

解 説

次世代シーケンサーデータの解析手法 第 17 回 バイオインフォマティクス教育の今後

門田 幸二^{1, 2, 3*}、大森 良弘²、寺田 朋子²、三浦 文²、
寺田 透^{1, 2, 3}、清水 謙多郎^{1, 2, 3}

¹ 東京大学 大学院情報学環・学際情報学府

² 東京大学 大学院農学生命科学研究科

³ 東京大学 微生物科学イノベーション連携研究機構

2020 年より続くコロナ禍の影響により、我々のバイオインフォマティクス（以下、バイオインフォ）教育の枠組みも大きな変革を強いられた。2020 年以降はオンライン講義が当たり前となり、2019 年まで売りにしてきた「1 人 1 台のノート PC を用いた講義」という特徴が消滅しただけでなく、多様な PC 環境に起因する様々な困難に直面した。2021 年は、チームコミュニケーションツールである Slack の併用によって「過去になされた質問とその顛末」の共有が可能となり、一定の省力化がなされた。また、コロナ禍とは無関係に、Python プログラミングに特化した講義の実施や、実験技術の進展に合わせた内容の更新にも取り組んでいる。本稿では、アグリバイオインフォマティクス教育研究プログラム（以下、アグリバイオ）のこれまでの取り組み、オンライン講義のメリット、そしてバイオインフォ教育の今後について述べる。

Key words : educational program, bioinformatics, programming language

はじめに

アグリバイオは、文部科学省の平成 16 年度科学技術振興調整費、新興分野人材養成の時限プロジェクト（2004～2008 年）としてスタートした。申請時の代表者が東京大学大学院農学生命科学研究科（以下、本研究科）の長であったことから、プログラム代表者は今日でも研究科長が務めている。運営主体は、プロジェクト申請者でもある最終著者（清水）をはじめとする兼任教員、および専任教員である。現在の兼任教員筆頭は、発足当初より筆頭著者（門田）らとともに運営・実務を担ってきた第 5 著者（寺田）である（2019 年までは清水）。開講科目の一部は、修士課程修了に必要な単位に加えることができる。これは、バイオインフォマティクス・バイオスタティスティクスの大学院修士課程相当の人材養成というカテゴリで採択されたことに

よる。丸 17 年におよぶ教育主体組織としての実績があり、バイオインフォの教育拠点として広く認識されている。これまで少人数のスタッフでどのように（オンライン）講義を行ってきたかや、本稿で述べるオンライン教育のメリットは、他の教育機関にとっても参考になる部分が多いのではと思われる。また、受講生側に対しては、運営側が望んでいること（受講の心得、よい質問の仕方、ノウハウ共有の重要性など）を知り、それを積極的に実践していただくことを期待している。

アグリバイオの概要

アグリバイオの設立により、本研究科の共通科目としてバイオインフォ関連科目が 10 以上新設された。概ね科目あたり 90 分×8 コマの講義時間であり、それが 1 単位となる。東京大学の場合は、他研究科でも一部科目の受講が可能である。取得単位のうち一定数が修了に必要な単位として認定される仕組みとなっており、これまでに医学・工学・理学・薬学など本研究科以外の様々な学生も受講している。また、発足当初から他大学の学生や社会人の受講

*To whom correspondence should be addressed.

Phone : +81-3-5841-8155

Fax : +81-3-5841-1136

E-mail : koji.kadota@gmail.com

も一定の条件下で認めており、その基本的なスタンスは2021年現在でも継承している。ただし、これらの外部生に対する東京大学としての正式な単位認定を行う枠組みは存在しないため、同一基準でアグリバイオ独自の単位認定を行っている。

アグリバイオの講義時間は、社会人の外部生が比較的受講しやすく既存の講義科目になるべく影響をおよぼさないよう、17:15-18:45に1コマ目、15分の休憩をはさんで19:00-20:30に同一科目の2コマ目を実施するスタイルを基本形としている。アグリバイオは2010年にカリキュラム再編を行い、3つのカテゴリ(基礎・方法論・先端トピックス)から構成される現在の講義科目名で実施している(表1)。例えば4科目から構成される基礎は、通常年度初めに実施される。1科目を週1回2コマこなしていくことで、1週間のうちどこかで講義のない日を設けつつも計4週間で基礎科目を終えることができる。こうすることで、遅くともGW後の最初の週で基礎科目が終わり、次週以降から方法論に移行することができる。受講生の多数派を占める修士1年の学生にとっては、修了に必要な単位を効率よく取得できるメリットがある。

表2は、年度ごとの開講科目数や合格者数などの各種統計情報である。初年度は、設立が2004年冬、常勤教員が6名になったのが2005年3月(筆頭著者の加入も2005年2月中旬)であった。次年度以降の本格的な体制構築作業に多くの労力が割かれていたこともあり、開講科目数(9

科目)も限定的であった。2008年と2009年間の境界線は、前述の外部予算の切れ目に相当する。一番右側の常勤教員数は、当該年度の途中で転出・加入する教員がいるため目安程度である。大まかには、繁忙期の在籍者数である。例えば初年度の繁忙期は2005年3月の集中講義時期であったため、当該年度の最大在籍者数が記されている。逆に、2018年8月加入の第2著者は、2018年の繁忙期(4-6月)以降である8月加入であったため、当該年度の常勤教員数(寺田・門田の2名のみ)には含めていない。

2004年から2020年の期間にわたってほぼ比較可能な情報は、「1科目以上の合格者数」および「のべ合格科目数」である。前者には、1科目だけ合格したヒトや4科目合格したヒトなどが含まれる(授業には出席したが単位認定されなかったヒトは含まれていない)。後者は、(1科目につき1単位なので)授与した総単位数という理解でもよい。例えば2007年はのべ合格科目数が343、1科目以上の合格者数の合計が100であることから、1科目以上の合格者1人あたり平均3.43科目合格していることになる。この平均合格科目数は開講科目数にもある程度依存するものの、概ね2.5~3.4の範囲に収まっている。のべ合格科目数(あるいは1科目以上の合格者数)は、2015年ごろまで増加傾向にある。しかし常勤教員数は減少傾向にあり、2012年以降は2~3名の常勤教員で運営している。連載第1回¹⁾の図1は、2014年のピーク時の講義風景である。アグリバイオの貸与用ノートPCは21台からスタートしたが、この頃には受講人数の増加に対応すべく、100台近いWindows PCをアグリバイオ単独で保有していた。し

科目名	2020	2021	利用言語・環境
【基礎】			
1. 生物配列解析基礎	開講	開講	Python
2. ゲノム情報解析基礎	開講	開講	R
3. バイオスタティクス基礎論	開講	開講	R
4. 構造バイオインフォマティクス基礎	開講	開講	
【方法論】			
5. 知識情報処理論	開講	開講	R
6. 生物配列統計学	開講	非開講	R
7. 分子モデリングと分子シミュレーション	開講	開講	Python, Perl
8. オーム情報解析	開講	非開講	R
9. 機能ゲノム学	開講	開講	R, Linux
10. システム生物学概論	開講	開講	R
16. フィールドインフォマティクス	開講	開講	R, Python
【先端トピックス】			
11. 農学生命情報科学特論 I	開講	開講	Python
12. 農学生命情報科学特論 II	非開講	非開講	
13. 農学生命情報科学特論 III	非開講	開講	Python
14. 農学生命情報科学特論 IV	開講	開講	R, Linux
15. 農学生命情報科学特別演習*	開講	開講	
【連携科目】			
生物測定学特論*	開講	開講	
情報生命工学*	開講	開講	

表 1. アグリバイオの教育プログラム

外部生も受講できるのは、*のついていない計15科目。1番下の生物測定学特論と情報生命工学は、兼任教員によってアグリバイオ設立以前より本研究科で開講されているバイオインフォ系の講義科目。2017年より連携科目としてアグリバイオ開講科目に含められるようになった。

年度 西暦	和暦	開講 科目数	1科目以上の合格者数				のべ合格 科目数	修了 者数	常勤 教員	
			学部	修士	博士	社会人				合計
2004	H16	9	-	12	3	5	20	53	8	6
2005	H17	15	-	65	7	3	75	251	14	6
2006	H18	15	-	73	11	8	92	297	23	7
2007	H19	15	-	83	13	4	100	343	26	7
2008	H20	15	-	68	6	1	75	237	20	7
2009	H21	12	6	64	9	1	80	185	2	3
2010	H22	15	5	100	14	11	130	362	12	4
2011	H23	15	7	91	23	19	140	461	20	4
2012	H24	14	7	113	16	32	168	510	13	3
2013	H25	15	9	115	20	26	170	430	15	2
2014	H26	13	4	105	21	55	185	561	22	2
2015	H27	13	9	139	25	34	207	629	20	2
2016	H28	12	5	140	25	20	190	486	9	2
2017	H29	14	2	143	17	28	190	487	4	2
2018	H30	16	1	145	20	26	192	541	20	2
2019	R1	15	8	150	10	18	186	491	9	3
2020	R2	16	11	150	15	0	176	482	13	2
2021	R3	15	-	-	-	-	-	-	-	2

表 2. アグリバイオの年度ごとの各種統計情報

2004~2008年は大学院向けの講義科目として開講していたため、学部生は全て社会人扱いでカウントしている。2017年以降の開講科目数には、2つの連携科目(表1)が含まれている。今年度(2021年)は、まだ講義が行われていない科目や現在進行中の科目も存在するため一部情報のみ記載している。



図 1. アグリバイオのオンデマンド配信用 Slack
 ①「ゲノム情報解析基礎」チャンネルの例。② Google ドライブ上に講義映像や資料をおき、その URL 情報を提示していることがわかる。③は受講可能期間。この科目のみリクエストに応じて配信期間を 1 か月延長した。

かし、受講生の持込 PC やネットワークの不具合対応、そして保有 PC の劣化など少人数での運営が限界に達していたため、2015 年以降は積極的な募集を行っていない²⁾。

表 1 の 3 番目のカテゴリである先端トピックスは、主に「農学生命情報科学特論 (以下、特論)」で構成されている。2010 年のカリキュラム再編前までの特論は、特論 I および II の 2 科目のみで構成され、企業や大学の研究者によるオムニバス形式のセミナーとして実施されていた。この流れは、再編後に特論 I ~ IV の 4 科目に拡大された今日でも一部科目で継承されている。例えば 2017 ~ 2021 年の特論 IV (主担当教員: 野尻秀昭 先生) は、環境微生物研究にどの様にインフォマティクスが利用されているかについて最新の研究成果を含めて紹介している。本学会誌の会員にとっては、おそらく最も分野的に近い科目であろう。また、2011 ~ 2017 年までの特論 III (主担当教員: 嶋田 透 先生) では、農業生物のインフォマティクスに関するトピックスを取り扱った。この取り組みは、2018 年に方法論の科目として新設されたフィールドインフォマティクス、およびフィールド系教員である第 2 著者 (大森) のアグリバイオへの参画につながっている。なお、本連載のテーマである次世代シーケンサー (以下、NGS) 関連の内容は、2011 ~ 2019 年まで主に特論 I で実施した。

R と Python (と Linux)

アグリバイオの特色の 1 つは、多くの科目でデータ解析環境 R³⁾ を利用したハンズオン講義が行われている点である。具体的には、年度によって開講科目や講義内容が多少異なるものの、2021 年は開講 15 科目中 7 科目 (予定含む) で R が利用されている (表 1)。これは、アグリバイオの設立当時、同様の人材養成プログラムがすでにいくつか存在していたこととも深く関係している。既存プログラムとの差別化を図るべく、明文化はされていないものの R をなるべく利用するという当初からの全体方針があったためである。

NGS を利用したトランスクリプトーム解析では、2021 年現在でも R を避けて行うことは困難である。本連載の第 5 回⁴⁾ では、「Linux 環境での R の利用法」を解説している。当時は NGS ハンズオン講習会 (2014 ~ 2017 年に主に NBDC とアグリバイオで実施した大規模な講習会) を円滑に行う目的で Bio-Linux⁵⁾ を利用したが、現在は Bio-Linux の提供サイトが消滅しているように思われる (2021 年 9 月 22 日確認)。これも時代の流れではあるが、最近では Windows での Linux 環境構築も容易になっている。Linux を用いた NGS 解析をこれから学ぶ読者は、2021 年

3月に出版された広島大学の坊農秀雅先生の著書⁶⁾などを参考にするとよいだろう。本連載の第10回頃までの情報は、相当古くなっているので注意してほしい。

アグリバイオの設立当初は、当時の主流であったPerlをプログラミング言語として教えていた。しかし近年はPythonが主流となっているため、2020年より順次Pythonに切り替えている(但し、2018年新設のフィールドインフォマティクスは当初からPythonである)。具体的には、特論Iの枠組みでPythonプログラミングに特化した講義を2020年に試行的に実施(担当教員:孫建強先生)して需要を確認し、2021年も継続的に実施している。また、フィールドインフォマティクスの源流である特論IIIも2021年に再度開講し、温室内に設置した栽培環境データのモニタリングや簡単なデータ処理を意識したPythonプログラミングの講義を実施した。Pythonプログラミング系の参考書としては、国立遺伝学研究所の黒川顕先生がとりまとめた書籍⁷⁾が挙げられる。執筆陣には、アグリバイオの非常勤講師として参画中の若手の先生方も含まれている。

多くの読者は、今でも「東大アグリバイオといえばR」というステレオタイプなイメージをもっているであろう。しかし実際には、Pythonプログラミング教育にもかなり力を入れている。今からプログラミング言語を習得する学生にとって有意義なのは、間違いなくPythonである。しかしRは、特に統計解析分野での地位を確立しており、将来的にも完全にPythonに取って代わられるとは考えにくい。また、複数の外部生の受講動機としてRを学びたいと明記されていることから、日本でRを学べるバイオインフォ教育機関として持続しつづけるのが東大アグリバイオの使命だと考えている。Rに特化した講義科目であるバイオスタティスティクス基礎論(担当教員:岩田洋佳先生)と知識情報処理論(担当教員:麻生川稔先生)は来年度以降も開講する予定である。

解析環境の構築

ここでの解析環境構築とは、受講に必要なソフトウェアのインストール作業のことを指す。アグリバイオでは、RについてはRStudio、PythonについてはGoogle ColaboratoryやJupyter Notebookを主に利用している。受講生に対しては、これらの講義に必要なソフトウェアを事前にインストールしておくよう通知している。しかしながら現実には、それを怠ってくるヒト、インストールを試みたがダメだったと講義開始後に伝えてくるヒトなど、講義中になんらかの不具合を抱えた受講生が一定数存在する。2019年までは環境構築済みのアグリバイオ所有PCを貸与することで事なきを得ることができたが、オンライン講義の現在は環境が多様な個人所有PCでの受講が基本形である。このため特に「インストールを試みたがダメだっ

た」場合に、進行中の講義の裏で少人数のスタッフが対応に追われ、講義自体に目配りできないという事態がアグリバイオでも生じた。たとえ講義(直)前の相談であったとしても、通常の講義準備や事務的な問い合わせ対応などに忙殺されているのが多くの教育機関の日常であろう。

アグリバイオでは、ウェブサイト上で「講義で使用予定のソフトウェア」の具体的なインストール手順を提供している。一部の受講生にとって、インストールは苦行である。そのため我々は、初めて見る画面でもどのボタンをクリックしてインストールを進め、最終的にどのようになっていれば無事インストールが完了していると判断してよいかまで丁寧を示すことを心掛けている。こうすることではじめて、その次の段階である講義内容の理解に向けた心の準備が穏やかにできる。特にRは、パッケージのインストールが鬼門である。インストール経験豊富なスタッフが、アグリバイオ所有の同じ型のPCで、ほぼ同じ時刻に同じ作業を行っても、一部パッケージのインストールに失敗することはよくある。それゆえ、R本体だけでなく、パッケージのインストール確認作業や、想定されるエラーメッセージとその対処法まで含めた情報を提供している。エラーメッセージを含む画面のスクリーンショットや、どこをどのように変えて自身の個別の問題に対処するかまで具体的に示した情報にしておくことで、受講生がその情報の在り処のみをたよりにして自力で問題解決できる可能性を極限まで引き上げることができる。

ハンズオン講義を円滑に行う上での最大の障壁である解析環境の構築ができれば、基本的に受講生を同じスタートラインに立たせることができたことになる。しかし現実には、諸事情による遅刻や(オンライン講義の場合は)一時的な接続切れにより、途中で脱落しかける受講生が一定数存在する。また、成功裏に構築できた解析環境であったとしても、講師側が構築した環境とプログラムのバージョンが異なる場合がありうるため、バージョンの違いに起因する不具合も十分に起こりうる。これらの問題を完全に回避するのは不可能であるが、たとえ自分の環境でその場ではうまくいかなかったとしても、成功裏に実行できた手順を完全に再現できるレベルの詳細な講義資料が提供されれば、受講生の理解度・満足度が各段に向上する。これは他稿²⁾で述べた「わざわざ講義に出席するのがばかばかしいと思えるほど詳細かつ丁寧な講義資料を提供する」と同義であり、一定の出席者数の減少に寄与する。実務的には、バージョンの違いなどに起因する講義資料の結果との違いや不具合は、全員で共有する価値のある情報である。多くの受講生は違いを失敗と判断して意気消沈する傾向にあるが、講師側がその点を正しく伝え、そのような結果を得た受講生が手を挙げやすい雰囲気をつくることも重要であろう。

コロナ禍のアグリバイオの活動

これまで(オンライン)講義を実施する側と受講する側が共通して実感した困難の多くは、個々人が所有する多様なPC環境に起因する事柄だったのではないだろうか。特にWindowsとMacというオペレーティングシステム(以下、OS)の違いは、アグリバイオでも講義の円滑な進行を幾度となく妨げた。この問題自体は以前から存在していたが、当時は持込PCで発生した問題に対して講義補助員が可能な範囲で対応し、時間がかかりそうな場合はアグリバイオ所有PCの貸与で完結していた。しかし、「問題を抱えた持込PC数>講義補助員数」となったときに講義がストップすることが多いため、2020年に初めて外部生の受け入れを断念した。理由は、我々はもちろん、受講生も講義補助員もオンライン講義が未経験であり、全員が持込PCで受講しているのと似た状態を想定せねばならなかったからである。受講生の絶対数を抑制することで、問題を抱えた受講生の数を極力減らす作戦はある程度奏功した。具体的には、講義内容も変わってきているため単純比較はできないものの、全体としてコロナ禍前の6~7割程度の受講人数で実施することができた。講義補助のメインは第1~3著者であったが、主に対応可能なOSが門田(Win)・大森(Mac)・寺田(WinとMac)というバランスであったことも、厳しい2020年を乗り切れた大きな要因であろう。

オンライン講義自体は、ウェブ会議サービスの1つであるZoomで実施した。講義資料はあらかじめダウンロードしておいてもらい、特に前半の多くの講義で画面共有を行わずに音声のみで行った。これは、2020年度当初、学生のインターネット回線の契約内容によっては、月末に通信制限がかかってしまうことが危惧されていたためである。質問はZoomのチャット機能を利用したが、一度に多くの質問が寄せられると対応が困難になる場面があった。Zoomのチャット機能は全員宛てと特定の個人宛てのどちらかが選べるが、後者の場合は寄せられた質問を読み上げないと他の受講生との共有ができないうえ、聞き逃してしまうこともよくある。それゆえ、できるかぎり全員宛てで質問をするように促すことも心掛けた。また、難聴などケアを必要とする学生からの質問に対する回答にも注意を払わねばならない。なお、後半の講義科目では、データ通信量が大学のキャンパス全体としても大丈夫であることが確認され、本研究科でも画面共有を制限なく行うことができるようになった。

講義映像(前半の多くは音声のみ)を録画したものを提供するオンデマンド配信は、おおむね講義の翌日から1か月程度視聴可能とした。2020年は東大生のみを受講対象としたこともあり、東大内部の学習管理システム(ITC-LMS)を利用して講義資料の配布、Zoom会議やオンデマンド配信URLの通知などを行った。2019年までのアグリバイオは、独自の教務管理システムのみを利用していた。

これは、ITC-LMSが年度末で全ての情報がリセットされること、外部生の情報も取り扱う必要があったこと、そして年度間にまたがる単位取得状況を記録しておく必要があるためである。新カリキュラム以降のアグリバイオは、8科目以上の合格(8単位以上の取得)で修了証を授与している。表2に示された修了者数に大きなばらつきが見られるが、理由はおそらく一定数が複数年かけて修了認定基準に達するためであろう。同一年度内で8科目以上合格したヒト、何年も頑張っってやっと修了認定となった社会人受講生、そしてその中間など多様である。

コロナ禍における一通りの業務を2020年に経験した結果、および次年度も原則としてオンライン講義を継続実施する東大全体の方針を踏まえ、2021年はオンデマンド配信のみとして外部生を受け入れる方針とした。Slackは、Zoomのチャット機能を利用した質問対応時の課題を補う目的で、2021年度初頭の東大生のみでの講義の中で試行的に利用した。東大農学部内の交流用として2021年初頭から運用されている既存Slackの中にアグリバイオのチャンネルを作成していただき、我々スタッフの利用経験を積んだ。外部生受け入れのお知らせはアグリバイオのホームページのみで行い、受講申し込みはGoogleフォームで行った(6月18日締め切り)。この間、外部生のみを対象としたアグリバイオ独自Slackの枠組みについての検討を同時並行で行い、6月中旬に立ち上げた。

図1は、基礎科目「①ゲノム情報解析基礎」チャンネルのスクリーンショットである。この科目を登録した受講生は、②全4回(計8コマ)の講義映像(ダウンロード不可)と講義資料(ダウンロード可)のURLが見られる。基本的に全て自習であり、通常の講義で提示したものと同一課題をこなしてもらう。同一科目内の全てのレポートを作成し終えた段階で、課題提出用のGoogleフォーム経由で提出するような枠組みとした。ただし、通常の講義においても聴講のみでレポートを提出しない学生がいるように、外部生向けのオンデマンド配信でもレポートを提出しない受講生は一定の割合で存在する。今年度の配信期間を終えた基礎4科目について調べた結果、レポートを提出したのはのべ受講登録者222名中93名(約42%)であった。この理由は、Slack内でとった受講目的に関するアンケート結果からも納得できる(図2)。外部生(73名中)50名が①お知らせチャンネル上に提示されたアンケートに気づいて回答し、②5名が単位取得目的、③残りの45名が単位取得目的ではなく単純に勉強したいに投票した。

このアンケートを行ったのは、当初基礎科目のオンデマンド配信期間を2か月としていたが、④雑談室チャンネル上の気軽なやりとりの中で、純粋に勉強したいだけのヒトにとっては配信期間が短いと感じるのでは?!と思ったためである。実際、⑤で希望する配信期間を3~6か月までの範囲で投票してもらくと、2か月では短いと確信できる結果となった。この投票結果を受けて、図1の③でも修正

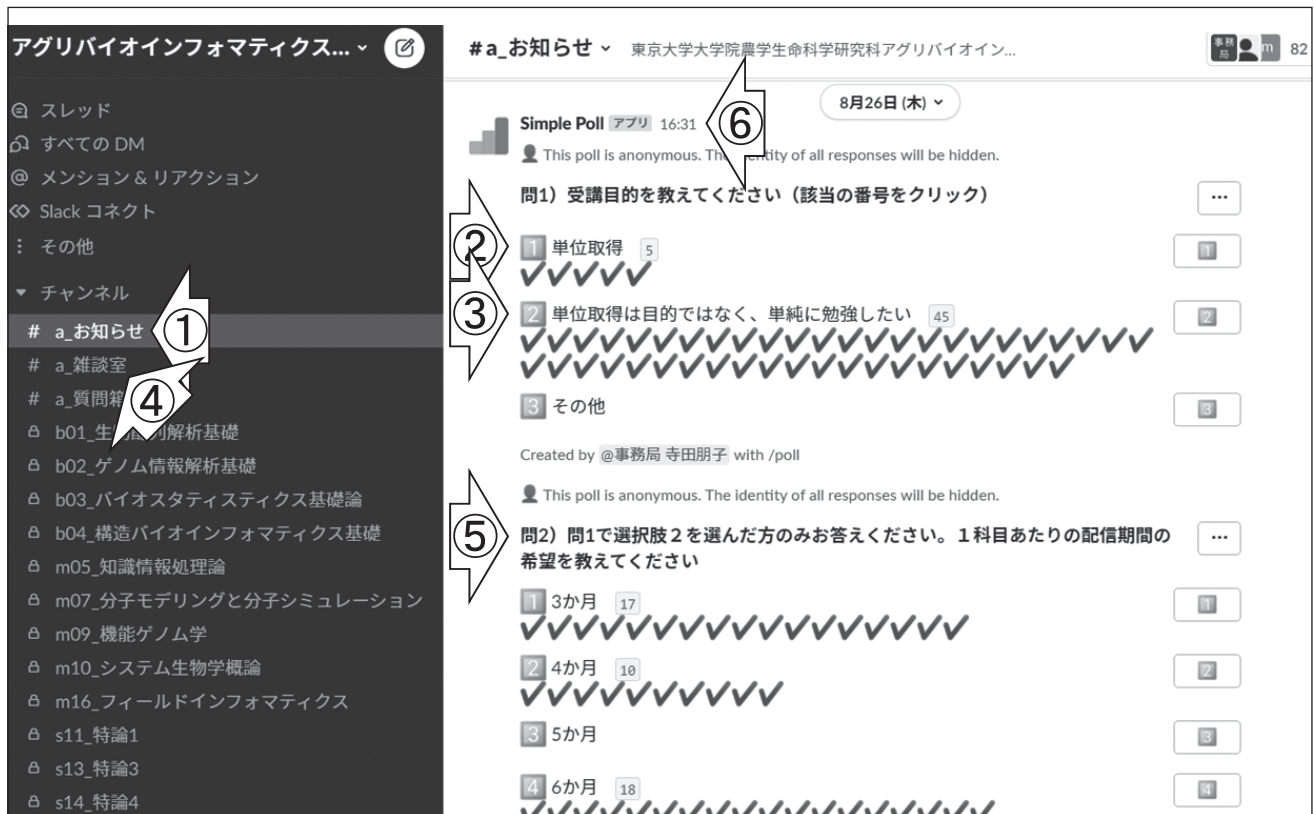


図 2. 外部生に対して行った受講目的のアンケート結果

①～⑤については本文を参照。このアンケートは、Slack と⑥ Simple Poll というアプリを連携させて作成している。受講生は該当する番号をクリックするだけでよく、その結果が即座に反映されるので便利である。

されているように、①の科目のみ配信期間を3か月に延長して対応した。来年度は、レポート提出までの期間のみ内部生と同程度としつつも、可能な科目についてはオンデマンド配信期間を最初から3～4か月程度として対応する予定である。制度上、非常勤講師に対して実際の講義時間 (+ a) 分の謝金しか出せないため、特にオンデマンド配信期間内に寄せられた質問に対しては、可能な限り運営側で答えるようにして非常勤講師の負担軽減に努めている。

図 3 は、Slack 上での質問と応答の例である。①の部分に受講生 ID、②の部分に質問内容が書かれている。これは「R の Biostrings というパッケージを用いて、塩基配列をアミノ酸配列に変換する translate 関数を実行した部分でエラーが出た」という内容である。Slack 上で質疑応答を行う一番のメリットは、③エラーメッセージが出た画面のスクリーンショット (以下、スクショ) を質問者が簡単に添付できる点である。エラーメッセージを含むスクショを見せてもらうだけで原因を特定できる場合が多いので、運営側にとってこの機能は非常に重要である。この科目の受講者は、誰でもこの質問に対して行われた計 6 件のやりとり (④スレッド) を見ることができる。つまり、過去になされた質問とそれに対する回答が蓄積されていくので、運営側が同じ質問に答えなくてよいという意味で負担軽減

効果がある。また、一部の受講生が自力解決した事柄や有用情報などを書き込んでくれたことも同様の効果を生んだ。我々だけでなく多くの組織・機関が慢性的な人手不足に悩まされているが、Slack のようなコミュニケーションツールの活用は、今後のバイオインフォマティクス教育の現場でも業界標準になっていくであろう。

オンライン講義のメリット

多くのバイオインフォ教育機関にとって、おそらく最も頭を悩ませている事柄は運営体制の維持であろう。今回のタイトルの基盤的な事柄ではあるが、現実には目を向けねばならない。丸 17 年の実績があるアグリバイオも、近年はこの問題を抱え続けている。外部予算プロジェクト (2004～2008 年) 後は、文部科学省の平成 21 年度特別教育研究経費「バイオインフォマティクスを基盤とした農学生命科学統合教育研究推進機構の創設」が認められ、第 2 期としてさらに 5 年間 (2009～2013 年) 継続した。しかし、2008～2009 年および 2011～2013 年の常勤教員の減少からも明らかのように、予算の切れ目や第 2 期の後半は非常に厳しいものがあつた。一人あたりの業務量の増加だけでなく様々な心労が重なるためである。2013 年以降は、次

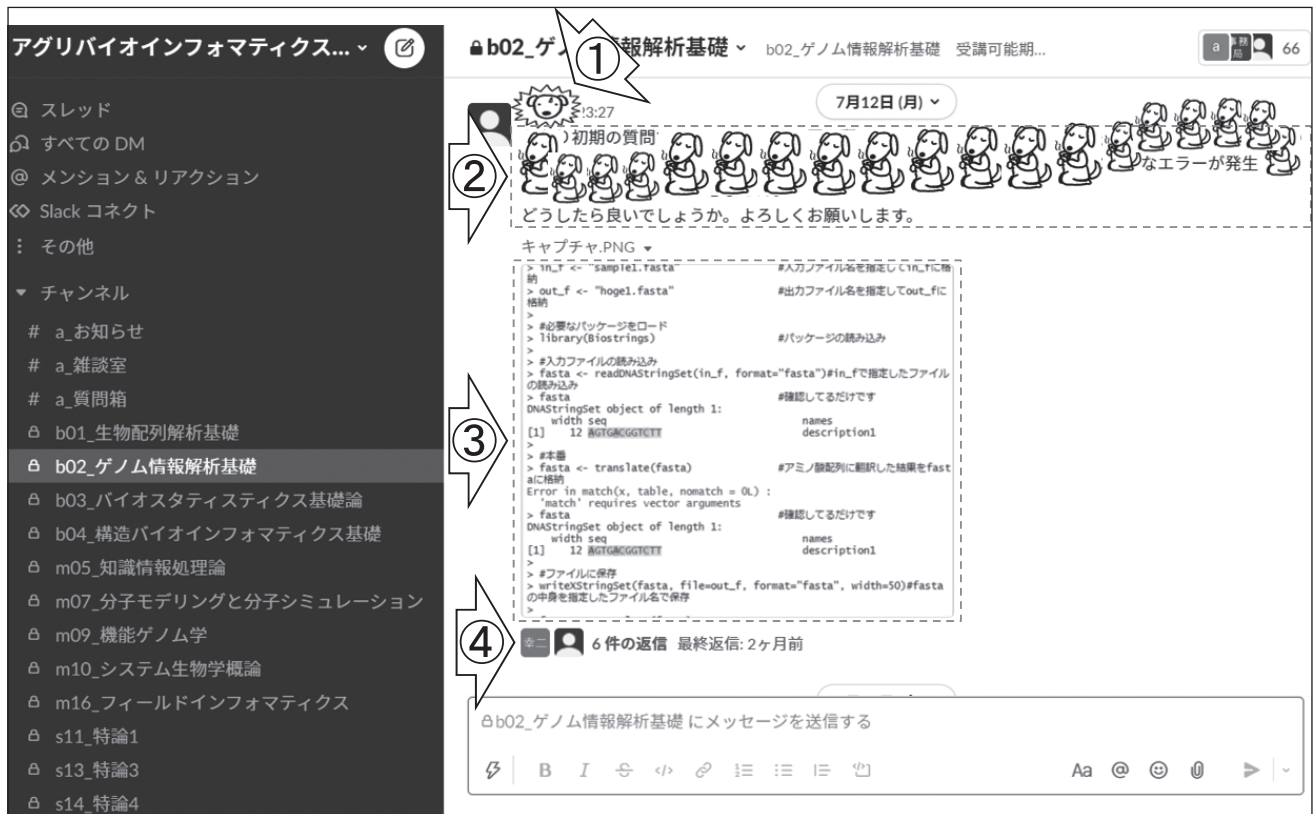


図 3. 受講生からの質問とスレッドを用いた応答の例。
 受講生は、③画面上に表示されているエラーメッセージを含む任意の領域のスクリーンショットをコピー&ペーストで Slack 上に貼り付けられる。運営側もエラーの状況が把握しやすいのでお互いにとって便利である。

年度の運営費確保（≒アグリバイオの存続）が一番の課題であった（主に清水が担当）。スタッフの人件費が多くを占めるものの、100台近いノートPCをはじめとする設備備品の維持管理費がかかるためである。

コロナ禍で強制的にスタートしたオンライン講義ではあるが、特にハンズオン講義との親和性が高いバイオインフォ教育においては、我々の実感としてメリットのほうが非常に大きいと感じている。1番のメリットは、コロナ禍前は数年後の破綻が確実視され、心労の大部分を占めていたノートPCの維持費が大幅に削減された点である。我々が2019年まで売りにしていた「1人1台のノートPCを用いた講義」という特徴は、事実上オンライン講義の前提条件として雲散した。しかし維持費の削減は、それを遥かに上回るメリットである。2点目は、労働条件の大幅な改善である。100台規模になると、かつてはソフトウェアのインストール作業や定期的な保守だけでも相当の時間がかかっていた。講義の準備や後片づけ、ノートPCの運び込みなどの肉体労働もなくなった。緊急事態宣言下で在宅勤務せざるを得ない状況でも講義を継続できたため、スタッフ自身の傷病や親の介護など、かつては休職または離職せざるを得ないような状況であったとしても（本人次第ではあるが）仕事を続けることができる。非常勤講師からも、

わざわざ東京大学に足を運ぶ必要がないだけでなく、人身事故などによる電車の大幅な遅延などの影響を受けないため、精神衛生上もよいと好評であった。

3点目は、非常勤講師選定の自由度向上である。かつては特に交通費の点で東京大学から一定の距離の範囲内では講師を選定することができなかったが、オンライン講義であれば距離の制約から解放される。これは受講生にとってもメリットが実感できる事柄であろう。4点目は、講義資料の印刷が不要になった点である。アグリバイオでは、計3時間の講義を1回で行うのが基本形である。必然的にスライドの枚数も多くなるため、講義資料の印刷だけでも数時間かかる。かつては、主に印刷機の不具合のため講義開始時間に間に合わないこともあった。受講人数規模の大きなアグリバイオならではのかもしれないが、紙代や印刷代が大幅に削減されるだけでなく、関連する様々な心労がなくなったメリットも実務担当の実感としては大きい。オンライン講義中は共有画面と手元の講義資料PDFの両方を見る必要があるものの、PDFファイル中にリンクを埋め込むことで解析データへのアクセスやスライド間の移動を容易にするような工夫も一部講義で実践した。

教育内容の更新および他機関との連携

本連載テーマである NGS 技術も継続的に発展している。教える内容は、従来のデノボアセンブリ・アノテーション・発現変動解析などから、時代の進展に合わせてパンゲノム・ハプロタイプフェージング・エピゲノム解析などに踏み込みこんでいかねばならない。また、得られるデータの高解像度化に伴い、バイオインフォ側も必要とされる生物学的な知識量が増えている印象を受ける。例えば畜産ゲノム解析系は、ENCODE プロジェクトの進展にアンテナを張ったほうがよいが、エンハンサ・インシュレータ・ヒストンマークなどがわからないと話にならない。ロングリード技術の進展によって、従来の SNP レベルの比較的小規模な変異だけではなく、広い意味での構造変異（逆位・転座・数千塩基レベルの広範な挿入や欠失）を捉えられる時代になっていることや、例えばそれと従来のマルチプルアラインメントの位置づけも正しく理解させねばならない。リードのマッピングは、従来のリファレンスゲノムではなくゲノムグラフに対して行われるようになっていく。その場合のマッピング結果ファイルの形式は、SAM/BAM ではなく GAM や GAF が普及しつつある。従来のゲノムブラウザから専用の可視化ツールへの移行も検討課題である。これらをただ単に増やすだけなら簡単であるが、実際には1つの科目の枠組みでうまく完結させねばならない。このため、教える内容から何を削って何を新たに追加すべきか、また現実的な問題として誰が教えられるかまで含めて考えねばならない。

幸い、バクテリアはノート PC レベルで解析できるので、講義で取り扱う材料として適している。同一生物種内の菌株間の多様性を調べる文脈で提唱されたパンゲノム解析⁸⁾は、比較ゲノム解析の一種とみなすことができる。そのため、古くから行われてきた比較するゲノム間での相互ベストヒットに基づくオーソログ同定などは、要素技術として今日でも教える価値がある（生物配列解析基礎で実施）。バクテリアのパンゲノム解析でよく示される「ゲノム数（個体数または菌株数）が増えるとパンゲノムが増えていくが、だんだん増加率（コアゲノムが減っていくがだんだん減少率）が下がっていき、やがて頭打ちになる」プロット⁹⁾は、分布の形状をヒープスの法則やべき乗則に当てはめて議論する意義を学生に伝えやすいと思われる。これは統計分野でよく議論されるモデル選択の話と密接に関連した事柄であるが、「ただ単にどの数式を用いるのが適切か」という議論をしているだけだということを一言添えるだけで、学生の苦手意識をかなり払しょくできるのではないだろうか。

コアゲノムとは、ある生物種を構成する様々な菌株が共通してもつ遺伝子セットのことである。ある特定の菌株は、コアゲノムに加えて、その菌株を特徴づけるユニークな遺伝子セット（dispensable ゲノム）を持つ¹⁰⁾。ある生物種を理解することは、その生物種を構成する様々な個体

（菌株）間の多様性を理解することとほぼ同義である。集合論的に表現すると、パンゲノムは和集合、コアゲノムは積集合となる。日本に点在するバイオインフォ教育機関についてもパンゲノム解析と同様の枠組みで整理することができる。例えば、アグリバイオの基礎科目で教えている内容の多くは、他の教育機関でも共通して教えているであろうことからコアゲノムとみなすことができる。また、アグリバイオならではの特徴に相当するもの（dispensable ゲノム）は、科目ではフィールドインフォマティクスや分子モデリングと分子シミュレーションが挙げられよう。逆に、アグリバイオにはないが他機関が教えているような内容もあろう。

今年度はアグリバイオ独自 Slack で外部生へのオンデマンド配信を行ったが、このような取り組みは特定の部局や大学単位ではなく、コンソーシアム/アライアンス化が原理的に可能である。数多くの経験を積んでいるはずの R パッケージのインストールですら、我々のみでは解決できないことも実際にあった。1つの大きな Slack を立て、質問とその回答の共有を行うだけでも有意義であろう。また、講義資料などもバイオインフォの“パン資料”として Slack 上で共有すれば、同じ内容を様々な視点で解説したバイオインフォ全体の知が集積されることが期待される。

おわりに

今回は大枠的なところを中心に述べたが、我々運営側が目指すべきところは受講生側にとって費用対効果の高い教育であろう。アグリバイオの講義を受けようと思えば受けられるが無理をしている学生、研究室ゼミなどによりリアルタイムで受けられない学生、実験の合間に自分のペースで勉強したい学生など様々である。したがって、オンデマンド配信の枠組みは、対面授業再開後もできる限り残した方がよいと思われる。しかし、オンデマンド配信での視聴は、無言の時間をもどかしく感じた学生も少なからずいたのではないだろうか。期待して視聴しつつ、最後に徒労感を味わった学生もいたかもしれない。運営側はできる限り詳細かつ丁寧な講義資料を提供し、学生側も視聴前に講義資料にざっと目を通しておくよいかもかもしれない。今回紹介したノウハウは、我々の時間が許す限り全て提供するので遠慮なく問い合わせしてほしい。

謝辞

これまでのアグリバイオの活動にご尽力・ご協力いただいた全ての関係者に感謝申し上げます。

利益相反 (COI)

門田幸二、大森良弘、寺田朋子、三浦文、寺田透、清水謙多郎：本論文発表の内容に関連して開示すべき COI 状態はない。

参 考 文 献

- 1) 門田幸二, 孫建強, 湯敏, 西岡輔, 清水謙多郎 (2014) 次世代シーケンサーデータの解析手法: 第1回イントロダクション, 日本乳酸菌学会誌 25: 87-94.
- 2) 門田幸二 (2020) バイオインフォ教育の実務に携わってきた一兵卒の私見, バイオインフォマティクス学会ニュースレター 37: 6.
- 3) R Development Core Team (2010) R: A language and environment for statistical computing, In R Foundation for Statistical Computing, Vienna, Australia.
- 4) 孫建強, 清水謙多郎, 門田幸二 (2015) 次世代シーケンサーデータの解析手法: 第5回アセンブル, マッピング, そしてQC, 日本乳酸菌学会誌 26: 193-201.
- 5) Field D, Tiwari B, Booth T, Houten S, Swan D, et al. (2006) Open software for biologists: from famine to feast, Nat Biotechnol 24: 801-803.
- 6) 坊農秀雅 (2021) Dr. Bono の生命科学データ解析 第2版, メディカル・サイエンス・インターナショナル
- 7) 先進ゲノム解析研究推進プラットフォーム編 (2021) 独習Python バイオ情報解析, 羊土社
- 8) Tettelin H, Massignani V, Cieslewicz MJ, et al. (2005) Proc Natl Acad Sci U S A 102: 13950.
- 9) Nourdin-Galindo G, Sánchez P, Molina CF, Espinoza-Rojas D, Oliver C, et al. (2017) Front Cell Infect Microbiol 7: 459.
- 10) Sherman RM, Salzberg SL (2020) Nat Rev Genet 21: 243-254.

Methods for analyzing next-generation sequencing data XVII. The Future of Bioinformatics Education.

Koji Kadota^{1, 2, 3}, Yoshihiro Ohmori², Tomoko Terada², Aya Miura², Tohru Terada^{1, 2, 3}, Kentaro Shimizu^{1, 2, 3}

¹*Interfaculty Initiative in Information Studies, The University of Tokyo.*

²*Graduate School of Agricultural and Life Sciences, The University of Tokyo.*

³*Collaborative Research Institute for Innovative Microbiology,
The University of Tokyo.*

Abstract

Due to the impact of COVID-19 disaster, our framework for bioinformatics education has been forced to undergo major changes since 2020. Online lectures have become the normal, and as a result, hands-on lectures that had been our main feature before the disaster disappeared. We also faced various difficulties mainly caused by the diverse PC environments. In 2021, the use of "Slack" enabled the sharing of "questions asked in the past and their consequences," which resulted in a certain amount of labor savings. We have also started lectures on the Python programming language in 2020 as well as updating our educational program. In this commentary, we review the efforts of our program and describe our vision for the future of bioinformatics education. Supplementary materials are available online at: http://www.iu.a.u-tokyo.ac.jp/kadota/r_seq2.html#about_book_JSLAB.