

日本進化学会 ニュース

Vol. 17 No. 1
March 2016

目次

- 1 会長挨拶
田村浩一郎 (首都大学東京)
- 2 2016年度役員紹介 (2016-2017年)
- 3 リレーエッセイ〈11〉
内臓逆位が進化するのなぜ巻貝だけなのか
浅見崇比呂 (信州大学)
- 7 進化学者に聞く！ 学生からの10の質問
宮 正樹 (千葉県立中央博物館)
- 9 学会賞受賞記
感謝のご挨拶
岸野洋久 (東京大学)
- 15 第4回 シリーズ「大量データと知見の架け橋」
全ゲノムシーケンズデータからのHLAタイピング手法
成相直樹 (カリフォルニア大学・サンディエゴ校)
- 23 どのアセンブリを使うか？：分子系統学的観点に基づくアセンブリの評価
原 雄一郎 (理化学研究所)
- 30 第24回 海外研究室だより
医学部に埋め込まれたハードコア分子進化の拠点：
コロラド大学デンバー校・David D. Pollock 研究室
福島健児 (コロラド大学)
- 38 第1回 ダーウィン研究室：国内にもある、Cutting-Edge Science!
東京工業大学・田中幹子研究室紹介
岡本恵里 (東京工業大学)
- ミーティングレポート
- 40 第24回 Plant and Animal Genome Conference (PAGXXIV) に参加して
川原善浩 (農業生物資源研究所)
- 43 CeMEB Eco-Evo国際カンファレンスinスウェーデン
荒木仁志 (北海道大学)
- 46 編集後記

会長挨拶

田村浩一郎（進化学会会長、首都大学東京）

日本進化学会会員の皆様

2016年1月より日本進化学会の会長を務めております田村浩一郎です。これから2年間、皆様のお役に立てるよう頑張る所存ですので、どうぞよろしくお願ひいたします。

執行部の顔ぶれは、前執行部人事に際して長谷部光泰前会長が大蛇を振り「私以外はほとんどの方が前執行部から交代致しました」とのことでしたので、今期は基本的にそのメンバーを継承し、慣れた職務をより一層深く進めてもらう方針としました。それでも副会長になった河村正二さんの後任として事務幹事長に大田竜也さん、庶務幹事に石川麻乃さん、渉外幹事に田中幹子さんを迎え、一段と若返り、ジェンダーバランスも改善しました。評議員ともども執行部一丸となって学会をより一層、盛り上げていきたいと思ひますので、学会員の皆様にも引き続きご協力をお願ひいたします。

多くのノーベル賞受賞者を輩出する一方、改竄・捏造・剽窃・研究費不正使用などスキャンダルもあり、サイエンスの世界は世間から一層注目される昨今ですが、学術の意義についてあらためて考え直すべき時に来ているのだと思ひます。自然科学研究機構のホームページを開いてみると、「『学術研究』とは自然、人間、社会におけるあらゆる現象の真理や基本原理の発見を目指して、人間が自由な発想、知的好奇心・探求心をもって行う知的創造活動です。」という言葉から始まります。学術の本質が明瞭、簡潔に記述されている一文だと思ひます。残念なことに、一般社会では応用分野への偏重が進んでおり、基礎と応用のバランスを欠いていると言わざるを得ません。進化生物学への理解も薄まってしまう危機に直面しています。基礎なくして応用はあり得ません。iPS細胞による再生医療や遺伝子治療のような分野でさえ、進化生物学から学ぶべき教訓は数多くあります。日本進化学会としてはこのような基礎科学の重要性を啓発し、何より知的好奇心を満たすための研究の面白さ、楽しさを、未来を担う青少年をはじめ広く社会に知らしめることがますます重要になります。幸いにも、次世代シーケンシングやゲノム編集などの新技術の恩恵は、モデル生物を使った研究だけでは済まされない進化生物学にとって他の分野以上に大きなものとなります。進化生物学の発展自体が社会貢献のための最善の道ですので、学会はその一助となるべく活動していきたいと思ひます。その中核となる今年の年大会は、8月25～28日に東京工業大学大岡山キャンパスで開催されます。学会員の皆さんの力で最大限に盛り上げていただきたいと思ひます。

日本進化学会は1999年10月に発足し、今年で17年目を迎えます。この間、会員増加や年大会充実のための試行錯誤が繰り返されてきました。財政的にも楽ではない時もずいぶんありました。現在、ようやく状況は安定してきたように感じますが、まだまだ安心はできません。幽霊会員の扱い方もあって一概には言えませんが、会員の年齢構成は年々高い方にシフトしており、一般会員に対して学生会員の割合は小さくなってきています。よく言われる若手の人材不足は日本進化学会もご多分に漏れず、今後少子化に伴いますますます深刻化すると予想されます。学会はこれまで高校生ポスター発表「みんなのジュニア進化学」を年大会で主催してきましたが、今後は学部学生に対しても進化生物学の面白さをアピールするようなことを考え、より多くの学部生に進化生物学関連の研究室を選択してもらえるような活動も行いたいと思ひます。また、近年の年大会主催者の努力により財政的に余裕ができてきました。社会的状況からも日本進化学会も法人となるべき段階に来たと考えます。今後数年の間に法人化することを検討したいと思ひます。

以上、今後2年間、微力ながら学会の活動を通して進化生物学の発展に尽力する所存です。学会員皆様のご協力をどうぞよろしくお願ひいたします。

2016年度役員紹介 (2016-2017年)

執行部

会長	田村 浩一郎	首都大学東京
副会長	河村 正二	東京大学
事務幹事長	大田 竜也	総合研究大学院大学
会計幹事	入江 直樹	東京大学
庶務幹事	石川 麻乃	国立遺伝学研究所
渉外幹事 (国内)	長田 直樹	北海道大学
渉外幹事 (国内)	田中 幹子	東京工業大学
編集幹事	荒木 仁志	北海道大学
web担当	野澤 昌文	国立遺伝学研究所
国外渉外担当	入江 直樹	東京大学
広報担当	土松 隆志	千葉大学
生物科学学会連合担当	寺井 洋平	総合研究大学院大学
生物科学学会連合教科書問題検討委員	和田 洋	筑波大学
生物科学学会連合ポスドク問題検討委員	寺井 洋平	総合研究大学院大学
日本分類学会連合担当	村上 哲明	首都大学東京
自然史学会連合担当	三中 信宏	農業環境技術研究所
男女共同参画委員会担当	榊原 恵子	東京大学

評議員

浅見 崇比呂	信州大学	田中 幹子	東京工業大学
荒木 仁志	北海道大学	野澤 昌文	国立遺伝学研究所
入江 直樹	東京大学	長谷川 真理子	総合研究大学院大学
巖佐 庸	九州大学	長谷部 光泰	基礎生物学研究所
遠藤 秀紀	東京大学	深津 武馬	産業技術総合研究所
長田 直樹	北海道大学	細 将貴	京都大学
倉谷 滋	理化学研究所	牧野 能士	東北大学
五條堀 孝	国立遺伝学研究所	真鍋 真	国立科学博物館
颯田 葉子	総合研究大学院大学	三中 信宏	農業環境技術研究所
高橋 文	首都大学東京	矢原 徹一	九州大学

編集委員

荒木 仁志 (編集長)	北海道大学
大島 一正 (副編集長)	京都府立大学
奥山 雄大	国立科学博物館
佐藤 行人	東北大学
真鍋 真	国立科学博物館
山道 真人	京都大学
工樂 樹洋 (~ 2016.3)	理化学研究所
石川 由希 (2016.4 ~)	名古屋大学

リレーエッセイ〈11〉

内臓逆位が進化するのなぜ巻貝だけなのか

浅見崇比呂 (信州大学学術研究院理学系)

Theodosius Dobzhansky (1973) は、晩年のエッセイを *Nothing in biology makes sense except in the light of evolution* と題した。いまでは周知の、この名文をはじめて目にしたのは1977年、刊行直後の *Evolution* の表紙をあけたときだった。雑誌 *Evolution* ではない。筆頭著者 Dobzhansky の弟子たち (Francisco J. Ayala, G. Ledyard Stebbins, James W. Valentine) による、たぶん史上はじめての進化生物学の教科書だった。それを手にした学部4年生のわたしは、40年先のことなど想像できるはずもなく、おそらくは想像する興味すらなかった。今年、前期・後期を通じての通年科目で Douglas J. Futuyma や Mark Ridley による同名の本を教科書に指定して10年になる。感慨深いのは、いまでは、ほかにも英語で書かれた進化生物学の教科書がいろいろにあり、ごく普通に版を重ねている。いわずもがな、進化生物学の今昔がここにも見てとれる。

元祖 *Evolution* が発刊された頃、少なくとも日本では、大学の専門科目名に *Evolution* や進化生物学はありえなかった。とはいえ Dobzhansky の名文が意味するところは、わたしが日本的に所属した集団遺伝学関連の研究室の面々には自明で、何ら新しい情報ではなかった。だから、その文章が、時を経るごとにますます有名になり、まさか遠い将来 (40年後) に、生命科学の多くの領域で共有されることになるとは、想像できるはずがなかった。進化 (生物の歴史) は考えなくてよかった (はずの) 他の生物学の領域で、名文の意味するところが肝になり、名文の引用がなにかと便利になって久しい。これを幾多の先輩がことあるごとに語るから、聞く耳をもつ若者はきつと聞き飽きている。だから、40年後の進化生物学の授業で、こんな時代が来るとは思わなかった…などと言わないように心がけているつもりなのに、数えきれないほど口にしてしまった。それほど「進化」を冠するのが生物学のほとんどの分野で何かと便利な時代、その到来を予測した専門家は、はたして70年代に日本列島のどこかにおられたらどうか。

Dobzhansky が亡くなったとき、わたしは学部2年生。彼の偉業は書物でしか知らない。その彼の、単純無比にして万能とされる種分化モデル。それを覆すといえば大げさだが、のちにその機会に恵まれたのは幸運だった。Dobzhansky-Muller (または Bateson-Dobzhansky-Muller) model とよばれる論理 (2座位モデル) は、明快な説得力に長けている。種分化に不可欠の生殖隔離をもたらす形質が進化するには、最低でも2座位の間でのエピスタシスが不可欠であり、単一遺伝子による種分化はありえないと予測するものだ。だが今では、巻貝のうち少なくとも有肺類 (カタツムリ・ナメクジ・モノアラガイなどのグループ) では、発生の左右極性を決める単一遺伝子が生殖隔離をもたらす種分化遺伝子として機能することは、たぶん、ごく普通の知識ではないだろうか。

有肺類では、単一遺伝子の突然変異で左右逆に発生し、巻き方向と内臓の配置が左右逆になる。しかも、交尾器が正中線ではなく体側にある。だから、左右反転した変異体は、野生型との交尾が物理的に難しくなる。それなら、多数派の野生型との交尾 (逆旋交尾) が難しいので、それだけで逆巻変異体は淘汰されることになる。だから、逆巻の集団が進化するには、集団が fitness landscape の斜面をくだり (淘汰されるはずの逆巻が増え)、適応度の谷を越え、集団が別の適応度の頂点にたどりつく (逆巻ばかりになる) プロセス、すなわちピークシフトをしなくてはならない。世界の常識、2座位モデルによれば、逆巻に進化するステップは、だから、それ以外の要因がなければありえない。

あらためて整理すると、巻型の一遺伝子 (座) だけでのピークシフトを2座位モデルが否定する根拠には二つある。一つは「逆旋交尾が物理的に難しい」。もう一つは「正の頻度依存淘汰で逆巻遺伝子が淘汰される」。19世紀末には、右巻種と左巻種が初期胚の段階から左右逆に発生することがわかっていた。にも

かかわらず、この単純な2点は、*Partula* (ポリネシマイマイ) 派が80年代初頭に実証するまでは、なんと巻貝(腹足類)を扱う専門家の間ですらも認識されていなかった。*Partula* 派とは、ポリネシアはタヒチ諸島に分布するカタツムリ*Partula* 属をもちい、かつて進化生物学・生態遺伝学の一大研究を展開した旧オックスフォードグループから分岐した、James Murray (バージニア大学)、Michael S. Johnson (バージニア大学、後に西オーストラリア大学) と Bryan Clarke (ノッティンガム大学) のことだ。

逆旋交尾が物理的に難しいことを知るのは、歴史的な難問である。交尾前隔離がない(交尾をしようとする)同種の右巻と左巻を使わなくてはならない。同じ巻型どうし(同旋交尾)より、逆旋交尾のほうが難しいか、それをテストしなくてはならない。そもそも野生集団に逆巻変異など、普通はみつからない。博士論文を提出した翌日にはじめてタヒチ諸島にむかった Clarke と Murray は、*Partula* 研究の当初からこの問題に関心があったわけではない。モーレア島の調査ですぐに左右二型の不思議な分布パターンに気がついてから10年後に、同一種にみつかると右巻集団と左巻集団をつかい、逆旋交尾の問題にいどむことになる。野生の、同一種の右巻と左巻をもちいた研究には前例がなく、逆旋交尾の定量的な実証研究は、世界の独壇場にあった。それだけに、*Partula* の逆旋交尾に関する一連の研究は、カタツムリの逆旋交尾に関するモデル研究として世界が気前よくうけとめた。だが、*Partula* 派の自信、意気込み、そして思い込みには、たぶん彼らが想像だにできなかったドラマチックな、傍からみればマニアックな展開がまっていた。

そもそも同一種のカタツムリに右巻集団と左巻集団がみつかることは、巻貝全体からすれば、あるいはカタツムリ全体からしても、例外にちかい。しかし、*Partula* 派は、Dobzhansky (進化総合説) のパラダイムのもと60年代後半に取り組んだ研究で、逆巻集団が自然選択(生殖的形質置換)により適応進化した例をミゾマイマイ (*Partula suturalis*) にみつけた。この場合、祖先型の左巻集団から分岐した右巻集団は島内で分布をひろげたものの、左巻集団と側所的に分布し、両集団のあいだには遺伝的差異がほとんどなかった。すなわち種分化していない。だが、境界域には右巻と左巻が混在し、たしかに巻き方向に対し正の頻度依存淘汰が生じていた。正の頻度依存淘汰により、少数派であるかぎり逆巻(右巻)の突然変異は淘汰される運命にある。ところが、ごく近縁な種(左巻)が分布する地域では、右巻のほうが不毛な雑種形成を減らせるため、右巻だけの集団が進化した。この生殖的形質置換を支持する結果を、*Partula* 派は、タヒチ諸島はモーレア島のミゾマイマイで得ることに成功した。

これに匹敵する鏡像進化の研究はほかに例がなかったため、20年以上にわたり、生殖的形質置換による逆巻(右巻)集団の進化は、あたかもカタツムリの逆巻進化のモデルケースのように独り歩きしていった。ミゾマイマイの場合、正の頻度依存淘汰により巻型変異は淘汰され、右巻集団の分布域には右巻だけが、左巻集団の分布域には左巻だけが存続していた(ここで過去形なのは、アフリカマイマイ退治のためにフランス政府がフロリダから移入したオカヒタチオビガイというカタツムリ専食のカタツムリが樹上性*Partula* を食べつくし、ミゾマイマイも80年代末に絶滅したからである。害虫のアフリカマイマイはいまもタヒチ諸島に健在)。ところが、両分布域の境界に右巻と左巻が混在し、遺伝子流動が継続するため、生殖的隔離が進化しない。巻型は母性遺伝(Toyama 1913 *sensu stricto*) するため、当初から右巻と左巻が混在すると(両者がたとえ交尾できなくても)右巻表現型と左巻表現型の間では理論的に遺伝子流動が継続し、核ゲノムが分化できない。

これら一連の研究をつみあげた彼らは、だから「巻貝では一般に」左右反転は種分化の原因にはならない、と主張するタカ派になっていった。じつはミゾマイマイの場合、逆旋交尾が同旋交尾よりむずかしいものの、逆旋交尾ができないわけではない。だが、わたしは右も左もわからぬ留学先で、*Partula* 派の主張するとおりに、ミゾマイマイで逆巻集団が種分化しない実態が、あたかも巻貝で普通のことのようにおもいこむことになった。ところが20年後に、Dobzhansky はもとより *Partula* 派恩師の到達点に真っ向から対抗する証拠を手にするようになる。決して意図したわけではない。そもそも、逆巻進化の実証研究が、国内外の現実の制約のもとで可能だとは到底おもえなかった。帰国してからいまにいたるまで、いくたびも恵まれた、偶然のあまりに大きな幸運と友人たちに感謝したい。

カタツムリ専食のセダカヘビ類は、左巻よりも圧倒的に多い右巻の肉をひきぬいて食べるように特化し、左巻はうまく捕食できない。セダカヘビの分布する東南アジアだけで頻繁に右巻から左巻への進化が生じたのは、これがため左巻変異が生存上有利だからと考えられる。細将貴さんの本で知られる、この左巻の対抗進化においても、ミゾマイマイでの右巻集団の進化と同様に、逆巻が有利になる環境が逆巻進化の十分条件として機能している。

さて、化石・現生種、海・陸・淡水圏をとわず、世界の巻貝の全体を見わたせばわかることだが、ほかの動物群（左右相称動物）とくらべ、巻貝では、左右逆に発生する内臓逆位（逆巻）の集団（種）が頻繁に進化している。左巻の種数ではなく、右巻のグループに左巻系統がふくまれる場合や、左巻の分類群に右巻種がふくまれる場合をかぞえてもそれが歴然である。こうしたカタツムリ全体での逆巻種の進化が、はたして本当に、2座位（以上）モデルがいうように、自然選択（野生型に対する自然淘汰）なしにはありえないだろうか。パラダイムにはじめて異を唱えたのは、Edmund Gittenberger (1988) と H. Allen Orr (1991) だった。*Partula* 派との論争のはじまりである。

有肺類で殻のあるカタツムリは、殻がひらたい種類と殻の背が高い種類のどちらかにわかれる。どちらも多系統で、くりかえし進化した。これは、*Partula* 派が派生した旧オックスフォードグループの棟梁 Arthur J. Cain が 1979 年に報告して以来、巻貝の殻に興味をもつ人ならだれでも知っていることだ。平型か尖り型のどちらか。大きく二つの形状にわかれる原因が議論されてきたが、いまもって説得力のある証拠は得られていない。

Gittenberger (1988) は、有肺類では、巻き方向（発生）の左右反転が種分化の原因となっている可能性を主張した。平型グループでは、まれに採集される左巻変異体が野生型の右巻とは交尾ができなかったことがくりかえし報告されている。尖り型グループではそうした報告がなく、かつ右巻祖先から左巻種が進化する頻度が尖り型のほうが高い。これらのことから、尖り型では逆巻突然変異に対する淘汰圧が低いことを予測した。だが Gittenberger は、尖り型と平型とでなぜ左右反転進化率や淘汰圧が異なるかには気がつかなかった。

この頃、わたしは論争にはまったくかわりなく、絶滅してしまったミゾマイマイのごとく同一種に右巻と左巻が見つかる事例（左右二型現象）が、カタツムリ全体でどれほどあるものかを知ろうとした。調べはじめるとすぐに気がついたのは、左右二型の報告がある種類（13 属）はどれも、殻がひらたくなく、背が高いか、細長いことだった。



図 右側：セトウチマイマイの対面交尾（同時正逆交尾、殻径 30mm）。左側：ヒクギセルの背面交尾（非同時正逆交尾、殻高 20mm）。

さらに、ふしぎなことに気がついた。ひらたいほうの分類群では、交尾する 2 個体が向きあい、一回の交尾で精包を交換する（同時正逆交尾）（図）。一方、殻が高いグループは、同時雌雄同体であるにもかかわらず、交尾するときには、オス役とメス役にわかれ、オス役がメス役の殻に乗って交尾する（非正逆交尾）。当時は、これらの異なる交尾様式に着目した先行研究はほとんどなく、さまざまな種の交尾をあらたに目視する実験は不可能にちかい。かぎられた情報を収集し、殻のかたちと交尾様式との対応関係をつきとめるのは容易ではなかった。

なぜ左右二型が見つかるのは殻の高い尖り型だけで、平型ではみつからないのか。突然変異で逆巻遺伝子が出現する確率がカタツムリの系統間で異なるとは考えられない。平型に左右二型が見つからないのは、逆巻が淘汰されやすいからにちがいない。そ

うおもって交尾様式をくらべ、物理的なメカニズムを直感するのは容易だった。非正逆交尾は、陰茎の挿入が上(雄役)から下(メス役)への一方向でよい。だが、ひらたいカタツムリの正逆交尾では、同時にたがい陰茎を挿入する際に、本来なら向きあう交尾体位をどちらか一方の個体が大きく変更しなくてはならないだろう。これは、右巻と左巻が乗っかり型の非正逆交尾をする場合とくらべ、物理的にはるかにむずかしいにちがいない。だから、逆巻に対する淘汰圧は、平型の集団のほうが高い。

この仮説を立証するには、平型のカタツムリで、右巻と左巻の交尾(逆旋交尾)と同じ巻型どうしの交尾(同旋交尾)のあいだで難易度をくらべ、その結果を尖り型の場合とくらべなくてはならない。しかし、平型の種類では、左右二型の集団は記録がなく、統計的検定にたえるだけの逆巻変異体を得るのは不可能にちかい。数年をへて、奇跡としかおもえない幸運により、この問題を史上はじめて解決する機会を得た。世界に誇るべき日本の自然史アマチュア研究者の一人、岡本正豊さんが自宅の庭でオナジマイマイの左巻変異1個体を発見し、生かしたまま研究用に持参してくださった。しかし、検証実験をするには逆巻個体が多数必要なので、逆巻の系統を確立・保存しなくてはならない。この問題をどのように克服したかは、あちこちに書いたので割愛する(たとえば、松本忠夫・長谷川真理子共編「生態と環境」5章、培風館)。

結果として、平型カタツムリの同時正逆交尾では、逆旋交尾が事実上不可能であることを立証できた。Johnson (1982)によれば、ミゾマイマイでは、逆旋交尾は26回中3回(約12%)成功している。これを同旋交尾の成功率とくらべると、逆旋交尾の相対成功率は22%となる。平型のオナジマイマイでの0%とくらべ、統計的検証がいらぬほど、ミゾマイマイでは右巻と左巻が簡単に交尾できたことがわかる。平型のカタツムリで生じる正の頻度依存淘汰が、いかに強烈か。これは、今日では世界でごく普通の知識となった。

その後、さらなる検証実験により、岡本さんが発見した左巻表現型の原因遺伝子は、左巻遺伝子ではなく、こどもの発生の左右極性を決定できないラセミ遺伝子であることがわかった。同様のラセミ遺伝子による逆巻表現型は、ナミギセルの野生集団でも生じることがわかった。これらの例からは、脊椎動物にみつかると内臓逆位の遺伝子と同様に、巻貝であっても、突然変異で左右極性決定能を欠く変異遺伝子が生じ、かならずしも右巻遺伝子が突然変異すれば左巻遺伝子になるのではないことがわかる。

歴史的におもしろいのは、逆旋交尾が物理的にむずかしいことを世界に知らしめたのが*Partula*派であったことだ。しかし、実験条件下で12%もの頻度で逆旋交尾が成功するから、側所的に分布するミゾマイマイの右巻集団と左巻集団のあいだでは、当然ながら種分化をさまたげる遺伝子流動がじゅうぶんに生じる。だが、カタツムリでは通常は不可能にちかい研究により、78%もの失敗率で正の頻度依存淘汰が生じることを実証したJohnsonをはじめ*Partula*派の歴史的な功績はいちじるしく大きい。その経緯もあってのことだろう、*Partula*研究に没頭するJohnson (1990)は、有肺類の全体をみわたして主張するGittenberger (1989)の、左右反転による種分化説に躍起になって反論した。結果、Johnson (1990)は、シミュレーションにより、巻型が母性遺伝すれば、理論的に逆旋交尾が不可能でも世代を越えて遺伝子流動が左右二型の間で継続することをしめた。だが、Johnson (1990)が使ったモデルは、集団サイズを無限大と仮定する簡単なものだった。

この問題に気づいたOrr (1991)はすぐに、遺伝的浮動の効果を取り入れたシミュレーションにより、母性遺伝する場合には逆に、小さな集団ほど逆巻遺伝子に固定しやすいことをしめすことに成功した。van Batenburg & Gittenberger (1996)も同様のシミュレーションでOrr (1991)の結果を支持。これらの成果は、結論は逆ながらJohnson (1990)のアプローチがあったからこそ得られたものだ。これらは、単一の巻型遺伝子(有肺類の発生の左右極性を決定する遺伝子)により逆旋交尾ができなくなるのであれば、集団が逆巻遺伝子に固定するだけで他の野生型集団からの交尾前隔離(種分化)が完成することを理論的に予測する。すなわち、Dobzhansky-Mullerのtwo-locus modelが否定する、単一遺伝子による種分化がカタツムリにありうるということが理論的にしめされた。

こうした歴史的な背景のもとで、single-gene speciationの予測を立証する成果をもたらしたのが、上

島 励氏 (東京大学) による *Euhadra* 属の形態分類と分子系統解析である。わたしたちの研究により、いくつもの定説が覆された。ほとんどは、いまでは世界のみとめる知識となっている。具体的な内容は割愛する (細胞工学23: 338-339 参照)。

比較対象を *Partula* 属から有肺類の全体、つぎに有肺類から巻貝の全体、さらには巻貝をふくめた左右相称動物の全体へひろげてゆくと、ここでまた、ふしぎなことに気がつく。たとえば、*Partula* では、多数派との交尾がむずかしいから少数派の巻型は淘汰されるものの、右巻と左巻は種分化しない。だが、有肺類全体をみわたすことで、すくなくとも平型 (同時正逆交尾) のカタツムリで、左右反転により交尾前隔離が完成する可能性があることがあきらかとなった。東南アジアではセダカヘビ類の右巻捕食圧が貢献しているが、世界全体では遺伝的浮動と正の頻度依存淘汰のたまものだろう。逆に、逆旋交尾が可能な尖り型カタツムリでは、逆巻遺伝子が淘汰されにくく、逆巻系統が進化しやすい。それなら、海の巻貝ではどうか。特に、放精放卵で体外受精する巻貝なら、交尾しなくてよいのだから頻度依存淘汰は生じないだろう。それなのに、体外受精して浮遊幼生期をへる巻貝では、逆巻の系統はまったく進化していない。左右相称動物の全体をみても同様である。交尾が正中線上にあれば、内臓逆位であっても野生型との交尾に支障はないはずだ。だから内臓逆位に対する頻度依存淘汰は生じない。ところが、左右相称動物では一般に内臓逆位が進化していない。進化しているのは、なんと交尾して繁殖する巻貝だけである。

むしろ、左右相称動物の全体からすれば、発生の左右極性の反転は純化淘汰されているとしか解釈のしようがない。このデフォルトの淘汰機構がまずあり、交尾する巻貝に特徴的なメカニズムが内臓逆位の系統を進化させているとみるのが順当だろう。モノアラガイでは、左巻変異は、成熟して正の頻度依存淘汰をうける前に、孵化する前に発生異常で純化淘汰をうけている。たぶん、他の左右相称動物でも同様だろう。交尾する巻貝の場合、内臓逆位は、表現型頻度で50%を超えるだけで頻度依存淘汰により繁殖上有利になれる。他の動物では、内臓逆位が繁殖上有利になることはない。これこそが、左右相称動物の全体で、交尾する巻貝にのみ、正常に発生する内臓逆位を選抜し、頻繁に進化させてきたメカニズムではなからうか。

(編集担当：荒木仁志)

進化学者に聞く！ 学生からの10の質問

宮 正樹 (千葉県立中央博物館)

「進化に興味はあるものの、周りには何をやっているのか分かってもらえないし、将来が不安…」と感じている読者も多いのではないのでしょうか。そこで、実際に進化学研究で世界をリードする日本進化学会の先輩方に、あれこれ聞いてみることにしました。今回は前・進化学会ニュース編集長、千葉県立中央博物館の宮 正樹さんです。

- Q1.** 進化学者になったきっかけを教えてください。
- A1.** 学部学生の頃から魚類進化の歴史の全容を自分自身の手で解き明かすことが夢でした。そのため、分岐学の理論を一生懸命勉強したのですが夢果たせず、後に36歳で分子系統学に出会ったことがこの道に入るきっかけになったのだと思います。
- Q2.** 進化学者になってよかった、と思った瞬間はいつですか？

- A2.** 自分が「進化学者」という意識はないのですが、「外洋における種分化」「深海魚の三つの科が一つに」「深海起源のウナギやマグロ」「パラオの海底洞窟から古代ウナギを発見」など、一般の方も含めて多くの人が驚くような進化仮説や発見を世界に先駆けて発表できたことでしょうか。
- Q3.** 小さい頃の夢を教えてください。
- A3.** 釣り少年だったので、プロの釣り師になることや釣り宿のオヤジになることが夢でした。ところが、学部2年の頃に自分如きがトップクラスになれる「釣り」そのものに対して（一生かけて追求すべきものなのか）疑問を持ち始め、紆余曲折あったのですがアカデミアの世界に興味を持ち始めました。その頃のことについては、釣りビジョンというチャンネルの「時代を超えた探究心が世界を変える・宮正樹」という番組で詳しく紹介されています。
- Q4.** 研究者になるのを諦めかけたこと、ありますか？もしあればどうやって克服を？
- A4.** 幸いにありません。とはいえ、博物館準備室に入った1987年当時、2年間は事務仕事のみで研究時間ゼロ。プロの研究者になるなんていうことを考えるゆとりも時間もありませんでした。92年から博物館が当時の文部省の指定研究機関になったわけですが、その後の競争的資金を獲る努力をしている中で、プロの研究者としての意識が出来てきました。その意味で、初めて獲得した科研費の奨励A「分子系統に基づく深海性オニハダカ属魚類の進化史の再構成」が獲れたのは大きかったですね。その後は連続して研究費を獲得していますから、これがなかったら今の私は確実にありません。
- Q5.** 進化学者になるのに必要な素質・スキルって何でしょう？
- A5.** 進化学者に限りませんが、物事に対する「探求心」と「情熱」は研究者にとって最も重要な資質（＝素質・スキル？）だと思います。その意味で、心から楽しそうに研究している人は、間違いなく研究者の資質をもっていますね。
- Q6.** 学部生・院生当時の一番の思い出は？
- A6.** 将来のことを何も考えずに博士課程2年目（1986年）に学生結婚したことでしょうか。今年結婚30周年になることを、この文章を書いていて気がつきました。後先を考えないというのも研究者にとって重要な資質かもしれません（笑）。
- Q7.** 現在の研究内容について教えてください。
- A7.** 魚類の大進化研究については一区切りつけて、現在は環境DNAによる魚類モニタリング技術の確立に執念を燃やしています。水の中を見てきたかのようなデータがとれるこの技術、新たな研究フロンティアを若手研究者と一緒に楽しんでいます。
- Q8.** 今後の研究展開や抱負を聞かせてください。
- A8.** 既成の特定分野にこだわらず、研究のフロンティアに立ち続けたいですね。今は進化学から少し離れていますが、環境DNAと次世代シーケンサを組み合わせたメタバーコーディングの技術は生物学の世界を大きく変える力をもっています。不可知であった地域や世界規模の生物相を誰でもモニタリングできる時代がやってきたのです。
- Q9.** 10年後、進化学はどこまで進んでいると思いますか？
- A9.** 人工知能、3Dプリンタ、一分子シーケンサなどワクワクするような科学技術がどんどん身近になり、考えられない分野が進化学にも誕生していると思います。どんな成果が出るのか予想もつき

ませんが、仮説検証型よりデータ駆動型・発見探索型のアプローチがメジャーになるのでは？

Q10. 未来の進化学者に一言。

A10. 世界を相手に活躍できる、こんなに楽しい世界はありません。プロの釣り師を目指していた私だって進化学者になれたくらいですから、探求心と情熱（と多少の頭脳）さえあれば誰だってこの世界で通用します。既成概念や分野にとらわれないパイオニアになって、研究のフロンティアを楽しんでください！

研究者紹介



名前：宮 正樹

所属：千葉県立中央博物館・自然誌歴史研究部・動物学研究科

最終学歴：東京大学大学院農学系研究科水産学専門課程博士課程修了

職歴（略歴）：1987年に農学博士を取得後、同年に千葉県教育庁文化課博物館準備室文化財主事となり、89年に千葉県立中央博物館がオープンすると同館の主任技師、上席研究員などを歴任。2011年に動物学研究科長に就任し、13年からは主席研究員を兼務。千葉大学大学院自然科学研究科客員教授（13～15年）、ドイツ・コンスタンツ大学の客員教授（05年）などを務めた経験もある。

（編集担当：真鍋 真）

学会賞受賞記

感謝のご挨拶



岸野洋久（東京大学）

この度、日本進化学会賞と木村資生記念学術賞をいただき、心から感謝いたします。私の研究生活はすべて、素晴らしい先生方と偶然に会い、これらの先生方の力強い仕事に導かれ、そして学ぶことによって支えられてきました。今回授かりました賞は諸先生方の仕事に向けられたものですので、いま思い出を振り返りながらそれらをご紹介し、この場をお借りしまして厚くお礼を申し上げたいと思います。

公開講座

いまからもう30年ほど前になりますが、1984年初頭、私は統計数理研究所の講堂で、長谷川政美先生による分子進化と分子系統樹の公開講座を聴いていました。勤め始めて4年が経とうとしていました。幸運にも修士を出てすぐ研究所の所員となりましたが、未熟でした。大学院では確率過程と確率制御を研究したものの、統計学に関する知識も経験も皆無のままのスタートでした。二人の巨匠、数量化理論を開発された林知己夫先生とAIC情報量規準を開発された赤池弘次先生がいらっしゃいました。私は統計数理セミナーを通して諸先輩の研究を学び、研究の方向を暗中模索しておりました。公開講座は、所員である私にとっても貴重な学びの場でした。遺伝学研究所の太田朋子先生や高畑尚之先生も講師としてお話をくださいました。公開講座に魅せられた私は、終了後すぐさま長谷川先生の研究室のドアをノックし、共同研究をさせていただけないか、お願いしました。

1981年にJoe Felsensteinが分子系統樹の最尤推定とそのためのアルゴリズムを提唱するといち早く、長谷川先生はその有効性に注目し、類人猿のミトコンドリアDNA配列のデータ解析を精力的に行ってお

られました。まだ扱う配列の数も少なく、長さも短く、分子系統樹の推定そのものも、始まったばかりのころでした。いまでは分子系統樹の最尤推定・ベイズ推定の強力なパッケージも容易に入手でき、配列データを手にすると、多くの人が当たり前のように利用していますが、当時これを利用する人は多くはありませんでした。DNA配列データを進化系統樹の推定に利用できる可能性が認知され、進化学者はその解析手法を手探りで探している状態でした。

長谷川先生は、同じデータに基づくのに、論文により結果が相矛盾していることに違和感を持っていらっしゃいました。いずれの結果も何らかの推定作業が入っているのに、矛盾が推定の不確実性を超えたものであるか、評価することが大切になってきます。Joeは当初より対数尤度の曲率で枝の長さの標準誤差を評価する方法を実装し、1985年には、サイトのブートストラップリサンプリングをすることにより枝の有意性を評価することを提案しています。

長谷川先生と私は、配列の対数尤度がサイトの対数尤度を足し合わせたものであることから、これを配列の長さで除した平均対数尤度は、その母平均である期待対数尤度を標本平均で推定したものであるという認識に至りました。標本平均の分散は、標本分散を標本サイズで割ることにより評価されますので、サイトの対数尤度の分散を計算し、それを配列の長さで割れば、平均対数尤度の推定精度が評価されたこととなります。系統樹に対するいくつかの仮説を統計的に比較することができるようになったわけですが、実際には多くの場合、最尤系統樹とその他を比較します。これは検定を何度も繰り返すことと同等の作業を行っていることとなります。大学院時代から統計的モデル比較の方法を追求していた下平英寿先生は、この点を鋭く指摘し、多重比較、多重検定の影響を評価しました。

2000年代になると、次第に複数の遺伝子座を同時解析する研究が主流になってきました。そうすると、平均対数尤度が遺伝子座の間で、対数尤度のサイト間のばらつきの大きさから期待される変動を超えてばらついていないか、見積もることができるようになってきます。Tae-Kun Seo (徐泰健) 先生は、複数遺伝子座のデータが遺伝子座のサンプリングとサイトのサンプリングからなる二段サンプリングによりゲノムから抽出されたものと捉え、遺伝子配列の持つ情報の量が遺伝子座の間で異なることを考慮に入れて精度評価する方法を提案しました。

さらに扱う遺伝子座の数が増えてくると、推定の偏りが目についてきます。どのような統計モデルも、どのような推定法も、完璧に実際に起きた現象を捉えることはできません。どうしても何らかの仮定をします。データが小さい時は、そうした仮定に基づく推定に多少の偏りがあっても、推定量が大きな不確実性を持つため、実害はありません。ところが、データが大きくなってくると、推定量の分散は小さく評価されます。そのとき、偏りに気づかないと、結果に瑕疵があるにもかかわらず、推定量の分散が小さくブートストラップ確率やベイズの事後確率が100%に近い結果を全面的に信ずることになってしまいます。

北添康弘先生は高知医科大学の医学情報センターに勤めていらっしゃいましたが、もともと原子核物理で学位を収めていらっしゃいました。配列進化を多次元ベクトル空間内を実現させることにより、配列間の距離行列の妥当性を評価する方法を考案しました。進化距離の平方根で系統樹の節の間の距離を定義します。違和感を持つかもしれませんが、配列間の距離行列はこの多次元空間内ではこれを結ぶ枝ベクトルの長さの2乗和となります。ピタゴラスの定理から各枝ベクトルは直交することが期待されるので、進化モデルの妥当性を種々の角度から診断することができるようになります。データが質的にも量的にも爆発的に向上してくると、これを過度に加工することなく、データそのものに耳を傾けて物語を語らせるという方向になっていくような気がします。

遅まきながらのシアトル

30代も半ばになろうとしていた私には、異文化に感激し吸収する素地を失う前に、外国の研究環境で学びたいという思いがひたひたと強くなっていました。1989年の春、長谷川先生にご相談したところ、Joeに紹介状を送っていただきました。私たち家族(8か月の赤ん坊と妻と私)を乗せた飛行機がシアトル

の空港で降り立ったのは、その年のクリスマス明けの12月26日でした。当日は濃霧で、一旦サンフランシスコに回るとのアナウンスがありました。サンフランシスコの空港に着陸すると、そのまま機内でしばらく待機した後、再び離陸しました。このため、シアトルへの到着はかなり遅れましたが、出口を出ると、Joeが暖かく待っていてくださいました。車に荷物を詰め込み、その日はお宅に泊めてくださいました。翌日、私を車に乗せて街中を回り、社会保障局で私たち家族の社会保障番号を取り、アパートの契約を結んでくださいました。

1月2日に新学期が始まると、彼は大学院生向けの集団遺伝の授業の準備などに忙殺されていました。学生時代は毎週開かれる研究室ゼミが切磋琢磨の核となっていましたので、こうした機会にJoeと情報交換することをもくろんでいましたが、そういうものはありませんでした。こうして自然に、ディスカッションの相手は博士課程のJeff Thorneになっていきました。彼と私はJoeの部屋と内扉で繋がる部屋で机を並べ、遺伝学科の他の大学院生とともにカフェテリアで昼食をとりました。疲れると一緒に散歩したり、家族ぐるみで夕方のパーティに誘われたり、と、すっかり親しい友人になっていきました。

Jeffはアラインメントの最尤推定の問題に取り組んでいました。ミスマッチの背景には塩基置換、ギャップの背景には挿入・欠失の進化イベントがあり、ミスマッチペナルティは塩基置換速度、ギャップペナルティは挿入・欠失の速度と関係します。塩基置換はマルコフ過程、挿入・欠失は出生・死滅過程で定式化されます。これらを併せて動的計画法で尤度を計算することになります。Jeffが夏のある日、尤度を出しました。Joeとは対照的に、Jeffの数式展開は一見泥臭く個性的です。が、手堅く重厚で、誤りがありません。塩基置換に基づく分子系統樹では、内部節の状態を潜在変数として扱います。こう見てくると、アラインメントは挿入・欠失の歴史を表現したものですので、直接観測できない潜在変数です。可能性のあるすべてのアラインメントの確率を足し合わせて尤度を計算します。これにより、アラインメントの不確実性を考慮に入れて進化系統樹を推定することができるようになります。その後強力なプログラムも開発されていますが、実際の挿入・欠失のパターンが複雑であることもあって、まだ十分広く実用に供されるどころまでには至ってないかもしれません。

JeffやJoeは、私をセミナーや授業に誘ってくださいました。JoeとElizabeth Thompsonが担当する統計遺伝のセミナーでは、学生さんが夏、田嶋文生先生が相次ぎ刊行した論文を紹介していました。昼食時に開かれるPopgen lunchはハンバーガー持参で、夕方開かれるPizza Evolutionは注文したピザと缶ビールを楽しみながら、担当者がOHPでプレゼンします。Jeffに誘われて参加した人類遺伝の授業は印象的でした。火曜と木曜の週2回の授業で、受講者はポスドクと大学院生合わせても数名でした。木曜日には性決定因子、色覚異常など、翌週に扱うテーマの鍵論文が配布され、次週の火曜日に担当者がその論文を紹介します。さらにその週の木曜日には論文の著者がやってきて、背景も含めて研究紹介をしてくださいます。これを踏まえて、1. その分野の知見の整理、2. 近い将来どこまでわかると期待されるか、3. 遠い将来どのように道が開けていくと期待されるか、というレポート課題が出されます。自由に議論する中で、先生が頷きながら、たとえばこんな実験を組んだらうまくいくかもしれない、などと黒板に書き込むのを見て、実験のデザインにはセンスが肝要と分野外の私でも実感させられました。

毎週鍵論文の著者を呼ぶほどのLee Hartwellは、酵母をやっているらしく、何やらすごそうだというのは想像がつかしました。が、後にノーベル賞を受賞されるほどの方とは知りませんでした。夏のある日、Joeがロッカーにしまっていた数多くの別刷りの中から1編とって、私に下さいました。それは、組み換えの進化的利点に関するGeneticsの1974年の論文でした。彼のこれまでの研究はどれも屑で、残せるようなものはこれ1本だとおっしゃっていました。私の知りえた研究分野に限っても次々に礎石を築き上げた先生の言葉を、よく理解する力が私にはありませんでした。当時まだ最尤法により分析できる配列の本数は数本程度と限られており、このことが最尤法の普及の障壁になっていましたが、Joeはすでに数千本の配列を最尤法により普通に解析できるようになるだろう、と唱えていらっしゃいました。四半世紀前にすでに今の時代を予想していたことになります。凡人である私には、想像もできませんでした。

このように素晴らしい環境に恵まれていたわけですが、統計学を自分の身の置き所にする自分が、分子系統学においてさえ自分自身のしていることを超えてはほとんど無知のままシアトルにきて、その素晴らしい環境の多くを価値に気づかぬままスルーしてしまいました。いまでも思い出すたびに恥ずかしくなり、汗が吹き出します。

■ 細い糸

当初1年半の予定でしたが、11か月で帰国し、1990年12月、東京大学海洋研究所に移ることになりました。資源解析研究室の田中昌一先生が退官され、沼知健一先生と小林敬典先生が大槌からやってこられていました。田中先生は数理と統計解析で水産資源の解析の道を切り開かれた先生で、国際捕鯨委員会の科学小委員会やその準備会合などでよくお目にかかっていました。沼知先生と小林先生は、種々の海産生物について日本各地から精力的に標本を集められ、その集団構造を推定していらっしゃいました。数理的な側面は松宮義晴先生が引き継いでいらっしゃいましたが、縁あって三重大学に移られました。そこで統計をやり、沼知先生の実験に近い仕事もしていた私が、松宮先生の代替を務めることになったのでしょうか。

統計数理研究所ではさまざまに共同研究がなされていましたが、多くの人は統計学をベースとして、その応用の場として共同研究を捉えていたように思います。海洋研究所は海のことを研究します。海に関わる物理、地質、化学、生物を総合的に研究する場です。資源解析研究室は水産資源の管理に関わる研究がひとつの柱になっています。研究所の中で一番人間社会に近いところを研究していると言えます。

資源の管理は突き詰めていくと、資源を回復するまで獲るのを控えよう、と唱えることになります。回復までに何十年もかかるようでしたら、一生我慢ということにもなりかねません。資源管理も、個体群動態の予測とコントロールのみに限定するのではなく、獲る・作る(料理する)・売ると多角的に捉え、資源の状態に応じて適応的に重心を移すようなスキームを考えれば、よりスムーズに機能するのではないかと考えました。そこで、どのようなところが問題となっており、どのような可能性があるか、知るために、全国の漁家の調査をしたいと思いました。当時栽培漁業協会で企画課長をされていた北田修一先生(現東京海洋大教授)にご相談したところ、調査の実施に向けてご尽力くださり、東京水産振興会のご理解を得て、二人で10の道と県のそれぞれから4漁協(秋田は合併により1漁協)を回りました。

北田先生は全国的に展開されている種苗法流について、有効性の評価に関わる仕事を精力的にされてきました。統計数理研究所にいた頃にお目にかかり、調査法のこと、サンプリング調査のこと、2000年代になってからは放流種苗の周囲への遺伝的影響のことなど、数多くのことを学ばせていただいています。

沼知先生が退官されて3か月後の1993年7月、縁あって駒場の教養学部の統計学教室に移ることになりました。統計学教室を率いる松原望先生はdecision makingを意思決定と訳し、広く日本に定着させました。先生の統計学の授業には定評があり、600人を収容する大講堂もあふれ返り、2回に分けて講義されていました。私は赴任するまで、統計学教室が相関社会科学を研究する先生方と国際関係論の先生方に囲まれていることの意味を理解していませんでした。経済学者は、社会は10年を超える教育を確度の高い人物評価の場と捉えていることを指摘し、政治学者は社会保障が社会を安定化するという役割を見抜きます。巧みな文章により種々の証拠を提示し、必要に応じて表を添えることにより、多くの人を説得します。これまでに全く経験したことのない手法でした。こうした素養の全くない私は、地域社会、長寿社会と福祉、リサイクル社会をテーマに調査をしました。よくデザインされていれば、無から有を生じ、社会に対してメッセージを提供できるかもしれないと考えたからです。いずれも社会科学の個々の分野のみでは問題は解決せず、それらが個性を持ちつつ融合した総会社会科学の場に学んで、初めて解決の糸口が見えてくることを期待しました。

進化学と関係のないお話になってしまい、申し訳ございません。実際、1990年代は、ほとんど進化学とは異なる世界で生活していました。オンラインジャーナルもなく、進化学の最新の論文に接することなく時が過ぎていきました。海洋研究所にいたときは定年まで無事に研究を続けるにはどうしたらいいか、

駒場にいたときは社会科学で生き残るにはどうしたらいいか、そのことのみを考えていました。そんな中、Jeffとのメールのみが細い糸のように私を進化学につなぎとめてくれました。

シアトルを離れるとき、すっかり友達になっていた私たちは、電子メールでの交信を続けようとお互いに約束しました。まだ電子メールは産声を上げたばかりでしたが、驚くほどよく機能しました。毎日のメールはほとんど “Hi”+ α 程度の簡単なものでしたが、たまに進化の話題が登場し、“I have a question for you.” などと書いて統計的な問題を議論してくれたりしました。実際に会えたのは2-3回でしたが、分子時計の揺らぎを考慮に入れて分岐年代を推定するというベイズ推定のアイデアが生まれました。その時はまだ、1999年4月にここ農学部にご縁をいただくことになるとは、想像していませんでした。たまたま農学部に移るお話があったときは、長い間自然科学から離れていたのだから、やっつけられるのか、とても不安でした。あれから20年近くになろうとしています、いまでも不安です。

農学に進化を学ぶ

農学部に来てから、進化のことを計り知れないくらい学んでいます。運が良かったと思います。子孫を受け継ぐ上で重要な脱粒性を失ったイネが効率的な収穫を可能としていたり、収量増が分げつを抑えられることの対価であったり、栄養価の付加が生合成経路の一部の故障の補償作用であったり、病原体への抵抗性の獲得の実態がその受容体となるたんぱく質が作られなくなったことであったりと、選択と適応、進化の舞台裏を耳にして、毎日が新鮮で、刺激的な生活を送っています。欠陥品も見人その時の環境によっては優れたものとなるという歴史的事実が、一番心にしみわたってきます。こうした環境に身を置いて、統計的なモデリングを通して適応の本質部分を定量的に評価するような仕事に触れたいと思うようになりました。RNAウィルスは世代の長さが短く、比較的単純な「生活環」を持つので、適応進化を決める因子を量的に掴む上で魅力的な材料に映りました。そうして、ウィルスの進化の源泉を追求する方々と巡り合うチャンスに恵まれました。

当時博士課程で長谷川先生の指導を受けていたTae-Kun Seoは、定期健診で得られたHIVの遺伝子配列の継時データを合体過程によるモデリングで分析しました。患者さんの間で分子進化速度と遺伝的多様度を比較すると、負の相関があることを見出しました。免疫系の力が強く細っているウィルス集団は、比較的淘汰圧の弱い集団よりも多くの変異を持ち、多様化の中から適応の芽が出るチャンスを広げていることとなります。太田朋子先生はすでに、突然変異の多くが弱有害であるときに分子進化速度が集団の大きさと負の相関を持つことを示していらっしゃいます。ウィルスのしぶとさの根源はスマートさにあるのではなく、困ったときの自暴自棄にあるのでした。ノンストラテジックであるだけに手ごわい相手であることを、改めて実感するのでした。

ノンストラテジックというと聞こえはいいですが、自暴自棄が適応の源泉というのは何かごまかされたような気持ちが残ったことは否めません。植物ウィルスを研究する宮下脩平先生は、宿主細胞内のウィルスゲノムのボトルネックに注目しました。ウィルスの感染した細胞の中には無数のウィルスゲノムが存在するため、突然変異により劣化したゲノムも他のゲノムの助けで子孫を残す可能性を持ちます。ところが、細胞間の移行や細胞質でのRNA分子の分解によりウィルスゲノムは強いボトルネックを受け、多様性を極度に落とします。他のゲノムの助けを持たない劣化ゲノムは次世代に受け継がれず、効率よく排除されていくこととなります。蛍光タンパクで色づけしたウィルスを同時感染させた細胞を継時的に観測することにより、ボトルネックの結果、感染細胞内には平均4-5個分のゲノムの多様性しかないことがわかりました。細胞間の不均質性を考えると、有効な集団の大きさは2程度にまで落ちます。2という数字は意味あり気ですが、ひょっとするとこれは、細胞内でゲノムの組換えを可能にすることと関係するかもしれません。点突然変異による微調整と異なり、ゲノムの組換えは複数の遺伝子のアレルの取り合わせを調節できるからです。ブラジルからの留学生であったLeonardo de Oliveira Martinsは来日前からHIVウィルスのゲノムの組換えに関心を持っていました。そして、ゲノム内で系統関係が変化するパターンをベイズ推

定することにより、組換えを起こす部位を検出する方法を開発しました。

ウィルスのゲノムは発現のためのツールを持たず、宿主のゲノムに組み込まれてタンパク質を作り出してもらっています。このため、彼らの適応進化の多くは、タンパク質の変化に原因を求めることができます。このことが、進化を研究する素材としてウィルスに魅力を感じたもう一つの特徴でした。渡部輝明先生は、高知大学医学情報センターに来られる前は、原子核物理で学位を収められ研究を邁進していらっしゃいました。渡部先生にとって、タンパク質は実体を掴みやすく、安心して取り組める素材だったそうです。タンパク質の立体構造は私には難解でハードルが高い対象でしたが、渡部先生はいとも容易くまな板の上に乗せられました。抗体と宿主の細胞がウィルス粒子を巡って席取り競争します。宿主細胞がうまくウィルス粒子と結合すれば、ウィルス粒子は細胞内に取り込まれます。一方、抗体と結合したウィルスは消滅します。二つの結合の強さのバランスがウィルスの適応度を決めます。タンパク質の立体構造の情報を納めたデータベースと配列情報を併せて分析することにより、多様化圧のかかる領域を特定し、突然変異による適応進化の可能性を定量的に評価する道が開かれます。

ウィルスの話が続きましたが、台湾からの留学生であった Hungyen Chen (陳虹諺さん) は、私を再び生態学と種間関係の分析へと導いてくれました。彼は、台湾北部の魚類群集の解析をさらに深めるべく来日しました。そして、生物群集の進化系統樹上の系統的歪みを測る方法を開発しました。ある生物群集の種組成を、たとえば開発前の群集など対照とする群集の種組成からの標本と解釈します。そして構成種の間分岐年代の分布を対照群集における分岐年代の分布と比較します。ある生物群集の構成種が仮に対照群集の構成種からの無作為標本であったとしても、分岐年代の分布は対照群集における分布に比べて古い方に偏ります。そこでこの偏り方の強さにより、構成種の系統的な偏りを評価することができます。まだまだ粗い尺度ですが、大規模なデータに基づく信頼性の高い分子系統樹が得られるようになるにつれ、この情報をフルに取り込んで生物群集の多様性を評価する方法も開発されてくるものと思われま

います。いまここでご紹介しきれなかった多くの方々にも、心から感謝いたします。田邊和祐先生にはヒトマリアリアが宿主とともに棲息領域を拡大してきたことを学び、Ze Zhang (張澤先生) には遺伝子の運命がその置かれたゲノムの背景に影響されることを学びました。Peter Waddellからは、哺乳類の進化の推定を通して、ゲノムデータの持つ情報の量と質について学びました。Avner Bar-HenとMahendra Mariadasouからは分子系統樹の頑健推定と偏りの評価を学びました。恥ずかしながら、私はこの齢になってまだ進化学を体系的に勉強しておりません。先生方は鋭いセンスで問題に切り込まれ、門外漢の私に統計的なモデリングで共同研究させていただく機会を下さいました。共同研究という形の on the job training を通して、進化学の1節1節を教えてくださいました。振り返るとあっという間で、不出来な弟子にはまだ断片的な知識しか持ち得ていません。どうかもう少し学ばせてくださいますよう、今後ともよろしく願い致します。

(編集担当：大島一正)

第4回 シリーズ「大量データと知見の架け橋」

全ゲノムシーケンスデータからの HLA タイピング手法



成相直樹 (カリフォルニア大学・サンディエゴ校)

はじめに

一昨年、Vol.15 No.2において本連載シリーズ第1回「RNA-Seqによる網羅的トランスクリプトーム解析」を担当させていただいた。RNA-Seqデータ解析においては得られたショートリードをcDNAリファレンス配列 (もしくはリファレンスゲノム配列) にアライメントすることにより、各転写産物由来のリード数を見積もり、転写量を推定する。注意すべき点は、転写産物のcDNA配列がお互い非常に良く似ているケースが多々あるため (例えば同じ遺伝子座から選択的スプライシングにより生成された転写産物アイソフォーム同士など)、多くのリードが複数のcDNA配列にアライメントし、リードがどの転写産物由来なのかを正しく判断することは必ずしも容易ではない、という事である。前回の記事では、各リードと転写産物のマッピング対応、及び各転写産物の転写量を統計的推定により同時に最適化することで、より正確な転写量推定が可能となることを紹介した。

いまや\$1000ゲノムの時代と言われるほど身近になった全ゲノムシーケンスデータ解析においても「ショートリードアライメントの曖昧さ」から由来する様々な困難な問題がある。例えばゲノム中に相同な配列を持つ遺伝子、パラログ、擬遺伝子等が複数存在する場合は、リードをリファレンスゲノムに正確にアライメントすることが困難となる。特に第6染色体短腕上には擬遺伝子も含めると200以上のHLA遺伝子座が存在し、全ての遺伝子座について合計すると10,000種類以上の多型 (HLAアレルの種類) がこれまでに知られている^[1]。従って、個人の全ゲノムシーケンスデータが得られたからといって、各リードを正しい場所にアライメントし、そのアライメント結果からHLA型を正確に決定することは必ずしも容易ではない。今回の解説記事では、統計的モデリング及び最適化手法を利用することで全ゲノムシーケンスから得られたショートリードをHLAアレルリファレンス配列に適切にアライメントを行い、その結果から全HLA遺伝子座のHLAタイピングを同時に行う手法HLA-VBSeq^[2]について紹介する。

HLA 遺伝子とは

HLA 遺伝子座群は第6染色体短腕6p21.3の約4 Mbの領域に渡って存在し、この領域はヒトゲノム中で最も遺伝子密度が高い領域である。また、200以上存在するHLA 遺伝子座はお互いに配列相同性が高く、かつヒト遺伝子の中で最も多様であることが知られている^[3]。その中でも特に重要なHLA 遺伝子がHLA-A、-B、-C (MHCクラスI分子をコードする)、及びHLA-DP、-DQ、-DR (MHCクラスII分子をコードする) である。MHCクラスI分子は内因性抗原を抗原提示し、クラスII分子は外来性抗原を抗原提示することで、免疫系が自己と非自己の認識を行う。従って、臓器移植においては拒絶反応を避けるためにドナーとレシピエントのこれらのHLA型をなるべくマッチさせることが重要となる。また、特定のHLA型が自己免疫疾患発症のリスクを高めること、薬の副作用に関わることが知られている。HLA 遺伝子座の複雑さ、多様性は進化的にも興味深いトピックであると共に、HLA型を簡便かつ正確に決定する手法が医療の現場でも必要とされている。

HLAアレル命名規則

図1 は日本組織適合性学会HLA標準化委員会によるHLAアレル命名規則である。第1区域 (もしくはtwo-digit resolution) は血清学的に分類されるアレルグループである。第2区域 (もしくはfour-digit resolution) は同じアレルグループのうち、アミノ酸変異を伴う非同義置換を区別したサブタイプである。

第3区域（もしくはsix-digit resolution）はアミノ酸変異を伴わない同義置換を区別したサブタイプである。第4区域（もしくはeight-digit resolution）は非コードDNA領域での変異を区別したサブタイプである。臓器移植におけるHLA型マッチングでは第2区域まで一致していることを条件にする場合が多いが、コーディング領域外の変異も含めた表現型への影響を調べるためにも、第4区域までの完全なHLA型を簡便かつ正確に決定できることが望ましい。

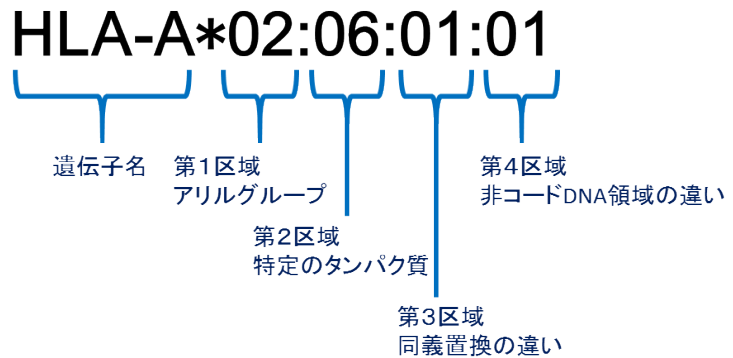


図1 HLAアリル命名規則

HLAタイピング手法の比較

最も古典的なHLAタイピング手法は血清学的にアリルグループを決定する方法である。例えば、ある検体由来の血清と、HLAの特異性が既知の抗血清を混ぜて抗原抗体反応を見ることで、アリルグループを決定する。しかしながら血清学的検査の欠点として、手間とコストが掛かる、第1区域までのHLA型までしか決定することが出来ない、といった点が挙げられる。そのため検体のHLA遺伝子領域のDNA配列を直接調べることによってHLA型を決定する手法が開発されてきた。比較的ハイスループットであるPCR-SSO (Sequence Specific Oligonucleotide) 法はPCRにより特定の遺伝子座を増幅した後、特定の相補的なプローブとのハイブリダイゼーションのパターンをみることでHLA型を決定する方法である^[4]。PCR-SSO法では既知のHLA型を第2区域まで決定することが可能であるが、新規のアリル、もしくは低頻度なアリルには対応していない。増幅した遺伝子座の塩基配列を直接シーケンシングにより決定するPCR-SBT (Sequence Based Typing) 法はサンガーシーケンシングの手間が掛かることが欠点である。

近年、次世代シーケンサ (NGS) の登場により、大量のシーケンシングをハイスループットで行うことが可能になった。しかしながら前述のPCR-SSO法、PCR-SBT法、そしてNGSによるターゲットシーケンス手法はPCR増幅を伴うが故の欠点がある。HLA遺伝子座は第6染色体上に200以上も存在し、かつお互いに似た塩基配列を持つため、特定の遺伝子座を特異的に増幅できるプライマー配列を設計することが難しい。そのため、特定のHLA遺伝子座をPCR増幅しようとしてもオフターゲットシーケンスが増幅されてしまう問題がある^[5]。また、HLAクラスII遺伝子は全長が1万塩基近くにもなるため、全長をPCR増幅することは難しく、実際は特定のexon領域のみのシーケンス結果からHLA型判定が行われることが多い。また仮に、ある特定のHLA遺伝子座を特異的にPCR増幅することが出来たと仮定しても、NGSで得られたショートリードの情報のみからではHLA型 (2つのHLAアリル) を正確に決定することは難しい。なぜなら変異を正確にフェージングして、かつその得られたフェーズ済みの塩基配列を元に、数百種類以上に及ぶHLAアリルのうち正しいと思われるアリルを2つ (ヘテロ接合の場合) 決定しなければならないからである。現在のところ、変異のフェージングに関しては、変異のフェージングをde novoアセンブリによって推定する手法^[5]、HLAアリル参照配列 (IMGT/HLAデータベース^[1]) へのアライメントによって推定する手法^[2,6,7]等が提案されている。

今ではPCR-freeのプロトコルを利用する全ゲノムシーケンシングが可能となっているため、ここから得られたデータを使用することで、200以上に及ぶ全てのHLA遺伝子座のHLAアリル (HLA型) を全て決定することが理論的には可能なはずである。しかしながら上述の通りHLA遺伝子座の複雑さ、多様性のため、通常的全ゲノム・リシーケンシング解析パイプラインではHLAタイピングを行うことは非常に難しい。一方、近年、著者らが開発したHLA-VBSeq^[2]は、全ゲノムシーケンシングから得られたショートリードからHLAタイピングを行う。本手法の利点は、全ゲノムシーケンシングデータを使用するためHLA遺伝子座

に特異的なプライマーを設計する必要が無いこと、検体の ethnicity 及び集団特異的なアレル頻度の事前情報が不要なこと、そして全HLA 遺伝子座について一貫通貫にタイピング出来ることである。一方、欠点は、全ゲノムシーケンスのコストが高いことである（しかし近い将来、コストが他の手法以下になる可能性は十分ある）。各手法の利点・欠点を表1 にまとめた。

HLA-VBSeq 概要

HLA-VBSeq^[2] は、全ゲノムシーケンスから得られたショートリードをIMGT/HLAデータベース^[1] に登録されている全てのHLAアレル参照配列に対してマルチマッピングを許したアライメントを行った後、統計的推定によってリードマッピングを最適化し、HLAタイピングを行う。IMGT/HLAデータベース^[1] に登録されている既知のHLAアレル参照配列は10,000種類以上も存在するが、HLAアレルはお互いに塩基配列が非常に良く似ているため、アライメント先が唯一に決まらない場合がほとんどである(図2)。

他の既存手法^[6,7] ではショートリードのHLAアレル参照配列へのベストアライメントのみを利用して、そのアライメント結果から各HLA 遺伝子座について最も適当だと思われるHLAアレルを2つ(ヘテ

表1 各HLAタイピング手法の比較

HLAタイピング手法	長所	短所
血清学的タイピング	抗原抗体反応を直接観察するため、正確	手間、コストが掛かる。第1区域までのタイピング
PCR-SSOP	比較的ハイスループット	第2区域までタイピング、新規アレルに対応していない。プライマー設計、PCR増幅が必要
PCR-SBT	新規アレルにも対応	手間が掛かる。プライマー設計、PCR増幅が必要
NGSによるターゲットシーケンス	ハイスループット、ターゲット領域の配列決定が可能	プライマー設計、PCR増幅が必要、全長の配列決定が困難
HLA-VBSeq (全ゲノムシーケンス)	ハイスループット、第4区域までタイピング可能、新規アレルにも対応、プライマー設計及びPCR増幅が不要	全ゲノムシーケンスのコスト

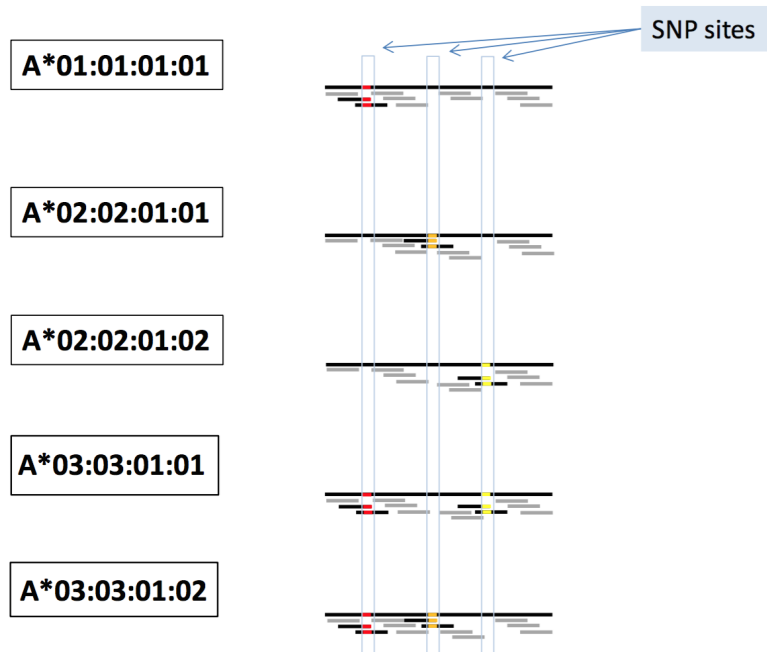


図2 HLA-A由来のショートリードのアライメント箇所が一意に決まらないケースの模式図。図の例では、どのリードも複数のHLAアレルにアライメント先候補が存在する

ロ接合の場合) 推定する。しかしながら図2の通り、ベストなアライメントが一意に定まらない場合がほとんどなので、ショートリードの大部分の情報を捨ててしまっていることになる。結果として、検体のHLAアレル由来のリード数を実際よりも過小評価することになるので、HLAアレルの決定、およびgene dosage (ヘテロ接合か、ホモ接合か) の推定パフォーマンスが下がる原因となる。

一方、HLA-VBSeqでは、各リードと各HLAアレルのマッピング対応を隠れ変数、各HLAアレルのgene dosage (相対的な存在量) を多項分布のパラメータとして、変分ベイズ法によって統計的推定を行う。本手法の特徴は、リードのアライメント先が複数ある場合でも、統計的推定により正しいアライメント先を見つけ、曖昧なアライメント状態にあるリードを一切捨てることなくHLAアレル及びgene dosageの推定を行うことである。以下、パラメータ推定のアルゴリズムの概要を説明する。

変分ベイズ推定アルゴリズム^[2]

ステップ1: 各リード n の各HLAアレル t への期待マッピング数 $E[Z_{nt}]$ を初期化する。リード n があるHLAアレル t に唯一アライメントする場合は $E[Z_{nt}]=1$ である。リード n が複数のHLAアレル t_1, \dots, t_m にアライメントする場合、それぞれの t_i ($i=1, \dots, m$) について、 $E[Z_{nt_i}]=1/m$ とする。

ステップ2: ステップ1で計算したリード毎のHLAアレルへの期待マッピング数 $E[Z_{nt}]$ を、HLAアレル毎に合計し、それぞれのHLAアレル t の合計期待マッピング数 r_t を求める。ここで $r_t = \sum_n E[Z_{nt}]$ である。合計期待マッピング数から、それぞれのHLAアレル t の期待アレル頻度 $E[\theta_t]$ を求める。

ステップ3: 現在のHLAアレル毎の期待アレル頻度を所与として、 $E[Z_{nt}]$ を再計算する。その際、リード n について複数のアライメント先の可能性がある場合、現在の期待アレル頻度が大きいHLAアレルに、より重み付けされる様に $E[Z_{nt}]$ がアップデートされる。

ステップ4: ステップ3で計算したリード毎の期待マッピング数を各HLAアレルごとに合計し、合計期待マッピング数及び期待アレル頻度を再計算する。全てのHLAアレルについて期待アレル頻度が前回の値から変化していなければ、アルゴリズムを終了する。そうでなければステップ3に戻る。

バイオインフォマティクスや機械学習に詳しい読者は、本アルゴリズムが、データを k 個にクラスタリングする k -means clusteringのアルゴリズムと類似していることに気が付くかもしれない(実際、関連がある)。しかしながら、 k -meansアルゴリズムは事前にクラスタの個数 k (この場合、検体に実際に存在するHLAアレルの種類数。考慮する遺伝子座、そしてそれぞれの遺伝子座がヘテロ接合か、ホモ接合かによって存在するHLAアレルの種類数は異なる)を指定する必要がある。本アルゴリズムでは変分ベイズ推定の枠組みの中でモデル選択(実際に検体に存在するHLAアレルの個数の選択)が自動的に行われるため、 k をあらかじめ指定する必要が無い。この「検体に実際に存在するHLAアレルの種類数が自動的に推定される」という特徴により、例えば検体のある遺伝子座のHLA型がホモ接合か、ヘテロ接合か、というgene dosageがモデル選択の枠組みの中で自動的に判定されることになる。HLA-VBSeqの統計的モデリングの詳細、及び変分ベイズ法についての詳細は、参考文献^[2,8,9]を参照いただきたい。

ここでは上記アルゴリズムをイメージ的に理解しやすくするため、リードが全てHLA-A遺伝子座から得られていることが分かっている場合を仮定して、HLA-Aのアレルのみを推定する場合を考える。最初に全てのリードを曖昧なアライメントを全て許した状態でHLA-Aアレル参照配列にアライメントを行った状態が前述の図2である。前述アルゴリズムのステップ3、4の繰り返しによって、各リード毎の各HLAアレルへの期待マッピング数と、各アレル毎の合計期待マッピング数を最適化し、パラメータ(各HLAアレルのgene dosage)推定値が収束した後の状態を図3に示す。

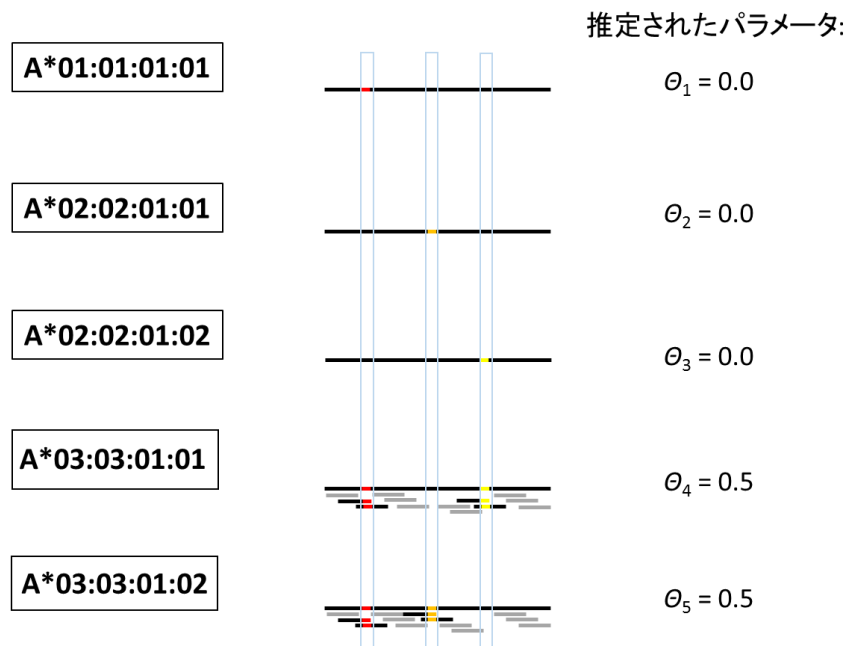


図3 各リードのアライメント先を隠れ変数、HLAアレルの存在量をパラメータとして、統計的推定を行った結果の模式図。

アルゴリズムが収束すると、各リード毎の正しいアライメント先、及び各HLAアレル毎のリードの合計期待マッピング数が推定された状態になっている。今回のケースではこの結果から、検体のHLA-A型はA*03:03:01:01/A*03:03:01:02のヘテロ接合である、と決定できる（一方、仮に検体がA*03:03:01:01のホモ接合の場合は、A*03:03:01:01の $\theta_4=1.0$ として推定されるはずである）。また、結果として、本来はショートリードの情報のみでは難しいリードのフェージングも自動的に解決されている。図2、3では簡単のためHLA-Aの遺伝子座についてのみ例示したが、全ゲノムシーケンズデータを利用する場合は全てのHLA遺伝子座の全てのHLAアレルについて同時に最適化を行う。以降、実際の全ゲノムデータを利用してHLA-VBSeqによりHLAタイピングを行う例を紹介する。

HLA-VBSeqの使用方法

イルミナ社Platinum Genomesウェブサイト (<http://www.illumina.com/platinumgenomes/>) ではHiSeq2000で読んだ高深度(40x)全ゲノムシーケンズデータが公開されている。今回、1000人ゲノムプロジェクトのサンプルとしても含まれているNA12878(子)、NA12891(父)、NA12892(母)のデータを利用して、HLA-VBSeqを利用してそれぞれの全HLA遺伝子座のHLAタイピングを行う。以下の例では、NA12878のFASTQデータをBWA-MEMによりhg19にアライメントして、位置でソート済みのNA12878.sorted.bamというファイルが作成されていると仮定する。

- 1) HLA遺伝子座 (HLA-A, B, C, DM, DO, DP, DQ, DR, E, F, G, H, J, K, L, P, V, MIC, 及びTAP) にアライメントしているリード名を取得する

```
$ samtools view NA12878.sorted.bam chr6:29907037-29915661 chr6:31319649-31326989
chr6:31234526-31241863 chr6:32914391-32922899 chr6:32900406-32910847
chr6:32969960-32979389 chr6:32778540-32786825 chr6:33030346-33050555
chr6:33041703-33059473 chr6:32603183-32613429 chr6:32707163-32716664
chr6:32625241-32636466 chr6:32721875-32733330 chr6:32405619-32414826
chr6:32544547-32559613 chr6:32518778-32554154 chr6:32483154-32559613
```

```
chr6:30455183-30463982 chr6:29689117-29699106 chr6:29792756-29800899
chr6:29793613-29978954 chr6:29855105-29979733 chr6:29892236-29899009
chr6:30225339-30236728 chr6:31369356-31385092 chr6:31460658-31480901
chr6:29766192-29772202 chr6:32810986-32823755 chr6:32779544-32808599
chr6:29756731-29767588 | awk '{print $$1}' | sort | uniq >
NA12878_partial_reads.txt
```

2) リード名でBAMファイルのインデックスを作る

```
$ java -jar -Xmx32g -Xms32g bamNameIndex.jar index NA12878.sorted.bam --indexFile
NA12878.sorted.bam.idx
```

3) インデックスを基に、HLA領域にアライメントしているリードをSAMファイルの形式で抽出

```
$ java -jar bamNameIndex.jar search NA12878.sorted.bam --name
NA12878_partial_reads.txt --output NA12878_partial.sam
```

4) SAMファイルから、FASTQファイルを作成

```
$ java -jar SamToFastq.jar I=NA12878_partial.sam F=NA12878_partial_1.fastq
F2=NA12878_partial_2.fastq
```

5) アンマップリードを抽出、FASTQファイルを作成

```
$ samtools view -bh -f 12 NA12878.sorted.bam > NA12878.sorted_unmapped.bam
$ java -jar SamToFastq.jar I=NA12878.sorted_unmapped.bam
F=NA12878_unmapped_1.fastq F2=NA12878_unmapped_2.fastq
```

6) HLA 遺伝子座由来のリードと、アンマップリードのFASTQファイルを結合する

```
$ cat NA12878_partial_1.fastq NA12878_unmapped_1.fastq > NA12878_part_1.fastq
$ cat NA12878_partial_2.fastq NA12878_unmapped_2.fastq > NA12878_part_2.fastq
```

7) IMGT/HLAに登録されているHLAアレルリファレンス配列 (hla_all.fasta) のインデックスを作成

```
$ bwa index hla_all.fasta
```

8) BWA-MEMで、HLAアレルリファレンス配列にアライメント

("-a" オプションにより、曖昧なリードアライメントを全て保持する)

```
$ bwa mem -t 8 -P -L 10000 -a hla_all.fasta NA12878_part_1.fastq
NA12878_part_2.fastq > NA12878_part.sam
```

9) HLA-VBSeqにより、HLAアレルのgene dosageを推定する

```
$ java -jar HLAVBSeq.jar hla_all.fasta NA12878_part.sam NA12878_result.txt
--alpha_zero 0.01 --is_paired
```

上記、パイプライン実行に必要なプログラム、及びマニュアルはウェブサイトにて公開している (<http://nagasakilab.csml.org/hla/>)。

結果

全てのリードを全てのHLAアレルにアライメントを行い、HLAアレル毎の合計期待マッピング数の初期状態を求めた結果 (アルゴリズムのステップ2の状態) が図4aである。この段階では、多くのリードが複数のHLAアレルにアライメントしているため、正しいHLAアレルを決定することは難しい。図4b, cは、変分ベイズ法のステップ3と4を10回、及び100回繰り返した後の、HLAアレル毎の合計期待マッピング数である。アルゴリズムが収束した後の結果から、各HLA遺伝座ごとに、上位2つのHLAアレル

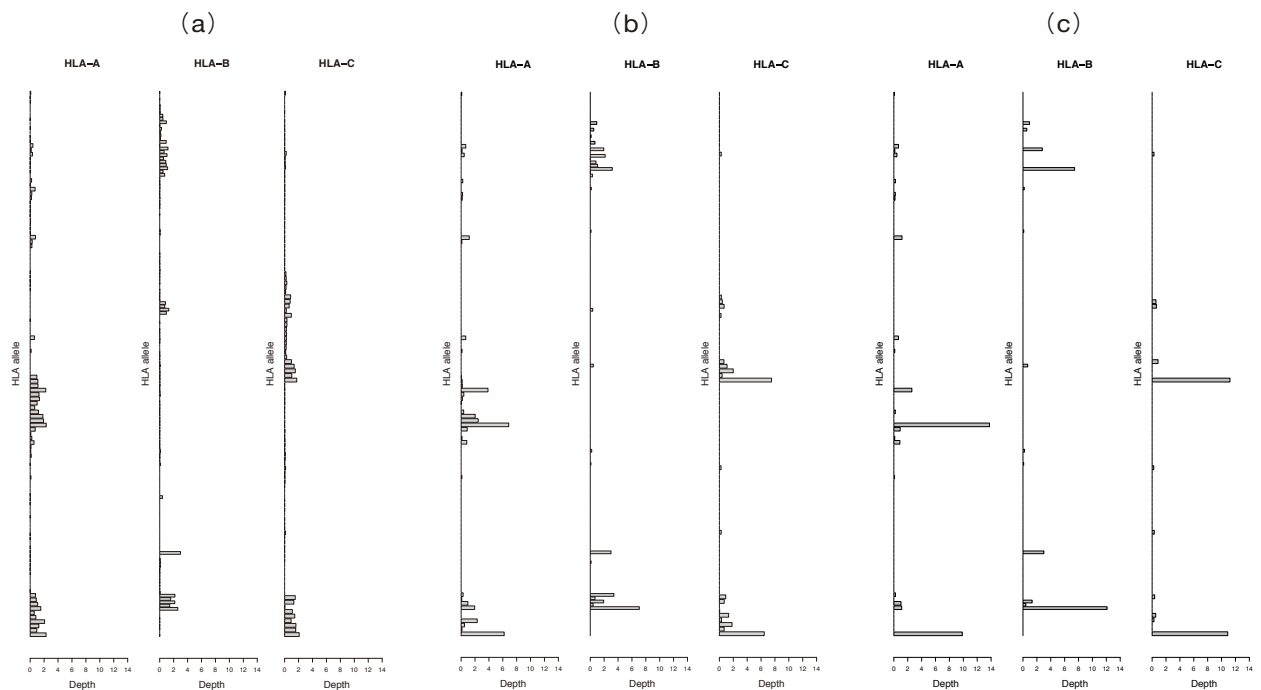


図4 変分ベイズ推定を行う前 (a)、10回ステップを繰り返した状態 (b)、100回ステップを繰り返した状態 (c) (x軸は各アレルに割り当てられたリードの深度、y軸は各遺伝子座ごとのHLAアレルを示す)

表2 HLA-VBSeqにより推定されたHLA型

検体	推定されたHLA型
NA12878 (子供)	A*01:01:01:01 / A*11:01:01 B*08:01:01 / B*56:01:01 C*01:02:01 / C*07:01:01:01
NA12891 (父親)	A*01:01:01:01 / A*24:02:01:01 B*07:02:01 / B*08:01:01 C*07:01:01:01 / C*07:02:01:03
NA12892 (母親)	A*02:01:01:01 / A*11:01:01 B*15:01:01:01 / B*56:01:01 C*01:02:01 / C*04:01:01:01

※太字は別の手法 (Erlich et al., 2011) により決定された第2区域までのHLA型と一致していることを示す

を選択することにより、HLA型を決定する（ホモ接合の場合は上位1つ）。本手法によりHLA-A、-B、-Cをタイピングした結果を表2にまとめた。本手法による推定結果は、過去に当該検体についてターゲットシーケンス法で決定されたHLA型と第2区域について一致していることが確かめられた（今回、HLA-VBSeqでは第4区域までHLA型を決定出来たが、第4区域までの実際のタイピングはこれまでに別の方法では行われていない）。

尚、本手法を健常な日本人集団（1,070人）の全ゲノムシーケンスデータに適用した結果、過去に別の大規模な日本人集団からPCR-SSOP手法によって得られたHLAアレル頻度と非常に似た分布が得られている^[10]。

まとめ

本記事では全ゲノムシーケンスデータからHLAタイピングを行うHLA-VBSeqのアルゴリズム概要、使用方法、実際のデータを使用した実行結果を紹介した。本手法は曖昧なりードアライメントを捨てることなく有効活用することで、全HLA遺伝子座について同時にHLAタイピングを実行することが可能である。今回紹介したショートリードを利用したHLAタイピング手法以外にも、例えばPacBio RS IIなどのロングリードシーケンサを利用したターゲットシーケンシングによるHLAタイピングが考えられる。ロングリードを利用することで、変異のフェージングの問題が解決出来る。しかしながら、上述の通りPCR増幅の問題、ロングリード特有の高いエラー率の扱い方など、別のチャレンジングな問題がある。また、HLA領域のハプロタイプに関する事前情報を所与として、SNPアレイデータの限られたジェノタイプ情報からHLA型をimputation（推定）する手法が提案されている^[11,12]。SNPアレイによるジェノタイプングはシーケンシングによる配列決定と比べてコスト、データの扱いやすさの観点で、利点がある。しかし上記のアプローチでは、検体のethnicityに応じて適切なHLAハプロタイプの事前情報を準備し、かつSNPアレイに関しては集団特異的なtagSNPの設計^[13]が成されていることが、高精度のHLA imputationのためには望ましい。また、PCR-SSOP法と同様、通常第2区域までの推定に限られており、新規のアレルの決定は出来ない。

HLAタイピングのために全ゲノムシーケンスデータを利用するHLA-VBSeqでは、1) HLA遺伝子座ごとに特異的なプライマーをデザインすることが不要、2) 検体のethnicity、及びHLAアレル頻度に関する事前知識が不要、3) 全HLA遺伝子座について同時にHLAタイピングが可能、という利点がある。一方で、本手法はIMGT/HLAデータベース等に登録されているHLAアレルリファレンス配列が正しく、かつ完全であると仮定してHLA型の推定を行うため、そうではない場合、推定精度が落ちる可能性がある（実際、現段階では完全なHLAアレルリファレンスデータベースは存在しないと考えられる）。また現在、IMGT/HLAデータベースに登録済みのHLAアレルはEuropean ancestryに高頻度なHLA型に偏っていると考えられるため、今後は日本人集団に特異的なHLAアレルリファレンス配列を整備していくことが望まれる。

参考文献

- [1] Robinson et al., *Nucleic Acids Research*. (2015) 43:D423-431.
- [2] Nariai et al., *BMC Genomics*. (2015) 16 Suppl 2:S7.
- [3] The MHC sequencing consortium, *Nature*. (1999) 401(6756):921-3.
- [4] Ito et al., *Immunogenetics*. (2005) 57(10):717-29.
- [5] Hosomichi et al., *BMC Genomics*. (2013) 14:355.
- [6] Warren et al., *Genome Medicine*. (2012) 4:95.
- [7] Bai et al., *BMC Genomics*. (2014) 15:325.
- [8] Nariai et al., *Bioinformatics*. (2013) 29 (18):2292-2299.
- [9] Bishop CM., *Science+Business Media, LLC, New York, NY, USA*. (2006).
- [10] Nagasaki et al., *Nature Communications*. (2015) 6:8018.
- [11] Okada et al., *Nature Genetics*. (2015) 47, 798-802.

[12] Khor et al., *The Pharmacogenomics Journal*. (2015) 15, 530-537.

[13] Kawai et al., *Journal of Human Genetics*. (2015) 60, 581-587.

(編集担当：佐藤行人)

第4回 シリーズ「大量データと知見の架け橋」

どのアセンブリを使うか？： 分子系統学的観点に基づくアセンブリの評価

原 雄一郎 (理化学研究所 ライフサイエンス技術基盤研究センター 分子配列比較解析ユニット)



はじめに

DNAシーケンサの超ハイスループット化、シーケンシングコストの大幅な低下、さらにバイオインフォマティクスの発展にともない、ヒトやモデル生物以外のいわゆる“非モデル生物”も大量配列解析の恩恵に与るようになった。地球上の様々な生命現象のなりたちを解き明かすことを使命とする進化学者にとって、自身が研究対象とする生物を分子配列のすみずみまで知りつくす時代が来たと言っても過言ではないだろう。一方で、リファレンスゲノムはもちろん、往々にしてゲノムサイズや核型の情報も乏しい生物から、ゲノムやトランスクリプトームの配列を再構築するには、多大な労力と様々な工夫が必要とされる。それに加えて、再構築された配列セット（アセンブリ）が実際のゲノムやトランスクリプトームにどの程度完全に迫れているかを、誰も正答を知らないまま評価しなければならない。アセンブリをめぐるこれらの問題は、多様な生物種のゲノム解読が進むにつれ多くの研究者を悩ませるようになってきた。これらを解決すべく、様々なアセンブリ作成方法ならびに評価方法をベンチマークし、より高品質なアセンブリの構築を目指す国際的なジャンボリー“*Assemblathon*（アセンブラソン）”が開催されてきた^[1]。

筆者は、爬虫類や軟骨魚類をはじめとする、様々な非モデル生物の配列データを産出するシーケンシングファシリティに所属している。これら生物のゲノム、トランスクリプトームアセンブリを作成するなかで、より正確にアセンブリを評価するためには、対象とする分類群に起きたゲノム進化を考慮した方法が必要であることに気づいた。そこで、*Assemblathon*によって提案された方法を出発点とし、様々な検討を経て、分子系統学的思考に基づくアセンブリの評価にたどり着いた。本稿では、ゲノム、トランスクリプトームアセンブリの完全度の評価を簡潔にレビューするとともに、筆者がアセンブリの完全度評価のために作成した「脊椎動物に特化した1-to-1オーソログセット（*Core Vertebrate Genes; CVG*）」を紹介する。

N50, NG50：配列長による指標

ゲノム・トランスクリプトームアセンブリを作成するには、超並列DNAシーケンサから得られた断片配列（リード）をつなぎ合わせる（*de novo*アセンブル）ことが現在の主流である。現在のシーケンシングおよび*de novo*アセンブリの技術では、染色体の全領域を1本の配列につなげることは困難であり、アセンブリは多数の部分配列（スキップフォールド、コンティグ）より構成されることがほとんどである。この断片的なアセンブリがどれくらい完全に近いかを評価するにあたり、つながり具合すなわち配列長に着目することは直感にもよくあっている。配列長に基づいて定義されるN50は、最も広く知られているアセンブリ評価の指標である。アセンブリの全配列を短い順（あるいは長い順）から長さを足していき、この値がアセンブリの全長の半分に達するときに加える配列がもつ長さ、これがN50である。一般的には、N50が大きいほどよくつながった高い完全度を持つアセンブリであると考えられる。

ゲノムサイズが推定されている生物のゲノムアセンブリには、N50の代わりにNG50（同様の手順で足していった配列長の和がゲノムサイズの半分に達したときのコンティグの長さ）を評価基準に用いること

ができる。アセンブリのサイズを基準とするN50とは異なり、ゲノムサイズを基準とするNG50は、同種のゲノムアセンブリに単一の基準を与えるため、同種アセンブリ間の比較により適している。ただし、トランスクリプトームにおいてNG50の基準に相当する「転写された遺伝子の配列長の和」をあらかじめ知ることは困難であるため、トランスクリプトームアセンブリにはN50の評価に限られる。N50, NG50を含むアセンブリの配列長に基づく様々な指標を計算するには、Assemblathon 2に用いられたPerlスクリプトが便利である (<https://github.com/ucdavis-bioinformatics/assemblathon2-analysis>)。

遺伝子の復元度：より生物学に寄り添った指標

N50やNG50は、その簡便さとの引き替えに、ゲノム、トランスクリプトーム配列に刻まれた生物学的な情報を失っている。このため、これらの尺度が持つ生物学的な意義はきわめて限定的である。ゲノム解読の第一の目的は遺伝子のカタログ化であり、推定された遺伝子情報をもとに分子生物学実験などによってその機能が解明される。したがって、アセンブリに復元される遺伝子数に基づいて完全度を評価することは、ゲノムが持つ生物学的特徴やアセンブリの有用性により即していると言えよう。この評価の手段として、カリフォルニア大学デービス校のIan Korfらが開発し2007年に発表したオーソログ同定パイプラインCEGMA (Core Eukaryotic Gene Mapping Approach) が広く知られている^[2,3]。

ここでCEGMAを簡単に紹介する。CEGMAは、アセンブリの配列にリファレンス遺伝子のホモログを予測し、類似度に基づきホモログがオーソログであるかを判定する。CEGMAの元来の目的は、遺伝子予測のトレーニングセットとして用いるための既知遺伝子のオーソログセットを得ることであり、真核生物に広く保存される458群の遺伝子 (Core Eukaryotic Genes, CEGs) が既知遺伝子のリファレンスセットとして用意されている。CEGMAをアセンブリの完全度の評価に用いるときには、458群のうち特に遺伝子重複が無い、もしくは少ないとされる248群 (248 CEG) をリファレンスとし、アセンブリ中に復元されるオーソログの数を測る。Assemblathonは、このCEGMAによる指標をアセンブリの完全度を示す尺度の一つとして提案している。さらに、この評価の基準 (248 CEG) はヒト、ハエ、線虫、分裂酵母、出芽酵母、シロイヌナズナという幅広い系統の真核生物に共有されるため、N50, NG50とは異なりアセンブリの完全度を種間で比較できる。これまでに、様々な真核生物のゲノム・トランスクリプトームアセンブリの評価にCEGMAは用いられてきた。

CEGMAには、ほぼ完全長をもって復元されたりファレンス遺伝子の割合 (Complete)、完全長、部分領域ともに含む復元されたりファレンス遺伝子の割合 (Partial) の2つの指標がある。今後本稿で述べる完全度は前者を指す。これら以外に、各リファレンス遺伝子の平均オーソログ数もゲノムアセンブリの評価に有用な指標である。平均オーソログ数が1に近づくほど冗長性が少ないアセンブリであると言える。CEGMAはウェブ上に公開されており、LinuxおよびMacのターミナル上で実行できる (<http://korflab.ucdavis.edu/datasets/cegma/>)。CEGMAの実行にはNCBI BLASTなどいくつかの既存プログラムを別途インストール必要がある。これらのうちGeneWiseのインストールが事前情報無しでは非常に難しく、そのことがCEGMAの使用を遠ざける要因となっている。GeneWiseのインストールログを公開しておくので、適宜参考にされたい (http://www2.clst.riken.jp/phylo/genewise_installation.html。インストールするGeneWiseは、最新版のバージョン2.4.1ではなくCEGMAのウェブページからダウンロードできるバージョン2.2.3-rc7を推奨する)。

筆者らは、遺伝子の復元度がアセンブリの評価に優れた尺度であると認識している。例えば、アセンブリに膨大に含まれる“うまくつながらなかった”ごく短い配列を取り除くことはよくある。本質的にはこの操作によってゲノムの完全度は変化しないにもかかわらず、N50は向上する。しかし、取り除いた短い配列に遺伝子が完全長をもって復元されることはまずないため、CEGMAに基づく評価に変化は無い。では、評価の基準となるリファレンス遺伝子も妥当であると言えるだろうか。筆者らはCEGのあまりにも広い分類群を対象にする有効性に疑問を感じていた。筆者らが配列を解読している脊椎動物のゲノムには、

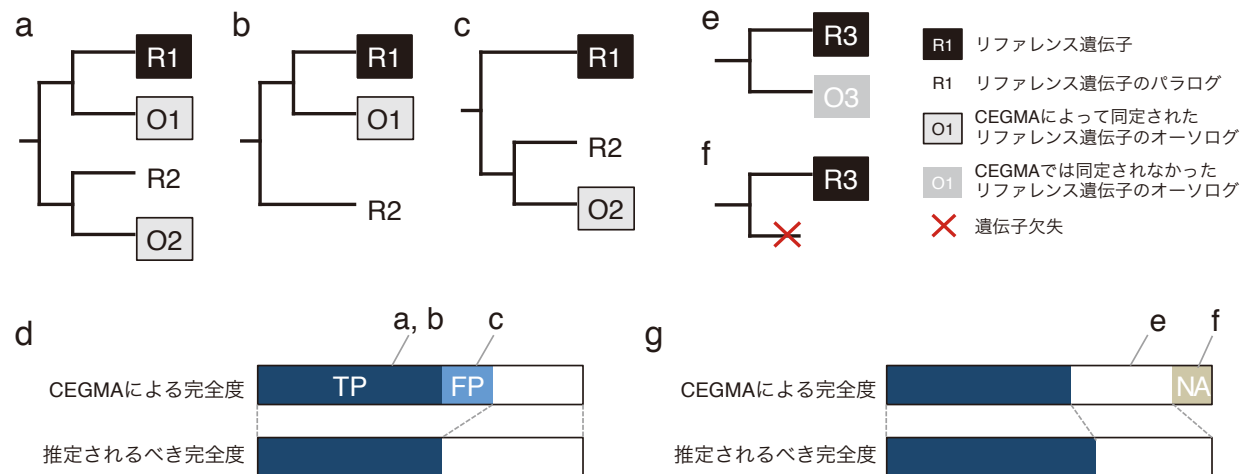
後生動物に至る過程や脊椎動物の初期に重複した遺伝子が多数存在することは言うまでもないが、遺伝子重複が少ない、あるいは無いという理由で選ばれた248 CEGMのうち64の遺伝子群にもそのような遺伝子重複が起きていることがわかったのである。CEGMAは、CEGMに含まれるパラログのうち1つのみリファレンス遺伝子として扱い、残りを無視する。アセンブリにリファレンス遺伝子のオーソログが含まれず「無視された重複遺伝子」のオーソログが存在する場合に、CEGMAはこの遺伝子を「リファレンス遺伝子のオーソログ」として誤検出し、結果として完全度の過大評価につながる(図1)。

また、CEGMを構成する生物種が代表する分類群は互いに遠縁であるため、それら以外の分類群に起きた遺伝子進化を考慮に入れない。たとえば、CEGMAは評価対象のアセンブリに特定の遺伝子が無いことを判定できるが、進化における遺伝子欠失によるものはたまたアセンブリの不完全性によるものを区別することはできない。したがって、進化においてリファレンス遺伝子を欠失したゲノムの完全度は過小に評価されてしまう(図1)。以上の理由から、オーソログ同定に用いるリファレンス遺伝子および生物種の選択は、完全度測定の精度を損ねないように慎重に行われるべきである。一方、これらの問題を考慮したリファレンス遺伝子セットを用いることにより、アセンブリ完全度を高精度に評価できるだろう。

CVG：脊椎動物での完全度測定に適した1-to-1 オーソログ遺伝子セット

改めて、アセンブリ完全度を評価するためにリファレンス遺伝子セットに理想的な条件を確認してみる。あらゆる既知のゲノム配列に1コピーのみ存在するオーソログ(1-to-1 オーソログ)であることがリファレンス遺伝子に求められる。だが、遺伝子重複や欠失という生物学的な理由、あるいはゲノムアセンブリや遺伝子アノテーションの不完全性という技術的な理由から、生物種を増やすほど得られる1-to-1 オーソログの数は減っていく。したがって、分類群の範囲とリファレンスに適する遺伝子数にはトレードオフが生じる。

筆者らのグループは円口類、軟骨魚類を含めた脊椎動物の多様な分類群の配列解析を行っているため、あらゆる脊椎動物のアセンブリの評価に通用するリファレンス遺伝子セットが必要であった。脊椎動物アセンブリの完全度をより正確に評価できるリファレンス遺伝子セットを作成するには、脊椎動物の初期進化に起きたとされる2回の全ゲノム重複の影響を受けていない、あらゆる脊椎動物の系統で1-to-1の関



【図1】 遺伝子重複・欠失がアセンブリ完全度の評価に影響する例

(a, b, c) リファレンス遺伝子が重複遺伝子を持つ場合、アセンブリにリファレンス遺伝子のオーソログ・パラログともに同定される(a)、オーソログのみ同定される(b)、パラログのみ同定される(c) ケースが考えられる。(d) CEGMAによって計算された完全度と実際の完全度。a, bは正しくオーソログを検出できているが(TP)、cはリファレンスのオーソログを見つけていないため誤検出(FP)であり過大評価を引き起こす。(e, f) CEGMAがオーソログを同定しない原因には、実際にはゲノム中に存在するもののアセンブリが不完全である場合(e)と、進化の過程でリファレンス遺伝子に欠失が生じた場合(f)が考えられる。後者は完全度の計算には適用すべきではない(NA)が、CEGMAはこの2つを区別できない。したがって遺伝子欠失が存在する場合には、CEGMAの完全度は実際より低く計算されてしまう(g)。

係をもつオーソログ群を用いることが求められる。そこでワークフロー (図2) に示すように、尾索類ホヤの遺伝子を外群として、主要な脊椎動物の分類群をすべて含む29種に厳密に1-to-1のオーソログ関係をもつ233群の遺伝子を同定し、これをCore Vertebrate Gene (CVG) と名付けた^[4]。この遺伝子セットはCEGMAのリファレンス遺伝子として用いられる。

CVGおよびCEGに基づくアセンブリ評価の精度を比較するために、CEGMAを用いて42種の脊椎動物ゲノムアセンブリ、およびソメワケササクレヤモリのトランスクリプトームアセンブリの完全度を解析した。このソメワケササクレヤモリは、その実験動物としての適性から、モデル爬虫類として確立すべく筆者らによって配列情報が整備されつつある (http://www.riken.jp/pr/press/2015/20151120_3/)。比較解析の結果、CVGに基づく評価はCEGに基づく評価より高精度であることがわかった。CVGによる完全度の評価の精度が高いことは、CVGの3つの特徴により説明できる。1つは、「厳密に1-to-1オーソログであるCVGはパラログによる誤検出の可能性が低い」ことが挙げられる。実際に、ソメワケササクレヤモリのトランスクリプトームアセンブリにおいてオーソログと同定された遺伝子の詳細を調べたところ、2~5%のCEGについてパラログがオーソログと誤検出されていたのに対し、CVGにはそのような誤検出は見られなかった。2つ目は、「CVGは様々な分類群に起きたゲノム進化を考慮できている」ことである。一般的に、CVGに基づく完全度はCEGに基づく完全度より低くなるが、例外的に鳥類ゲノムアセンブリでは、CEGの方がCVGより低い完全度を示す (図3a)。最近発表された鳥類ゲノムの大規模解析により、鳥類の初期進化において多数の遺伝子がゲノムから失われたことが示された^[5]。CEGにはこの欠失が起きた遺伝子が含まれることにより、鳥類のアセンブリ完全度が過小評価されてしまったと考えられる。一方、CVGのすべての遺伝子群は鳥類ゲノム (ニワトリ、ゼブラフィンチ) にも存在するため、鳥類特有の過小評価は起き得ない。3つ目は、「CVGによる完全度はより高い解像度をもつ」ことである。CEGに基づく完全度が85~95%を示すゲノムアセンブリにおいて、CVGに基づく完全度はより多様な値を示す (図3a)。この特徴はトランスクリプトームアセンブリにも同様に観察される (図3b)。CEGの構成遺伝子より長い配列を持つCVGの構成遺伝子を復元するには、よりつながったアセンブリを必要とするため、保守的かつ高解像度な完全度評価を実現できた。

CVGとCEGは、ゲノムだけではなくトランスクリプトームアセンブリの完全度評価にも十分適用できる遺伝子セットであることを付け加えておきたい。転写される遺伝子のレパートリーは採取した組織やタイミングによって大きく異なるため、遺伝子レパートリーに偏りがあるトランスクリプトームアセンブリ

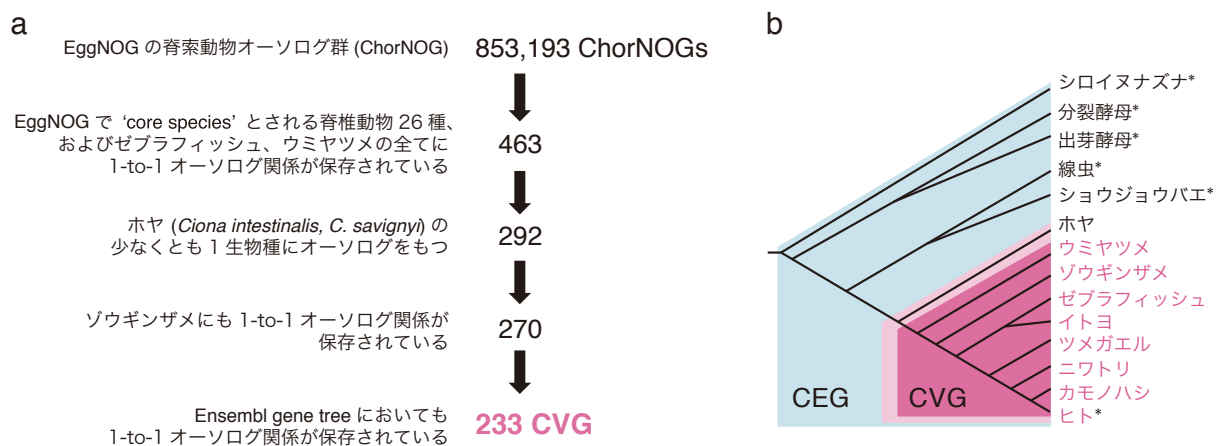


図2 CVGの概要

(a) CVG作成フローチャート。「脊椎」動物オーソログ群ではなく「脊索」動物オーソログ群を用いた理由は、ホヤを外群として脊椎動物初期に起きた遺伝子重複を持つ遺伝子群を除くためである。ゾウギンザメ遺伝子はオーソログデータベースEggNOGに含まれていないため、相同検索における双方向ベストヒットに基づきオーソログを同定した。CVGに含まれるオーソログは、29種全てにおいて1-to-1のオーソログ関係を保持している。(b) CEGとCVGの分類学的範囲。*印がつく生物種によってCEGMAが構成されている。図中の脊椎動物8種は、主要な分類群の代表種としてCVGのCEGMA用データセット作成に用いた。

に普遍的なリファレンス遺伝子を用いて完全度を評価することに意義があるのかと疑問に思われる方もいるだろう。ところが、CVGやCEGのようにレパートリーの変化に保守的な遺伝子には、あらゆる組織に発現するハウスキーピング遺伝子が多く存在する。筆者らが解析してきたトランスクリプトームアセンブリも、適切なシーケンシング、アセンブリ方法を適用すれば、CVG・CEGともに100%もしくは100%に近い完全度を示すことがほとんどであった。また、発現遺伝子が極端に偏るトランスクリプトームであっても、同一のシーケンスデータを用いて作成したアセンブリ方法（プログラムやそのパラメータ）の比較を目的とする、リファレンス遺伝子を用いた完全度の評価は有効である。CVGの作成手順およびトランスクリプトームアセンブリに関する詳細は筆者らが出版した論文^[4]を参照されたい。また後述するように、CVGのデータセットおよびカスタム遺伝子セットに対応したCEGMAプログラムを研究室ホームページで公開している。

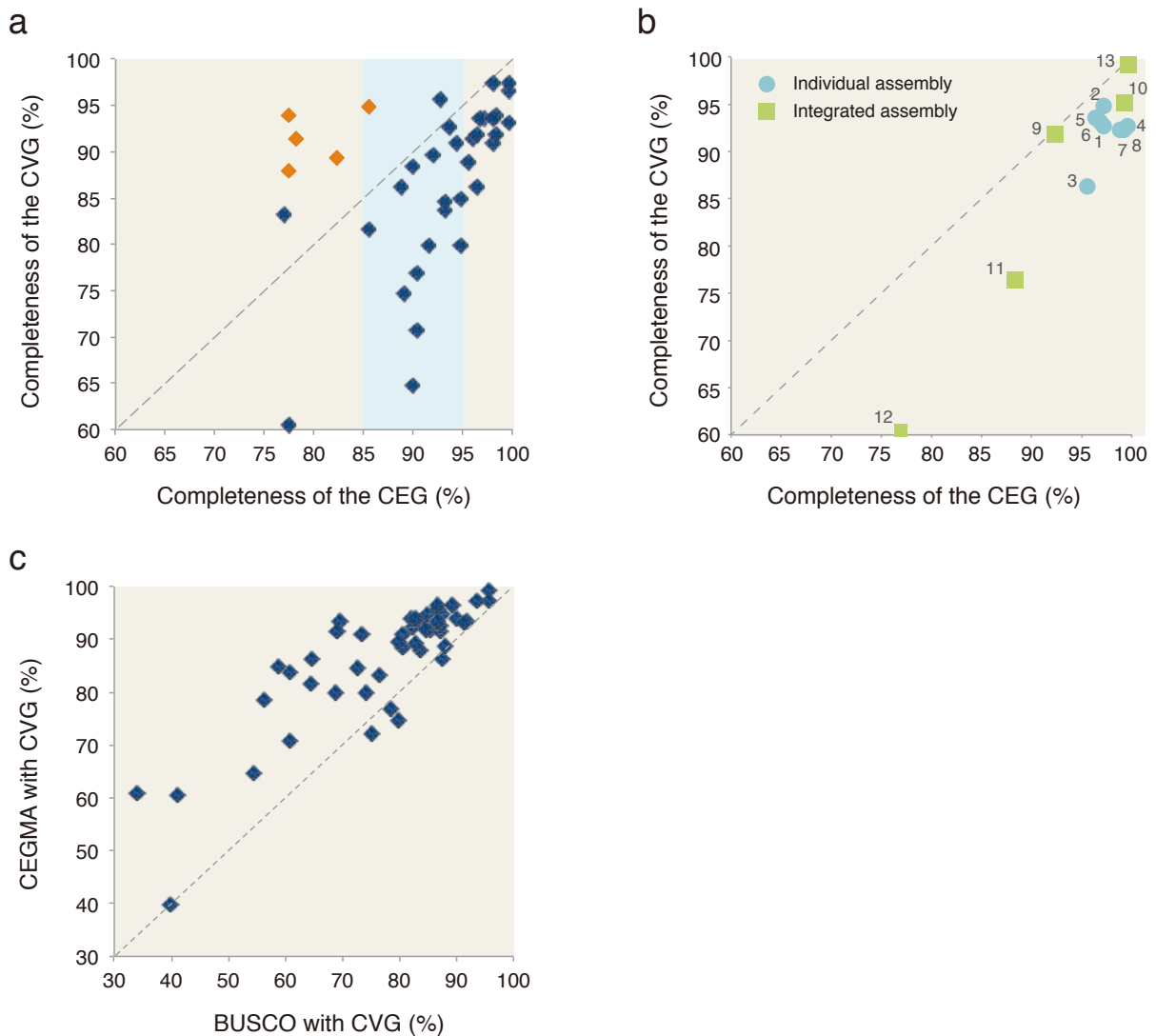


図3 CVGおよびCEGのベンチマーク

(a) 42種の脊椎動物ゲノムアセンブリを用いたCVG, CEGの評価。各点が1つのアセンブリを表している。アセンブリの完全度には、CEGMAによってほぼ完全に復元された(Complete)リファレンス遺伝子の割合を用いた。CEGに基づく完全度が85~90%の領域を水色の背景、鳥類ゲノムアセンブリを橙色で示している。図の見やすさのため完全度が低いアセンブリを除いている。(b) ソメワケササクレヤモリのトランスクリプトームアセンブリを用いたCVG, CEGの評価。アセンブリ1~8は、RNAサンプルあるいはシーケンス方法が異なる個別のアセンブリであり、9~13はこれら8つのアセンブリを異なる方法を用いて統合したアセンブリである。9~13のアセンブリ完全度は1~8より高くなるはずだが、この条件を満たすのはアセンブリ13(1~8のアセンブリをマージして類似性に基づきクラスタリングする)だけであった。アセンブリ方法の詳細については[4]を参照のこと。(c) 42種の脊椎動物ゲノムアセンブリを用いたCEGMA+CVG、BUSCO+CVGによる完全度。BUSCOでは'--long'オプションを用い、self-trainingモデルによってAUGUSTUSによる遺伝子予測を改善している。

■ 遺伝子探索パイプラインの選択：CEGMAとBUSCOのどちらにするか？

上記のURLからCEGMAのウェブサイトにアクセスすると、CEGMAはこれ以降開発・サポートしない旨が記されている。一方、CEGMAの代替となりうる、アセンブリ完全度評価のためのパイプラインBUSCOの論文が最近出版された^[6]。スイス・ジュネーブ大学のZdobnovらが開発したBUSCOは、アセンブリ中にリファレンス遺伝子のオーソログを同定し、復元された遺伝子の割合を計算する。つまり、BUSCOの基本設計はCEGMAに非常に類似している。Zdobnovらのチームは、オーソログデータベースOrthoDBを開発している特色を活かし、真核生物、後生動物、昆虫、脊椎動物などそれぞれの分類群に特化したBUSCO用のリファレンス遺伝子セットを提供している。BUSCOのリファレンス遺伝子は対象とする生物種にパラログを持たないオーソログに限られるが、10%の生物種には見つからなくてもよいという条件もあわせ持つ。そのため、脊椎動物のリファレンス遺伝子数はCVGの10倍近く(3023個)に達する。またBUSCOには、ゲノム配列だけではなくトランスクリプトーム配列あるいは予測アミノ酸配列に特化したパイプラインも備わっている。これらの多様な機能を持ち合わせたBUSCOは、CEGMAはおろかCVGすら代替できるように見える。

しかし、BUSCOとCEGMAのパフォーマンスを比較したところ、BUSCOのデータセットやアルゴリズムにはいくつか問題があることがわかった。BUSCOの脊椎動物リファレンス遺伝子を構成する生物種は、いわゆる「硬骨脊椎動物」に限られるため、系統的にこれらの外側に位置する軟骨魚類や円口類のアセンブリの評価には、BUSCOのリファレンス遺伝子ではなくCVGを用いるべきである。実際に、円口類であるカワヤツメのゲノムアセンブリ(LetJap1.0)の完全度を測定したところ、CVGを用いたCEGMAでは83%であったのに対し、上述した脊椎動物リファレンスを用いたBUSCOでは、オーソログが正しく判定されず23%の完全度しか示さなかった。我々は、BUSCOに適用したCVGデータセットも用意し、この問題を解決している。またCVGには、リファレンス遺伝子数が少ないことによる計算時間の短縮という効果もある。

リファレンス遺伝子の問題に加えて、BUSCOのゲノムアセンブリ評価用パイプラインにはCEGMAより精度が劣る箇所が見られた。概要のみ簡潔に記す。

- ・BUSCOが用いる遺伝子推定プログラム(Augustus)は、対象種の遺伝子構造によくフィッティングされたモデルが無い限り、遺伝子予測の感度を大きく低下させてしまう
- ・この問題の対策として、Augustusに自己学習によるモデルを適用し遺伝子予測を改善できるが、それでもBUSCOの精度はCEGMAを上回らないことが多い(図3c)。また、自己学習の計算に時間がかかるため、計算時間が短いというBUSCOの長所も失われる
- ・BUSCOの脊椎動物用リファレンス遺伝子セットは、鳥類に起きた大規模な遺伝子欠失を考慮できていないため、鳥類での完全度が過小評価となってしまう
- ・予測遺伝子とリファレンス遺伝子のオーソログ・パラログ判定について、BUSCOは遺伝子の進化速度を考慮せず一律の基準を持つため、進化が速い遺伝子のオーソログを同定しにくい

CEGMAの開発が終わってしまったのは残念ではあるが、特にゲノムアセンブリの評価においてCEGMAはいまだ優れたツールであることには変わりない。

■ おわりに

CEGMA、BUSCOに用いるCVGデータ、およびカスタムリファレンス遺伝子用に改良したCEGMAプログラムは我々の研究室ホームページ(<http://www2.clst.riken.jp/phylo/resource.html>)からダウンロードして使用できる。BUSCO用のCVGデータにもCEGMAと同等のオーソログ・パラログ判定条件を採用し、予測の精度を上げている。現状では、脊椎動物のゲノムアセンブリにはCEGMA+CVG、そして、脊椎動物のトランスクリプトームおよび予測アミノ酸配列にはBUSCO+CVGを用いるのが良いだろう。CVGを用いたCEGMAの実行に要する環境設定は、CEGを用いたCEGMAの実行と同等である。た

だし、CVGを構成する遺伝子はCEGを構成する遺伝子に比べてイントロンが長い傾向があることに注意する必要がある。カスタムリファレンス遺伝子用のCEGMAプログラムには、イントロン長、ならびに遺伝子予測の対象領域のマージンを自由に設定できるオプションを加えている。筆者は、CVGを用いたCEGMAの実行において、哺乳類ゲノムの最大イントロン長を100kb、哺乳類以外の脊椎動物ゲノムの最大イントロン長を50kbとし、遺伝子予測の対象領域の前後に10kbのマージンを取るよう設定している。コマンドの例を以下に示す（\$から始まる一連のコマンドは1行の入力を表す）。

```
$ ln -s /PATH/TO/CVG_20062015/cegma-extended/CVG .
```

```
$ cegma.mod.pl --protein CVG/CVGs.fa --hmm_profiles CVG/hmm_profiles --prot_num
8 --hmm_prefix chorNOG --cutoff_file CVG/profiles_CVG_cutoff.tbl --complete_file
CVG/completeness_CVG_cutoff.tbl --interlen 50000 --boundaries 10000 -T 8 -v
--genome genome.fa -o output --ext
```

同様に、トランスクリプトームおよび予測アミノ酸配列の完全度を評価するためのBUSCOコマンドを以下に例示する。BUSCOにはユーザが準備したリファレンス遺伝子を用いるオプションがあるため、これを用いてCVGを適用する。

```
$ ln -s /PATH/TO/CVG_20062015/busc0s/other CVG
```

(トランスクリプトーム配列)

```
$ python3 /PATH/TO/BUSCO_v1.1b1/BUSCO_v1.1b1.mod.py -l CVG -Z 6757 -m trans -c 8
-g transcript.fa -o BUSCO_transcript_cvg
```

(予測アミノ酸配列)

```
$ python3 /PATH/TO/BUSCO_v1.1b1/BUSCO_v1.1b1.mod.py -l CVG -Z 6757 -m OGS -c 8
-g gene.pep.fa -o BUSCO_pep_cvg
```

筆者らは、CVG + CEGMAに基づく完全度の評価を様々なゲノムアセンブリに行っており、現在世に出ている脊椎動物ゲノム配列の有用性の指標とすべく、データベース化する道を探っている。また、本稿で解説した個別の分類群に最適化した遺伝子セットは、脊椎動物以外の系統でももちろん作成でき、CEGより精度の高いアセンブリ評価に適用できると期待される。筆者らがCVGの作成に用いたEggNOGやEnsembl Comparaのように様々な分類群におけるオーソログデータベースが充実してきており、目的の分類群の1-to-1オーソログも得やすいだろう。

引用文献

- [1] Bradnam, K., et al. "Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species." *GigaScience*. 2013, 2(1):1-31.
- [2] Parra G., Bradnam K., Korf I. "CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes." *Bioinformatics*. 2007, 23(9):1061-1067.
- [3] Parra G., et al. "Assessing the gene space in draft genomes." *Nucleic acids research*. 2009, 37(1):289-297.
- [4] Hara Y., et al. "Optimizing and benchmarking de novo transcriptome sequencing: from library preparation to assembly evaluation." *BMC genomics*. 2015, 16(1):977.
- [5] Zhang G. et al. "Comparative genomics reveals insights into avian genome evolution and adaptation." *Science*. 2014, 346:1311-1320.
- [6] Simão F., et al. "BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs." *Bioinformatics*. 2015, 31(19):3210-3212.

(編集担当：佐藤行人)

医学部に埋め込まれたハードコア分子進化の拠点： コロラド大学デンバー校・David D. Pollock研究室

福島健児 (コロラド大学・日本学術振興会海外特別研究員)

1955年、当時の合衆国大統領アイゼンハワーは旅先で心臓発作に見舞われ、コロラド州オーロラ市に位置するフィッツシモンズ陸軍病院へと担ぎ込まれた。彼は幸運にも一命をとりとめ、回復までの7週間のあいだは病室から執務をこなしたと伝えられている。それから40年の後、陸軍病院はその役目を終え、重厚な建物は大学施設へと転用された(図1)。現在のオーロラ市には、この建物を中心として研究棟や病棟が林立し、コロラド大学デンバー校アンシュッツ医療キャンパス (Anschutz Medical Campus) を形成している。デンバー校のメインキャンパスから車で30分ほどの郊外にあり、豊富な芝地には野ウサギが跳ねる(図2)。キャンパス名は、設立資金として巨額を寄付した実業家Anschutz夫妻への敬意を表している。私がデスクを置くDavid D. Pollock研究室はこの医療キャンパスにあり、School of Medicineすなわち医学部に属している。私の学位研究テーマは食虫植物の進化であった。そのような人間がどうして医学部で研究を続けているのか、その経緯から紹介したい。

収斂進化の魅力

私は総合研究大学院大学の5年一貫博士課程に在籍するあいだ、自然科学研究機構・基礎生物学研究所にある長谷部光泰教授の研究室で食虫植物の研究に取り組んだ。食虫植物は被子植物中で複数回独立に出現しており (Albert et al., 1992)、収斂進化を研究する好材料となる。研究を進めるうちに、「収斂進化を“反復実験”のように捉え、統計的に解析することでまだ見ぬ適応進化のメカニズムを解明できるのではないか」と考えるようになった。進化生物学においては、人工進化系など一部の例を除いて実験的反復をとれないため、それが研究を進める上での大きな制限となっている (Bak and Paczuski, 1995)。収斂進化には、そのような現状を打破する鍵が秘められていると考えたのである。

近年、表現型レベルでの収斂進化が遺伝子配列レベルでの収斂進化、すなわち分子収斂を基盤に持つとする報告が相次いでおり (Stern, 2013)、私も次第に興味を寄せていった。しかし、様々な理由から分子収斂の検出は困難を極め、標準となるような検出法はいまだ現れていない。たとえば、Parker et al. (2013) は遺伝子系統樹の樹形をもとにした分子収斂検出法を使い、エコーロケーションの能力を独立に獲得したコウモリとイルカの間にも多数の収斂遺伝子が存在すると報告して注目を集めたが、のちに複数



図1 元陸軍病院。現在はBuilding 500もしくは“pyramid”の愛称で親しまれ、アイゼンハワーの病室は今もなお観光用に保存されている。



図2 アンシュッツ医療キャンパスの芝地。7羽の野ウサギが写っている。

の技術的批判に曝されている (Zou and Zhang, 2015b; Thomas and Hahn, 2015; Zou and Zhang, 2015a; Goldstein et al., 2015)。そんな状況の中、私の目にはDavid D. Pollock研究室で確立された手法 (Castoe et al., 2009) が最も妥当なように映った。この方法を出発点にするのが次の研究目標を達成するのに一番の近道であろうと判断し、当時、全く面識がなかったPollock博士に海外学振の受入先になってくれないかとしたためのEメールを送信したのである。しかしながら、この話にはよくあるオチがついてくる。返事がこない。

スパムメール

メールが返ってこない。そんなとき、どのような原因が想定されるだろうか。先方が忙しいのかもしれないし、運悪くスパムメールとして振り分けられてしまったのかもしれない。後者の可能性を想定するならば、よりスパムらしくない内容・送信方法を心掛けるべきであろう。

当時、返事がないことを研究室で嘆いていたところ、長谷部研究室助教の玉田洋介博士から助言を受けた。曰く、「アメリカの教授は忙しすぎるから、メールなんか全然見ていない。5回、いや、10回くらいなら再送しても全く失礼にあたらぬ。」と。それではスパムメールそのものではないか。しかし、アメリカで5年を生き抜いた猛者が言うのであるから真実味を帯びている。私は、そのスパムメール戦術を採用することにした。ダメ元で何度も送ってみよう。そう意気込んで2度目のメールを送信すると、あっけなくも次の日には歓迎の返事が来てしまった。

Pollock研究室に参加してすぐ、「最初のメールに返事をくれなかったじゃないか」と訊くと、Davidは「私宛に来るメールのうち90%は私の興味を引かないからタイトルをちらっと見て瞬時に捨てているんだ。返事が必要なときはスパムメールのように何通も送ってくれ。」と言っていた。幸運にも私は2度目で返事をもらえたが、玉田博士の助言は全面的に正しかったようである。

ホワイトボード

最初の返信から数往復のやりとりののち、私から希望してコロラドを訪れた。日本学術振興会海外特別研究員へ応募するにあたって、その申請内容について議論するためである。当時Pollockラボでポスドク職にあったRobert P. Ruggiero博士の家に泊めてもらい、研究室メンバー全員の前でのプレゼンテーション、メンバー各個人とのディスカッション、大学院生とのランチ、PI+ポスドクとのディナーと忙しい一日を過ごしたが、これはポスドク面接の標準的な流れのようだ。最後に、「もしフェローシップが取

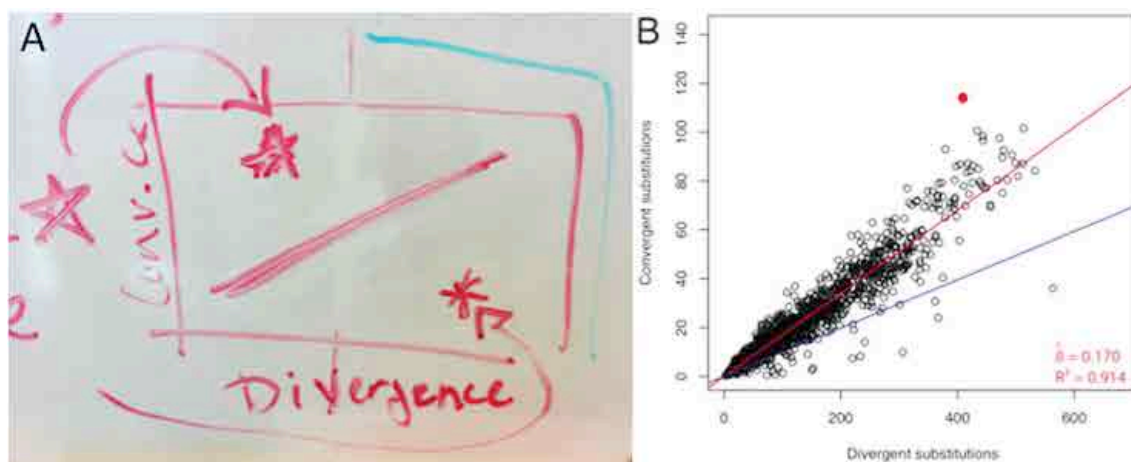


図3 分子収斂の検出法を示した図。Aがホワイトボードに残ったメモ、BがCastoe et al. (2009) のFig. 3Bである。収斂的なアミノ酸置換数(縦軸)が多様化置換数(横軸)によって経験的に予測できることを示している(赤線)。収斂置換の多寡は、赤色の回帰直線からのずれを指標に推定する。また、既存のアミノ酸置換モデルをもとに収斂置換数を予測した場合、その数を過小評価してしまうことも示している(青線)。図はDavid D. Pollock博士の許可を得て引用した。

れなかったら雇ってください」とDavidにお願いしてコロラドを後にした。このお願いが実効性を有していたかどうかは結局不明であるが、幸運にも海外学振に採用され、2015年4月からPollock研究室に参加することができた。

Pollockラボ参加の翌々月、ポスドクのRobertは次の職場であるニューヨーク大学アブダビ校へと旅立ち、私は仮のデスクをたたんで彼が使っていたオフィスへと取まった。彼は部屋を掃除していってくれたが、壁に掛かったホワイトボードだけは消さずにおいてくれとお願いし、今も当時のまま残してある。そこには、Robertの前の住人であるTodd A. Castoe博士のメモが残っていたのである。私がPollock研究室を選ぶ決め手となった論文(Castoe et al., 2009)における最も重要な図の原案だ(図3)。それが残る部屋を引き継げたのはとても感慨深い。

準備は余裕をもって

異動の経緯について、私生活寄りの視点からも紹介しておきたい。まず、私の妻は神経科学者だ。現在は同じキャンパス内のAbigail Person研究室で、学振海外特別研究員として研究に勤しんでいる。渡米の前年度まで、私がいた基礎生物学研究所と隣接する生理学研究所に在籍し、私と同学年の大学院生であった。学位取得後の進路はそれぞれ思惑があったわけであるが、妻からの譲歩によって私の第一希望であったコロラド大学へ二人揃って異動することにした。しかしながら、学位審査、引越し、ビザ取得など、思いつく限りほぼ全ての必要手続きを私が後手に回してしまったため、妻には多大な心配と苦勞をかけてしまった。

殊に、海外学振の受入先決定が遅すぎたのはまずかった。海外学振募集の機関内締切りは概ね毎年4月であるが、私がDavidへ打診メールを送ったのはその年の2月であった。ギリギリまで複数の研究室を天秤にかけていたためである。私の申し出が断られてしまったら妻がコロラド大学へ行く理由がなくなるので、Davidからの返事が来るまで彼女の職探しに必要なアクションが取れなかった。スパムメール戦術だのなんだのとやっているあいだ、私以上に妻のキャリアを危険に晒してしまったわけである。さらに、学位審査を通過できなければ私の海外学振内定は取り消され、妻の単身渡米が決定してしまう。このような体たらくでは「結婚詐欺!」との誹りを受けたとしても甘んじるほかない。というか実際に受けた。現在の私たちの家庭が成り立っているのは、当時の妻の寛容さの賜物である。妻は現在の所属研究室と研究テーマを非常に気に入っているのが結果オーライではあるのだが、だからといって当時のことは免罪されない。準備は余裕をもって。

医学部に埋め込まれたハードコア分子進化研究室

この記事を書いている最中、「日本の進化学コミュニティへ何か研究室紹介のメッセージはないか?」とDavidに聞いてみたところ、以下のように一筆したためてくれた。

『This is an exciting time to be working in molecular evolution in the Pollock laboratory. The massive bursts of adaptation, coevolution and convergence in ancestral snake lineages (Castoe et al., 2009, 2008) point to a different mode of evolution under adaptive pressure compared to nearly neutral evolution. However, we have also come to understand that functional molecules evolve with fluctuating constraint and decreasing convergence over time (Goldstein et al., 2015) due to the “evolutionary Stokes shift” (Pollock et al., 2012; Pollock and Goldstein, 2014). Current work involves using new methods to evaluate temporal and spatial fluctuations in evolutionary processes (de Koning et al., 2012, 2010), and developing the statistical mechanics that gives rise to evolutionary processes (manuscripts in preparation).』

上に述べられている通り、Pollockラボは分子進化学を専門にする研究室である。タンパク質進化モデル、反復配列、エピスタシス、分子収斂、集団遺伝、ゲノム解読などが具体的なキーワードになる。ここ数年に論文化された主な研究成果として、ヒトゲノムにおける反復配列の高感度な検出(de Koning et

al., 2011)、系統仮説における塩基・アミノ酸置換過程の“partial sampling”による尤度計算の高速化 (de Koning et al., 2012, 2010)、ビルマニシキヘビ (Burmese python) のゲノム解読 (Castoe et al., 2013)、分子収斂の分岐時間依存性の発見 (Goldstein et al., 2015) などが挙げられる。研究目的ごとに独自の解析プログラムを開発・維持しているが、外部への周知にあまり積極的ではないのは惜しい限りだ。たとえば、反復配列検出ソフトウェアのデファクトスタンダードである RepeatMasker は既知の反復配列と類似性を示さないものを見逃す傾向にあるが、Pollock ラボで開発されたプログラム P-clouds は、ゲノムの k-mer 頻度から反復配列を予測し、既存の反復配列データベースに依存しないため、その弱点を補完し得る性能を有する。

私が参加するのに前後して人員の転出が重なったため、現在、Pollock ラボはポスドク 1 名 (私) に大学院生が 3 名と、小さな規模で運営している。一時期は第二世代シーケンサーの波に乗って配列データ生産へと研究活動の舵を切っていたが、ここ数年の興味は分子進化の理論構築へとシフトしてきている。タンパク質を熱力学的な物体と捉え、そこから“メカニスティックに裏打ちされた”分子進化モデルを構築しようという試みはその例だ (Pollock et al., 2012)。Pollock ラボが占有する実験スペースは今や埃をかぶり、ラボメンバーは UNIX サーバーへのコマンド送信により多くの時間を費やす。ただし、大学院生の一人がサンプル集めから始める集団遺伝学的研究を計画しているので、その埃が払われる日はそう遠くなくさそうだ。

少し大雑把な David の性格を反映して、研究室の管理運営はゆるやかである。David のスケジュールに余裕があるときはラボメンバーの誰かをカフェへ連れ出し、コーヒーを飲みながら一対一で研究アイデアをインプットしていく。ラボに参加したての頃、3 時間以上をカフェでのディスカッション (と雑談) に費やしたときは少々面食らった。そうやって外へ出て“conversational”に研究を進めるのが David の理想のようであるが、書き仕事に取り組むときは小さなオフィスに窮屈そうに収まる。よく机に足を投げ出してラップトップのキーボードを叩いているが、そういうときはきっと面白くない仕事なのだろう。申請書提出や論文投稿など、節目となる仕事を終えた日は決まってみんなでビールを飲みに行く。デンバー近郊は地ビールメーカーが多数存在することで知られ、多彩なクラフト・ビールを楽しめる。キャンパスから徒歩圏内にあるパブは、我々のようにひとときの解放を求める客で平日から賑わいを見せる (図 4)。

週一回のラボミーティングも然りで、ルールは緩い。日時は決まっているものの、たいていは前日、あるいは直前まで何をやるか周知されない。一度など、議論が弾むようにとの目論見で上述のパブで開催した (ただし、「みんな研究の話をしなさい」との理由で二度と同じ形式が採用されることはなかった)。そのような例外を除けば、ラボ内で最近出たデータをネタに議論したり、David が構想している研究アイデアを紹介することが多い。後者の場合、ideal gas (理想的な気体) だとか Boltzmann equation (ボルツマン方程式) といった、生物学者にはあまり馴染みのない語句と数学的な記法を抵抗なく咀嚼できなければ

議論についていくのは難しい。私はまだうまく咀嚼できない側の人間だ。大学院生の背景知識も様々で、David の解説に対して物理系の学部教育を受けた学生がうんうんと頷く横で、生物系出身の学生が私と同じ顔で硬直していることもしばしばである。

他方、医学部という環境にあって、応用研究とはうまく距離感を保っているように見える。David は“ハードコア分子進化学者”と言ってよい根っからの基礎志向研究者であり、その興味と重ならない共同研究でがんじがらめになっている様子はない。それでも、「インフルエンザウイルスのここ数年分の配列データがあるんだが・・・」などと相談に訪れる研究者はち



図 4 大学近くのパブ。奥に身の丈よりも大きなビールタンクが見える。

らほらいて、多くの場合、Davidから助言を受けて懸案に活路を見出すようだ。

“日本版NIH”医療研究開発機構の発足に代表されるように、応用研究重視の波は既に日本にも押し寄せている。重視に留まらず応用偏重となるようならば流れを引き戻す努力も必要であろうが、潮目を見極めてこそ開く活路もあるはずだ。Davidのように医学系研究者の中にあって進化学の有用性を示すこともまた、強い選択圧に晒されつつある進化学研究者が今後採りうる適応戦略のひとつであるだろう。

Mechanisms of Protein Evolution

前述の通り、医療系の学科が集まる性質上、アンシュッツ医療キャンパスには進化学周りを専門にする研究室はそう多くない。しかしながら、およそ2年に一度、ここに多くの分子進化系研究者が集う。Society of Molecular Biology and Evolution (SMBE)のサテライトミーティング“Mechanisms of Protein Evolution”が開催されるのである (URL: <http://www.proteinevolution.org/>)。Davidを筆頭に数名がコアとなって企画している研究会だ。直近では2015年11月に開催され、Davidに比肩する“ハードコアな”分子進化学者が最新の研究成果を持ち寄って賑わいを見せた (図5)。参加者数50名程度で、全員の顔が見えるちょうどよい規模の会である。私にとって特筆すべき点の一つが、分子収斂の研究者が多数参加する点だ。思いがけず、周辺分野の動向を知る良い機会となった。

この会の独特な基調講演スタイルについても紹介しておきたい。2015年開催の研究会では1時間程度の基調講演が11題あり、それぞれに分子進化学の発展に大きく貢献した研究者の名が冠せられている。Ohta Keynote、Dobzhansky Keynote、McClintock Keynoteといった具合だ。講演者は、それらの研究者が成した偉業に軽く触れ、自身の研究内容との関連を示したうえで研究発表を行う。例えば、Jacob Keynoteの講演者であったAnne-Ruxandra Carvunis博士は、François Jacobによる記事『Evolution and Tinkering』からの引用で発表を始めた。Jacob曰く、『The probability that a functional protein would appear de novo by random association of amino acids is practically zero.』であると (Jacob, 1977)。Jacobの業績についてももう少し触れたのちに、彼女自身の研究から明らかになった“*de novo* gene birth” (Carvunis et al., 2012)の普遍性と新たな知見について詳しく述べられた。偉業を成した一人の研究者を軸に分野の成立と変遷がコンパクトに紹介されるため、前提知識のないトピックでも非常に聞きやすかったのを覚えている。

The Mile-High City

コロラド州の州都デンバー市 (Denver) とそれに隣接するオーロラ市 (Aurora) は、アメリカ合衆国西部でも有数の都市圏を形成している。デンバー市は標高およそ1マイル (1609メートル) に位置することから、The Mile-High Cityの愛称で親しまれる。気圧が低いためにカップ麺などのパッケージは膨張してしまい、スーパーマーケットで積み上げて陳列するのが難しそうだ。オーロラ市が日本のメディアで報じられる機会はそう多くないが、もし聞き覚えがあったら、2012年の映画館銃乱射事件を想起することだろう。美しい名の街であるだけに残念だ。そのような凄惨な事件からくるイメージとは裏腹に、治安はさほど悪くない。もちろん地区によるのだが、閑静な住宅街では男女問わず日没後のジョギングを楽しむ姿も見られる。

デンバー近郊は風光明媚な土地柄と言ってよいだろう。西のロッキー山脈はほぼ年中雪をかぶり (図



図5 SMBE Satellite Meeting “Mechanisms of Protein Evolution III 2015: Origins”の冒頭でウェルカムスピーチを行うDavid D. Pollock博士。

6A)、東にはグレートプレーンズが果てなく広がる。デンバー市内北部にはロッキーマウンテン兵器工場跡地に自然保護区が広がっている。大戦中はありとあらゆる環境汚染の限りを尽くして工場を稼働させていたようであるが、保護地に指定されて数十年経つ今では自然豊かな場所だ (図6B)。ビジターセンターの説明書によると、ここで作られたナパーム弾は東京大空襲に使われたらしい。サリンのような毒ガスも作っていたらしく、そういった危険度の高い兵器はほとんど使われずに廃棄されたようである。廃液は地中深くへの高圧注入により捨てられ、それが人工地震として有名なデンバー地震の原因になったと伝えられている。デンバーから車で1時間ほどかけて南下すれば、Garden of the Godsと呼ばれる自然公園が広がる (図6C)。巨大な岩が多数山肌に露出した美しい場所だ。ロッキー山脈の景色を楽しみながら国道70号線を西に走り、首なし鶏マイクの逸話が残るフルータ (Fruita) を越えれば、デンバーから5時間ほどでグランドジャンクション (Grand Junction) へと到着する。ここにあるMuseums of Western Coloradoでは化石掘りツアーが開催されており、素人でも骨の欠片くらいは掘り当てられる (図6D)。

デンバー近郊はスノースポーツが盛んで、山岳リゾートでの余暇を目当てに訪れる観光客も少なくない。残念ながら私はまだこの地でスノースポーツに興じる機会を持っていないが、スキーストックは頻繁に手にする。Pollock研究室がよく使う会議室に、指示棒やレーザーポインターの代用として置いてあるのだ。外部からセミナー発表に訪れた研究者は、ほぼ例外なくこれに戸惑う。ただし、スキーストックを指示棒代わりにしているのは研究棟全体で見ると稀で、所属する学科が所有する2つの会議室以外ではまだ確認できていない。

■ 出産・子育て

渡米から5ヶ月後、私達夫婦の生活に大きな変化が訪れた。第一子となる娘が誕生したのである。妻は産休ギリギリまで実験を詰め込んでいたが、予定していたよりも早く陣痛が始まってしまい、コロラド大学病院へと駆け込んだ。診察台の上に乗せられ、まだ実験が終わっていないだのなんだのと悪あがきをす



図5 (A) オフィスからの眺望。6月末の撮影であるが、奥に見えるロッキー山脈には雪が残る。(B) ロッキーマウンテン兵器工場跡地で草を食むバイソン。(C) Garden of the Godsで巨大な岩を支える筆者。(D) グランドジャンクションで筆者が掘り当てた骨の化石(黒色の部分)。博物館職員は一瞥して恐竜のものだと言っていたが、真偽は不明。

る妻は、看護師さんに諫められてようやく観念した。出産に立ち会って印象的だったのは、医療スタッフの多さである。陣痛が強くなってくると、医師・看護師が次々と現れ自己紹介をしてくれる。その数は両手の指では足りず、分娩台の前にできた人だかりに、私たちは完全に圧倒されてしまった。ある医師など、私達の緊張を見かねて「産まれてくる娘さんは何歳になるまでデート禁止なんだい？」とジョークを投げたが、それに対応する余裕は当時の私たちには残っておらず、乾いた愛想笑いしか返せなかった（結局、その医師はバツの悪そうな顔で部屋を出たきり二度と戻ってこなかった）。その後数時間の格闘の末、娘は無事生まれてきた。コロラド大学病院は、私が勤務する研究棟のすぐ横にある。今でもときおり、オフィスの窓から病棟を眺め、出産の日に泊まった病室はどの窓だったかと探してしまう（図7）。

研究室は子育てに協力的だ（図8）。妻のお腹が目立つようになってきた頃、妻のボスAbbyは彼女の家でベビーシャワーを開いてくれた。これはアメリカ発祥の出産前祝いパーティである。赤ん坊が産まれると何かと忙しくなるので、その前にお祝いを済ませようというのは実に合理的だ。Abby自身も幼い2児の母親であり、子育ての気苦労をよく慮ってくれる。クリスマスにはDavidの自宅でパーティが開かれたが、参加者の半数以上が子連れで大変賑やかなものであった。

今日まで、娘の存在が生活をより一層鮮やかにしたことは言うまでもない。今では娘を保育園に預け、妻も仕事に復帰している。また、妻子はキャンパス内の被験者募集に応募し、母親の体調が子供に与える影響を調べる研究に参加している。ときおり身体測定に出向いたり、娘のおむつを提供したりといった具合だ。このような募集が多数あるのは医療系キャンパスならではだろう。

留学で研究活動の谷を越える？

前任のポストドクRobertは、次の職場があるアブダビへと旅立つ直前、「新天地で働くのはいいもんだろ？」と私に聞いてきた。正直、留学に伴う困難など売るほどある。私を含め大多数の日本人にとって、その最たるものは英語での意思疎通であるだろう。さらに、フェローシップを取ってビザを申請して…と必要手続きをこなすだけで、論文を一報書けてしまうくらいの時間と労力を要するかもしれない。しかしながら、私のような駆け出しの研究者が研究環境を大きく変えることのポジティブな効果も強調しておきたい。

少なくとも私にとって、研究活動は視野狭窄下での山登りのようなものだ。懸命に登ってひとつのピークに到達したからといって、そこが大域的に見たときに最も高い峰であるとは限らない。手元で大切に温めているアイデアが最適解であるという保証はないのだ。しかし、研究環境が変われば新たなインプットがあり、より良いアイデアが浮かんでくることもある。生物が適応度地形の谷を越えるとき、環境変動が引き金になると似た話だ（Steinberg and Ostermeier, 2016）。海外学振の研究計画を仕上げた当時はそ



図7 アンシュッツ医療キャンパスの一部。左手奥にPollock研究室が入る研究棟、右手奥にコロラド大学病院が見える。



図8 Pollock研究室メンバーと対面する娘。生後一ヶ月ほど。

れに大変満足していたが、こちらのラボメンバーと議論を交えながら研究を進めてみると、少し陳腐な部分が見えてきた。このラボでなければ出てこないだろうと思わせるアイデアで軌道修正することも少なくない。留学に伴う諸々を天秤にかけたとき、少なくとも私や妻の場合、僅かばかりポジティブな側面へ針が振れているように思う。サンプル数2（しかも独立性が疑われる）で判断を下すのは心許ないが、当時のRobertの問いには今現在も肯定で返したい。新天地で働くのはいいものだ。

謝 辞

著作権を保持する図の掲載を許可し、一部の原稿作成に協力してくれたDavid D. Pollock博士にお礼申し上げます。また、数点の写真掲載を快く許可してくれたPollock研究室メンバーにも感謝したい。最後に、「結婚詐欺！」のくだりを載せることに好意的な返事をくれた田淵紗和子博士に感謝の意を述べたい。

参考文献

- ・ Albert, V.A., Williams, S.E., and Chase, M.W. (1992). Carnivorous plants: phylogeny and structural evolution. *Science* 257: 1491–1495.
- ・ Bak, P. and Paczuski, M. (1995). Complexity, contingency, and criticality. *Proc Natl Acad Sci U S A* 92: 6689–6696.
- ・ Carvunis, A.-R. et al. (2012). Proto-genes and de novo gene birth. *Nature* 487: 370–4.
- ・ Castoe, T.A. et al. (2013). The Burmese python genome reveals the molecular basis for extreme adaptation in snakes. *Proc. Natl. Acad. Sci. U. S. A.* 110: 20645–50.
- ・ Castoe, T.A., Jiang, Z.J., Gu, W., Wang, Z.O., and Pollock, D.D. (2008). Adaptive evolution and functional redesign of core metabolic proteins in snakes. *PLoS One* 3: e2201.
- ・ Castoe, T.A., de Koning, A.P., Kim, H.M., Gu, W., Noonan, B.P., Naylor, G., Jiang, Z.J., Parkinson, C.L., and Pollock, D.D. (2009). Evidence for an ancient adaptive episode of convergent molecular evolution. *Proc Natl Acad Sci USA* 106: 8986–8991.
- ・ Goldstein, R.A., Pollard, S.T., Shah, S.D., and Pollock, D.D. (2015). Nonadaptive amino acid convergence rates decrease over time. *Mol. Biol. Evol.* 32: 1373–1381.
- ・ Jacob, F. (1977). Evolution and tinkering. *Science* 196: 1961–1966.
- ・ de Koning, A.P.J., Gu, W., Castoe, T.A., Batzer, M.A., and Pollock, D.D. (2011). Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet.* 7: e1002384.
- ・ de Koning, A.P.J., Gu, W., Castoe, T.A., and Pollock, D.D. (2012). Phylogenetics, likelihood, evolution and complexity. *Bioinformatics* 28: 2989–90.
- ・ de Koning, A.P.J., Gu, W., and Pollock, D.D. (2010). Rapid likelihood analysis on large phylogenies using partial sampling of substitution histories. *Mol. Biol. Evol.* 27: 249–65.
- ・ Parker, J., Tsagkogeorga, G., Cotton, J.A., Liu, Y., Provero, P., Stupka, E., and Rossiter, S.J. (2013). Genome-wide signatures of convergent evolution in echolocating mammals. *Nature* 502: 228–231.
- ・ Pollock, D.D. and Goldstein, R.A. (2014). Strong evidence for protein epistasis, weak evidence against it. *Proc. Natl. Acad. Sci. U. S. A.* 111: E1450.
- ・ Pollock, D.D., Thiltgen, G., and Goldstein, R.A. (2012). Amino acid coevolution induces an evolutionary Stokes shift. *Proc. Natl. Acad. Sci. U. S. A.* 109: E1352–9.
- ・ Steinberg, B. and Ostermeier, M. (2016). Environmental changes bridge evolutionary valleys. *Sci. Adv.* 2: e1500921–e1500921.
- ・ Stern, D.L. (2013). The genetic causes of convergent evolution. *Nat Rev Genet* 14: 751–764.
- ・ Thomas, G.W.C. and Hahn, M.W. (2015). Determining the null model for detecting adaptive convergence from genomic data: a case study using echolocating mammals. *Mol. Biol. Evol.* 32: 1232–6.
- ・ Zou, Z. and Zhang, J. (2015a). Are convergent and parallel amino acid substitutions in protein evolution more prevalent than neutral expectations? *Mol. Biol. Evol.* 32: 2085–2096.
- ・ Zou, Z. and Zhang, J. (2015b). No genome-wide protein sequence convergence for echolocation. *Mol. Biol. Evol.* 32: 1237–41.

(編集担当：佐藤行人)

第1回

ダーウィン研究室：国内にもある、Cutting-Edge Science!
東京工業大学・田中幹子研究室紹介

岡本恵里 (東京工業大学大学院 生命理工学研究科 生体システム専攻 博士後期課程1年)

2016年最初の進化学会ニュースとなる本号より、海外ばかりでなく国内のユニークな研究室も紹介すべく、本企画を立ち上げました。記念すべき第1回は、東京工業大学大学院・生命理工学研究科・発生生物学分野・田中幹子研究室博士後期課程1年の岡本恵里さんによる「マイラボ」紹介です。

東工大 田中幹子研究室は、進化発生学を専門として、脊椎動物の四肢の対象に研究している研究室です。私は田中研所属 博士課程1年の岡本恵里です。進化学会ニュースの国内版研究室だよりの第一回目を書かせていただくことになりました。いろいろな方々に、田中研の魅力やちょっと変わったところ、研究の面白さなどをお伝えすることができれば幸いです。

まず、田中研でどのような研究をおこなっているのかを簡単にご説明させてください。

田中研は、進化発生生物学 (Evolutionary Developmental Biology 略してエボデボ) を専門としてい

ます。エボデボでは、「生物の多様な形態がどのようにして進化したのか」という問について、発生プログラムの変化から理解しようとしています。具体的には、異なる形態をもつ生物間で、それぞれの形態の発生を制御している分子システム (遺伝子の発現タイミングや発現領域の違いなど) を比較し、その違いを明らかにすることで形態の進化を理解しようとしています。

生物の持つ多様な形態のなかで、田中研は脊椎動物の四肢を研究対象にしています。四肢は、対鰭とよばれる左右で対になった鰭 (胸鰭・腹鰭) から進化しましたが、原始的な脊索動物であるナメクジウオなどは対鰭をもちません。田中研では、対鰭がどのようにして獲得されたのか、また、対鰭から四肢への形態変化がどのように生じたのかということを中心にしています。さらに、四肢や対鰭にみられる形態的な多様性がどのようにして生じたのかというテーマや、四肢や対鰭の適切な発生が、どのような発生プログラムによってコントロールされているのかなどのテーマについて、様々な実験生物を用いて日々研究に取り組んでいます。

田中研のボスである田中幹子准教授は、ハイヒールで颯爽と歩く姿が印象的な先生です。他の研究室的先生や学生が、「ハイヒールの音がすると田中先生が近づいてきたことが分かる」と言うほど、田中先生のトレードマークとなっています。田中先生はかなり迫力のあるマシンガントークで、私が初めて田中研のセミナーに参加した際には、矢継ぎ早に飛んでくる先生からの質問に、答えられないでいると強制終了、という場面を見て、田中研のセミナーは戦場のように感じられました。今では戦場にも慣れ、それなりに応戦できるようになったとは思いますが、新入生が戦々恐々としているのを見ては、自分の入学当初を思い出しています。また、発表の仕方や論文の書き方に関しては特に厳しく、原稿を何度もダメ出しされることも多いですが、かなり細かい部分までの確かなアドバイスとともに指導してもらえるので、プレゼン能



図1 田中研のメンバー。学生は、学部時代からの東工大生、高専からの編入生、他大から入学した大学院生、留学生とメンバーの背景は様々。中列右から二番目が田中幹子准教授。前列左が著者。

力はかなり鍛えられると思います。厳しそうな面ばかり書いてしまいましたが、その分、いつも親身にアドバイスしてくれる先生です。また、田中先生はいつもかなりハイテンションで明るい先生で（怖い先生と誤解を生みそうなので）、生物の進化発生の面白さを語る際の熱量にはいつも圧倒されます。「○○○の論文見た？すごいよねー！」「○○○さんの実験ですごくいい結果が出たよ！」と話しかけられることも多いです。また、ケーキやゆるキャラが好きといったチャーミングな一面もあります。

そんな先生の人柄からか、田中研のメンバーは、とにかく面倒見が良いです。私が入学した当初は、田中研には年齢の離れた先輩が多く所属していました。質問するのも躊躇われるくらい、すごい集中力で実験に取り組む姿に、下手なことは聞けないぞと少し怖く感じるくらいでした。しかし、先輩たちは非常によく面倒を見てくださり、配属後の数か月間は、実験の基本的な操作、セミナー発表のアドバイスなどを様々なことを教えてくれました。実験や発表準備で行き詰まった際には、問題が解決するまで付き合ってもらったことも多く、配属後の数か月で大きく成長することができました。また、静かながら熱量を持って真面目に研究に向き合う姿に刺激を受け、自分もこんな風になりたいと思ったものでした。

田中研では先生と学生が皆同じ居室ですごしています。いつも机に向かって真剣に作業をしている先輩や先生に囲まれて、今思えば最初の数か月はかなり緊張しながら過ごしていたと思います。しかし今では、先生に質問しやすいという点では非常に良い環境であると思っています。田中先生が席を立つと、いろいろな学生から一気に質問の声がかかることが多々あります。また逆に、実験が行き詰っていきそうな新入生に、先生や先輩が声をかけるということも多いです。最近、研究室見学に来た学部生に「居室が先生と一緒に、居づらくないですか？」とこっそり聞かれたことがありますが、この環境は居づらさ以上に、研究生活にとってはメリットであることのほうが多いと思います。

エボデボでは形態の異なる生物を比較するという方法が主なことから、扱う実験生物の種類も多いですが、現在田中研で研究に使用している生物種は、研究機関等からいただいている胚や卵等を含めると、スタンダードなモデル生物であるゼブラフィッシュ、メダカ、ニワトリ、アフリカツメガエルをはじめ、ナイルティラピア、サラマンダー、ゼブラフィンチ、エミュー、アカハライモリ、フグ、ヤツメウナギ、トラザメ、ゾウギンザメなど実に多様です。そんな中で、私は軟骨魚類ハナカケトラザメを主な対象として、対鰭筋形成システムの進化について研究しています。

四肢の祖先形質である対鰭は無顎類で獲得されたと考えられていますが、現在も生き残っている無顎



図2 先生がケーキ好きな田中研では、メンバーの誕生日だけでなく、イベントやお祝い事（就職が決まったり、試験に合格したりしたときなど）があると頻繁にケーキを食べます。なぜか12月にメンバーの誕生日が集中しているため、クリスマスも含め、12月は毎週ケーキを食べることになります。



図3 田中研で扱っている動物たち。エミューの卵（左）。右のサインペンと比較するとニワトリの卵よりもかなり大きいことが分かります。どこか抜けた顔のサラマンダー（中央）とナイルティラピア（右）。実験室の水槽で飼育しており、実験やデスクワークで疲れた学生が癒しを求めにやってきます。

類は、対鰭の獲得以前に分岐したと考えられている、ヤツメウナギとヌタウナギのみです。そのため、現在も生き残っている脊椎動物の中で、対鰭をもつ最も古い系統である軟骨魚類は、対鰭・四肢の進化を見るうえで非常に重要な位置にあり、対鰭の獲得進化の生き証人ともいえます。

私は、この軟骨魚類のハナカケトラザメを主な対象とし、対鰭筋の形成システムの進化について研究しています。四肢動物の四肢の筋肉は、遊離筋とよばれる移動能を持つ筋芽細胞によって形成されます(図5)。一方で、軟骨魚類は四肢動物とは異なり、遊離筋によらない原始的なシステムによって対鰭筋が形成されると報告されています。私は、軟骨魚類の対鰭筋形成システムについて明らかにすることで、どのようなシステムで対鰭に筋肉がもたらされたのかを理解しようとしています。

研究を進めるにつれて、対鰭筋形成システムの進化の新たな可能性がだんだん見えてきました。エボデボの研究では、生物がある形態を手に入れるまでに辿ってきた長い歴史やストーリーを明らかにすることができるという点で、非常に魅力的でやりがいがあり、毎日充実した研究生生活を過ごしています。

最後に、拙い原稿にアドバイスをくださった荒木仁志先生にお礼申し上げます。

研究室の先生からの一言：私は熱量(?) だけなのですが、優秀な学生さんとスタッフに支えられて、なんとかやっています。研究は楽しく真剣に、がモットーです。

(編集担当：荒木仁志)



図4 ハナカケトラザメの卵(殻を一部切除したところ)。巾着型の卵殻が特徴的。卵黄は鮮やかな青緑色をしており、左上部にヨークサックで卵黄とつながった胚が見えます。

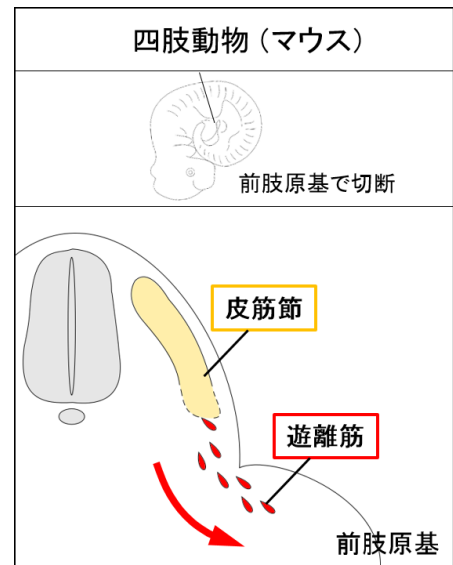


図5 四肢動物の四肢筋形成：マウス胚の前肢原基レベルでの断面の模式図。四肢動物の四肢筋は、移動能をもつ筋芽細胞である遊離筋(赤色部)が皮筋節(黄色部)から分離し、四肢の原基へと遊走することで形成される。

ミーティングレポート①

第24回 Plant and Animal Genome Conference(PAGXXIV)に参加して



川原善浩 (農業生物資源研究所)

国際動植物ゲノム・カンファレンス Plant and Animal Genome Conference (PAG) は、米国・サンディエゴのタウン&カントリー リゾート&コンベンションセンターにて、毎年1月中旬に開催されています(写真)。サンディエゴは米国西海岸のカリフォルニア州の南端のメキシコとの国境にあり、一年を通して穏やかな気候であり、観光地としても人気が高い都市です。この時期のサンディエゴは日本と同じ冬ですが、天気の良い日は20度近くまで気温が上がり、寒い冬の日本からの参加者にとっては春(初夏?)のよ

うな暖かさでホッとさせてくれます。

この学会、「動植物ゲノム」というタイトルではありますが、モデル生物を対象とした基礎研究は含まれず、主に農業・水産業や畜産分野の研究者や関連企業が数多く参加する学会となっています。特に、近年の超並列シーケンサーの登場によるゲノミクス研究の急速な進展はこの分野にも新たな研究展開や市場を生み出しており、シーケンシング関連の機器、試薬、受託関係の企業の参加が増え、非常に活気のある学会になっています。2年前からはPAG Asiaという分会が毎年シンガポールで開催されるようになり、アジアにおける農業・水産業や畜産分野の研究についての情報交換の場となっています。

第24回のPAGは今年の1月9～13日に開催されました。70を超える国から約3,100人の参加者が集まり、約2,400件の研究発表（約1,300件のポスター発表を含む）があったと報告されています。例年、日本からの参加者も多く、主に農水系の独立行政法人やその他の研究機関、大学などからが中心となっており、今年も100人近くの方が参加されていたようです。PAGのウェブサイトから提供されている統計情報によると^[1]、参加者や発表件数は年々増加傾向にあるようで、特に2009年頃からの急激な増加の要因は超並列シーケンサーの登場に他ならないと思われまます。私自身、2010年からほぼ隔年で参加していますが、当時は新型シーケンサーの登場に沸いたお祭りのような雰囲気で、毎年のように新機種の発表があるシーケンサーのベンダー主催のワークショップは立ち見が出るほどの大盛況でした。それから5年以上の時を経た今年のPAGは、シーケンサー開発については大分落ち着いたような印象を受けました。実際に、会期中に発表された新しいシーケンサーは、Illumina社のMiniSeqぐらいで、他にはPacific Biosciences (PacBio) 社がRoche社と共同で開発して昨年の秋に発表したSequelシステムの紹介や実機がブースに展示されていたぐらいで、技術的に目新しいものはほとんどありませんでした。一時期注目を集めていたいわゆるナノポア技術を用いたOxford Nanopore社のシーケンサーについても特に目立った発表はなかったように思います。超並列シーケンサーの登場以降、ベンダー各社がしのぎを削る戦国時代が続いていましたが、現在はロングリードタイプのPacBio社、ショートリードタイプのイルミナ社の二強時代であると言えるかと思えます。

現在、私はイネを中心とした作物のオミックス研究を主におこなっており、今回のPAGでは私達のグループから今年の1月に論文を発表した、mRNA-Seq法による様々なストレス環境下におけるイネのトランスクリプトームデータベース「TENOR」に関するデモンストレーションとポスター発表をおこないました^[2]。PAGにはComputer Demoというセッションがあり、データベースや解析ツールの使い方を実演しながら紹介することができます。PAGの口頭発表は基本的に招待講演のみですが、このComputer Demoは採択されれば誰でも口頭発表できるので、そういった成果をお持ちの方にはお勧めです。



写真 会場となるコンベンションセンターの入口（左）と南国の雰囲気が漂うタウン&カントリーホテルの敷地内の様子（右）。前日まで嵐のような天気だったようですが、会期中は天気にも恵まれ暖かい日が続きました。

ここからは、私自身の興味から農学関連の分野におけるシーケンシング技術を活用した研究の動向について紹介したいと思います。作物の育種(品種改良)は古くから人間にとって望ましい性質をもった植物を見つけて増やしたり、有用な形質をもった品種同士を交配させてより良い品種を作出したりする「従来育種」によって行われてきました。しかしながら、現在ではゲノム情報を有効に利用して交配をより計画的、効率的におこない、品種改良のスピードを大幅に向上させることが可能な「分子育種」が広まってきています。特に超並列シーケンサーの登場によって急速に広まった分子育種手法として、「ゲノムワイド関連解析 (Genome Wide Association Study: GWAS)」と「ゲノミックセレクション (Genomic Selection: GS)」が挙げられます。今回のPAGでも、GWASやGS関連の研究(タイトルや要旨を対象とし、「GWAS」や「GS」等のキーワードでヒットしたものは約340件(全体の約15%)もあり、非常に注目されている技術であることが分かります。GWASはヒトでもSNPマーカー等を用いて特定の病気と関連するSNPを見つけ出し、その近傍にあると推測される疾患関連遺伝子候補を同定する手法として用いられています。農学分野でも様々な品種を対象としたDNAマーカーと表現型の関係から、重要な農業形質(開花期や収量性、食味など)に関わる遺伝子を同定する目的で用いられています。一方のGSは、トレーニング用の集団を用いてDNAマーカーと表現型の関係の予測モデルを予め構築した上で、交配後代の選抜集団のゲノムワイドな遺伝子型情報を予測モデルに当てはめ、最も望ましい表現型をもつことが予測される優良個体を予測、選抜する方法です。GWASが表現型を支配する原因遺伝子を検出することを目的としているのに対し、GSは原因遺伝子については分からなくても良く、とにかくゲノムワイドなDNAマーカーから表現型を予測し、優良個体を選抜する手法である、という違いがあります。どちらも育種スピードを大幅に向上させることができる技術として期待されていますが、これらの手法がここまで広く利用されるようになったのは、シーケンシング技術の進歩によって多数の品種や系統の遺伝子型情報が容易に得られるようになったからに他ならないでしょう。

リシーケンシング解析などによって遺伝子型情報が容易に得られるようになった一方で、表現型情報をいかに効率よく得るかということが重要な問題になってきています。これまで、作物の生長形質(草丈、葉の大きさや枚数、開花期など)や収量形質(種子数や穂の数や大きさなど)に関する情報は、栽培している作物を手作業で計測して集められており、多数の品種や系統数をこなすには限界がありました。そこで期待されているのが効率よく表現型情報を収集するためのフェノーム解析技術です。今回のPAGでもワークショップや企業セッションでフェノーム解析に関わる研究発表が多数あり、特に栽培現場でドローンや大型の撮影機材によって取得した大量の画像データから形質情報を抽出するための技術開発やその普及に関して議論されていました。

様々な生物種において多数の系統、品種のゲノム配列データを用いた多様性解析が進められるようになったことで、リファレンスゲノムや遺伝子アノテーション情報の不足や質の低さが問題になってきています。イネを例に挙げると、アジアイネ(*Oryza sativa*)の栽培種はジャポニカとインディカの2種類の亜種に分類されますが、高精度なゲノム情報が利用できるのは、国際コンソーシアムが解読したジャポニカの「日本晴」という品種のみです(ちなみにイネのゲノムサイズは約390 Mbp)。そのため、例えばショートリードタイプのシーケンサーで解読した多数のインディカ品種のゲノム配列情報を用いて多様性解析をしようとした場合、遠縁の日本晴のゲノム配列をリファレンスとして用いたのでは多くのリードがマップされず、別途解析する必要があるなどの扱いにくさが生じていました。そういった現状を打開するため、現在複数の研究グループによって、第2、第3のイネのリファレンスゲノム配列の構築が進行中であり、将来的にGWASやGS研究への利用が予定されているとのことでした。そのような目的で主に利用されているのはPacBio社のロングリードタイプのシーケンサーです。これまでもゲノムサイズの小さな生物ならロングリードによるゲノムアセンブルが有効だと言われていました。しかしながら、スループットが向上しリード長が伸びた今では、ゲノムサイズが大きくて複雑なゲノムこそPacBio社のロングリードタイプのシーケンサーが威力を発揮するという状況になってきています。あくまで目安にはなりますが、

ゲノムサイズの40～50x相当量の配列データを得ることができるのであれば、PacBio社のシーケンサーを用いることで、もっとも効率よく、精度の高いゲノムアセンブリが得られるようです。さらに、トウモロコシにおいては、リファレンスゲノムの高精度化と平行して、遺伝子アノテーション情報の高度化も進められており、様々な組織から作成した完全長cDNAライブラリをPacBio社のシーケンサーで解読する、Iso-Seq法という手法が用いられていました。ショートリードタイプのシーケンサーによるRNA-Seq法でも遺伝子構造予測が可能ですが、ショートリードのアラインメントから複雑なアイソフォーム構造を予測するのは容易ではなく、完全長cDNAの全長を丸々と読んでしまう、アセンブルいらずのIso-Seq法とは得られる遺伝子構造の精度は比べ物にならないでしょう。10年前には数十億円もの予算が必要だったトウモロコシのリファレンスゲノムと完全長cDNA配列の決定が、今では数百万円の予算でより高精度に出来てしまうという話は衝撃的でした。

以上のような最新の研究動向から、ここ暫くの間は、系統的、もしくは育種上の重要な品種についてはロングリードタイプのシーケンサーを用いて高精度なリファレンスゲノム配列と遺伝子アノテーションを整備し、多様性解析に用いる多数の品種や系統のゲノム解読はショートリードタイプのシーケンサーでこなうといった戦略が主流になるだろうと思われれます。超並列シーケンサーが普及し、誰もがPCRをするぐらいの感覚で大量のオミックス情報を手に入れられるような時代になりました。もはやシーケンシングをすることは当たり前で、シーケンシングをして何をするのか、というアイデアが重要だということであらためて強く感じました。

最後に、本稿を読んでいただくことで、PAGの雰囲気至少在りでも伝わり、来年以降の参加を検討して頂ければ幸いです。ちなみに、すでに2019年までのPAGの日程が会議公式ウェブサイトでご公表されており、来年は2017年1月14～18日に開催されるようです。

参考文献

- [1] <http://www.intlpag.org/2016/images/pdf/PAGXXIII-demographics.pdf>
 [2] Kawahara Y, et al., 2016. *Plant Cell Physiol.* Jan; 57(1):e7

(編集担当：工楽樹洋)

ミーティングレポート②

CeMEB Eco-Evo 国際カンファレンス in スウェーデン

荒木仁志 (北海道大学)

— 全ての始まりは、見知らぬ送り主からの一通のメールだった — という推理小説の書き出しのようだが、そうではない。昨年6月に送られてきたそのメールには、「自分はスウェーデン・イエテボリ大学の研究者で、10月に大学のマリンステーションで国際カンファレンスを開きたい。あなたの専門分野にピッタリのセッションがあるので、都合がつくようなら是非お招きしたい。」と書かれていた。

通常その手のメールは大方一日何十通にも及ぶ未開封メールの中に埋もれるのだが、イエテボリ大学のマリンステーション (The Linnaeus Centre for Marine Evolutionary Biology、通称CeMEB) については以前、水圏生物進化の先駆的な研究をしている、と耳にしたことがあった。何だか面白そうなイベントへの招待、という予感があったのかもしれない。ホームページで情報を確認し、身元を確かめた送り主と何度かメールでやり取りした後、何とか都合をつけてカンファレンス参加を決めた。(最近フィッシング詐欺ばりの「招待講演依頼メール」で蓋を開けてみると参加費自腹、お金目的などというものも多いので十分ご注意ください。)

“Applications of Evolutionary Biology in the Marine Environment”と名付けられたそのカンファレンスは、主にヨーロッパの水圏生物研究者を対象とした3日間のイベントで、その内容は微生物の適応進化から魚類生態まで、水圏生物の生態進化を幅広くカバーするものだった。Tjärnö (シャルノ) というノルウェーとの国境近くの海辺の町(村?)で、人里離れたマリンステーションの施設に合同宿泊しての3日間だったため、60名ほどの参加者は皆同じ宿舎で寝起きを共にし、朝晩の食事はもとよりフィールド視察や深夜のビアパーティーなど、その気になれば24時間お互いに顔を合わせる事の出来る理想的な環境だった(写真参照)。そのため、フィヨルドの海を眼前に眺める静かな周辺環境とはいえ、カンファレンス期間中は場所や時間を問わず、施設内のあちこちで時に熱くディープな研究談義が交わされていた。比較的若い研究者も多く、短い秋のサマースクール、といえは全体の雰囲気を表しているだろうか。

カンファレンスは「Harvest induced evolution」「Colonization genetics / invasive species」「Breeding to support stock production」「Biotechnological applications」および「Short talks」の5つのセッションで構成されていた。それぞれのセッションごとに招待演者2~3人がまず研究紹介し、その後5つほどのグループに分かれて関連する課題についてグループディスカッションを実施するスタイルだ。各グループには私を含め、それなりの年齢の専門家が割り振られているのだが、実際に議論の中軸となったのは主に30代前後の若手研究者や大学院生であった。もちろん文化的な側面や個人差はあるが、それだけ欧米の若い研究者や学生が幅広い知識をよく身に付け、それを日々テストされている、ということだろう。私も研究の場をアメリカやヨーロッパから日本に移して3年近く経つが、一般に若い学生の専門知識量には日本と欧米間での差があるように感じられる。(自分が若い頃にどれほど勉強していたか、についてはさておき。)

「Harvest induced evolution」セッションではフィンランド・ヘルシンキ大のAnna Kuparinenによる「Fisheries-induced evolution and its ecological feedbacks」と題した講演があった。漁網の目の細かさが捕獲される魚の体長に特異的な人為選択の要因となっており、その結果、漁業対象種の成熟年齢と成熟サイズに経年変化が見られる、といった事例を紹介しつつ、対象種の生活史と表現型への選択圧がどのように個体群動態と相互作用しながら変化していくのか、を理論的に予測するという内容だった。漁網サイズによる人為選択はRapid evolutionの例として魚類の進化学では有名な話だが、対象種が適応進化することで個体群の体サイズや年齢の構造、ひいては世代時間が変化し、それが漁獲に影響することによって人為選択スキームにもフィードバックがかかる、という点が複雑で、それゆえ面白い。ちなみにこの演者とは空港までの帰りの車と一緒に個人的にもいろいろと研究の話をする機会があったのだが、彼女曰く日本にも水産学系の研究者で関連研究をしている方がいらっしゃるとのことで、いずれ直接お会いしてお話を伺うつもりだ。

続いての講演は同じくフィンランドのトゥルク大、Silva Uusi-Heikkiläで、ゼブラフィッシュのトラン

スクリプトーム解析を用いて体サイズ依存の人為選択の分子基盤を解明しよう、という話だった。この研究においては5世代にわたる人工飼育で実際に体サイズにおける人為選択をかけ、発現量に変化が見られた遺伝子群を同定しようとしていた。その後のグループディスカッションでは特にこのアプローチの一般性や有用性が議論となり、賛否両論が展開された。個人的な見解を述べれば、このアプローチそのものの有効性は理解しつつも、魚はともかくショウジョウバエやとうもろこしなどでは昔からsize-selective artificial selectionの実験は行われていたわけだし、トランスクリプ



写真1 カンファレンス会場でもある宿舎からの眺望。

トーム解析によりどれほど新規性の高い結果が得られるか、具体的な成果には疑問を残す内容だった。

ユニークだったのは「Colonization genetics / invasive species」セッションで講演したスイス・ベルン大のLars Bosshardの話で、バクテリアが増殖し、空間的に広がりを見せる中で、中立・選択的・弱有害遺伝子がそれぞれどのように進化していくか、という話だった。こう表現すると完全に集団遺伝学の話だが、彼らが念頭に置いているのは高等生物の移入といった生態学的な事象で、何故外来種は既存種よりその環境にさらされている期間が短いにもかかわらず生息域を広げられるのか、といった疑問に答えようとしているのである。このアプローチの有用性についてもその後のディスカッションでは意見が分かれたが、私個人も集団遺伝学の考え方・ツールを使って生態学の問題に答えを探す立場にあり、より多くの人がこういったアプローチについて真剣に考えるよい機会になったように思う。

「Colonization genetics / invasive species」は演者が多く、2日に分けてのセッションとなったが、その後も侵略的外来種のRAD-Seqによる適応進化の話や地球温暖化関連で高温適応の分子進化メカニズムの話など、進化系カンファレンスの面目躍如となる興味深い話が印象的だった。

後半の「Breeding to support stock production」「Biotechnological applications」セッションについては私自身の講演も含め、やや水産学や保全学研究の色合いが強いため説明を割愛する。しかし、全体を通して今や次世代シーケンサーを用いた進化プロセスの分子基盤解析は完全に「当世代」の解析技術であり、個々の研究対象種ごとにアプローチの差はあるものの、ビッグデータに基づく解析は不可欠となりつつあることを改めて痛感するセッションとなった。これは日本の進化学会において「卒論研究に次世代シーケンサーを使った解析をやってみました」という類の話を聞く頻度が高まっていることとも対応している。ただし水産業にはお金が絡むためか、水産学分野でのこのアプローチによる問題解決の「本気度」や「熱」は進化学の上を行っているようだ。

ちなみに、中日となる2日目のランチ後には、スウェーデン北部のフィヨルドとフィールド実験サイトをボートで体感するフィールド視察も実施された(写真参照)。フィヨルドというとノルウェーの深く削られた山々を想像してしまうが、シャルノ周辺の様子はそれとは違い、低い岩肌が点々と存在する、やや殺風景な雰囲気だ。しかし、水面から下は深くそり立っていて、浅い部分はもとより深海性の魚を含む高い生物多様性のゆりかごになっているという。また周辺の森には国境を越えてノルウェーの富裕層が建てるという大きく立派な家が点在し、独特の雰囲気を醸し出していた。地元の学生に話を聞くとシャルノには何もなくて退屈、という答えが帰ってきたが、研究施設としてはしっかりした設備があり、研究に専念したい学生や夏にサバティカルで滞在するような研究者には快適な場所だろう。ちなみに、帰りに車でノルウェーの空港に向かって途中、それまでは現地の人ですら見たこともなかったという検問に出くわした。我々は無事通過出来たが、これも難民問題の影響だろうか。こんな清閑な場所にすら戦争の余波が及んでいるかと思うと気が滅入るが、CeMEBのイベントは毎年行われるようなので、機会があればまた



写真2 フィールド視察の様子。スウェーデンのフィヨルド生態系の説明に聞き入る一行。



写真3 海側から見たイエテボリ大・シャルノ・マリンステーション

訪れたい場所の一つとなった。

* CeMEBについては

<http://cemeb.science.gu.se/>

* カンファレンスプログラムについては

http://cemeb.science.gu.se/activities/CeMEB_Assemblies/cemeb-14th-assembly/programme

(編集担当：工樂樹洋)

編集後記

編集幹事 荒木仁志

進化学会は今年から田村新会長を迎え、新体制での活動開始となりました。私の住む札幌でも今年は例年になく雪解けが早く、春の訪れが近いことを感じさせます。進化学会ニュースでも何か新しいことを、ということで、長年続けてきた「海外研究室だより」に加え、学生視点で国内のユニークな研究室を紹介する「ダーウィン研究室」を立ち上げました。学生・研究者、それぞれの視点から楽しんで読んでいただける企画に育てていけたらと考えております。

日本進化学会ニュース Vol. 17, No. 1

発行：2016年3月11日

発行者：日本進化学会（会長 田村浩一郎）

編集：日本進化学会ニュース編集委員会（編集幹事：荒木仁志 副編集長：大島一正

編集委員：奥山雄大/工樂樹洋/佐藤行人/真鍋 真/山道真人）

発行所：株式会社クバプロ 〒102-0072 千代田区飯田橋3-11-15 UEDA ビル 6F

TEL: 03-3238-1689 FAX: 03-3238-1837

<http://www.kuba.co.jp> e-mail: kuba@kuba.jp