

日本進化学会 ニュース

Vol. 14 No. 2
July 2013

目次

- 01 リレーエッセイ〈4〉
植物の形の安定性について
長谷部光泰 (基礎生物学研究所)
- 02 第15回大会 (つくば大会) のお知らせ
- 04 シリーズ 第6回
「私と進化学」・下
長谷川政美 (復旦大学生命科学学院)
- 10 海外研究室だより 第16回
**楽都ウィーンで進化を学ぶ：
グレゴール・メンデル研究所・Magnus Nordborg 研究室**
土松隆志 (グレゴール・メンデル研究所 EMBO フェロー)
- 16 **MAFFT 多重配列アラインメントプログラムの紹介**
加藤和貴 (大阪大学 IFRc、産総研 CBRC)
- 21 ソフトウェア紹介
raxmlGUI 1.3 : RAxML を快適なユーザインタフェースで使おう
宮 正樹 (千葉県立中央博物館)

リレーエッセイ〈4〉

植物の形の安定性について

長谷部光泰 (基礎生物学研究所)

前回、このコーナーが会長と副会長の間での交換エッセイだということに気づかずに、あまりリレーになっていないエッセイで失礼しました。そういえば、10年ほど前に当時は裳華房から出ていた「遺伝」という雑誌で倉谷さんと10回ほどリレー対談をしたのを思い出しました。恐竜の作り方を縦糸に反復説など、発生進化学の問題点について楽しく議論させてもらいました。お暇のある方は図書館などで探してみてください。

6月に南仏Hyereで花発生と進化についてのワークショップがあったのですが、倉谷さんの記事を読んで、シャルルドゴール空港からオルロ空港に移動する途中にパリ植物園があるのを思い出しました。秘書の「また乗り遅れますよ」という暖かい助言を気にしつつ、2時間ほど滞在して講義用の資料収集をしました。

私は講義で、シジクモ藻類から始まって被子植物まで各系統でどのような形態が進化したのか、どんな問題があるのかを話します。分子系統学の進展で陸上植物の系統関係はかなりの部分がわかってきましたので、系統樹に基づいてどの分枝でどんな形質が進化したのかをトレースでき、とても説明しやすくなりました。ところが、明らかになった系統関係は、従来の系統学、形態学の常識を逸している場合が多々あり、新しく定義された科や目の共有派生形質の多くは、これまであまり注目されてこなかった形質が多いのです。ですから、講義で説明しようにも文献やネットで共有派生形質をよく示す写真が出てこないことがまあまり困ります。また、自分で一度も見たことが無い植物だと、どうしても話が表面的になってしまいます。

そんなわけで、出張に行くときは多少無理をしても植物園やフィールドに立ち寄って講義用の写真を撮ることにしています。もちろん、講義だけでなく、研究上のアイデアがうかんでくることも多々あります。パリ植物園では、日本ではなかなかお目にかかれない地中海性気候の地域周辺に分布するモクセイウ科の花が咲いていました(図1)。おかげで共有派生形質だと言われている花弁上の突起の写真を撮ることができました。しかし、なんでまた花弁の途中にこんな突起ができるのか、どうやってこんな花弁ができるのかなんとも不思議に思いました。

個々の分類群に固有な形質でも発生進化学的、形態学的に面白いものがあります。ここにはスギナモ(オオバコ科)の花がありました(図2左)。オオバコ科はオオイヌノフグリやキツネノテブクロなど綺麗な虫媒花が多いのですが、オオバコのように風媒に進化し、花弁やガク片が目立たない種もあります。スギナモは水草で花序を水上に立ち上げます。風媒花で雌蕊1本、



図2



図1



雄蕊1本、そして、両者を1枚の葉的器官が包んでいます(図2中)。普通の花は雌蕊と雄蕊は花弁とガク片に包まれるのに、スギナモの花は1枚の葉的器官しかありません。この器官は花弁なのでしょうかガク片なのでしょう。一部の器官が退化すると相同性がわからなくなるのは動物も植物も同じです。さらに、スギナモではもう一つわからない点があります。被子植物の多くは「葉」とは形態が異なる葉的器官「苞」の腋に、花を形成します。スギナモの場合は、花を抱えている葉

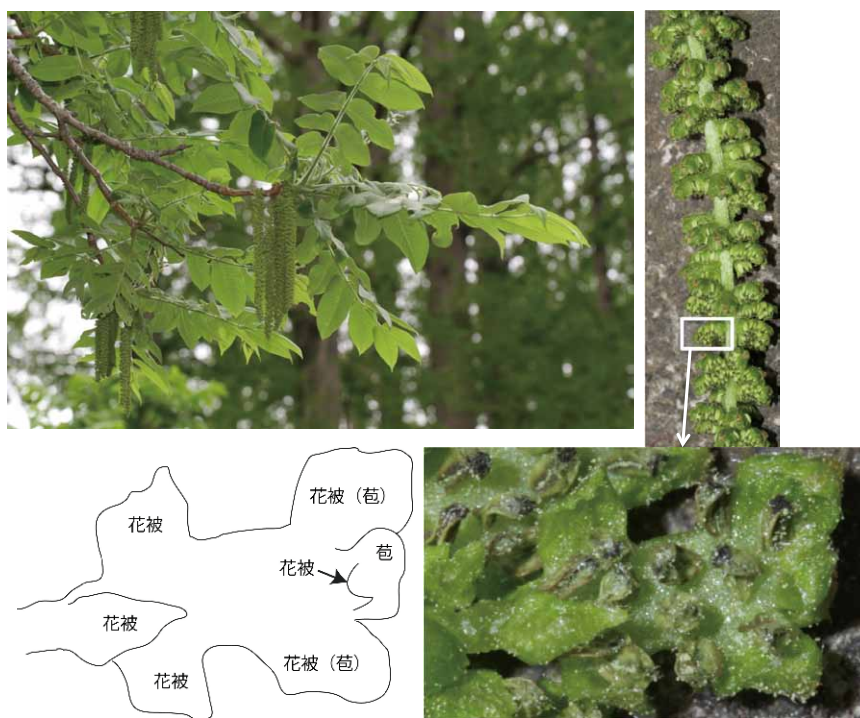


図3

的器官(図2右下)は外見上、「葉」(図2右上)と区別できません。ですから、花を抱えている葉的器官が苞なのか葉なのかはわからないのです。

器官の相同性がはっきりしない花は他にもたくさんあります。例えば、日本の川沿いに良く生えているクルマミ(オニグルミ)があります。クルマミは5月初旬に垂れ下がる長い雄花序(花がついた茎を花序と呼びます)を形成します(図3左上)。図3右上の図で白い四角で囲んだ部分が花と苞が融合したものだと考えられています(図3右下が苞と花が癒合した器官の拡大写真、左下がその輪郭を示した図。ごま粒のように見えるのが雄蕊で12本あり、輪郭図では省略してあります)。この器官、苞が3枚で4枚の花被(花弁かガク片かわからないのでまとめて花被と呼びます)があると考えられたり、苞は1枚で6枚花被があると考えられたりしています。発生過程を見ても、維管束の走行を見ても、どちらが正しいのかはよくわかりません。

スギナモやクルマミの花を見ていると個別種における器官の相同性という、重箱の隅的な面白さに加えて、より一般的な問題点も気になります。つまり、一度確立されたボディープラン(この場合は「花」)をどこまで崩せるのか。崩すといっても種内や属内では形態が保存されているのですから、崩したなりに安定した状態になっているはず。形態変異空間を仮定するなら、いくつかの安定点があるのかもしれない。だとすると安定点はどのように決まるのでしょうか。この問題は、結局、発生制約とは何なのかという問題ともつながります。スッポンや食虫植物や、その気になればどんな材料でも研究できる時代になったので、今後、いろいろな生物の研究がさらに進むとますますいろいろな事がわかってこないかなと楽しみです。

第15回大会(つくば大会)のお知らせ

8月28日(水)から8月31日(土)の期間、日本進化学会第15回大会を筑波大学で開催します。シンポジウム、ワークショップの詳細が決まりました。一般発表のプログラムもまもなくHPでお知らせできると思います。夏のつくばで皆様とお会いできることを楽しみにしています。つくばへのアクセスなど、詳細は、HPで。
<https://sites.google.com/site/tsukubacce/>

国際プレナリーシンポジウム

“Bottleneck in development and evolution”

Shigeru Kuratani (Riken CDB)

Naoki Irie (Riken CDB)

Mitsuyasu Hasebe (NIBB)

Pavel Tomancak (MPI-CBG)

Marcel Quint (Leibniz-Institut für Pflanzenbiochemie)

シンポジウム・ワークショップ

8月28日(水)

Methods for molecular evolutionary studies in special reference to massive data

(企画者 斎藤成也：国立遺伝学研究所)

昆虫の形態多様性をもたらす遺伝子の機能 —最近のトピックスと今後の展望—

(企画者 行弘研司、畠山正統、瀬筒秀樹：独立行政法人農業生物資源研究所)

高校生物の進化分野はどこが変わったか？—新しい教科書『生物』を比較する—(無料公開)

(企画者 嶋田正和：東京大、中井咲織：立命館宇治中高)

ウズベン葉緑体の進化 (企画者 稲垣祐司：筑波大学)

海外での生物調査、生物標本・サンプル=生物遺伝資源の取得とABS問題

(企画者 村上哲明：首都大学東京)

8月30日(金)

進化的原動力としての共生

(企画者 稲垣祐司：筑波大学)、神川龍馬：京都大学、松崎素道：東京大学、菊池義智：産総研)

DNAバーコーディングで何ができるか (企画者 熊澤慶伯：名古屋市立大学、伊藤元己：東京大学)

古代ゲノム学：地球科学と生命科学の融合 (企画者 遠藤一佳：東京大学)

形と進化 (企画者 荒木仁志：北海道大学、細将貴：京都大学)

Sexual selection of hermaphroditic animals: One further step from “So what?”

(企画者 中寺由美：VU University Amsterdam)

種分化を伴う進化と伴わない進化—環境傾度に対する植物の適応進化とその生態・遺伝的背景

(企画者 田中健太：筑波大学、北山兼弘：京都大学)

進化学会夏の学校(一般無料公開) 8月31日(土) 9:00～12:00

橋本哲男(筑波大)「生命の樹—昔と今」

石田健一郎(筑波大)「葉緑体—細胞内共生が駆動する細胞進化」

長谷川真理子(総研大)「ヒトはなぜヒトになったか？—ヒト固有の性質の進化—」

遠藤一佳(東京大)「バイオミネラル化の起源と進化」

公開講演会(一般無料公開) 8月31日(土) 13:30～15:00

植物多様性を生かして活かす植物園の研究 奥山雄大(国立科学博物館植物研究部)

自然史標本からわかること ～標本を利用した研究例の紹介～

佐藤 崇(国立科学博物館標本資料センター)

標本・観察データのデータベース化～意義と利用～ 福田知子(国立科学博物館植物研究部)

15:00～ 一般参加者から40名限定で、国立科学博物館 施設見学ツアーを開催

「私と進化学」・下

長谷川政美 (復旦大学生命科学学院)

2007年以降—上海・復旦大学に移ってから

2007年3月に私は統数研の定年を迎えた。以前からの共同研究者であった上海の復旦大学生命科学学院のYang Zhongさん(図31)からの誘いがあり、4月から上海に移ることにした。大学院生の研究指導以外の義務はほとんどなく、中国語が喋れないので会議も免除され、研究に専念できることになった。中国には1999年以来度々訪れて各地を回ったが、中国は生物多様性の宝庫であり、この国の生物のことをもっと知りたいということも、復旦大に移った動機であった。私の定年前にちょうど総研大で学位をとった米澤君を講師として復旦大で採ってもらい、研究室は最初この2人でスタートした。



図31 右からYang Zhongさん、Bojian Zhong、筆者(2010年、チベット・ラサのポタラ宮殿を望むホテルにて)

間もなくして、Yang Zhongさんの修士の学生のBojian Zhong(図31)がわれわれの研究室で研究することになる。Yang Zhongさんは植物学者なので、Bojianも植物の研究をやることにした。彼は実験的な研究よりも、データ解析のほうを希望したので、それまでにデータベースに蓄積していたイネ科植物の葉緑体ゲノムデータの解析をやってもらうことにした。1990年代初頭に葉緑体DNAの解析が始まった頃から、イネ科植物はほかの系統よりも進化速度が高いことが指摘されており、その理由について多くの議論があった。Bojianの研究により、確かにイネ科植物の進化速度は高いが、特に高いのはイネ科植物の共通祖先の枝においてであり、現生のイネ科植物ではそれほどでないことが示された。このように時間による進化速度の変化を知るには、進化の時間スケールを正しく把握することがまず必要となる。分子時計を仮定しないベイズ法による年代推定では、隣り合った枝での進化速度には相関があるとするCorrelated rate modelを使うのが一般であった。つまりあまり急激には進化速度は変化しないものであるという仮定である。しかしながら、イネ科の共通祖先で起こった進化速度の上昇は、このモデルではフォローできないほど急激なものであることが示された。さらに、イネ科共通祖先の枝で適応進化と思われるアミノ酸置換がたくさん起こっていることも示された。これらの研究成果は、Zhong et al. (*PLoS One* 4(4): e5297 (2009))として発表した。

その後Bojianは、グネツム目という植物の系統的な位置づけの問題に挑戦する。グネツム目には東南アジアのグネツム科、漢方薬麻黄のマオウ科、それにアフリカのナミブ砂漠にしか分布しないサバクオモト1種の奇想天外科の3科しか含まれない。グネツム目が種子植物のなかのどこに位置づけられるかという問題は、長い間植物系統学者の間で論争の的であった。形態学からは顕花植物の姉妹群と考えられたが、分子系統学のほうからは球果植物との近縁性が示唆されてきた。Bojianは、当時データベースに登録されていた葉緑体ゲノムデータを集めて分子系統樹解析をした。球果植物にはマツ科とスギなどのヒノキ科があるが、Bojianの最初の解析ではグネツム目がヒノキ科に近いという系統樹が100%のブートストラップ確率で支持された。ところが、解析を進めてみるとこの結果はどうも間違いであることが明らかになってきた。

この解析では、ヒノキ科の代表としてはスギのデータしか入っていなかったが、明らかにグネツム目とスギの系統ではほかよりも進化速度が高くなっていた。Long-branch attraction (LBA) によるartifactが疑われるのである。LBAは進化速度の高い枝ほど多重置換が過小評価される傾向が強いので、その結果、間違っ

い枝同士が組んでしまうことである。進化速度の高い遺伝子ほど、多重置換が多くなるので、LBAの影響を受け易い。そこで、進化速度の高い遺伝子を解析から除いていくと、次第にグネツム目がスギではなくマツと組むようになってきた。このことは、最初に得られた系統樹がLBAのartifactによるものであることを示唆する。もう1つの問題は、ヒノキ科の葉緑体ゲノムデータとしては日本のスギしか登録されていなかったということである。ところが、ゲノム全部ではないが、部分配列であればもっと多くの種からデータが得られていた。そのような部分配列を解析してみると、意外なことが明らかになった。それは、グネツム目の共通祖先の枝とヒノキ目の間で多くのアミノ酸平行置換が起こっていたのである。一方、マツ科との間ではそのようなことが見られない。ところが、ゲノムの解析では、ヒノキ科の代表としてスギしか入っていないので、平行置換は正しく平行置換だとは評価されず、その分だけ誤って強くヒノキ科がグネツム目に近縁であるように見えていたのである。

従来、分子レベルでは収斂進化や平行進化はまれにしか起こらないものだと考えられていたが、実際には考えられていた以上に高い頻度で起こっていることが分かってきた。この研究は、Zhong et al. (*Mol. Biol. Evol.* **27**: 2855-2863 (2010)) として発表した。Bojianは、このような研究成果を残して修士課程を修了したあと、New ZealandのDavid Penny (図22) のもとで博士後期課程に進学した。中国の学生は一般に、日本にくらべて海外に積極的に進出しようという傾向があるようだ。

Bojianが去った後、新しい院生としてJiaqi Wu (図32) がわれわれの研究室に入ってきた。彼女は哺乳類の進化に興味をもっていたので、ミトコンドリアのゲノムデータによる真獣類の系統解析を行なった。その頃までに、ミトコンドリアから得られた系統樹が核遺伝子によるものと食い違っている例がいくつか知られていた。それらのなかには交雑によるミトコンドリアの他種への移入 (introgression) や祖先多型のincomplete lineage sortingなどによるものもあるが、様々な原因でミトコンドリアによる遺伝子系統樹の推定が間違っていると思われるものが多かった。そのため、これまでもっぱらミトコンドリアを使って系統解析をしていた人々の間でも、ミトコンドリアは系統樹推定のマーカーとしてふさわしくないのではないかという疑念が湧いてきていた。Jiaqiはデータベースから真獣類のミトコンドリアゲノムのデータを集めたところ、463種の真獣類の配列データが集まった。このような大量データを使ったら、これまで明らかに間違っていた関係を示していたミトコンドリア系統樹がどうなるか、という点に興味があった。

少し前までは到底考えられないことであったが、Alexandros Stamatakisが開発したRAxMLを使えば463本の配列データから最尤系統樹を得ることが可能なのである (ただしこれが厳密に最尤系統樹だという保証はないが)。このような計算能力の進歩により新しい展望が開けてきたのである。分子系統樹解析において種のサンプリングを密にすることは、最尤法を使えば祖先配列が正しく推定できるということで、大事なことである。従来は、まばらなサンプリングしか行われなかったので、祖先配列が曖昧なままであったために、系統樹推定がうまくいかなかったということがあったようである。Jiaqiが463本のミトコンドリア配列データを解析した結果、それまでまばらな種サンプリングではおかしなところに位置づけられていたいくつかの種が、核遺伝子系統樹と同じ位置にくるようになった。データの量が増えるということは、質的にも新しい状況を生み出すものである (Wu et al., 投稿中)。種のサンプリングが疎であると間違っていた系統樹が、密にすると正しい系統樹になるというのは、最尤法のように祖先配列を推定しながら系統樹推定を行うという方法を用いた場合に成り立つもので、距離行列法ではうまくいかない。距離行列法では祖先配列を考慮



図32 復旦大学の筆者らの研究室のメンバー、左から院生のJiaqi Wu、筆者、副教授の米澤隆弘君 (2012年、筆者らの研究室の入口にて)

しないので、種のサンプリングを密にしたというメリットが生かせないからだと思われる。また最節約法でもだめである。この方法では祖先配列は考慮されるが、多重置換の効果が入っていないからであろう。最尤法であればどんな置換モデルでもよいかというと、そうではなく、正しい系統樹を得るためには現実の過程をある程度よく近似したモデルが必要である。

復旦大学に移ってから米澤隆弘君 (図32) が中心になって始めた研究に、家畜化に伴う遺伝的な変化についての研究がある。これは甘肅省・蘭州大学のJianquan Liuさんのグループとの共同研究として始められたものである。

青海チベット高原で家畜化されたウシ科のヤクという動物がいる (図33)。彼らは、野生のヤクと家畜のヤクの間の遺伝的な違いを調べるために、21頭の野生ヤクと51頭の家畜ヤクのミトコンドリアゲノム配列を決定していた。米澤君がこのデータを解析したところ、家畜化された集団では野生集団とくらべて非同義(アミノ酸)置換と同義置換の比がはっきりと高い値になっていることが分かった (Wang, Yonezawa, et al. (2011) *Mol. Biol. Evol.* **28**: 1553-1556)。このことは、ヤクの家畜化に伴ってミトコンドリアゲノムにかかる自然選択圧が低下しているためと解釈される。

野生のヤクは標高4,000メートル以上の特に冬は過酷な環境を生き抜いているため、それに耐えられる遺伝的な形質をもつことは必須である。家畜化の過程では、人間による選択圧がかかる形質もあるが、過酷な自然環境を生き抜くための形質は多少劣化しても許容されるようになったものと考えられる。Björnerfeldtらの研究により、同じような傾向はオオカミから家畜化されたイヌにおいても報告されている。家畜化はCharles Darwinが生物進化を考える際にモデルとしたものであり、進化学と共通する側面が多い。Jianquan Liuさんのグループは更に核ゲノムの解析により、ヤクの高所適応の手掛かりを得ている (Qiu et al. (2012) *Nature Genetics* **44**: 946-949)。

2010年になって、昔関わっていた懐かしい問題に再会することになる。1989年のIwabe論文で扱った真正細菌、古細菌、真核生物に関する問題である。Iwabe論文はこれら地球上のあらゆる生物の共通祖先、つまりUniversal common ancestor (UCA) の存在を仮定した上で、真正細菌、古細菌、真核生物の間の関係を議論したものであるが、Douglas TheobaldはそもそもUCAがいたかどうかを統計的に検定することを試

みた論文を2010年に*Nature*に発表した。現在地球上に生きている生物は一つの共通祖先から進化したという考えは、多くのひとたちによって支持されている。その根拠の一つは遺伝暗号表があらゆる生物でほぼ共通(普遍暗号)だということがある。大澤省三先生 (図34) らは1980年代から非普遍暗号を多数報告されたが、そのいずれもが普遍暗号のちょっとした修正版に過ぎないので、非普遍暗号の存在がむしろUCAの存在を支持するものであった。大澤先生は普遍暗号から非普遍暗号が



図33 チベットの家畜ヤク (2001年、チベットにて)



図34 大澤省三先生の講演風景、司会者は五條堀孝さん、進化学会の現在のメンバーが多く見える(統数研での研究会にて、1998年)

進化する過程をうまく説明する理論を提唱された (Osawa, S. "Evolution of the Genetic Code". Oxford Univ Press (1995) (日本語訳: 大澤省三「遺伝暗号の起源と進化」渡辺ら訳、共立出版 (1997))。

このようにUCAの存在は広く受け入れられているが、それを支持する証拠は状況証拠だけで、直接的にその存在を証明しようとする試みはなかった。Douglas Theobaldは多くの保存的な蛋白質のアミノ酸配列データの統計的検定により、これを証明しようとした。真正細菌、古細菌、真核生物が一つの系統樹上でまとまるという仮説 (UCA仮説あるいは共通祖先仮説) と真正細菌は古細菌+真核生物とは独立の祖先から進化したとする仮説 (独立起源仮説) とを比較したのである。独立起源仮説にはこのほかにさまざまな組み合わせのものも比較の対象になる。ただし、これらの仮説 (モデル) のパラメータ数は違うので、尤度で直接比較することはできない。そこで彼は赤池情報量規準AICで比較した。その結果、共通祖先仮説がどのような独立起源仮説よりも圧倒的に良い仮説であることが分かった。そのことから、Theobaldは独立起源仮説が棄却できたと主張した。

ところが、彼の議論にはいくつかの問題があることに気がついた。米澤隆弘君が少し皮肉っぽい例を見つけてきた。ウシ、シカ、カバのミトコンドリアのCyt-bとNd2をもってきて、この2つの遺伝子が共通祖先から進化した仮説と独立に進化した仮説とを比較してみたのである。アラインメントは共通祖先仮説に有利な偏りを与えるはずだから、アラインメントはせずに、5'端から読み始めコドンだけを除いて、同じ長さに揃えたデータを使った。その結果、なんと共通祖先仮説のほうが支持されたのである (T. Yonezawa and M. Hasegawa (2010) *Nature* 468, E9)。でもそうだからといって、Cyt-bとNd2が共通祖先から進化したと考えるひとはいないだろう。同じミトコンドリア・ゲノム中にある遺伝子であるために、塩基組成やコドン使用頻度が似ていることから、実際には独立に進化したものが共通祖先から進化したように見えただけと思われる。

このことは、Theobaldの議論の問題点をあぶり出した。いろいろな制約のもとで、遺伝子は進化してきたものであり、たまたま共通の制約がかかった状況で独立に進化したものが、見かけ上共通祖先から進化したように見えることもあるだろう。特に、最近遺伝子レベルでの収斂進化の例が多く報告されている。Theobaldの解析では、このような可能性はまったく考慮されていない。更にもっと致命的な欠陥は、彼がアラインメントした配列データを使っているということである。そもそもアラインメントは、共通祖先を仮定した上で行われる作業なので、このことはあらかじめ共通祖先仮説に有利な偏りを与えている (T. Yonezawa and M. Hasegawa (2012) *The Scientific World Journal*, 2012: 479824)。

それでは、Theobaldが志したことを、もっとちゃんとやるにはどうしたらよいだろうか? 現在の分子系統学では通常、アラインメントと系統樹推定が別々の作業として行われている (少なくとも最尤法で通常行われているやりかたでは)。本当は、この2つは密接に関連したものであるから、同時に進めることが望ましい。将来最尤法の枠組みの中で、アラインメントと系統樹推定が同時に行われるようになれば、共通祖先仮説と独立進化仮説とがどちらかに偏ることなく、公平に比較できるようになるであろう。ところが、収斂進化の可能性を考慮して、2つの仮説を比較するのは現在の分子系統学の枠組みのなかでは困難である。分子系統樹推定のための置換モデルは通常発散的なものであり、収斂進化が検出されるのは、ほかの遺伝子の情報から確立された系統樹を仮定して行われるのである。収斂進化の可能性を考慮した検定を行うためには、全く新しい理論的な枠組みが必要である。いずれにしても、仮説検定はいくつかの仮定の上で成り立っているという認識が必要であろう。

この議論の際、1990年頃のArchaea論争が少し再燃した。Theobaldは古細菌、真正細菌をそれぞれArchaea, Bacteriaと呼んでいたが (若いTheobaldは多分昔の論争は知らないと思う)、*Nature*に出したTheobaldに対するコメントのなかでわれわれは古くからのArchaeobacteria, Eubacteriaを使った。*Nature*の編集部はなぜArchaeaを使わないのかと言ってきたが、1990年頃のこれに関する論争を説明し、Archaeobacteria, Eubacteriaで通してしまった。Woeseは1989年のIwabe論文で古細菌が真核生物に近縁であることが判明したことから、古「細菌」はおかしいとして、Archaeaを提唱したのであった。しかしこの議論は、硬骨

魚類が軟骨魚類よりも四足動物に近いから硬骨「魚類」と呼ぶのはおかしいと言っているようなものである。

復旦大学に移ってからは自由な時間がとれるようになったので、一般向けに哺乳類の系統進化の本を書きたいと思うようになった。1998年に国際哺乳類シンポジウムを開いて以来10年以上が経過していたが、その間の分子系統学の進歩には目覚ましいものがあり、系統学の多くの問題が解明された。もちろん分子系統学には未だにいろいろな問題があり、解明されたと思っているものであっても、将来間違いが明らか

になるものがあるかもしれない。しかしながら、それよりも数年前にくらべるとかなり安定した解析結果が得られるようになっていたので、今がそのような本を書くにはちょうどよいタイミングではないかと思った。2010年の夏頃のことである。その年の9月から集中して3か月ほど書き上げた。

哺乳類のほかに鳥類の系統に関してもかなりのことが分かってきていたので、この2つの動物群を中心にした本にすることにした。系統樹には動物の写真を入れてヴィジュアルな本にしようと思った。1990年頃から学会発表等で使う動物の写真はなるべく自分で撮ったものを使いたいと考えていたので、海外・国内を問わず旅行した際にはいろいろな動物の写真を撮るように心掛けていた。本には結局合計すると450枚以上の写真を使ったが、どうしても自分のコレクションだけでは間に合わない15枚くらいは上野動物園長だった小宮輝之さん(写真4)など何人かのかたに頼んで提供してもらった。2011年の2月に「新図説・動物の起源と進化」というタイトルで八坂書房から出版された(図35)。小宮さんからは、この本は新しい時代の図鑑のひな型になるだろうというコメントをいただいた。生物多様性を理解する基本は系統樹であり、様々な生物群である程度信頼することのできる系統樹が得られつつあるのだから、生物図鑑もそれを反映したものになる必要があるだろう。系統樹のかたちで様々な生物種が写真で示されれば、形態がどのようにして進化してきたのかについて読者はいろいろと思いを巡らすことができるであろう。

分子系統学の新しい展開

21世紀に入り、生物学はポストゲノム時代を迎えた。ポストゲノムとは、ゲノムの時代が終わったという意味ではなく、ゲノム・データが生物学研究の前提になってきたということである。多くの生物種でゲノムのデータが得られるようになり、ヒトなどいくつかの種については多くの個体のゲノムが解読されるようになってきた。このような潮流は分子系統学にも大きな影響を与えている。哺乳動物の場合、一個体のもつ核ゲノムには父親由来と母親由来の2セットがあるが、組み換えがあるために様々な由来をもつ遺伝子が混在している。ミトコンドリアの場合は、組み換えがなく母性遺伝なので、現在70億人いるといわれる現生人類のミトコンドリアの共通祖先を遡っていくと、いわゆる「ミトコンドリアのイヴ」にたどり着くが、核遺伝子に関してはそのように単純な図式は成り立たない。「ミトコンドリアのイヴ」の同時代に生きていた多くの人々の遺伝子が、現代人一人一人のゲノムを作っているのであり、ネアンデルタール人からの遺伝子の寄与もあっていわれている。このように遺伝子の進化といっても、その基本は集団レベルの現象であり、これを理解するには集団遺伝学が必須である。分子系統学の目的は、種分化の歴史をたどることであるが、種分化とはあくまでも集団として分かれていく過程であり、遺伝子系統樹は集団としての種系統樹と必ずしも一致しないし、

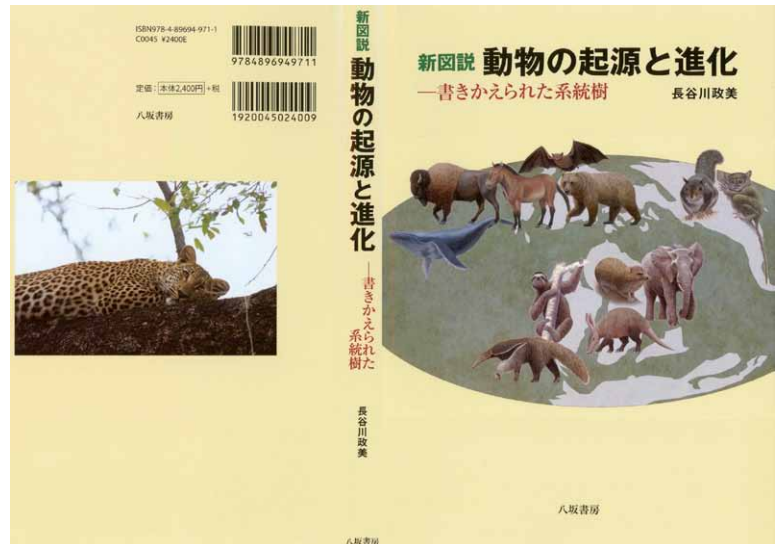


図35 2011年に出版した「新図説・動物の起源と進化」(八坂書房)のカバー

遺伝子ごとにも食い違いがある。このように多くの遺伝子系統樹を統合して種の歴史をたどるのが、ポストゲノム時代の分子系統学の役割であろう。

私が統計数理研究所を定年で辞めた数年後の2010年に、集団遺伝学の理論家の間野修平さん(図36)が名古屋市立大学から研究所に准教授として赴任して来られた。定年後も特命教授として研究所に籍を置かせていただいているので、私も日本にいるときには時々研究所に行き、いろいろな議論をさせていただいている。間野さんのポストクに以前ヒグマの



図36 右列、手前から間野修平、中込滋樹、Peter Lockhart、左列、手前から下平英寿、二階堂雅人、Andy Shedlock (2011年、チベット・ラサにて、Andyの背後にうっすらと見えるのがポタラ宮殿)

研究をしていた中込滋樹君(図36)がいる。ちょうどその頃、*Science*誌に面白い論文が掲載された(Hailer et al. (2012) *Science* 336, 344)。これまでのミトコンドリアDNAの解析からは、ホッキョクグマはヒグマの内部系統から進化したとされてきた。ヒグマはユーラシアから北アメリカまで非常に広い範囲に分布するが、ミトコンドリアDNAを見ると、ホッキョクグマはアラスカ太平洋側の一番南のABC Islandsのヒグマと特別に近縁なのであった。ところがHailerたちは多くの核遺伝子を解析した結果、これと食い違う結果を得た。彼らの系統樹では、ABC Islandsのものも含めて世界中のヒグマが系統的に一つのグループとしてまとまり、ホッキョクグマはその姉妹群になったのである。Hailerたちの解釈は、核遺伝子の示す系統樹が種分化の歴史を表しており、最近の交雑によってABC Islandsのミトコンドリアがホッキョクグマの集団に移入されたというものである。これは一見、説得力のある解釈だと思われるが、中込君が彼らのデータを解析しなおした結果、様々な問題があることが明らかになった。

Hailerたちは14個の核遺伝子を解析しているが、実は遺伝子系統樹は遺伝子ごとに違っているのである。ところが彼らはお互いに矛盾した系統関係をもつ遺伝子を無理やり連結して一本の配列にまとめた上で系統樹解析を行い、ヒグマ単系統の系統樹を得たのである。明らかに矛盾した系統関係をもつ遺伝子を連結して、1つの系統樹を描くことに果たしてどんな意味があるのだろうか？中込君は集団遺伝学的な解析を行った結果、Hailerたちが主張するモデルよりも、ホッキョクグマがヒグマの地域集団から進化したというモデルのほうがデータをよく近似していることを示した(Nakagome, Mano & Hasegawa, *Science* 339, 1522)。遺伝子ごとに系統樹が食い違うということは系統学にとって困った問題として捉えられがちであるが、むしろこの問題に正面から取り組むことによって、種分化に関するわれわれの理解が深まるであろう。このように大量データを扱うポストゲノム時代の分子系統学において、集団遺伝学の果たす役割がますます大きくなってきているように思う。

最後に

私は日本ではすでにとくに定年を迎えた年齢であり、自分自身がいつまで研究を続けられるか分からないが、この歳になって思い出されるのは、Ernst Mayr(図37)が100歳になって*Science*に書いた回想論文である(Mayr (2004) *Science* 305: 46-47)。彼は進化の総合説を打ち立てるのに貢献した人たちのなかの最後の生き残りであった。彼は総合説の黄金時代を生きてきたことに満足した気持ちを表明する一方で、最後に進化生物学のこれからの発展を自分は見届けることができないう寂しい気持ちの表明で文章を終えている。

分子系統学の歴史を振り返ってみると、確かに私が研究してきた時代は分子系統学の黄金時代だったと言えるかもしれない。そのような時代に研究生を送ることができたことは、非常に幸運だったと思っている。その間に分子系統学の方法が急速な進歩を遂げ、系統学の多くの問題が解明されてきた。しかし、系統関係を知ることは研究の目的ではなく、出発点に過ぎない。生物のもつさまざまな性質がどのように進化したかを調べる第一歩は、まず信頼できる系統樹を構築し、その上に生物の性質(形質)をプロットすることであろう。そうすることによって、どのような順番で形質の進化が起こったのかが分かる。分子系統学が発展する前は、形態形質だけで系統樹を構築していたので、収斂進化が起こっていてもそれを見抜くことができず、議論が同義反復になっていた。分子系統学が進歩したおかげで、初めて形態進化の歴史をしっかりとした基盤の上で議論できるようになってきたのである。その意味で、進化生物学の黄金時代はこれからである。ゲノムデータの出現は、遺伝情報の全貌が明らかになったという意味で重要であるし、発生生物学の進歩はこれからの進化生物学の原動力になるであろう。

図37は、1987年に私がカリフォルニア大学のBerkeley校にAllan Wilsonを訪ねたときに、Ernst Mayrがたまたま講演していたので、聴講させてもらった際のものである。講演の内容は、彼自身の種分化の理論であった。講演後、若い研究者が彼に対して厳しい批判的な質問を浴びせていたが、彼は力強い口調で反論していたのが印象的であった。そのとき彼はすでに83歳であった。私自身もせめてあのときの彼の歳くらいまでは進化生物学の発展を見届けることができるとすれば、と考えると今後が楽しみである。

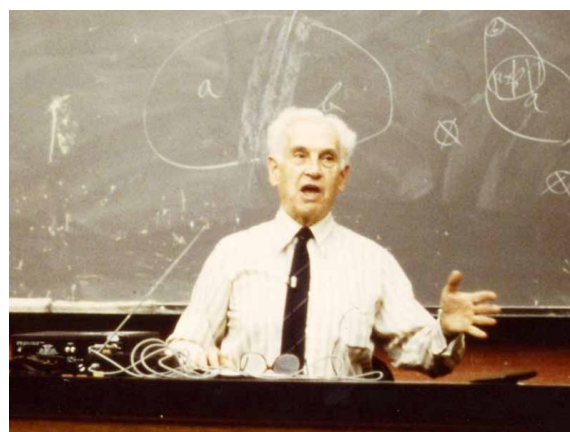


図37 Ernst Mayr (1904 ~ 2005)。1987年に私がBerkeleyにAllan Wilsonを訪ねた際にたまたまそこで講演していた。

第16回



海外研究室だより

楽都ウィーンで進化を学ぶ： グレゴール・メンデル研究所・ Magnus Nordborg 研究室

土松隆志 (グレゴール・メンデル研究所EMBOフェロー)

オーストリアのウィーンに私がポスドクとして異動したのは2012年1月、もう1年以上も前のことである。院生・ポスドクとして都合4年近くもお世話になったスイス・チューリッヒ大学を離れ、2箇所目のポスドク先としてグレゴール・メンデル研究所という植物科学の研究所を選んだ。遺伝学の父であるメンデルの名を冠するだけあって、植物遺伝学を主な専門とする9人のグループリーダーが研究室を構え総勢100人程度の比較的小規模な研究所である。

私は、その9人のうちのひとり、Magnus Nordborg博士の研究室に所属している。Nordborg研の主なテーマは、ゲノムデータを用いた集団遺伝学である。とくに、新しい遺伝子マッピングの手法として注目されているゲノムワイド相関解析(GWAS)を用いた、モデル植物シロイヌナズナにおける野生変異の解析が中心テーマだ。GWASの統計手法自体の開発にも積極的に取り組んでいる上、最近ではシロイヌナズナの1001系統のゲノム配列を決定する国際プロジェクトの一端を担うなど、集団ゲノムに関する多様なテーマに取り組んでいる。研究室メンバーが筆頭著者の論文を2012年だけで3本*Nature Genetics*誌に掲載させるなど、ゲノミ

クス・集団遺伝学分野で勢いのあるグループのひとつである。なお私自身の研究テーマは、シロイヌナズナにおける自家受精(自殖)の進化である。大学院生の頃から長く取り組んでいるテーマなのだが、1001ゲノムやGWASなどのNordborg研のリソースや技術を生かしつつ、自殖に関わる形質の進化を集団ゲノムの観点からより詳細に解明することを目指している。

私のこれまで所属した研究室は、東大の伊藤元己先生、チューリッヒ大学の清水健太郎先生とこれで3つ目になるが、Nordborg研は研究室の文化や雰囲気、研究の進め方などあらゆる点でこれまでのいずれとも大きく異なっていた。それは刺激的で、知的な興奮に満ち、楽しい経験だが、決して楽なことばかりでもなかった。この記事では、そのNordborg研ライフの紹介を中心に、研究所のこと、ウィーンの日常生活も交えて伝えられたらと思う。



図1 スウェーデンの砂浜に生えるシロイヌナズナ。モデル生物とはいえ、私たちの研究対象はこのように野外に生きるシロイヌナズナである。

カジュアルで気ままなNordborg研ライフ

研究室の概要を説明しよう。ボスMagnusのもと、技官が3人、プログラマーが2人、ポスドクが9人、博士課程の院生が6人の計21人と、比較的大所帯だ。この他、数ヶ月～半年程度のインターンやサマースクールが滞在する。国籍は多様で、スウェーデン、オーストラリア、アメリカ、中国等、計12カ国に及ぶ。地元オーストリア人は全21人のうちのわずか3人だ。構成員のバックグラウンドは多様であり、生物情報科学、進化遺伝学、統計遺伝学を中心に、植物生理学、農学、生態学、数理生物学にまでおよぶ。

Nordborg研の第一印象は、ありとあらゆる形式を極限まで排除した超カジュアル研究室、といったところだった。研究室というよりむしろ、にぎやかな動物園。およそこの研究室には、ルールと呼べるものはほとんど存在しない。皆好きな時間に来て、好きな場所、格好、姿勢で研究をする。ソファで寝っ転がってプログラムを書いている人や、机に足を乗っけて完全にリラックスしてプレゼンをする人など、この世には実にいろんな格好で研究をする人がいるものと思った。そして、メンバーたちは好きなように仲間たちと議論して、好きな時間に帰る。

最初に来てまず驚いたことは、研究室セミナーの予定表が存在しないということだった¹。週に1回セミナーがあるのだが、前日に喋りたい人をメール等で募り、喋りたい人が適当に名乗りを上げてプレゼンを行う。特に喋る予定のなかった人であっても、「こんな面白い結果が出たんだけど、どう思う?」とその場の飛び入りでスライドを一枚見せるのももちろんOKだ。これでよく研究室が回るものだと思うのだが、形式に気をかけない分、研究の中身にかかるメンバーたちの情熱は特別大きい。研究室セミナーでは皆本当によく喋り、厳しいコメントを容赦なく発表者に投げかける。所属して早1年強、計6回ほど私はこのセミナーで発表してきたが、まだポジティブなコメントをもらえたことはほとんどない。ここでこれだけ叩かれればどんな論文査読者も怖くないと言い聞かせ、小心者の自分はいつも胃を痛めて発表に臨んでいる。

データ駆動型の発見的生物学

「シロイヌナズナの1,000個体以上のゲノム配列がある。その全個体について遺伝子の発現量やメチル化、表現型のデータがある。野外での適応度もわかっている。GWASなどの統計手法も持っている。あなたはそこからどんなbiologyを見つける?」…ボスMagnusがよく我々に問いかける質問だ。Nordborg研の志向を一言で述べるなら、データ駆動型の発見的生物学である。理論予測やモデルを先験的にはあまり考えず、大量

のデータを眺めてよく考え、様々な統計的な解析を行い、そこから見えてくるbiologyに学ぼうというスタンスだ。

研究室がこういった志向の生物学を進める背景には、Magnusの来歴が関係している。Magnusは、もともと数理生物学者Marcus Feldmanのもとで学位を取得した生粋の理論家だ。ポストドクから集団遺伝学へ移行し、やがてシロイヌナズナの集団ゲノミクスに取り組むようになる。2005年に*PLoS Biology*誌に出版した論文が、彼にとってひとつの転機となったようだ。この論文で彼は、ゲノム上の876遺伝子座の配列を96系統について解読し、その塩基多型を調べた。するとそのパターンは、既存のあらゆる集団遺伝学的モデルにうまく当てはまらなかった。多型のパターンは我々が考えるよりずっと複雑だ。シンプルで一般的なモデルを考えるために我々はまだ知らないことが多すぎる。そう確信した彼はついに「すべてのモデルは死んだ」と発言するに至る。もちろんこれは言い過ぎだが、ひととおり理論を極めたMagnusが言うのとたしかに説得力がある。彼の研究室はその後、実証研究へとさらに大きく舵を切るようになる。

大量のデータを眺めてよく考え、そこから何らかの生物学的知見をあぶり出すという発見的アプローチ。着実な仮説検証型研究と異なるのは、問うべきQuestionが最初から必ずしも定まっているわけではない点だ。やや逆説的だが、問うべきQuestionを探ること自体が重要な研究過程のひとつであり、それが見つかればゴールが見えてくる。このプロセスの鍵になっているのが、さまざまなバックグラウンドを持つ人達との対話と議論であるように思う。Nordborg研におけるメンバー間の共同作業について、次の節で詳しく紹介したい。

大規模プロジェクトをどう進めるか

Nordborg研のこれまでの主要な業績のいくつかは、著者が20人近くも並ぶような大規模プロジェクトだ。筆頭著者が4、5人同等貢献として並ぶ論文も珍しくない。研究室に来る前に私が興味を持っていたことは、こういうプロジェクトが一体どのように進んでいくのかということだった。ボスの圧倒的なリーダーシップか、あるいは筆頭著者たちの優れた協働か。1年間見てきた限りでは、その両方の側面はあるものの、むしろ後者が強いと感じている。ボスは重要な局面での意思決定に非常に優れてはいるものの、普段から「これをやってください」などと指示を出すことはまずなく、基本的に完全放任である。メンバーたちは、研究室にいくつか設置されているホワイトボードを前に日々ワーワーと賑やかに議論を繰り広げる。もちろん個人差はあるものの、基本的にこの研究室では、一人だけでコツコツと研究を進めるというスタイルはあまり見られない。と

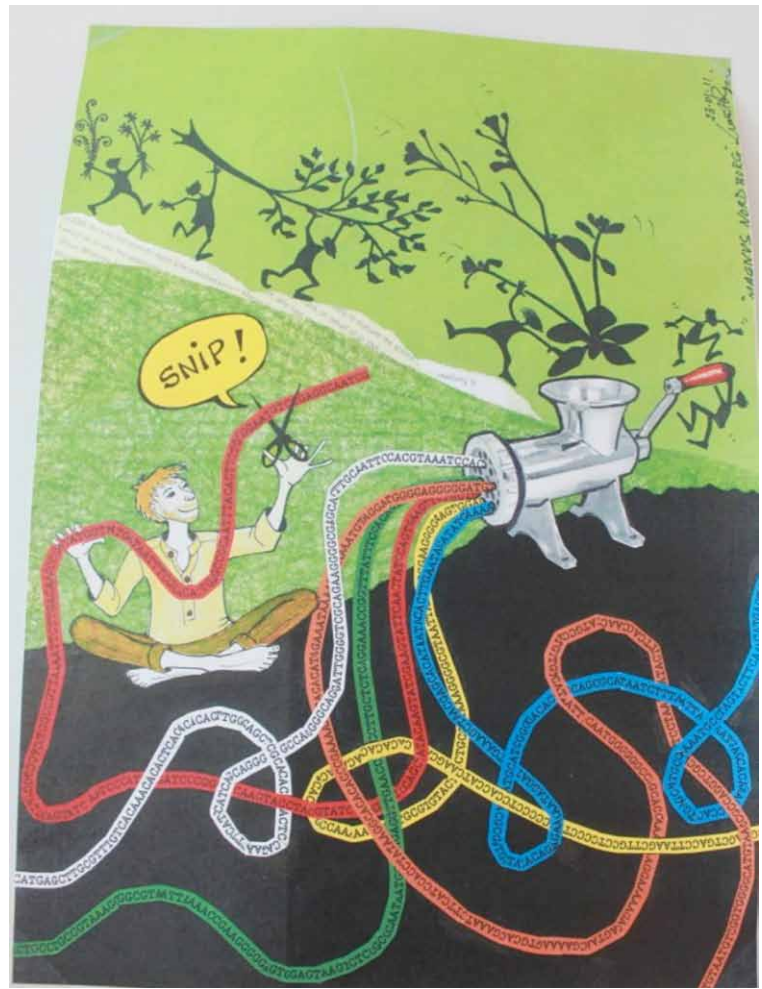


図2 研究所のキッチンに貼られた切り絵、作品名は「Magnus Nordborg」。研究所のメンバーが制作した彼の肖像画だ。手回しの機械に植物を放り込むとそこからDNA配列のテープが生成される。彼は長いテープから一塩基多型 (SNP) を切り出して喜んでいる。

にかく、みんな本当によく喋るのだ。その内実は、時には馬鹿げたアイデアであったり誤りであったりもするのだけれど、とりあえず気にしない。喋って喋って、ディスカッションの中から解析方法を模索し、ストーリーをビルドアップしていく。これが Nordborg 研の標準的な研究スタイルであるようだ。

私自身、この研究室に来て研究の進め方がだいぶ変わってきたと思う。私の研究はどれもそう大規模なものではないのだが、それでも何人かのメンバーたちに助けられて進めている。というのも、次世代シーケンサーの吐き出す膨大なデータの解析や集団遺伝学にはじまり、GWASの統計的方法論の開発、分子遺伝学的なウェットの実験や野外での生態学に至るまで、現在の最先端の進化・生態ゲノミクスに求められる守備範囲は、もはや一人でなんとかできるレベルを越えている。それなら、餅は餅屋である。わからないことはどんどん聞いて回って議論して意見をもらい、必要であれば手伝ってもらえばいい。論文の著者の数は増えていくが、それでマイナスになることは実際ほとんどない。それより大事なことはサイエンスとしての質とスピード。Nordborg 研スタイルの背景にあるのは、こういった信念であるように思う。

研究スタイルに対する賛否はあるだろう。しかし、進化・生態ゲノミクス分野のスタンダードのひとつになりつつあることは確かだ。この研究室で実感したことがある。それは、大規模プロジェクトを主導することそれ自体が、ひとつの重要な研究能力であるということだ。議論を仕切り、リードして合意を得る。多くの人に効率よく作業・解析してもらえよう環境や基盤を整える。そこには、一人でコツコツと研究を進めることとは明らかに別の資質が求められる。Nordborg 研では現在、数万個体のシロイヌナズナを野外で育てる大規模なフィールドプロジェクトが進行中だ。主導するポスドクたちの仕事は、実験のデザインや解析、フィールドワーク、論文執筆はもちろんのこと、関わる人達のスケジュール管理や、プロジェクトが回るための様々な手配・調整など、マネジメント業務が主要な部分になっている。彼らがいつか独立して研究室を持つようになったら運営はお手の物だろうと、傍で見ながら思っている。

ハードルの高かった Nordborg 研スタイル

メンバー間での協働を前提に研究を進めるこのスタイルには、しかしながら、高いコミュニケーション能力と語学力が求められる。もともとあまり喋り好きでもなく、ろくな英語も喋れない自分は、研究室に来た当初かなり苦労した。皆周りを気にせずマイペースにワーワーと喋るので、なかなか議論に割り込めない。まあいいかと黙っていると話はどんどん進んでいき、しまいにはフォローできなくなる。海外にいる方にはある程度わかってくれると思うが、この疎外感はなかなかしんどいものである。もっと話をして周りの人の技術や意見を自分の研究に取り入れなくてはいけないのに、なかなかうまく進められない。ボスも、特に何かしてくれるわけでもない。最初の数ヶ月はほとんど何もできず、何も進まず、話もできず、結構落ち込んでいた。

「Nordborg 研スタイル」に少しずつ慣れ、研究がうまく回るようになってきたと思うのは比較的最近のこと



図3 研究室で議論するメンバーたち。



図4 打ち合わせを兼ねたフィールド遠足にて議論するメンバーたち。

だ。大きな理由は気のもち方の変化だったと思う。1つ目は、思ったことがあれば、何でも口に出して言うべきだと気づいたこと。そして、その場で口に出さない限りは、自分の意見や存在をなんら周りに認識してもらえないと思い知ったこと。自分が思っているほど周りは自分の発言など気にしていない。誤りを恐れるあまり何も言えずにいるのは本当につまらなかったと思う。2つ目は、自分のこれまでやってきたこと、今やっていることに自信を持つこと。たしかに、Nordborg研のポスドクたちは、これまでに優れた論文を出版してきたプライドの高い強者揃いだ。しかし、自分も自殖の進化に関することであれば一家言ある。自分のやってきたこと、できること、知っていることを皆に示すことで、だんだんとメンバーに助言や議論を求められるようになった。頼るばかりでなく頼ってもらえるようになったということは、メンバーとして受け入れられたように感じて、本当に嬉しかったのを覚えている。

クラウドで研究室の情報共有

メンバー間の共同研究での技術的な工夫についても多少触れたい。これまで述べてきた通り、Nordborg研は約20人と比較的大規模で、研究室内外での共同研究が多い。加えて、ボスはしばしば学会等で不在だ。よって、メンバー間で情報をどのように共有するかが、研究室が効率的に回るためのひとつの鍵になっている。このために我々が今活用しているのが、インターネット上でのデータ共有、いわゆるクラウドである。具体的には、Google Driveと呼ばれるGoogle社の提供するオンラインストレージサービスを使っている。Google Drive上のファイルをメンバー間で共有すると、その共同閲覧・編集が可能になる。これを利用して、各プロジェクトの進行ログ、会議の履歴、新しい解析結果、データセットの情報、面白かった論文など、ありとあらゆる情報を可能な限りGoogle Drive上で共有し、全メンバーがアクセス可能にする。各ファイルにはコメントが付けられるので、新しい実験結果にフィードバックを得ることも可能だ（もちろん顔を付き合せて話すのには敵わないのだけれど）。人によっては論文もGoogle Drive上で書き上げてしまう。共著者からのインプットを得るには、たしかに便利なやり方である。論文の査読コメントを共有しておけば、共同研究者同士で分担して改訂するのも役立つ。

背景にある思想は、各人が持っているデータや情報、技術を徹底的にオープンにしてメンバー間で共有するということだ。メンバー間でのフィードバックを考えると、オープンにするメリットはそのデメリットよりも大抵ずっと大きい。もう一点、クラウドを利用する実際的な理由として、E-mail上でのやりとりは多くの人が関わる共同作業にあまり向いていない点が挙げられる。E-mailでの情報交換は後から見返しにくいと感じる方は多いだろう。すべての情報や議論の履歴を、誰でもアクセス可能な形でわかりやすく残すための方法を模索した結果、現在のGoogle Driveでの共有に落ち着きつつある。まだまだ試行錯誤は続いているが、この方法は割とうまくいっているように思う。

子だくさん研究室

Nordborg研のもうひとつの特筆すべき点は、多くのメンバーが現在子育て中であるということだ。託児所が研究所に付設されており、たいいてい皆そこに子供を預けて研究室に出勤する。子供が体調を崩した時もすぐピックアップできるので、助かるだろうなと思う。研究室にゲストが来ると夜食事にしかけるわけだが、そういった時もたいいてい子供たち歓迎、kids-friendlyだ。というわけで、研究室での食事はいつも非常に賑やか、というよりしばしばカオスの状況を呈する。夫婦ともに研究者の場合、もし食事がkids-friendlyでないと、どちらかが家にはいなくてはいけない。食事が子供たち歓迎なのはこういった配慮もあるからだろう。ボスのMagnus自身が共働きで子育てをしていることもあり、子供に関することには全体的に非常に理解がある。研究室の行事や日程調整はいつも皆の子供の都合が優先だ。

研究所内外との盛んな交流

研究所のことについても紹介したい。冒頭でも少し述べたが、私の所属するグレゴール・メンデル研究所

は、植物の遺伝学・分子生物学の基礎研究を行うための研究所である。設立から13年、形式上「会社」という形の研究所だが、オーストリア科学アカデミーという国の組織に属しており、実質的に国立の研究所だ。

この研究所で驚いたことは、ポストドクはもちろん、グループリーダーも含めて科学者はみな任期付きである一方、技官やバイオインフォマティシャン、ラボマネージャーなど、研究をサポートする人達は任期無しのパーマネント契約であることが多いということだ。技官は学位を持っていることが多

く、それぞれが専門性の高い技術を持っている。前にいたスイスでもこういった職の方たちがパーマネントである例は多かったので、オーストリアだけでなく、ヨーロッパではある程度普遍的なシステムなのかもしれない。高い技術をもつ人材を研究所として囲い込める点で優れた仕組みだと思う。また、技官やラボマネージャーがポストドク後の一般的なキャリアとして定着しているのも、特筆すべき点だ。

グレゴール・メンデル研究所は、Vienna Bio Center (VBC) という研究所群の一角にある (図5)。VBCはウィーン中心部からトラムで10分ほどの便利な場所にあり、グレゴール・メンデル研究所以外にも、分子生物工学研究所 (IMBA)、ウィーン分子病理学研究所 (IMP) などの基礎研究を行う研究所が並んでいる。各研究所の建物は内部で繋がっており、広いカフェテリアを共有している。研究所間の交流は盛んで、各種セミナーのほか、各研究室が持ち回りで準備する毎週金曜日のBeer Hourや、クリスマスパーティー、演劇、オーケストラ、スキー旅行、サッカー大会など様々な催しが企画される。

VBC自体は、いわゆる分子生物学系の研究室が中心で、進化生物学の研究者はあまり多くない。しかし、ウィーン全体を見渡すと、Adaptive Dynamics理論の先導者Ulf Dieckmann、進化ゲーム理論の教科書の著者としても知られるKarl Sigmund、集団遺伝学分野ではNick Bartonなど名だたる研究室がいくつもあり、欧州における進化生物学の一大拠点都市だ。街中に研究所・大学が散在していて研究室間の行き来がやや面倒ではあるものの、交流は盛んである。例えば私たちの研究室では、ウィーン獣医科大学のChristian Schlötterer研究室と合同セミナーを不定期に開催しているほか、論文紹介セミナーはマックスペルーツ研究所のJoachim Hermisson研究室と合同で行なっている。ウィーンの集団遺伝学の研究室が集まってできたPhDプログラムもあり²、大学院生間の交流もかなりあるようだ。

外国人が住み良い街ウィーン

ウィーンはドイツ語圏だが、研究所の公用語は英語だ。よって、研究には一切ドイツ語は必要ない。日常生活でドイツ語が必要かということ、多少でも話せるとだいぶ快適だろうと思う。ありがたいことに、研究所内には無料のドイツ語講座があるのだが、無精な自分はまだ受講していない。妻が仕事の関係上ドイツ語を勉強しているので、私は彼女のおかげで門前の小僧的に覚えた怪しげなドイツ語をつぶやく程度である。まあ、話せなくてもなんとかなるのが国際都市ウィーンのいいところなのである。

ウィーンは治安や交通の便がよく、外国人にも親切なとても暮らしやすい街である。「住み良い街ランキング」ではいつも世界最上位だ。短期滞在者も含めると在留邦人はそれなりに多い。研究者のほか、観光業、国連や政府関係機関の職員、音楽関係者など、日本に住んでいてはなかなか知り合えない方と簡単に仲良くなれるのは海外に住んでいて楽しいことのひとつだ。オーストリア日本人会の主催する年2回のソフトボール



図5 グレゴール・メンデル研究所の外観。分子生物工学研究所 (IMBA) と建物を共有している。

大会は、ウィーンに住む日本人たちの楽しみのひとつである。国連チーム、日本人学校チーム、大使館チームなど、100人以上の日本人が集って練習の成果を競い合う。私も研究所の同僚たちと参加しているのだが、如何せん皆強すぎてぜんぜん勝てない(日本人学校のエースは元高校球児だ。これはずい)。ウィーンを離れるまでに一度くらい優勝したいものである。

ところで、ウィーンというとケーキやカフェが有名かもしれないが、私はワインをお勧めしたい。ウィーン郊外の丘陵地(いわゆるウィーンの森)には、ホイリゲと呼ばれる小さなワイン酒場がたくさんある。夏の夕べは自家製ワインを求めて多くの地元民で賑わう。暑い夏の盛りに、ウィーン市街を望むテラスに腰を掛け、葡萄棚の日陰で飲むよく冷えた辛口の白ワインは、もう言葉にできない。酒好きには最高の街である(図6)。

以上、しんどいこともあるけれど、楽しいことや嬉しいこともそれ以上にたくさんあるウィーンでのポストドク生活を紹介してきた。スイス、オーストリアと、私の海外生活は計6年目で、もはやこちらでの研究キャリアのほうが長くなってしまった。いつか日本に帰れる日が来るのかはわからないけれども、世界中どこであっても、楽しく研究できるところで自分のやりたいことをマイペースに続けられたら、それが一番だと思っている。

欧州圏で研究をしたいと考えていらっしゃる若い研究者の方など、私でよければご相談に乗りますので、お気軽に以下までご連絡ください(takashi.tsuchimatsu@gmi.oeaw.ac.at)。

グレゴール・メンデル研究所ウェブサイト <http://www.gmi.oeaw.ac.at/>

謝辞：佐々木江理子博士(グレゴール・メンデル研究所)には本稿に貴重な意見を頂いた。篤くお礼申し上げます。

- 1) ただ、これではあまりに発表しない人がいるということで、最近少しずつ状況は変わりつつある。
- 2) <http://www.popgen-vienna.at/>



図6 ウィーン郊外の丘陵地から市街を望む。目の前には葡萄畑が広がる。

MAFFT 多重配列アラインメントプログラムの紹介

加藤和貴(大阪大学 IReC、産総研 CBRC)

塩基配列やアミノ酸配列の比較解析において、多重配列アラインメントは基本的な操作である。その計算手法の研究には、過去数十年にわたって多くの研究者が取り組んできた。今後も配列解析が行われる限り、アラインメントはその基本であり続けると予想される。筆者らは、2002年にMAFFT 多重配列アラインメントプログラム^[1]をリリースし、以後継続的に改良を行ってきた。その性能は、第三者による多くのベンチマークにおいて高評価を得、最近では、MAFFTプログラムはClustalシリーズ^[2,3]などと並んで多重配列アライン

メントのための標準的なツールの一つと見なされるようになった。Pfamやprositeなどバイオインフォマティクスにおける主要なデータベースの構築にもMAFFTは使われている。

多重配列アラインメントが使われる研究の範囲は広いが、中でも進化解析は重要な適用分野であると考えている。そのためMAFFT最新版に関する論文は*Molecular Biology and Evolution*誌で公表した^[4]。この号では、筆者らの提供したイラスト(大阪大学Amada Karlou Mar Santander氏による)が表紙に選ばれた。本稿では、MAFFTプログラムの基本的な使用法、および、分子進化的解析において有用と思われる、既存のアラインメントに新しい配列を加える機能について説明する。RNAやタンパク質の構造を考慮したアラインメント、大規模なアラインメントなど、特殊な使い方については述べない。

基本的な使用法

MAFFTはUNIX的なコマンドラインプログラムであり、LinuxとMacintosh上で動く。Windows用のバイナリも用意しているが、機能は制限される。いずれのパッケージも、産総研生命情報工学研究センター(CBRC)のウェブサイトからダウンロードできる。

<http://mafft.cbrc.jp/alignment/software/>

インストール後、multi FASTA形式で入力配列を与え、

```
mafft --auto input > output
```

のようにコマンドを実行すると、multi FASTA形式でアラインメントが返る。出力形式の変更やその他のオプションについては、ウェブページをご覧ください。

GUIが必要な場合は、いくつかの研究機関によるMAFFTサービスが利用できる。

<http://mafft.cbrc.jp/alignment/server/>

<http://www.genome.jp/tools/mafft/>

<http://www.ebi.ac.uk/Tools/msa/mafft/>

<http://toolkit.tuebingen.mpg.de/mafft/>

<http://myhits.isb-sib.ch/cgi-bin/mafft/>

<http://mobylye.pasteur.fr/cgi-bin/portal.py#forms::mafft>

産総研CBRCのサービスのスクリーンショットを図1に示す。筆者の管理しているこのサーバではだいたい最新版が動いているが、他のサーバが追隨しているとは限らない。また、Javaで書かれたアラインメントエディタJalview^[5]の中でMAFFTアラインメントを実行することもできる。

<http://www.jalview.org/>

なお、MEGAプログラム^[6]からMAFFTを利用可能にすることは、Sudhir Kumar博士らにより計画されているが、Windowsの制約があり進んでいない。

前提

MAFFTは、配列上のすべての文字の順番が保存されてきたと仮定して計算を行う。したがってゲノム配列の再編成には対応していない。長い配列に対しては、FFTアルゴリズムによって類似性の高い領域を高速に発見し、その後、その間の領域をダイナミックプログラミング(DP)でつなぐという戦略をとる。類似性の高い領域が存在するけれども出現順序が配列によって異なる場合は、このような方法はまったく有効ではない。ウェブ版では、塩基配列が入力された場合、LAST^[8]によるドットプロットを表示する(図1C左)。LASTとは、産総研CBRCのMartin C. Frith博士によって開発されている高性能なローカルアラインメントプログラムである。このドットプロットを見ることによって、ゲノム再編成などにすぐ気づくようになっていく。そのような問題に対しては、ローカルアラインメントによってそれを考慮した方法を適用する必要がある。

MAFFTはすべての配列がホモログスであることを仮定する。仮に、進化的に無関係な多数の配列を入力

すると、たくさんのギャップが導入され、計算に長い時間がかかる。結果は返るかもしれないが、意味のある結果ではない。アラインメントに含める配列の選択は、基本的には、相同性検索プログラムとユーザの手作業に任せるしかない。近年、配列データベースが膨張していることと、信頼性の低い配列を含むようになってきたことにより、この段階の処理が難しい問題になってきている。

この問題にある程度対応するために、ウェブ版は、対話的な配列選択機能を持っている(図1C右)。MaxAlign^[9]やCD-HIT^[10]による自動的な配列の選択、アラインメント、近隣結合法^[11]による系統樹推定、Archaeopteryx^[7]上の配列の手動選択、といった操作が循環的にできる。さらにこの機能は、aLeaves^[12]というツールとコーディネートされている。

<http://aleaves.cdb.riken.jp/>

これは、動物のタンパク質コード遺伝子の系統樹推定を行うための配列データを、ウェブ上に散在しているゲノムデータベースから、できるだけ過不足なく収集するためのツールである。aLeavesについては、開発者である工樂樹洋博士による紹介記事が進化ニュースの次号に掲載される予定である。将来、動物だけでなく、植物、真核生物全体、真正細菌、古細菌といった、いろいろなレベルのグループについて、同様のサービスを構築したいと考えている。そのためには、それぞれのグループに特異的な生物学的知識を持った共同研究者が不可欠である。もし、ご協力いただける方がいらっしゃいましたら、ぜひご連絡ください。

配列の追加機能

既存のアラインメントを保持しながら、新しい配列を付け加えたいことがある。たとえば、手作業によるアラインメントや、構造情報などを用いて信頼性が高いと考えられるアラインメントがあり、それにホモログな新しい配列を決定した時、このような必要が生じる。この時、既存のアラインメントを一つのプロファ

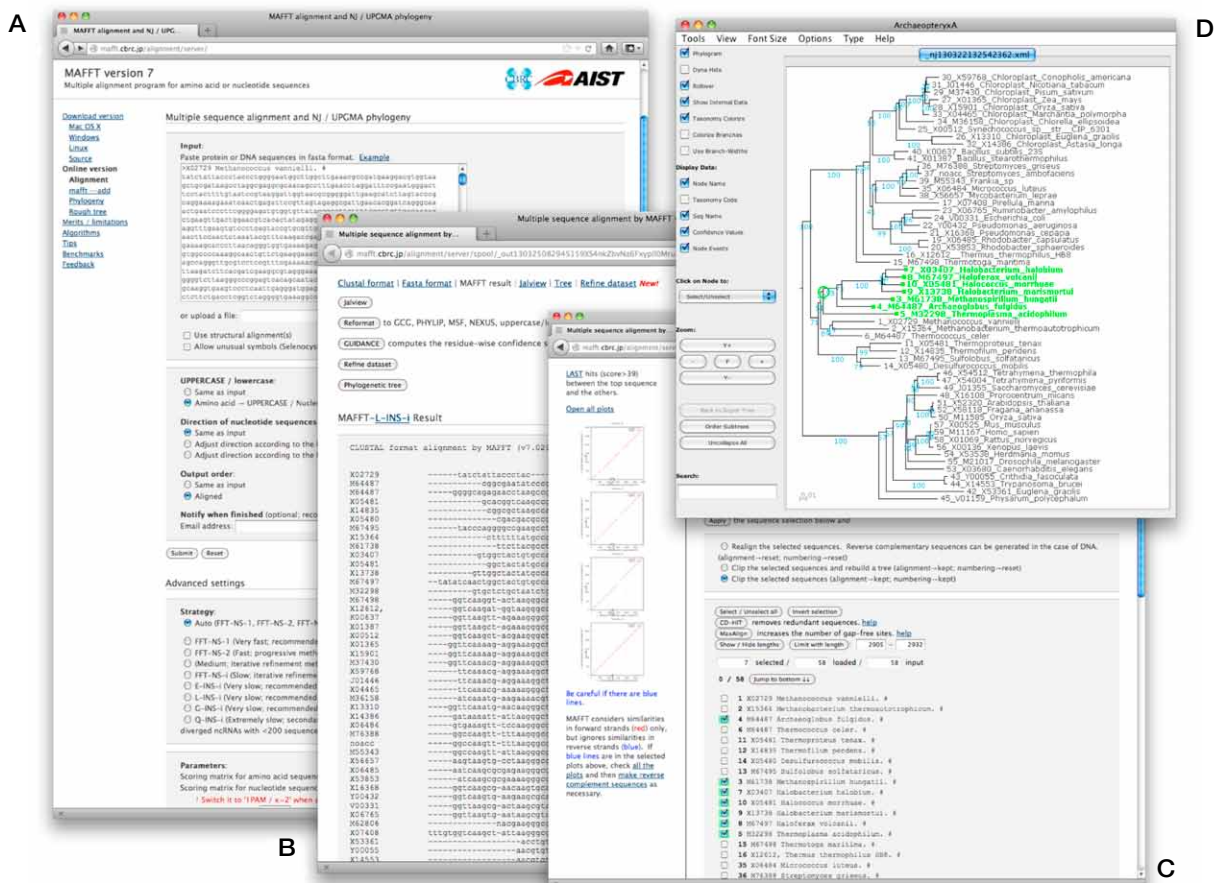


図1 産総研CBRCのMAFFTウェブサービスのスクリーンショット。A. 入力画面；B. 計算結果；C. 配列選択画面；D. Archaeopteryx^[7]を用いたNJ系統樹表示

イルに変換し、それと新しい配列のあいだのアラインメントを計算するのは、いうまでもなく不適切である。**図2A**のような系統関係を仮定することになるからである。その代わりに、**図2B**のように個々の新しい配列と既存のアラインメントに含まれる配列の間の系統関係を考慮する必要がある。この目的のためのツールは、PaPaRa^[13]、PAGAN^[14]など、最近いくつか続けて開発された。MAFFTの--add, --addfragmentsという二つの新しいオプションも、この目的に使うことができる^[15]。

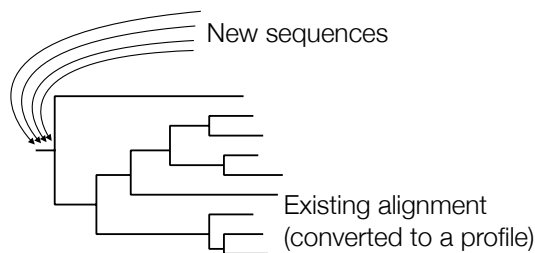
--addオプションのアルゴリズムは特別なものではなく、通常の累進法アラインメントとほぼ同じである。違いは、累進法の各段階で比較するグループの両者が既存のアラインメントに含まれている場合はその計算を省略し、既存のアラインメントを採用するという点だけである。

--addfragmentsオプションはさらに単純で、個々の新しい配列を--addで別々に処理して、最後に一つのアラインメントにまとめる。これは、新しい配列が元のアラインメントに比べてとても短い場合を想定しているが、そうでもない場合にも使えるようである。以上二つのオプションのそれぞれについて、DPに基づいて比較的厳密に距離を推定するオプション(--multipair)と、共有6merの数に基づく近似的な距離を用いる高速なオプション(--6merpair)が選択できる。

これらのオプションの性能をテストした結果を**表1**に示す。--addfragmentsオプションは、高い効率で並列化できるので、多くのコアを持っている計算機の上では、1万本程度の配列からなる既存のアラインメントに、10万本程度の新しい配列を加えるといった、比較的大きな問題にも適用できる。ただし、これは近似的な距離を用いた場合(--6merpair)であり、より厳密な方法(--multipair)をこのような大きな問題に適用することはできない。配列の類似性が高い場合には距離の計算の厳密性はアラインメントの結果の正確さにはほとんど影響しないが、配列の類似性が低い場合、影響する^[15]。厳密な方法のスケラビリティの向上は、今後の課題である。

また、既存のアラインメントに新しい配列を加えることは、低品質の配列データの処理にも役立つことを期待している。例えば、ある生物種の一つの遺伝子ファミリーについて、データの質は低いと生物学的に興味

A. Misuse of mafft-profile



B. --add, --addfragments

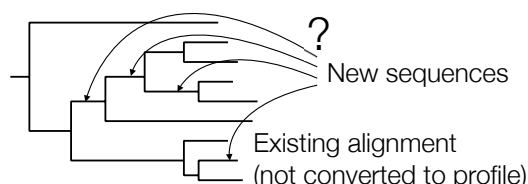


図2 既存のアラインメントに配列を追加する時に仮定する系統関係。

表1 Comparison of different options using the 16S.B.ALL dataset^[16]

Data	Method	Accuracy	CPU time	Actual time ‡
Case 1* mafft --multipair --addfragments frags existingmsa		0.9969	6.67 days	18.3 hours
mafft --6merpair --addfragments frags existingmsa		0.9949	3.76 hours	36.2 minutes
mafft --localpair --add frags existingmsa		0.9707	23.4 days †	2.43 days †
mafft --6merpair --add frags existingmsa		0.9604	1.32 hours	1.44 hours
profile alignment		0.2779	15.5 hours	1.60 hours
Case 2* mafft --6merpair --addfragments frags existingmsa		0.9969	4.54 hours	33.8 minutes
Case 3* mafft --6merpair --addfragments frags existingmsa		0.9949	1.79 days	5.91 hours

16S rRNAの実際のデータに基づくベンチマークの結果。データのサイズは以下の通り。

* Case 1: 13,822 sequences in the existing alignment × 13,821 fragments;

* Case 2: 1,000 sequences in the existing alignment × 138,210 fragments;

* Case 3: 13,822 sequences in the existing alignment × 138,210 fragments;

詳細は、文献^[4]参照。

深い配列が多数あり、既知の配列との比較解析を行いたいとする。それらの配列全部まとめて多重アライメントを計算しようとしても、データのエラーにより収拾がつかなくなることが多い。それに対して、まず、適切にアノートされた高品質のデータから多重配列アライメントを計算し、それに、質の低いデータをまとめて、または一つずつ、加えることによって、エラーの影響を抑えながらそれらを比較することができる。

■フィードバックの重要性について

MAFFTプログラムの開発は、筆者が京都大学宮田隆教授の研究室の大学院生であった時に、三澤計治博士、隈啓一博士らの指導を受けて開始した。実際に配列解析を行っている研究者とのコミュニケーションを通して現実のデータを扱う難しさを知ることができ、それが開発に役立った。このことから、ユーザのフィードバックはきわめて重要な情報源であると考えている。現在は、世界各地から寄せられるフィードバックに助けられている。ユーザの地理的分布は、ウェブサービスのアクセス解析からおおまかに推定すると、ヨーロッパ38%、南北アメリカ33%、アジア23%、その他6%である。

最後に、国内的状況について述べる。MAFFTユーザ全体の中で日本のユーザの占める割合は9%程度と推定され、国別に見れば米国(22%程度)に次いで2番目に多いはずである。しかし、日本からのフィードバックは、米国はもとより、ドイツ、英国、インドなどに比べても少ない。このような日本のユーザのおとなしさが、日本を拠点とした研究ツール開発を難しくしている一因であると感じる。日本のバイオインフォマティクス分野の研究者は多くの有望な計算手法や理論を考案している。それらを研究コミュニティに普及させるために、開発者側の努力に加えてユーザ側の協力、つまりフィードバックを通してツールを育てることを期待したい。新しい方法を現実のデータに適用した時に生じる問題を開発者が知ることが、その方法の改良につながる。その結果として、広い範囲の研究を支えるインフラストラクチャ的ツールが、日本からも多く発信されることを願う。

引用文献

- [1] Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.*, **30**, 3059-3066.
- [2] Higgins, D. G. and Sharp, P. M. (1988) CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene*, **73**, 237-244.
- [3] Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J. D., and Higgins, D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*, **7**, 539.
- [4] Katoh, K. and Standley, D. M. (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol*, **30**, 772-780.
- [5] Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., and Barton, G. J. (2009) Jalview version 2 – a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, **25**, 1189-1191.
- [6] Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*, **28**, 2731-2739.
- [7] Zmasek, C. M. and Eddy, S. R. (2001) ATV: display and manipulation of annotated phylogenetic trees. *Bioinformatics*, **17**, 383-384.
- [8] Kielbasa, S. M., Wan, R., Sato, K., Horton, P., and Frith, M. C. (2011) Adaptive seeds tame genomic sequence comparison. *Genome Res*, **21**, 487-493.
- [9] Gouveia-Oliveira, R., Sackett, P. W., and Pedersen, A. G. (2007) MaxAlign: maximizing usable data in an alignment. *BMC Bioinformatics*, **8**, 312.
- [10] Li, W. and Godzik, A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, **22**, 1658-1659.
- [11] Saitou, N. and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*, **4**, 406-425.
- [12] Kuraku, S., Zmasek, C. M., Nishimura, O., and Katoh, K. (2013) aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. *Nucleic Acids Res*, **41**: W22-W28 (doi:10.1093/nar/gkt389).

- [13] Berger, S. A. and Stamatakis, A. (2011) Aligning short reads to reference alignments and trees. *Bioinformatics*, **27**, 2068-2075.
- [14] Löytynoja, A., Vilella, A. J., and Goldman, N. (2012) Accurate extension of multiple sequence alignments using a phylogeny-aware graph algorithm. *Bioinformatics*, **28**, 1684-1691.
- [15] Katoh, K. and Frith, M. C. (2012) Adding unaligned sequences into an existing alignment using MAFFT and LAST. *Bioinformatics*, **28**, 3144-3146.
- [16] Mirarab, S., Nguyen, N., and Warnow, T. (2012) SEPP: SATé-Enabled phylogenetic placement. *Pac Symp Biocomput*, **17**, 247-258.



ソフトウェア紹介

raxmlGUI 1.3 : RAxMLを 快適なユーザインタフェースで使おう

宮 正樹 (千葉県立中央博物館)

はじめに

最尤樹推定ソフトである RAxML^[1] の登場により分子系統学の世界は大きく変わった。それまで考えられなかったような多数の長い塩基配列を (たとえば数十 kb に及ぶ数百本の配列を)、現実的な時間内 (たとえば数日) で解析できるようになったのである。私事になるが、魚類ミトコンドリアゲノムの 10 kb ほどの塩基配列を 100 本解析した論文^[2] を発表したのが 10 年前の 2003 年。当時、この規模のデータセットで解析可能だったのは、形質ベースでは最節約法のみ。ところが、今や 100 本の配列なんて小さな数字に見えてくるくらいソフトとハードが進歩した。その急激な進歩は、最尤法のパイオニアの一人である長谷川政美先生が「少し前までは到底考えられなかった」と本号の記事で述べられているくらいのすさまじいもの。RAxML はそれだけ高い性能をもったソフトで、今や分子系統解析には欠かせないツールの一つになっている。

ただ、RAxML で残念なのは実質的にユーザインタフェースというものがなく、すべてをコマンドラインで操作しなくてはならないこと。代替で用意されているウェブ版の動作は不安定だし、最新版が常に使えるわけではない。インフォマティクスの高度な専門家である作者の Alexandros Stamatakis にしてみれば、ユーザフレンドリーなんてソフトの性能とまったく関係ないのだからどうでもよいことなのかもしれないが、ある程度 RAxML を使い慣れている私でさえ、間隔をあけて使うと戸惑うことが多かったのも事実。実際、今回紹介する raxmlGUI が出るまでは、よく使うコマンドラインをテキストファイルで用意しておき、それを使い回していた。ユーザインタフェースの向上は (ある意味) RAxML のさらなる普及にとっては必須の事項であった。

このような背景の下に登場したのが、raxmlGUI^[3] である。Python というプログラミング言語をつかって書かれたこのソフトは、RAxML を快適なユーザインタフェースで走らせることができるだけでなく、さらに便利な機能をいくつか備えている。私もこのソフトを使い出してからは、コマンドラインの RAxML とはおさらばできた。ウェブ版の不安定な動作と限定的な機能しか使えない不便さを考えれば、もっと多くの人がこのソフトを使って良いと思ひ、今回の紹介文を書いてみた。以下の説明は私が使っている環境下 (Mac OS 10.7.5) でのものとなるが、Windows でも Linux でも問題なく動作するとのことを知っている。

インストールと起動

ソフトは <https://sites.google.com/site/raxmlgui/> から sourceforge 経由で入手できる。昨年の 12 月にバージョンアップされてから 2 か月で、わずか 63 回しかダウンロードされていないところを見ると、依然として知名度が低いことがわかる。ダウンロードして解凍すると、ソフトが OS を自動的に判断してくれて、そこ

から先は指示に従ってインストールすれば良い。ちなみに、動作が確認されているのはMacではIntel 64bitマシンでOSは10.6か10.7、Windowsの場合はXPか7、そしてLinuxはUbuntu。いずれもPython 2.5～2.7の環境下で動作するが、Python 3以上の場合は動作しないので注意が必要。実装されているRAxMLは7.4.2となっており、現在利用できるものの中では一番新しい。

raxmlGUIのアイコンをダブルクリックすると図1のような画面が現れる。Macの場合はPythonの環境を利用するためにターミナルが自動的に立ち上がる。

データの読み込みと各種の設定

raxmlGUIが読み込めるのは、RAxMLと同じでPhylip拡張形式 (OTUの名前が10文字を超えても可)。ただし、DendroPy^[4]という同じくPythonで書かれたライブラリーを読み込むことによりNexus形式のファイルをPhylip形式に変換してくれる (詳細についてはマニュアルを読んでほしい)。

データを読み込むには、画面左上の「Load alignment」をクリックし、ポップアップメニューが現れたらそこからファイルを選択する。データの形式が正しくなかったり、OTUの数や塩基数が実際のものとは違ったりするとデータを読み込めない。

なお、今回の例では塩基配列のみを扱うが、アミノ酸も形態形質データも、あるいはそれらが混ざったものであれ、形質状態の変化に対してモデルを設定できれば何でも扱えるのはRAxMLと同様である。Phylip形式ではデータの種類までは指定できないので、メニューバーから「File」を選び、そこから「Set Data Type」を選ぶことになる。

外群の設定

実際にアラインメント済みのデータを読み込んだところが図2になる。正しくデータを読み込むと、外群設定のボタン「Outgroup」が現れる。ルーティングを1種で行う場合にはこのボタンを使って外群を指定するのだが、複数の場合にはメニューバーの「Analysis」から「Select multiple outgroup」を設定しなければならない (図3)。

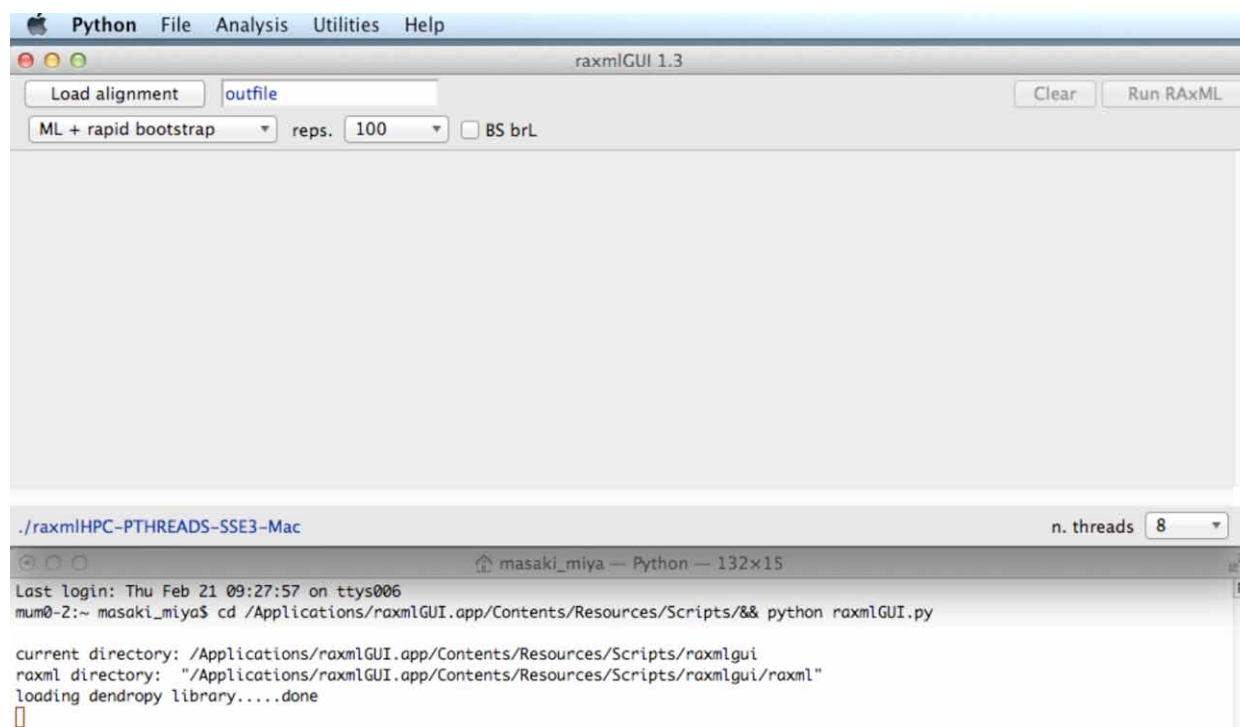


図1 raxmlGUI 1.3のインターフェース。アラインメント済みの塩基配列を読み込むまでは、このように大変素っ気ない画面。下に見えるのはUNIX環境のウィンドーとなる「ターミナル」の画面

解析法の設定

次に解析法の設定を行う。二段目左端にあるボタンがそれ。このボタンをクリックすると、**図4**のように7種類の解析法を選ぶことができる。それぞれの解析法の詳細についてはマニュアルを参照してほしいが、通常の解析であれば最尤推定とブートストラップ解析を同時に行う「ML + rapid bootstrap」でリーズナブルな結果が得られるだろう（あくまでも経験レベルの話）。ブートストラップサンプルの各回ごとにより網羅的な樹形探索を行いたいのであれば、「ML + thorough bootstrap」を選ぶべきかもしれないが、こちらのアルゴリズムを使ったからといってブートストラップ値が大きく変化することはないようだ。「Bootstrap + consensus」は、最尤樹ではなく合意樹上にブートストラップ値をプロットするもので、これがあると最尤樹が奇妙なトポロジーになってしまった場合にチェックができる。

他にも、最尤樹の探索のみを行う「Fast tree search」と「ML search」が用意されているが、ブートストラップ樹を利用した上記アルゴリズムの能力が非常に高いため、個人的にはつかったことがない（したがってこれらを使った場合に出てくるサブメニューについても知識がない）。さらに、祖先配列の推定 (ancestral states) や最節約樹（あるいはユーザが指定した系統樹）に基づく最尤距離の計算 (pairwise distance) を行うこともできる。

ブートストラップに基づく解析を行う場合には、その回数を指定しなくてはならない。最近の複数のCPUをもつコンピュータの場合は、multithreadsに特化したRAxMLを使えるので、1,000回くらいのブートストラップサンプリングを行っても大した時間はかからないし、回数を増やせばそれだけ真の最尤樹に近づく確率は増す。私の場合、どんな大きさのデータセットでも必ずブートストラップサンプリングを1,000回はやるようにしている。

塩基置換モデルの設定

RAxMLの作者の信念により、選択できるモデルはGTR (general time reversible)^[5]をベースにしたも

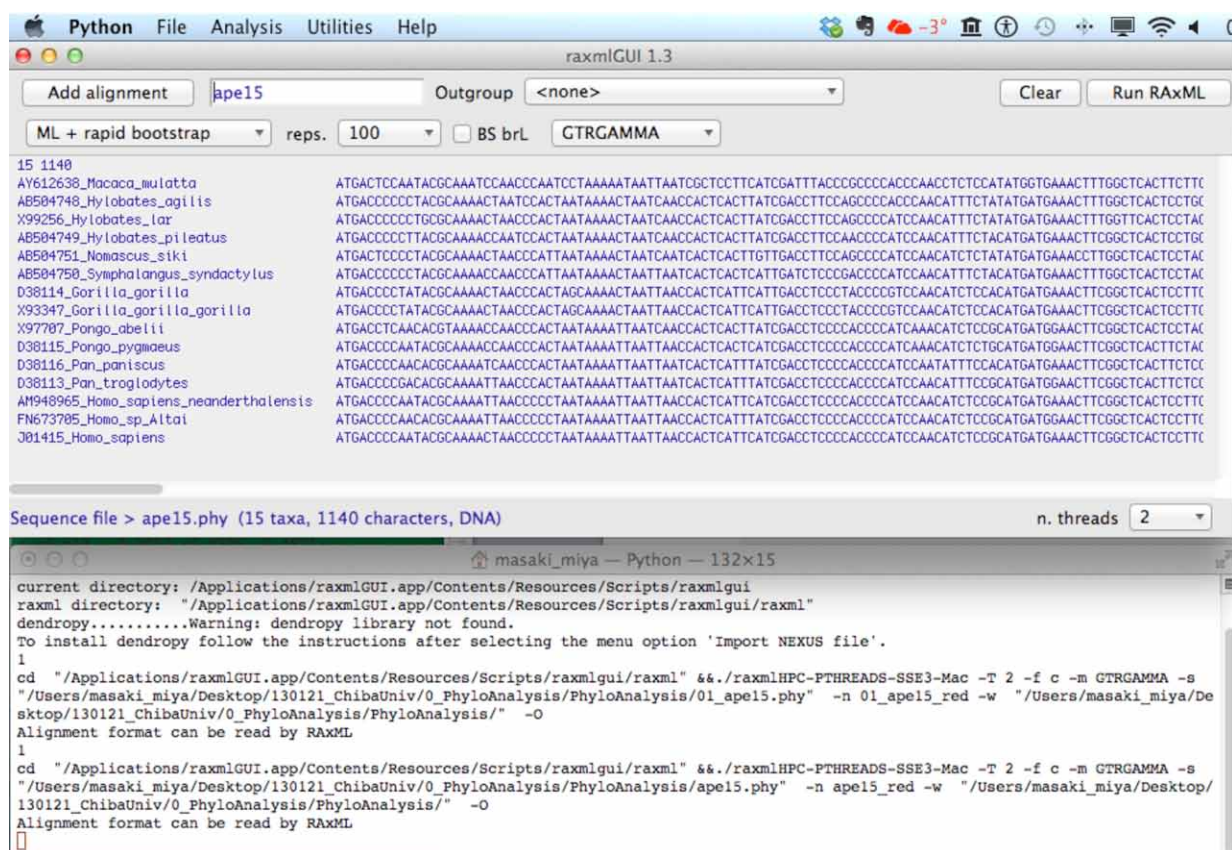


図2 「Load alignment」ボタンを押してデータを読み込んだところ。ログがデータを読み込んだ画面の下に残される

のみである。したがって、モデル選択ソフトでそれよりシンプルなもの指定されても対処できないのは RAxMLと同じ。昨今の分子系統学に於けるデータの大きさを考えると実用的にはまったく問題はないと思われるし、実際にパラメータ過剰で奇妙な推定になったという経験も(少なくとも個人的には)ない。また、不変サイトの補正パラメータ (I) を使うべきではないというのが RAxML の作者の立場なのだが (RAxML のマニュアルの 20 ページにある FAQ を参照)、一応使えるように用意されている。なお、巨大データのために GTRCAT という選択肢も用意されているが、これはよほどデータが大きくない限り無視しても構わないだろう

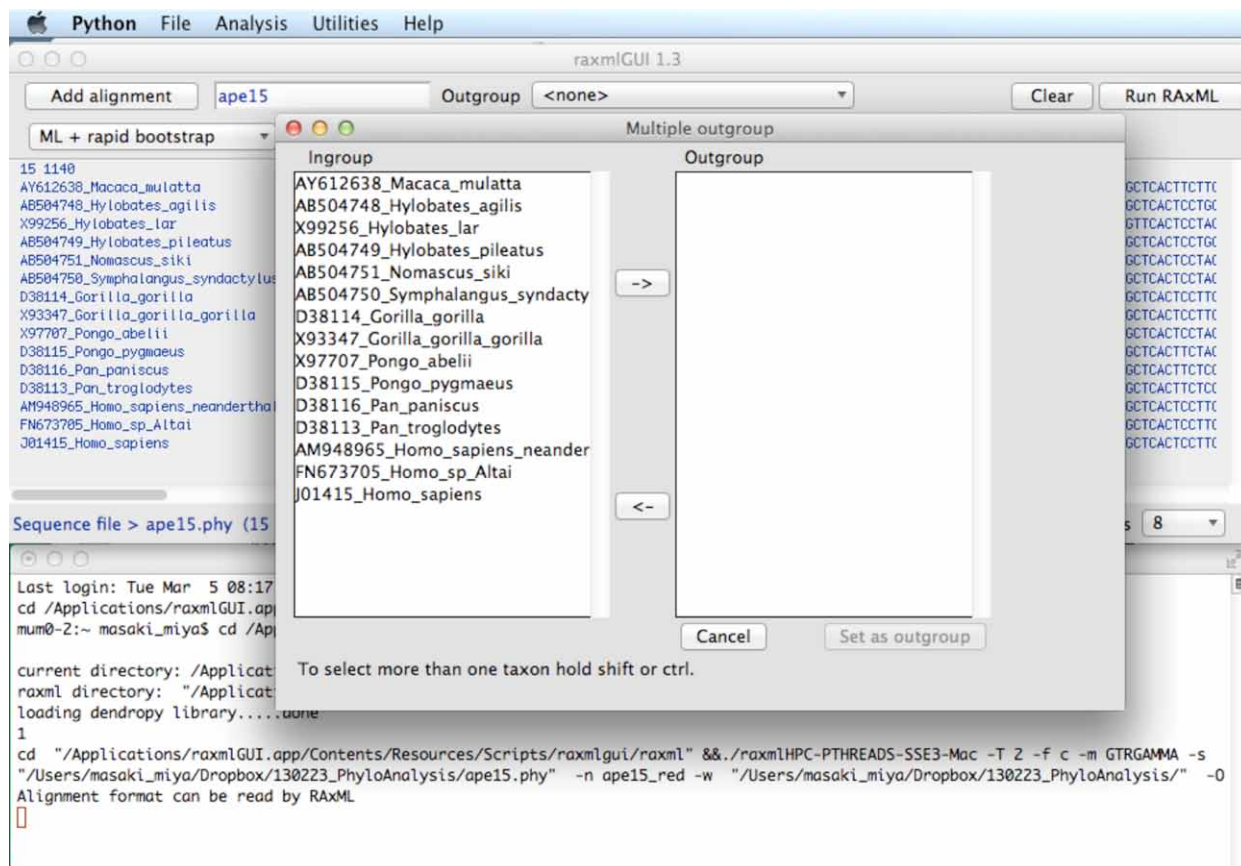


図3 複数の外群を設定するときには、メニューバーの「Analysis」から「Select multiple outgroup」を設定しなければならない

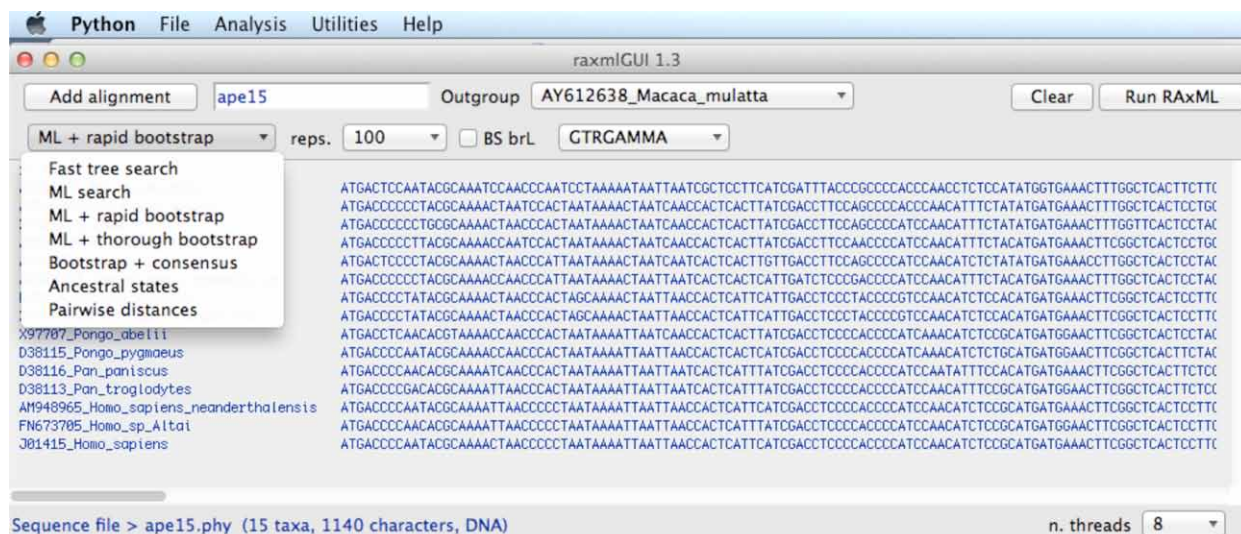


図4 解析方法を選択するところ。デフォルトは ML + rapid bootstrap になっている。通常の解析であればこれで問題ない

う。結局、私はRAxMLの作者のお薦めに従って、塩基配列の場合には常にGTR + gammaを使って解析を進めている。

パーティションの設定

ここまで来るともう解析を始めても構わないのだが、実は大事なことが一つ抜けている。塩基配列がタンパク質をコードしているものの場合、コドンによって変異のパターンが大きく異なることがよく知られている。このような場合に、コドンごとにパーティションを区切ると、それぞれに置換モデルのパラメータを与えられるのだから、より現実に即した推定ができるはずである。

さらに、核遺伝子を複数連結した解析も最近では当然のように行われるようになってきた。このような場合には、遺伝子ごとに変異のパターンや速度が大きく異なる。コドンに加えてさらに遺伝子ごとにパーティションを区切った方が良いことも最近の研究で示された^[6]。

raxmlGUIではパーティションの設定エディターが用意されており、簡単にパーティションファイルを作成することができる。今回、例として使ったデータは霊長類のミトコンドリアのcyt b遺伝子計1,140 bp。これをコドンごとに簡単にパーティションに分けてみよう。

まずは「Analysis」から「Set/Edit partitions」を選択する。すると図5のようなエディターが現れる。データタイプはDNA、全長は1,140 bpだから、「to」のところに1140と記入し、「model」はcodon specificを選択する。次に左下の「Add」のボタンをクリックしていくつかの問いに答えると、図6のような画面になりコドンごとのパーティションができる。右の「Options」ボタンをクリックすると、この設定をエクスポートしたり、これをさらに編集したりできる。実際には右下の「Set」を押すことにより、このパーティションが実際にデータに対して設定される。

もちろん、外部からのすでに作成済みのパーティションファイルを読み込むことも可能。私の場合はrRNA/tRNAを含むパーティションも設定するが多いため、実際にはこのようにしてraxmlGUIを使っている。

解析の開始

解析の開始は右上にある「Run RAxML」のボタンを押すだけである。ただし、CPUを複数備えたコンピュータを使う場合には、必ず右下にある「n. threads」を確認すること。raxmlGUIは自動的にCPUの数を判断し、このボタンを押すと何個のCPUが使用可能かわかる。系統解析専用機であればすべてのCPUを使えば最速で解析を行えるが、そうでない場合は一つか二つ余らせておいて、バックグラウンドで解析を行う

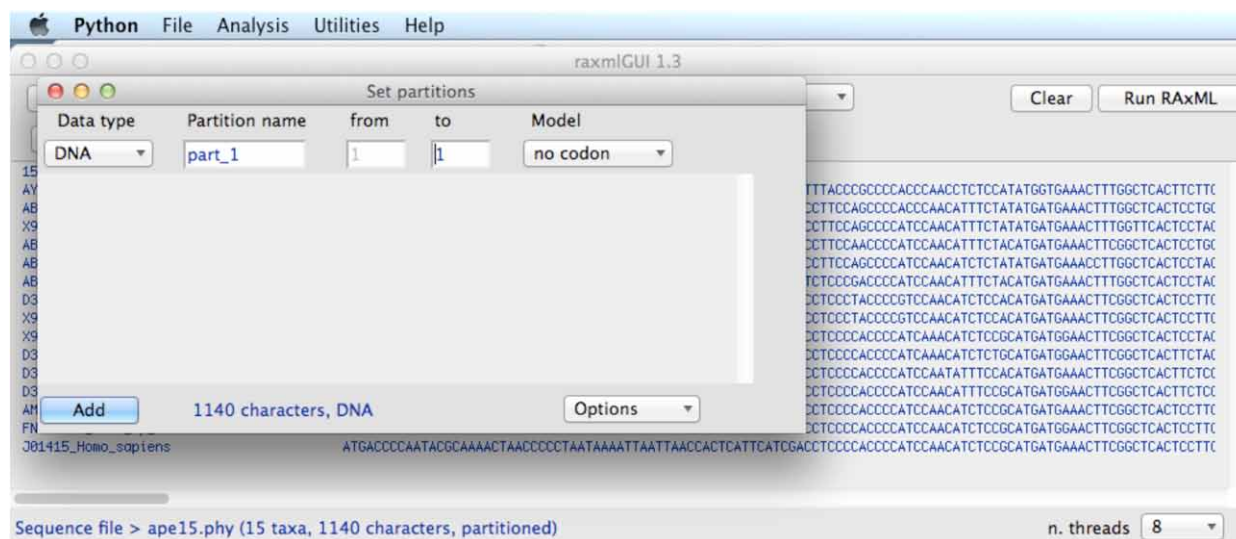


図5 パーティションを設定するエディター

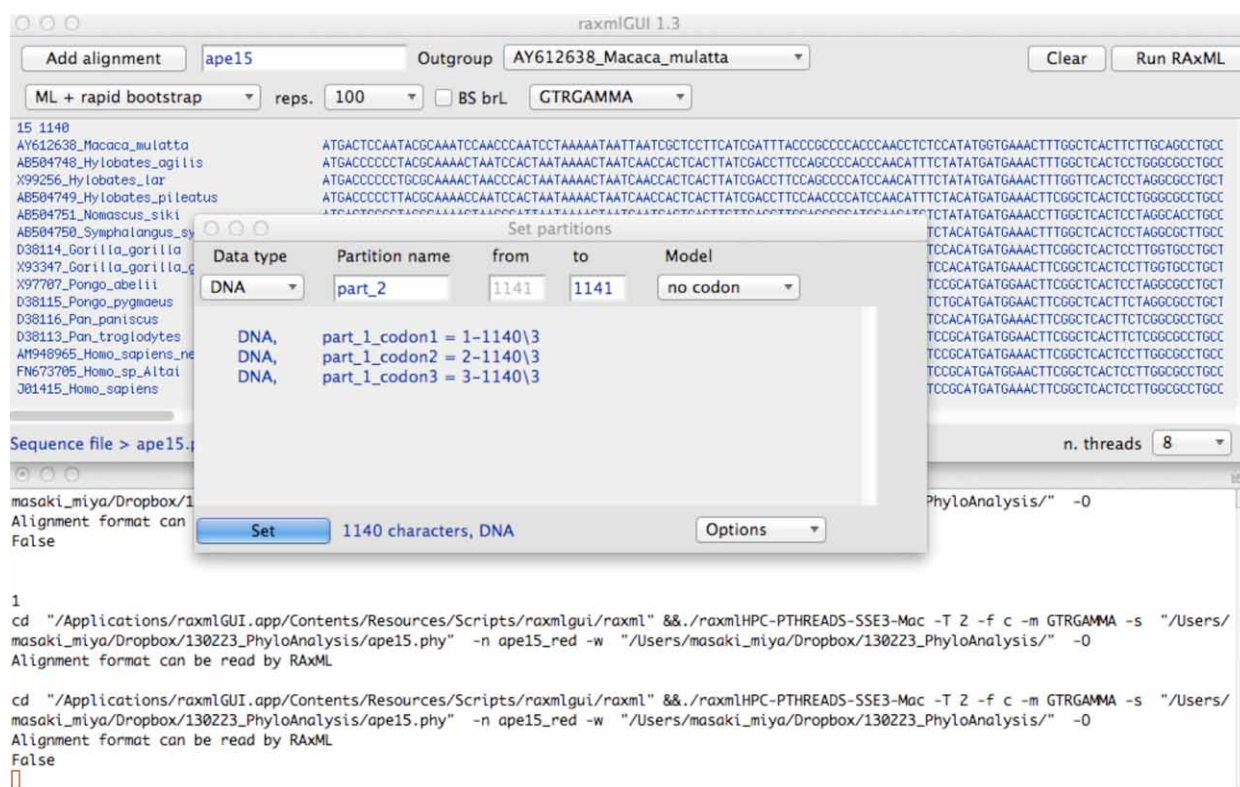


図6 エディターをつかえば簡単にコドンごとのパーティションを設定することができる

とよい。ちなみに、15種で1,140 bpという小さいデータでは100回的高速ブーストラップ解析が8つのCPUを使うと数分で終了する。10,039塩基を4つのパーティションに分けた437種の解析では16個のCPUで111時間(4.6日)かかった。

その他の便利な機能

最近話題になっている「ならず者分類群」^[7](系統樹の中を飛び回って系統推定の信頼度を下げる分類群)を見つけ出す「Dropset」という機能が「Analysis」の中の「Consensus tree」に入っている。これを用いると、ブーストラップに基づく解析中に奇妙なふるまいをする分類群を容易に特定できる。また、この分類群を除いた解析をしてみることによって、ブーストラップ値の過小推定を防げることもある。さらに「Analysis」の中の「Additional analyses」には、系統樹同士の距離(共通するクレードの割合)を計算したり、統計検定ソフトCONSELを使うためのサイトごとの尤度を計算したりなど、便利な機能が他にも数多くある。すべてを紹介することはできないが、一通りのことはraxmlGUIでできるようになっている。

おわりに

まだ、十分に普及しているソフトとはいいがたいが、RAxMLを日常的に使っている人であれば、本ソフトを導入することにより、より快適な解析環境を構築できることは間違いない。また、使い方の難しさを理由にRAxMLを使っていなかった人にとっては、福音とも言えるソフトである。

引用文献

- [1] Stamatakis, A. 2006 RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688-2690. (DOI: 10.1093/bioinformatics/btl446)
- [2] Miya, M., Takeshima, H., Endo, H., Ishiguro, N. B., Inoue, J. G., Mukai, T., Satoh, T. P., Yamaguchi, M., Kawaguchi, A., Mabuchi, K., et al. 2003 Major patterns of higher teleostean phylogenies: a new perspective based on 100 complete mitochondrial DNA sequences. *Mol. Phylogenet. Evol.* **26**, 121-138.

- [3] Silvestro, D., Michalak, I. 2012 raxmlGUI: a graphical front-end for RAxML. *Org Divers Evol* **12**, 335-337. (10.1007/s13127-011-0056-0)
- [4] Sukumaran, J., Holder, M. T. 2010 DendroPy: a Python library for phylogenetic computing. *Bioinformatics* **26**, 1569-1571.
- [5] Yang, Z. 1994 Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J. Mol. Evol.* **39**, 306-314.
- [6] Li, C., Lu, G., Orti, G. 2008 Optimal data partitioning and a test case for ray-finned fishes (Actinopterygii) based on ten nuclear loci. *Syst. Biol.* **57**, 519-539. (10.1080/10635150802206883)
- [7] Sanderson, M. J., Shaffer, H. B. 2002 Troubleshooting molecular phylogenetic analyses. *Annu. Rev. Ecol. Syst.* **33**, 49-72.

日本進化学会ニュース Vol. 14, No. 2

発行：2013年7月31日

発行者：日本進化学会（会長 倉谷 滋）

編集：日本進化学会ニュース編集委員会（編集幹事 宮 正樹）

編集委員：荒木仁志/奥山雄大/大島一正/工樂樹洋/真鍋 真)

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