2)</sup> スキャフォールディングという手法を用いました (図2)。

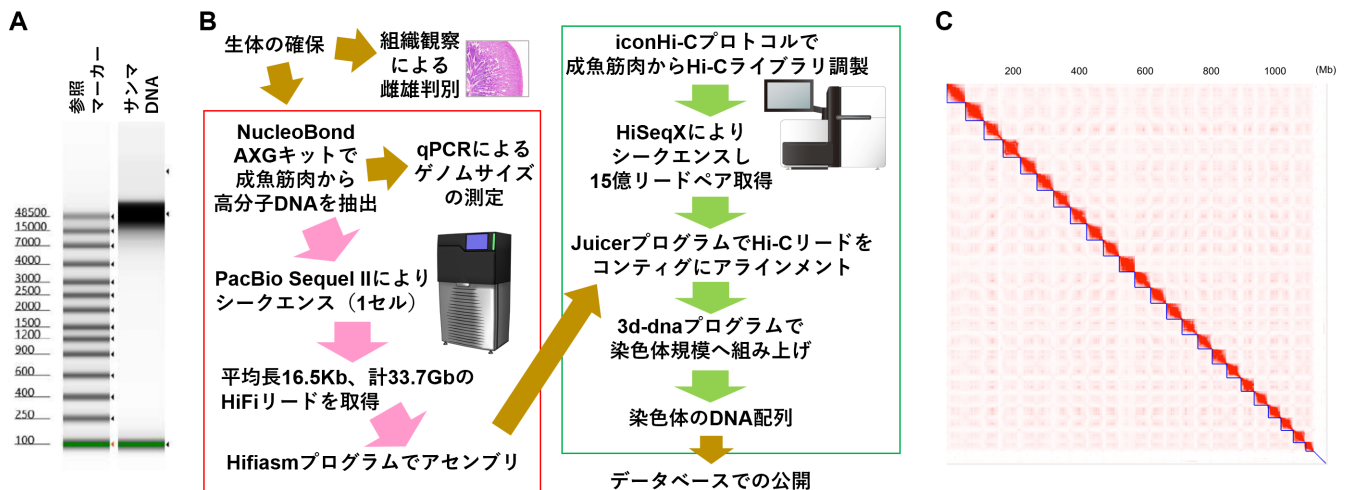


図2: サンマの全ゲノム DNA 配列構築

- A. サンマの組織から抽出した DNA 分子のサイズチェック結果。左に添えた数字は DNA 長のスケール (単位は塩基対)。
- B. 試料の確保から配列情報の仕上げまでのおもな手順。赤で囲んだ左側のロングリード取得に、緑で囲んだ Hi-C データ取得が加わってはじめて染色体規模の DNA 配列情報を構築できることが多い。
- C. サンマのクロマチン相互作用マトリクス。Hi-C データに基づいて作成し、染色体規模へのゲノム配列の繋ぎ合わせに利用した。対角線上の赤い正方形それぞれが 1 本の染色体に対応している。24 個の正方形が認められるため、核型は  $2n=48$  であると推定される。

この手法は、細胞の核の中に存在するクロマチン、すなわち、DNA とタンパク質の複合体において、別々の染色体の間に比べて同じ染色体内で DNA がより高い頻度で近接して相互作用するという現象を利用して染色体構成を推測するものです。この Hi-C スキャフォールディングの結果、核ゲノムを構成する 24 本の染色体の

DNA 配列情報を組み上げることに成功しました。この手順で使用したサンマの Hi-C データは、研究チームにおいて最適化した実験手法(iconHi-C プロトコル<sup>(6)</sup>)により比較的安価に取得でき、また、Hi-C データを用いたコンピュータでの計算は、これまで他生物でノウハウを蓄積してきたおかげで極めてスムーズに進めることができました。

研究チームは、さまざまな客観的な指標に基づいて、得られた全ゲノム配列情報を精査しました。従来から、種間で広く共通に保持される遺伝子の網羅度が高いことや配列がより長いことを善しとする風潮がありましたが、全体的に高い精度での読み取りが可能になった現在では遺伝子の網羅度は飽和しがちであるなどの理由で、むしろ、対象とする生物のゲノムサイズや核型、つまり染色体構成をどれだけ忠実に反映しているか、そして、DNA 配列をどれだけもれなく収録できているかが究極的に大事とされています。

今回取得したサンマのゲノム情報は、これらの指標を含めて評価した結果、長年の努力が注ぎ込まれていると考えられる伝統的な実験動物のゲノム情報と比較しても、遜色がない、あるいは、それを凌ぐほどの完成度を示していました(図3)。

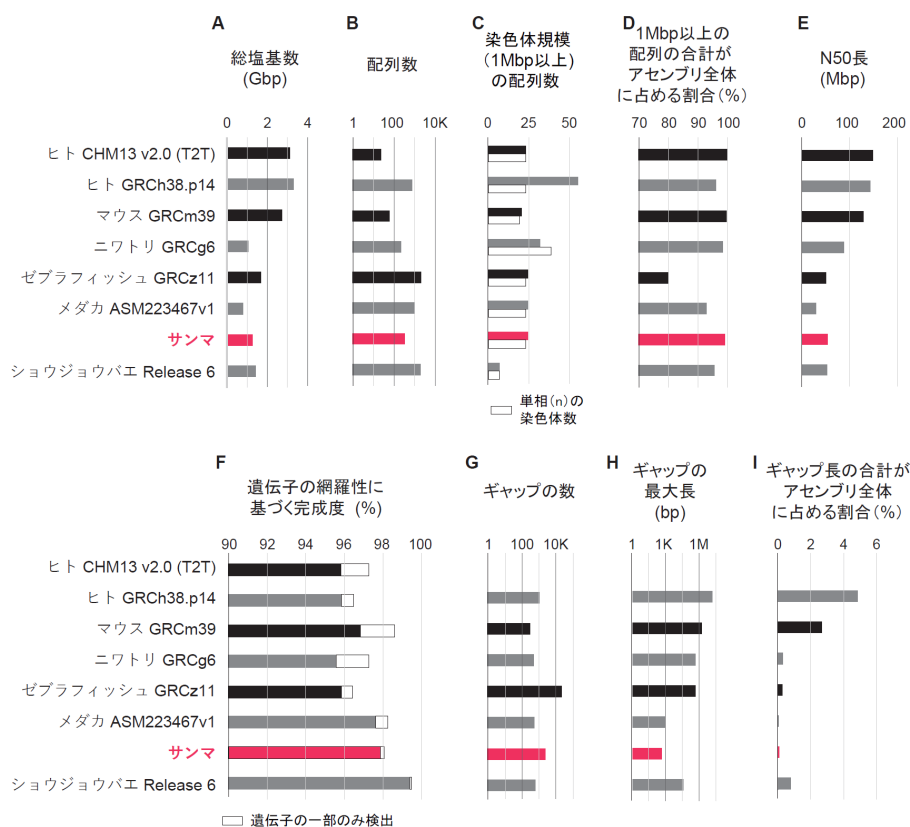


図3 サンマの全ゲノム DNA 配列情報の他生物との完成度の比較  
ゲノム情報の完成度は様々な指標を用いて評価されるが(本文を参照)、取得したサンマの情報は、核型を再現できていること(C)、ほとんどのゲノム領域を染色体配列に収録できていること(D)、そして配列が未決定な領域が極少であること(I)から、伝統的な実験動物について

広く利用されている情報の完成度に匹敵する、あるいは、それを凌ぐものであると考えられる。一方、ヒトで 2022 年に完全解読(図中「T2T」)がついに達成されたことに次いで、ゼブラフィッシュやメダカなどでも、完全解読を目指した努力が続いており、今回得たサンマのゲノム情報についても、さらなる改善の余地は大きいともいえる。1Gbp は 10 億塩基対、1Mbp は 100 万塩基対に対応する。

基盤情報の早期公開を目標として進めたことから、本研究では、基本的なゲノム構造の分析に絞って取得したゲノム配列の解析を行いました。数百～数千塩基からなる繰り返し配列からなる反復配列の分量をメダカと比較したところ、サンマにおけるその種類の全体的な割合はメダカと同程度であるものの、絶対的な塩基数が大きく異なっており、この差異がゲノムサイズを変動させた大きな要因であったことが明らかになりました(図4)。一方で、染色体規模の配列の比較により、硬骨魚類のダツ目において、サンマとメダカの系統が約 7000 万年

前に分岐した後に起きた染色体間の再構成は稀であり、サンマとメダカの系統間では、染色体の構成は大部分維持されてきたことがわかりました(図5)。

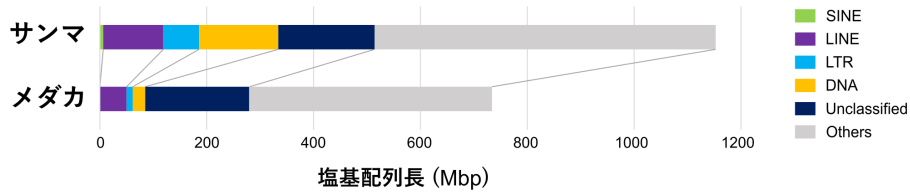


図4: ゲノム配列中の反復配列の種類と量

他の脊椎動物のゲノム解析から知られている反復配列の種類 (SINE や LINE など) のそれぞれについて、サンマとメダカの全ゲノム配列中の量を図示。これら2種ともにゲノム全体の約半分弱が反復配列で占められていることに加えて、サンマでは、それぞれの種類の絶対量がメダカに対して大幅に増加していることがわかる。

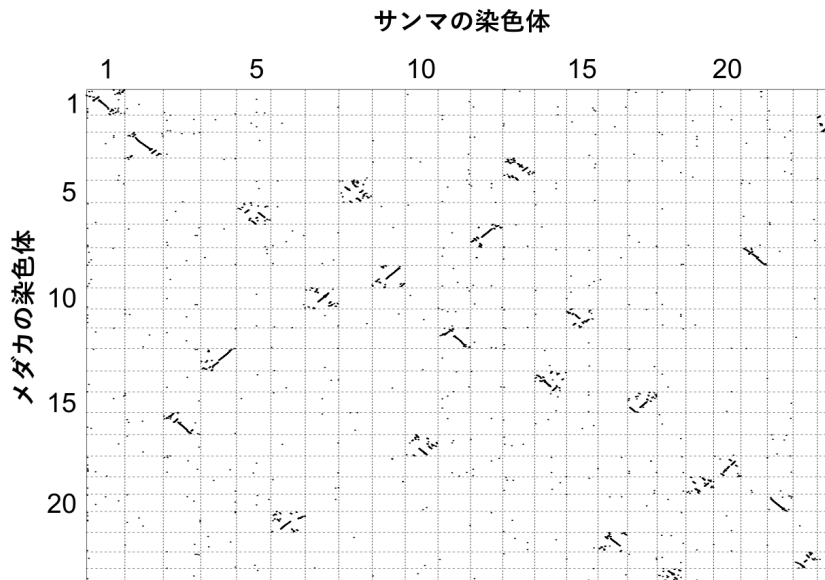


図5: サンマの染色体のDNA配列のメダカとの比較

横軸 (サンマ) と縦軸 (メダカ) は、第1番染色体の端から染色体を並べた際の累積塩基数を表し、格子の内部の斜めの線が、そのDNA配列の類似性が高い部位を表す。サンマと同様にメダカの核型も  $2n=48$  である。広い領域に渡る類似部位は種間で1対1に対応し、それが染色体を単位に生じていることから、共通の祖先に存在した染色体が若干の変更を受けながらもそれぞれの種へ引き継がれてきたことが分かる。

構築したサンマの全ゲノム配列情報が、さまざまなゲノム領域や個々の遺伝子の解析を行う目的に足る品質のものであることを例証するために、研究チームは、水チャンネルとして機能するアクアポリンタンパク質をコードする遺伝子群を調べました。サンマのゲノムに含まれるアクアポリン遺伝子をすべて検出するとともに、対応するメダカ遺伝子と比較した結果、サンマのゲノム情報には、メダカと同数のアクアポリン遺伝子が含まれていて、それらのエキソンの数にも目立った違いはなく、イントロンの大まかな長さも共通していることが示されました。また、研究チームがサンマのゲノム全体を対象として推定した遺伝子構造は、NCBI データベースによっ

て独立に行われた推定と近いこともデータの品質を裏付けられています。

今回取得したサンマの全ゲノム情報とそれに基づく遺伝子の情報(上述)は、論文出版前から、DDBJ (DNA Data Bank of Japan) および NCBI にて公開しています(図6)。これらの情報へのリンクは、研究チームにおいて情報解析を主導する国立遺伝学研究所の分子生命史研究室が運営するウェブサイト (<https://treethinkers.nig.ac.jp/saira/>) に集積されています。水産資源としてのサンマの生態や繁殖についての研究に加え、魚類のさまざまな系統を比較する分子レベルの解析に利用できるほか、DNA 情報についての教育のための教材としても活用することが可能です。

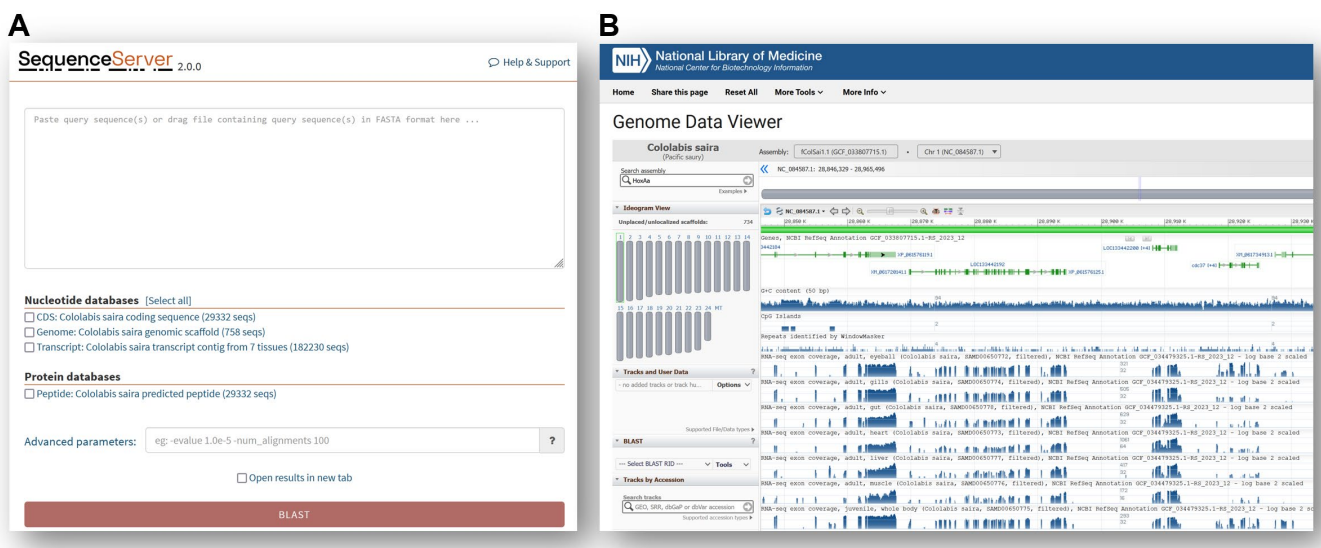


図6 公開されているサンマの全ゲノム情報

A. サンマの遺伝子配列検索用ウェブサイト。国立遺伝学研究所分子生命史研究室において運営している (<https://treethinkers.nig.ac.jp/saira/>)。

B. サンマの第1番染色体の中ほどの領域。NCBIにおいて近年整備されたビューワー Genome Data Viewer を通して、ゲノム配列上の遺伝子構造や GC 含量や反復配列の構成、そして転写産物の組織ごとの量などを総覧することが可能である

([https://www.ncbi.nlm.nih.gov/datasets/genome/GCF\\_033807715.1/](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_033807715.1/))。

## ● 今後の期待

本研究で得られたサンマのゲノム情報は、外洋でのサンマの集団構造や遺伝的多様性の把握、そして、他の生物種との関係や集団の進化的な来歴だけでなく、サンマの種を特徴づける形質の成り立ちを知るために特定の遺伝子の機能を調べる際の基礎情報になります。

海外に目を向けると、分類群や地域の生物相を対象とした多様な生物のゲノム情報の体系的な取得が、Earth BioGenome プロジェクト (EBP)<sup>(7)</sup>により世界規模で精力的に進められています。この動きに対して、日本から直接協力しているのは、国立遺伝学研究所 分子生命史研究室とかずさ DNA 研究所が率いるプロジェクトにほぼ限られます。国内の情勢を盛り立てるため、サンマをはじめ日本の食を彩る生物のゲノム情報を集積する取り組みが、国立遺伝学研究所とかずさ DNA 研究所を中心に Washoku Biogenome コンソーシアム

(8)として動き出しています。

今回の成果は、サンマの飼育繁殖を長年手掛けてきた水族館のオンリーワンのリソースに対して、各機関の研究者が意気投合して他の生物を対象に培ってきた解析技術を持ち寄ることで可能になったものです。今回の発表は、塩基配列を読み取ったに過ぎず、あくまで新たな研究基盤の一部が整ったことの報告であり、読み取った配列情報の「解読」のための生物学的な研究としてはまだ出発点に過ぎません。サンマのように、エビデンスに基づく科学的な研究のための分子生物学的情報整備がままならず、それでいて日本で取り組む意義が大きいと考えられる他の多数の生物種に対しても、体系的な情報整備が進むことが期待されます。

## ■ 用語解説

### (1) 一分子リアルタイム(SMRT)塩基読み取り技術

並列化された一分子 DNA シーケンスの手法の一つであり、Pacific Biosciences of California 社(アメリカ)から発売されているシーケンサープラットフォーム上で動作する。塩基読み取りの際に DNA 鎖の増幅を伴わず、シーケンスの進行に伴う精度の低下などの影響を受けないため、数万塩基という長鎖配列の読み取り(ロングリード)が可能となった。しかし、塩基読み取りにおける高めのエラー率がネックとなり、精度の面では短鎖読み取り(ショートリード)のプラットフォームにはひけをとっていた。今回使用した同社の Sequel IIe シーケンサーでは、同じ DNA 鎖を複数回読み取る(HiFi リード)ことで平均 99.9%以上の精度(HiFi リード)を実現しており、さらに先行機と比べて大幅に費用が削減された。多量の反復配列を含むなどの複雑性の高いゲノムについて、完成度の高い情報読み取りへの道が拓かれたうえ、HiFi リードを生成する計算がシーケンサー内部ですで行われるため、その出力結果を用いた計算は比較的小規模で済むという利点も大きい。

### (2) Hi-C

3C 法(chromosome conformation capture)を応用した方法で、クロマチン相互作用によって細胞の核内で空間的に近接しているゲノム領域を検出する。空間的に近接するゲノム領域は、同じ染色体上の近くに存在することが多い。この特性を活かして、断片的なゲノム配列を染色体上の実際の位置関係の順に並べることにより(この過程をスキュフォールディングと呼ぶ)、染色体スケールのゲノム配列につなぎ合わせることができる。

### (3) 硬骨魚類

脊椎動物亜門から四肢動物を除外した動物群を魚類と呼び、3 万を超える生物種を含む巨大な分類群。そのうち、円口類と軟骨魚類を除いた分類群を硬骨魚類と呼ぶ。フグ、マグロ、サケ、コイ、ウナギの仲間など一般によく知られる魚類に加え、実験動物として利用されるメダカやゼブラフィッシュも含まれる。

### (4) ダツ目

硬骨魚類のうち、間舌骨(interhyal)を欠くという特徴を共有する 200 種余りを含む、分子系統学によって支持される分類群の一つ。メダカ亜目とダツ亜目で構成され、前者には小型の淡水魚メダカ類が、後者には外洋で遊泳するサンマ・トビウオ・サヨリなど、水産資源として重要な種が多く含まれる。なお、メダカ亜目は、独立のメダカ目、あるいは、カダヤシ目の一部とする見方もある。

### (5) sQuantGenome プロトコル

ゲノムサイズを測定するための実験手順。ゲノムサイズとは、特定の種の一つの細胞の核に含まれる DNA の総量。通常は単相(n)に相当する量を表示する。sQuantGenome プロトコルは、ゲノムサイズを測定するため

の複数ある実験手順の中でも、生きた個体の細胞を測定に供することが困難なケースに備えて、リアルタイム PCR 法で単一コピー遺伝子領域の増幅を定量することでゲノムサイズの正確な測定を可能にした手法 (<https://doi.org/10.12688/f1000research.136385.1>)。

#### (6) iconHi-C(アイコニック)プロトコル

Hi-C データ取得のための実験手順を指す。Hi-C データ取得には、種によって確保できる生体組織が様々であり、実験条件を画一化しにくいことや、多額になりがちな費用など、難点が多いため、汎用性を重視しつつ、コストダウンを目指したカスタムプロトコルを自前で用意する価値が大きい。研究チームでは、汎用性と廉価性を追求した手順の最適化を行い、その成果を 2020 年に iconHi-C プロトコルとして公表した (<https://doi.org/10.1093/gigascience/giz158>)。このプロトコルは、これまで数十の生物種に適用され、他の水産対象種クサフグ ([https://www.a.u-tokyo.ac.jp/topics/topics\\_20220603-1.html](https://www.a.u-tokyo.ac.jp/topics/topics_20220603-1.html)) に加えて、創薬研究に用いられるマーモセットやカニクイザル ([https://www.amed.go.jp/news/release\\_20200325-03.html](https://www.amed.go.jp/news/release_20200325-03.html))、爬虫類やモリ的一种 ([https://www.riken.jp/press/2018/20180416\\_1/index.html](https://www.riken.jp/press/2018/20180416_1/index.html)) そしてジンベエザメなど ([https://www.nig.ac.jp/nig/ja/2023/10/research-highlights\\_ja/rh20230817b.html](https://www.nig.ac.jp/nig/ja/2023/10/research-highlights_ja/rh20230817b.html)) についてその実績が既に発表されている。

#### (7) Earth BioGenome プロジェクト (EBP)

2018 年からの 10 年間で真核生物の全生物種について高精度な全ゲノム配列情報を集積することを目標として、それに賛同した世界のさまざまな活動を束ねているコンソーシアム。

#### (8) Washoku(和色) BioGenome コンソーシアム

食用種を中心とした日本を象徴する生物種を対象にゲノム情報の集積を目指すコンソーシアム (<https://washoku.kazusa.or.jp/>)。Earth BioGenome Project (EBP) の活動に協力しているかずさ DNA 研究所と国立遺伝学研究所の有志の研究者を中心として進めている。食べ物だけを対象とするわけではないため、敢えて「食」ではなく「色」の字をあてている。

### ■ 研究体制と支援

本研究は、情報・システム研究機構 国立遺伝学研究所 ゲノム・進化研究系 分子生命史研究室 教授 工楽樹洋(くらく しげひろ)(理化学研究所 生命機能科学研究センター 客員研究員)、北里大学 海洋生命科学部 増殖生物学講座 水族増殖学研究室 助教 福田 和也(ふくだ かずや)、公益財団法人ふくしま海洋科学館 (アクアマリンふくしま) 上席技師の山内信弥(やまうち しんや)、理化学研究所生命機能科学研究センター技師の門田満隆(かどた みつたか)らの研究グループにより遂行されました。

本研究は、情報システム研究機構の「戦略的研究プロジェクト(日本を象徴する重要生物種の分子研究と保全を加速させる生命情報基盤の構築)」の支援を受けて行われました。



## ■ 問い合わせ先

<研究内容のうち、DNA 情報や遺伝子の進化に関すること>

- 国立遺伝学研究所 ゲノム・進化研究系 分子生命史研究室  
教授 工樂 樹洋(くらくしげひろ)  
(理化学研究所 生命機能科学研究センター 客員研究員)

<研究内容のうち、サンマの養殖の可能性に関すること>

- 北里大学 海洋生命科学部 増殖生物学講座 水族増殖学研究室  
助教 福田 和也(ふくだ かずや)

<研究内容のうち、サンマの飼育や繁殖に関すること>

- 公益財団法人ふくしま海洋科学館(アクアマリンふくしま)  
上席技師 山内 信弥(やまうち しんや)

<報道担当>

- 国立遺伝学研究所 リサーチ・アドミニストレーター室 広報チーム
- 学校法人北里研究所 総務部広報課
- 理化学研究所 広報室 報道担当
- 環境水族館アクアマリンふくしま 学習企画営業部

※Zoom 会議での取材にも対応できますので、Zoom 会議をご希望の場合には、その旨お知らせください。

### 配付先

文部科学記者会、科学記者会、三島記者クラブ、神戸市政記者クラブ、神戸民間放送記者クラブ、大阪科学・大学記者クラブ、京都大学記者クラブ、いわき市記者クラブ・記者会会員など