2016.09.01版

バイオインフォマティクス
～LinuxでNGS解析（の基礎）～

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
門田幸二（かどた　こうじ）
kadota@iu.a.u-tokyo.ac.jp
http://www.iu.a.u-tokyo.ac.jp/~kadota/

Contents

- イントロダクション
  - 概要、背景（NGS用カリキュラム、講習会）、Linuxスキル習得の意義
  - ウェブ情報（日本乳酸菌学会誌のNGS連載やNGS講習会資料）
- 実習環境に慣れる
  - 仮想環境での作業に慣れる
  - GUIとCUI（マウス操作かコマンド入力操作か）
  - ターミナルでの作業
  - 共有フォルダの概念を理解
- 練習
  - 作業ディレクトリの変更、練習用NGSデータファイルのダウンロード
  - ファイルの確認、de novoゲノムアセンブリ
  - BLAST検索
- 課題
  - グループごとに異なる課題ファイルを入力として、「ダウンロード、de novoアセンブリ、BLAST検索」を実行し、得られた結果をレポートにまとめて発表せよ
  - グループ1はkadai1.fasta、グループ2はkadai2.fasta、etc.

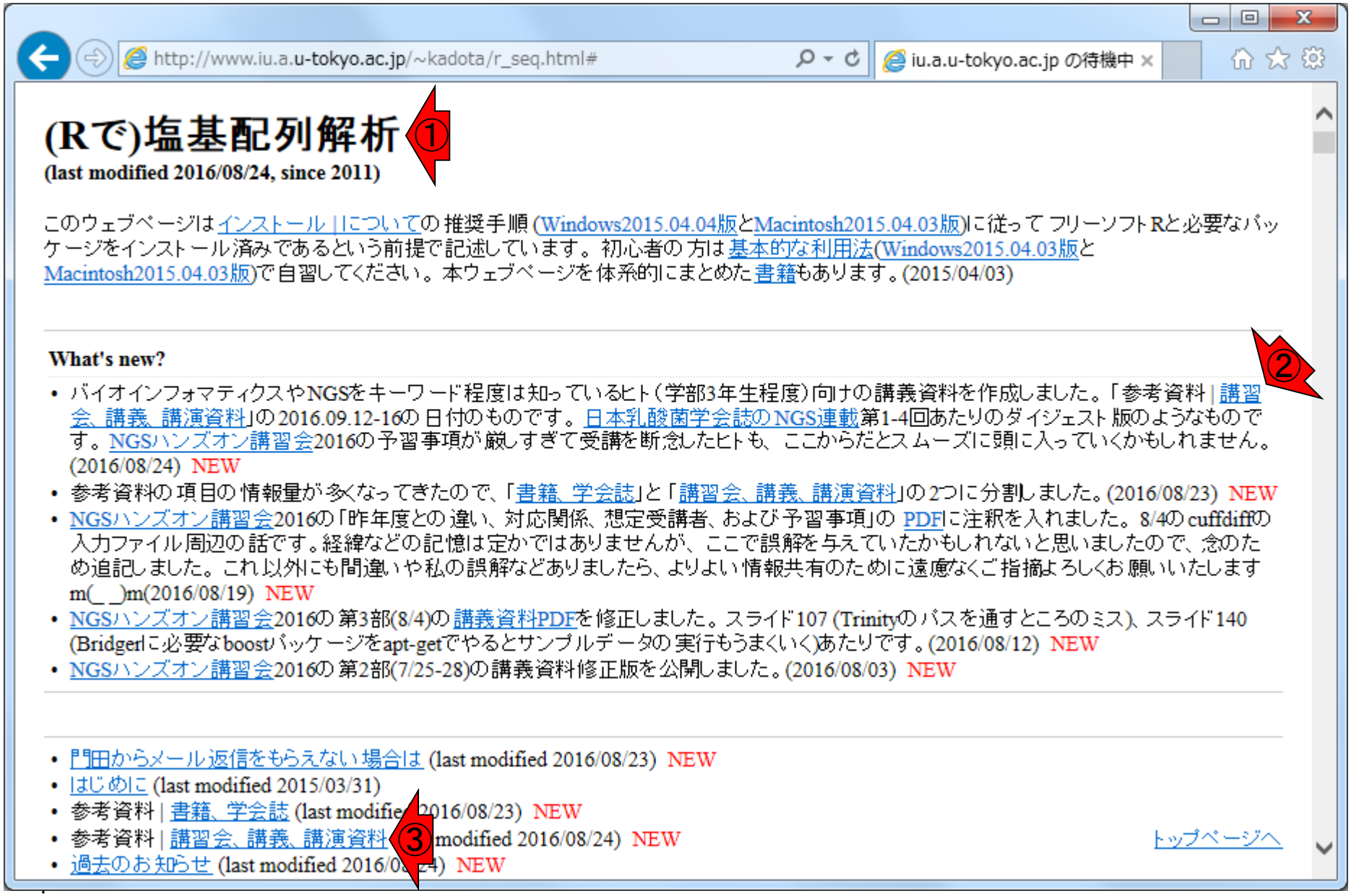
概要

（主にNGS解析を意識した）バイオインフォマティクス**スキルの習得がメイン**だが、何かをやったという達成感も得られるように実際のNGSデータの一部を用いてゲノムアセンブリまで行う

- キーワード
  - NGS, **Linux**, バイオインフォマティクス, 仮想環境, Bio-Linux, ゲノムアセンブリ
- Linux
  - WindowsやMacintoshと同じく、OSの一種
  - バイオインフォマティクス分野でよく利用される
  - 「Windowsのコマンドプロンプト」や「Macintoshのターミナル」と同じく、lsやcdなどのLinuxコマンドを知らなければ何もできないため、慣れるまでが大変
  - 使いこなせれば、最先端の解析用プログラムを自在にインストール・利用可能となり、効率的かつ**通り一辺倒でない**データ解析も可能となる
- 次世代シークエンサ(NGS)解析
  - NGSとは、大量の塩基配列を出力する実験機器(Next-Generation Sequencer)またはその技術を指す。主にゲノム解析やトランスクリプトーム解析と呼ばれる分野で利用されている
  - 塩基配列解析用プログラムは、UNIX(今のLinux)環境で動作するものが多かった歴史的背景などから、現在でもLinux上で動くプログラムがまず最初に開発される場合が多い

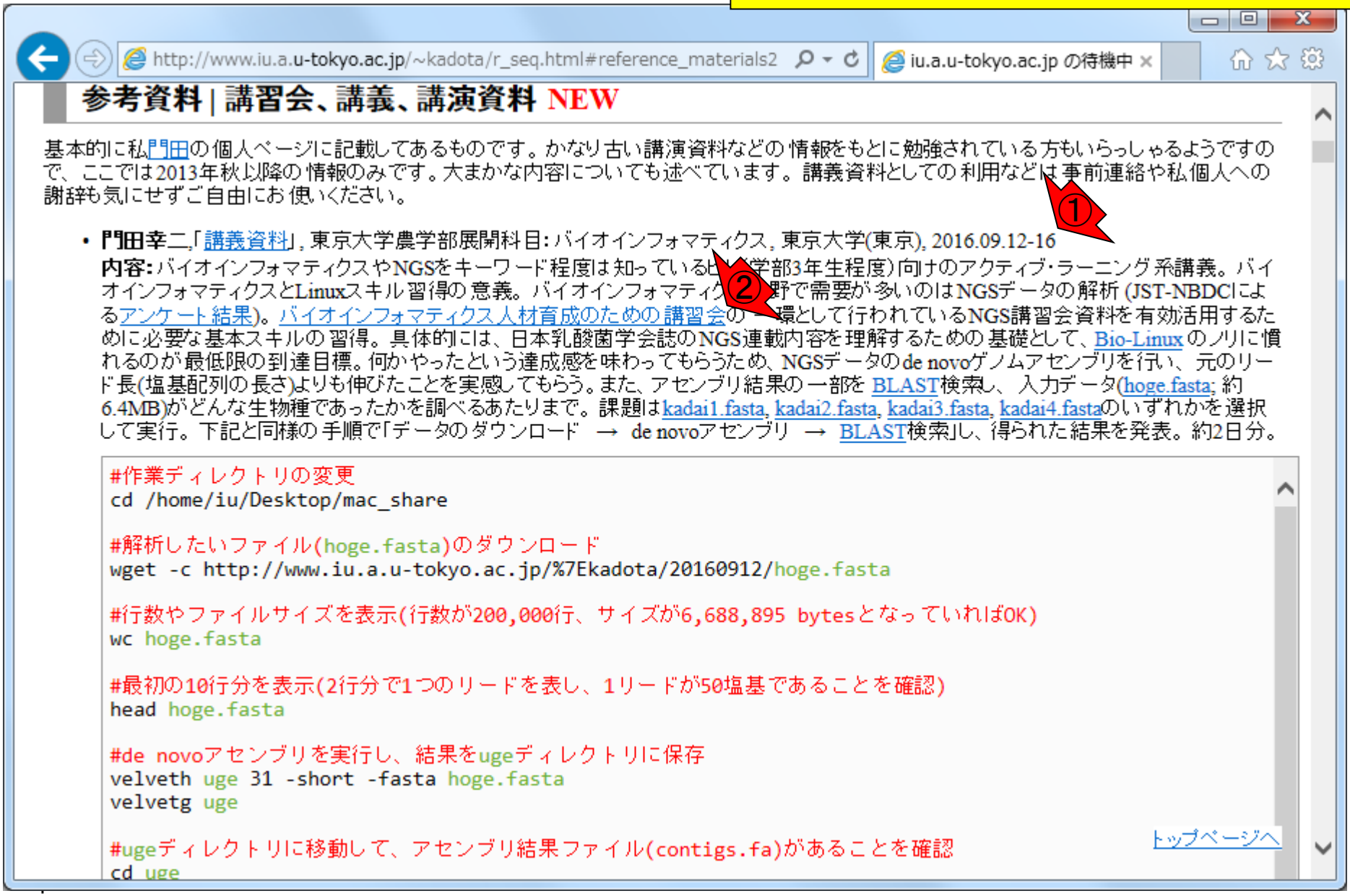
背景

「R NGS」などでググり、①のウェブページへ。②または③のあたりをクリック



背景

①2016.09-12-16の講義資料に辿りつく。ここは、私の講習会、講義、講演資料が公開されています。②をクリック



背景

①需要の多い次世代シークエンサ(Next-Generation Sequencer; NGS)から得られる大量塩基配列データを効率的に解析するためのバイオインフォマティクス人材育成カリキュラム（NGS用カリキュラム）。平成26年3月公開

NGS用カリキュラム

①

NGS用カリキュラムの中身。NGSデータ解析に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」にわかれている。ここで示しているのは①「速習」

【速習】

大項目	日数		No.	項目	習得技術		
1. コンピュータリテラシーとサーバー設計	4日	2日	1-1	OS, ハード構成	・コンピュータの基本の理解	初級	講義
			1-2	ネットワーク基礎	・インターネット、セキュリティの基本の理解	初級	講義
			1-3	UNIX I	・UNIXの基礎の理解 ・Linux導入	中級	実習
		2日	1-4	スクリプト言語	・Perl ・シェルスクリプト	中級	実習
2. 配列インフォマティクス	1日		2-1	配列解析基礎	・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	初級	実習
			2-2	バイオ系データベース概論	・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	初級	実習
3. データ解析基礎	2日		3-1	R 基礎1	・R言語の基礎(インストールから利用まで)	初級	実習
			3-2	R 基礎2	・ファイルの読み込み、行列演算の基本	初級	実習
			3-3	R 各種パッケージ	・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の利用法	中級	実習
			3-4	R bioconductor I	・bioconductorの利用法	中級	実習
			3-5	R bioconductor II	・FASTA and FASTQ形式ファイルの読み込み。ファイル形式の変換(FASTQ -> FASTA)、クオリティチェック、リード配列長分布、フィルタリングやトリミング、GC含量計算など	中級	実習
4. 次世代シークエンサ	2.5日	0.25日	4-1	次世代シークエンサ基礎I	・原理の理解	初級	講義
		0.25日	4-2	次世代シークエンサ基礎II	・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)	初級	講義
		0.5日	4-3	次世代シークエンサ実習I	・ファイル形式、可視化、quality check、マッピング、アセンブル	初級	実習
		1.5日	4-4	次世代シークエンサ実習II	・代表的なパイプラインについての実習(多型解析(IGV)、RNA-seq、ChIP-seq、及び統合解析、定番のツールを利用)	初級	実習
5. ゲノム関連の倫理・法律	0.5日		5-1	ゲノム情報倫理概論	・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、法的問題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	初級	講義
6. 分子生命科学	0.5日		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	初級	講義
			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	初級	講義
			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	初級	講義

NGS用カリキュラム

①

【速習以外】

大項目	No.	項目	習得技術	レベル	形式
1. コンピュータリテラシーとサーバー設計	1-5	UNIX II	・Windows OS上のUNIX環境構築	中級	実習
	1-6	プログラミング	・C、C++, Java	中級	講義
	1-7	ウェッブ開発技術 I	・サーバ構築(apache, XML, JavaScript, Ajaxなど)、公開に関する技術	中級	実習
	1-8	並列計算技術	・並列計算プログラミング技術大規模データ解析のために利用することができるクラスターマシン、クラウド環境(遺伝研スパコン、amazon EC2など)の選定 ・遺伝研スパコンなどの大型計算機システム利用のためのキューイングシステム活用法	上級	講義
	1-9	ウェッブ開発技術 II	・ビッグデータに対応したサーバ構築技術(Hadoopなど)	上級	実習
2. 配列インフォマティクス	2-3	分子系統解析	・分子系統解析ツールの導入や解析技術	中級	実習
	2-4	遺伝子機能解析	・COG,KEGG,GOなどを利用して,遺伝子配列から遺伝子機能の推定方法	中級	実習
4. 次世代シークエンサ	4-5	次世代シークエンサ　最新技術	・企業からの最新情報 ・機器のスペック、適した利用例及び標準的な導入価格の紹介	中級	講義
	4-6	アセンブル、マッピング	・次世代シークエンサデータ処理技術 ・アセンブル、マッピング	中級	実習
	4-7	発現解析	・次世代シークエンサデータ処理技術 ・RNA-seq 発現解析 ・BAMファイルの読み込み ・BAM → BED形式への変換 ・カウント情報取得(リファレンス配列の違い、アノテーション情報利用の有無など)	中級	実習
	4-8	クロマチン構造解析	・次世代シークエンサデータ処理技術 ・ChIP-seq解析技術 ・Hi-C	中級	実習
	4-9	環境ゲノム解析	・次世代シークエンサデータ処理技術 ・環境ゲノム解析技術の導入および利用技術 ・環境ゲノム解析技術 ・16SrRNAデータ解析:RDPデータベースやqiimeを利用した系統組成推定 ・メタゲノム解析:COG,KEGGを利用した機能組成推定	中級	実習
	4-10	比較ゲノム解析	・次世代シークエンサデータ処理技術 ・比較ゲノム解析 ・ゲノム内比較，種間比較に必要なツール(BLAST, HMMER, MAFFTなどを含む)の導入や解析技術	上級	実習
	4-11	個人ゲノム解析	・次世代シークエンサデータ処理技術 ・個人ゲノム解析 ・正の選択検出、SNVのリスク予測のツールの導入や解析技術	上級	実習
	4-12	疾患ゲノム解析	・次世代シークエンサデータ処理技術 ・CNV解析 ・エキソーム解析 ・変異同定	上級	実習
5. ゲノム関連の倫理・法律	5-2	ゲノム情報倫理応用	・匿名化、暗号化、情報セキュリティ詳細(技術的内容など)	上級	講義

NGS用カリキュラム

NGS用カリキュラムの中身。NGSデータ解析に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①「速習」の内容をとりあえずやってみたのが…

①

【速習】

大項目	日数		No.	項目	習得技術		
1. コンピュータリテラシーとサーバー設計	4日	2日	1-1	OS, ハード構成	・コンピュータの基本の理解	初級	講義
			1-2	ネットワーク基礎	・インターネット、セキュリティの基本の理解	初級	講義
			1-3	UNIX I	・UNIXの基礎の理解 ・Linux導入	中級	実習
		2日	1-4	スクリプト言語	・Perl ・シェルスクリプト	中級	実習
2. 配列インフォマティクス	1日		2-1	配列解析基礎	・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	初級	実習
			2-2	バイオ系データベース概論	・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	初級	実習
3. データ解析基礎	2日		3-1	R 基礎1	・R言語の基礎(インストールから利用まで)	初級	実習
			3-2	R 基礎2	・ファイルの読み込み、行列演算の基本	初級	実習
			3-3	R 各種パッケージ	・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の利用法	中級	実習
			3-4	R bioconductor I	・bioconductorの利用法	中級	実習
			3-5	R bioconductor II	・FASTA and FASTQ形式ファイルの読み込み。ファイル形式の変換(FASTQ -> FASTA)、クオリティチェック、リード配列長分布、フィルタリングやトリミング、GC含量計算など	中級	実習
4. 次世代シークエンサ	2.5日	0.25日	4-1	次世代シークエンサ基礎I	・原理の理解	初級	講義
		0.25日	4-2	次世代シークエンサ基礎II	・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)	初級	講義
		0.5日	4-3	次世代シークエンサ実習I	・ファイル形式、可視化、quality check、マッピング、アセンブル	初級	実習
		1.5日	4-4	次世代シークエンサ実習II	・代表的なパイプラインについての実習(多型解析(IGV)、RNA-seq、ChIP-seq、及び統合解析、定番のツールを利用)	初級	実習
5. ゲノム関連の倫理・法律	0.5日		5-1	ゲノム情報倫理概論	・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、法的問題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	初級	講義
6. 分子生命科学	0.5日		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	初級	講義
			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	初級	講義
			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	初級	講義

NGS速習コース講習会

①平成26年9月に行われた「NGS速習コース」講習会

バイオインフォマティクス人材育成のための講習会

NGS解析

ライフサイエンス分野の研究現場においては、莫大・多様な研究データが産出され、取り扱うデータ量が飛躍的に増えています。一方で、それらのデータを整備・活用するための人材は不足している状況です。 そのような研究現場の状況を踏まえ、NBDC運営委員会人材育成分科会において、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に基づき実施した講習会です。本カリキュラムは、対応が急務であると思われる次世代シークエンサデータに焦点をあてた内容となっております。

● 人材育成分科会で策定したカリキュラム

バイオインフォマティクス人材育成カリキュラム（次世代シークエンサ）

カリキュラムで習得できる技能

カリキュラム　フロー図

● H28年度NGSハンズオン講習会（2016年7月19日～8月4日）

※受講者アンケートにご協力をお願いいたします（予めメールでお知らせした認証IDとパスワードをご用意ください）。

受講者アンケート回答はこちら
回答期間：7/19（火）～9/16（金）

● H27年度NGSハンズオン講習会（2015年7月22日～8月6日）

● H26年度NGS速習コース講習会（2014年9月1日～12日）①

NGS速習コース講習会

①カリキュラム通りに行ったので、座学（講義）のみの時間もあった。また、計10日間にもおよぶため②担当講師数も多く連携をとりきれなかった。結果として③報告書中の受講生アンケートの主な要望は「実習のみで全体の連携」をとってほしい、であった

平成26年度NGS速習コース講習会（2014年9月1日～12日）

「バイオインフォマティクス人材育成カリキュラム（次世代シークエンサ）」の速習部分について、講習会を開催しました。

H26年度概要

H26年度講義日程・参考資料

H26年度講習会時の情報についてはこちらもご覧ください。

H26年度実施報告書・講義資料・動画等

● 講習会実施報告書（PDFファイル：3.7MB）

● 講義資料・動画

実施日	実施時間	大項目	項目番号および項目	レベル	形式	担当講師（敬称略）	講義資料・動画
9月1日	10:40-12:00	1.コンピュータリテラシーとサーバー設計	1-1. OS、ハード構成	初級	講義	中村 保一（DDBJ）	統合TV
	13:15-14:45		1-2. ネットワーク基礎				
	15:00-16:30		1-3. UNIX I	初級	実習	仲里 猛留（DBCLS）	統合TV
	16:45-18:15						
9月2日	10:30-12:00			中級	実習	仲里 猛留（DBCLS）	統合TV
	13:15-14:45						
	15:00-16:30						
	16:45-18:15						
9月3日	10:30-12:00		1-4. スクリプト言語	中級	実習	服部 恵美(アメリエフ)	統合TV
	13:15-14:45						

NGSハンズオン講習会

①平成27年7-8月に行われた「NGSハンズオン講習会」では、実習に特化した内容で実施

バイオインフォマティクス人材育成のための講習会

NGS解析

ライフサイエンス分野の研究現場においては、莫大・多様な研究データが産出され、取り扱うデータ量が飛躍的に増えています。一方で、それらのデータを整備・活用するための人材は不足している状況です。 そのような研究現場の状況を踏まえ、NBDC運営委員会人材育成分科会において、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に基づき実施した講習会です。本カリキュラムは、対応が急務であると思われる次世代シークエンサデータに焦点をあてた内容となっております。

● 人材育成分科会で策定したカリキュラム

バイオインフォマティクス人材育成カリキュラム（次世代シークエンサ）

カリキュラムで習得できる技能

カリキュラム　フロー図

● H28年度NGSハンズオン講習会（2016年7月19日～8月4日）

※受講者アンケートにご協力をお願いいたします（予めメールでお知らせした認証IDとパスワードをご用意ください）。

受講者アンケート回答はこちら
回答期間：7/19（火）～9/16（金）

● H27年度NGSハンズオン講習会（2015年7月22日～8月6日）①

● H26年度NGS速習コース講習会（2014年9月1日～12日）

NGSハンズオン講習会

①Linux基礎の項目は1日分しかないが、1日でLinuxの基礎を習得可能というわけではない！

平成27年度NGSハンズオン講習会

平成27年度は、平成26年度の実績を踏まえ、講義内容の改善等を行い、ハンズオンに特化した、より効果的なNGS講習会を開催しました。

H27年度概要

H27年度講義日程・参考資料

H26年度講習会の情報についてはこちらをご覧ください。

H27年度実施報告書・講義資料・動画等

● 講習会実施報告書（PDF：2.17MB）および受講者アンケート集計結果（データ集）（PDF：662KB）

● 講義資料・動画　*講義資料一覧のファイル名をクリックすると資料ファイル（PDF等）がダウンロードできます。

実施日	実施時間	大項目	項目	レベル	習得技術	担当講師(敬称略)	講義資料・動画(統合TV)
7月22日（水）	10:30-12:00 13:15-14:45 15:00-16:30 16:45-18:15	PC環境の構築	Bio-Linux8とRのインストール状況確認		・Linux導入 ・R導入 ・NGS解析に必要な環境構築技術	門田 幸二（東京大学） 寺田 透（東京大学）	事前予習資料一覧（PDF:52KB） 講義資料一覧（PDF:108KB）
7月23日（木）	10:30-12:00 13:15-14:45 15:00-16:30 16:45-18:15	UNIX/Linuxとスクリプト言語	Linux基礎 ①	初級 中級	UNIXの基礎の理解	門田 幸二（東京大学）	講義資料一覧（PDF:32KB） 統合TV
7月24日（金）	10:30-12:00		スクリプト言語	中級	シェルスクリプト	服部 恵美（アメリエフ）	講義資料（PDF:1.8MB）

NGSハンズオン講習会

①Linux基礎は、②事前予習事項の復習という位置づけ。講習会受講者の大半は、(Windows上で)Linuxコマンドを利用可能な③Bio-Linux8という解析環境を自力で構築するところからスタートして、1週間程度はかかる自習をしてきたヒト

平成27年度NGSハンズオン講習会

平成27年度は、平成26年度の実績を踏まえ、講義内容の改善等を行い、ハンズオンに特化した、より効果的なNGS講

H27年度概要

H27年度講義日程・参考資料

H26年度講習会の情報についてはこちらをご覧ください。

H27年度実施報告書・講義資料・動画等

● 講習会実施報告書（PDF：2.17MB）および受講者アンケート集計結果（データ集）（PDF：662KB）

● 講義資料・動画　*講義資料一覧のファイル名をクリックすると資料ファイル（PDF等）がダウンロードできます。

実施日	実施時間	大項目	項目	レベル	習得技術	担当講師(敬称略)	講義資料・動画(統合TV)
7月22日（水）	10:30-12:00 13:15-14:45 15:00-16:30 16:45-18:15	PC環境の構築	Bio-Linux8とRのインストール状況確認		・Linux導入 ・R導入 ・NGS解析に必要な環境構築技術	門田 幸二（東京大学） 寺田 透（東京大学）	事前予習資料一覧（PDF:52KB） 講義資料一覧（PDF:108KB）
7月23日（木）	10:30-12:00 13:15-14:45 15:00-16:30 16:45-18:15	UNIX/Linuxとスクリプト言語	Linux基礎	初級 中級	UNIXの基礎の理解	門田 幸二（東京大学）	講義資料一覧（PDF:32KB） 統合TV
7月24日（金）	10:30-12:00		スクリプト言語	中級	シェルスクリプト	服部 恵美（アメリエフ）	講義資料（PDF:1.8MB）

NGSハンズオン講習会

①事前予習事項のLinux部分は、日本乳酸菌学会誌に連載中のNGS解析記事をベースとしており、ウェブページ「（Rで）塩基配列解析」から全情報を取得可能

平成27年度NGSハンズオン講習会

平成27年度は、平成26年度の実績を踏まえ、講義内容の改善等を行い、ハンズオンに特化した、より効果的なNGS講習会を開催しました。

H27年度概要

H27年度講義日程・参考資料

H26年度講習会の情報についてはこちらをご覧ください。

H27年度実施報告書・講義資料・動画等

● 講習会実施報告書（PDF：2.17MB）および受講者アンケート集計結果（データ集）（PDF：662KB）

● 講義資料・動画　*講義資料一覧のファイル名をクリックすると資料ファイル（PDF等）がダウンロードできます。

実施日	実施時間	大項目	項目	レベル	習得技術	担当講師(敬称略)	講義資料・動画(統合TV)
7月22日（水）	10:30-12:00 13:15-14:45 15:00-16:30 16:45-18:15	PC環境の構築	Bio-Linux8とRのインストール状況確認		・Linux導入 ・R導入 ・NGS解析に必要な環境構築技術	門田 幸二（東京大学） 寺田 透（東京大学）	事前予習資料一覧（PDF:52KB） 講義資料一覧（PDF:108KB）
7月23日（木）	10:30-12:00 13:15-14:45 15:00-16:30 16:45-18:15	UNIX/Linuxとスクリプト言語	Linux基礎	初級 中級	UNIXの基礎の理解	門田 幸二（東京大学）	講義資料一覧（PDF:32KB） 統合TV
7月24日（金）	10:30-12:00		スクリプト言語	中級	シェルスクリプト	服部 恵美（アメリエフ）	講義資料（PDF:1.8MB）

Contents

- イントロダクション
  - 概要、背景（NGS用カリキュラム、講習会）、Linuxスキル習得の意義
  - ウェブ情報（日本乳酸菌学会誌のNGS連載やNGS講習会資料）
- 実習環境に慣れる
  - 仮想環境での作業に慣れる
  - GUIとCUI（マウス操作かコマンド入力操作か）
  - ターミナルでの作業
  - 共有フォルダの概念を理解
- 練習
  - 作業ディレクトリの変更、練習用NGSデータファイルのダウンロード
  - ファイルの確認、de novoゲノムアセンブリ
  - BLAST検索
- 課題
  - グループごとに異なる課題ファイルを入力として、「ダウンロード、de novoアセンブリ、BLAST検索」を実行し、得られた結果をレポートにまとめて発表せよ
  - グループ1はkadai1.fasta、グループ2はkadai2.fasta、etc.

（Rで）塩基配列解析

「R NGS」などでググる。①連載原稿やウェブ資料PDFはここからダウンロード可能

(Rで)塩基配列解析

(last modified 2016/08/12, since 2011)

このウェブページはインストール||についての推奨手順 (Windows2015.04.04版とMacintosh2015.04.03版) に従ってフリーソフトRと必要なパッケージをインストール済みであるという前提で記述しています。初心者の方は基本的な利用法(Windows2015.04.03版とMacintosh2015.04.03版)で自習してください。本ウェブページを体系的にまとめた書籍もあります。(2015/04/03)

What's new?

- NGSハンズオン講習会2016の第3部(8/4)の講義資料PDFを修
を通すところのミス)、スライド140 (Bridgerに必要なboostパッケ
実行もうまくいく)あたりです。(2016/08/12) NEW
- NGSハンズオン講習会2016の第2部(7/25-28)の講義資料修正
- NGSハンズオン講習会2016の第3部(8/01-04)の講義資料を一
- Twitterハッシュタグ AJACSはこちら。(2016/07/27) NEW
- NGSハンズオン講習会2016の第2部(7/25-28)の コピペ用コマン
デートする可能性はあります。(2016/07/21) NEW

- 書籍 | トランスクリプトーム解析 | 4.3.2 データの正規化(応用編) (last modified 2014/04/27)
- 書籍 | トランスクリプトーム解析 | 4.3.3 2群間比較 (last modified 2014/04/28)
- 書籍 | トランスクリプトーム解析 | 4.3.4 [illegible]の実験デザイン(3群間) (last modified 2014/04/28)
- 書籍 | 日本乳酸菌学会誌 | について ① modified 2016/07/01)
- 書籍 | 日本乳酸菌学会誌 | 第1回イントロダクション (last modified 2015/09/11)
- 書籍 | 日本乳酸菌学会誌 | 第2回GUI環境からコマンドライン環境へ (last modified 2015/11/26)
- 書籍 | 日本乳酸菌学会誌 | 第3回Linux環境構築からNGSデータ取得まで (last modified 2015/12/07)
- 書籍 | 日本乳酸菌学会誌 | 第4回クオリティコントロールとプログラムのインストール (last modified 20
- 書籍 | 日本乳酸菌学会誌 | 第5回アセンブル、マッピング、そしてQC (last modified 2016/07/05)
- 書籍 | 日本乳酸菌学会誌 | 第6回ゲノムアセンブリ (last modified 2016/06/17)
- 書籍 | 日本乳酸菌学会誌 | 第7回ロングリードアセンブリ (last modified 2016/07/01)
- イントロ | 一般 | ランダムに行を抽出 (last modified 2014/07/17)
- イントロ | 一般 | 任意の文字列を行の最初に挿入 (last modified 2014/07/17)
- イントロ | 一般 | 任意のキーワードを含む行を抽出(基礎) (last modified 2016/04/20)
- イントロ | 一般 | ランダムな塩基配列を生成 (last modified 2014/06/16)

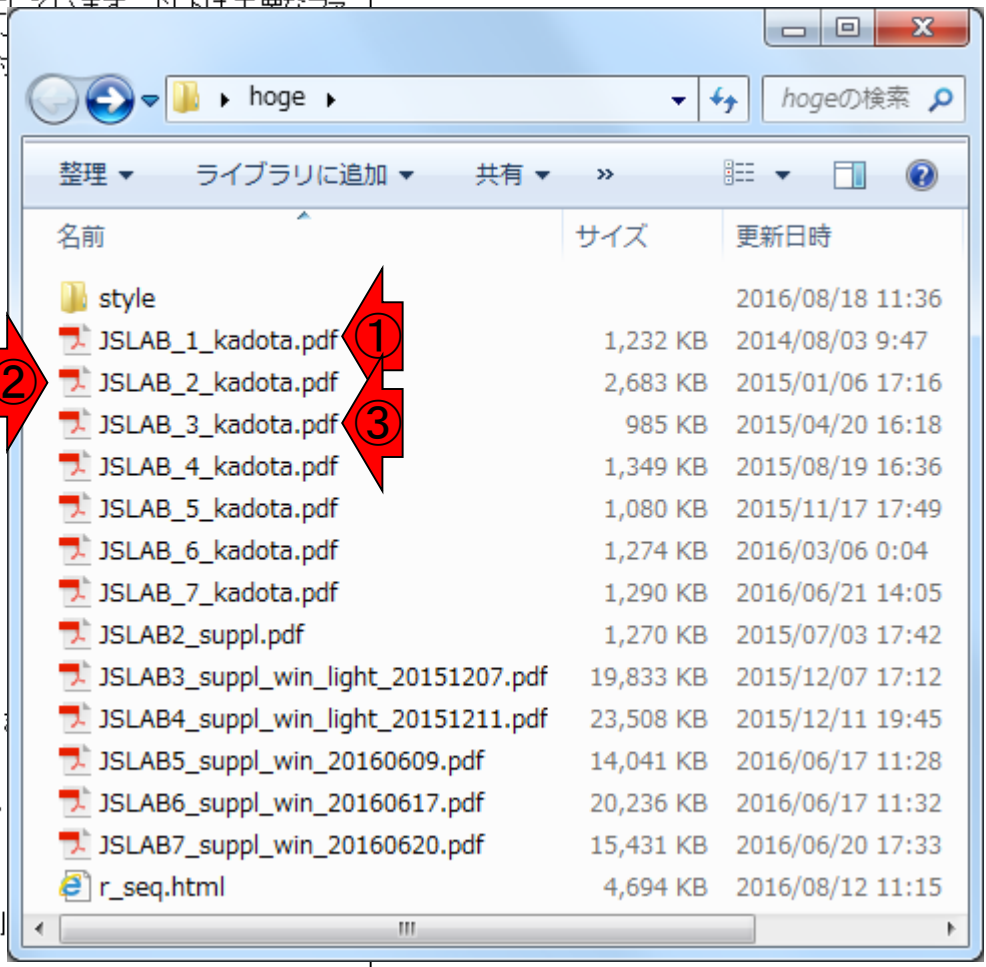
乳酸菌NGS連載

①②③で示す各回の原稿PDF（JSLAB_*_kadota.pdf）は、デスクトップ上にあるhogeフォルダ内にあります

書籍 | 日本乳酸菌学会誌 | について NEW

(このウェブページの取扱い上、書籍としていますが学会誌です)日本乳酸菌学会誌の連載原稿を書いています。NGSデータ解析初心者用に、各種情報収集先、Linux環境構築、Linuxコマンドなど、講習会などに出なくても十分な学習効果が得られるような情報提供を目指して執筆しています。情報もできるだけWindows用とMacintosh用の両方を作成しています。原稿PDF、ウェブ資料を含めフリーでダウンロード可能です。本文中で触れたウェブサイトのリンク先などの情報も辿れるようにしています。以下は主要なファイルのみリストアップしています。ダウンロードしたPDFファイルのトップページ右上にある日付のバージョンがいるウェブページのキャッシュに残っているのが表示されてしまう現象に遭遇してしまっています。対処法は「キャッシュを削除」です。

- 第1回イントロダクション(2014年07月):
  - 原稿PDF ①
- 第2回GUI環境からコマンドライン環境へ(2014年11月):
  - 原稿PDF ②
  - ウェブ資料PDF(2015.07.03版; 約2MB)
  - 1. VirtualBox、および2. Extension Packのインストール手順:
    - Windows用(2015.11.18版; 約3MB)
    - Macintosh用(2015.11.18版; 約8MB)
  - 3. 仮想マシンの作成、および4. Bio-Linux 8のisoファイルからのインストール手順:
    - Windows用(2015.11.19版; 約6MB)
    - Macintosh用(2015.11.19版; 約5MB)
  - Bio-Linux 8のovaファイルからのインストール手順(2015/11/25追加):
    - Windows用(2015.11.24版; 約2MB)
    - Macintosh用(2015.11.25版; 約4MB)
- 第3回Linux環境構築からNGSデータ取得まで(2015年03月):
  ユーザからの要望を踏まえ、ウェブ資料を更新しました。念のためオリジナル版も残してはいますが、軽量版を利用ください(2015.12.07追加)。
  - 原稿PDF ③
  - ウェブ資料PDF(オリジナル版; 原稿PDFと同じく、HDD150GB、約1.3億の全リードのダウンロード):
    - Windows用(2015.07.01版; 約21MB; 非推奨)
    - Macintosh用(2015.04.27版; 約23MB)
  - ウェブ資料PDF(軽量版; 原稿PDFと異なり、HDD100GB、最初の150万リードのみに制限):
    - Windows用(2015.12.07版; 約20MB; 推奨)
    - Macintosh用(未着手)



乳酸菌NGS連載

原稿PDFのページ数には制限があるため、①や②のウェブ資料PDFでより詳細な情報を提供しています

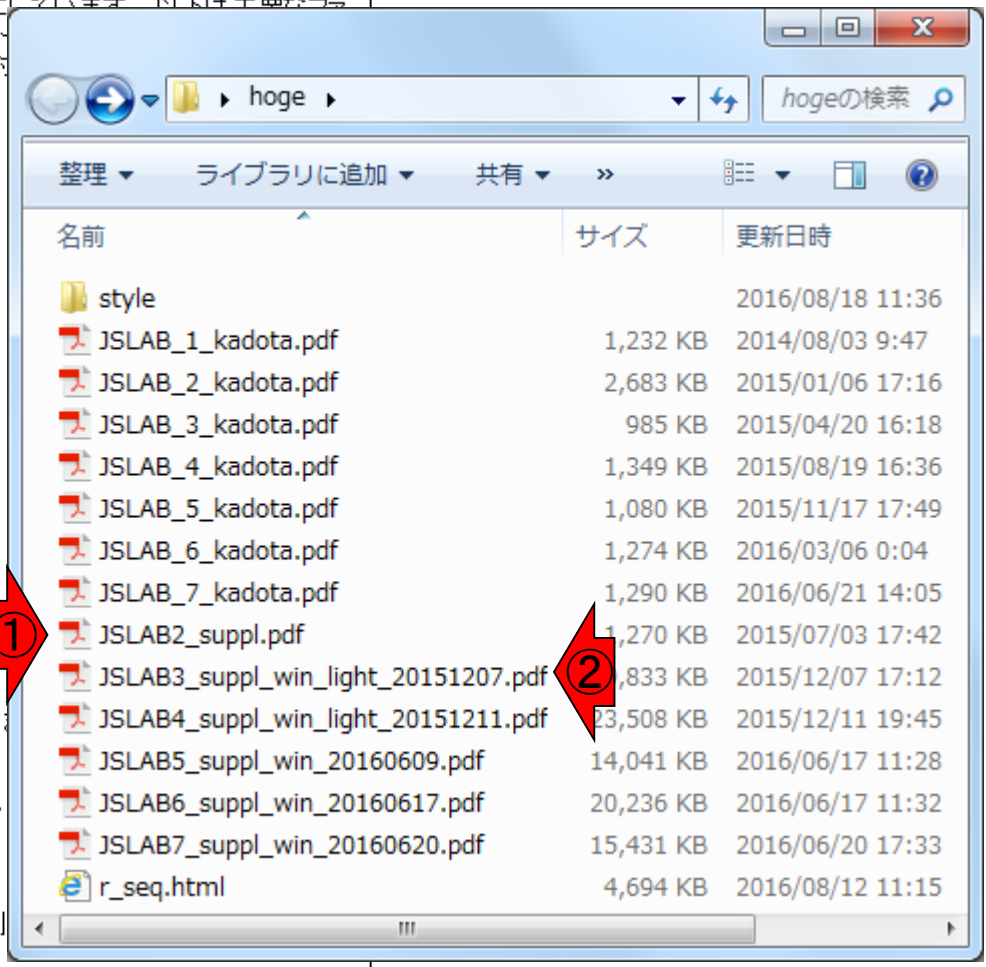
書籍 | 日本乳酸菌学会誌 | について NEW

(このウェブページの取扱い上、書籍としていますが学会誌です)日本乳酸菌学会誌の連載原稿を書いています。NGSデータ解析初心者用に、各種情報収集先、Linux環境構築、Linuxコマンドなど、講習会などに出なくても十分な学習効果が得られるような情報提供を目指して執筆しています。情報もできるだけWindows用とMacintosh用の両方を作成しています。原稿PDF、ウェブ資料を含めフリーでダウンロード可能です。本文中で触れたウェブサイトのリンク先などの情報も辿れるようにしています。以下は主要なファイルのみリストアップしています。ダウンロードしたPDFファイルのトップページ右上にある日付のバージョンが古いままの場合、ブラウザがウェブページのキャッシュに残っているのが表示されてしまう現象に遭遇してしまっています。対処法は「キャッシュを削除」です。

- 第1回イントロダクション(2014年07月):
  - 原稿PDF
- 第2回GUI環境からコマンドライン環境へ(2014年11月):
  - 原稿PDF
  - ウェブ資料PDF(2015.07.03版; 約2MB)
  - 1. VirtualBox、および2. Extension Packのインストール手順:
    - Windows用(2015.11.18版; 約3MB)
    - Macintosh用(2015.11.18版; 約8MB)
  - 3. 仮想マシンの作成、および4. Bio-Linux 8のisoファイルからのインストール手順:
    - Windows用(2015.11.19版; 約6MB)
    - Macintosh用(2015.11.19版; 約5MB)
  - Bio-Linux 8のovaファイルからのインストール手順(2015/11/25追加):
    - Windows用(2015.11.24版; 約2MB)
    - Macintosh用(2015.11.25版; 約4MB)
- 第3回Linux環境構築からNGSデータ取得まで(2015年03月):
  ユーザからの要望を踏まえ、ウェブ資料を更新しました。念のためオリジナル版も残してはいますが、通常は更新版を利用ください(2015.12.07追加)。
  - 原稿PDF
  - ウェブ資料PDF(オリジナル版; 原稿PDFと同じく、HDD150GB、約1.3億の全リードのダ
    - Windows用(2015.07.01版; 約21MB; 非推奨)
    - Macintosh用(2015.04.27版; 約23MB)
  - ウェブ資料PDF(軽量版; 原稿PDFと異なり、HDD100GB、最初の150万リードのみに制
    - Windows用(2015.12.07版; 約20MB; 推奨)
    - Macintosh用(着手)

①

②



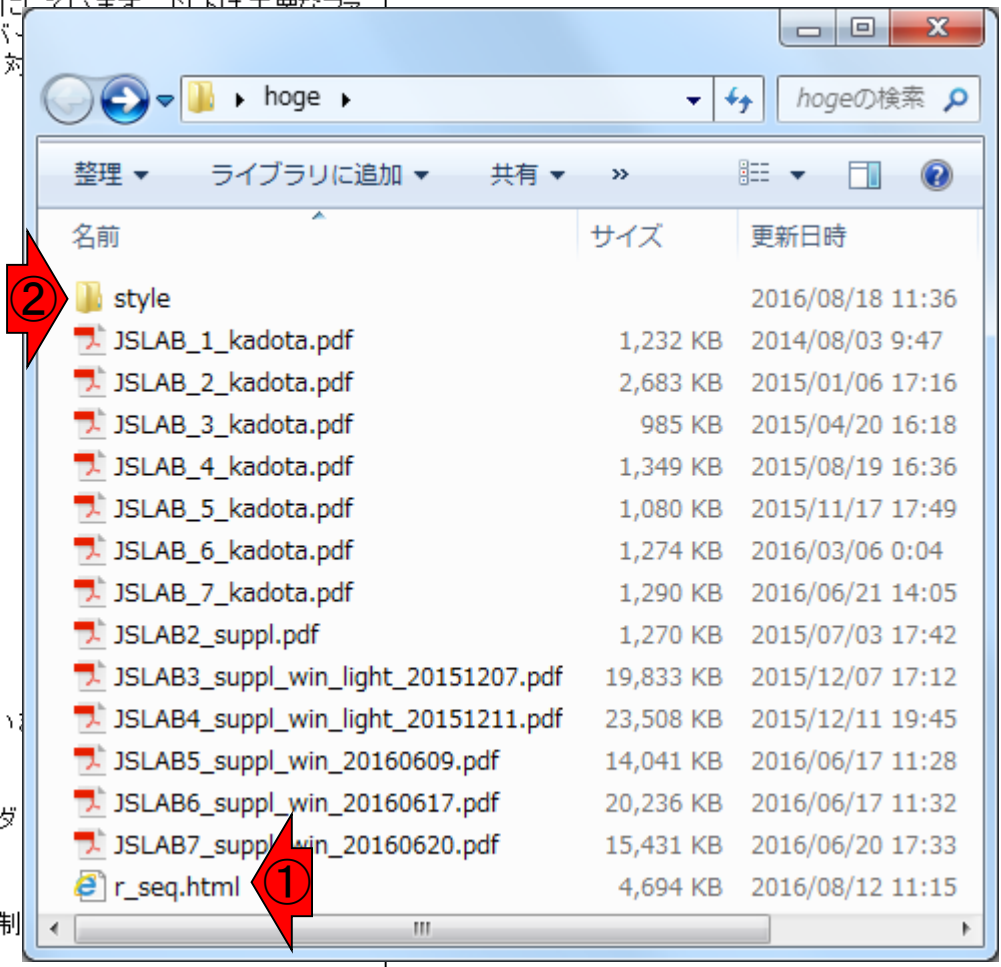
乳酸菌NGS連載

①は「(Rで)塩基配列解析」のソースファイル。ネットワーク不調時にダブルクリックで開くことで、ローカル環境でウェブページを開くことができます。②はウェブページの各種設定情報を含むフォルダです

書籍 | 日本乳酸菌学会誌 | について NEW

(このウェブページの取扱い上、書籍としていますが学会誌です)日本乳酸菌学会誌の連載原稿を書いています。NGSデータ解析初心者用に、各種情報収集先、Linux環境構築、Linuxコマンドなど、講習会などに出なくても十分な学習効果が得られるような情報提供を目指して執筆しています。情報もできるだけWindows用とMacintosh用の両方を作成しています。原稿PDF、ウェブ資料を含めフリーでダウンロード可能です。本文中で触れたウェブサイトのリンク先などの情報も辿れるようにしています。以下は主要なファイルのみリストアップしています。ダウンロードしたPDFファイルのトップページ右上にある日付のバージョンのウェブページのキャッシュに残っているのが表示されてしまう現象に遭遇してしまっています。対処法は「キャッシュを削除」です。

- 第1回イントロダクション(2014年07月):
  - 原稿PDF
- 第2回GUI環境からコマンドライン環境へ(2014年11月):
  - 原稿PDF
  - ウェブ資料PDF(2015.07.03版; 約2MB)
  - 1. VirtualBox、および2. Extension Packのインストール手順:
    - Windows用(2015.11.18版; 約3MB)
    - Macintosh用(2015.11.18版; 約8MB)
  - 3. 仮想マシンの作成、および4. Bio-Linux 8のisoファイルからのインストール手順:
    - Windows用(2015.11.19版; 約6MB)
    - Macintosh用(2015.11.19版; 約5MB)
  - Bio-Linux 8のovaファイルからのインストール手順(2015/11/25追加):
    - Windows用(2015.11.24版; 約2MB)
    - Macintosh用(2015.11.25版; 約4MB)
- 第3回Linux環境構築からNGSデータ取得まで(2015年03月):
  ユーザからの要望を踏まえ、ウェブ資料を更新しました。念のため**オリジナル版**も残してはいますが、通常は新版を利用ください(2015.12.07追加)。
  - 原稿PDF
  - ウェブ資料PDF(**オリジナル版**; 原稿PDFと同じく、HDD150GB、約1.3億の全リードのダ
    - Windows用(2015.07.01版; 約21MB; 非推奨)
    - Macintosh用(2015.04.27版; 約23MB)
  - ウェブ資料PDF(**軽量版**; 原稿PDFと異なり、HDD100GB、最初の150万リードのみに制
    - Windows用(2015.12.07版; 約20MB; 推奨)
    - Macintosh用(未着手)



連載第1回原稿

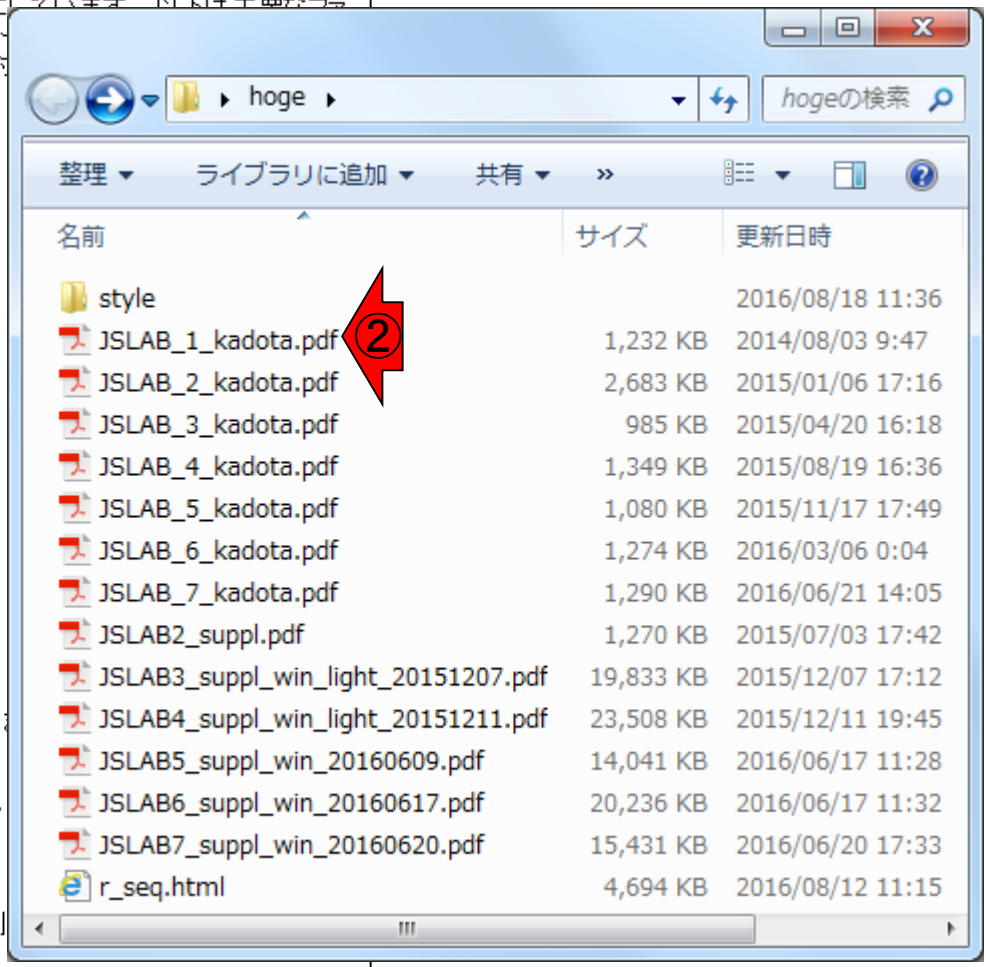
①乳酸菌NGS連載第1回原稿に相当する、②をダブルクリックで開いてみましょう

書籍 | 日本乳酸菌学会誌 | について NEW

(このウェブページの取扱い上、書籍としていますが学会誌です)日本乳酸菌学会誌の連載原稿を書いています。NGSデータ解析初心者用に、各種情報収集先、Linux環境構築、Linuxコマンドなど、講習会などに出なくても十分な学習効果が得られるような情報提供を目指して執筆しています。情報もできるだけWindows用とMacintosh用の両方を作成しています。原稿PDF、ウェブ資料を含めフリーでダウンロード可能です。本文中で触れたウェブサイトのリンク先などの情報も辿れるようにしています。以下は主要なファイルのみリストアップしています。ダウンロードしたPDFファイルのトップページ右上にある日付のバージョンが表示されていないウェブページのキャッシュに残っているのが表示されてしまう現象に遭遇してしまっています。対処法は「キャッシュを削除」です。

- 第1回イントロダクション(2014年07月):
  - 原稿PDF
- 第2回GUI環境からコマンドライン環境へ(2014年11月):
  - 原稿PDF
  - ウェブ資料PDF(2015.07.03版; 約2MB)
  - 1. VirtualBox、および2. Extension Packのインストール手順:
    - Windows用(2015.11.18版; 約3MB)
    - Macintosh用(2015.11.18版; 約8MB)
  - 3. 仮想マシンの作成、および4. Bio-Linux 8のisoファイルからのインストール手順:
    - Windows用(2015.11.19版; 約6MB)
    - Macintosh用(2015.11.19版; 約5MB)
  - Bio-Linux 8のovaファイルからのインストール手順(2015/11/25追加):
    - Windows用(2015.11.24版; 約2MB)
    - Macintosh用(2015.11.25版; 約4MB)
- 第3回Linux環境構築からNGSデータ取得まで(2015年03月):
  ユーザからの要望を踏まえ、ウェブ資料を更新しました。念のため**オリジナル版**も残してはいますが、通常は軽量版をご利用ください(2015.12.07追加)。
  - 原稿PDF
  - ウェブ資料PDF(**オリジナル版**; 原稿PDFと同じく、HDD150GB、約1.3億の全リードのダ
    - Windows用(2015.07.01版; 約21MB; 非推奨)
    - Macintosh用(2015.04.27版; 約23MB)
  - ウェブ資料PDF(**軽量版**; 原稿PDFと異なり、HDD100GB、最初の150万リードのみに制
    - Windows用(2015.12.07版; 約20MB; 推奨)
    - Macintosh用(未着手)

①



②

連載第1回原稿

こんな感じのものが見えるはずです。例えば原稿中の①「統合TV」のサイトはググってもよいが、各回のサイトからも辿れるようにしているので、②第1回のサイトをクリック

書籍 | 日本乳酸菌学会誌 | につい

(このウェブページの取扱い上、書籍としていますが学
初心者用に、各種情報収集先、Linux環境構築、Linux
提供を目指して執筆しています。情報もできるだけWi
めフリーでダウンロード可能です。本文中で触れたウェ
イルのみリストアップしています。ダウンロードしたPDF
いるウェブページのキャッシュに残っているのが表示さ
キャッシュを削除」です。

② 第1回イントロダクション(2014年07月):
- 原稿PDF

- 第2回GUI環境からコマンドライン環境へ(2014年
  - 原稿PDF
  - ウェブ資料PDF(2015.07.03版; 約2MB)
  - 1. VirtualBox、および2. Extension Packの
    - Windows用(2015.11.18版; 約3MB
    - Macintosh用(2015.11.18版; 約8MB
  - 3. 仮想マシンの作成、および4. Bio-Linux
    - Windows用(2015.11.19版; 約6MB
    - Macintosh用(2015.11.19版; 約5MB
  - Bio-Linux 8のovaファイルからのインスト
    - Windows用(2015.11.24版; 約2MB
    - Macintosh用(2015.11.25版; 約4MB

- 第3回Linux環境構築からNGSデータ取得まで(
  ユーザからの要望を踏まえ、ウェブ資料を更新
  用ください(2015.12.07追加)。
  - 原稿PDF
  - ウェブ資料PDF(**オリジナル版**; 原稿PDFと
    - Windows用(2015.07.01版; 約21M
    - Macintosh用(2015.04.27版; 約23M
  - ウェブ資料PDF(**軽量版**; 原稿PDFと異な
    - Windows用(2015.12.07版; 約20M
    - Macintosh用(未着手)

NGS解析周辺の講義内容を中心に
ンを行う。ウェブサイト（Rで）塩基配列解析（URL：http://www.iu.a.u-tokyo.ac.jp/~kadota/r_seq.html）中に本連載で述べるリンク先を掲載してあるので効率的に活用してほしい。

Key words：NGS, RNA-seq, Bioinformatics, R, Linux

情報収集先（講義、講習会、ウェブサイトなど）

筆者らの所属機関である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）では、数年前よりNGS関連ハンズオン講義（ノートPCを用いた実習を含む講義のこと）の割合を徐々に高めている。大学院講義ではあるものの、東京大学以外の学生、一般企業の社会人、ポスドクも受講可能である。アグリバイオの主な目標は、高度なバイオインフォマティシャン養成ではなく、手元にあるデータを自在に解析する技術を身につけたい実験系研究者の養成である。例年20%程度の受講生が東京大学以外であり、受講費用もかからないため、本誌読者向けの講義プログラムといえる。

*To whom correspondence should be addressed.
Phone ：+81-3-5841-2395
Fax ：+81-3-5841-1136
E-mail ：kadota@bi.a.u-tokyo.ac.jp

NGSデータ解析手法に関わるバイオインフォマティシャンの多くは、日本バイオインフォマティクス学会（JSBi）かNGS現場の会に所属している。これらの年会や研究会への参加を通じた情報収集も有意義であろう。例えば、2014年度のJSBi年会は10月に仙台で開催されるが、いくつか講習会も企画されている。HPCI人材養成プログラムも比較的規模の大きな活動を実施している。お台場にある産業技術総合研究所・CBRCを拠点として、NGSに特化した内容ではないものの、セミナー、ワークショップ、チュートリアル、e-learningなどが精力的に実施されている。

① e-learning系の代表格としては統合TV[1]が挙げられる。文字通り様々なウェブツールやデータベースなどの使い方を紹介する番組である。"NGS"というキーワード検索でリストアップされるものは全776番組中9番組と意外に少ない（2014年5月4日調べ）ものの、"次世代シーケンサ"で検索すると20番組程度がヒットする。また、バイオインフォマティシャンの多くが利用しているLinux（り

第1回のサイト

原稿中の①「統合TV」のサイトは、②から辿れます

書籍 | 日本乳酸菌学会誌 | 第1

日本乳酸菌学会誌の 第1回分です。

- 原稿PDF

要約:

- (Rで)塩基配列解析はここ

情報収集先(講義、講習会、ウェブサイトなど):

- アグリバイオインフォマティクス 教育研究
- 日本バイオインフォマティクス学会(JSBi)
- NGS現場の会
- 産総研CBRC
- HPCI人材養成プログラム
  - e-learning
  - tutorial(講習会)
  - セミナー
  - ワークショップ

② 統合TV: Kawano et al., Brief Bioinform.,
  - WindowsでUNIX! 1. Cygwin イン
  - WindowsでUNIX! 2. ファイル操作
  - WindowsでUNIX! 3. ファイル操作
  - WindowsでUNIX! 4. ファイル操作
  - Perlの使い方 インストール編(Wi
  - Perlの使い方 事例編(前編)
  - Perlの使い方 事例編(後編)

NGS解析周辺の講義内容を中心に基礎から応用まで幅広く述べる。第1回は、全体のイントロダクションを行う。ウェブサイト（Rで）塩基配列解析（URL：http://www.iu.a.u-tokyo.ac.jp/~kadota/r_seq.html）中に本連載で述べるリンク先を掲載してあるので効率的に活用してほしい。

Key words：NGS, RNA-seq, Bioinformatics, R, Linux

情報収集先（講義、講習会、ウェブサイトなど）

筆者らの所属機関である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）では、数年前よりNGS関連ハンズオン講義（ノートPCを用いた実習を含む講義のこと）の割合を徐々に高めている。大学院講義ではあるものの、東京大学以外の学生、一般企業の社会人、ポスドクも受講可能である。アグリバイオの主な目標は、高度なバイオインフォマティシャン養成ではなく、手元にあるデータを自在に解析する技術を身につけたい実験系研究者の養成である。例年20%程度の受講生が東京大学以外であり、受講費用もかからないため、本誌読者向けの講義プログラムといえる。

*To whom correspondence should be addressed.
Phone ：+81-3-5841-2395
Fax ：+81-3-5841-1136
E-mail ：kadota@bi.a.u-tokyo.ac.jp

NGSデータ解析手法に関わるバイオインフォマティシャンの多くは、日本バイオインフォマティクス学会（JSBi）かNGS現場の会に所属している。これらの年会や研究会への参加を通じた情報収集も有意義であろう。例えば、2014年度のJSBi年会は10月に仙台で開催されるが、いくつか講習会も企画されている。HPCI人材養成プログラムも比較的規模の大きな活動を実施している。お台場にある産業技術総合研究所・CBRCを拠点として、NGSに特化した内容ではないものの、セミナー、ワークショップ、チュートリアル、e-learningなどが精力的に実施されている。

① e-learning系の代表格としては統合TV[1]が挙げられる。文字通り様々なウェブツールやデータベースなどの使い方を紹介する番組である。"NGS"というキーワード検索でリストアップされるものは全776番組中9番組と意外に少ない（2014年5月4日調べ）ものの、"次世代シーケンサ"で検索すると20番組程度がヒットする。また、バイオインフォマティシャンの多くが利用しているLinux（り

第1回のサイト

各回のウェブサイトを用意することで、統合TVの①原著論文へのリンクや、②統合TVで提供している具体的な番組名やそのリンク先を示すことができます。また、③ページ下部に移動して眺めると、提供している情報量も膨大であることがわかります

書籍 | 日本乳酸菌学会誌 | 第1回イントロダクション

日本乳酸菌学会誌の第1回分です。

- 原稿PDF

要約:

- (Rで)塩基配列解析はここ

情報収集先(講義、講習会、ウェブサイトなど):

③

- アグリバイオインフォマティクス教育研究プログラム
- 日本バイオインフォマティクス学会(JSBi)
- NGS現場の会
- 産総研CBRC
- HPCI人材養成プログラム
  - e-learning
  - tutorial(講習会)
  - セミナー
  - ワークショップ

①

- 統合TV: Kawano et al., Brief Bioinform., 2012
  - WindowsでUNIX! 1. Cygwinインストール編
  - WindowsでUNIX! 2. ファイル操作編
  - WindowsでUNIX! 3. ファイル操作応用編
  - WindowsでUNIX! 4. ファイル操作発展編
  - Perlの使い方 インストール編(Windows)
  - Perlの使い方 事例編(前編)
  - Perlの使い方 事例編(後編)

②

ちなみに①は、②のリンク先と同じです

第1回のサイト

書籍 | 日本乳酸菌学会誌 | 第1回イントロダクション

①

日本乳酸菌学会誌の 第1回分です。

- 原稿PDF

要約:

- (Rで)塩基配列解析はここ

情報収集先(講義、講習会、ウェブサイトなど):

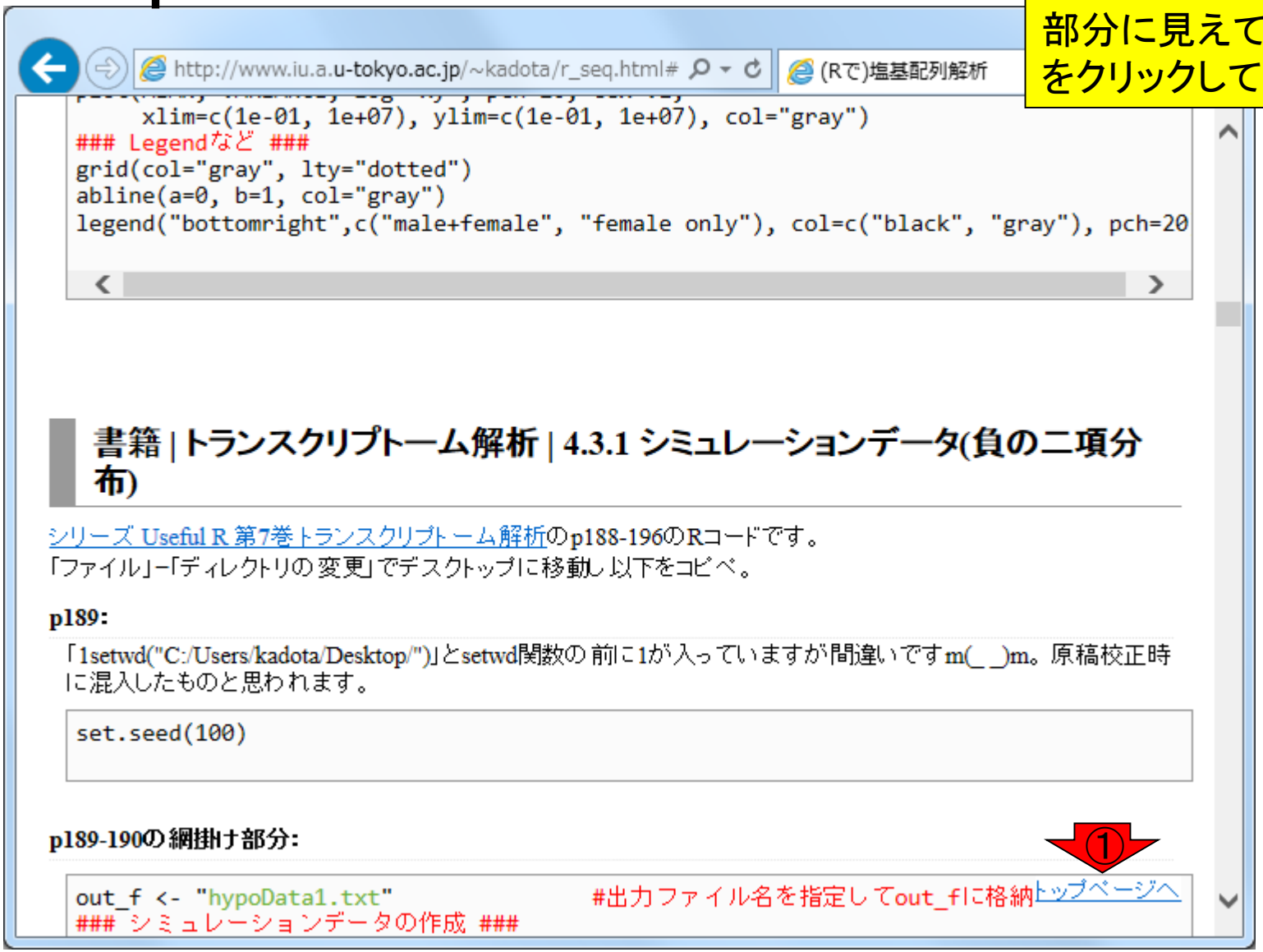
- アグリバイオインフォマティクス教育研究プログラム
- 日本バイオインフォマティクス学会(JSBi)
- NGS現場の会
- 産総研CBRC
- HPCI人材養成プログラム
  - e-learning
  - tutorial(講習会)
  - セミナー
  - ワークショップ
- 統合TV: Kawano et al., Brief Bioinform., 2012
  - WindowsでUNIX! 1. Cygwin インストール編
  - WindowsでUNIX! 2. ファイル操作編
  - WindowsでUNIX! 3. ファイル操作応用編
  - WindowsでUNIX! 4. ファイル操作発展編
  - Perlの使い方 インストール編(Windows)
  - Perlの使い方 事例編(前編)
  - Perlの使い方 事例編(後編)

②

- 書籍 | トランスクリプトーム解析 | 4.3.2 データの 正規化(応用編) (last modified 2014/04/27)
- 書籍 | トランスクリプトーム解析 | 4.3.3 2群間比較 (last modified 2014/04/28)
- 書籍 | トランスクリプトーム解析 | 4.3.4 他の実験デザイン(3群間) (last modified 2014/04/28)
- 書籍 | 日本乳酸菌学会誌 | について (last modified 2015/07/01)
- 書籍 | 日本乳酸菌学会誌 | 第1回イントロダクション (last modified 2015/09/11)
- 書籍 | 日本乳酸菌学会誌 | 第2回GUI環境からコマンドライン環境へ (last modified 2015/11/26)
- 書籍 | 日本乳酸菌学会誌 | 第3回Linux環境構築からNGSデータ取得まで (last modified 2015/12/07)
- 書籍 | 日本乳酸菌学会誌 | 第4回クオリティコントロールとプログラムのインストール (last modified 20
- 書籍 | 日本乳酸菌学会誌 | 第5回アセンブル、マッピング、そしてQC (last modified 2016/07/05)
- 書籍 | 日本乳酸菌学会誌 | 第6回ゲノムアセンブリ (last modified 2016/06/17)
- 書籍 | 日本乳酸菌学会誌 | 第7回ロングリードアセンブリ (last modified 2016/07/01)
- イントロ | 一般 | ランダムに行を抽出 (last modified 2014/07/17)
- イントロ | 一般 | 任意の文字列を行の最初に挿入 (last modified 2014/07/17)
- イントロ | 一般 | 任意のキーワードを含む行を抽出(基礎) (last modified 2016/04/20)
- イントロ | 一般 | ランダムな塩基配列を生成 (last modified 2014/06/16)

Tips

ウェブブラウザのサイズを変更したりすると、自分がどこにいるのかよくわからなくなります。その場合は、常に右下部分に見えている①「トップページへ」をクリックして…



Tips

ウェブブラウザのサイズを変更したりすると、自分がどこにいるのかよくわからなくなります。その場合は、常に右下部分に見えている①「トップページへ」をクリックして、②「(Rで)塩基配列解析」のタイトルが見える一番上まで移動したのち、例えば③NGS連載第2回のページをクリックするなどすればよい

(Rで)塩基配列解析

(last modified 2016/08/12, since 2011) ②

このウェブページはインストール||についての推奨手順 (Windows2015.04.04版とMacintosh2015.04.03
に従って フリーソフトRと必要なパッケージをインストール済みであるという前提で記述しています。初
の 方は 基本的な利用法(Windows2015.04.03版と Macintosh2015.04.03版)で自習してください。本ウェ
ページを体系的にまとめた書籍もあります。(2015/04/03)

What's new?

- NGSハンズオン講習会2016の 第3部(8/4)の 講義資料PDFを修
を通すところのミス)、スライド 140 (Bridgerに必要なboostパッケ
実行もうまくいく)あたりです。(2016/08/12) NEW
- NGSハンズオン講習会2016の 第2部(7/25-28)の講義資料修正
- NGSハンズオン講習会2016の 第3部(8/01-04)の講義資料を一
- Twitterハッシュタグ AJACSはこちら。(2016/07/27) NEW
- NGSハンズオン講習会2016の 第2部(7/25-28)の コピペ用コマン
デートする可能性はあります。(2016/07/21) NEW

- 書籍 | トランスクリプトーム解析 | 4.3.2 データの 正規化(応用編) (last modified 2014/04/27)
- 書籍 | トランスクリプトーム解析 | 4.3.3 2群間比較 (last modified 2014/04/28)
- 書籍 | トランスクリプトーム解析 | 4.3.4 他の実験デザイン(3群間) (last modified 2014/04/28)
- 書籍 | 日本乳酸菌学会誌 | について (last modified 2016/07/01)
- 書籍 | 日本乳酸菌学会誌 | 第1回イントロダクション (last modified 201[illegible]/11)
- 書籍 | 日本乳酸菌学会誌 | 第2回GUI環境からコマンドライン環境へ ③ modified 2015/11/26)
- 書籍 | 日本乳酸菌学会誌 | 第3回Linux環境構築からNGSデータ取得まで (last modified 2015/12/07)
- 書籍 | 日本乳酸菌学会誌 | 第4回クオリティコントロールとプログラムのインストール (last modified 20
- 書籍 | 日本乳酸菌学会誌 | 第5回アセンブル、マッピング、そしてQC (last modified 2016/07/05)
- 書籍 | 日本乳酸菌学会誌 | 第6回ゲノムアセンブリ (last modified 2016/06/17)
- 書籍 | 日本乳酸菌学会誌 | 第7回ロングリードアセンブリ (last modified 2016/07/01)
- イントロ | 一般 | ランダムに行を抽出 (last modified 2014/07/17)
- イントロ | 一般 | 任意の文字列を行の最初に挿入 (last modified 2014/07/17)
- イントロ | 一般 | 任意のキーワードを含む行を抽出(基礎) (last modified 2016/04/20)
- イントロ | 一般 | ランダムな塩基配列を生成 (last modified 2014/06/16)

Tips

あるいは、「①CTRL + ②F」キーを押して、③「コマンドライン」などの任意のキーワードを入力し、ページ内検索をしてもよい。1つの項目中で示されている情報量が膨大なため、実際問題としてこのサイト利用時にはキーワード検索もよく用いる

NGSハンズオン講習会

①H28年度の講習会（のLinux部分）は、②乳酸菌NGS連載第1-4回を予習として課した。予習事項は大まかに「仮想環境構築、Bio-Linux上での作業、共有フォルダやLinux系用語に慣れる、Linuxコマンドを一通り習得、…」

バイオインフォマティクス人材育成のための講習会

NGS解析

ライフサイエンス分野の研究現場においては、莫大・多様な研究データが産出され、取り扱うデータ量が飛躍的に増えています。一方で、それらのデータを整備・活用するための人材は不足している状況です。 そのような研究現場の状況を踏まえ、NBDC運営委員会人材育成分科会において、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に基づき実施した講習会です。本カリキュラムは、対応が急務であると思われる次世代シークエンサデータに焦点をあてた内容となっております。

● 人材育成分科会で策定したカリキュラム

バイオインフォマティクス人材育成カリキュラム（次世代シークエン

カリキュラムで習得できる技能

カリキュラム　フロー図

● H28年度NGSハンズオン講習会（2016年7月19日～8月4日）

※受講者アンケートにご協力をお[illegible]たします（予めメールでお知らせし

受講者アンケート回答はこちら
回答期間：7/19（火）～9/16（金）

● H27年度NGSハンズオン講習会（2015年7月22日～8月6日）

● H26年度NGS速習コース講習会（2014年9月1日～12日）

- 書籍 | トランスクリプトーム解析 | 4.3.2 データの 正規化(応用編) (last modified 2014/04/27)
- 書籍 | トランスクリプトーム解析 | 4.3.3 2群間比較 (last modified 2014/04/28)
- 書籍 | トランスクリプトーム解析 | 4.3.4 他の 実験デザイン(3群間) (last modified 2014/04/28)
- 書籍 | 日本乳酸菌学会誌 | について (last modified 2016/07/01)
- 書籍 | 日本乳酸菌学会誌 | 第1回イントロダクション (last modified 2015/09/11)
- 書籍 | 日本乳酸菌学会誌 | 第2回GUI環境からコマンドライン環境へ (last modified 20[illegible]1/26)
- 書籍 | 日本乳酸菌学会誌 | 第3回Linux環境構築からNGSデータ取得まで (last modi[illegible]015/12/07)
- 書籍 | 日本乳酸菌学会誌 | 第4回クオリティコントロールとプログラムのインストール (last modified 20
- 書籍 | 日本乳酸菌学会誌 | 第5回アセンブル、マッピング、そしてQC (last modified 2016/07/05)
- 書籍 | 日本乳酸菌学会誌 | 第6回ゲノムアセンブリ (last modified 2016/06/17)
- 書籍 | 日本乳酸菌学会誌 | 第7回ロングリードアセンブリ (last modified 2016/07/01)
- イントロ | 一般 | ランダムに行を抽出 (last modified 2014/07/17)
- イントロ | 一般 | 任意の文字列を行の最初に挿入 (last modified 2014/07/17)
- イントロ | 一般 | 任意のキーワードを含む行を抽出(基礎) (last modified 2016/04/20)
- イントロ | 一般 | ランダムな塩基配列を生成 (last modified 2014/06/16)

NGSハンズオン講習会

乳酸菌NGS連載第1-4回の予習事項をマスターしておけば、後は①で公開されている講義資料や動画（統合TVで今年度中に公開予定）で独習可能。時代はe-learningでハンズオン

H28年度 NGSハンズオン講習会カリキュラム

H28年度日程・講義資料・動画等

＊講義資料・動画については、順次公開いたします（2016.8.18更新）。

カリキュラム（PDF:72KB）

①

実施日	実施時間	大項目	タイトル	内容（予定）	担当講師 (敬称略)	講義資料・ 動画(統合TV)
7月19日 （火）	10:30- 18:15	はじめに（講習会参加者必読） PC環境の構築	Bio-Linux8とRのインストール状況確認	・Bio-Linux8（第2部および3部で利用するovaファイル）の導入確認 ・共有フォルダ設定完了確認 ・基本的なLinuxコマンドの習得状況確認 ・R本体およびパッケージのインストール確認 ・講師指定の事前予習内容の再確認 ・講習会期間中に貸与されるノートPCを用いた各種動作確認	主催・共催機関	講義資料 (PDF:4MB)
7月20日 （水）	10:30- 18:15	第1部 統計解析 (農学生命情報科学特論I)	ゲノム解析、塩基配列解析	・NGS解析手段、ウェブツール(DDBJ Pipeline)との連携 ・k-mer解析（k個の連続塩基に基づく各種解析）の基礎と応用 ・塩基ごとの出現頻度解析(k=1)、2連続塩基の出現頻度解析(k=2) ・塩基配列解析を行うための基本スキルの復習や作図 ・de novoアセンブリ時のエラー補正やゲノムサイズ推定の基本的な考え方	門田 幸二 (東京大学)	講義資料 (PDF:7.3MB) 解析データ (ZIP:2.2MB)
7月21日	10:30-		トランスクリプトー	・カウントデータ取得以降の統計解析(RNA-seq)		講義資料

ちなみに

アグリバイオ単体で行う大学院講義では、Linux環境でのデータ解析系講義は行われません（と思っておけば間違いありません）。受講人数が多すぎること（①最大で130名）、受講生の意識レベルや習熟度の差が大きく、講義として成立させることが困難なためです

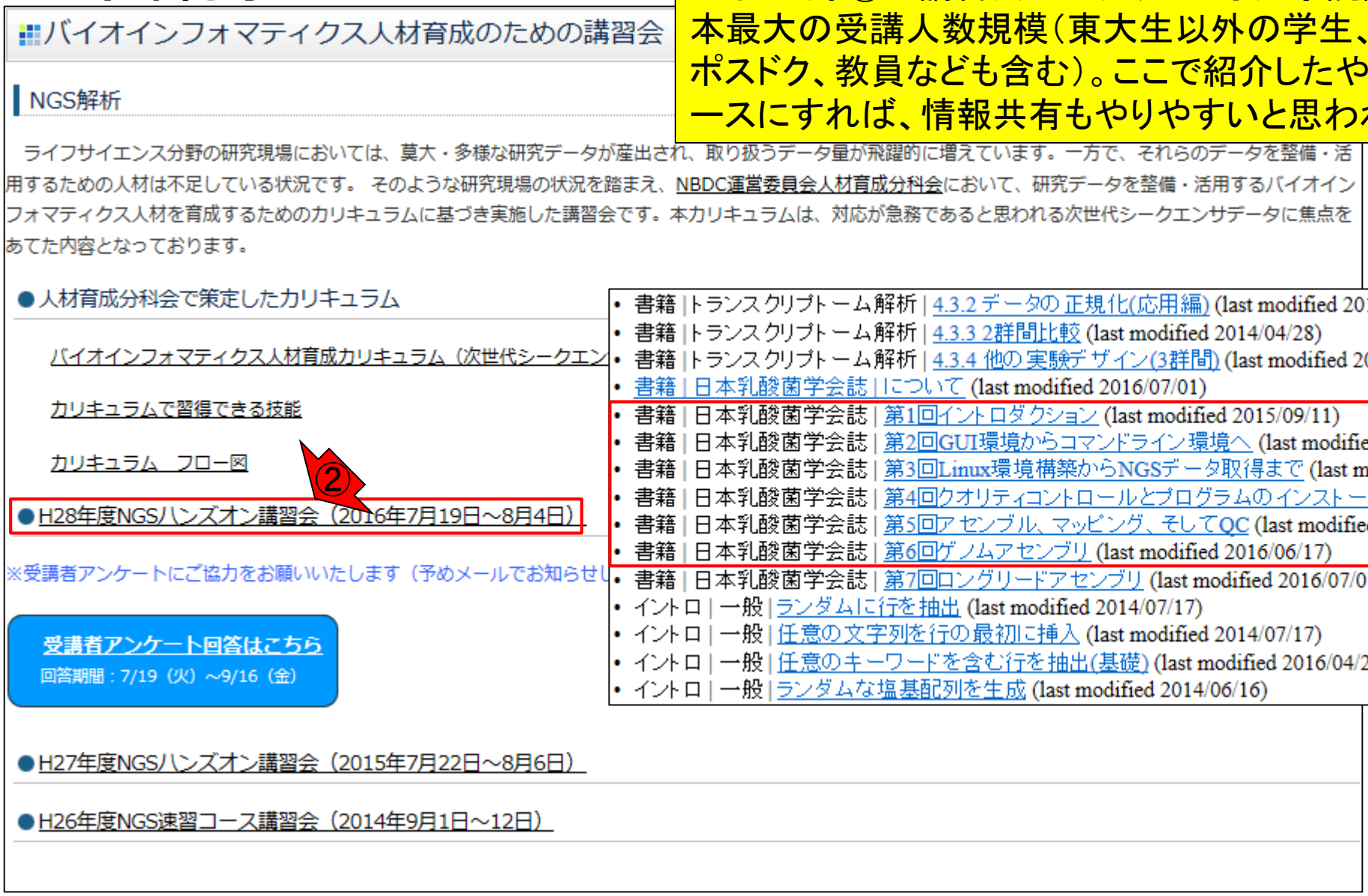
本講義では…

門田担当分は、(主にNGS解析を意識した)**Linuxスキルの習得が主目的**。何かをやったという達成感も得られるように、実際のNGSデータの一部を用いてゲノムアセンブリまで行う

- キーワード
  - NGS, **Linux**, バイオインフォマティクス, 仮想環境, Bio-Linux, ゲノムアセンブリ
- Linux
  - WindowsやMacintoshと同じく、OSの一種
  - バイオインフォマティクス分野でよく利用される
  - 「Windowsのコマンドプロンプト」や「Macintoshのターミナル」と同じく、lsやcdなどのLinuxコマンドを知らなければ何もできないため、慣れるまでが大変
  - 使いこなせれば、最先端の解析用プログラムを自在にインストール・利用可能となり、効率的かつ**通り一辺倒でない**データ解析も可能となる
- 次世代シークエンサ(NGS)解析
  - NGSとは、大量の塩基配列を出力する実験機器(Next-Generation Sequencer)またはその技術を指す。主にゲノム解析やトランスクリプトーム解析と呼ばれる分野で利用されている
  - 塩基配列解析用プログラムは、UNIX(今のLinux)環境で動作するものが多かった歴史的背景などから、現在でもLinux上で動くプログラムがまず最初に開発される場合が多い

本講義では…

実際に行うのは①の一部。それでも実際に手を動かし門田提供教材のノリに慣れておくことで、②の自習にもつながる。②の講習会やアグリバイオ大学院講義は、日本最大の受講人数規模（東大生以外の学生、社会人、ポスドク、教員なども含む）。ここで紹介したやり方をベースにすれば、情報共有もやりやすいと思われます



Contents

- イントロダクション
  - 概要、背景（NGS用カリキュラム、講習会）、Linuxスキル習得の意義
  - ウェブ情報（日本乳酸菌学会誌のNGS連載やNGS講習会資料）
- 実習環境に慣れる
  - 仮想環境での作業に慣れる
  - GUIとCUI（マウス操作かコマンド入力操作か）
  - ターミナルでの作業
  - 共有フォルダの概念を理解
- 練習
  - 作業ディレクトリの変更、練習用NGSデータファイルのダウンロード
  - ファイルの確認、de novoゲノムアセンブリ
  - BLAST検索
- 課題
  - グループごとに異なる課題ファイルを入力として、「ダウンロード、de novoアセンブリ、BLAST検索」を実行し、得られた結果をレポートにまとめて発表せよ
  - グループ1はkadai1.fasta、グループ2はkadai2.fasta、etc.

VirtualBoxを起動

連載第1回原稿のp88。Windows PC上でLinuxを動かす際に、①仮想化ソフトを利用。貸与PCには②VirtualBoxという仮想化ソフトがインストールされています。デスクトップ上にある③のアイコンをダブルクリックで起動

NGS データ解析を高速かつ効率的に実行するプログラムの多くは、Linux というデータ解析環境上でのみ動作する。それゆえ、バイオインフォマティシャンを目指す場合には、しばしば Linux 環境構築が最初の課題として与えられる。最近では、普段利用している Windows または Macintosh マシン上で Linux 環境を同時に利用するための手軽な手段も提供されている。この種の勉強を本格的に始めようとすると、慣れない用語に戸惑う読者は意外に多い。まず Linux とは何か？ Linux は、Windows や Mac OS などと同じオペレーティングシステム（以下、OS）の一種である。多少の誤解を恐れずに説明する。Windows や Macintosh が安定して使いやすい大衆車だとすると、Linux は乗りこなせると速いスポーツカーのようなものである。次に戸惑うのは Linux の種類（ディストリビューション）に関する記述である。Windows に Windows 7 や Windows 8.1 があるように、Linux にも様々な種類がある。Linux（スポーツカー）に関する情報収集をしているつもりでも、いつのまにか説明内容が Ubuntu（うぶんつ、と読む）とか LinuxMint とか CentOS などに切り替わっていてだんだん混乱してくる場合が多いだろう。Ubuntu と

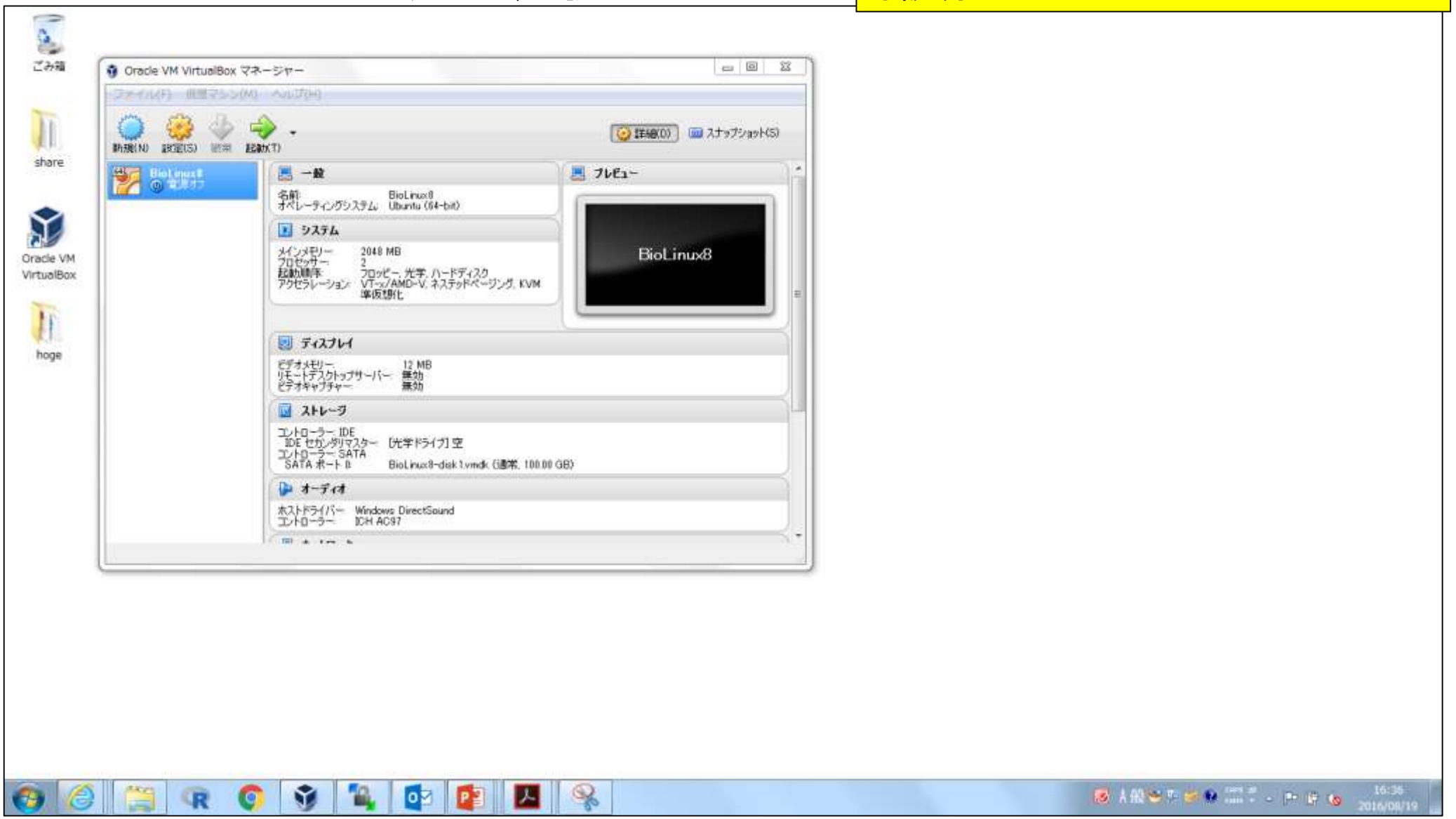
一種のコントローラと解釈してもよいだろう。通常は 1 台の車しか運転できないが、仮想化ソフトというコントローラのおかげで 2 台の車を同時に操作することができる。しばしば動作が不安定になるのは 2 台の車を同時に操作するという複雑なことをしているためであると解釈すればよい。

代表的な無料の仮想化ソフトとしては、VMware 社の VMware Player とオラクル社の VirtualBox の 2 つが挙げられる。VMware Player は非営利に限りフ[illegible]であり、歴史が古く安定している。VirtualBox は比較的最近開発されたソフトウェアで利便性が高い一方、やや動作が不安定だという声をきく。アグリバイオの講義で使用する PC には VMware Player をインストールしている。理由は、ウェブ上で収集可能な情報量が多くトラブル対応が比較的容易だったからである（図 1）。

アセンブルなど多くの NGS 解析手法を自在に操りたいバイオインフォマティクス中級～上級を目指したい場合には、是非 Linux 環境構築に挑戦してほしい。最初は手持ちのノート PC で十分である。cd, ls, grep などの独特なコマンド操作、見慣れない GUI 画面、Windows 向けのソフトウェアなどとは異なりダブルクリックでプログラムのイ

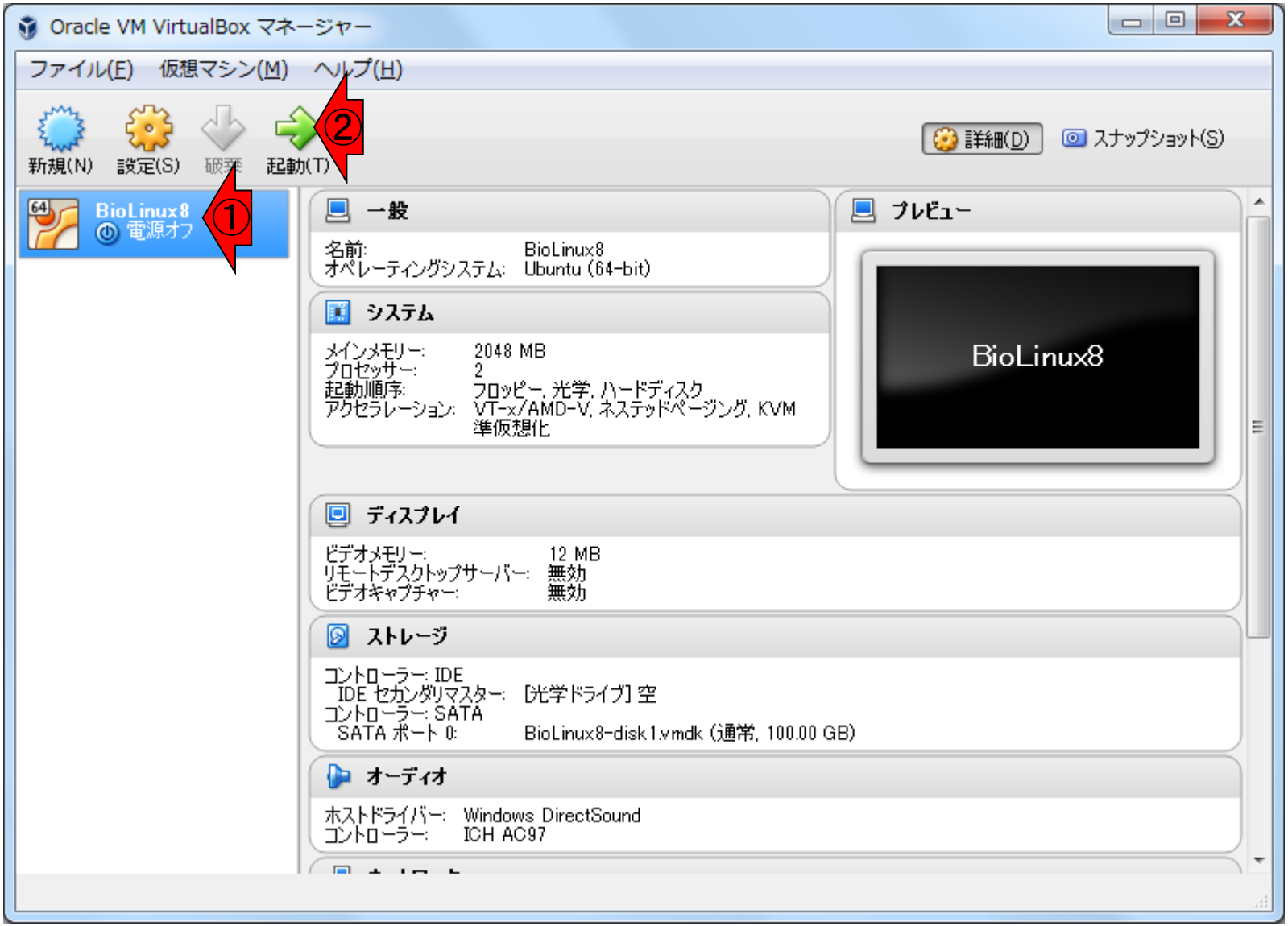
VirtualBox起動後の状態

こんな感じになります。連載第3回ウェブ資料(JSLAB3_suppl_....pdf)と同じような説明



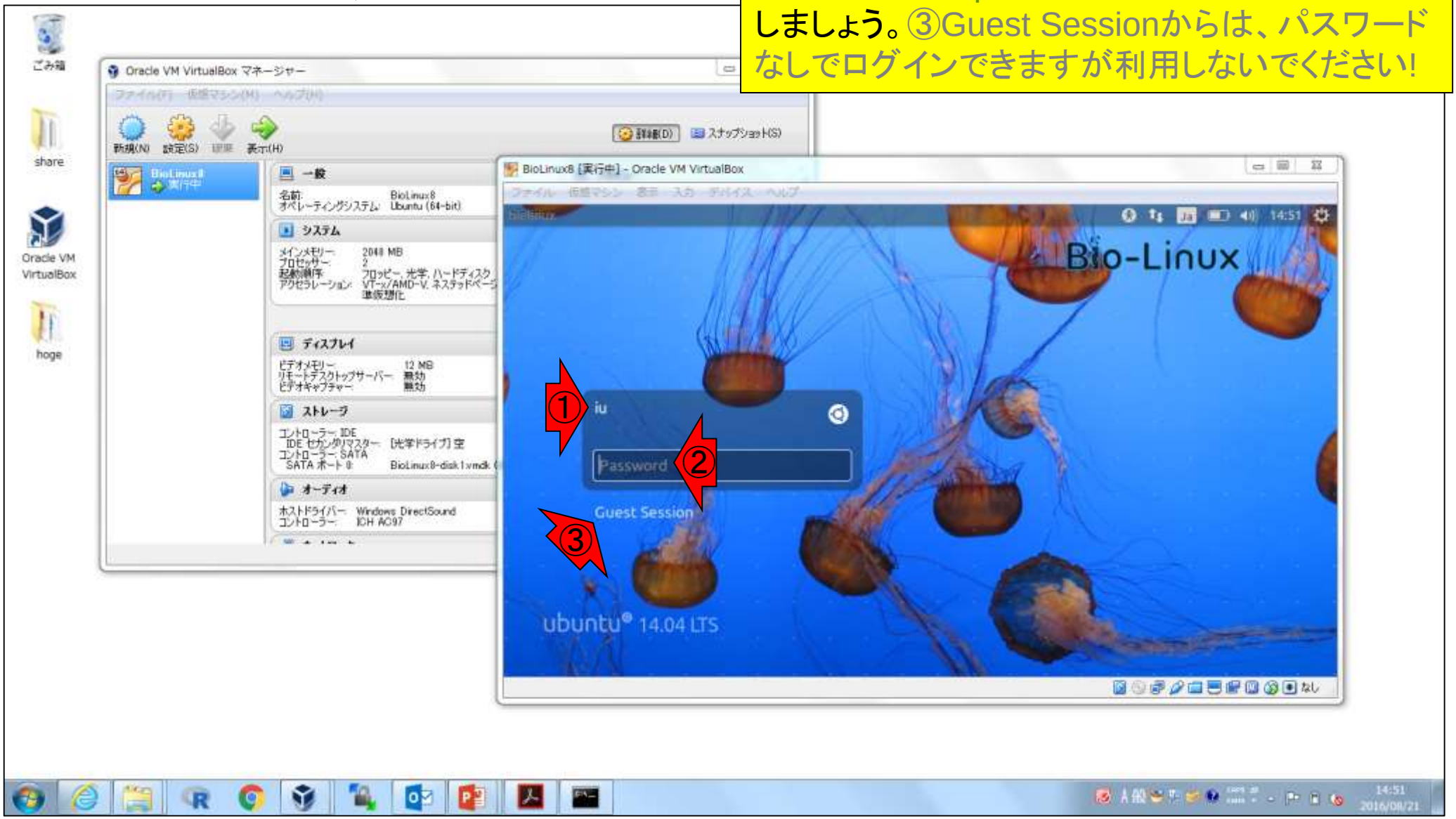
①BioLinux8というのが見えるので、②起動

Linuxを起動



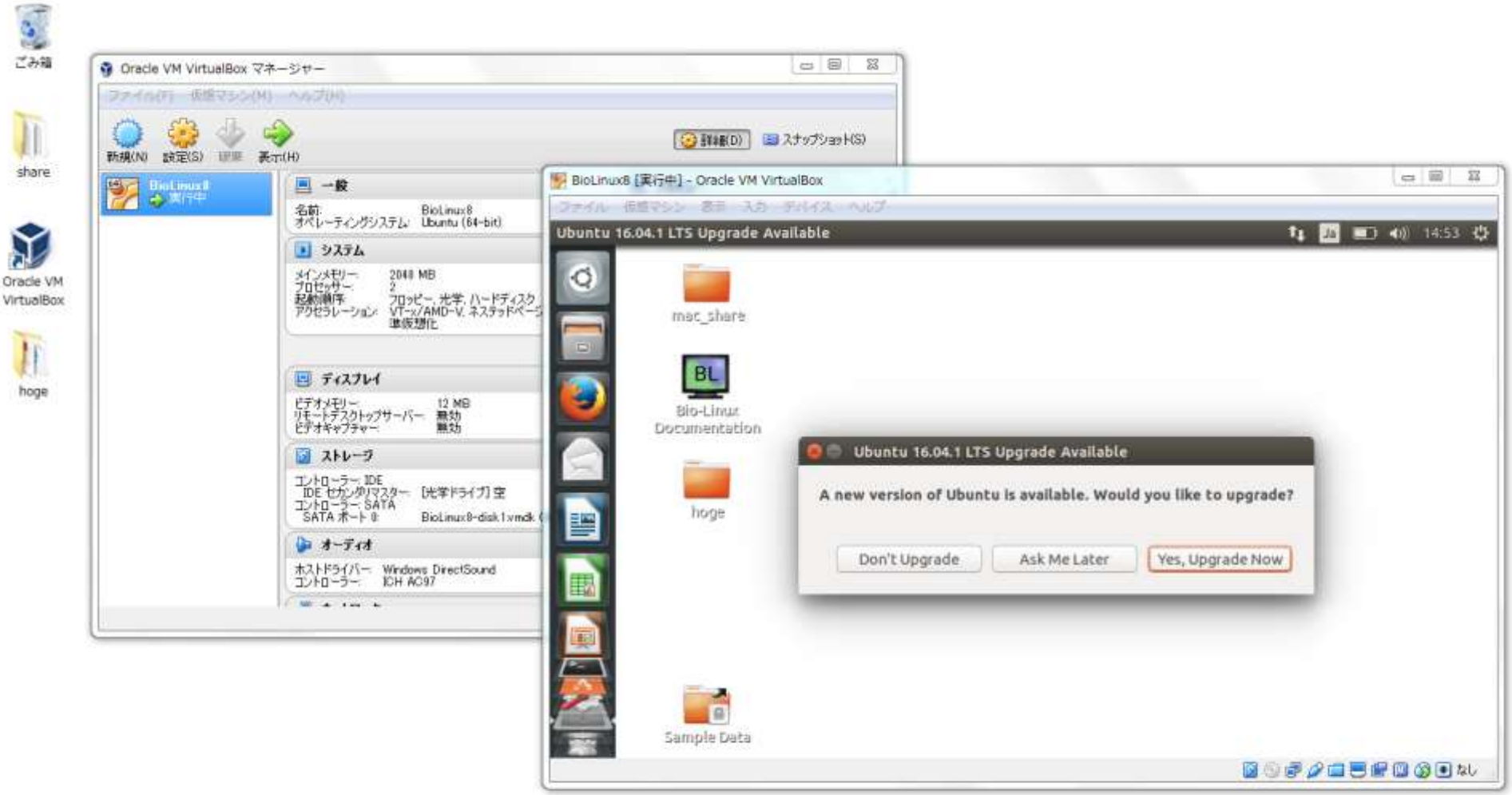
Linux起動後の状態

約1分でこのような状態になります。Windowsのログイン画面と同じ状態です。①ユーザ名はiu、②パスワードはpass1409です。Linuxにログインしましょう。③Guest Sessionからは、パスワードなしでログインできますが利用しないでください!



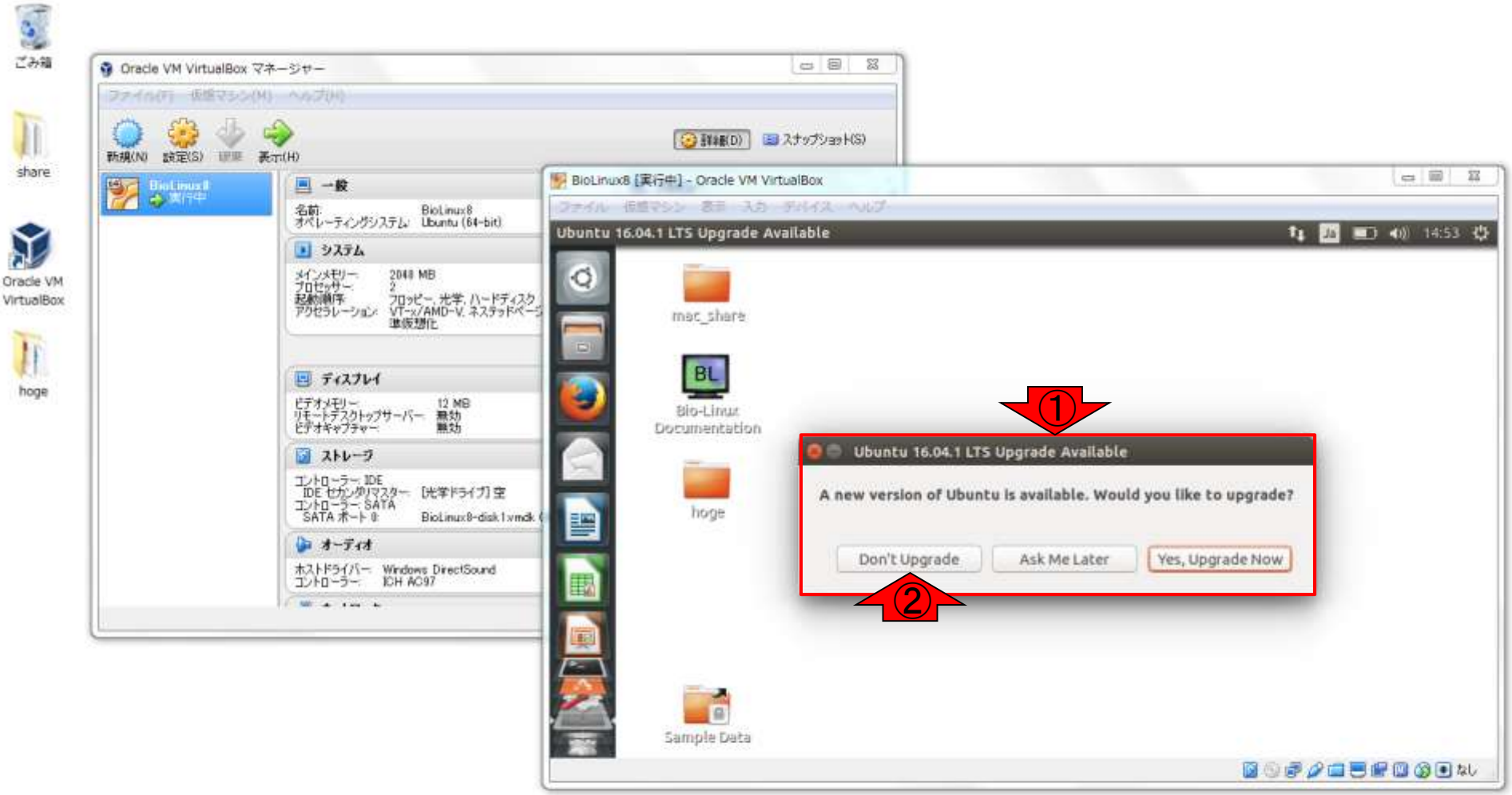
約1分でこのような状態になります

ログイン後の状態



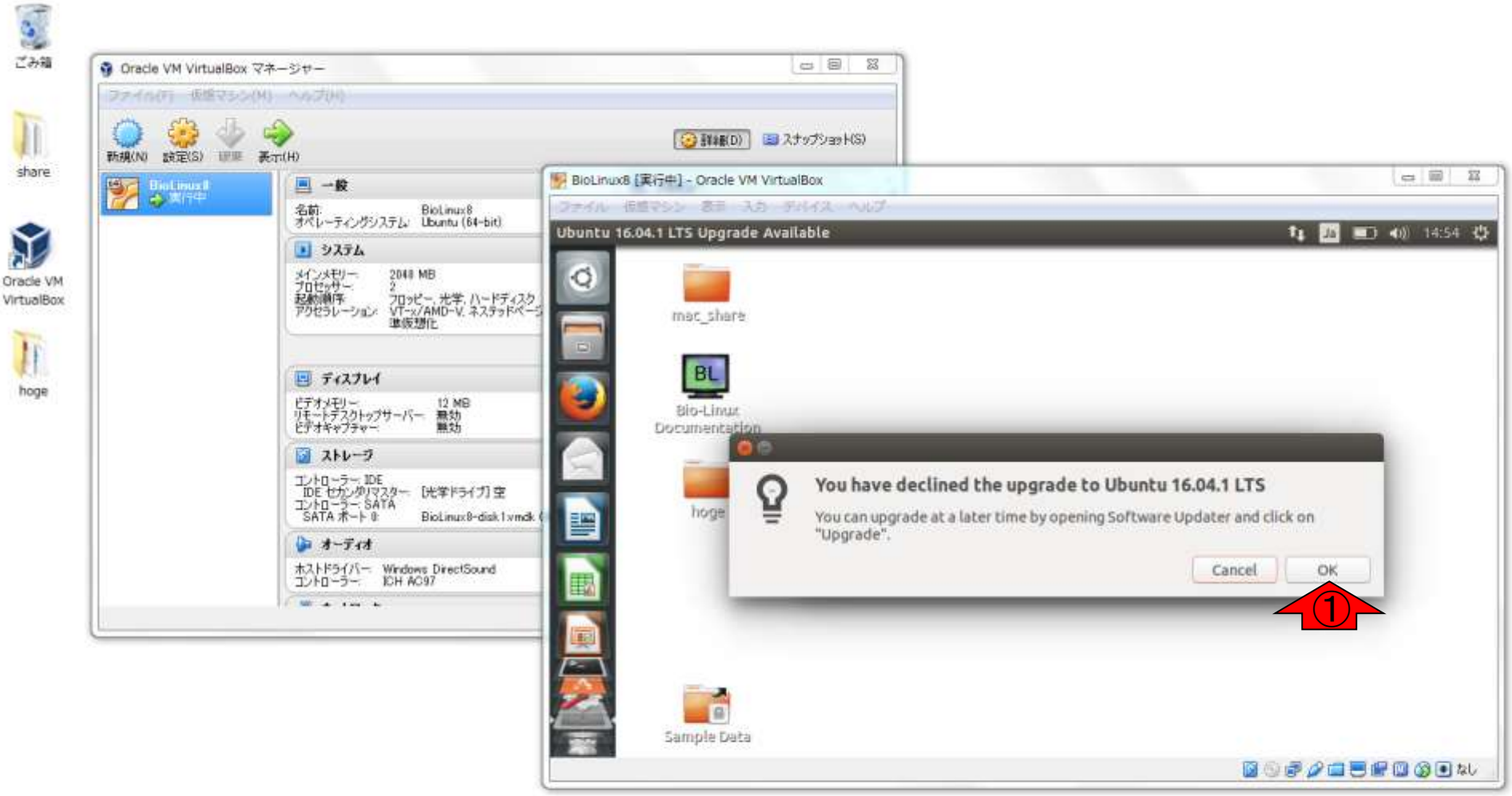
ログイン後の状態

①のポップアップは「Windows10にアップグレードするか？」という類のものです。②Don't Upgrade



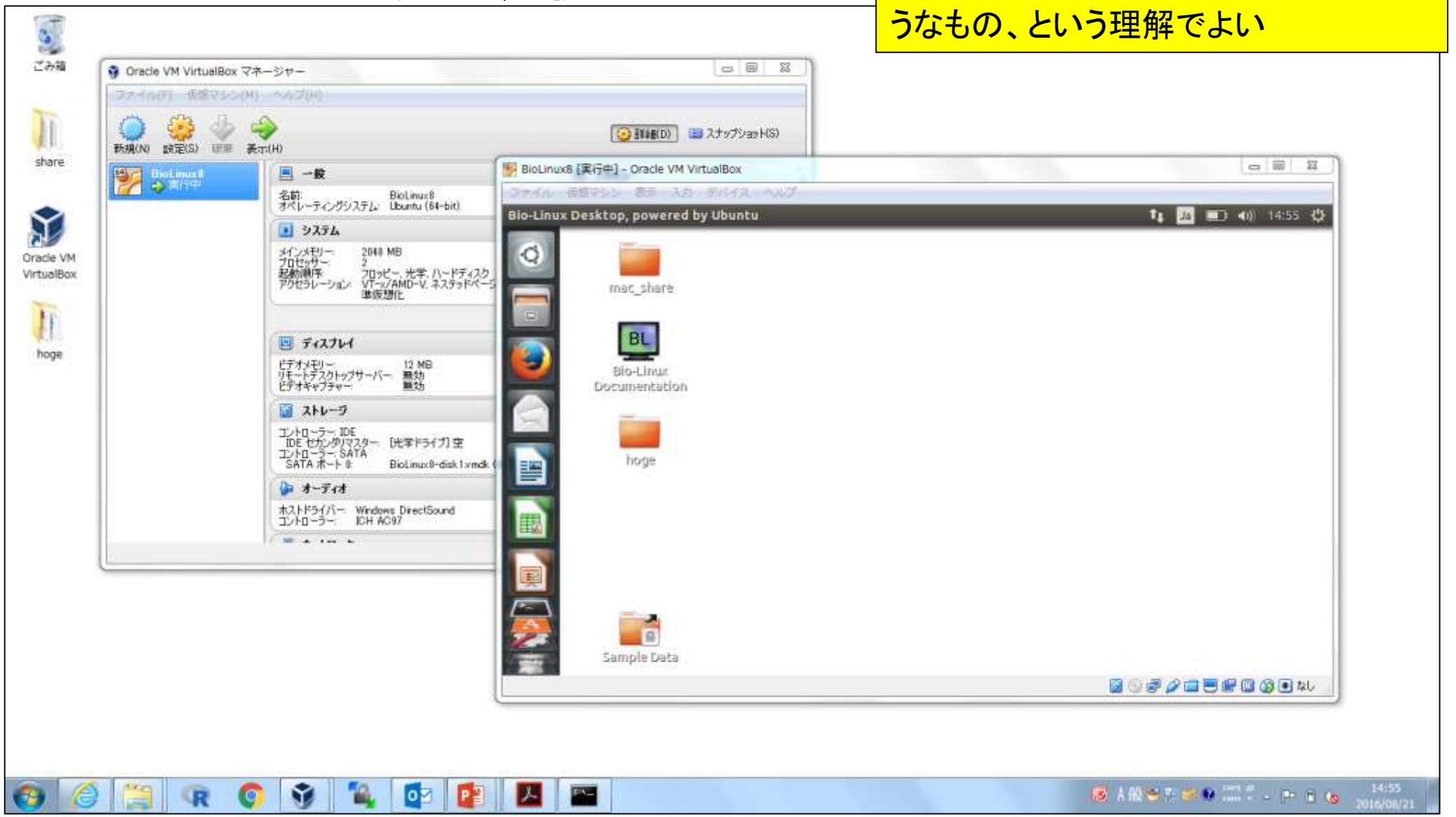
①OK

ログイン後の状態



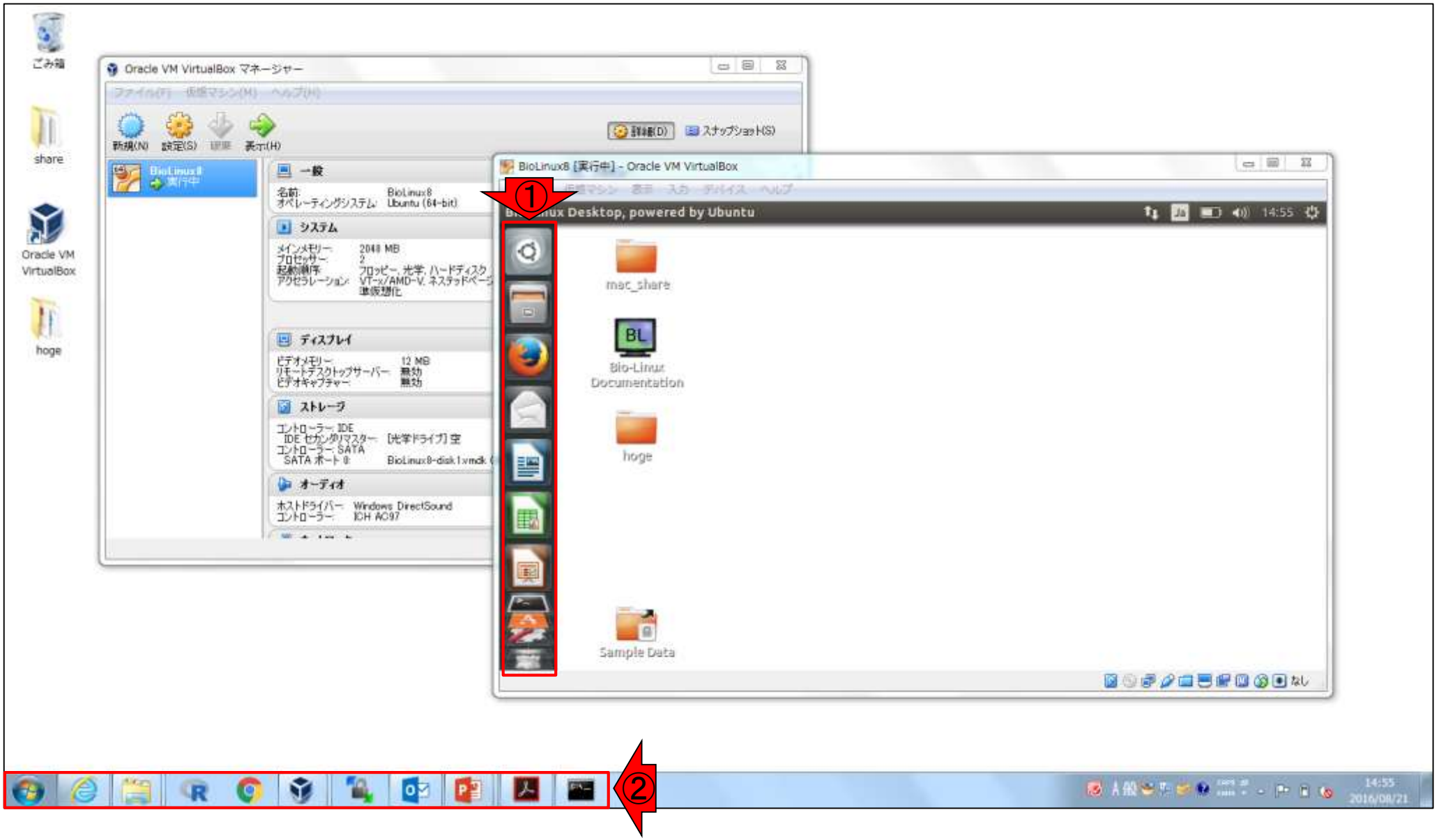
BioLinux8起動後の状態

Windows（ホストOS）上で、BioLinux8というLinux（ゲストOS）が立ち上がっている状態。VirtualBoxは仲介役のようなもの、という理解でよい



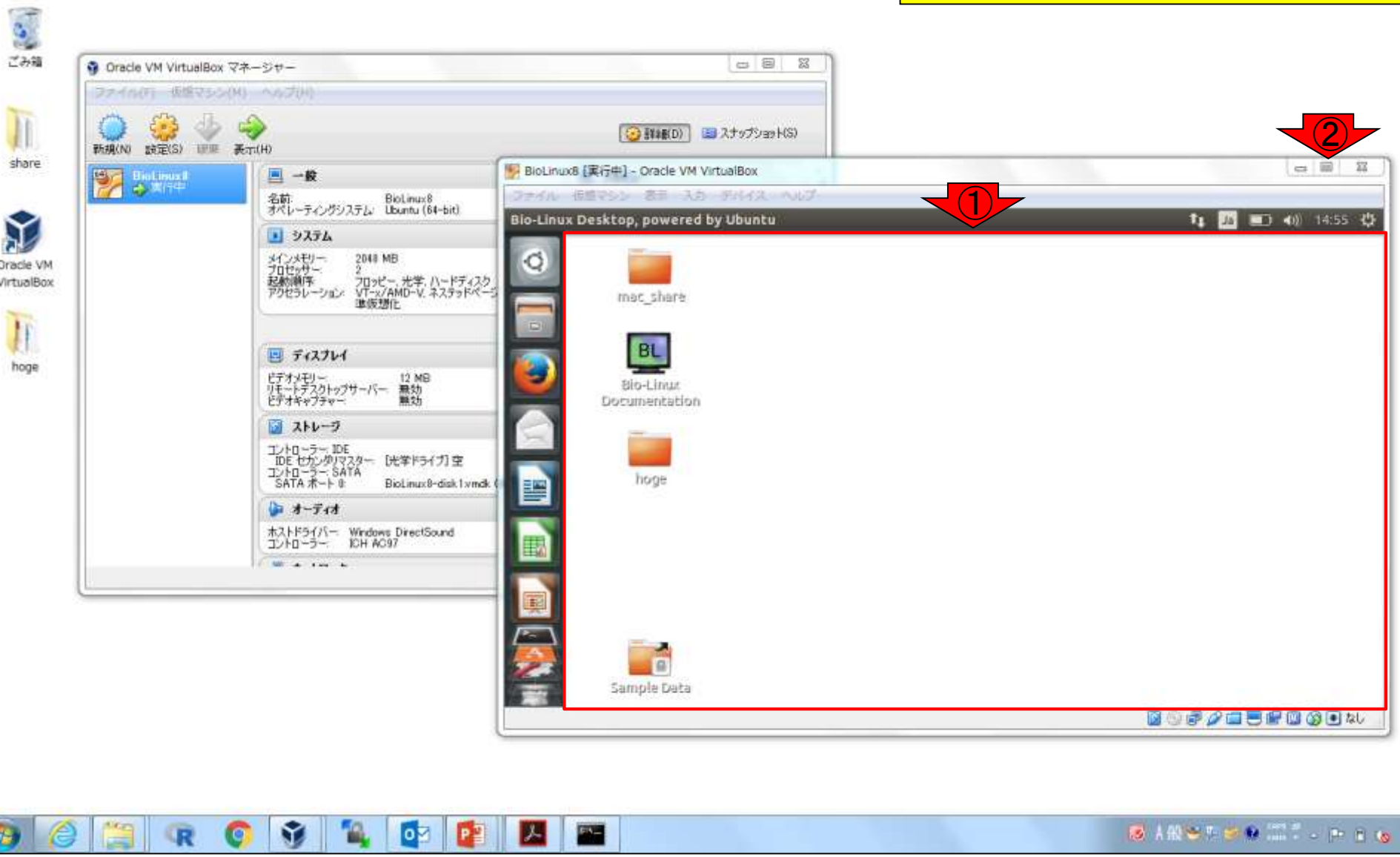
対応関係

①BioLinux8の赤枠部分は、②Windowsのタスクバーと同じようなもの



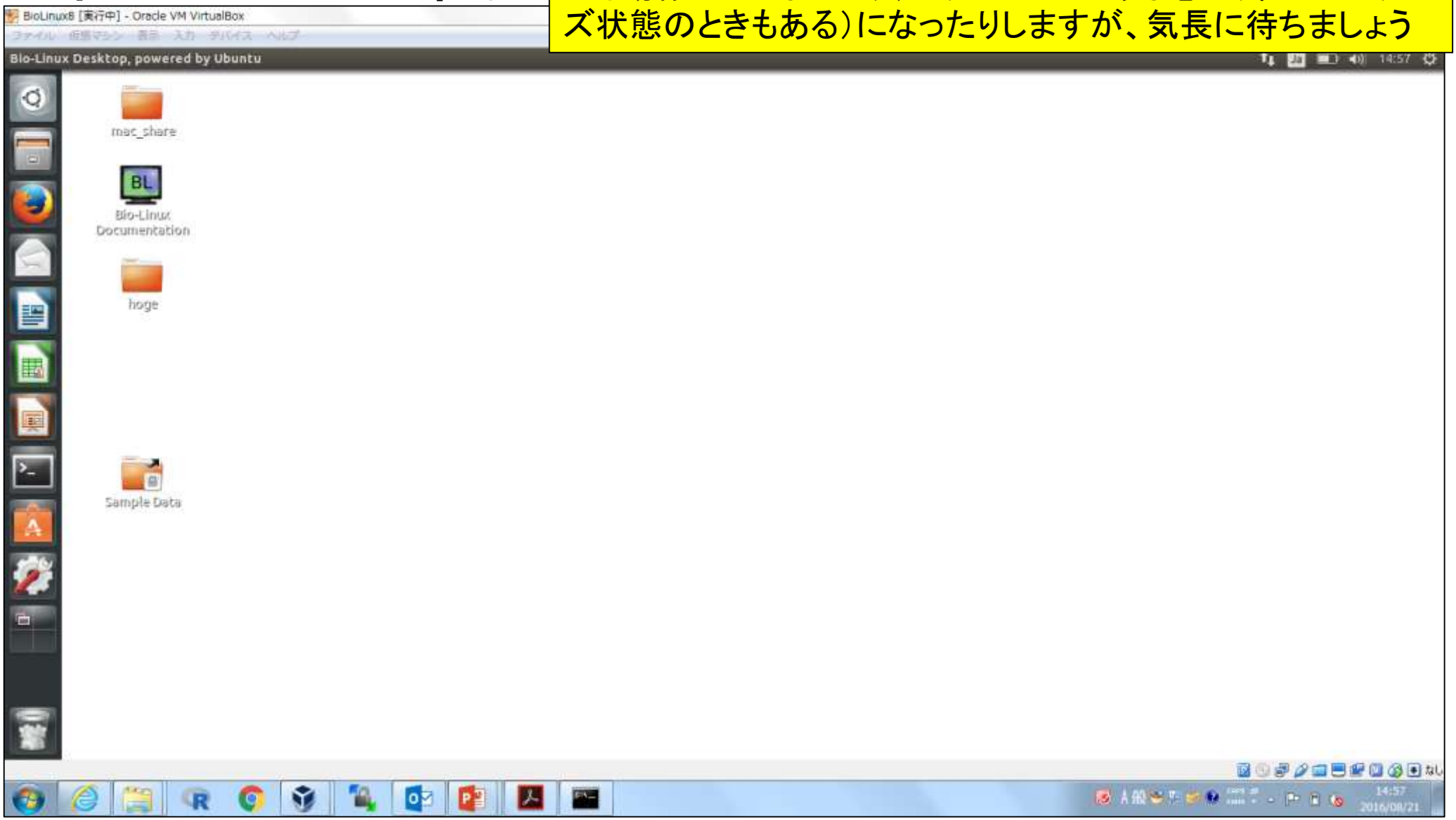
対応関係

①赤枠部分はBioLinux8のデスクトップ画面に相当します。②の部分を押してBioLinux8の画面を最大化すれば…



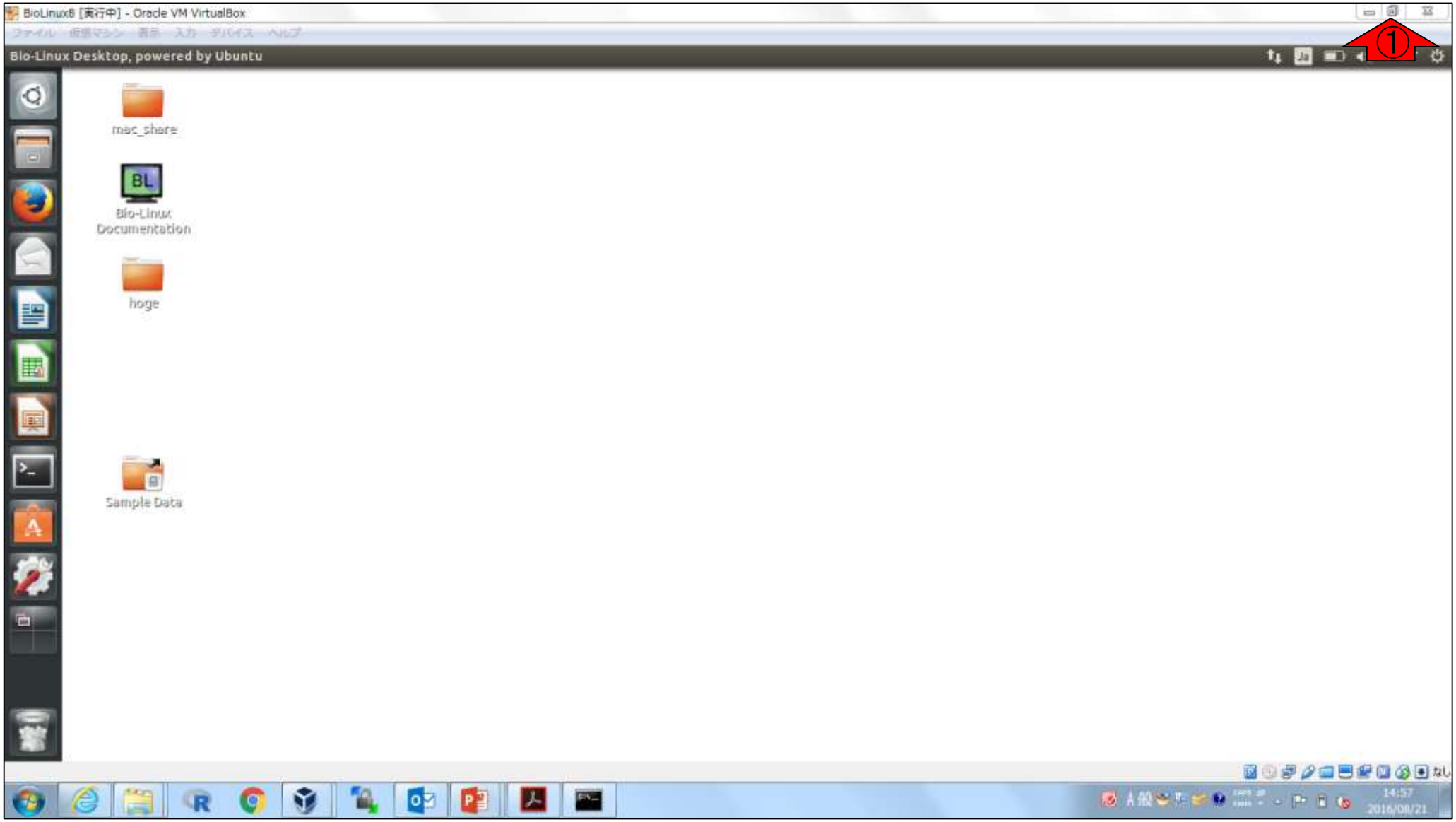
仮想Linux環境

仮想的にLinux環境で仕事をしているのと同じような感じになります。ただし、Windows上でLinuxを動かしているので、どうしても動作が重くなったり、フリーズしたような感じ（本当にフリーズ状態のときもある）になったりしますが、気長に待ちましょう



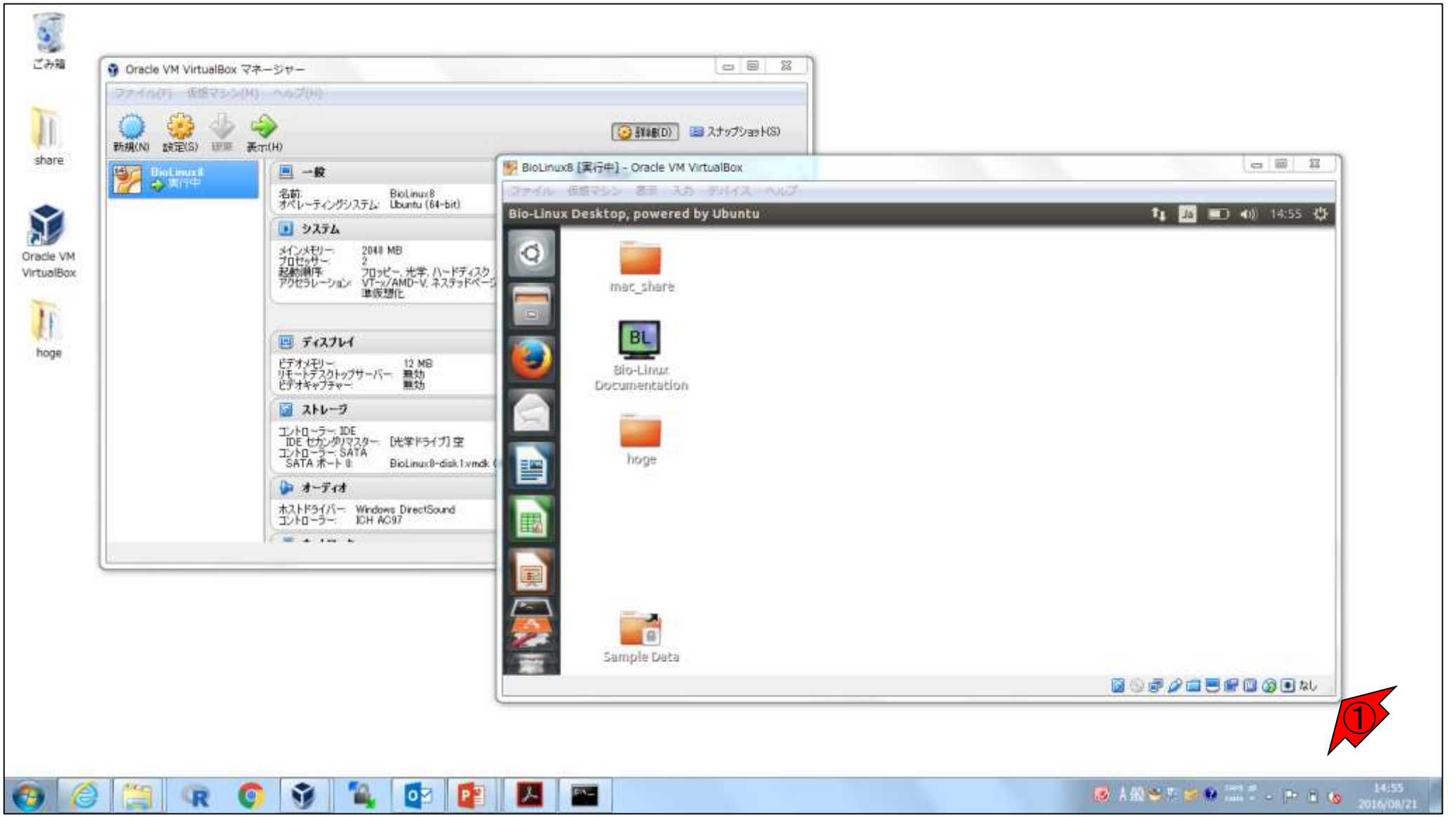
①一旦画面サイズを元に戻す

仮想Linux環境



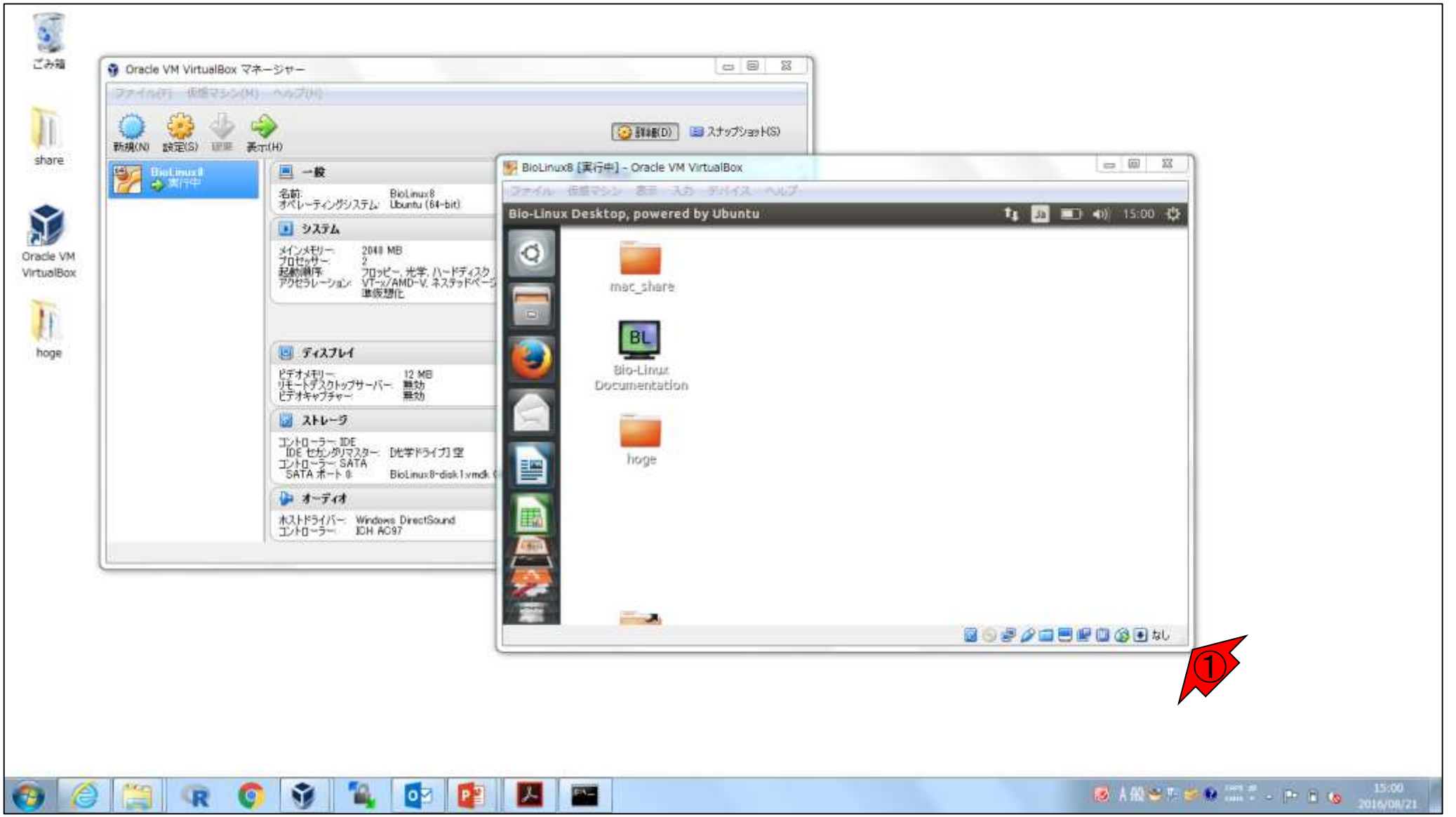
いろいろと…慣れです

BioLinux8のGUI画面サイズを変更
すべく、①の部分を動かしてみよう



いろいろと…慣れです

こんな感じにしたり、ガスガス変えまくっていると…動作が不安定になって落ちます

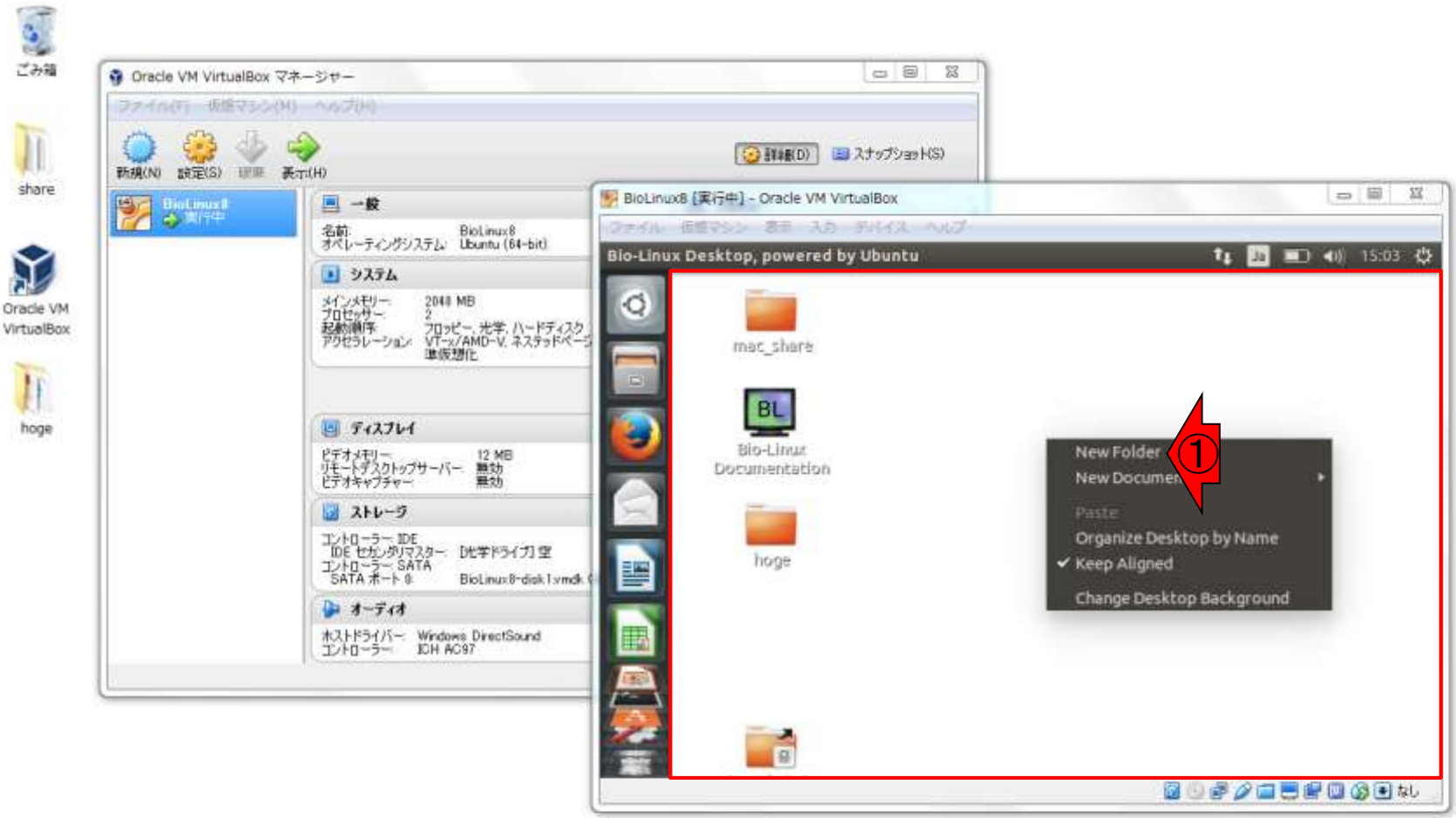


いろいろと…慣れです

ログイン画面に戻った状態です。最初はこの程度で落ちる不安定さにイラッとしますが、慣れです。パスワード(pass1409)を打ち込んでログインし直しましょう。この経験から、画面サイズを変更するときにはこういうことも起こるということを学んだ

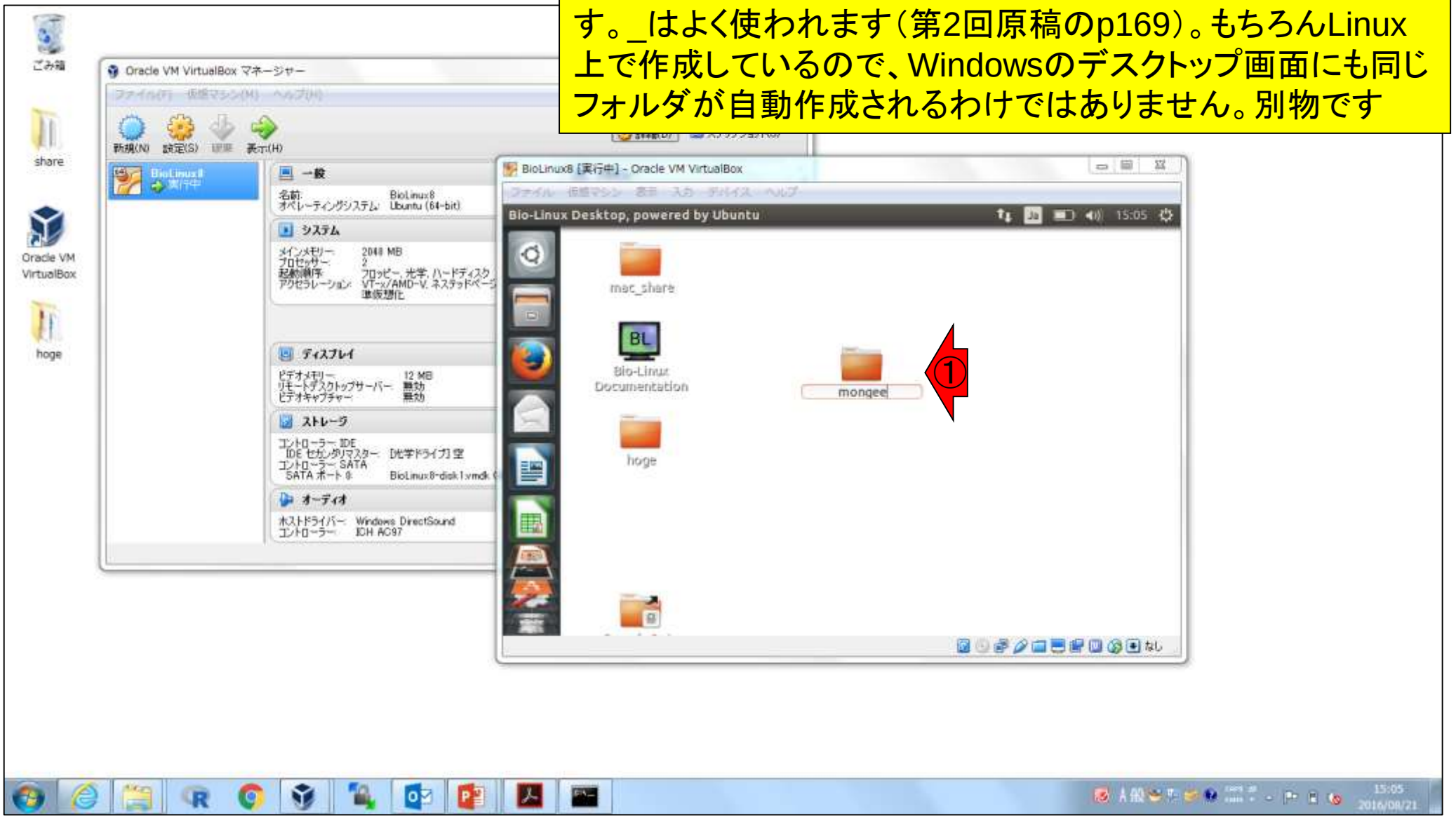
フォルダ作成

任意の名前のフォルダを作成してみましょう。赤枠内で右クリックし、①New Folder



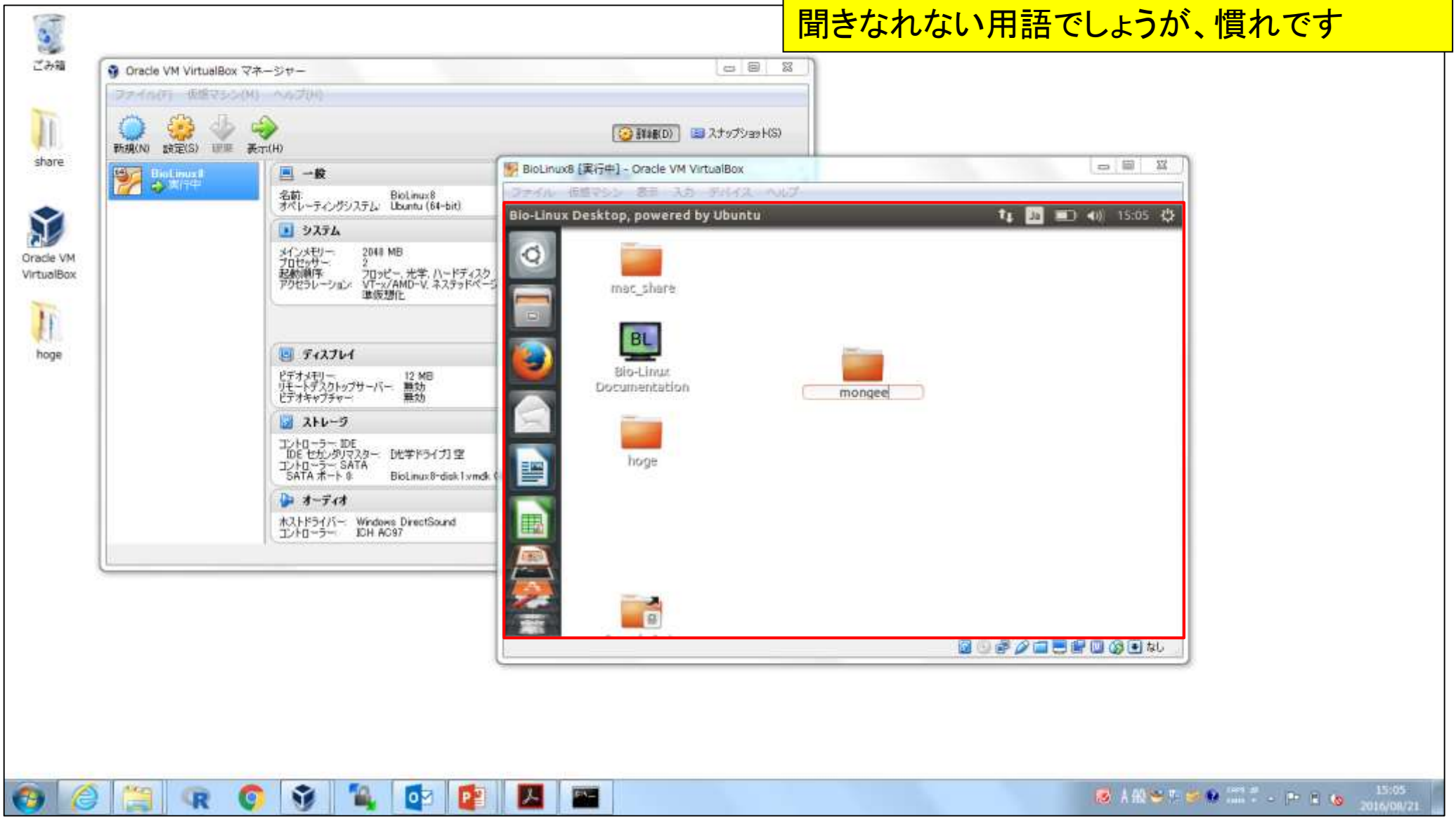
フォルダ作成

私はmongeeというフォルダ名にしました。Linuxの世界では、フォルダ名やファイル名に、通常日本語は利用しません。また、'&%¥*?などの特殊文字やスペースも使わないのが常識です。_はよく使われます（第2回原稿のp169）。もちろんLinux上で作成しているので、Windowsのデスクトップ画面にも同じフォルダが自動作成されるわけではありません。別物です



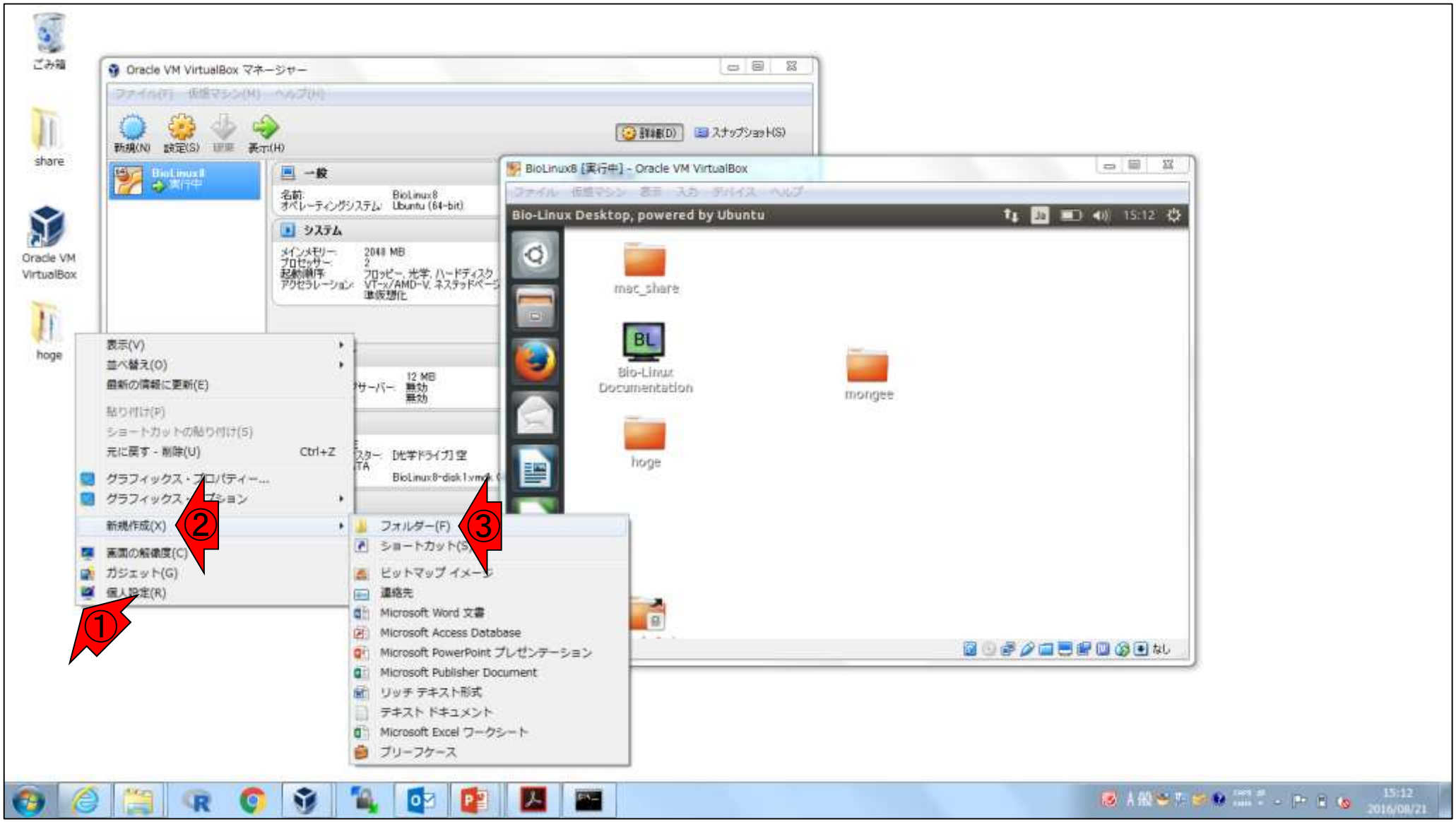
ゲストとホスト

今はWindows上でLinuxを動かしています。
赤枠内がLinux環境（ゲストOS環境）で、それ以外がWindows環境（ホストOS環境）です。
聞きなれない用語でしょうが、慣れです



念のため

Windows（ホストOS環境）上で、「①右クリック、②新規作成、③フォルダー」の流れで新規フォルダの作成が可能です

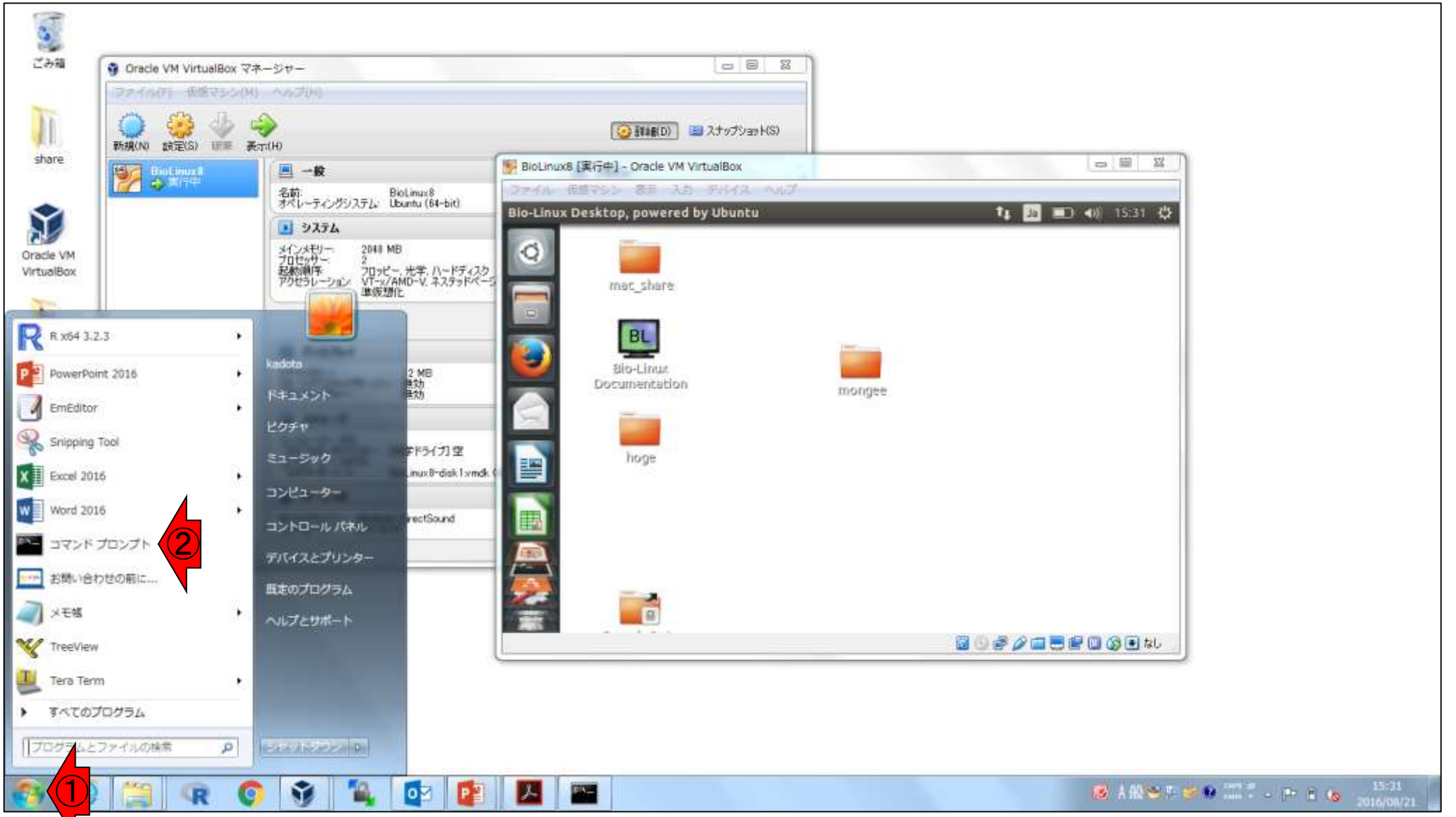


Contents

- イントロダクション
  - 概要、背景（NGS用カリキュラム、講習会）、Linuxスキル習得の意義
  - ウェブ情報（日本乳酸菌学会誌のNGS連載やNGS講習会資料）
- 実習環境に慣れる
  - 仮想環境での作業に慣れる
  - GUIとCUI（マウス操作かコマンド入力操作か）
  - ターミナルでの作業
  - 共有フォルダの概念を理解
- 練習
  - 作業ディレクトリの変更、練習用NGSデータファイルのダウンロード
  - ファイルの確認、de novoゲノムアセンブリ
  - BLAST検索
- 課題
  - グループごとに異なる課題ファイルを入力として、「ダウンロード、de novoアセンブリ、BLAST検索」を実行し、得られた結果をレポートにまとめて発表せよ
  - グループ1はkadai1.fasta、グループ2はkadai2.fasta、etc.

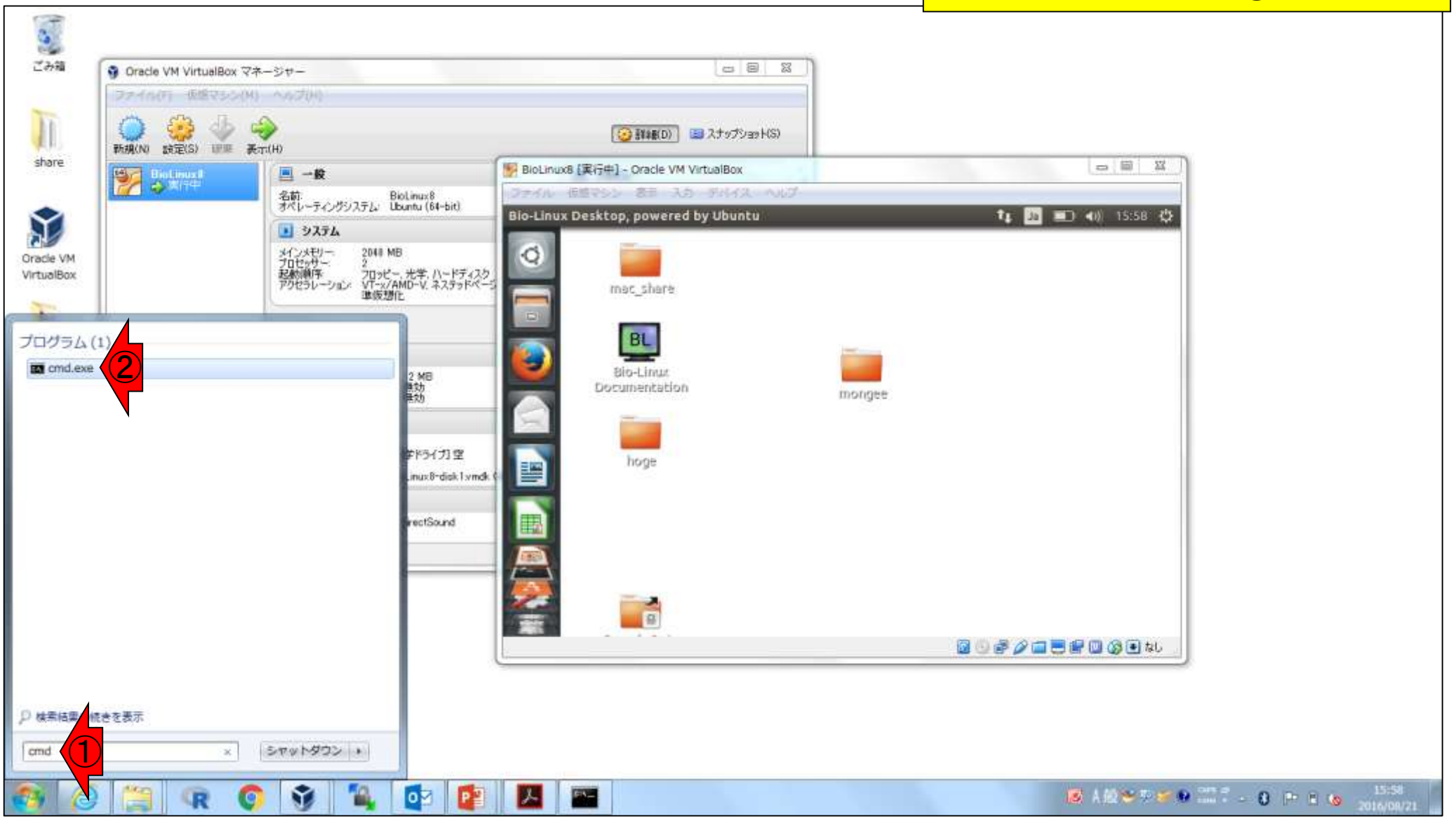
GUIとCUI

①スタートメニューから、②コマンドプロンプトを選んで起動しましょう



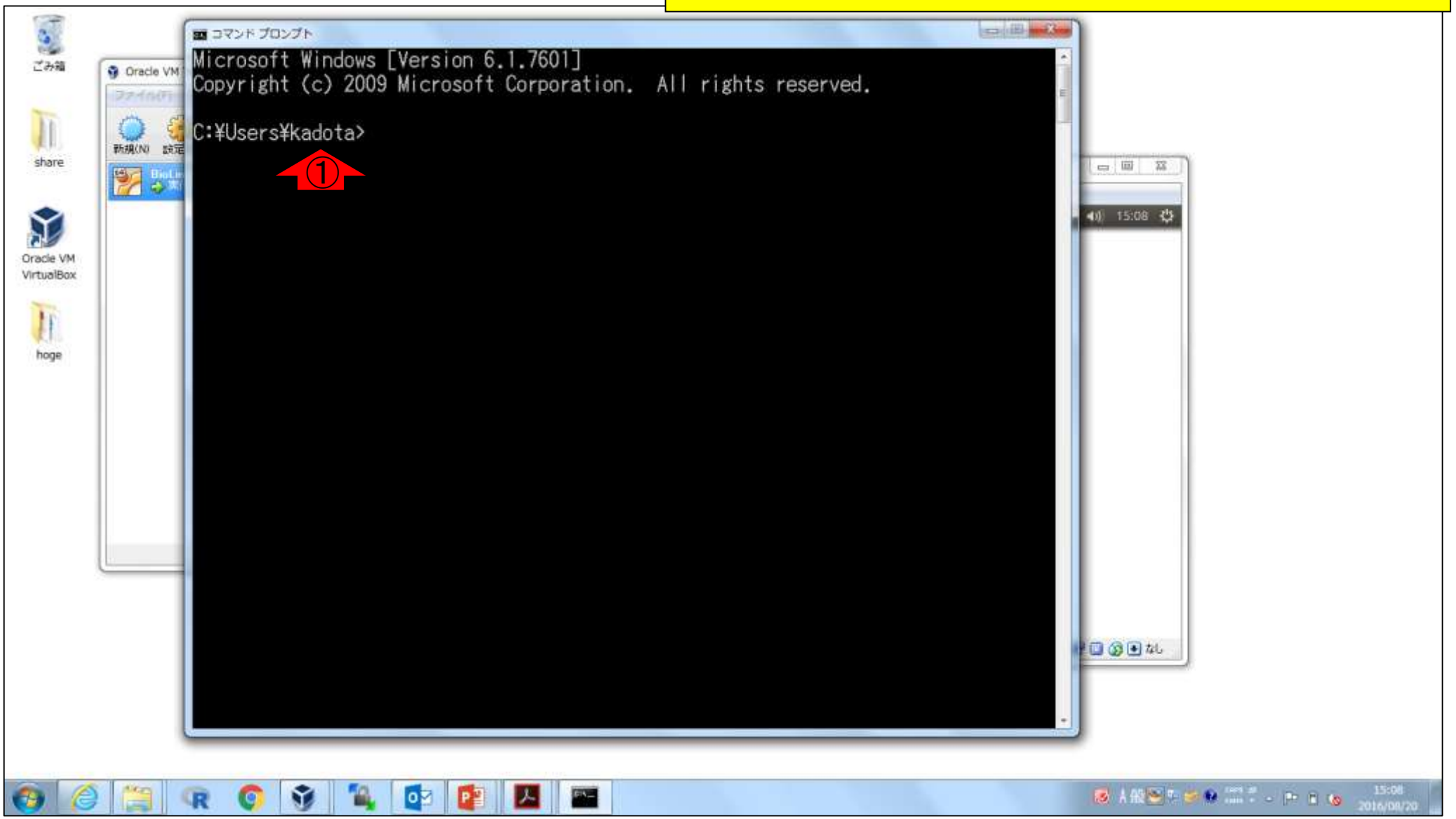
GUIとCUI

コマンドプロンプトがすぐに見つからない場合は、①検索窓で、cmdと打つのでもよいです。②cmd.exe



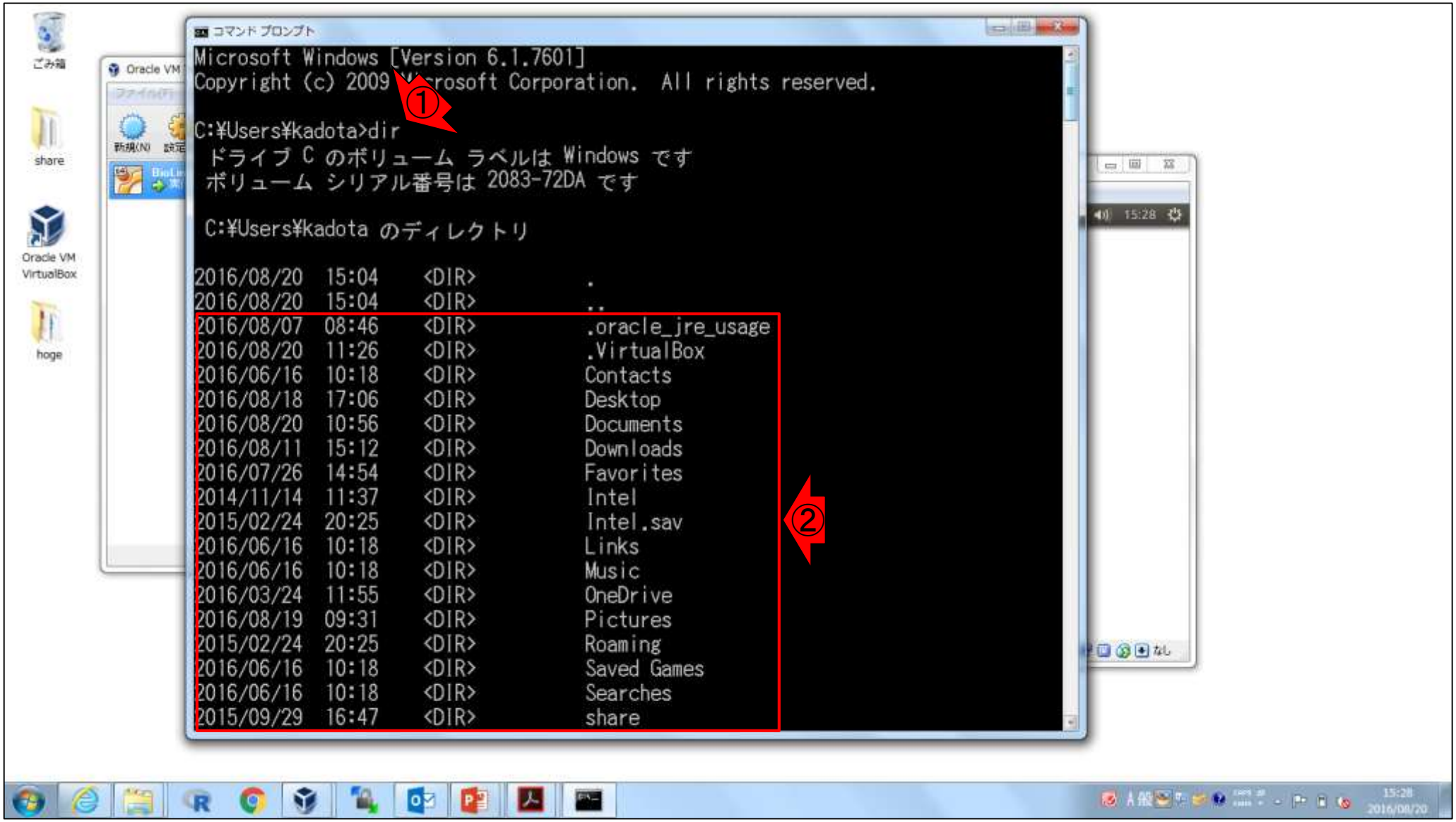
GUIとCUI

コマンドプロンプト起動後の状態。貸与PCはユーザ名iuなので、①の部分が「C:Users¥iu」。Macintoshのヒトは、「ターミナル」を起動するのと同じと思えばよい



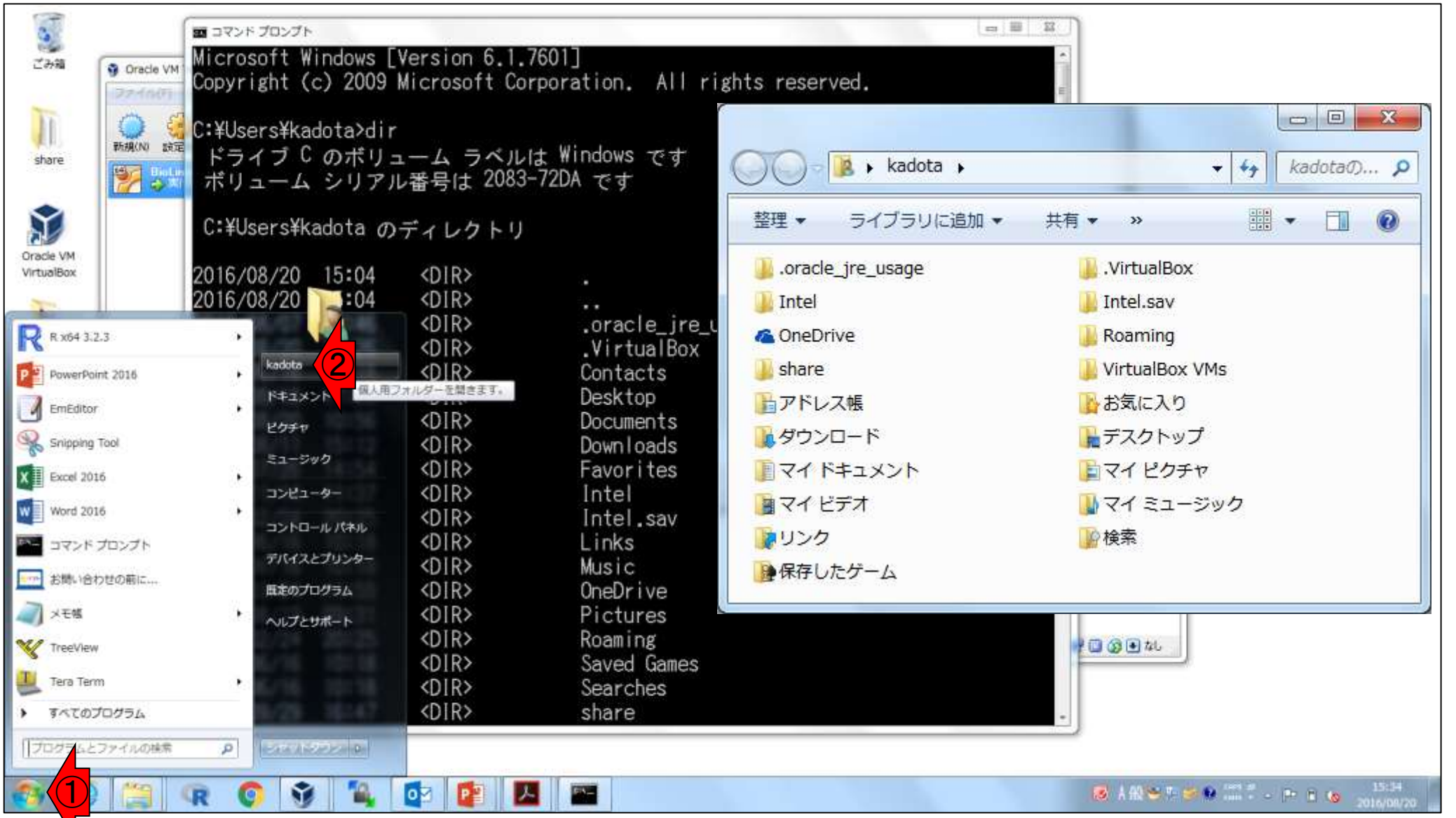
GUIとCUI

①dirと打って、リターンキーを押す。
②赤枠で見ているものは…



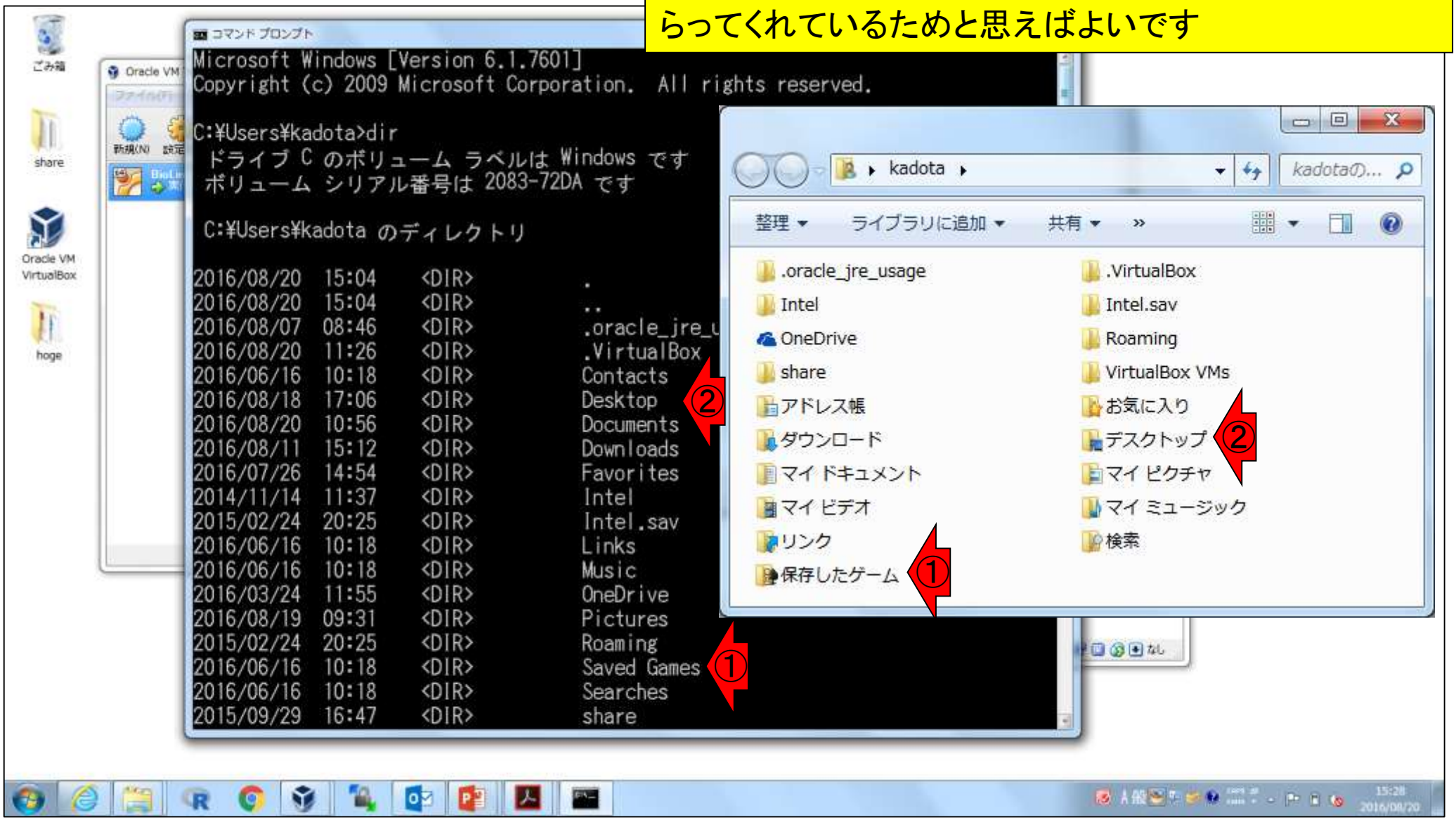
GUIとCUI

①スタートメニューの右上にある、②ユーザ名kadotaの「ホームディレクトリ」の中身です



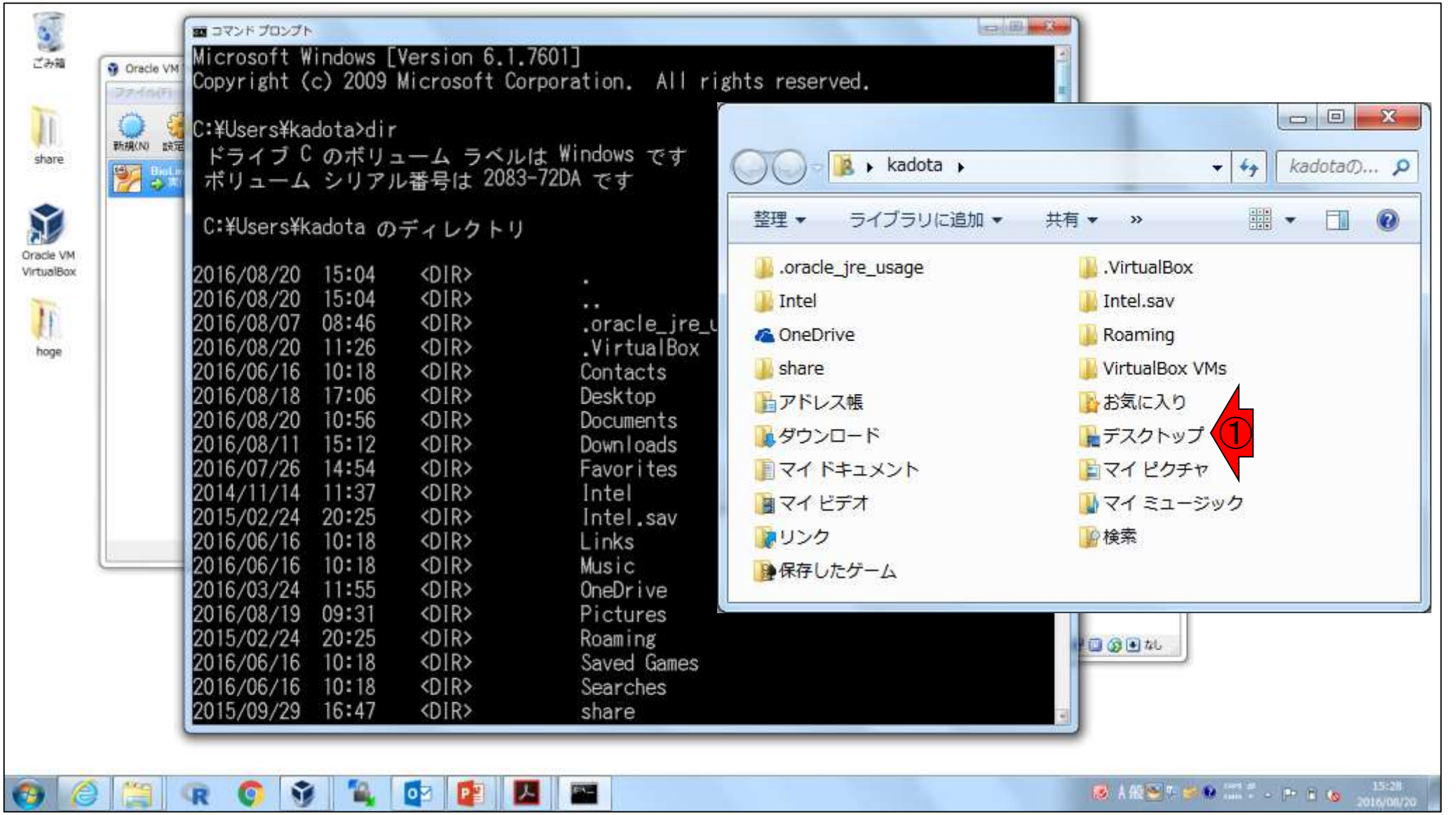
GUIとCUI

①「Saved Games ⇔ 保存したゲーム」、②「Desktop ⇔ デスクトップ」などと「English ⇔ 日本語」の変換が自動でなされていますが、これはWindows側でよきに計らってくれているためと思えばよいです



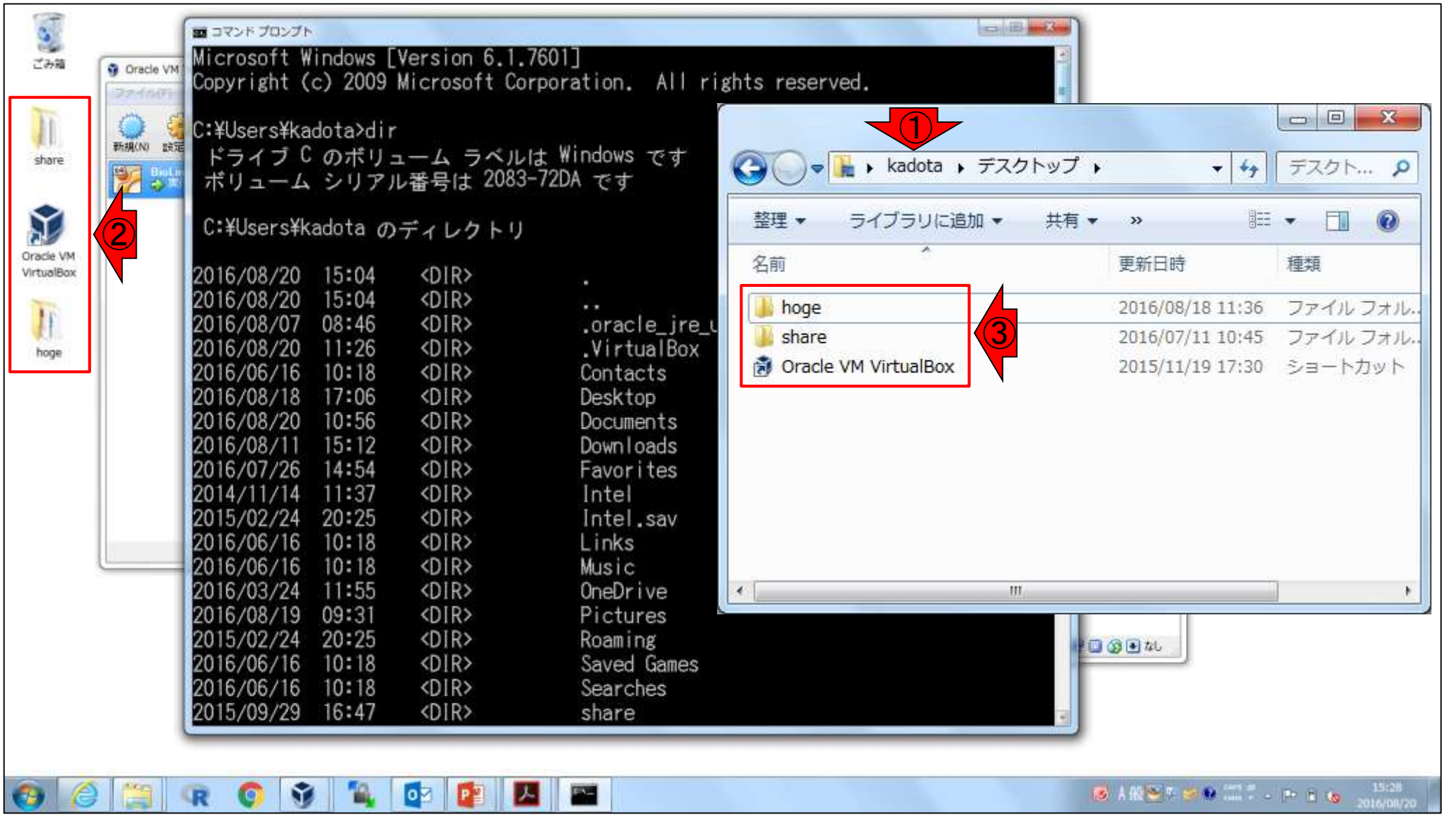
①デスクトップをダブルクリックすると…

GUIとCUI



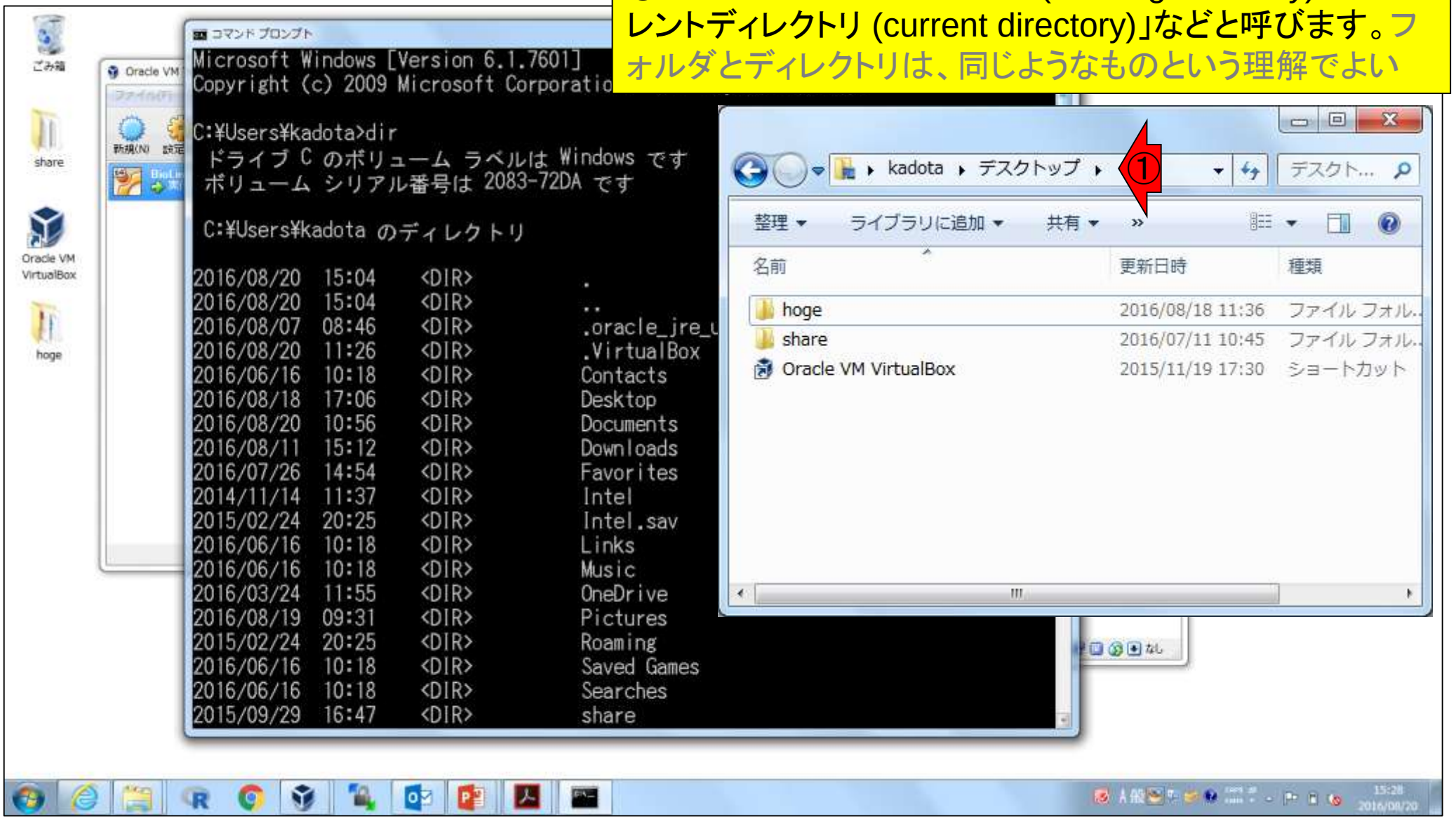
GUIとCUI

①kadotaのPC環境では、②赤枠の3つしかないので、それに相当するものが③で見えています



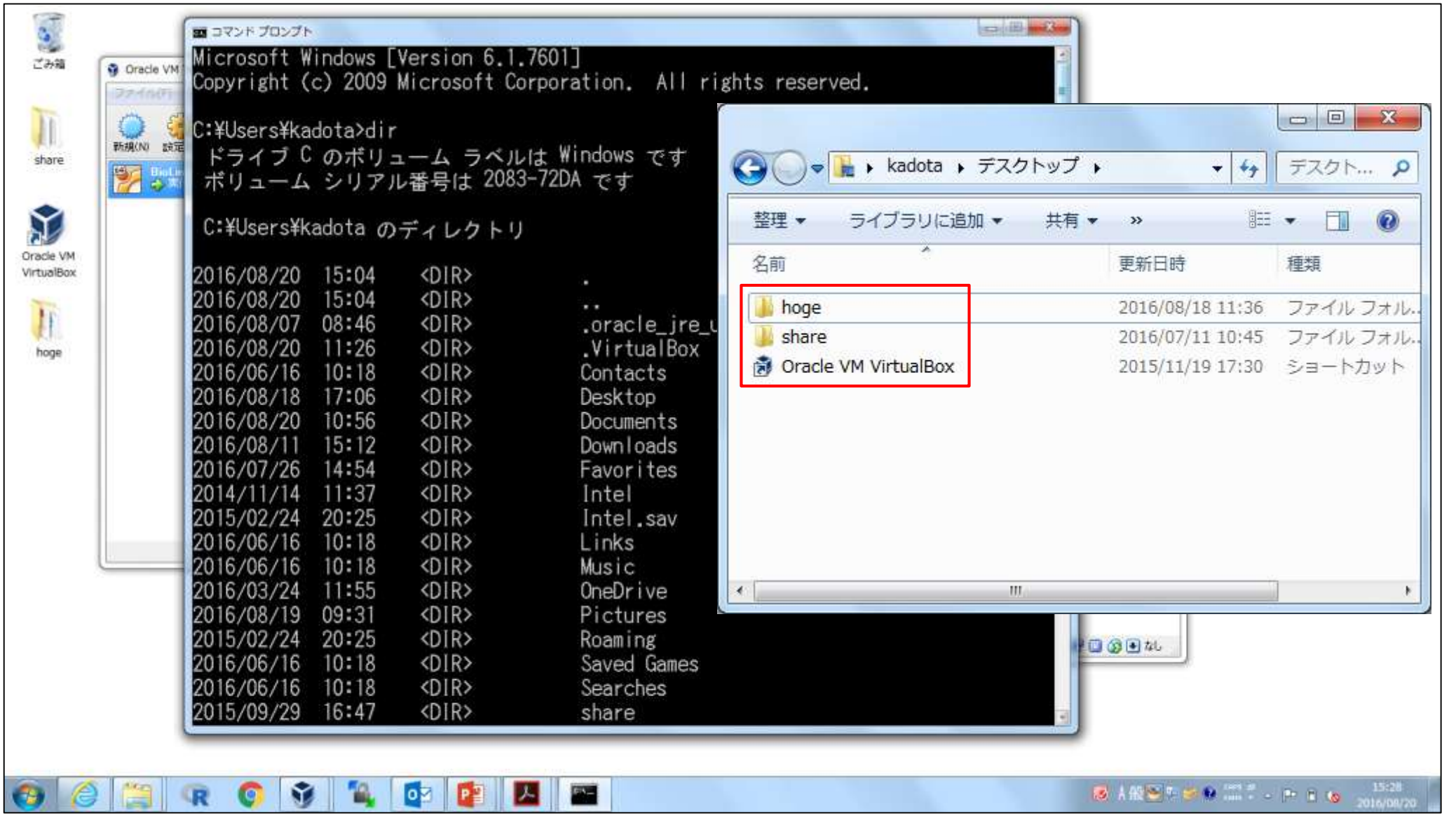
GUIとCUI

「kadotaさんのデスクトップ」であることが①で明示されているので、現在どこで作業をしているかがよくわかります。①の場所を「作業ディレクトリ (working directory)」や「カレントディレクトリ (current directory)」などと呼びます。フォルダとディレクトリは、同じようなものという理解でよい



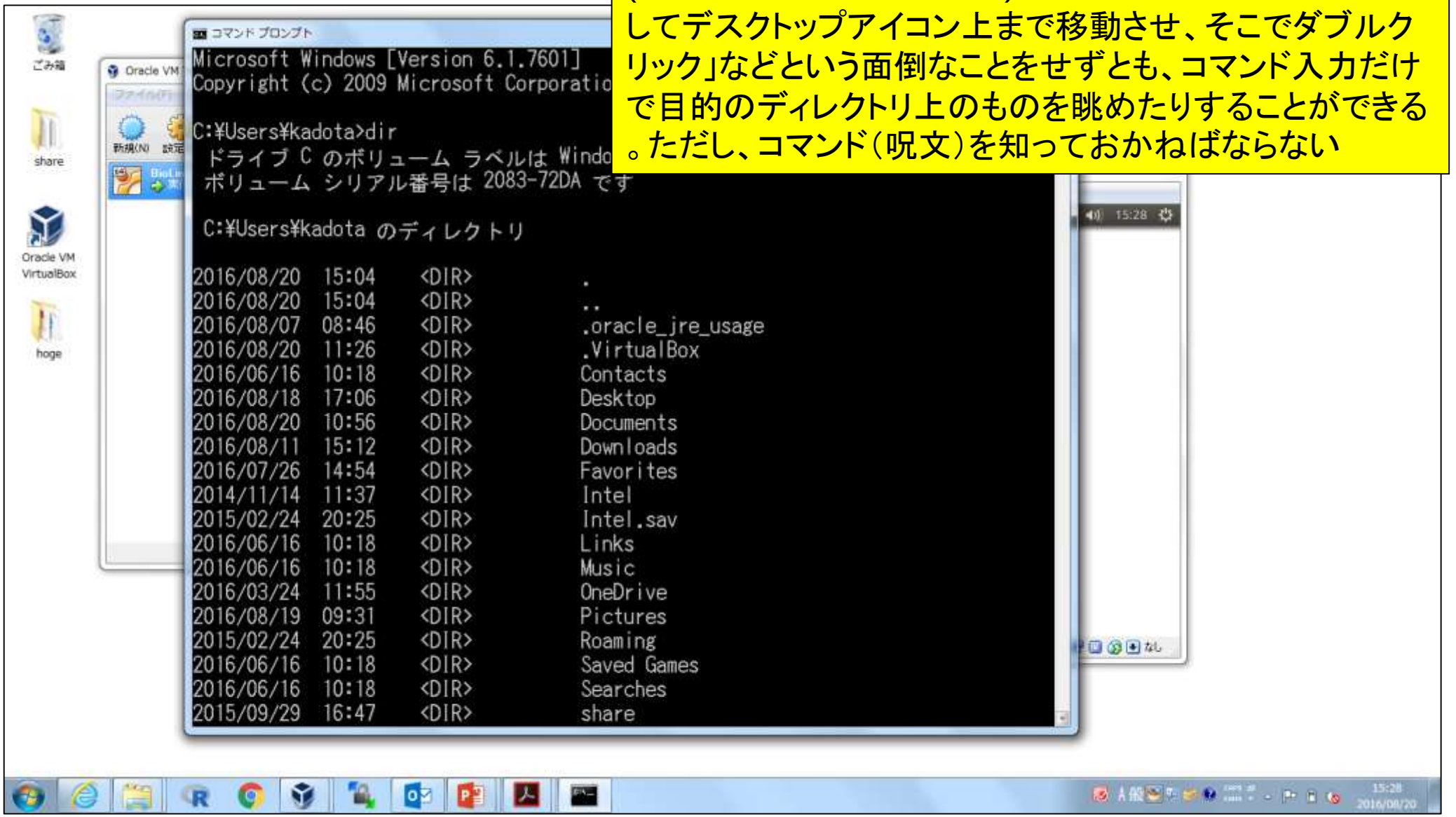
GUIとCUI

GUI (Graphical User Interface)での作業は、デスクトップというフォルダをダブルクリックして、そこを眺めるという流れ



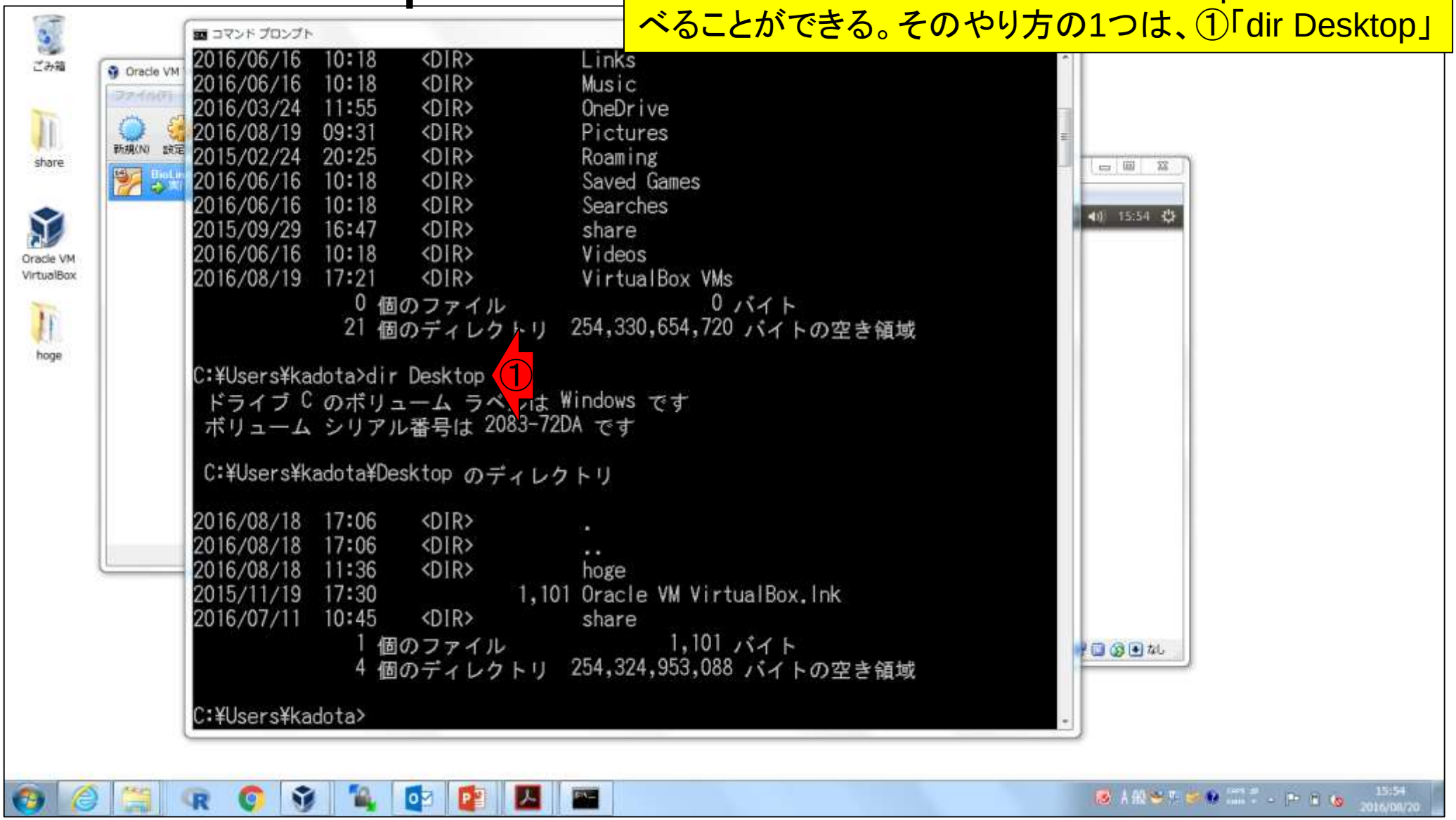
GUIとCUI

コマンドプロンプト上での作業は、CUI (Character User Interface; Console User Interface)での作業に相当。CLI (Command Line Interface)などともいう。「マウスを操作してデスクトップアイコン上まで移動させ、そこでダブルクリック」などという面倒なことをせずとも、コマンド入力だけで目的のディレクトリ上のものを眺めたりすることができる。ただし、コマンド(呪文)を知っておかねばならない



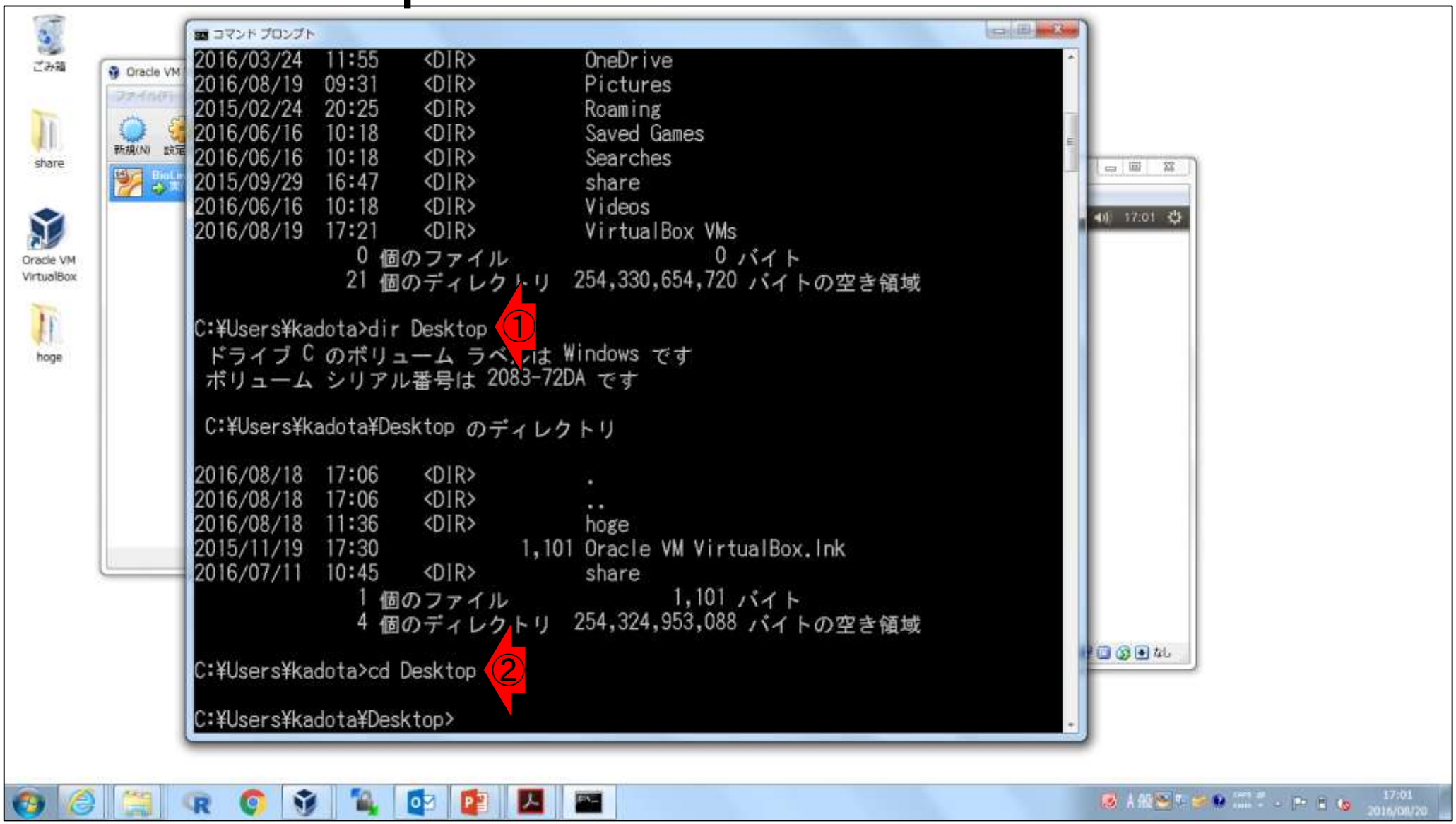
dir Desktop

例えば、コマンドプロンプト起動直後の場所（ホームディレクトリという）は、この場合「C:¥Users¥kadota」に相当する。この場所にいながらにして、Desktop上のものを調べることができる。そのやり方の1つは、①「dir Desktop」



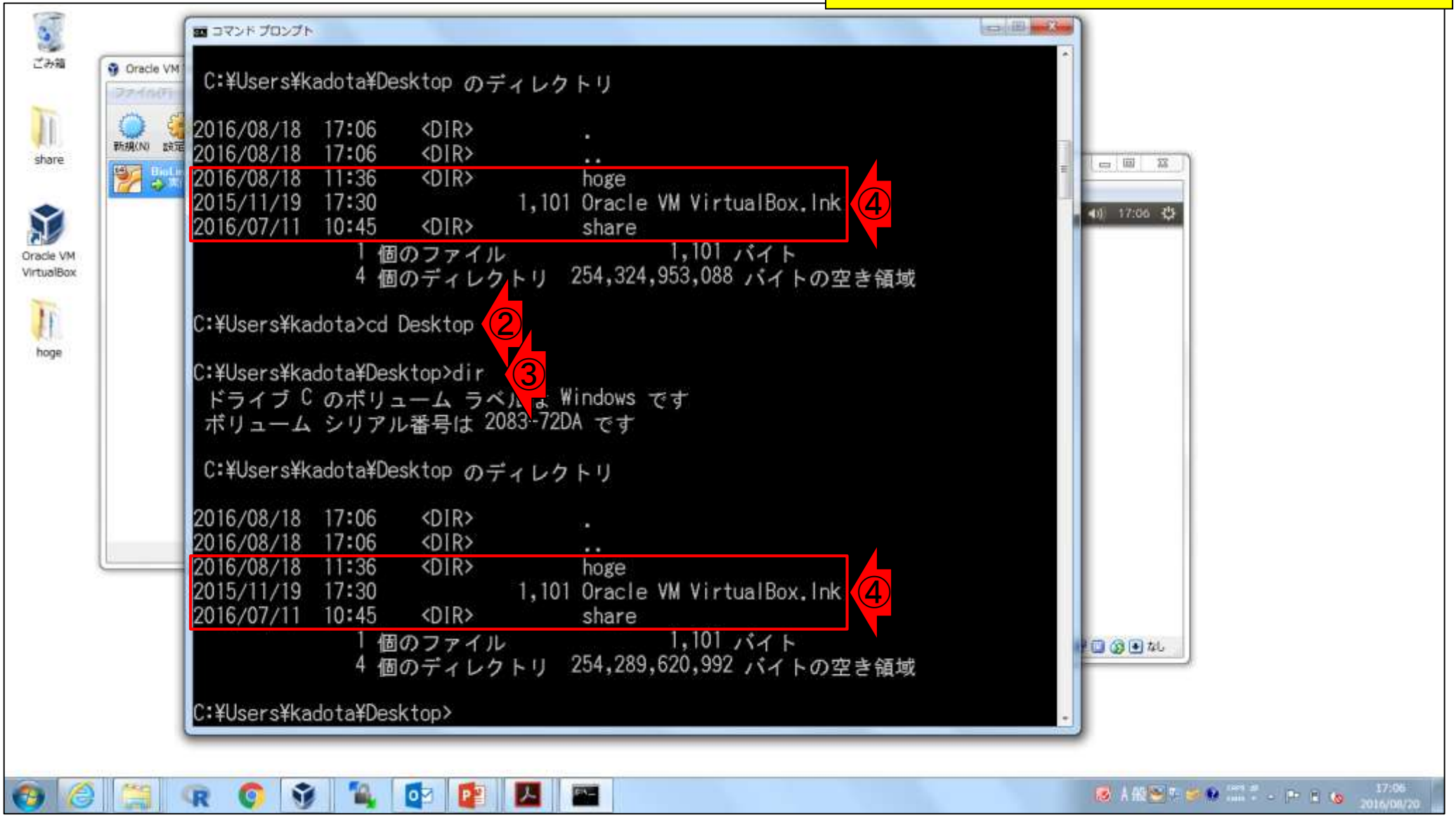
cd Desktop

もう1つのやり方は、②「cd Desktop」で作業ディレクトリをDesktopに移動してから…



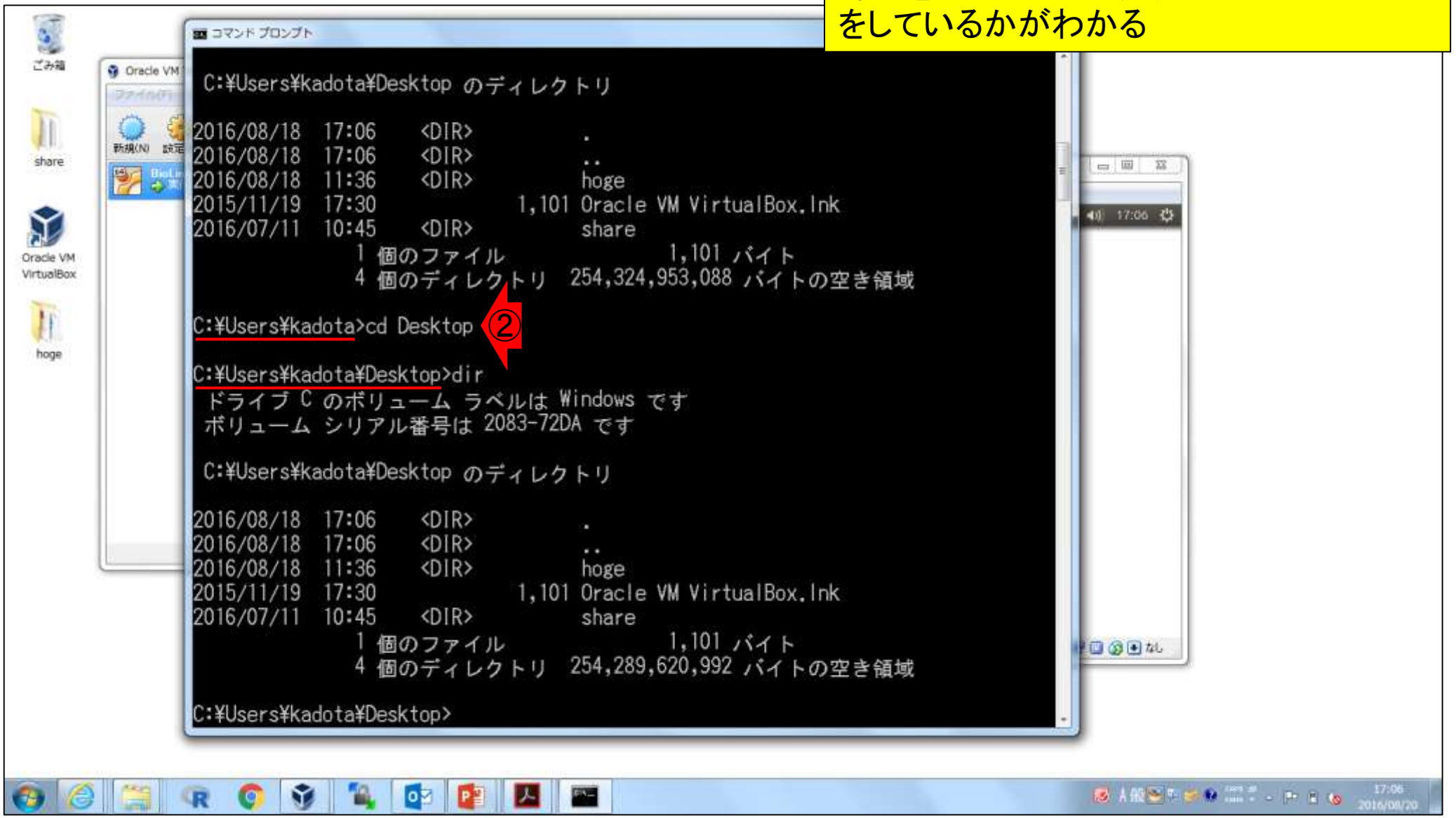
dir

もう1つのやり方は、②「cd Desktop」で作業ディレクトリをDesktopに移動してから…
③「dir」。確かに④同じ結果になっている



作業ディレクトリの把握

②「cd Desktop」実行前後で、赤下線部分が変わっていることがわかる。つまり、この部分を眺めることで、今自分がどこで作業をしているかがわかる

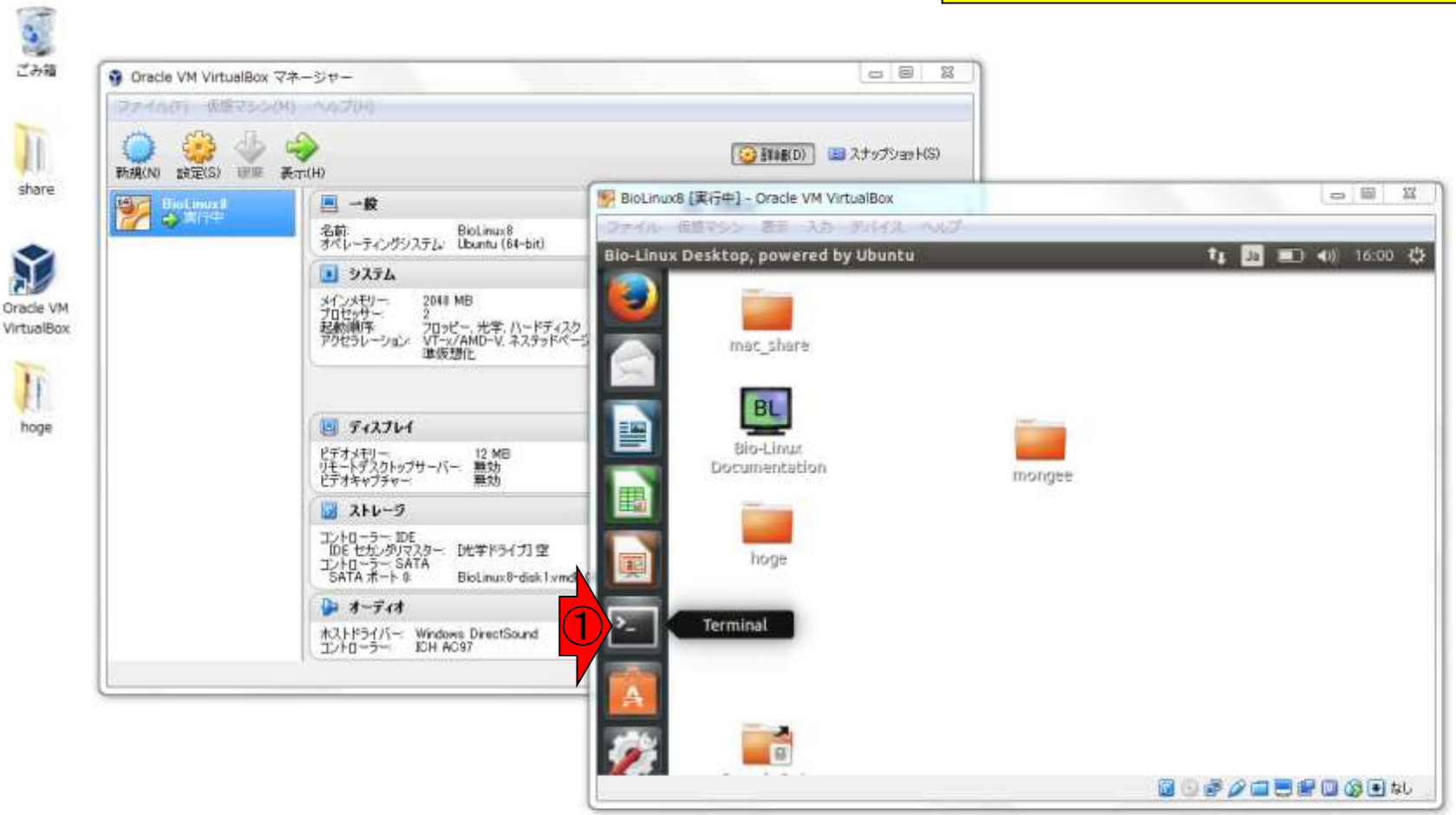


Contents

- イントロダクション
  - 概要、背景（NGS用カリキュラム、講習会）、Linuxスキル習得の意義
  - ウェブ情報（日本乳酸菌学会誌のNGS連載やNGS講習会資料）
- 実習環境に慣れる
  - 仮想環境での作業に慣れる
  - GUIとCUI（マウス操作かコマンド入力操作か）
  - ターミナルでの作業
  - 共有フォルダの概念を理解
- 練習
  - 作業ディレクトリの変更、練習用NGSデータファイルのダウンロード
  - ファイルの確認、de novoゲノムアセンブリ
  - BLAST検索
- 課題
  - グループごとに異なる課題ファイルを入力として、「ダウンロード、de novoアセンブリ、BLAST検索」を実行し、得られた結果をレポートにまとめて発表せよ
  - グループ1はkadai1.fasta、グループ2はkadai2.fasta、etc.

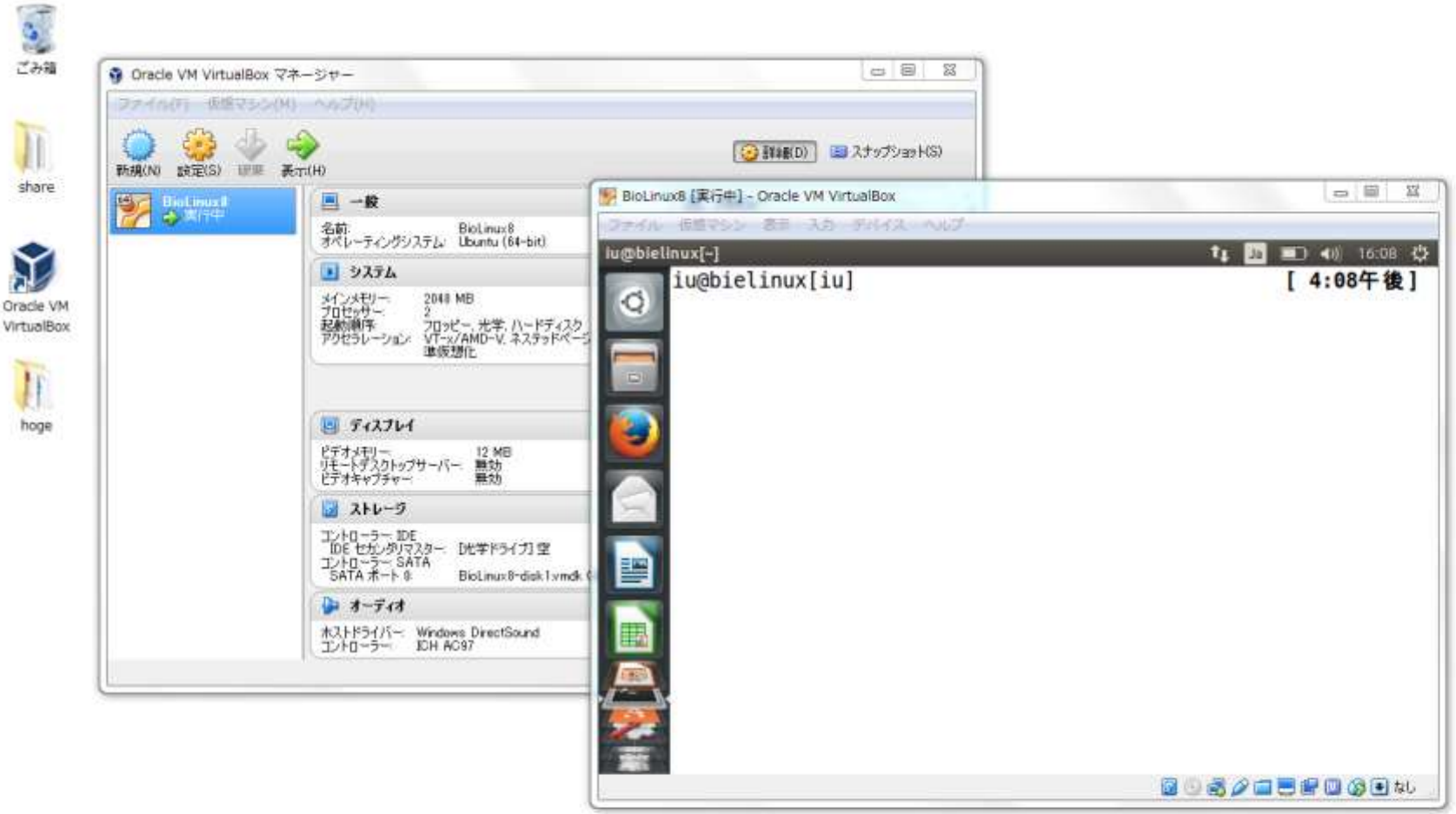
ターミナル

Windowsのコマンドプロンプトに対応するものは、Linuxでは（Macintosh同様）①ターミナル。第3回ウェブ資料（W8-3；スライド50）あたり



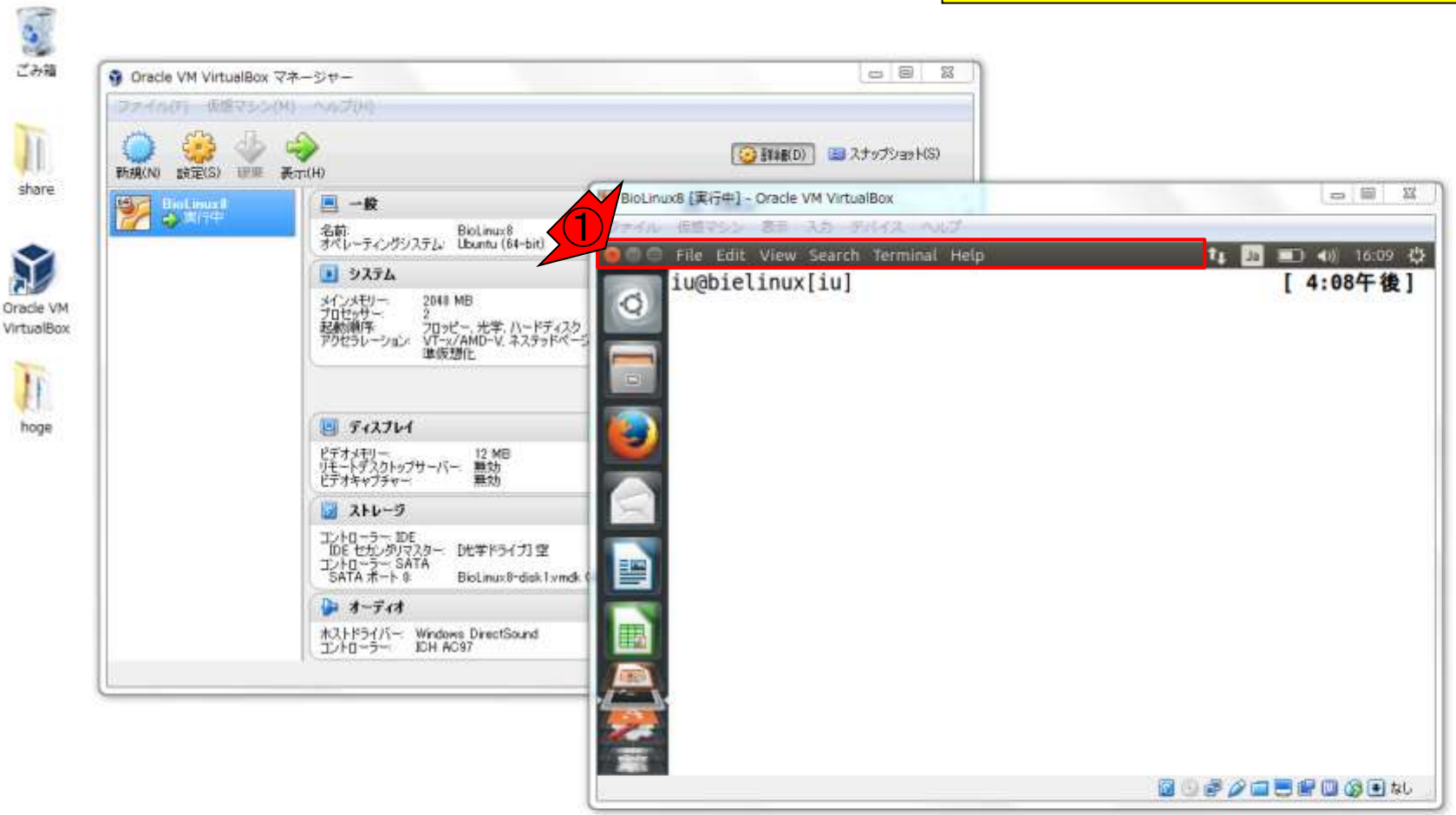
ターミナル

こんな感じになります。これはターミナルがLinux画面いっぱいに開いている状態です



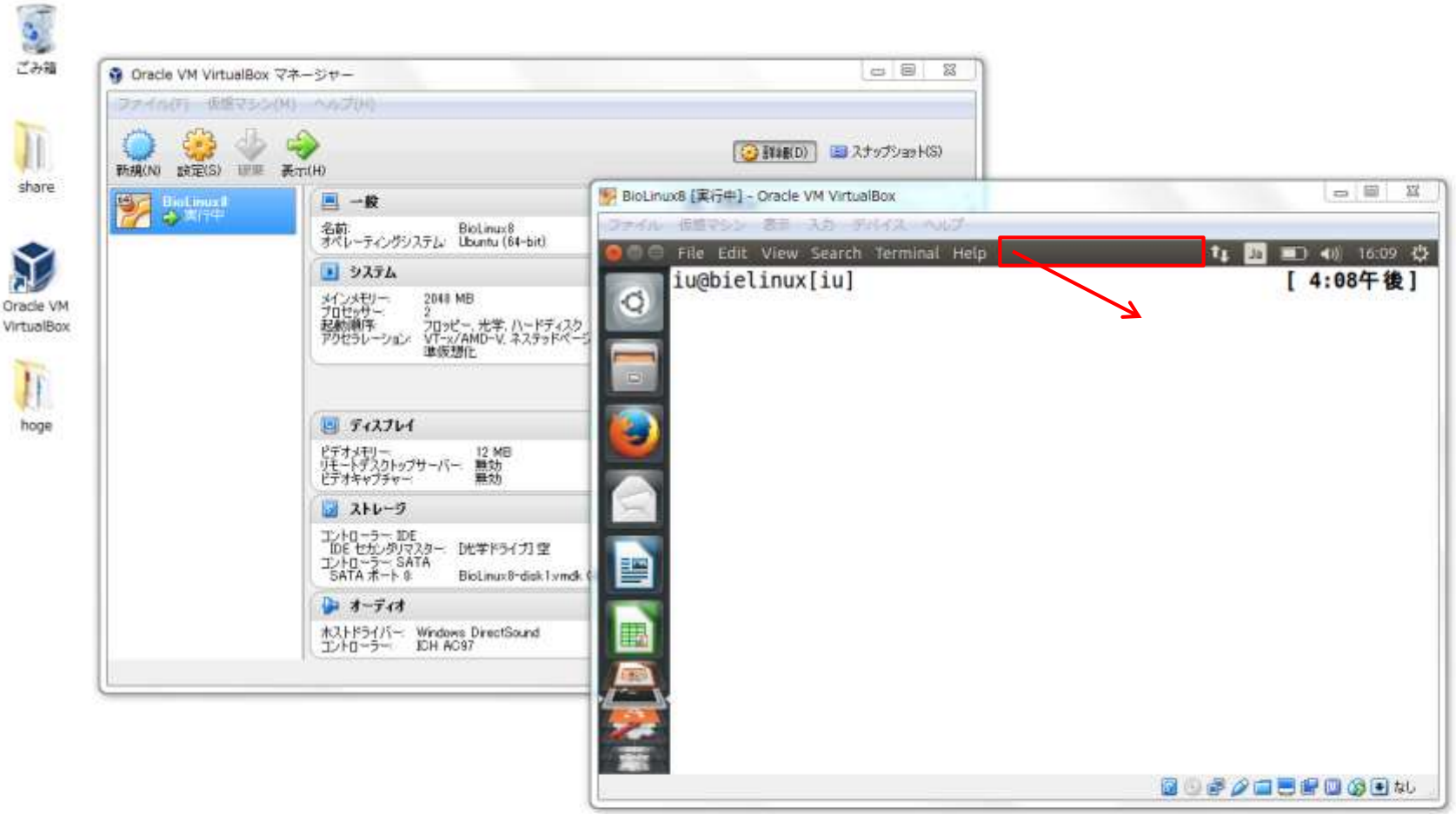
ターミナル

赤枠あたりにカーソルをもっていくと、メニューバーが見られます。①一番左の×ボタンを押すと、ターミナルを終了できます（が押さない）



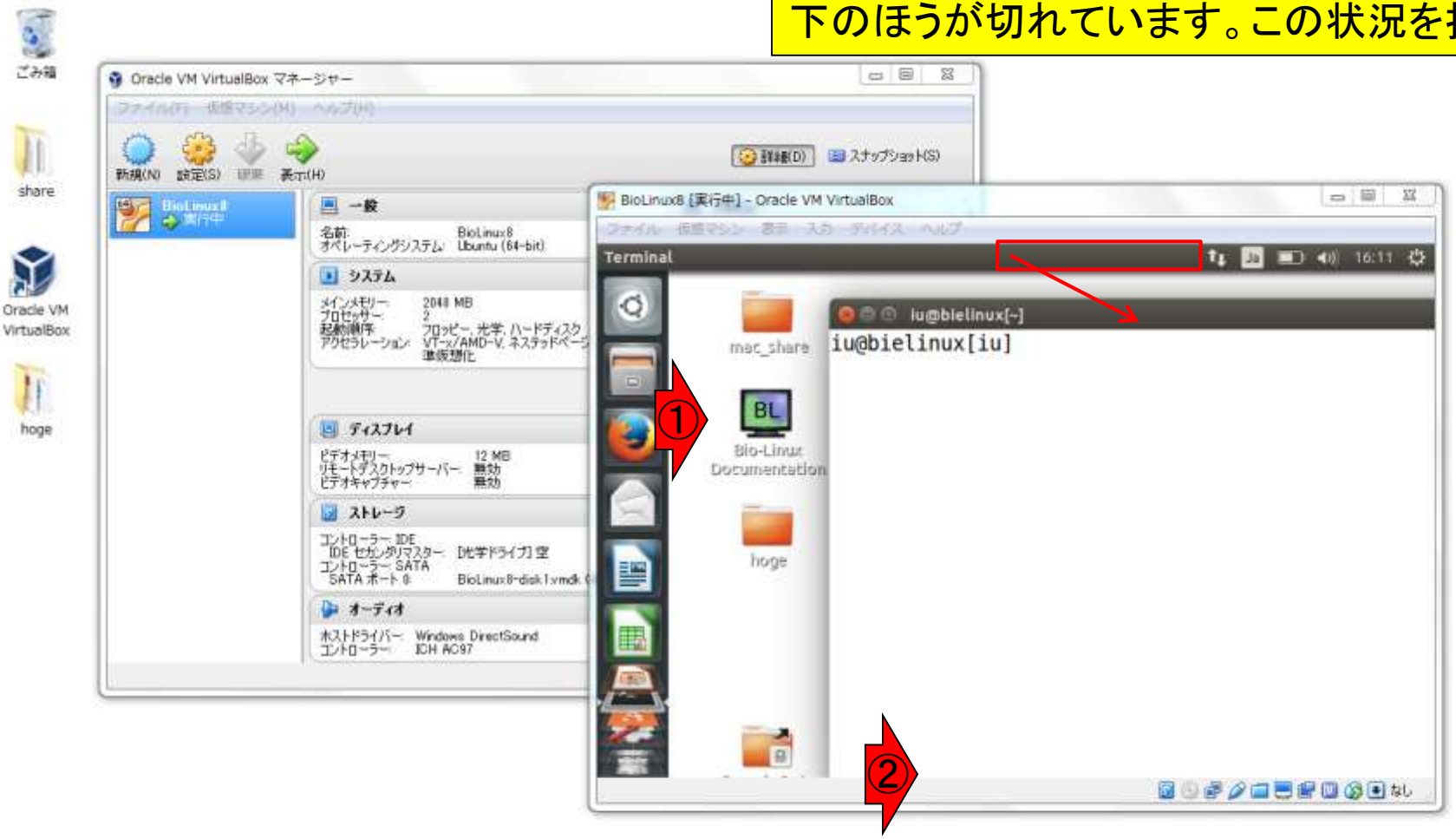
ターミナル

赤枠あたりで、矢印の始点から終点に向かってドラッグ &ドロップすると…



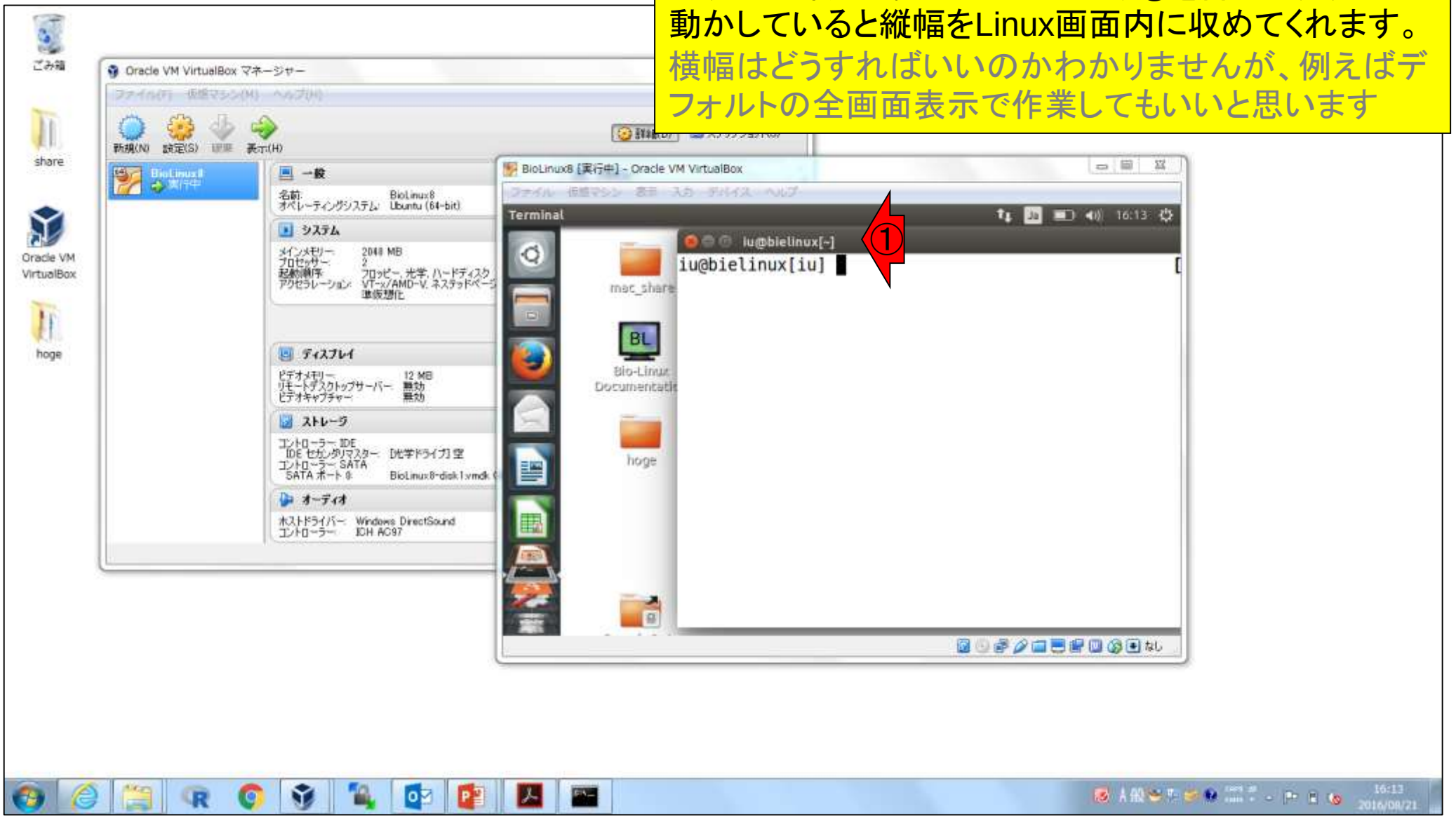
ターミナル

こんな感じになって、ターミナル画面を移動させることができます。そのおかげでデスクトップ画面上の①アイコンも見えるようになります。しかし、②ターミナル画面の下のほうが切れています。この状況を打破するには…



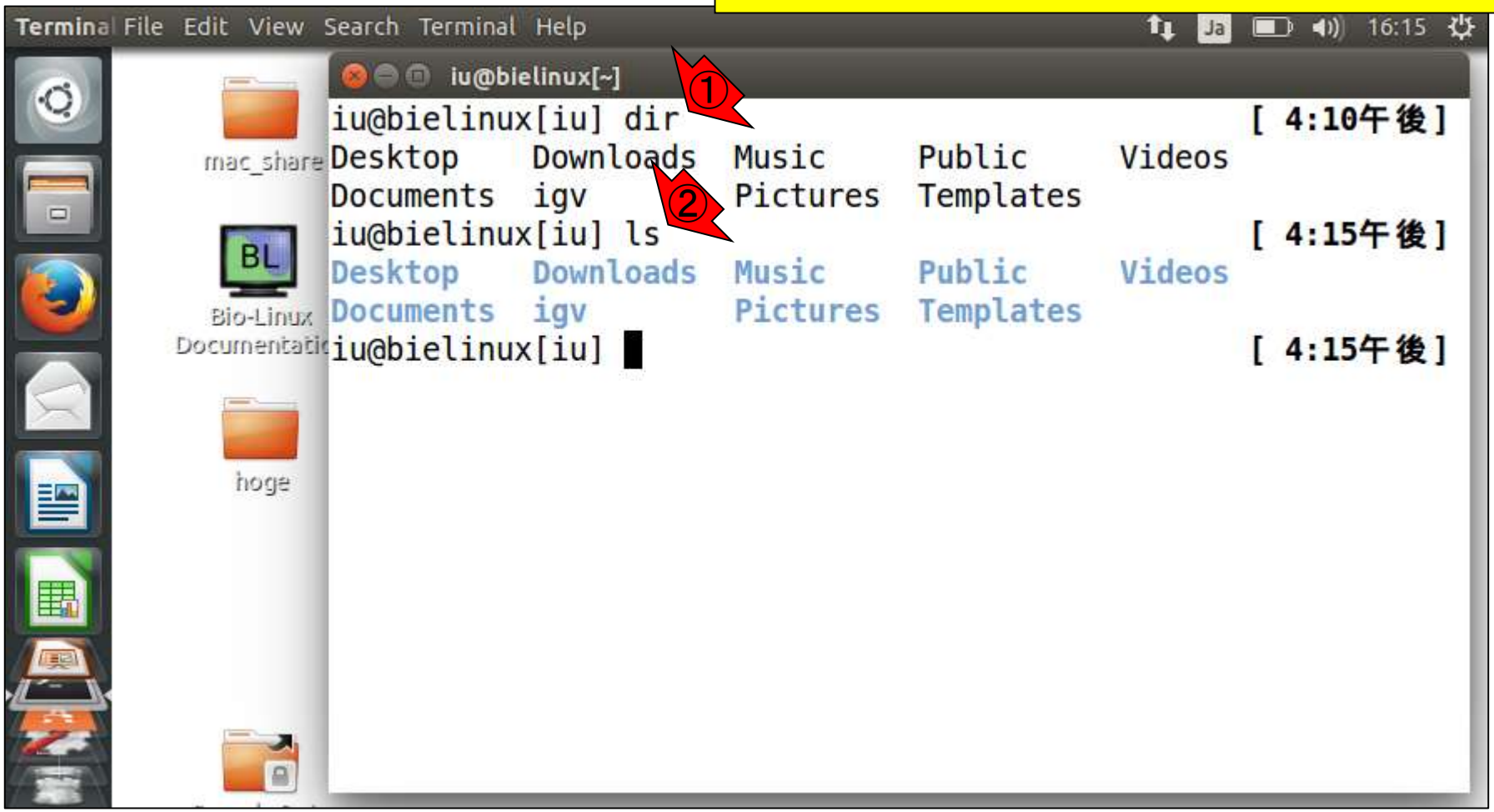
ターミナル

ターミナル画面の縦幅をLinux画面内に収めるためには、通常はターミナル画面の右下あたりで調整しますが、右下部分が見えていません。①を持って、あちこち動かしていると縦幅をLinux画面内に収めてくれます。
横幅はどうすればいいのかわかりませんが、例えばデフォルトの全画面表示で作業してもいいと思います



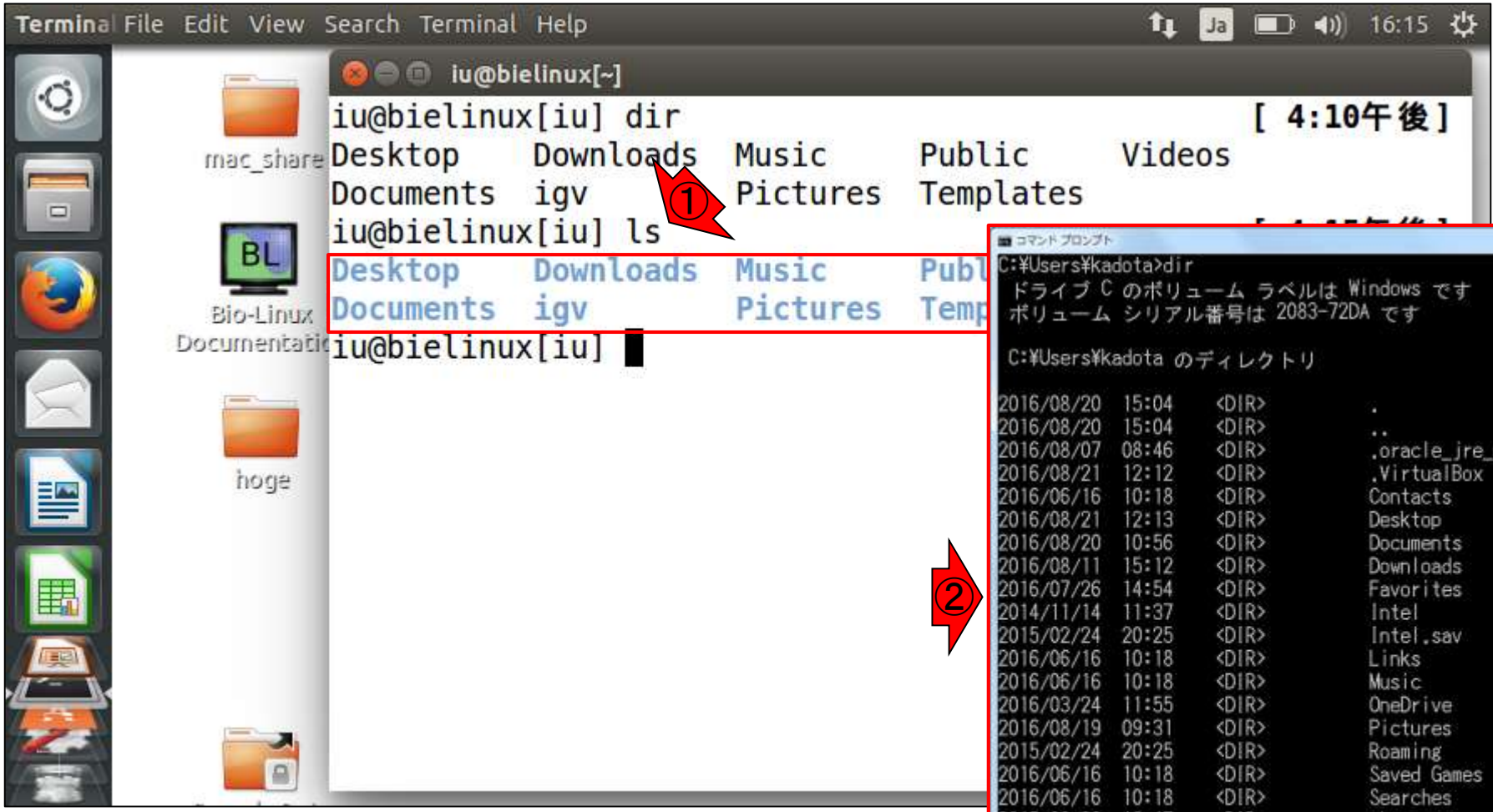
dirではなくls

作業ディレクトリの中身を確認するのに、Windowsのコマンドプロンプト上では、dirと打ち込みました。Linux環境でも一応①dirで動作しますが、通常は②ls（えるえす）と打つ



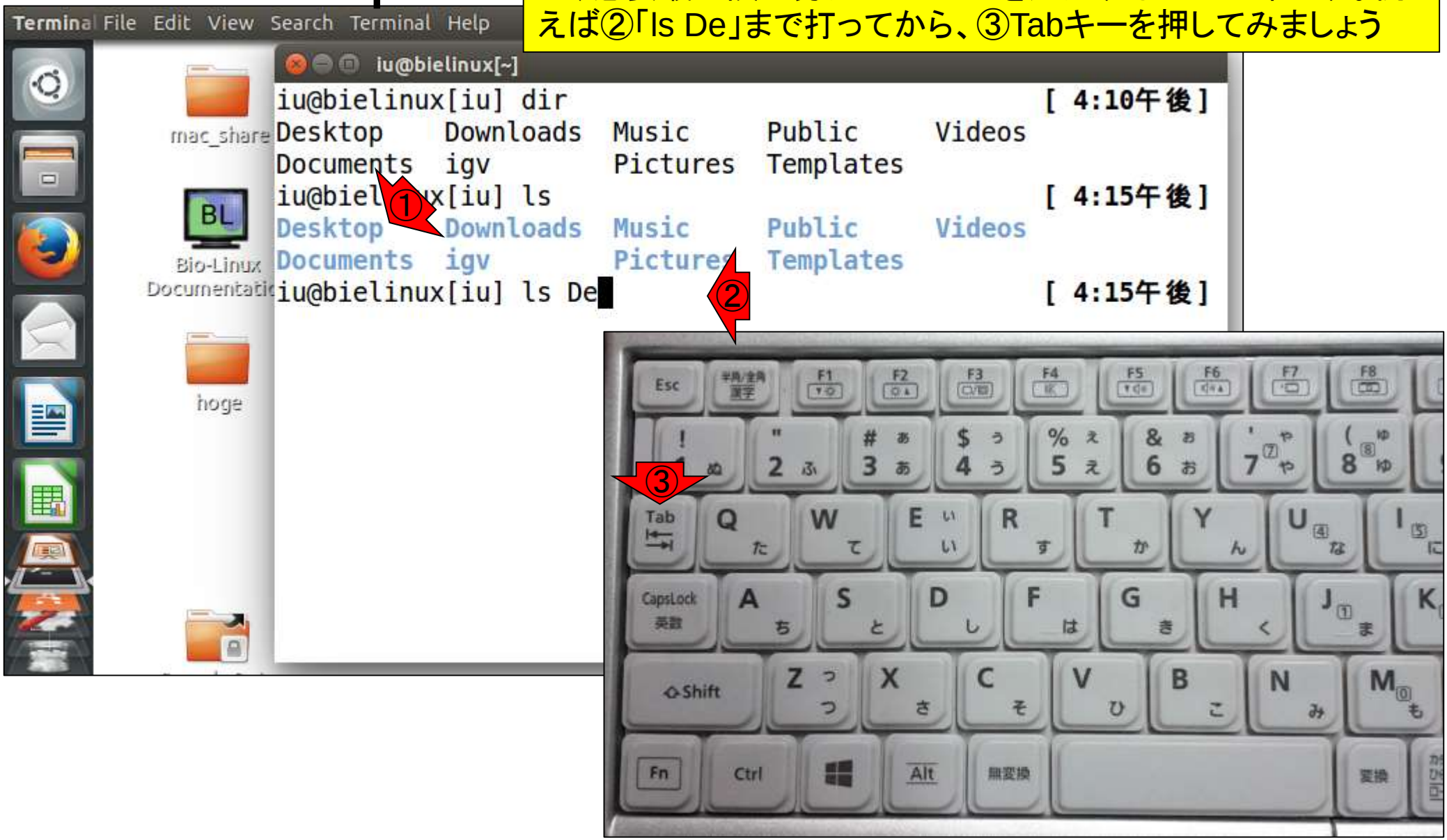
ls実行結果

Linux（ホストOS）上での①ls実行結果は、②Windows上でのdir実行結果と似たような感じであることがわかります

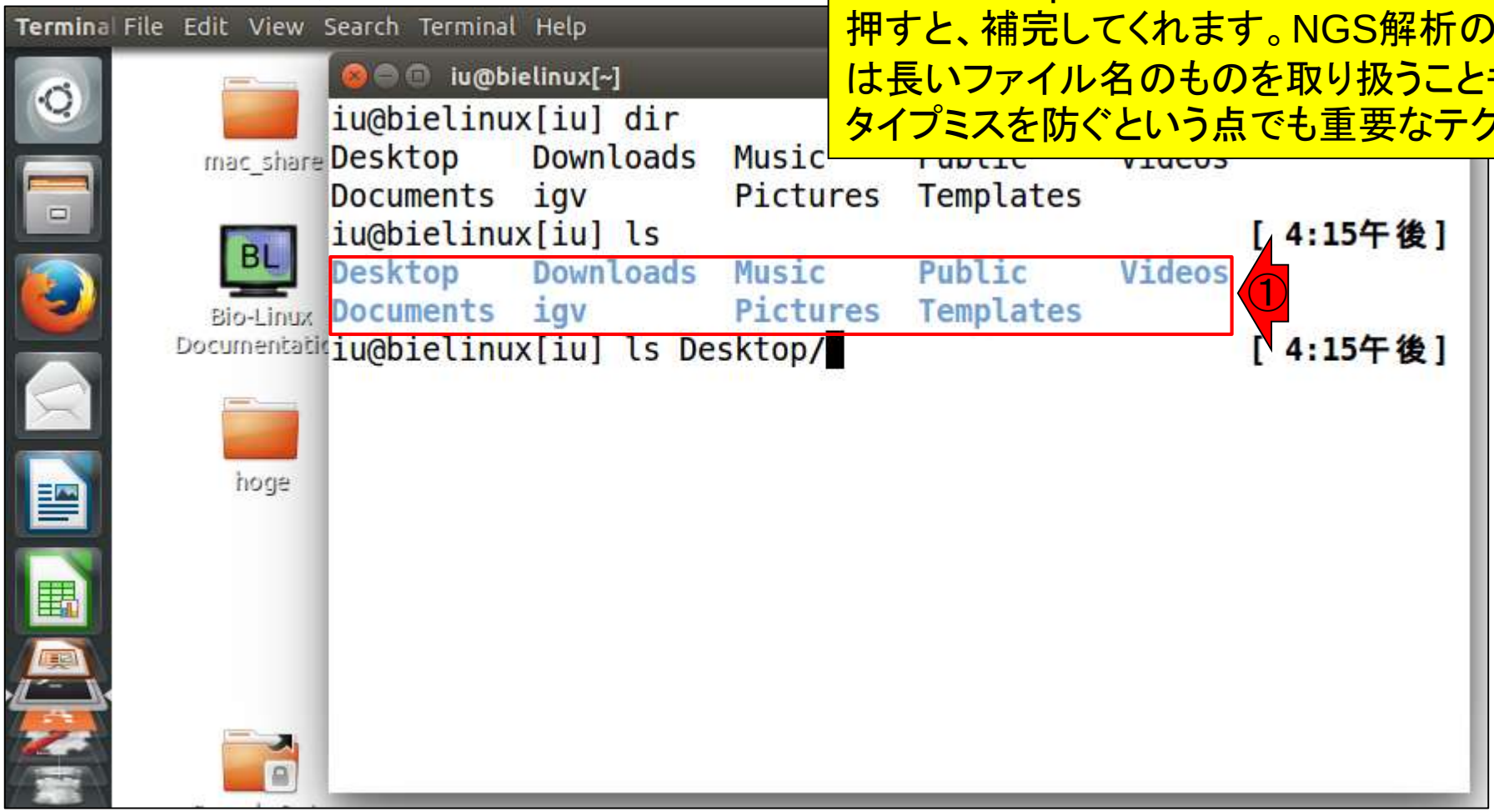


ls Desktop

①Desktopというディレクトリが見えているので、そこの中身を表示させます。「ls Desktop」と打てばいいですが、Linuxの世界では、必要最小限の労力でコマンドを入力するのが基本です。例えば②「ls De」まで打ってから、③Tabキーを押してみましょう



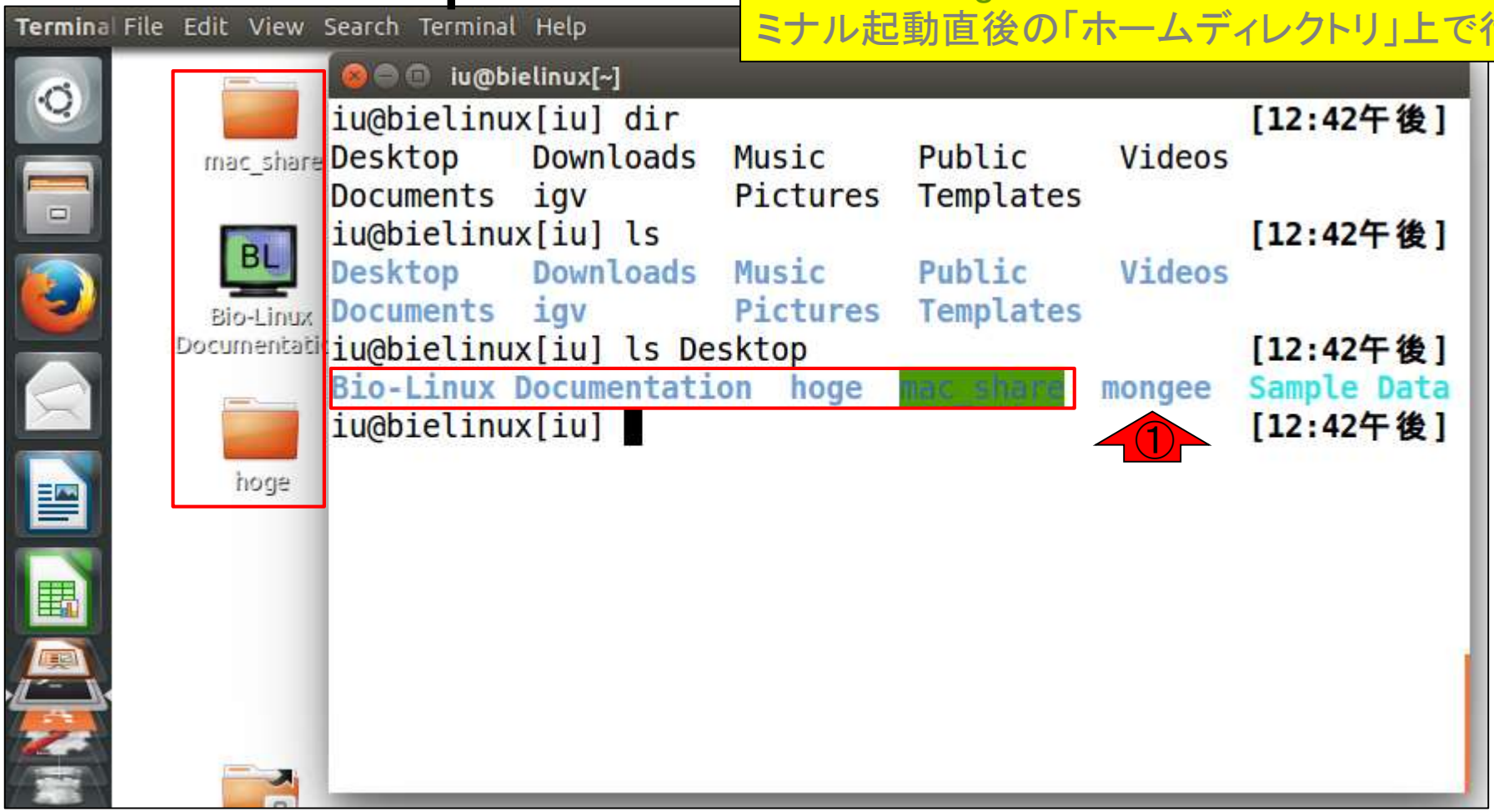
タブ補完



「ls Desktop/」となります。このテクニックを「タブ補完」などと呼ぶ。①赤枠を眺めると、Deから始まるものはDesktopしかない。このような状況でTabキーを押すと、補完してくれます。NGS解析の実務局面では長いファイル名のものを取り扱うこともあるので、タイプミスを防ぐという点でも重要なテクニックです

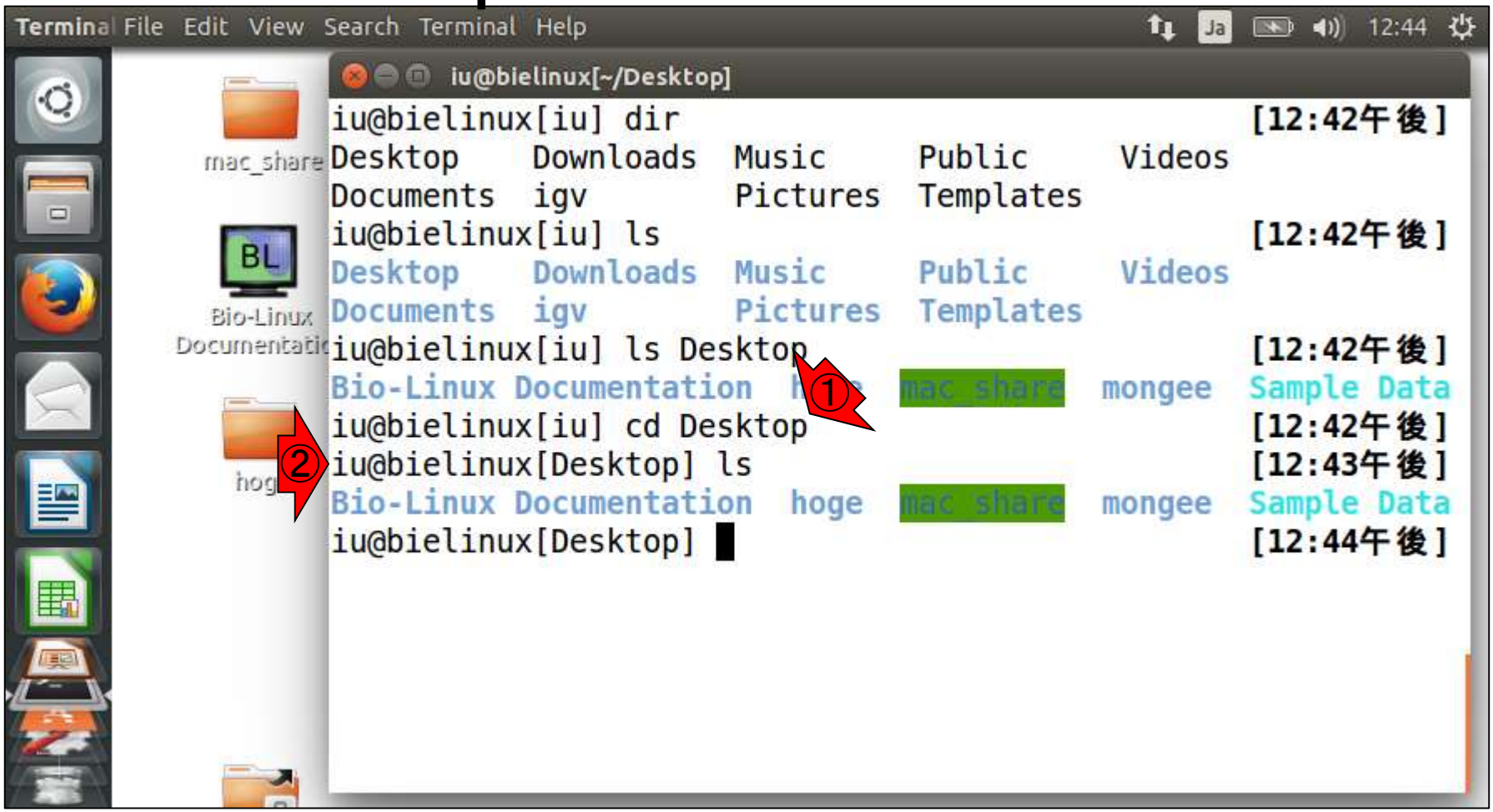
ls Desktop

「ls Desktop」実行結果。確かに赤枠で示すように、Linuxのデスクトップ画面に見えているものと同じものが見えている。①mongeeはヒトそれぞれ。ここまでの作業はターミナル起動直後の「ホームディレクトリ」上で行いました



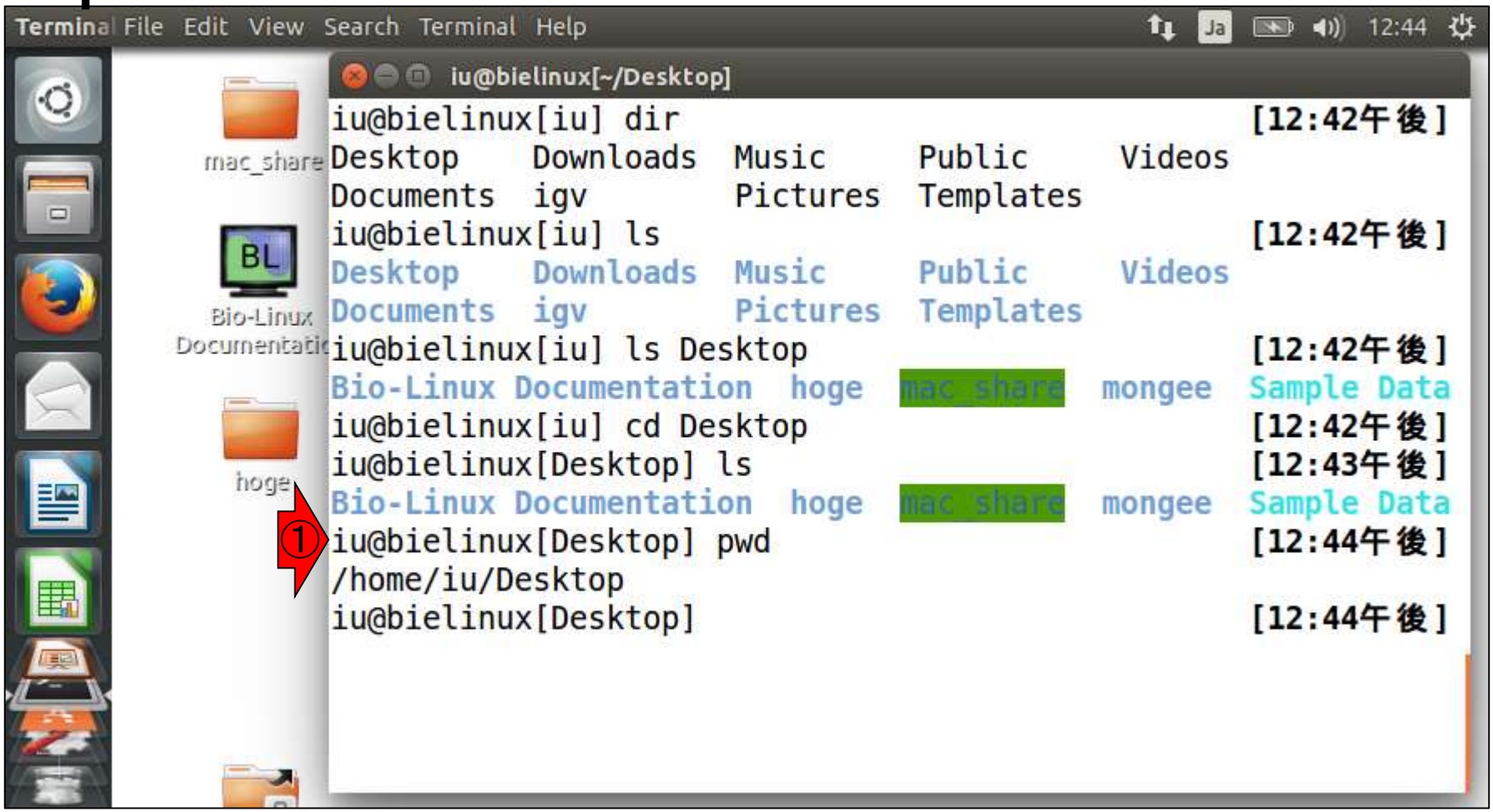
cd Desktop

もちろん①cd Desktopとして、Desktopディレクトリに移動してから、②lsするのでも構いません



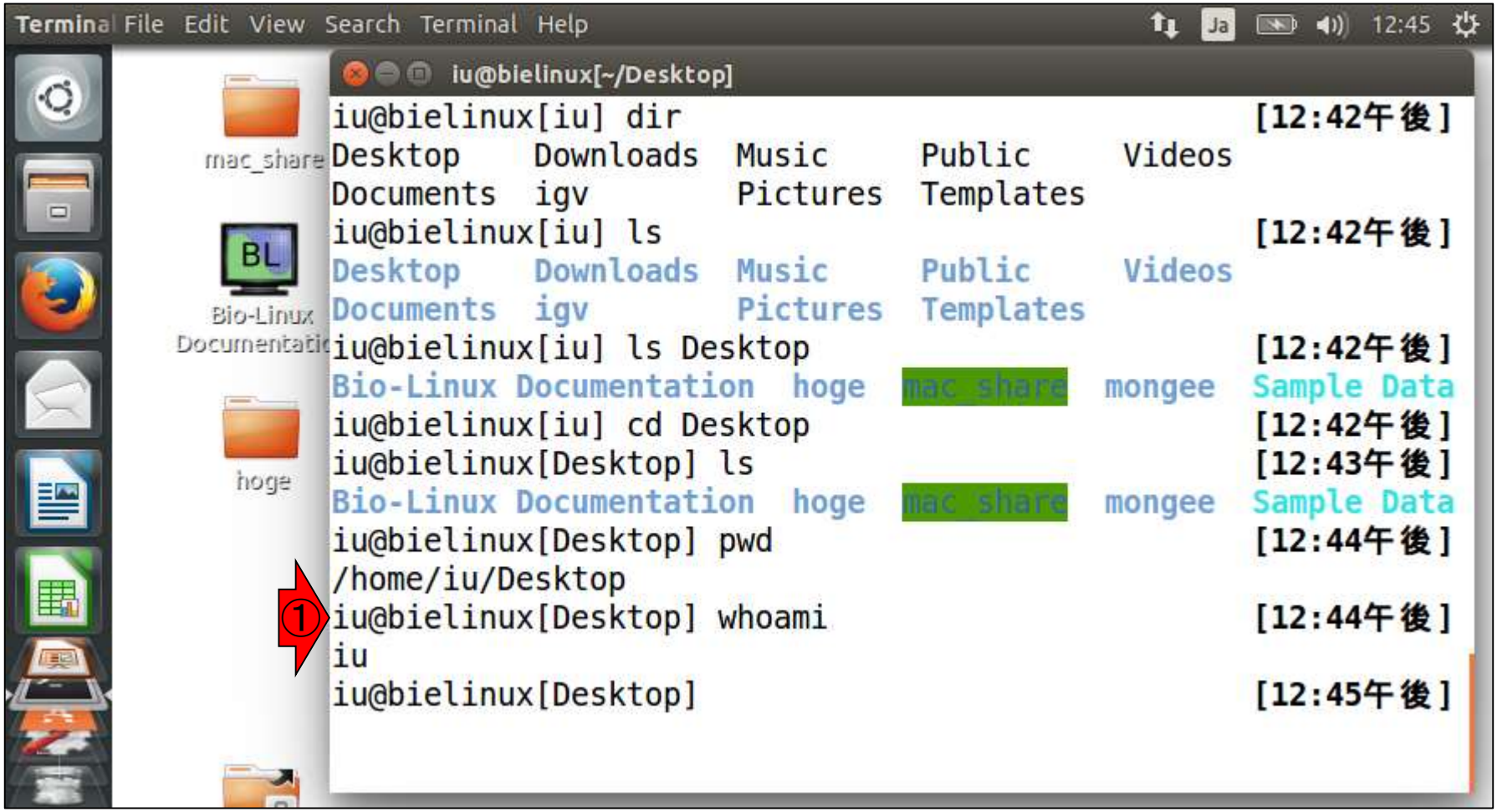
pwd

①pwdで現在の作業ディレクトリを表示させています(print working directory)



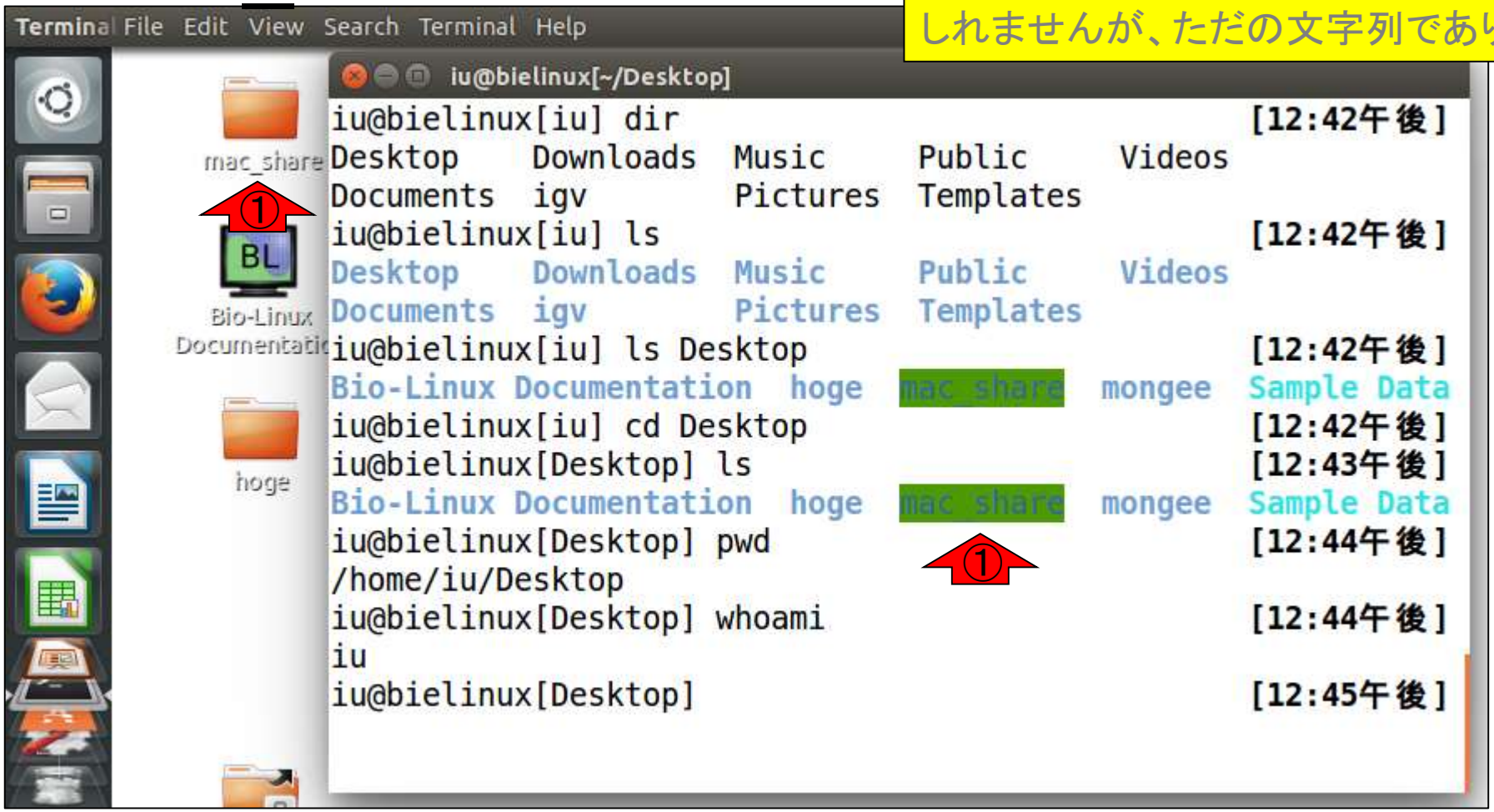
whoami

①whoamiでユーザ名(iu)を調べることができます



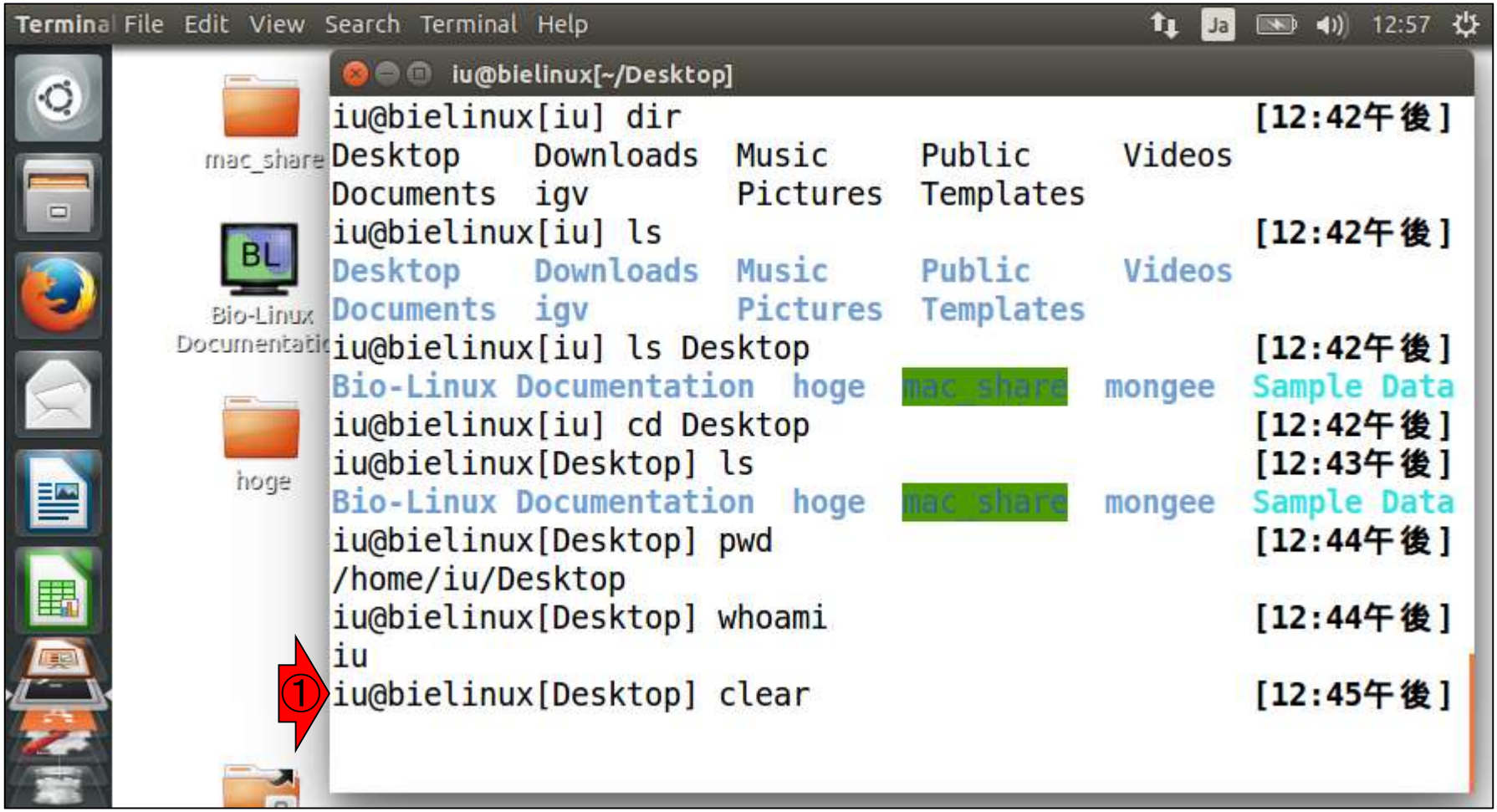
mac_share

①貸与PCは、mac_shareというディレクトリが反転されていると思います。macというキーワードから、Macintoshを連想するヒトがいるかもしれませんが、ただの文字列であり無関係です



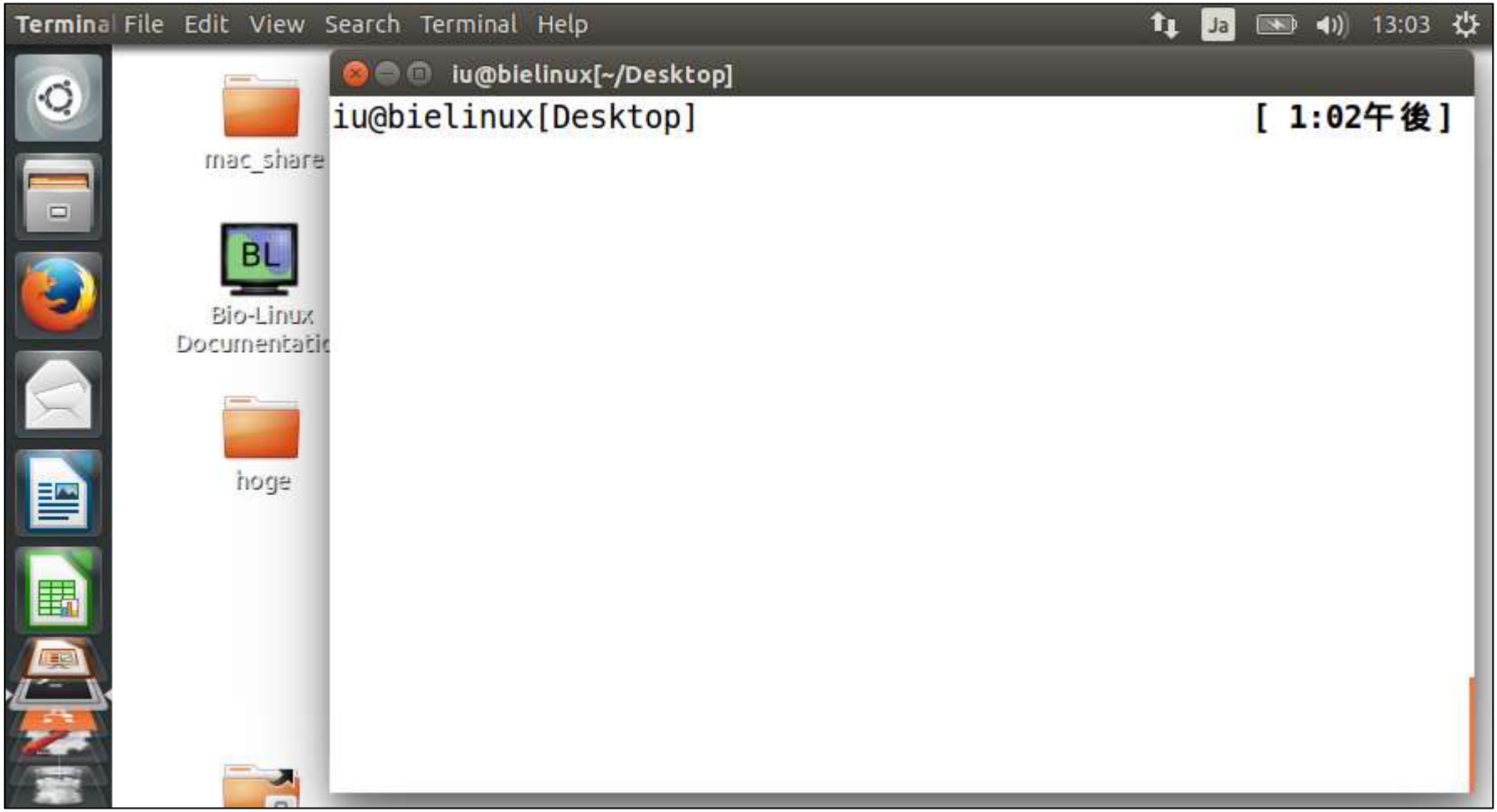
clear

①clearと打つことで、ターミナル画面をリフレッシュすることができます



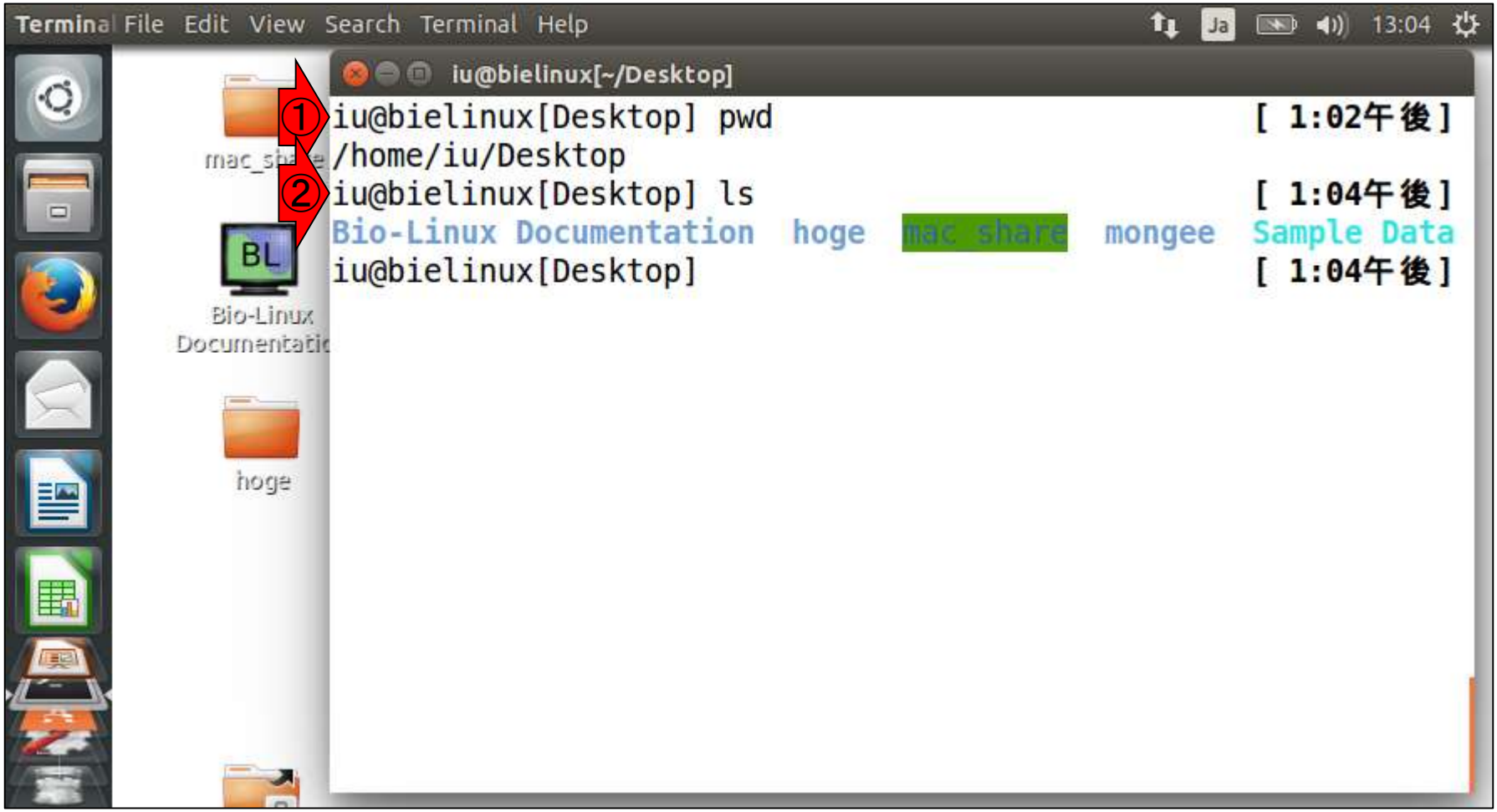
こんな感じになります

clear



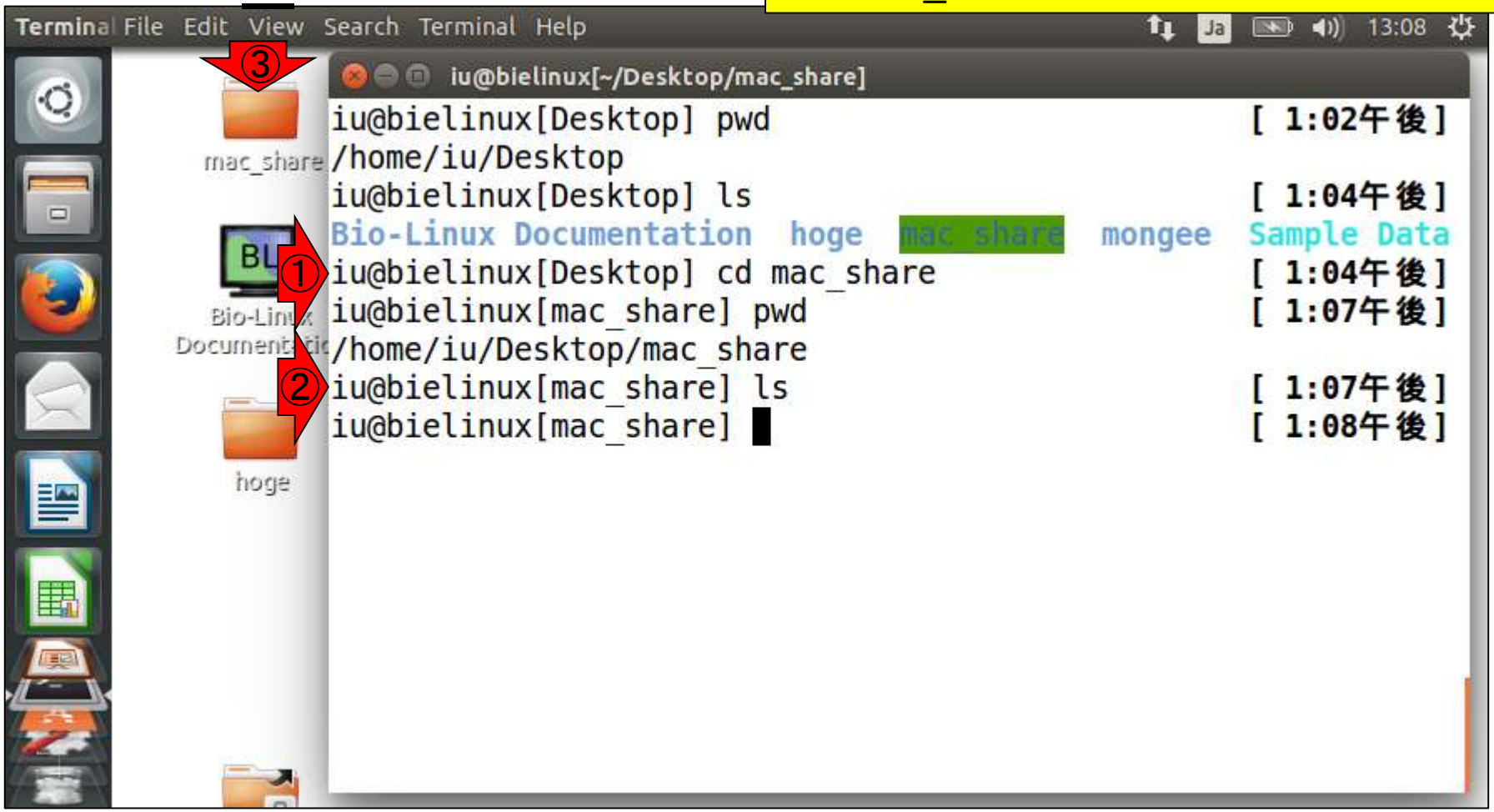
clear

①pwd、②ls。作業ディレクトリはclear実行前と同じです



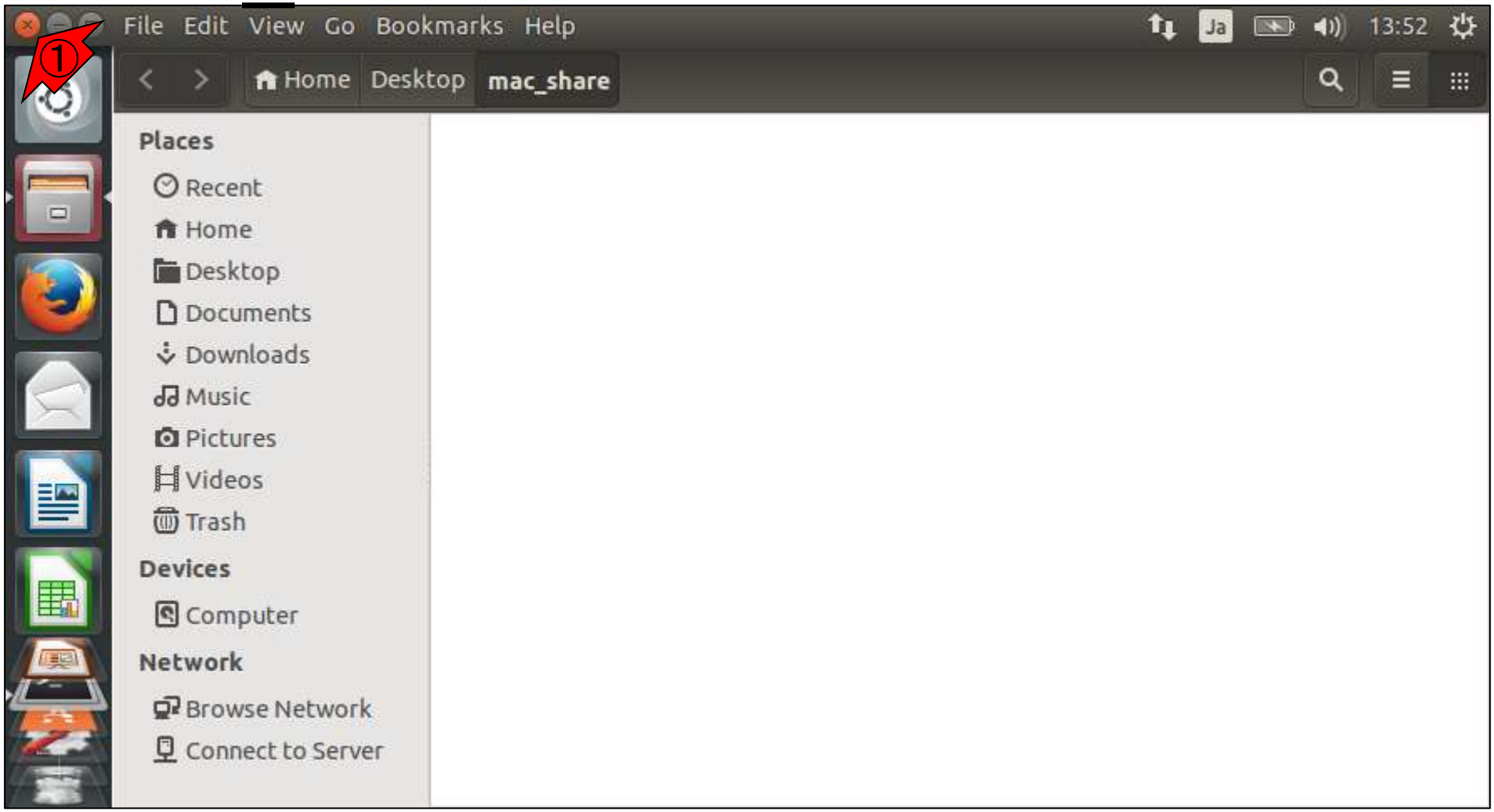
mac_share

①mac_shareディレクトリに移動して、(pwdで確認し)
②ls。このディレクトリ内には何もないことがわかります
。③mac_shareフォルダをダブルクリックして開くと…



mac_share

こんな感じになって、このフォルダ中には何もないことがわかります。①×で終了しておく

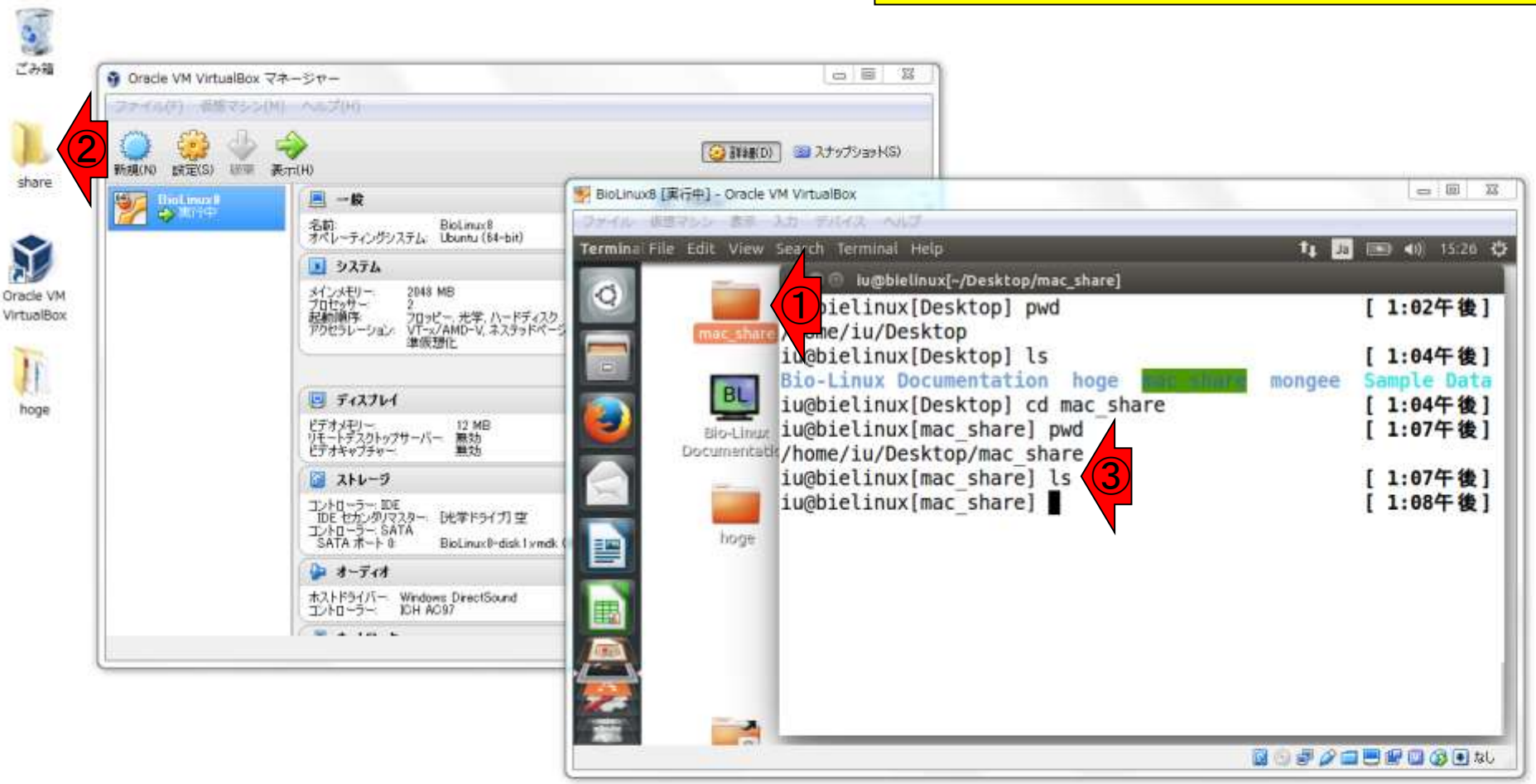


Contents

- イントロダクション
  - 概要、背景（NGS用カリキュラム、講習会）、Linuxスキル習得の意義
  - ウェブ情報（日本乳酸菌学会誌のNGS連載やNGS講習会資料）
- 実習環境に慣れる
  - 仮想環境での作業に慣れる
  - GUIとCUI（マウス操作かコマンド入力操作か）
  - ターミナルでの作業
  - 共有フォルダの概念を理解
- 練習
  - 作業ディレクトリの変更、練習用NGSデータファイルのダウンロード
  - ファイルの確認、de novoゲノムアセンブリ
  - BLAST検索
- 課題
  - グループごとに異なる課題ファイルを入力として、「ダウンロード、de novoアセンブリ、BLAST検索」を実行し、得られた結果をレポートにまとめて発表せよ
  - グループ1はkadai1.fasta、グループ2はkadai2.fasta、etc.

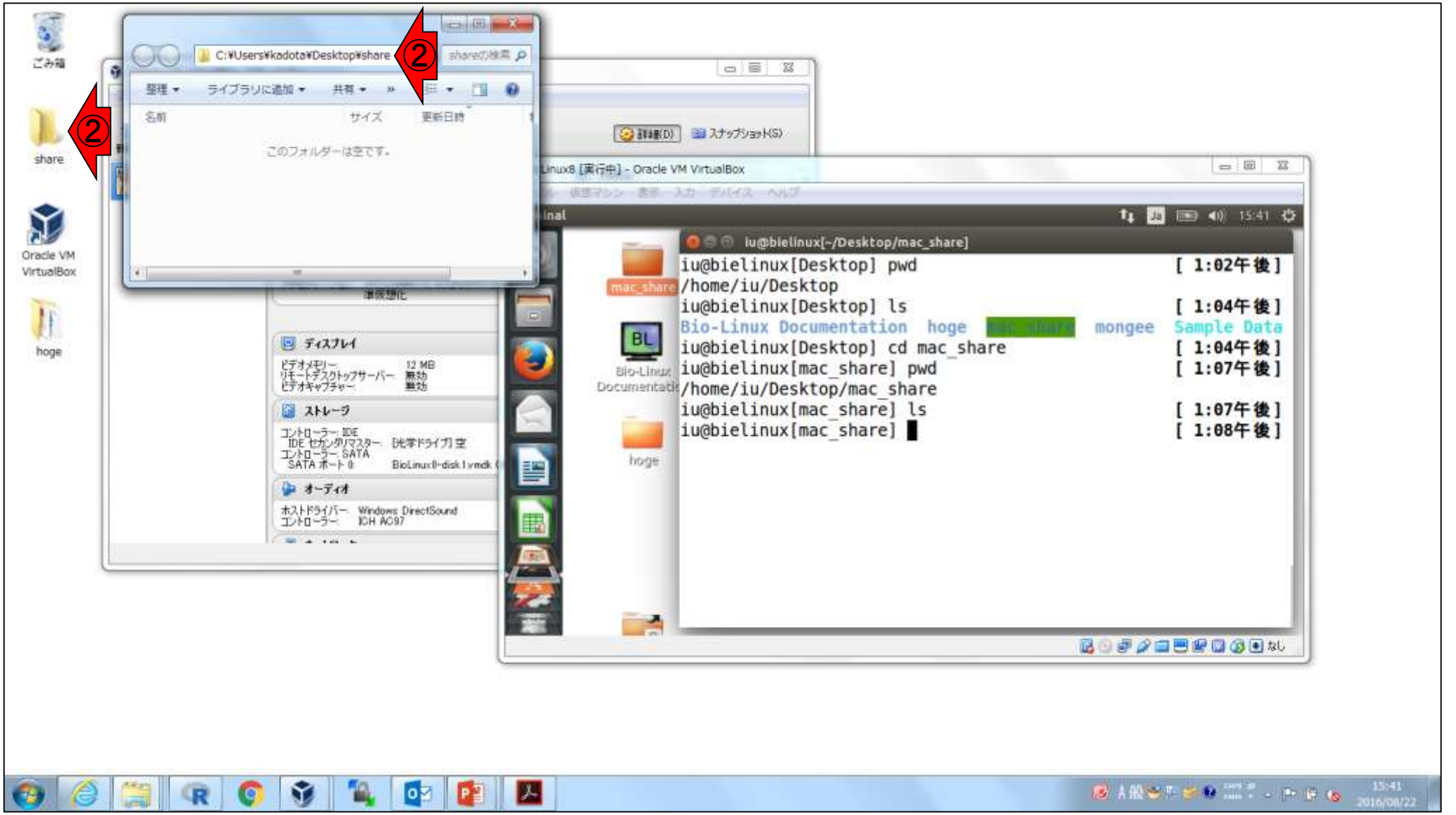
共有フォルダ

①Linuxのmac_shareと、②Windowsのshareは、共有フォルダです。③mac_share上でlsした結果何もなかったので、②には何もないはず



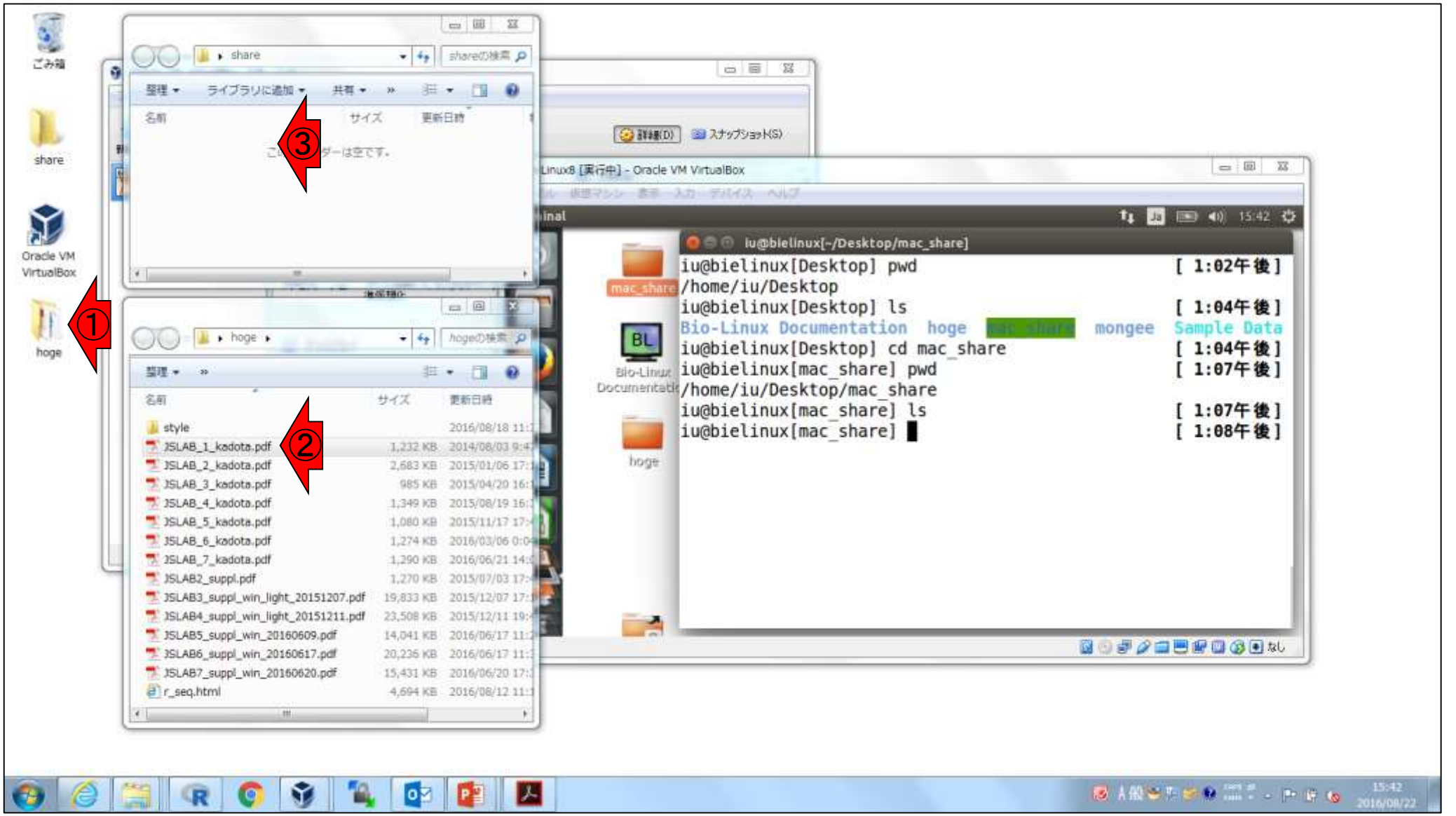
②確かに何もありません

共有フォルダ



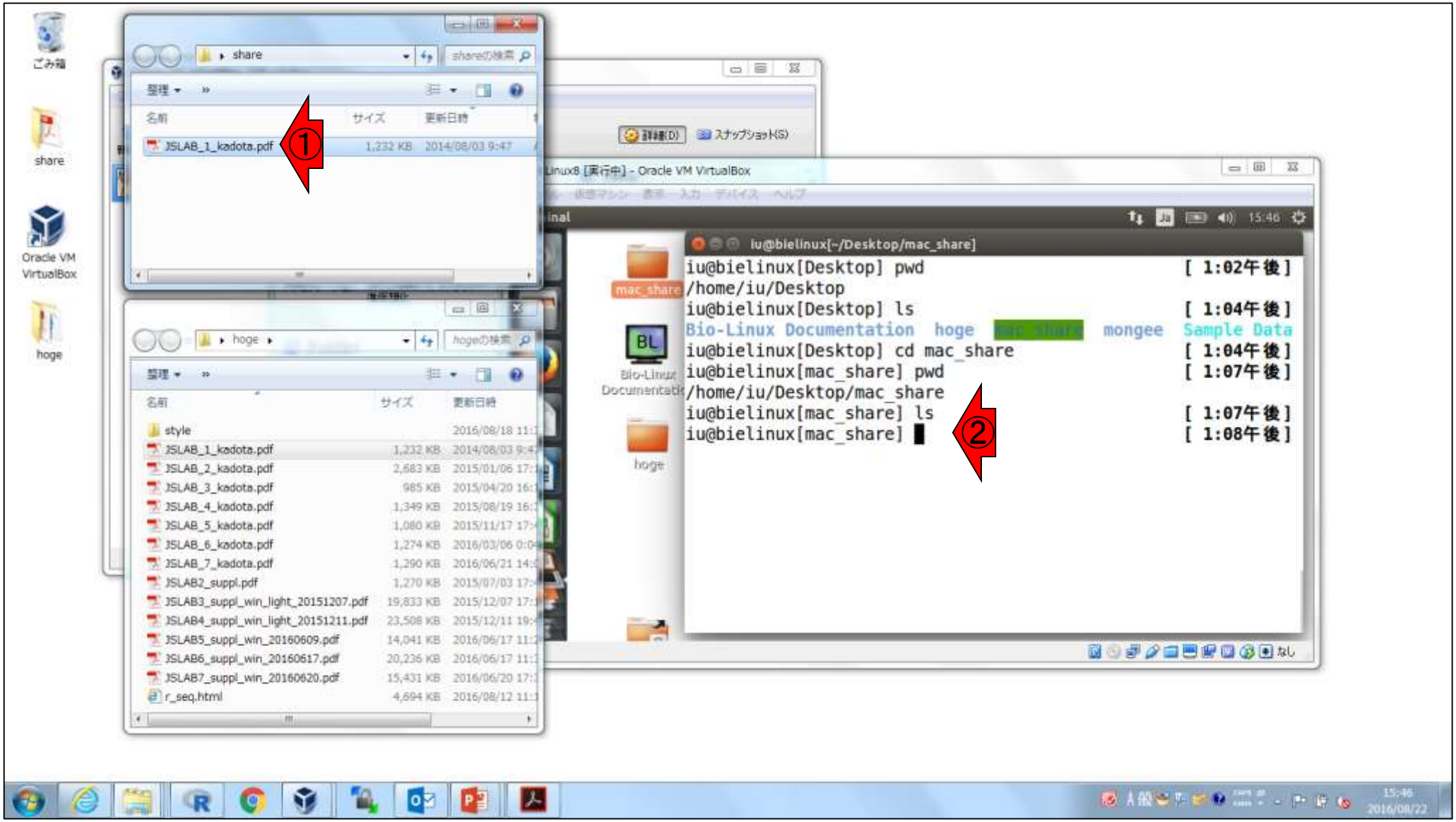
共有フォルダ

①hogeフォルダ中の②pdfファイルを、③shareフォルダ内にコピーしてみましょう



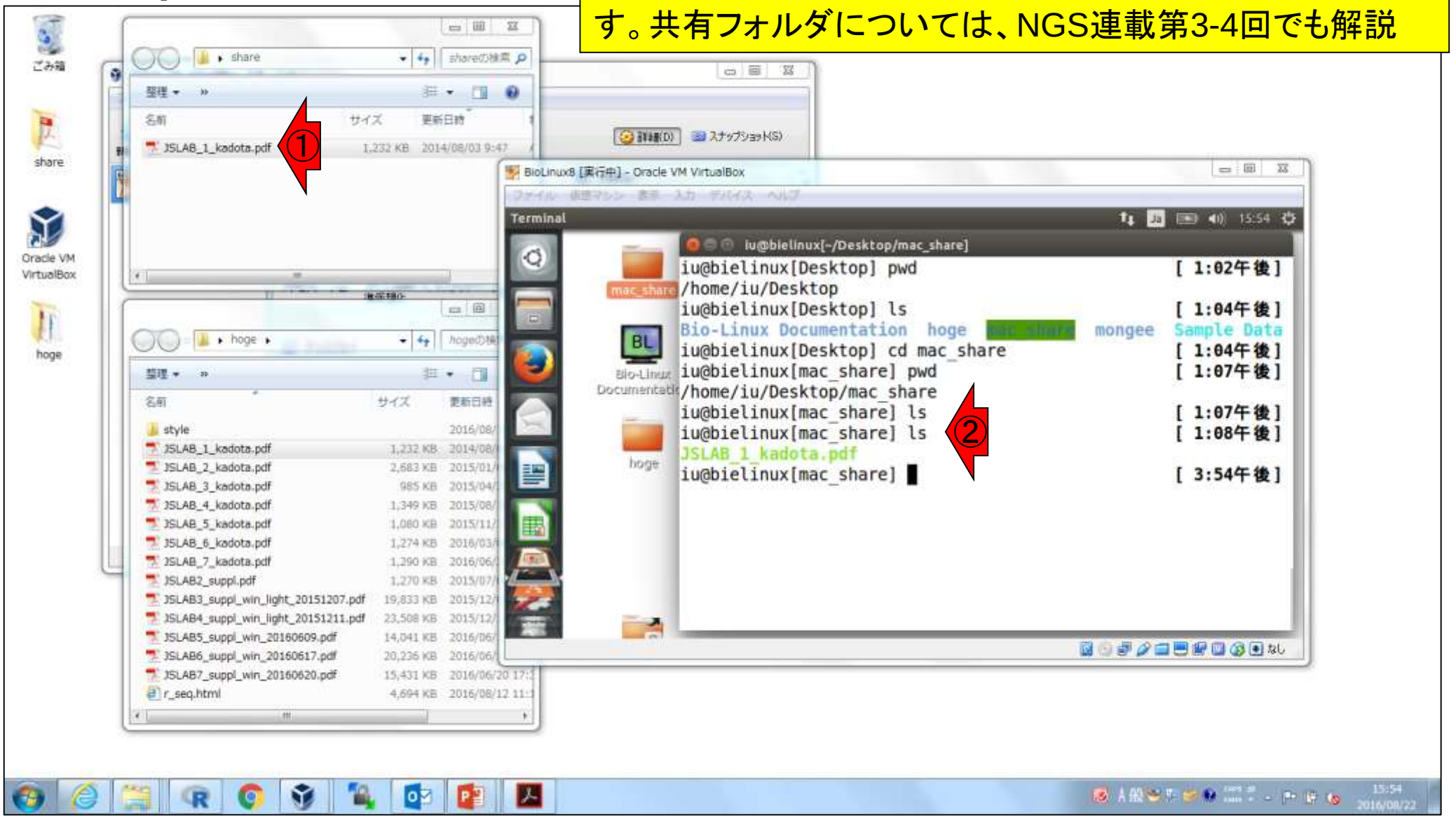
共有フォルダ

①こんな感じになります。共有フォルダなので、②mac_share上でも同じファイルが見えるはずです



共有フォルダ

②lsした結果、確かに見えました。こんな感じで、Linux上でのプログラム実行結果を共有フォルダ経由でWindowsに移動またはコピーし、Windows上で結果を整形するなどできます。共有フォルダについては、NGS連載第3-4回でも解説



Contents

- イントロダクション
  - 概要、背景（NGS用カリキュラム、講習会）、Linuxスキル習得の意義
  - ウェブ情報（日本乳酸菌学会誌のNGS連載やNGS講習会資料）
- 実習環境に慣れる
  - 仮想環境での作業に慣れる
  - GUIとCUI（マウス操作かコマンド入力操作か）
  - ターミナルでの作業
  - 共有フォルダの概念を理解
- 練習
  - 作業ディレクトリの変更、練習用NGSデータファイルのダウンロード
  - ファイルの確認、de novoゲノムアセンブリ
  - BLAST検索
- 課題
  - グループごとに異なる課題ファイルを入力として、「ダウンロード、de novoアセンブリ、BLAST検索」を実行し、得られた結果をレポートにまとめて発表せよ
  - グループ1はkadai1.fasta、グループ2はkadai2.fasta、etc.

練習

練習用として①仮想NGSデータファイル（hoge.fasta）の、②ダウンロードから③de novoアセンブリ、および④BLAST検索から、アセンブリ結果として得られた塩基配列が乳酸菌ゲノム配列であることの確認までを行います

- 門田幸二,「講義資料」, 東京大学農学部展開科目: バイオインフォマテ
  内容: バイオインフォマティクスやNGSをキーワード程度は知っているヒ
  オインフォマティクスとLinuxスキル習得の意義。バイオインフォマティクス分野で需要が多いのはNGSデータの解析 (JST-NBDCによるアンケート結果)。バイオインフォマティクス人材育成のための講習会の 一環として行われているNGS講習会資料を有効活用するために必要な基本スキルの習得。具体的には、日本乳酸菌学会誌のNGS連載内容を理解するための基礎として、Bio-Linux の に慣れるのが最低限の到達目標。何かやったという達成感を味わってもらうため、NGSデータのde novoゲノムアセンブリを行い、 ド長(塩基配列の長さ)よりも伸びたことを実感してもらう。また、アセンブリ結果の一部を BLAST検索 、入力データ(hoge.fasta; 約6.4MB)がどんな生物種であったかを調べるあたりまで。課題はkadai1.fasta, kadai2.fasta, kadai , kadai4.fastaのいずれかを選択して実行。下記と同様の手順で「データのダウンロード → de novoアセンブリ → BLAST検 、得られた結果を発表。約2日分。

①

④

②

③

```
#作業ディレクトリの変更
cd /home/iu/Desktop/mac_share

#解析したいファイル(hoge.fasta)のダウンロード
wget -c http://www.iu.a.u-tokyo.ac.jp/%7Ekadota/20160912/hoge.fasta

#行数やファイルサイズを表示(行数が200,000行、サイズが6,688,895 bytesとなっていればOK)
wc hoge.fasta

#最初の10行分を表示(2行分で1つのリードを表し、1リードが50塩基であることを確認)
head hoge.fasta

#de novoアセンブリを実行し、結果をugeディレクトリに保存
velveth uge 31 -short -fasta hoge.fasta
velvetg uge

#ugeディレクトリに移動して、アセンブリ結果ファイル(contigs.fa)があることを確認
cd uge
ls

#行数やファイルサイズを表示(入力のhoge.fastaは200,000行だったが、出力のcontigs.faは4,038行となっている)
wc contigs.fa
```

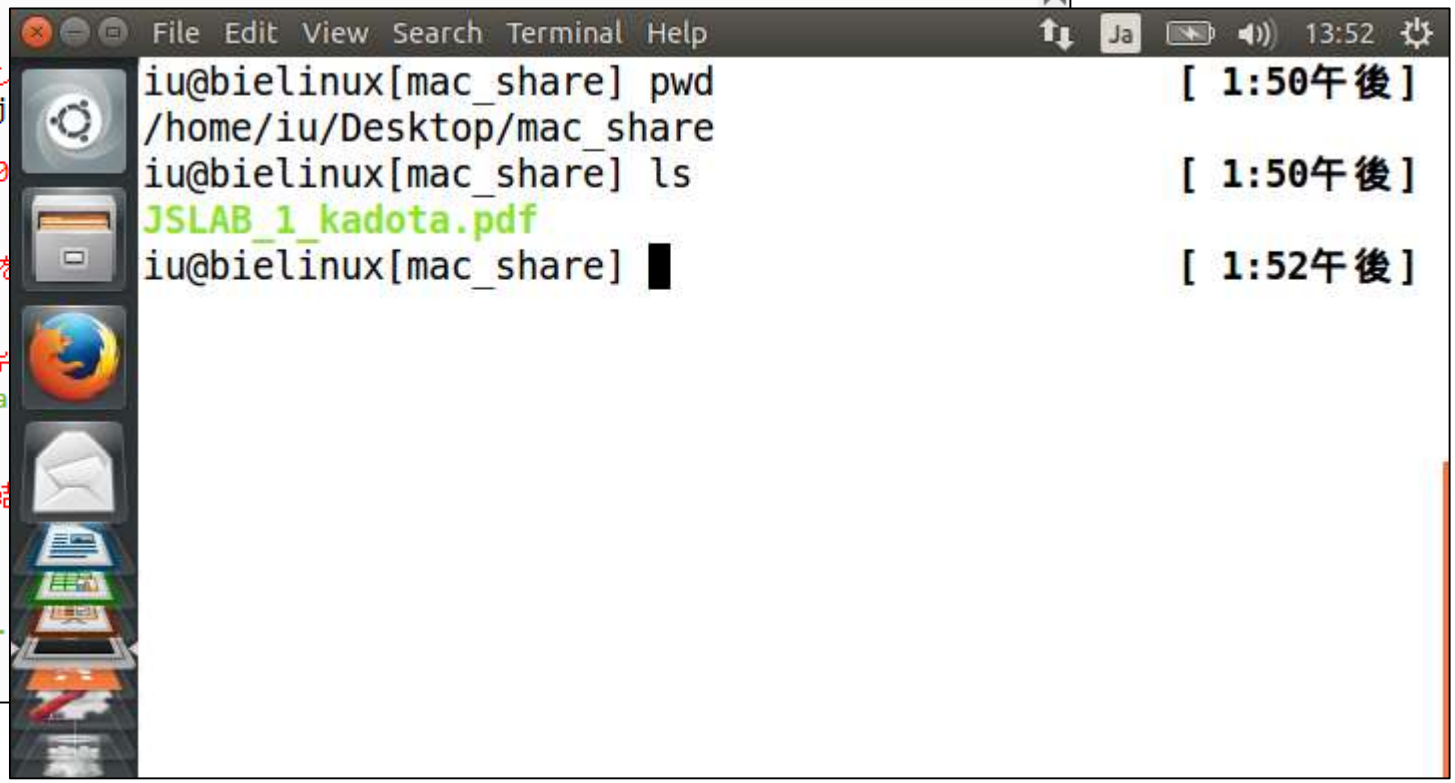
作業ディレクトリ

手順通りにやったヒトは、作業ディレクトリがmac_shareのままであり、pdfファイルが1つある状態。この場合、①をやる必要はないが、やってもよいのでやってみる

- 門田幸二,「講義資料」, 東京大学農学部展開科目: バイオインフォマティクス, 東京大学(東京), 2
  内容: バイオインフォマティクスやNGSをキーワード程度は知っているヒト(学部3年生程度)向け
  オインフォマティクスとLinuxスキル習得の意義。バイオインフォマティクス分野で需要が多いのはNGSデータの解析 (JST-NBDCによるアンケート結果)。バイオインフォマティクス人材育成のための講習会の 一環として行われているNGS講習会資料を有効活用するために必要な基本スキルの習得。具体的には、日本乳酸菌学会誌のNGS連載内容を理解するための基礎として、Bio-Linux のノリに慣れるのが最低限の到達目標。何かやったという達成感を味わってもらうため、NGSデータのde novoゲノムアセンブリを行い、元のリード長(塩基配列の長さ)よりも伸びたことを実感してもらう。また、アセンブリ結果の一部を BLAST検索し、入力データ(hoge.fasta; 約6.4MB)がどんな生物種であったかを調べるあたりまで。課題はkadai1.fasta, kadai2.fasta, kadai3.fasta, kadai4.fastaのいずれかを選択して実行。下記と同様の手順で「データのダウンロード → de novoアセンブリ → BLAST検索」し、得られた結果を発表。約2日分。

```
#作業ディレクトリの変更
cd /home/iu/Desktop/mac_share

#解析したいファイル(hoge.fasta)のダウン
wget -c http://www.iu.a.u-tokyo.ac.j

#行数やファイルサイズを表示(行数が200,0
wc hoge.fasta

#最初の10行分を表示(2行分で1つのリードを
head hoge.fasta

#de novoアセンブリを実行し、結果をugeデ
velveth uge 31 -short -fasta hoge.fa
velvetg uge

#ugeディレクトリに移動して、アセンブリ結
cd uge
ls

#行数やファイルサイズを表示(入力のhoge.
wc contigs.fa
```

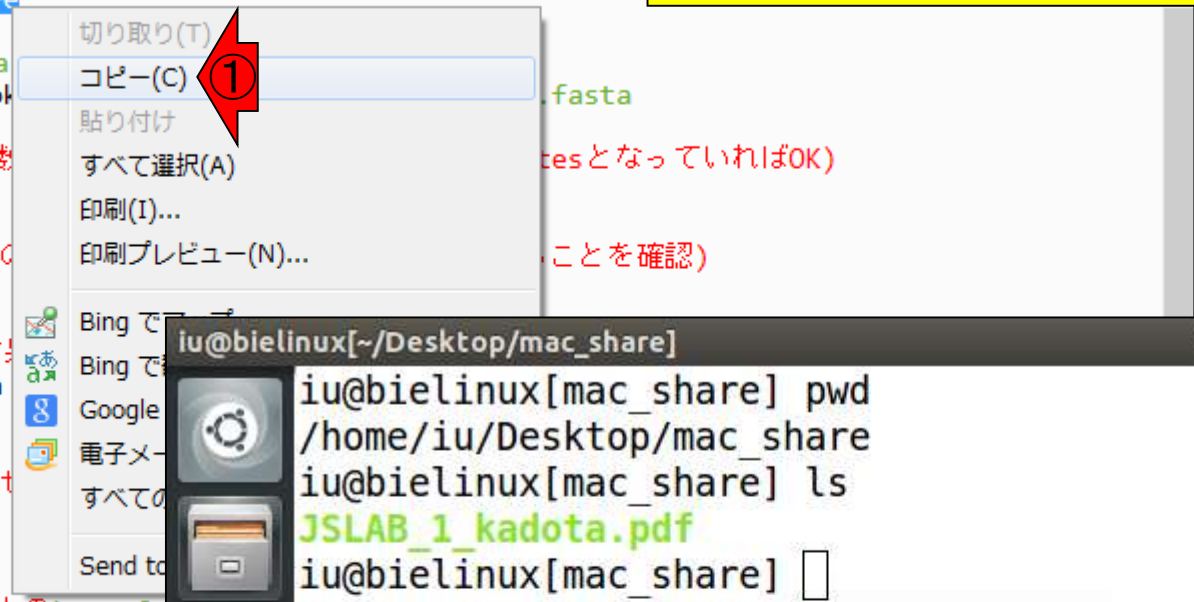
①

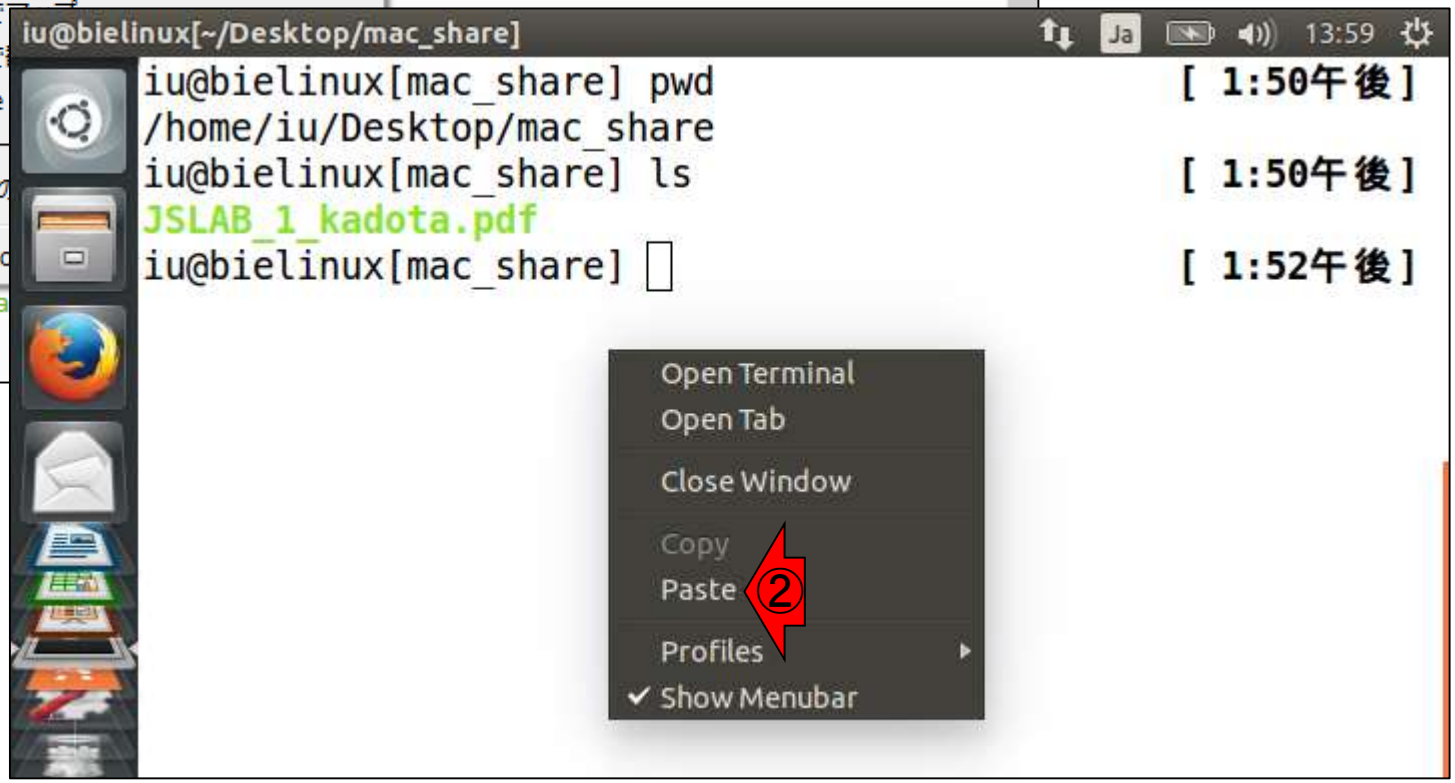


コピペ

①入力したいコマンド部分を反転させてコピー。②ターミナル画面上で右クリックしてペースト。ターミナル画面はLinuxの世界。Windows上で有効な「CTRL + v」は通用しません

```
#作業ディレクトリの変更
cd /home/iu/Desktop/mac_share

#解析したいファイル(hoge.fasta
wget -c http://www.iu.a.u-tok

#行数やファイルサイズを表示(行数
wc hoge.fasta

#最初の10行分を表示(2行分で1つ
head hoge.fasta

#de novoアセンブリを実行し、結
velveth uge 31 -short -fasta
velvetg uge

#ugeディレクトリに移動して、ア
cd uge
ls

#行数やファイルサイズを表示(入力のhoge.fa
wc contigs.fa
```





同じディレクトリなので何も変化はありません

コピペして確認

```
#作業ディレクトリの変更
cd /home/iu/Desktop/mac_share

#解析したいファイル(hoge.fasta
wget -c http://www.iu.a.u-tok

#行数やファイルサイズを表示(行数
wc hoge.fasta

#最初の10行分を表示(2行分で1つ
head hoge.fasta

#de novoアセンブリを実行し、結
velveth uge 31 -short -fasta
velvetg uge

#ugeディレクトリに移動して、ア
cd uge
ls

#行数やファイルサイズを表示(入力のhoge.fa
wc contigs.fa
```

```
.fasta

tesとなっていればOK)

ことを確認)
```

切り取り(T)
コピー(C)
貼り付け
すべて選択(A)
印刷(I)...
印刷プレビュー(N)...
Bing で
Bing で
Google
電子メー
すべての
Send to

```
File  Edit  View  Search  Terminal  Help                    Ja  14:05
iu@bielinux[mac_share] pwd                                  [ 1:50午後]
/home/iu/Desktop/mac_share
iu@bielinux[mac_share] ls                                   [ 1:50午後]
JSLAB_1_kadota.pdf
iu@bielinux[mac_share] cd /home/iu/Desktop/mac_share
iu@bielinux[mac_share] pwd                                  [ 2:04午後]
/home/iu/Desktop/mac_share
iu@bielinux[mac_share] ls                                   [ 2:04午後]
JSLAB_1_kadota.pdf
iu@bielinux[mac_share]                                      [ 2:04午後]
```

wgetでダウンロード

次は、①wgetというコマンドを用いて任意のURL上にあるファイル(hoge.fasta)のダウンロードです。②赤下線部分を丸々コピペでもいいのですが、せっかくなのでより汎用性の高い方法を伝授

- 門田幸二,「講義資料」, 東京大学農学部展開科目: バイオインフォマティクス, 東京大
  内容: バイオインフォマティクスやNGSをキーワード程度は知っているヒト(学部3年生
  オインフォマティクスとLinuxスキル習得の意義。バイオインフォマティクス分野で需要が多いのはNGSデータの解析 (JST-NBDCによるアンケート結果)。バイオインフォマティクス人材育成のための講習会の一環として行われているNGS講習会資料を有効活用するために必要な基本スキルの習得。具体的には、日本乳酸菌学会誌のNGS連載内容を理解するための基礎として、Bio-Linuxのノリに慣れるのが最低限の到達目標。何かやったという達成感を味わってもらうため、NGSデータのde novoゲノムアセンブリを行い、元のリード長(塩基配列の長さ)よりも伸びたことを実感してもらう。また、アセンブリ結果の一部をBLAST検索し、入力データ(hoge.fasta; 約6.4MB)がどんな生物種であったかを調べるあたりまで。課題はkadai1.fasta, kadai2.fasta, kadai3.fasta, kadai4.fastaのいずれかを選択して実行。下記と同様の手順で「データのダウンロード → de novoアセンブリ → BLAST検索」し、得られた結果を発表。約2日分。

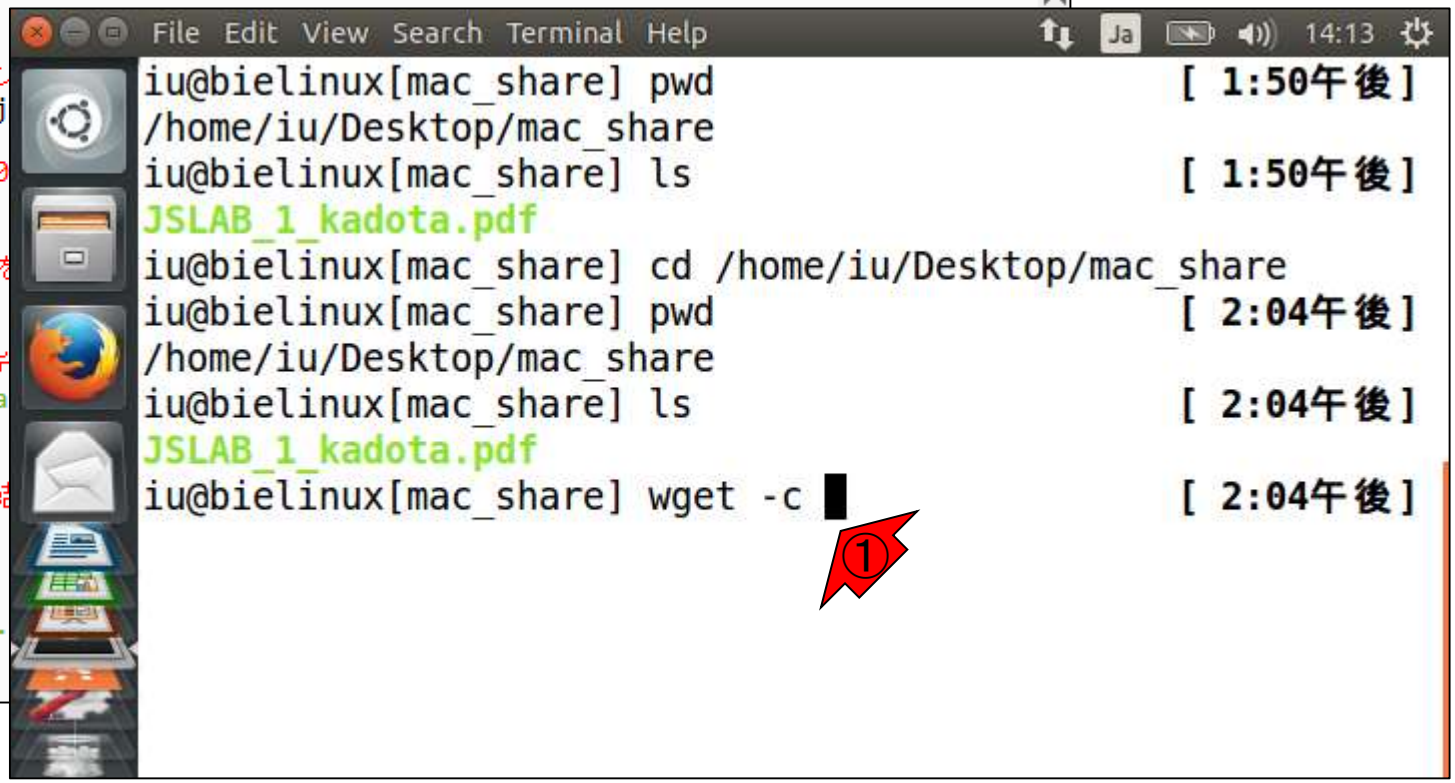
```
#作業ディレクトリの変更
cd /home/iu/Desktop/mac_share

#解析したいファイル(hoge.fasta)のダウンロード
wget -c http://www.iu.a.u-tokyo.ac.jp/%7Ekadota/20160912/hoge.fasta

#行数やファイルサイズを表示(行数が200,000行、サイズが6,688,895 bytesとなっていればOK)
wc hoge.fasta

#最初の10行分を表示(2行分で1つのリードを表し、1リードが50塩基であることを確認)
head hoge.fasta

#de novoアセンブリを実行し、結果をugeディレクトリに保存
velveth uge 31 -short -fasta hoge.fasta
velvetg uge

#ugeディレクトリに移動して、アセンブリ結果ファイル(contigs.fa)があることを確認
cd uge
ls

#行数やファイルサイズを表示(入力のhoge.fastaは200,000行だったが、出力のcontigs.faは4,038行となっている)
wc contigs.fa
```

wgetでダウンロード

①「wget -c 」(ダブルゲット、スペース、ハイフン、スペース)まで打ってから、②ダウンロードしたいファイル(hoge.fasta)のURL情報を取得

- 門田幸二,「講義資料」, 東京大学農学部展開科目: バイオインフォマティクス, 東京大学(東京), 2016.09.12-16
  内容: バイオインフォマティクスやNGSをキーワード程度は知っているヒト(学部3年生程度)向けのアクティブ・ラーニング系講義。バイオインフォマティクスとLinuxスキル習得の意義。バイオインフォマティクス分野で需要が多いのはNGSデータの解析 (JST-NBDCによるアンケート結果)。バイオインフォマティクス人材育成のための講習会の一環として行われているNGS講習会資料を有効活用するために必要な基本スキルの習得。具体的には、日本乳酸菌学会誌のNGS連載内容を理解するための基礎として、Bio-Linuxの使用に慣れるのが最低限の到達目標。何かやったという達成感を味わってもらうため、NGSデータのde novoゲノムアセンブリを行い、ノード長(塩基配列の長さ)よりも伸びたことを実感してもらう。また、アセンブリ結果の一部をBLAST検索し、入力データ(hoge.fasta, 約6.4MB)がどんな生物種であったかを調べるあたりまで。課題はkadai1.fasta, kadai2.fasta, kadai3.fasta, kadai4.fastaのいずれかを選択して実行。下記と同様の手順で「データのダウンロード → de novoアセンブリ → BLAST検索」し、得られた結果を発表。約2日分。

```
#作業ディレクトリの変更
cd /home/iu/Desktop/mac_share

#解析したいファイル(hoge.fasta)のダウン
wget -c http://www.iu.a.u-tokyo.ac.j

#行数やファイルサイズを表示(行数が200,0
wc hoge.fasta

#最初の10行分を表示(2行分で1つのリードを
head hoge.fasta

#de novoアセンブリを実行し、結果をugeデ
velveth uge 31 -short -fasta hoge.fa
velvetg uge

#ugeディレクトリに移動して、アセンブリ結
cd uge
ls

#行数やファイルサイズを表示(入力のhoge.
wc contigs.fa
```

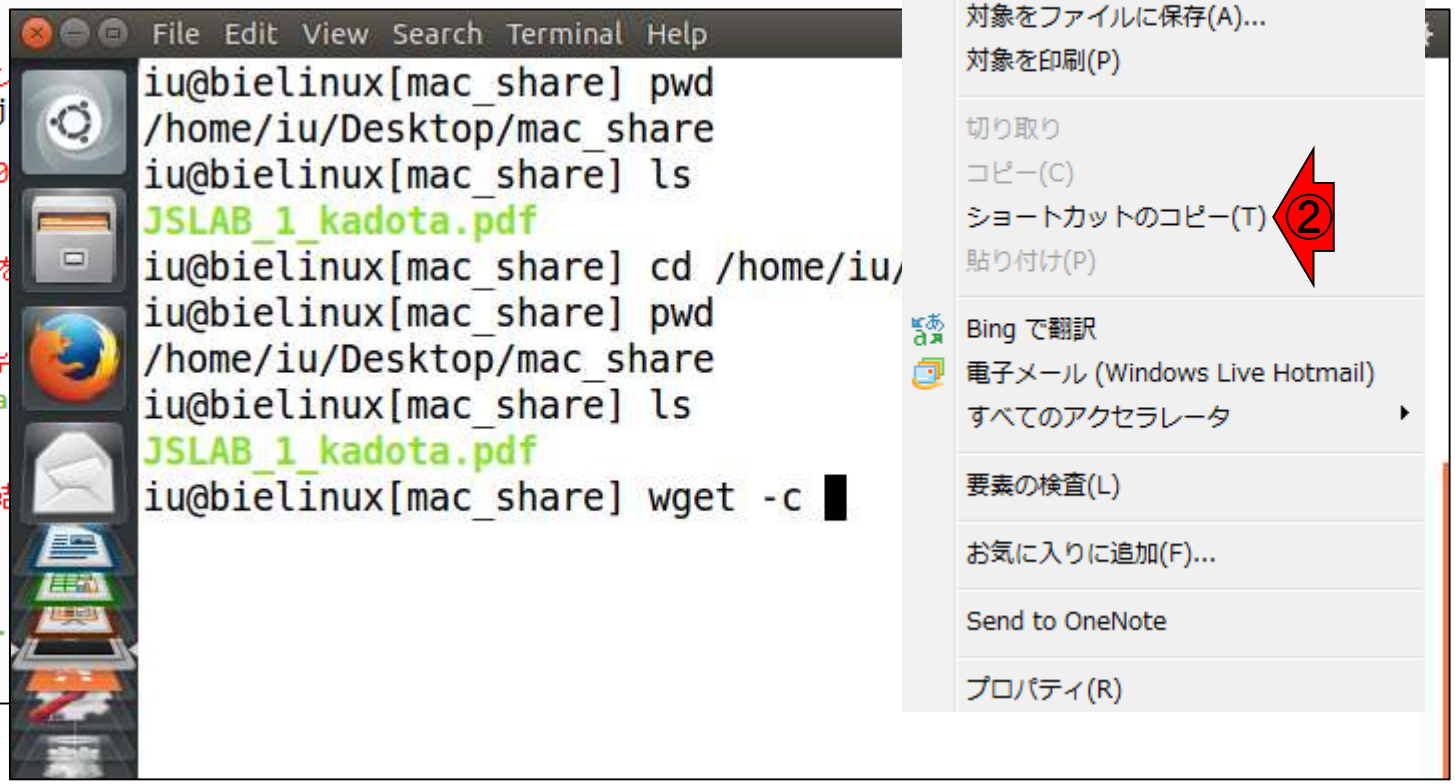


wgetでダウンロード

①ダウンロードしたいファイル(hoge.fasta)上で右クリックして、②「ショートカットのコピー」。Macintoshの場合は「リンク先のコピー」だったと思います

- 門田幸二,「講義資料」, 東京大学農学部展開科目: バイオインフォマティクス, 東京大学(東京), 2016.09.12-16
  内容: バイオインフォマティクスやNGSをキーワード程度は知っているヒト(学部3年生程度)向けのアクティブ・ラーニング系講義。バイオインフォマティクスとLinuxスキル習得の意義。バイオインフォマティクス分野で需要が多いのはNGSデータの解析 (JST-NBDCによるアンケート結果)。バイオインフォマティクス人材育成のための講習会の一環として行われているNGS講習会資料を有効活用するために必要な基本スキルの習得。具体的には、日本乳酸菌学会誌のNGS連載内容を理解するための基礎として、Bio-Linuxの[illegible]に慣れるのが最低限の到達目標。何かやったという達成感を味わってもらうため、NGSデータのde novoゲノムアセンブリを行い、[illegible]ド長(塩基配列の長さ)よりも伸びたことを実感してもらう。また、アセンブリ結果の一部をBLAST検索し、入力データ(hoge.fasta [illegible] 6.4MB)がどんな生物種であったかを調べるあたりまで。課題はkadai1.fasta, kadai2.fasta, kadai3.fasta, kadai4.fastaのいずれ[illegible]して実行。下記と同様の手順で「データのダウンロード → de novoアセンブリ → BLAST検索」し、得られた結果を発表.

```
#作業ディレクトリの変更
cd /home/iu/Desktop/mac_share

#解析したいファイル(hoge.fasta)のダウン
wget -c http://www.iu.a.u-tokyo.ac.j

#行数やファイルサイズを表示(行数が200,0
wc hoge.fasta

#最初の10行分を表示(2行分で1つのリードを
head hoge.fasta

#de novoアセンブリを実行し、結果をugeデ
velveth uge 31 -short -fasta hoge.fa
velvetg uge

#ugeディレクトリに移動して、アセンブリ結
cd uge
ls

#行数やファイルサイズを表示(入力のhoge.
wc contigs.fa
```

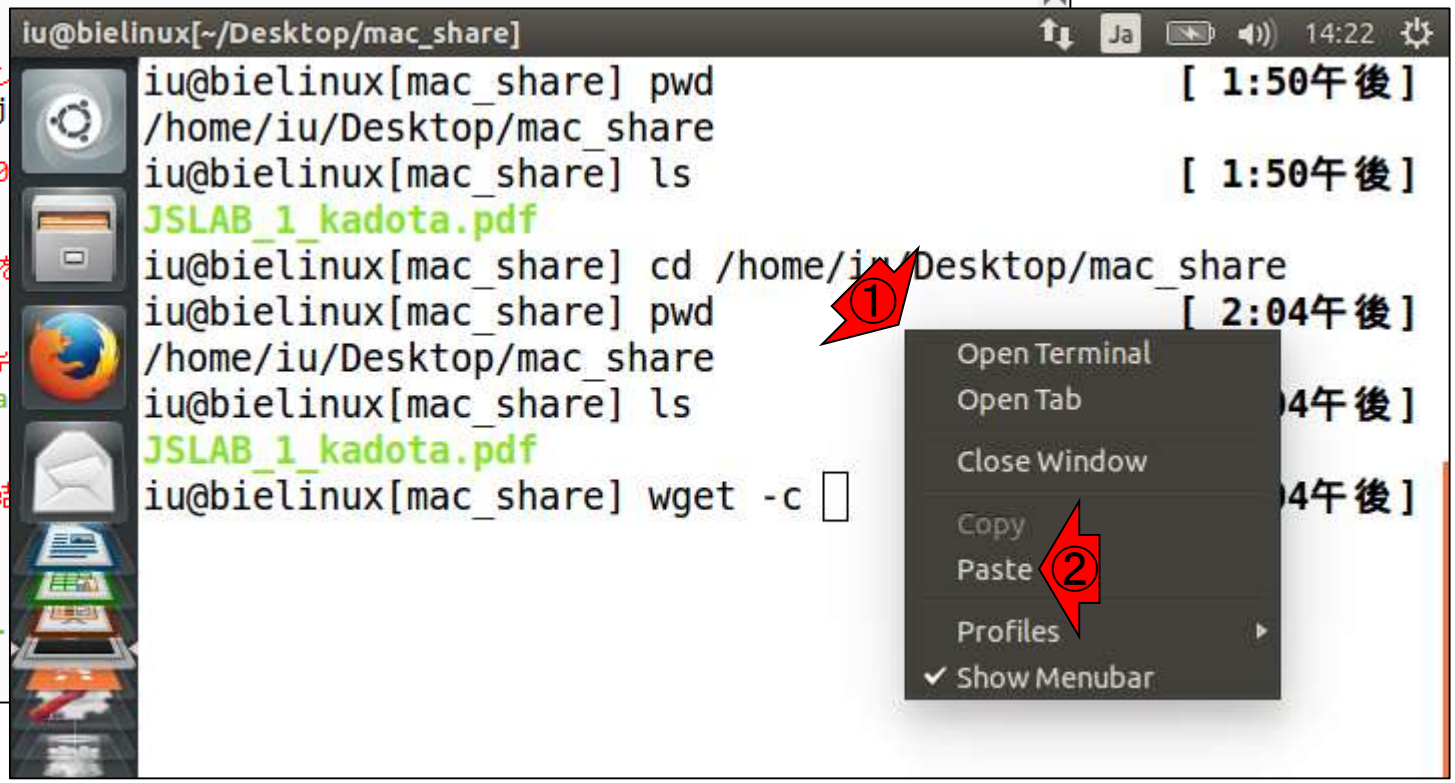


①ターミナル画面上で右クリックして、②ペースト

wgetでダウンロード

- 門田幸二,「講義資料」, 東京大学農学部展開科目: バイオインフォマティクス, 東京大学(東京), 2016.09.12-16
  内容: バイオインフォマティクスやNGSをキーワード程度は知っているヒト(学部3年生程度)向けのアクティブ・ラーニング系講義。バイオインフォマティクスとLinuxスキル習得の意義。バイオインフォマティクス分野で需要が多いのはNGSデータの解析 (JST-NBDCによるアンケート結果)。バイオインフォマティクス人材育成のための講習会の一環として行われているNGS講習会資料を有効活用するために必要な基本スキルの習得。具体的には、日本乳酸菌学会誌のNGS連載内容を理解するための基礎として、Bio-Linuxのノリに慣れるのが最低限の到達目標。何かやったという達成感を味わってもらうため、NGSデータのde novoゲノムアセンブリを行い、元のリード長(塩基配列の長さ)よりも伸びたことを実感してもらう。また、アセンブリ結果の一部をBLAST検索し、入力データ(hoge.fasta; 約6.4MB)がどんな生物種であったかを調べるあたりまで。課題はkadai1.fasta, kadai2.fasta, kadai3.fasta, kadai4.fastaのいずれかを選択して実行。下記と同様の手順で「データのダウンロード → de novoアセンブリ → BLAST検索」し、得られた結果を発表。約2日分。

```
#作業ディレクトリの変更
cd /home/iu/Desktop/mac_share

#解析したいファイル(hoge.fasta)のダウン
wget -c http://www.iu.a.u-tokyo.ac.j

#行数やファイルサイズを表示(行数が200,0
wc hoge.fasta

#最初の10行分を表示(2行分で1つのリードを
head hoge.fasta

#de novoアセンブリを実行し、結果をugeデ
velveth uge 31 -short -fasta hoge.fa
velvetg uge

#ugeディレクトリに移動して、アセンブリ結
cd uge
ls

#行数やファイルサイズを表示(入力のhoge.
wc contigs.fa
```



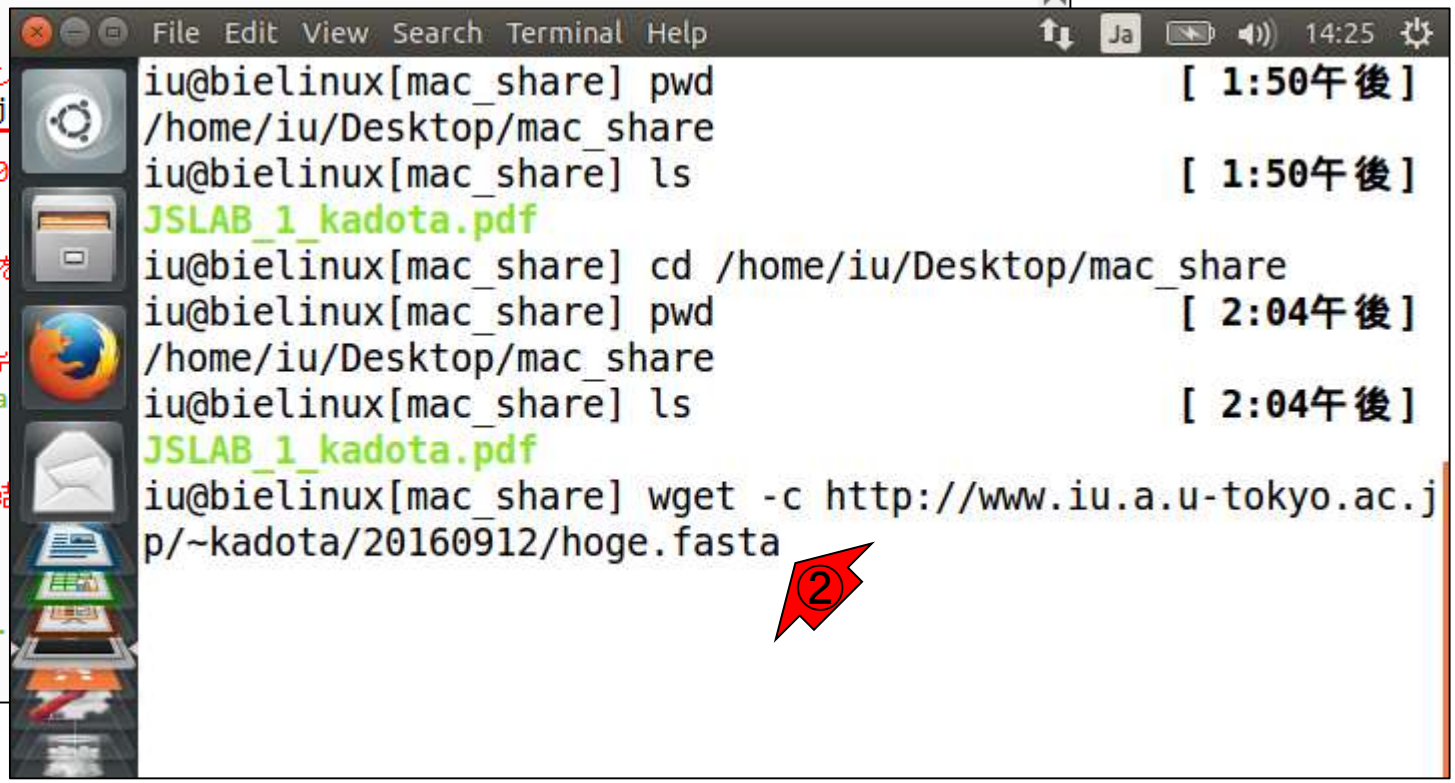
wgetでダウンロード

①赤下線部分と同じURL情報を、②ペーストできていることがわかります。リターンキーを押すとダウンロードが始まります

- 門田幸二,「講義資料」, 東京大学農学部展開科目: バイオインフォマティクス, 東京大学(東京), 2016.09.12-16
  内容: バイオインフォマティクスやNGSをキーワード程度は知っているヒト(学部3年生程度)向けのアクティブ・ラーニング系講義。バイオインフォマティクスとLinuxスキル習得の意義。バイオインフォマティクス分野で需要が多いのはNGSデータの解析 (JST-NBDCによるアンケート結果)。バイオインフォマティクス人材育成のための講習会の一環として行われているNGS講習会資料を有効活用するために必要な基本スキルの習得。具体的には、日本乳酸菌学会誌のNGS連載内容を理解するための基礎として、Bio-Linuxのノリに慣れるのが最低限の到達目標。何かやったという達成感を味わってもらうため、NGSデータのde novoゲノムアセンブリを行い、元のリード長(塩基配列の長さ)よりも伸びたことを実感してもらう。また、アセンブリ結果の一部をBLAST検索し、入力データ(hoge.fasta; 約6.4MB)がどんな生物種であったかを調べるあたりまで。課題はkadai1.fasta, kadai2.fasta, kadai3.fasta, kadai4.fastaのいずれかを選択して実行。下記と同様の手順で「データのダウンロード → de novoアセンブリ → BLAST検索」し、得られた結果を発表。約2日分。

```
#作業ディレクトリの変更
cd /home/iu/Desktop/mac_share

#解析したいファイル(hoge.fasta)のダウン
wget -c http://www.iu.a.u-tokyo.ac.j

#行数やファイルサイズを表示(行数が200,0
wc hoge.fasta

#最初の10行分を表示(2行分で1つのリードを
head hoge.fasta

#de novoアセンブリを実行し、結果をugeデ
velveth uge 31 -short -fasta hoge.fa
velvetg uge

#ugeディレクトリに移動して、アセンブリ結
cd uge
ls

#行数やファイルサイズを表示(入力のhoge.
wc contigs.fa
```

①

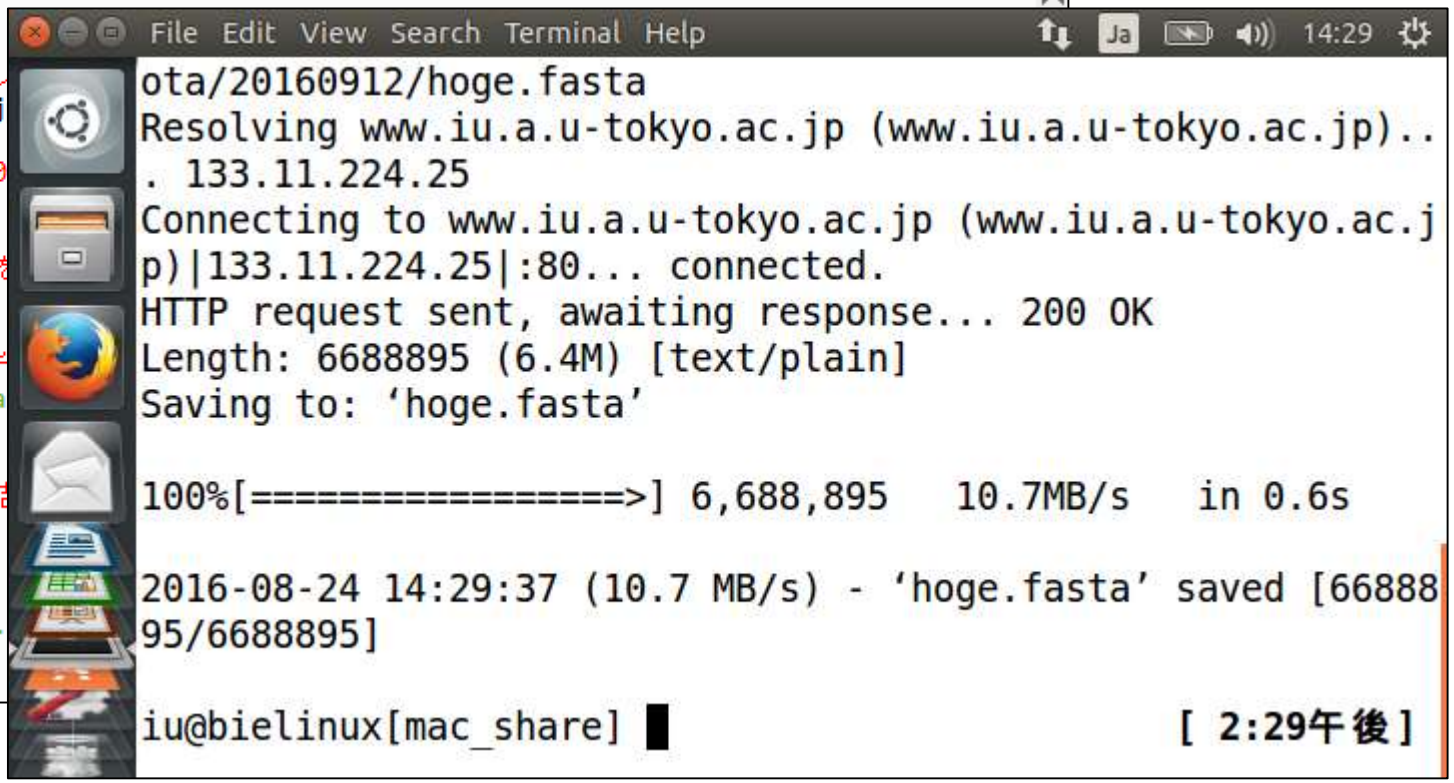


wget実行直後

無事ダウンロードが終了すると、こんな感じになります

- 門田幸二,「講義資料」, 東京大学農学部展開科目: バイオインフォマティクス, 東京大学(東京), 2016.09.12-16
  内容: バイオインフォマティクスやNGSをキーワード程度は知っているヒト(学部3年生程度)向けのアクティブ・ラーニング系講義。バイオインフォマティクスとLinuxスキル習得の意義。バイオインフォマティクス分野で需要が多いのはNGSデータの解析 (JST-NBDCによるアンケート結果)。バイオインフォマティクス人材育成のための講習会の一環として行われているNGS講習会資料を有効活用するために必要な基本スキルの習得。具体的には、日本乳酸菌学会誌のNGS連載内容を理解するための基礎として、Bio-Linuxのノリに慣れるのが最低限の到達目標。何かやったという達成感を味わってもらうため、NGSデータのde novoゲノムアセンブリを行い、元のリード長(塩基配列の長さ)よりも伸びたことを実感してもらう。また、アセンブリ結果の一部をBLAST検索し、入力データ(hoge.fasta; 約6.4MB)がどんな生物種であったかを調べるあたりまで。課題はkadai1.fasta, kadai2.fasta, kadai3.fasta, kadai4.fastaのいずれかを選択して実行。下記と同様の手順で「データのダウンロード → de novoアセンブリ → BLAST検索」し、得られた結果を発表。約2日分。

```
#作業ディレクトリの変更
cd /home/iu/Desktop/mac_share

#解析したいファイル(hoge.fasta)のダウン
wget -c http://www.iu.a.u-tokyo.ac.j

#行数やファイルサイズを表示(行数が200,0
wc hoge.fasta

#最初の10行分を表示(2行分で1つのリードを
head hoge.fasta

#de novoアセンブリを実行し、結果をugeデ
velveth uge 31 -short -fasta hoge.fa
velvetg uge

#ugeディレクトリに移動して、アセンブリ結
cd uge
ls

#行数やファイルサイズを表示(入力のhoge.
wc contigs.fa
```



Contents

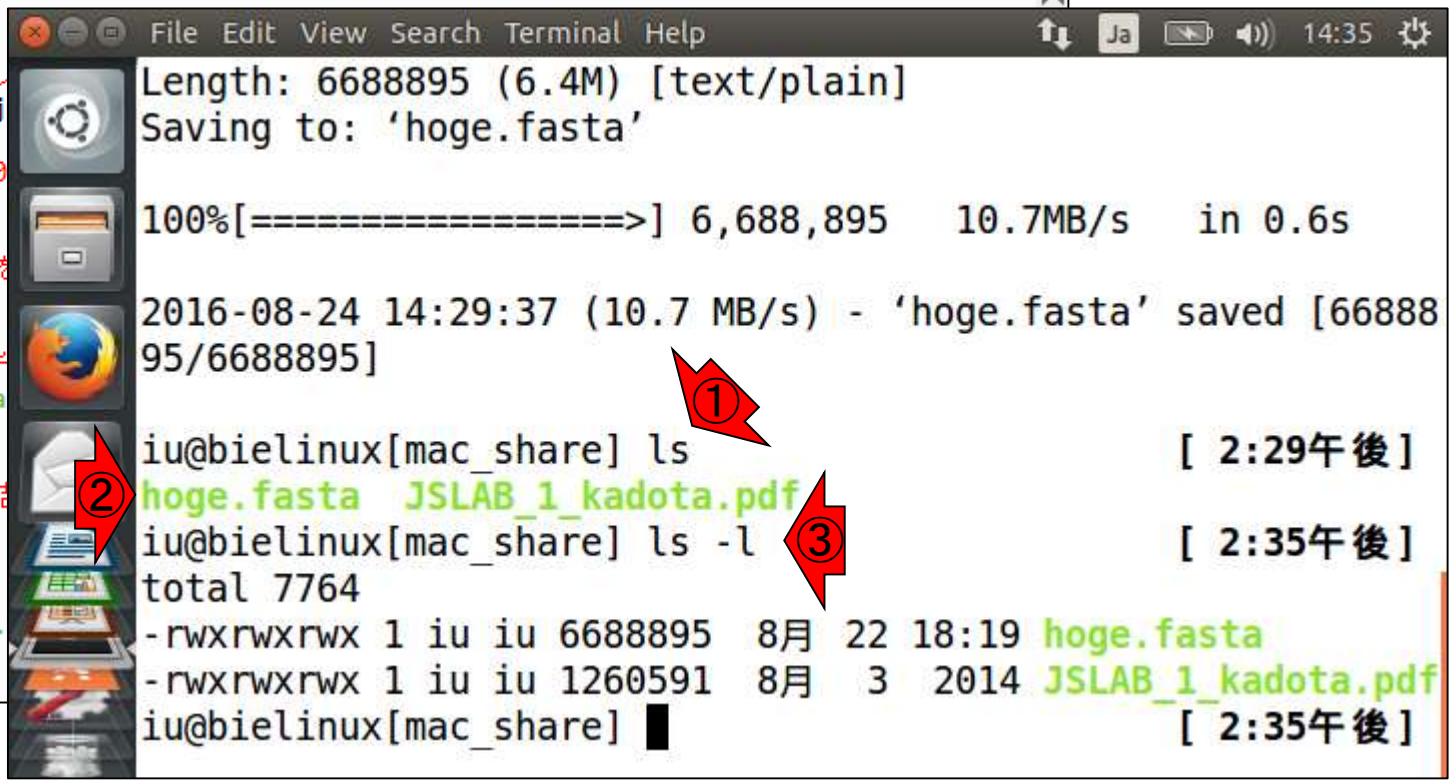
- イントロダクション
  - 概要、背景（NGS用カリキュラム、講習会）、Linuxスキル習得の意義
  - ウェブ情報（日本乳酸菌学会誌のNGS連載やNGS講習会資料）
- 実習環境に慣れる
  - 仮想環境での作業に慣れる
  - GUIとCUI（マウス操作かコマンド入力操作か）
  - ターミナルでの作業
  - 共有フォルダの概念を理解
- 練習
  - 作業ディレクトリの変更、練習用NGSデータファイルのダウンロード
  - ファイルの確認、de novoゲノムアセンブリ
  - BLAST検索
- 課題
  - グループごとに異なる課題ファイルを入力として、「ダウンロード、de novoアセンブリ、BLAST検索」を実行し、得られた結果をレポートにまとめて発表せよ
  - グループ1はkadai1.fasta、グループ2はkadai2.fasta、etc.

lsで確認

①ls（えるえす）で確認。②確かにダウンロードしたhoge.fastaがあります。③ls -l（エルエス、スペース、ハイフンえる）でより詳細な情報を見ることもできます

- 門田幸二,「講義資料」, 東京大学農学部展開科目: バイオインフォマティクス, 東京大学(東京), 2016
  内容: バイオインフォマティクスやNGSをキーワード程度は知っているヒト（学部3年生程度）向けのア
  オインフォマティクスとLinuxスキル習得の意義。バイオインフォマティクス分野で需要が多いのはNGSデータの解析 (JST-NBDCによるアンケート結果)。バイオインフォマティクス人材育成のための講習会の 一環として行われているNGS講習会資料を有効活用するために必要な基本スキルの習得。具体的には、日本乳酸菌学会誌のNGS連載内容を理解するための基礎として、Bio-Linux のノリに慣れるのが最低限の到達目標。何かやったという達成感を味わってもらうため、NGSデータのde novoゲノムアセンブリを行い、元のリード長(塩基配列の長さ)よりも伸びたことを実感してもらう。また、アセンブリ結果の一部を BLAST検索し、入力データ(hoge.fasta; 約6.4MB)がどんな生物種であったかを調べるあたりまで。課題はkadai1.fasta, kadai2.fasta, kadai3.fasta, kadai4.fastaのいずれかを選択して実行。下記と同様の手順で「データのダウンロード → de novoアセンブリ → BLAST検索」し、得られた結果を発表。約2日分。

```
#作業ディレクトリの変更
cd /home/iu/Desktop/mac_share

#解析したいファイル(hoge.fasta)のダウン
wget -c http://www.iu.a.u-tokyo.ac.j

#行数やファイルサイズを表示(行数が200,0
wc hoge.fasta

#最初の10行分を表示(2行分で1つのリードを
head hoge.fasta

#de novoアセンブリを実行し、結果をugeデ
velveth uge 31 -short -fasta hoge.fa
velvetg uge

#ugeディレクトリに移動して、アセンブリ結
cd uge
ls

#行数やファイルサイズを表示(入力のhoge.
wc contigs.fa
```



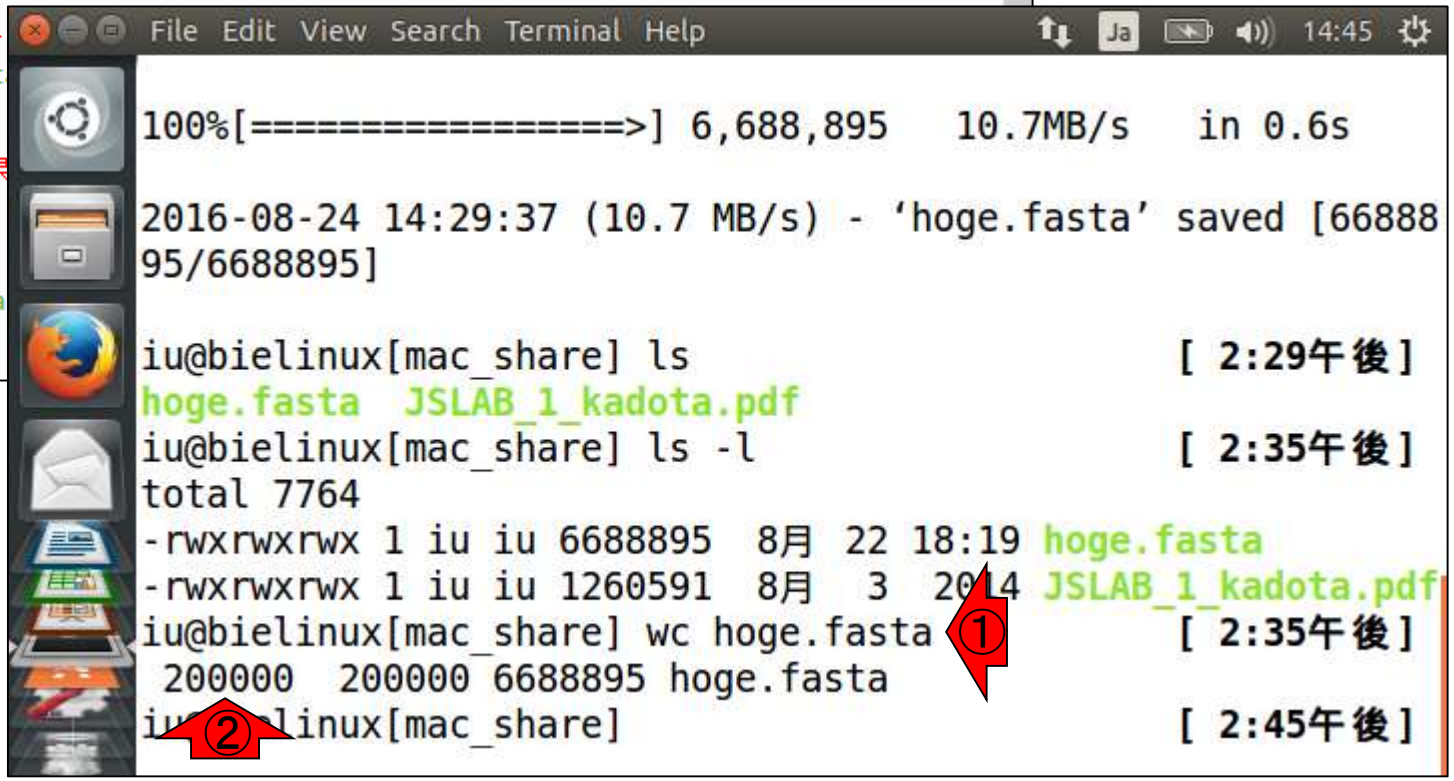
wcで確認

①wcコマンドは、主にファイルの行数を調べる目的で利用します。②確かに200,000行になっていることがわかります。NGSデータの場合は、行数から「(リードと呼ばれる)塩基配列の数」を調べることができます

```
#作業ディレクトリの変更
cd /home/iu/Desktop/mac_share

#解析したいファイル(hoge.fasta)のダウンロード
wget -c http://www.iu.a.u-tokyo.ac.jp/%7Ekadota/20160912/hoge.fasta

#行数やファイルサイズを表示(行数が200,000行、サイズが6,688,895 bytesとなっていればOK)
wc hoge.fasta

#最初の10行分を表示(2行分で1つのリードを表し、1リードが50塩基であることを確認)
head hoge.fasta

#de novoアセンブリを実行し、結果をugeディ
velveth uge 31 -short -fasta hoge.fast
velvetg uge

#ugeディレクトリに移動して、アセンブリ結果
cd uge
ls

#行数やファイルサイズを表示(入力のhoge.fa
wc contigs.fa
```

①

②

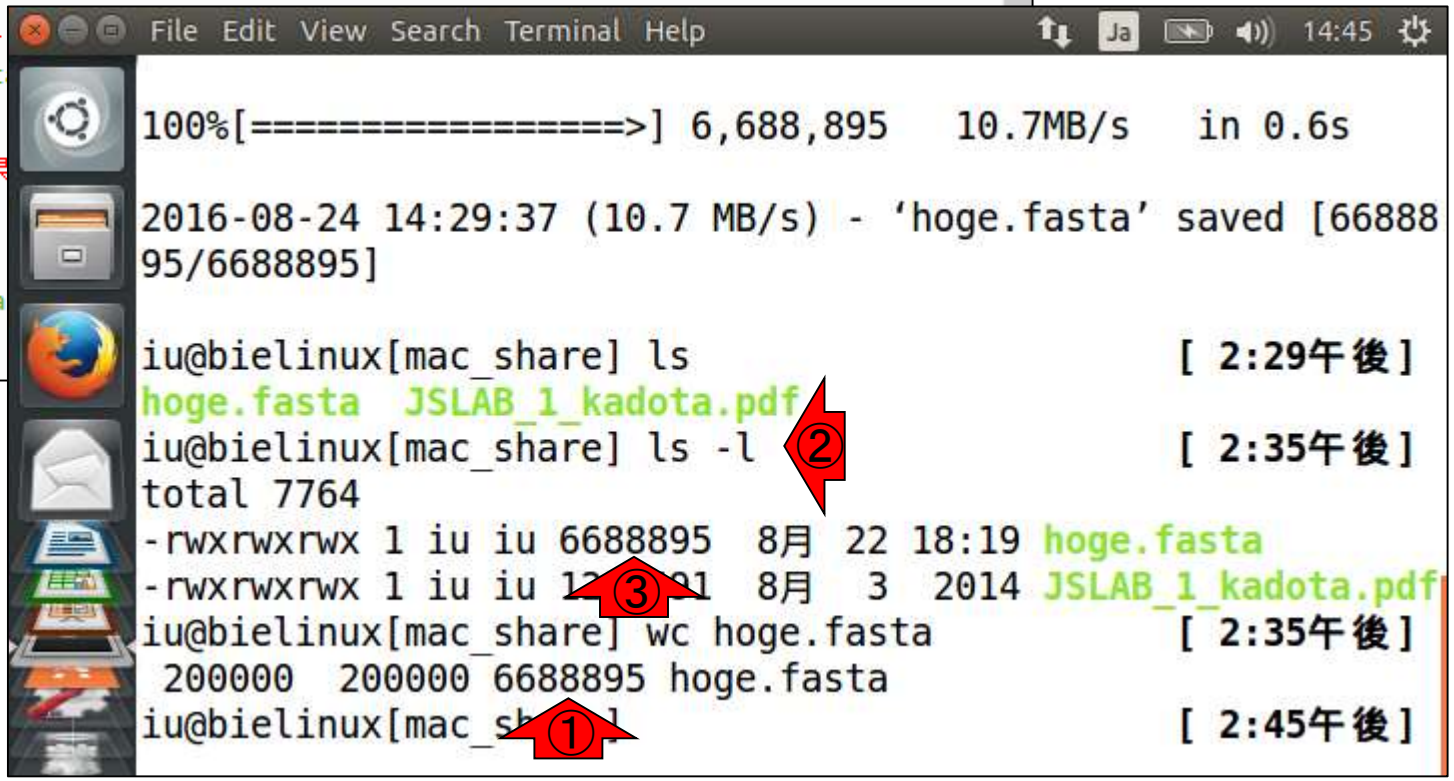


wcで確認

①ファイルサイズ情報。②ls -l実行結果として得られる、③の値と同じです

```
#作業ディレクトリの変更
cd /home/iu/Desktop/mac_share

#解析したいファイル(hoge.fasta)のダウンロード
wget -c http://www.iu.a.u-tokyo.ac.jp/%7Ekadota/20160912/hoge.fasta

#行数やファイルサイズを表示(行数が200,000行、サイズが6,688,895 bytesとなっていればOK)
wc hoge.fasta

#最初の10行分を表示(2行分で1つのリードを表し、1リードが50塩基であることを確認)
head hoge.fasta

#de novoアセンブリを実行し、結果をugeディ
velveth uge 31 -short -fasta hoge.fast
velvetg uge

#ugeディレクトリに移動して、アセンブリ結果
cd uge
ls

#行数やファイルサイズを表示(入力のhoge.fa
wc contigs.fa
```

①



headで確認

①headは、(デフォルトでは)ファイルの最初の10行分を表示させるコマンドです。このファイルは、FASTA形式と呼ばれるもので、2行で1つのリードを表します

```
#作業ディレクトリの変更
cd /home/iu/Desktop/mac_share

#解析したいファイル(hoge.fasta)のダウンロード
wget -c http://www.iu.a.u-tokyo.ac.jp/%7Ekadota/20160912/hoge.fasta

#行数やファイルサイズを表示(行数が200,000行、サイズが6,688,895 bytesとなっていればOK)
wc hoge.fasta

#最初の10行分を表示(2行分で1つのリードを表し、1リードが50塩基であることを確認)
head hoge.fasta

#de novoアセンブリを実行し、結果をugeディ
velveth uge 31 -short -fasta hoge.fast
velvetg uge

#ugeディレクトリに移動して、アセンブリ結果
cd uge
ls

#行数やファイルサイズを表示(入力のhoge.fa
wc contigs.fa
```

①

```
File  Edit  View  Search  Terminal  Help                    Ja  15:19
-rwxrwxrwx 1 iu iu 1260591  8月  3  2014 JSLAB_1_kadota.pdf
iu@bielinux[mac_share] wc hoge.fasta                    [ 2:35午後]
 200000  200000 6688895 hoge.fasta
iu@bielinux[mac_share] head hoge.fasta                  [ 2:45午後]
>sequence.1
ATGNATCGAAACAGTATTTACAAGATTTGCATACTGAAATTGAAGCTGAT
>sequence.2
GTCNGAACACATGAATGGTGAAACGGCGCTGAACTTTTCACGGACGCGGC
>sequence.3
CAANGATACAATCATTATCATGAACTCTAATGCCGGTTCTGGTGATGCAG
>sequence.4
AGANAGTATTTATCGTAAAATCTTAATTCATTCACAAGGTGGGGCGAACC
>sequence.5
AGTNCGGTTAGTGGAAGCTGCTAGGCATGTGCACACACCATTGTTTGCAC
iu@bielinux[mac_share]                                  [ 3:19午後]
```

①

headで確認

10塩基ごとに灰色の縦線を入れています。このNGSデータは、(少なくともここで見えている最初の5リード分については)50塩基の長さであることがわかります。①このファイル(hoge.fasta)は、②200,000行からなるので100,000リード。このように、大量の短いリード(short read)からなるのが典型的なNGSデータ

```
#作業ディレクトリの変更
cd /home/iu/Desktop/mac_share

#解析したいファイル(hoge.fasta)のダウンロード
wget -c http://www.iu.a.u-tokyo.ac.jp/%7Ekadota/20160912/hoge.fasta

#行数やファイルサイズを表示(行数が200,000行、サイズが6,688,895 bytesとなっていればOK)
wc hoge.fasta

#最初の10行分を表示(2行分で1つのリードを表し、1リードが50塩基であることを確認)
head hoge.fasta

#de novoアセンブリを実行し、結果をugeディ
velveth uge 31 -short -fasta hoge.fast
velvetg uge

#ugeディレクトリに移動して、アセンブリ結果
cd uge
ls

#行数やファイルサイズを表示(入力のhoge.fa
wc contigs.fa
```

```
File  Edit  View  Search  Terminal  Help                    Ja  15:19
-rwxrwxrwx 1 iu iu 1260591  8月  3  2014 JSLAB_1_kadota.pdf
iu@bielinux[mac_share] wc hoge.fasta                    [ 2:35午後]
 200000  200000 6688895 hoge.fasta
iu@bielinux[mac_share] head hoge.fasta                  [ 2:45午後]
>sequence.1
ATGNATCGAAACAGTATTTACAAGATTTGCATACTGAAATTGAAGCTGAT
>sequence.2
GTCNGAACACATGAATGGTGAAACGGCGCTGAACTTTTCACGGACGCGGC
>sequence.3
CAANGATACAATCATTATCATGAACTCTAATGCCGGTTCTGGTGATGCAG
>sequence.4
AGANAGTATTTATCGTAAAATCTTAATTCATTCACAAGGTGGGGCGAACC
>sequence.5
AGTNCGGTTAGTGGAAGCTGCTAGGCATGTGCACACACCATTGTTTGCAC
iu@bielinux[mac_share]                                  [ 3:19午後]
```

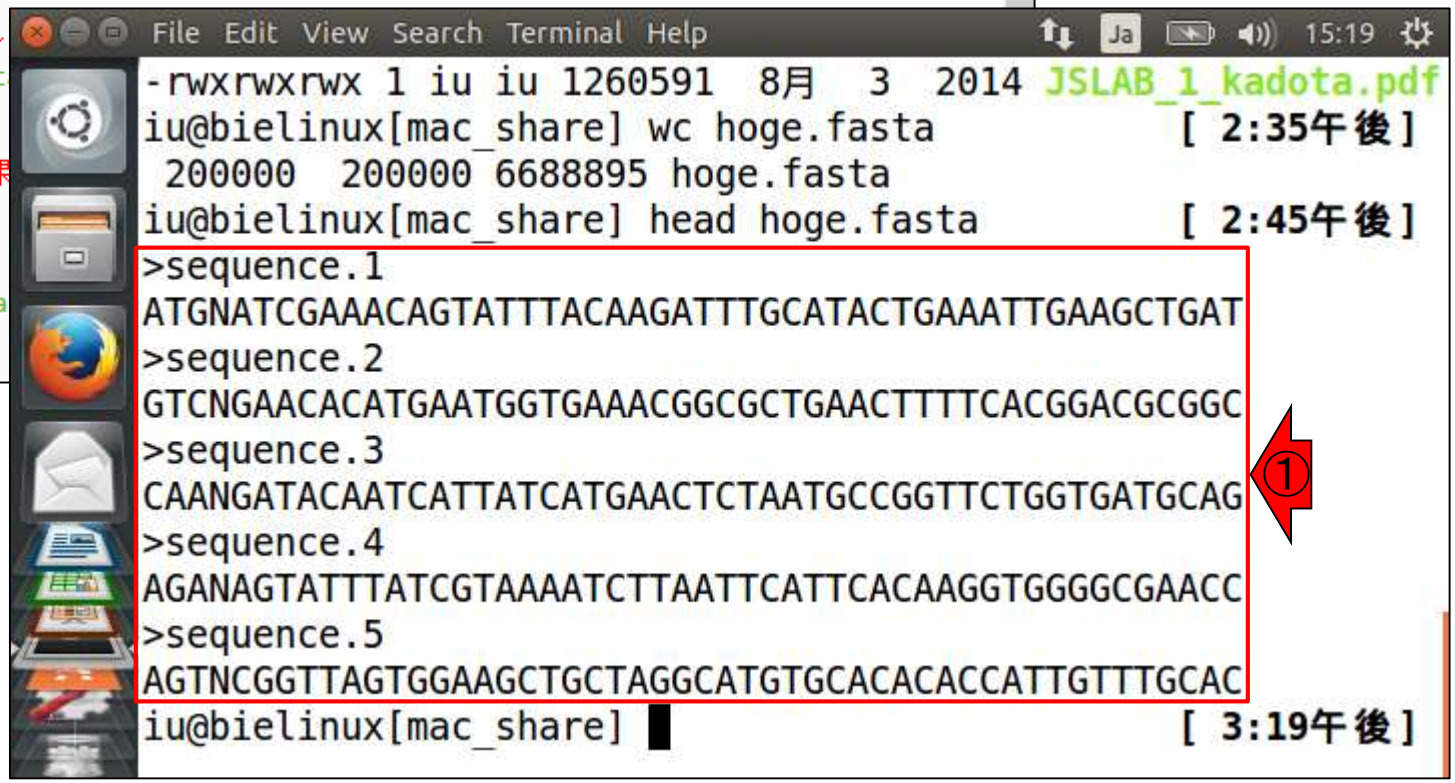
Contents

- イントロダクション
  - 概要、背景（NGS用カリキュラム、講習会）、Linuxスキル習得の意義
  - ウェブ情報（日本乳酸菌学会誌のNGS連載やNGS講習会資料）
- 実習環境に慣れる
  - 仮想環境での作業に慣れる
  - GUIとCUI（マウス操作かコマンド入力操作か）
  - ターミナルでの作業
  - 共有フォルダの概念を理解
- 練習
  - 作業ディレクトリの変更、練習用NGSデータファイルのダウンロード
  - ファイルの確認、de novoゲノムアセンブリ
  - BLAST検索
- 課題
  - グループごとに異なる課題ファイルを入力として、「ダウンロード、de novoアセンブリ、BLAST検索」を実行し、得られた結果をレポートにまとめて発表せよ
  - グループ1はkadai1.fasta、グループ2はkadai2.fasta、etc.

de novoアセンブリ

①このデータは乳酸菌(*Lactobacillus hokkaidonensis*)ゲノムの実際のNGSデータの一部。NGSデータ解析の目的の1つは、このような短い塩基配列データを入力として、より長い元のゲノム配列を再構築すること

```
#作業ディレクトリの変更
cd /home/iu/Desktop/mac_share

#解析したいファイル(hoge.fasta)のダウンロード
wget -c http://www.iu.a.u-tokyo.ac.jp/%7Ekadota/20160912/hoge.fasta

#行数やファイルサイズを表示(行数が200,000行、サイズが6,688,895 bytesとなっていればOK)
wc hoge.fasta

#最初の10行分を表示(2行分で1つのリードを表し、1リードが50塩基であることを確認)
head hoge.fasta

#de novoアセンブリを実行し、結果をugeディ
velveth uge 31 -short -fasta hoge.fast
velvetg uge

#ugeディレクトリに移動して、アセンブリ結果
cd uge
ls

#行数やファイルサイズを表示(入力のhoge.fa
wc contigs.fa
```



*de novo*アセンブリ

入出力のイメージ。*de novo*アセンブリとは、リードの塩基配列情報のみを頼りに、元のリード長よりも長い配列（コンティグ）を出力する作業。この例の場合、赤下線が一致部分。出力は、元のリード長よりも2塩基長いコンティグとなる

入力：NGSリードファイル

リード1: CACCAGGACATGAAGACGCG

リード2: CCAGGACATGAAGACGCGTT

出力：コンティグ（より長くなった塩基配列）

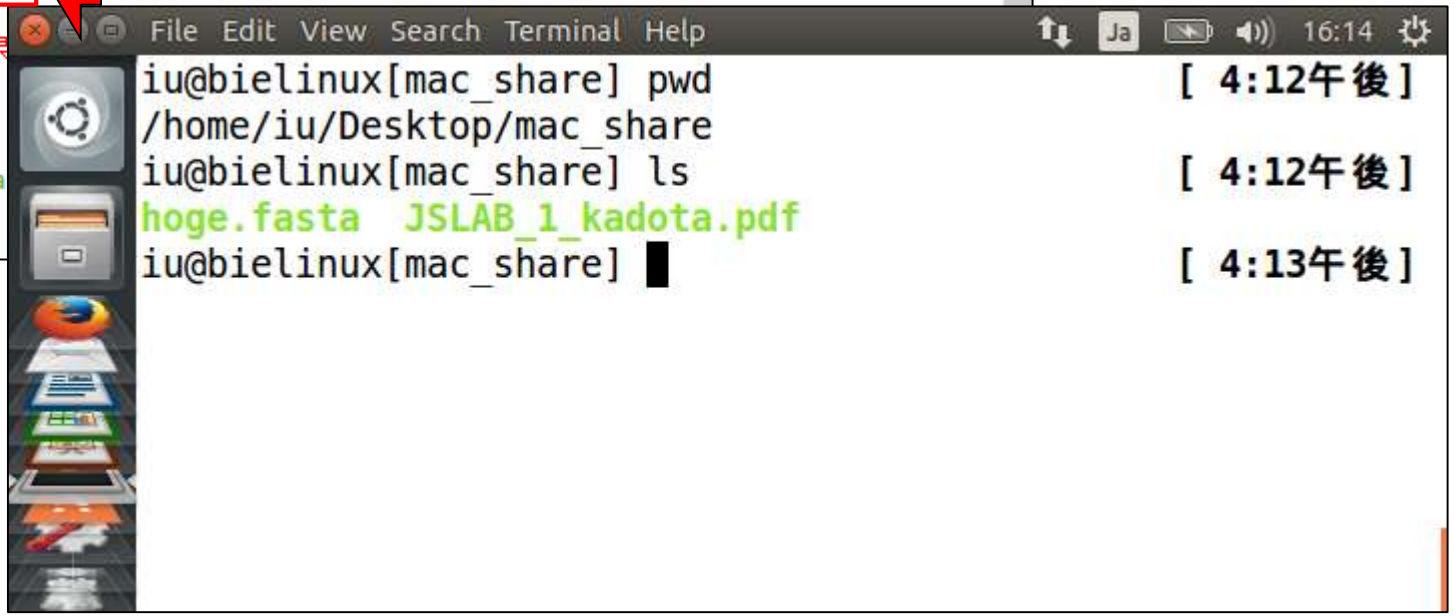
CACCAGGACATGAAGACGCGTT

de novoアセンブリ

①赤枠部分をコピペ実行。Velvetというアセンブリプログラムを実行しているが、細かいコマンドの意味などはここでは気にしなくてよい。ここで重要なのは、入力はhoge.fastaであり、プログラムを実行するとugeというディレクトリが作成されるということのみ。そしてugeディレクトリ内にあるcontigs.faが主なアセンブリ結果ファイル

```
#作業ディレクトリの変更
cd /home/iu/Desktop/mac_share

#解析したいファイル(hoge.fasta)のダウンロード
wget -c http://www.iu.a.u-tokyo.ac.jp/%7Ekadota/20160912/

#行数やファイルサイズを表示(行数が200,000行、サイズが6,688,895 bytesとなっていればOK)
wc hoge.fasta

#最初の10行分を表示(2行分で1つのリードを表し、1リードが50塩基であることを確認)
head hoge.fasta

#de novoアセンブリを実行し、結果をugeディレクトリに保存
velveth uge 31 -short -fasta hoge.fasta
velvetg uge

#ugeディレクトリに移動して、アセンブリ結果
cd uge
ls

#行数やファイルサイズを表示(入力のhoge.fa
wc contigs.fa
```

①

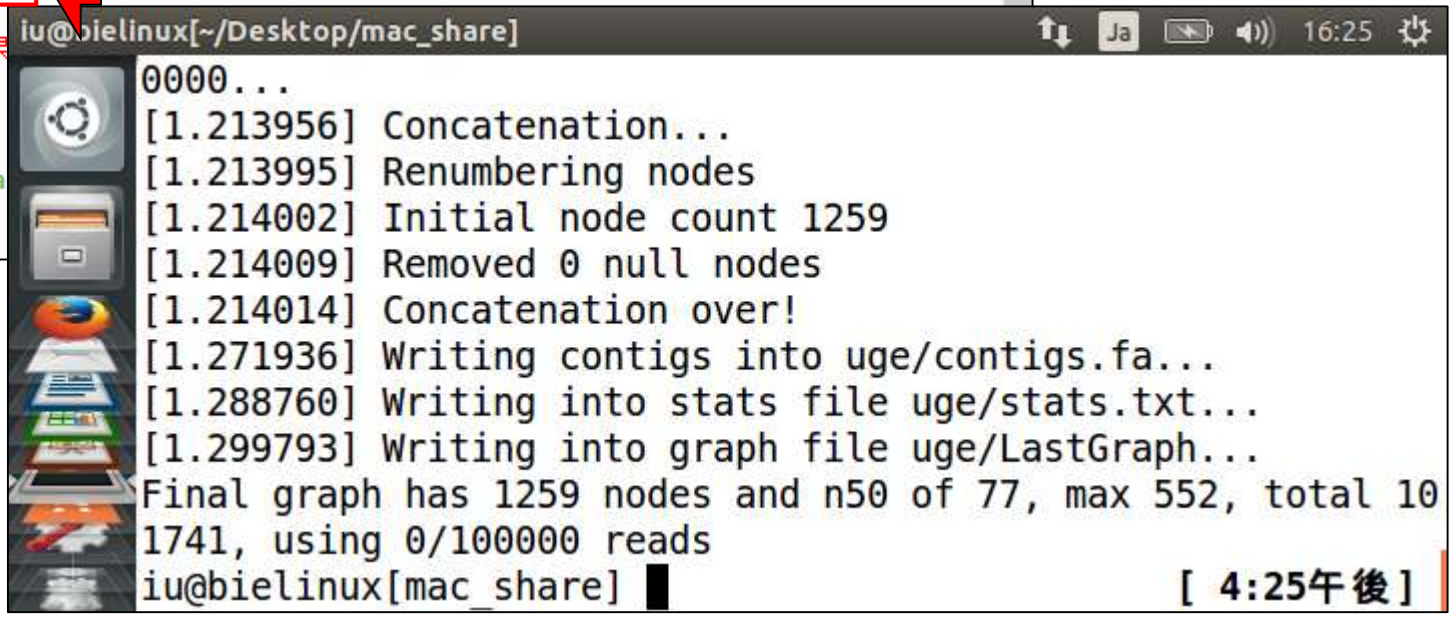


コピペ実行直後

①コピペ実行後の状態。計算自体は10秒程度で終わります

```
#作業ディレクトリの変更
cd /home/iu/Desktop/mac_share

#解析したいファイル(hoge.fasta)のダウンロード
wget -c http://www.iu.a.u-tokyo.ac.jp/%7Ekadota/20160912/hoge.fasta

#行数やファイルサイズを表示(行数が200,000行、サイズが6,688,895 bytesとなっていればOK)
wc hoge.fasta

#最初の10行分を表示(2行分で1つのリードを表し、1リードが50塩基であることを確認)
head hoge.fasta

#de novoアセンブリを実行し、結果をugeディレクトリに保存
velveth uge 31 -short -fasta hoge.fasta
velvetg uge

#ugeディレクトリに移動して、アセンブリ結果
cd uge
ls

#行数やファイルサイズを表示(入力のhoge.fa
wc contigs.fa
```

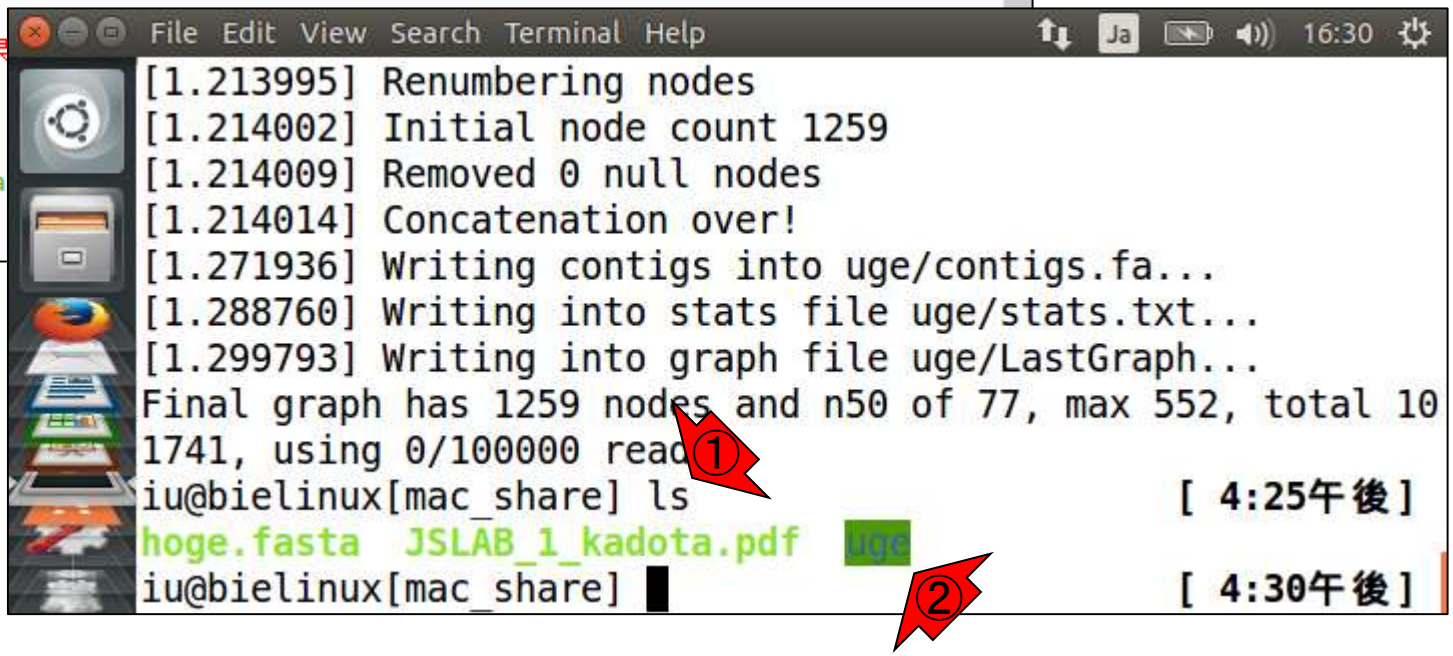
①



①ls。確かに②ugeディレクトリが作成されています

lsで確認

```
#作業ディレクトリの変更
cd /home/iu/Desktop/mac_share

#解析したいファイル(hoge.fasta)のダウンロード
wget -c http://www.iu.a.u-tokyo.ac.jp/%7Ekadota/20160912/hoge.fasta

#行数やファイルサイズを表示(行数が200,000行、サイズが6,688,895 bytesとなっていればOK)
wc hoge.fasta

#最初の10行分を表示(2行分で1つのリードを表し、1リードが50塩基であることを確認)
head hoge.fasta

#de novoアセンブリを実行し、結果をugeディレクトリに保存
velveth uge 31 -short -fasta hoge.fasta
velvetg uge

#ugeディレクトリに移動して、アセンブリ結果
cd uge
ls

#行数やファイルサイズを表示(入力のhoge.fa
wc contigs.fa
```

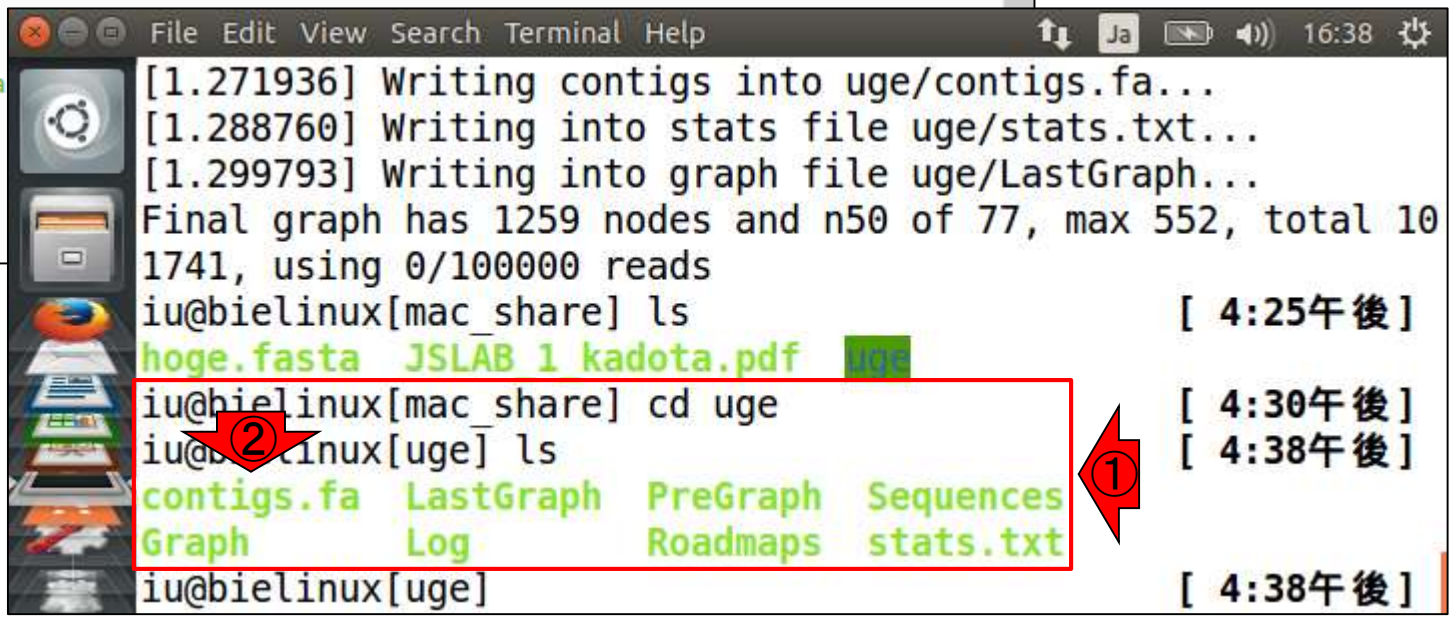


移動して確認

①ugeディレクトリに移動してls。②contigs.faが主なアセンブリ結果ファイルです

```
#解析したいファイル(hoge.fasta)のダウンロード
wget -c http://www.iu.a.u-tokyo.ac.jp/%7Ekadota/20160912/hoge.fasta

#行数やファイルサイズを表示(行数が200,000行、サイズが6,688,895 bytesとなっていればOK)
wc hoge.fasta

#最初の10行分を表示(2行分で1つのリードを表し、1リードが50塩基であることを確認)
head hoge.fasta

#de novoアセンブリを実行し、結果をugeディレクトリに保存
velveth uge 31 -short -fasta hoge.fasta
velvetg uge

#ugeディレクトリに移動して、アセンブリ結果ファイル(contigs.fa)があることを確認
cd uge
ls

#行数やファイルサイズを表示(入力のhoge.fa
wc contigs.fa

#最初の10行分を表示(アセンブリ結果として、
head contigs.fa
```

①

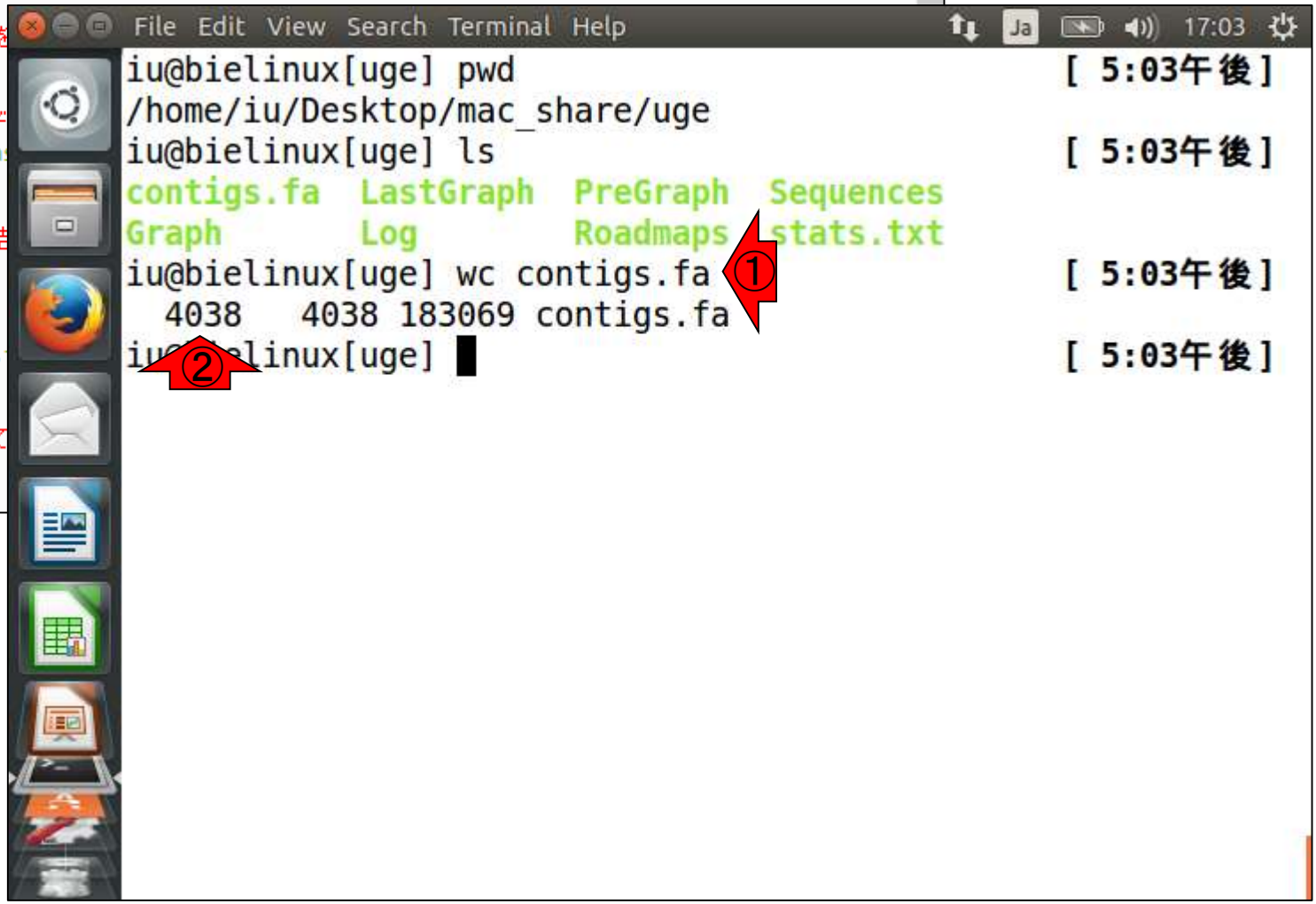


wcで確認

①wcでアセンブリ結果ファイル（contigs.fa）の行数を確認。②4,038行。入力（hoge.fasta）は200,000行であることから、行数が大幅に減ったことがわかる

```
#解析したいファイル(hoge.fasta)のダウンロード
wget -c http://www.iu.a.u-tokyo.ac.jp/%7Ekadota/20160912/hoge.fasta

#行数やファイルサイズを表示(行数が200,000行、サイズが6,688,895 bytesとなっていればOK)
wc hoge.fasta

#最初の10行分を表示(2行分で1つのリードを
head hoge.fasta

#de novoアセンブリを実行し、結果をugeデ
velveth uge 31 -short -fasta hoge.fa
velvetg uge

#ugeディレクトリに移動して、アセンブリ結
cd uge
ls

#行数やファイルサイズを表示(入力のhoge.
wc contigs.fa

#最初の10行分を表示(アセンブリ結果として
head contigs.fa
```

①



headで確認

①headでアセンブリ結果ファイル（contigs.fa）の最初の10行分を表示。パッと見で、入力（50塩基の長さのリードが100,000個）よりも長い塩基配列（コンティグという）が得られていることがわかる

```
#解析したいファイル(hoge.fasta)のダウンロード
wget -c http://www.iu.a.u-tokyo.ac.jp/%7Ekadota/20160912/hoge.

#行数やファイルサイズを表示(行数が200,000行、サイズが6,688,895 bytesとなっていればOK)
wc hoge.fasta

#最初の10行分を表示(2行分で1つのリードを
head hoge.fasta

#de novoアセンブリを実行し、結果をugeデ
velveth uge 31 -short -fasta hoge.fa
velvetg uge

#ugeディレクトリに移動して、アセンブリ結
cd uge
ls

#行数やファイルサイズを表示(入力のhoge.
wc contigs.fa

#最初の10行分を表示アセンブリ結果として
head contigs.fa
```

①

```
File  Edit  View  Search  Terminal  Help                    Ja  17:16
iu@bielinux[uge] pwd                                        [ 5:03午後]
/home/iu/Desktop/mac_share/uge
iu@bielinux[uge] ls                                         [ 5:03午後]
contigs.fa  LastGraph  PreGraph  Sequences
Graph       Log        Roadmaps  stats.txt
iu@bielinux[uge] wc contigs.fa                              [ 5:03午後]
  4038   4038 183069 contigs.fa
iu@bielinux[uge] head contigs.fa                            [ 5:03午後]
>NODE_1_length_124_cov_1.596774
CGGATTAGCTCAATTGATTATGGGAAATGCTGTACCCGGAGCCGTACTTGGTATTTTAGT
GGGTAAAGGTGTTGATGAACTTGGCTGGACCAAAACAACTAAAATTATGTTAACAGCTGT
AATCTTGCTATTTGTCTTGAGTGGGTTCTTCCGC
>NODE_2_length_81_cov_1.530864
ATCCAACCGCACCAAAGTTGGCTGTCATGACTATCCGCGTTACTTGTGCGGCATCAGCAG
CTTTCAAGATGCGAATAATACCTTTTTTTGCAGCCTGGTCCATTTCTTCTG
>NODE_3_length_94_cov_2.340425
TGGAGTTGTAATAATGGCAAATAATAATGATTCTACTATGAGTCGTGAAGAATTACGTAA
CAGTAAAAAGAATCAACAACAAAATGAACAAAACCAATCCGGACCAGTTAAATCAAAGAA
iu@bielinux[uge]                                            [ 5:16午後]
```

Contents

- イントロダクション
  - 概要、背景（NGS用カリキュラム、講習会）、Linuxスキル習得の意義
  - ウェブ情報（日本乳酸菌学会誌のNGS連載やNGS講習会資料）
- 実習環境に慣れる
  - 仮想環境での作業に慣れる
  - GUIとCUI（マウス操作かコマンド入力操作か）
  - ターミナルでの作業
  - 共有フォルダの概念を理解
- 練習
  - 作業ディレクトリの変更、練習用NGSデータファイルのダウンロード
  - ファイルの確認、de novoゲノムアセンブリ
  - BLAST検索
- 課題
  - グループごとに異なる課題ファイルを入力として、「ダウンロード、de novoアセンブリ、BLAST検索」を実行し、得られた結果をレポートにまとめて発表せよ
  - グループ1はkadai1.fasta、グループ2はkadai2.fasta、etc.

BLAST検索

世界中から得られた塩基配列のデータベース(の一部)に対して、手元にある塩基配列をBLASTというプログラムを用いて検索する作業。配列相同性検索ともいいます。詳細については秋の講義科目「生物情報科学」で説明がなされると思います。ここでは詳細はすっ飛ばして、必要最小限の作業を行う

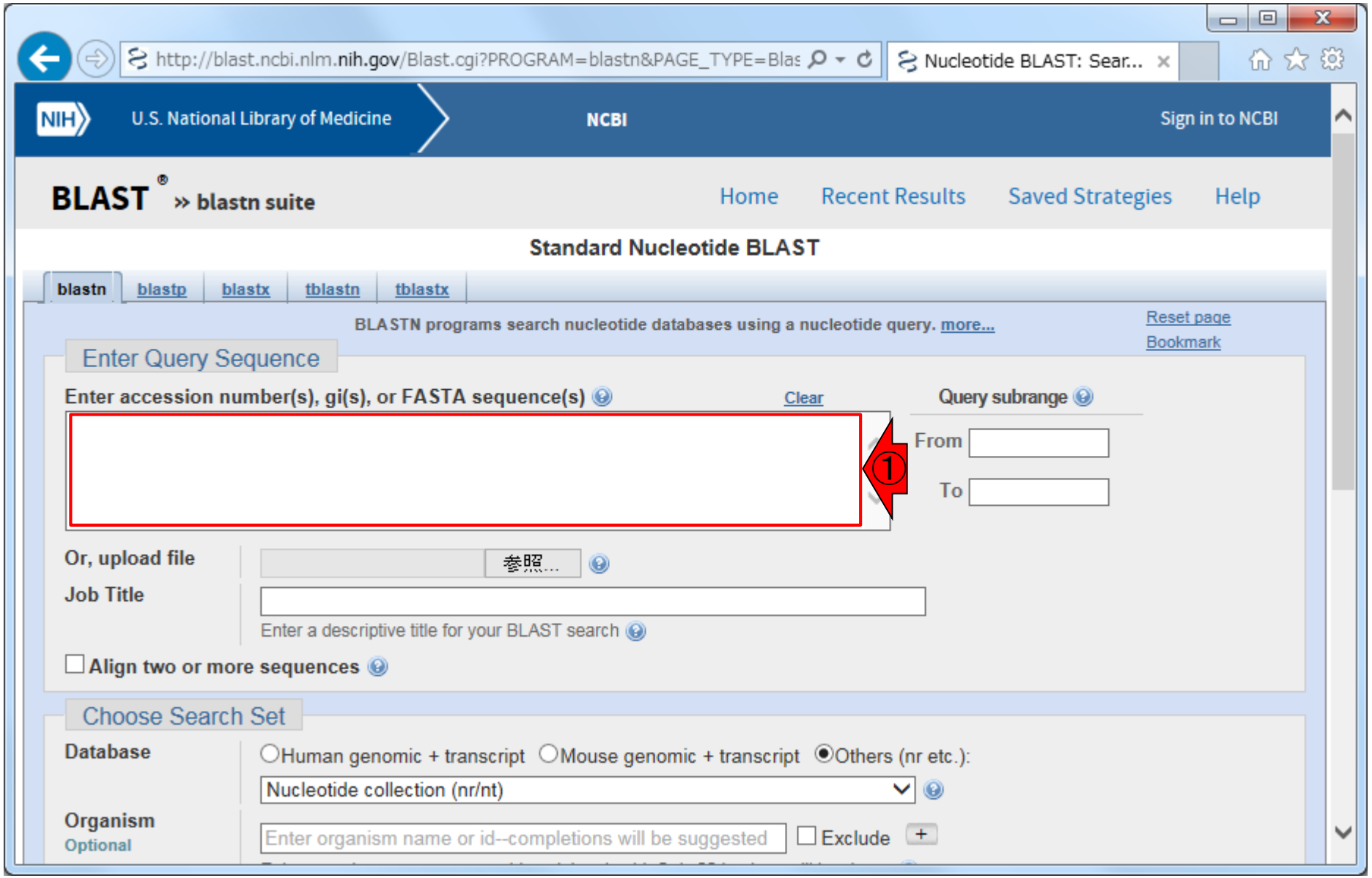
- 門田幸二,「講義資料」, 東京大学農学部展開科目: バイオイン
  内容: バイオインフォマティクスやNGSをキーワード程度は知っ
  オインフォマティクスとLinuxスキル習得の意義。バイオインフ
  るアンケート結果)。バイオインフォマティクス人材育成のための講習会の一環として行われているNGS講習会資料を有効活用するた
  めに必要な基本スキルの習得。具体的には、日本乳酸菌学会誌のNGS連載内容を理解するための基礎として、Bio-Linuxのノリに慣
  れるのが最低限の到達目標。何かやったという達成感を味わってもらうため、NGSデータのde novoゲノムアセンブリを行い、元のリー
  ド長(塩基配列の長さ)よりも伸びたことを実感してもらう。また、アセンブリ結果の一部をBLAST検索し、入力データ(hoge.fasta; 約
  6.4MB)がどんな生物種であったかを調べるあたりまで。課題はkadai1.fasta, kadai2.fasta, kadai3.fasta, kadai4.fastaのいずれかを選択
  して実行。下記と同様の手順で「データのダウンロード → de novoアセンブリ → BLAST検索」し、得られた結果を発表。約2日分。

①

```
#作業ディレクトリの変更
cd /home/iu/Desktop/mac_share

#解析したいファイル(hoge.fasta)のダウンロード
wget -c http://www.iu.a.u-tokyo.ac.jp/%7Ekadota/20160912/hoge.fasta

#行数やファイルサイズを表示(行数が200,000行、サイズが6,688,895 bytesとなっていればOK)
wc hoge.fasta

#最初の10行分を表示(2行分で1つのリードを表し、1リードが50塩基であることを確認)
head hoge.fasta

#de novoアセンブリを実行し、結果をugeディレクトリに保存
velveth uge 31 -short -fasta hoge.fasta
velvetg uge

#ugeディレクトリに移動して、アセンブリ結果ファイル(contigs.fa)があることを確認
cd uge
ls

#行数やファイルサイズを表示(入力のhoge.fastaは200,000行だったが、出力のcontigs.faは4,038行となっている)
wc contigs.fa
```

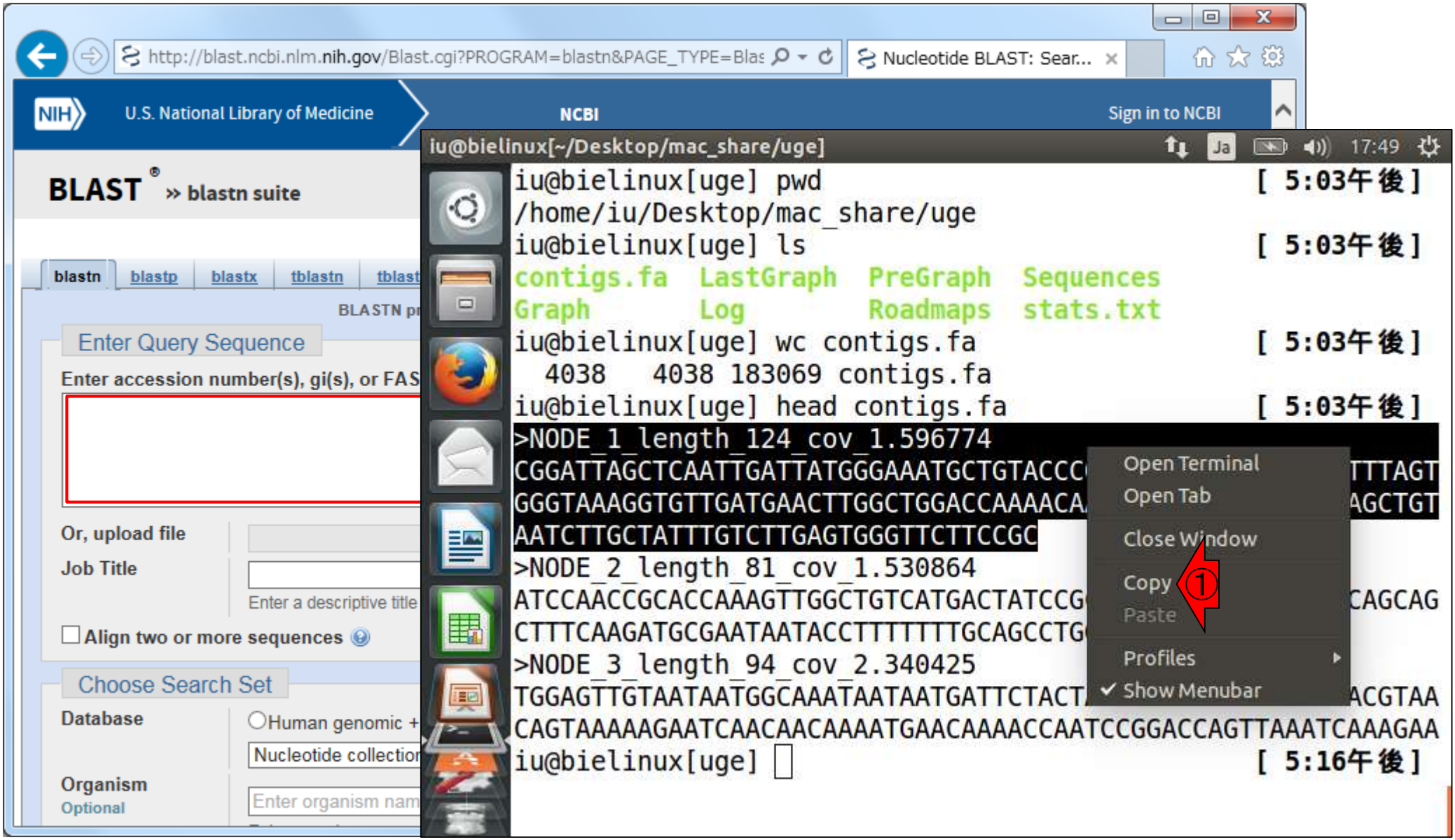
BLAST検索

BLASTのトップ画面。①の部分にアセンブリ結果として得られた配列の一部を入力としてBLASTを実行する



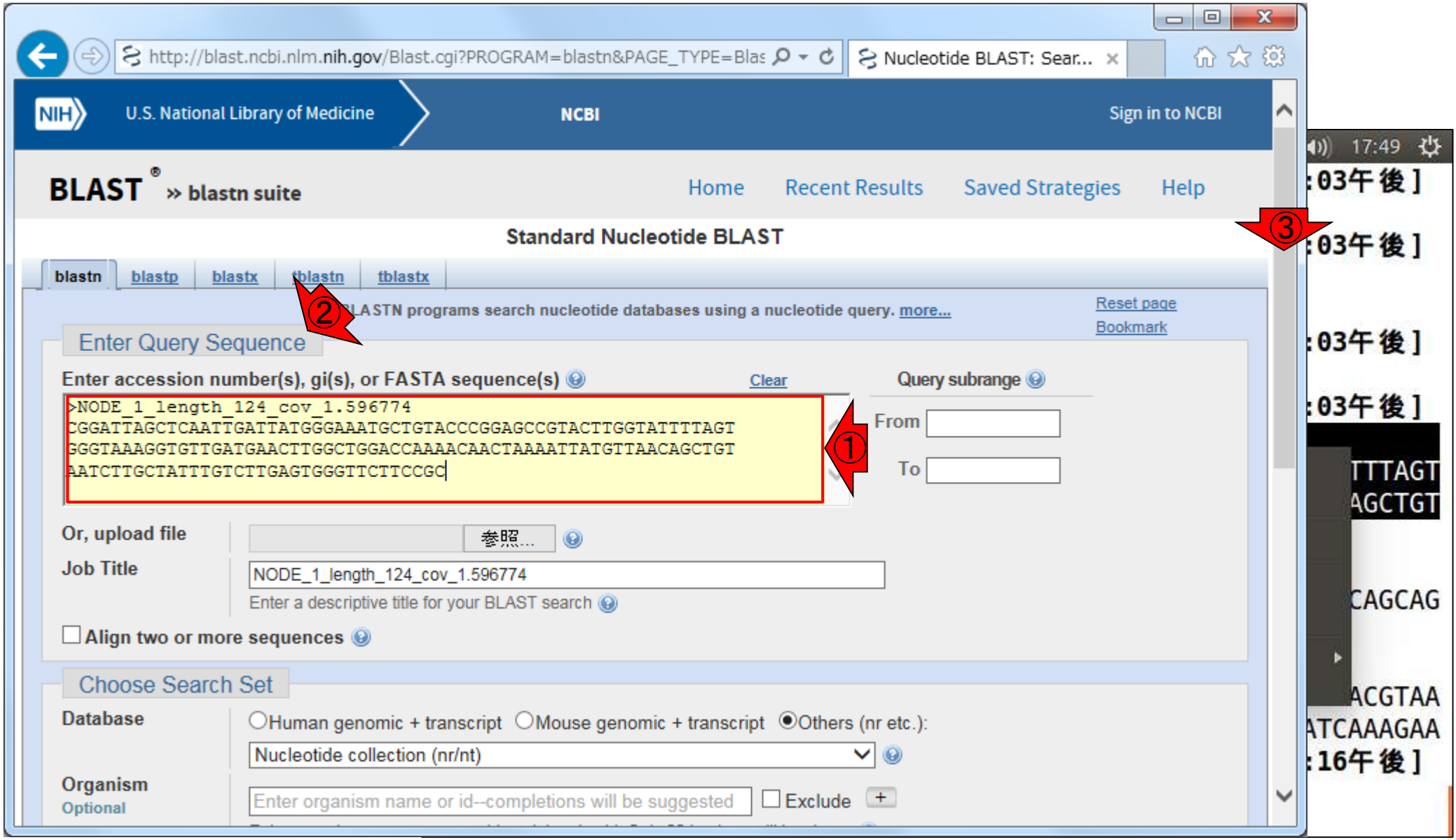
BLAST検索

アセンブリ実行結果の、①最初のコンティグ（反転部分）をコピー



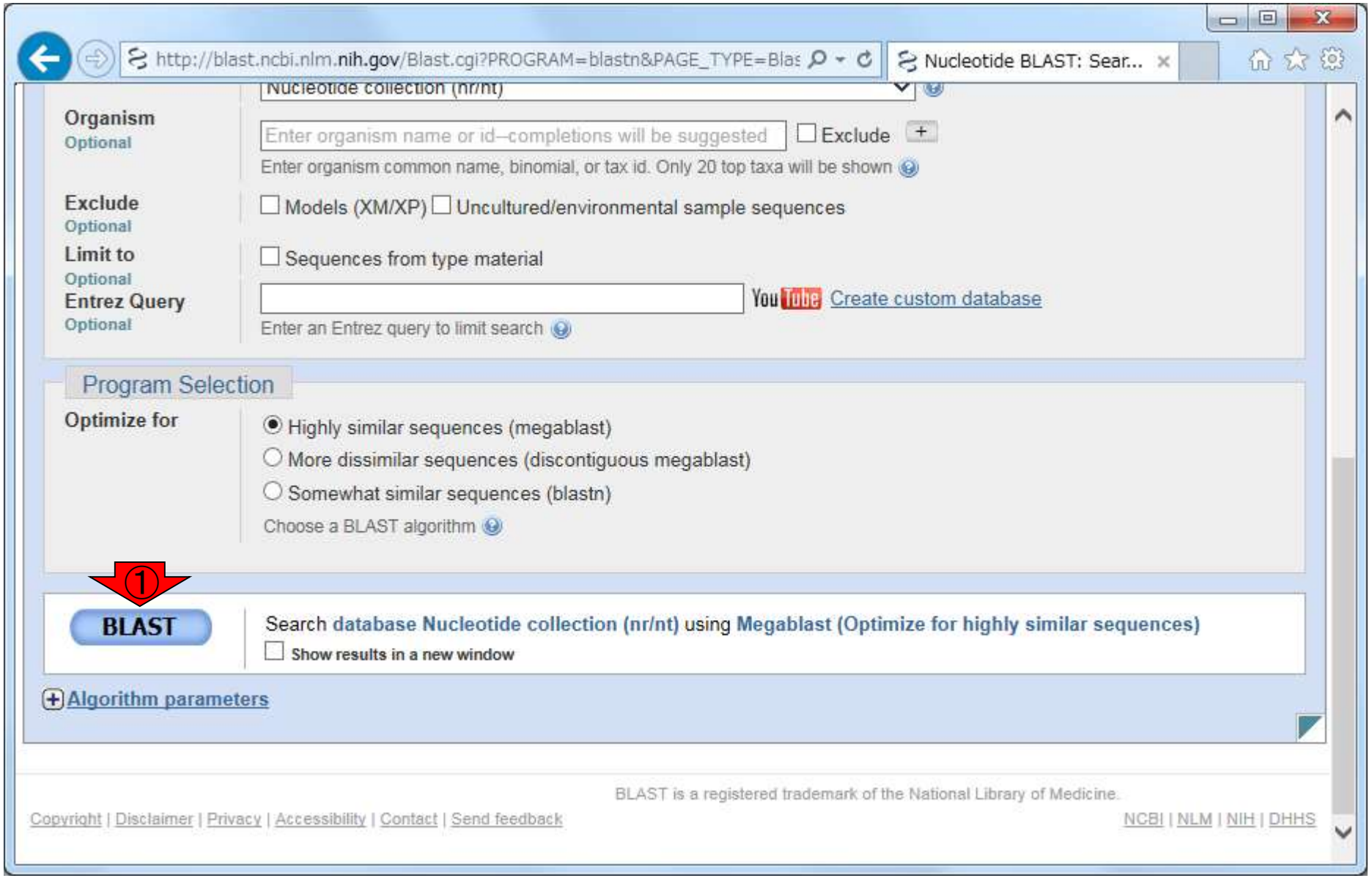
BLAST検索

①赤枠内でペースト。これが②問い合わせしたい塩基配列（Query Sequecne）です。③ページ下部にスクロール



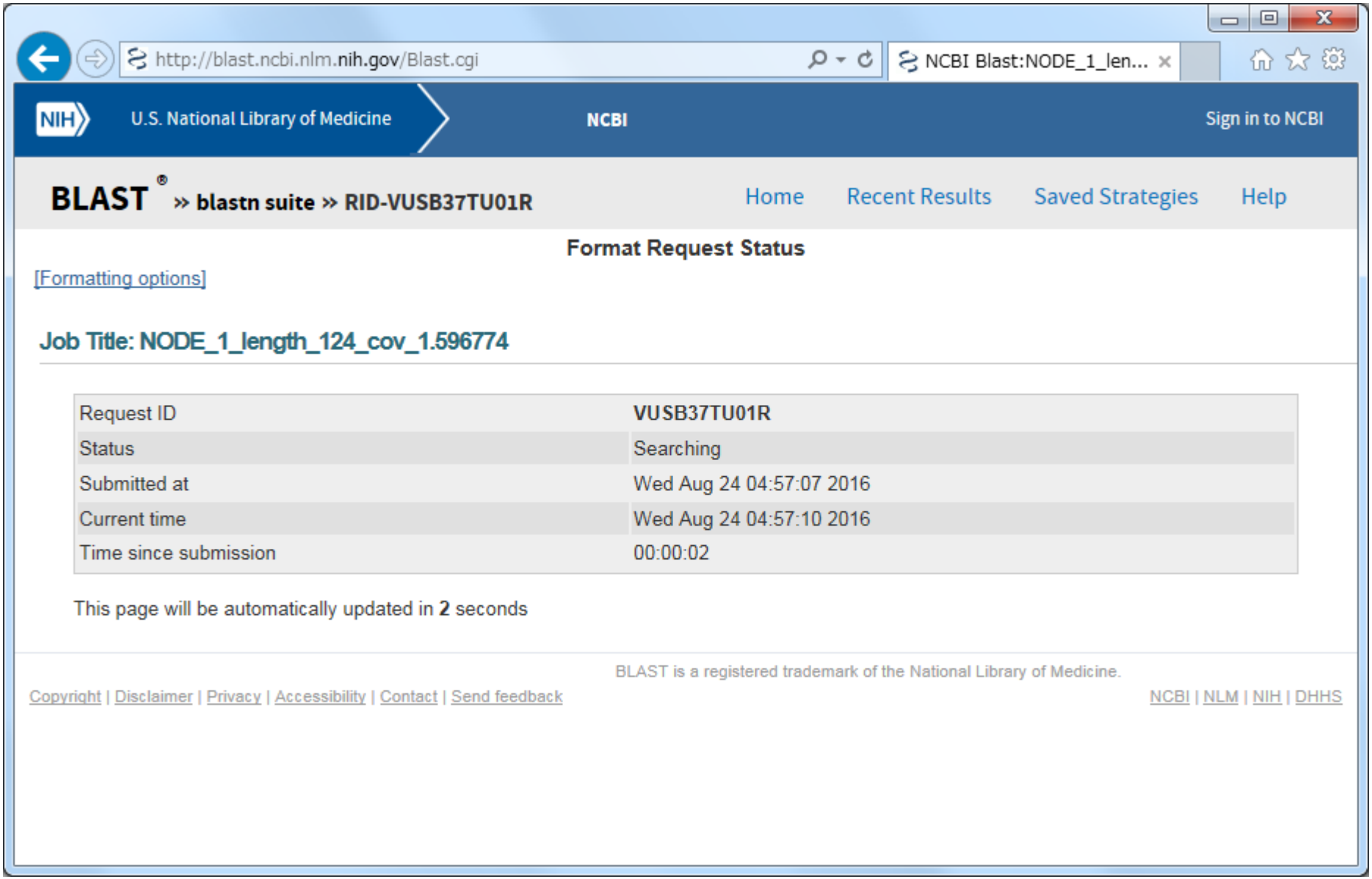
①BLASTボタンを押して実行

BLAST検索



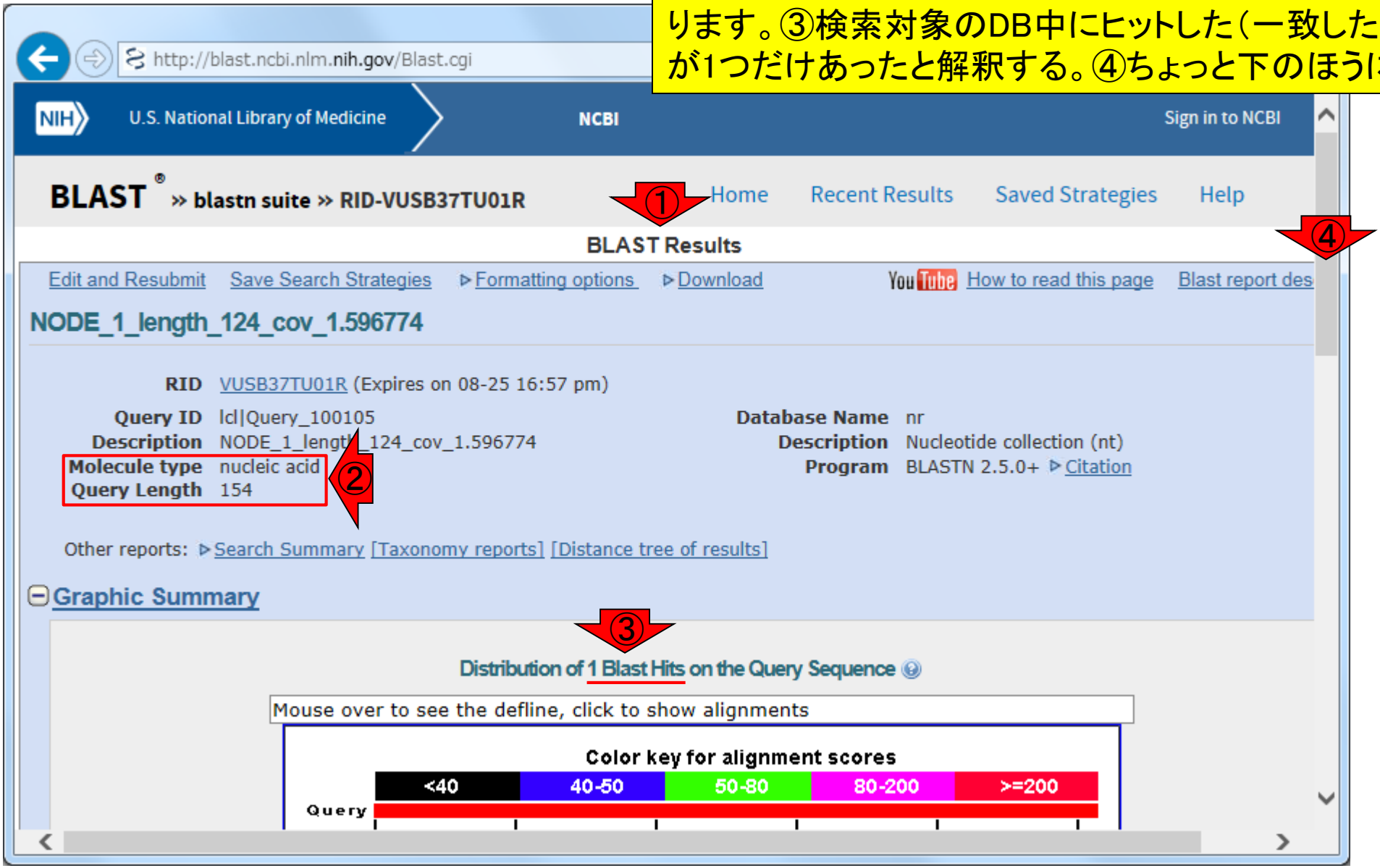
こんな画面に切り替わります

BLAST検索



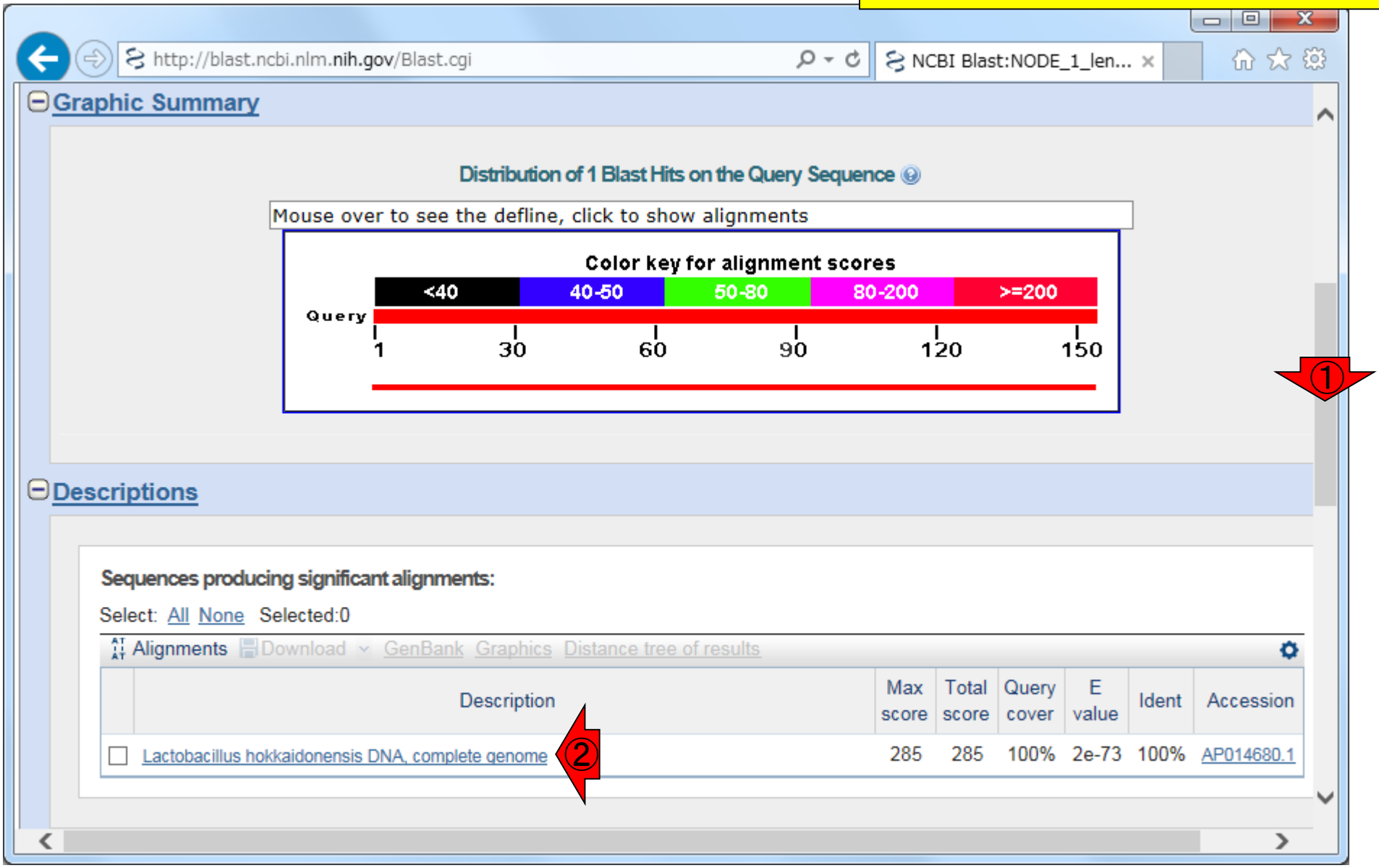
BLAST検索

サーバの混み具合にも依存しますが、概ね1分以内にこのような①BLAST実行結果が得られます。②問い合わせ配列は塩基配列で、長さは154塩基だったことが分かります。③検索対象のDB中にヒットした（一致した）ものが1つだけあったと解釈する。④ちょっと下のほうに移動



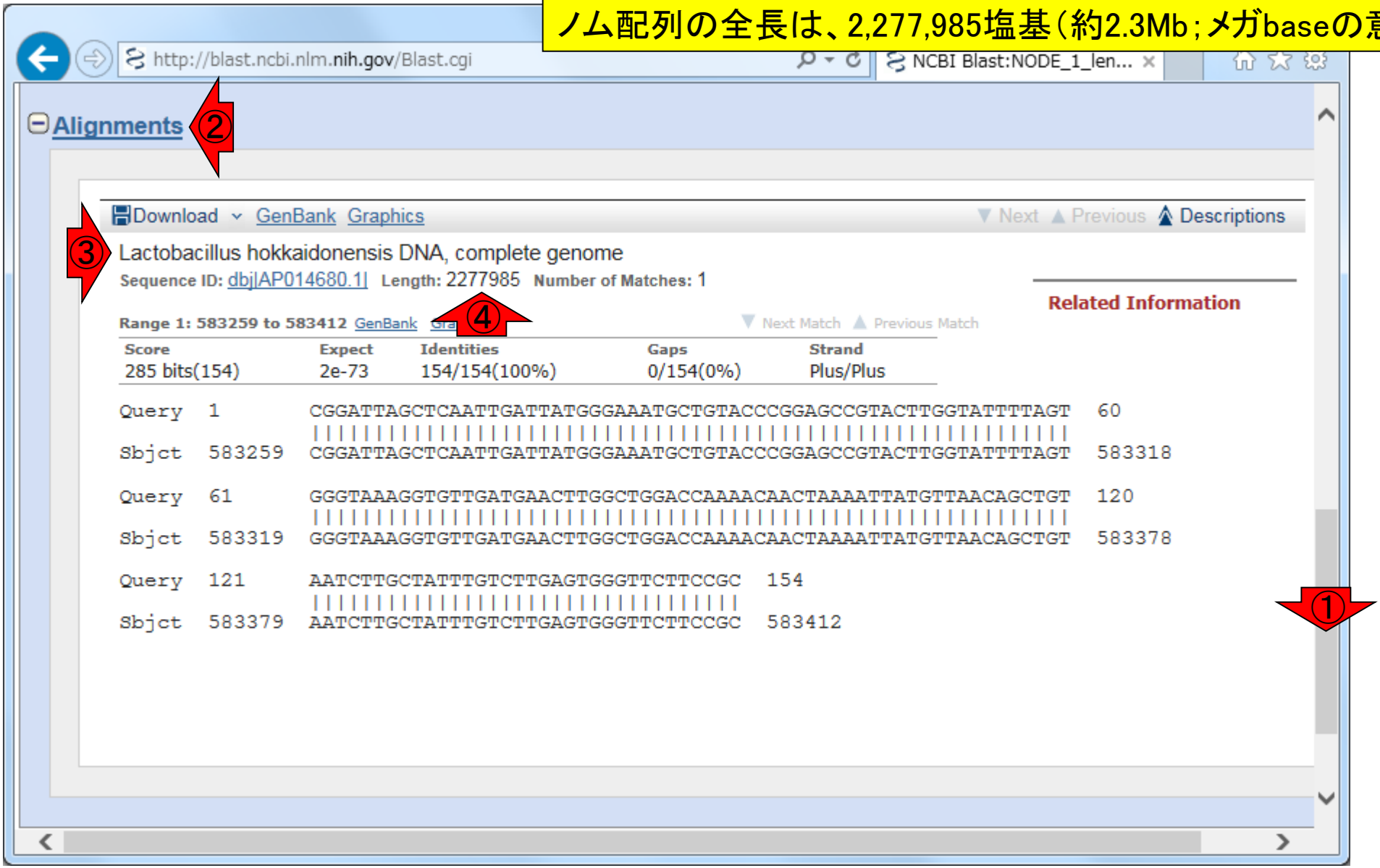
BLAST検索

①このあたり。検索対象のDB中でヒットしたものは、②*Lactobacillus hokkaidonensis*（ある乳酸菌株）の完全なゲノム配列



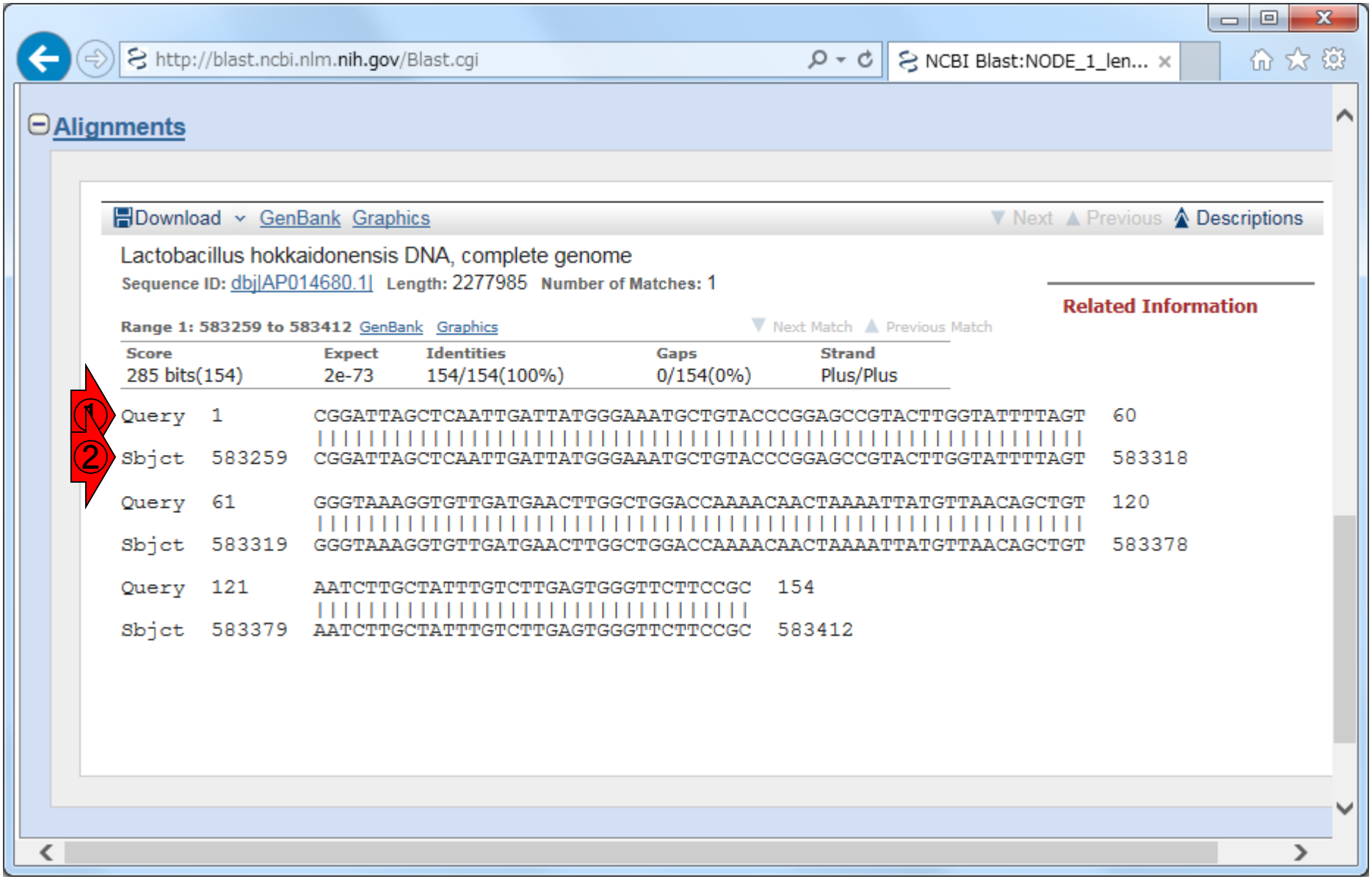
BLAST検索

①さらにページ下部に移動。②Alignmentsというところ。154塩基の問い合わせ配列(Query)が、③乳酸菌ゲノム配列のどのあたりにヒットしたのかを並べて(alignして)示した結果。④乳酸菌ゲノム配列の全長は、2,277,985塩基(約2.3Mb;メガbaseの意味)



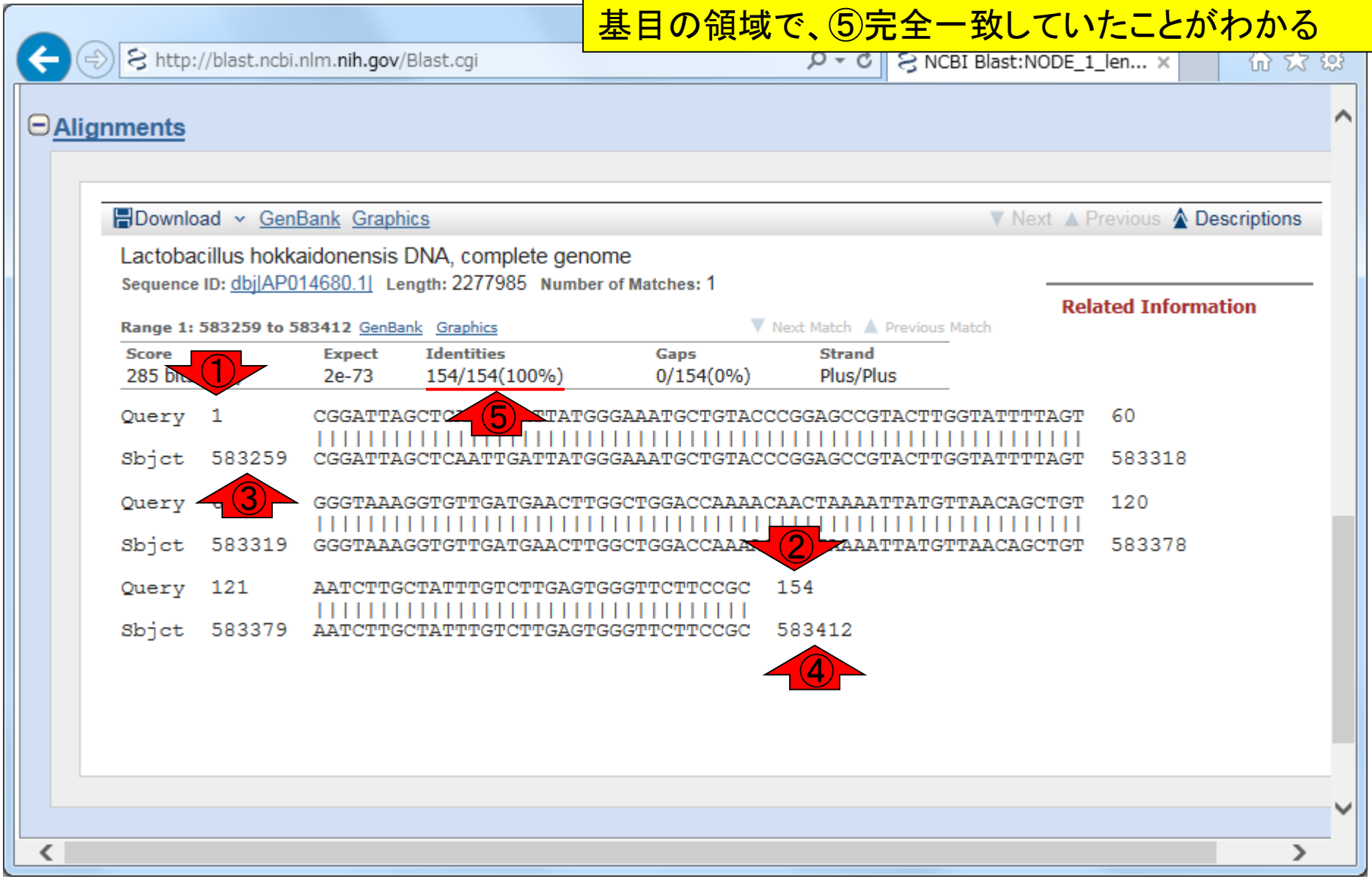
BLAST検索

①154塩基の問い合わせ配列(Query sequence)が上、②ヒットした乳酸菌ゲノム配列（Subject sequenceの略でSbjct）が下



BLAST検索

154塩基の問い合わせ配列(Query sequence)の①1塩基目から②154塩基目が、②ヒットした乳酸菌ゲノム配列(Subject sequenceの略でSbjct)の③583,259塩基目から④583,412塩基目の領域で、⑤完全一致していたことがわかる



Contents

- イントロダクション
  - 概要、背景（NGS用カリキュラム、講習会）、Linuxスキル習得の意義
  - ウェブ情報（日本乳酸菌学会誌のNGS連載やNGS講習会資料）
- 実習環境に慣れる
  - 仮想環境での作業に慣れる
  - GUIとCUI（マウス操作かコマンド入力操作か）
  - ターミナルでの作業
  - 共有フォルダの概念を理解
- 練習
  - 作業ディレクトリの変更、練習用NGSデータファイルのダウンロード
  - ファイルの確認、de novoゲノムアセンブリ
  - BLAST検索
- 課題
  - グループごとに異なる課題ファイルを入力として、「ダウンロード、de novoアセンブリ、BLAST検索」を実行し、得られた結果をレポートにまとめて発表せよ
  - グループ1はkadai1.fasta、グループ2はkadai2.fasta、etc.

課題用ファイルは、このあたりにあります

課題

- **門田幸二**,「講義資料」, 東京大学農学部展開科目: バイオインフォマティクス, 東京大学(東京), 2016.09.12-16
  内容: バイオインフォマティクスやNGSをキーワード程度は知っているヒト(学部3年生程度)向けのアクティブ・ラーニング系講義。バイオインフォマティクスとLinuxスキル習得の意義。バイオインフォマティクス分野で需要が多いのはNGSデータの解析 (JST-NBDCによるアンケート結果)。バイオインフォマティクス人材育成のための講習会の一環として行われているNGS講習会資料を有効活用するために必要な基本スキルの習得。具体的には、日本乳酸菌学会誌のNGS連載内容を理解するための基礎として、Bio-Linuxのノリに慣れるのが最低限の到達目標。何かやったという達成感を味わってもらうため、NGSデータのde novoゲノムアセンブリを行い、元のリード長(塩基配列の長さ)よりも伸びたことを実感してもらう。また、アセンブリ結果の一部をBLAST検索し、入力データ(hoge.fasta; 約6.4MB)がどんな生物種であったかを調べるあたりまで。課題はkadai1.fasta, kadai2.fasta, kadai3.fasta, kadai4.fastaのいずれかを選択して実行。下記と同様の手順で「データのダウンロード → de novoアセンブリ → BLAST検索」し、得られた結果を表。約2日分。

①

```
#作業ディレクトリの変更
cd /home/iu/Desktop/mac_share

#解析したいファイル(hoge.fasta)のダウンロード
wget -c http://www.iu.a.u-tokyo.ac.jp/%7Ekadota/20160912/hoge.fasta

#行数やファイルサイズを表示(行数が200,000行、サイズが6,688,895 bytesとなっていればOK)
wc hoge.fasta

#最初の10行分を表示(2行分で1つのリードを表し、1リードが50塩基であることを確認)
head hoge.fasta

#de novoアセンブリを実行し、結果をugeディレクトリに保存
velveth uge 31 -short -fasta hoge.fasta
velvetg uge

#ugeディレクトリに移動して、アセンブリ結果ファイル(contigs.fa)があることを確認
cd uge
ls

#行数やファイルサイズを表示(入力のhoge.fastaは200,000行だったが、出力のcontigs.faは4,038行となっている)
wc contigs.fa
```

実習用PC環境を自力で

実習用PCは、既にVirtualBoxをインストールし、Bio-Linuxというものを導入（インポート）し、共有フォルダを設定した後の状態です。この環境を自力で構築したいヒトは、①第6回ゲノムアセンブリ、②のインストール手順を参考にしてください。

(Rで)塩基配列解析

(last modified 2016/08/12, since 2011)

このウェブページはインストール||についての推奨手順 (Windows2015.04.04版とMacintosh2015.0
に従って フリーソフトRと必要
の 方は 基本的な利用法(Win
ページを体系的にまとめた書

What's new?

- NGSハンズオン講習会201
を通すところのミス)、スラ
実行もうまくいく)あたりです
- NGSハンズオン講習会201
- NGSハンズオン講習会201
- Twitterハッシュタグ AJAC
- NGSハンズオン講習会201

- 書籍 | トランスクリプトーム解析 | 4.3.2 データの 正規化(応用編) (last modified 2014/04/27)
- 書籍 | トランスクリプトーム解析 | 4.3.3 2群間比較 (last modified 2014/04/28)
- 書籍 | トランスクリプトーム解析 | 4.3.4 他の 実験デザイン(3群間) (last modified 2014/04/28)
- 書籍 | 日本乳酸菌学会誌 | について (last modified 2016/07/01)
- 書籍 | 日本乳酸菌学会誌 | 第1回イントロダクション (last modified 2015/09/11)
- 書籍 | 日本乳酸菌学会誌 | 第2回GUI環境からコマンドライン環境へ (last modified 2015/11/26)
- 書籍 | 日本乳酸菌学会誌 | 第3回Linux環境構築からNGSデータ取得まで (last modified 2015/12/07)
- 書籍 | 日本乳酸菌学会誌 | 第4回クオリティコントロールとプログラムのインストール (last modified 20
- 書籍 | 日本乳酸菌学会誌 | 第5回アセンブル、マッピング、そしてQC (last modified 2016/07/05)
- 書籍 | 日本乳酸菌学会誌 | 第6回ゲノムアセンブリ (①) modified 2016/06/17)
- 書籍 | 日本乳酸菌学会誌 | 第7回ロングリードアセンブリ (last modified 2016/07/01)
- イントロ | 一般 | ランダムに行
- イントロ | 一般 | 任意の文字列
- イントロ | 一般 | 任意のキーワ
- イントロ | 一般 | ランダムな塩

書籍 | 日本乳酸菌学会誌 | 第6回ゲノムアセンブリ

日本乳酸菌学会誌の第6回分です。Linuxコマンドのリンク先は主に日経BP社様です。原稿PDFの「はじめに」項目のところにtypoがあります。誤:するであれば、正:する"の""であれば、ですねm(_ _)m。また、「配列長によるフィルタリング」項目の最後の文章「これらについては、第7回で詳述する予定である。」についてですが、これは「これらについては、"第8回以降"で詳述する予定である。」と読み替えてください。第7回ドラフト原稿作成時点で、これらの内容は含まれていないためです（2016年4月23日追加）。

- 原稿PDF
- ウェブ資料PDF
  - Windows用(2016.06.17版; 約20MB)
  - Macintosh用
- (共有フォルダ設定情報を含む)連載第3回終了時点以降のovaファイルからのインストール手順:
  - Windows用(2015.12.28版; 約2MB)
  - Macintosh用(2015.12.28版; 約2MB)

②

実習用PC環境を自力で

実習用PCと完全に同じ環境にしたいヒトは、①のスライド5のところで私宛にメールする際に、件名を「乳酸菌連載第4回終了時点のovaファイル希望」としてください

書籍 | 日本乳酸菌学会誌 | 第6回ゲノムアセンブリ

日本乳酸菌学会誌の第6回分です。Linuxコマンドのリンク先は主に日経BP社様です。原稿PDFの「はじめに」項目のところにtypoがあります。誤:するであれば、正:する"の"であれば、ですねm(__)m。また、「配列長によるフィルタリング」項目の最後の文章「これらについては、第7回で詳述する予定である。」についてですが、これは「これらについては、"第8回以降"で詳述する予定である。」と読み替えてください。第7回ドラフト原稿作成時点で、これらの内容は含まれていないためです(2016年4月23日追加)。

- 原稿PDF
- ウェブ資料PDF
  - Windows用(2016.06.17版; 約20MB)
  - Macintosh用

- (共有フォルダ設定情報を含む)連載第3回終了時点以降のovaファイルからのインストール手順:
  - Windows用(2015.12.28版; 約2MB)
  - Macintosh用(2015.12.28版; 約2MB)

①