

平成27年度 NGSハンズオン講習会 PC環境の構築 (7月22日) 講義資料

http://www.iu.u-tokyo.ac.jp/~kadota/r_seqhtml#bioinfo.ngs_sokushu.2015

前座プレゼン資料：受講生の心積入(PDF:995KB)

事前予習チェックリスト

- 2015年7月22日：Bio-Linux 8とRのインストール状況確認
- ・書籍「日本乳酸菌学会誌」についてで示した通りのPC環境を構築しておきましょう。
- ・連載第1-3回、および第4回のウェブ資料W6-5までの予習は必須です。
- ・Rについても同様です。
- ・一週間程度はきっちり時間をかけて予習しておきましょう。
- ・7/22は、以下に示すようなことができる(わかる)ようになっていることの確認を自分でしてもらいます。
- ・ウェブ資料のページ数も膨大ですので、どこにどういことが書かれていたかの全体像の俯瞰やチェックリストという位置づけでもあります。

Bio-Linux 8

1	(第4回だけでなく第2-3回のウェブ資料PDFにも目を通している)
2	ゲストOS、ホストOSなどと言われてうろたえない(第1回)
3	Linux環境では、一般にダブルクリックでソフトウェアのインストールはしない(第1回)
4	Perl, Ruby, Pythonはプログラミング言語である(第1回)
5	「円(yen)マーク」と「バックslash」の違いは見栄えの問題であり、文字コードとしては同じなので気にしない(第2回)
6	コマンドとオプションの間には「スペース」が入るものだ(第2回)
7	.(ドット) から始まるファイル名のもの(の多く)は、環境設定ファイルである(第2回)
8	全角文字をコマンドとして打ち込むのは非常識である(第2回)
9	ファイルの拡張子は常識的に使う。例えばFASTQ形式ファイルの拡張子を.fastaや.faなどにしない(第2回)
10	「Tabキーによる補完」、「上下左右の矢印キー」を駆使して効率的にコマンドを打てるようになっている(第2回)
11	リダイレクトやパイプを含むコマンドを眺めて、何をやっているものが大体わかるようになっている(第2回)
12	Bio-Linuxの左側の引き出しみたいなアイコンまたはターミナル上で、ホスト <-> ゲスト間のファイルのやり取りが自在にできる(第3回)
13	ホストからゲストへはドラッグ&ドロップ可能だが、ゲストからホストへはドラッグ&ドロップ不可能[W8-1](第3回)
14	ターミナルの画面サイズや不具合は自分で変更および対処可能[W9-2-4](第3回)
15	共有フォルダの概念を理解している[W9-3](第3回)
16	rmコマンドは「ゴミ箱」への移動ではなく本当に消える[W10-4](第3回)
17	ダウンロードしたいファイルのURL情報取得ができる[W11-1](第3回)
18	ゲストOSのネットワーク接続が不調の場合は、つながるまで待つのではなく、ホストOS上で必要なファイルをダウンロードして共有フォルダ経由でゲストOS上の適切な場所に置くことができる[W11-1](第3回)
19	ダウンロードしたいファイルのURL情報取得ができる[W11-1](第3回)
20	圧縮・解凍コマンド(zip, unzip, gzip, gunzip, bzip2, and bunzip2)の使い方をマスターしている[W13](第3回)
21	headやtailコマンドでファイル中の最初と最後の部分を確認することができる[W14](第3回)
22	cat, more, lessコマンドの基本的な使用方法(脱出したときはCTRL + Cなど)を理解している[W14-4からW14-6](第3回)
23	多くのLinuxコマンドには様々なオプションがあり、-hや--helpなどで利用可能なオプションを眺めることができる[W15-2](第3回)
24	PDFファイルやPowerPointファイル中のコマンドをコピーするときは気をつける[W16-1](第3回)
25	コマンドオプション利用時の「ハイフン」と「ダッシュ」の気をつける[W16-3](第3回)
26	初心者向けテキストエディタgeditの基本的な利用法は知っている[W17](第3回)
27	仮想環境全体のディスク使用状況把握(dfコマンド)や、カレントディレクトリのディスク使用量(duコマンド)を把握し、適切に不要なファイルを削除できる[W18-2からW18-4](第3回)
28	「.(ドット)」はカレントディレクトリを表し、「..(ドットとドット)」は1つ上の階層のディレクトリを意味する[W18-3](第3回)
29	複数のコマンドを組み合わせたパイプ()利用の基本を知っている[W19](第3回)
30	連載原稿中では乳酸菌RNA-seqデータ(SRR618288)を例に解説しているが、NGSハンズオン講習会(または予習)では、(自己責任で)自分が興味あるデータで平行して解析してもよいことを知っている[W20以降](第3回)
31	bzip2はgzipに比べて圧縮率は高く、圧縮に要する時間も早いですが、解凍は遅い[W22-5からW22-6](第3回)
32	ファイルサイズが大きいものは、一つ一つの処理に要する時間がとてつもなく長くなることを知っている[W23](第3回)
33	コマンドオプション、パイプ、リダイレクトを駆使することでNGSデータのサブセットを抽出できることを知っている[W25-2](第3回)
34	アスタリスク(*)などのワイルドカードというものがあり、それらの挙動がある程度わかる[W1-2からW1-3](第4回)
35	ゲストOSがフリーズ状態に遭遇しても、自力で再起動するなりして復旧できる[W2-1](第4回)
36	Permission問題はchmodで自力で解決できる[W3-1](第4回)
37	同じ目的を達成する上で様々なやり方があることを知っている[W3-2](第4回)
38	ファイルのダウンロード時に拡張子がオリジナルと変わっていることがある[W4-4](第4回)
39	自動マウント(オートマウント)の設定ができていない[W4-5; W4-7](第4回)
40	絶対パス、相対パスの概念が理解できている[W4-6](第4回)
41	文字コード(改行コード)をodやfileコマンドで確認し、perlで変換できる[W5-2; W5-3](第4回)

R	
1	インストールについてをよく読み、ここに書いてある手順に従って2015年4月4日以降にインストールを行った
2	インストールについて書いてある内容はBio-Linux8(ゲストOS)とは無関係であり、WindowsやMacintosh(つまりホストOS)上で行う作業である
3	ファイルの拡張子(.txtや.docxなど)はちゃんと表示されている
4	Rの起動と終了ができる。終了時に表示されるメッセージにうろたえない
5	R本体だけでなくRパッケージ群のインストールもちゃんと行った
6	Rパッケージ群のインストール確認も行き、エラーが出ないことを確認した
7	library関数を用いたRパッケージのロード中に、別のパッケージがないことに起因するエラーメッセージが出ることもあるが、必要なパッケージを個別にインストールするやり方を知っている
8	基本的な利用法をよく読み、予習を行った
9	作業ディレクトリの変更ができる
10	例題ファイルのダウンロード時に、拡張子が勝手に変わることがあるので注意する
11	慣れないうちは、getwd()とlist.files()を打ち込むことで、作業ディレクトリと入力ファイルの存在確認を行う
12	エラーに遭遇した際、「ありがちなミス1-4」に当てはまっていないかどうか自分で確認