頭足類の比較ゲノムプラットフォームの展望

小 倉 淳*

頭足類は発達した眼と非常に高度な脳神経をもち、無脊椎動物の中で最も進化した生物の一つである。古くから発生学・進化学・神経生理学など様々な分野にわたるアプローチがなされており、脊椎動物と形態類似性を持つカメラ眼や神経内分泌系、閉鎖血管系は、頭足類の研究のみならず、脊椎動物の進化プロセスを研究する上で非常に重要な生物種である。しかしこれまで、頭足類ゲノムは解読されていない。この理由としては、頭足類ゲノムは全般的に2-5Gbとヒトを超えるゲノムサイズを持つことが珍しくなく、ゲノムを決定することが難しかったことがあげられる。こうした状況の中、2012年5月に様々な国における頭足類研究者およびゲノム研究者が集まり、頭足類ゲノム決定を推進するコンソーシアム (Cephalopod Sequencing Consortium: CephSeq) を立ち上げた。本稿では、比較ゲノム研究・比較トランスクリプトーム研究の最近の動向をふまえながら、頭足類ゲノムプラットフォームの今後の展開を紹介する。

1. はじめに

生物のゲノムやトランスクリプトームデータは生物学的研究の基礎データである。ゲノムデータは遺伝子の総合的なリストであり、遺伝子発現制御機構を解読するのに必要である²⁸⁾。トランスクリプトームデータは遺伝子の発現プロファイルであり、様々な細胞・組織・発生ステージ・生理的条件で異なり、生体機能を解析するのに必要である^{12, 40)}。ここまでは1個体内における基礎データであるが、ゲノムリシーケンシングなどにより集団内におけるゲノム比較を行うことにより、ゲノムレベルでの集団遺伝学的解析や生態学的解析、あるいはゲノムレベルでの系統解析も可能になる^{34, 41)}。また、トランスクリプトームデータを生物種間で比較することにより、生物種に特異的な機能の解明や発生過程の進化に迫ることができる。

また、ゲノムが明らかになって初めて可能になった研究も数多く存在する。non-coding RNAとよばれる翻訳されない転写産物に関しても最初には哺乳類 (Xist, H19) やショウジョウバエ (BX-G) で発見されたが、それがゲノム解析により様々な生物に保存されていることが確認された¹⁷⁾。実際、ゲノムをもとにしてヒトのすべての遺伝子を明らかにする Encode プロジェクトでは、ゲノムの7-8割もの領域が一度は転写されており、タンパク質コード遺伝子のほかにも non-coding RNAが非常にウエイトを占めること、また、様々な遺伝子に関してもある程度転写され続けているものが多く、常に何らかの発現調節が働いているという

ことが明らかになった。また、Fireと Melloによって発見された microRNAもマイナーな存在ではなく、microRNAの影響範囲や常在性の解明に関してはゲノムの果たした役割は大きい^{11, 43)}。発生や組織特異性に大きな役割を果たすエピジェネティック修飾もゲノムなしにはその全貌解明は不可能であり、最近の Chip-seqや FAIRE-seqなどの技術の発達により、エピジェネティックな変化の起きる場所やその機能が明らかにされつつある^{8, 25)}。転写後に塩基が置換される RNA editingに関しては哺乳類で数多くの研究事例があるが、トランスクリプトームとゲノムの比較研究から、想像されていたよりも幅広く様々な機能に関わっていることが報告されている^{13, 42)}。本稿では、比較生理学分野において役立つと考えられる比較ゲノムプラットフォームの現状と、頭足類ゲノムの意義と展望、頭足類における具体的な比較トランスクリプトーム・比較ゲノム研究の紹介を行う。

2. 次世代シーケンス技術と比較ゲノムプラットフォーム

比較ゲノムでは、様々な生物の持つ固有の特徴を支えるゲノム基盤を比較することが可能になり、特定の表現型を持つ生物と持たない生物のゲノムを比較することで、表現型に関わる情報をゲノム領域から引き出すことが可能になる^{5, 26)}。それではどのような生物のゲノムがこれまでに決定されているのだろうか。

次世代シーケンス技術の発達に従い、近年ゲノムが読まれる生物種は飛躍的に増えている。世界中のゲノムプロジェクトに関しては、Genomes Online Database (<http://www.genomesonline.org>) にまとめられており、2012年12月時点で3,800近いゲノムが決定されていて、進行中のプロジェクトも14,000を超えている。ただし、これらのゲノムの大半はゲノムサイズの小さな細菌や古細菌であり、真核生物は180種ほどしか含まれていない。アノテート (注釈付け) されたゲノム情報は、NCBI (<http://www.ncbi.nlm.nih.gov/genome>)、Ensembl (<http://www.ensembl.org/>) などでもまとめられており、検索やデータマイニングが行えるようになっている。また、新しい試みとして、Genome10K プロジェクト (<http://genome10k.soe.ucsc.edu>) では、Genomic Zooをうたい、数多くの生物のゲノムプロジェクトを進行させている。ただし、この場合も対象は脊椎動物であり、生物の大進化、特に左右相称動物のゲノム構造の変化を探る目的には不十分である。一方、無脊椎動物では、モデル生物として研究が盛んなショウジョウバエや線虫では姉妹種を含むゲノムが決定されており、昆虫ではi5K プロ

*Atsushi OGURA, 徳島大学疾患プロテオゲノム研究センター (〒770-8503 徳島県徳島市蔵本町3-18-15)

Received 15 November 2012, Accepted 28 December 2012

ジェクト (<http://www.arthropodgenomes.org/wiki/i5K>) という5,000種の昆虫ゲノムを決定しようとする動きも存在する。しかし、他の動物門に関しては十分なゲノム情報が蓄積していない。比較ゲノムのデータセットがそろっていないということは、情報にバイアスがあるということである。実際、タンパク質コード遺伝子の遺伝子ファミリーの研究では、ほとんどは脊椎動物での機能解析を基本としているもので、遺伝子多様性の全貌を明らかにしているとは言い難い⁵⁸⁾。比較ゲノムプラットフォームに様々な生物が加われば、より多様な遺伝子機能や未知のドメインなどが見つかってくると考えられ、発生・進化・生理学などを生物横断的に比較していくことが可能になる。

次世代シーケンス技術は454 life sciences社（現在はRoche社）のパイロシーケンス法がきっかけとなり開発されてきた⁵⁹⁾。技術詳細に関しては、専門のレビュー等を参照してもらうとして、基本的なアイデアとしてはスライドやプレート上にシーケンス反応を行う無数のウェルを用意し、そこで順次取り込まれる核酸を高精度スキャナで読み取り、配列データを構築するというものである。基本的には、ショートリードと呼ばれるようにサンガー法などに比べ短い数十bpから数百bpの配列が数千万から数億個を得られるという特徴がある。最近では一度のランで数百Gbpもの配列データを得ることが可能になり、サンガー法とは比べものにならないスループットを持つ。一度に生成可能なデータ量が増えただけでなく、ランニングコストが非常に低下したことも特徴である。ゲノム既知の生物のリシーケンシングであれば、ヒトほどのゲノムサイズであっても数十万円ですみ、1Gbのゲノムサイズを持つ生物の新規ドラフトゲノム決定であっても数百万円ですむようになってきている（2012年時点）。今後はより低価格化がすすみ、医療などにも応用可能な個人ゲノムシーケンシングが10万円以下で可能になるといわれている。

具体的なゲノム解析手法として、既知ゲノム・近縁ゲノムが利用可能な場合は、そのゲノムにサンプルのショートリードを貼付けていくマッピングという手法が用いられる。この方法では、アセンブル（短い配列をつなぎあわせて長い配列を得る作業）を行わなくても、既知のゲノム配列との一致箇所を確認してゲノムを再構成することができ、個体ゲノムや近縁ゲノムの再構築のみならず、疾患モデルや集団遺伝学的モデルにおける塩基多型や染色体構造、遺伝的浮動の変化をとらえる研究などで活用されている。

既知ゲノム・近縁ゲノムが利用しにくい新規ゲノムに関しては、染色体ごとにライブラリを構築してからシーケンスを行う手法と、完全にde novoでショートリードを決定して、インフォマティクスでゲノムを再構築するWhole Genome shotgun (WGS) 手法およびそれらの混合手法がとられる。完全にde novoでショートリードを読むときには、いくつかの異なるライブラリを構築してショートリードを読むというのが現在の技術では一般的である。ひとつは、ペアエンドライブラリ・フラグメントライブラリとよばれるライブラリで、ゲノムDNAを200-800bpほどに分断して、その両端から配列を決定していく手法である。もうひとつは、メイトペアライブラリ・ジャンピングライブラリと呼ばれるライブラリで、ゲノムを2000bp-10000bpほどに分断した後、環状DNAにした上で、そのつなぎ目の配列を決定する。これにより、数千bpはなれた場所にある1

対の配列が決定されるため、ゲノム地図を作成しやすくなる。こうして作られた複数種類のライブラリをもとのゲノムサイズの80倍から200倍程度配列決定することで、実際に堪えうるゲノム配列を構築することが可能になる。

ゲノム決定アルゴリズムに関しては、サンガー法のデータに関しては、塩基配列をのり付けして伸長する手法をもとに計算して行ってきたが(Phrapなど)、ショートリードではのりしろの情報が少なく計算量も膨大なため、de Bruijn Graphとよばれるアルゴリズムを利用したアセンブルに移り変わってきている。このアルゴリズムでは、ショートリードをk-merと呼ばれる短い配列に分割し、隣り合う頂点を結ぶことでde Bruijn Graphとよばれるグラフを作成する。これはオイラーパス問題であり、効率的な解法が存在することが知られており、ゲノムアセンブリに適用されている。こうした完全にde novoでアセンブルする手法は安価で手軽であるが、リピート配列が多かったり、Heterozygosityが高かったりすると高精度なゲノムを決定することがむずかしく、完全にGapのないゲノム情報を構築することは極小のゲノムサイズでない限りほぼ不可能である。ショートリードを用いたドラフトゲノムでは、完全につながった配列が得られる代わりに、コンティグとスキヤフォールドという2種類の配列が得られる。コンティグとは主としてペアエンドライブラリを利用してショートリードをアセンブルして構築したギャップのない配列群のことで、通常はそれほど長い配列は得られない。スキヤフォールドはメイトペアライブラリやジャンピングライブラリと呼ばれるような比較的遠い距離をつなぐ目的で作られたライブラリをもとにコンティグをつないで作製する配列で、その中にはジャンピングライブラリで読めていない中間配列がNとして記されている。ゲノムの精度はプロジェクトごとに多様なので、その品質を評価するとき、一般的にはN50と呼ばれる指標でゲノムの精度を測る。これはすべてのスキヤフォールドを大きい順に並べ、合計値をカウントしていく過程で、全ゲノムサイズの半分に到達するスキヤフォールドサイズをN50とよび、よりギャップの少ない優れたゲノムの場合にはN50が大きくなる。主要なアセンブラーとしては、Assemblathonというアセンブリの正確さなどを競う大会にて好成績を収めたAllpaths-LG、BGIが開発しパンダゲノムや様々なゲノム決定に使われるSOAP-de novo、商用ながら高精度なゲノムを構築することが知られているCLC Genomics work bench、ほかにもABYSSやVelvetなどがよく用いられている¹⁵⁾。

トランスクリプトームに関しては、やはりマッピングベースの発現解析手法と、de novo assembleを行いながら発現頻度をReads per kilo base per million reads (RPKM) と呼ばれる数に置き換えて解析をする手法がある。トランスクリプトームの場合はゲノムと違って異なるSplicing variantsが存在する場合も多いため、多少異なるアルゴリズムが要求されている。現在よく用いられるRNA-seq用のアセンブラーとしてはBROAD instituteが開発したTrinityが挙げられる¹⁶⁾。

今後の次世代技術の将来展望としては第三世代、次々世代と呼ばれる新しい技術に基づくシーケンスが利用可能になってくる。1分子シーケンスや長鎖シーケンスが可能なPacBio、半導体を用いてシーケンスを行うIon Proton、プロテイン・ナノポアを通過する時のイオン電流変化から塩

基配列を決定するナノポアシーケンサなど新しい技術に基づくシーケンサが次々と利用可能になる。現在のショートリードを前提とした解析パイプラインも新技術にあわせて変化していくと考えられる。

3. 頭足類ゲノムコンソーシアムの設立と展望

頭足類とはイカやタコ、オウムガイなどからなる軟体動物に属し、すくなくとも先カンブリア後期から存在する歴史を持ち、現在では700種以上を擁する動物群である^{24, 38)}。成体で1 cmほどのヒメイカから18m 以上にもなる無脊椎動物最大のダイオウイカまで幅広いサイズが存在している。また、無脊椎動物において最も高等な動物ともいわれ、高度に発達したカメラ眼や巨大で分化した脳神経系、ジェット推進を利用した移動手法などを持ち、発生学、進化学、認知科学など様々な分野にわたる興味を惹いてきた¹⁸⁾。また、カメラ眼や神経内分泌系や閉鎖血管系は脊椎動物と相同性を持ち、脊椎動物の進化プロセスを論ずる上で非常に重要な比較対象となっている^{7, 35, 49)}。さらに、体色を自在に変化させることのできる種も存在し、摂食や逃亡に重要な役割を果たしている。これらの体色変化システムは擬態行動・擬態メカニズムの研究のターゲットともなっている。頭足類の生物学的研究はさまざまな分野にわたるが、金字塔としてはイカの巨大軸索の生理学的解析からノーベル賞を受賞した A. Hodgkin と A. Huxley の研究が挙げられ

る²¹⁾。また J.Z. Young らによる頭足類の脳を用いた学習と記憶に関する研究も著名である⁵⁵⁾。また、発生学の分野においても頭足類は貴重な題材を与えてくれる。オウムガイ以外の頭足類では殻を消失し、外套膜を用いた移動や呼吸を獲得している。こうした進化プロセスはほかの軟体動物から急速な選択圧がかかったことが予測され、高度な脳神経系の獲得にいたったと考えられる。

軟体動物の中ではカサガイやアコヤガイ、牡蠣などのゲノムは決まっているが、頭足類ゲノムはまだ決まっていない^{48, 56)}。この理由としては、頭足類のゲノムが他の軟体動物と比べて2～5 Gb と巨大であることが挙げられる(図1)⁵⁴⁾。今後上記に挙げたような様々な分野で頭足類を用いた研究がなされていくと考えられるため、高精度なゲノム情報を整備する必要がある。また、頭足類研究者以外にも、頭足類ゲノムプラットフォームが整備される意義は大きい。頭足類ゲノムの利点の一つは、遺伝子の取得が容易になることである。興味がある遺伝子を探すのも楽になるし、表現型からターゲットとなる遺伝子を探索するのも簡単になる。こうした利点は頭足類の研究者たちだけではなく、比較ゲノムをターゲットとする研究者にも非常に重要なデータとなる。

そこで、筆者らを含む国際共同研究グループは、2012年5月の NESCent Catalysis Group Meetingにて The Cephalopod Sequencing Consortium (CephSeq) を立ち上げた。CephSeq で議論された内容は White paper としてまとめられ

図1 軟体動物の系統とゲノム 頭足類の中でもイカタコの仲間である Coleoid cephalopod でカメラ眼や高度な脳神経系を獲得している。ゲノムサイズは他の軟体動物と比べて高めであり、染色体数も多い。

図2 CephSeq CephSeqには図のような役割がある。

Standards in Genomic Sciencesにて発表している¹⁾。議論の内容としては、頭足類の中でどの生物を決定するかということである。軟体動物の中で、5億年以上前に頭足類は共通祖先から分岐したと考えられている^{2, 22, 24, 30)}。タコイカなどを含むColeoideaとオウムガイのNautiloideaという2系統が現存している。この中で、特定の種をゲノムシーケンシングのターゲットとすることは、多様な研究者が存在する中では難しいが、貴重性、重要性、入手性、系統関係を考慮して初期ターゲットとして9種の頭足類が選定されている。コンソーシアムの役割としては、共通ストラテジーの策定としてどのようなシーケンス手法を用いてゲノムプロジェクトを推進するかなどの方針決定、データシェアリングポリシーとして研究者間の発表前データをどのように取り扱いコミュニケーションを向上させるための方針策定、研究グラントへの応募支援、WIKIなどを通じた情報発信、ミーティングの開催が挙げられる(図2)。コミュニティの活発な研究の促進、データや情報の共有化、共同研究の促進、ゲノムアノテーションへの尽力を目的とした運営委員会が設置されており、初代メンバーとしてはLaure Bonnaud (Muséum National d'Histoire Naturelle, France), C. Titus Brown (Michigan State University, USA), Roger Hanlon (Marine Biological Laboratory, USA), Atsushi Ogura (University of Tokushima, Japan), Clifton Ragsdale (Chicago University, USA), Goujie Zhang (BGI, China) が任にあたっている。

現在ゲノムプロジェクトのターゲットとして考えられている種は下記の通りである(表1)。

マダコ (*Octopus vulgaris*) は、世界中に生息する一般的なタコで、体長は60cmほどである。皮膚には色素細胞が分布しており、周囲の環境に合わせて体色などを変化させ擬態することが可能である。繁殖期は春から初夏で、交尾したメスは長径2.5mmほどの楕円形の卵を数万から数十万個産む。歴史的に頭足類研究の主要な対象生物であり、解剖学、行動学、神経生理学などによく用いられてき

た^{18, 37, 46, 47, 49, 55)}。また、日本では重要な水産資源であり、水産業の観点からも重要性は高い。

カリフォルニア・ツースポット・タコ (*Octopus bimaculoides*) は近年頭足類のモデル生物として北米で用いられるようになってきた種である。成体は20cm程度で、眼に似た模様を持つことで知られる。卵は大きく扱いが容易で発生過程を観察するのに非常によいためよく用いられている¹⁹⁾。

ヒョウモンダコ (*Hapalochlaena maculosa*) はオーストラリアから日本にかけての潮溜まりや珊瑚礁に生息する10-20cmほどのタコである。唾液に猛毒のテトロドトキシンを持つことが知られ、毒性の研究によく用いられる³⁸⁾。

ヨーロッパコウイカ (*Sepia officinalis*) は神経生理学の研究によく用いられてきたと同時に水産業の重要な種目である。複雑な色素胞を持つことで知られ擬態やコミュニケーションに利用している^{3, 10, 18)}。ヨーロッパコウイカもまた卵の扱いが楽で、発生過程もよく調べられている^{4, 6, 28, 36)}。ヨーロッパではエコーエボデボ研究のモデルとして研究が始まっている。

ヤリイカは神経インパルス伝導に関する研究によく用いられる。巨大軸索は神経生理の研究によく用いられている。*Loligo paeleii*が最もよく用いられる種であり、水産業の重要品種であり入手性がよいことで知られる^{14, 33)}。また、海洋生物研究で著名なMarine Biological Laboratory (Woods Hole, MA, USA) 近くで通年採取可能なため、研究者人口が多いという点も重要である。

ミミイカ (*Euprymna scolopes*) は海洋性発光バクテリアの *Vibrio Fischeri* と共生することでよく知られる。共生、クオラムセンシング、生物発光の研究モデルとしてよく用いられる^{20, 22, 26, 30, 31, 38)}。2005年には *Vibrio Fischeri* のゲノムが決まっており、ホストであるダンゴイカのゲノムが決定されればより研究は進展する⁴⁴⁾。

ヒメイカ (*Idiosepius paradoxus*) は2.1Gbと知られる中で頭足類最小のゲノムサイズであり、体長は1-2cmと非常に小さくライフサイクルも短いため研究室内で飼育し

表1 ゲノムプロジェクトの対象となっている頭足類

生物種	推定ゲノムサイズ (Gb)	現在のシーケンスカバー率	分布	ライフスタイル 幼生/成体	コメント
マダコ <i>O. vulgaris</i>	2.5-5	46 ×	全世界	プランクトン/ 底生	脳神経系・行動・水産の古典的モデル
カリフォルニアツースポットダコ <i>O. bimaculoides</i>	3.2	50 ×	カリフォルニア/ メキシコ	底生	発生・行動・水産の最近のモデル
ヒョウモンダコ <i>H. maculosa</i>	4.5	10 ×	インド-太平洋	底生	毒性
ヨーロッパコウイカ <i>S. officinalis</i>	4.5	—	東大西洋-地中海	遊泳-底生	行動・発生・水産の古典的モデル
ヤリイカ <i>L. paeleii</i>	2.7	—	東-北大西洋	遊泳	神経生理・水産のモデル
ミミイカ <i>E. scolopes</i>	3.7	—	ハワイ	遊泳-底生	動物-バクテリア共生モデル
ヒメイカ <i>I. paradoxus</i>	2.1	80 ×	日本	遊泳-底生	発生のモデル, 小さなゲノムサイズ
ヒメイカ <i>I. notoides</i>	—	50 ×	オーストラリア	遊泳-底生	発生のモデル, 小さなゲノムサイズ
ダイオウイカ <i>A. dux</i>	4.5	60 ×	全世界	遊泳	巨大な体長
オウムガイ <i>N. pompilius</i>	2.8-4.2	10 ×	インド-太平洋	遊泳	生きた化石, イカタコのアウトグループ

やすくモデル生物として優れている^{52,54)}。

ダイオウイカ (*Architeuthis dux*) は深海に生息する巨大なイカであるが、世界中に生息するダイオウイカの系統関係は不明である⁵¹⁾。

オウムガイ (*Nautilus pompilius*) は生きた化石と呼ばれるからを持つ頭足類であり、ピンホール眼、外殻、無数の触腕、などの数多くの原始的で祖先型の特徴を持つ。同じ頭足類であるが、タコイカのアウトグループとして貴重であり、タコイカがどのようにしてその特徴的な表現型を獲得してきたかを明らかにする上で非常に重要な生物種である⁴⁵⁾。

4. 頭足類比較ゲノムおよび比較トランスクリプトーム

筆者らは軟体動物5種の眼のトランスクリプトームデータをもとにした比較トランスクリプトーム解析や、次世代シーケンスデータを用いて得た部分ゲノムを利用した頭足類3種の比較ゲノム研究を行っており、最後にこれを紹介したい。

眼の形態には様々な種類がある。カメラ眼、複眼、単眼、ほかにもミラー眼やピンホール眼なども存在する。これらすべての眼は同一の起源をもつといわれている(図3)。眼の形態形成には非常に保存された遺伝子セットが働くことがわかってきたが、眼の形態多様性を担保するような分子メカニズムに関してはほとんどわかっていない。軟体動物は動物にみられる様々な眼の形態を動物門内に保持しているため眼の形態多様化研究モデルとして優れている。そこで、分子レベルで眼の形態多様化メカニズムを探るには、眼の発生過程における遺伝子発現プロファイルを経間比較

する必要がある。

イカタコの仲間は非常に発達したカメラ眼を持ち、その構造は脊椎動物のカメラ眼と非常によく似ていることから古典的に収斂進化の例とされてきた。しかし、同じ軟体動物ではイカタコ以外ではカメラ眼は持たず、異なる眼形態を持っていることから、イカタコが他の軟体動物と分岐してからカメラ眼を発達させる遺伝子およびそのネットワークを進化させてきたのだと考えられる。そこで我々は、5種の軟体動物における遺伝子発現プロファイルを比較解析し、頭足類のカメラ眼に特異的な遺伝子を抽出する解析を行った。用いた軟体動物の遺伝子発現プロファイルは、ヤリイカ、マダコ、ヒメイカ、のカメラ眼を持つ頭足類3種と、カメラ眼を持たない頭足類であるオウムガイ、二枚貝類のホタテガイの2種をアウトグループとして用いた。これらの遺伝子プロファイルの比較解析から、5,707個の遺伝子を頭足類カメラ眼特異的な遺伝子として得た。5,707個の遺伝子はヒメイカの発生過程の眼における遺伝子を用いたマイクロアレイにより発生過程の頭足類のカメラ眼において確かに発現していることを確認した(図4)。次に、分子進化解析により頭足類の系統で特異的にポジティブセクションを受けている遺伝子を探出し、156個の遺伝子を抽出した(図5)。これらの遺伝子は、頭足類において選択圧を受けて構造が変化した可能性のある遺伝子であり、頭足類におけるカメラ眼獲得において重要な役割を果たした可能性が考えられる。そして、5,707個の遺伝子のうち脊椎動物の眼においても発現している遺伝子を選び出し、カメラ眼の獲得に双方の系統で関与した可能性のある遺伝子を1571個抽出した。これらのカメラ眼獲得に寄与したと考えられる遺伝子のうち3つに関しては *in situ hybridization* によ

図3 動物の眼の多様性 軟体動物にはカメラ眼の他にもピンホール眼、複眼など多岐にわたる眼の形態多様性を保持している。

図4 ヒメイカにおける頭足類カメラ眼特異的な遺伝子の発現プロファイル 左)縦軸は遺伝子数で、横軸は発生ステージ。ヒメイカでカメラ眼が形成されるステージ25近辺で遺伝子数が多くなっている。これらがカメラ眼の形成に関わると考えられる。右)ステージ特異的な遺伝子解析のためのベン図。ステージ25だけで発現する遺伝子は163個であった。

る確認も行い、発生過程の optic lobe の outer layer において局在していることを確認した (図6)⁵⁴⁾。

ゲノムを比較することにより、新規遺伝子の同定、リピート領域の評価、遺伝子重複などのゲノム構造の違いを明らかにすることが可能である。種特異的な表現型につながる種特異的なゲノム構造の変化を明らかにするために我々は3種の軟体動物：ヒメイカ、オウムガイ、ホタテガイの部分ゲノムを Partial Genome Shotgun (PGS) 解析した。ゲノムサイズの推定は、部分ゲノムであったため K-mer を用いた推定では誤差が大きいと考えられたため、2x coverage の頻度をもとに、Sequencing theory に準じて、ゲノムサイズと 2x coverage になる期待値をもとに算出した。その結果、ゲノムサイズは、ヒメイカで 2.1Gb であったのに対し、オウムガイは 4.2Gb、ホタテガイは 1.8Gb であると推定された (図7)。この結果は、FACS を用いた実験からの推定値がそれぞれ 2.1Gb、2.8Gb、1.63Gb であったことから、オウムガイ以外では妥当だと考えられた。オウムガイのゲノムサイズ推定値が大きく違う理由としては、表に示す通り、オウムガイだけリード数が少なかったことが原因の一つと考えられる。このことは、この中で最も先進的な脳神経を持つヒメイカは必ずしも大きなゲノムを持つ訳ではないことが示された。一方、FACS でのゲノムサイズ推定に勝る PGS 解析の利点としてはゲノムサイズ以外にも、リピート配列の分類と評価、遺伝子重複などの割合と多重遺伝子族の推定などが行える点である。通常ゲノムの巨大化には、リピート配列の増加かゲノム重複が大きな役

割を果たすと考えられているが、リピート配列の構成や割合を比較解析した結果、リピート配列の占める割合はヒメイカで 9.2% であり、ほかの2種 (オウムガイで 4.0%、ホタテガイで 3.8%) に比べて高いことがわかった。さらに、カサガイとの分岐以降の種特異的な遺伝子重複に関する解析も行った。遺伝子重複がランダムにおこっている場合は、重複遺伝子のアミノ酸同義置換の頻度がなだらかに減衰していくことが知られているが、ヒメイカ、オウムガイ、ホタテガイでは、中央にピークがあることが示され、特定の時期にカサガイに比べて大規模な遺伝子重複がヒメイカおよびオウムガイで発生していたことが示唆された (図8)。カサガイは軟体動物の中でもゲノムサイズが小さいことで知られており、今回調査した生物種ではカサガイとの分岐以降にゲノム重複でゲノムサイズを増やし遺伝子レパートリーを増やしていたことが示唆された。以上の結果はまとめて、Gene 誌に報告済みである³¹⁾。

5. むすび

これまでゲノムプロジェクトが進んでいなかった頭足類において、国際的なコンソーシアムが立ち上がり、様々な頭足類ゲノムプロジェクトが平行して進行している。今後は質の高いゲノムとアノテーションをコミュニティ全体でサポートしながら完成させていくことが必要である。頭足類のゲノムを用いることで、脊椎動物の進化の異なる視点で再評価や、ゲノム構造の変化を幅広い動物間における解

図5 頭足類カメラ眼特異的な遺伝子の dN/dS 比 ポジティブセレクションを受けた遺伝子として、特に高い dN/dS 比を示すものは24個存在した。全体的には非同義置換は抑えられ遺伝子の構造が保たれる傾向にあると考えられる。

- Stage 24におけるin situ hybridization
- Centaurin-gamma homolog
- dN/dS = 2.49
- 網膜光受容体を形成する Optic lobesに発現

図6 Centaurin-gamma homolog の in situ hybridization カメラ眼特異的な遺伝子と推定されたもののうちポジティブセレクションを受けている遺伝子をいくつか in situ で確認した。ステージ24において、網膜光受容体を形成する Optic lobes に発現している。

	Pygmy squid	Nautilus	Scallop
Reads after quality control (reads)	401,050	296,439	310,444
RE proportions (%)	9.15%	4.04%	3.82%
2x coverage frequency	0.0429	0.0165	0.0462
Estimated genome size -elementary probability model	2.08Gb	4.24Gb	1.76Gb
Estimated genome size -experiments	2.1Gb	2.8Gb	1.63Gb*

図7 ヒメイカ (Pygmy squid), オウムガイ (Nautilus), ホタテガイ (Scallops) における Partial Genome Shotgun 解析 Readsは解析に用いた454シーケンサによるリード数, RE proportionはリピート配列のゲノムに占める割合, 2x coverage frequencyは同一遺伝子座をカバーした readsの割合でここからゲノムサイズを elementary probability modelで推定した。下図はリピート配列の種類。

図8 カサガイ-ヒメイカ/オウムガイ/ホタテガイの重複遺伝子の Ks 値から重複年代を推定 重複遺伝子数としては ヒメイカ186個, オウムガイ230個, ホタテガイ506個を用いた。X軸は Ks 値で, Y軸はその頻度 (個数)

析が可能になると考えられ, 頭足類研究者以外への波及効果も高いと考えられる。

6. 謝辞

本稿で紹介した研究は, 吉田真明博士, 滋野修一博士, 瀬々潤博士, 清水健太郎博士らとの共同研究として遂行されたものです。ここに深く感謝申し上げます。最後に, このような寄稿の機会をいただきましたことに, 多くの関係者の方々に感謝いたします。

文 献

- Albertin CB, Bonnaud L, Brown CT, Crookes-Goodson WJ, da Fonseca RR, Cristo CD, Dilkes BP, Edsinger-Gonzales E, Freeman RJ, Hanlon RT, Koenig KM, Lindgren AR, Martindale MQ, Minx P, Moroz LL, Nödl MT, Nyholm SV, Ogura A, Pungor JR, Rosenthal JJC, Schwarz EM, Shigeno S, Strugnell JM, Wollesen T, Zhang G, & Ragsdale CW: Stand. Genomic Sci. 7, 175-188 (2012)
- Allcock AL, Cooke IR, & Strugnell JM: Zool J Linn Soc 161:573-586 (2011)
- Barbosa A, Mathger LM, Chubb C, Florio C, Chiao CC, & Hanlon RT: J Exp Biol 210:1139-1147 (2007)
- Boletzky SV, Erlwein B, & Hoffmann DK: Vieetmilieu-Life and environment 56:191-201 (2006)
- Boussau B, & Daubin V: Trends Ecol Evol 25:224-232 (2010)
- Boycott BB: Proc R Soc Lond B Biol Sci 153:503-534 (1961)
- De Lisa E, Paolucci M, & Di Cosmo A: J Neuroendocrinol 24:275-284 (2012)
- Eissenberg JC, & Shilatifard A: Dev Biol 339:240-249 (2010)
- Ewing B, & Green P: Genome Research 8:186-194 (1998)
- Ferguson GP, & Messenger JB: Proc R Soc Lond B Biol Sci 243:63-67 (1991)
- Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, & Mello CC: Nature, 391, 809-811 (1998)
- Garber M, Grabherr MG, Guttman M, & Trapnell C: Nat Methods 8:469-477 (2011)
- Garrett S, & Rosenthal JJ: Science 335:848-851 (2012)
- Gilbert DL, Adelman WJ, & Arnold JM: Squid as Experimental Animals. New York Plenum Press (1990)

- 15) Gnerre S, MacCallum I, Przybylski D, Ribeiro F, Burton J, Walker B, Sharpe T, Hall G, Shea T, Sykes S, Berlin A, Aird D, Costello M, Daza R, Williams L, Nicol R, Gnirke A, Nusbaum C, Lander ES, & Jaffe DB: *Proc Natl Acad Sci USA* 25:108(4):1513-8 (2010)
- 16) Grabherr, M. G. et al: *Nat Biotechnol* 29, 644–652 (2011)
- 17) Guttman M. et al: *Nature* 477:295-300 (2011)
- 18) Hanlon RT, & Messenger JB: *Cephalopod Behaviour*, Cambridge University Press (1996)
- 19) Hanlon RT, & Forsythe JW: *Advances in the laboratory culture of octopuses for biomedical research*. *Lab Anim Sci* 35:33-40 (1985)
- 20) Hanlon RT, Claes MF, Ashcraft SE, & Dunlap PV: *Biol Bull* 192:364-374 (1997)
- 21) Hodgkin AL, & Huxley AF: *J Physiol*. 117(4):500-544 (1952)
- 22) Koropatnick TA, Engle JT, Apicella MA, Stabb EV, Goldman WE, & McFall-Ngai MJ: *Science* 306:1186-1188 (2004)
- 23) Kröger B, Vinther J, & Fuchs D: *Bioessays* 33:602-613 (2011)
- 24) Kwon SH, & Workman JL: *Bioessays* 33:280-289 (2011)
- 25) Lalueza-Fox C, & Gilbert MT: *Curr Biol* 21:R1002-R1009 (2011)
- 26) Lee PN, Callaerts P, De Couet HG, & Martindale MQ: *Nature* 424:1061-1065 (2003)
- 27) Lemaire J: *Bull. Soc. Zool. France* 95:773-782 (1970)
- 28) Lander ES: *Nature* 470:187-197 (2011)
- 29) Lindgren AR: *Mol Phylogenet Evol* 56:77-90 (2010)
- 30) McFall-Ngai MJ, & Ruby EG: *Science* 254:1491-1494 (1991)
- 31) McFall-Ngai MJ, Nyholm SV, & Castillo MG: *Semin Immunol* 22:48-53 (2010)
- 32) Moreno H, Yu E, Pigino G, Hernandez AI, Kim N, Moreira JE, Sugimori M, & Llinas RR: *Proc Natl Acad Sci USA* 106:5901-5906 (2009)
- 33) Morrell PL, Buckler ES, & Ross-Ibarra J: *Nat Rev Genet* 13:85-96 (2011)
- 34) Muñoz-Chápuli R: *Int J Dev Biol* 55:345-351 (2011)
- 35) Naef A: *Fauna Flora Golfo Napoli* 35:1-357 (1928)
- 36) Nixon M, & Young JZ: *The Brains and Lives of Cephalopods*. Oxford University Press (2003)
- 37) Norman MD: *Cephalopods: A World Guide*. Hackenheim, ConchBooks (2000)
- 38) Nyholm SV, & McFall-Ngai MJ: *Nat Rev Microbiol* 2:632-642 (2004)
- 39) Ozsolak F, & Milos PM: *Nat Rev Genet* 12:87-98 (2011)
- 40) Podlaha O, Riester M, De S, & Michor F: *Evolution of the cancer genome*. *Trends Genet* 28:155-163 (2012)
- 41) Rosenthal JJ, & Seeburg PH: *Neuron* 74:432-439 (2012)
- 42) Ruvkun G: *Nat Med* 14:1041-1045 (2008)
- 43) Ruby EG. et al: *Proc Natl Acad Sci USA* 102:3004-3009 (2005)
- 44) Saunders WB, & Landman NH: *Nautilus: The Biology and Paleobiology of a Living Fossil*. 2nd ed. New York: Springer (2009)
- 45) Shomrat T, Zarrella I, Fiorito G, & Hochner B: *Curr Biol* 18:337-342 (2008)
- 46) Shomrat T, Graindorge N, Bellanger C, Fiorito G, Loewenstein Y, & Hochner B: *Curr Biol* 21:1773-1782 (2011)
- 47) Takeuchi T. et al: *Draft genome of the pearl oyster Pinctada fucata: a platform for understanding bivalve biology*. *DNA Res* 19, 117-130 (2012)
- 48) Wells MJ: *Octopus. Physiology and Behaviour of an Advanced Invertebrate*, Chapman and Hall (1978)
- 49) Wheeler, DA. et al: *Nature* 452 (7189): 872–876 (2008)
- 50) Winkelmann I, & Gilbert MT: *Ocean Sciences Meeting Salt Lake City, Utah* 503-504 (2012)
- 51) Yamamoto M: *Zoolog Sci* 5:989-998 (1998)
- 52) Yoshida MA, Ishikura Y, Moritaki T, Shoguchi E, Shimizu KK, Sese J, & Ogura A: *Gene* 483:63-71 (2011)
- 53) Yoshida, MA, & Ogura, A.: *BMC Evol Biol*, 180 (2011)
- 54) Young JZ: *The Anatomy of the Nervous System of Octopus vulgaris*. Oxford: Clarendon Press (1971)
- 55) Young RS, Marques AC, Tibbit C, Haerty W, Bassett AR, Liu JL, & Ponting CP: *Genome Biol Evol* 4:427-442 (2012)
- 56) Zmasek CM, & Godzik A: *Genome Biol* 12(1):R4 (2011)

Abstract

Genome analysis platform for Cephalopod studies and its prospects

Atsushi OGURA

Institute for Genome Research, The University of Tokushima, Tokushima, 770-8503, Japan

Cephalopods have evolved an elaborate camera eye and sophisticated central nervous system, and are believed to be the most advanced invertebrates. The heart-blood vascular systems as well as camera type eye in cephalopods has also been recognized as homologous organs found in vertebrates. Therefore, genomes of the cephalopods should be essential resources not only for cephalopods researchers but also for biologists studying evolution, development and comparative genomics. However, there is no genome data available so far, due to relatively large-sized genomes that are comparable with mammals (3-5Gb). In this situation, the Cephalopod Sequencing Consortium (CephSeq Consortium) has established at a NESCent meeting held on May, 2012 and agreed on a set of cephalopod species of particular importance for initial sequencing and developed strategies, and an organization (CephSeq Consortium) to promote this sequencing. Here, I introduce recent movements of cephalopod genomics, and discuss prospects of cephalopod studies and applications to outside the cephalopod researchers community.

Key words: Genome, Transcriptome, Comparative genomics, Next generation sequencer, Cephalopods, CephSeq