

平成 26 年度

バイオインフォマティクス人材育成カリキュラム

(次世代シーケンサ) 速習コース

実施報告書

門 田 幸 二

(東京大学 大学院農学生命科学研究科 特任准教授)

概要

近年の次世代シーケンサ(NGS)解析技術の飛躍的な発展は、ライフサイエンス分野におけるデータ量の急増を生み、データの整備や活用がボトルネックとなっている。研究現場では、NGS データを自在に解析できるバイオインフォマティクス人材が求められている。これらの状況を踏まえ、バイオサイエンスデータベースセンター (NBDC) 運営委員会 人材育成分科会では、NGS を利用した様々な研究目的に対応可能な人材を効率的に育成するためのカリキュラム (<http://biosciencedbc.jp/gadget/chousa/generation-sequencer.pdf>) の策定を行った。

この NGS 用カリキュラムは、最低限必要とされる知識・技術を 2 週間程度で身につけることを想定した「速習」と、時間をかけて習得することを想定した「速習以外」に分かれている。本講習会は、「速習」コースの試行的実施、および教材の整備・充実を目的とした共同研究として実施した。具体的には、NBDC と連携して、2 週間の速習コースを試行的に実施し、受講者アンケート等によりその有効性を検証する。また、可能な範囲でカリキュラムに関する講義内容を公開する。これらの活動を通じて、NGS データの解析において必要な人材の育成に資する教材の整備・充実や公開および知識の共有を目指す。

1. 実施体制およびスケジュール

本研究では、NGS 解析に特化したバイオインフォマティクス人材育成を目的として、1) NGS 用カリキュラムの試行的実施、および 2) 教材の整備について研究開発を行う。

グループ名	研究代表者氏名	所属機関・部署・役職名	研究題目
東京大学	門田幸二	東京大学・大学院農学生命科学研究科・特任准教授	次世代シーケンサ解析に特化したバイオインフォマティクス人材育成カリキュラム実施および教材の整備

全体スケジュールは以下の通りである。

研究項目	H26年度
1. NGS用カリキュラム ・講師の選定、講義内容の検討など ・「速習」コース実施(9/1~12)	4~8月 9/1-12 ↔ ↔

<p>2. 教材の整備</p> <ul style="list-style-type: none"> ・ポータルサイトの構築支援 ・「Rで塩基配列解析」の内容充実 	
---	--

2. 実施結果

2-1. 「速習」コースの概要

NBDC 運営委員会・人材育成分科会で策定された NGS 用カリキュラムの「速習」コース (<http://biosciencedbc.jp/gadget/chousa/generation-sequencer.pdf>) は、6つの大項目(1. コンピュータリテラシーとサーバー設計、2. 配列インフォマティクス、3. データ解析基礎、4. 次世代シーケンサ、5. ゲノム関連の倫理・法律、6. 分子生命科学) および 10.5 日分の日数からなる。今回の試行実施では、1日を4コマ(1コマあたり90分)に分け、全部で10日(9/1-5 および 9/8-12)の期間で行った。0.5日超過分は、「5. ゲノム関連の倫理・法律」および「6. 分子生命科学」の講義時間を半分に圧縮することで対応した。

実施日	実施時間	項目番号および項目	レベル	形式	担当講師(所属)
9月1日	10:40-12:00	1-1. OS、ハード構成	初級	講義	中村保一(DBBJ)
	13:15-14:45	1-2. ネットワーク基礎	初級	講義	中村保一(DBBJ)
	15:00-18:15	1-3. UNIX I	初級	実習	仲里猛留(DBCLS)
9月2日	10:30-18:15	1-3. UNIX I	中級	実習	仲里猛留(DBCLS)
9月3日	10:30-18:15	1-4. スクリプト言語(シェルスクリプト)	中級	実習	服部恵美(Amelieff)
	10:30-18:15	1-4. スクリプト言語(Perl)	中級	実習	服部恵美(Amelieff)
9月5日	10:30-14:45	2-1. 配列解析基礎	初級	実習	坊農秀雅(DBCLS)
	15:00-18:15	2-2. バイオ系データベース概論	初級	実習	小野浩雅(DBCLS)
9月8日	10:30-12:00	3-1. R基礎1	初級	実習	門田幸二(東京大学)
	13:15-14:45	3-2. R基礎2	初級	実習	門田幸二(東京大学)
	15:00-18:15	3-3. R各種パッケージ	中級	実習	門田幸二(東京大学)
9月9日	10:30-14:45	3-4. R Bioconductor I	中級	実習	門田幸二(東京大学)
	15:00-18:15	3-5. R Bioconductor II	中級	実習	門田幸二(東京大学)
9月10日	10:30-12:00	4-1. 次世代シーケンサ基礎I	初級	講義	倉田哲也(NAIST)
	13:15-14:45	4-2. 次世代シーケンサ基礎II	初級	講義	倉田哲也(NAIST)
	15:00-18:15	4-3. 次世代シーケンサ実習I	初級	実習	山口昌雄(Amelieff)
9月11日	10:30-14:45	4-4. 次世代シーケンサ実習II (ChIP-seq)	初級	実習	河岡慎平(ATR)
	15:00-16:30	6. 分子生命科学	初級	講義	河岡慎平(ATR)
	16:45-18:15	5-1. ゲノム情報倫理概論	初級	講義	箕輪真理(NBDC)

					川嶋実苗(NBDC)
9月12日	10:30-14:45	4-4. 次世代シーケンサ実習II (多型解析)	初級	実習	山口昌雄(Amelieff)
	15:00-18:15	4-4. 次世代シーケンサ実習II (RNA-seq)	初級	実習	山口昌雄(Amelieff)

表 1. 実施された NGS 用カリキュラム「速習」コース概要。

2-2. 「速習」コースの実施体制

本コースの世話人は、NGS 用カリキュラムをとりまとめた藤博幸先生（産総研・ゲノム情報解析センター）および門田幸二（東京大学）の2名である。藤博幸先生は、コース初日の講義に先立ち、世話人を代表して趣旨説明を行った。研究代表者の所属機関である東京大学、および JST-NBDC の2つの機関による主催として実施した。また、他に3つの機関が共催として参画した。具体的な各機関の担当業務は以下の通りである。

主催機関1	東京大学・大学院農学生命科学研究科・アグリバイオインフォマティクス教育研究ユニット
担当業務	NGS用カリキュラム試行実施の統括、講師の選定、具体的な講義内容の検討、実習用PCへのソフトウェアインストール作業、講義実施および補助（Teaching Assistant; TA）、受講者アンケートの実施、教材に関する情報収集、ポータルサイトの構築支援。
参考URL	http://www.iu.a.u-tokyo.ac.jp/~kadota/r_seq.html#bioinfo_ngs_sokushu_2014
主催機関2	科学技術振興機構・バイオサイエンスデータベースセンター（JST-NBDC）
担当業務	NGS用カリキュラムの統括、企画・推進、講義実施および補助（ビデオ撮影）、ポータルサイトの構築、受講者への追跡調査。
参考URL	http://events.biosciencedbc.jp/training/ajacs47
共催機関1	HPCI戦略プログラム分野1「予測する生命科学・医療および創薬基盤」人材養成プログラム（産総研・ゲノム情報研究センター）
担当業務	受講申込業務全般（受付ウェブシステムの開発・運用、事務連絡）、講義補助（TA）。
参考URL	https://hpci.cbrc.jp/modules/tutorial/NBDC2014.html
共催機関2	情報・システム研究機構 ライフサイエンス統合データベースセンター（DBCLS）
担当業務	講義実施および補助（ビデオ撮影およびTA）、講義映像編集および統合TVでの公開業務全般。
参考URL	http://togotv.dbcls.jp/ja/
共催機関3	日本DNAデータベース（DDBJ）
担当業務	講義実施および補助（TA）。
参考URL	http://www.ddbj.nig.ac.jp/index-j.html

表 2. 「速習」コース実施体制。

実施場所は、東京大学農学部 2 号館 2 階第一講義室（最大 120 名収容）とした（図 1）。この講義室は、主催機関 1 の通常の大学院講義で使用しており、100 人規模のノート PC を用いたハンズオン講義において安定した実績（無線 LAN、100 アンペア以上の電源容量）を誇るのが主な理由である。また、主催機関 1 は 90 台程度のノート PC（以下、アグリバイオ PC）を所有しており、「速習」コース受講者が持ち込む PC の各種トラブルに対しても、講義室そばに保管しているアグリバイオ PC の貸与で速やかに対応可能という利点も考慮した（図 2）。



図 1. 「速習」コース会場風景 1。講義室に向かって左後方（図 2 の白矢印付近）から撮影。白矢印付近から図 2 を撮影した。3 人掛け×10 列×4 ブロック = 最大 120 名。



図 2. 「速習」コース会場風景 2。講義室に向かって右前方（図 1 の白矢印付近）から撮影。白矢印付近から図 1 を撮影した。灰色矢印の講義室に隣接する小部屋にアグリバイオ PC が保管されている。80 アンペア分の電源容量は、この小部屋のコンセントから延長コードを用いて確保した。

2-3. 「速習」コースの受講対象、受講定員、受講条件など

受講対象は、学生・社会人問わず可能とし、共催機関 1 のウェブサイト上で受講を受け付けた。受講優先順位は以下のようにした：

1. 全日程参加可能かつ修士卒以上の方
2. 全日程参加可能な方
3. 参加可能な日程の多い方

また、以下の 3 つを受講必須条件として挙げ、チェックボックスにチェックを入れないと受講申込ができないようにした：

1. キーボード操作等（テキスト入力、マウスによるウィンドウ操作やコピー・ペースト等）PC の基本操作をスムーズに行うことが可能な方
2. 本プログラムは講習会場後方よりビデオ撮影を行い、後日公開予定です。このビデオに映る可能性があることを了承していただける方
3. 受講前および受講期間中に、受講に必要なソフトウェアのインストールやデータのダウンロードをお願いすることがありますが、これら事前準備を行っただけで受講可能な方

さらに、以下のお願い事項についても了承ボタンを押さないと受講申込ができないようにした：

1. スケジュールおよび内容は、やむを得ない事情等により変更される場合があります。

2. 万が一受講できなくなった場合は速やかに事務局に連絡をお願いします。当日講義開始 15 分後までに事前連絡なく会場にいらっしゃらなかった場合には、他の方に席を譲ることがあります。
3. パイプ椅子のみでの受講となる場合がありますので、ご了承ください。
4. 周囲に困っている人がいた場合、わかる範囲でアシストをお願いいたします。
5. アンケートへのご協力をお願いいたします。アンケート結果は後日公表する予定です。

受講申込受付サイトは5月7日に公開し、ウェブサイトやメーリングリストなどで告知を行った。受講定員は30名程度とし、応募状況によって最大80名程度まで受け付けることとした。6月2日より全日程受講希望者の受付を開始し、当初は申込状況によって7月以降に一部日程受講希望者の受付を開始する予定であった。しかし、開始初日に42名の参加申込があり、申込み開始12日後の6月14日には当初の上限である80名に達した。この中には、優先順位1である修士卒に該当しない、「修士課程在学中、学部在学中、学部卒」の受講希望者が含まれていた。しかし、事実上申込者が受付システム上で自己申告にてチェックを入れるだけの性善説に立っていること、修了証明書の提出などは非現実的であることから、自己申告で優先順位1に該当する人数が80名以上となるまで全日程受講希望者の募集を継続した。最終的に、募集開始から約3週間後の6月24日に全日程受講申込の受付を修士修了未満の学生13名を含む93名で終了し、一部日程の受講申込は行わないこととした。また、多少キャンセルが出たとしても講習会開始時点で80名以下にはならないという想定のもと、多くの要望が寄せられたものの全日程受講希望のキャンセル待ち受付は行わなかった。結果として、講習会前日までに93名中7名がキャンセルした。残りの86名の中にも、その後体調不良などの理由による一部日程の欠席連絡があったが、連絡なしの不参加者はゼロであった。受講生86名の各種内訳は以下の通りである。

所属	大学：59名 ¹⁾ 、研究所：12名、企業：12名、その他：3名
年齢	20代：36名、30代：38名、40代：7名、50代：5名
職種	学生：35名、研究：32名、教職：6名、その他：13名
バックグラウンド	バイオ系：67名、IT系：7名、その他：12名 ²⁾
Linux利用経験	全く触ったことがない：21名、使ったことはあるがよく覚えていない：22名、簡単なコマンドなら知っている：27名、時々or頻繁に使っている：16名
R利用経験	全く触ったことがない：24名、使ったことはあるがよく覚えていない：24名、簡単なコマンドなら知っている：22名、時々or頻繁に使っている：16名
NGS利用経験	今のところ使用予定はない：5名、使ったことはないが近いうちに使用予定がある：27名、いずれ使用すると思う：27名、時々or頻繁に使っている：27名
実習PC環境	Win持込：32名、Mac持込：29名、アグリバイオPC貸与希望：25名 ³⁾

表 3. 受講生の各種内訳。 ¹⁾59名のうち最も多かったのは東京大学の28名。 ²⁾その他のカテゴリには、数学、化学、医学、歯学、環境工学、バイオとIT両方、などが含まれる。 ³⁾受講申込時の希望

調査結果。持込 PC へのソフト未インストールや各種不具合対応によって、最終的に 40 台程度のアグリバイオ PC が利用された。

2-4. 「速習」コースのオブザーバー制度、Twitter について

経験上、Linux や R などのコマンドライン環境での作業経験が浅い受講生に対するハンズオン講義（ノート PC を用いた実習を含む講義）では、優秀な Teaching Assistant (TA) の確保が成否に大きな影響を与える。そのため本コースでは、受講生募集と並行してオブザーバーの募集も行った。オブザーバーは、主に日頃指導する側の立場の人を対象として、講義の仕方や指導内容を学んでもらいつつ、実習の補助を TA として担ってもらうことを期待したものである。交通費、宿泊費、TA 代などは出ないため自腹での参加であったが、福岡、岡山、広島、京都、名古屋、千葉、東京から若手を中心に 9 名の申込みがあった（主催機関、共催機関からの TA を除く）。

オブザーバー制度は、非常にうまく機能したと思われる。例えば、講師が答えられないような質問に対して、オブザーバーがその場ですぐにコメントする場面が幾度となくあった。休憩時間中に解決策を見出して、休憩後にオブザーバーが解説する場面もみられた。また、講師の講義スライドを映したスクリーンのほかに、別のスクリーンも利用してオブザーバーの PC 画面で実際のコマンド打ち込み例を示すことにも貢献した。

本コースでは Twitter (<https://twitter.com/hashtag/AJACS>) も活用された。質疑応答ですぐに解決策が見つからなかった場合でも、オブザーバー、受講生、および他の講師らが発見した解決策を #AJACS のハッシュタグつきでツイートすることで情報の共有を効率的に行うことができた。

2-5. 「速習」コースの実習用 PC 環境構築

ハンズオン講義は、大きく 2 つの形態に分けることができる。一つは主催側が用意した同一 OS、同一バージョンの統一的な PC 環境で実習を行うものであり、もう一つは受講生が普段利用している PC を持ち込んでヘテロな PC 環境で行うものである。前者は、開催側の事前準備は大変であるものの、一通りの動作確認を行ってさえいれば、本番（実習期間）中に不具合に遭遇する確率は非常に低い。受講生にとっては、提供された PC 環境で手順通りに行えば一通りの結果（成功体験）が得られるというメリットがある。デメリットとしては、普段利用する PC ではないため、受講後にいざ自分の PC 環境でやろうと思っても様々なトラブルに遭遇してうまくいかない場合が多い点である。

後者は、ラボに戻ってからも引き続き環境構築済みの同じ PC を利用できるというメリットがある。デメリットとしては、ヘテロな PC 環境で実習を行うため、本番中に様々なトラブルに遭遇する場合が多い。そして、TA 数が少ないとトラブル対応に忙殺され、講義全体がストップしがちである。結果として、自前 PC で速やかに成功体験を得ることができた受講生にとっては時間を持て余し、トラブルへの対処ができないまま実習が終わってしまった受講生にとっては不満しか残らない場合が多い。そのため、一般にこの講義形態は多くの受講生にとって不平不満の感想しか残らない傾向にある。しかしながら、様々なトラブルへの遭遇および対処法を身につけることこそが本当の実力を身につける上で肝要である。うまくいった受講生は、トラブルに遭遇している他の受講生

の状況を把握し「この OS、このバージョンの PC 環境ではこのようなことも起こりうる」という事例を経験し助け合うのが講習会を円滑に進める上で重要である。つまり、受講生側の心構えも講習会の成否に大きな影響を与えることが、「2-3.の受講条件」中にお願い事項 4.を明記している理由である。

このカリキュラム (<http://biosciencedbc.jp/gadget/chousa/generation-sequencer.pdf>) には、「速習」コースの 1-3.UNIX I の習得技術欄に「Linux 導入」がある。また、「速習以外」の 1-5. UNIX II の習得技術欄に「Windows OS 上の UNIX 環境構築」がある。つまり、自力で Linux 環境の構築ができる人材の育成を目指している。本コースでは、持込 PC を利用することを基本とし、できた受講生は TA として振る舞ってもらうことでトラブル対応時間をできるだけ短くするという方針を掲げた。できる受講生向けのアドバンスドな内容を教える時間を確保し、全体としてより多くの受講生の満足度を高めることを目指した。

通常、ヘテロな PC 環境での Linux 導入は半日程度を要する。これを避けるべく、本コースでは詳細なインストール手順を公開して、Linux を含む各種 PC 環境構築作業を事前に行っておくことを持込み PC の受講生にお願いした。OS の違いによらず統一的な GUI 環境で Linux を学べるように、NGS 解析用プログラムを含む様々な解析ソフトが一通り組み込まれた Bio-Linux 7 を採用することとした。本コースの講師も務めるアメリエフ株式会社が、Bio-Linux 7 上にサンプルデータおよび以下のプログラムを追加インストールしたイメージファイル(BioLinux.ova)の作成を担当した。

プログラム名	原著論文
FIRE	Elemento et al., <i>Mol Cell</i> , 2007
ngs.plot	Shen et al., <i>BMC Genomics</i> , 2014
DiffBind	Ross-Innes et al., <i>Nature</i> , 2012
WebLogo	Crooks et al., <i>Genome Res.</i> , 2004
SRA Toolkit	-
IGV	Thorvaldsdóttir et al., <i>Brief. Bioinform.</i> , 2013
Genome Analysis Toolkit (GATK) ver. 1.6	McKenna et al., <i>Genome Res.</i> , 2010

表 4. Bio-Linux 7 に追加インストールしたプログラムリスト。

イメージファイルおよび各種インストール手順は、主催機関 1 のウェブサイトから提供した。Bio-Linux 7 については、仮想環境構築（仮想化ソフト VirtualBox のインストール）からイメージファイルの導入までの一連の手順を OS ごと（Windows 用および Macintosh 用）に分けて提供した。講習会初日の仲里猛留先生の担当部分で、Linux 関連の環境構築確認を行ったが、確認作業開始 10 分足らずで個別対応を要する受講生が 2 名となり、次のステップへと講義を速やかに進めることができた。アンケート結果でも、詳細な手順書のおかげで無事インストールができたという記述も見られた。以下の手順書の説明内容は、自力でインストールできる一つのモデルケースといえよう。

- ・ Bio-Linux 7 関連インストール手順 (Windows 用)

http://www.iu.a.u-tokyo.ac.jp/~kadota/bioinfo_ngs_sokushu_2014/install_NGSsokushu_windows.pdf

- ・ Bio-Linux 7 関連インストール手順 (Macintosh 用)

http://www.iu.a.u-tokyo.ac.jp/~kadota/bioinfo_ngs_sokushu_2014/install_NGSsokushu_macintosh.pdf

また、R および各種 R パッケージのインストール手順についても、トラブルシューティング情報を含めて事前準備用として以下に公開した。しかしながら、Linux 関連よりも R 関連のインストール作業でつまづく受講者が多く見られた。これについては、パッケージ提供サイトの一つである Bioconductor がしばしばアクセスできないことがあること、長時間有線 LAN でネットワーク接続しておく必要があること、Macintosh の最新 OS (Marvericks) が R の最新バージョン (R ver. 3.1.0) に対応していなかったこと、64 bit PC を想定していたが持ち込める PC が 32 bit 版しかなかったこと、インストール作業自体が慣れないコマンドライン環境であるためエラー対応が比較的難しかったことなどが挙げられる。パッケージ間の依存関係やファイルサイズなどの理由により、講習会当日にその場でインストール作業を行うのは非現実的であるが、結果的に講習会当日にインストールできなかった受講生に対しては、速やかにインストール済みのアグリバイオ PC の貸与を行うことで、全員が一通りの解析を行うことができた。また、スタッフがその場で対応できるソフトについてはインストールを行い、できる限り持ち込み PC で実習できるよう努めた。

- ・ R および各種 R パッケージのインストール手順 (Windows 用)

http://www.iu.a.u-tokyo.ac.jp/~kadota/bioinfo_ngs_sokushu_2014/R_install.pdf

- ・ R および各種 R パッケージのインストール手順 (Macintosh 用)

http://www.iu.a.u-tokyo.ac.jp/~kadota/bioinfo_ngs_sokushu_2014/R_install_mac.pdf

2-6. 「速習」コースの講義内容

以下のような型式で示す。統合 TV または講義資料のリンク先がないものは、非公開または公開に向けた準備中である。講義時間は当初予定分を示す。

<ul style="list-style-type: none"> ・ 日時、「項目番号および項目」、レベル(初級または中級)、形式(講義または実習) ・ 講師 (所属) ・ 統合TVリンク先 ・ 講義資料リンク先 ・ 講義内容およびコメント
<ul style="list-style-type: none"> ・ 2014年9月1日10:40-12:00、「1-1. OS, ハード構成」、初級、講義 ・ 中村保一 (DDBJ)

• https://www.youtube.com/watch?v=cv_9yTryIv8

• http://www.iu.a.u-tokyo.ac.jp/~kadota/bioinfo_ngo_sokushu_2014/20140901_1-1_1-2.pdf

• コンピューターの基本の理解。必要なスキル。ハードウェア（計算機の歴史、現代のハードウェア構成）。遺伝研スパコンを例に。ソフトウェア（オペレーティング・システム(OS)、言語・プログラム・ライブラリ）。UNIX OSや、UNIXの構成、OSのご利益。シェル環境（bash, tcsh, zsh）。スクリプト言語。米国ではPythonが人気。コメント：講義資料自体は、項目1-1と1-2の両方を含む。講義が順調に進んだため、「1-2. ネットワーク基礎」の大半をこのコマで話した。

• 2014年9月1日13:15-14:45、「1-2. ネットワーク基礎」、初級、講義

• 中村保一 (DDBJ)

• https://www.youtube.com/watch?v=cv_9yTryIv8

•

http://www.iu.a.u-tokyo.ac.jp/~kadota/bioinfo_ngo_sokushu_2014/20140901_1-1_1-2_additional.pdf

• インターネット、セキュリティの基本の理解。ネットワーク（よく使うインターネットプロトコル）。インターネットの始まり、TCP/IP。よく使う通信プロトコル（HTTP, FTP, SSH, SFTP, SMTP/POP）。スパコンへのログインはWindowsはTera Term、Macはターミナルを起動。外部サーバとのファイル送受信（sftp, Cyberduck, Asperaなど）。DDBJ DRAの話。データ量の推移や検索の仕方、解析パイプラインも用意しており、主要な生物種ではLinuxを覚えなくてもどうにかなる。遺伝研スパコンは無料。GNU/Linux。DDBJパイプラインを用いて公開データ納豆菌をVelvetでアセンブルする話。SRAメタデータの説明。コメント：ここの講義資料は中村保一先生が他所で話されたNGS関連資料をかき集めて急遽追加で作成したものである。

• 2014年9月1日15:00-18:15、「1-3. UNIX I」、初級、実習

• 仲里猛留 (DBCLS)

• <https://www.youtube.com/watch?v=0hH3GFczAfE>

•

http://www.iu.a.u-tokyo.ac.jp/~kadota/bioinfo_ngo_sokushu_2014/20140901_1-3_nakazato.pdf

• UNIXの基礎の理解、VirtualBoxとBioLinux環境構築。UNIXはOSであり、Windows 7やMacOS Xと同じようなもの。マルチユーザー、マルチタスク、super user, root。Linuxは個人で使うレベルのUNIX。ディストリビューション（Debian系、RedHat系、OpenSuSe）。カーネル(kernel)とシェル(shell)。BioLinuxの起動と終了、画面説明、ターミナルの起動。ディレクトリとファイル。lsコマンド、オプションの使い分け。プロンプトは自在に設定可能。引数はカタマリごとに入れ替えてもよい。ディレクトリ構造。cdコマンド。ホームディレクトリは「ちるだ」、カレントディレクトリは「どっと」、一つ上のディレクトリは「どっとが二つ」、ディレクトリの区切りは「/」。絶対パスと相対パス。基本コマンド：ls, cd, pwd, mv, cp, rm, mkdir, cat, more/less, head/tail。質疑応答で様々なオプションやTipsを共有。rmコマンドの-fや-rオプション。headコマンドの-15オプションで15行分見られるとか。「CTRL + L」ボタンでターミナルをリフレッシュ。lsコマンドでワイルドカードの意味を説明。「/home/admin1409/amelieff/perl/pl_answer」

ディレクトリ上で「ls -l perl?.pl」とか「ls -l perl*.pl」での挙動の違いを確認。ゲストOS上のコマンドのコピーはホストOSの「CTRL+C」などはだめなので、右クリックでやるべし。ホストOS (WindowsやMacintosh)上のファイルをゲストOS (BioLinux)上にコピーするのはドラッグアンドドロップで可能。その逆はデフォルトでは不可。日本語を含む20140911_4-4_kawaoka.txtをBioLinux上でみると文字化けするが、下記のようにiconvコマンドで「iconv -f SJIS 20140911_4-4_kawaoka.txt > hoge.txt」で解決可能。nkfコマンドがあればそれでも解決できるらしい。コメント：順調に進んだため、9月2日分の講義を引き続いて行った。

- ・ 2014年9月2日10:30-18:15、「1-3. UNIX I」、中級、実習
- ・ 仲里猛留 (DBCLS)
- ・ https://www.youtube.com/watch?v=z5qwZ_QHk48

http://www.iu.a.u-tokyo.ac.jp/~kadota/bioinfo_ngs_sokushu_2014/20140902_1-3_nakazato.pdf

- ・ 権限(Permission)、「Read, Write, eXecute」、「ユーザ, グループ, その他」。chmod, su, sudoコマンド。~/genome/yeast.ntを用いてgrep, sort, uniqコマンドを実行。「grep GAATCC genome/yeast.nt」、「grep ^GAA genome/yeast.nt」、「grep GAA\$ genome/yeast.nt」など。パイプ(縦棒)の話。「grep GAATCC genome/yeast.nt | wc -l」などいろいろ。より高機能なegrepもある。正規表現、シングルクォーテーションとダブルクォーテーションの違い。「echo '\$SHELL' 任意のファイル名」と「echo "\$SHELL" 任意のファイル名」。「wc -l genome/yeast.nt」、「sort genome/yeast.nt | uniq | wc -l」、「sort genome/yeast.nt | uniq -c | sort -n」、「sort genome/yeast.nt | uniq -c | sort -rn | head -3」、「sort genome/yeast.nt | uniq -c | sort -rn | head -3 | less -S」。リダイレクトでファイルに保存。「sort genome/yeast.nt | uniq > genome/yeast.uniq.nt」。「sort genome/yeast.nt | uniq」、画面表示を途中で止めたい場合は「CTRL+C」。Cufflinksというプログラムをwgetして*.tar.gzファイルが存在するところからの代表的なインストール例。「tar xvzf cufflinks-2.2.1.tar.gz」、「cd cufflinks-2.2.1」、「whereis make」。今回の実習環境ではnkfコマンドが入ってなかったので、みんなで「sudo apt-get install nkf」、「apt-cache search nkf」。そして文字化けしていたファイルの文字コード変換を実行。「nkf 20140911_4-4_kawaoka.txt > hoge2.txt」。パスの話。「echo \$PATH」。デフォルトで「bwa」と打つとbwa ver. 0.6.1-r104となるが、ホームディレクトリ以下にbwa ver. 0.7.10が既にインストールされている。bwa最新版を「bwa」で利用したいときにパス指定をしたり、lnコマンドでシンボリックリンクを作成したりする。エディタの紹介。viやemacsがあるが、emacsはターミナルで利用可能な「CTRL+e」や「CTRL+a」などが共通。マルチタスクの実例。「CTRL+z」で終了、「Jobs」で確認、「bg %1」や「fg %2」など。「ssh admin1409@localhost」で、自分自身のマシンに新たにログイン。「screen」コマンド。コメント：9/2講義分は前日にほとんど終えていたためどうなるかと思ったが、かなり詳細な質疑応答が行われ、結果的に時間を持て余すどころか初日より疲労感の漂う一日であった。また、急遽仲里猛留先生が作成したLinuxコマンド集 (http://www.iu.a.u-tokyo.ac.jp/~kadota/bioinfo_ngs_sokushu_2014/20140902_1-3_nakazato_linux.pdf) や エディタのキーバインド集

(http://www.iu.a.u-tokyo.ac.jp/~kadota/bioinfo_ngo_sokushu_2014/20140902_1-3_nakazato_editor.pdf)はこの後の講習会全体でも使えるものであり、A4一枚で手軽に利用できる点が非常に好評であった。

- ・ 2014年9月3日10:30-18:15、「1-4. スクリプト言語 (シェルスクリプト)」、中級、実習
- ・ 服部恵美 (アメリエフ)
- ・ <https://www.youtube.com/watch?v=MJ-vUCTIirM>

http://www.iu.a.u-tokyo.ac.jp/~kadota/bioinfo_ngo_sokushu_2014/20140903_1-4_shellscript.pdf

・ シェルスクリプトの基礎 (変数、引数、条件付き処理、繰り返し処理、標準出力と標準エラー出力、シバンなど)。変数を使うと値を一元管理できる。変数にしたほうがいいのは、入力ファイル名 (リファレンスゲノム、塩基配列のファイル名) とか。変数名に使わないほうがいいものは「変数名 予約語」などで検索するとよい。引数を使うとスクリプトの中身をいちいち書き換えなくてよくなる。\$1や\$2以外に「\$0」もある。例えば、\$0はUsageのところなどで利用される。条件付き処理でmkdir。複数の条件を指定したいときは「-aや-r」を利用。bashのif文で使用する-eの意味は、指定した名称のディレクトリまたはファイルが存在するかである。「bash -n ファイル名」でシェルのコーディング確認ができる。echoで実行コマンドを出力させて結果をわかりやすくする。標準出力と標準エラー出力。「>&2」が標準エラー出力。「bash test7.sh 1>log 2>err」は、標準出力はlogというファイルに出力し、標準エラー出力はerrというファイルに出力する。プログラム内部を理解しやすくするため、スクリプト中に#から始まるコメントを入れた方がわかりやすい。「シバン」というのは、一般的にプログラムの一行目を書く。このプログラムはシェルスクリプトである、ということを示す場合は「#!/bin/bash」と書く。こうすることで例えば「bash test.sh」と書く代わりに「./test.sh」で済む。sequence logos (weblogo)の紹介をして、mature.faやhairpin.faを入力としてレベルに応じた最終課題(mkdir, grep, weblogoなど)。コメント: アメリエフ株式会社は、トータル10日のうち3.5日分を担当しました。当然ながら、少なくともこの3.5日分の講義資料は統一感があり好評であった。長丁場ではあるものの、意思統一のできる少人数の講師陣でできるだけ多くのコマ数を担当するのが受講生にとってはよいと思われる。

- ・ 2014年9月4日10:30-18:15、「1-4. スクリプト言語 (Perl)」、中級、実習
- ・ 服部恵美 (アメリエフ)、 統合TV、講義資料 (20140908, 09:15)
- ・ <https://www.youtube.com/watch?v=IYMYIujHrhM>

http://www.iu.a.u-tokyo.ac.jp/~kadota/bioinfo_ngo_sokushu_2014/20140904_1-4_perl.pdf

・ Perlの基礎 (変数、配列、引数、ハッシュ、条件付き処理、繰り返し処理、ファイル入出力、シバン、正規表現など)。一般に、Perlのほうがシェルスクリプトよりも複雑な処理に向いていると言われる。例えば、ファイル形式の変換や結果ファイルの独自解析など。達成目標は、hairpin.faの特定のRNAの配列を切り出して別のファイルに書き出すPerlプログラムを書けるようになる。Perlの記載方法。値を出力するにはprintを実行、文字列はダブルクォートかシングルクォートで囲む、行の末尾に「;」をつける、全角記号や全角空白文字は使えない。シェルスクリ

プト同様、Perlでも「変数」を使うことができる。「\$変数」と書くと、変数に入っている値を呼び出すことができる。最初に変数が出てくるときには「my」をつける。シェルスクリプトと違って、代入の「=」の前後に空白が入っても大丈夫。Perlのprintはシェルスクリプトのechoと違って最後が改行されない。そのため、改行したい場合は明示的に「\n」と改行コードを書く必要がある。「配列」を使うと、たくさんの値をまとめて扱うことができる。配列からの値の取り出し方。「添字」は0から始まる。push関数で配列に値を追加。配列の要素数の調べ方(Rのlength関数に相当)。配列を結合して文字列にするのはjoin関数。文字列を分割して配列にするのはsplit関数。Perlでも引数が使え。Perlでは、引数を「@ARGV」という名前の専用の配列で受け取る。ハッシュの話。「\$ハッシュ名{キー}」でハッシュから値を取り出す。配列とハッシュの違い。Perlでも条件付き処理や繰り返し処理が可能。シェルスクリプトと書き方が似ているが微妙に異なるので注意。条件付き処理は、if-elsif-else構文を使用。Perlの比較演算子。複数の条件指定。Perlの繰り返し処理にはforeach, for, whileなどがある。繰り返し処理でハッシュにアクセス。繰り返し処理の中で次の要素スキップするにはnextを使用。繰り返し処理自体を中止するにはlastを使用。{}囲んだ範囲をスコープと呼ぶ。ファイル入出力「use strict;」と「use warnings;」でバグを見つけやすくする。#で始まる行はコメント扱いとなり、処理に影響しない。Perlプログラムの1行目に「#!/usr/bin/perl」を記述すると、このファイルがPerlスクリプトであることが明示的になる。正規表現は文字列のパターンを表現する方法であり、Perlでは、正規表現を//で定義する。「~/正規表現パターン/置換文字列/オプション」で、正規表現を用いた文字列の置換を行うことができる。最終課題は、hairpin.faの塩基配列行を1行にするなど。コメント：今まで一度も触れたことのない初心者から非常に好評であった。NGS解析に直接リンクした例題があると尚よかったという意見があった。

- ・ 2014年9月5日 10:30-14:45、「2-1. 配列解析基礎」、初級、実習
- ・ 坊農秀雅 (DBCLS)
- ・ <https://www.youtube.com/watch?v=Zw91OjJGR18>
- ・ http://www.iu.a.u-tokyo.ac.jp/~kadota/bioinfo_ngo_sokushu_2014/20140905_2-1_bono.pdf
- ・ 配列、ゲノムデータ記述のフォーマット、アラインメント(DP)、データベース検索(BLAST、BLAT)等の基礎的な配列比較解析の原理と実習。コメント：相同性と配列類似性の違いからNGS解析への橋渡しの講義内容であったが、前日までの内容から一気にバイオ寄りになったため、専門用語についていけない受講生も何人かいたようである。来年度以降は、NGS解析やこの項目などで頻出する用語をリストアップしておき、この講義あるいは講習会前に一通り眺めてもらうほうがいいのかも。また、gitがBio-Linux 7で使えなかったため、gitのインストール手順も示しておく必要がある。

- ・ 2014年9月5日 15:00-18:15、「2-2. バイオ系データベース概論」、初級、実習
- ・ 小野浩雅 (DBCLS)
- ・ https://www.youtube.com/watch?v=G6E_I3UG3vY
- ・ http://www.iu.a.u-tokyo.ac.jp/~kadota/bioinfo_ngo_sokushu_2014/20140905_2-2_hono.pdf
- ・ 基本的な各種バイオ系データベースの理解、統合DBの利用法。Google Chromeだと、統合TV

を倍速指定で見ることができる。UCSC Genome Browserでwigファイル関連、httpでうまくいかない場合はftpでやるべし。UCSC Genome Browser, Biomart, IGVの統合TV番組を見せながら概説。Integbio データベースカタログ、生命科学データベース横断検索、生命科学系データベース アーカイブ。統合データベース講習会 AJACS, MotDB、新着論文レビュー、領域融合レビュー、Allie、inMeXes、RefEx、GGRNA、GGGenome、diff、DBCLS SRA、ライフサイエンスQA、togo picture gallery。Allieは、NGSなど略号が最初に出てきた論文を知ることができる。inMeXesでは、文献に頻出する英語表現を、一文字の入力ごとに高速に検索。前置詞の出現頻度を調べる場合などに便利。ライフサイエンス辞書オンラインサービスと連携しているらしい。コメント：様々なDBを総花的に紹介する内容であったが、受講生にとってはいくつか存在する有用なDBやツールを知ることができて有意義だったようである。統合TVの3番組を1時間ほどかけて自分で眺める時間があったが、実習時間としてわざわざ確保しなくてもよいという意見が少なからず存在した。これは、統合TVがあまりにも有名であり、存在を知らないヒトがほとんどいないということに起因するのであろう。

・2014年9月8日10:30-12:00、「3-1. R 基礎1」、初級、実習

・門田幸二 (東京大学)

・<https://www.youtube.com/watch?v=7PM7qsmVNZk>

・http://www.iu.a.u-tokyo.ac.jp/~kadota/bioinfo_ngo_sokushu_2014/20140908_3-1_kadota.pdf

・Rインストール自体は基本的に終了した状態を想定しているものの、最初にlibrary(Biostrings)などいくつかの利用予定パッケージのロードを行い、パッケージのインストールがうまくいっているかどうかを確認（できていなかったヒトの同定および対処）。Rの一般的な利用法。log関数などの基本的かつ挙動を完全に把握できる関数を例として、関数内部のオプション変更や「?関数名」で利用法の幅を広げる基本テクを概観。exp, mean, median, sort, length関数。

・2014年9月8日13:15-14:45、「3-2. R 基礎2」、初級、実習

・門田幸二 (東京大学)

・<https://www.youtube.com/watch?v=7PM7qsmVNZk>

・http://www.iu.a.u-tokyo.ac.jp/~kadota/bioinfo_ngo_sokushu_2014/20140908_3-2_kadota.pdf

・翻訳配列の取得を例に「(Rで)塩基配列解析」の基本的な利用法を紹介。塩基配列中にNを含む場合のエラー例とその対処法。RGui画面中のSTOPボタンやRの終了手順。行列形式のタブ区切りテキストファイルからの情報抽出。Perlのハッシュに相当する部分。ありがちなミスとその対処法。入力ファイルのヘッダ行の有無とread.table関数を利用した読み込み時のheaderオプションの使い分け。論理値ベクトルの理解：is.element関数。集合演算：union, intersect, setdiff関数。その他：sort, table, toupper, tolower関数。

・2014年9月8日15:00-18:15、「3-3. R 各種パッケージ」、中級、実習

・門田幸二 (東京大学)

・<https://www.youtube.com/watch?v=7PM7qsmVNZk>

・http://www.iu.a.u-tokyo.ac.jp/~kadota/bioinfo_ngo_sokushu_2014/20140908_3-3_kadota.pdf

・Rの各種パッケージのインストール法と代表的なパッケージBiostringsの利用法。library,

search, objects, ls関数。作業スペースの概念。Biostringパッケージで利用可能な関数を概観 (Bioconductorのイントロダクション)。 source関数の利用 : Rでもシェルスクリプトっぽいことができる、他人がウェブ上で公開している関数の読み込み、 Bioconductor提供パッケージインストールのためのbiocLite関数。

- 2014年9月9日10:30-14:45、「3-4. R bioconductor I」、中級、実習
- 門田幸二 (東京大学)
- https://www.youtube.com/watch?v=_H95NWHeD9Y
- http://www.iu.a.u-tokyo.ac.jp/~kadota/bioinfo_ngs_sokushu_2014/20140909_3-4_kadota.pdf

• setwd関数利用で効率的に解析。塩基配列を入力としてアミノ酸配列(翻訳配列)を取得する translate関数を用いて、アミノ酸配列を入力として実行しエラーを出させることを通じて、データの型の基本概念を説明。 rm, ls, objects関数を用いたオブジェクトの消去および確認。「?translate」で引数(Arguments)部分の説明。翻訳配列取得コードの中身を説明。DNAStrngSet形式、width列とnames列の説明。自分のR環境を把握し、バージョンの違いが結果の違いとなることを認識。 sessionInfo関数、昔のバージョンのRインストール法。定期的なバージョンアップの重要性。 TCCパッケージを例として関数名が変わりうることやBioconductorサイトの見方を説明。 Biostringsパッケージのread.DNAStrngSet関数がreadDNAStrngSet関数に変わっていること。 Bsgenome.Hsapiens.UCSC.hg19パッケージで、R ver. 3.1.0で上流配列取得時に警告メッセージが出ており、将来的には使えなくなること。 Bioconductorのサイトを概観。いくつかのカテゴリを眺めて、CRISPRseekパッケージを発見することなど。

- 2014年9月9日15:00-18:15、「3-5. R bioconductor II」、中級、実習
- 門田幸二 (東京大学)
- https://www.youtube.com/watch?v=_H95NWHeD9Y
- http://www.iu.a.u-tokyo.ac.jp/~kadota/bioinfo_ngs_sokushu_2014/20140909_3-5_kadota.pdf

• multi-FASTAファイルからの情報抽出。GC含量計算の詳細な説明。alphabetFrequency, apply関数、行列演算の基本。 FASTQ形式ファイルの読み込み、ファイル形式の変換 (FASTQからFASTA)、クオリティチェック(QC)。クオリティスコアでのフィルタリング、動作確認用のサブセット作成。FASTA/FASTQファイルのdescription行の整形。 sample, set.seed関数、gzip圧縮ファイルの読み込みや書き出しが可能であることなど。

- 2014年9月10日10:30-12:00、「4-1. 次世代シーケンサ基礎I」、初級、講義
- 倉田哲也 (NAIST)
- 統合TV (非公開)
- 講義資料 (非公開)

• 高速シーケンサ概略、シーケンサの歴史、サンガー法。各種高速シーケンサの原理。 454 Life Sciences社のemulsion PCR、Illumina社のbridge PCRの説明。single-end readとpaired-end read。pyrosequencingの原理。ピコタイタープレート (picotiter plate;)、生データの画像イメージから数値化 (デジタルデータファイル; composite wells format; cwf) して塩基配列ファイル (.fna)とクオリティスコア情報のファイル(.qual)を作成。Illuminaのsequence by synthesis (SBS

法)の説明。phasingとprephasingの補正、クラスターの純度を示す指標quality filtering (chastity filtering)、クオリティスコアの説明や昔はいくつかのバリエーションがあったことなど。マルチプレックス(多サンプルをシーケンスしたいときや、1レーンに複数サンプルを混ぜてシーケンスしたいとき)。SOLiDのtwo-base encodingを含む原理の説明。Ion Torrent Personal Genome Machine (PGM)は蛍光を使わず塩基の取り込み時のpH変化を検出するため、高速、低コストでシーケンスが可能であることなど。シーケンス用ライブラリ調整の話。

- ・2014年9月10日13:15-14:45、「4-2. 次世代シーケンサ基礎II」、初級、講義
- ・倉田哲也 (NAIST)
- ・統合TV (非公開)
- ・講義資料 (非公開)
- ・応用分野とそのための計測技術の理解 (RNA-seq, ChIP-seq, がんゲノム, 個人ゲノム, 環境ゲノム, Hi-C)。シーケンスデータの情報解析の流れ。リファレンス配列へのマッピング、マップされたリード数の計測、アセンブリ、統計解析ソフト、ゲノムブラウザなど。Re-sequencing (SNP解析), de novo sequencing (ゲノム配列決定), bisulfite sequencing (メチル化)データ特有のマッピング問題など。False discovery rate (FDR)の概念説明。RNA-seqはマイクロアレイとの比較や、directional RNA-seq、各種統計、Gene Ontology (GO)解析、single cell RNA-seqなどの説明。ChIP-seq, 3C, Hi-C, メタゲノム解析。第3世代シーケンサー(PacBioやナノポア)の原理説明。

- ・2014年9月10日15:00-18:15、「4-3. 次世代シーケンサ実習I」、初級、実習
- ・山口昌雄 (アメリエフ)
- ・<https://www.youtube.com/watch?v=s83cHzvydUM>
- ・http://www.iu.a.u-tokyo.ac.jp/~kadota/bioinfo_ngs_sokushu_2014/20140910_4-3_NGS.pdf
- ・ファイル形式、データの可視化、データのクオリティチェック、マッピング、アセンブル。NGS解析でよく使われるファイル形式は、fastq, bam/sam, vcf, bed, fasta形式。fastqは4行で1リードを表し、クオリティスコアは「記号のASCIIコード - 33」と対応する。sam/bamはリードをリファレンス配列にマッピングしたアラインメント情報を含み、samはテキストデータ、bamはそのバイナリ版である。samtoolsコマンドでsamからbamに変換できる。vcfファイルは変異情報を含み、#で始まるヘッダ行と、1行に1つの変異の情報がタブ区切りで記載されている。bedは、主にゲノム情報の領域情報を含み、エクソームシーケンスなどのターゲットシーケンスで解析範囲を指定するために用いられるほか、ChIP-seqで検出されたピークを示すのに用いられる。bamToBedコマンドでbam形式からbed形式に変換できる。データの可視化手段として、IGVを説明。Bio-Linux上でIGVを起動し、インデックスの作成から、リファレンスゲノムの選択、および可視化まで。データのクオリティチェックとしてFastQCの説明および実行。マッピングの概念説明、および分野(Reseq, RNA-seq, Methyl-seq)ごとに使われる代表的なプログラム。用途(ゲノム用、トランスクリプトーム用)ごとのアセンブルプログラムのリストアップ。

- ・2014年9月11日10:30-14:45、「4-4. 次世代シーケンサ実習II」、初級、実習
- ・河岡慎平 (ATR)

• <https://www.youtube.com/watch?v=USESqse-kHg>

•

http://www.iu.a.u-tokyo.ac.jp/~kadota/bioinfo_ngs_sokushu_2014/20140911_4-4_kawaoka.pdf

• ChIP-seqの代表的なパイプラインについての実習。iPS細胞の山中因子(OSKM factor; Oct4, Sox2, Klf4, and c-Myc) のヒトゲノム上での局在を調べたというChIP-seqデータ(Soufi et al., Cell, 2012) の取得、SRA toolkitを用いたファイル形式の変換(SRA形式からFASTQ形式)。“Input”の重要性。FastQCを用いたクオリティチェック、ゲノムへのbowtie2でのマッピングおよび結果の解釈。複数個所にマップされるリードの取り扱い。grep, wcなどのLinuxコマンドを駆使したフィルタリングや結果の概要把握。samtoolsを用いたファイル形式の変換(SAMからBAM)。ngs.plotを用いて任意の座標におけるマップされたリードの分布を調査。MACS14を用いたピークコーリング(peak calling)。UCSCゲノムブラウザで結果を眺める。Wigファイル、BEDファイル。モチーフ解析。

• 2014年9月11日15:00-16:30、「6. 分子生命科学」、初級、講義

• 河岡慎平 (ATR)

• <https://www.youtube.com/watch?v=v8I3uM9WDA>

• http://www.iu.a.u-tokyo.ac.jp/~kadota/bioinfo_ngs_sokushu_2014/20140911_6_kawaoka.pdf

• 6-1. 分子生命科学概論 (複製、転写、翻訳、代謝、シグナル伝達などの基礎知識)。 6-2. オミクス概論(ゲノム以外のオミクスデータの基礎知識)。 6-3. 遺伝/進化概論(ゲノムデータを扱う上での遺伝学、進化学の基礎知識)。

• 2014年9月11日16:45-18:15、「5-1. ゲノム情報倫理概論」、初級、講義

• 箕輪真理・川嶋実苗 (NBDC)

• <https://www.youtube.com/watch?v=Np9WD3lsFPU>

• http://www.iu.a.u-tokyo.ac.jp/~kadota/bioinfo_ngs_sokushu_2014/20140911_5-1_minowa.pdf

• ゲノム情報を扱う上で、プライバシー保護などの必要な倫理的問題、法的问题の国内外の状況を理解し、ゲノム情報を適切に利用できるようにする。匿名化、暗号化、情報セキュリティ概要。

• 2014年9月12日10:30-14:45、「4-4. 次世代シーケンサ実習II (Reseq解析)」、初級、実習

• 山口昌雄 (アメリエフ)

• <https://www.youtube.com/watch?v=t1cctdz8eXM>

•

http://www.iu.a.u-tokyo.ac.jp/~kadota/bioinfo_ngs_sokushu_2014/20140912_4-4_NGS_reseq.pdf

• Reseqの一般的な解析手順 (公開データ取得、クオリティコントロール、マッピング、変異検出)。Reseqの一般的な解析手順として、公開データ取得、クオリティコントロール(QC)、マッピング、変異(SNVとIndel)検出を概観。BreakDancer, VariationHunter, Pindel, CREST, DELLYなどのプログラムが存在する。QCからマッピングは、FASTQファイルを入力として「QCleaner (QC) -> BWA (マッピング) -> Picard (重複するリードの除去) -> GATK (リアライメントおよび塩基クオリティのリキャリブレーション)」でマッピング結果のBAMファイルを得る。マッピング結果

から変異検出は、「GATK (SNV/Indel検出およびフィルタリング) -> snpEff (アノテーション)」までで得たVCFファイルを入力としてQMergeを実行することでGenotypeファイルを得る。例題データはIllumina iGenomesから取得した酵母ゲノムファイル。NGSデータはERR038793。paired-endのbzip2圧縮ファイルの解凍から最初の1,000リード分を抽出し、一連の解析を実行。

- ・ 2014年9月12日15:00-18:15、「4-4. 次世代シーケンサ実習II (RNA-seq解析)」、初級、実習
- ・ 山口昌雄 (アメリエフ)
- ・ <https://www.youtube.com/watch?v=jRPDNwJ9x-o>
- ・

http://www.iu.a.u-tokyo.ac.jp/~kadota/bioinfo_ngs_sokushu_2014/20140912_4-4_NGS_RNAseq.pdf

- ・ RNA-seqの一般的な解析手順 (公開データ取得、クオリティコントロール、マッピング、発現定量)。

2-7. 「速習」コースの講義風景

講師ごとの講義風景を示す。2週間連続ということでも脱落者が相当数出る可能性も想定したが、最初から最後までほぼ全員出席していることがわかる。



図 3. 中村保一 講師



図 4. 仲里猛留 講師



図 5. 服部恵美 講師



図 6. 坊農秀雅 講師



図 7. 小野浩雅 講師



図 8. 門田幸二 講師



図 9. 倉田哲也 講師



図 10. 河岡慎平 講師



図 11. 箕輪真理 講師



図 12. 山口昌雄 講師

2-8. 「速習」コースのアンケート結果

本コースでは、講義項目（担当講師）ごとの難易度や講義内容に関するアンケート、およびコース全般に関する質問を行った。各種追跡調査を行うことを予定していたため、記名の有無および提出媒体（紙または電子）を選択可能とし、受講生 86 名のうち 51 名（記名回答者は 46 名）がアンケートに回答した。アンケート最後の設問は、「追跡調査に協力する」が 43 名、「協力しない」が 6 名、回答なしが 2 名であった。「追跡調査に協力する」とした無記名回答者が 1 名いたため、実際に追跡調査可能な人数は 42 名である。実際の電子版アンケートは以下の URL よりダウンロード可能である。

PDF 形式：http://www.iu.a.u-tokyo.ac.jp/~kadota/bioinfo_ngs_sokushu_2014/questionnaire_2014.pdf

Word 形式：http://www.iu.a.u-tokyo.ac.jp/~kadota/bioinfo_ngs_sokushu_2014/questionnaire_2014.docx

以下では、講習会全体の Q0～Q5 までのアンケート結果を示す。自由記述方式の Q6 および Q7 の結果は長文もあるため、付録に示す。

Q0. どのようにしてこの講習会を知りましたか。（当てはまるもの全てに☑をつけて下さい。）

友人・知人等：20名¹⁾

アグリバイオインフォマティクス（Rで塩基配列解析を含む）：12名

メーリングリスト：6名²⁾

SNS：5名

キーワード検索の結果：4名

分子生物学会：3名 NBDC：3名

¹⁾ 20名のうち3名は「先生からの紹介」と明記。²⁾ 6名のうち2名は「NGS-field」と明記。これ以外は、CBRC：1名、DBCLS：1名、生物物理学会：1名、その他（東大農学部キャンパス内の張り紙）：1名。複数回答可能としたため、総数はアンケート回答人数（51名）を超える。

Q1. 速習コースを2週間で実施いたしました、期間は適切でしたか？（3択）
適切：39名 ¹⁾ 長い：8名 ²⁾ 短い：2名

¹⁾ 1名が「しかし結構きつかった」とコメント。²⁾ 1名が「東北大からの参加は大変でした」とコメント。総数（49名）がアンケート回答人数（51名）以下なのは、アンケートの全項目に回答がなされているわけではないためである。

Q2. 講義、実習の内容から考えて、受講人数は適切でしたか？（2択）
適切：45名 多い：4名 ¹⁾

¹⁾ 1名が「海外の同様なコースは、～80名まで制限する」とコメント。

Q3. 速習コースの内容量は適切でしたか？（3択）
適切：39名 多い：5名 少ない：5名

Q4. 速習コースのカリキュラム（習得技術項目）は、適切でしたか？（3択）
適切：38名 他の習得技術項目を追加した方がよい：10名 不要な習得技術項目があった：6名

複数回答可能としたため、総数はアンケート回答人数（51名）を超える。

追加希望項目は、R 関連：3名、プログラミング言語 Python：2名、de novo アセンブル：2名であった。以下は実際のアンケート記載内容（一部抜粋）である。

- バイオ系：Rによるグラフ化？グラフ化できる形へのデータの改変+メモリ量を意識した計算速度を早くするための工夫のしかた。参加者の様々なバックグラウンド・習得状況の違いを考えた中では適切なカリキュラムだったとおもいます。個人的には速習であれば、もう少し中級の内容まで踏み込んで早くプロのバイオインフォマティクンに近づける内容だとうれしく思います。
- バイオ・医学系：えを書くだけでなく統計解析もあるとありがたい

- バイオ系：発現変動遺伝子群の全体的特徴を把握する方法についても、もう少し多く加えていただけますと有り難く存じます。
- IT系：Python とエピゲノム解析
- バイオ系：Python
- バイオ系：計算時間がネックになると思いますが…de novo アセンブルについての実習もほしかった
- バイオ系：系統ゲノム解析、比較ゲノム、分子進化、シーケンスアセンブル
- バイオ系：環境構築の方法
- バイオ系：できれば metagenome も…
- バイオ系：モデル生物以外の解析の進め方の例など。

不要だと思う項目は、講義のみの科目が多くを占めた。バイオ系であったとしても速習コース初日のコンピュータリテラシー関連の心得があるものにとっては不要かもしれない。またゲノム倫理関連は、事実上ヒトを取り扱わない受講生にとっては無関係であることも影響しているのであろう。

- バイオ・医学系：生命倫理、コンピューターとネットワーク、分子生命科学
- バイオ系：ChIP シークエンスは勉強にはなったのですが、論文は読むことはあっても実際行う予定がなかったので、実習か講義かの選択制にさせていただけるとありがたかったです。
- バイオ系：ヒトを扱っていない方も多いと思います。ゲノム倫理などは自習でも良いのではないかと感じました。
- バイオ系：ゲノム倫理
- IT系：コンピュータリテラシーとサーバー設計。IT系で何らかのプログラミング言語を習得してきているものには自力で十分習得できる内容なので不要。
- バイオ系：インターネットやセキュリティ基礎

Q5. 今回の速習コース全般について、満足度をご記入ください。(不満足1～満足5の5段階評価)
5 : 26名
4 : 20名
3 : 0名
2 : 0名
1 : 0名

アンケート無記名回答者の5名中3名がこの項目に回答し、全員が満足度4であった。記名回答者46名中43名がこの項目に回答し、満足度5が26名、満足度4が17名であった。

2-9. 「速習」コースの総評

謝辞

本「速習」コースは、2週間にわたる長丁場のハンズオン講義であった。それゆえ、どれだけのニーズがあるか？引き受けてくれる講師がいるか？どれだけの脱落者が出るか？PC 関連問題は発

生しないか?など、どう転ぶか全く見通せない状態での文字通り「試行的実施」であった。通常の数十分程度の講演や90分1コマの講義とは異なり、NGS用カリキュラムに沿った講義資料作成からのスタートであったにもかかわらず、快く引き受けてくださった講師の先生方に深く感謝したい。講師が見つからず開催できない項目もあるのではないかと危惧していたが、中村保一先生、仲里猛留先生、小野浩雅先生の3氏に声をかけ、とりまとめていただいた坊農秀雅先生の功績は非常に大きい。一番の懸案であった1,2日目の講師が4月中旬までに早々と決まったおかげで、並行して話を進めていた倉田哲也先生、河岡慎平先生、箕輪真理先生、川嶋実苗先生以外の残り枠をアメリエフ株式会社(山口昌雄先生、服部恵美先生)が引き受ける形で講師選定問題が一気に決着した(4月16日確定)。アメリエフ株式会社は、本コースで用いたBio-Linux 7のカスタマイズ、およびイメージファイルの作成でも尽力いただいた。川嶋実苗先生は主に本コースの講義風景撮影部隊として、小野浩雅先生は本コースの講義風景撮影および「統合TV」での映像公開全般の取り仕切りで講義以外でも尽力いただいた。

受講申込関連業務は、ノウハウのある共催機関1のHPCI人材養成プログラム(杉原氏、寺田氏、坂井氏)が一手に引き受けた。受講受付システムの構築、受講申込状況報告など緊密なやりとりのおかげでウェブ掲載情報の更新が速やかに行うことができた。3氏の尽力に感謝したい。

本コースでは、一人当たりの担当時間が丸二日にわたる講師も存在した。そのため、講義資料作成が間に合わない、あるいは体調不良によって当日の講義が不可能という状況も想定した。万が一の事態への備え(補欠講師)や若手講師への講義内容のアドバイス要員として支えて下さったイルミナ株式会社の鈴木健介氏、東京大学の鈴木穰先生および新村芳人先生にも感謝したい。

今回は全日程参加およびアンケートに回答するという縛りを設けてはいたが、受講生86名のうち35名がアンケート未回答であった。これらの受講生の感想は不明であるが、一般に講義内容に不満があれば、独学のほうが有意義だと判断して欠席するのが世の常である。しかし講義風景写真からもわかるように、コース終盤の体調不良による事前連絡済みの欠席者数名を除くと、無断欠席者は実質数名以下だったと思われる。今回のような大規模な講習会の場合、(想定通りではあったものの)もし実質TAとして手助けしてくれるオブザーバーを一定数確保できなかったなら、ここまでの出席者数を維持できなかったであろう。実際、後述するアンケートQ6.の回答においても、非常に多くの受講生から好意的なコメントが寄せられた。9名の公募オブザーバー(TA)諸氏に深く感謝したい。9名中8名は、研究代表者との面識すらなく文字通りのガチ公募であったが、主催者側の意図通り手を挙げた受講生への行き届いたケアが行われていた。

受講期間、人数について

2週間程度という受講期間は、適切という評価が多く妥当だと思われる。当初、受講可能な「職種」は、比較的時間にゆとりのある学生に限定されるのではという懸念もあった。実際には、86名の受講生のうち学生が35名と、むしろ学生以外の受講生数(51名)のほうが多かった。また、大学や研究機関が71名であったが、企業からの参加者も12名いた。受講申込締切り後のキャンセル待ちに関する問い合わせもいくつかあり、この種の講習会の需要は多いと判断できる。

受講生80名程度(TAを含めて90名程度)で開催した受講人数については、49名中45名が適切

と回答した。これは 80 名という人数だけでなく、講習会会場の規模との兼ね合いによるものだと思われる。今回使用した会場は、図 1 や図 2 に示すように最大 120 名収容可能であった。にもかかわらず最大 80 名程度に制限したのは、図 4 の仲里猛留講師の講義風景からもわかるように 3 人掛けの真ん中部分を空けることで、荷物置き場や持込 PC 以外の貸与 PC での動作確認作業用のスペース確保を考慮したためである。このように比較的ゆとりのある座席配置にしたことが、ほとんどの受講生が不満を持たなかった理由であろう。比較のため、主催機関 1 (アグリバイオインフォマティクス教育研究プログラム) による同会場での通常の大学院講義風景を図 13 に示す。最大収容人数ぎりぎりまで受け入れていた場合の受講風景はこのようになる。もし 2 週間連続の講義がこの状態で行われていたら、相当の不満が出ていたであろう。



図 13. アグリバイオインフォマティクスの大学院講義風景。3 人掛けの真ん中の列も受講生ではほぼ埋まっていることがわかる。

少数派意見として、49 名中 4 名が「受講人数が多い」と回答したが、これは事実上無視できるレベルであろう。共催機関 1 が実施している講習会 (<https://hpci.cbrc.jp/modules/tutorial/>) は、受講人数を 40 名に限定している。他の NGS 系講習会も 40 名以下が一般的である。もし本コースにおいても先着 40 名を厳密に適用していたら、「受講人数が多い」と回答した 4 人中追跡可能な 3 名全てが受講資格を与えられなかったことを確認済みである。

受講定員を一般的な 40 名以下にせず 86 名まで引き上げたのは、事実上研究代表者の独断である。しかし、もし本コースを定員 40 名での開催としたら、今回受講できた人数の半分以下しか受けられなかったことになる。開催にかかる諸費用、NGS 解析分野の進展、実際のアンケート結果などを勘案しても、大人数に対してタイムリーに受講機会を与えることができたという点において正しい判断であったと思われる。

講習会会場、開始・終了時間について

本コースは、9月1日～12日 10:30-18:15、東京大学農学部キャンパス内で行った。図13で示すように、研究代表者が通常利用している講義室にした。地方のホテルか旅館で行うという案も出たが、実績のある通常利用講義室を放棄してまで、未知会場での開催による不安要素（ネットワーク環境、PCの不具合対応、電源容量の問題など）を増やす選択肢が研究代表者にはなかったためである。最寄り駅（地下鉄南北線東大前駅）徒歩1分という交通アクセスのよさも、2週間連続で通う受講生にとっては重要なポイントであろう。首都圏の自宅から通う受講生に配慮して開始時間を若干遅めの10:30スタートにしたことを好意的にコメントする受講生も存在した。終了時間は18:15ではあったが、実際には少し早めの18:00前に終了した。それでもコース終盤は多くの受講生に疲労感が漂っていたことから、このように若干ゆとりを持たせた時間配分で正解だったと思われる。

ネットワーク環境は、アグリバイオインフォマティクスの大学院講義と同じ「無線ルーター（Buffalo Air Station Pro）3台体制」を提供した。本コースにおいても終始安定していたことから、強力な無線LAN環境を提供できるのであれば、有線LAN環境構築は必要ないと思われる。しかし、80名程度の大人数で一度に特定のウェブサイトアクセスするのは、攻撃とみなされかねない。これを避けるため、できるだけ時間差をつけてアクセスするよう徹底した。

電源容量は120アンペア分を確保した。本コースで貸与したPC（アグリバイオPC）は、アグリバイオインフォマティクスが保有するPanasonic社のLet's noteシリーズである。このシリーズは、PC1台あたり消費電力が最大でも100W以下、つまり $100W/100V = 1A$ 以下ということになる。理論上、このPCのみだと120人が同時利用しても問題ない。しかし本コースでは、持込・貸与PCを問わず、実習時間の半分程度を目安になるべくランダムにコンセントを利用して充電するようにお願いした。これは、一般に古いPCほど消費電力が大きいこと、受講生の持込PCのスペックを把握しきれないこと、携帯電話やスマートフォンなどPC以外の電子機器の充電がなされる可能性や、20Aのコンセントごとの負荷分散の仕組みを完全には達成できないことなどを考慮したためである。

図14は、コンセントの配置図である。受講生2人で1つのACタップ（4個口～7個口）を利用できるようにし、万が一コードを引っかけてPCを落としてしまった場合でも、それが実質的に自身のPCになるようにした。また、図6などからもわかるように、通路にコードがなるべくはみ出さないようにした。結果として、講習会期間中にそのようなトラブルは起きなかった。



図 14. コンセントの配置。3人掛けの机 1 つにつき AC タップ 1 個を配置した。

PC の事前準備について

上記項目とも関連するが、今回受講者が大人数であるにもかかわらず無線 LAN 環境で滞りなく講習会を終えることができたのは、Bio-Linux 7 を利用したコマンドライン環境での実習としたことが大きいと思われる。全 10 日間で 5.5 日分の実習形式の講義の大半が、ネットワークを経由せずローカル PC 環境で行えたからである。2-5. 「速習」コースの実習用 PC 環境構築でも述べているように、今回は、Linux を含む各種 PC 環境構築作業を事前に行うことを持込み PC の受講生にお願いした。NGS 用カリキュラムに厳密に従えば、本来この作業は「速習」コース内で行うものであり、「速習以外」の 1-5. UNIX II の習得技術欄にも「Windows OS 上の UNIX 環境構築」と記載されている内容でもある。

通常、ヘテロな PC 環境での Linux 導入は半日程度を要し、データ解析環境の構築自体が多くのユーザにとって乗り越えることのできない壁である。そのため、詳細な手順を公開して自力でのインストール作業を受講生に要求しつつも、当初は Linux 関連の PC 環境構築ができているかどうかの確認を行うところ（講習会初日の仲里猛留先生の担当部分）で事実上半日を費やす予定であった。しかし実際にはほとんどの受講生が PC 環境構築に成功しており、前述のように開始 10 分不足で事実上確認作業は終了した。むしろ、理想を掲げつつも現実的な対応をとったこと（事前作業を要求しておきながら講習時間内で Linux 導入確認作業を行うこと）に対する不満がアンケートに記載されていた。

Bio-Linux 以外の UNIX/Linux 環境構築や各種インストール系実習は、次回以降も詳細な手順書を公開して自力で行ってもらうことを基本としたほうが効率的であろう。もちろん、本コース事前

準備期間においても、数名の受講生からのインストールに関する質問メールは受けた。それゆえ、以下の NGS 用カリキュラム改善提案とも関連するが、「速習」コース初日（あるいはコース前日）は、対応可能なスタッフや TA を数名常駐させて以下のような各種準備や確認作業を自主的に行ってもらおうほうが効率的だと思われる。

- 持込 PC のソフトウェアインストールや各種解析環境構築状況の確認
- IT 系の受講生は、各自でバイオ系の過去の講義を統合 TV などでも自習
- バイオ系の受講生は、コマンドライン環境に慣れる

NGS 用カリキュラムについて

本講習会は、バイオサイエンスデータベースセンター（NBDC）運営委員会・人材育成分科会で策定された、研究データを整備・活用するバイオインフォマティクス人材を育成するためのカリキュラム（<http://biosciencedbc.jp/gadget/chousa/generation-sequencer.pdf>）を試行的に実施し、受講者アンケート等によりその有効性を検証することを目的としている。

全体としては、試行実施であったものの講習会も滞りなく終了し、受講生のアンケート結果も好意的であった。「速習」コース自体は、最低限必要とされる知識・技術を 2 週間程度で身につけることを想定したものであり、初心者想定している。実際、上記 PDF ファイル中でもレベルは初級と中級のみからなり、初級の項目が多くを占める。そのため、**主催側の意図通り本来の速習コース受講対象者である初心者にとっては非常に満足度が高かった**ようである。その一方で、理解度が早いか中級レベル以上の講義内容を期待していた受講生にとっては不満足な内容だったようである。後者は、「速習」コースの本来意図した受講対象ではない。しかし、それでも受講したのは、よりアドバンスドな内容からなる「速習以外」コースの前準備としての目的か、上記 PDF で比較的曖昧に記述されている「速習」コースの習得技術欄の記載内容によるのかもしれない。

受講生による全般的な要望は、NGS に直接的なハンズオン講習の時間をもっと確保してほしいというものが多くを占めた。具体的には、カリキュラム中の大項目「4. 次世代シーケンサ」の部分をもっと多くとってほしかったというものである。逆に、**2-8. 「速習」コースのアンケート結果の Q4** で不要だと思われる項目として挙げられていたものの多くは、カリキュラム中の大項目「1. コンピュータリテラシーとサーバー設計」、「5. ゲノム関連の倫理・法律」、および「6. 分子生命科学」であった。この二つは密接に関連している。この傾向は、アグリバイオインフォマティクスの講義の偏り（実習を含む講義科目のほうが受講人数が多い）とよく似ている。つまり、**受講生の要望は講義ではなく実習**である。平たく言えば、NGS 解析で実際に行われるコマンドをもっと打ち込んで手を動かしたいということであろう。今回不要であると指摘された部分の講義映像は全て統合 TV で配信されるため、各自で自習してもらおうという形でもいいのではないだろうか。したがって、もし NGS 用カリキュラムの枠組み内に残しておくとしても、来年度以降「速習」コースの枠組みで実際に時間を確保する必要性は高くはないだろう。

次に、講義の中身、難易度、および項目間の連携について述べる。最初の 4 日分の「1. コンピュータリテラシーとサーバー設計」から「2. 配列インフォマティクス」への移行時に、話の内容がインフォマティクス系からバイオ系となり、また話のテンポも急に早くなったという感想が多く寄せ

られた。また、NGS 分野で用いられる用語に不慣れなため講義についていくのが困難であったという声や、プログラミング言語 Perl が実際に NGS 解析の分野でどのように使われているかがわかりづらかったといった、**講義間の連携や順番の改善を期待する声**も寄せられた。

具体例として、今回のコースでは、講師の日程調整上および大項目の番号の順番から、疑問に思いつつも「6. 分子生命科学」をコース終盤に配置した。しかし、NGS 用カリキュラムのフロー図 (<http://biosciencedbc.jp/gadget/chousa/flow-diagram.pdf>) から明らかなように、本来であれば「6. 分子生命科学」の講義が少なくとも「2. 配列インフォマティクス」よりも前にこななければいけないはずである。バイオ関連の知識に自信のない受講生は、「6. 分子生命科学」の代替手段として、NGS 解析分野で頻用される用語やバイオ関連の基礎知識を事前に自習してもらっておけば大分感想は違っていたであろう。講師陣が挙げた参考書に目を通しておくことを徹底させるのもいいかもしれない（共催機関 1 のウェブサイト：<https://hpci.cbrc.jp/modules/tutorial/NBDC2014.html>）。

今回の「速習」コースの講義内容は、一部カリキュラムの「速習以外」との重なりが見られた。例えば、「速習以外」の 4-7 発現解析において、習得技術に BAM ファイルの読み込みや BED 形式への変換が書かれているが、これは「速習」の 4-3 で述べられていた。もちろん実際の講義では復習的な位置づけで行われるであろうが、「速習以外」の習得技術欄に記載しなくてもいいように思われる。要望の多い「速習以外」のコース実施と合わせて、習得技術欄の記載内容も随時変更していく必要があると思われる。

項目（講義科目）間の**連携を向上させる最も効果的な戦略は、できる限り少人数の講師陣でハンズオン講義を行うこと**である。一人あたりの担当コマ数が増えるため講師の負担は増えるものの、難易度のばらつきや実習の進展度合いに応じたフレキシブルな対応も容易になる。予算申請書やプレゼン資料などにありがちな「大人数の講師陣で充実した講義を提供」とは対極の方向性であり、ごちんまりとした印象を与え一見地味かもしれない。しかし受講生側の視点に立てば、実質的に連携をとりづらい大人数ではなく、少人数の教育能力の高い講師陣からなる体制に移行していくのが望ましいと思われる。

ハンズオンに特化した継続的な人材育成事業の必要性について

本コースの主催機関 1（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）の前身であるアグリバイオインフォマティクス人材養成プログラム設立当時（2005 年）は、各所にバイオインフォマティクス人材養成プログラムが存在していたため、農学分野への応用を全面に打ち出していた。しかし、2014 年現在、本コースの共催機関 1 が行っている人材養成プログラムなど一部を除き、ほぼ壊滅状態である。この主な理由は、**これまでの人材養成プログラムの成果として大勢の優秀なバイオインフォマティシャンが養成されて成功を収めたこと**になっているため、もはや同種の人材養成プログラムは立ち上げづらい状況にあるためであろう。しかし、ライフサイエンスデータベース統合推進事業と同様、本コースの試行結果が明白に示すように**人材養成事業もまたライフサイエンスの中心的な課題そのものである**。

どのような NGS を利用した研究プロジェクトであっても、今回の「速習」コースの基本スキルを身につけた「バイオインフォマティクス技術者」の需要は多い。本来、この種の人材養成は大学

の役割であり、それが正論である。しかし、研究成果も求められる大学教員が、本コースのような NGS の基本スキルに相当する内容も教えつつ、アップデートの激しい NGS の最先端分野にアンテナを張り続けるのは現実的ではない。また、NGS 解析の基本スキルを身につけたい受講生が望む、**ハンズオン講義を実施可能な機関は極めて限定**されている。

主催機関 1 (アグリバイオ) においても、昨今の NGS 解析スキル習得の要望に応えるべく、数年前より NGS 関連講義を増やしてきた。意思統一できる少人数の教員でフレキシブルに講義内容をアップデートし、実習系講義を増やすとともに、常時 2 名以上の TA を配置して受講生の理解度向上に努めてきた。本コースにおいても、主催機関 1 の設備やノウハウを主に用い、共催機関 1 などと連携して実施することで大人数の需要を満たすことに成功した。しかしながら、主催機関 1 の平成 27 (2015) 年度以降の存続は未定である。また、共催機関 1 も平成 27 年度末で一区切りとなる予定である。つまり、早急に検討すべき課題の一つとして、**本コース実施主体は恒久的な組織ではない**ことが挙げられる。

本コースの講義内容は、NGS 分野に特化しているものの、大部分はバイオインフォマティクス分野全般の教育に資するものである。一つの提案として、本コースの「速習」や「速習以外」の**ハンズオン講義に特化した教育専門機関の設立**が挙げられる。既存の複数のバイオインフォマティクス拠点の基礎教育や教材整備エフォートの軽減にもつながるため、高度な研究レベルの維持も期待される。「京」などのスーパーコンピュータを用いた高度な解析に必要な基本スキルも含めた講義も教育専門機関が担当し、本コースで実施したように共催機関 2 (DBCLS) と連携して講義映像を統合 TV で配信する枠組みを確立させることで、全受講生に平等に教育機会を提供することができる。プログラムのインストール関連相談窓口としての機能も持たせることができれば、近くに気軽に相談できるヒトがいない実験系研究室に所属する一人バイオインフォマティシャンにとっても心強い存在になるであろう。

2-10. よりよいハンズオン講義のために

受講生は、自身の研究を遂行する時間を削ってハンズオン講義に参加している。ハンズオン講義実施主体は、予習用のよりよい教材を整備するのはもちろんのこと、当日出席した受講生の満足度をできる限り高めるための最大限の努力をすべきであろう。ここでは、主催機関 1 および共催機関 1 のハンズオン講義経験、そして本コース受講生の要望を踏まえ、よりよいハンズオン講義のために必要な上記記載事項以外の事柄を備忘録的にまとめる。開催規模、場所、期間、確保できる TA 数などにも左右されるが、以下に示す心がけは重要であろう。

主催側：

- **受講者数の 1/3 程度以上の USB メモリを用意。** ネットワーク (有線、無線 LAN) の不具合に対応するため。当日ウェブ経由でダウンロード予定のファイルを USB メモリに置いておき、万が一ネットワークが使えなくてもハンズオン講義を継続できる体制にしておく。本コースでは、このほかにもネットワーク負荷軽減のため、個人所有の無線 LAN をできるかぎり利用するよう呼びかけた。

- 文字の大きさが確保された配布資料を用意。本コースでは、講師に 18 pt 以上の大きさをスライドを作成するようお願いし、図 15 に示すような「A4 横、4 アップ、両面印刷」の印刷物を配布した。また、印刷に伴うインクの消費量を減らすため、講師にはスライドの背景を白とするようにお願いした。

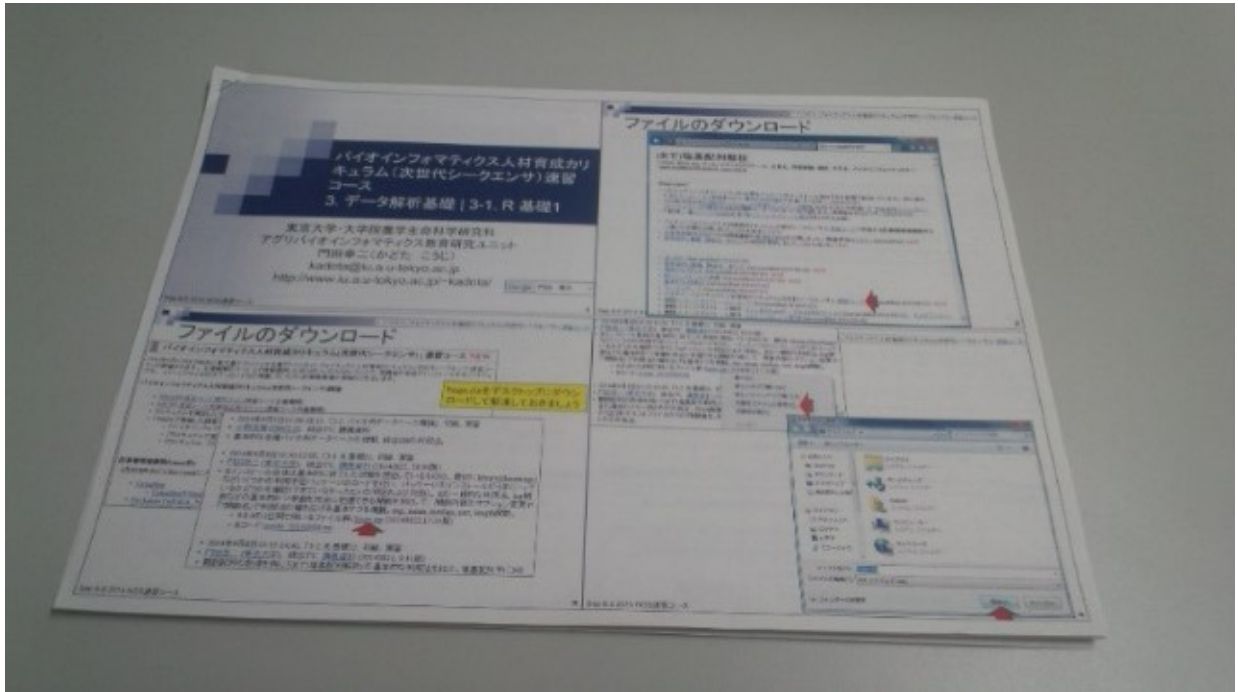


図 15. 本コース配布資料概観。

- 質問の切り分け。実は配布資料が間違っていただけというオチだったとしても、受講生は自分だけが間違っているのではという不安感をもつ。受講生の質問も、全体で共有可能な講師への質問と、動作確認程度の (TA への) も質問に切り分けると効率的である。本コースでは、「講師側に手を挙げる全体質問」と「後ろ向きで手を挙げる個別質問」に切り分けた。
- 質問しやすい雰囲気づくり。本コースでは公式な交流会は開催しなかったため、最終日に有志のみ 20 名程度で飲み会を開いたようである。アンケートでも、数人の受講生から本コースの比較的早い段階で交流会を開くことを希望するコメントがあった。個人情報保護の観点から、名札程度以外の情報開示は非現実的であるが、交流会場所の提供だけでも受講生同士の交流が生まれ、講習会全体の一体感も生まれるのではと思われる。
- 質問時のマイクの使用徹底。特に規模の大きな講習会の場合、講師側に近い講義室前方にいる受講生の質問は、講義室後方の受講生には聞こえないことが多い。講師以外用に複数のマイクを用意しておき、質問者には“速やかに”マイクを向けることを徹底させたほうがよいだろう。
- スクリーン下部に注意を払う。これは、講師の演壇が低い位置にあつて、受講生が講師を見下ろすような講義室の場合には特に気にする必要はない。しかし、本コース会場のような平面の講義室の場合には、スライドを投影するスクリーンの下部は、前方の受講生の頭に隠れてしまい、後方の受講生には見えないことが多い。本コースでは、Linux コマンド実行画面を見せる場合に、ターミナル画面下部がスクリーン画面の上から 2/3 程度の位置になるように配慮した。

講師が無意識に画面を広げる場合があるので、ターミナル画面下部がなるべく講義室前方の受講生の頭より上になるように気をつけたほうがよい。同様な理由から、講師側の心構えとも重なるが、ターミナル画面に限らずスライド中の記載情報はなるべく上のほうにシフトさせるように講師に徹底してもらったほうがよい（つまり後方の受講生が見づらい位置に重要な情報を載せない）。

- **予備の（ノート）PC を可能な限り確保。** 本コースを滞りなく終えることができたのは、代替 PC を十分量確保していたことが大きい。Linux 関連のインストールは、一般に理不尽な問題に直面することが多い。本コースにおいても、受講生全員が Bio-Linux 7 上で同じコマンド (`sudo apt-get install nkf`) を打ち込んで文字コード変換コマンド `nkf` のインストールを試みたが、原因不明の理由により数名インストールできなかった。もちろん文字コード変換程度であれば他の代替手段で回避可能であるが、それ以外のすぐに解決不可能なエラーに遭遇した場合でも、代替 PC 貸与によって実習を継続可能である。本コースでは、Linux 関連以外にも、R のパッケージがインストールされていなかった 5~6 名に対しても、代替 PC 貸与で事なきを得た。
- **OS の違いに気をつける。** 上記の代替 PC とも関連するが、講師の PC と受講生持込 PC の OS の違いによる不具合も存在する。具体的には、R のコードを実行する際に Windows OS では `gzip` 圧縮ファイルの読み込みができる一方で、Macintosh OS では読み込みができないことを確認済みである（R ver. 3.1.0 調べ）。このため、動作確認は可能な限り Windows と Macintosh の両方で主催側が行っておき、講師にエラー事例報告するという形がよいだろう。特に R は GUI 画面が OS 間で若干異なるため、できれば両方の OS 下での操作説明資料を用意しておくとうまい。本コースでは、研究代表者（門田幸二）による Windows OS 環境下での R の実習が行われたが、Mac ユーザにも Windows OS の代替 PC を貸与することで、Mac 環境下でのうまくいかない事例と Win 環境下でのうまくいく事例の両方を確認させることができた。

講師側：

- **主催側の意図や希望に沿った実習を心がける。** 上記主催側の心構え記載事項と関連して、配布資料の文字の大きさ、スクリーン下部に重要情報をなるべく記載しない、OS の違いに気を配るなど、主催側が挙げた講師側の注意事項をどれだけ忠実に徹底できるかが、複数の講師からなるハンズオン講義の、全体としての整合性や統一性に大きく寄与する。
- **講義資料送付などの締切を厳守。** 多少認識のずれがあった場合でも、主催側のチェックや修正依頼に対応する余裕が生まれる。
- **受講生への事前ソフトウェアインストール指示は早めに。** ソフトウェアのバージョンや、OS 指定も忘れずに。主催側は、ハンズオン講義直前に受講生からのインストール失敗報告・対処依頼や、必要なメモリや HDD 容量確認などの確認メールに忙殺される。できるだけ丁寧なインストール手順書を作成し、想定されるエラーへの対処法なども示しておくとなおよい。
- **実習当日のネットワーク接続不良への事前対応。** 事前確認ではうまくアクセス・ダウンロードできて、当日相手先サーバのダウンや講習会会場のネットワーク不具合などにより、うまくいかないことが現実にはよく起こる。利用予定のデータファイルなどは、予め主催側に URL

とともに告知しておき、当日 USB でデスクトップにコピーさせるなどしておくとうよい。

- **丁寧な講義資料を作成。** 詳細かつ丁寧な講義資料は、復習にも役立つほか実習時においていかれた場合でも自力での巻き返しを可能にするため、実習の成否を握るといっても過言ではない。受講生がノートをとらなくてもいいように、Tips を含めて全て記載しておいたほうがよい。コマンド実行結果も含めてスクリーンショットをとったものを講義資料に載せておくとうよい。これは、初心者は自分の打ったコマンドが正しく動作しているかどうかの確認ができないため、「このような画面になっていれば正解」というものがあると安心するからである。よくある「研究に答えはない」という議論とは切り離して、一通りのコマンド入力周辺の手続き上の正しさ（スペルミスやパスが通っているかなど）の観点をもつべきであろう。
- **コピペ用コード集も用意。** 例えば NGS データのマッピングなどは、一連のコマンドが数行にも及ぶことがある。講師が手入力で打ち込むというポリシーだったとしても、受講生自身が遅れを挽回しやすいように、実習で用いる一連のコマンド集を別に用意しておくとうよい。また、パワーポイントなどで作成したコードは、オートコレクト機能のため、オプションの「ハイフン」が「ダッシュ」に変更されていて、コピペでエラーが出る場合が多い。これでハマる受講生も一定の割合で存在するため、注意喚起する必要がある。
- **「手を動かしながら聞いてください」は極力避ける。** 受講生はコマンド入力に慣れていないことに気を配るべきである。講師はコマンド入力にかかる時間を十分に与え、コマンド実行後の説明は、なるべく足並みが揃ったことを確認してから進めるべきであろう。同様に、コマンド実行結果の画面はなるべくいじらないのが鉄則である。入力ミスなどで遅れがちな受講生は一定の割合で存在する。エラーに遭遇した場合、受講生は自身の PC 画面と前で示しているスクリーン画面との違いを確認しようとする。しかし、講師が実行後の時間を十分に取らずに別のことを実行すると、画面がスクロールし、正しい入力かわからない、正しい動作結果かわからないといったことがおきる。いくら TA がついていても、TA ごと置きざりにされてしまうこともある。実行直後に結果の説明をしても、エラーに遭遇した受講生は、エラーの原因究明で精いっぱいのため、エラー関連のノートをとっている間に説明を聞きそびれる場合が多い。それは結局、正しいコマンドの打ち込みだけに追われて、肝心な中身の理解にまでたどり着けないという不幸な結果を生む。ハンズオン講義に慣れていない講師は、通常の講義のように自分のペースのみで実習が進むわけではないということを肝に銘じておくべきである。尚、丁寧な講義資料作成は、この種の不満を緩和する効果がある。
- **「これ」や「あれ」のような指示語は使わない。** 指示語は説明時に前の画面を見ていないと何のことを言っているのかわからなくなることがある。反射的に使ってしまう傾向にある場合は、手を動かすのを一旦やめさせて受講生の視線をスクリーンに向けさせてから話すように心がけるとよい。
- **受講生の PC 環境は多様だということを認識する。** これは OS の違いに気をつけることと本質的には同じである。同じ Bio-Linux 環境で同じインストールコマンドを打ち込んでも `nkf` コマンドのインストールに失敗する受講生が現実存在する。同じ一連のコードを打ち込んでも実行結果が微妙に違うことはしばしばある。講師がスクリーン上で示す結果のみが、唯一無二の

結果とは限らない。「スクリーンと同じ結果になっていないヒト？」と確認しながら進めていくとよいだろう。

2-11. 付録（速習コースのアンケート結果）

2-8 で示した以外の講習会全体のアンケート結果（自由記述方式の Q6 および Q7）を示す。ここでは、受講生のバックグラウンド（バイオ系、IT 系など）と実際の記述内容を示す。Q6 は 47 名、Q7 は 41 名が回答。同じ通し番号が同じ受講生とは限らない。

Q6（速習コース全般について、ご意見や改善提案をご記入ください。）

1. バイオ系

今回、これまで体系的に学ぶことのできなかつたことを（独学がメイン or 周囲の人にきく）いろんな風でしることができてよかった。自分ができることが広がったように感じた。もっと深く学んでいきたいです。

2. バイオ系

非常に参考になる Tips を紹介されていました。しかし、メモる余裕がないので、メモ程度でいいのでテキストに書いてもらえると助かると思います。初心者には、Linux ソフト自体のインストール実習を 2-3 個やってもらえるとよかったと思います。また、シェルスクリプトや Perl 講義の時間数を増やしてもらえるといいと思いました。

3. バイオ系

非常に満足できる内容でした。要望としては講義・実習の順番を概論的なところから始め、NGS について全体を俯瞰したうえで実習に臨めると理解が深まると思います。

4. バイオ系

様々な方々の講義を横断的に聞くことができ大変有意義だった。ただ、対象がどのレベルなのか、正直わかりづらかった。一週目と二週目で、内容のレベルに差を感じた。

5. バイオ系

開始時刻 10:30 で大変助かりました（これよりも早いと 2 週間通うのが困難になっていたと思います。）環境構築は、持込 PC を利用した講習では非常に難しいところだと思いますが、フォローする意気込みが十分に感じられ、ありがたく思いました。VirtualBox 上の BioLinux が、1 年ほど前に試したときよりも有用で、かつアメリエフさんのパッケージが非常によかったです。是非継続して使いたいのですが、VB の仕様の問題ですが、ディスク容量の増量（私は不成功）、共有ディスク化（こちらは成功）が難しいのが残念です。多様な専門性を持つ先生方から、さまざまな流派の解析法を短期間に集中して触れさせていただき、私は大いに満足しています。通常ならば自己流で解決しなければならぬところ、非常に効率よく多くのことを学ぶことができたこと心から感謝しております。

6. バイオ系

全般を通して非常に有意義な講習会であったと考えている。早速復習を兼ねて業務に活用していくつもりである。可能であれば録画していた講義内容を閲覧出来るようにしていただきたい（復習を兼ねて）

7. バイオ系

NGS 解析に必要な技術をすべて独学でやろうとするとかなりの時間と労力が必要だと思うので、速習コースが今後自分で勉強する足がかりとなるよいきっかけとなった。10 日間午前・午後の長丁場であったが実習メインだったので眠くてくたびれるようなことはなく、どの講義も楽しく勉強できた。Twitter での疑問点・トラブルの共有や web 上でソフト等のちょっとした Tips を公開するなどの情報系の方々の研究に対するオープンな姿勢は見習いたいと思った。

8. バイオ系

ラボにサーバーもソフトも全くないが、生物を扱っていて ngs 解析をこれからはじめたい、という人、設備はそろっているがまだ先輩に教えられたことしか出来ない人、インフォマティクス系でバイオを始めたい人、などかなりスタートが違う状態だったと思う。それぞれの講師の方の内容をより親和性のあるように改善すれば、すべてのターゲットをフォローできるような講義内容にもう少しでできる気がする。生物を扱っていて ngs 解析をこれから始めたい自分としては、内容は非常に分かりやすかったが、実際にスパコンにログインして始めることはできそうにないのがおしかった。

9. バイオ系

大変良く企画されたコースで有り大変満足いたしました。世話人の門田先生、諸先生方の御尽力に大変感謝いたします。最後の次世代シーケンス結果の実習を初めの部分でも導入編として入れていただき、以下に続く講義との連携を示していただけると良かったように思います。

10. バイオ系

質問とそれに対する答えを講義室にいる人全員で共有するという姿勢、質問をしやすい雰囲気、この 2 点にかなり助けられました。これは座長門田先生のカラーの賜物と思われます。一方で個別の質問の中でいい質問を受講者全員に伝達するようになっていたのかどうかは、あやふやな印象を受けたので、次回はその辺を徹底したほうがいいかもしれません。コース全般の容量に関するアンケートについて、適切としましたが、全般としてではなく個別に設問を設けたほうがいいかもしれません。

11. バイオ系

コンテンツの内容はよかったです。出てくる順番を工夫すればもっと分かりやすかったと思います。また、TA の方のサポートで助かった部分が多くありました。丁寧な対応ありがとうございました。

した。なお、このアンケート全般で言えることですが、「学部 3 年次」の意見としてとらえていただければうれしいです。

12. バイオ系

私のような初心者にとっては非常にありがたい機会でした。また別に、ChIP-seq や RNA-seq などに特化した中級 (?) コースも期待しています。

13. バイオ系

初心者でも一通りの知識と解析の基礎を学べてよかったです。甘い物のさし入れがとても有難かったです。

14. バイオ系

完全な基礎知識 0 で受講しました。たまに講師の方が「やったことある方? けっこういるんですね。」という言葉が恐怖でした。すべて 4 にしたのは、自分で事前に基礎知識をいれとけば、あるいは、そーゆーさらに基礎知識の講義があれば…というところです。最後の 2 日間だけをまた十分な時間をやってほしい気がします。しかし、貸与 PC ですが、成功体験は得ました。河岡さんの論文データの再現は wet 系の身としては、わかりやすかったです。

15. バイオ系

TA の方が多く、とても助かりました。ありがとうございました。

16. バイオ系

NGS 解析は初心者であるため、本コースを受講することができ、漠然としていたものがどのような解析を行うのかイメージできるようになりました。テキストなどの資料は初心者を設定されており、講師の説明も非常にわかり易かったため、NGS 解析に必要な一連の解析技術の入り口にスムーズに入れたと思っています。質疑応答では、初心者にはついていけない細かい質問が多かったため個別対応でもいいのではとも思いましたが、質問は何でも受け付けるという姿勢に安心感 (?) を覚えました。TA の方々も適切にご指導くださり、ついていくことができました。

17. バイオ系

バイオインフォマティクス関連の講義をこれだけ広く集中的に学習出来たことは非常に有意義でした。受講以前は Linux や解析関連ツールの名称なども区別がつかないような状態でしたが、講義と演習により全体像が朧げではありますが、イメージ出来る様になったと思います。実際に自らが扱うデータを用いて、更に細かい部分をつめていけば自分でも何とかデータを扱うことが出来るような気になったのが、受講して最も良かった点です。特にアメリエフさんの講義は分かりやすく、良かったです。当初、2 週間の実習期間は長いような気がしておりましたが、体系的に基礎から学ぶ為には、これぐらいの期間は最低限必要であると考え直しました。ただし、実習開催は個人的に

は学会シーズン直前の9月よりは7～8月の開催にして頂けたら、もう少し余裕をもって受講できたと思います。分野によって学会シーズンも異なると思いますが。講義資料をほぼ全て up して頂いたのは、実習についていく上で大変たすかりました。また、再度見直す際にも有効でした。ただ、(R で)塩基配列解析の web page はテキスト主体のページのため、目的の項目やリンクを探すのが若干困難ではありました。

18. バイオ系

CUI 初心者ですが、実習では時間を多くとっていただいたことと、TA の方々の助けもあって、大きなトラブルもなくついていくことができました。当初受講人数が多いと思いましたが、そのぶん質問や意見が多く出て、結果として良かったと思います。

19. IT 系

全体的なレベルを低めに設定しすぎ。

20. バイオ系

wet バックグラウンドと dry バックグラウンドの受講者に同等に教えるより、(希望を言えば) 分けて短い日程又は、より中級の内容までの速習コースだとよいと思います。2 週間研究室を離れることで参加を見送るヒトがいることと、参加を希望する人はある程度初級の内容は知っているのではないのでしょうか。家で復習しないと追いつけないくらいの強度の講習を希望します。

21. バイオ系

なぜこのような素晴らしい講習を無料でしていただけるのか分かりませんが、非常に密度の濃い、十分な量の内容で、このまま続けていただければ若手研究者にとってありがたいと思います。講習会の対象者としてウェットの研究者を想定する場合は、日程をもう少し改良することができるかもしれません。講習の日程が詰まっていたため、両立が若干困難でした。

22. バイオ系

素人すぎて何がよいか悪いかはよくわかりませんが、コースの内容は難しすぎず二週間ほぼ勉強になったなあ！！と達成感を得ていました。理解の足りない箇所はすぐ復習したいと思います。

23. バイオ系

躓いた時に TA さんにすぐに対応していただけたのは非常に助かりました。

24. バイオ系

世話人の門田さんがとてもていねいで、分からない時も手をあげやすかったです。私は本当の初心者で、専門用語すらわからなかったので、ていねいに皆さん講義して下さいました。遠方から来ている方や実験の都合とかもあるかもしれませんが、このような回をもっとたくさんやって

ほしいです。

25. バイオ系

これだけの内容を独学したら時間がどれほどかかったかわかりません。教えてもらえてラッキーでした。有給休暇を取って参加した甲斐がありました。また、宿題がなかったので助かりました。休暇を取ったとはいえ、仕事のメールがきますので、宿題があったらできなかったと思います。先生方および関係者は準備が大変だったことは想像に難くありません。十分お礼が言えませんでした。感謝の気持ちをお伝えしたいです。

26. バイオ系

マイナーで大量な情報をどのようにアノテーションするか、今回の講義をもとに **BLAST** などをかけてみようと思う。

27. バイオ系

多忙な先生方が講師を務められているので難しいこととは思いますが、事前に配布資料の間違いがチェックされていると良いと感じた。また、重複して講義される部分があったので、講師間の事前の連帯、または総括で内容を吟味する方が居れば尚良いと感じた。

28. その他

横のつながりができるようなシチュエーション（コマ）があってもよかったです。と思います。

29. バイオ系

もう少し実的な解析技術を習得できることを期待して申し込んだが、受講してよかった。特に **Linux** の基礎操作やシェルスクリプト、**Perl**、**R** の基本についてしっかり習うことができ、独習では判断しにくいような、「重要なこと」「よく使う方法」を身につけることができた。また、今までなんとなく知っていただけでのデータベースの使い方やデータ解析の原理について理解することができ、勉強になった。

30. バイオ系

バイオインフォマティクスに関わる講義・実習は初めてでしたが、初心者レベルにあわせた内容であったため、ゼロからやろうとしている人間にとって、非常に満足のいくカリキュラムでした。また、配付資料を事前・事後に **web** で見られたことで、予習・復習を非常に効率よくできました。改善点としては、もう少し受講者の人数とレベルを絞った方が良いような気がしました。

31. バイオ系

10日間お疲れ様でした。また本当にありがとうございました。

32. IT 系

どの講義とも、その分野のさわりとなるものでしたので、ある受講者にとっては易し過ぎるし、ある受講者にとっては初めてのことばかりで難しいという、受講者ごとに満足度の分かれる講義だったのではないかと思います。IT系のものからしますと、上記4でも不要としたのですが、(得られる知識だけで短絡的に不要と判断できるものでもありませんが。)前半4日目の Perl までは受講参加に二の足を踏む日程でしたし、現にたとえ初めて学ぶ言語であっても自力で十分習得できる内容でした。受講前にスキル背景を回答させていたので、その回答に合わせて、講習前半をIT系とバイオ系にクラス分けして、背景スキルのギャップをならすことができないものか、または、今回の日程ならば、途中から参加のIT系を受け入れていただける枠があればと思いました。(当方の周囲には大項目2、つまり5日目以降の一部参加を考えていて、満席で参加できなかったIT系同僚がおりました。)また、大項目に「サーバー設計」とありましたので、あまり経験のないサーバー設計を前半でご指導していただけることを個人的には期待していましたので、それがなく残念でした。しかし、業務を離れてまとまった時間で、講義を通して既得知識の整理をさせていただきましたし、プログラミング初心者の方達に、どのような手助けができるかを知る機会ともなりました。

一方、後半のツール紹介や解析フローに合わせた講義は、別の解析業務でコマンドを打ちなれている当方にも、講義通りにコマンドを実行していくのが精一杯の、まさしく「速習」でしたが、自分で覚えたり学んだりすることの端緒となるラインナップとして満足のいくものでした。また、講師の先生方、TAの方々が、質問しやすい、ヘルプを依頼しやすい雰囲気を作ってくださっていて、非常に活気のある実習形式の講義になっていて、大変充実した2週間でした。VirtualBox の活用については、とてもよいものでしたが、共有フォルダの設定までご指導したらどうでしょうか？講習で使用した BioLinux を使ってスキルアップを図る方にとって、ゲストOSで出力したファイルをホストOSのアプリで利用したい状況がすぐに生じると思うからです。(河岡先生の講習会中にも受講者の方がエクセルシートを BioLinux 上の慣れないアプリで処理している時がありましたが、この時も共有フォルダに移動させれば苦勞なく処理できたはずです。)
「ググれば」自己解決できることと言ってしまえばそれまでですが、BioLinux 側で Permission を意識したり、sudo コマンドを使用したりという局面が出てきますので、手慣れたものでないと自力でやるには難しいと思います。コマンドに慣れた後に、講習会後のことを意識したワンポイントレッスンとして組み込んだらどうでしょうか？速習コースの申し込み方法について3段階の優先順位をつけて受け付けていましたが、全日程満席締切の通知が公開されたのが、全日程申し込み期限終了前でした。優先順位1の方だけで定員に達していたとしても、同じ条件なら先着順という注意書きはありませんでしたから、全日程期限終了まで募集を続けるべきだったのではないのでしょうか？

33. バイオ系

今回、たまたま twitter 経由でこのような講習会が開かれることを知り、参加させて頂いた。ただ、実は、私達がこれまでやっていた実験は NGS でも環境微生物叢の解析を目的としたアンプリコンシーケンスが殆どだったので、私の事前調査が足りずにミスマッチする点多かった。ゲノム解析

に関しては知識が殆ど無かったので少し的外れな質問などもしてしまいご迷惑をおかけしてしまったと反省している。今、環境微生物でもどんどんゲノムが解読されはじめているので、ゲノム解析からの実験を組み立てるなど応用を考えていきたい。TAの方も沢山居て初心者にも優しい講習会だったと感謝している。改善点を上げられなくて申し訳ないです。個人的には非常に勉強になった。

34. バイオ系

80名のバックグラウンドが異なる参加者を同じレクチャーで講義するのは非常に難しい取り組みと感じましたが、講師・TAの方々が一歩進めて下さった。内容はちょうど良かった。質問内容も含め、メーカー側お客様側など様々な角度からNGSを考えるいい機会になった。末尾になりますが、先生方・TAの方に心より感謝いたします。

35. バイオ系

門田先生の資料がベストです。説明も細かく書いていますので、後でながめてかえして勉強・復習するのに大変つかえる。

36. バイオ系、その他

NGS解析が目的であるため、直接関係ない部分は減らしてほしい。実践的な演習をもっと増やしてほしい。今回の速習コースは満足することができました。ありがとうございました。

37. バイオ系

もう少し実践的な講義が欲しかったです。門田先生のホームページのお陰で、BioLinuxのダウンロード、インストールの方法や、講義資料に目を通し予習することができました。また講義中にトラブルが発生した時にTAの方々に助けて頂きました。ありがとうございました。

38. バイオ系

Googleとコピペ（最近は何XXるというらしい？）は偉大だと再確認できました。膨大な講義と実習に対して開校期間が短いと思います。

注）研究代表者の判断で、XXと変更した。

39. バイオ系

実習についてはある程度予習してから行ったほうが、効率が良いように思いました。

40. バイオ系

大変有意義でした。是非これからも同様のコースを開催して頂けると幸いです。門田先生の講義は解説と実習のバランスがすごくよかったです。あと、他の先生の講義中にも必要な際にみなをフォローされていて、そのおかげで理解を深めることができました。坊農秀雅先生の講義も特に勉強に

なりました。

41. バイオ系

講義ごとのアンケートにも書きましたが、講義の時間配分に改善の余地がある気がします。

42. バイオ系

受講間で、受講開始時点での理解度に大きな開きがあるように感じました。理解度ごとに細かくクラス分けされた方が、理解がより高まるのではないかと思います。

43. バイオ系

丁寧に準備されており、TAさんも親切で助かりました。受講生の交流がしやすいとよかったです。個人情報の問題等あるのかもしれませんが、所属と専門分野を事前に集計・配布するなど。

44. バイオ系

講義の内容について一部のヒトは **Twitter** 使ってスクリプトなどをツイートしていたが、全体で取り上げるほどでもないと感じた疑問などを気軽に書けるので、積極的に利用すべきだと思う。個別PC持参の人の対応時間はもう少し短くてもいいと思う。講習来る人にあらかじめインストールするようアナウンスしてあるし、ほとんどの人が準備できていた。もし準備ができていない人がいても貸出し用PCを使用すればいいと思う。

45. バイオ系

受講者の知識/技能レベルを均一にしていなかったため、序盤の内容は既知の事柄が多かった。初心者向け、解析経験者向けなど、細分化しての開催に期待したい。

46. バイオ系

受講者全員を気にかけて1回、1回足並みをそろえたりするのは良かった。また資料もコマンドや結果も載っているため、家で見直したときも何をしていたかわかるため、今後参考にもできる。

47. バイオ系

このようなコースを、開催頂き、大変感謝しております。研究上の問題を解決するために、**R**、**Linux**も全く初めてで、約1年前から今まで実質一人でインターネットや書籍から情報を得て、トライ&エラーを繰り返していたのですが、要する時間と労力の割に、成功率が低く、また、得られる情報が信頼性の確認にも多大な労力を要し、辛く、苦しい思いをしており、有償でもよいので教えてくれる方はいないか、と考えていたところでした。教育・研究共に、特に**NGS**データ解析において、進んでいる部署・分野の先生方から、はるかに効率よく、有用かつ汎用されている情報をご教授いただき、また実習という手間のかかる形態でのご指導を無償でいただき、大変有意義でした。実際は実習のように、うまくいかないこともあると思われ、今後が本番ではありますが、ありがとうございます。

ございました。教育という重要な役割が、日本でももっと高く評価されるようになるとうよいと思いをしました。

受講者側の意識の問題だと思いますが、質問が、全体にとってプラスになるか、という視点がやや不足した、個人的なこだわりなどについてのものが少なからずあり、講師の先生方がお困りになっているように見え、申し訳なく思いました。正直、質問内容・回答共に、レベルが高すぎてよく理解できないことが多かったです。河岡先生の時に、講師の先生方の使われている画面を、中央と、教壇から見て右側の両方で見せて頂けて、教壇から見て左端の列の座席でも見やすかったです。こちらが初心者であるためだと思いますが、なるべくゆっくり話していただけると有り難いです。また、コマンドを打つ指示を出された後、新たな情報を口頭で追加して頂いてしまうと聞き逃してしまうので、善意で行っていただいていることは十分承知しておりますが、できれば無言でいただけると、有り難かったです。蛇足ですが、准教授の門田先生が、全ての質問にマイクを持って走らなくてよいように、複数個所にマイクを置いて、発言したい受講者はそこに行く、とか、雇用されている先生が、分担エリアを決めてマイクを運ぶ、とかしてもよいように思いました。質問内容の理解のため、発言者の PC 画面も確認したい、という意図(メリット)があったのならよいのですが……。教育・研究共に、特に NGS データ解析において、進んでいる部署・分野の先生方から、はるかに効率よく、有用かつ汎用されている情報をご教授いただき、また実習という手間のかかる形態でのご指導を無償でいただき、大変有意義でした。実際は実習のように、うまくいかないこともあると思われ、今後が本番ではありますが、ありがとうございました。教育という重要な役割が、日本でももっと高く評価されるようになるとうよいと思いをしました。

Q7 (速習コースの次のステップとして、どのような講習会があれば参加したいと思えますか。具体的にご記入ください。)

1. バイオ系

RNA-seq など各分野に特化した講義。解析をもっとくわしくなど。

2. バイオ系

1) RNA-seq において、フリーソフトを用いて、Pair end を考慮しながらクリーニングする方法を紹介してもらえませんか。

2) Python の講習してもらえないでしょうか。

3. バイオ系

実際の生データから論文の figure になるまでの過程を体験する講習会がほしいです。

4. バイオ系

今回基礎は学ぶことができたので、今後、論文化における pitfall に重点をおいた講習 (または講習会) があれば参加したいと思えます。

5. バイオ系

まずは“速習以外”として扱われている講義内容を扱う講習会に参加したい。短期間決戦でなく、できれば週イチくらいの日程で開催していただければと考えている（勤務先での業務調整が困難で、短期間コースでは参加が難しいため）

6. バイオ系

今回の速習コースではシェルスクリプト・Perlの実習が1日だけで今すぐ使えるほど覚えられなかったので、プログラミング言語を数日間～1週間程度でみっちり習得する講習会。Reseq、RNA-seq、ChIP-seq等の自分に取り組んでいる、取り組む予定の解析法の習得に特化した数日間～1週間程度の講習会。

7. バイオ系

自分のデータ（FASTQ など。一部を取り出してもよい）を用いた実習形式の実践的な講習会。パソコンの登録方法などを含むとありがたいと思う。

8. バイオ系

マシンスペックが問題になってきますが、実際のヒトの結果を扱う実習などがあれば興味深いです。

9. バイオ系

今回は、各講師の方が教材データをそれぞれ用意してくれました。次のステップとしては、同じ教材データを R, Perl, LINUX プログラム、で扱うような講義もいいのではと素人は考えます。そうすることによって NGS 解析全体における各プログラム言語やプログラムの立ち位置を理解しやすくなると思うからです。今回の RNA-seq や ChIP-seq 実習のようなレベルと容量の講義をもっとたくさん受講したいとも思っています。

10. バイオ系

基本的な内容をもう一度見直し、実際にうまく活用できるレベルになれるようなものやさらなる応用力を身につけられるもの、生物学的な NGS の利用例（NGS を用いた研究）を多く（深く）知ることができるものがあれば参加してみたいです。

11. バイオ系

Metagenome

12. バイオ系

自分の実験に当てはめた時に応用がきくかトラブルシューティングができるかは難しいと思うので、実際のデータ解析を一通り行うものがあったら参加したいです。

13. バイオ系

実際にいろいろなパッケージを使用（実習）し、違い（長所・短所など）を実感するような講習があると嬉しいです。

14. バイオ系

速習コースではヒトやモデル生物を前提とした面があったため、速習コース以外では非モデル生物での適応例などを紹介して頂けると、私の場合、非常にうれしく、NGS 解析の裾野もより広がると思います。

15. バイオ系

プログラム言語(perl,python)、データ検索/NGS 解析に特化した実習コース

16. バイオ系

今回の講習で習得したような Perl や R の技術を踏まえて、次世代シーケンサデータ処理実習に時間をかけた講習会があれば参加したいと思います。

17. IT 系

R、統計、データベースについてのより高いレベルの講習

18. バイオ系

各解析技術があるのはわかってきたので、実際にその技術を使って解析を長く行っている人の現場での細かいトラブルシューティングや、解析方法の比較、オプション設定の仕方、エラーの対処を教えていただきたいと思います。（特に ChIP-seq、RNA-seq、data の可視化）

19. バイオ系

「速習以外」のコースの一部である、遺伝研スパコンなどの計算機の利用についての講習会です。

20. バイオ系

SNPs 解析

21. バイオ系

今回は割とレベルがバラバラだったような気がします。例えば学生向け（本当に初心者向け）にやってもらえば、初心者の学生でも分かりやすいし、インフォをやるきっかけになると思います。

22. バイオ系

統計学の講習会を希望します。難しいとは思いますが、土曜日開催ならさらに有難いです。もし無理なら東大の授業に参加させていただきたいです。

23. バイオ系

高速シーケンスの利用法別の研究者が集まる講習会。De novo シーケンシングをやっている人、RNA-seq をやっている人など、それぞれの利用法を実践している研究者が別々に集まって、その分野に特化したトラブルシューティングや、より良い解析法について意見を交換する会。

24. その他

私はウイルスのメタゲノム解析を計画しています。この研究テーマに即して Distance learning の指導教官のような形で指導していただければ心強いです。ご興味ご経験のある先生おられましたらご連絡ください。

25. バイオ系

集団遺伝学解析、ゲノムワイド関連解析、非モデル生物ゲノム解析（アセンブルやアノテーション付与の方法など）の内容の講習会があれば参加したい。マッピングや SNP 検出の際のクオリティフィルタリングについて、詳細な解説を聞きたい。

26. バイオ系

今回のコースは、時間的な制約もあって講義の後半部分についてはより踏み込んだ形で実習ができなかったものもありました。このため、内容的には今回と同じレベルぐらいで良いので、RNA-Seq や Reseq 等について時間を取ることで、より応用的に解析を行っていただければと思います。

27. バイオ系

速習以外も行うのであれば、ぜひ参加してみたい。この手の要望多数、でも人手もマネーも足りないでしょうが・・・

28. IT 系

BioLinux の構築などサーバー設計。R 言語でのクラスタリングやヒートマップ作成などの統計処理及び生物統計学

29. バイオ系

サーバの構築やスパコンの利用などに特化した短期講習会：スパコンの利用などは結構敷居が高いイメージであった。ただ、今回の講習会で無償でアクセス出来るということを知り、必要な機会があれば是非利用してみたいと思う。より詳しい利用法の研修などあったら参加してみたい。

30. バイオ系

まだまだシェルスクリプト・perl などは使用が難しいと感じたので、その部分を追加で学んでみたいと感じました。

31. バイオ系

応用変として、アセンブル、RAD-seq、Phylogenomics など。ヒトやモデル生物以外の生物についても触れてほしい。

32. バイオ系、その他

実際のデータを使用した演習形式（速習コースの後半部分）があれば、是非参加したいです。次のステップということなので、基本操作ができることが前提ですが、事前に課題（NGS データと解析）を提示し、講習会で演習するなど。事前準備も可能なので作業に没頭することなく、疑問点や解決できない点に集中できる。

33. バイオ系

Exome や RNA-Seq などの解析に特化した講義があれば積極的に受講したいと考えています。今回は名古屋から参加したのですが、2 週間ラボを空けるのは難しいので、各地方の勉強会などの勉強会があれば参加したいと思います。

34. バイオ系

自分のデータをもって、実際に解析するような講習会があるとうれしいです。

35. バイオ系

未経験者や少し触った程度の人を対象にするならば、実習時間をもっととるべきだと思います。理論よりもとにかく触って覚えるのがベターではないでしょうか。アグリバイオインフォマティクスの講義よりも実習が多くて頭に入りやすかったです（何度もエラー、ワーニングの表記が出ましたが・・・）

36. バイオ系

スクリプト言語の実習、GATK for non-model species

37. バイオ系

プログラミングや解析ツールの実際の運用法をさらに練習できるとありがたいです。

38. バイオ系

課題があって各自自習するような（ただ TA にいつでも聞ける）コースがあるといいかと思います。

39. バイオ系

以前統計数理研のセミナーで、少数の受講者が自分の研究データを発表し、それにどのような統計的手法を使うのがよいかを専門家と議論していました。このような実践的な（バイオインフォ版の）

ワークショップがあるとうれしいです。

40. バイオ系

実際のデータを想定した、計算サーバーでの実技実習（今回のようにデータを部分使用するのではなく、全利用）。細菌研究者向けの実習（ヒトゲノムをサンプルとして使用されても実感がわかなくなったり、取り入れなくてよいものがあるため）

41. バイオ系

ツールの比較による結果の違いや様々なツールの使い方など