

2015.12.07版

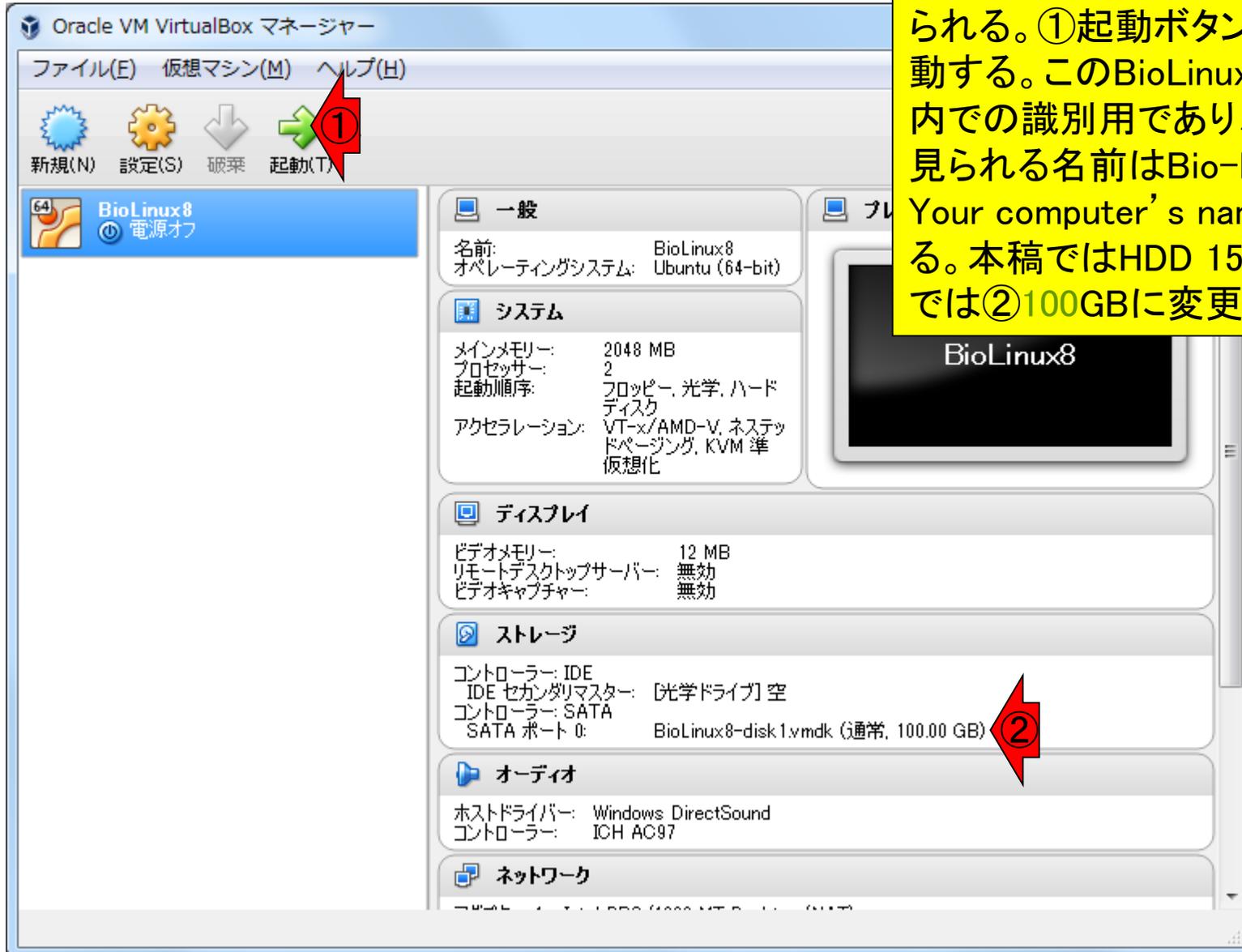
本稿ではHDD 150GBとなっているが、それだけのHDD容量を確保できないヒト用に、100GBに変更しています。また、*L. casei* 12A 株のRNA-seqデータ(SRR616268)が解凍前で約15GBだったため、ダウンロード自体が困難なヒトがいたようです。それゆえ、全リード(約1.3億)ではなく、ダウンロードの段階から最初の150万リードのみにする方法に変更しています。難しいオプションが最初の段階で出ますが、それをクリアすれば、その後のハンズオンがサクサク進みます

次世代シーケンサーデータの解析手法 第3回Linux環境構築からNGSデータ取得まで ウェブ資料 (Windows版の軽量版)

東京大学・大学院農学生命科学研究科
孫 建強、三浦 文、清水 謙多郎、門田 幸二
kadota@bi.a.u-tokyo.ac.jp
<http://www.iu.a.u-tokyo.ac.jp/~kadota/>

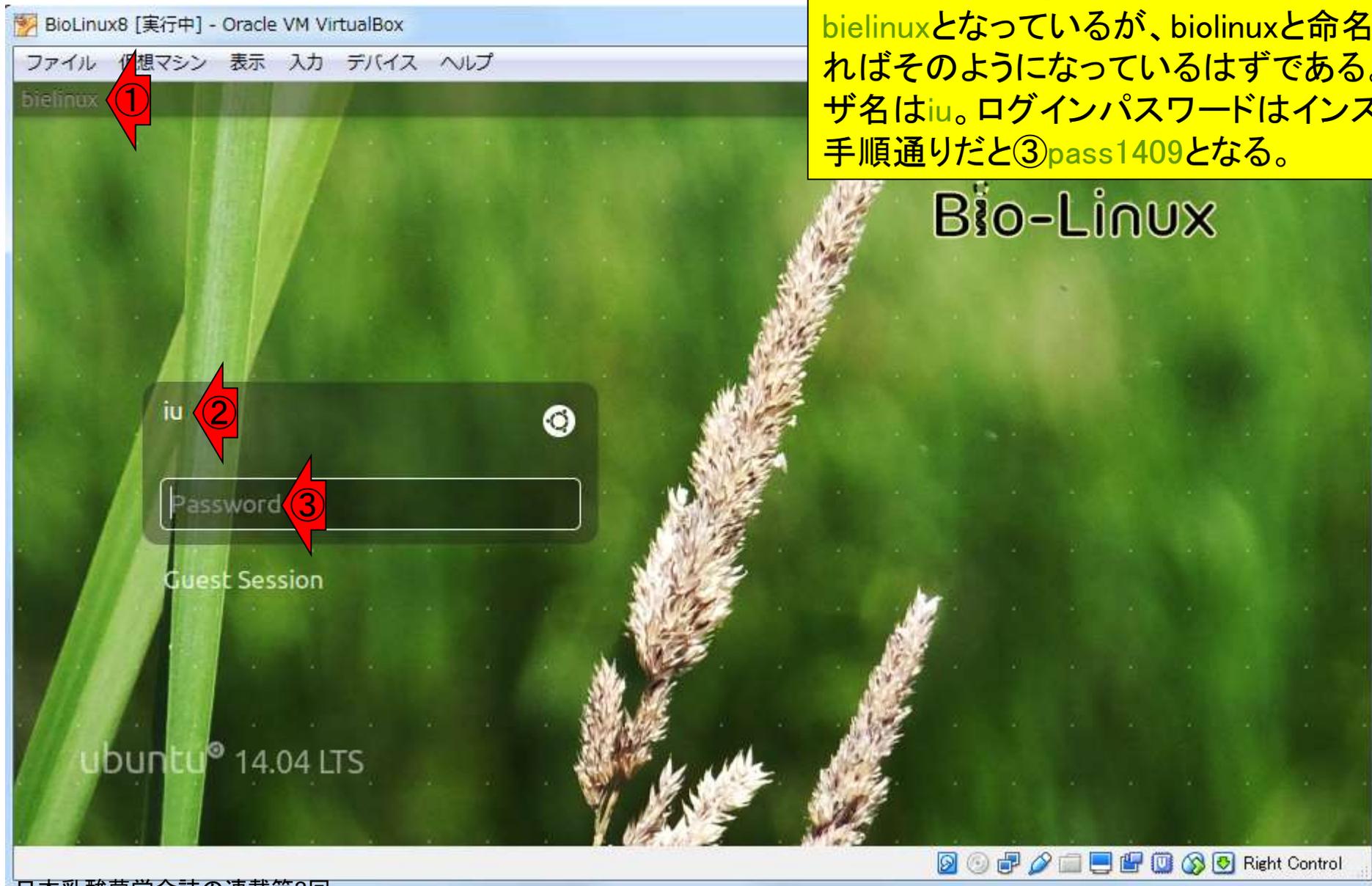
W1

仮想化ソフトVirtualBox起動直後の画面。BioLinux8という名前の仮想PC本体(仮想マシン環境)を導入済みであるため、このように見られる。①起動ボタンを押すと、BioLinux8が起動する。このBioLinux8という名前はVirtualBox内での識別用であり、実際にネットワーク上で見られる名前はBio-Linux 8インストール時にYour computer's nameで指定したbielinuxである。本稿ではHDD 150GBとなっているが、ここでは②100GBに変更しています。



W2-1

BioLinux8という名前の仮想PC本体を起動すると、WindowsやMacintoshで見られるようなログイン画面が現れる。PC名はこの場合① `bielinux`となっているが、`biolinux`と命名していればそのようになっているはずである。②ユーザ名は`iu`。ログインパスワードはインストール手順通りだと③`pass1409`となる。



W2-2

連載第2回で用いたWindows PC (Panasonic Let's note)とMacintosh PC (MacBook) との対応関係としては、概ね以下のように解釈すればよい。

①

iu

②

Password

③

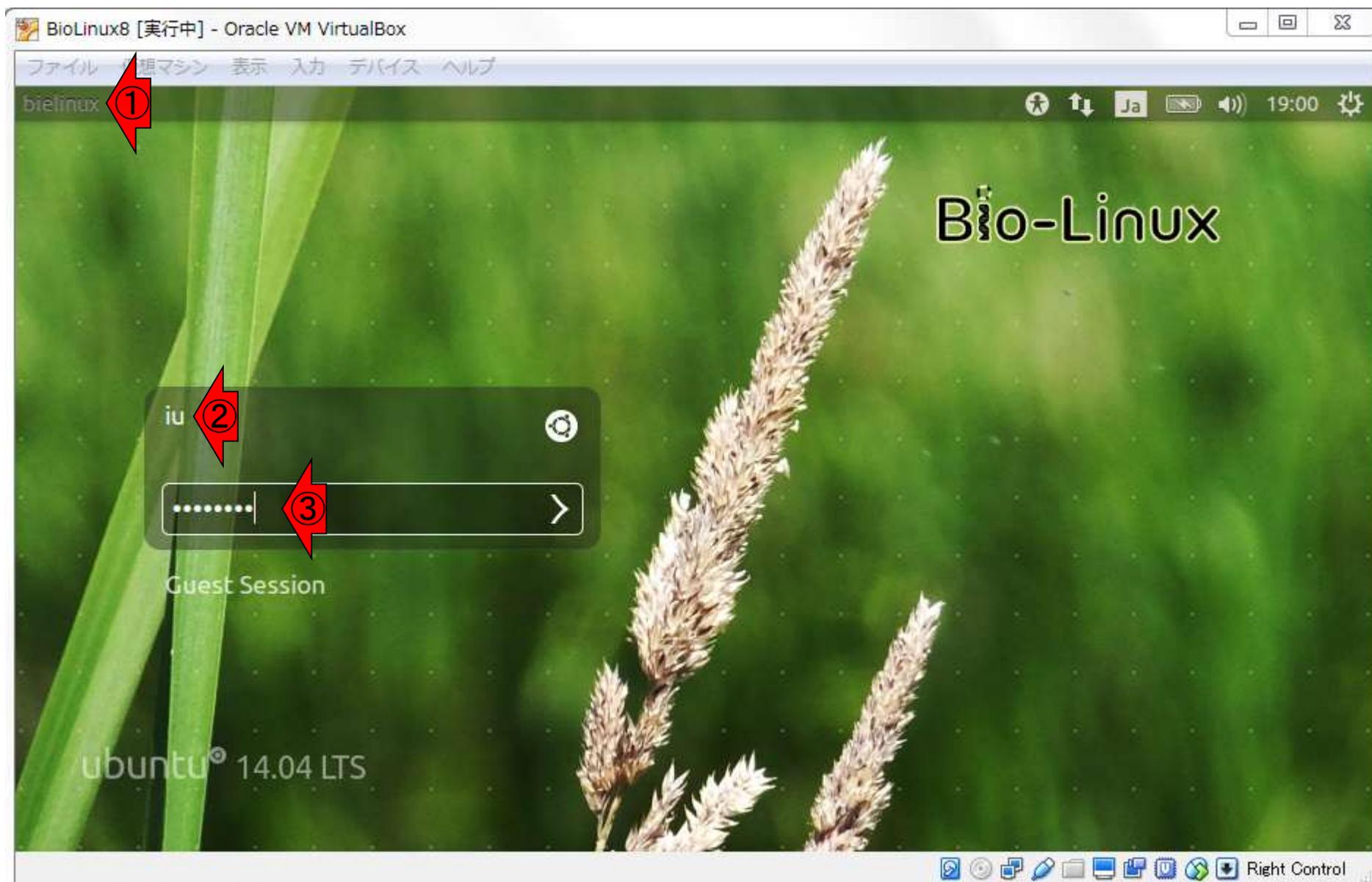
Guest Session

ubuntu® 14.04 LTS

PC本体の名前	BioLinux8	Let's note	MacBook
ホストOS or ゲストOS	ゲストOS	ホストOS	ホストOS
OSの種類	Ubuntu (64 bit)	Windows 7 (64 bit)	Mac OS X (32 bit)
(任意につける)ネットワーク上で識別されるPC名	bielinux or biolinux	kadota-pc	agribio-macbook
ユーザ名	iu	kadota	kadota

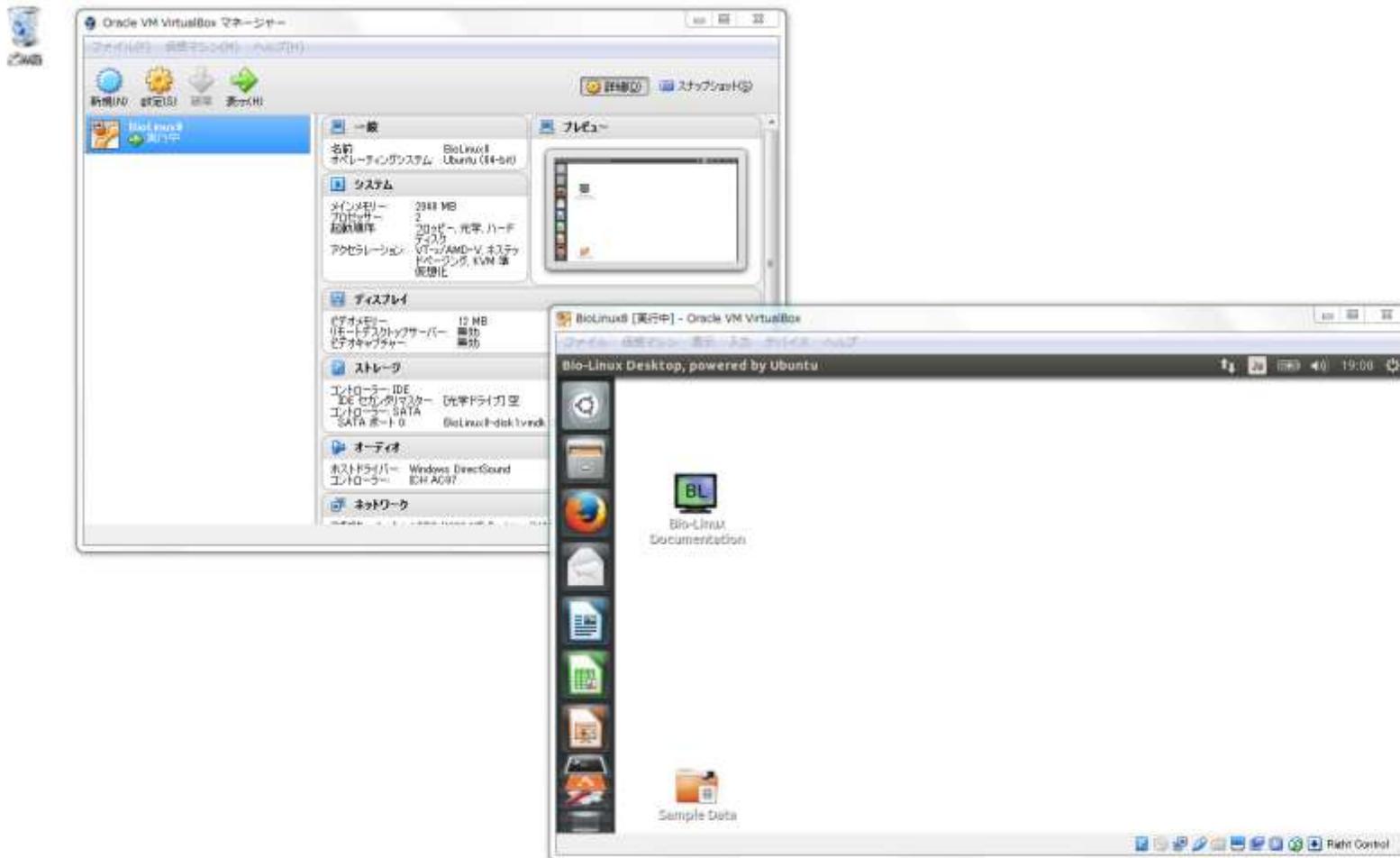
W2-3

ログインパスワード(③pass1409)を打ち込んで、「Enter」キーを押す。



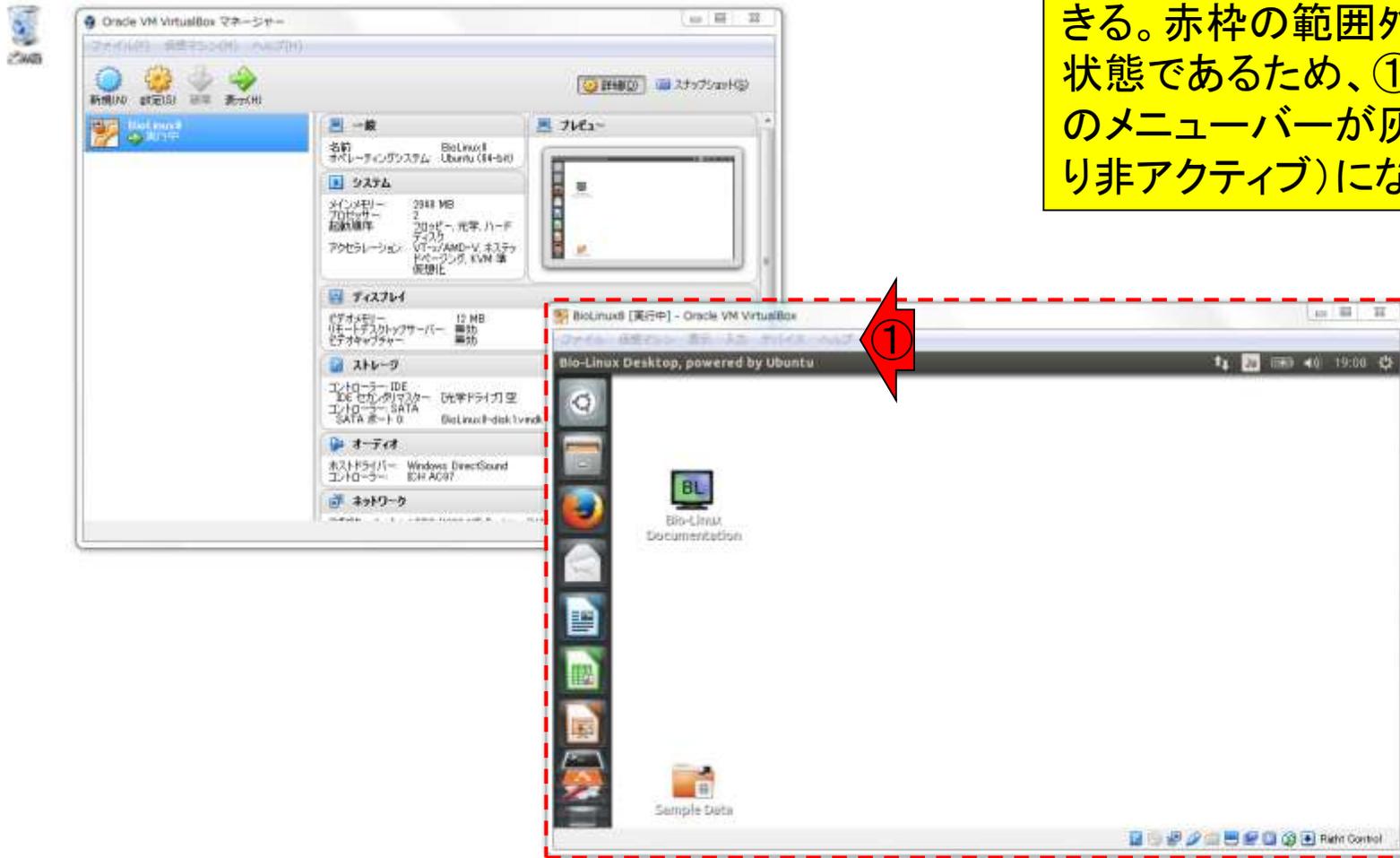
W2-4

ホストOSがWindows環境下での
Bio-Linux 8起動直後の状態。



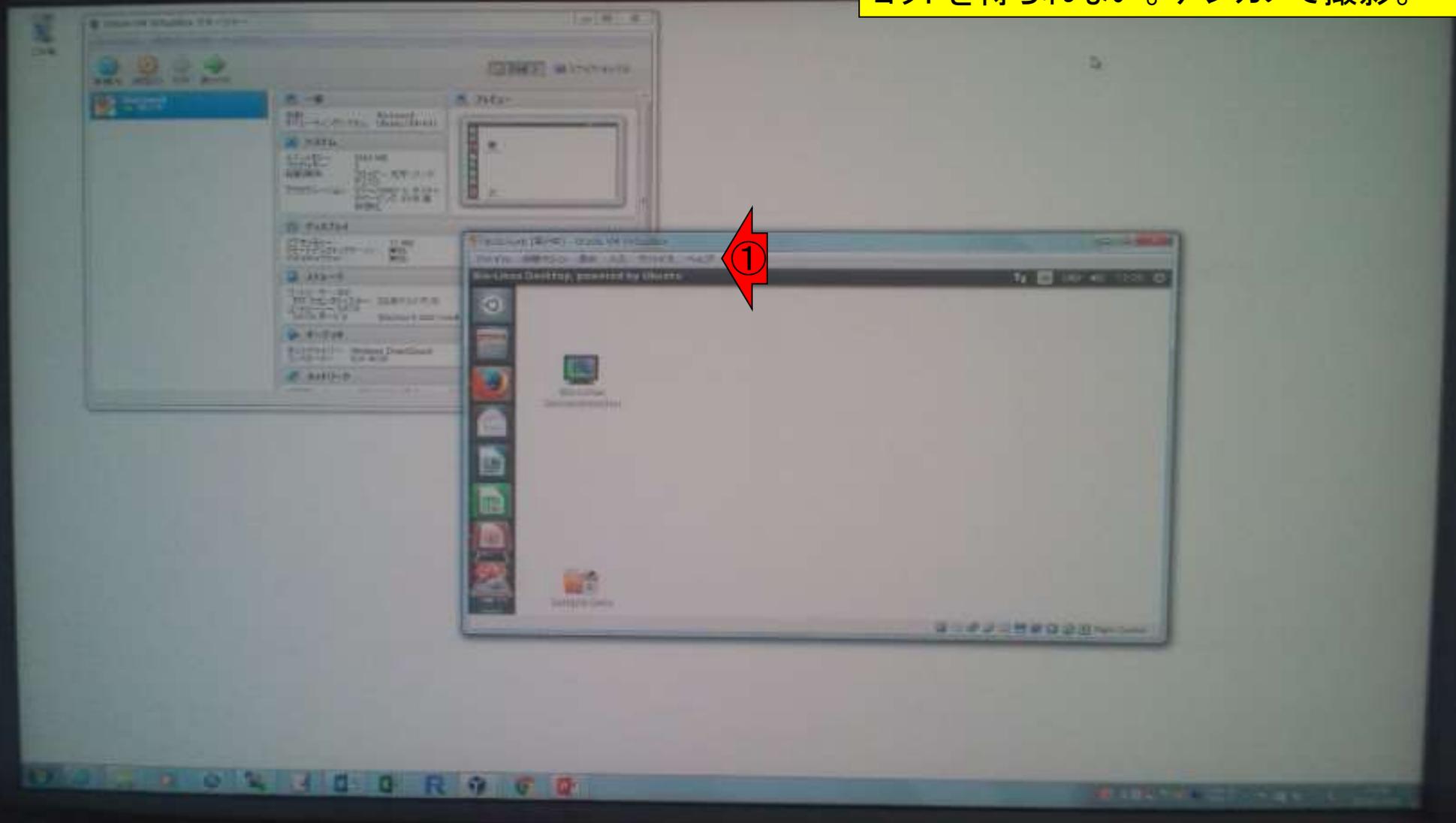
W3-1

点線の赤枠部分に相当するBioLinux8のウィンドウがアクティブでない場合に、W2-4のスクリーンショットを得ることができる。赤枠の範囲外でマウスクリックした状態であるため、①BioLinux8ウィンドウのメニューバーが灰色になっている(つまり非アクティブ)になっていることが分かる



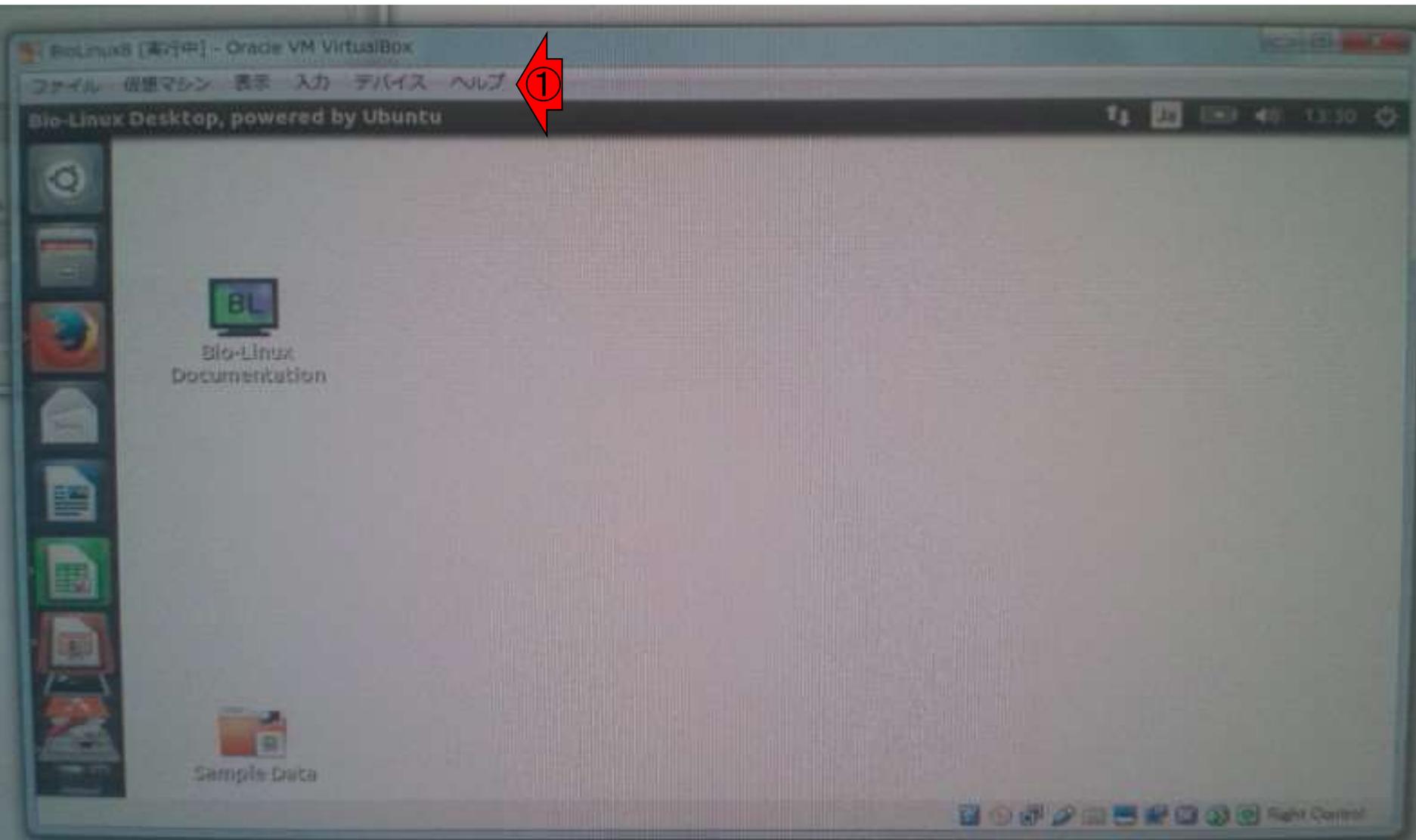
W3-2

BioLinux8のウィンドウがアクティブな状態だと①メニューバーが黒くなる。この状態ではW2-4のようなスクリーンショットを得られない。デジカメで撮影。



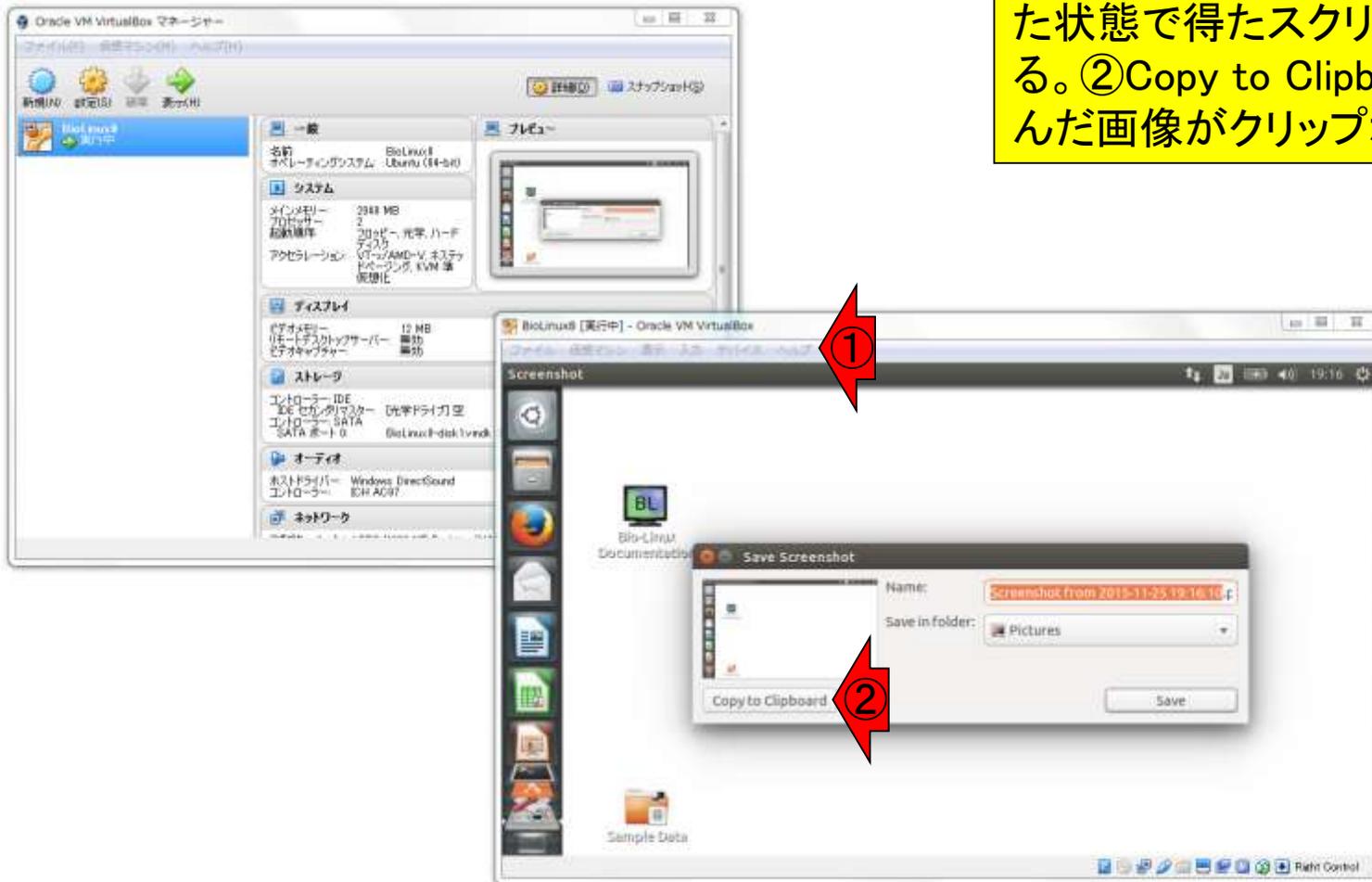
W3-3

BioLinux8ウィンドウがアクティブ状態の拡大図。



W4-1

BioLinux8ウィンドウ(ゲストOS)がアクティブな状態でスクリーンショットを撮った結果。
①BioLinux8ウィンドウを非アクティブにした状態で得たスクリーンショットを示している。
②Copy to Clipboardを押すと、取り込んだ画像がクリップボードにコピーされる



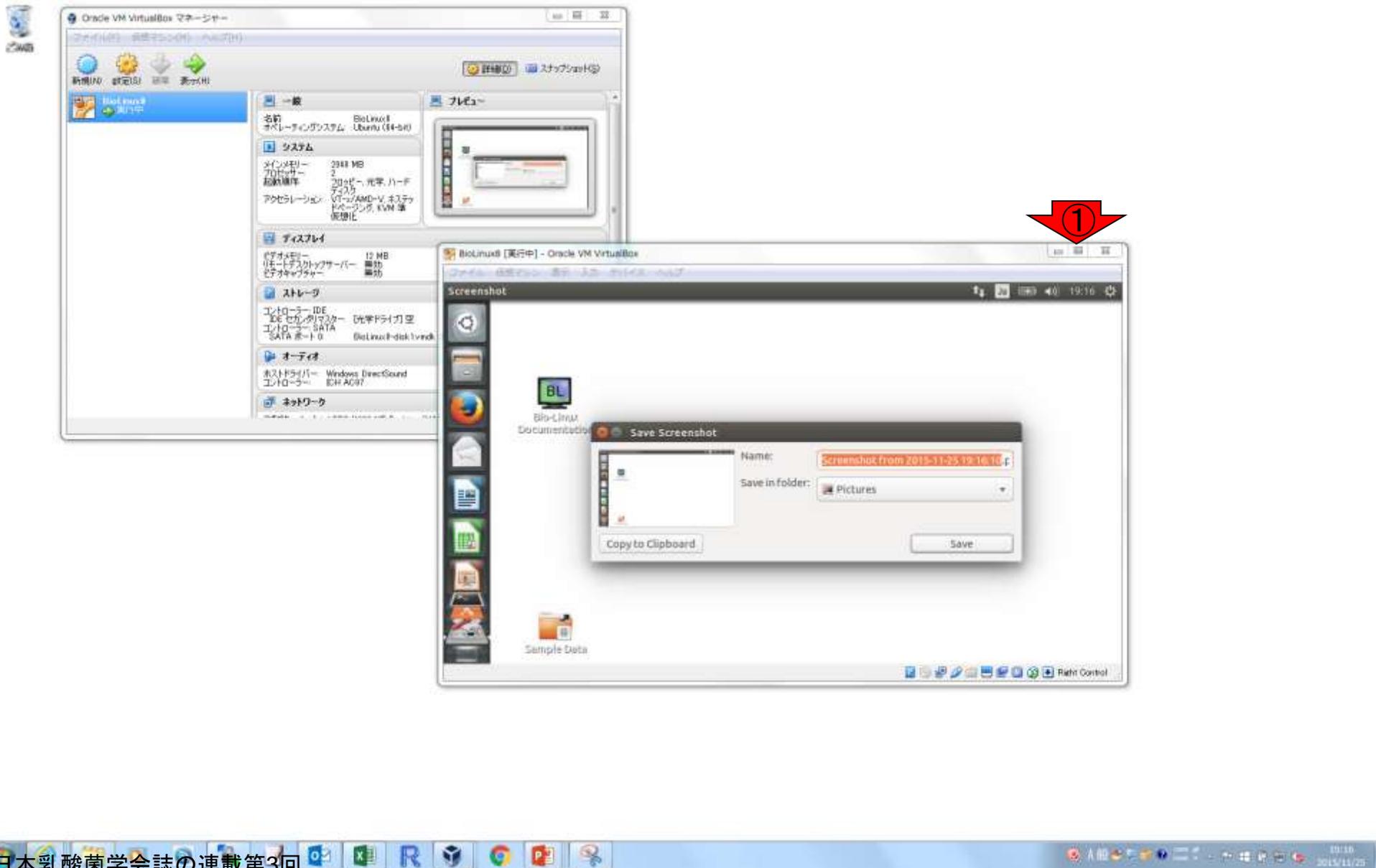
W4-2

クリップボードにコピーされた画像
(取り込んだゲストOSのPC画面)。



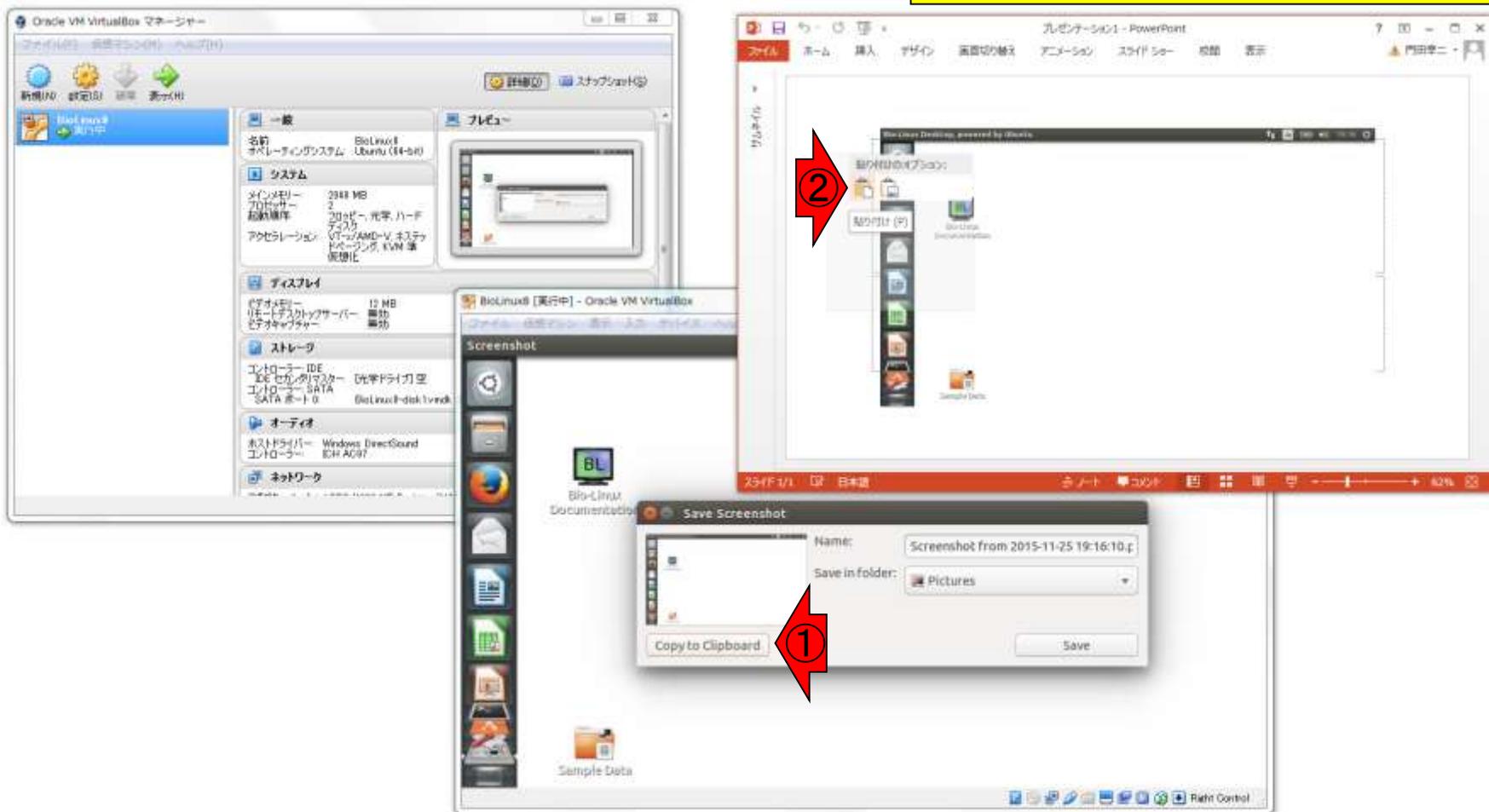
W4-3

①画面サイズを最大にすれば、
ほぼ完全なLinux環境となる



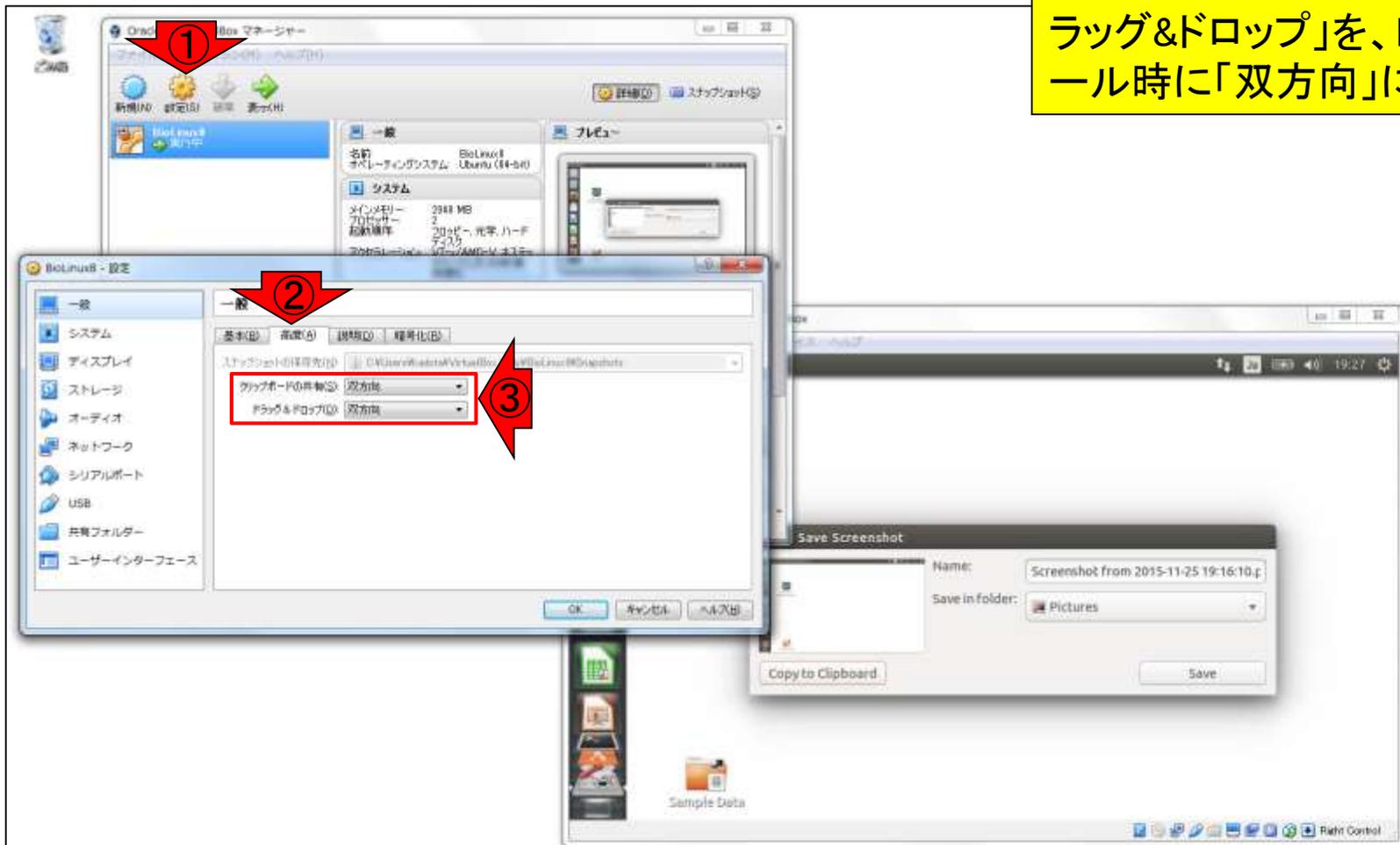
W5-1

ゲストOSのスクリーンショットを①クリップボードにコピーし、②ホストOSのMicrosoft PowerPointに張りつけているところ。



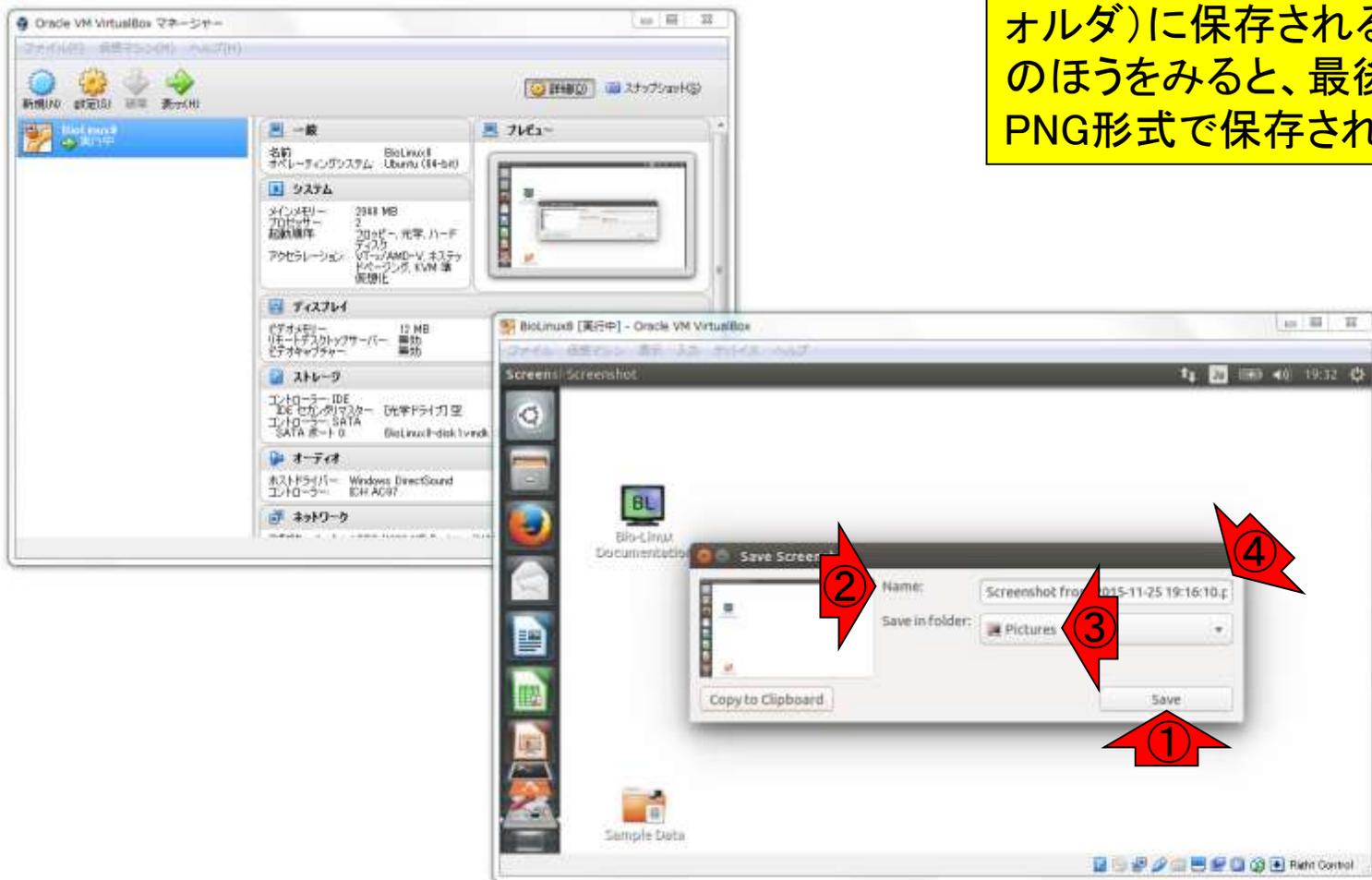
W5-2

ゲストOSのクリップボード情報をホストOSで利用できたのは、③デフォルトでは無効になっていた「クリップボードの共有」と「ドラッグ&ドロップ」を、Bio-Linux 8のインストール時に「双方向」に変更していたから。



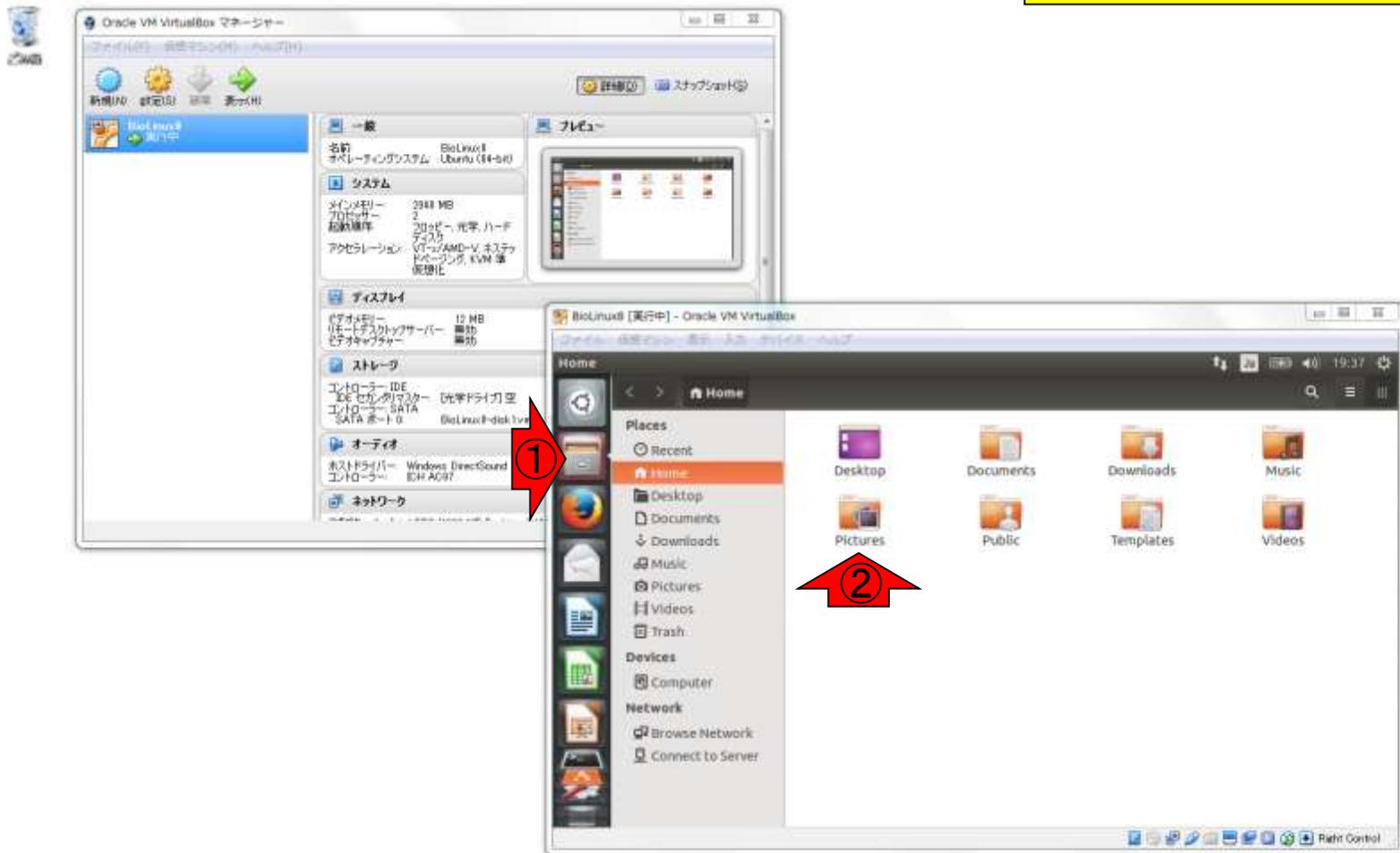
W6-1

①Saveボタンを押すと、②Nameに記載されているファイル名で、Save in folderに記載されているフォルダ(③つまりPicturesフォルダ)に保存される。④ファイル名の右のほうをみると、最後が.pngとなっており、PNG形式で保存されることが分かる。



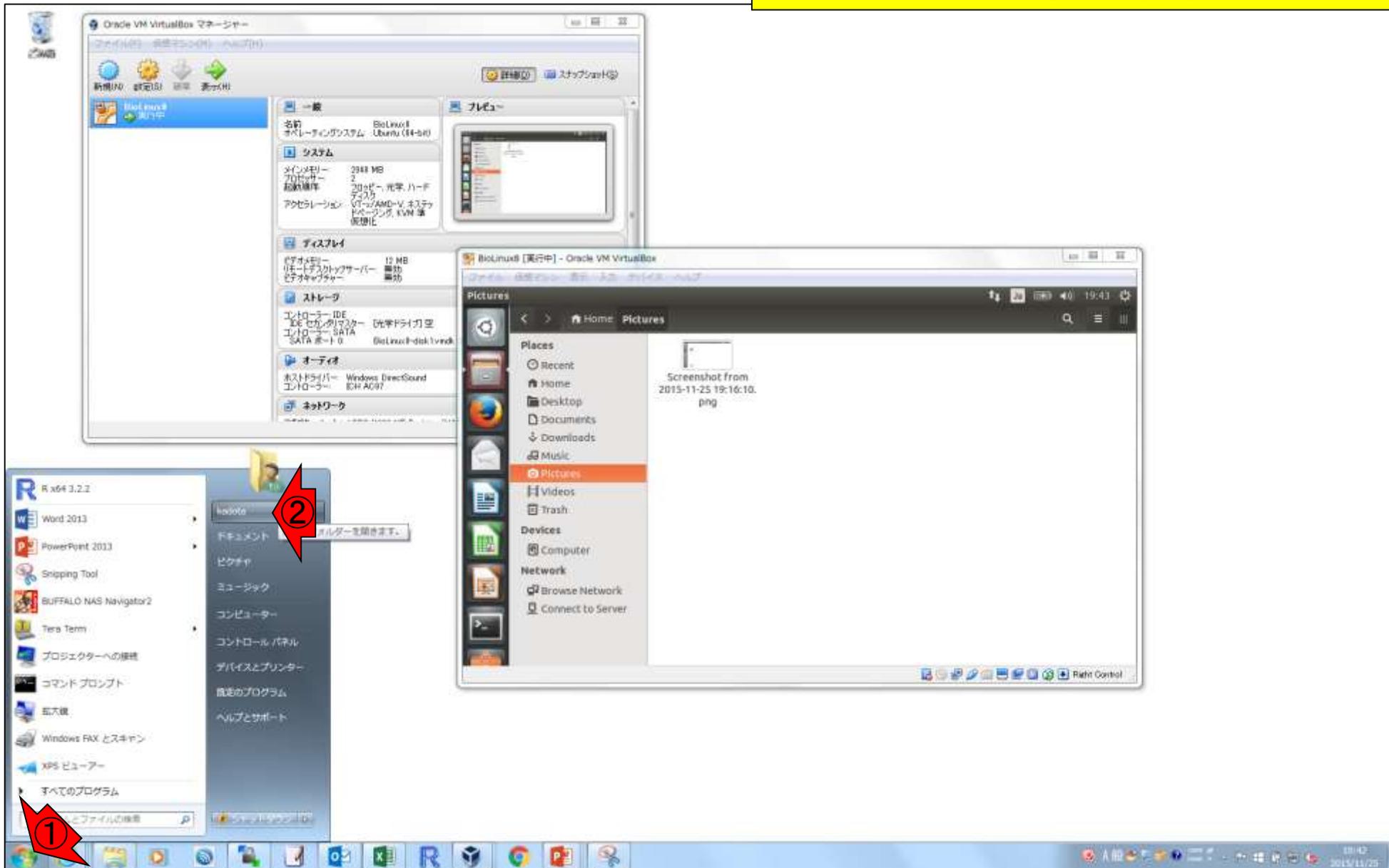
W6-2

①の引き出しみたいなのアイコンをクリックし、②Picturesフォルダをクリックすると、保存したpngファイルを見ることができる。



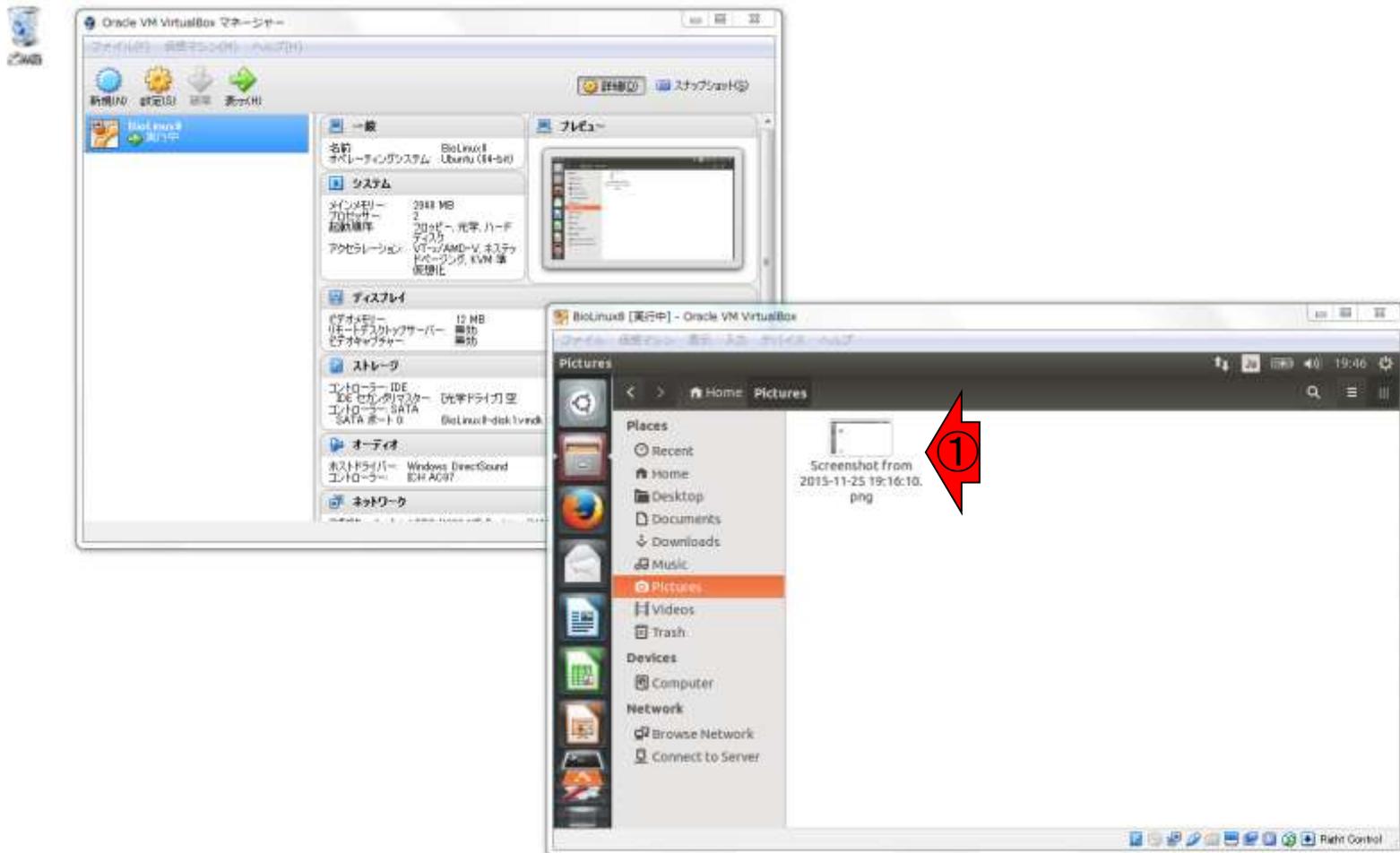
W6-2

ホストOS(Windows 7)の場合、①左下のスタートメニューを押して、②ログインユーザ(この場合 kadota)の個人用フォルダを開く操作に相当する



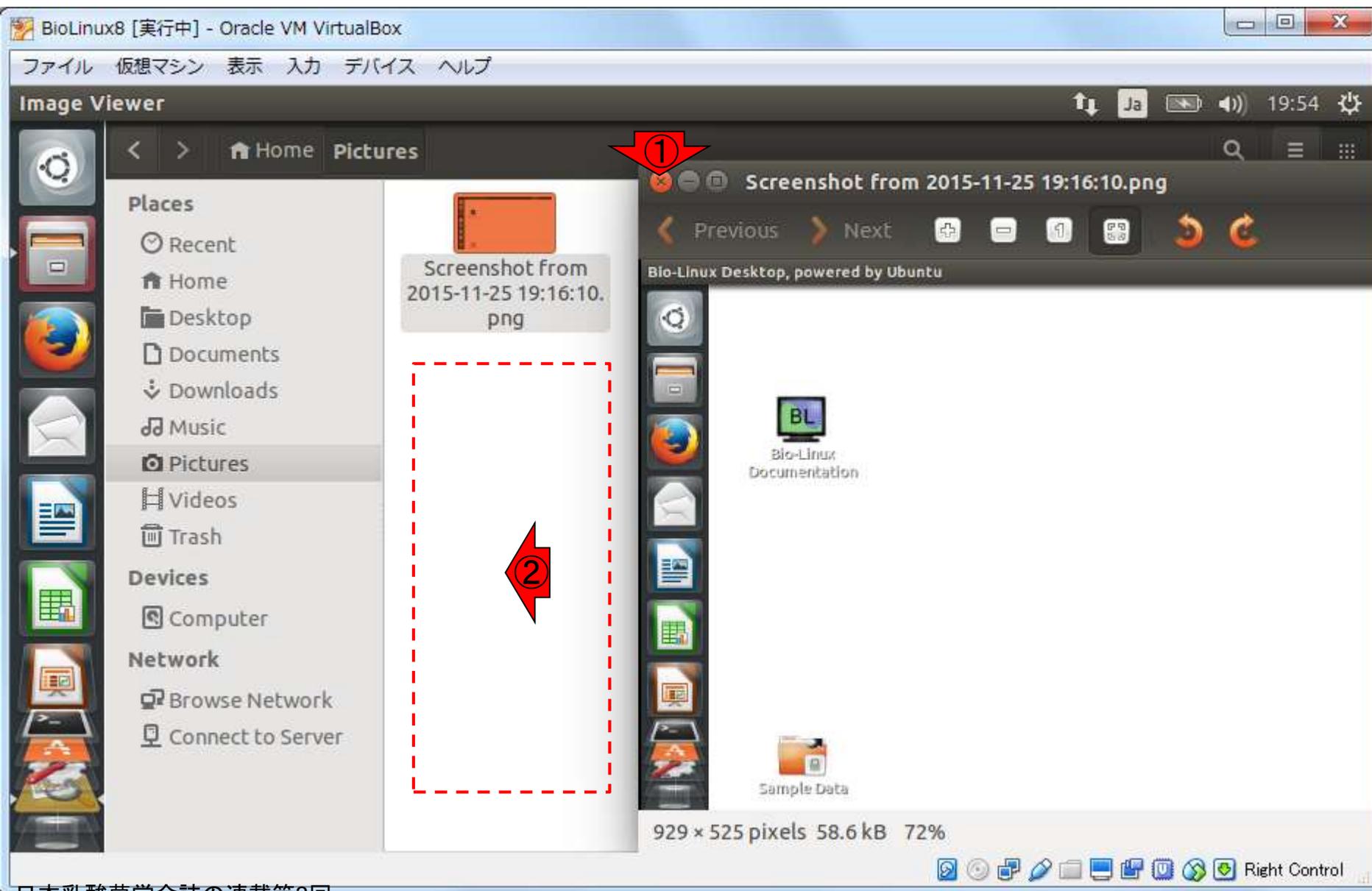
W6-2

①ダブルクリックでpngファイルを開くことができる



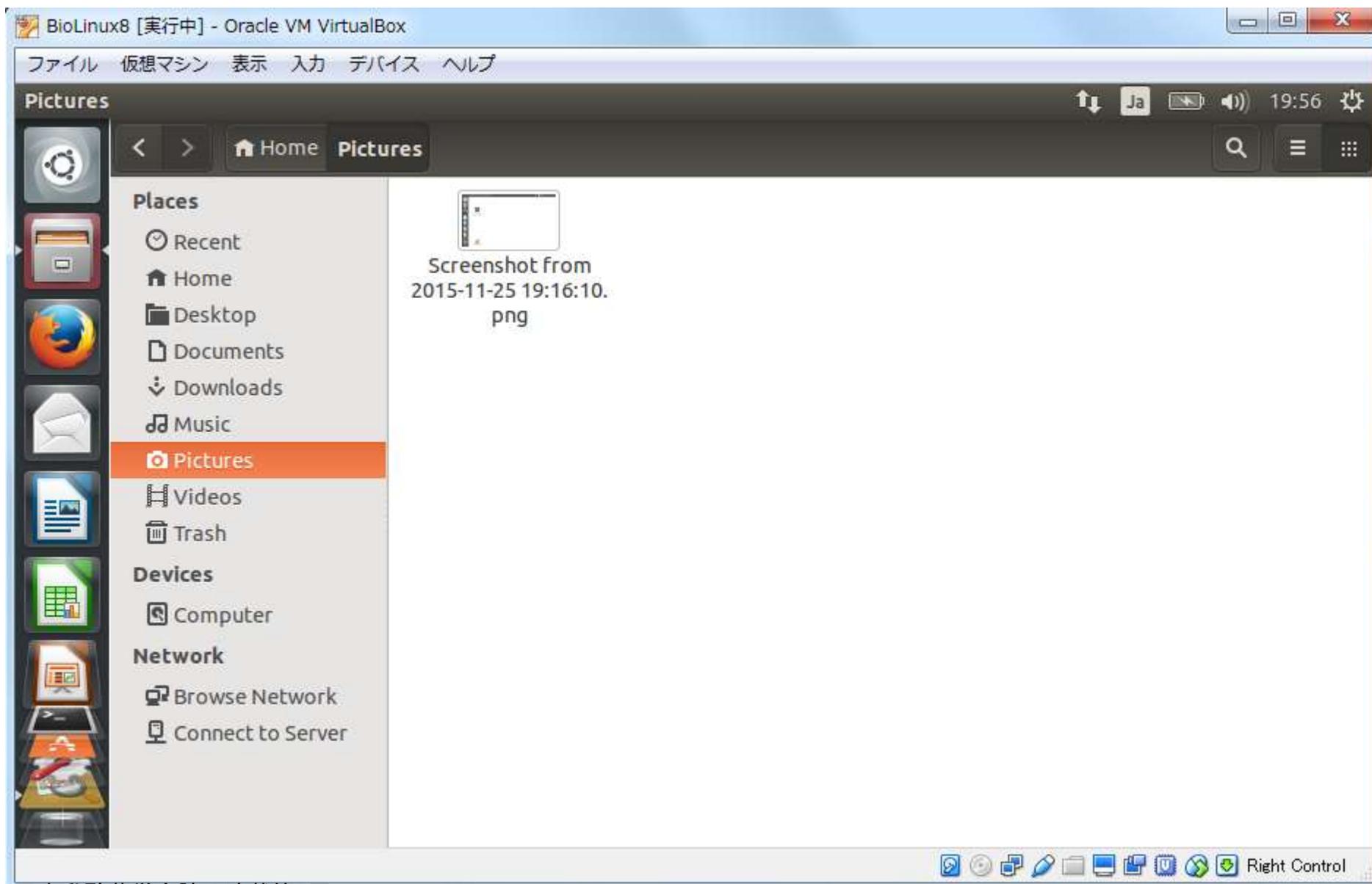
W6-2

①×ボタンを押すと閉じることができるが、
やらずに②赤枠あたりでクリックすると…



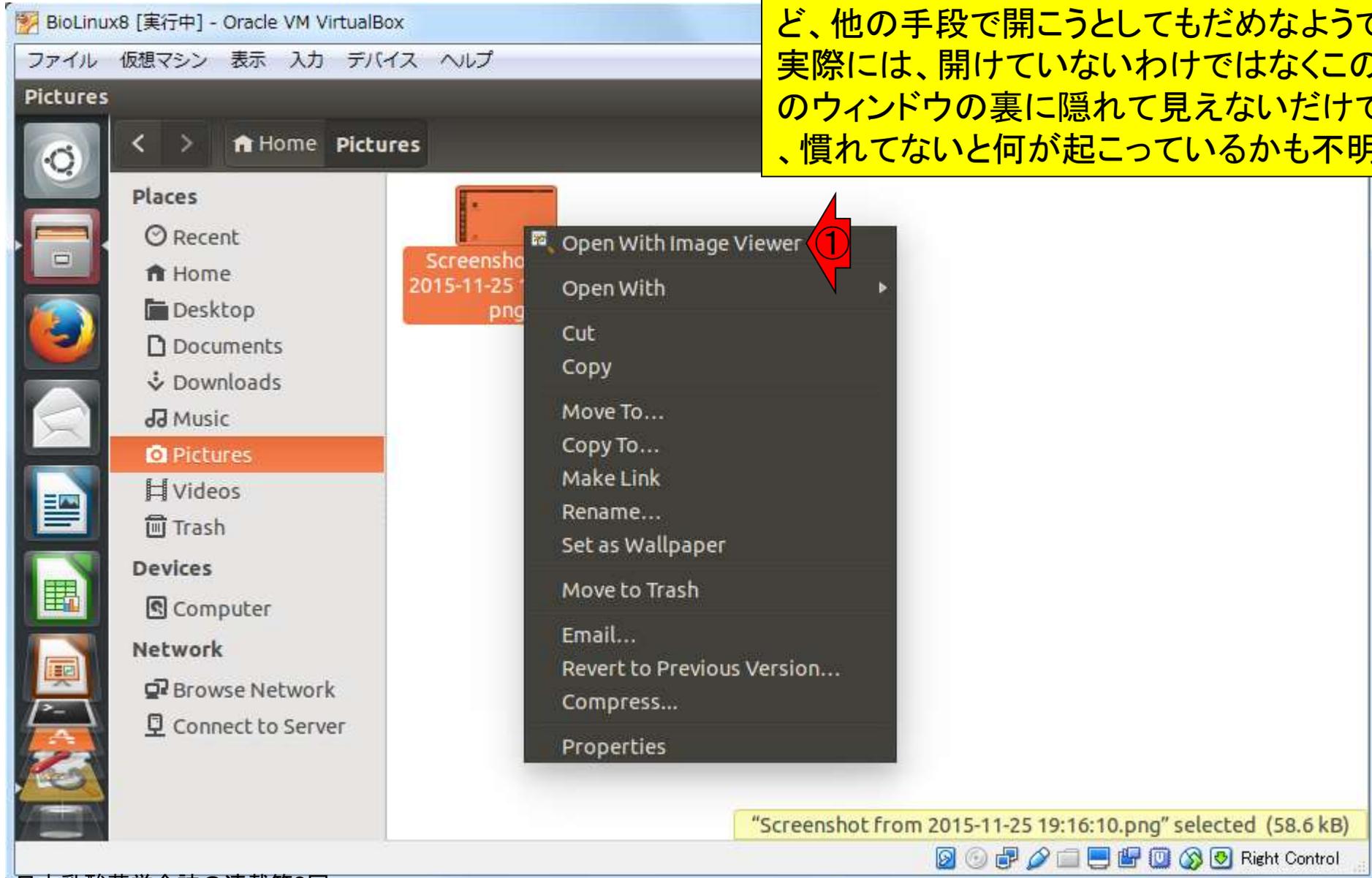
W6-2

開いていたpngファイルが消える(ように見えて、実はPicturesフォルダの裏に隠れているだけ)。



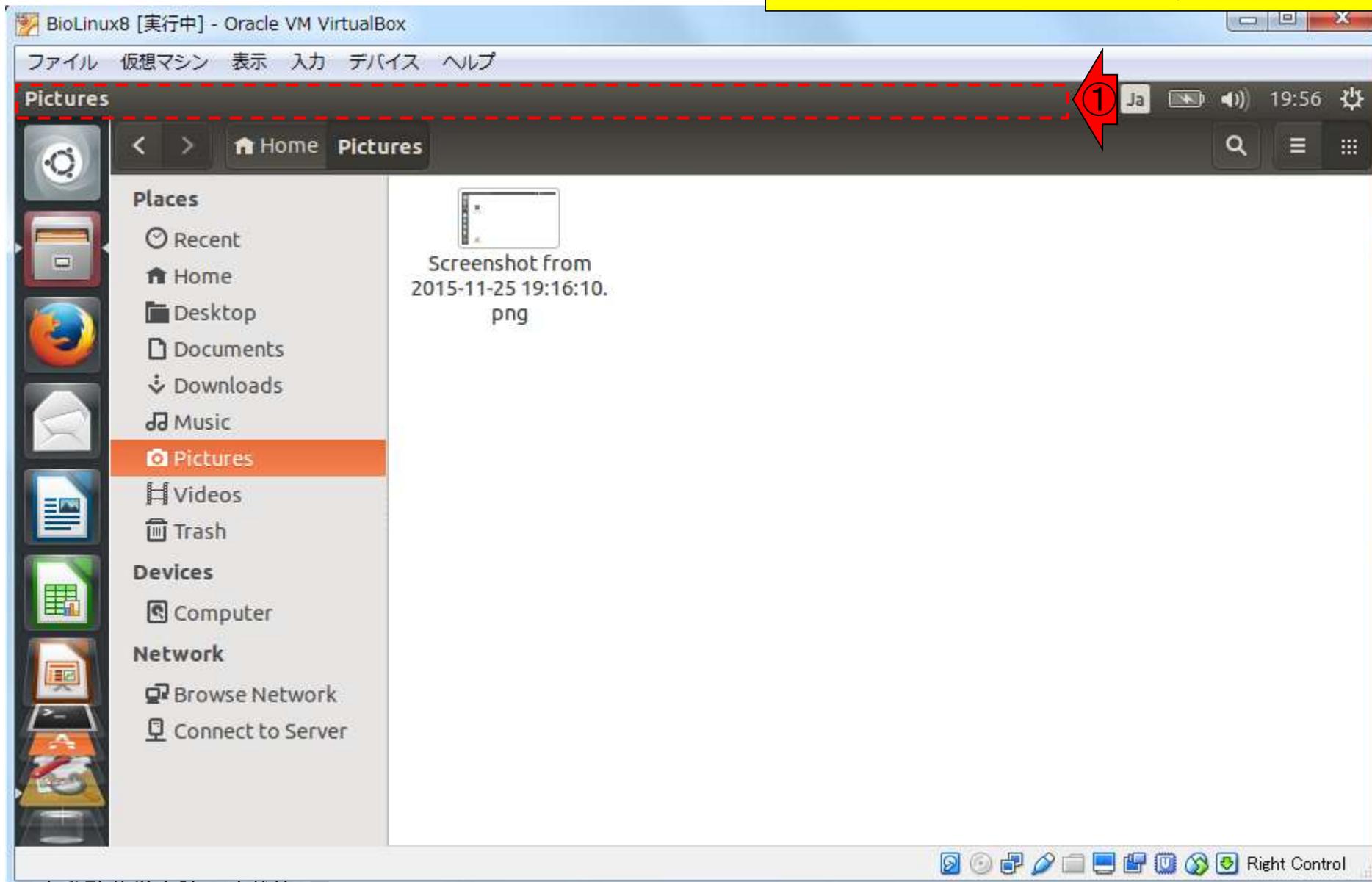
W6-3

もう一度ダブルクリックでpngファイルを開こうとしても開けないことがある。こうなると、例えば右クリックで①Open with Image Viewerを選択するなど、他の手段で開こうとしてもだめなようである。実際には、開けていないわけではなくこの作業中のウィンドウの裏に隠れて見えただけであるが、慣れてないと何が起きているかも不明…。

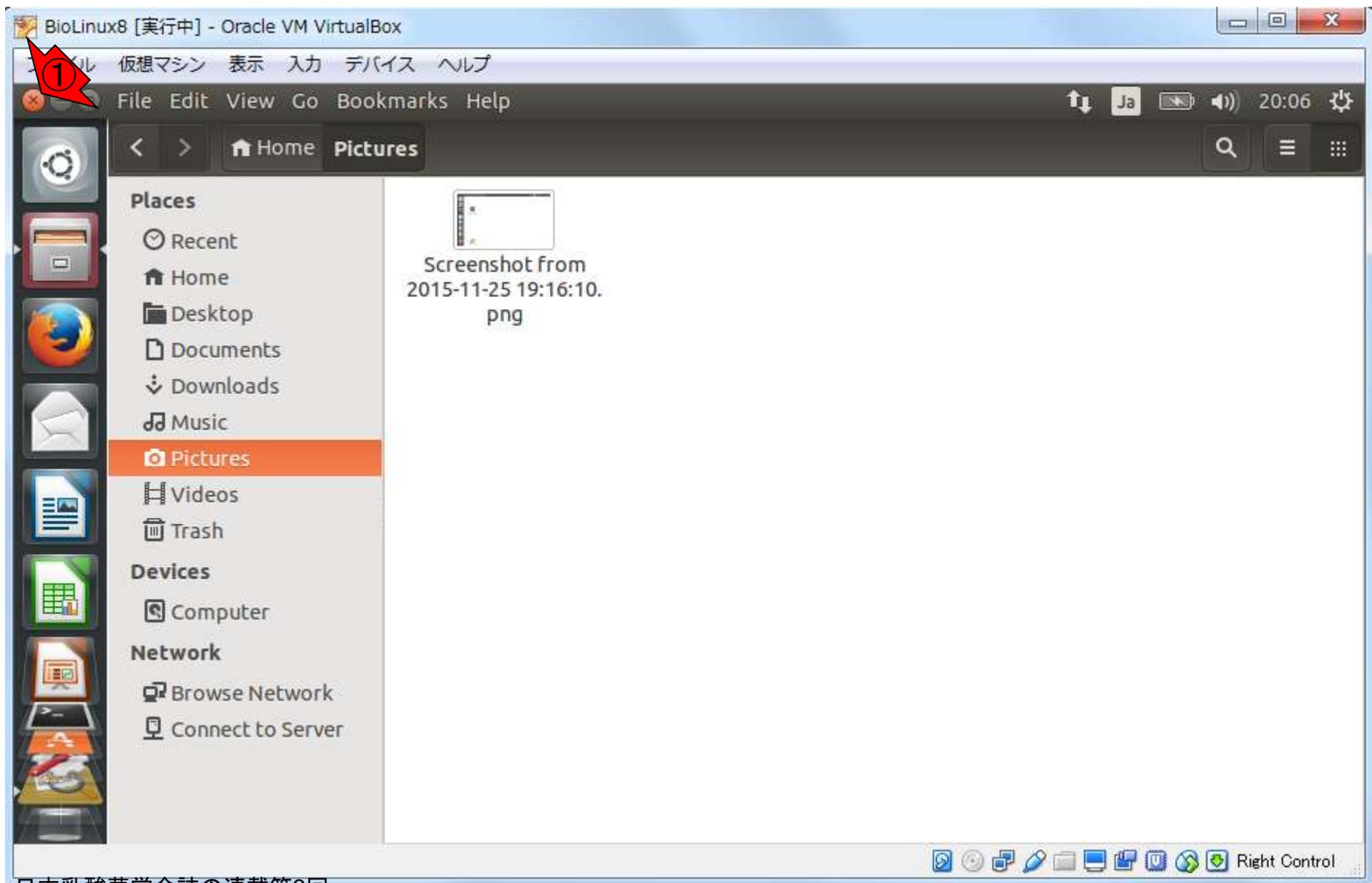


W6-3

Picturesフォルダを開く前の状態に戻したい場合は、まず①赤い点線の枠内にカーソルを移動させる。そうするとメニューが表示される。

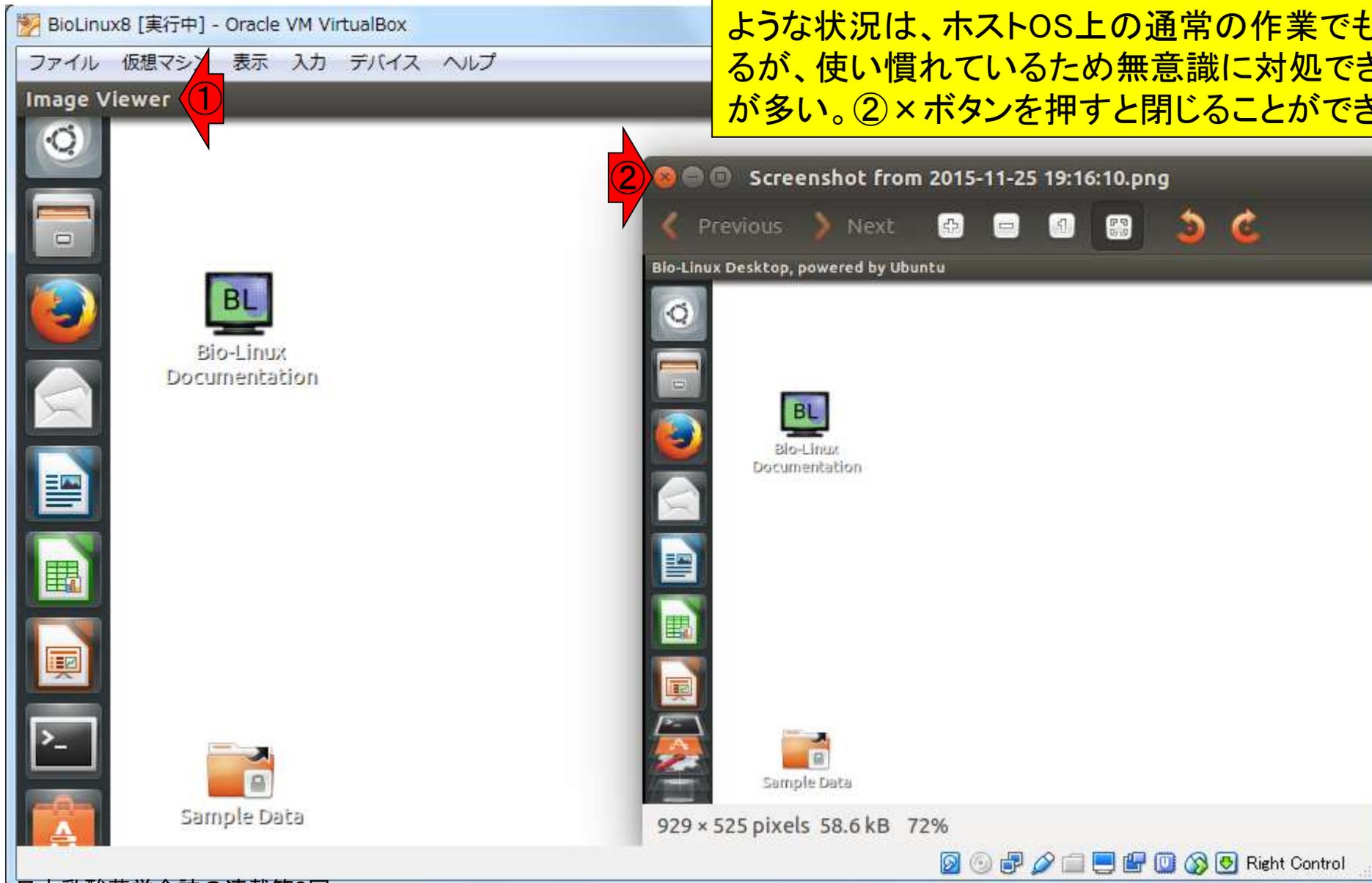


W6-3



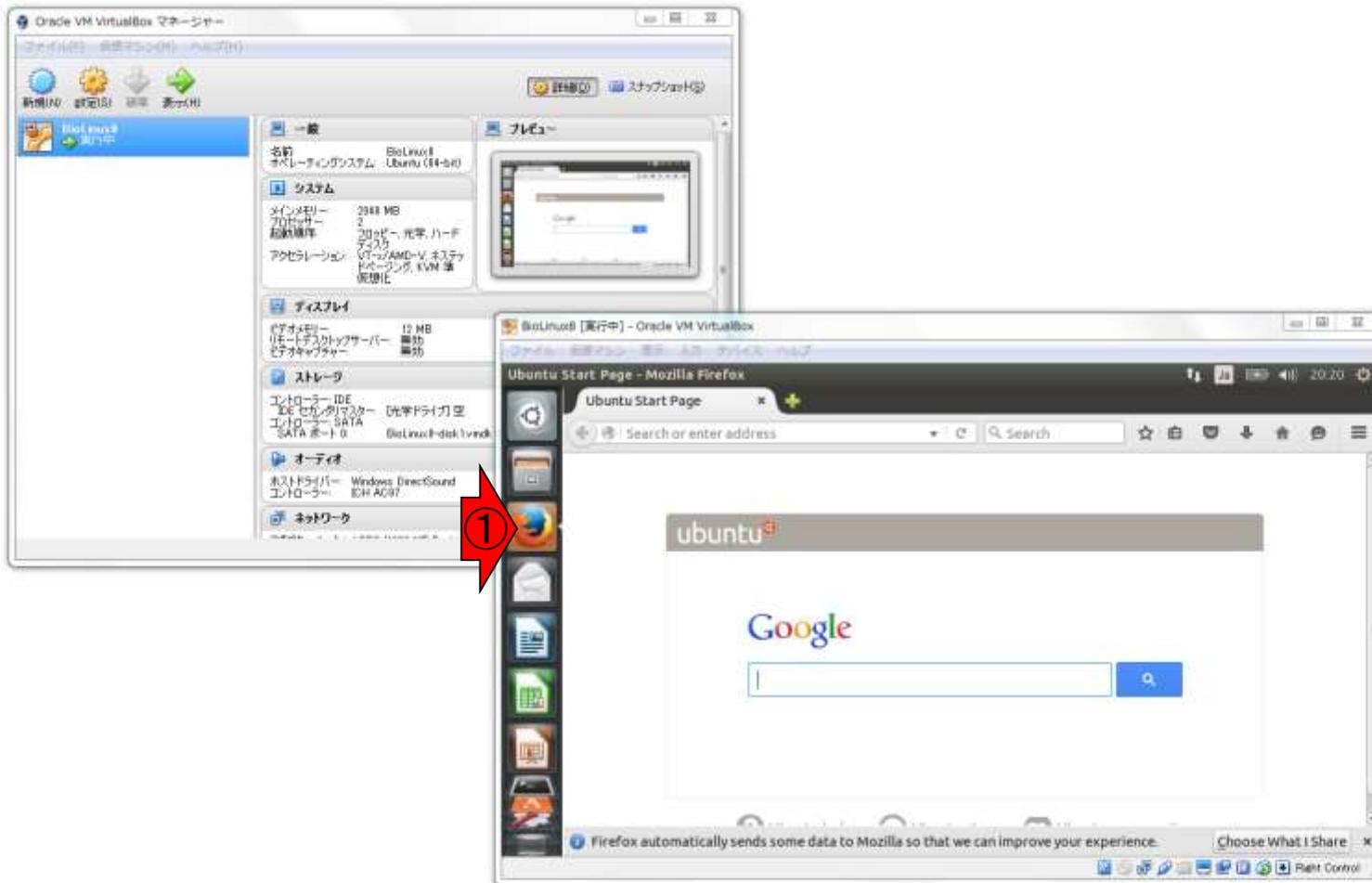
W6-3

この段階で初めて、①Image Viewerで画像ファイルを開いていたことに気づく。つまり、さきほどのPicturesフォルダのウィンドウに隠れていただけである。このような状況は、ホストOS上の通常の作業でも遭遇するが、使い慣れているため無意識に対処できる場合が多い。②×ボタンを押すと閉じることができる。



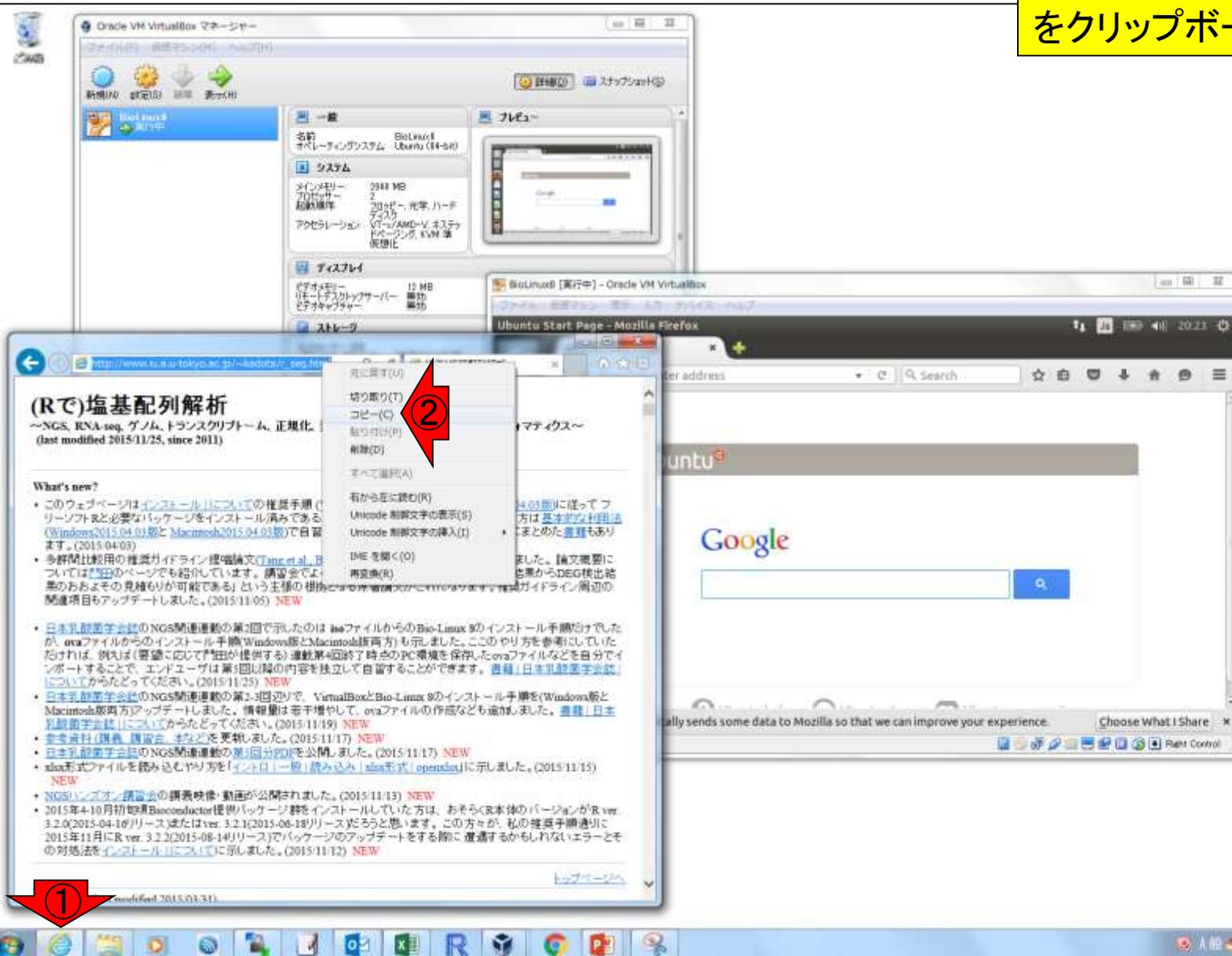
W7-1: 付属ソフト

①ウェブブラウザFirefox起動後の状態。これはゲストOS上のブラウザ。



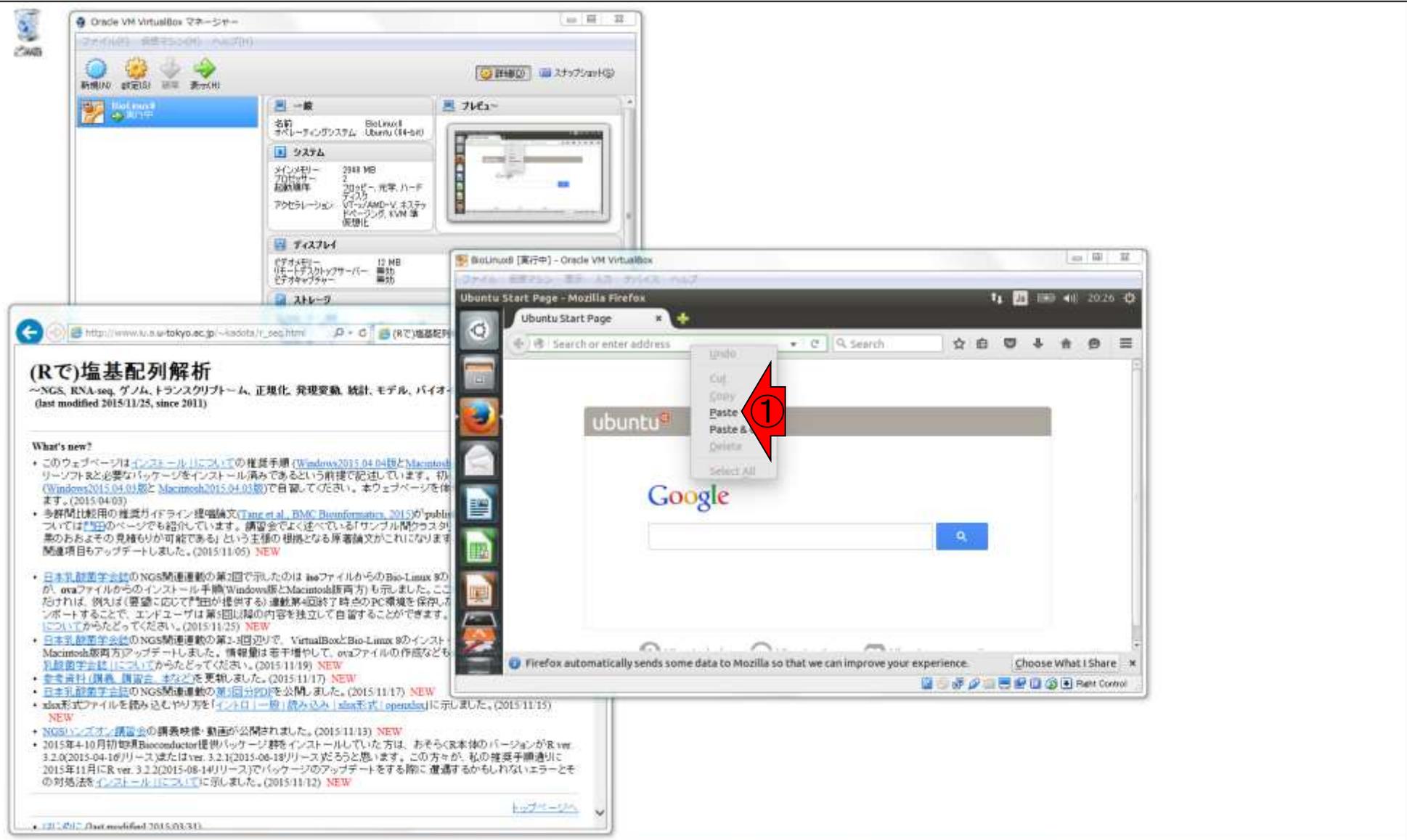
W7-1: 付属ソフト

①ホストOS上のウェブブラウザ
Internet Explorerを起動し、本連載
の参考ウェブページを起動。②URL
をクリップボードにコピーして…



①ゲストOS上のFirefoxで
ペーストすることができる

W7-1: 付属ソフト



W7-1: 付属ソフト

The screenshot shows a Windows 7 desktop environment. In the background, the Oracle VM VirtualBox Manager is open, displaying the configuration for a virtual machine named 'BioLinux8'. The VM is based on the Ubuntu (64-bit) operating system. The configuration details include:

- 名前:** BioLinux8
- オペレーティングシステム:** Ubuntu (64-bit)
- システム:**
 - メモリ: 2048 MB
 - ハードディスク: 2
 - 起動順序: 2000px, 光学, ハードディスク
 - アクセラレーション: VT-x/AMD-V, ネステドモード, KVM 準拠
- ディスプレイ:**
 - ビデオメモリ: 12 MB
 - ハードウェアアクセラレーション: 無効
 - ビデオアダプター: ビデオアダプター
- ストレージ:**

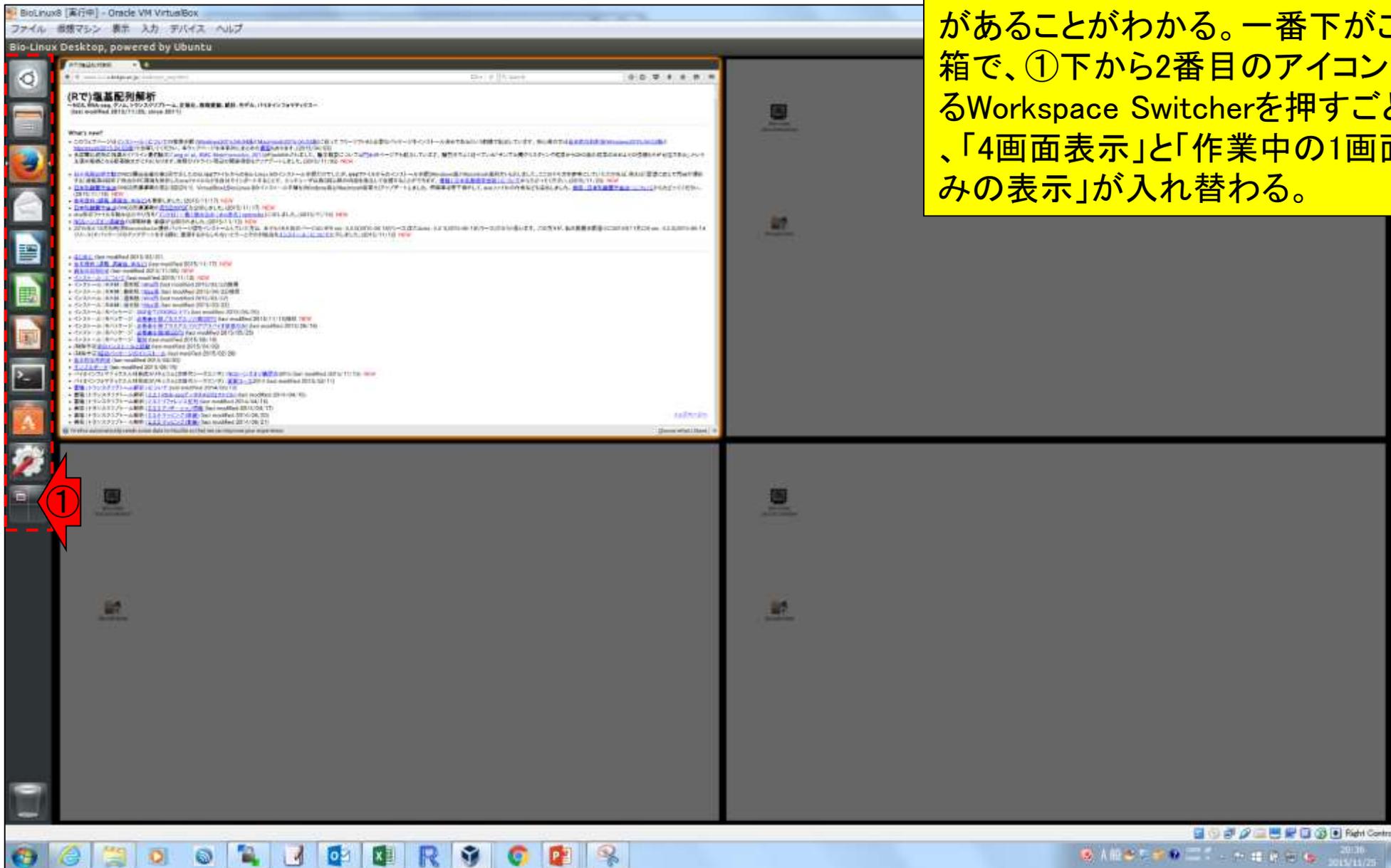
In the foreground, a Mozilla Firefox browser window is open, displaying the webpage 'http://www.iu.u-tokyo.ac.jp/~kadota/r_seq.html'. The page title is '(Rで)塩基配列解析' (Base sequence analysis with R). The content includes a 'What's new?' section with several updates:

- このウェブページはインストール1についての推奨手順 (Windows 2015.04.04版と Macintosh 2015.04.03版) 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](#) (Windows 2015.04.03版と Macintosh 2015.04.03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)
- 多群間比較用の推奨ガイドライン提唱論文 (Tang et al., BMC Bioinformatics, 2015) が publish されています。このページでも紹介しています。講習会でよく述べている「サンプル間クラスター化のおおよその見積りも可能である」という主張の根拠となる原著論文がこれになります。関連項目もアップデートしました。(2015/11/05) **NEW**
- 日本乳がん学会の NGS 関連連載の第2回で示したのは iso ファイルからの Bio-Linux 8 のインストール手順 (Windows 版と Macintosh 版両方) も示しました。この iso ファイルは (要望に応じて門田が提供する) 連載第4回終了時点の PC 環境を保存したイメージです。エンドユーザは第4回以降の内容を独立して自習することができます。インストールから始めてください。(2015/11/25) **NEW**
- 日本乳がん学会の NGS 関連連載の第2回 (2015/11/19) で、VirtualBox と Bio-Linux 8 のインストール手順 (Windows 版と Macintosh 版両方) をアップデートしました。情報量は若干増やして、ova ファイルの作成なども [インストール1](#) に追加してご紹介ください。(2015/11/19) **NEW**
- 本ウェブページ (構成、講習会、本などを) を更新しました。(2015/11/17) **NEW**
- 日本乳がん学会の NGS 関連連載の第2回 (2015/11/17) を公開しました。(2015/11/17) **NEW**
- xlsx 形式のファイルを読み込む方法を「インストール1」の「読み込み」>「xlsx 形式」>「openxlsx」に示しました。(2015/11/15) **NEW**
- NGS イングレブの講習会の講義映像・動画が公開されました。(2015/11/13) **NEW**
- 2015年4-10月初期 Biconductor 提供パッケージ群をインストールしていた方は、おそらく (ご自身のバージョンが R ver. 3.2.0 (2015-04-14) リリースまたは R ver. 3.2.1 (2015-06-18) リリース) だるうかと思えます。この方が、私の推奨手順通りに2015年11月に R ver. 3.2.2 (2015-08-14) リリースでパッケージのアップデートをする際に遭遇するかもしれないエラーとその対処法を [インストール1](#) に追加して示しました。(2015/11/12) **NEW**

A red arrow with the number '1' points to the top of the browser window, indicating the instruction to maximize the page.

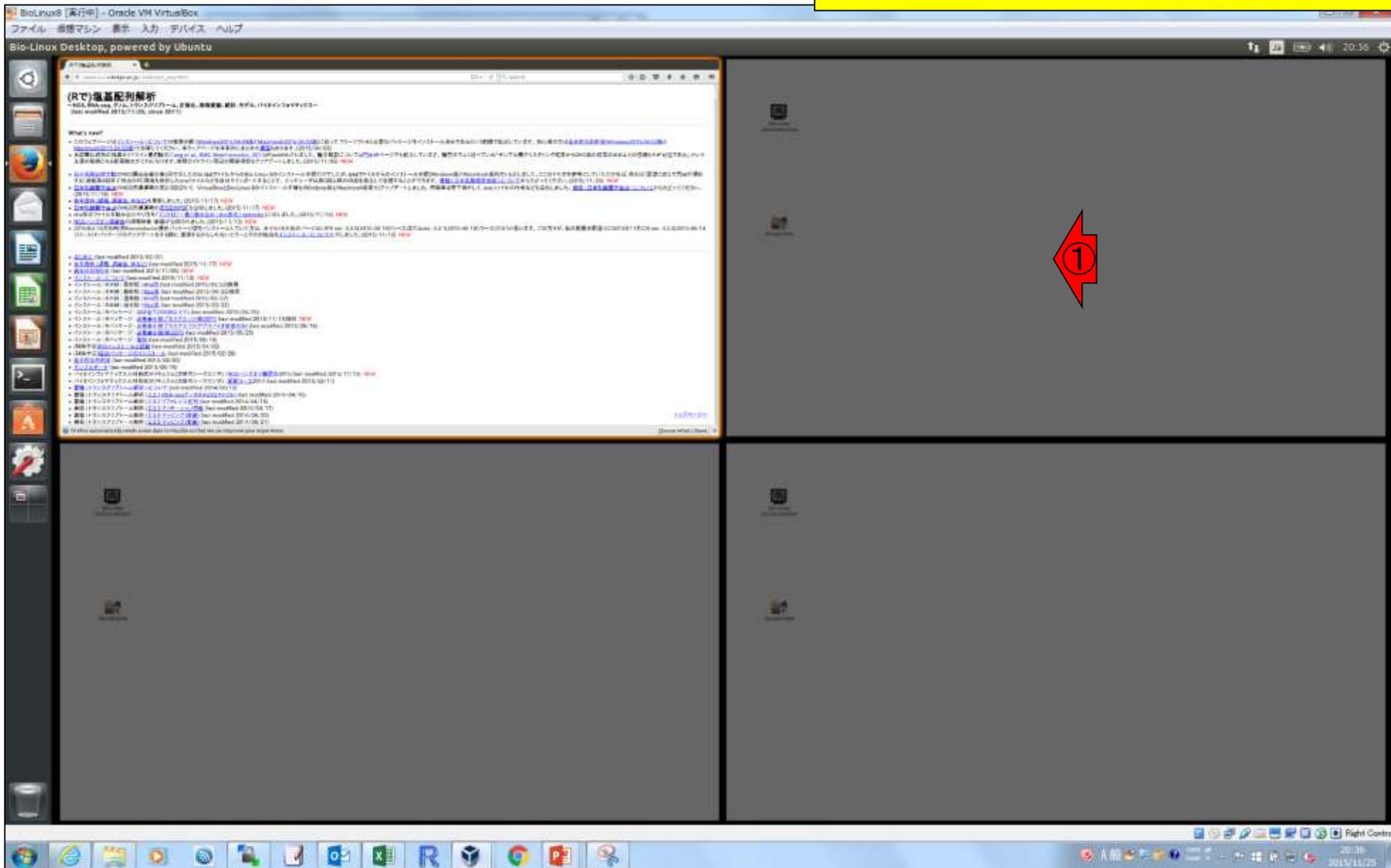
W7-2: マルチ画面

赤枠内がメニューバーのようなもの。この枠内にカーソルを移動させると、下のほうにもいくつかのアイコンがあることがわかる。一番下がごみ箱で、①下から2番目のアイコンであるWorkspace Switcherを押すごとに、「4画面表示」と「作業中の1画面のみの表示」が入れ替わる。



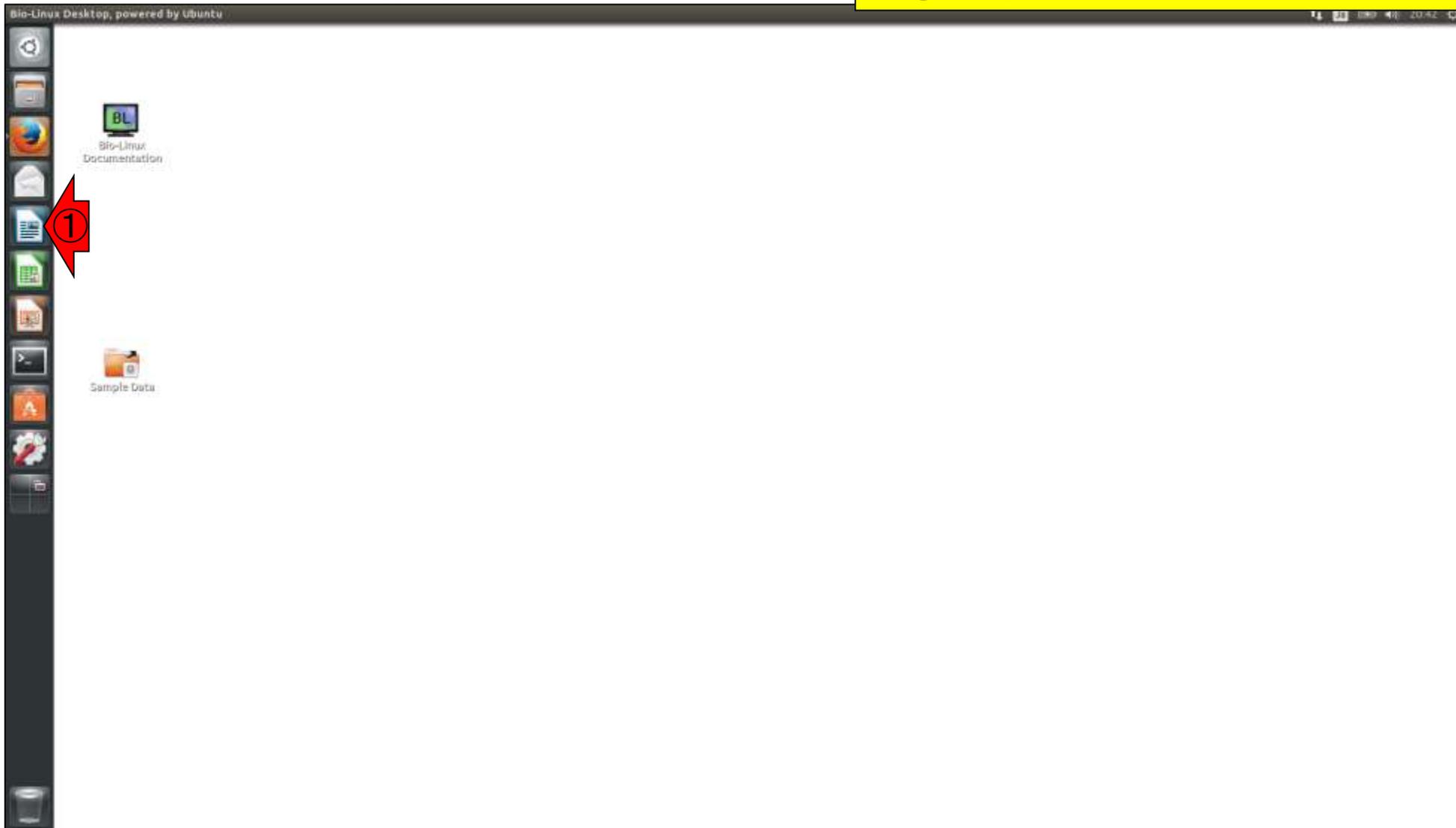
W7-2: マルチ画面

①右上の画面上でクリックすると、ウェブブラウザが表示されない新しい画面となる。このようにマルチ画面を駆使するのが一般的



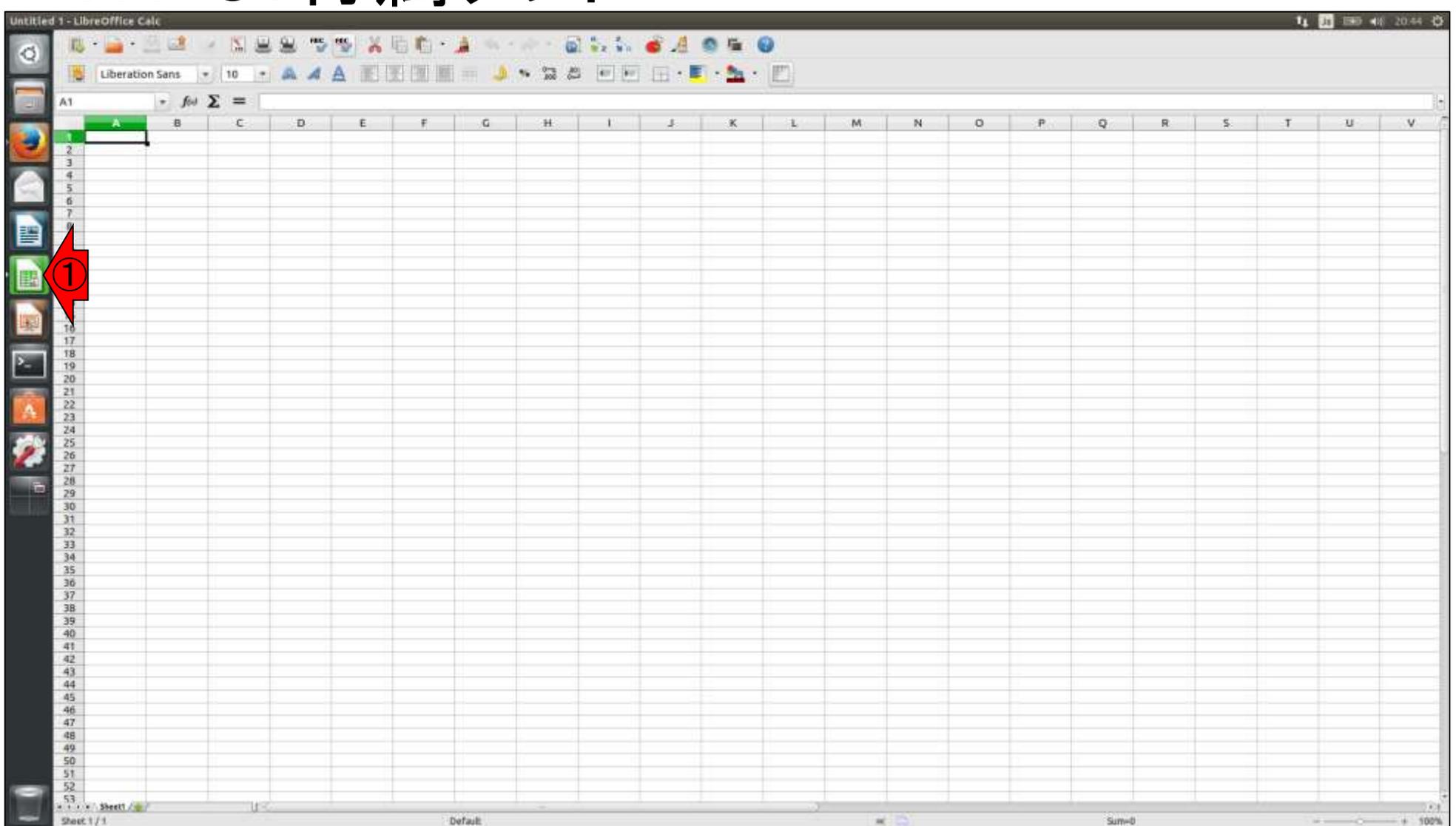
W7-3: 付属ソフト

4画面分の右上を表示させた結果。テキストエディタ(LibreOffice Writer)を開きたい場合は①ここをクリック。実際にはしなくてよい。



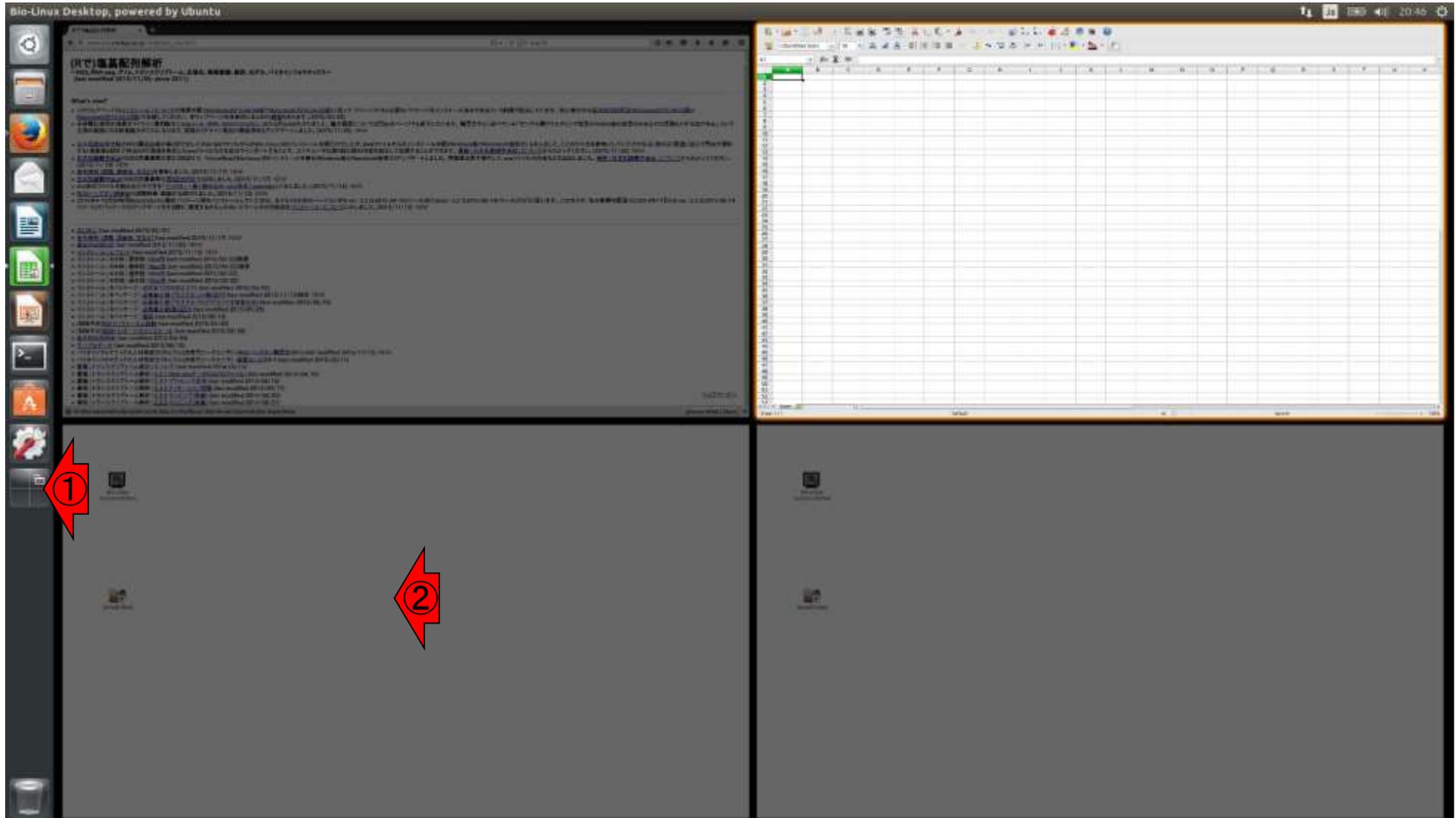
W7-3: 付属ソフト

①表計算ソフト(LibreOffice Calc)
起動後の状態。



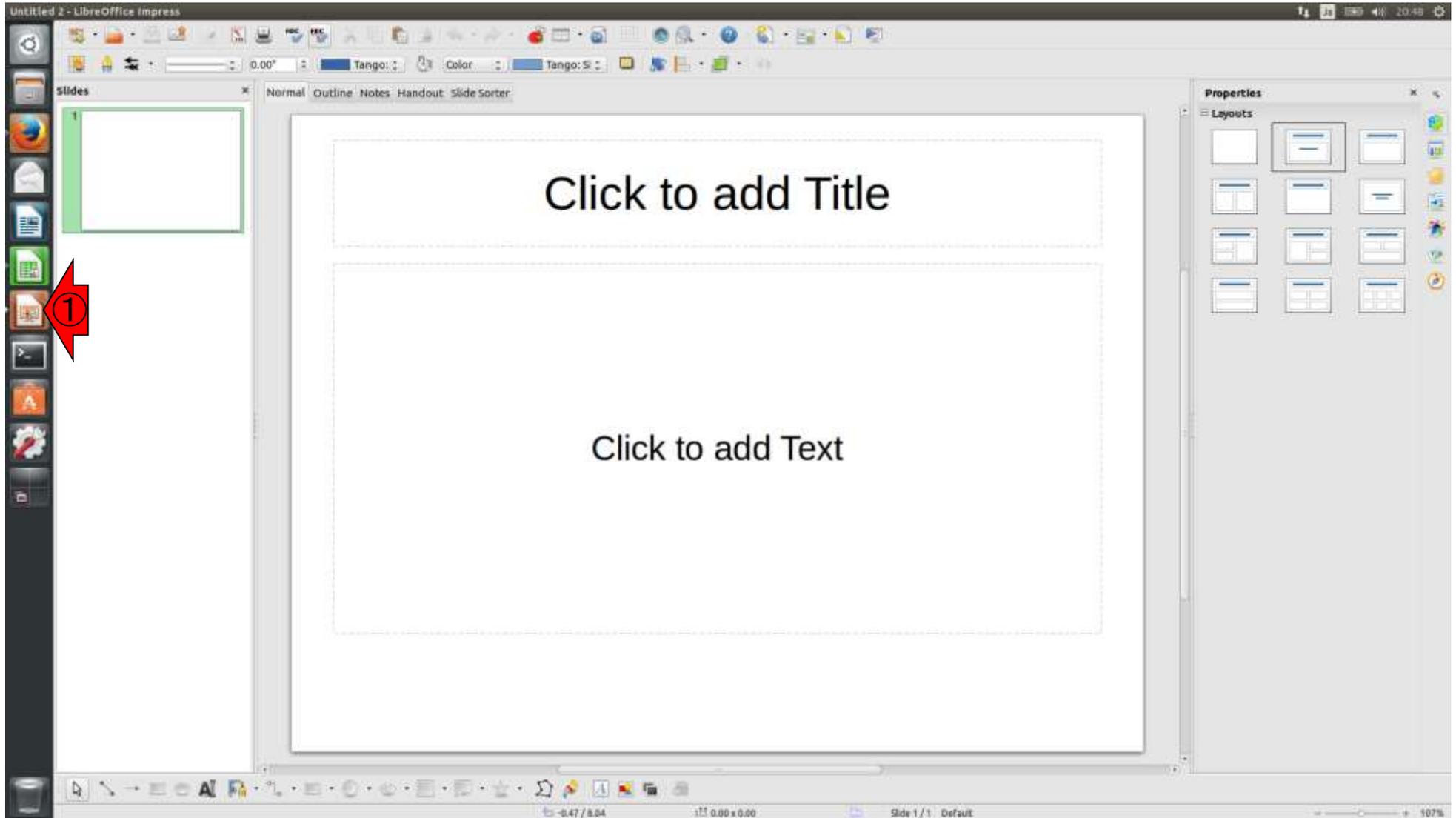
W7-3: 付属ソフト

①Workspace Switcherアイコンをクリックした結果。②4画面分の左下を表示させる。



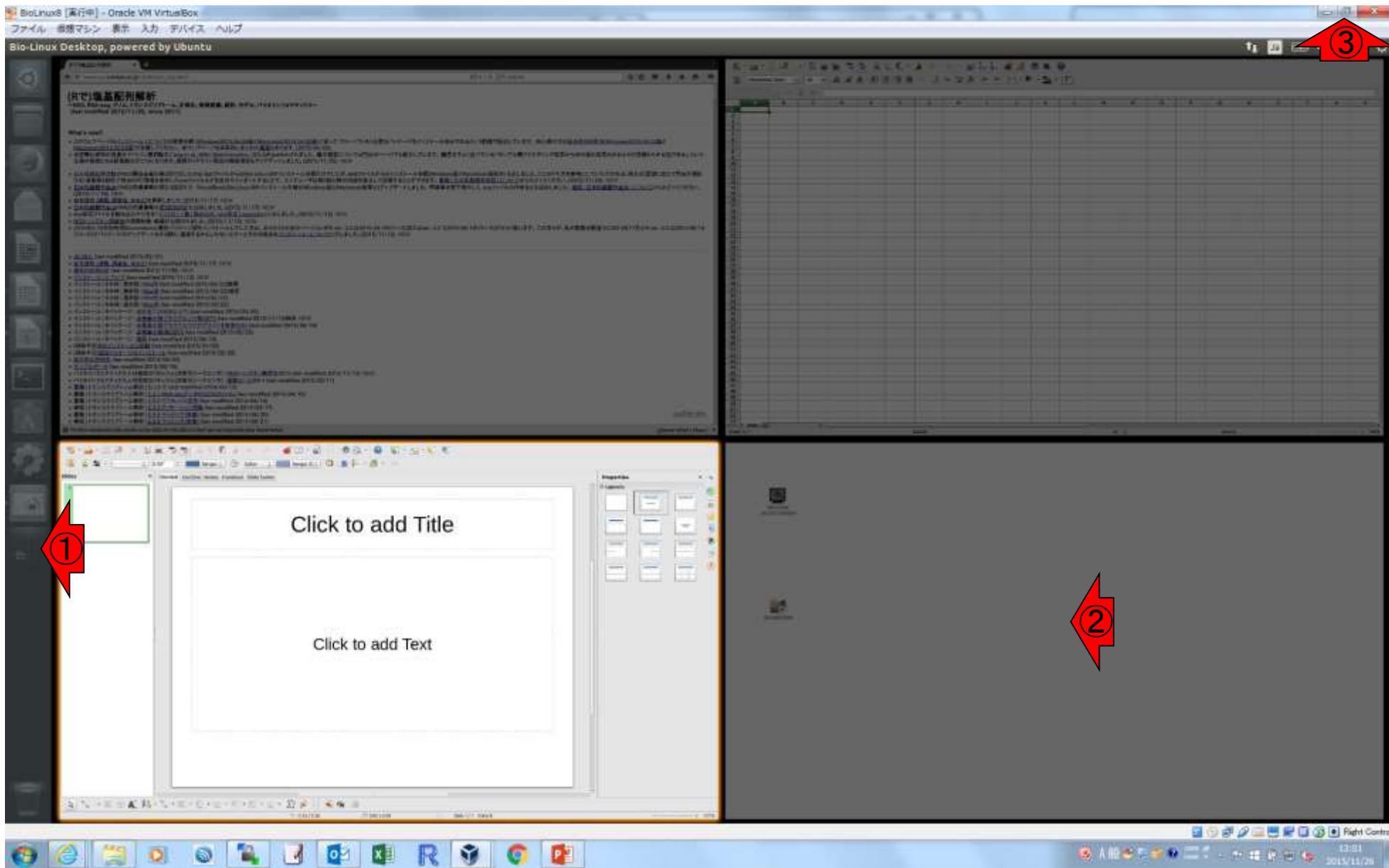
W7-3: 付属ソフト

4画面分の左下を表示させ、①プレゼンテーションソフト(LibreOffice Impress)を起動した結果。



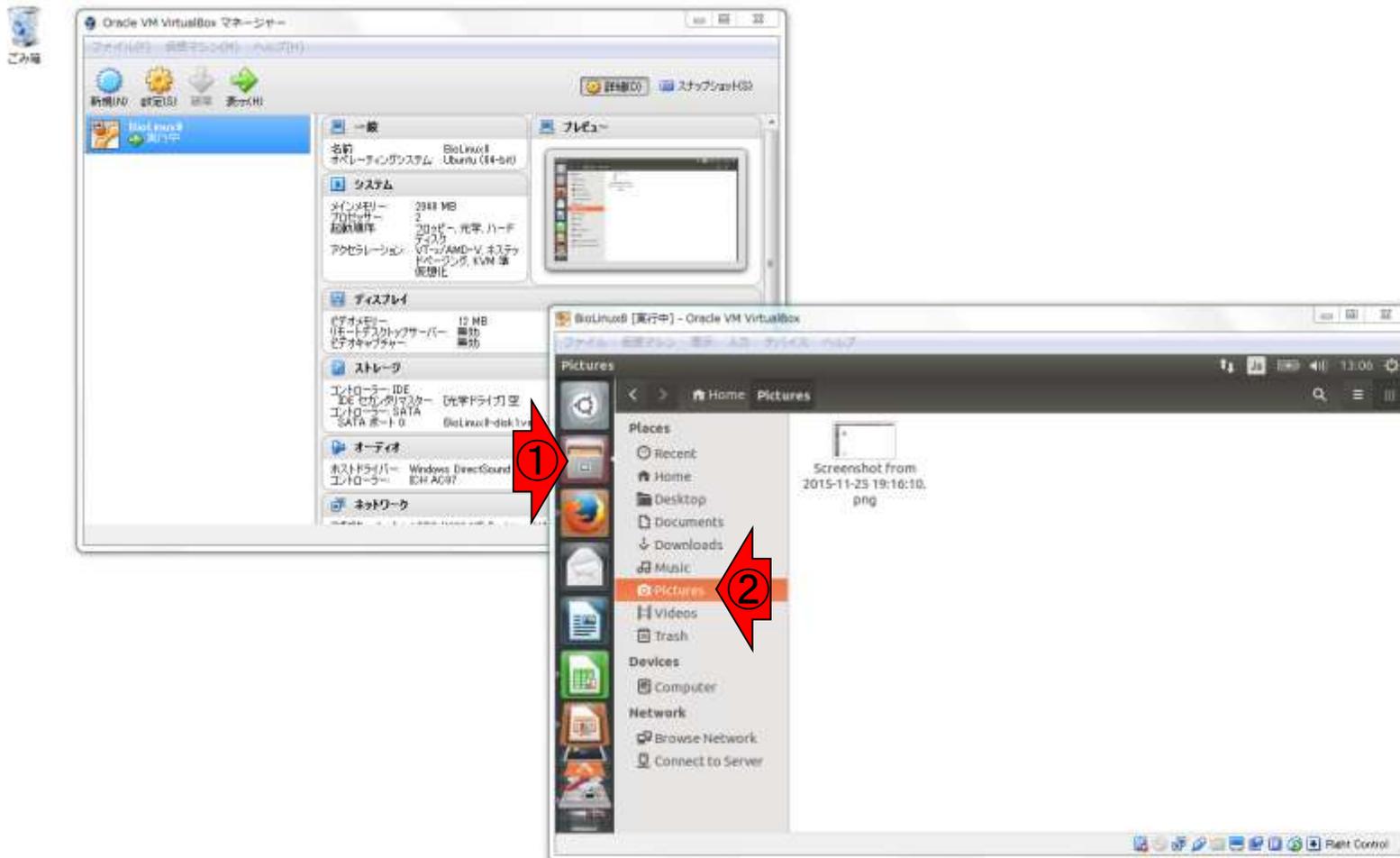
W8-1: ホストからゲスト

①4画面分の②右下を表示させ、③全画面表示を解除



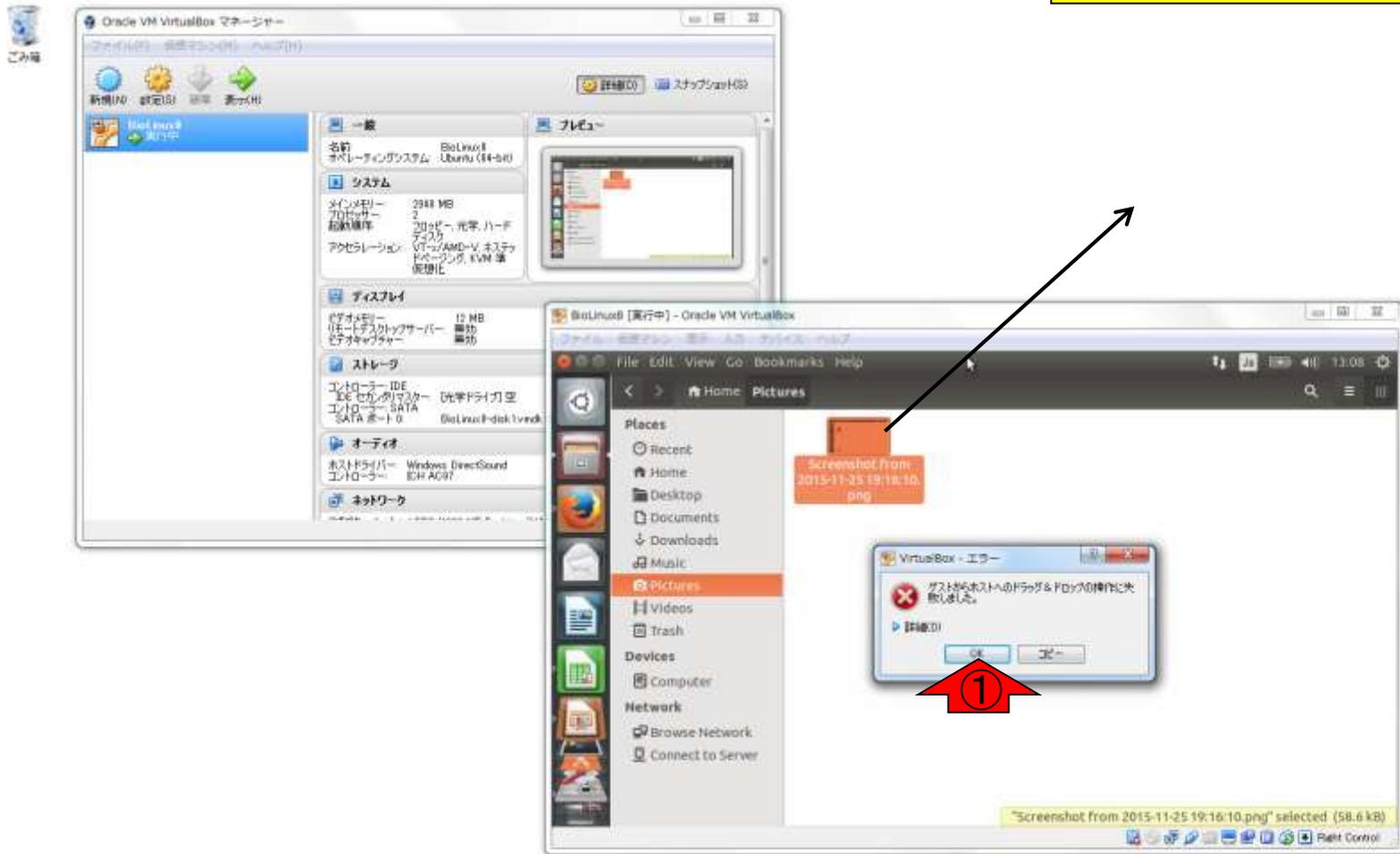
W8-1: ホストからゲスト

- ①フォルダアイコンをクリックし
- ②Picturesフォルダをクリック



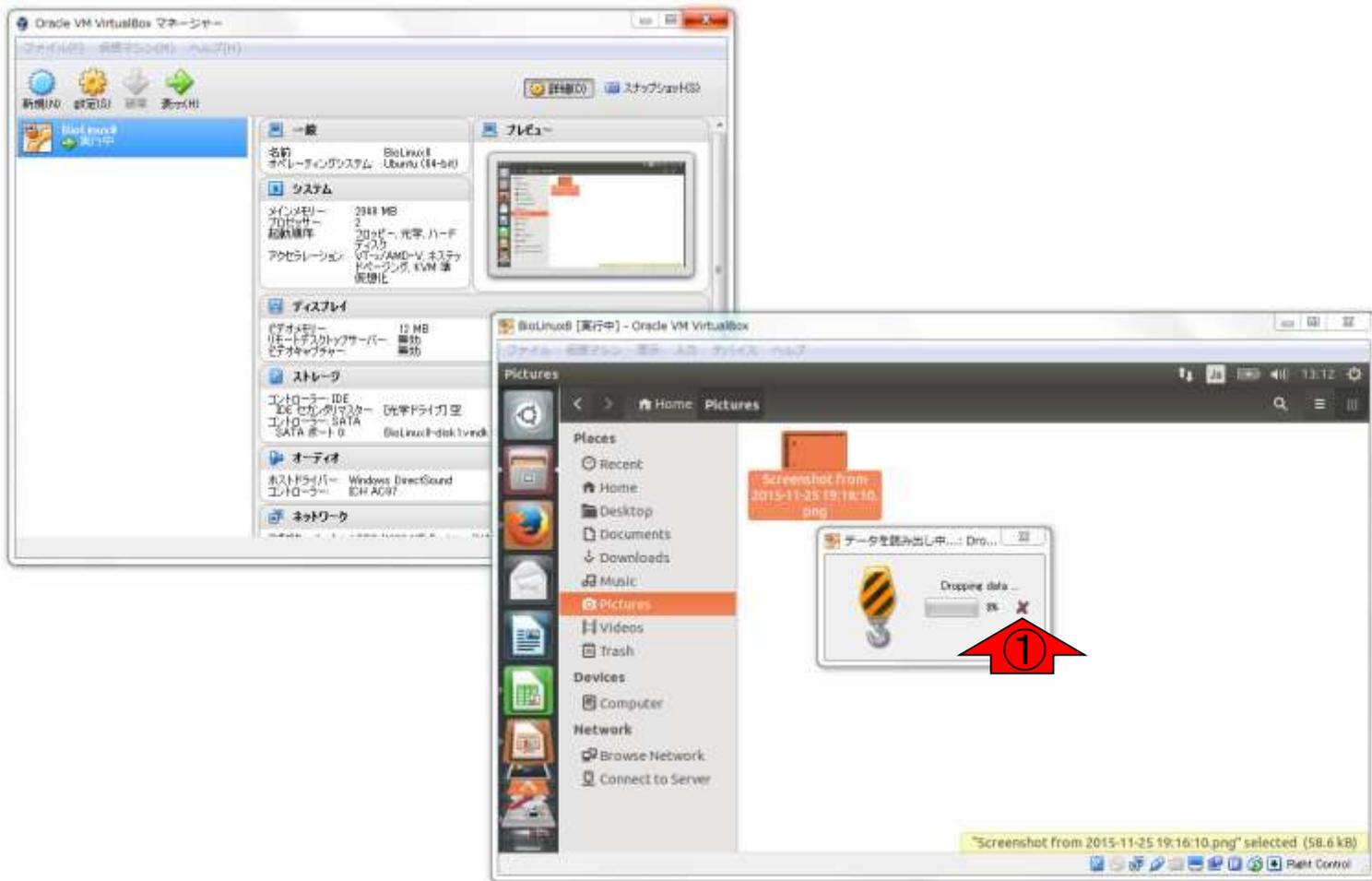
W8-1: ホストからゲスト

ゲストOSのフォルダ内にあるpngファイルを、ホストOSのデスクトップにドラッグ&ドロップすることはできないようだ。①OK。



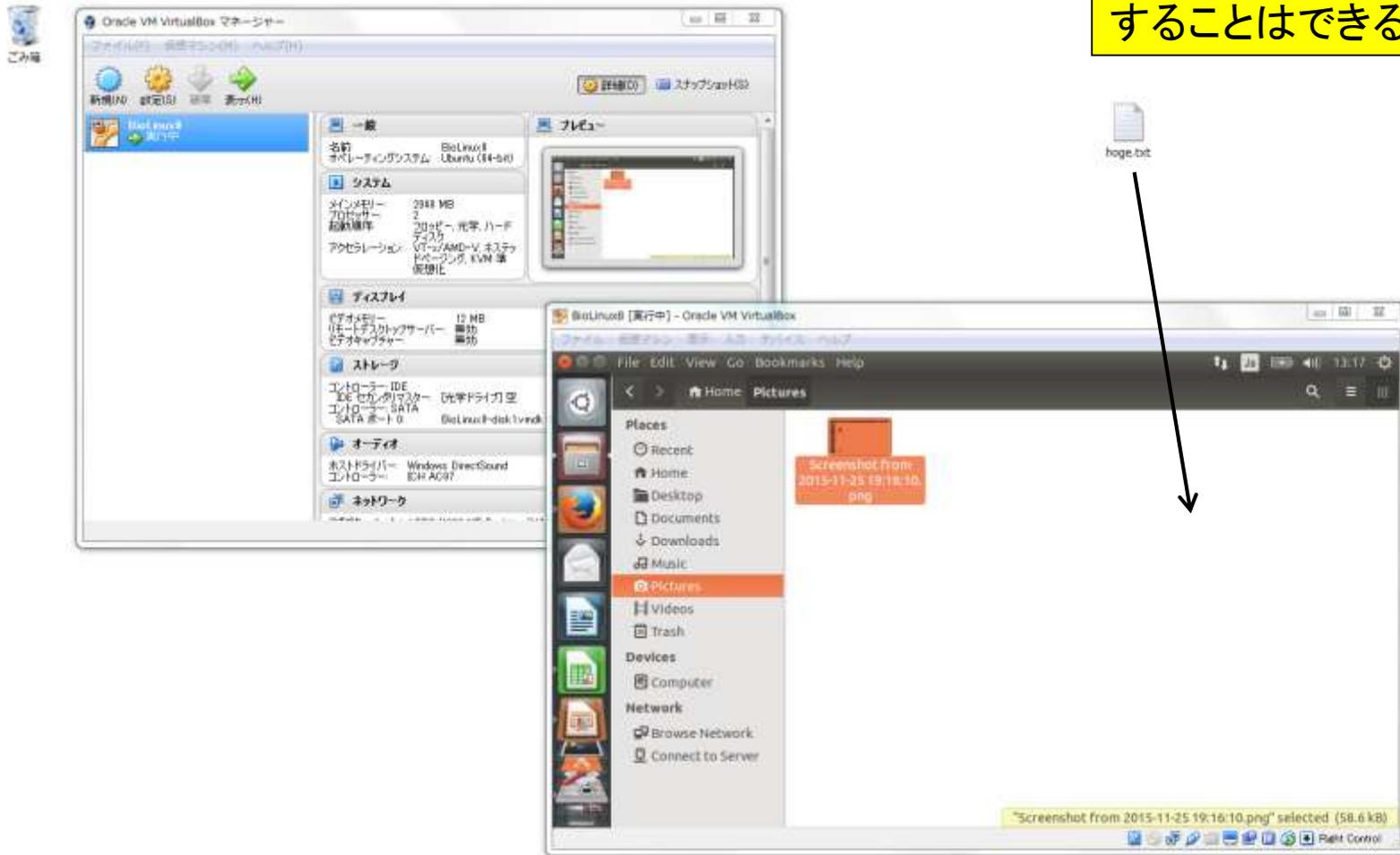
なんか変なことになってフリーズしたようなので、①×。

W8-1: ホストからゲスト



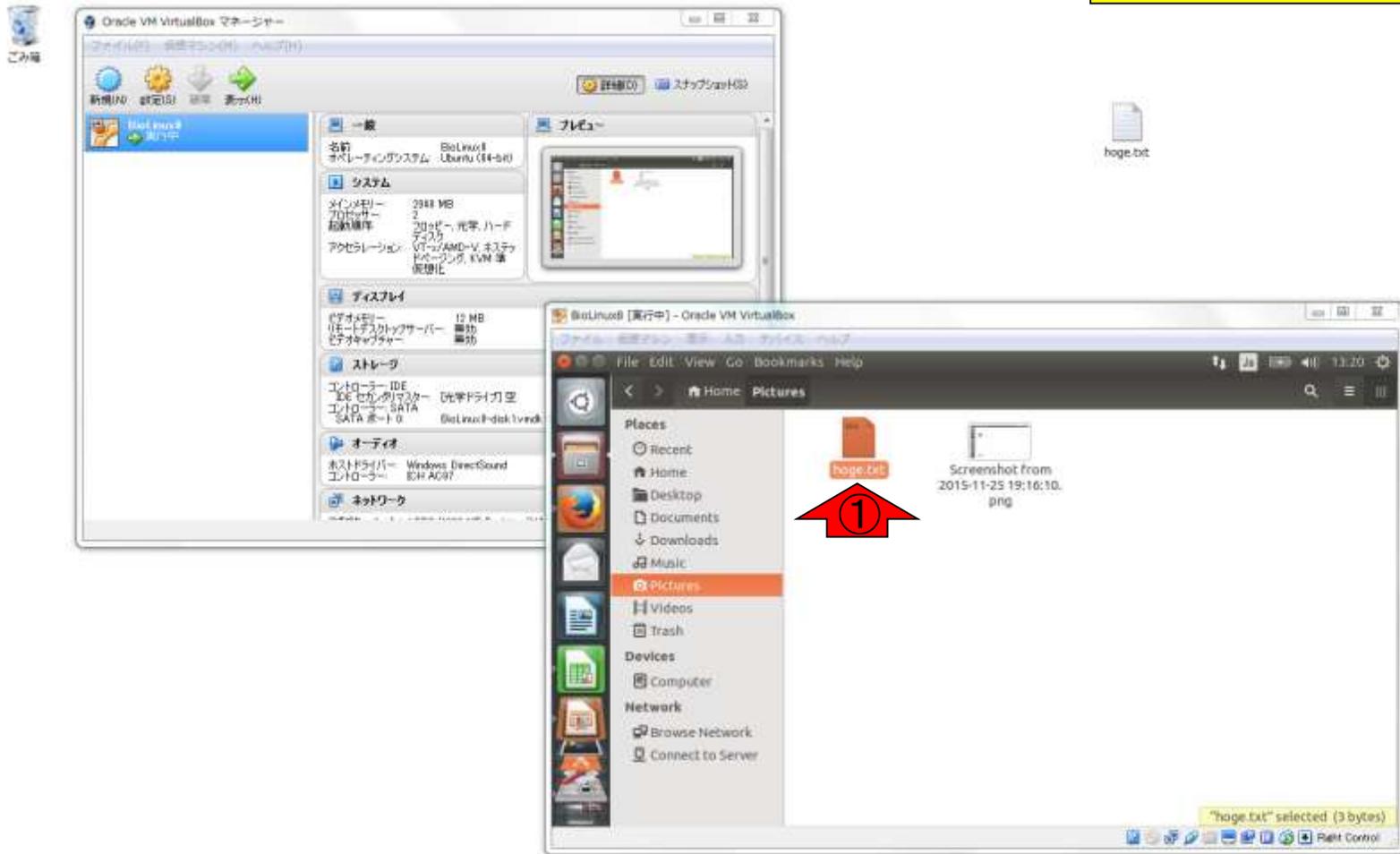
W8-1: ホストからゲスト

ホストOSのデスクトップに置いた任意のファイル(ここではhoge.txt)を、ゲストOSのフォルダ内にドラッグ&ドロップすることはできるようだ



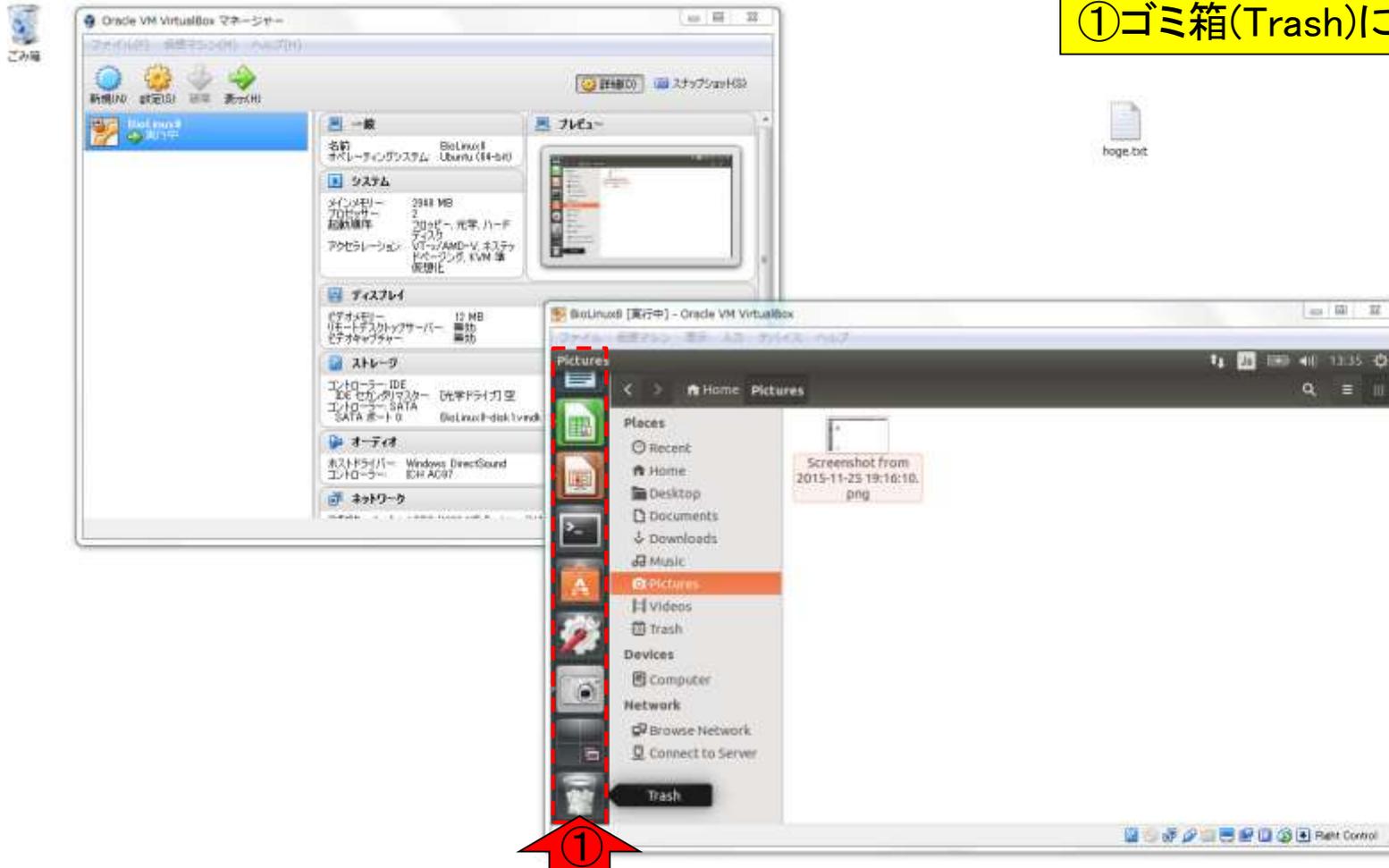
W8-1: ホストからゲスト

ホストOS→ゲストOSへのドラッグ&ドロップ実行結果。①ファイルの移動ではなくコピーになっていることがわかる。



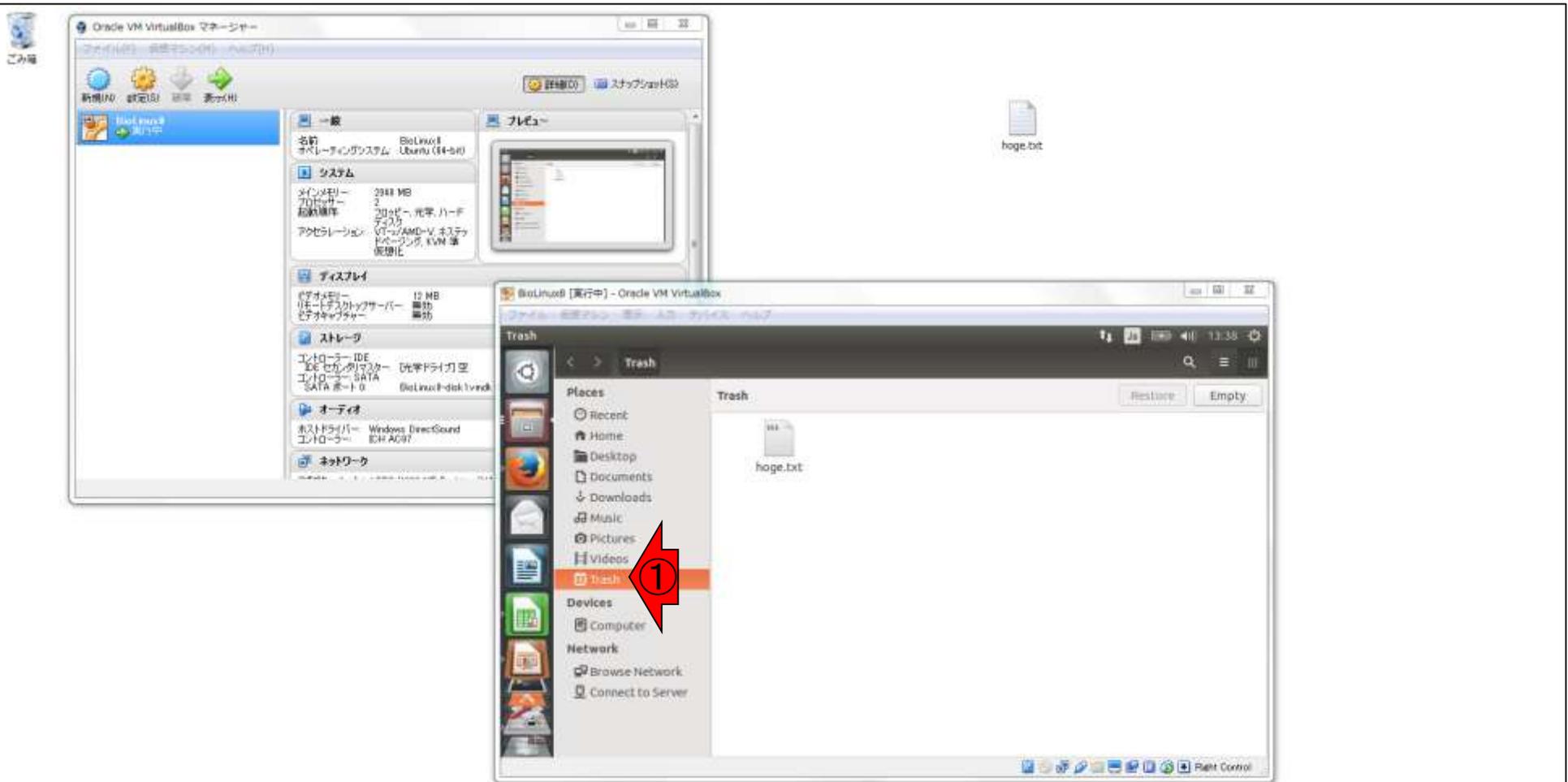
W8-2: ファイルの削除

ファイルの削除は、通常のGUI環境と基本的に同じ。削除したいファイル(例: hoge.txt)を赤枠アイコンの一番下にある①ゴミ箱(Trash)に移動させればよい。



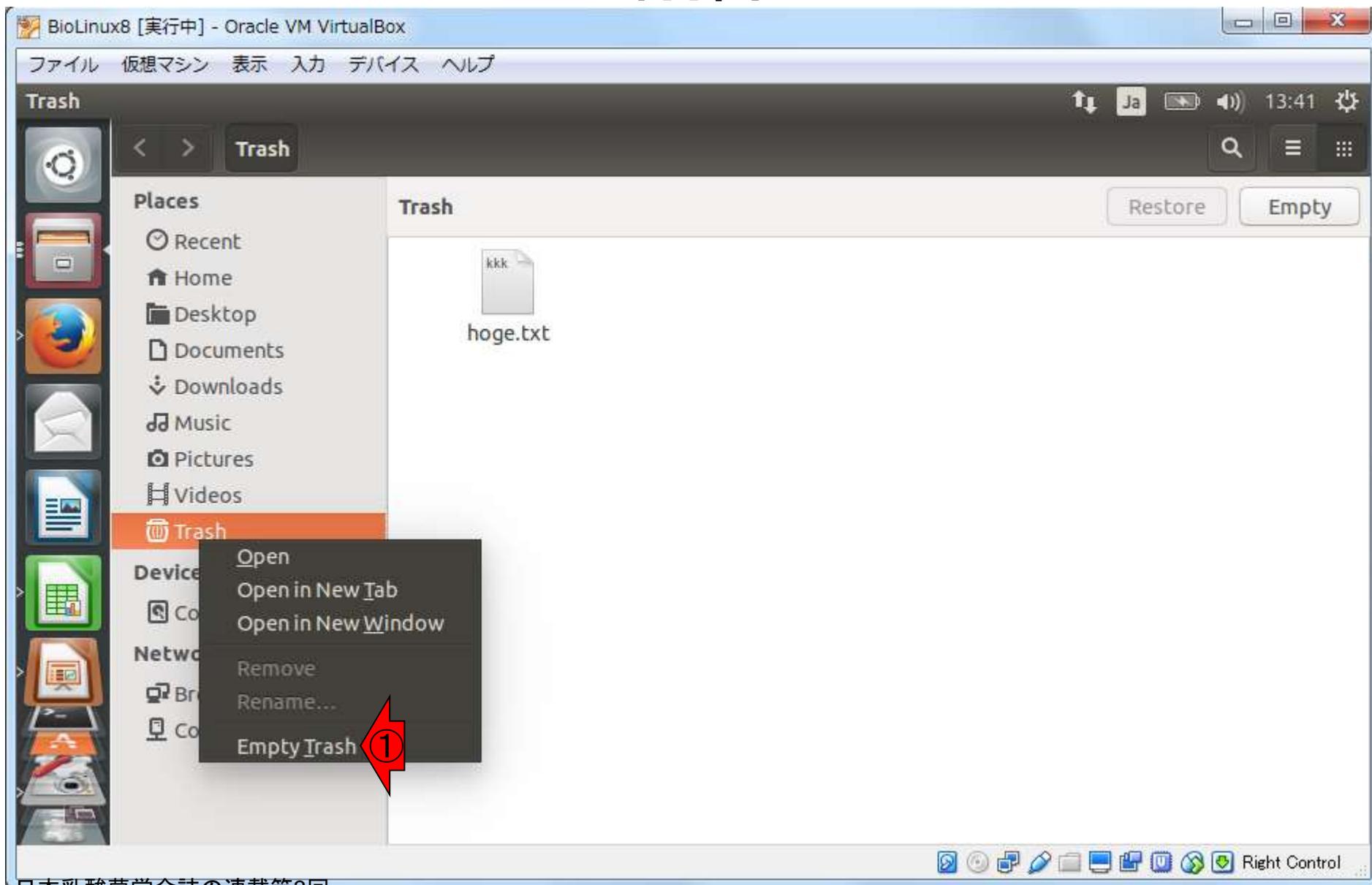
W8-2: ファイルの削除

ゴミ箱をクリックするなり、①赤矢印部分を
クリックすると、ごみ箱の中身を見られる。

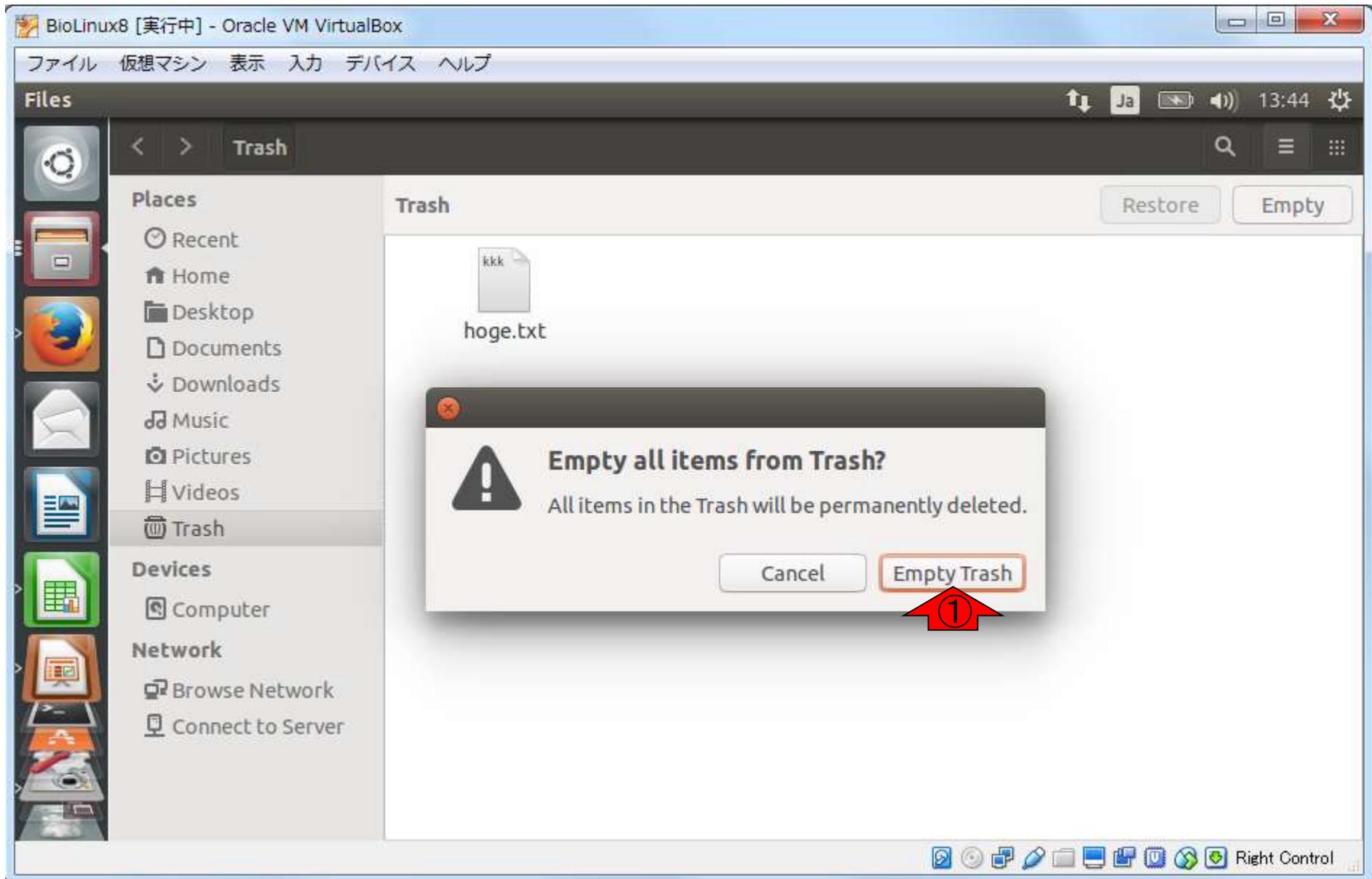


W8-2: ファイルの削除

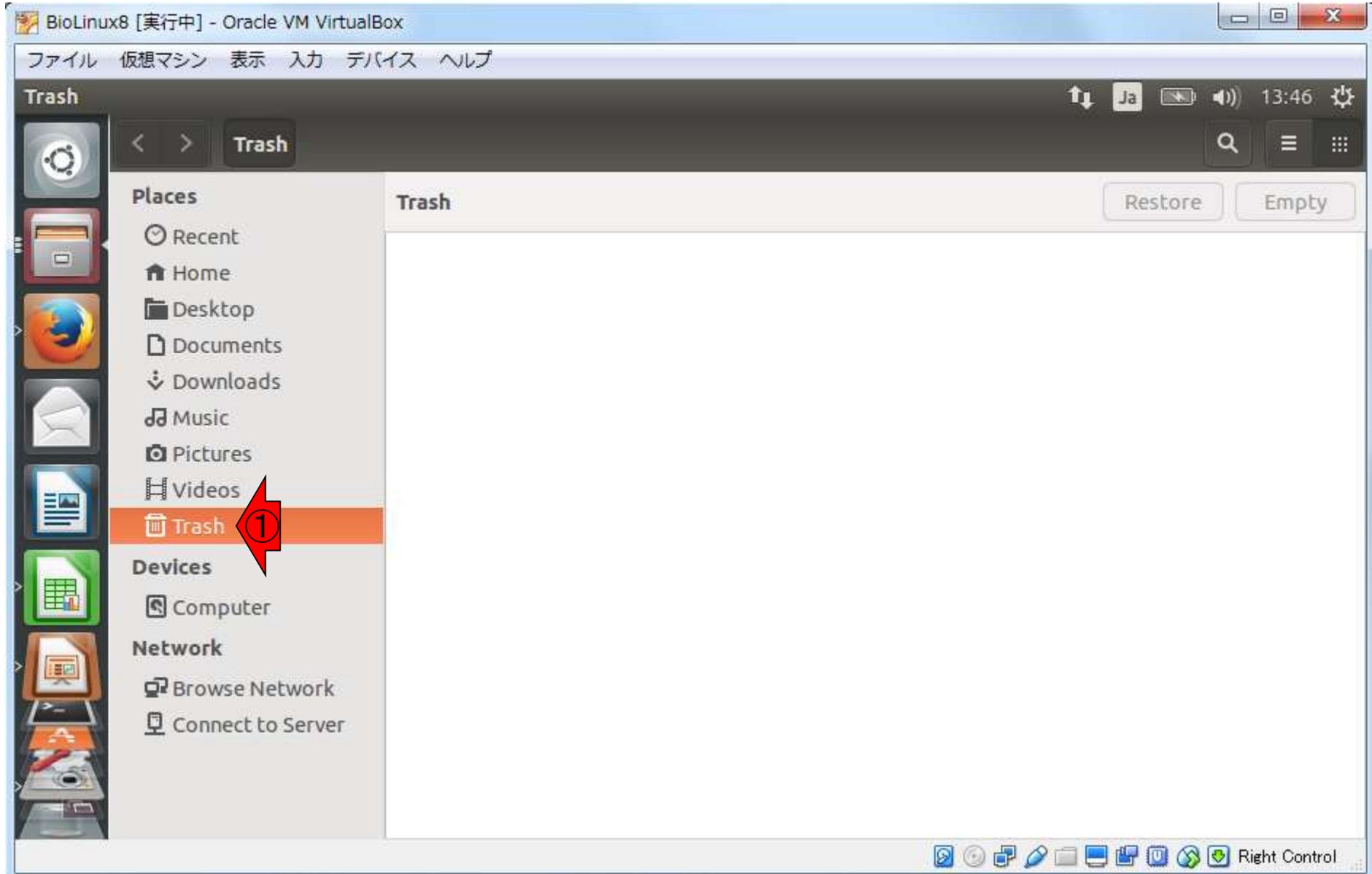
ゴミ箱アイコン上で右クリックし、① Empty Trashでゴミ箱を空にできる



W8-2: ファイルの削除

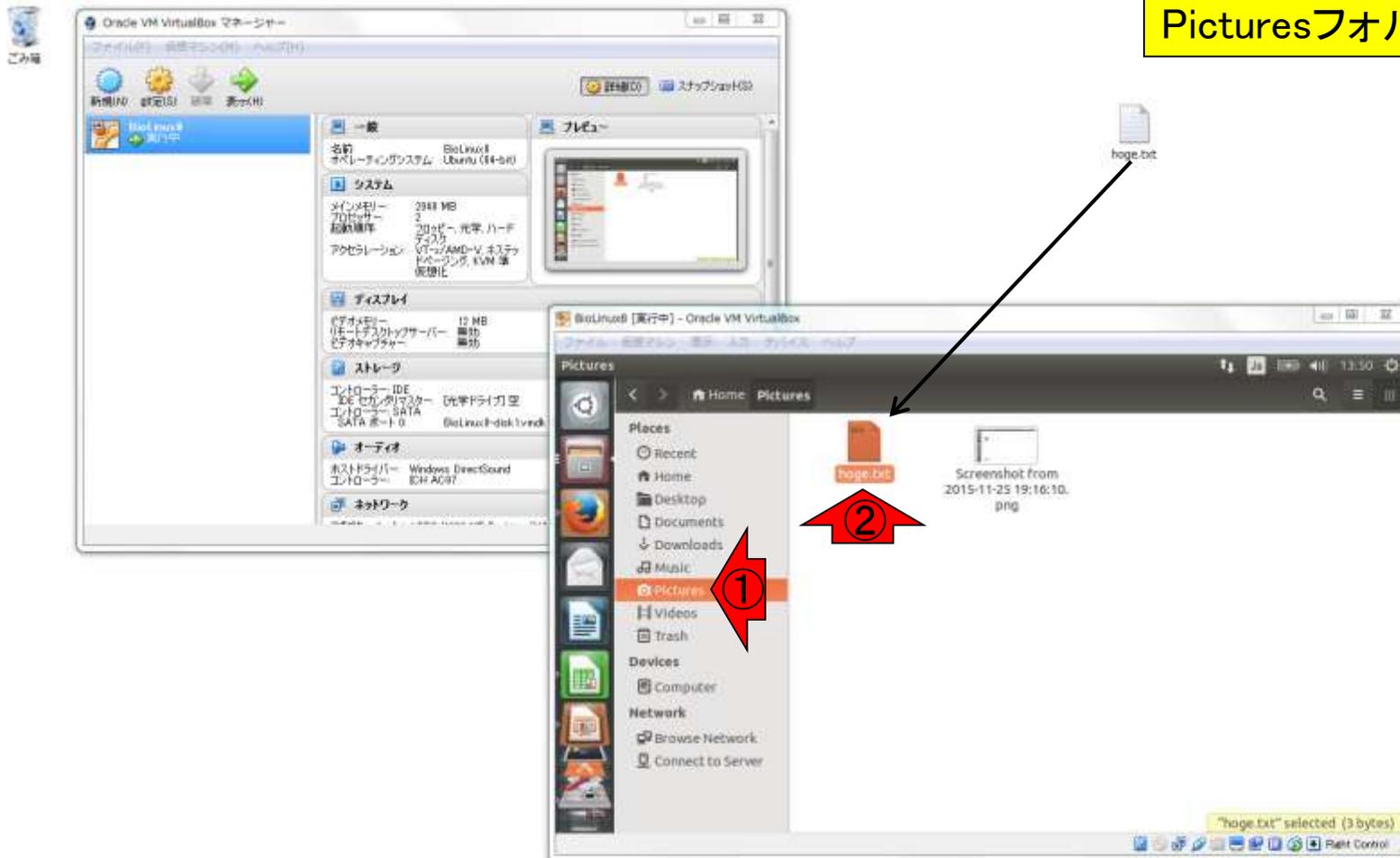


W8-2: ファイルの削除



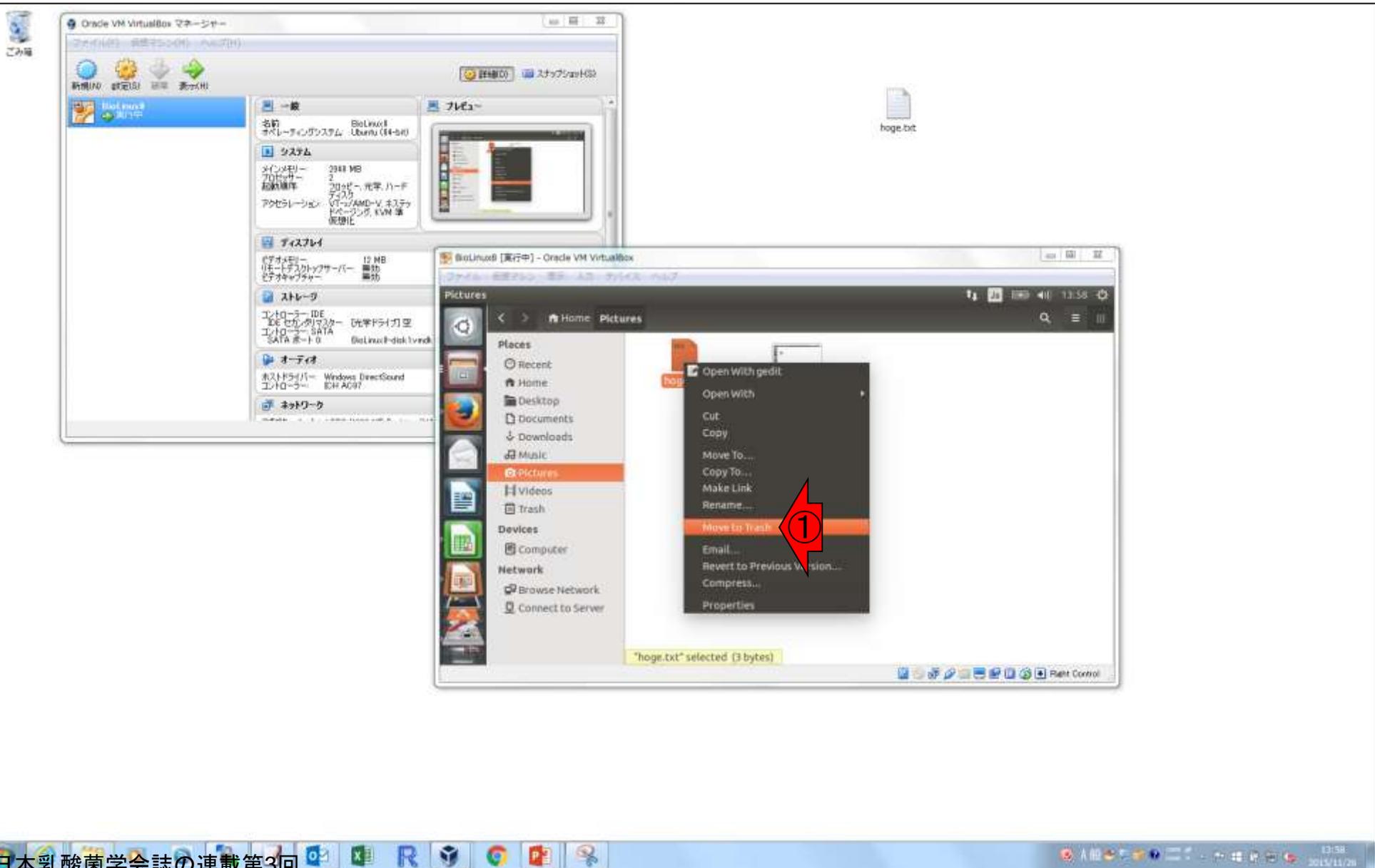
W8-2: ファイルの削除

①もう一度、さきほどのPicturesフォルダに移動して、②ホストOSのデスクトップにあるhoge.txtをゲストOSのPicturesフォルダにドラッグ&ドロップ



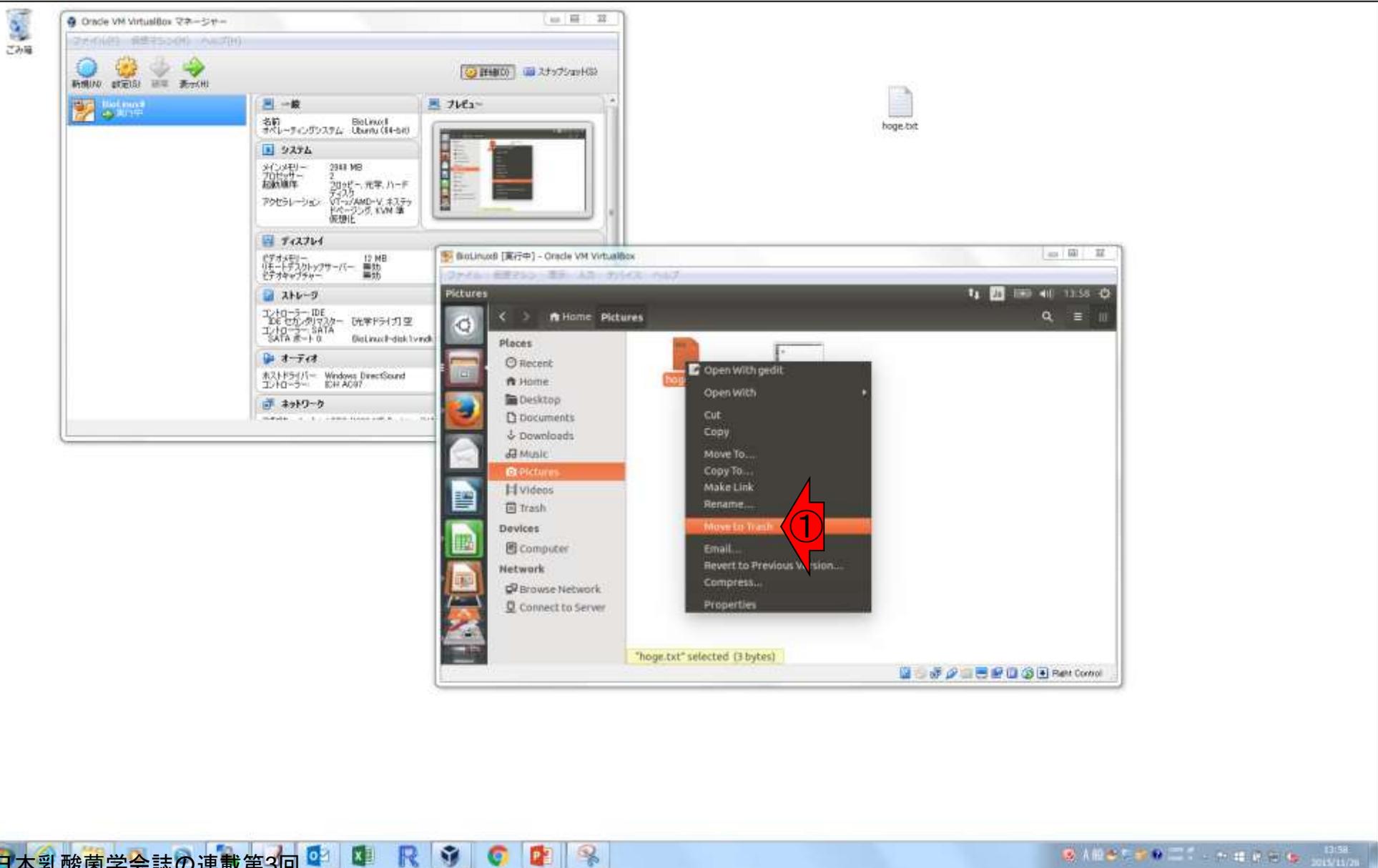
W8-2: ファイルの削除

削除したいファイルのアイコン上で右クリックして、①Move to Trashでもよい。



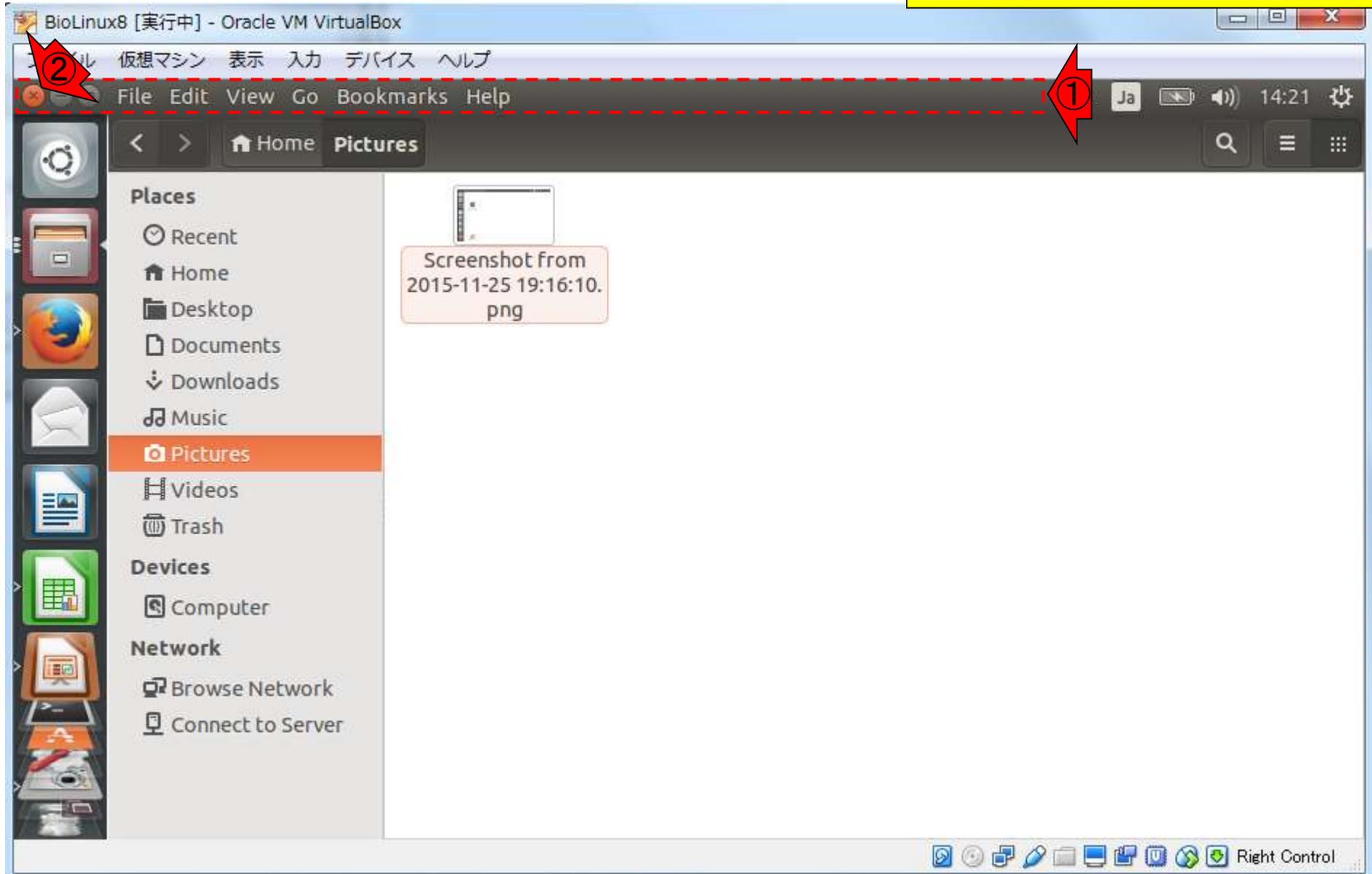
W8-3: ターミナル

削除したいファイルのアイコン上で右クリックして、①Move to Trashでもよい。



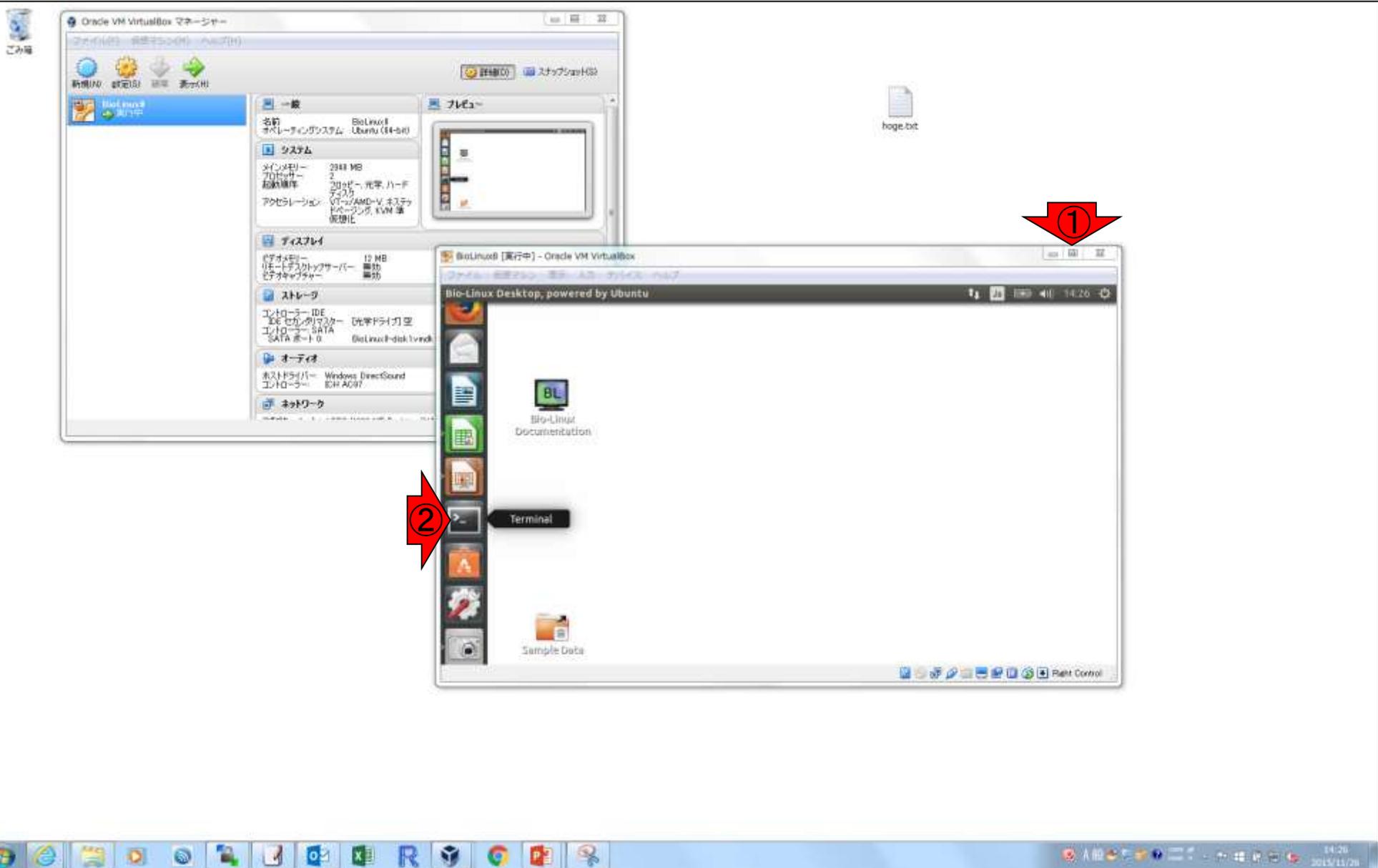
W8-3: ターミナル

ターミナルの説明に入る前に、①赤枠内にカーソル移動し、②一番左側の×ボタンを押してフォルダ画面を閉じる。



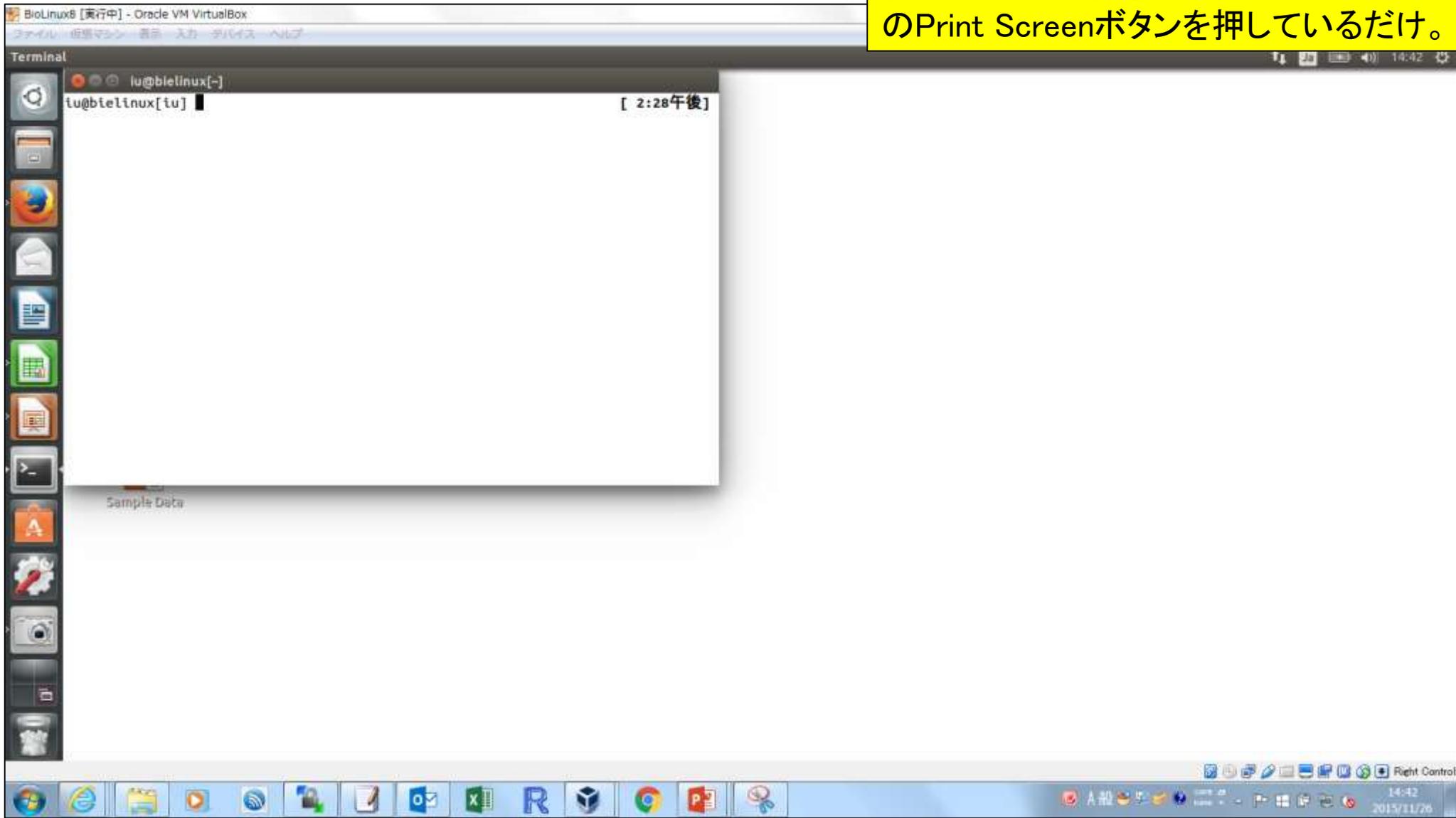
W8-3: ターミナル

①全画面表示に切り替えて、②ターミナルを起動。



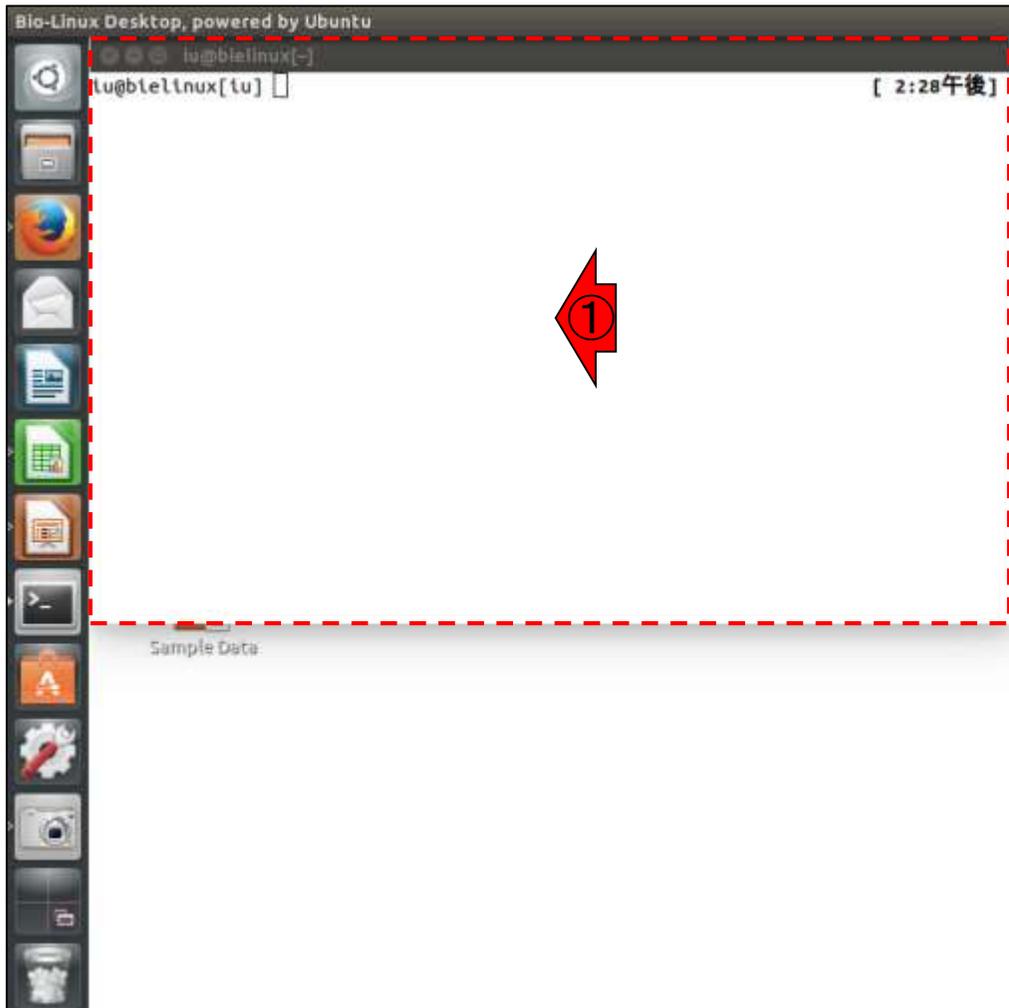
W8-3: ターミナル

ターミナル起動後の状態。ホストOSでのスクリーンショット。BioLinux8ウィンドウを非アクティブ状態にして、キーボードのPrint Screenボタンを押しているだけ。



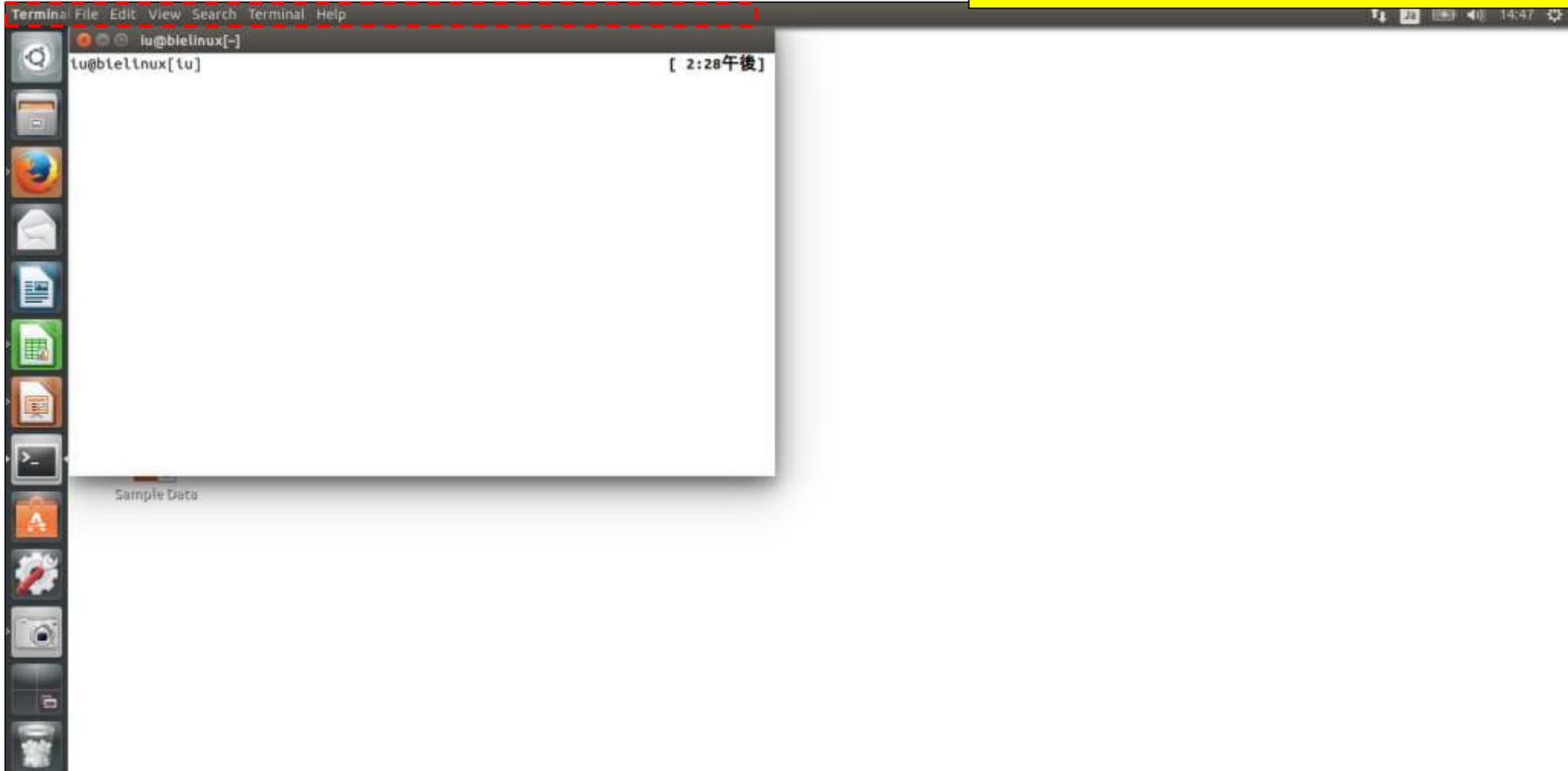
W8-4: ターミナル設定

ターミナル起動後の状態。ゲストOSでのスクリーンショット。キーボードのPrint Screenボタンを押しているだけ。ボタンを押してもスクリーンショットがとれないこともあるがそんなものです。①赤枠内でクリックし、ターミナルをアクティブにする。

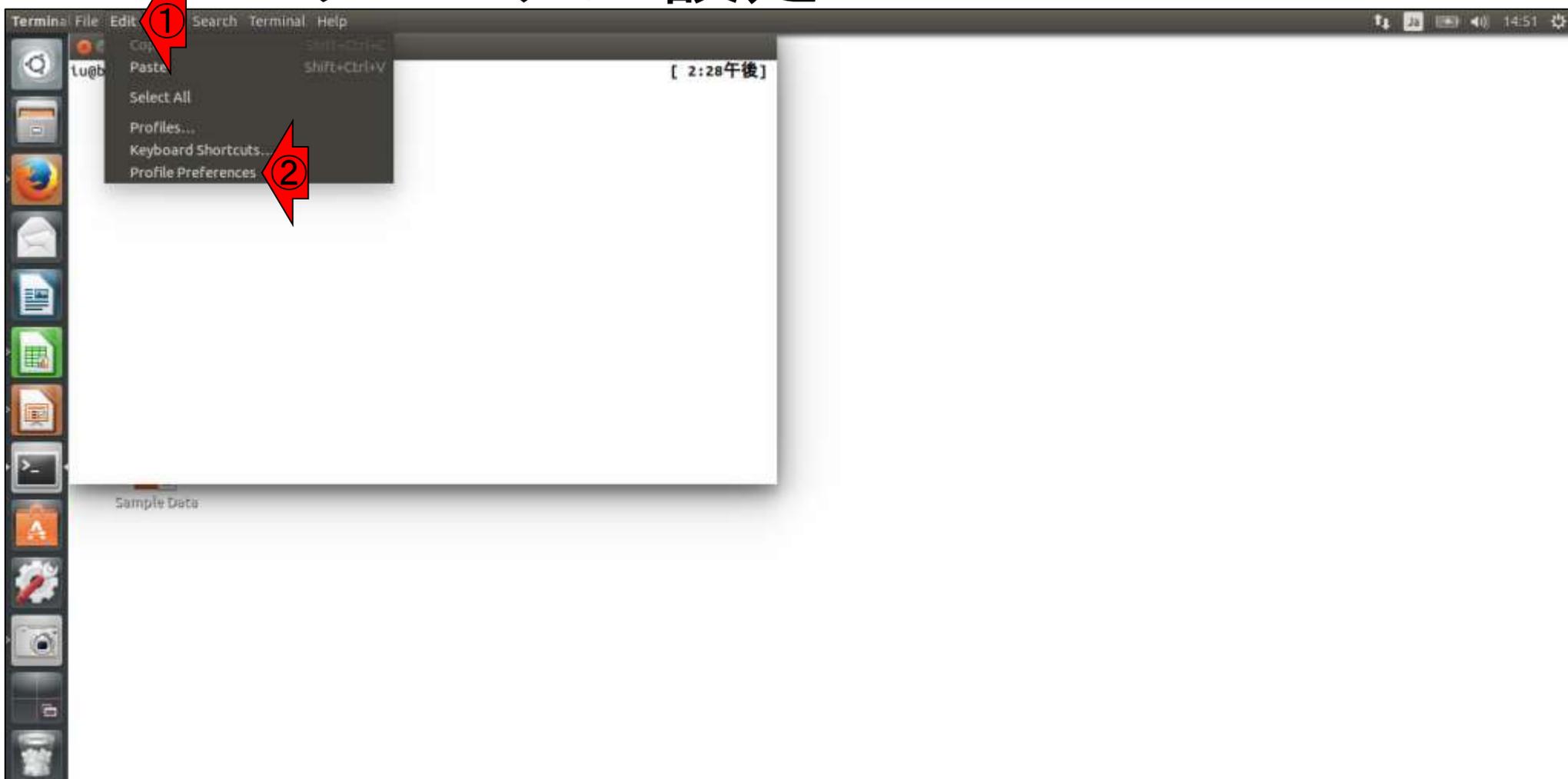


W8-4: ターミナル設定

ターミナルの文字の大きさを変える作業を行う。①赤枠部分にカーソル移動させるとメニューが見られるようになる

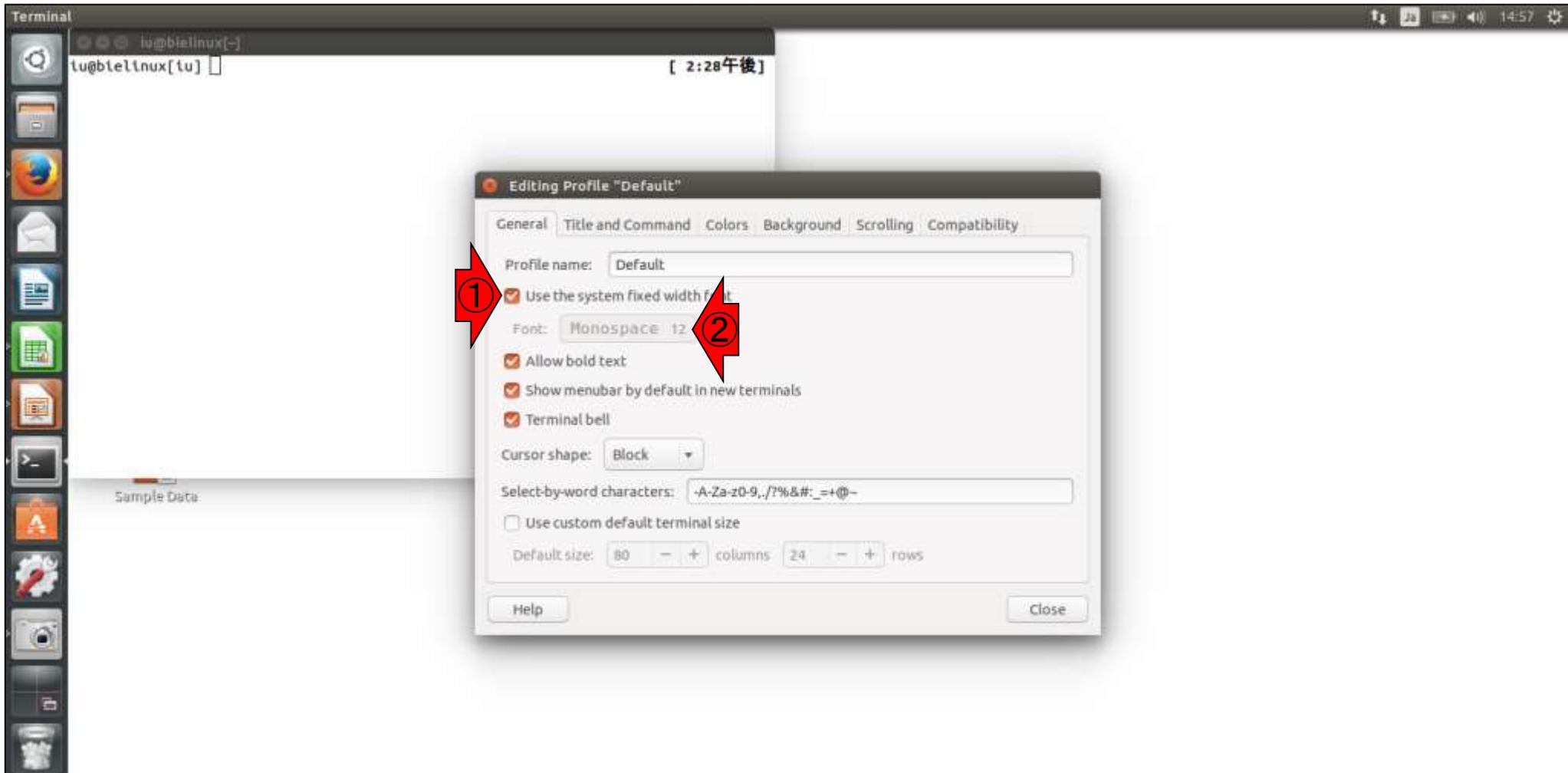


W8-4:ターミナル設定



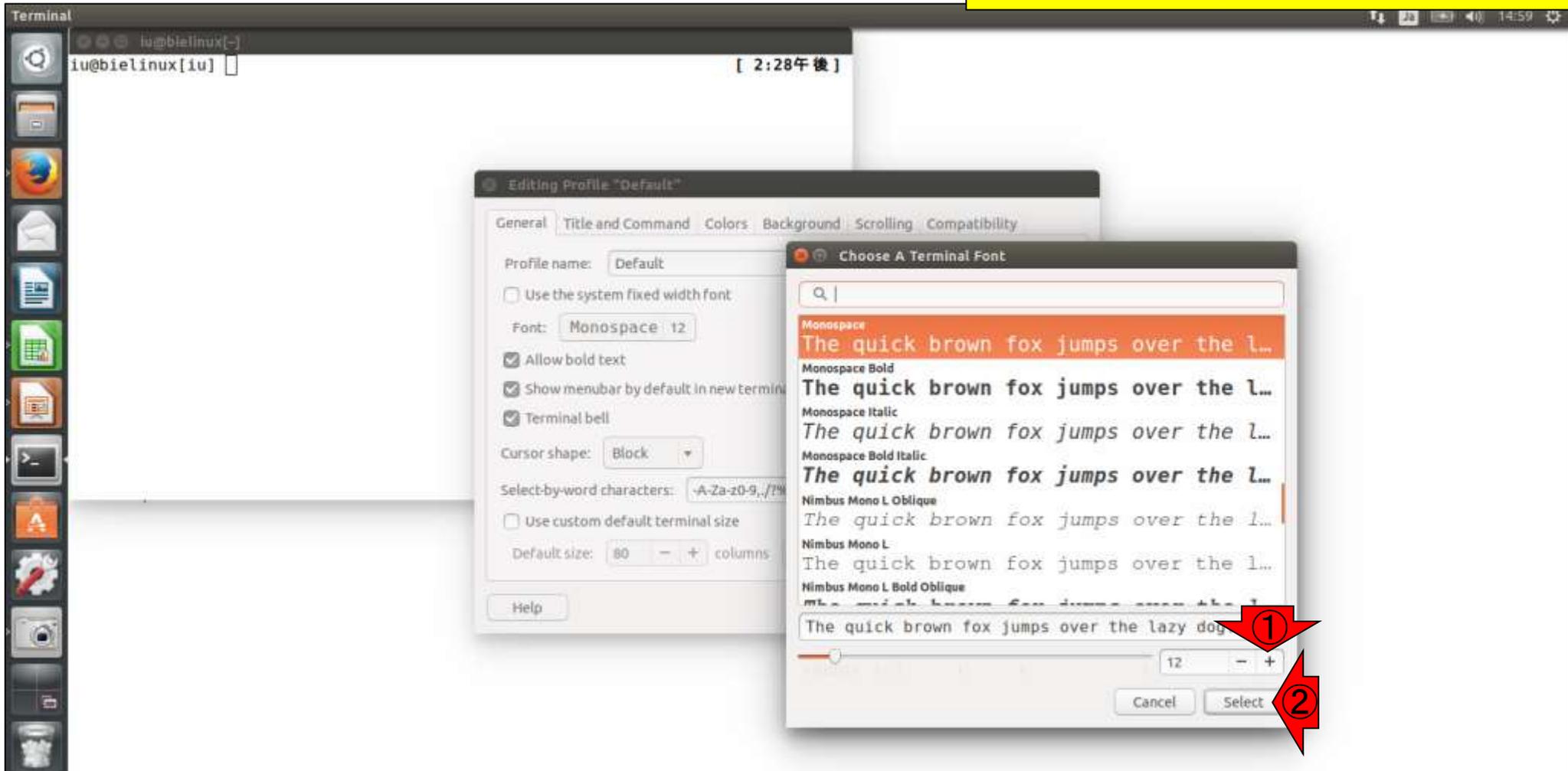
W8-4: ターミナル設定

①「Use the system fixed width …」のチェックを外し、②「Monospace 12」のあたりをクリック



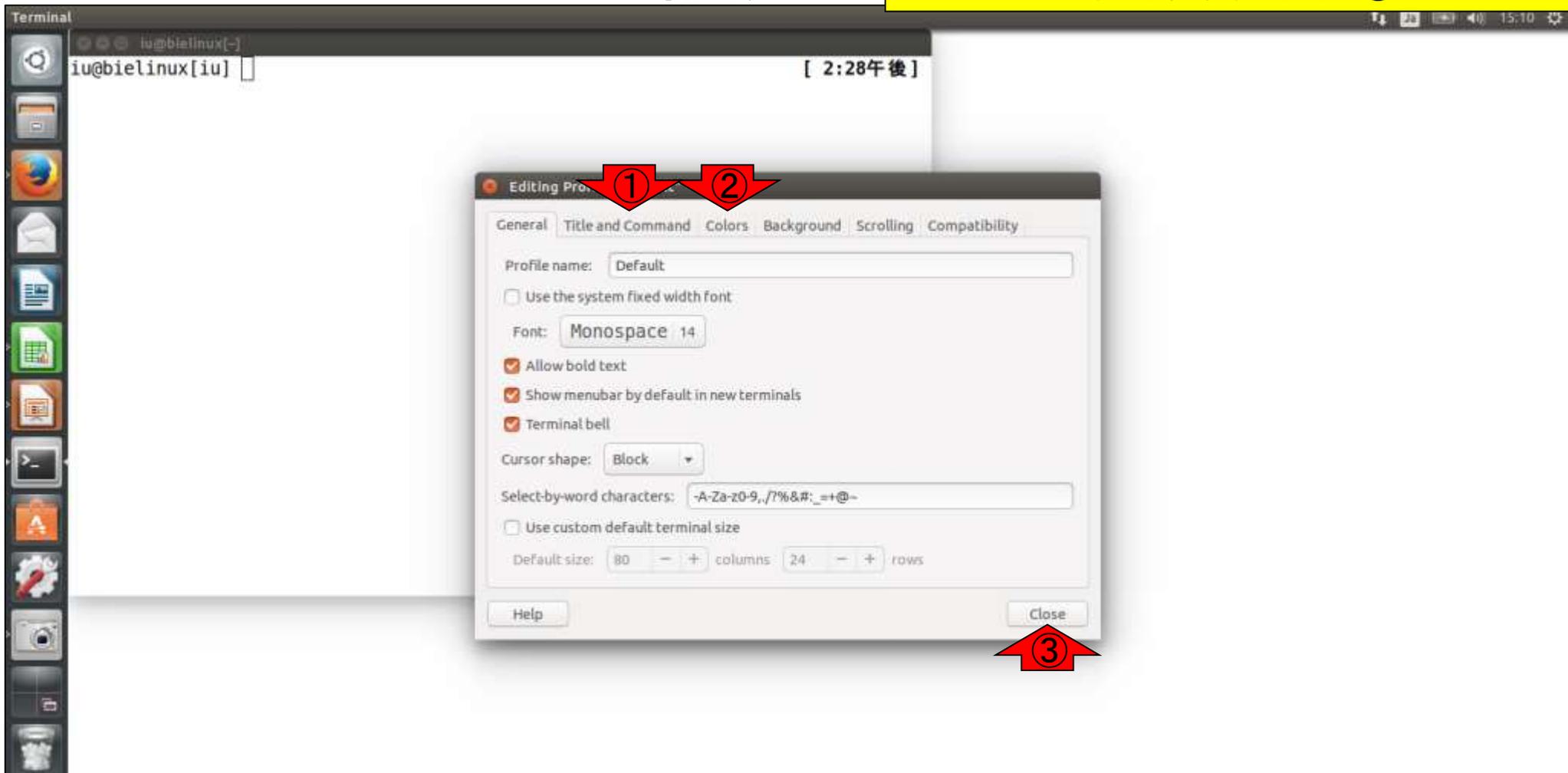
W8-4: ターミナル設定

フォントと文字の大きさを選択できます。ここでは、①+を2回押して大きさを14にしてから、②Selectボタンを押す



W8-4: ターミナル設定

確かに文字が大きくなったことがわかります。他にも①Title and Commandや②Colorsなどいくつか設定変更できる。③Close。



ターミナル画面以外にも、Bio-Linux本体の設定も変更できます。①System Settings、②Appearance。

W8-5: その他設定



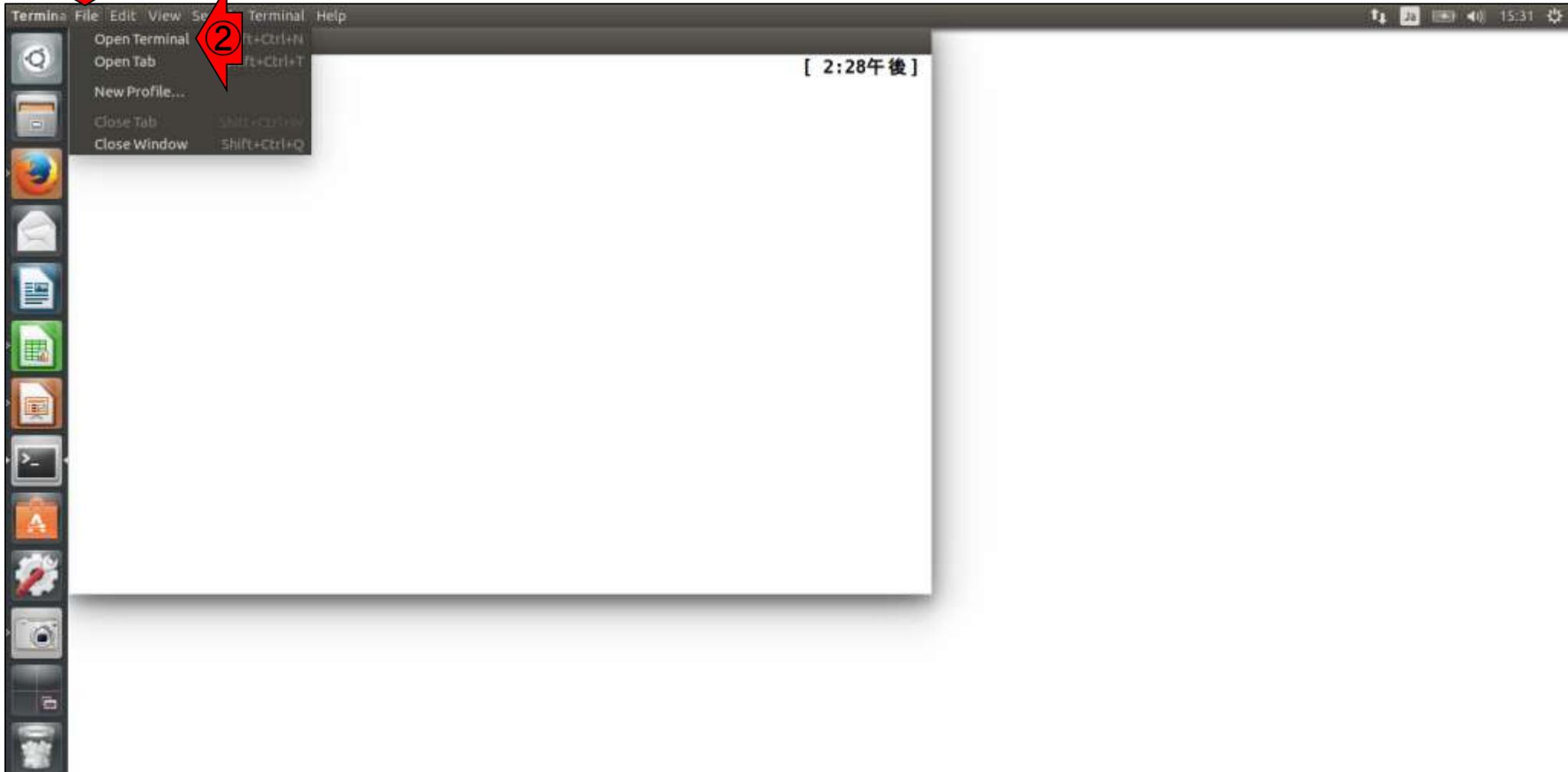
W8-5: その他設定

①の部分などをいじって、いろいろ変更可能。連載第2回の「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」の最後のほうにもTipsとして載せています。②一旦閉じる。



①File、②Open Terminalで新しいターミナルを開くことができる

W8-6: 新規ターミナル



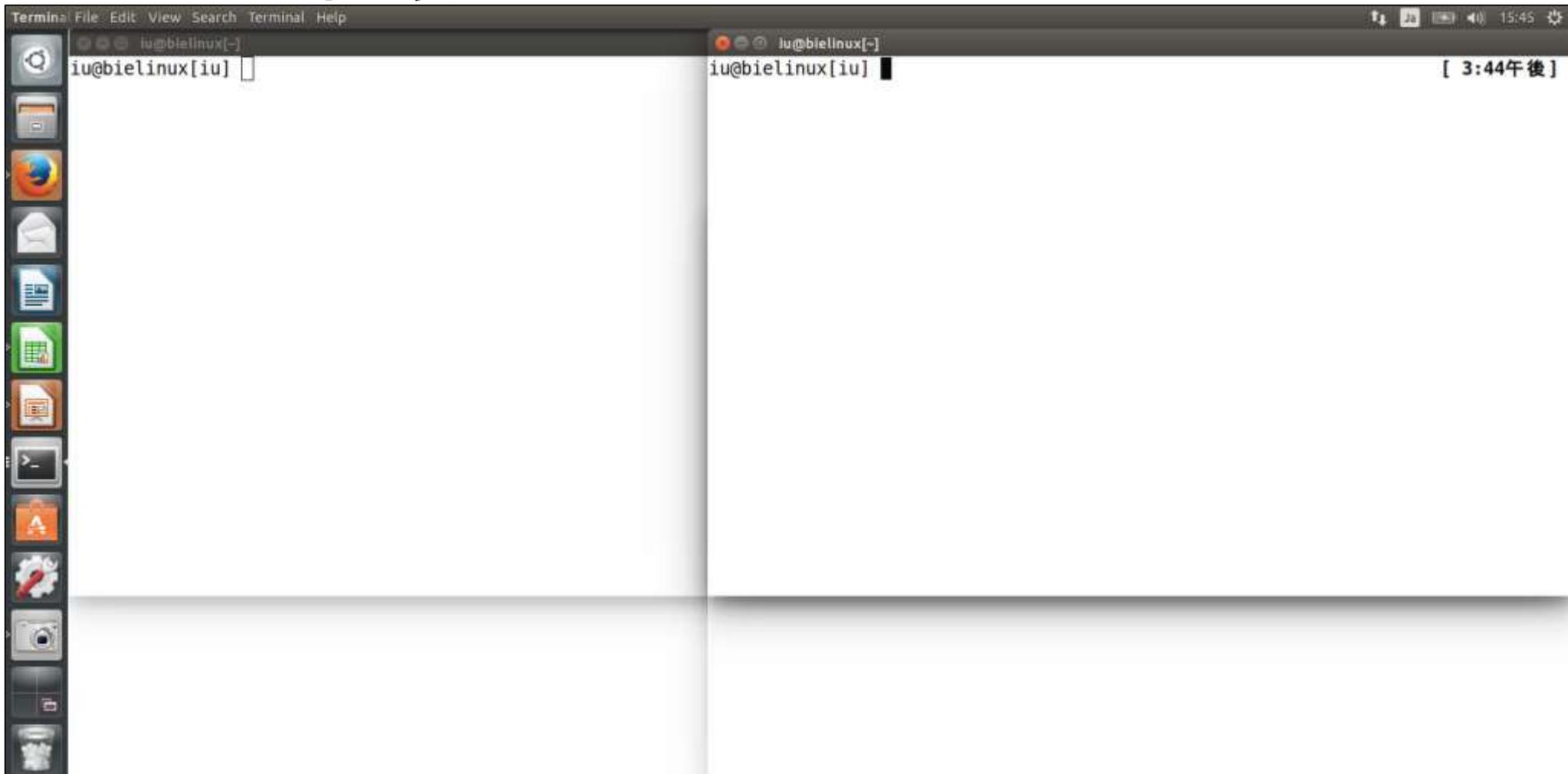
W8-6: 新規ターミナル

①ターミナルアイコン上で右クリックし、②New Terminalでもよい



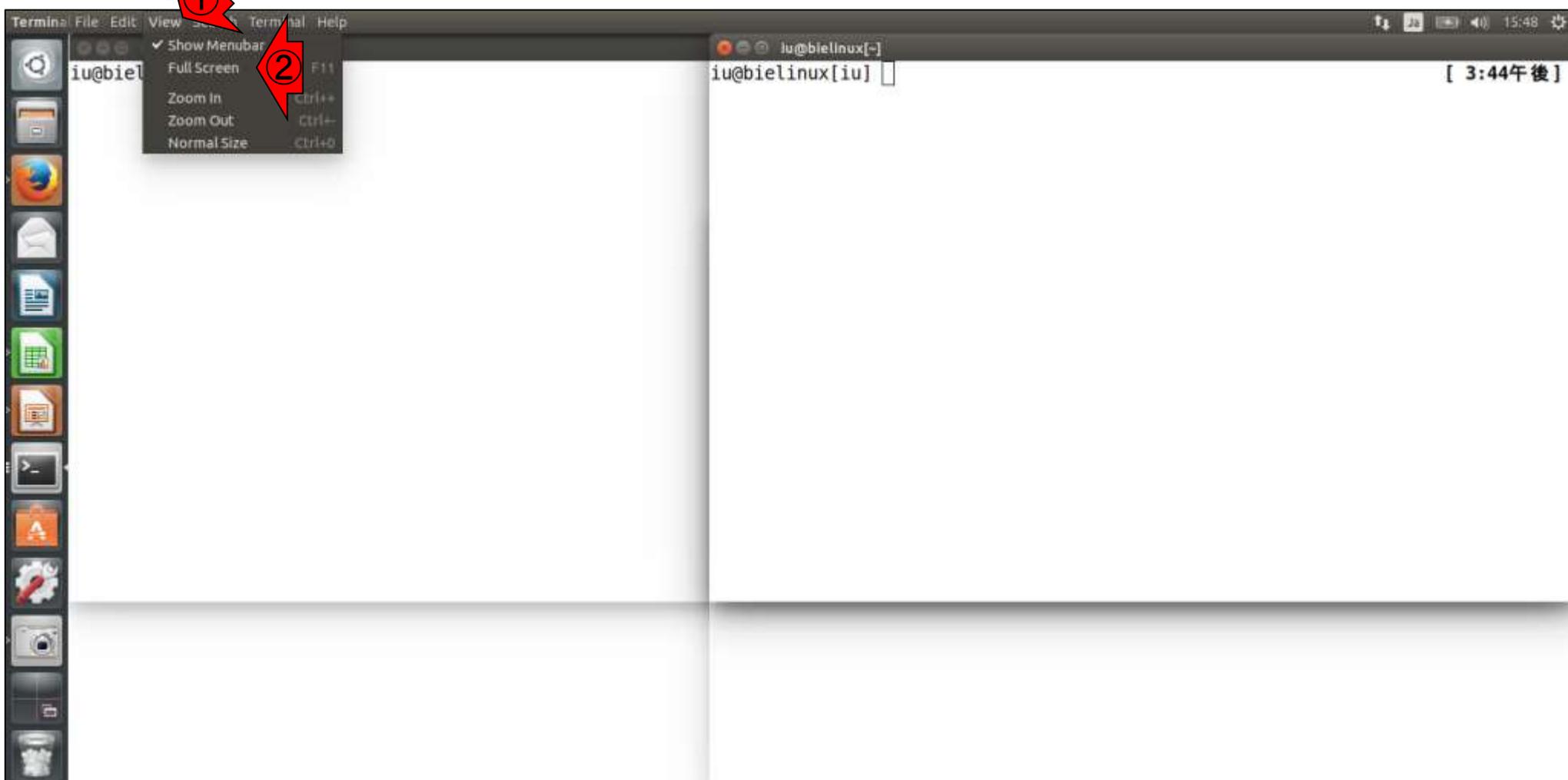
W8-6: 新規ターミナル

このように複数のターミナルを同時に開いて作業を行うのが一般的



メニューバーをいろいろいじってみる。例えば①View、②Full Screen

W8-7: メニューバー



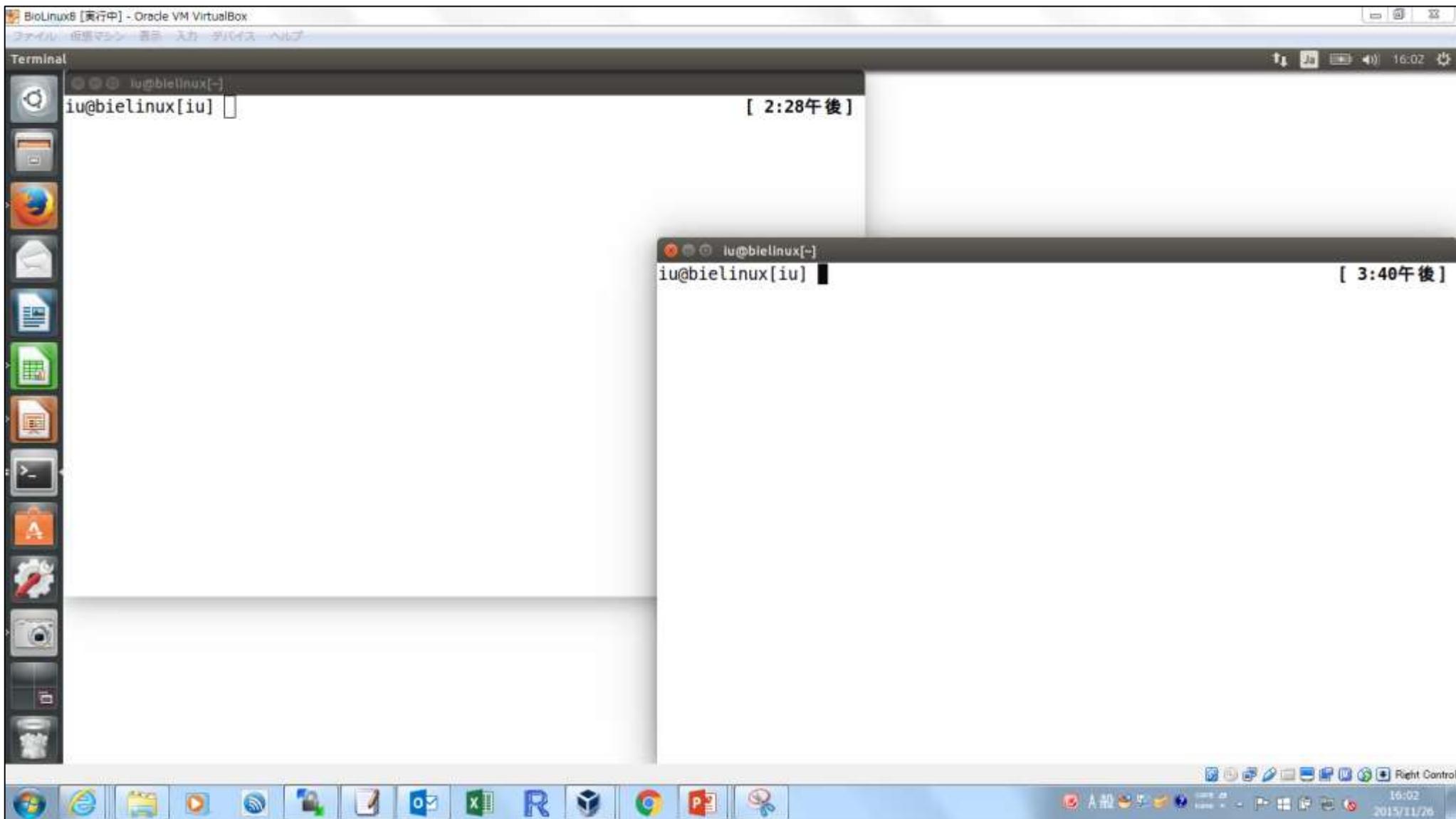
W8-7: メニューバー

これはホストOS上でスクリーンショットをとったもの。このような状況からどのようにして元に戻せばいいのか?1つの対処法は、右クリックで①Close Window



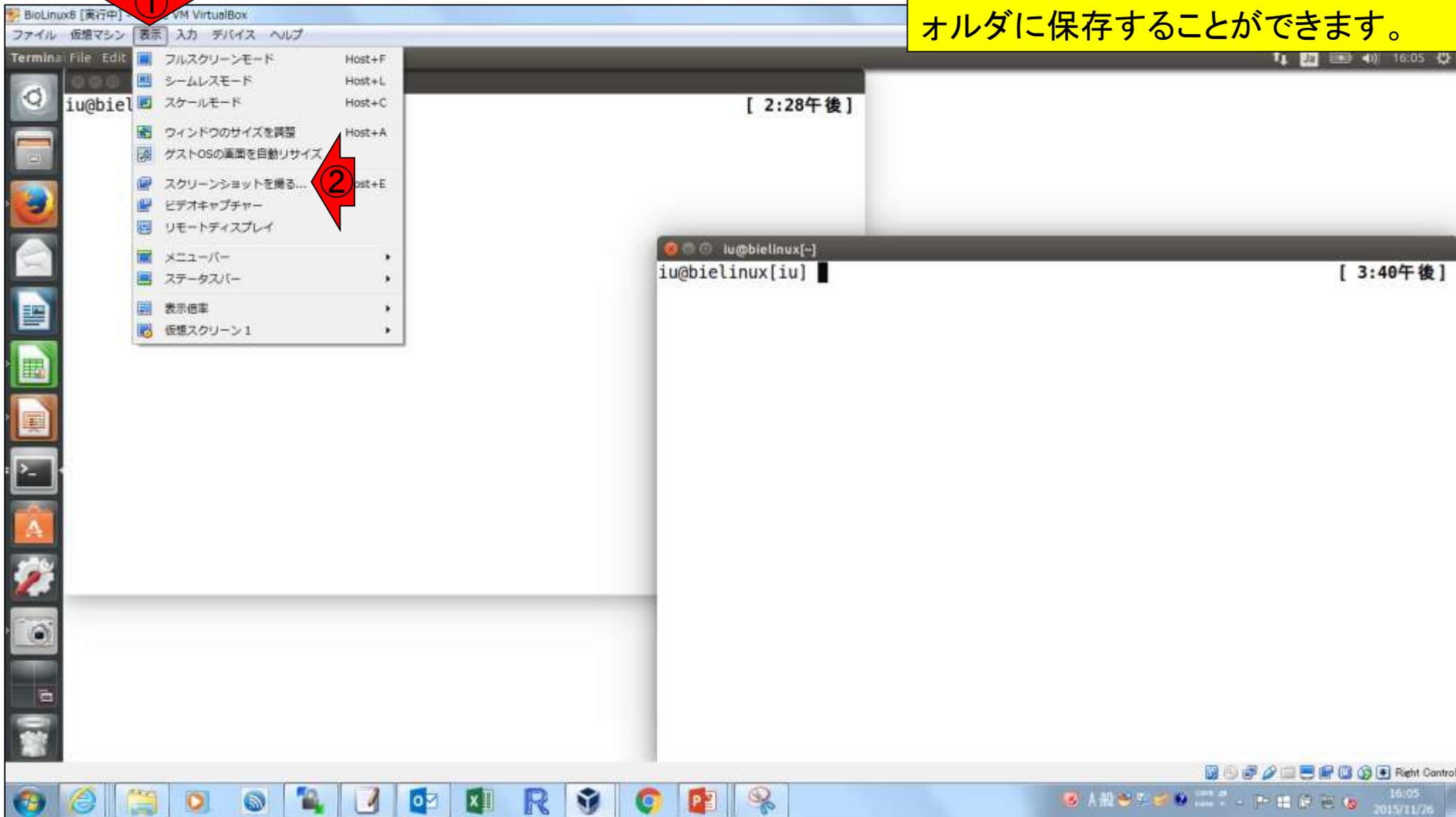
W8-7: メニューバー

該当ターミナルはクローズされますが、一応元の画面に戻ります。



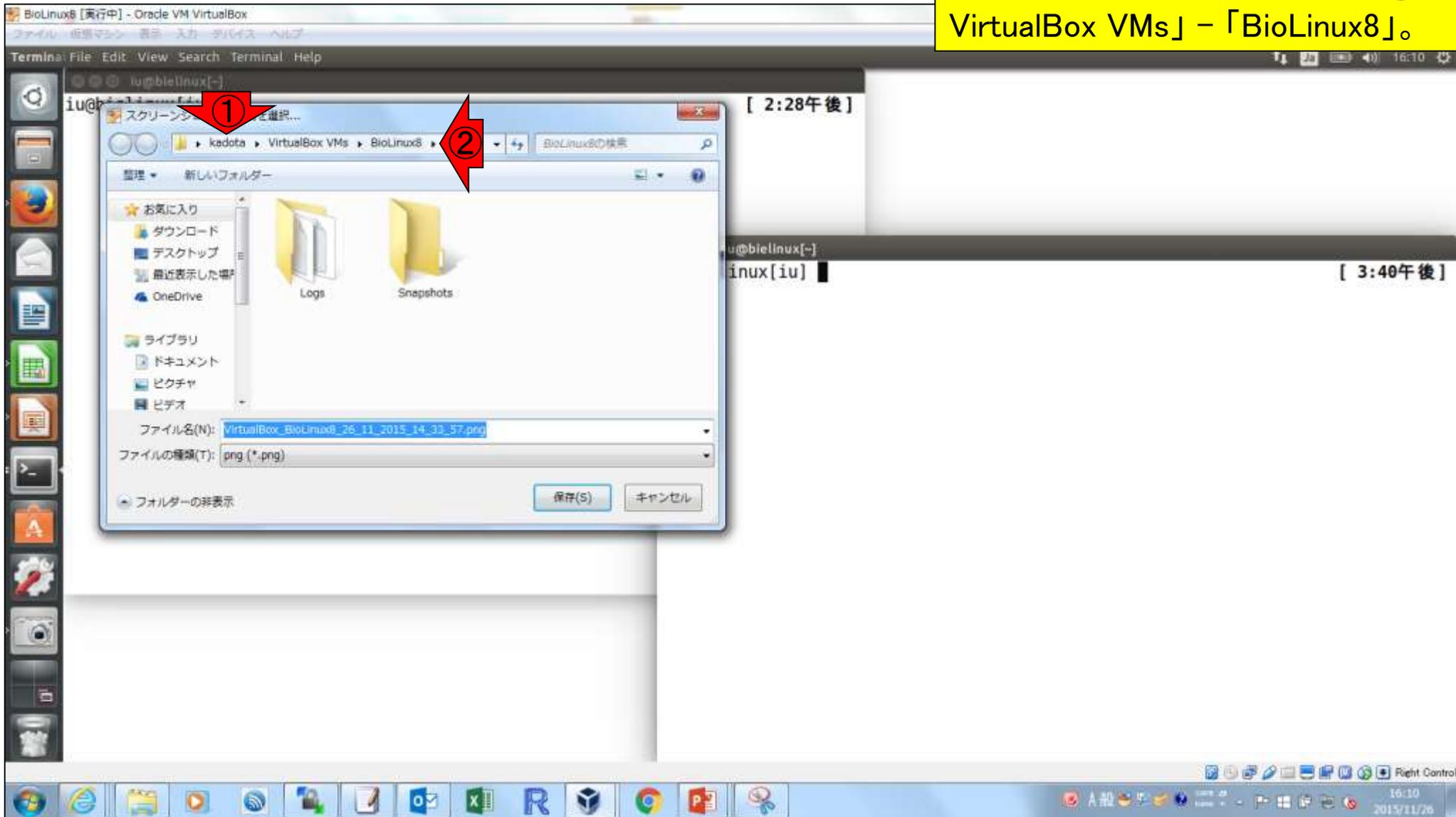
W8-8: スクリーンショット

①表示、②スクリーンショットを撮る。
ゲストOSのスクリーンショットのpng形式ファイルをホストOS中の任意のフォルダに保存することができます。



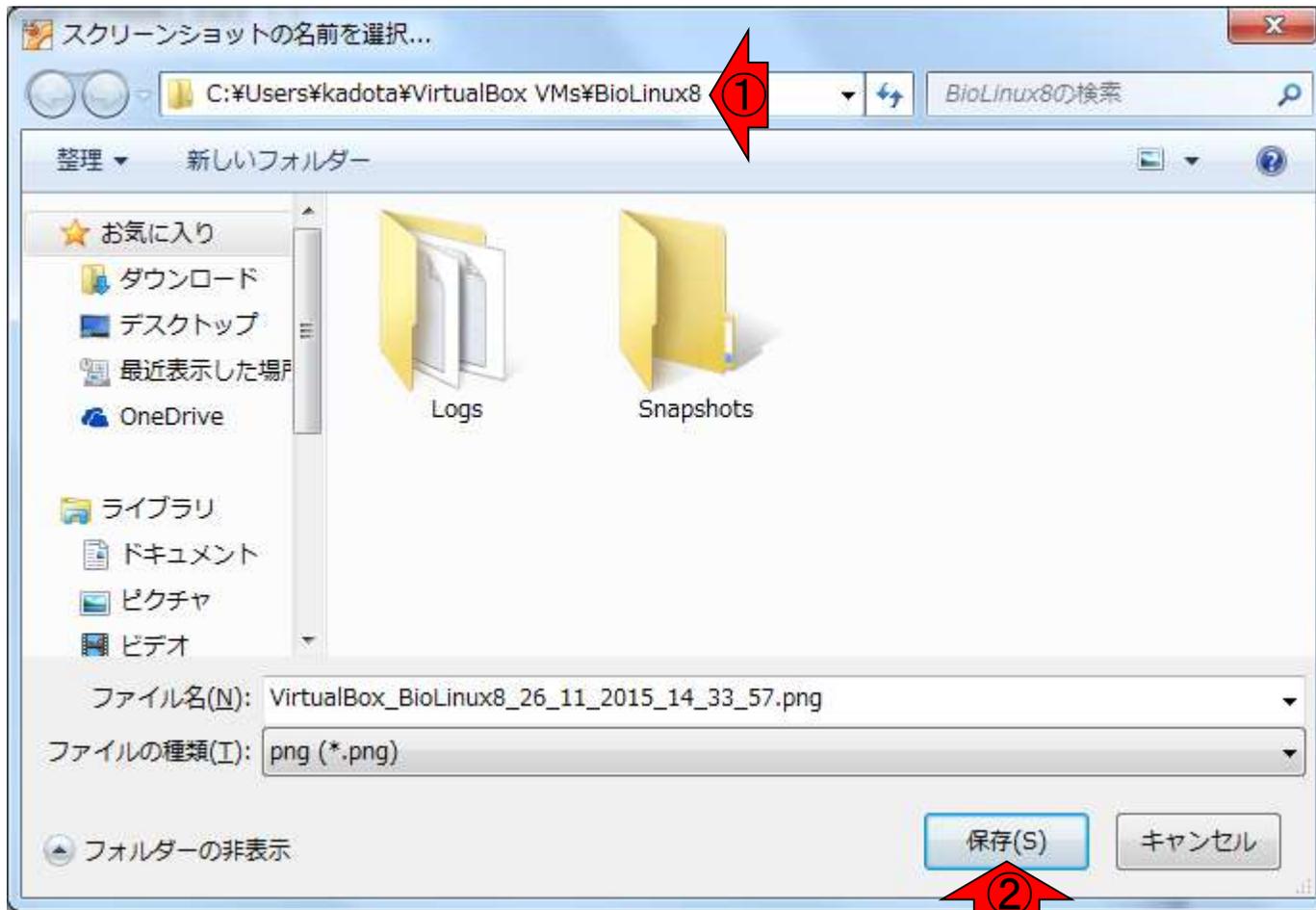
W8-8: スクリーンショット

デフォルトの保存先は、ホストOSのログインユーザ(この場合①kadota)の個人用フォルダ直下にある②「VirtualBox VMs」 - 「BioLinux8」。



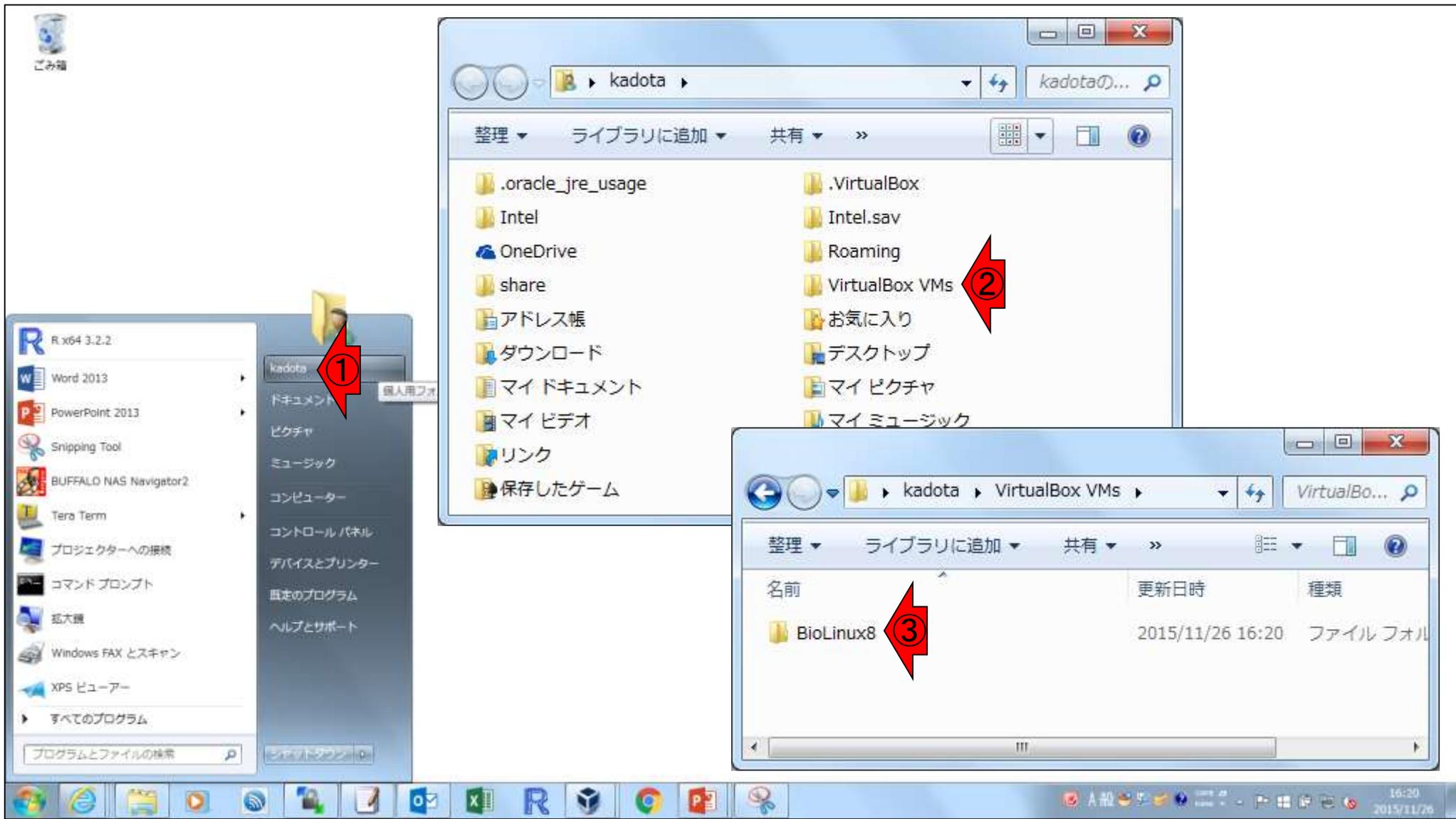
W8-8: スクリーンショット

①をクリックするとフルパスがわかる。②保存。



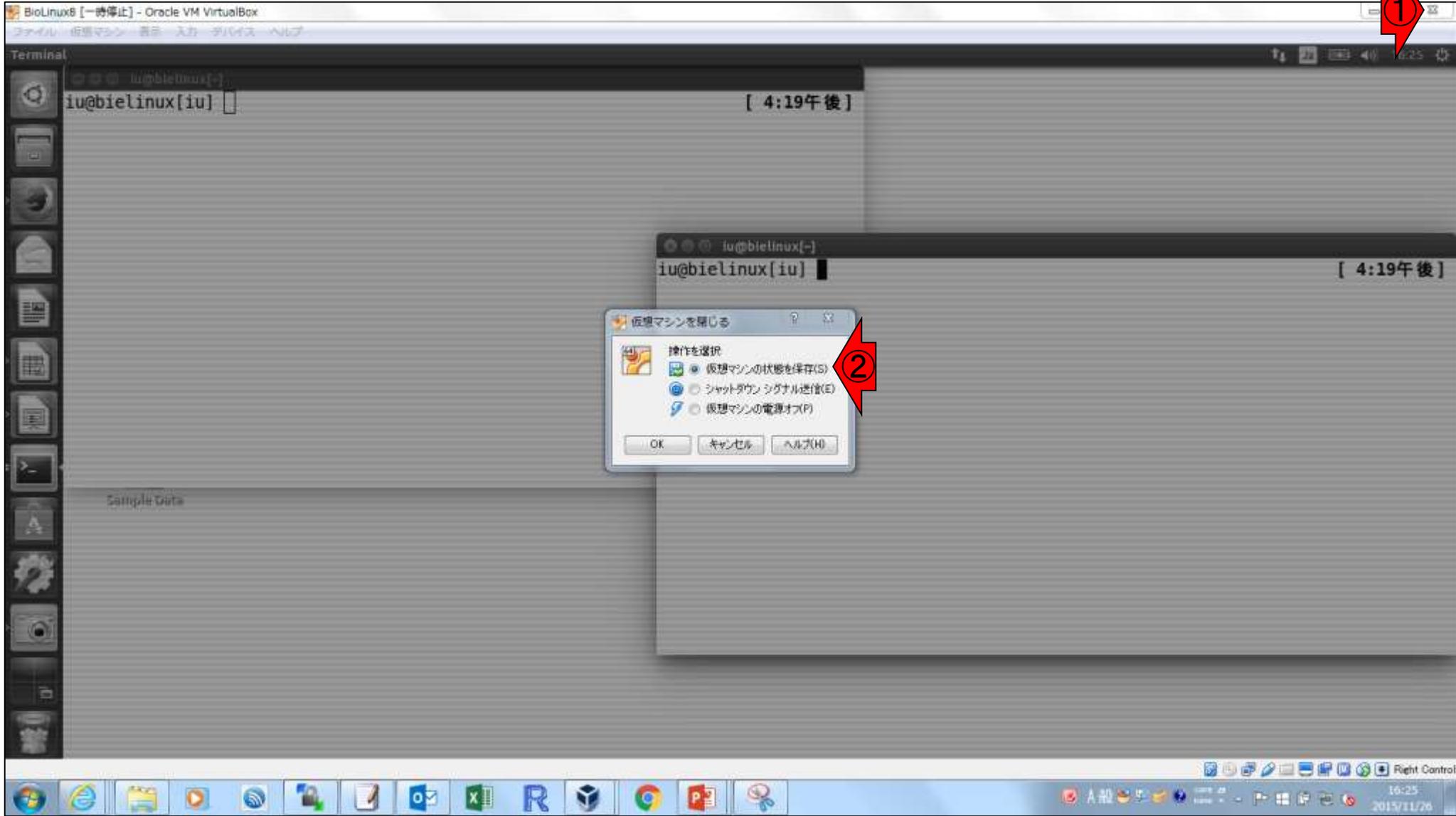
保存したpngファイルは、
以下のようにして送れます

W8-8: スクリーンショット



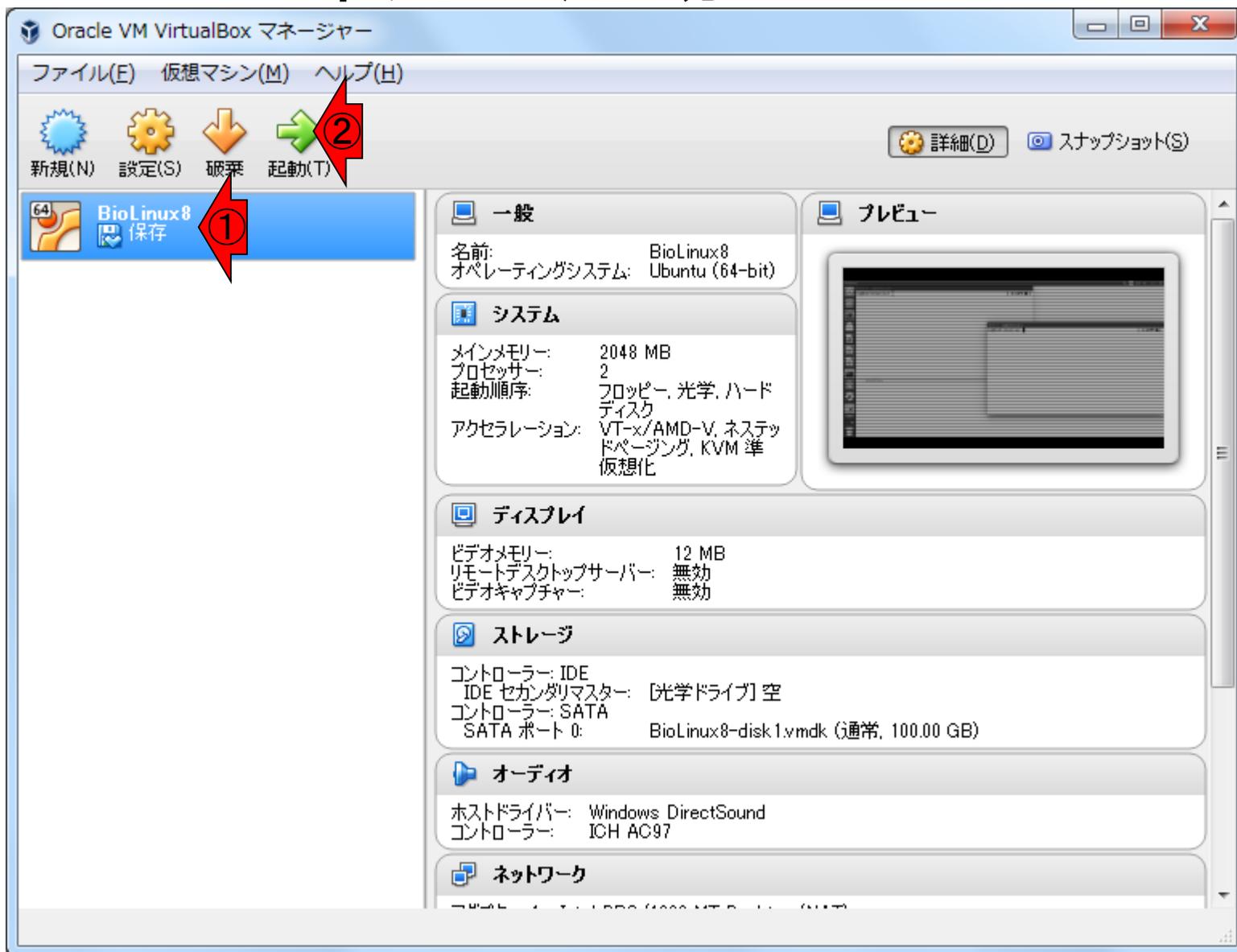
①ゲストOSの終了時に②仮想マシンの状態を保存で、次回起動時に続きから作業することができる

W8-9: 終了と起動



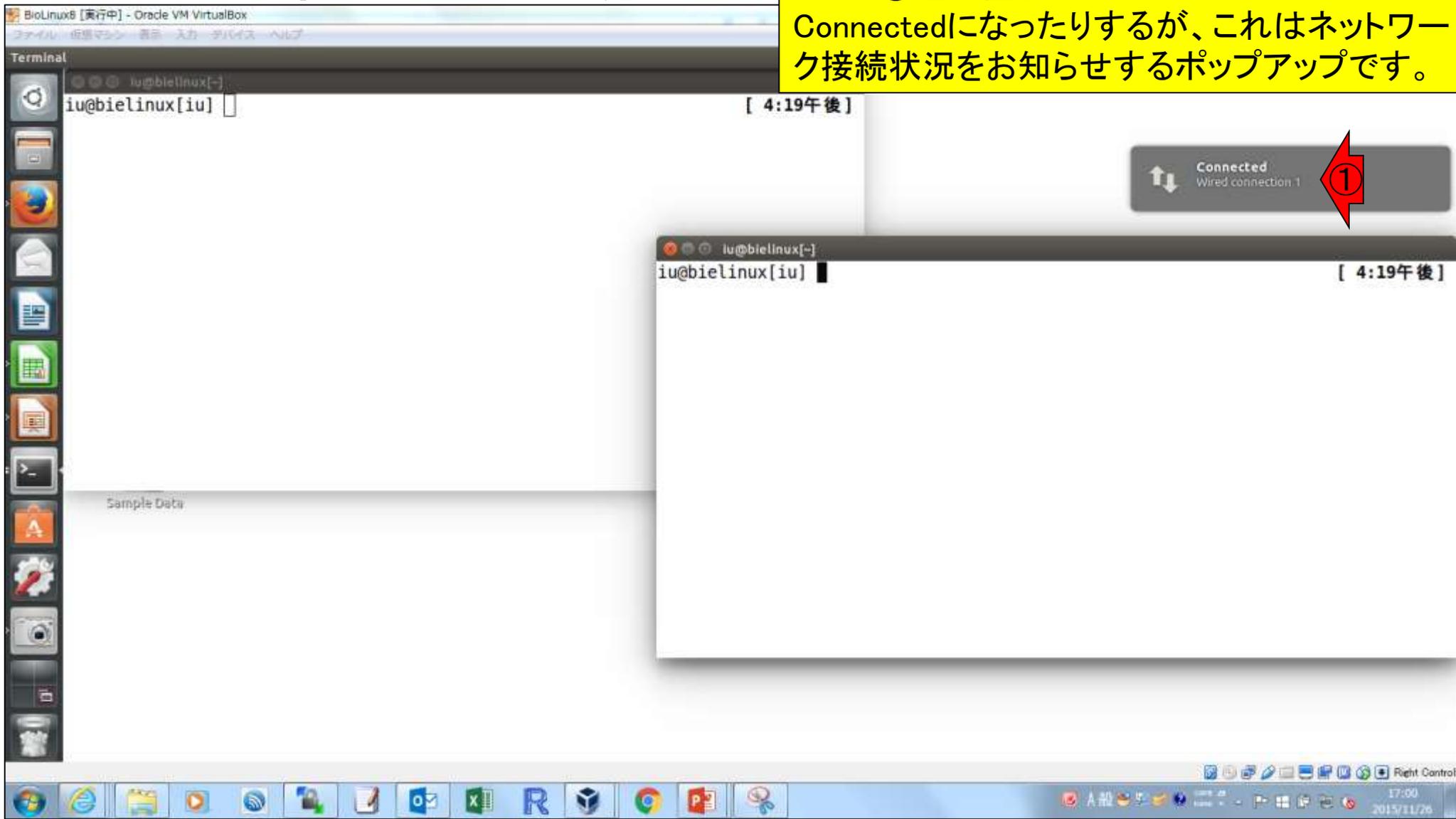
W8-9: 終了と起動

ゲストOS終了直後のVirtualBoxマネージャー。
①保存、になっていることがわかる。②起動。



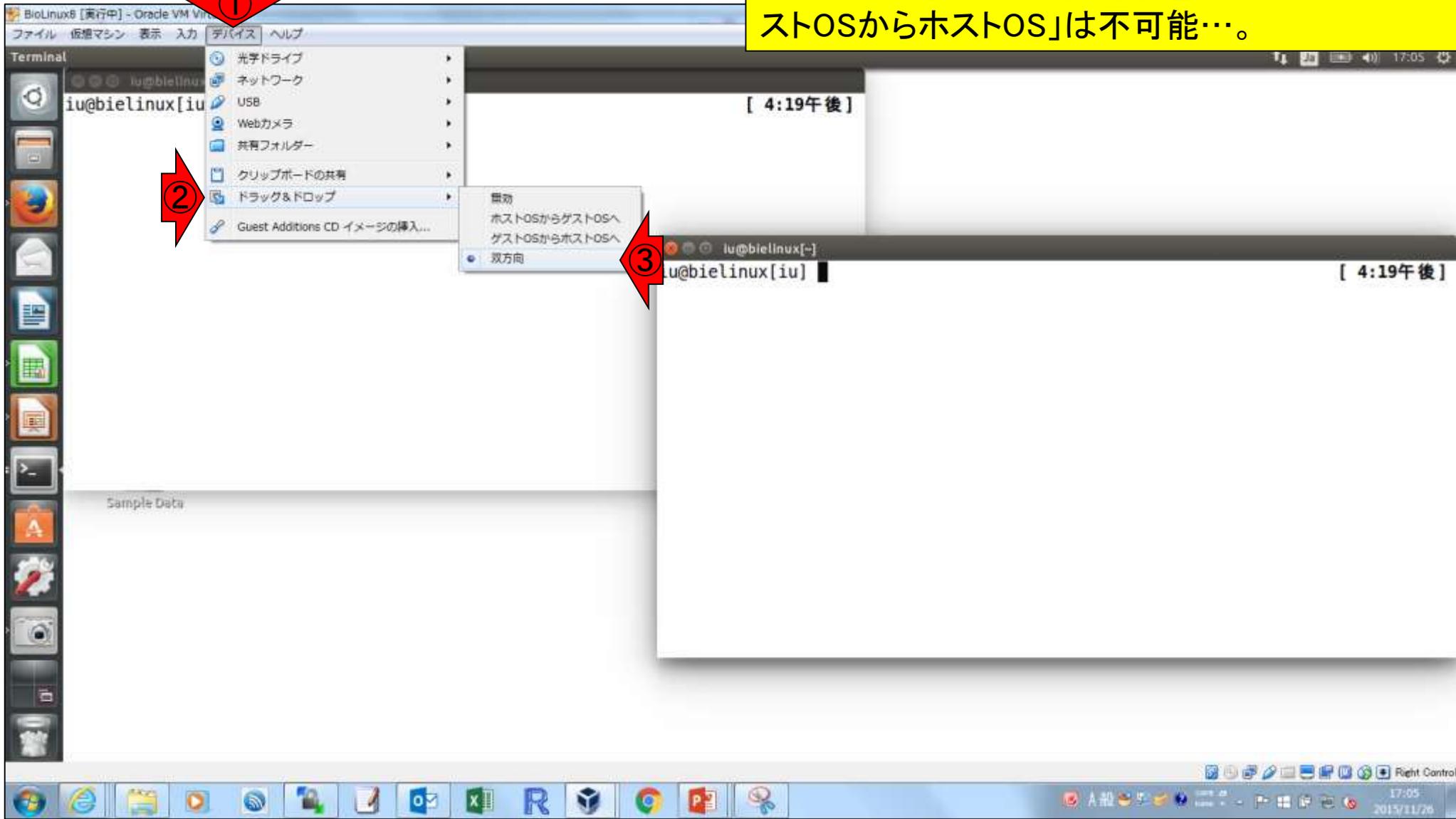
W8-9: 終了と起動

起動後の状態。終了前と同じ(全画面表示でターミナルが2つ立ち上がった)であることがわかる。①起動直後にDisconnectedだったり、Connectedになったりするが、これはネットワーク接続状況をお知らせするポップアップです。



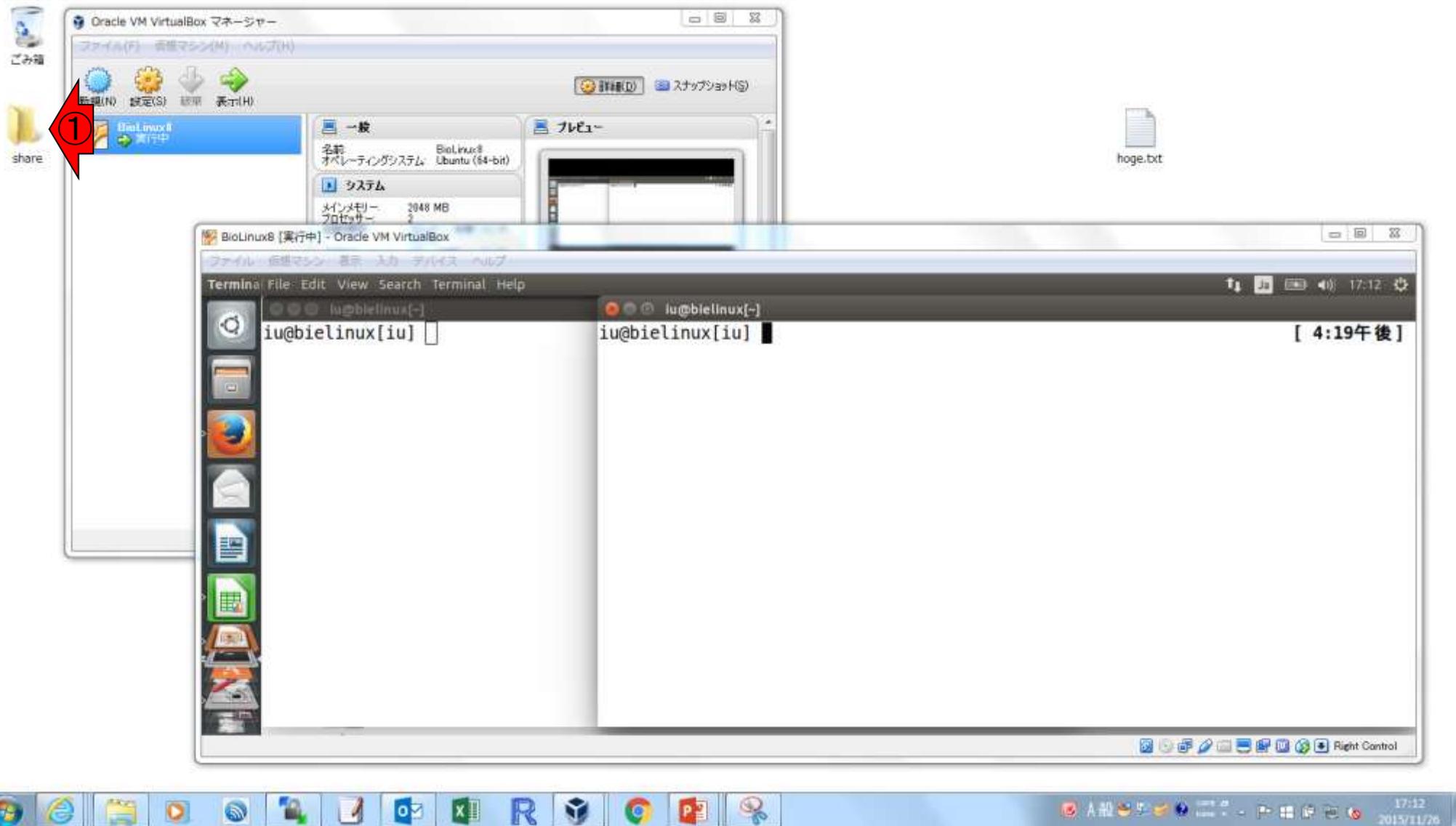
W9-1: 共有フォルダ

VirtualBoxウィンドウメニューの①デバイス、②ドラッグ&ドロップでも設定変更可能。VirtualBox ver. 5.0.10現在、「双方向」を選択していても「ゲストOSからホストOS」は不可能…。

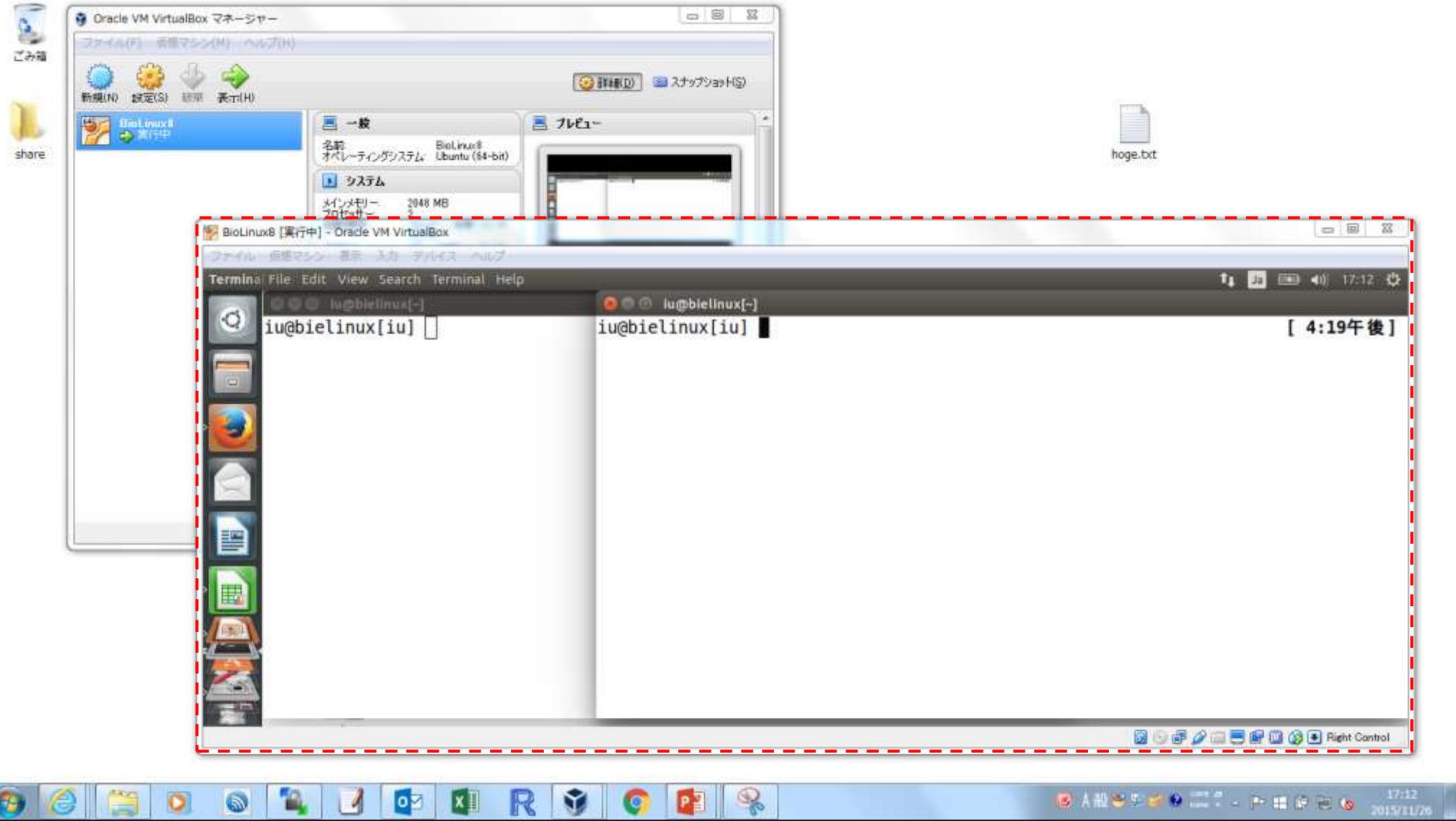


W9-2-1: 共有フォルダ

共有フォルダの設定を行う。ホストOSのデスクトップ上に①shareフォルダを新規作成

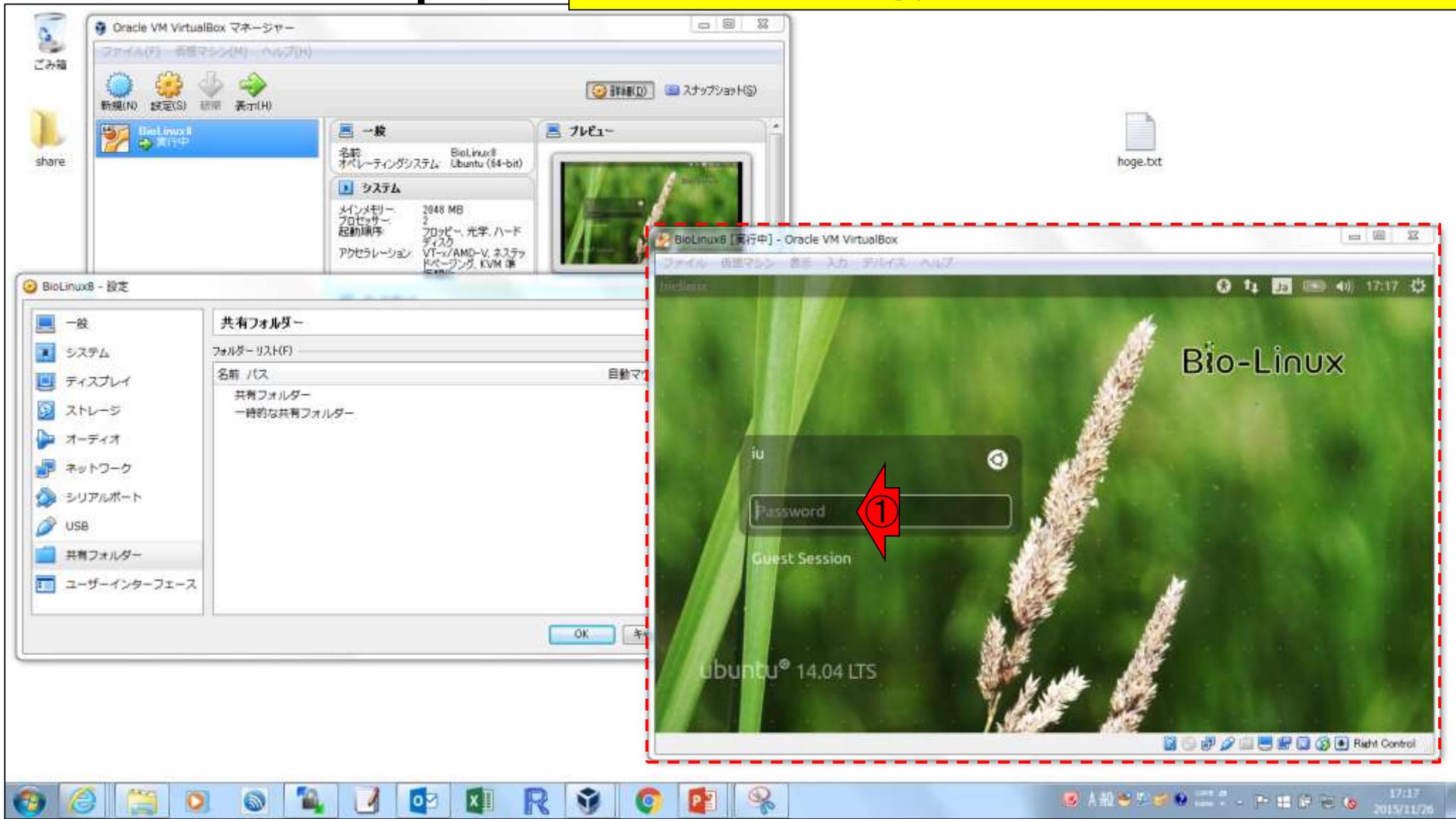


W9-2-1 : Tips



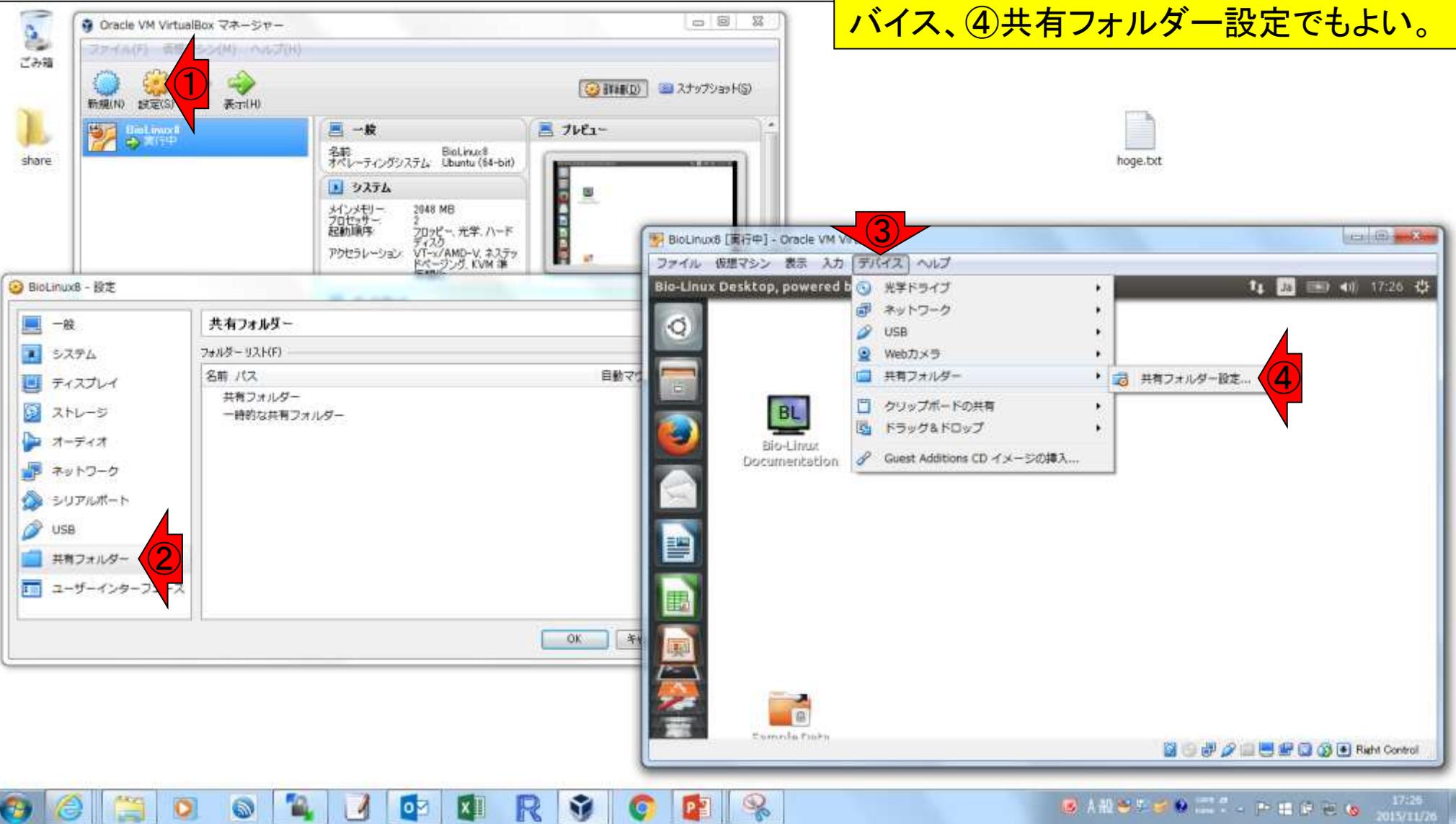
W9-2-1 : Tips

こんな具合に、保存しないで終了し、勝手にログイン画面に戻ってしまうことも多々ありますので注意しましょう。こういうときは、気を取り直して①pass1409と打ち込んで再度ログイン



W9-2-2: 共有フォルダ

VirtualBoxマネージャーウィンドウの①設定、②共有フォルダ。別の手段として、ゲストOS(BioLinux8)ウィンドウ中で③デバイス、④共有フォルダ設定でもよい。



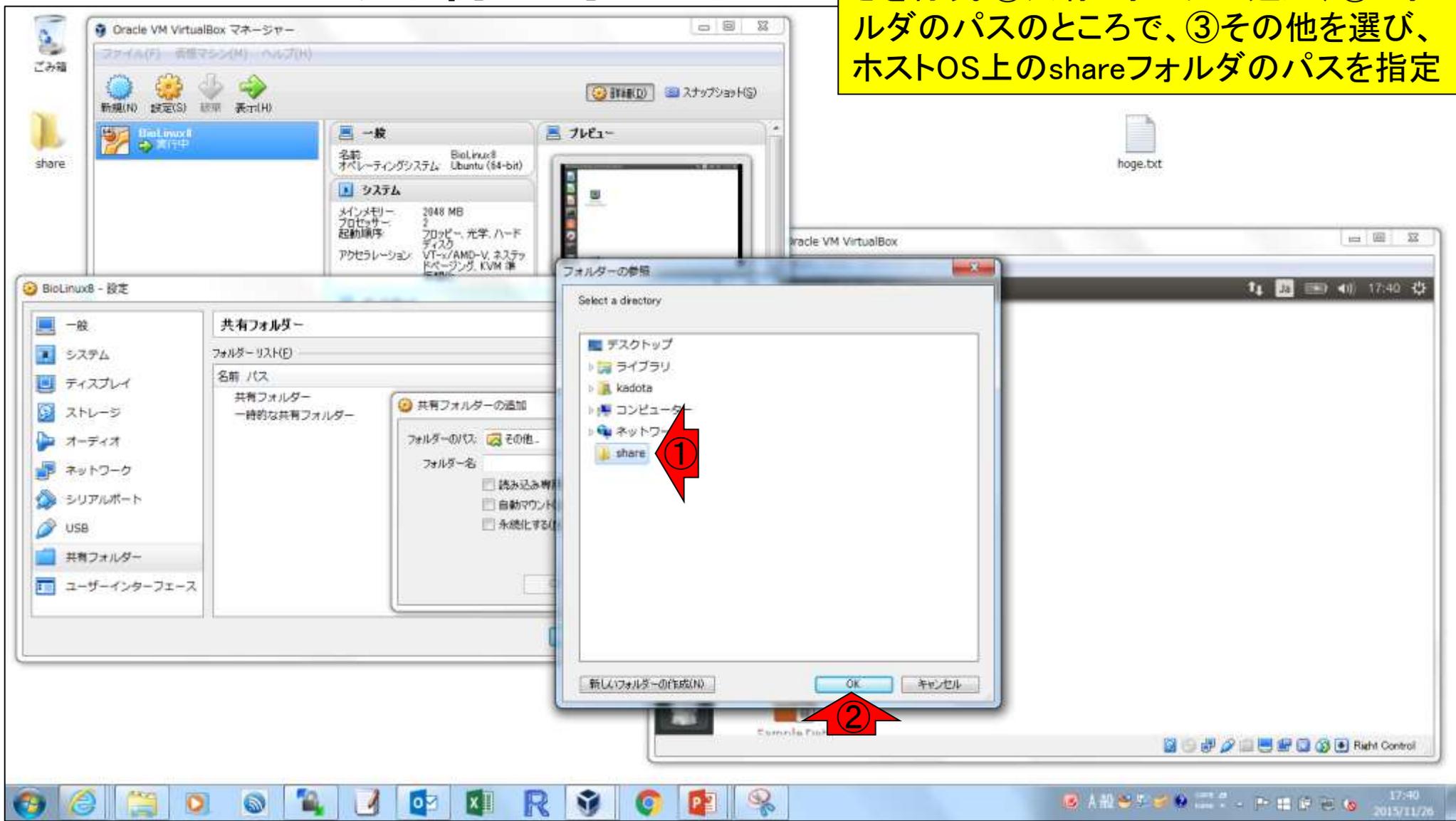
W9-2-2: 共有フォルダ

ホストOSのデスクトップ上に作成したshareフォルダを共有フォルダとするための手続きを行う。①共有フォルダの追加、②フォルダのパスのところ、③その他を選び、ホストOS上のshareフォルダのパスを指定

The screenshot shows the 'BioLinux8 - 設定' (BioLinux8 - Settings) window. The '共有フォルダ' (Shared Folders) section is active. The '共有フォルダの追加' (Add Shared Folder) dialog box is open, showing the 'フォルダのパス' (Folder Path) field and the 'フォルダ名' (Folder Name) dropdown menu. The 'その他...' (Other...) option is selected in the dropdown. Red arrows point to the '共有フォルダの追加' button (1), the 'フォルダのパス' field (2), and the 'その他...' option (3).

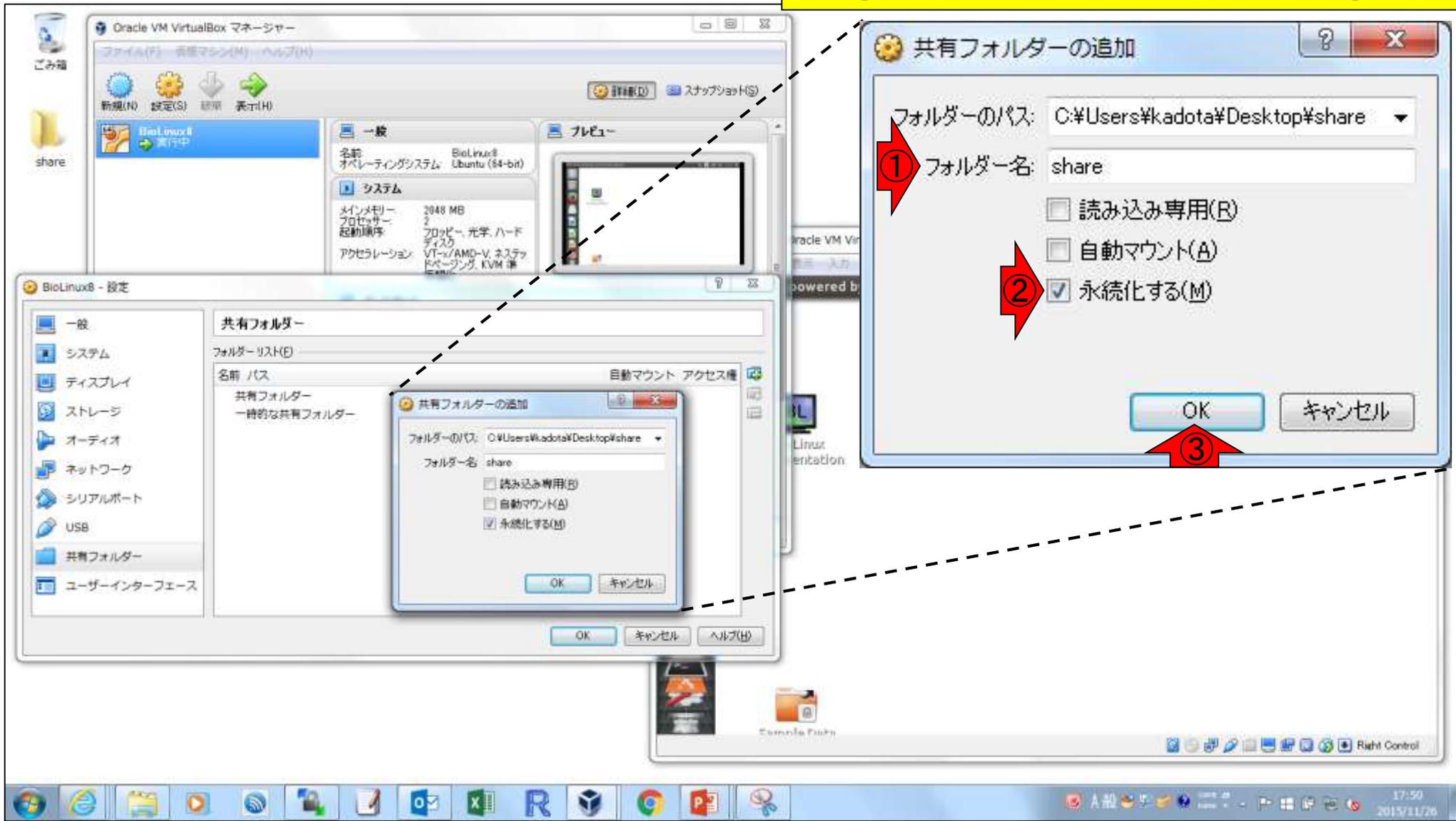
W9-2-2: 共有フォルダ

ホストOSのデスクトップ上に作成したshareフォルダを共有フォルダとするための手続きを行う。①共有フォルダの追加、②フォルダのパスのところで、③その他を選び、ホストOS上のshareフォルダのパスを指定

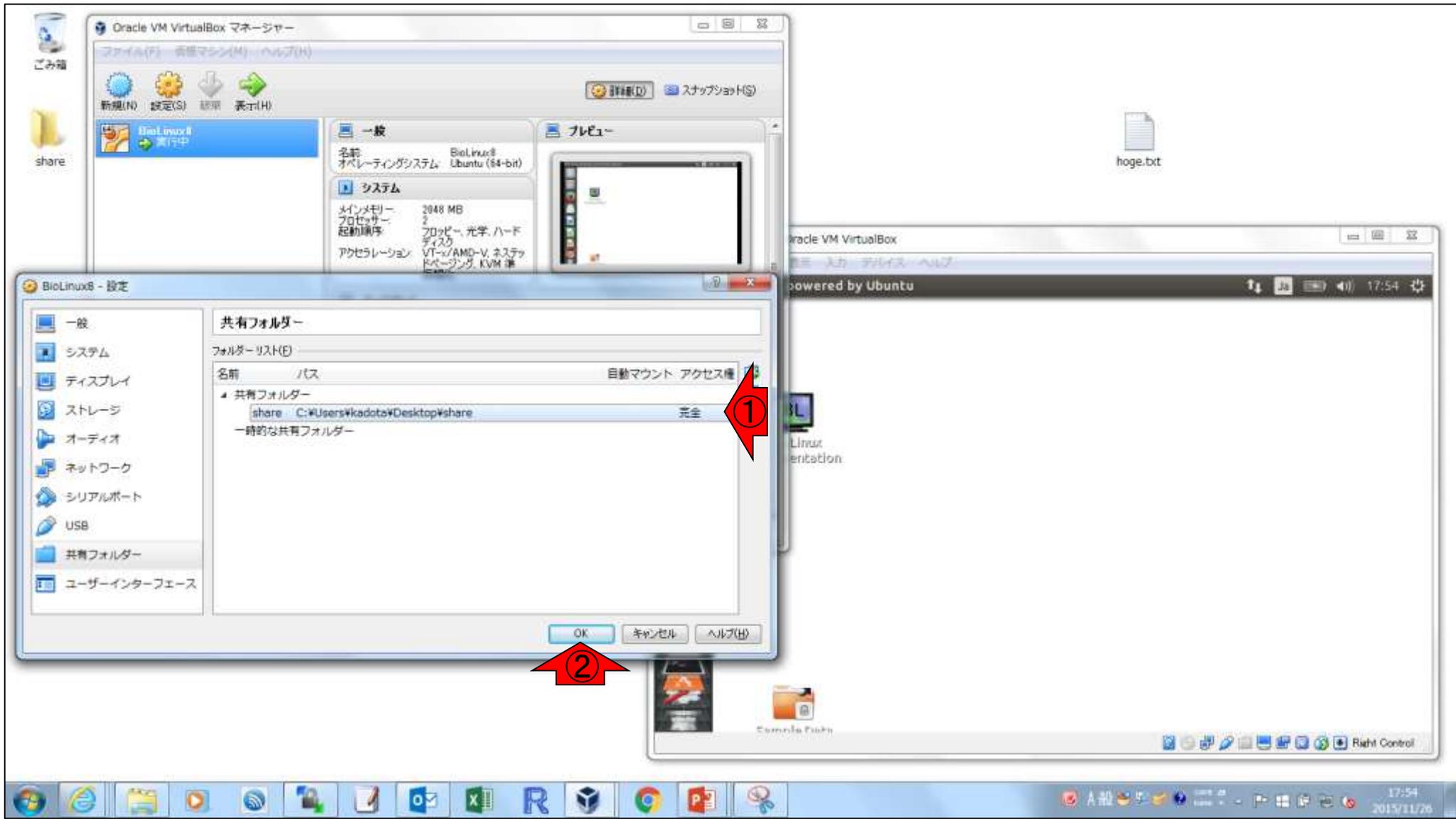


W9-2-2: 共有フォルダ

ホストOS上のshareフォルダのパスを指定すると、①フォルダ名のところが自動的にshareとなる。②永続化するにチェックを入れて、③OK。

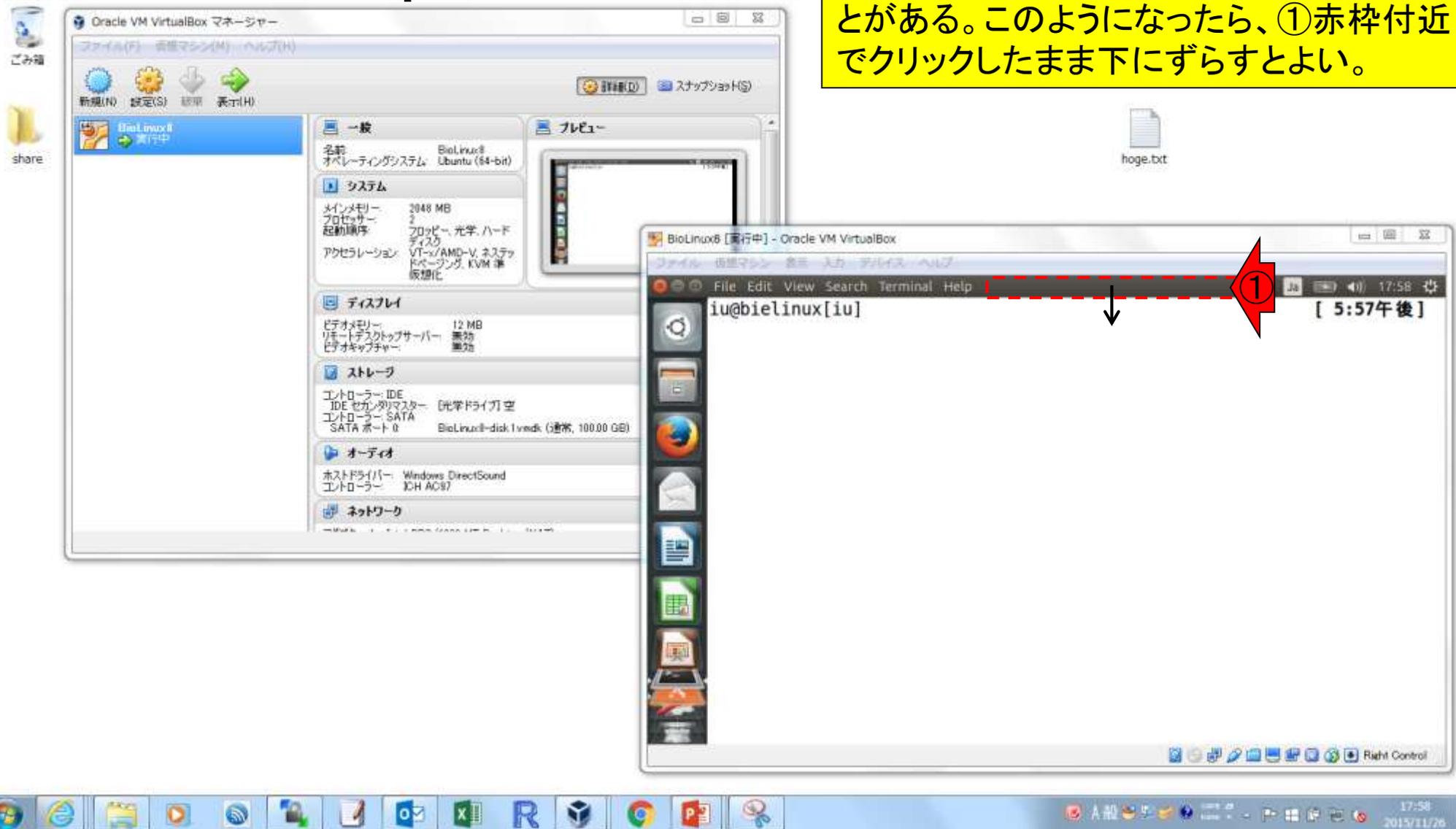


W9-2-2: 共有フォルダ



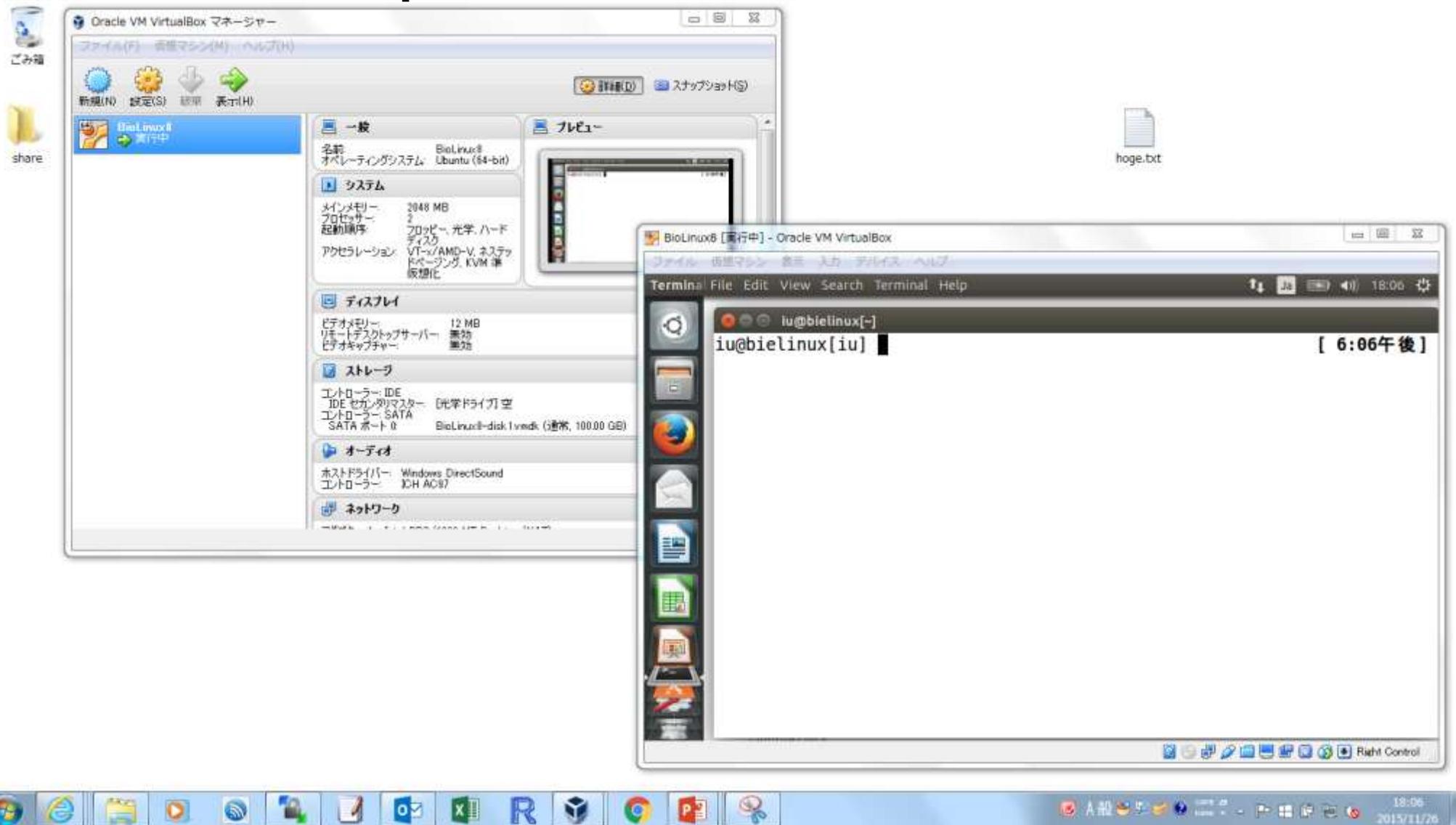
W9-2-3: Tips

ターミナルを起動したときなど、ときどきこのようなホストOS上で全画面表示になって他のターミナルが見られなくなったりすることがある。このようになったら、①赤枠付近でクリックしたまま下にずらすとよい。



W9-2-3: Tips

こんな感じになります。Firefoxを開いているときなども同様です。



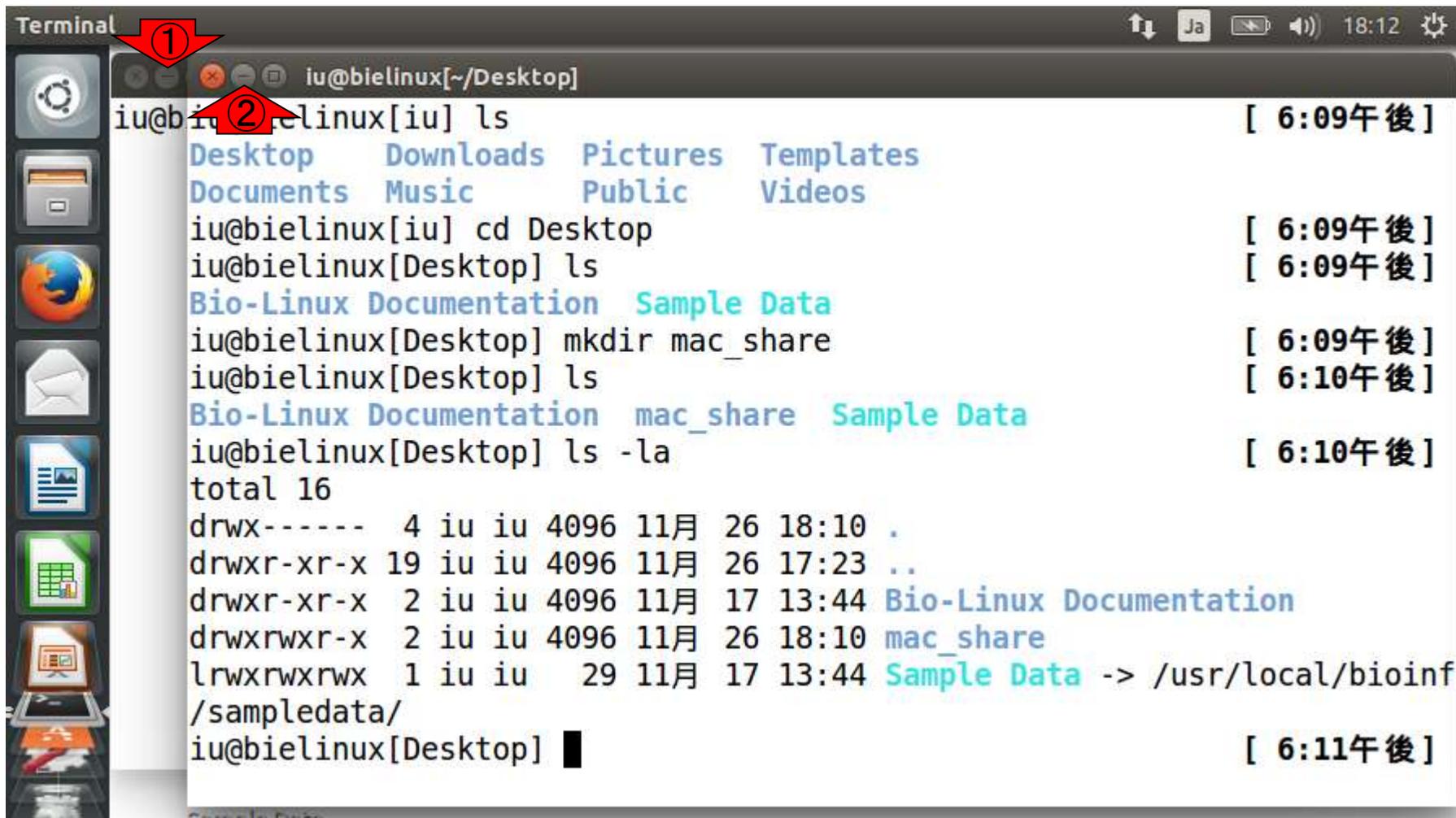
W9-2-3

多少見栄えが異なっても気にしなくてよい。ゲストOSのターミナル上で、①Desktopに移動し、②mac_shareという名前のフォルダを作成している。それ以外のlsコマンドはただの確認。赤枠部分が作成されたmac_shareフォルダ。

```
Terminal
iu@bielinux[~/Desktop]
iu@bielinux[iu] ls [ 6:09午後 ]
Desktop Downloads Pictures Templates
Documents Music Public Videos
① iu@bielinux[iu] cd Desktop [ 6:09午後 ]
iu@bielinux[Desktop] ls [ 6:09午後 ]
Bio-Linux Documentation Sample Data
② iu@bielinux[Desktop] mkdir mac_share [ 6:09午後 ]
iu@bielinux[Desktop] ls [ 6:10午後 ]
Bio-Linux Documentation mac_share Sample Data
iu@bielinux[Desktop] ls -la [ 6:10午後 ]
total 16
drwx----- 4 iu iu 4096 11月 26 18:10 .
drwxr-xr-x 19 iu iu 4096 11月 26 17:23 ..
drwxr-xr-x 2 iu iu 4096 11月 17 13:44 Bio-Linux Documentation
drwxrwxr-x 2 iu iu 4096 11月 26 18:10 mac_share
lrwxrwxrwx 1 iu iu 29 11月 17 13:44 Sample Data -> /usr/local/bioinf
/sampledata/
iu@bielinux[Desktop] [ 6:11午後 ]
```

W9-2-3

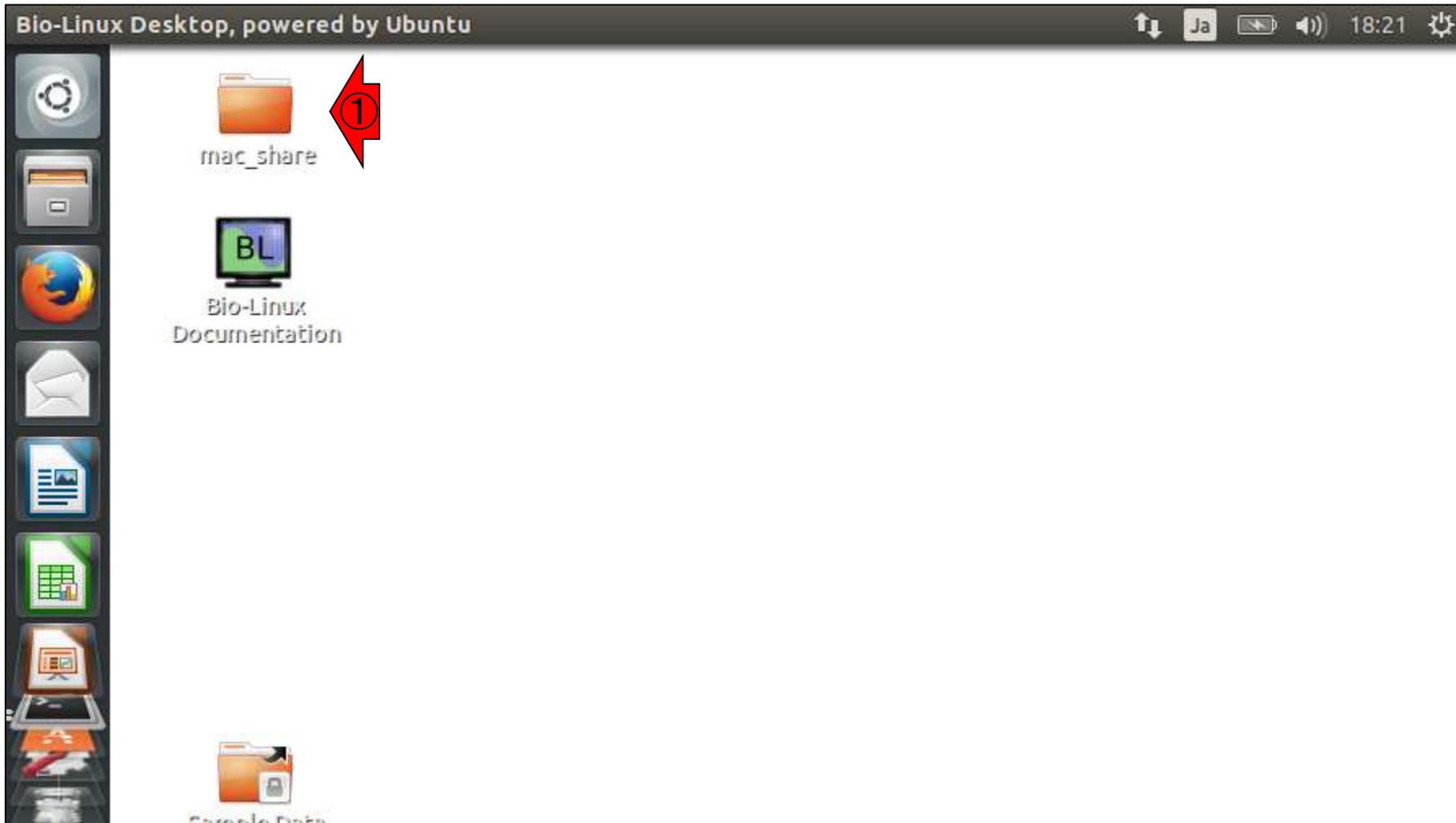
現在開いている2つのターミナル画面を「最小化」することでデスクトップを眺めることができる



```
Terminal
iu@bielinux[~/Desktop]
iu@bielinux[iu] ls [ 6:09午後 ]
Desktop Downloads Pictures Templates
Documents Music Public Videos
iu@bielinux[iu] cd Desktop [ 6:09午後 ]
iu@bielinux[Desktop] ls [ 6:09午後 ]
Bio-Linux Documentation Sample Data
iu@bielinux[Desktop] mkdir mac_share [ 6:09午後 ]
iu@bielinux[Desktop] ls [ 6:10午後 ]
Bio-Linux Documentation mac_share Sample Data
iu@bielinux[Desktop] ls -la [ 6:10午後 ]
total 16
drwx----- 4 iu iu 4096 11月 26 18:10 .
drwxr-xr-x 19 iu iu 4096 11月 26 17:23 ..
drwxr-xr-x 2 iu iu 4096 11月 17 13:44 Bio-Linux Documentation
drwxrwxr-x 2 iu iu 4096 11月 26 18:10 mac_share
lrwxrwxrwx 1 iu iu 29 11月 17 13:44 Sample Data -> /usr/local/bioinf
/sampledata/
iu@bielinux[Desktop] [ 6:11午後 ]
```

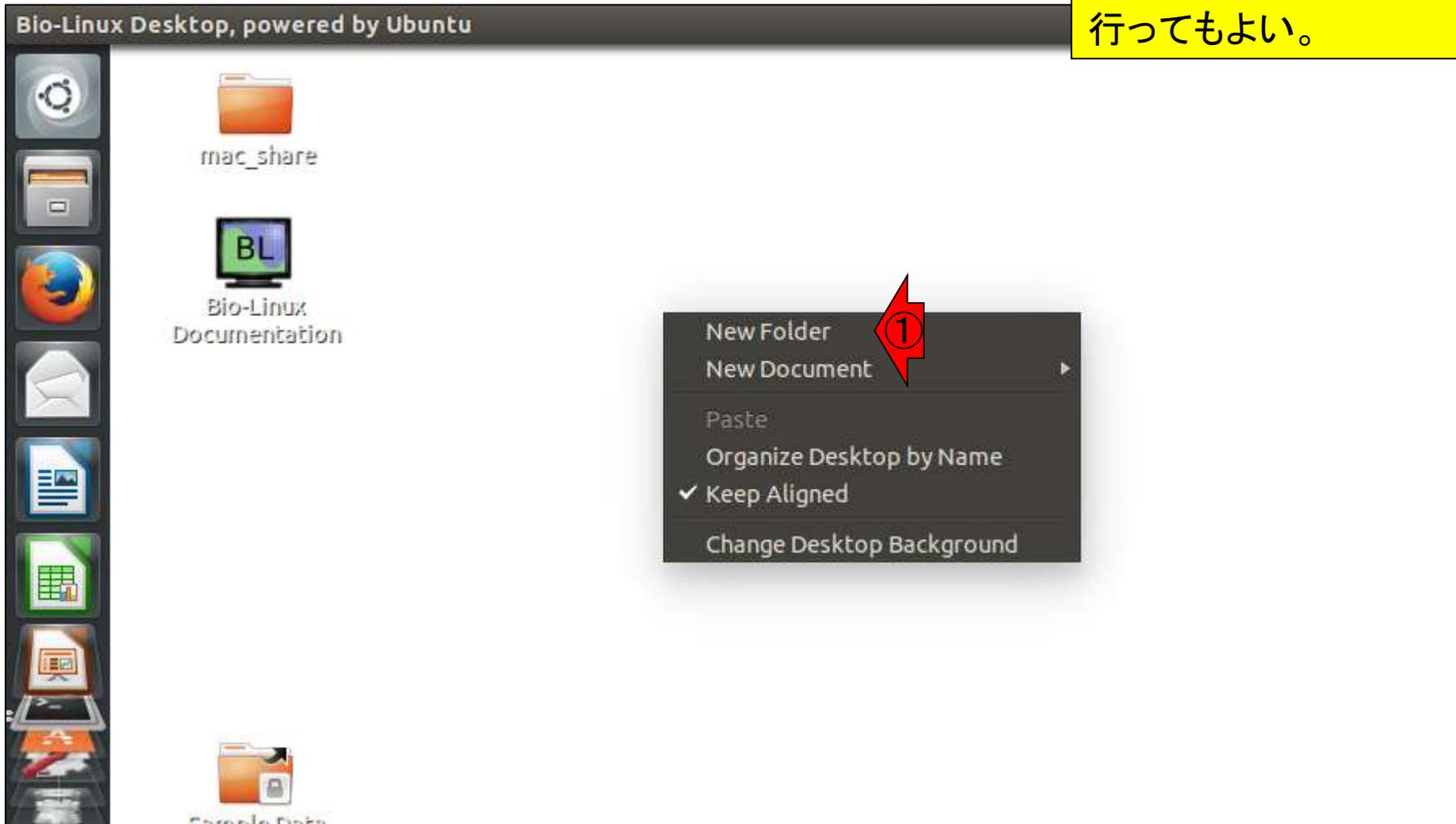
確かにゲストOSのデスクトップ上に①mac_share
というフォルダが作成されていることがわかる。

W9-2-3: 確認



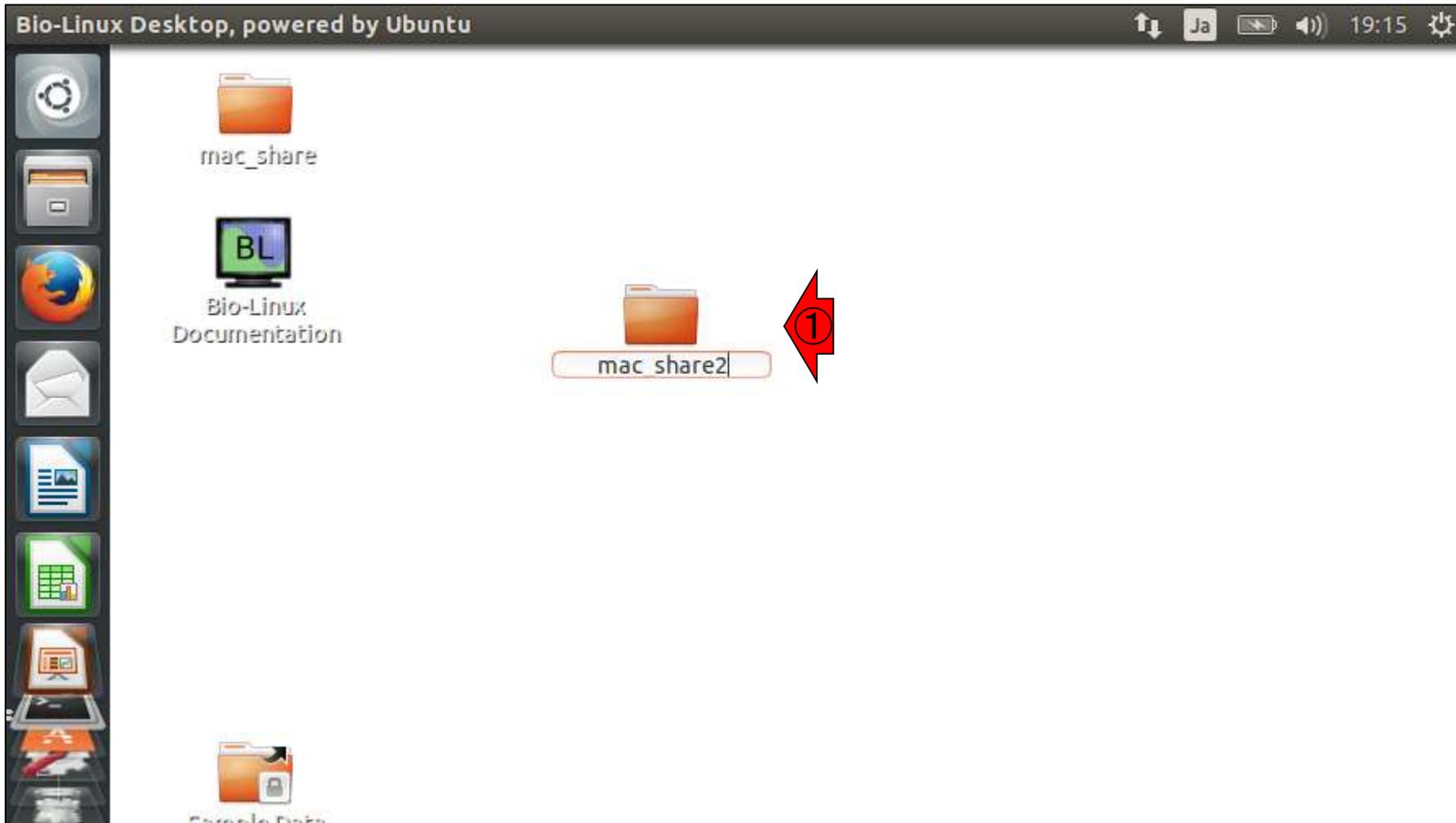
W9-2-3: GUIベースの場合

デスクトップ画面上で①右クリック、
②New Folderで通常のGUI画面と
同じ感覚で新規フォルダの作成を
行ってもよい。



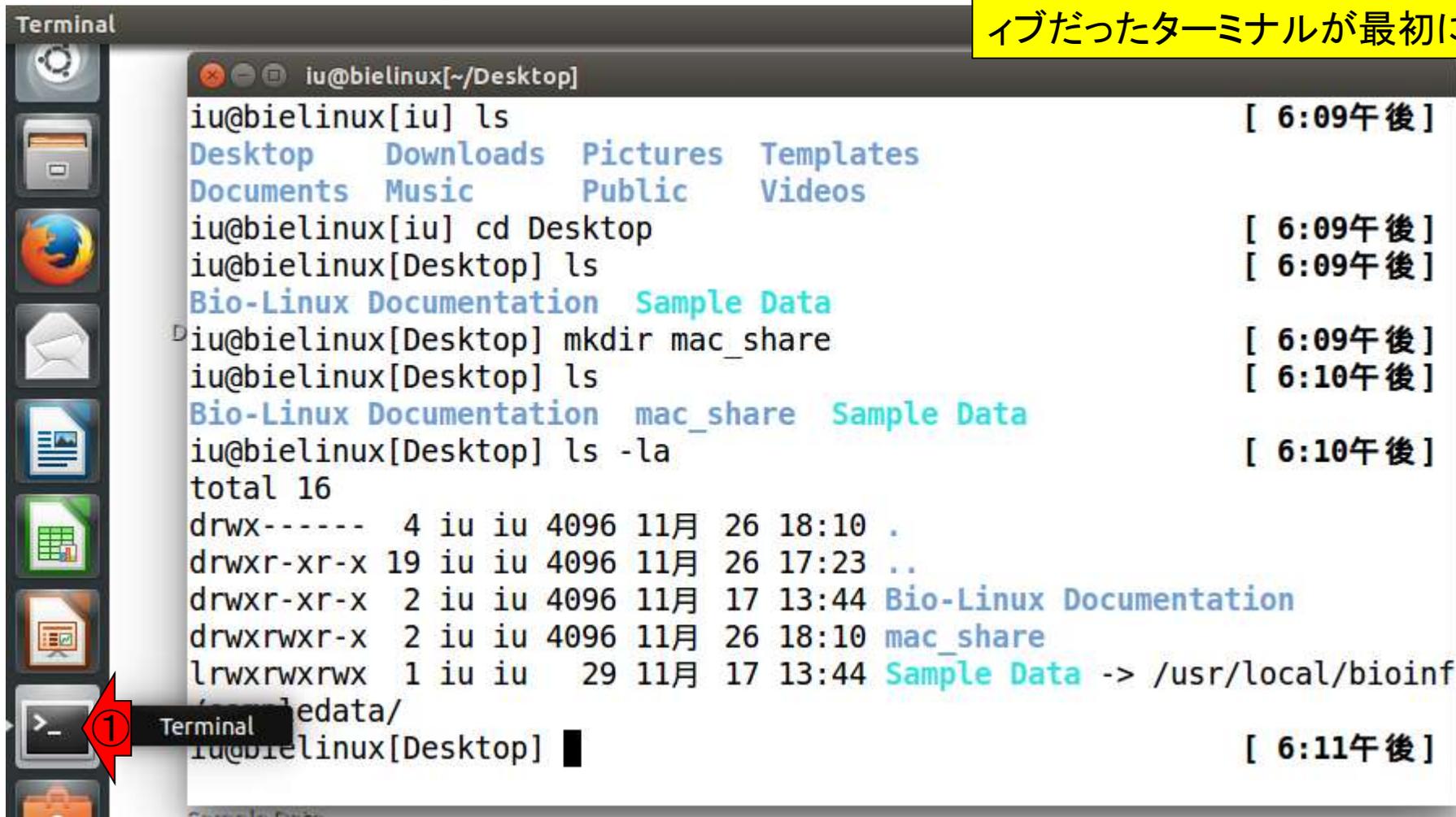
ここでは①mac_share2
フォルダを作成

W9-2-3: GUIベースの場合



W9-2-4: 確認

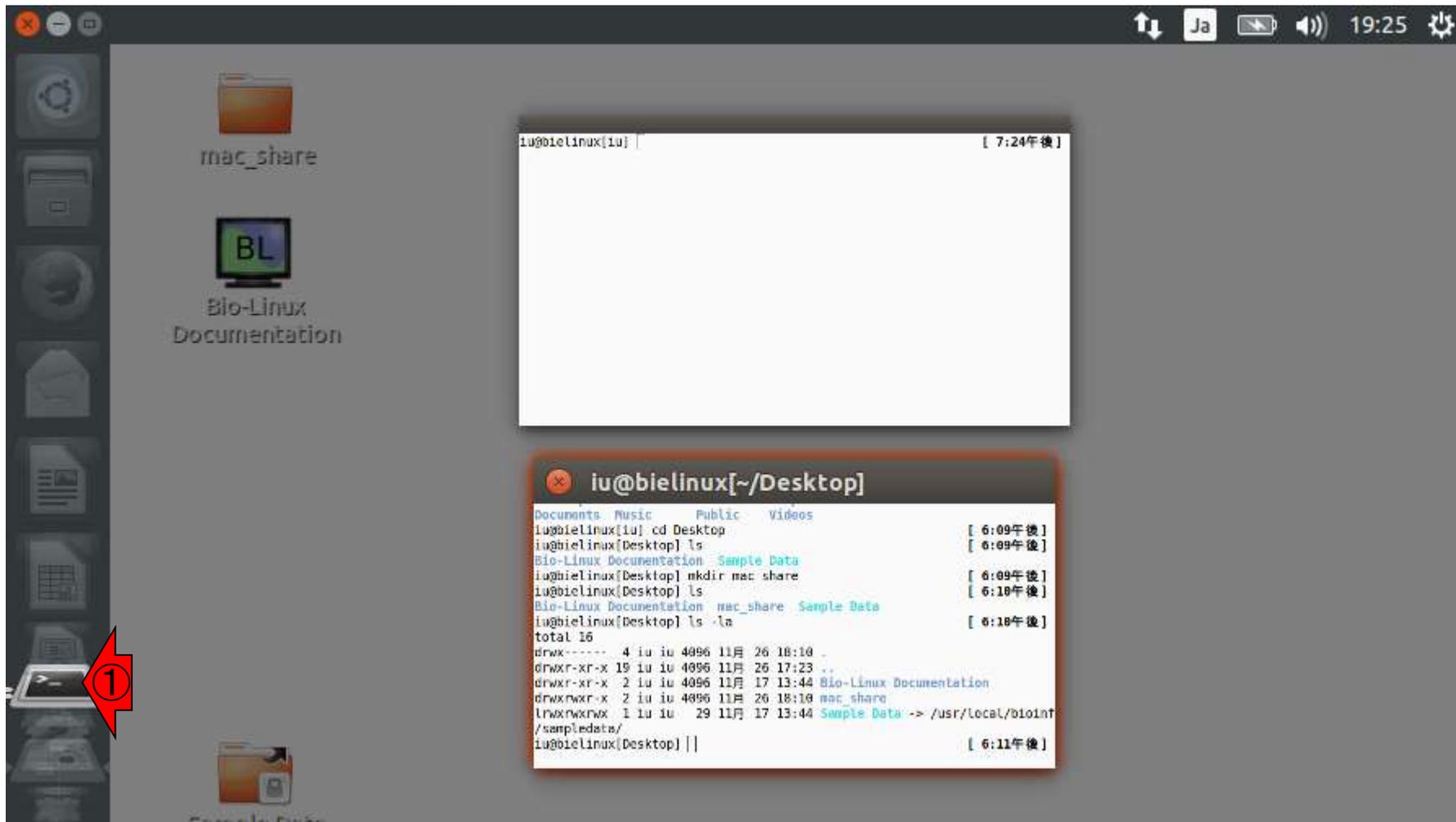
①ターミナルアイコンをクリックすると、最小化されていたターミナルが表示される。ターミナルは2つあったが、最後にアクティブだったターミナルが最初に表示される



```
Terminal
iu@bielinux[~/Desktop]
iu@bielinux[iu] ls [ 6:09午後]
Desktop Downloads Pictures Templates
Documents Music Public Videos
iu@bielinux[iu] cd Desktop [ 6:09午後]
iu@bielinux[Desktop] ls [ 6:09午後]
Bio-Linux Documentation Sample Data
iu@bielinux[Desktop] mkdir mac_share [ 6:09午後]
iu@bielinux[Desktop] ls [ 6:10午後]
Bio-Linux Documentation mac_share Sample Data
iu@bielinux[Desktop] ls -la [ 6:10午後]
total 16
drwx----- 4 iu iu 4096 11月 26 18:10 .
drwxr-xr-x 19 iu iu 4096 11月 26 17:23 ..
drwxr-xr-x 2 iu iu 4096 11月 17 13:44 Bio-Linux Documentation
drwxrwxr-x 2 iu iu 4096 11月 26 18:10 mac_share
lrwxrwxrwx 1 iu iu 29 11月 17 13:44 Sample Data -> /usr/local/bioinf
data/
iu@bielinux[Desktop] [ 6:11午後]
```

W9-2-4: 確認

①もう一度ターミナルアイコンをクリックすると、もう一つのターミナルも選択可能になる。



W9-2-4: 確認

mac_share2フォルダ作成後に再び①「ls -la」した結果。②mac_share2が見えるようになった。コマンドは、改めて打ち込まなくても「上矢印ボタン」を1回押せばよい

```
Terminal File Edit View Search Terminal Help
iu@bielinux[~/Desktop]
iu@bielinux[Desktop] ls -la [ 6:10午後 ]
total 16
drwx----- 4 iu iu 4096 11月 26 18:10 .
drwxr-xr-x 19 iu iu 4096 11月 26 17:23 ..
drwxr-xr-x 2 iu iu 4096 11月 17 13:44 Bio-Linux Documentation
drwxrwxr-x 2 iu iu 4096 11月 26 18:10 mac_share
lrwxrwxrwx 1 iu iu 29 11月 17 13:44 Sample Data -> /usr/local/bioinf
/sampledata/
① iu@bielinux[Desktop] ls -la [ 6:11午後 ]
total 20
drwx----- 5 iu iu 4096 11月 26 19:24 .
drwxr-xr-x 19 iu iu 4096 11月 26 19:24 ..
drwxr-xr-x 2 iu iu 4096 11月 17 13:44 Bio-Linux Documentation
drwxrwxr-x 2 iu iu 4096 11月 26 18:10 mac_share
drwxrwxr-x 2 iu iu 4096 11月 26 19:15 mac_share2 ②
lrwxrwxrwx 1 iu iu 29 11月 17 13:44 Sample Data -> /usr/local/bioinf
/sampledata/
iu@bielinux[Desktop] [ 7:28午後 ]
```

W9-2-4: 画面サイズ

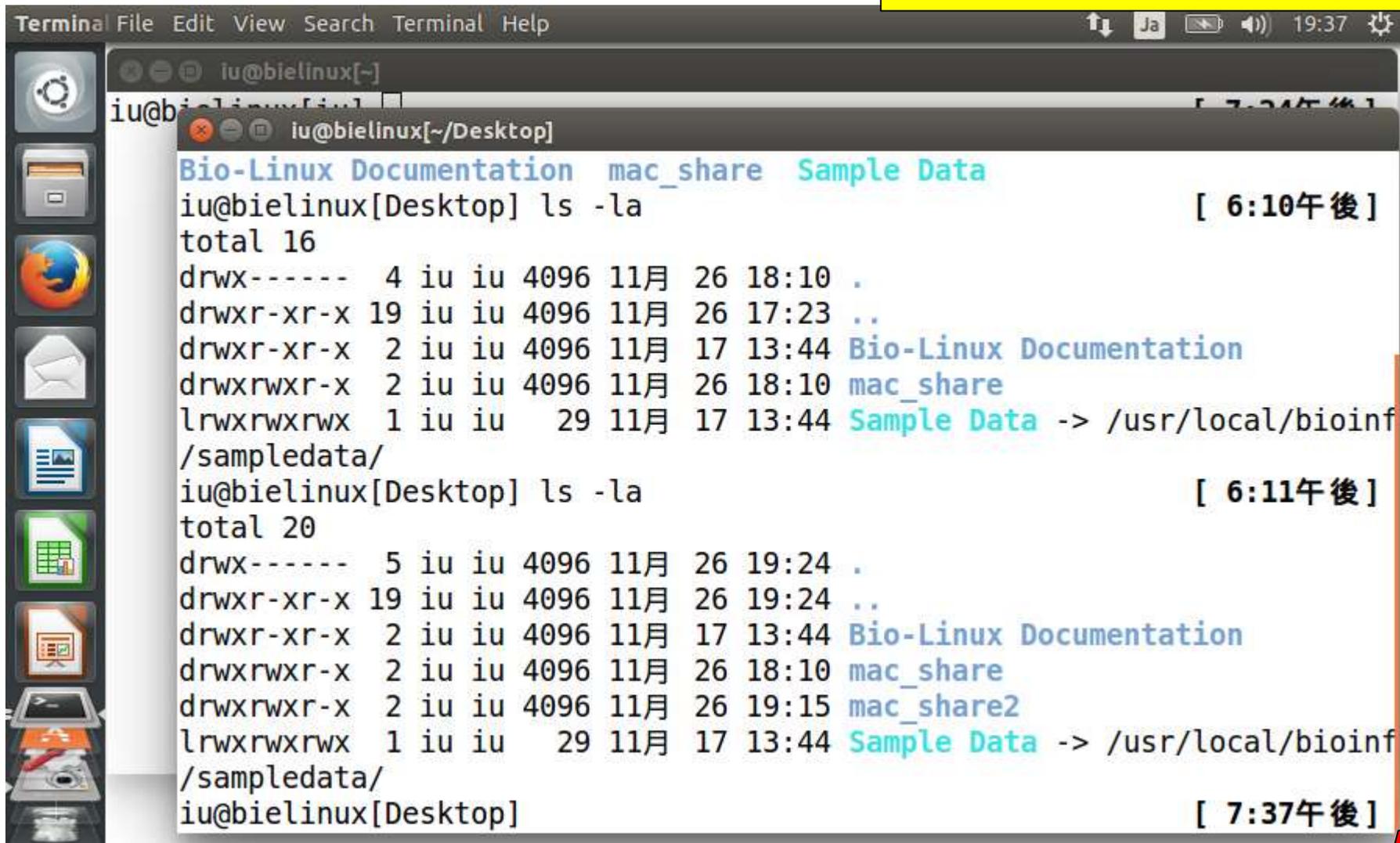
Tips。ターミナルの画面サイズを変えるのは右下が基本だが、①右下部分にマウスポインタを合わせてもサイズ変更可能な状態にならないときがある。そのような場合は赤枠付近でマウスクリックしたまま下にずらすとよい

```
Terminal File Edit View Search Terminal Help
iu@bielinux[~/Desktop]
iu@bielinux[Desktop] ls -la [ 6:10午後 ]
total 16
drwx----- 4 iu iu 4096 11月 26 18:10 .
drwxr-xr-x 19 iu iu 4096 11月 26 17:23 ..
drwxr-xr-x 2 iu iu 4096 11月 17 13:44 Bio-Linux Documentation
drwxrwxr-x 2 iu iu 4096 11月 26 18:10 mac_share
lrwxrwxrwx 1 iu iu 29 11月 17 13:44 Sample Data -> /usr/local/bioinformatics/
/sampledata/
iu@bielinux[Desktop] ls -la [ 6:11午後 ]
total 20
drwx----- 5 iu iu 4096 11月 26 19:24 .
drwxr-xr-x 19 iu iu 4096 11月 26 19:24 ..
drwxr-xr-x 2 iu iu 4096 11月 17 13:44 Bio-Linux Documentation
drwxrwxr-x 2 iu iu 4096 11月 26 18:10 mac_share
drwxrwxr-x 2 iu iu 4096 11月 26 19:15 mac_share2
lrwxrwxrwx 1 iu iu 29 11月 17 13:44 Sample Data -> /usr/local/bioinformatics/
/sampledata/
iu@bielinux[Desktop] █ [ 7:28午後 ]
```



W9-2-4: 画面サイズ

こんな感じになるので、①ウィンドウの右下を左クリックボタンを押しながら移動させることでウィンドウサイズを任意に変更できる。



The image shows a Linux desktop environment with a terminal window open. The terminal displays the output of the 'ls -la' command in the Desktop directory at three different times: 6:10 PM, 6:11 PM, and 7:37 PM. The file listings are as follows:

```
iu@bielinux[Desktop] ls -la [ 6:10午後 ]
total 16
drwx----- 4 iu iu 4096 11月 26 18:10 .
drwxr-xr-x 19 iu iu 4096 11月 26 17:23 ..
drwxr-xr-x 2 iu iu 4096 11月 17 13:44 Bio-Linux Documentation
drwxrwxr-x 2 iu iu 4096 11月 26 18:10 mac_share
lrwxrwxrwx 1 iu iu 29 11月 17 13:44 Sample Data -> /usr/local/bioinf
/sampledata/

iu@bielinux[Desktop] ls -la [ 6:11午後 ]
total 20
drwx----- 5 iu iu 4096 11月 26 19:24 .
drwxr-xr-x 19 iu iu 4096 11月 26 19:24 ..
drwxr-xr-x 2 iu iu 4096 11月 17 13:44 Bio-Linux Documentation
drwxrwxr-x 2 iu iu 4096 11月 26 18:10 mac_share
drwxrwxr-x 2 iu iu 4096 11月 26 19:15 mac_share2
lrwxrwxrwx 1 iu iu 29 11月 17 13:44 Sample Data -> /usr/local/bioinf
/sampledata/

iu@bielinux[Desktop] [ 7:37午後 ]
```



W9-2-5: mount

①「`sudo mount -t vboxsf share mac_share`」。ホストOSのデスクトップ上にあるshareフォルダを、ゲストOSのデスクトップ上のmac_shareフォルダとして認識させ、利用可能な状態にしている(この操作をマウントという)。②「`[sudo] password for iu:`」とログインユーザiuのパスワードを聞かれているので、この場合pass1409と打ち込む。③見ため上、特に何も起こらず、通常のコマンド入力待ち状態になるだけである。

```
Terminal File Edit View Search Terminal Help
iu@bielinux[~]
iu@bielinux[~/Desktop]
total 16
drwx----- 4 iu iu 4096 11月 26 18:10 .
drwxr-xr-x 19 iu iu 4096 11月 26 17:23 ..
drwxr-xr-x 2 iu iu 4096 11月 17 13:44 Bio-Linux Documentation
drwxrwxr-x 2 iu iu 4096 11月 26 18:10 mac_share
lrwxrwxrwx 1 iu iu 29 11月 17 13:44 Sample Data -> /usr/local/bioinf
/sampledata/
iu@bielinux[Desktop] ls -la
total 20
drwx----- 5 iu iu 4096 11月 26 19:24 .
drwxr-xr-x 19 iu iu 4096 11月 26 19:24 ..
drwxr-xr-x 2 iu iu 4096 11月 17 13:44 Bio-Linux Documentation
drwxrwxr-x 2 iu iu 4096 11月 26 18:10 mac_share
drwxrwxr-x 2 iu iu 4096 11月 26 19:15 mac_share2
lrwxrwxrwx 1 iu iu 29 11月 17 13:44 Sample Data -> /usr/local/bioinf
/sampledata/
iu@bielinux[Desktop] sudo mount -t vboxsf share mac_share
[sudo] password for iu:
iu@bielinux[Desktop]
```

[6:11午後]

[7:37午後]

[8:01午後]

W9-2-5: mount

③もう一度「ls -la」した結果。赤の下線で示すように、①「sudo mount -t vboxsf share mac_share」実行前後でmac_shareフォルダのところの記述が変わっていることがわかる。

```
Terminal File Edit View Search Terminal Help
iu@bielinux[~]
iu@bielinux[~/Desktop]
drwx----- 5 iu iu 4096 11月 26 19:24 .
drwxr-xr-x 19 iu iu 4096 11月 26 19:24 ..
drwxr-xr-x 2 iu iu 4096 11月 17 13:44 Bio-Linux Documentation
drwxrwxr-x 2 iu iu 4096 11月 26 18:10 mac_share
drwxrwxr-x 2 iu iu 4096 11月 26 19:15 mac_share2
lrwxrwxrwx 1 iu iu 29 11月 17 13:44 Sample Data -> /usr/local/bioinf
/sampledata/
iu@bielinux[Desktop] sudo mount -t vboxsf share mac_share [ 7:37午後 ]
[sudo] password for iu:
iu@bielinux[Desktop] ls -la [ 8:01午後 ]
total 16
drwx----- 5 iu iu 4096 11月 26 19:24 .
drwxr-xr-x 19 iu iu 4096 11月 26 19:24 ..
drwxr-xr-x 2 iu iu 4096 11月 17 13:44 Bio-Linux Documentation
drwxrwxrwx 1 root root 0 11月 26 17:10 mac_share
drwxrwxr-x 2 iu iu 4096 11月 26 19:15 mac_share2
lrwxrwxrwx 1 iu iu 29 11月 17 13:44 Sample Data -> /usr/local/bi
oinf/sampledata/
iu@bielinux[Desktop] [ 8:05午後 ]
```



W9-2-5: permission

```
Terminal File Edit View Search Terminal Help
iu@bielinux[~]
iu@bielinux[~/Desktop]
drwx----- 5 iu iu 4096 11月 26 19:24 .
drwxr-xr-x 19 iu iu 4096 11月 26 19:24 ..
drwxr-xr-x 2 iu iu 4096 11月 17 13:44 Bio-Linux Documentation
drwxrwxr-x 2 iu iu 4096 11月 26 18:15 mac_share
drwxrwxr-x 2 iu iu 4096 11月 26 19:15 mac_share2
lrwxrwxrwx 1 iu iu 29 11月 17 13:44 Sample Data -> /usr/local/bioinf/sampled/
/sampled/
iu@bielinux[Desktop] sudo mount -t vfat /dev/sdb1 /mnt
[sudo] password for iu:
iu@bielinux[Desktop] ls -la
total 16
drwx----- 5 iu iu 4096 11月 26 19:24 .
drwxr-xr-x 19 iu iu 4096 11月 26 19:24 ..
drwxr-xr-x 2 iu iu 4096 11月 17 13:44 Bio-Linux Documentation
drwxrwxrwx 1 root root 0 11月 26 17:10 mac_share
drwxrwxr-x 2 iu iu 4096 11月 26 19:15 mac_share2
lrwxrwxrwx 1 iu iu 29 11月 17 13:44 Sample Data -> /usr/local/bioinf/sampled/
iu@bielinux[Desktop]
```

特に重要なのは赤枠で囲った部分。「drwxrwxrwx」のようになっていればOK。一番左のdはディレクトリの意味。それ以外の9文字のrwxrwxrwxは、mac_shareディレクトリの「所有者、特定のグループ、その他のヒト」を含め全てのヒトに、このディレクトリ中のファイルを読むこと(r)、書き込み(w)、実行(x)の権限を与えるという意味です。もしここがrwxrwxrwxになっていなければ、共有フォルダとしてゲストOS上に作成したユーザiuが、mac_shareディレクトリを介して「ホストOSのデスクトップ上に作成したshareフォルダ」に書き込みが行えないなどの不具合が生じます。Windowsユーザは、特にいじっていないければこのデフォルトはrwxrwxrwxとなって問題ないはず。

[8:05午後]

W9-2-5: permission

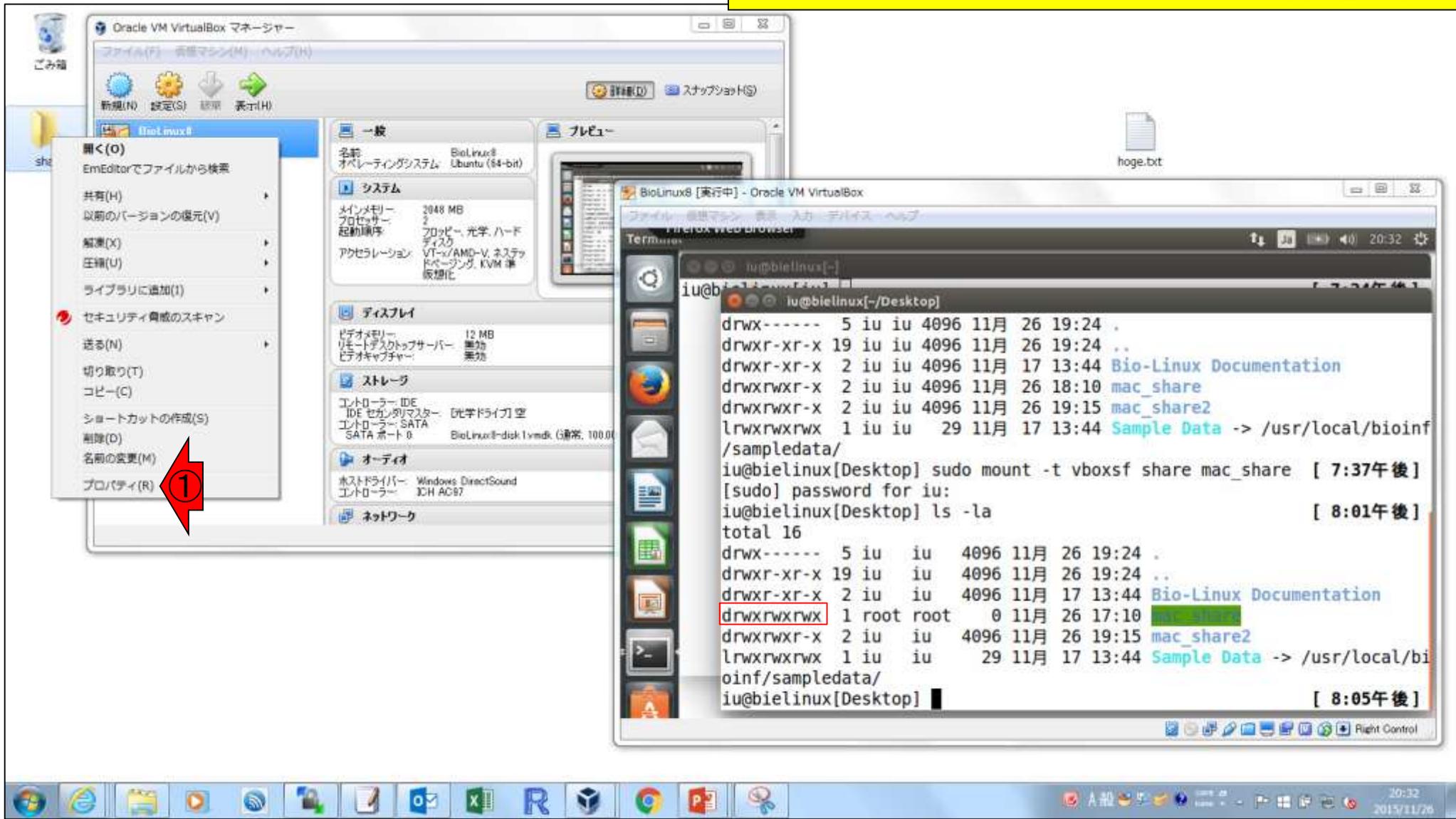
ゲストOSのデスクトップに作成したmac_shareディレクトリの実体は、ホストOSのデスクトップ上に作成したshareフォルダです。「sudo mount -t vboxsf share mac_share」コマンドを含む一連の作業でそうなるようにしているからです。そのため、mac_shareディレクトリの所有者やアクセス権限の変更は、ゲストOS上でいくら頑張ってもできません

The image shows the Oracle VM VirtualBox Manager interface on the left and a terminal window on the right. In the VM Manager, the 'BioLinux8' VM is selected, and a red arrow points to the 'share' folder in the host file system. The terminal window shows the following commands and output:

```
iu@bielinux[~/Desktop]
iu@bielinux[~/Desktop] sudo mount -t vboxsf share mac_share [ 7:37午後 ]
[sudo] password for iu:
iu@bielinux[~/Desktop] ls -la [ 8:01午後 ]
total 16
drwx----- 5 iu iu 4096 11月 26 19:24 .
drwxr-xr-x 19 iu iu 4096 11月 26 19:24 ..
drwxr-xr-x 2 iu iu 4096 11月 17 13:44 Bio-Linux Documentation
drwxrwxr-x 2 iu iu 4096 11月 26 18:10 mac_share
drwxrwxr-x 2 iu iu 4096 11月 26 19:15 mac_share2
lrwxrwxrwx 1 iu iu 29 11月 17 13:44 Sample Data -> /usr/local/bioinf
/sampledata/
iu@bielinux[~/Desktop]
iu@bielinux[~/Desktop] sudo mount -t vboxsf share mac_share [ 7:37午後 ]
[sudo] password for iu:
iu@bielinux[~/Desktop] ls -la [ 8:01午後 ]
total 16
drwx----- 5 iu iu 4096 11月 26 19:24 .
drwxr-xr-x 19 iu iu 4096 11月 26 19:24 ..
drwxr-xr-x 2 iu iu 4096 11月 17 13:44 Bio-Linux Documentation
drwxrwxrwx 1 root root 0 11月 26 17:10 mac_share
drwxrwxr-x 2 iu iu 4096 11月 26 19:15 mac_share2
lrwxrwxrwx 1 iu iu 29 11月 17 13:44 Sample Data -> /usr/local/bi
oinf/sampledata/
iu@bielinux[~/Desktop]
```

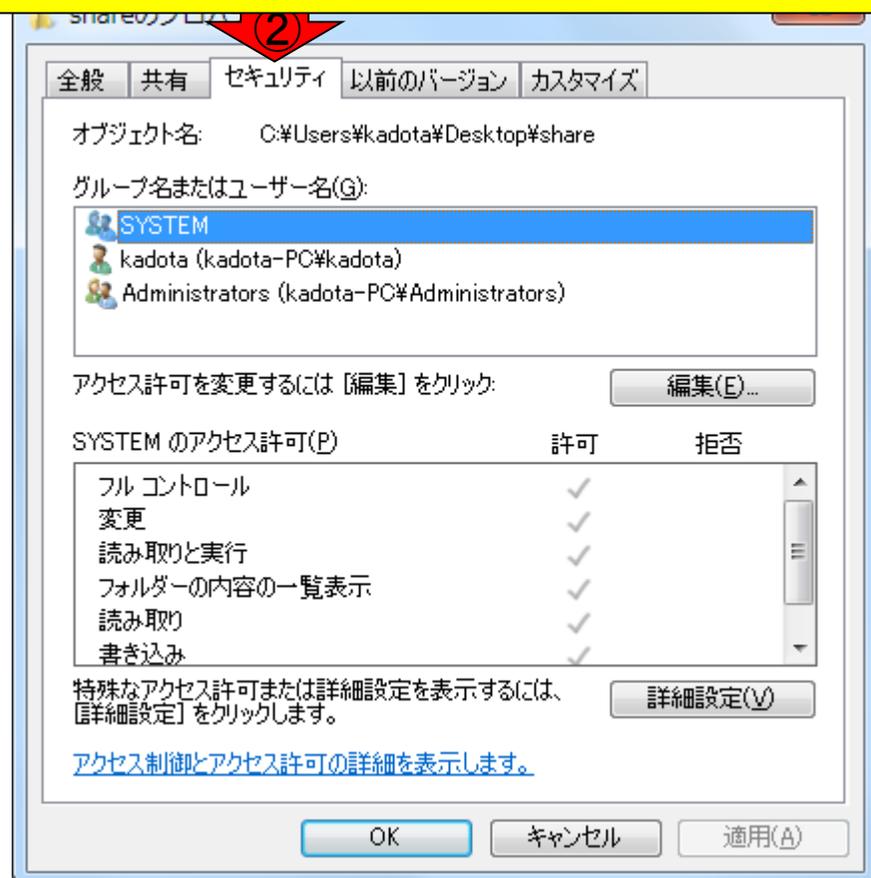
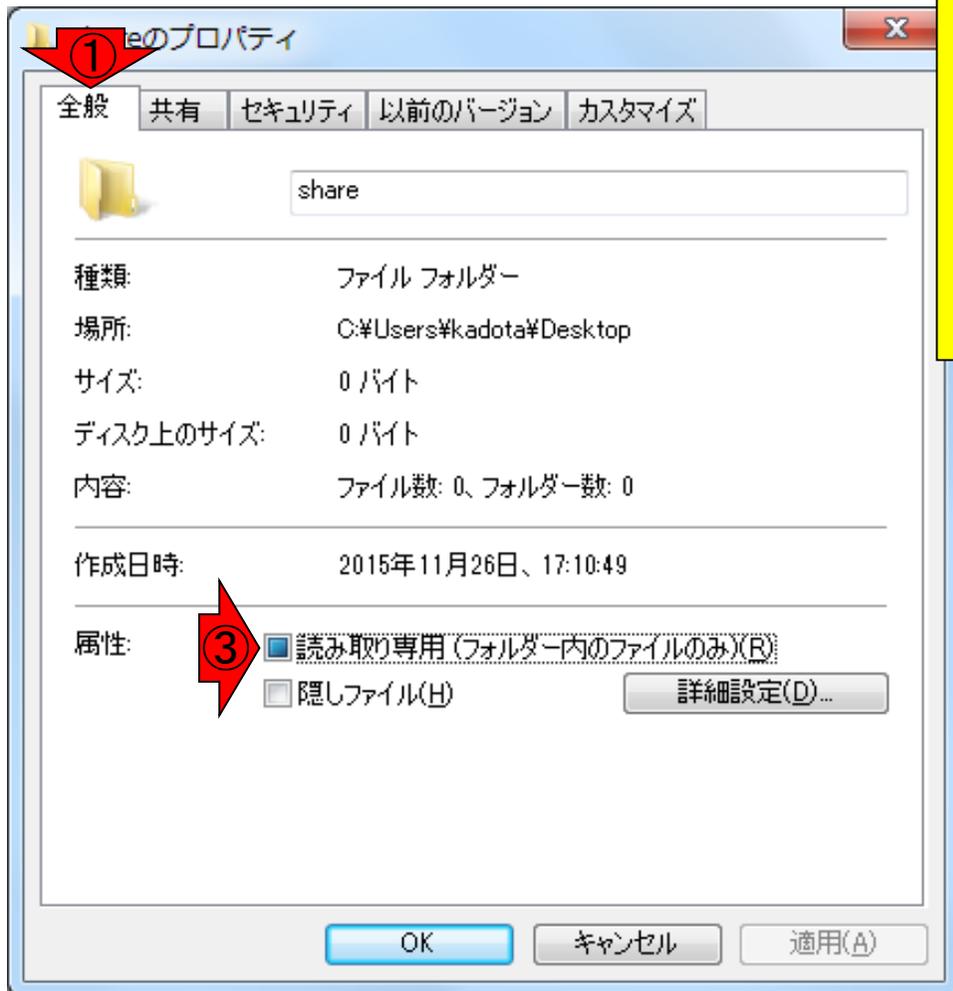
W9-2-6: プロパティ

赤枠部分が「drwxrwxrwx」になっていない場合は、
ホストOSのデスクトップ上にあるshareフォルダの
アイコン上で右クリックして「プロパティ」を選択。



W9-2-6: プロパティ

①全般や②セキュリティタブなどの項目中に「drwxrwxrwx」になっていない原因があるはずですので、どうにかして変更しましょう。その後、またターミナル上で「ls -la」で「drwxrwxrwx」となっているかどうかの確認。その繰り返しです。私の場合、「drwxrwxrwx」となっている環境でも③読み取り専用となっていました。うまくいっている状況でわざわざこのチェックを外してハマったことがあります。うまくいっているときにはいじらないのが基本。



W9-2-7: 確認

共有フォルダの状況確認。①mac_shareディレクトリに移動し、②lsしてゲスト側で何もないことを確認。③ホスト側にも何もない。

The image shows a screenshot of the Oracle VM VirtualBox Manager interface. The main window displays the configuration for a virtual machine named 'BioLinux8'. The 'System' tab is selected, showing details like '名前: BioLinux8', 'オペレーティングシステム: Ubuntu (64-bit)', 'メインメモリー: 2048 MB', and '起動順序: フロッピー、光学、ハードディスク'. The 'ディスプレイ' tab shows 'ビデオメモリー: 12 MB' and 'リモートデスクトップサーバー: 無効'. The 'ストレージ' tab shows 'コントローラー: IDE', 'IDE セカンダリマスター: 光学ドライブ 空', and 'BioLinux8-disk1.vmdk (6.6GB, 100.0%)'. The 'オーディオ' tab shows 'ホストドライバー: Windows DirectSound' and 'コントローラー: ICH AC97'. The 'ネットワーク' tab is also visible.

In the foreground, a Windows File Explorer window is open to 'C:\Users\Wkadota\Desktop\share', showing an empty folder. A red arrow labeled '③' points to the address bar.

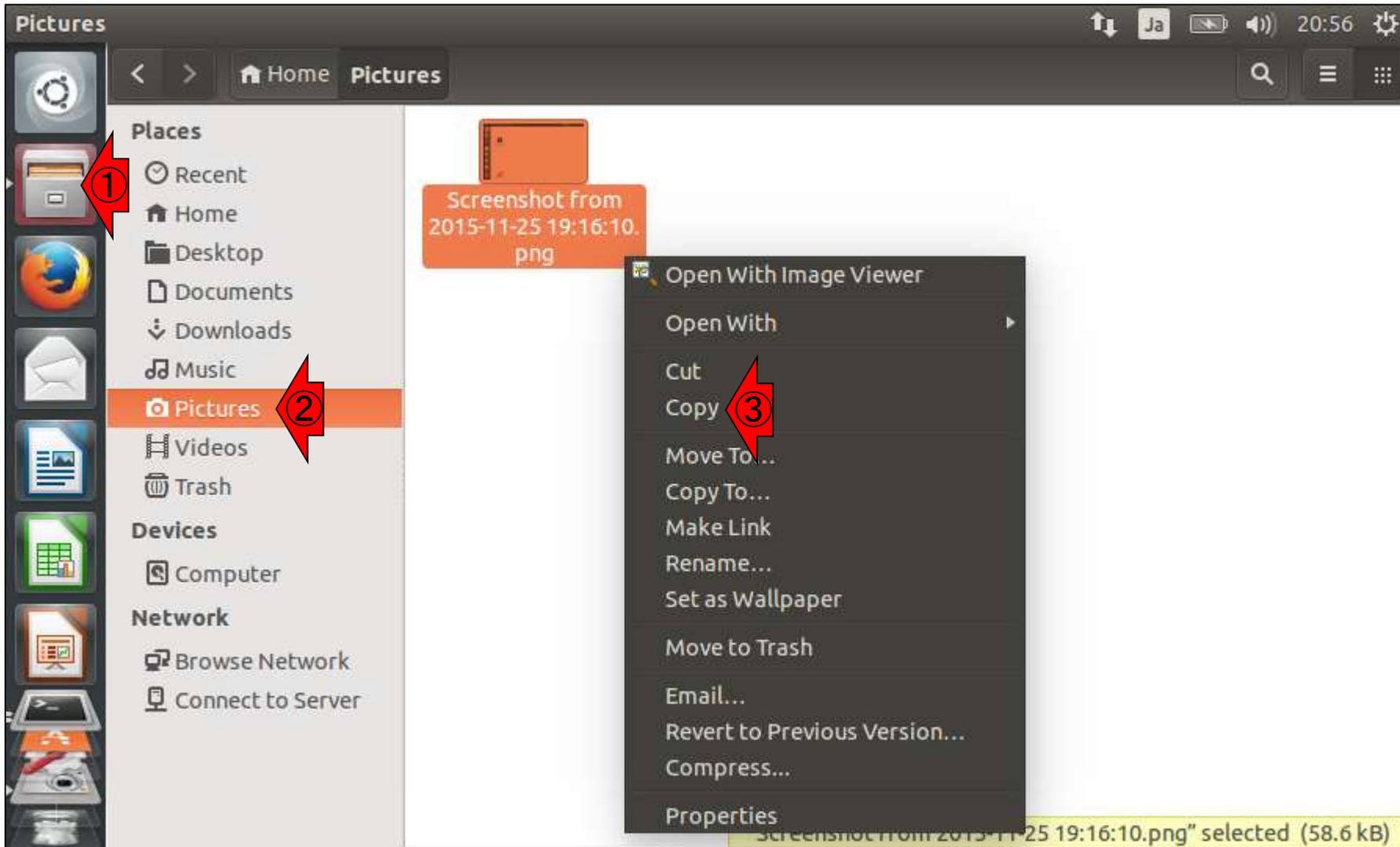
Overlaid on the VirtualBox window is a terminal window titled 'BioLinux8 [実行中] - Oracle VM VirtualBox'. The terminal shows the following commands and output:

```
iu@bielinux[~/Desktop/mac_share]
drwxr-xr-x  2 iu iu 4096 11月 17 13:44 Bio-Linear Documentation
drwxrwxr-x  2 iu iu 4096 11月 26 18:10 mac_share
drwxrwxr-x  2 iu iu 4096 11月 26 19:15 mac_share2
lrwxrwxrwx  1 iu iu  29 11月 17 13:44 Sample Data -> /usr/local/bioinf
/sampled/
iu@bielinux[Desktop] sudo mount -t vboxsf share mac_share [ 7:37午後 ]
[sudo] password for iu:
iu@bielinux[Desktop] ls -la [ 8:01午後 ]
total 16
drwx-----  5 iu  iu  4096 11月 26 19:24 .
drwxr-xr-x  19 iu  iu  4096 11月 26 19:24 ..
drwxr-xr-x  2 iu  iu  4096 11月 17 13:44 Bio-Linear Documentation
drwxrwxrwx  1 root root  0 11月 26 17:10 mac_share
drwxrwxr-x  2 iu  iu  4096 11月 26 19:15 mac_share2
lrwxrwxrwx  1 iu  iu  29 11月 17 13:44 Sample Data -> /usr/local/bi
oinf/sampled/
iu@bielinux[Desktop] cd mac_share [ 8:05午後 ]
iu@bielinux[mac_share] ls [ 8:40午後 ]
iu@bielinux[mac_share] [ 8:40午後 ]
```

Red arrows labeled '①' and '②' point to the 'cd mac_share' and 'ls' commands in the terminal, respectively.

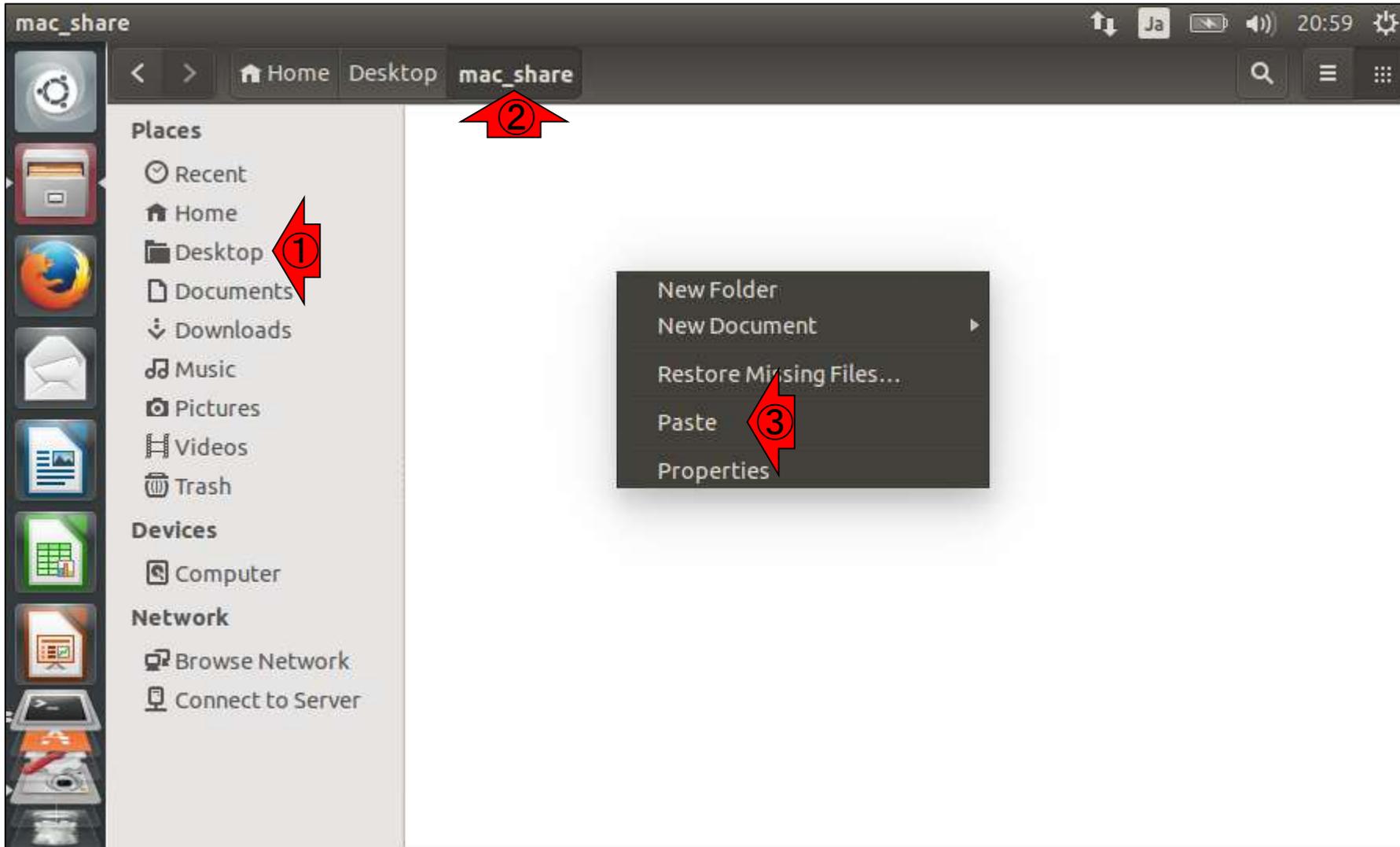
以前にスクリーンショットを保存していた②
Picturesフォルダ中のpngファイルを③コピー。

W9-2-8: ファイル名...

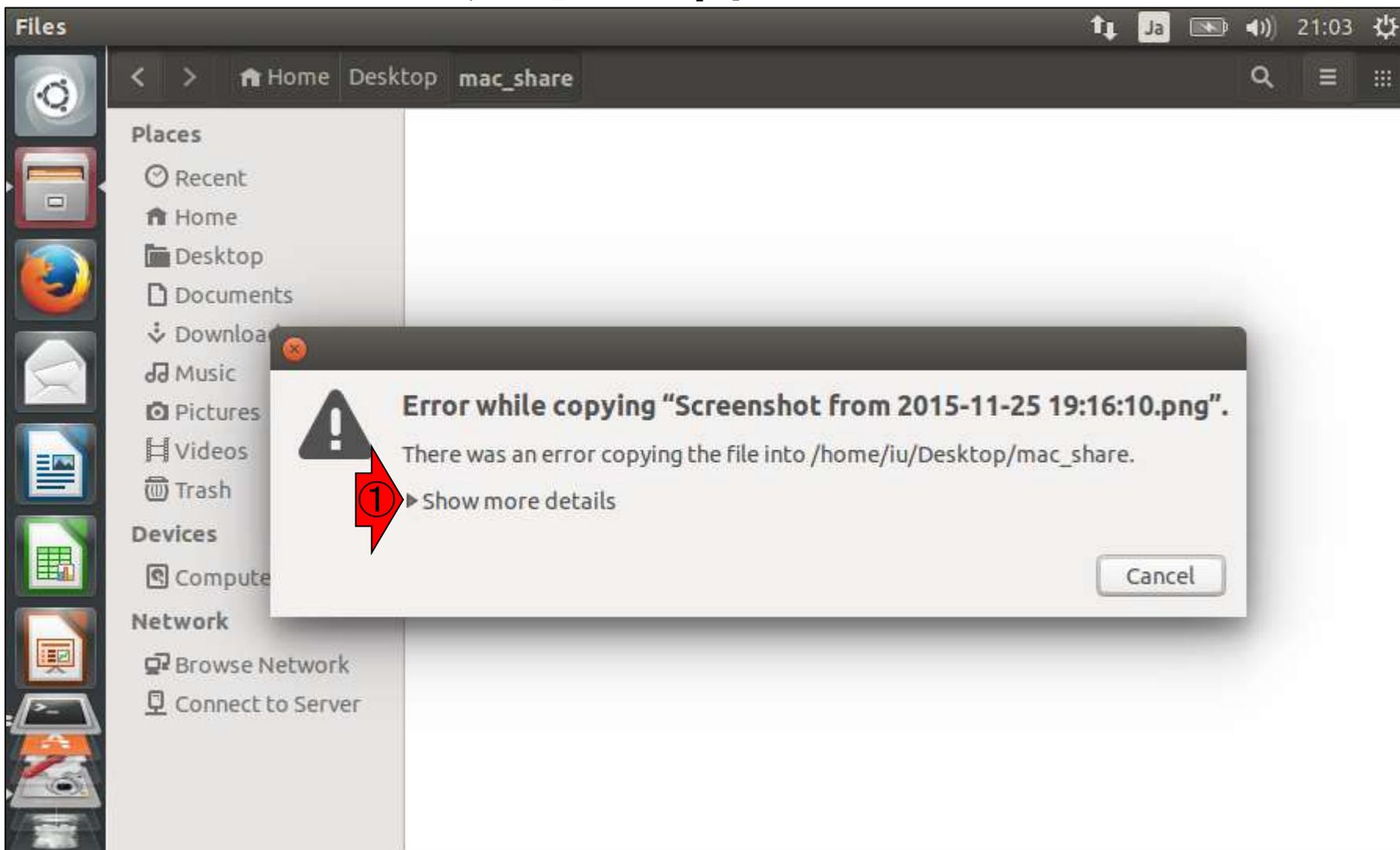


ゲストOSの①デスクトップ上にある
②mac_shareフォルダ上で③ペースト

W9-2-8: ファイル名...

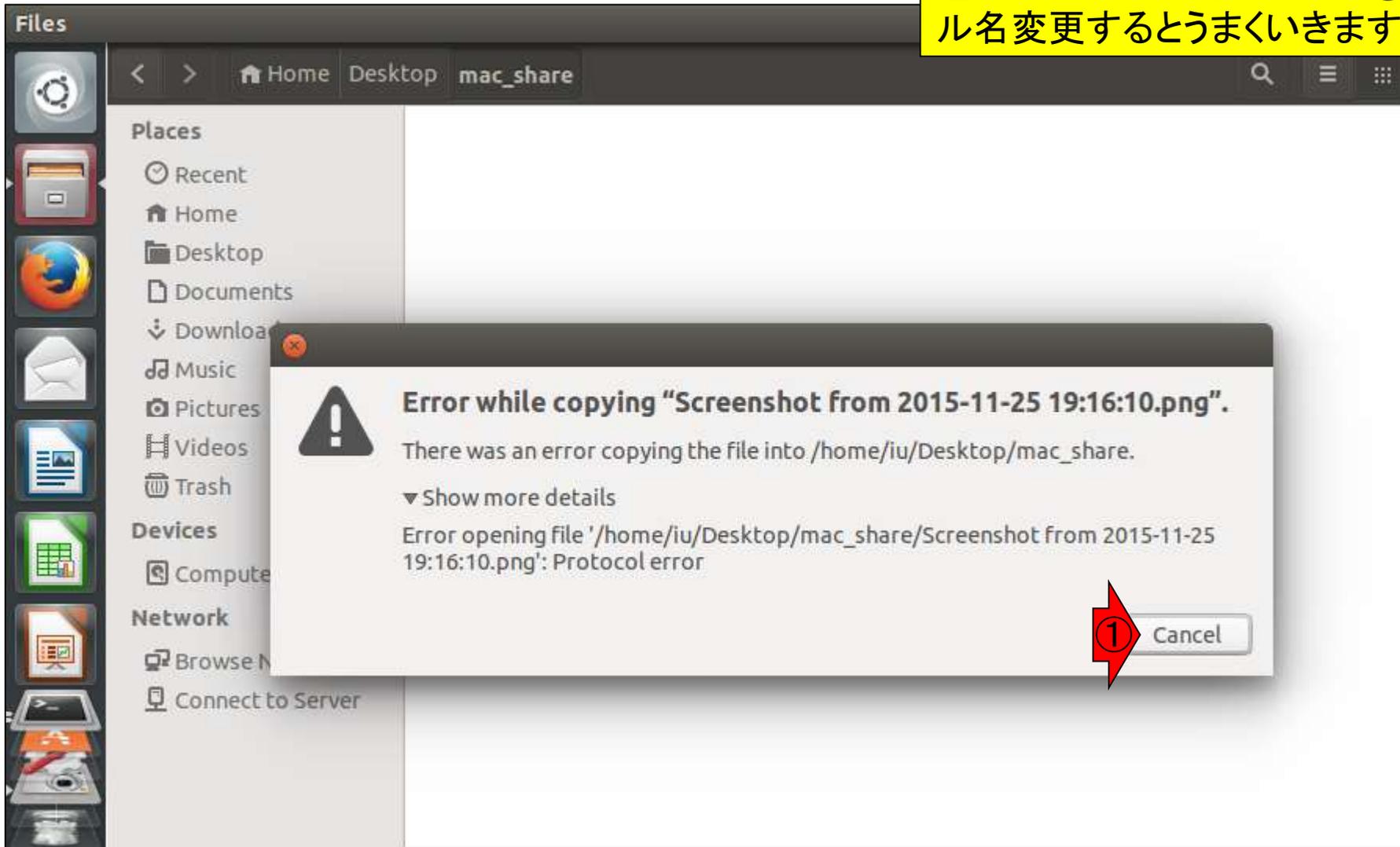


W9-2-8: ファイル名...



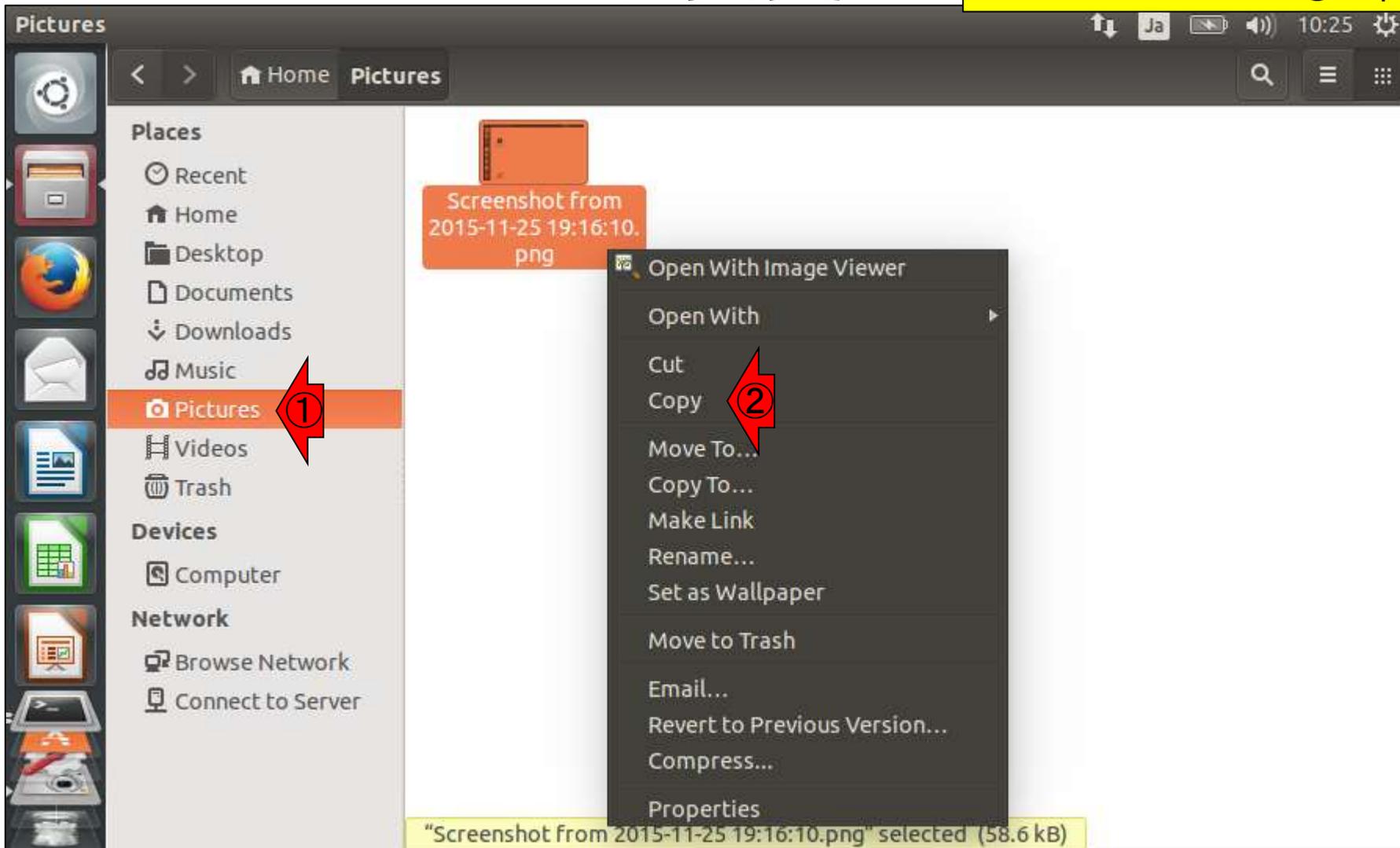
W9-2-8: ファイル名...

結論としては、「Screenshot from 2015-11-25 19:16:10.png」というファイル名に問題があるようです。例えばhoge.pngにファイル名変更するとうまくいきます。①Cancel

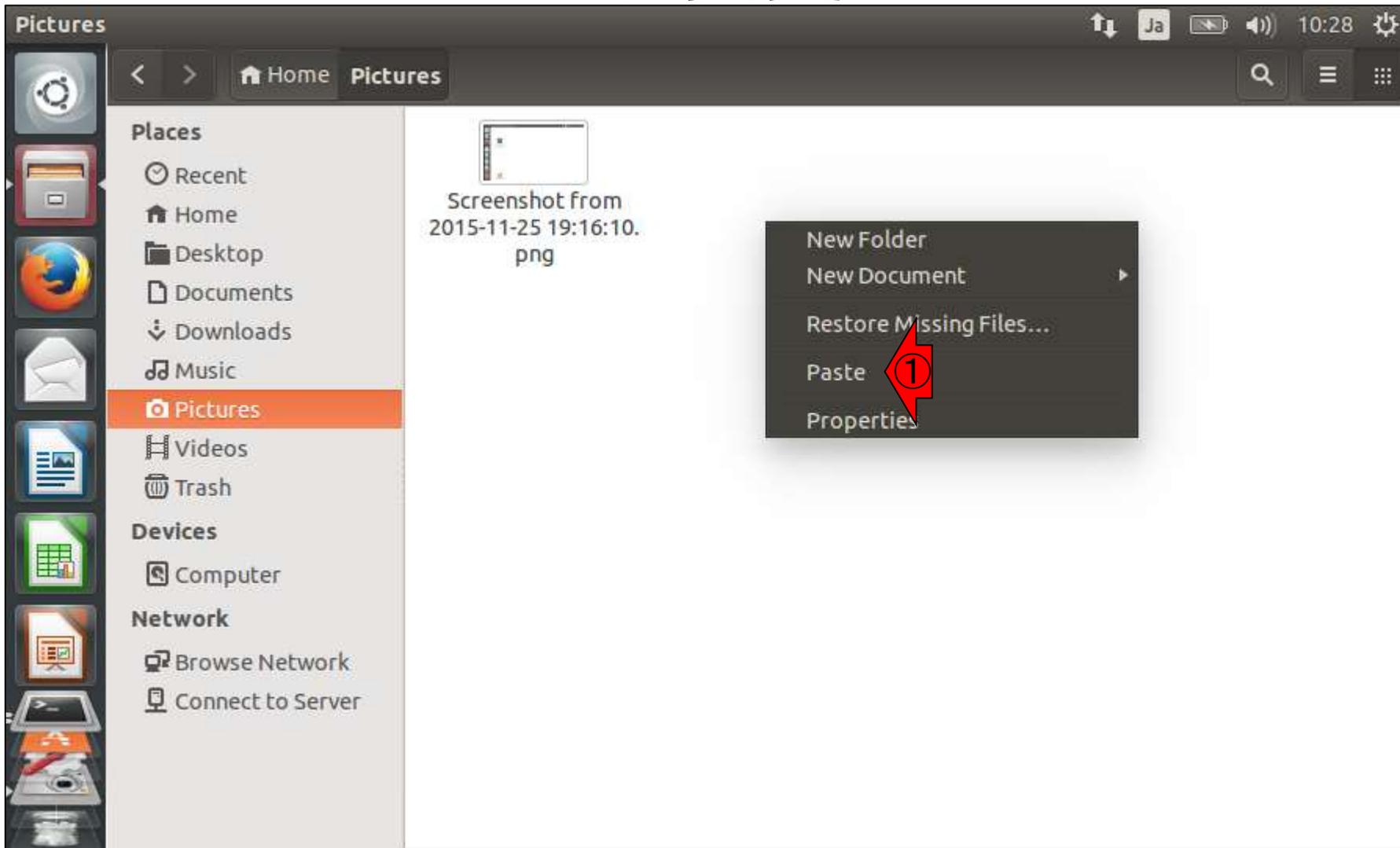


W9-3: ファイル名変更

① Pictures フォルダ中の「Screenshot from 2015-11-25 19:16:10.png」のアイコン上で右クリックして② Copy。

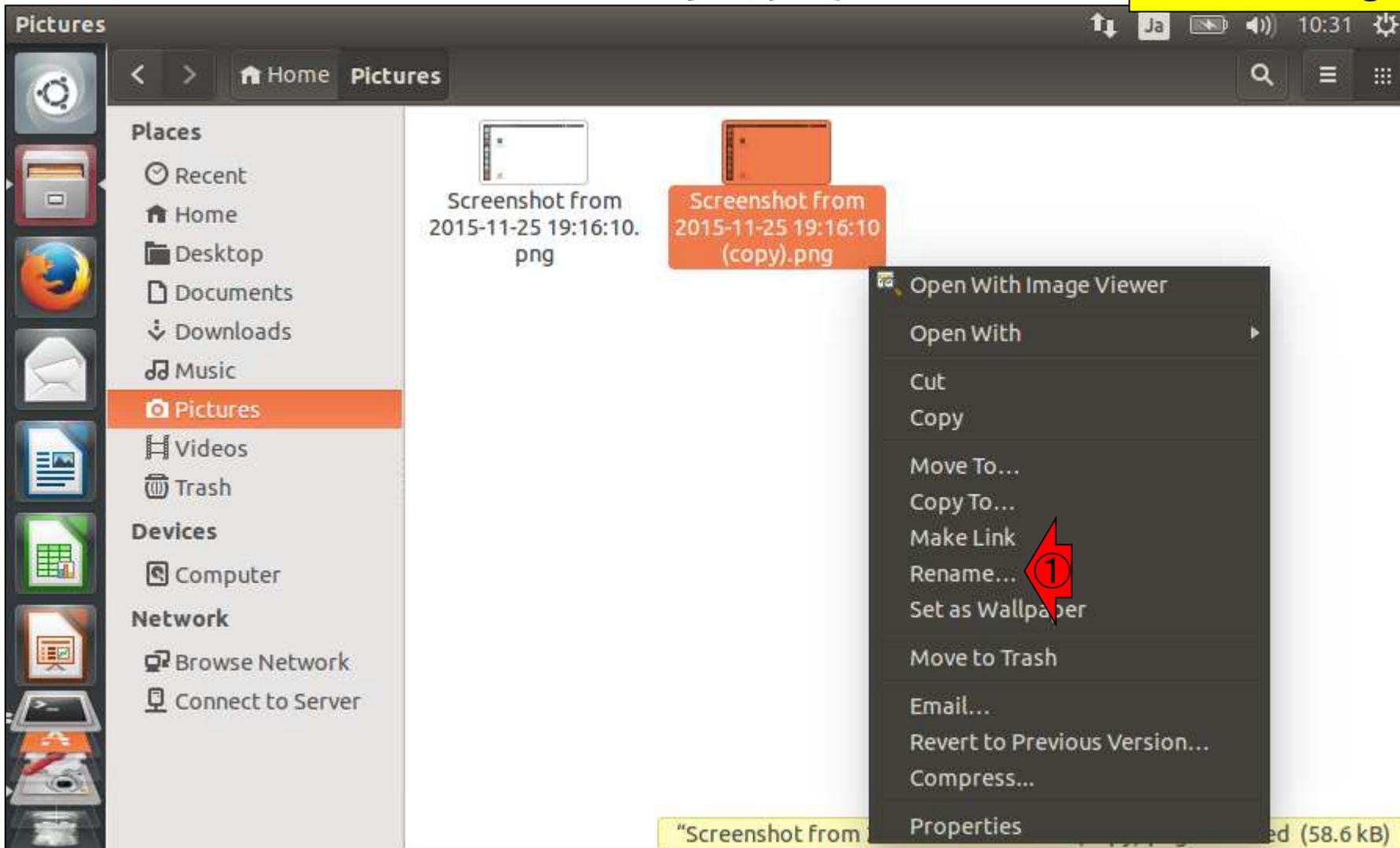


W9-3: ファイル名変更

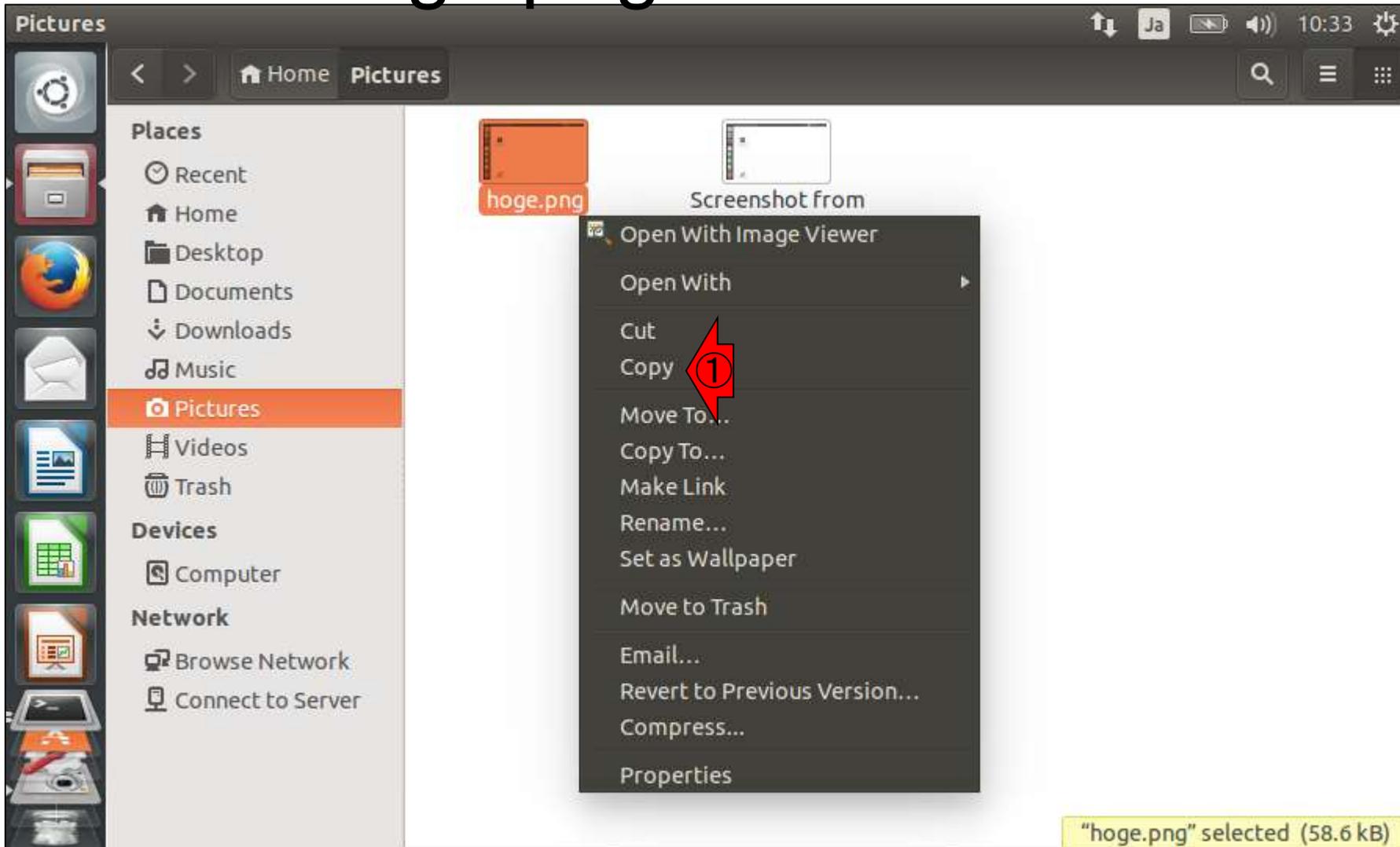


W9-3: ファイル名変更

作成されたファイルのアイコン上で、再び右クリックして①Renameし、hoge.pngとする。

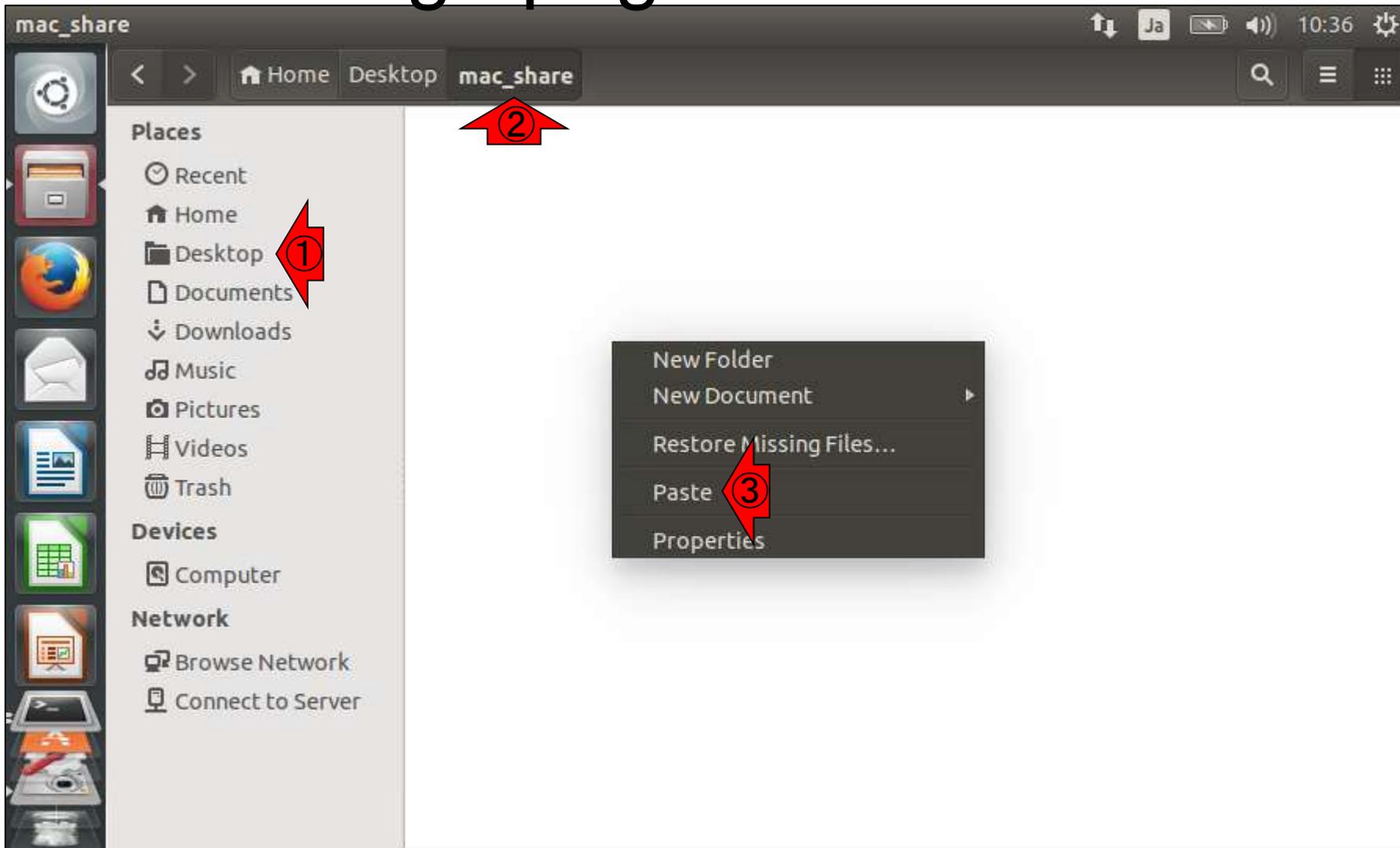


W9-3: hoge.pngでリトライ



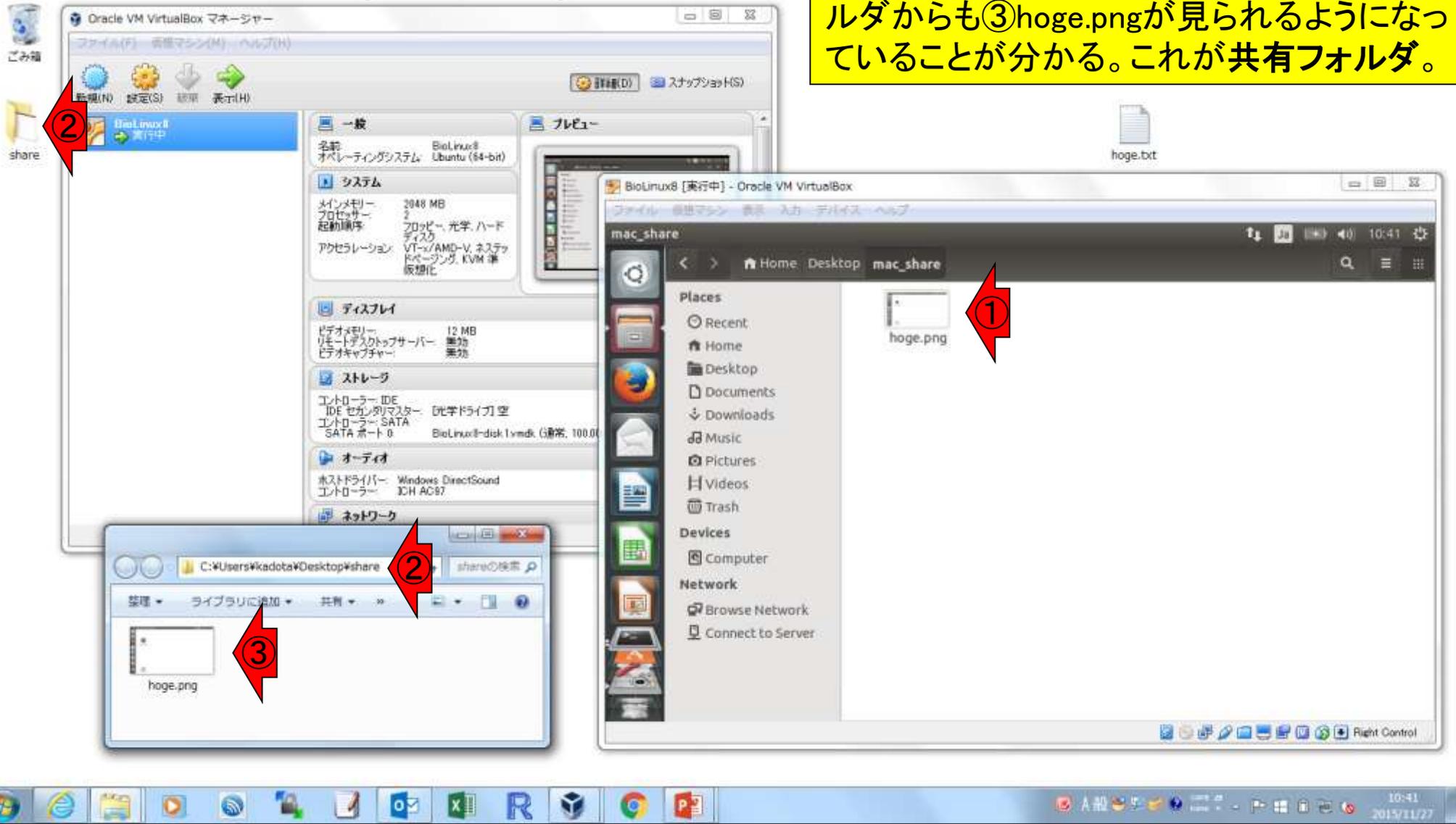
ゲストOSの①デスクトップ上にある
②mac_shareフォルダ上で③ペースト

W9-3: hoge.pngでリトライ

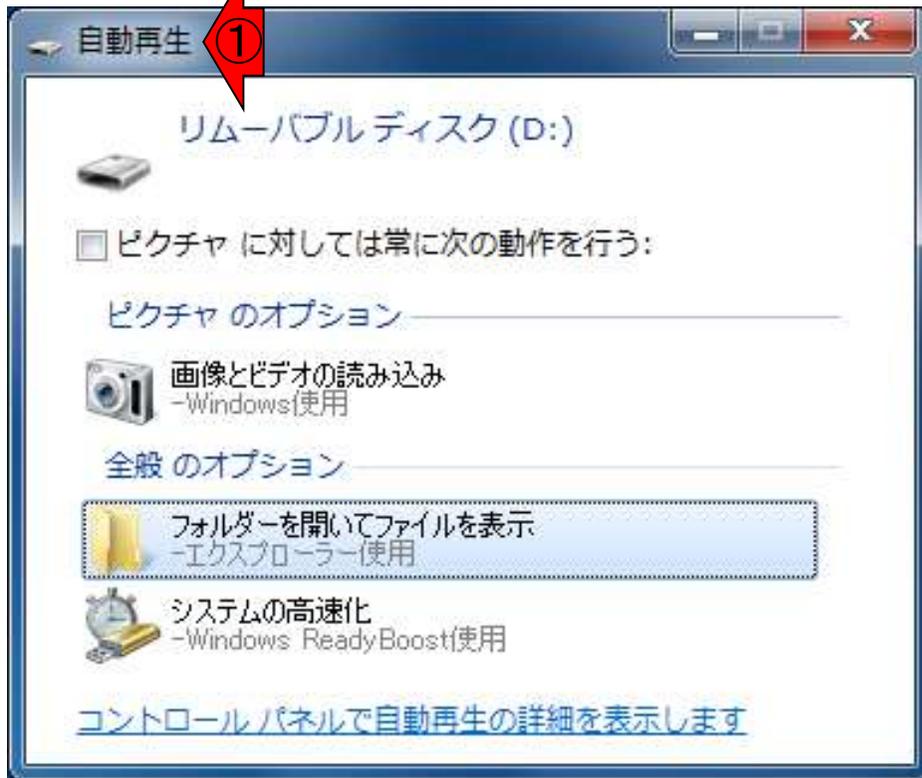


W9-3: hoge.pngで成功

ゲストOSのデスクトップ上のmac_shareフォルダに①hoge.pngが確かに存在する。また、ホストOSのデスクトップ上の②shareフォルダからも③hoge.pngが見られるようになっていいることが分かる。これが共有フォルダ。



W9-4: マウント?!



マウントの説明。USBメモリなどを差し込んだときに①自動再生というウィンドウが出る。Windowsの場合は、このようにデフォルトで自動的にマウント(オートマウント、とも呼ばれる)してくれる。

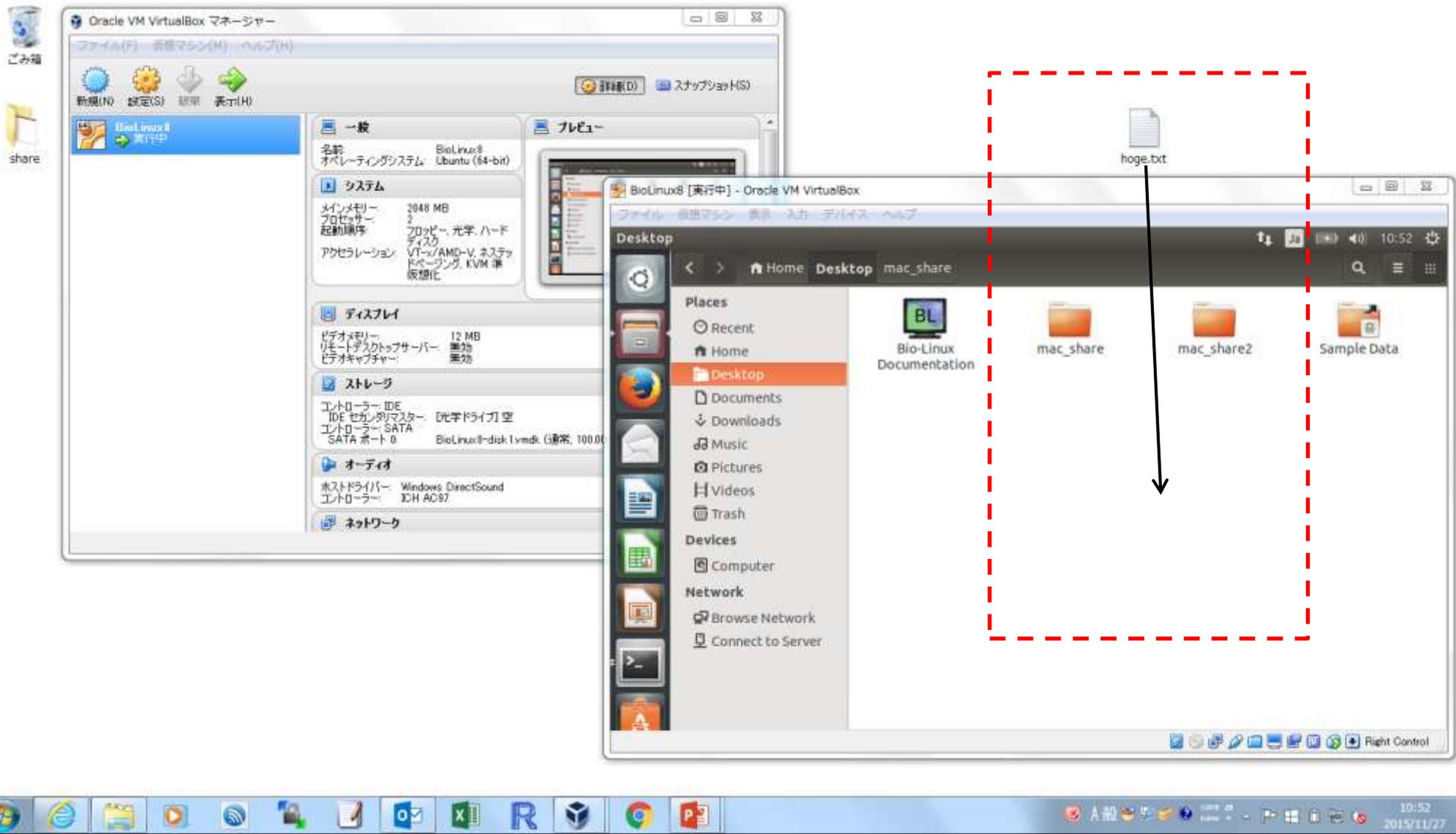
W9-4: マウント?!



USBメモリとPC間で何かデータのやり取りをしている最中に、手でいきなり引き抜くという物理的な取り外しは危険である。そのため、USBメモリを取り外したい場合には、一般にいきなり手で引き抜く(物理的な取り外し)前に、画面右下にあるタスクトレイから目的の「USBメモリの取り出し」を行う。これはPC側からの認識を解除(アンマウント、という)していることに相当し、物理的な取り外しの対比的な用語として、論理的な取り外しという表現がなされる。

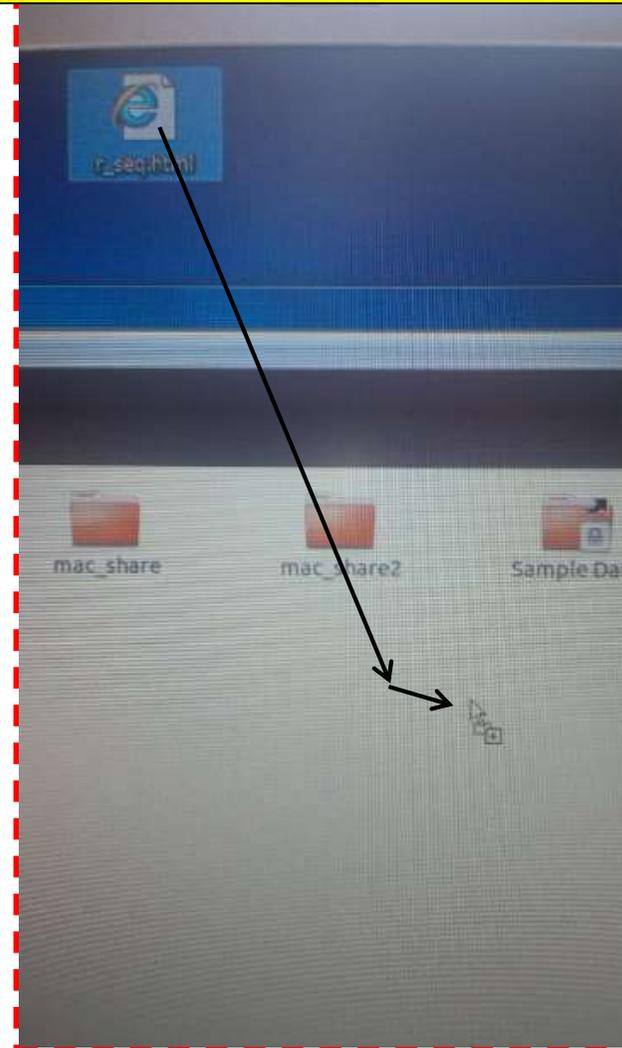
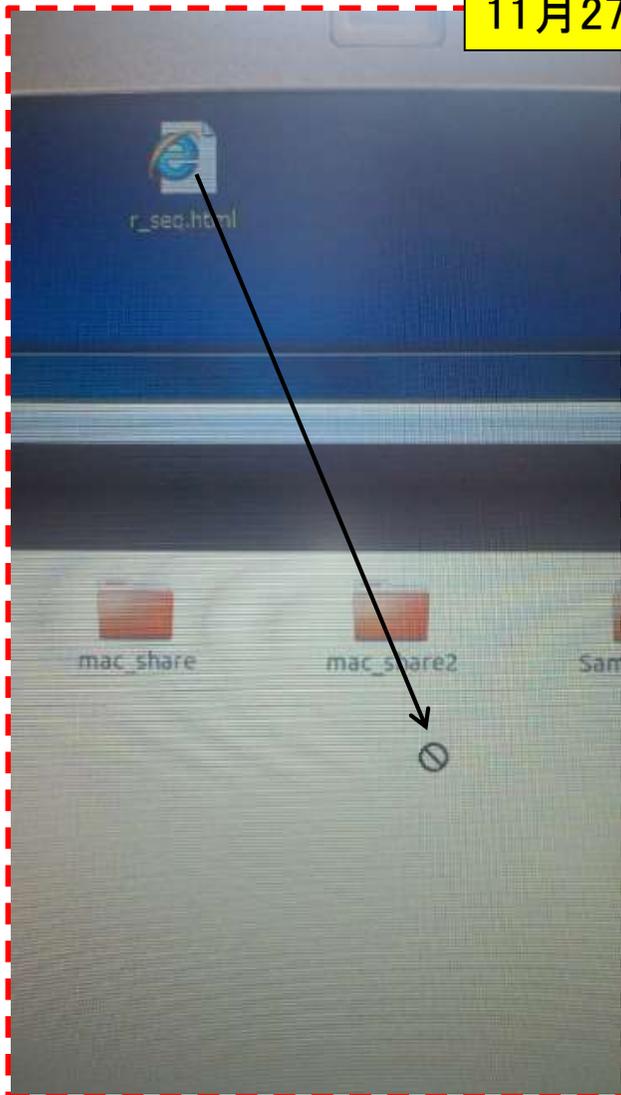
W9-5: D & D

ホストOSからゲストOSへドラッグ&ドロップ (D & D)ができるときとできないときがあるようだ。



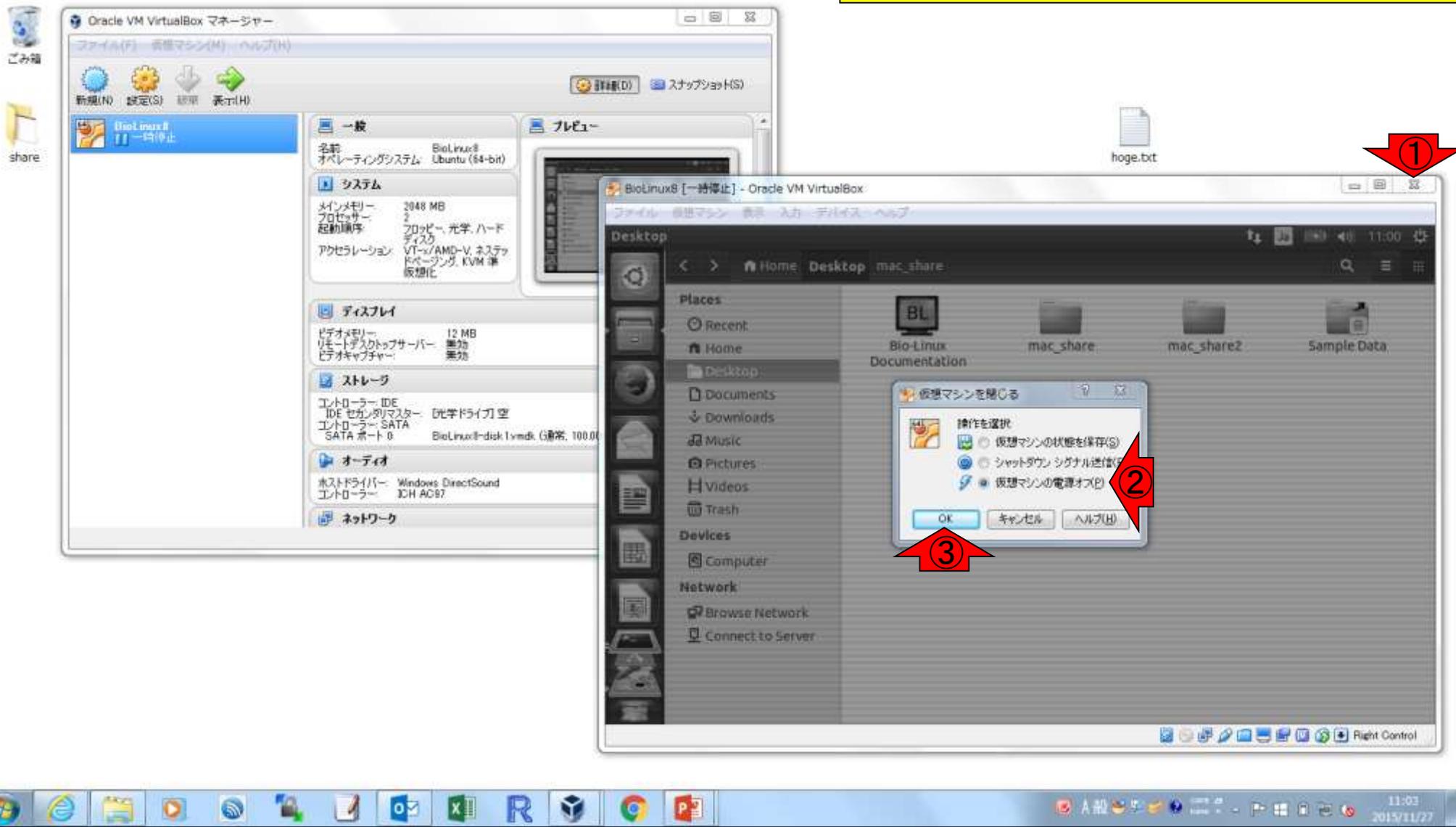
W9-5:D & D

ゲストOSのウィンドウ上で、左図のようにマウスカーソルが駐車禁止マークのようになっていると、ドロップ(ファイルを置く)することができないが、ちょっとカーソルの位置をずらしたりすれば右図のように置くことができるマークに変わる。(2015年11月27日にうまく再現できなかったので、昔のスライドを流用)



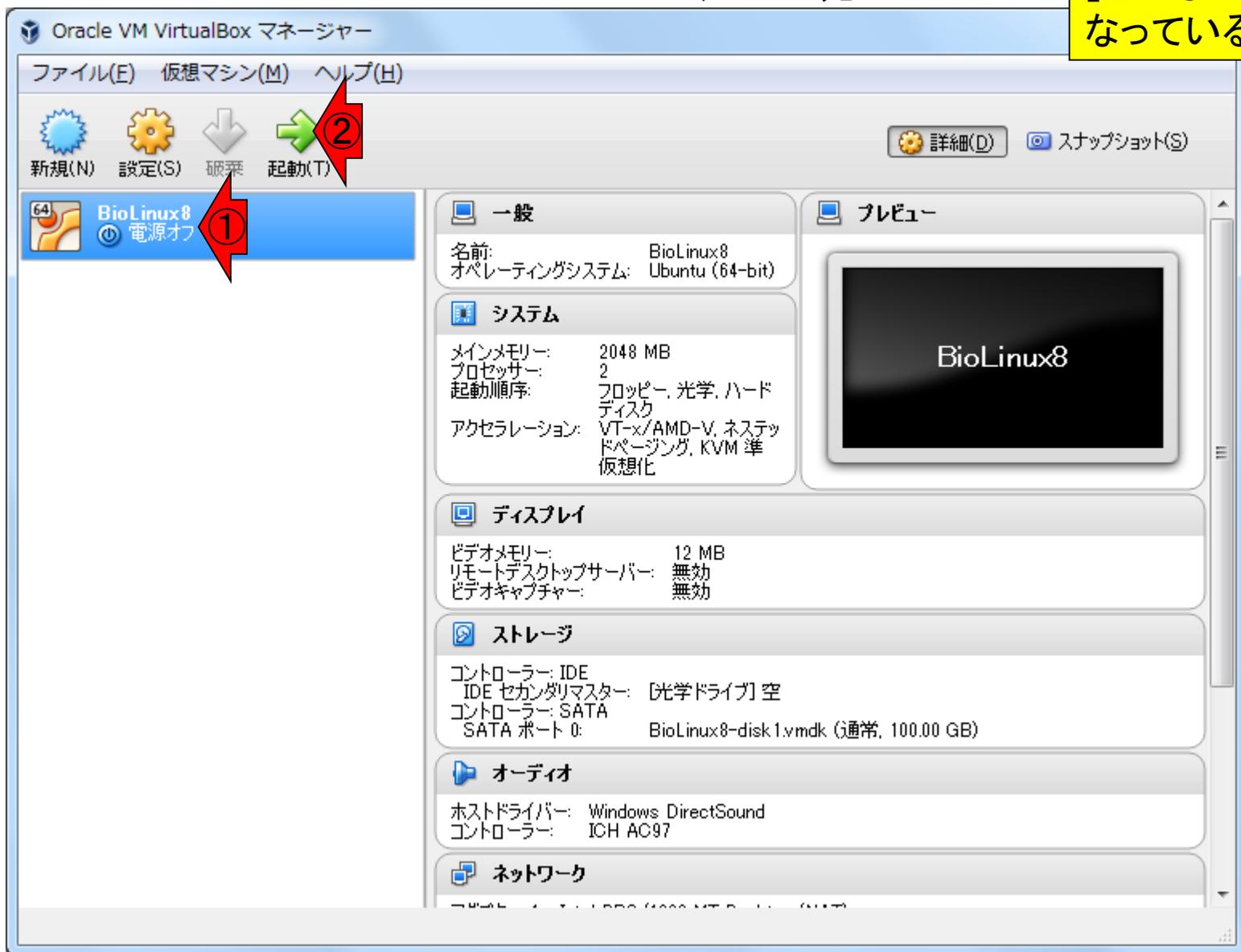
W9-6: 保存せずに終了

ゲストOSの①終了時に「仮想マシンの状態を保存」にせず、②仮想マシンの電源オフを選択して終了してみる。



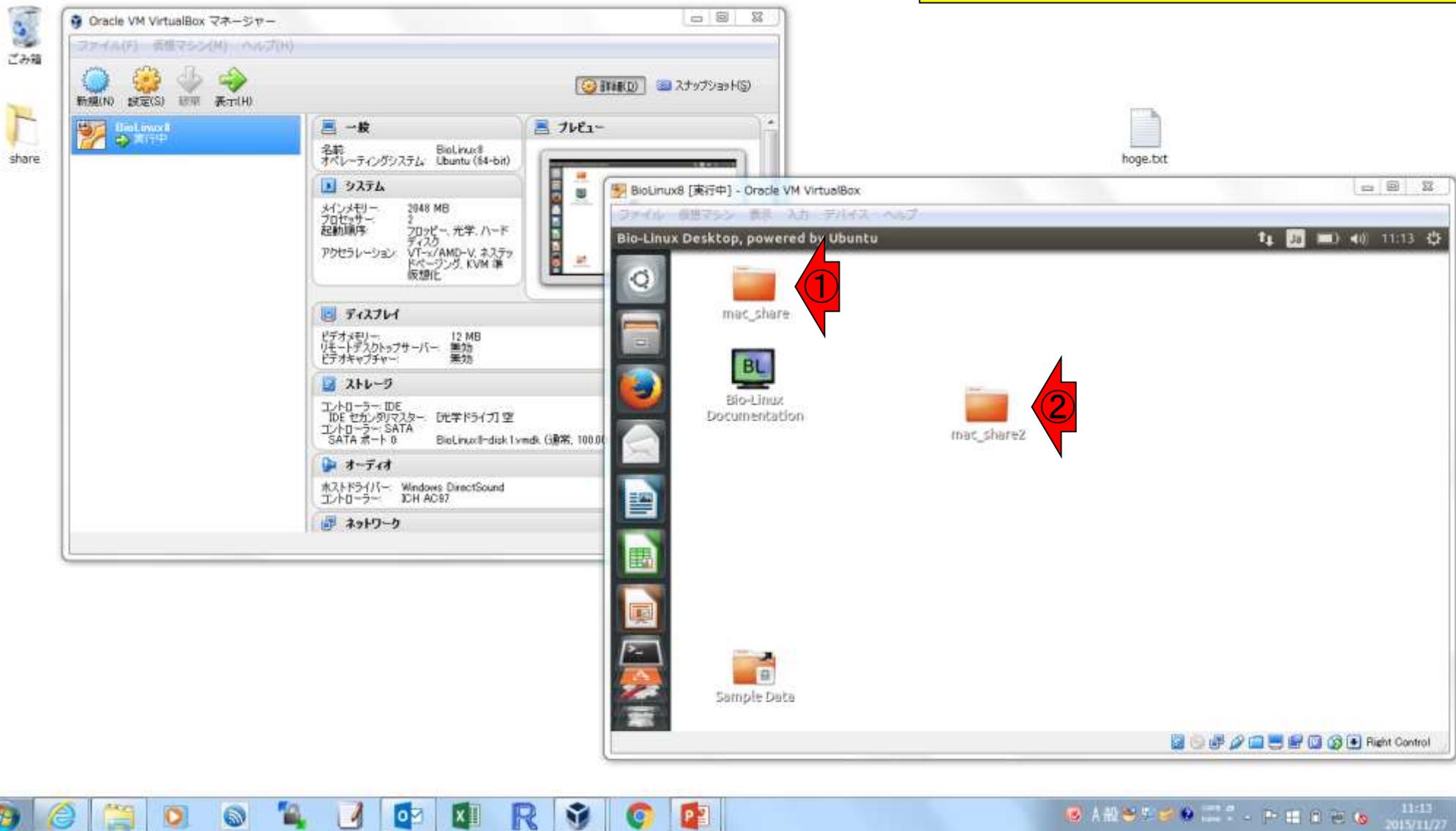
W9-6: BioLinux8起動

ゲストOS終了直後のVirtualBoxマネージャー。W8-9と違って、「保存」とはなっておらず、①電源オフになっていることがわかる。②起動。



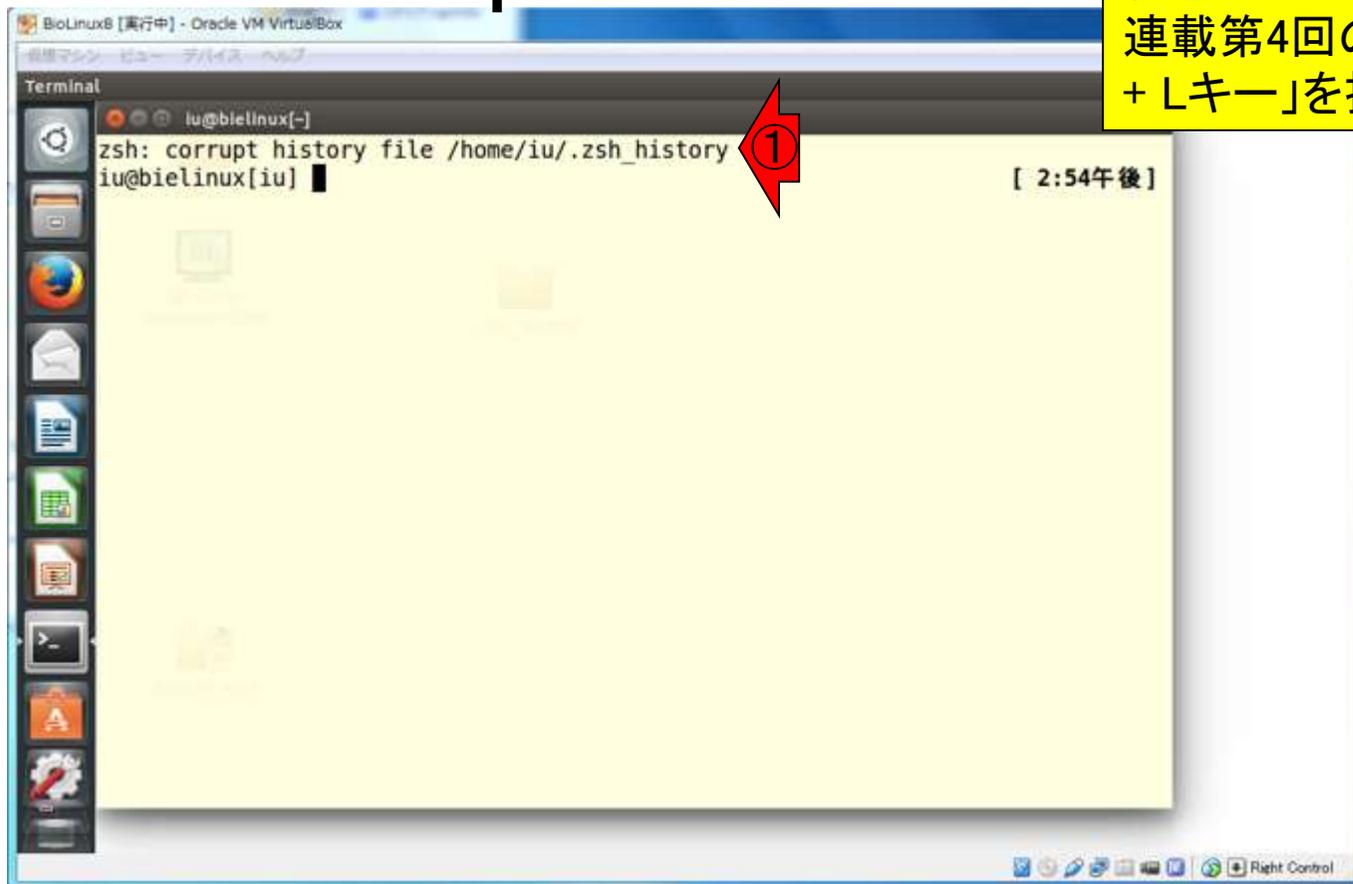
W9-6: 起動直後の状態

終了前に作成した①mac_shareや②mac_share2フォルダは残っていることがわかる。



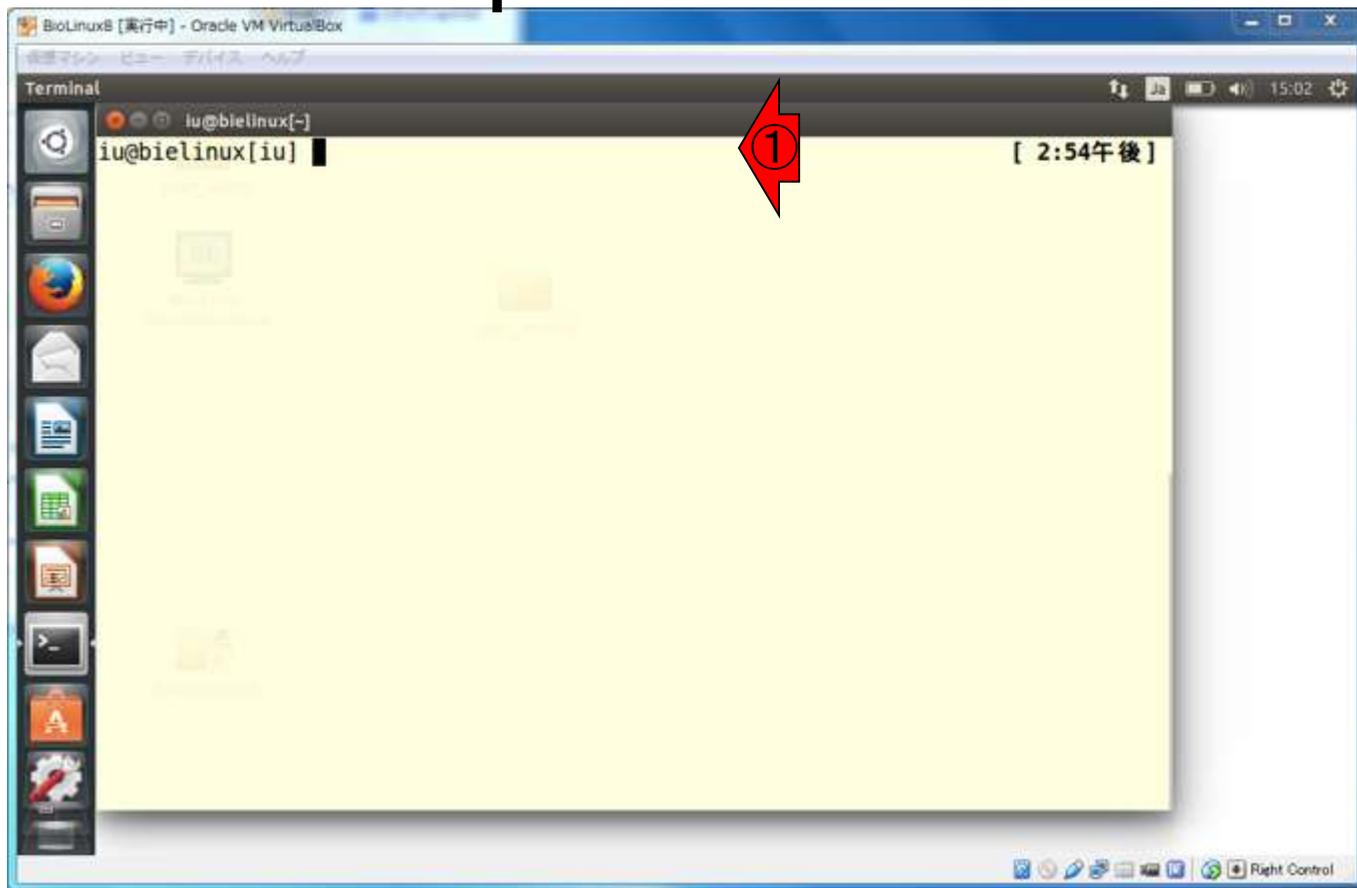
W9-6: Tips

ターミナル起動後にこのようなメッセージが出ることもあるが、本質的な問題ではないので初心者のうちは気にしなくてよい。対処法は、連載第4回のW10-2に示している。「CTRLキー + Lキー」を押すと画面がリフレッシュされる。



W9-6: Tips

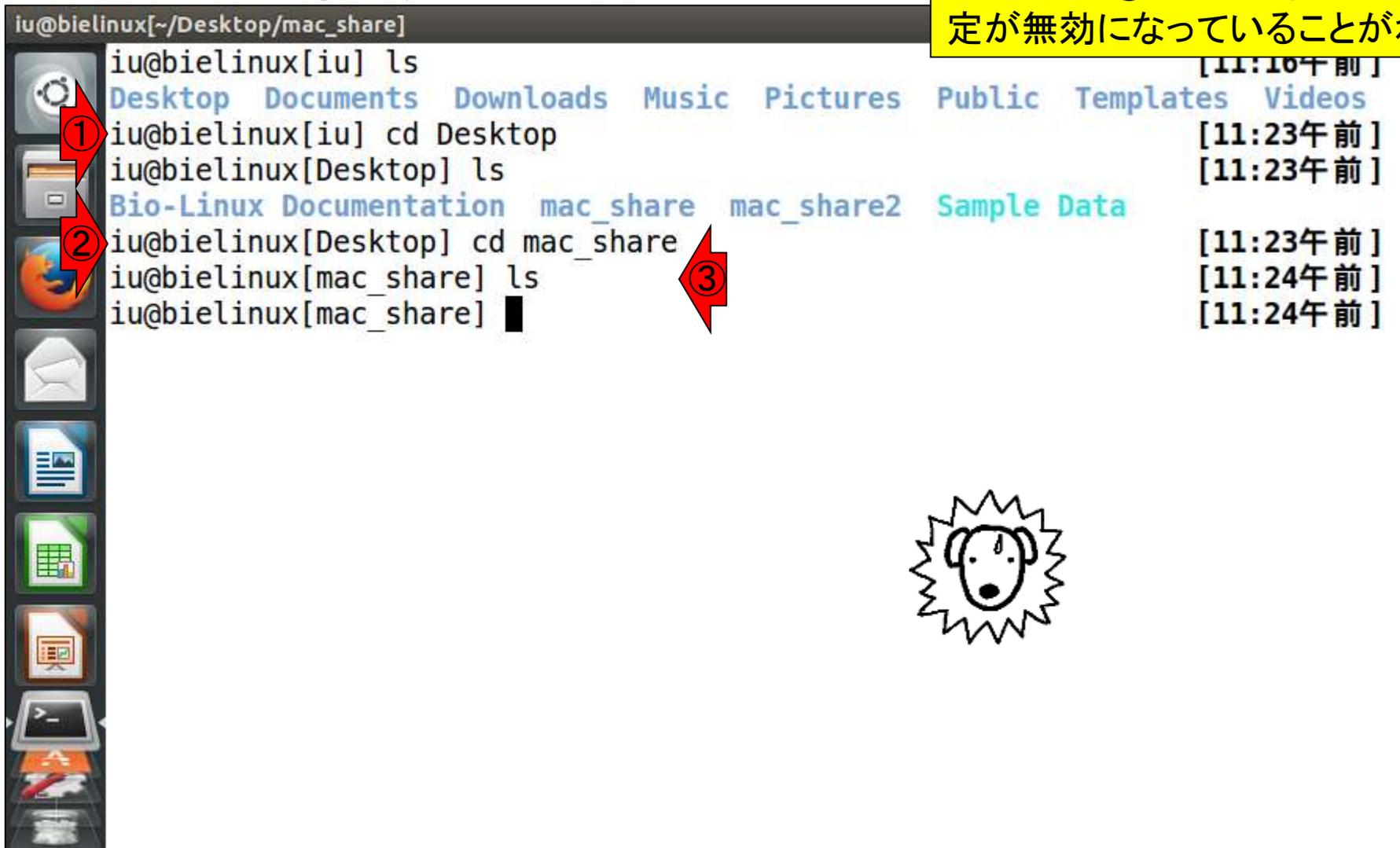
「CTRLキー + Lキー」を押して
画面がリフレッシュされた状態



W9-6: 設定が消える...

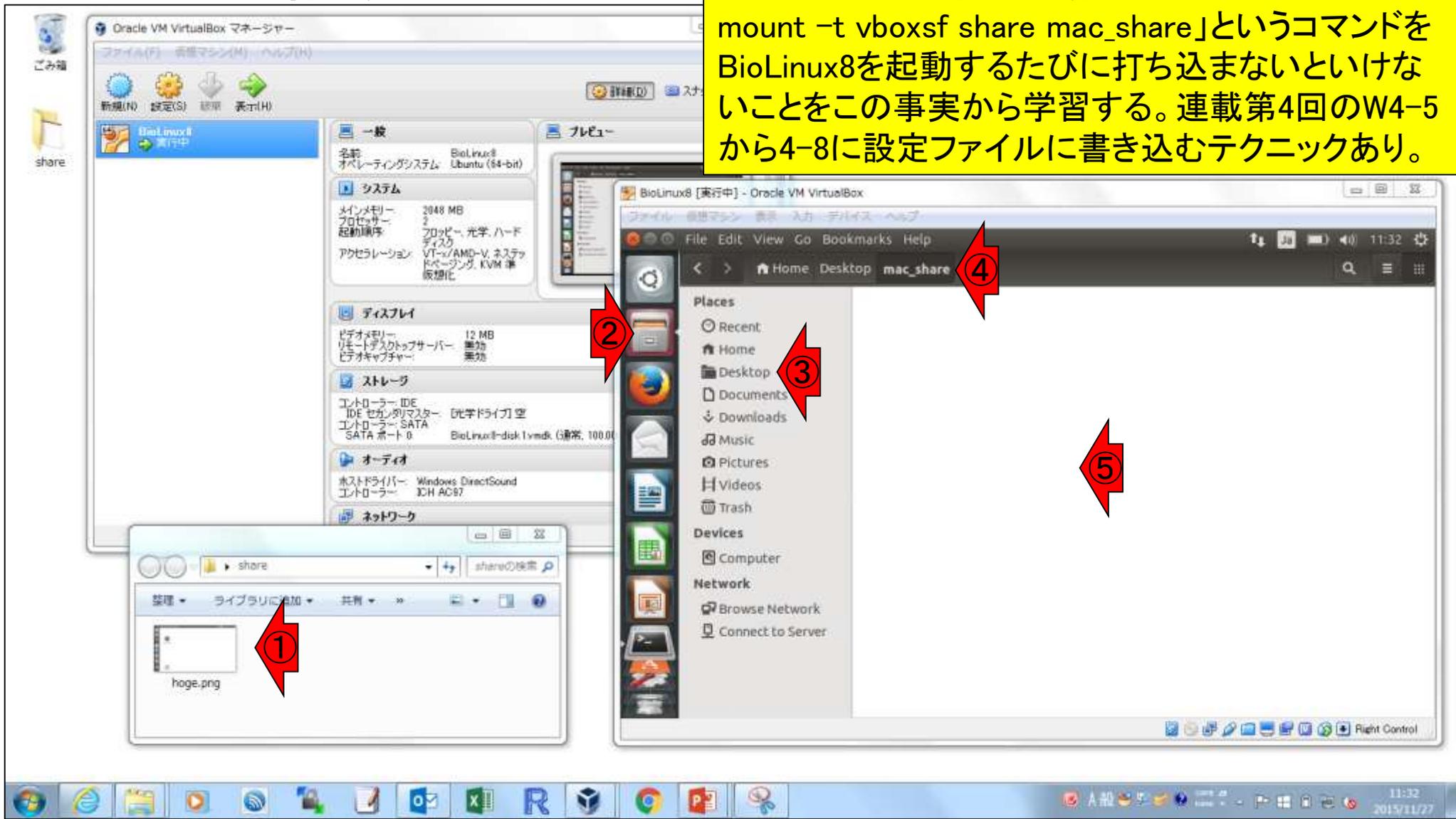
ターミナルを起動して前回共有フォルダとして設定した①Desktop上の②mac_shareに移動し、③lsした結果。共有フォルダ設定が無効になっていることがわかる。

```
iu@bielinux[~/Desktop/mac_share]
iu@bielinux[iu] ls
Desktop Documents Downloads Music Pictures Public Templates Videos
iu@bielinux[iu] cd Desktop
iu@bielinux[Desktop] ls
Bio-Linux Documentation mac_share mac_share2 Sample Data
iu@bielinux[Desktop] cd mac_share
iu@bielinux[mac_share] ls
iu@bielinux[mac_share] █
```



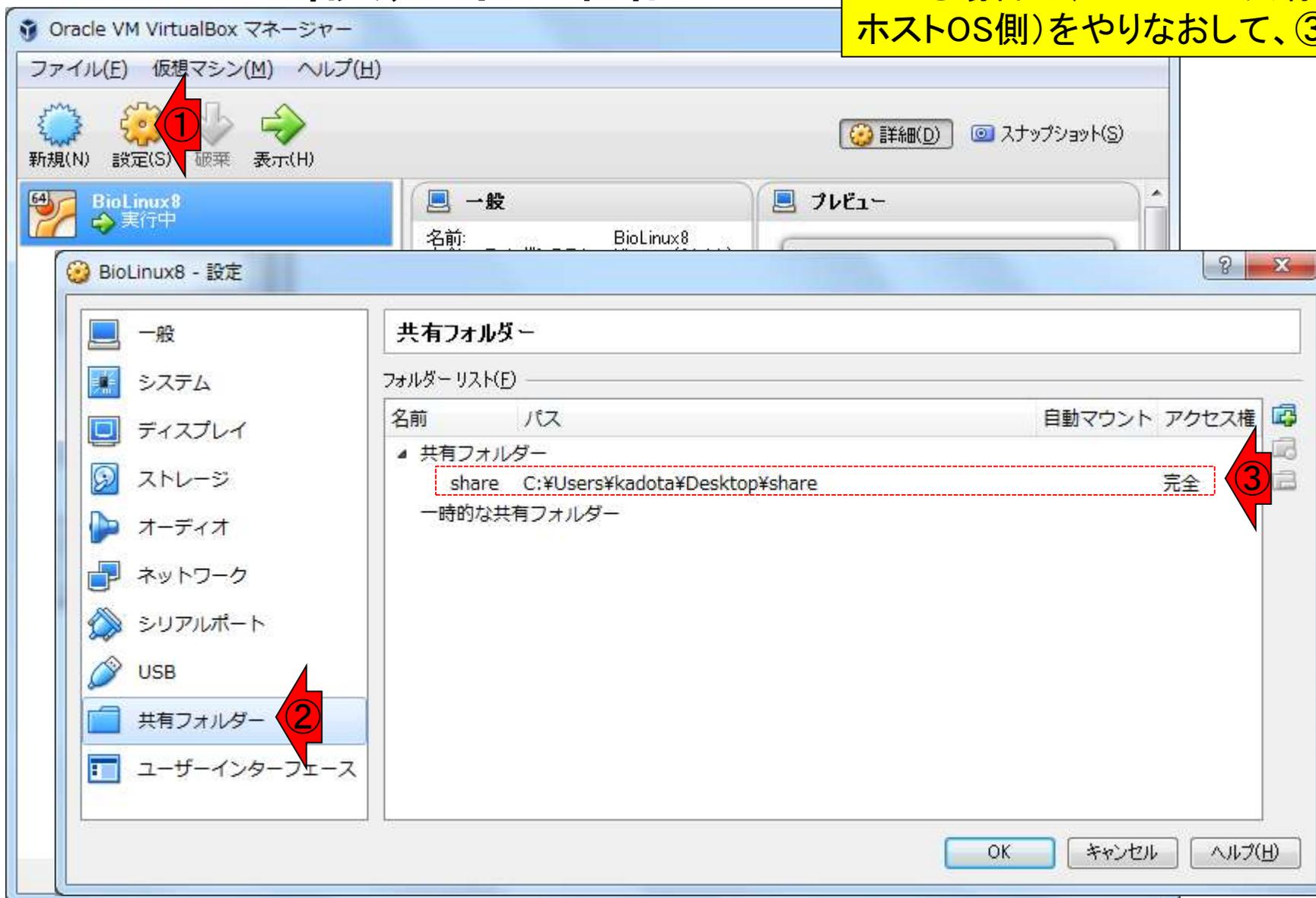
W9-6: 設定が消える

ホストOS上の「Desktop」 - 「share」にある①pngファイルが、ゲストOSのGUI画面でも⑤確かにない。共有フォルダを利用したい場合は、W9-2-5の「sudo mount -t vboxsf share mac_share」というコマンドをBioLinux8を起動するたびに打ち込まないといけないことをこの事実から学習する。連載第4回のW4-5から4-8に設定ファイルに書き込むテクニックあり。



W9-6: 設定再確認

②までやった段階で、③の赤枠で示された「共有フォルダ名とそのパス」が見られなくなっている場合は、W9-2-2の共有フォルダ設定(ホストOS側)をやりなおして、③のようにする。



W9-6: ゲスト側再設定

①「`sudo mount -t vboxsf share mac_share`」という赤矢印部分のコマンドを打ち込む。②パスワードを聞かれて打ち込んでも何も表示されないが、遠慮なく打ち込んでリターンキーを押す。ホストOSのデスクトップ上にあるshareフォルダをゲストOSのデスクトップ上のmac_shareフォルダとしてマウントさせているつもりだが赤下線のようにエラーが出ていることがわかる。

```
File Edit View Search Terminal Help
iu@bielinux[iu] ls
Desktop Documents Downloads Music Pictur
iu@bielinux[iu] cd Desktop
iu@bielinux[Desktop] ls
Bio-Linux Documentation mac_share mac_shar
iu@bielinux[Desktop] cd mac_share
iu@bielinux[mac_share] ls
① iu@bielinux[mac_share] sudo mount -t vboxsf share mac_share [11:24午前]
[sudo] password for iu:
/sbin/mount.vboxsf: mounting failed with the error: No such file or directo
ry
iu@bielinux[mac_share] [12:09午後]
```

W9-6: エラーの説明

「そんなファイルまたはディレクトリはない」というエラーメッセージが出ている。確かにその通り。①はゲストOSの「デスクトップ」-「mac_share」フォルダに移動してからmountコマンドを打ち込んでいるが、実際には、1つ上のディレクトリ(つまりデスクトップ)に移動してmac_shareディレクトリを見ることができ的状态で打ち込むべきコマンドだから。

```
File Edit View Search Terminal Help
iu@bielinux[iu] ls
Desktop Documents Downloads Music Pictur
iu@bielinux[iu] cd Desktop
iu@bielinux[Desktop] ls
Bio-Linux Documentation mac_share mac_shar
iu@bielinux[Desktop] cd mac_share
iu@bielinux[mac_share] ls
iu@bielinux[mac_share] sudo mount -t vboxsf share mac_share
[sudo] password for iu:
/sbin/mount.vboxsf: mounting failed with the error: No such file or directo
ry
iu@bielinux[mac_share]
```

[11:23午前]
[11:24午前]
[11:24午前]
[12:09午後]



W9-6: 正しい場所で...

①1つ上のディレクトリ(つまりデスクトップ)に移動して、②マウントコマンドを再度実行。③エラーは出なくなった。

```
iu@bielinux[iu] ls [11:16午前]
Desktop Documents Downloads Music Pictures Public Templates Videos
iu@bielinux[iu] cd Desktop [11:23午前]
iu@bielinux[Desktop] ls [11:23午前]
Bio-Linux Documentation mac_share mac_share2 Sample Data
iu@bielinux[Desktop] cd mac_share [11:23午前]
iu@bielinux[mac_share] ls [11:24午前]
iu@bielinux[mac_share] sudo mount -t vboxsf share mac_share [11:24午前]
[sudo] password for iu:
/sbin/mount.vboxsf: mounting failed with the error: No such file or directory
iu@bielinux[mac_share] pwd [12:09午後]
/home/iu/Desktop/mac_share
① iu@bielinux[mac_share] cd .. [ 7:04午後]
iu@bielinux[Desktop] pwd [ 7:04午後]
/home/iu/Desktop
② iu@bielinux[Desktop] sudo mount -t vboxsf share mac_share [ 7:04午後]
[sudo] password for iu:
iu@bielinux[Desktop] █ [ 7:04午後]
③
```

W9-6: 正しい場所で...

- ① mac_shareディレクトリに移動して、
- ②確かにホストOSのshareフォルダ中に存在するpngファイルを確認。

```
iu@bielinux[iu] ls [11:16午前]
Desktop Documents Downloads Music Pictures Public Templates Videos
iu@bielinux[iu] cd Desktop [11:23午前]
iu@bielinux[Desktop] ls [11:23午前]
Bio-Linux Documentation mac_share mac_share2 Sample Data
iu@bielinux[Desktop] cd mac_share [11:23午前]
iu@bielinux[mac_share] ls [11:24午前]
iu@bielinux[mac_share] sudo mount -t vboxsf share mac_share [11:24午前]
[sudo] password for iu:
/sbin/mount.vboxsf: mounting failed with the error: No such file or directory
iu@bielinux[mac_share] pwd [12:09午後]
/home/iu/Desktop/mac_share
iu@bielinux[mac_share] cd .. [ 7:04午後]
iu@bielinux[Desktop] pwd [ 7:04午後]
/home/iu/Desktop
iu@bielinux[Desktop] sudo mount -t vboxsf share mac_share [ 7:04午後]
[sudo] password for iu:
iu@bielinux[Desktop] cd mac_share [ 7:04午後]
iu@bielinux[mac_share] ls [ 7:10午後]
hoge.png
iu@bielinux[mac_share] [ 7:11午後]
```



W9-6: Protocol error...

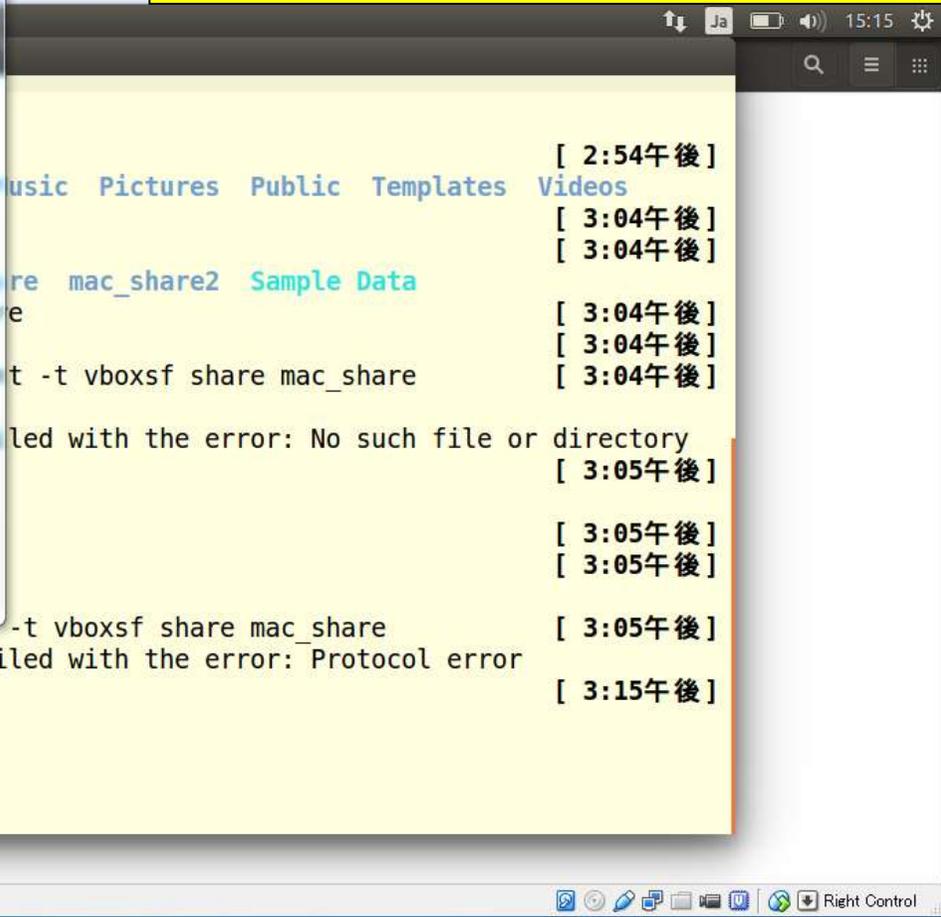
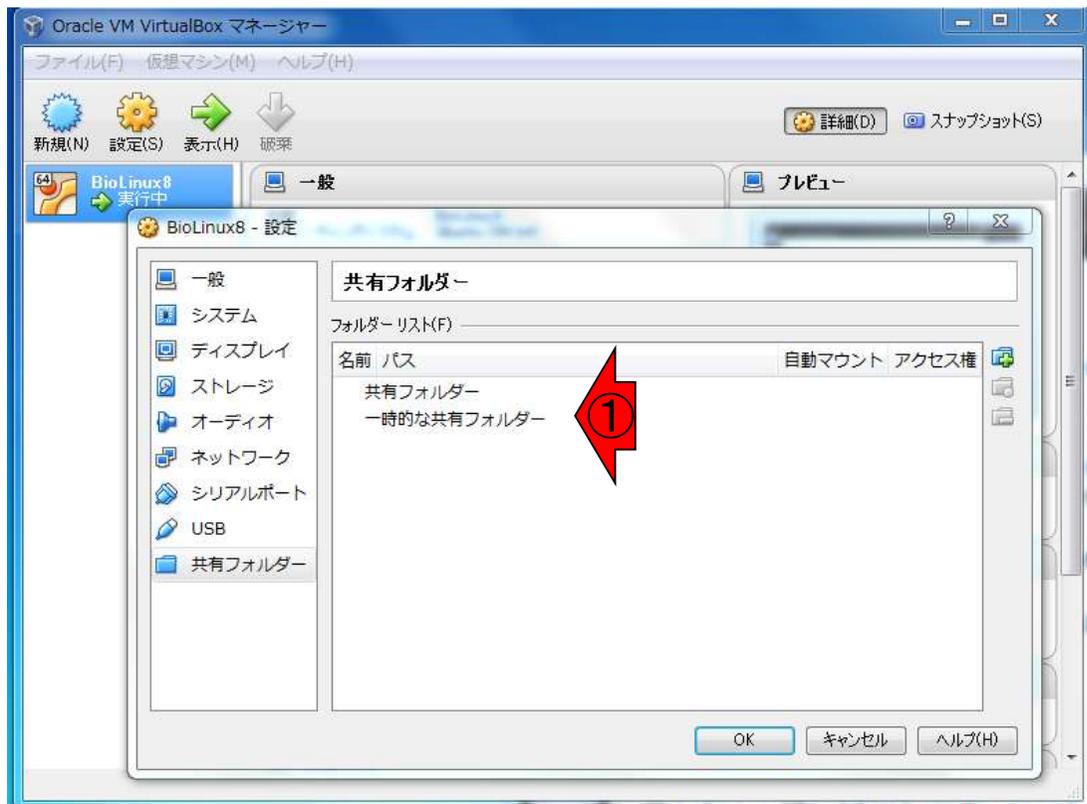
打つべき場所は正しいはずなのに、
①Protocol errorというエラーが出るときは、VirtualBoxマネージャー画面の「設定」-「共有フォルダー」でshareフォルダを共有フォルダにできているかを再度チェック。

```
Terminal File Edit View Search Terminal Help
iu@bielinux[~/Desktop]
iu@bielinux[iu] ls
Desktop Documents Downloads Music Pictures Public Te [ 3:04午後]
iu@bielinux[iu] cd Desktop [ 3:04午後]
iu@bielinux[Desktop] ls
Bio-Linux Documentation mac_share mac_share2 Sample Data [ 3:04午後]
iu@bielinux[Desktop] cd mac_share [ 3:04午後]
iu@bielinux[mac_share] ls [ 3:04午後]
iu@bielinux[mac_share] sudo mount -t vboxsf share mac_share [ 3:04午後]
[sudo] password for iu:
/sbin/mount.vboxsf: mounting failed with the error: No such file or directory [ 3:05午後]
iu@bielinux[mac_share] pwd
/home/iu/Desktop/mac_share
iu@bielinux[mac_share] cd .. [ 3:05午後]
iu@bielinux[Desktop] pwd [ 3:05午後]
/home/iu/Desktop
iu@bielinux[Desktop] sudo mount -t vboxsf share mac_share 3:05午後]
/sbin/mount.vboxsf: mounting failed with the error: Protocol error ①
iu@bielinux[Desktop] █ 3:06午後]
```

①

W9-6: Protocol error...

打つべき場所は正しいはずなのに、Protocol errorというエラーが出るときは、VirtualBoxマネージャー画面の「設定」-「共有フォルダー」でshareフォルダを共有フォルダにできているかを再度チェック。①は設定が消えている例。



W10-1: CTRL + L

新規ターミナル起動直後と同じ状態にするやり方。①cdでホームディレクトリに移動(その後のpwdはただの確認)。②「CTRLキー + Lキー」を押すと画面がリフレッシュされる。

```
iu@bielinux[Desktop] ls [11:23午前]
Bio-Linux Documentation mac_share mac_share2 Sample Data
iu@bielinux[Desktop] cd mac_share [11:23午前]
iu@bielinux[mac_share] ls [11:24午前]
iu@bielinux[mac_share] sudo mount -t vboxsf share mac_share [11:24午前]
[sudo] password for iu:
/sbin/mount.vboxsf: mounting failed with the error: No such file or directory
iu@bielinux[mac_share] pwd [12:09午後]
/home/iu/Desktop/mac_share
iu@bielinux[mac_share] cd .. [ 7:04午後]
iu@bielinux[Desktop] pwd [ 7:04午後]
/home/iu/Desktop
iu@bielinux[Desktop] sudo mount -t vboxsf share mac_share [ 7:04午後]
[sudo] password for iu:
iu@bielinux[Desktop] cd mac_share [ 7:04午後]
iu@bielinux[mac_share] ls [ 7:10午後]
hoge.png
iu@bielinux[mac_share] cd [ 7:11午後]
iu@bielinux[iu] pwd [ 7:23午後]
/home/iu
iu@bielinux[iu] [ 7:23午後]
```



W10-1: CTRL + L

「CTRLキー + Lキー」をした後の状態。「clear」というコマンドを打ち込むのもよい。



W10-2: 図1

```
iu@bielinux[~/Desktop/hoge]
iu@bielinux[iu] ls
Desktop Documents Downloads Music Pictu
iu@bielinux[iu] cd Desktop
iu@bielinux[Desktop] ls
Bio-Linux Documentation mac share mac_sha
iu@bielinux[Desktop] mkdir hoge
iu@bielinux[Desktop] cd hoge
iu@bielinux[hoge] ls
iu@bielinux[hoge] pwd
/home/iu/Desktop/hoge
iu@bielinux[hoge] whoami
iu
iu@bielinux[hoge] █
```

本文中の図1の詳細説明。①赤下線部分は「ユーザ名@PC名」となっている。②はホームディレクトリ(~)から、カレントディレクトリまでの相対パスを示している。この表示はディレクトリ移動のたびに変遷する。③角括弧[]内の文字(iu, Desktop, hoge)はカレントディレクトリの名前を示し、②の部分の表示と連動している。④のpwd実行結果(/home/iu/Desktop/hoge)がカレントディレクトリの絶対パスに相当する。

```
[11:03午後]
[11:03午後]
[11:03午後]
[11:03午後]
```

図1 Bio-Linux 8のターミナル画面。PC名はbielinux、ユーザ名はiu。ゲストOSのデスクトップ上にhogeフォルダを作成する④のmkdirコマンドを含め、計8つのコマンドを実行している。

W10-3-1 : download

ホストOSのウェブブラウザIEを開いて、ホストOSのデスクトップのhogeに保存しようとしているところ。ここでは①Ensembl release 22の*L. casei* 12Aへのリンクを貼っているが、2015年11月現在はrelease 29となっていてファイル名も異なるので注意。Lhaplusのようなフリーの圧縮/解凍ソフトウェアで拡張子gzがついた②のgzip圧縮ファイルを解凍。

データ取得とチェックサム:

- [Ensembl Genomes: Flicek et al., Nucleic Acids Res., 2014](#)
- [Lactobacillus casei 12A \(Taxonomy ID: 1051650\): Broadbent et al., BMC Genomics, 2012](#)
- [Ensembl Bacteria - Lactobacillus casei 12A - Download DNA sequence](#)
[Lactobacillus casei 12a.GCA_000309565.1.22.dna.toplevel.fa.gz](#)
- [Toplevelの概念を説明しているところ](#)
- [Release note](#)
- [Ensembl Bacteria - Lactobacillus casei](#)
- [Ensembl Bacteria](#)
 - [Lactobacillus](#)
 - [Escherichia](#)
 - ...
- [Ensembl Fungi](#)
 - [Saccharomyces](#)
 - [Schizosaccharomyces](#)

FTP ディレクトリ /pub/bacteria/release-22/fasta/bacteria_15_collection/lactobacillus_casei_12a/dna/ / ftp.ensemblgenomes.org

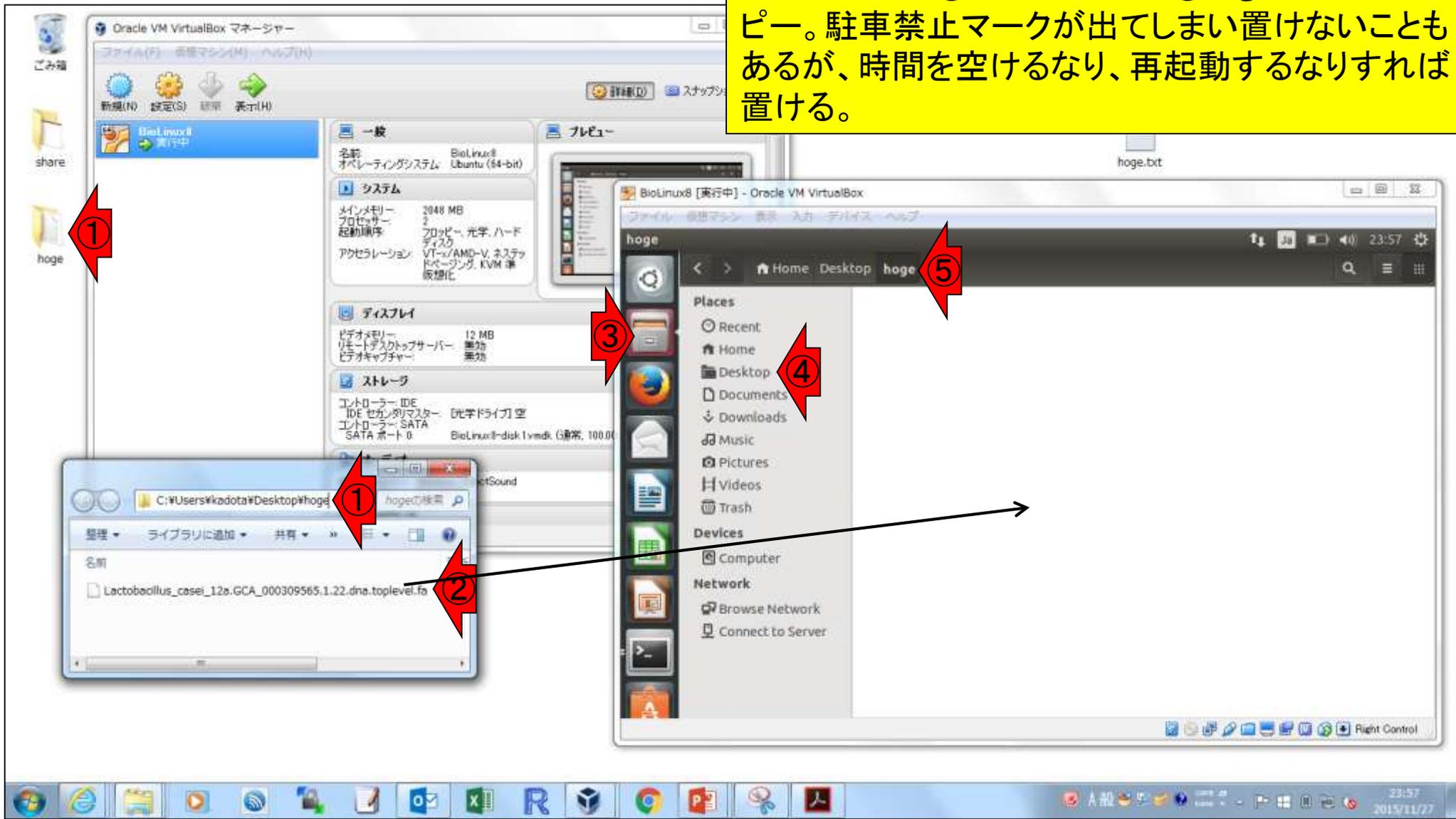
エクスプローラーでこの FTP サイトを表示するには、Alt キーを押して、[表示]をクリックし、[エクスプローラーで FTP サイトを開く]をクリックしてください。

1 階層上のディレクトリへ

04/03/2014 12:00午前	469	CHECKSUMS
03/21/2014 12:00午前	902,847	Lactobacillus casei 12a.GCA_000309565.1.22.dna.genome.fa.gz
03/21/2014 12:00午前	902,847	Lactobacillus casei 12a.GCA_000309565.1.22.dna.toplevel.fa.gz
03/21/2014 12:00午前	902,952	Lactobacillus casei 12a.GCA_000309565.1.22.dna.rm.genome.fa.gz
03/21/2014 12:00午前	902,952	Lactobacillus casei 12a.GCA_000309565.1.22.dna.rm.toplevel.fa.gz
03/21/2014 12:00午前	902,953	Lactobacillus casei 12a.GCA_000309565.1.22.dna.sm.genome.fa.gz
03/21/2014 12:00午前	902,953	Lactobacillus casei 12a.GCA_000309565.1.22.dna.sm.toplevel.fa.gz
03/21/2014 12:00午前	3,111	README

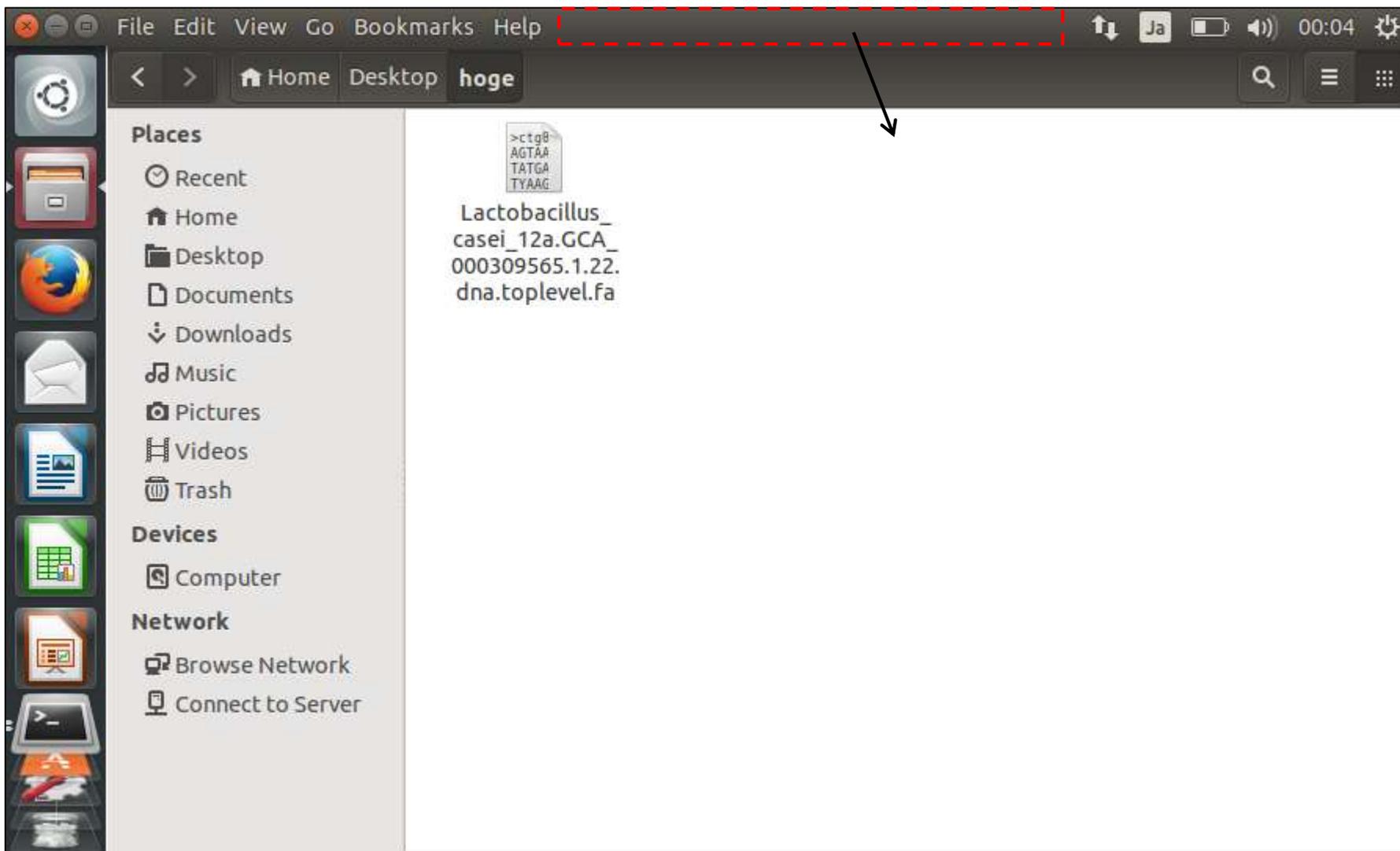
W10-3-2: D&D

やり方1: ①ホストOSのデスクトップのhogeフォルダ中にある②FASTAファイルを、ドラッグ&ドロップでゲストOSの④デスクトップの⑤hogeフォルダにコピー。駐車禁止マークが出てしまい置けないこともあるが、時間を空けるなり、再起動するなりすれば置ける。



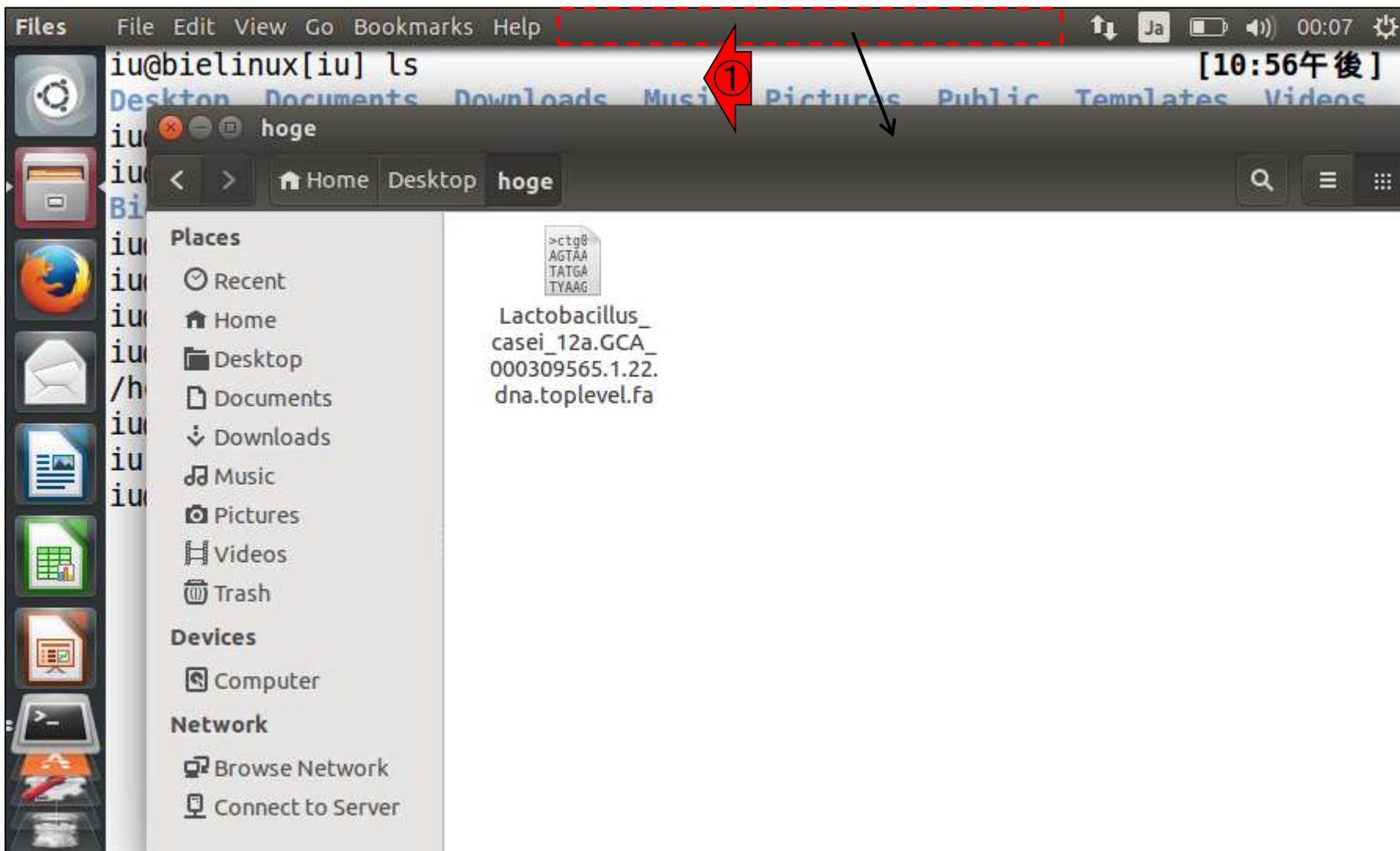
W10-3-2: D&D

Tips。赤枠付近でクリックしたまま下にずらすと、hogeフォルダの裏に隠れていた他のターミナルを見ることができる

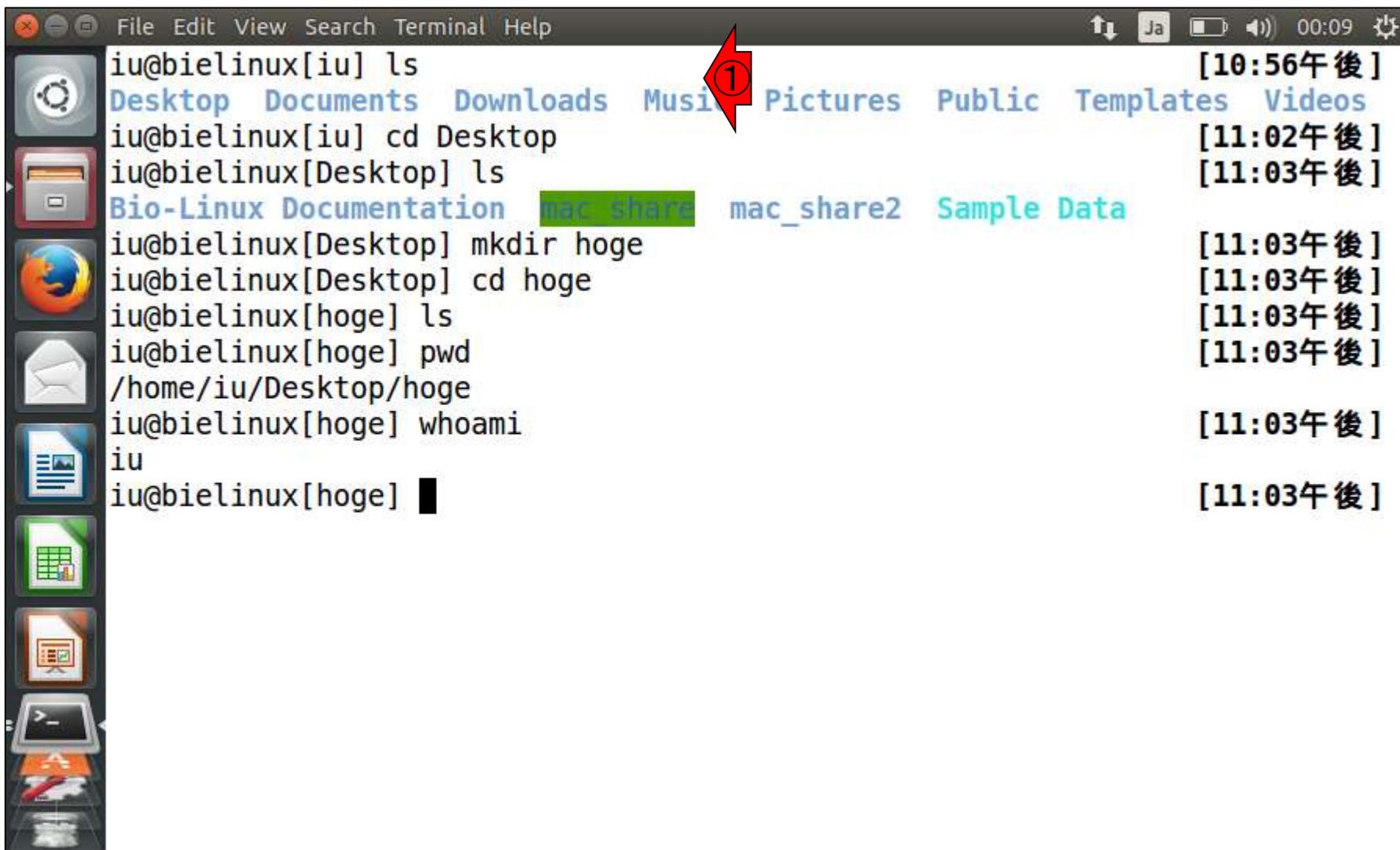


W10-3-2: D&D

Tips。赤枠付近でクリックしたまま下にずらすと、hogeフォルダの裏に隠れていた他のターミナルを見ることができる



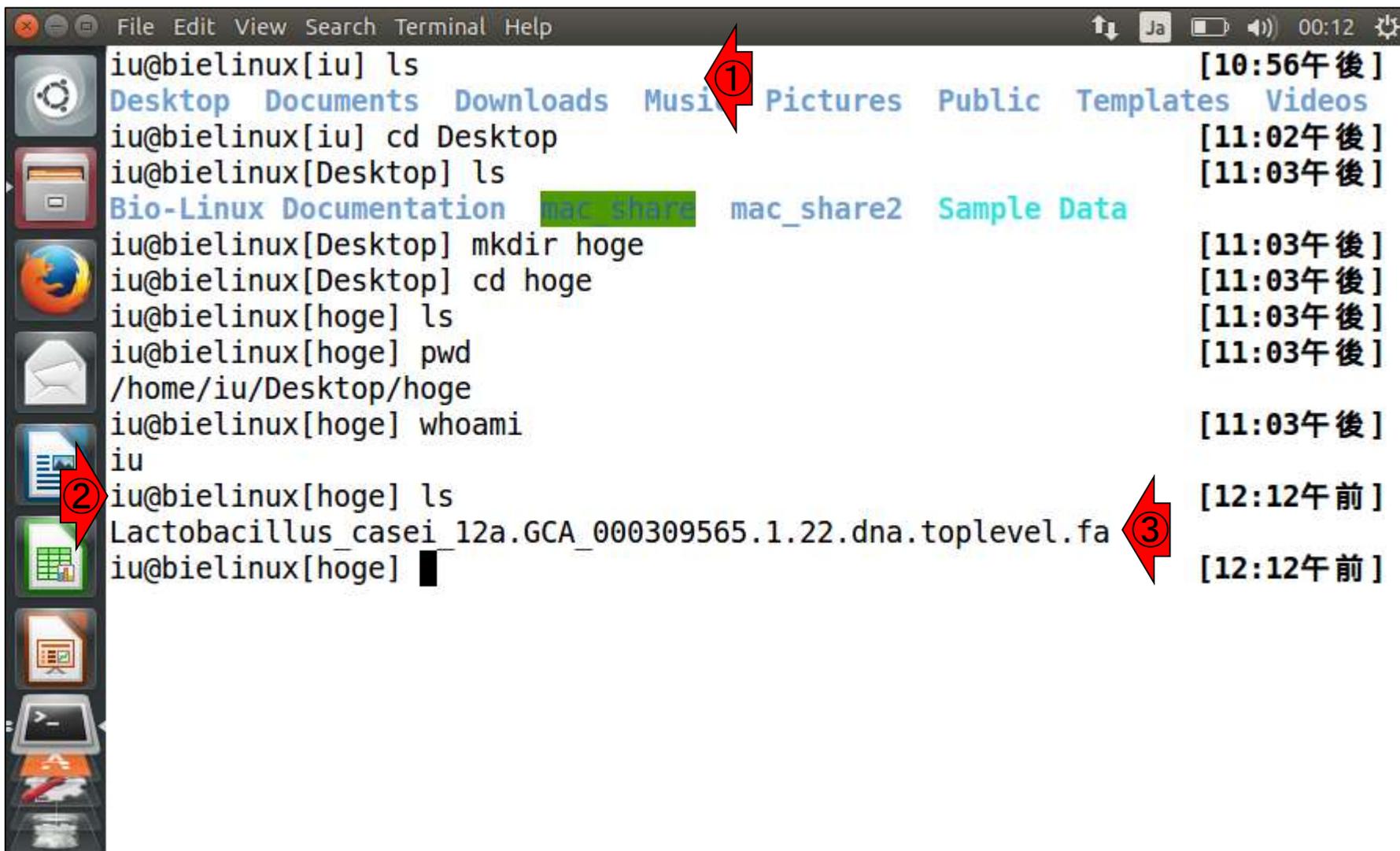
W10-3-2: D&D



```
iu@bielinux[iu] ls [10:56午後]
Desktop Documents Downloads Musi Pictures Public Templates Videos
iu@bielinux[iu] cd Desktop [11:02午後]
iu@bielinux[Desktop] ls [11:03午後]
Bio-Linux Documentation mac share mac_share2 Sample Data
iu@bielinux[Desktop] mkdir hoge [11:03午後]
iu@bielinux[Desktop] cd hoge [11:03午後]
iu@bielinux[hoge] ls [11:03午後]
iu@bielinux[hoge] pwd [11:03午後]
/home/iu/Desktop/hoge
iu@bielinux[hoge] whoami [11:03午後]
iu
iu@bielinux[hoge] [11:03午後]
```

①あたりをクリックしてターミナルをアクティブにし...
もう一度②lsすると、③FASTAファイルが見られる。

W10-3-2: D&D



```
iu@bielinux[iu] ls [10:56午後]
Desktop Documents Downloads Musi Pictures Public Templates Videos
iu@bielinux[iu] cd Desktop [11:02午後]
iu@bielinux[Desktop] ls [11:03午後]
Bio-Linux Documentation mac share mac_share2 Sample Data
iu@bielinux[Desktop] mkdir hoge [11:03午後]
iu@bielinux[Desktop] cd hoge [11:03午後]
iu@bielinux[hoge] ls [11:03午後]
iu@bielinux[hoge] pwd [11:03午後]
/home/iu/Desktop/hoge
iu@bielinux[hoge] whoami [11:03午後]
iu
iu@bielinux[hoge] ls [12:12午前]
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
iu@bielinux[hoge] [12:12午前]
```

W10-4: 一旦削除

FASTAファイルを共有フォルダ経由でもおけることを示すべく、一旦FASTAファイルを①rmコマンドを用いて削除。赤枠の「rm L」まで打ってからタブ補完を有効に利用すべし。

```
iu@bielinux[iu] ls [10:56午後]
Desktop Documents Downloads Music Pictures Public Templates Videos
iu@bielinux[iu] cd Desktop [11:02午後]
iu@bielinux[Desktop] ls [11:03午後]
Bio-Linux Documentation mac share mac_share2 Sample Data
iu@bielinux[Desktop] mkdir hoge [11:03午後]
iu@bielinux[Desktop] cd hoge [11:03午後]
iu@bielinux[hoge] ls [11:03午後]
iu@bielinux[hoge] pwd [11:03午後]
/home/iu/Desktop/hoge
iu@bielinux[hoge] whoami [11:03午後]
iu
iu@bielinux[hoge] ls [12:12午前]
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
① iu@bielinux[hoge] rm Lactobacillus_casei_12a.GCA_000309565.1.22.dna.topleve [12:12午前]
l.fa
rm: remove regular file 'Lactobacillus_casei_12a.GCA_000309565.1.22.dna.top [12:12午前]
level.fa'?
```

W10-4: 一旦削除

「rm: remove regular file '...' ?」は、本当に消してもいいか?と聞いてきているので、yesに相当する②yを打ち込んでリターンキーを押す。その後lsすると確かにFASTAファイルが消えていることがわかる。注意点として、この作業はゴミ箱への移動を意味するわけではなく、本当に消えてしまうので、重要なファイルは気をつけるべし

```
iu@bielinux[iu] ls
Desktop Documents Downloads Musi
iu@bielinux[iu] cd Desktop
iu@bielinux[Desktop] ls
Bio-Linux Documentation mac share mac_share2 Sample Data
iu@bielinux[Desktop] mkdir hoge
iu@bielinux[Desktop] cd hoge
iu@bielinux[hoge] ls
iu@bielinux[hoge] pwd
/home/iu/Desktop/hoge
iu@bielinux[hoge] whoami
iu
iu@bielinux[hoge] ls
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
iu@bielinux[hoge] rm Lactobacillus_casei_12a.GCA_000309565.1.22.dna.topleve
l.fa
rm: remove regular file 'Lactobacillus_casei_12a.GCA_000309565.1.22.dna.top
level.fa'? y
```

[11:03午後]

[11:03午後]

[11:03午後]

[11:03午後]

[11:03午後]

[12:12午前]



W10-5: 共有フォルダ経由

①「ホストOS - デスクトップ - hoge」中の②FASTAファイルを、③共有フォルダとして設定した「ホストOS - デスクトップ - share」フォルダに移動。

The image shows a Windows 10 desktop environment with Oracle VM VirtualBox Manager. A Linux VM named 'BioLinux8' is running. The host OS desktop has a folder named 'hoge' containing a FASTA file 'Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa'. A shared folder named 'share' is also visible on the host. The terminal window in the VM shows the following commands and output:

```
iu@bielinux[iu] ls [10:56午後]
Desktop Documents Downloads Music Pictures Public Templates Videos
iu@bielinux[iu] cd Desktop [11:02午後]
iu@bielinux[Desktop] ls [11:03午後]
Bio-Linux Documentation mac_share2 Sample Data
iu@bielinux[Desktop] mkdir hoge [11:03午後]
iu@bielinux[Desktop] cd hoge [11:03午後]
iu@bielinux[hoge] ls [11:03午後]
iu@bielinux[hoge] pwd [11:03午後]
/home/iu/Desktop/hoge
iu@bielinux[hoge] whoami [11:03午後]
iu
iu@bielinux[hoge] ls [12:12午前]
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
iu@bielinux[hoge] rm Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa? y
rm: remove regular file 'Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa'? y
iu@bielinux[hoge] ls [11:04午前]
iu@bielinux[hoge] █ [11:04午前]
```

Red arrows and numbers indicate the steps: ① points to the 'hoge' folder on the host desktop; ② points to the FASTA file within 'hoge'; ③ points to the 'share' folder on the host desktop. A black arrow points from the 'hoge' folder to the terminal window, indicating the file transfer process.

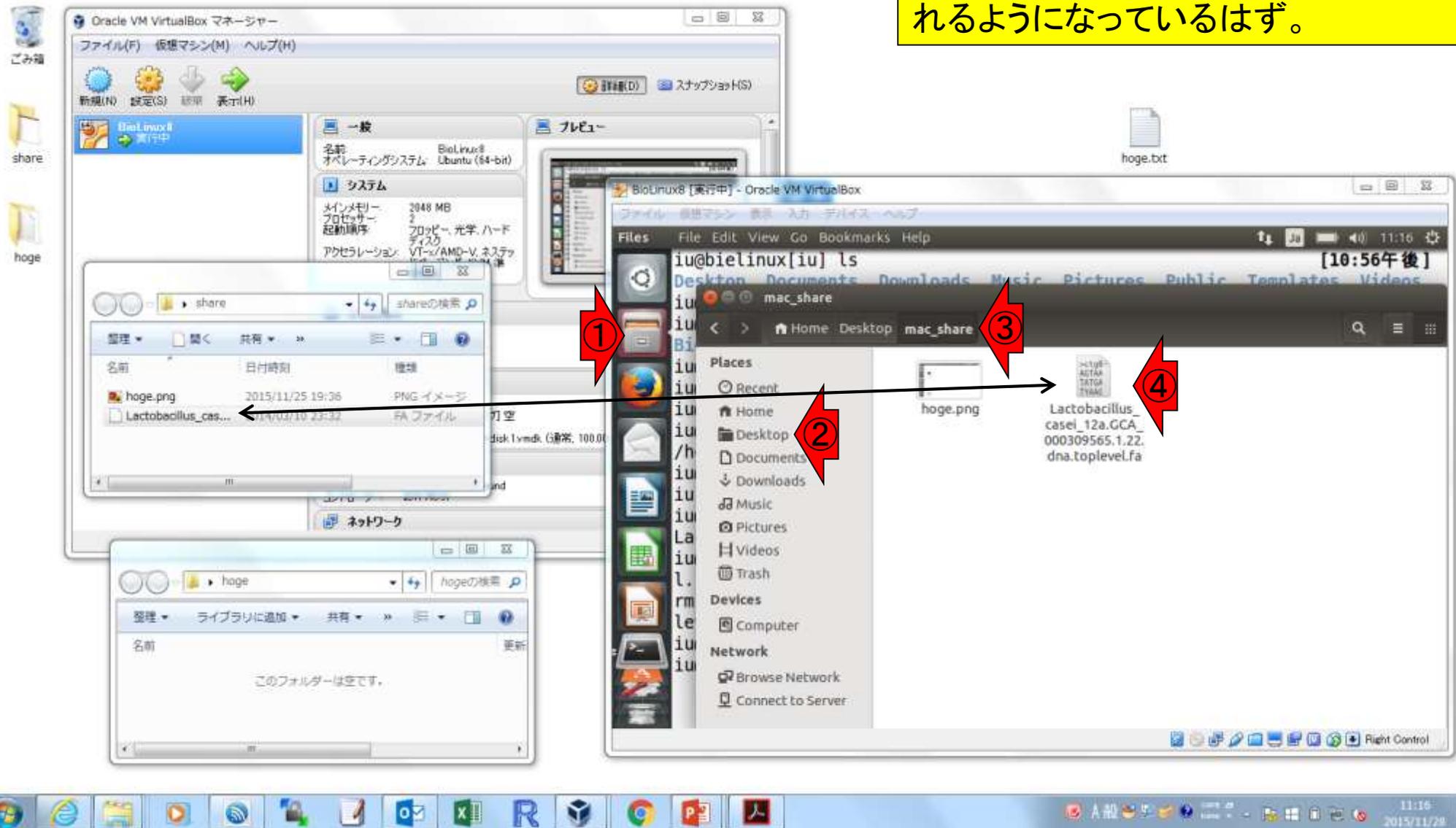
W10-5: 共有フォルダ経由

The screenshot displays the Oracle VM VirtualBox interface. In the foreground, a Windows 10 desktop is visible. A file explorer window shows the 'share' folder containing 'hoge.png' and 'Lactobacillus_cas...'. Another window shows the 'hoge' folder, which is currently empty. A terminal window titled 'BioLinux8 [実行中]' shows the following commands and outputs:

```
iu@bielinux[~/Desktop/hoge]
iu@bielinux[iu] ls [10:56午後]
Desktop Documents Downloads Music Pictures Public Templates Videos
iu@bielinux[iu] cd Desktop [11:02午後]
iu@bielinux[Desktop] ls [11:03午後]
Bio-Linear Documentation mac_share2 Sample Data
iu@bielinux[Desktop] mkdir hoge [11:03午後]
iu@bielinux[Desktop] cd hoge [11:03午後]
iu@bielinux[hoge] ls [11:03午後]
/home/iu/Desktop/hoge
iu@bielinux[hoge] pwd [11:03午後]
/home/iu/Desktop/hoge
iu@bielinux[hoge] whoami [11:03午後]
iu
iu@bielinux[hoge] ls [12:12午前]
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
iu@bielinux[hoge] rm Lactobacillus_casei_12a.GCA_000309565.1.22.dna.topleve
l.fa
rm: remove regular file 'Lactobacillus_casei_12a.GCA_000309565.1.22.dna.top
level.fa'? y
iu@bielinux[hoge] ls [11:04午前]
iu@bielinux[hoge] █ [11:04午前]
```

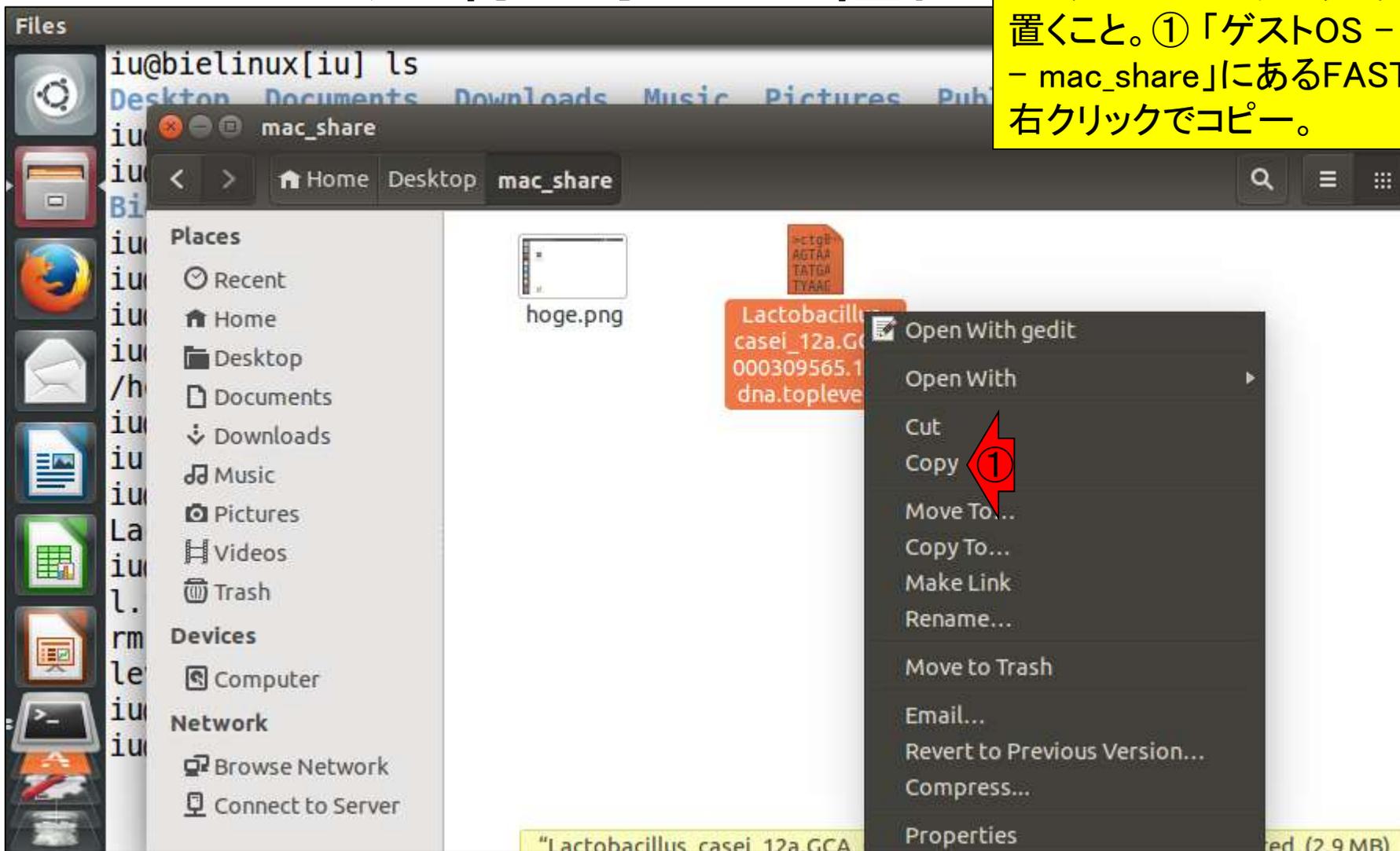
W10-5: 共有フォルダ経由

共有フォルダが正しく機能していれば、FASTAファイルが「ゲストOS - ②デスクトップ - ③mac_share」上でも見られるようになっているはず。



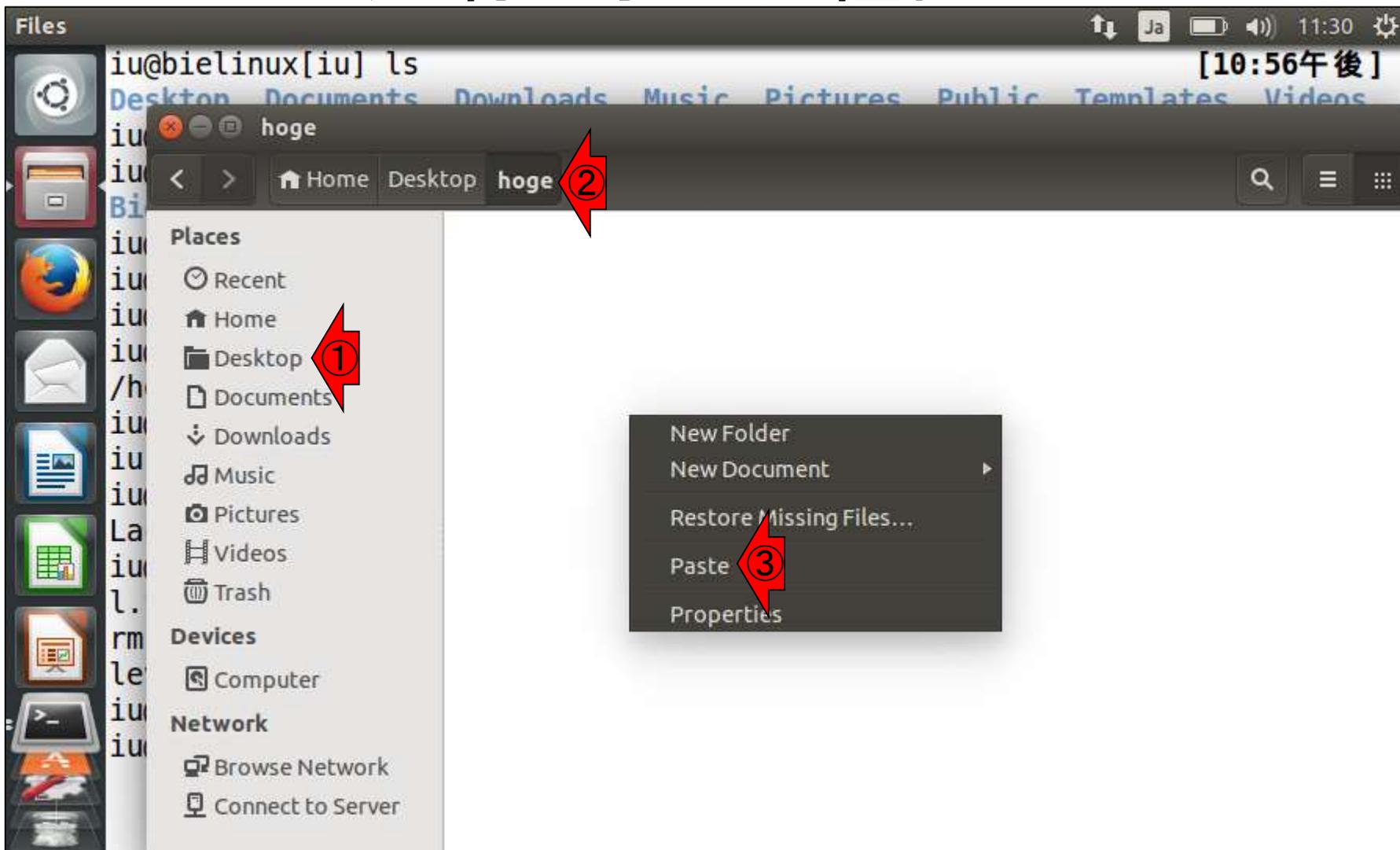
W10-5: 共有フォルダ経由

ここでの目的をおさらい。やりたかったことは、FASTAファイルをどうにかして「ゲストOS - デスクトップ - hoge」に置くこと。①「ゲストOS - デスクトップ - mac_share」にあるFASTAファイルを右クリックでコピー。

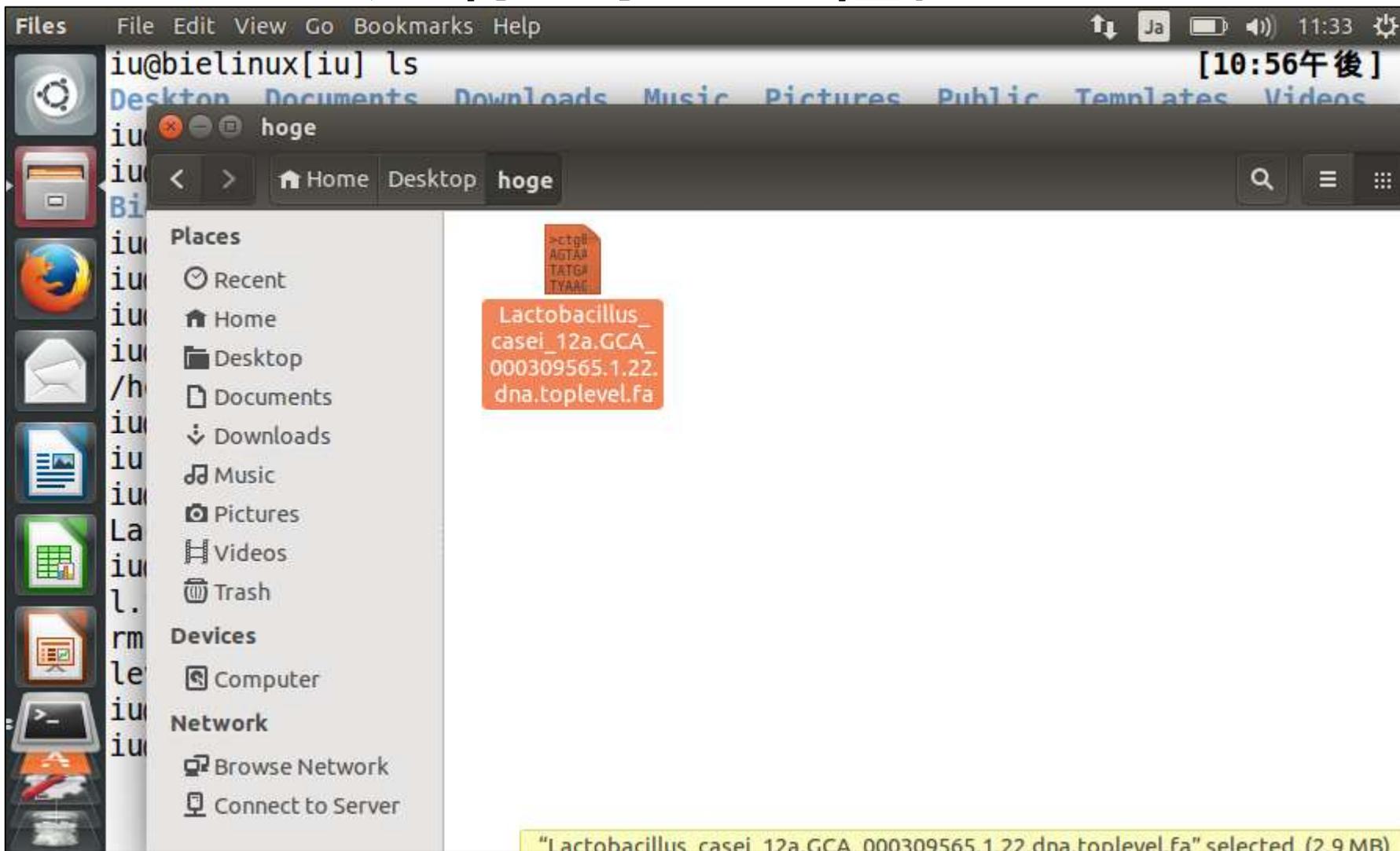


「ゲストOS - ①デスクトップ - ②hoge」フォルダ上で、右クリックで③Paste。

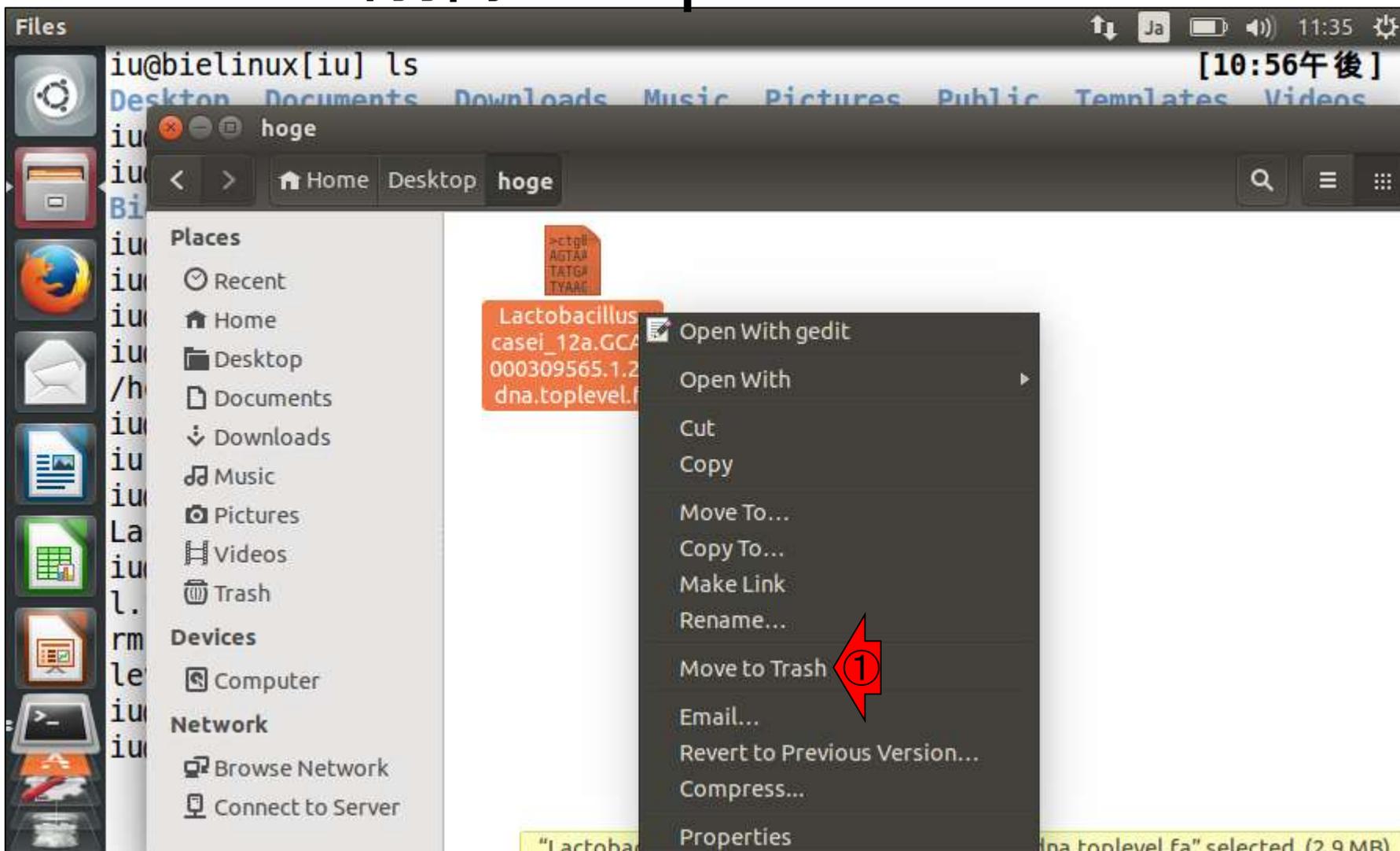
W10-5: 共有フォルダ経由



W10-5: 共有フォルダ経由

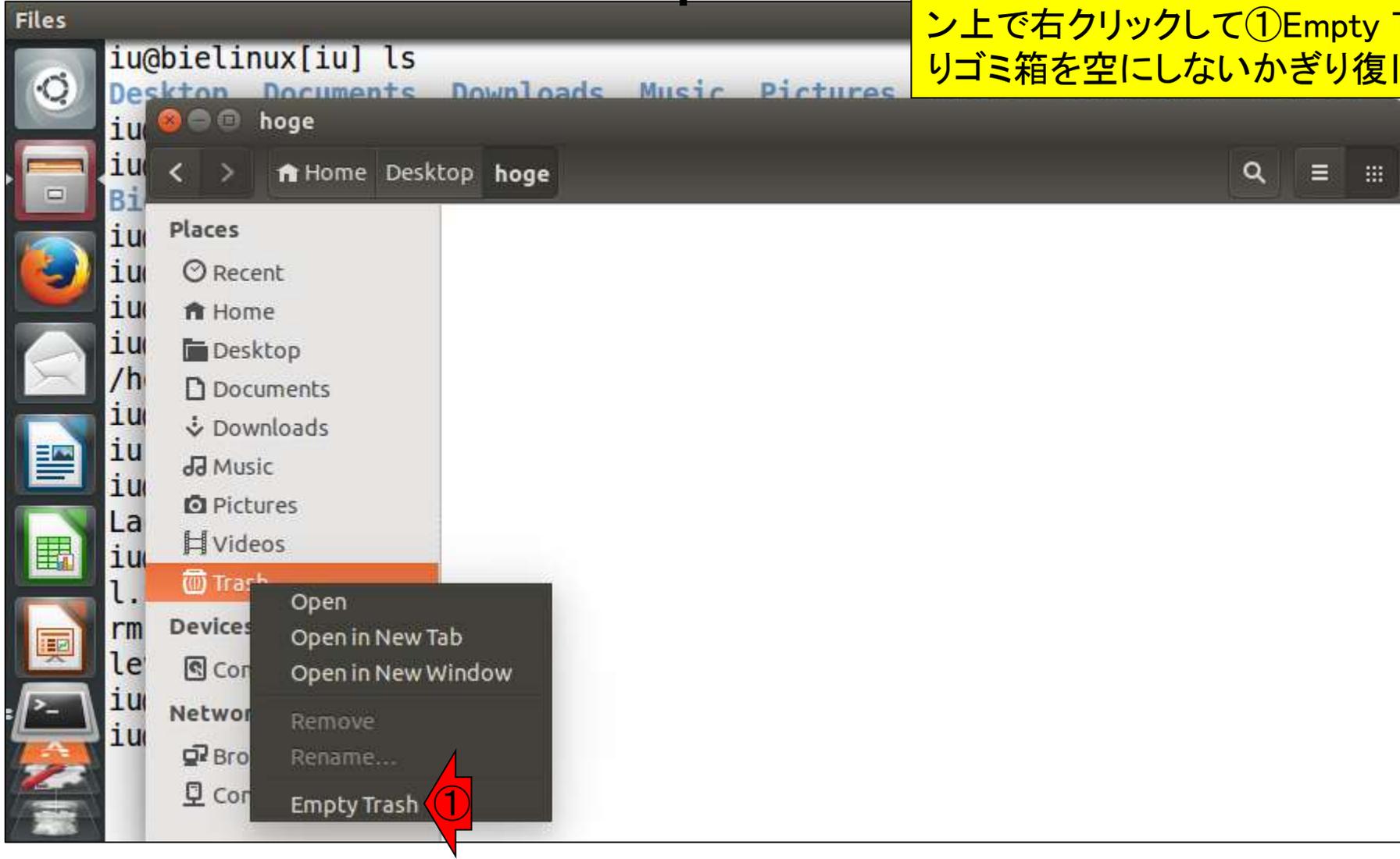


W10-5: 削除のTips



W10-5: 削除のTips

FASTAファイル移動後のごみ箱はMacと同様膨れる。rmコマンドと違って、この作業自体はゴミ箱への移動なので、ごみ箱アイコン上で右クリックして①Empty Trash、つまりゴミ箱を空にしないかぎり復旧できる。



W11-1: URL情報取得

ホストOS上での通常のダウンロード作業は、目的のgzip圧縮ファイル上で右クリックし②「対象をファイルに保存」だが、それだとURL情報を取得できない。

データ取得とチェックサム:

- [Ensembl Genomes: Flicek et al., Nucleic Acids Res., 2014](#)
 - [Lactobacillus casei 12A \(Taxonomy ID: 1051650\): Broadbent et al., BMC Genomics, 2012](#)
 - [Ensembl Bacteria - Lactobacillus casei 12A - Download DNA sequence Lactobacillus casei 12a.GCA_000309565.1.22.dna.toplevel.fa.gz](#)
 - [Toplevelの概念を説明しているところ](#)
 - [Release note](#)
 - [Ensembl Bacteria - Lactobacillus casei 12A - Download Lactobacillus casei 12a.GCA_000309565.2.24.dna...](#)
- [Ensembl Bacteria](#)
 - [Lactobacillus](#)
 - [Escherichia](#)
 - ...
- [Ensembl Fungi](#)
 - [Saccharomyces cerevisiae](#)

FTP ディレクトリ /pub/bacteria/release-22/fasta/bacteria_15_collection/lactobacillus_casei_12a/dna/ / ftp.ensemblgenomes.org

エクスペローラーでこの FTP サイトを表示するには、Alt キーを押して、[表示]をクリックし、[エクスペローラーで FTP サイトを開く]をクリックしてください。

1 階層上のディレクトリへ

04/03/2014 12:00午前	469	CHECKSUMS
03/21/2014 12:00午前	902,847	Lactobacillus casei 12a.GCA_000309565.1.22.dna.genome.fa.gz
03/21/2014 12:00午前	902,847	Lactobacillus casei 12a.GCA_000309565.1.22.dna.toplevel.fa.gz
03/21/2014 12:00午前	902,952	Lactobacillus casei 12a.GCA_000309565.1.22.dna.rm.genome.fa.gz
03/21/2014 12:00午前	902,952	Lactobacillus casei 12a.GCA_000309565.1.22.dna.rm.toplevel.fa.gz
03/21/2014 12:00午前	902,953	Lactobacillus casei 12a.GCA_000309565.1.22.dna.sm.genome.fa.gz
03/21/2014 12:00午前	902,953	Lactobacillus casei 12a.GCA_000309565.1.22.dna.sm.toplevel.fa.gz
03/21/2014 12:00午前	3,111	README

- 開く(O)
- 新しいタブで開く(W)
- 新しいウィンドウで開く(N)
- 対象をファイルに保存(A)...
- 対象を印刷(P)
- 切り取り
- コピー(C)
- ショートカットのコピー(T)
- 貼り付け(P)
- Bing で翻訳
- 電子メール (Windows Live Hotmail)
- すべてのアクセラレータ
- 要素の検査(L)
- お気に入り追加(F)...
- Send to OneNote
- プロパティ(R)

W11-1: URL情報取得

ブラウザがInternet Explorerの場合は、右クリックの①「ショートカットのコピー」でURL情報を取得できる。

The screenshot shows an Internet Explorer browser window displaying an FTP directory listing from `ftp://ftp.ensemblgenomes.org/pub/bacteria/release-22/fasta/bacteria_15_collection/lactobacillus_casei_12a/dna/`. The listing includes files like `CHECKSUMS`, `Lactobacillus casei 12a.GCA 000309585.1.22.dna.genome.fasta.gz`, and `README`. A context menu is open over the listing, with the option `ショートカットのコピー(T)` (Shortcut Copy) highlighted by a red arrow with the number 1. The menu also includes options like `開く(O)`, `新しいタブで開く(W)`, `新しいウィンドウで開く(N)`, `対象をファイルに保存(A)...`, `対象を印刷(P)`, `切り取り`, `コピー(C)`, `貼り付け(P)`, `Bingで翻訳`, `電子メール (Windows Live Hotmail)`, `すべてのアクセラレータ`, `要素の検査(L)`, `お気に入りに追加(F)...`, `Send to OneNote`, and `プロパティ(R)`.

W11-1: URL情報取得

ブラウザが(ゲストOSデフォルトの)Firefoxの場合は、右クリックの「Copy Link Location」でURL情報を取得できる。ここでは、その後の説明も兼ねて1つ上の階層の①「Download DNA sequence」のページから行う。

(Rで)塩基配列解析

www.iu.a.u-tokyo.ac.jp/~kadota/r_seq.t

◦ [rm](#) (ファイルやディレクトリの削除)

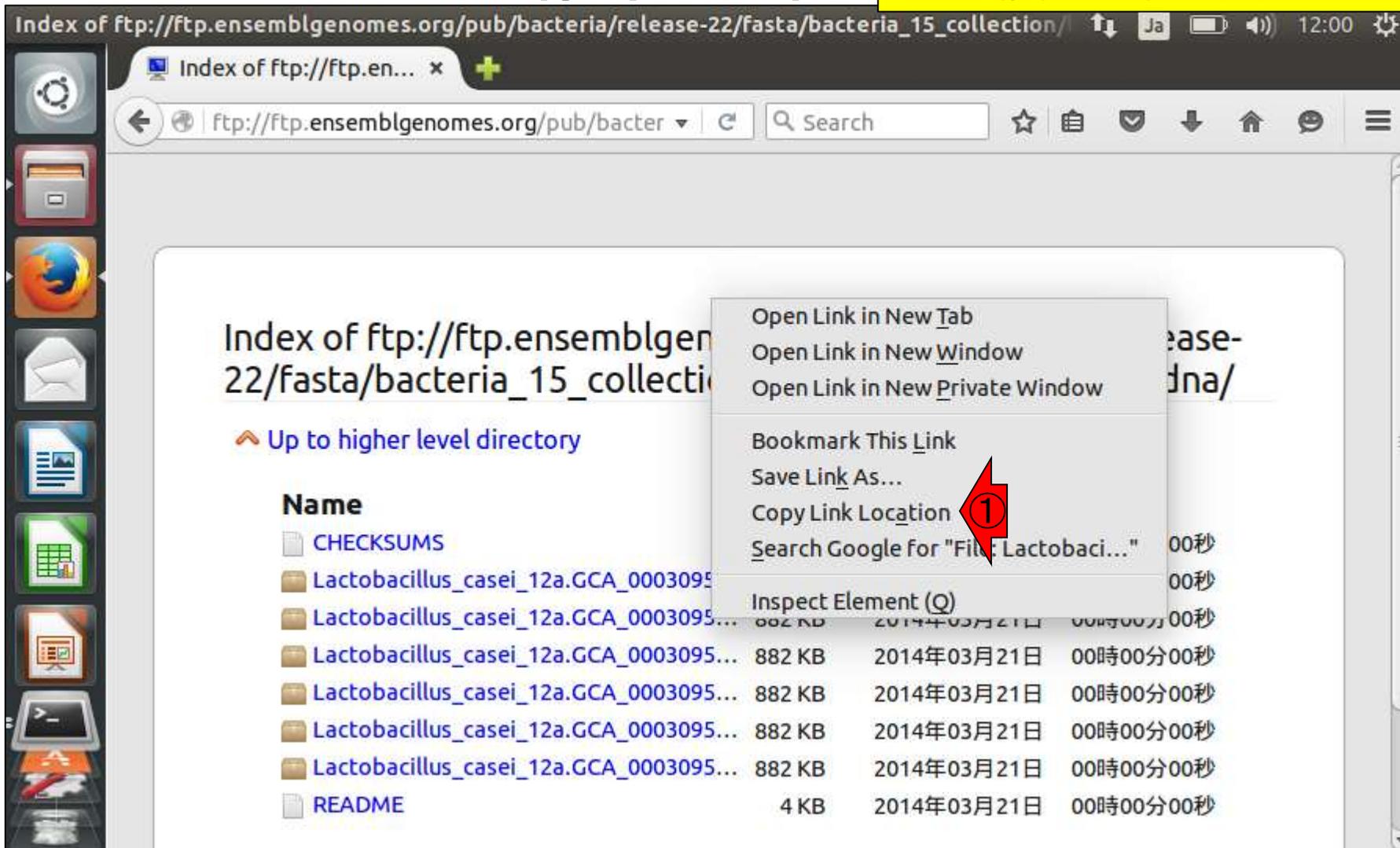
データ取得とチェックサム:

- [Ensembl Genomes: Flicek et al., Nucleic Acids Res., 2014](#)
 - [Lactobacillus casei 12A \(Taxonomy ID: 1051650\): Broadbent et al., BMC Genomics, 2012](#)
 - [Ensembl Bacteria - Lactobacillus casei 12A - Download DNA sequence Lactobacillus casei 12a.GCA_000309565.1.22.dna.toplevel.fa.gz](#) ①
 - [Toplevelの概念を説明しているところ](#)
 - [Release note](#)
 - [Ensembl Bacteria - Lactobacillus casei 12A - Download DNA sequence - Lactobacillus casei 12a.GCA_000309565.2.24.dna.toplevel.fa.gz \(release 24\)](#)
- [Ensembl Bacteria](#)
 - [Lactobacillus](#)
 - [Escherichia](#)
 - ...
- [Ensembl Fungi](#)
 - [Saccharomyces cerevisiae](#)
 - [Schizosaccharomyces pombe](#)
 - [Aspergillus nidulans](#)

[トップページへ](#)

W11-1: URL情報取得

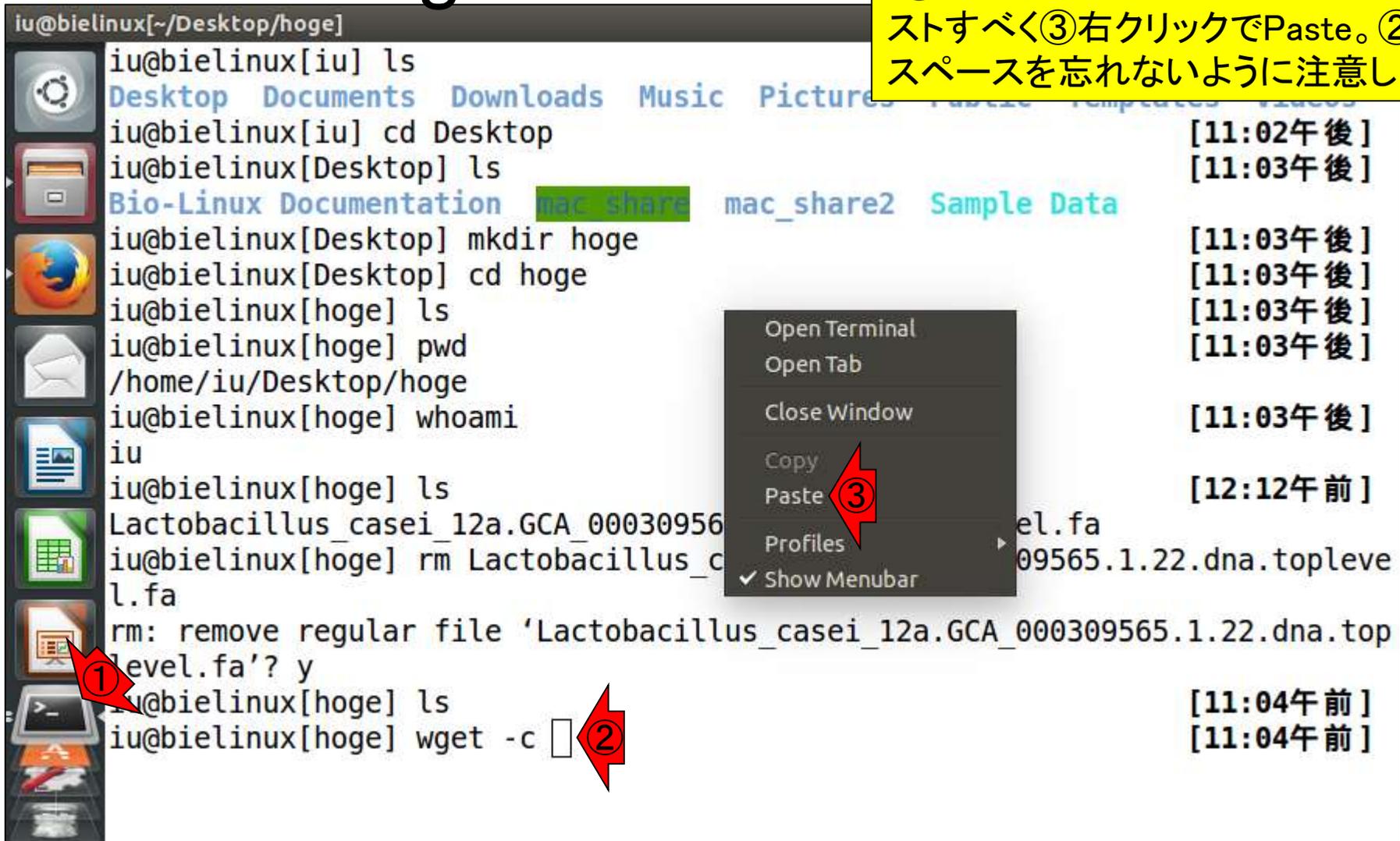
ブラウザが(ゲストOSデフォルトの)Firefoxの場合は、右クリックの①「Copy Link Location」でURL情報を取得できる。



W11-1: wget

①ターミナル画面にして、「FASTAファイル」をダウンロードしたい「Desktop - hoge」上で、②「`wget -c`」と打ち込んでから、URL情報をペーストすべく③右クリックでPaste。②のところでスペースを忘れないように注意しよう。

```
iu@bielinux[~/Desktop/hoge]
iu@bielinux[iu] ls
Desktop Documents Downloads Music Pictures Public Templates Videos
iu@bielinux[iu] cd Desktop
iu@bielinux[Desktop] ls
Bio-Linux Documentation mac_share mac_share2 Sample Data
iu@bielinux[Desktop] mkdir hoge
iu@bielinux[Desktop] cd hoge
iu@bielinux[hoge] ls
iu@bielinux[hoge] pwd
/home/iu/Desktop/hoge
iu@bielinux[hoge] whoami
iu
iu@bielinux[hoge] ls
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
iu@bielinux[hoge] rm Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
rm: remove regular file 'Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa'? y
iu@bielinux[hoge] ls
iu@bielinux[hoge] wget -c [ ]
```



W11-1: wget

ペースト直後の状態。①リターンキーを押すとダウンロードが始まる

```
iu@bielinux[iu] ls [12:14午後]
Desktop Documents Downloads Music Pictures Public Templates Videos
iu@bielinux[iu] cd Desktop [12:14午後]
iu@bielinux[Desktop] ls [12:14午後]
Bio-Linux Documentation mac_share mac_share2 Sample Data
iu@bielinux[Desktop] mkdir hoge [12:14午後]
iu@bielinux[Desktop] cd hoge [12:14午後]
iu@bielinux[hoge] ls [12:14午後]
iu@bielinux[hoge] pwd [12:14午後]
/home/iu/Desktop/hoge
iu@bielinux[hoge] whoami [12:14午後]
iu
iu@bielinux[hoge] ls [12:14午後]
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
iu@bielinux[hoge] rm Lactobacillus_casei_12a.GCA_000309565.1.22.dna.topleve
l.fa
rm: remove regular file 'Lactobacillus_casei_12a.GCA_000309565.1.22.dna.top
level.fa'? y
iu@bielinux[hoge] ls [12:15午後]
iu@bielinux[hoge] wget -c ftp://ftp.ensemblgenomes.org/pub/bacteria/release
-22/fastq/bacteria_15_collection/lactobacillus_casei_12a/dna/Lactobacillus_
casei_12a.GCA_000309565.1.22.dna.toplevel.fa.gz
```



W11-1 : wget

```
iu@bielinux[hoge] wget -c ftp://ftp.ensemblgenomes.org/pub/bacteria/release-22/fasta/bacteria_15_collection/lactobacillus_casei_12a/dna/Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa.gz
--2015-11-28 12:22:17-- ftp://ftp.ensemblgenomes.org/pub/bacteria/release-22/fasta/bacteria_15_collection/lactobacillus_casei_12a/dna/Lactobacillus_asei_12a.GCA_000309565.1.22.dna.toplevel.fa.gz
=> 'Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa.gz'
Resolving ftp.ensemblgenomes.org (ftp.ensemblgenomes.org)... 193.62.197.94
Connecting to ftp.ensemblgenomes.org (ftp.ensemblgenomes.org)|193.62.197.94|:21... connected.
Logging in as anonymous ... Logged in!
==> SYST ... done.      ==> PWD ... done.
==> TYPE I ... done.   ==> CWD (1) /pub/bacteria/release-22/fasta/bacteria_15_collection/lactobacillus_casei_12a/dna ... done.
==> SIZE Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa.gz ... 902847
==> PASV ... done.    ==> RETR Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa.gz ... done.
Length: 902847 (882K) (unauthoritative)

14% [====>                ] 131,236      10.4KB/s  eta 72s
```

W11-1: wget

無事ダウンロードが終わると通常のコマンド入力待ち状態となる。ここでは、①lsして確認したところまでを示す

```
z'  
Resolving ftp.ensemblgenomes.org (ftp.ensemblgenomes.org)... 193.62.197.94  
Connecting to ftp.ensemblgenomes.org (ftp.ensemblgenomes.org)|193.62.197.94  
|:21... connected.  
Logging in as anonymous ... Logged in!  
==> SYST ... done.      ==> PWD ... done.  
==> TYPE I ... done.    ==> CWD (1) /pub/bacteria/release-22/fasta/bacteria_1  
5_collection/lactobacillus_casei_12a/dna ... done.  
==> SIZE Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa.gz ...  
902847  
==> PASV ... done.      ==> RETR Lactobacillus_casei_12a.GCA_000309565.1.22.d  
na.toplevel.fa.gz ... done.  
Length: 902847 (882K) (unauthoritative)  
  
100%[=====>] 902,847      13.4KB/s   in 65s  
  
2015-11-28 12:23:26 (13.6 KB/s) - 'Lactobacillus_casei_12a.GCA_000309565.1.  
22.dna.toplevel.fa.gz' saved [902847]  
  
iu@bielinux[hoge] ls [12:23午後]  
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa.gz  
iu@bielinux[hoge] [11:27午前]
```



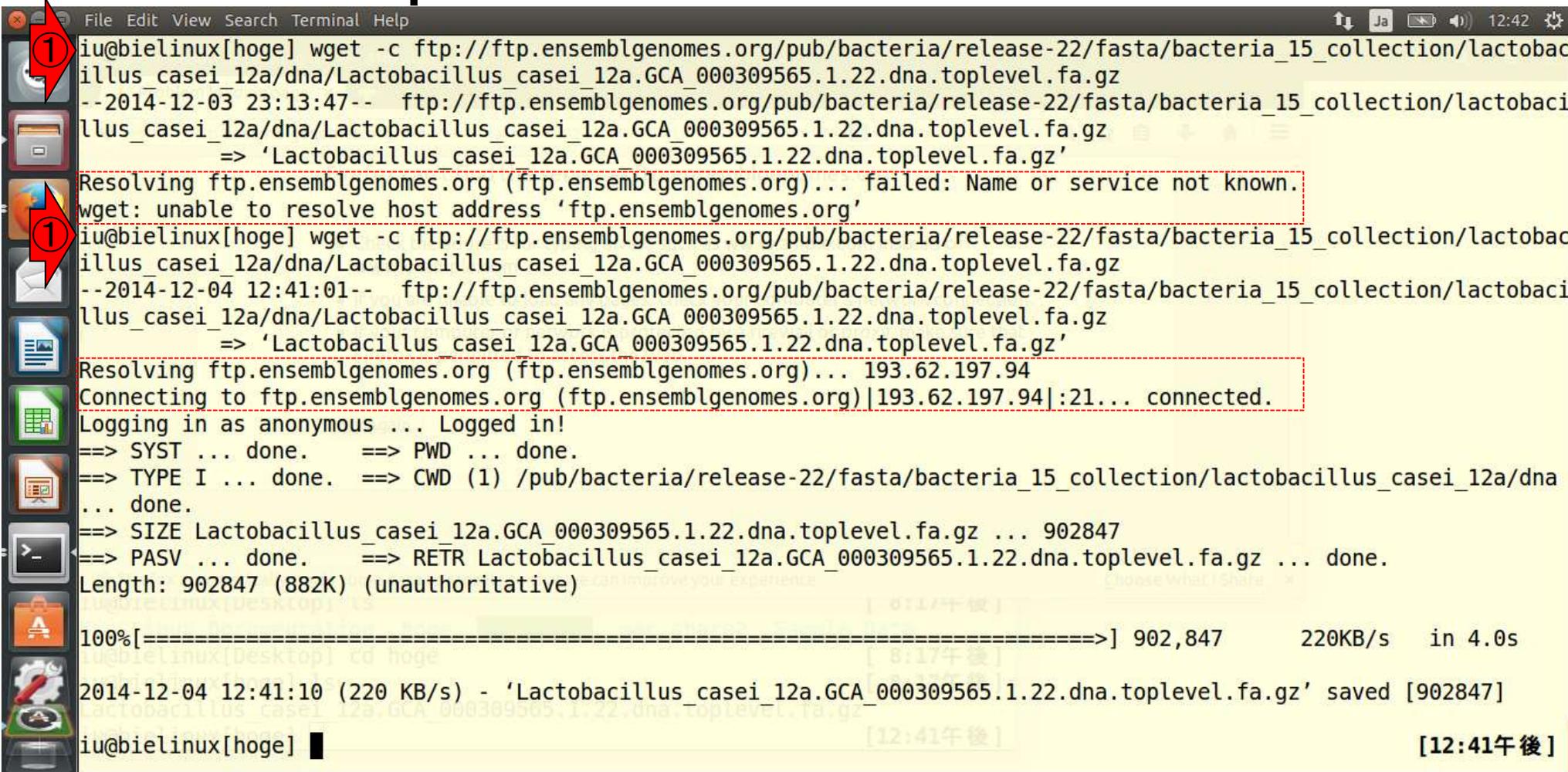
W11-1: Tips

ときどきダウンロードに失敗した(failed)というメッセージに遭遇することがある。ホストOSのウェブ経由だとうまくいくが、このような場合はゲストOSのウェブもなぜか利用不可能になっている。理由は不明であるが、数十分、あるいは数時間後にリトライするとうまくいく場合が多い。ここでは①と②で2回トライして失敗している

```
iu@bielinux[~/Desktop/hoge]
① iu@bielinux[hoge] wget -c ftp://ftp.ensemblgenomes.org/pub/bacteria/release-22/fasta/bacteria_15_collection/lactobacillus_casei_12a/dna/Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa.gz
--2014-12-03 22:53:52-- ftp://ftp.ensemblgenomes.org/pub/bacteria/release-22/fasta/bacteria_15_collection/lactobacillus_casei_12a/dna/Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa.gz
=> 'Lactobacillus casei 12a.GCA_000309565.1.22.dna.toplevel.fa.gz'
Resolving ftp.ensemblgenomes.org (ftp.ensemblgenomes.org)... failed: Name or service not known.
wget: unable to resolve host address 'ftp.ensemblgenomes.org'
① iu@bielinux[hoge] wget ftp://ftp.ensemblgenomes.org/pub/bacteria/release-22/fasta/bacteria_15_collection/lactobacillus_casei_12a/dna/Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa.gz
--2014-12-03 22:54:16-- ftp://ftp.ensemblgenomes.org/pub/bacteria/release-22/fasta/bacteria_15_collection/lactobacillus_casei_12a/dna/Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa.gz
=> 'Lactobacillus casei 12a.GCA_000309565.1.22.dna.toplevel.fa.gz'
Resolving ftp.ensemblgenomes.org (ftp.ensemblgenomes.org)... failed: Name or service not known.
wget: unable to resolve host address 'ftp.ensemblgenomes.org'
iu@bielinux[hoge] █
```

[10:54午後] Choose What I Share [10:54午後]

W11-1: Tips



```
iu@bielinux[hoge] wget -c ftp://ftp.ensemblgenomes.org/pub/bacteria/release-22/fasta/bacteria_15_collection/lactobacillus_casei_12a/dna/Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa.gz
--2014-12-03 23:13:47-- ftp://ftp.ensemblgenomes.org/pub/bacteria/release-22/fasta/bacteria_15_collection/lactobacillus_casei_12a/dna/Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa.gz
=> 'Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa.gz'
Resolving ftp.ensemblgenomes.org (ftp.ensemblgenomes.org)... failed: Name or service not known.
wget: unable to resolve host address 'ftp.ensemblgenomes.org'

iu@bielinux[hoge] wget -c ftp://ftp.ensemblgenomes.org/pub/bacteria/release-22/fasta/bacteria_15_collection/lactobacillus_casei_12a/dna/Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa.gz
--2014-12-04 12:41:01-- ftp://ftp.ensemblgenomes.org/pub/bacteria/release-22/fasta/bacteria_15_collection/lactobacillus_casei_12a/dna/Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa.gz
=> 'Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa.gz'
Resolving ftp.ensemblgenomes.org (ftp.ensemblgenomes.org)... 193.62.197.94
Connecting to ftp.ensemblgenomes.org (ftp.ensemblgenomes.org)|193.62.197.94|:21... connected.
Logging in as anonymous ... Logged in!
==> SYST ... done.      ==> PWD ... done.
==> TYPE I ... done.   ==> CWD (1) /pub/bacteria/release-22/fasta/bacteria_15_collection/lactobacillus_casei_12a/dna
... done.
==> SIZE Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa.gz ... 902847
==> PASV ... done.     ==> RETR Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa.gz ... done.
Length: 902847 (882K) (unauthoritative)

100%[=====] 902,847      220KB/s   in 4.0s

2014-12-04 12:41:10 (220 KB/s) - 'Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa.gz' saved [902847]

iu@bielinux[hoge] █
```

W11-2: ファイルサイズ

ファイルサイズは、①ホストOS上で眺めると902,847、②ゲストOSだと882 KBと微妙に異なっているように見えるが、実は表示のさせ方の違いだけのようである。

FTP ディレクトリ /pub/bacteria/release-22/fasta/bacteria_15_collection/lactobacillus_casei_12a/dna/ /
ftp.ensemblgenomes.org

エクスプローラーでこの FTP サイトを表示するには、Alt キーを押して、[表示]をクリックし、[エクスプローラーで FTP サイトを開く]をクリックしてください。

1 階層上のディレクトリへ

04/03/2014 12:00午前	469	CHECKSUMS
03/21/2014 12:00午前	902,847	Lactobacillus casei_12a.GCA_000309565.1.22.dna.genome.fa.gz
03/21/2014 12:00午前	902,847	Lactobacillus casei_12a.GCA_000309565.1.22.dna.toplevel.fa.gz
03/21/2014 12:00午前	902,952	Lactobacillus casei_12a.GCA_000309565.1.22.dna.rm.genome.fa.gz
03/21/2014 12:00午前	902,952	Lactobacillus casei_12a.GCA_000309565.1.22.dna.rm.genome.fa.gz
03/21/2014 12:00午前	902,953	Lactobacillus casei_12a.GCA_000309565.1.22.dna.rm.genome.fa.gz
03/21/2014 12:00午前	902,953	Lactobacillus casei_12a.GCA_000309565.1.22.dna.rm.genome.fa.gz
03/21/2014 12:00午前	3,111	README

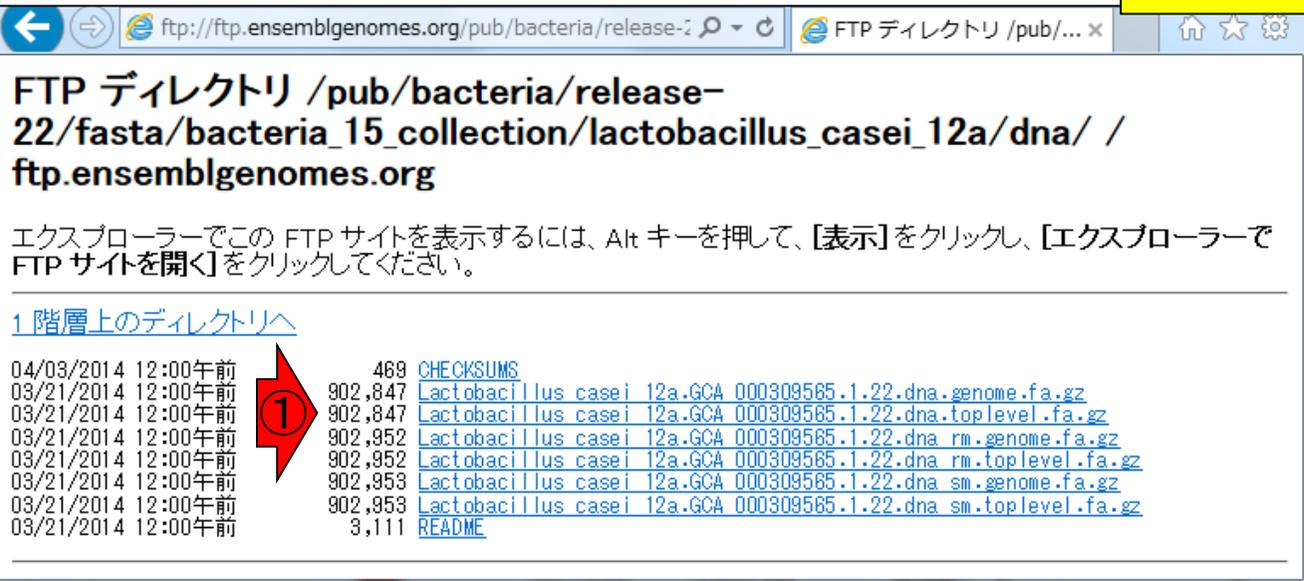


Name	Size	Last Modified ↓
Lactobacillus_casei_12a.GCA_0003095...	882 KB	2014年03月21日 00時00分00秒
Lactobacillus_casei_12a.GCA_0003095...	882 KB	2014年03月21日 00時00分00秒
Lactobacillus_casei_12a.GCA_0003095...	882 KB	2014年03月21日 00時00分00秒
Lactobacillus_casei_12a.GCA_0003095...	882 KB	2014年03月21日 00時00分00秒
Lactobacillus_casei_12a.GCA_0003095...	882 KB	2014年03月21日 00時00分00秒
Lactobacillus_casei_12a.GCA_0003095...	882 KB	2014年03月21日 00時00分00秒
README	4 KB	2014年03月21日 00時00分00秒
CHECKSUMS	1 KB	2014年04月03日 00時00分00秒



W11-2: ファイルサイズ

実際にダウンロードしたファイルを②「ls -la」で眺めると902847となり、①ホストOS上のIEで表示されている数値(902,847)と同じ。

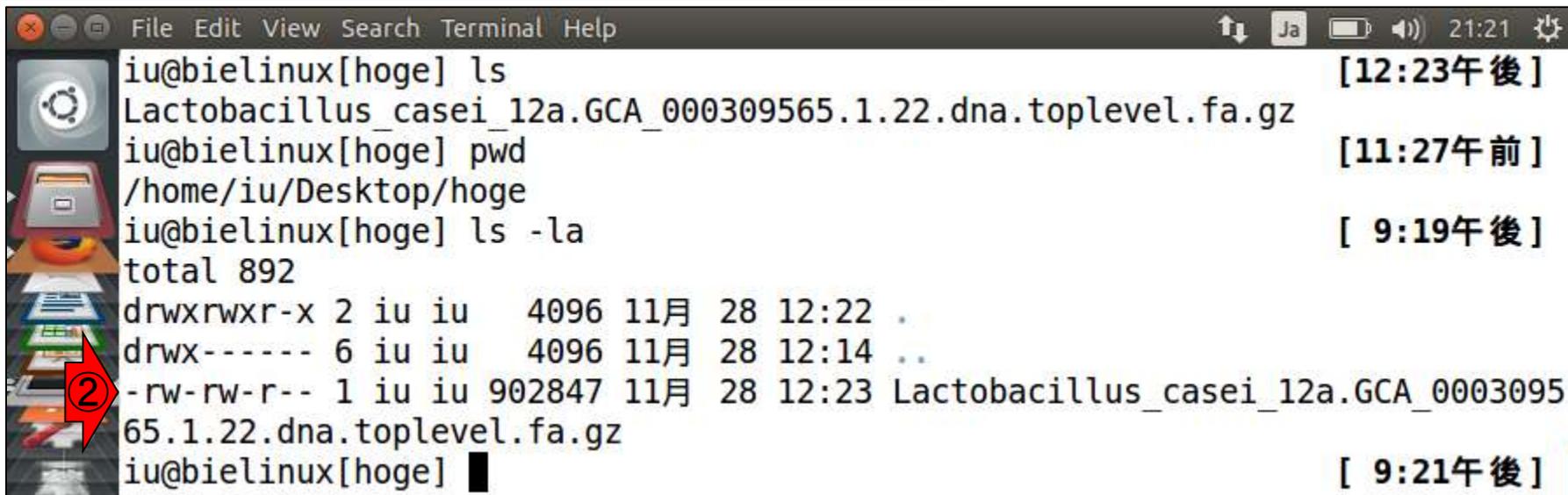


FTP ディレクトリ /pub/bacteria/release-22/fasta/bacteria_15_collection/lactobacillus_casei_12a/dna/ / ftp.ensemblgenomes.org

エクスプローラーでこの FTP サイトを表示するには、Alt キーを押して、[表示]をクリックし、[エクスプローラーで FTP サイトを開く]をクリックしてください。

1 階層上のディレクトリへ

04/03/2014 12:00午前	469	CHECKSUMS
03/21/2014 12:00午前	902,847	Lactobacillus casei 12a.GCA_000309565.1.22.dna.genome.fa.gz
03/21/2014 12:00午前	902,847	Lactobacillus casei 12a.GCA_000309565.1.22.dna.toplevel.fa.gz
03/21/2014 12:00午前	902,952	Lactobacillus casei 12a.GCA_000309565.1.22.dna.rm.genome.fa.gz
03/21/2014 12:00午前	902,952	Lactobacillus casei 12a.GCA_000309565.1.22.dna.rm.toplevel.fa.gz
03/21/2014 12:00午前	902,953	Lactobacillus casei 12a.GCA_000309565.1.22.dna.sm.genome.fa.gz
03/21/2014 12:00午前	902,953	Lactobacillus casei 12a.GCA_000309565.1.22.dna.sm.toplevel.fa.gz
03/21/2014 12:00午前	3,111	README



```
iu@bielinux[hoge] ls [12:23午後]
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa.gz
iu@bielinux[hoge] pwd [11:27午前]
/home/iu/Desktop/hoge
iu@bielinux[hoge] ls -la [ 9:19午後]
total 892
drwxrwxr-x 2 iu iu 4096 11月 28 12:22 .
drwx----- 6 iu iu 4096 11月 28 12:14 ..
-rw-rw-r-- 1 iu iu 902847 11月 28 12:23 Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa.gz
iu@bielinux[hoge] [ 9:21午後]
```

W11-2: ファイルサイズ

実際にダウンロードしたファイルを②「ls -lh」で眺めると882Kとなり、①ゲストOS上のFirefoxで表示されている数値(882K)と同じ

Index of ftp://ftp.ensemblgenomes.org/pub/bacteria/release-22/fasta/bacteria_15_collection/lactobacillus_casei_12a/dna/

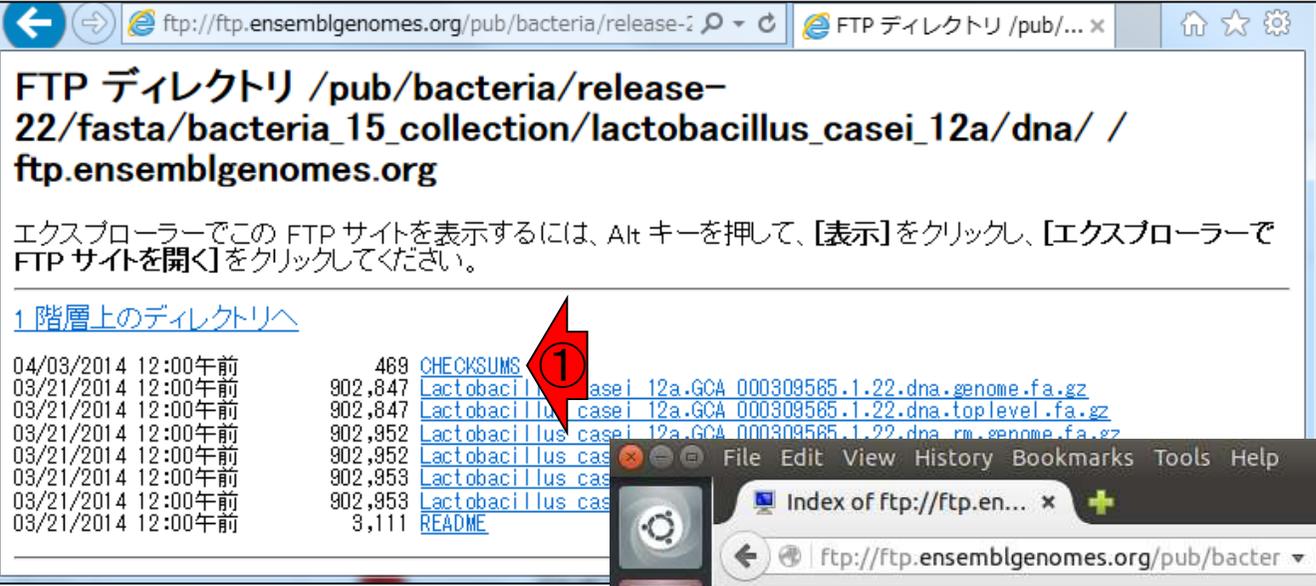
[Up to higher level directory](#)

Name	Size	Last Modified ↓
Lactobacillus_casei_12a.GCA_0003095...	882 KB	2014年03月21日 00時00分00秒
Lactobacillus_casei_12a.GCA_0003095...	882 KB	2014年03月21日 00時00分00秒
Lactobacillus_casei_12a.GCA_0003095...	882 KB	2014年03月21日 00時00分00秒
Lactobacillus_casei_12a.GCA_0003095...	882 KB	2014年03月21日 00時00分00秒

```
iu@bielinux[hoge] ls -la [ 9:19午後 ]
total 892
drwxrwxr-x 2 iu iu 4096 11月 28 12:22 .
drwx----- 6 iu iu 4096 11月 28 12:14 ..
-rw-rw-r-- 1 iu iu 902847 11月 28 12:23 Lactobacillus_casei_12a.GCA_0003095
65.1.22.dna.toplevel.fa.gz
iu@bielinux[hoge] ls -lh [ 9:21午後 ]
total 884K
-rw-rw-r-- 1 iu iu 882K 11月 28 12:23 Lactobacillus_casei_12a.GCA_000309565
.1.22.dna.toplevel.fa.gz
iu@bielinux[hoge] [ 9:24午後 ]
```

Checksum。データ提供側で算出した値は、①や②で見られる

W12: チェックサム

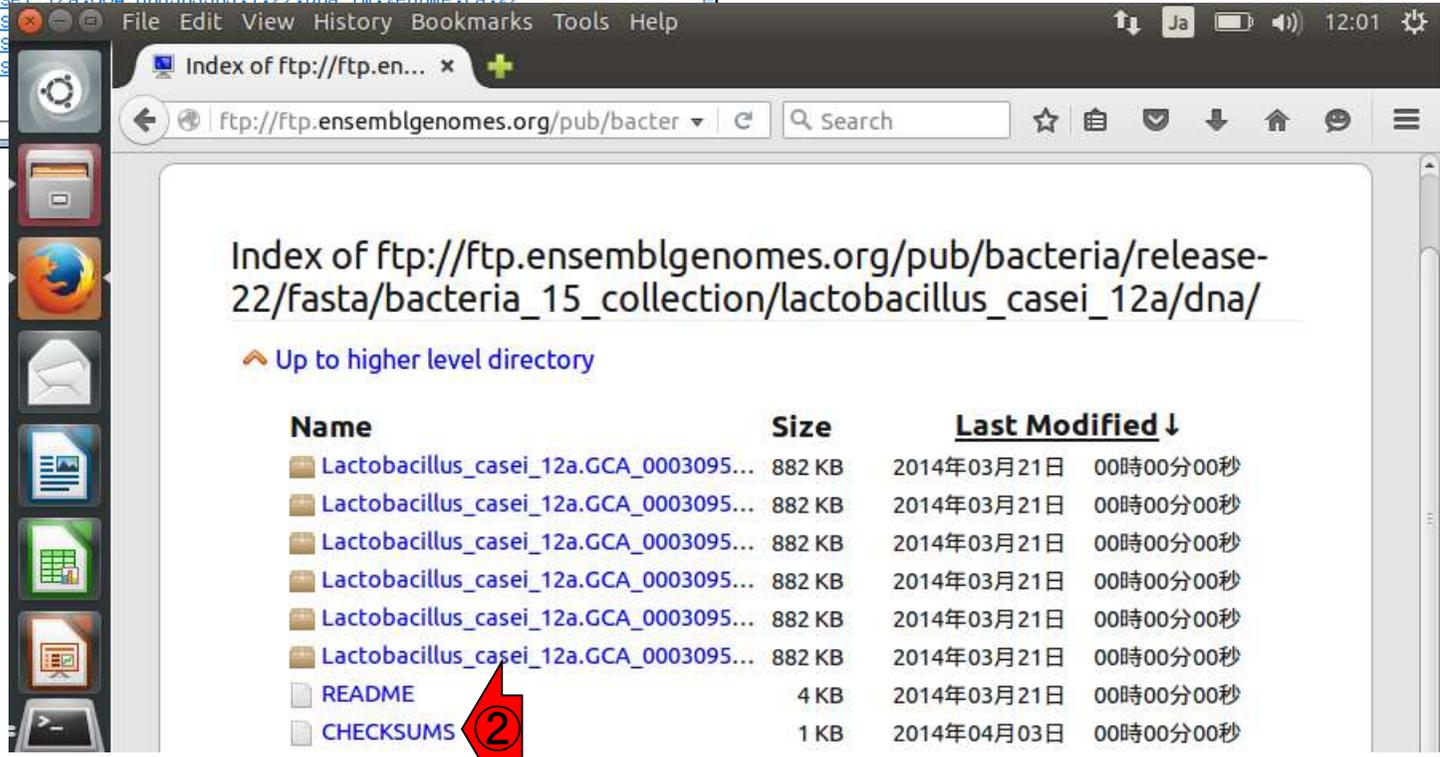


FTP ディレクトリ /pub/bacteria/release-22/fasta/bacteria_15_collection/lactobacillus_casei_12a/dna/ / ftp.ensemblgenomes.org

エクスプローラーでこの FTP サイトを表示するには、Alt キーを押して、[表示]をクリックし、[エクスプローラーで FTP サイトを開く]をクリックしてください。

1 階層上のディレクトリへ

04/03/2014 12:00午前	469	CHECKSUMS	①
03/21/2014 12:00午前	902,847	Lactobacillus_casei_12a.GCA_000309565.1.22.dna.genome.fa.gz	
03/21/2014 12:00午前	902,847	Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa.gz	
03/21/2014 12:00午前	902,952	Lactobacillus_casei_12a.GCA_000309565.1.22.dna.rm.genome.fa.gz	
03/21/2014 12:00午前	902,952	Lactobacillus_casei_12a.GCA_000309565.1.22.dna.rm.genome.fa.gz	
03/21/2014 12:00午前	902,953	Lactobacillus_casei_12a.GCA_000309565.1.22.dna.rm.genome.fa.gz	
03/21/2014 12:00午前	902,953	Lactobacillus_casei_12a.GCA_000309565.1.22.dna.rm.genome.fa.gz	
03/21/2014 12:00午前	3,111	README	



Index of ftp://ftp.ensemblgenomes.org/pub/bacteria/release-22/fasta/bacteria_15_collection/lactobacillus_casei_12a/dna/

Up to higher level directory

Name	Size	Last Modified ↓
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.genome.fa.gz	882 KB	2014年03月21日 00時00分00秒
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa.gz	882 KB	2014年03月21日 00時00分00秒
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.rm.genome.fa.gz	882 KB	2014年03月21日 00時00分00秒
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.rm.genome.fa.gz	882 KB	2014年03月21日 00時00分00秒
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.rm.genome.fa.gz	882 KB	2014年03月21日 00時00分00秒
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.rm.genome.fa.gz	882 KB	2014年03月21日 00時00分00秒
README	4 KB	2014年03月21日 00時00分00秒
CHECKSUMS	1 KB	2014年04月03日 00時00分00秒

②

W12: チェックサム

sumコマンドとmd5sumコマンド実行結果。赤枠部分のみ打ち込んでタブ補完するテクニックを有効活用すべし。

```
iu@bielinux[hoge] pwd [ 9:32午後 ]
/home/iu/Desktop/hoge
iu@bielinux[hoge] ls -lh [ 9:32午後 ]
total 884K
-rw-rw-r-- 1 iu iu 882K 11月 28 12:23 Lactobacillus_casei_12a.GCA_000309565
.1.22.dna.toplevel.fa.gz
iu@bielinux[hoge] sum Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplev
el.fa.gz
34752 882
iu@bielinux[hoge] md5sum Lactobacillus_casei_12a.GCA_000309565.1.22.dna.top
level.fa.gz
28224d6642574444b2b1c9f615f107c3 Lactobacillus_casei_12a.GCA_000309565.1.2
2.dna.toplevel.fa.gz
iu@bielinux[hoge] [ 9:32午後 ]
```

W12: チェックサム

②データ提供元(Ensembl)のウェブページでは
①sumコマンド実行結果の「チェックサム」が与えられていることがわかる。同じ値(34752 882)となっていれば、ダウンロードしたファイルの中身はデータ提供元と同じと判断してよい。

```
File Edit View Search Terminal Help
iu@bielinux[hoge] pwd
/home/iu/Desktop/hoge
iu@bielinux[hoge] ls -lh
total 884K
-rw-rw-r-- 1 iu iu 882K 11月 28 12:23 Lactobacillus_casei_12a.GCA_000309565
.1.22.dna.toplevel.fa.gz
iu@bielinux[hoge] sum Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplev
el.fa.gz
34752 882
iu@bielinux[hoge] md5sum Lactobacillus_casei_12a.GCA_000309565.1.22.dna.top
level.fa.gz
28224d6642574444b2b1c9f615f107c3 Lactobacillus_casei_12a.GCA_000309565.1.2
2.dna.toplevel.fa.gz
iu@bielinux[hoge]
```

[9:32午後]

[9:32午後]

```
ftp://ftp.ensemblgenomes.org/pub/bacteria/release-22/f
ftp.ensemblgenome... x
34752 882 Lactobacillus_casei_12a.GCA_000309565.1.22.dna.genome.fa.gz
34752 882 Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa.gz
14283 882 Lactobacillus_casei_12a.GCA_000309565.1.22.dna_rm.genome.fa.gz
14283 882 Lactobacillus_casei_12a.GCA_000309565.1.22.dna_rm.toplevel.fa.gz
20112 882 Lactobacillus_casei_12a.GCA_000309565.1.22.dna_sm.genome.fa.gz
20112 882 Lactobacillus_casei_12a.GCA_000309565.1.22.dna_sm.toplevel.fa.gz
12836 4 README
```

W12: チェックサム

Linuxコマンドもタブ補完可能。「MD5チェックサム」を実行するコマンドは「md…なんだっけ?!」という場合は、①「md」まで打ち込んでからTabキーを押す。すると、mdから始まるLinuxコマンドの候補がリストアップされるので、それを眺めながら残りの「md5sum」を打ち込めばよい。

```
File Edit View Search Terminal Help
total 884K
-rw-rw-r-- 1 iu iu 882K 11月 28 12:23 Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa.gz
iu@bielinux[hoge] sum Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa.gz
34752 882
iu@bielinux[hoge] md5sum Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa.gz
28224d6642574444b2b1c9f615f107c3 Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa.gz
iu@bielinux[hoge] md
md5pass          mdatopbm        mdir
md5sum           mdel            mdu
md5sum.textutils mdeltree        mdust
```

[9:32午後]



W13-1 : gzip

①gzip圧縮ファイルの解凍、②gzip圧縮ファイルの作成、および③チェックサム値の確認。ダウンロード直後のチェックサム値やファイルサイズとは異なることもあるが、基本的に気にしなくてもよい

```
iu@bielinux[hoge] pwd [ 9:49午後 ]
/home/iu/Desktop/hoge
iu@bielinux[hoge] ls -lh [ 9:49午後 ]
total 884K
-rw-rw-r-- 1 iu iu 882K 11月 28 12:23 Lactobacillus_casei_12a.GCA_000309565
.1.22.dna.toplevel.fa.gz
① iu@bielinux[hoge] gzip -d Lactobacillus_casei_12a.GCA_000309565.1.22.dna.to
plevel.fa.gz
iu@bielinux[hoge] ls -lh [ 9:49午後 ]
total 2.9M
-rw-rw-r-- 1 iu iu 2.8M 11月 28 12:23 Lactobacillus_casei_12a.GCA_000309565
.1.22.dna.toplevel.fa
② iu@bielinux[hoge] gzip Lactobacillus_casei_12a.GCA_000309565.1.22.dna.tople
vel.fa
iu@bielinux[hoge] ls -lh [ 9:49午後 ]
total 884K
-rw-rw-r-- 1 iu iu 882K 11月 28 12:23 Lactobacillus_casei_12a.GCA_000309565
.1.22.dna.toplevel.fa.gz
③ iu@bielinux[hoge] sum Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplev
el.fa.gz
25013 882
iu@bielinux[hoge] [ 9:50午後 ]
```

W13-2: gunzipなど

①gzip圧縮ファイルの解凍は「gzip -d」以外に「gunzip」コマンドでも可能。②bzip2コマンドを用いたbzip2圧縮（拡張子は.bz2）、③bunzip2コマンドを用いたbzip2圧縮ファイルの解凍。bunzip2の代わりに「bzip2 -d」を用いるのもよい。

```
iu@bielinux[hoge] pwd
/home/iu/Desktop/hoge
iu@bielinux[hoge] ls -lh
total 884K
-rw-rw-r-- 1 iu iu 882K 11月 28 12:23 Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa.gz
iu@bielinux[hoge] gunzip Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa.gz
iu@bielinux[hoge] ls -lh
total 2.9M
-rw-rw-r-- 1 iu iu 2.8M 11月 28 12:23 Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
iu@bielinux[hoge] bzip2 Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
iu@bielinux[hoge] ls -lh
total 824K
-rw-rw-r-- 1 iu iu 822K 11月 28 12:23 Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa.bz2
iu@bielinux[hoge] bunzip2 Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa.bz2
iu@bielinux[hoge] ls -lh
total 2.9M
-rw-rw-r-- 1 iu iu 2.8M 11月 28 12:23 Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
iu@bielinux[hoge]
```

[10:00午後]

[10:01午後]

[10:01午後]

[10:01午後]

[10:01午後]



①gzip圧縮(882K)よりも、②bzip2圧縮(822K)のほうが圧縮率が高いことがわかる。

W13-3: 圧縮効率

```
File Edit View Search Terminal Help
iu@bielinux[hoge] pwd [10:00午後]
/home/iu/Desktop/hoge
iu@bielinux[hoge] ls -lh [10:00午後]
total 884K
-rw-rw-r-- 1 iu iu 882K 11月 28 12:23 Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa.gz
iu@bielinux[hoge] gunzip Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa.gz
iu@bielinux[hoge] ls -lh [10:01午後]
total 2.9M
-rw-rw-r-- 1 iu iu 2.8M 11月 28 12:23 Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
iu@bielinux[hoge] bzip2 Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
iu@bielinux[hoge] ls -lh [10:01午後]
total 824K
-rw-rw-r-- 1 iu iu 822K 11月 28 12:23 Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa.bz2
iu@bielinux[hoge] bunzip2 Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa.bz2
iu@bielinux[hoge] ls -lh [10:01午後]
total 2.9M
-rw-rw-r-- 1 iu iu 2.8M 11月 28 12:23 Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
iu@bielinux[hoge] [10:01午後]
```

W14-1: mvとgrep

①mvコマンドでファイル名変更、②grepコマンドで">"を含む行の数を表示、③headコマンドでgenome.faの最初の10行を表示。

```
iu@bielinux[hoge] pwd [10:08午後]
/home/iu/Desktop/hoge
iu@bielinux[hoge] ls [10:08午後]
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
① iu@bielinux[hoge] mv Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa genome.fa
iu@bielinux[hoge] ls [10:08午後]
genome.fa
② iu@bielinux[hoge] grep -c ">" genome.fa [10:08午後]
28
③ iu@bielinux[hoge] head genome.fa [10:08午後]
>ctg09_00003 dna:supercontig supercontig:GCA_000309565.1:ctg09_00003:1:899:1
AGTAAGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCT
TATGACCTGGGCTACACACGTGCTACAATGGATGGTACAACGAGTTGCGAGACCGCGAGG
TYAAGCTAATCTCTTAAAGCCATTCTCAGTTCGGACTGTAGGCTGCAACTCGCCTACACG
AAGTCGGAATCGCTAGTAATCGCGGATCAGCACGCCGCGGTGAATACGTTCCCGGGCCTT
GTACACACCGCCCGTCACACCATGAGAGTTTGTAAACACCCGAAGCCGGTGGCGTAACCCT
TTTAGGGAGCGAGCCGTCTAAGGTGGGACAAATGATTAGGGTGAAGTCGTAACAAGGTAG
CCGTAGGAGAACCTGCGGCTGGATCACCTCCTTTCTAAGGAAACAGACTRAAAGTCTGAC
GGAAACCTGCACACACGAACTTTGTTTGTAGTTTGGAGGGATCACCTCAAGCACCTAA
CGGGTGCGACTTTGTTCTTTGAAACTGGATATCATTGTATTAATTGTTTTAAATTGCCG
iu@bielinux[hoge] [10:08午後]
```

①headコマンドでgenome.faの最初の4行分を表示
、②tailコマンドでgenome.faの最後の7行分を表示

W14-2: headとtail

```
iu@bielinux[~/Desktop/hoge]
iu@bielinux[hoge] pwd [10:12午後]
/home/iu/Desktop/hoge
iu@bielinux[hoge] ls [10:12午後]
genome.fa
① iu@bielinux[hoge] head -n 4 genome.fa [10:12午後]
>ctg09_00003 dna:supercontig supercontig:GCA_000309565.1:ctg09_00003:1:899:1
AGTAAGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCT
TATGACCTGGGCTACACACGTGCTACAATGGATGGTACAACGAGTTGCGAGACCGCGAGG
TYAAGCTAATCTCTTAAAGCCATTCTCAGTTCGGACTGTAGGCTGCAACTCGCCTACACG
② iu@bielinux[hoge] tail -n 7 genome.fa [10:12午後]
ACCTACTCAGTGTCCATGCTATGCTTGTTAGCGTCGAAACCAAAGCAAGCGAGGTTATG
AGTAAATGTCCCAATACGATCCTACACTGTCCGTCCTTGGAATACCAGACCATAATATCA
AAGTAGCCTTTGTTTCGTCATGAATATCGCGGCAACGGGGTACGTGCGCCGCCAGTATCATG
TGATTGATGCCGAGCTGACTTACCGGTTAACCGGTGCCACTGTGTGGCTTTGAGGCCT
TGCACCCTAACGGGTTTTATACGGCCACGTGCGCGTCCTCAACGGGGTTGAAATACCGA
CAGTCATTGACTTGCACAAGCAACGATGGCGCTGTCATAACTGTTACCACACAGTCAGTG
CCAAGACGCCACTCGTGCAAC
iu@bielinux[hoge] [10:12午後]
```

W14-3: wc

wcコマンドで「行数、単語数、ファイルサイズ」を表示。行数は48138、単語数は48194、ファイルサイズは2935947である。塩基配列部分は実質的に1行1単語となるので、行数と単語数の若干の違いはdescription行のところで出る。②Description行は3単語、③計28コンティグがあるので、行数(48138) + 2 × 28 = 48194という計算となる。このようにファイル全体を眺めなくてもwc, grep, headコマンドの組合せでdescription行の記述形式が同じであることを予想して確認もできる

```
iu@bielinux[~/Desktop/hoge]
iu@bielinux[hoge] pwd
/home/iu/Desktop/hoge
iu@bielinux[hoge] ls -l
total 2868
-rw-rw-r-- 1 iu iu 2935947 11月 28 12:23 genome.fa
① iu@bielinux[hoge] wc genome.fa [10:14午後]
  48138   48194 2935947 genome.fa
iu@bielinux[hoge] grep -c ">" genome.fa [10:15午後]
28
③ iu@bielinux[hoge] head -n 5 genome.fa [10:15午後]
>ctg09 00003 dna:supercontig supercontig:GCA 000309565.1:ctg09 00003:1:899:1 ②
AGTAAGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCT
TATGACCTGGGCTACACACGTGCTACAATGGATGGTACAACGAGTTGCGAGACCGCGAGG
TYAAGCTAATCTCTTAAAGCCATTCTCAGTTCGGACTGTAGGCTGCAACTCGCCTACACG
AAGTCGGAATCGCTAGTAATCGCGGATCAGCACGCCGCGGTGAATACGTTCCCGGGCCTT
iu@bielinux[hoge] █ [10:15午後]
```

grepコマンド実行時に">"を含む行数をカウントするのが「-c」オプションだが、さらに①「-v」オプションを追加することで">"を含まない行数をカウントすることができる

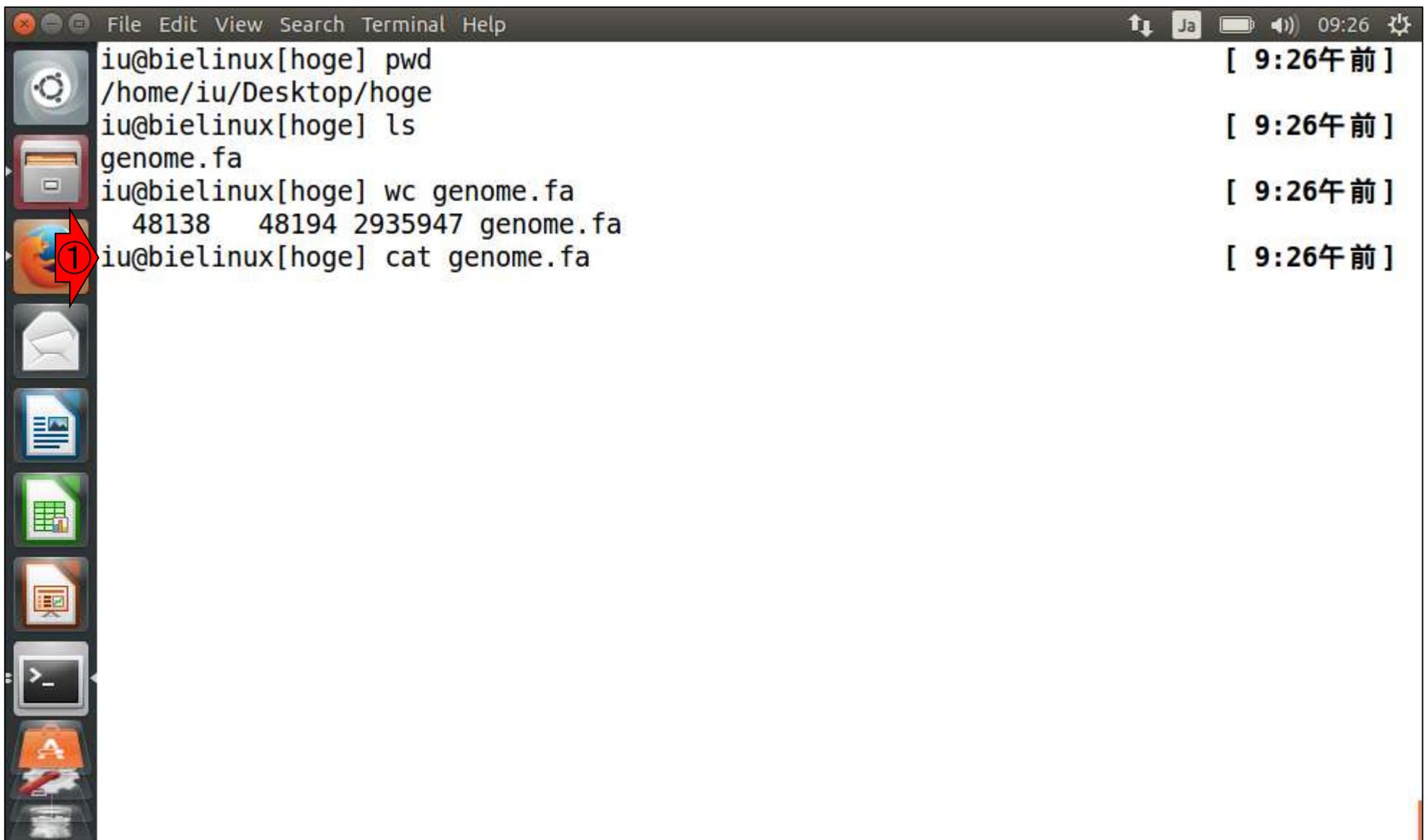
W14-3: grep -v

```
iu@bielinux[hoge] pwd [ 9:22午前 ]
/home/iu/Desktop/hoge
iu@bielinux[hoge] ls -l [ 9:22午前 ]
total 2868
-rw-rw-r-- 1 iu iu 2935947 11月 28 12:23 genome.fa
iu@bielinux[hoge] wc genome.fa [ 9:22午前 ]
 48138  48194 2935947 genome.fa
iu@bielinux[hoge] grep -c ">" genome.fa [ 9:22午前 ]
28
iu@bielinux[hoge] head -n 5 genome.fa [ 9:22午前 ]
>ctg09_00003 dna:supercontig supercontig:GCA_000309565.1:ctg09_00003:1:899:1
AGTAAGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCT
TATGACCTGGGCTACACACGTGCTACAATGGATGGTACAACGAGTTGCGAGACCGCGAGG
TYAAGCTAATCTCTTAAAGCCATTCTCAGTTCGGACTGTAGGCTGCAACTCGCCTACACG
AAGTCGGAATCGCTAGTAATCGCGGATCAGCACGCCGCGGTGAATACGTTCCCGGGCCTT
iu@bielinux[hoge] grep -c -v ">" genome.fa [ 9:22午前 ]
48110
iu@bielinux[hoge] █ [ 9:22午前 ]
```



W14-4: cat

catコマンド実行前。リターンキーを押すと一気にファイルの全内容が表示される



```
iu@bielinux[hoge] pwd [ 9:26午前 ]
/home/iu/Desktop/hoge
iu@bielinux[hoge] ls [ 9:26午前 ]
genome.fa
iu@bielinux[hoge] wc genome.fa [ 9:26午前 ]
 48138  48194 2935947 genome.fa
iu@bielinux[hoge] cat genome.fa [ 9:26午前 ]
```

W14-4: cat

catコマンド実行後。私の環境では5秒程度で表示が完了する。このファイルは5万行弱なので、概ね1万行につき1秒程度表示に時間がかかっていることがわかる。①「tail -n 3」でファイルの最後の3行分を表示させて確認。



```
File Edit View Search Terminal Help
AAGGCGATTAAGTTCTCCACCATTGATTCCAACCTGCAACTGACCATGGCCTTCATTGTG
AACTCACTGCTGCTGATTCTGGGCGCAGCGCTGTTCTTCGGCACCAGTAGCTCAGTTGGC
CGTTTCGTTGATCTTTTCAACGCACTAAGCAATTCTCAGATCGTCGGTGGCGATCGCCAGC
CCGATGCTTTCCATGTTGTTGCGGGTTGCGCTGCTCGCATCCGGCCAAAGCTCCACGATT
ACAGGAACACTTGCCGGTCAGATTATCATGGAAGGCTTCATTCATTTGAAAATGCCGCTC
TGGGCCCAACGACTTTTGACTCGGCTCATGTCCGTGACCCAGTGCTCATCTTCGCGATT
TACTACCATGGCAACGAAGCCAAAATCGAAAACCTGTTGACCTTCTCGCAGGTCTTCCTC
AGTATCGCATTACCATTGCGCGTCATCCCGCTAGTGCTCTACACTAGCGATAAAAAAATC
ATGGGTGAGTTCGCCAACCGTGCTTGGGTGAAGTGGACAGCATGGTTCATCAGTGGGGTC
CTGATCATTTTGAACCTTTACCTGATTGCGCAGACGTTAGGGTTTGTAAATAAGTCATT
GTAGCAAGCAAAGCCGCTCGATTGCTGATCGAGCGGCTTTTAGTGGGTCTACACTTTC
TGGTGTTGAGAAATAATCAACACTGAACGTGTAGGCCCTTTTGGATAACGCAAAAAGAC
ACCTACTCAGTGTCCATGCTATGCTTGTAGCGTCGAAACCAAAGCAAGCGAGGTTATG
AGTAAATGTCCCAATACGATCCTACACTGTCCGTCCTTGAATACCAGACCATAATATCA
AAGTAGCCTTTGTTGTCATGAATATCGCGGCAACGGGGTACGTCGCCGCCAGTATCATG
TGATTGATGCCGAGCTGACTTACCGGTTAACCCGGTGCCCACTGTGTGGCTTTGAGGCCT
TGCACCCTAACGGGTTTTATACGGCCACGTGCGCGTCCTCAACGGGGTTGAAATACCGA
CAGTCATTGACTTGCACAAGCAACGATGGCGCTGTCATAACTGTTACCACACAGTCAGTG
CCAAGACGCCACTCGTGCAAC
iu@bielinux[hoge] tail -n 3 genome.fa
TGCACCCTAACGGGTTTTATACGGCCACGTGCGCGTCCTCAACGGGGTTGAAATACCGA
CAGTCATTGACTTGCACAAGCAACGATGGCGCTGTCATAACTGTTACCACACAGTCAGTG
CCAAGACGCCACTCGTGCAAC
iu@bielinux[hoge]
```

[9:28午前]

[9:28午前]

Tips。catコマンド実行後、全内容表示を途中で止めたい場合には、「CTRLキー + Cキー」を押せばよい。

W14-4: 中断



```
TTAGTCGCGAGATCATCAATCTTTTTGGCTAACACCGGCGAGATGTTTCGGTGCAGGCGCA
AGCGTGAGATGCAAAACATGGTGCTGATGGCGATTCTCAACGATCGGAAAAACACGCACT
TGATCGCGGCCATCACGTGTCATCATAACGGAGACTTCACGATCAAATTGCTGCCGTTTT
TCGAGAATGCAAGGACCATGCGTCAAAGTTCGCGGGCGGCGGGAAAATCATCTGGTGTA
TTGATATCTTGCTGACCATGGCCATCATAACCGCCAGTGGTTCGTCCTTCAGGATGCTTGGT
AAGCCGATGGTCGCGACTGCTTGTGCCAACTCAGCGGCACGGTTGACAGCGGCAAAGAC
GTCACCGGTACCCCATGTGTCTTTAAAAAAGTCTTTTCTTTGATCCGATCGCGGGTTATC
GCCAAAATGGCGGTACCTTGAGGAATCCGCGTCTCGCTGCGAACGGTTTCCAACGCGGCT
AAGTCAACATTCTCAAACATCATATGTGAGCACATCTGACGCATCAGCCAGTTGCTTTAAG
GCGGCTTGATCGTTATAGGGCGCAGTGATTTGAAAATCCGCTACCTGACCGGCTGGCGAA
TTGGGTGTCGGCTCTAAAATACCGACGCGATAACCCATCGACTTGGCGCTTTGTGTCAGC
ATGCGTCCCAACTGACCGCCGCCAATGATGCCAATTGTTGCTGGGGACGTCACTACATCA
ATCATTGAAGCGCGGCCTCGCTTTCTTTGGCAGCTTAGTTTGAGCATCGCGAAAAGCTT
GCAGTCGGTCAGCCAGTTCTGAATCGTTCAGCGCTAAAATAGCGGCAGCGAACAACGCCG
CATTGACAGCTCCGGCATCACCGATGGCCATCGTGCCAACCGGGACCCCGCCAGGCATTT
GCACGATTGACAGGAGACTATCAACCCCATTTAATGTTTCGTGTTTTGATGGGAACCCCGA
TAACTGGCAGCAATGTATTAGCGGCGAGCATTCCAGGTAAATGAGCAGCGCCGCCTGCAC
CGGCGATAATTGCACCAAACCCAGTGCCACTGCCTGTTTACCAAATCCGCCAGCTCGT
CTGGCATCCGATGAGCCGAAATAACCCGCTTCTCATAAGGAATAGCCAATGACGTTAACT
GAGCCGTCACCTTGCTGCATGGTTCGGCCAATCAGAGGTTGACCCCATGACCACTGCTACTT
TTTGCATCTCGTTCCCGCTTTCTTTAAGATGTTTTGAGAATAGCACGTCAGCGCGACAG
ATTCAATAAAAATGAGAACAATTCTCAATGTTATATTAATAATCGTTCGTGTTTCTCTTT
^C%
iu@bielinux[hoge] █
```

[10:35午前]

W14-5: more

「more genome.fa」実行直後。スペースキーを押すと、次ページに移動

```
File Edit View Search Terminal Help 11:01
>ctg09_00003 dna:supercontig supercontig:GCA_000309565.1:ctg09_00003:1:899:1
AGTAAGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCT
TATGACCTGGGCTACACACGTGCTACAATGGATGGTACAACGAGTTGCGAGACCGCGAGG
TYAAGCTAATCTCTTAAAGCCATTCTCAGTTCGGACTGTAGGCTGCAACTCGCCTACACG
AAGTCGGAATCGCTAGTAATCGCGGATCAGCACGCCGCGGTGAATACGTTCCCGGGCCTT
GTACACACCGCCCGTCACACCATGAGAGTTTGTAAACCCGAAGCCGGTGGCGTAACCCT
TTTAGGGAGCGAGCCGTCTAAGGTGGGACAAATGATTAGGGTGAAGTCGTAACAAGGTAG
CCGTAGGAGAACCTGCGGCTGGATCACCTCCTTTCTAAGGAAACAGACTRAAAGTCTGAC
GGAAACCTGCACACACGAACTTTGTTTAGTTTTGAGGGGATCACCTCAAGCACCTAA
CGGGTGC GACTTTGTTCTTTGAAAACCTGGATATCATTGTATTAATTGTTTTAAATTGCCG
AGAACACAGCGTATTTGTATGAGTTTCTGAAAAAGAAATTCGCATCGCATAACCGCTGAC
GCAAGTCAGTACAGGTTAAGTTRCAAAGGGCGCACGGTGGATGCCCTTGGCACTAGGAGCC
GATGAAGGACGGAACATAACCGATATGCTTCGGGGAGCTATAAGTAAGCTTTGATCCGG
AGATTTCCGAATGGGGGAACCCAGTACACATCAGTGTATTGCTTGTGAGTGAATACATAG
CTGGCCGGCGGCAGACGCGGGGAACTGAAACATCTCAGTACCCGCAGGAAGAGAAAGAAA
ACTCGATTCCCATAGTAGCGGCGAGCGAAGTGGGAAAAGCCCAAACCGAGAAGCTTGCT
>ctg03_00032 dna:supercontig supercontig:GCA_000309565.1:ctg03_00032:1:2417:1
CGACTGTAAACATCCGCACTGGTGCCGGCACCGGCTATGCATCCGTTGGTAGCTATACAC
TCGGTGAAAGTGTGATTTATGATCACGTGTATATCCGTGGCACATATGTTTGGGCGCGTT
ATCTCAGCTACTCAGGCAGGTATCACTATGTTGCGCTAGGCGTAAATGGTGGTGAGAGCT
ATGGTTCGCGCAGTTCAAATGCGCAAACCTATTCGCACACGTA CTACACAGTCCGCTCTG
GTGACAGCTTCTGGAGTATTGCCAGCAAGTATGGCATTAGCATGTACACGCTGGCTGCCA
ACAACGGAAAATCAATCTACAGCCTGATCTATCCGGGCGAAAGCCTGTATATTAGGTAAC
--More-- (0%)
```

W14-5: more

スペースキーを1回だけ押して次ページに移動した状態。スペースキーを押し続けると①%のところの数値がどんどんあがっていき、100%までいくと最後まで閲覧した状態になって、コマンド入力待ち状態(コマンドプロンプト)となる。途中でやめたい場合は、「qキー」を押すとよい。

iu@bielinux[~/Desktop/hoge]

```
AAACAGCCCTCTGCTCGCTAACGCGGGTGGAGGAC
GTGGATAAGTGGCTTTGTGAAGATTTGCTAGGAAC
AACCAGATTAGTACCGATTTACGGTTAAAACGTTGTTATATCAGGCATACTATGCCATT
CACGCTAACTTATCCACAGGGGTTGACATCTAGAATGATATAATATTCATAGAAAGGAGA
GGGAACACATGGATGTCATGAAAATTGCAAATTGGTTTCGCGTCAGGAATTATTCTGATA
TGCAAGTTGATGAGAGTATAGATGAGCTCACTCAGCTAAAGCTGATGAAGCTACTGTATT
TTTCCCAAGGTGTTTTTCTCGCCGCATATCATAAGAGATTATTTGATGATCCCATTCTTG
CGTGGAATTATGGTCCGGTGGTCAAACAAGTACATGATAAGTACGTAGGAAAACGTGGAA
TTGTCAATGACAATGTGCCAATTACGAATGCTGACTTGAGCGACTATGAAAATATAAATT
CCGATGAAGAGGGCGTCAGCAGTTCTTGAAGCTGTTGATCGTGCCTATGGTTCAATGTCTT
CAAGCGATTTGGTCAACATTACGCACTCGCAAACACCGTGGAAAACGACTCCGCAGAGCG
GCGAAATTAGTACTGATAAAATAGAGAAGTACTTTAAGGATCACGTTGTGACATATGGCT
AAGAAAAAGAAAATTTATATTTCTCAGCCTAGTGTTACAGATCCGGGCGTTTTTTTGTCT
AACTCAGTTCGGACAGAGCGGCTCACGTTAACTTTTCGTTTCTCACAGCTGACAAAAG
TACTCGATAGCCGGCTTGTACACGAAGACAGTAAAATCCACAAGAAAATTCTGGAACGA
ATAGAGGCATTGTGAGAGAAGAGATAAACTCTAGTATTTATTGACGGCAAGGAGTCTGGG
CTGGAGACTTTACCTCAAGAAGAAGTTCATTTTCGATTAATAGCGAGTATATCGATAGT
AAAAGAGATGCACAGGCAGGGACTAAATATTGGGTCTTCCGGTTAAACAAGCAAGGCAGA
GTAATCGGAAAGATTATTGGAAATATTTTTTATGTCCTTGCCGTTGACACTTCATTCGAT
GCATACAGACATTAAGATAAATAAGCCTCCTACCAGCAATGGCGGGAGGCTTATTTTTG
TGCATAAATCACATTGTTTACAAAAATAGTTAAAAGATGTAATACGGTTTAAACGATGTA
GAGAGACAAGCAGGTCAATAATTGCAACCGGTTGACAAAAGTACCGTGAAAAGGCCAAAAG
GCTTACATACGCCGGCTTAGGTAACCTATTATGAGCAAATTGTTTGCATACCCAGCATGG
--More-- (0%)
```

①

W14-5: more



```
File Edit View Search Terminal Help
AAACAGCCCTCTGCTCGCTAACGCGGGTGGAGGACTTTTTTAGAACCAAGTAAATTTCT
GTGGATAAGTGGCTTTGTGAAGATTTGCTAGGAACATTGATACAACGGTGTTTTGTAA
AACCAGATTAGTACCGATTTACGGTAAAACGTTGTTATATCAGGCATACTATGCCATT
CACGCTAACTTATCCACAGGGGTTGACATCTAGAATGATATAATATTCATAGAAAGGAGA
GGGAACACATGGATGTCATGAAAATTGCAAATTGGTTTCGCGTCAGGAATTATTCTGATA
TGCAAGTTGATGAGAGTATAGATGAGCTCACTCAGCTAAAGCTGATGAAGCTACTGTATT
TTTCCAAGGTGTTTTCTCGCCGCATATCATAAGAGATTATTTGATGATCCCATTCTTG
CGTGGAATTATGGTCCGGTGGTCAAACAAGTACATGATAAGTACGTAGGAAAACGTGGAA
TTGTCAATGACAATGTGCCAATTACGAATGCTGACTTGAGCGACTATGAAAATATAAATT
CCGATGAAGAGGCGTCAGCAGTTCTTGAAGCTGTTGATCGTGCCTATGGTTCAATGTCTT
CAAGCGATTTGGTCAACATTACGCACTCGCAAACACCGTGGAAAACGACTCCGCAGAGCG
GCGAAATTAGTACTGATAAAATAGAGAAGTACTTTAAGGATCACGTTGTGACATATGGCT
AAGAAAAAGAAAATTTATATTTCTCAGCCTAGTGTTACAGATCCGGGCGTTTTTTGTCT
AACTCAGTTCGGACAGAGCGGCTCACGTTAACTTTTCGTTTCTCACAGCTGACAAAAG
TACTCGATAGCCGGCTTGTACACGAAGACAGTAAAATCCACAAGAAAATTCTGGAACGA
ATAGAGGCATTGTGAGAGAAGAGATAAACTCTAGTATTTATTGACGGCAAGGAGTCTGGG
CTGGAGACTTTACCTCAAGAAGAAGTTCATTTTCGATTAATAGCGAGTATATCGATAGT
AAAAGAGATGCACAGGCAGGGACTAAATATTGGGTCTTCCGGTTAAACAAGCAAGGCAGA
GTAATCGGAAAGATTATTGGAAATATTTTTATGTCCTTGCCGTTGACACTTCATTCGAT
GCATACAGACATTAAGATAAATAAGCCTCCTACCAGCAATGGCGGGAGGCTTATTTTTG
TGCATAAATCACATTGTTTACAAAAATAGTTAAAAGATGTAATACGGTTAAACGATGTA
GAGAGACAAGCAGGTCAATAATTGCAACCGGTTGACAAAAGTACCGTGAAAAGGCCAAAAG
GCTTACATACGCCGGCTTAGGTAACCTATTATGAGCAAATTGTTTGCATACCCAGCATGG
iu@bielinux[hoge]
```

[11:14午前]

「less genome.fa」実行直後。スペースキーを押すと、次ページに移動。

W14-6-1 : less



```
>ctg09_00003 dna:supercontig supercontig:GCA_000309565.1:ctg09_00003:1:899:1
AGTAAGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCT
TATGACCTGGGCTACACACGTGCTACAATGGATGGTACAACGAGTTGCGAGACCGCGAGG
TYAAGCTAATCTCTTAAAGCCATTCTCAGTTCGGACTGTAGGCTGCAACTCGCCTACACG
AAGTCGGAATCGCTAGTAATCGCGGATCAGCACGCCGCGGTGAATACGTTCCCGGGCCTT
GTACACACCGCCCGTCACACCATGAGAGTTTGTAAACCCGAAGCCGGTGGCGTAACCCT
TTTAGGGAGCGAGCCGTCTAAGGTGGGACAAATGATTAGGGTGAAGTCGTAACAAGGTAG
CCGTAGGAGAACCTGCGGCTGGATCACCTCCTTTCTAAGGAAACAGACTRAAAGTCTGAC
GGAAACCTGCACACACGAAACTTTGTTTAGTTTTGAGGGGATCACCTCAAGCACCTAA
CGGGTGC GACTTTGTTCTTTGAAAACCTGGATATCATTGTATTAATTGTTTTAAATTGCCG
AGAACACAGCGTATTTGTATGAGTTTCTGAAAAAGAAATTCGCATCGCATAACCGCTGAC
GCAAGTCAGTACAGGTTAAGTTRCAAAGGGCGCACGGTGGATGCCTTGGCACTAGGAGCC
GATGAAGGACGGAACATAACCGATATGCTTCGGGGAGCTATAAGTAAGCTTTGATCCGG
AGATTTCCGAATGGGGGAACCCAGTACACATCAGTGTATTGCTTGTGAGTGAATACATAG
CTGGCCGGCGGCAGACGCGGGGAACTGAAACATCTCAGTACCCGCAGGAAGAGAAAGAAA
ACTCGATTCCCATAGTAGCGGCGAGCGAAGTGGGAAAAGCCCAAACCGAGAAGCTTGCT
>ctg03_00032 dna:supercontig supercontig:GCA_000309565.1:ctg03_00032:1:2417:1
CGACTGTAAACATCCGCACTGGTGCCGGCACCGGCTATGCATCCGTTGGTAGCTATACAC
TCGGTGAAAGTGTGATTTATGATCACGTGTATATCCGTGGCACATATGTTTGGGCGCGTT
ATCTCAGCTACTCAGGCAGGTATCACTATGTTGCGCTAGGCGTAAATGGTGGTGAGAGCT
ATGGTTCGCGCAGTTCAAATGCGCAAACCTATTCGCACACGTA CTACACAGTCCGCTCTG
GTGACAGCTTCTGGAGTATTGCCAGCAAGTATGGCATTAGCATGTACACGCTGGCTGCCA
ACAACGGAAAATCAATCTACAGCCTGATCTATCCGGGCGAAAGCCTGTATATTAGGTAAC
genome.fa
```

W14-6-1 : less

①「Gキー」を押して、最終ページに移動した状態。「gキー」を押すと、最初のページに移動する



```
GTTGCGACCTTAGTGCTGGTTATCTTATTGGTTTTCGCTTATGAAGTGATTCTCGCGCAA
CCAAATGTCCCGGAACTGCTGAAAGGCTATCTGCCACACGCGGATATCGTGACCAACAAA
TCAATGCTCTATCTCTCTTTGGGGATTGTTGGTGCCACGGTGATGCCGCATGACCTGTTCT
CTCGGCTCCTCGATTTGCGAAACACGAAAAATTGATCGCACGAAGCACGAAGAAGTCAA
AAGGCGATTAAGTTCTCCACCATTGATTCCAACCTGCAACTGACCATGGCCTTCATTGTG
AACTACTGCTGCTGATTCTGGGCGCAGCGCTGTTCTTCGGCACCAGTAGCTCAGTTGGC
CGTTTCGTTGATCTTTTCAACGCACTAAGCAATTCTCAGATCGTCGGTGCGATCGCCAGC
CCGATGCTTTCCATGTTGTTGCGGGTTGCGCTGCTCGCATCCGGCCAAAGCTCCACGATT
ACAGGAACACTTGCCGGTCAGATTATCATGGAAGGCTTCATTCATTTGAAAATGCCGCTC
TGGGCCCAACGACTTTTGACTCGGCTCATGTCCGTGACCCAGTGCTCATCTTCGCGATT
TACTACCATGGCAACGAAGCCAAAATCGAAAACCTGTTGACCTTCTCGCAGGTCTTCCTC
AGTATCGCATTACCATTGCGCGTCATCCCGCTAGTGCTCTACACTAGCGATAAAAAAATC
ATGGGTGAGTTCGCCAACCGTGCTTGGGTGAAGTGGACAGCATGGTTCATCAGTGGGGTC
CTGATCATTTTGAACCTTTACCTGATTGCGCAGACGTTAGGGTTTGTTAAATAAGTCATT
GTAGCAAGCAAAGCCGCTCGATTCGCTGATCGAGCGGCTTTTAGTGGGTCTACACTTTC
TGGTGTTGAGAAATAATCAACACTGAACGTGTAGGCCCTTTTTGGATAACGCAAAAAGAC
ACCTACTCAGTGTCCATGCTATGCTTGTTAGCGTCGAAACCAAAGCAAGCGAGGTTATG
AGTAAATGTCCAATACGATCCTACACTGTCCGTCCTTGGAATACCAGACCATAATATCA
AAGTAGCCTTTGTTGTCATGAATATCGCGGCAACGGGGTACGTCGCCGCCAGTATCATG
TGATTGATGCCGAGCTGACTTACCGGTTAACCCGGTGCCCACTGTGTGGCTTTGAGGCCT
TGCACCCTAACGGGTTTTATACGGCCACGTGCGCGTCCTCAACGGGGTTGAAATACCGA
CAGTCATTGACTTGACAAGCAACGATGGCGCTGTCATAACTGTTACCACACAGTCAGTG
CCAAGACGCCACTCGTGCAAC
(END)
```

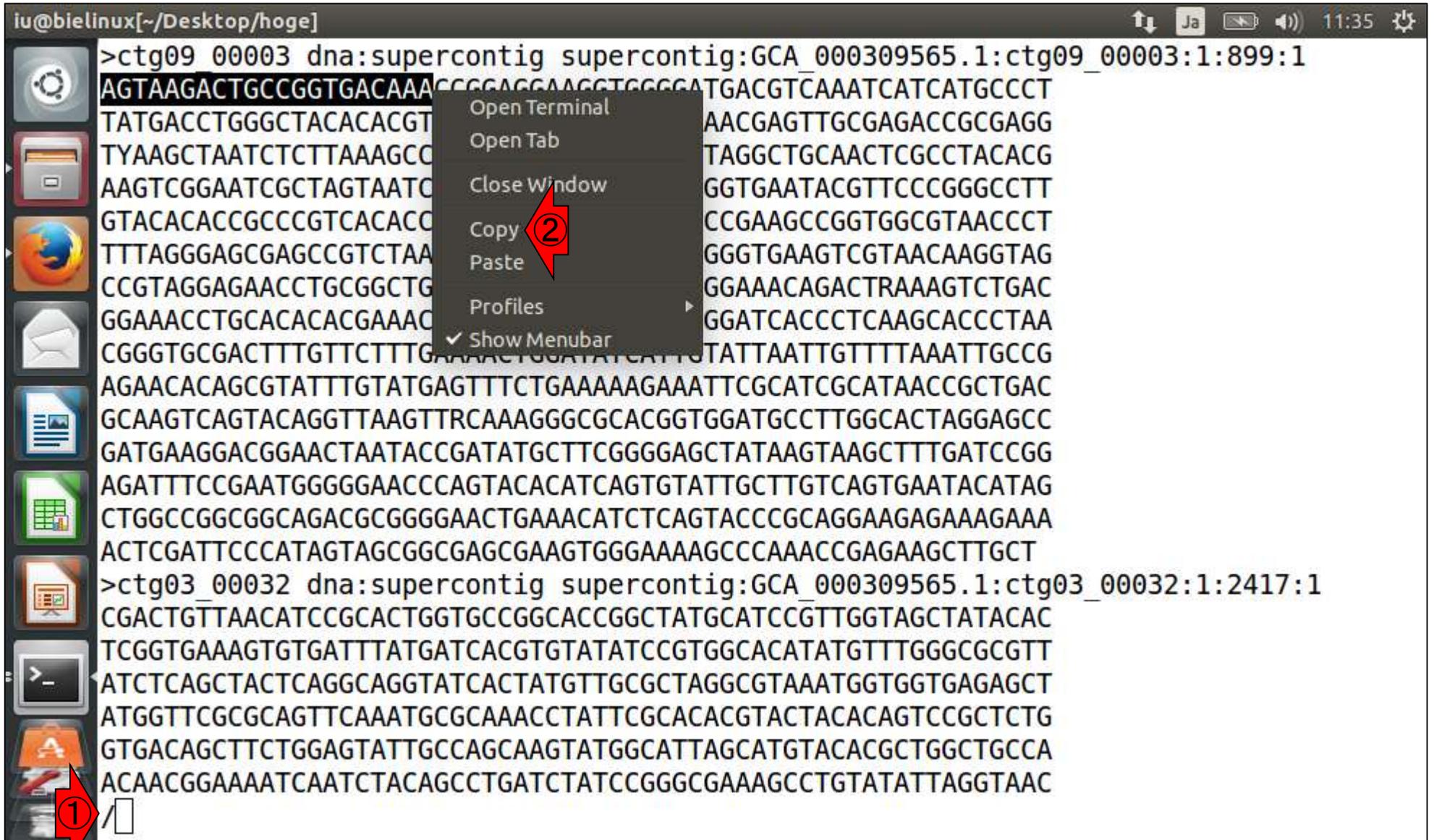
W14-6-1 : less

①「gキー」を押して、最初のページに移動した状態。
moreコマンドと同様に、「qキー」を押すとコマンド入力
待ち状態に戻る(がここではやらずに次のスライドへ)。

```
File Edit View Search Terminal Help 11:33
>ctg09_00003 dna:supercontig supercontig:GCA_000309565.1:ctg09_00003:1:899:1
AGTAAGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCT
TATGACCTGGGCTACACACGTGCTACAATGGATGGTACAACGAGTTGCGAGACCGCGAGG
TYAAGCTAATCTCTTAAAGCCATTCTCAGTTCGGACTGTAGGCTGCAACTCGCCTACACG
AAGTCGGAATCGCTAGTAATCGCGGATCAGCACGCCGCGGTGAATACGTTCCCGGGCCTT
GTACACACCGCCCGTCACACCATGAGAGTTTGTAAACACCCGAAGCCGGTGGCGTAACCCT
TTTAGGGAGCGAGCCGTCTAAGGTGGGACAAATGATTAGGGTGAAGTCGTAACAAGGTAG
CCGTAGGAGAACCTGCGGCTGGATCACCTCCTTTCTAAGGAAACAGACTRAAAGTCTGAC
GGAAACCTGCACACACGAACTTTGTTTAGTTTTGAGGGGATCACCTCAAGCACCTAA
CGGGTGCAGCTTTGTTCTTTGAAAACCTGGATATCATTGTATTAATTGTTTTAAATTGCCG
AGAACACAGCGTATTTGTATGAGTTTCTGAAAAAGAAATTCGCATCGCATAACCGCTGAC
GCAAGTCAGTACAGGTTAAGTTRCAAAGGGCGCACGGTGGATGCCCTTGGCACTAGGAGCC
GATGAAGGACGGAACATAACCGATATGCTTCGGGGAGCTATAAGTAAGCTTTGATCCGG
AGATTTCCGAATGGGGGAACCCAGTACACATCAGTGTATTGCTTGTGAGTGAATACATAG
CTGGCCGGCGGCAGACGCGGGGAACTGAAACATCTCAGTACCCGCAGGAAGAGAAAGAAA
ACTCGATTCCCATAGTAGCGGCGAGCGAAGTGGGAAAAGCCCAAACCGAGAAGCTTGCT
>ctg03_00032 dna:supercontig supercontig:GCA_000309565.1:ctg03_00032:1:2417:1
CGACTGTAAACATCCGCACTGGTGCCGGCACCGGCTATGCATCCGTTGGTAGCTATACAC
TCGGTGAAAGTGTGATTTATGATCACGTGTATATCCGTGGCACATATGTTTGGGCGCGTT
ATCTCAGCTACTCAGGCAGGTATCACTATGTTGCGCTAGGCGTAAATGGTGGTGAGAGCT
ATGGTTCGCGCAGTTCAAATGCGCAAACCTATTCGCACACGTAACACAGTCCGCTCTG
GTGACAGCTTCTGGAGTATTGCCAGCAAGTATGGCATTAGCATGTACACGCTGGCTGCCA
ACAACGGAAAATCAATCTACAGCCTGATCTATCCGGGCGAAAGCCTGTATATTAGGTAAC
:
```

W14-6-2: 文字列検索

①スラッシュ(/)を打ち込んでから、そこにあることがわかっている文字列を②コピー

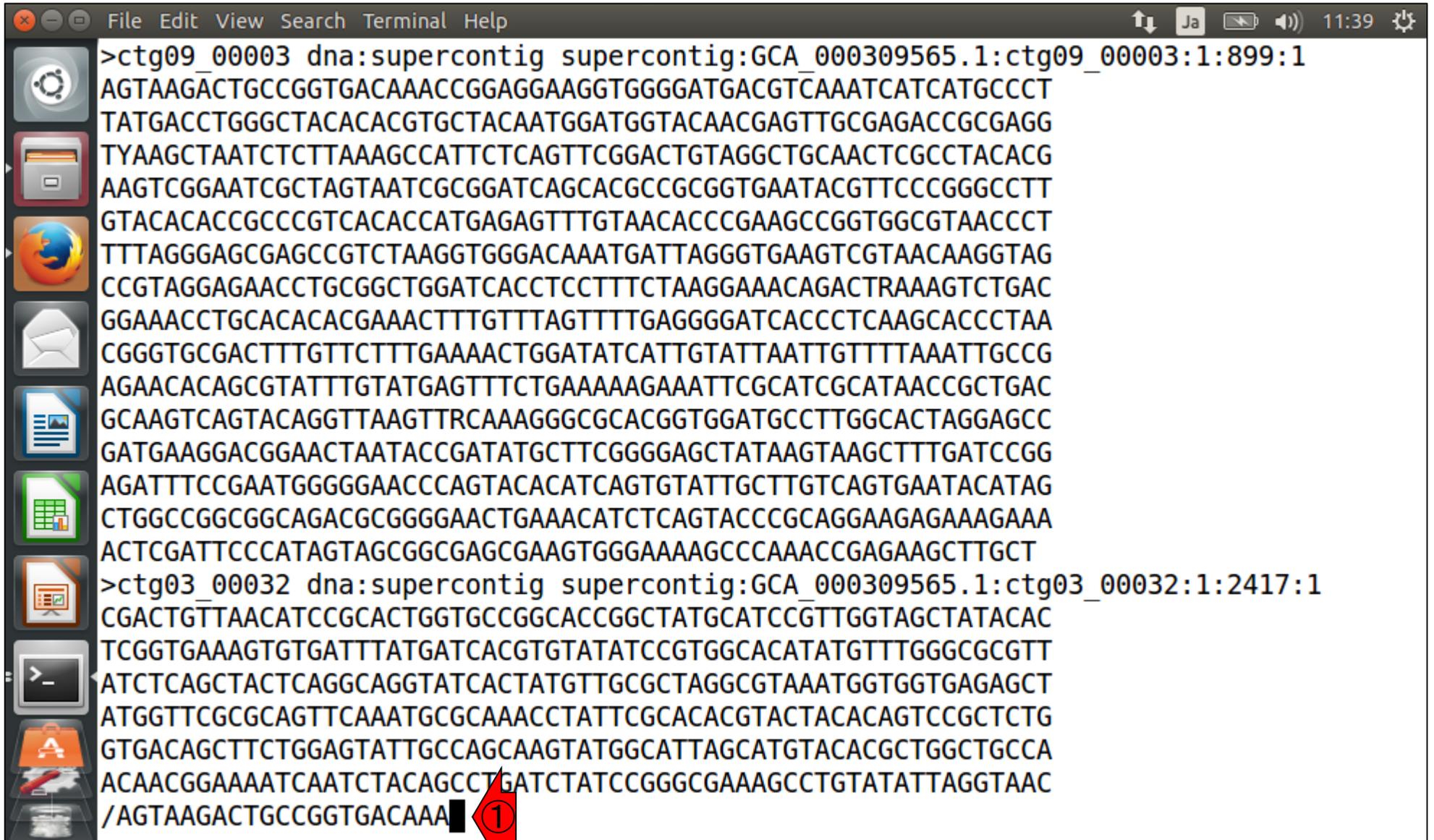


The image shows a terminal window with a dark background and light text. The terminal prompt is `iu@bielinux[~/Desktop/hoge]`. The first command is `>ctg09_00003 dna:supercontig supercontig:GCA_000309565.1:ctg09_00003:1:899:1`. The output is a long DNA sequence. A context menu is open over the first line of the sequence, with the `Copy` option highlighted by a red arrow and a circled '2'. A red arrow and a circled '1' point to the terminal icon in the desktop environment on the left side of the window.

```
iu@bielinux[~/Desktop/hoge]
>ctg09_00003 dna:supercontig supercontig:GCA_000309565.1:ctg09_00003:1:899:1
AGTAAGACTGCCGGTGACAAACCGGAGCAAGCTGGGATGACGTCAAATCATCATGCCCT
TATGACCTGGGCTACACACGTTAACGAGTTGCGAGACCGCGAGG
TYAAGCTAATCTCTTAAAGCCTAGGCTGCAACTCGCCTACACG
AAGTCGGAATCGCTAGTAATCGGTGAATAACGTTCCCGGGCCTT
GTACACACCGCCCGTCACACC CGAAGCCGGTGGCGTAACCCT
TTTAGGGAGCGAGCCGTCTAA GGGTGAAGTCGTAACAAGGTAG
CCGTAGGAGAACCTGCGGCTG GAAACAGACTRAAAGTCTGAC
GGAAACCTGCACACACGAAAC GGATCACCTCAAGCACCTAA
CGGGTGC GACTTTGTTCTTTGATTAATTGTTTTAAATTGCCG
AGAACACAGCGTATTTGTATGAGTTTCTGAAAAAGAAATTCGCATCGCATAACCGCTGAC
GCAAGTCAGTACAGGTTAAGTTRCAAAGGGCGCACGGTGGATGCCCTTGGCACTAGGAGCC
GATGAAGGACGGA ACTAATACCGATATGCTTCGGGGAGCTATAAGTAAGCTTTGATCCGG
AGATTTCCGAATGGGGGAACCCAGTACACATCAGTGTATTGCTTGTGAGTGAATACATAG
CTGGCCGGCGGCAGACGCGGGGAACTGAAACATCTCAGTACCCGCAGGAAGAGAAAGAAA
ACTCGATTCCCATAGTAGCGGCGAGCGAAGTGGGAAAAGCCCAAACCGAGAAGCTTGCT
>ctg03_00032 dna:supercontig supercontig:GCA_000309565.1:ctg03_00032:1:2417:1
CGACTGTAAACATCCGCACTGGTGCCGGCACCGGCTATGCATCCGTTGGTAGCTATACAC
TCGGTGAAAGTGTGATTTATGATCACGTGTATATCCGTGGCACATATGTTTGGGCGCGTT
ATCTCAGCTACTCAGGCAGGTATCACTATGTTGCGCTAGGCGTAAATGGTGGTGAGAGCT
ATGGTTCGCGCAGTTCAAATGCGCAAACCTATTCGCACACGTA CTACACAGTCCGCTCTG
GTGACAGCTTCTGGAGTATTGCCAGCAAGTATGGCATTAGCATGTACACGCTGGCTGCCA
ACAACGGAAAATCAATCTACAGCCTGATCTATCCGGGCGAAAGCCTGTATATTAGGTAAC
```

右クリックでペーストした直後。①のようになっているはず。リターンキーを押す。

W14-6-2: 文字列検索



```
File Edit View Search Terminal Help 11:39
>ctg09_00003 dna:supercontig supercontig:GCA_000309565.1:ctg09_00003:1:899:1
AGTAAGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCT
TATGACCTGGGCTACACACGTGCTACAATGGATGGTACAACGAGTTGCGAGACCGCGAGG
TYAAGCTAATCTCTTAAAGCCATTCTCAGTTCGGACTGTAGGCTGCAACTCGCCTACACG
AAGTCGGAATCGCTAGTAATCGCGGATCAGCACGCCGCGGTGAATACGTTCCCGGGCCTT
GTACACACCGCCCGTCACACCATGAGAGTTTGTAAACCCGAAGCCGGTGGCGTAACCCT
TTTAGGGAGCGAGCCGTCTAAGGTGGGACAAATGATTAGGGTGAAGTCGTAACAAGGTAG
CCGTAGGAGAACCTGCGGCTGGATCACCTCCTTTCTAAGGAAACAGACTRAAAGTCTGAC
GGAAACCTGCACACACGAAACTTTGTTTAGTTTTGAGGGGATCACCTCAAGCACCTAA
CGGGTGCGACTTTGTTCTTTGAAAACCTGGATATCATTGTATTAATTGTTTTAAATTGCCG
AGAACACAGCGTATTTGTATGAGTTTCTGAAAAAGAAATTCGCATCGCATAACCGCTGAC
GCAAGTCAGTACAGGTTAAGTTRCAAAGGGCGCACGGTGGATGCCCTTGGCACTAGGAGCC
GATGAAGGACGGAACATAACCGATATGCTTCGGGGAGCTATAAGTAAGCTTTGATCCGG
AGATTTCCGAATGGGGGAACCCAGTACACATCAGTGTATTGCTTGTCAGTGAATACATAG
CTGGCCGGCGGCAGACGCGGGGAACTGAAACATCTCAGTACCCGCAGGAAGAGAAAGAAA
ACTCGATTCCCATAGTAGCGGCGAGCGAAGTGGGAAAAGCCCAAACCGAGAAGCTTGCT
>ctg03_00032 dna:supercontig supercontig:GCA_000309565.1:ctg03_00032:1:2417:1
CGACTGTAAACATCCGCACTGGTGCCGGCACCGGCTATGCATCCGTTGGTAGCTATACAC
TCGGTGAAAGTGTGATTTATGATCACGTGTATATCCGTGGCACATATGTTTGGGCGCGTT
ATCTCAGCTACTCAGGCAGGTATCACTATGTTGCGCTAGGCGTAAATGGTGGTGAGAGCT
ATGGTTCGCGCAGTTCAAATGCGCAAACCTATTCGCACACGTAACACAGTCCGCTCTG
GTGACAGCTTCTGGAGTATTGCCAGCAAGTATGGCATTAGCATGTACACGCTGGCTGCCA
ACAACGGAAAATCAATCTACAGCCTGATCTATCCGGGCGAAAGCCTGTATATTAGGTAAC
/AGTAAGACTGCCGGTGACAAA
```



①

W14-6-2: 文字列検索

確かにそこが最初にハイライトされるべきところがそうになっていることがわかる。他の箇所にも同じ文字列があれば「nキー」を押すとそこに飛ぶ。



```
AGTAAGACTGCCGGTGACAAAACCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCT
TATGACCTGGGCTACACACGTGCTACAATGGATGGTACAACGAGTTGCGAGACCGCGAGG
TYAAGCTAATCTCTTAAAGCCATTCTCAGTTCGGACTGTAGGCTGCAACTCGCCTACACG
AAGTCGGAATCGCTAGTAATCGCGGATCAGCACGCCGCGGTGAATACGTTCCCGGGCCTT
GTACACACCGCCCCTCACACCATGAGAGTTTGTAAACACCCGAAGCCGGTGGCGTAACCCT
TTTAGGGAGCGAGCCGTCTAAGGTGGGACAAATGATTAGGGTGAAGTCGTAACAAGGTAG
CCGTAGGAGAACCTGCGGCTGGATCACCTCCTTTCTAAGGAAACAGACTRAAAGTCTGAC
GGAAACCTGCACACACGAAACTTTGTTTAGTTTTGAGGGGATCACCTCAAGCACCTAA
CGGGTGCGACTTTGTTCTTTGAAAACCTGGATATCATTGTATTAATTGTTTTAAATTGCCG
AGAACACAGCGTATTTGTATGAGTTTCTGAAAAAGAAATTCGCATCGCATAACCGCTGAC
GCAAGTCAGTACAGGTTAAGTTRCAAAGGGCGCACGGTGGATGCCTTGGCACTAGGAGCC
GATGAAGGACGGAACATAACCGATATGCTTCGGGGAGCTATAAGTAAGCTTTGATCCGG
AGATTTCCGAATGGGGGAACCCAGTACACATCAGTGTATTGCTTGTGAGTGAATACATAG
CTGGCCGGCGGCAGACGCGGGGAACTGAAACATCTCAGTACCCGCAGGAAGAGAAAGAAA
ACTCGATTCCCATAGTAGCGGCGAGCGAAGTGGGAAAAGCCCAAACCGAGAAGCTTGCT
>ctg03_00032 dna:supercontig supercontig:GCA_000309565.1:ctg03_00032:1:2417:1
CGACTGTAAACATCCGCACTGGTGCCGGCACCGGCTATGCATCCGTTGGTAGCTATACAC
TCGGTGAAAGTGTGATTTATGATCACGTGTATATCCGTGGCACATATGTTTGGGCGCGTT
ATCTCAGCTACTCAGGCAGGTATCACTATGTTGCGCTAGGCGTAAATGGTGGTGAGAGCT
ATGGTTCGCGCAGTTCAAATGCGCAAACCTATTCGCACACGTAACACAGTCCGCTCTG
GTGACAGCTTCTGGAGTATTGCCAGCAAGTATGGCATTAGCATGTACACGCTGGCTGCCA
ACAACGGAAAATCAATCTACAGCCTGATCTATCCGGGCGAAAGCCTGTATATTAGGTAAC
AAACAGCCCTCTGCTCGCTAACGCGGGTGGAGGACTTTTTTAGAACCAAGTAAATTTCTC
:
```

「nキー」を1回押して、次の一致箇所を眺めているところ。

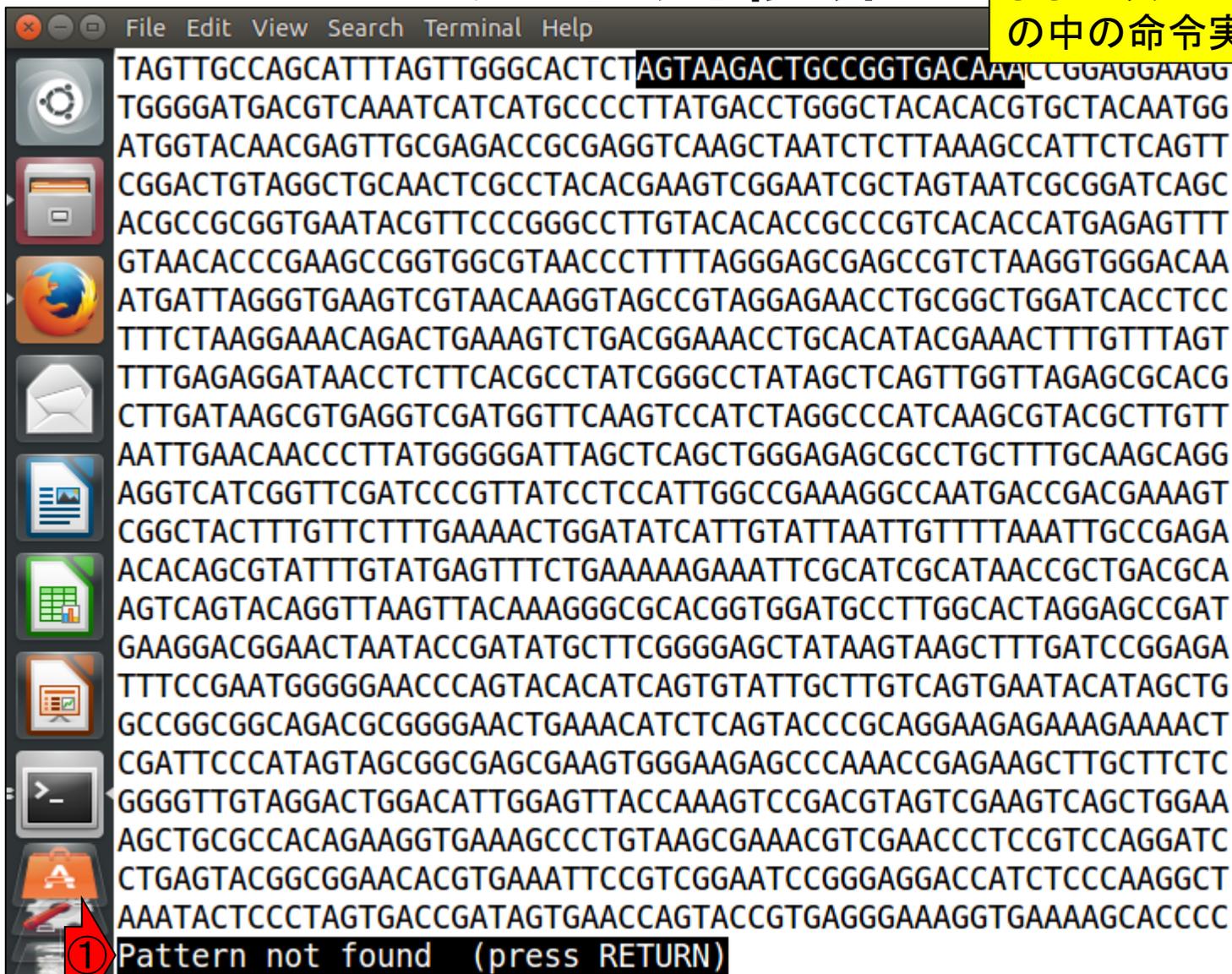
W14-6-2: 文字列検索



```
File Edit View Search Terminal Help 11:43
TAGTTGCCAGCATTTAGTTGGGCACTCTAGTAAGACTGCCGGTGACAAA CCGGAGGAAGG
TGGGGATGACGTCAAATCATCATGCCCTTATGACCTGGGCTACACACGTGCTACAATGG
ATGGTACAACGAGTTGCGAGACCGCGAGGTCAAGCTAATCTCTTAAAGCCATTCTCAGTT
CGGACTGTAGGCTGCAACTCGCCTACACGAAGTCGGAATCGCTAGTAATCGCGGATCAGC
ACGCCGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGAGAGTTT
GTAACACCCGAAGCCGGTGGCGTAACCCTTTTAGGGAGCGAGCCGTCTAAGGTGGGACAA
ATGATTAGGGTGAAGTCGTAACAAGGTAGCCGTAGGAGAACCTGCGGCTGGATCACCTCC
TTTCTAAGGAAACAGACTGAAAGTCTGACGGAAACCTGCACATACGAAACTTTGTTTAGT
TTTGAGAGGATAACCTCTTCACGCCTATCGGGCCTATAGCTCAGTTGGTTAGAGCGCAGC
CTTGATAAGCGTGAGGTCGATGGTTCAAGTCCATCTAGGCCATCAAGCGTACGCTTGTT
AATTGAACAACCCTTATGGGGGATTAGCTCAGCTGGGAGAGCGCCTGCTTTGCAAGCAGG
AGGTCATCGGTTTCGATCCCGTTATCCTCCATTGGCCGAAAGGCCAATGACCGACGAAAGT
CGGCTACTTTGTTCTTTGAAAACCTGGATATCATTGTATTAATTGTTTTAAATTGCCGAGA
ACACAGCGTATTTGTATGAGTTTCTGAAAAAGAAATTCGCATCGCATAACCGCTGACGCA
AGTCAGTACAGGTTAAGTTACAAAGGGCGCACGGTGGATGCCTTGGCACTAGGAGCCGAT
GAAGGACGGAACATAACCGATATGCTTCGGGGAGCTATAAGTAAGCTTTGATCCGGAGA
TTTCCGAATGGGGGAACCCAGTACACATCAGTGTATTGCTTGTGAGTGAATACATAGCTG
GCCGGCGGCAGACGCGGGGAACTGAAACATCTCAGTACCCGCAGGAAGAGAAAGAAAAC
CGATTC CATAGTAGCGGCGAGCGAAGTGGGAAGAGCCCAAACCGAGAAGCTTGCTTCTC
GGGGTTGTAGGACTGGACATTGGAGTTACCAAAGTCCGACGTAGTCGAAGTCAGCTGGAA
AGCTGCGCCACAGAAGGTGAAAGCCCTGTAAGCGAAACGTCGAACCCCTCCGTCCAGGATC
CTGAGTACGGCGGAACACGTGAAATTCCGTGGAATCCGGGAGGACCATCTCCCAAGGCT
AAATACTCCCTAGTGACCGATAGTGAACCAGTACCGTGAGGGAAAGGTGAAAAGCACCCC
:
```

W14-6-2: 文字列検索

もう1回「nキー」を押すと、①「一致箇所はない」というメッセージが出る。いわれるがままにリターンキーを押すと、lessコマンドの中の命令実行待ち状態(:)になる。



```
File Edit View Search Terminal Help
TAGTTGCCAGCATTAGTTGGGCACTCTAGTAAGACTGCCGGTGACAACCGGAGGAAGG
TGGGGATGACGTCAAATCATCATGCCCTTATGACCTGGGCTACACACGTGCTACAATGG
ATGGTACAACGAGTTGCGAGACCGCGAGGTCAAGCTAATCTCTTAAAGCCATTCTCAGTT
CGGACTGTAGGCTGCAACTCGCCTACACGAAGTCGGAATCGCTAGTAATCGCGGATCAGC
ACGCCGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGAGAGTTT
GTAACACCCGAAGCCGGTGGCGTAACCCTTTTAGGGAGCGAGCCGTCTAAGGTGGGACAA
ATGATTAGGGTGAAGTCGTAACAAGGTAGCCGTAGGAGAACCTGCGGCTGGATCACCTCC
TTTCTAAGGAAACAGACTGAAAGTCTGACGGAAACCTGCACATACGAAACTTTGTTTAGT
TTTGAGAGGATAACCTCTTCACGCCTATCGGGCCTATAGCTCAGTTGGTTAGAGCGCAGC
CTTGATAAGCGTGAGGTCGATGGTTCAAGTCCATCTAGGCCATCAAGCGTACGCTTGTT
AATTGAACAACCCTTATGGGGGATTAGCTCAGCTGGGAGAGCGCCTGCTTTGCAAGCAGG
AGGTCATCGGTTTCGATCCCGTTATCCTCCATTGGCCGAAAGGCCAATGACCGACGAAAGT
CGGCTACTTTGTTCTTTGAAAACCTGGATATCATTGTATTAATTGTTTTAAATTGCCGAGA
ACACAGCGTATTTGTATGAGTTTCTGAAAAAGAAATTCGCATCGCATAACCGCTGACGCA
AGTCAGTACAGGTTAAGTTACAAAGGGCGCACGGTGGATGCCTTGGCACTAGGAGCCGAT
GAAGGACGGAACATAACCGATATGCTTCGGGGAGCTATAAGTAAGCTTTGATCCGGAGA
TTTCCGAATGGGGGAACCCAGTACACATCAGTGTATTGCTTGTGAGTGAATACATAGCTG
GCCGGCGGCAGACGCGGGGAACTGAAACATCTCAGTACCCGCAGGAAGAGAAAGAAAAC
CGATTCCCATAGTAGCGGCGAGCGAAGTGGGAAGAGCCCAAACCGAGAAGCTTGCTTCTC
GGGGTTGTAGGACTGGACATTGGAGTTACCAAAGTCCGACGTAGTCGAAGTCAGCTGGAA
AGCTGCGCCACAGAAGGTGAAAGCCCTGTAAGCGAAACGTCGAACCCCTCCGTCCAGGATC
CTGAGTACGGCGGAACACGTGAAATTCCGTGGAATCCGGGAGGACCATCTCCCAAGGCT
AAATACTCCCTAGTGACCGATAGTGAACCAGTACCGTGAGGGAAAGGTGAAAAGCACCCC
① Pattern not found (press RETURN)
```

W14-6-2: 文字列検索

いわれるがままにリターンキーを押して、
①lessコマンドの中の命令実行待ち状態(:)
になったところ。



```
File Edit View Search Terminal Help
TAGTTGCCAGCATTTAGTTGGGCACTCTAGTAAGACTGCCGGTGACAAA CCGGAGGAAGG
TGGGGATGACGTCAAATCATCATGCCCTTATGACCTGGGCTACACACGTGCTACAATGG
ATGGTACAACGAGTTGCGAGACCGCGAGGTCAAGCTAATCTCTTAAAGCCATTCTCAGTT
CGGACTGTAGGCTGCAACTCGCCTACACGAAGTCGGAATCGCTAGTAATCGCGGATCAGC
ACGCCGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGAGAGTTT
GTAACACCCGAAGCCGGTGGCGTAACCCTTTTAGGGAGCGAGCCGTCTAAGGTGGGACAA
ATGATTAGGGTGAAGTCGTAACAAGGTAGCCGTAGGAGAACCTGCGGCTGGATCACCTCC
TTTCTAAGGAAACAGACTGAAAGTCTGACGGAAACCTGCACATACGAAACTTTGTTTAGT
TTTGAGAGGATAACCTCTTCACGCCTATCGGGCCTATAGCTCAGTTGGTTAGAGCGCAGC
CTTGATAAGCGTGAGGTCGATGGTTCAAGTCCATCTAGGCCATCAAGCGTACGCTTGTT
AATTGAACAACCCTTATGGGGGATTAGCTCAGCTGGGAGAGCGCCTGCTTTGCAAGCAGG
AGGTCATCGGTTTCGATCCCGTTATCCTCCATTGGCCGAAAGGCCAATGACCGACGAAAGT
CGGCTACTTTGTTCTTTGAAAACGGATATCATTGTATTAATTGTTTTAAATTGCCGAGA
ACACAGCGTATTTGTATGAGTTTCTGAAAAAGAAATTCGCATCGCATAACCGCTGACGCA
AGTCAGTACAGGTTAAGTTACAAAGGGCGCACGGTGGATGCCTTGGCACTAGGAGCCGAT
GAAGGACGGAACATAACCGATATGCTTCGGGGAGCTATAAGTAAGCTTTGATCCGGAGA
TTTCCGAATGGGGGAACCCAGTACACATCAGTGTATTGCTTGTGAGTGAATACATAGCTG
GCCGGCGGCAGACGCGGGGAACTGAAACATCTCAGTACCCGCAGGAAGAGAAAGAAAAC
CGATTC CATAGTAGCGGCGAGCGAAGTGGGAAGAGCCCAAACCGAGAAGCTTGCTTCTC
GGGGTTGTAGGACTGGACATTGGAGTTACCAAAGTCCGACGTAGTCGAAGTCAGCTGGAA
AGCTGCGCCACAGAAGGTGAAAGCCCTGTAAGCGAAACGTCGAACCCCTCCGTCCAGGATC
CTGAGTACGGCGGAACACGTGAAATTCCGTGGAATCCGGGAGGACCATCTCCAAGGCT
AAATACTCCCTAGTGACCGATAGTGAACCAGTACCGTGAGGGAAAGGTGAAAAGCACCCC
:
```

W14-6-3: 行またぎは.

lessコマンドの中の命令実行待ち状態(:)で、「gキー」を押して先頭ページに移動した状態(全く同じ状態にしなくてもよい)。次は、行をまたがる文字列検索がうまくいかない例を紹介。

```
File Edit View Search Terminal Help
>ctg09_00003 dna:supercontig supercontig:GCA_000309565.1:ctg09_00003:1:899:1
AGTAAGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCT
TATGACTCTGGGCTACACACGTGCTACAATGGATGGTACAACGAGTTGCGAGACCGCGAGG
TYAAGCTAATCTCTTAAAGCCATTCTCAGTTCGGACTGTAGGCTGCAACTCGCCTACACG
AAGTCGGAATCGCTAGTAATCGCGGATCAGCACGCCGCGGTGAATACGTTCCCGGGCCTT
GTACACACCGCCCGTCACACCATGAGAGTTTGTAAACACCCGAAGCCGGTGGCGTAACCCT
TTTAGGGAGCGAGCCGTCTAAGGTGGGACAAATGATTAGGGTGAAGTCGTAACAAGGTAG
CCGTAGGAGAACCTGCGGCTGGATCACCTCCTTTCTAAGGAAACAGACTRAAAGTCTGAC
GGAAACCTGCACACACGAACTTTGTTTAGTTTTGAGGGGATCACCTCAAGCACCTAA
CGGGTGCAGCTTTGTTCTTTGAAACTGGATATCATTGTATTAATTGTTTTAAATTGCCG
AGAACACAGCGTATTTGTATGAGTTTCTGAAAAAGAAATTCGCATCGCATAACCGCTGAC
GCAAGTCAGTACAGGTTAAGTTRCAAAGGGCGCACGGTGGATGCCTTGGCACTAGGAGCC
GATGAAGGACGGAACATAACCGATATGCTTCGGGGAGCTATAAGTAAGCTTTGATCCGG
AGATTTCCGAATGGGGGAACCCAGTACACATCAGTGTATTGCTTGTGAGTGAATACATAG
CTGGCCGGCGGCAGACGCGGGGAACTGAAACATCTCAGTACCCGCAGGAAGAGAAAGAAA
ACTCGATTCCCATAGTAGCGGCGAGCGAAGTGGGAAAAGCCCAAACCGAGAAGCTTGCT
>ctg03_00032 dna:supercontig supercontig:GCA_000309565.1:ctg03_00032:1:2417:1
CGACTGTAAACATCCGCACTGGTGCCGGCACCGGCTATGCATCCGTTGGTAGCTATACAC
TCGGTGAAAGTGTGATTTATGATCACGTGTATATCCGTGGCACATATGTTTGGGCGCGTT
ATCTCAGCTACTCAGGCAGGTATCACTATGTTGCGCTAGGCGTAAATGGTGGTGAGAGCT
ATGGTTCGCGCAGTTCAAATGCGCAAACCTATTCGCACACGTAACACAGTCCGCTCTG
GTGACAGCTTCTGGAGTATTGCCAGCAAGTATGGCATTAGCATGTACACGCTGGCTGCCA
ACAACGGAAAATCAATCTACAGCCTGATCTATCCGGGCGAAAGCCTGTATATTAGGTAAC
:
```

W14-6-3: 行またぎは...

```

File Edit View Search Terminal Help
>ctg09_00003 dna:supercontig supercontig:GCA_000309565.1:ctg09_00003:1:899:1
AGTAAGACTGCCGGTGACAAAACCGGAGGAAGGTGGGGATGACGTCAAATCATCATGTGCCCT
TATGACCTGGGCTACACACGTGCTACAATGGATGGTACAACGAGTTGCGAGACCGCGAGG
TYAAGCTAATCTCTTAAAGCCATTCTCAGTTCGGACTGTAGGCTGCAACTCGCCTACACG
AAGTCGGAATCGCTAGTAATCGCGGATCAGCACGCCGCGGTGAATACGTTCCCGGGCCTT
GTACACACCGCCCGTCACACCATGAGAGTTTGTAAACACCCGAAGCCGGTGGCGTAACCCT
TTTAGGGAGCGAGCCGTCTAAGGTGGGACAAATGATTAGGGTGAAGTCGTAACAAGGTAG
CCGTAGGAGAACCTGCGGCTGGATCACCTCCTTTCTAAGGAAACAGACTRAAAGTCTGAC
GGAAACCTGCACACACGAACTTTGTTTAGTTTTGAGGGGATCACCTCAAGCACCTAA
CGGGTGC GACTTTGTTCTTTGAAAACCTGGATATCATTGTATTAATTGTTTTAAATTGCCG
AGAACACAGCGTATTTGTATGAGTTTCTGAAAAAGAAATTCGCATCGCATAACCGCTGAC
GCAAGTCAGTACAGGTTAAGTTRCAAAGGGCGCACGGTGGATGCCTTGGCACTAGGAGCC
GATGAAGGACGGAACATAACCGATATGCTTCGGGGAGCTATAAGTAAGCTTTGATCCGG
AGATTTCCGAATGGGGGAACCCAGTACACATCAGTGTATTGCTTGTGAGTGAATACATAG
CTGGCCGGCGGCAGACGCGGGGAACTGAAACATCTCAGTACCCGCAGGAAGAGAAAGAAA
ACTCGATTCCCATAGTAGCGGCGAGCGAAGTGGGAAAAGCCCAAACCGAGAAGCTTGCT
>ctg03_00032 dna:supercontig supercontig:GCA_000309565.1:ctg03_00032:1:2417:1
CGACTGTAAACATCCGCACTGGTGCCGGCACCGGCTATGCATCCGTTGGTAGCTATACAC
TCGGTGAAAGTGTGATTTATGATCACGTGTATATCCGTGGCACATATGTTTGGGCGCGTT
ATCTCAGCTACTCAGGCAGGTATCACTATGTTGCGCTAGGCGTAAATGGTGGTGAGAGCT
ATGGTTCGCGCAGTTCAAATGCGCAAACCTATTCGCACACGTA CTACACAGTCCGCTCTG
GTGACAGCTTCTGGAGTATTGCCAGCAAGTATGGCATTAGCATGTACACGCTGGCTGCCA
ACAACGGAAAATCAATCTACAGCCTGATCTATCCGGGCGAAAGCCTGTATATTAGGTAAC
/TGCCCTTATGAC
  
```



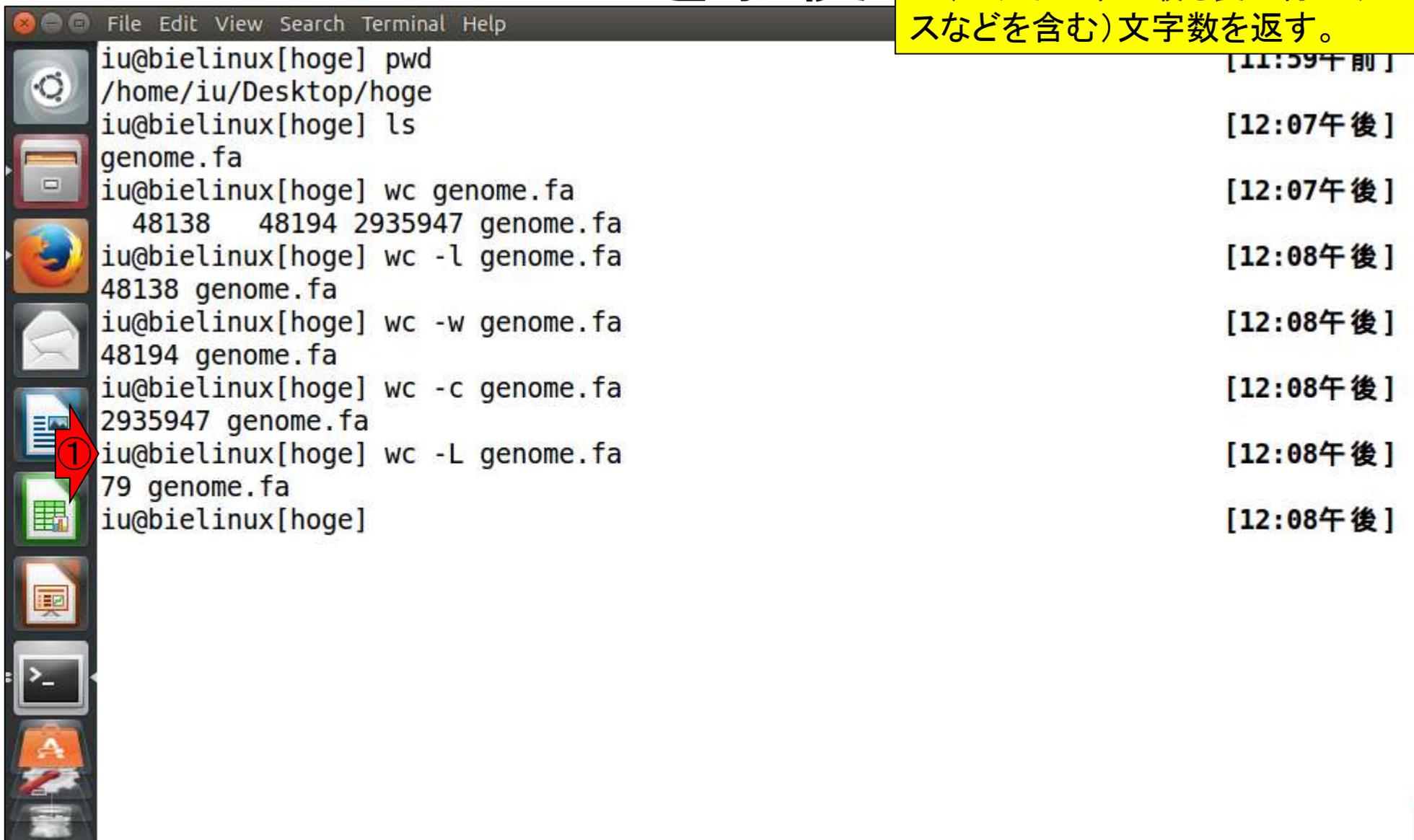
W14-6-3: 行またぎは...

確かに行をまたがる文字列検索がうまくいかないことがわかります。「リターンキー」を押した後に、「qキー」を押すと、コマンド入力待ち状態に戻る。

```
File Edit View Search Terminal Help
>ctg09_00003 dna:supercontig supercontig:GCA_000309565.1:ctg09_00003:1:899:1
AGTAAGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCT
TATGACTGGGCTACACACGTGCTACAATGGATGGTACAACGAGTTGCGAGACCGCGAGG
TYAAGCTAATCTCTTAAAGCCATTCTCAGTTCGGACTGTAGGCTGCAACTCGCCTACACG
AAGTCGGAATCGCTAGTAATCGCGGATCAGCACGCCGCGGTGAATACGTTCCCGGGCCTT
GTACACACCGCCCGTCACACCATGAGAGTTTGTAAACACCCGAAGCCGGTGGCGTAACCCT
TTTAGGGAGCGAGCCGTCTAAGGTGGGACAAATGATTAGGGTGAAGTCGTAACAAGGTAG
CCGTAGGAGAACCTGCGGCTGGATCACCTCCTTTCTAAGGAAACAGACTRAAAGTCTGAC
GGAAACCTGCACACACGAACTTTGTTTAGTTTTGAGGGGATCACCTCAAGCACCTAA
CGGGTGCAGCTTTGTTCTTTGAAAACCTGGATATCATTGTATTAATTGTTTTAAATTGCCG
AGAACACAGCGTATTTGTATGAGTTTCTGAAAAAGAAATTCGCATCGCATAACCGCTGAC
GCAAGTCAGTACAGGTTAAGTTRCAAAGGGCGCACGGTGGATGCCTTGGCACTAGGAGCC
GATGAAGGACGGAACATAACCGATATGCTTCGGGGAGCTATAAGTAAGCTTTGATCCGG
AGATTTCCGAATGGGGGAACCCAGTACACATCAGTGTATTGCTTGTGAGTGAATACATAG
CTGGCCGGCGGCAGACGCGGGGAACTGAAACATCTCAGTACCCGCAGGAAGAGAAAGAAA
ACTCGATTCCCATAGTAGCGGCGAGCGAAGTGGGAAAAGCCCAAACCGAGAAGCTTGCT
>ctg03_00032 dna:supercontig supercontig:GCA_000309565.1:ctg03_00032:1:2417:1
CGACTGTAAACATCCGCACTGGTGCCGGCACCGGCTATGCATCCGTTGGTAGCTATACAC
TCGGTGAAAGTGTGATTTATGATCACGTGTATATCCGTGGCACATATGTTTGGGCGCGTT
ATCTCAGCTACTCAGGCAGGTATCACTATGTTGCGCTAGGCGTAAATGGTGGTGAGAGCT
ATGGTTCGCGCAGTTCAAATGCGCAAACCTATTCGCACACGTAACACAGTCCGCTCTG
GTGACAGCTTCTGGAGTATTGCCAGCAAGTATGGCATTAGCATGTACACGCTGGCTGCCA
ACAACGGAAAATCAATCTACAGCCTGATCTATCCGGGCGAAAGCCTGTATATTAGGTAAC
① Pattern not found (press RETURN)
```

W15-1: オプションを駆使

オプションを駆使することで様々な結果を得ることができる。①「-L」オプションは、ファイル中で最も長い行の(スペースなどを含む)文字数を返す。

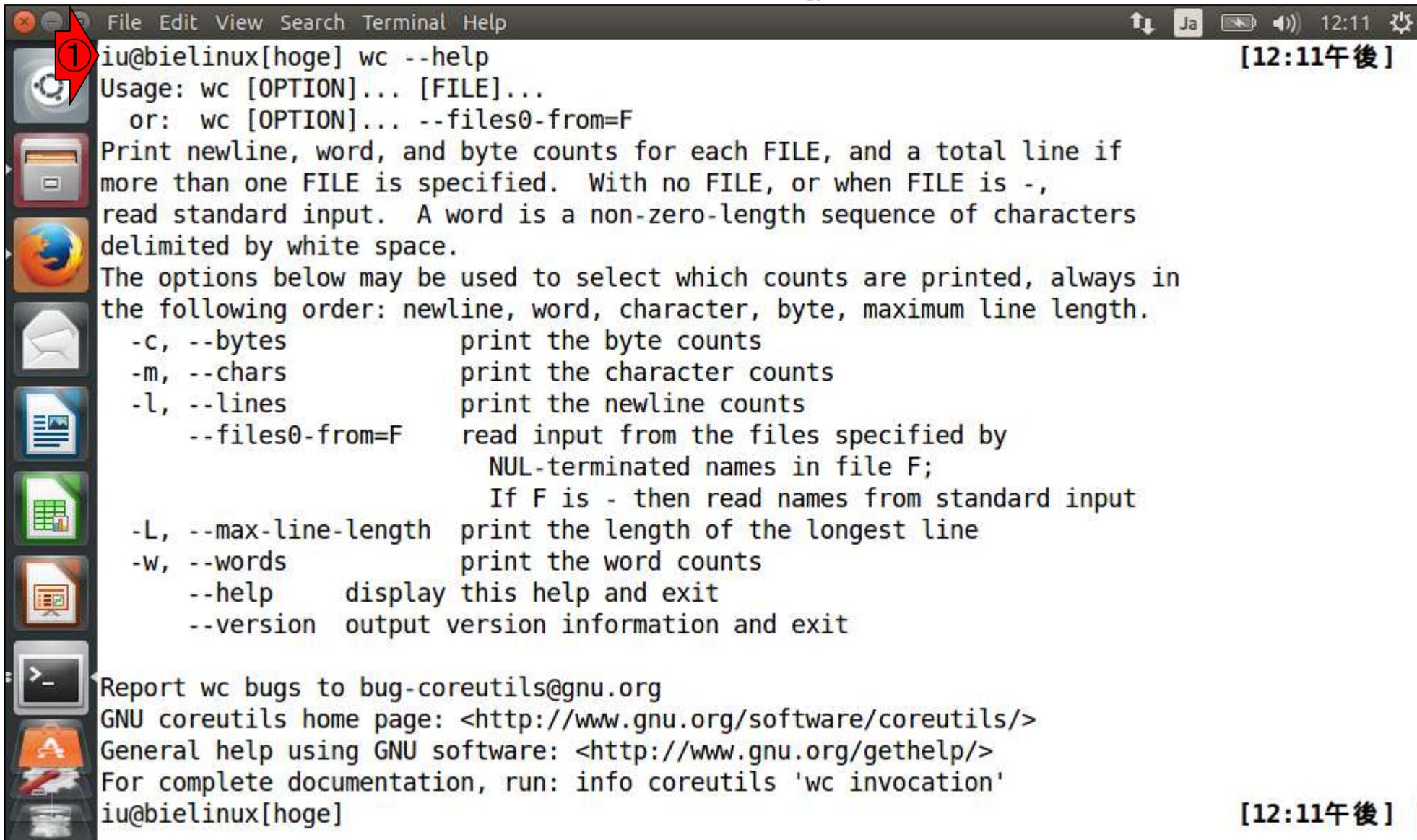


```
iu@bielinux[hoge] pwd
/home/iu/Desktop/hoge
iu@bielinux[hoge] ls
genome.fa
iu@bielinux[hoge] wc genome.fa
 48138  48194 2935947 genome.fa
iu@bielinux[hoge] wc -l genome.fa
48138 genome.fa
iu@bielinux[hoge] wc -w genome.fa
48194 genome.fa
iu@bielinux[hoge] wc -c genome.fa
2935947 genome.fa
① iu@bielinux[hoge] wc -L genome.fa
79 genome.fa
iu@bielinux[hoge]
```

[11:59午前]
[12:07午後]
[12:07午後]
[12:08午後]
[12:08午後]
[12:08午後]
[12:08午後]
[12:08午後]

W15-2: オプションを駆使

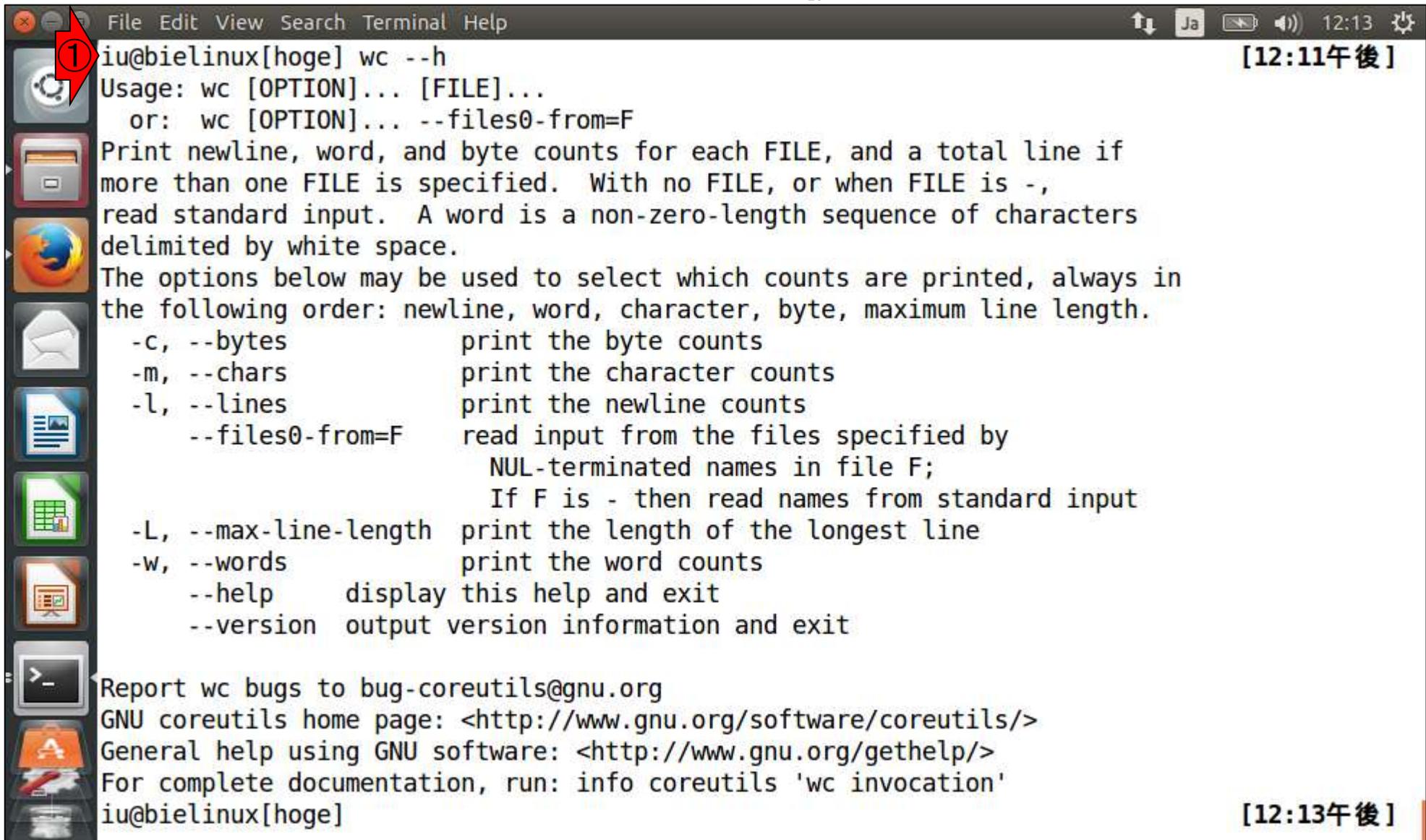
①「wc --help」で、利用可能なオプションやその説明を見ることができる。



```
iu@bielinux[hoge] wc --help
Usage: wc [OPTION]... [FILE]...
  or: wc [OPTION]... --files0-from=F
Print newline, word, and byte counts for each FILE, and a total line if
more than one FILE is specified.  With no FILE, or when FILE is -,
read standard input.  A word is a non-zero-length sequence of characters
delimited by white space.
The options below may be used to select which counts are printed, always in
the following order: newline, word, character, byte, maximum line length.
  -c, --bytes          print the byte counts
  -m, --chars          print the character counts
  -l, --lines          print the newline counts
  --files0-from=F     read input from the files specified by
                      NUL-terminated names in file F;
                      If F is - then read names from standard input
  -L, --max-line-length print the length of the longest line
  -w, --words          print the word counts
  --help              display this help and exit
  --version           output version information and exit

Report wc bugs to bug-coreutils@gnu.org
GNU coreutils home page: <http://www.gnu.org/software/coreutils/>
General help using GNU software: <http://www.gnu.org/gethelp/>
For complete documentation, run: info coreutils 'wc invocation'
iu@bielinux[hoge]
```

W15-2: オプションを駆使



```
File Edit View Search Terminal Help
iu@bielinux[hoge] wc --h
Usage: wc [OPTION]... [FILE]...
  or: wc [OPTION]... --files0-from=F
Print newline, word, and byte counts for each FILE, and a total line if
more than one FILE is specified.  With no FILE, or when FILE is -,
read standard input.  A word is a non-zero-length sequence of characters
delimited by white space.
The options below may be used to select which counts are printed, always in
the following order: newline, word, character, byte, maximum line length.
-c, --bytes          print the byte counts
-m, --chars          print the character counts
-l, --lines          print the newline counts
  --files0-from=F    read input from the files specified by
                    NUL-terminated names in file F;
                    If F is - then read names from standard input
-L, --max-line-length print the length of the longest line
-w, --words          print the word counts
--help              display this help and exit
--version           output version information and exit

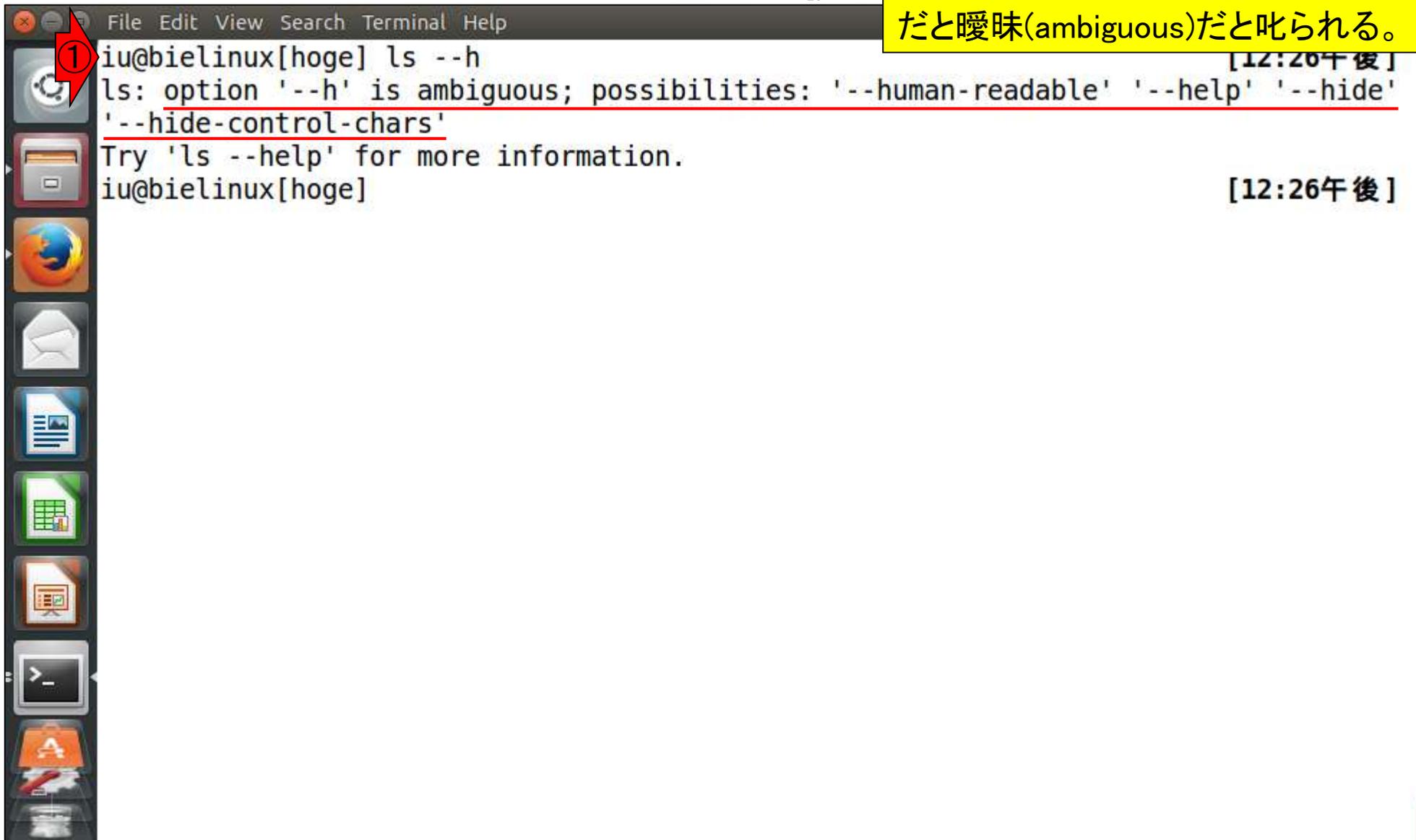
Report wc bugs to bug-coreutils@gnu.org
GNU coreutils home page: <http://www.gnu.org/software/coreutils/>
General help using GNU software: <http://www.gnu.org/gethelp/>
For complete documentation, run: info coreutils 'wc invocation'
iu@bielinux[hoge]
```

[12:11午後]

[12:13午後]

W15-2: オプションを駆使

①lsコマンドは「--h」はだめ。理由は赤下線にあるように「--h」から始まる複数のオプションが存在するので、--hだけだと曖昧(ambiguous)だと叱られる。



```
File Edit View Search Terminal Help
iu@bielinux[hoge] ls --h [12:26午後]
ls: option '--h' is ambiguous; possibilities: '--human-readable' '--help' '--hide'
'--hide-control-chars'
Try 'ls --help' for more information.
iu@bielinux[hoge] [12:26午後]
```

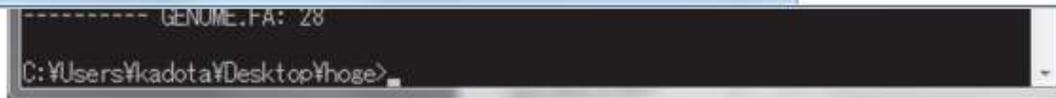
W15-3: man

「man wc」実行直後の状態。詳細なマニュアルを見ることができる。使い方はmoreやlessコマンドと同じ。マニュアル表示画面からの脱出は「qキー」。

```
WC(1) User Commands WC(1)
NAME
wc - print newline, word, and byte counts for each file
SYNOPSIS
wc [OPTION]... [FILE]...
wc [OPTION]... --files0-from=F
DESCRIPTION
Print newline, word, and byte counts for each FILE, and a total line if
more than one FILE is specified. With no FILE, or when FILE is -, read
standard input. A word is a non-zero-length sequence of characters
delimited by white space. The options below may be used to select which
counts are printed, always in the following order: newline, word, charac-
ter, byte, maximum line length.
-c, --bytes
    print the byte counts
-m, --chars
    print the character counts
Manual page wc(1) line 1 (press h for help or q to quit)
```

W16-1: "" に注意!

PDFファイル中のコマンドをコピーして実行しようとしてもうまくいかない場合がある。これは連載第2回のPDFファイル。反転部分で説明。



```
----- GENOME.FA: 28
C:\Users\kadota\Desktop\fhoge>
```

図5. コマンドプロンプトでの作業例。① cd コマンドを用いた作業ディレクトリの変更、② rename コマンドを用いたファイル名の変更、③ find コマンドを用いた">"を含む行数のカウント。

れるファイル形式である。grep は、文字列検索コマンドである。FASTA 形式ファイルである genome.fa を読み込んで、">" から始まる description 行を抽出し、出力を行数表示にする -c オプションを与えているため、結果が 28 (行) と返される。行数表示の -c オプションをつけなかった場合、つまり「grep ">" genome.fa」を実行すると、description 情報がそのまま表示される [ウェブ資料 21]。grep は応用範囲の極めて広い非常に強力なコマンドである。例えば、アノテーションファイルから特定の染色体名を含む行のみを抽出したり、grep 出力結果をパイプでつないで他のコマンドで処理する作業が実際に行われる。また最近では、RNA-seq データからのキメラ転写物の同定に用いられるなど、grep コマンドの NGS 解析での有効性も報告されている⁷⁾。もちろんここで示した程度のことであれば、Win のコマンドプロンプト上でも実現可能である (図 5; ウェブ資料 22-24)。例えば、Linux の grep に相当する Win のコマンドは find である。しかし、Linux にも find コマンドは存在し、(文字列検索ではなく) ファイル

ことを「grep する」といい、通常「find する」はファイルやディレクトリ検索のことを指す。つまり、一般にコマンド名で会話が成立するのは Linux コマンドのみである。

Bio-Linux の導入

連載第 2 回は、Win のコマンドプロンプトおよび Mac のターミナルという通常利用 PC 環境を用いて、GUI 環境とコマンドライン環境の見栄えの違い、および Win と Linux の基本コマンドを紹介した。Mac は Linux コマンドをすぐに使えるため、Linux 入門としてはおすすめである。しかし連載第 1 回でも述べたように、仮想ソフトをインストールして、様々な NGS 解析用プログラムが一通り組み込まれた Bio-Linux⁸⁾ を導入すれば、Win と Mac というホスト OS の違いによらず同じ Linux 環境で解析が可能である。2014 年 7 月末にリリースされた Bio-Linux 8 が最新版である (2014 年 10 月 15 日調べ)。連載第 3 回は、Bio-Linux 8 (ゲスト OS) 環境での解説を行う予定である。

http://www.iu.a.u-tokyo.ac.jp/~kadota/JSLAB_2_kadota.pdf

W16-1: "" に注意!

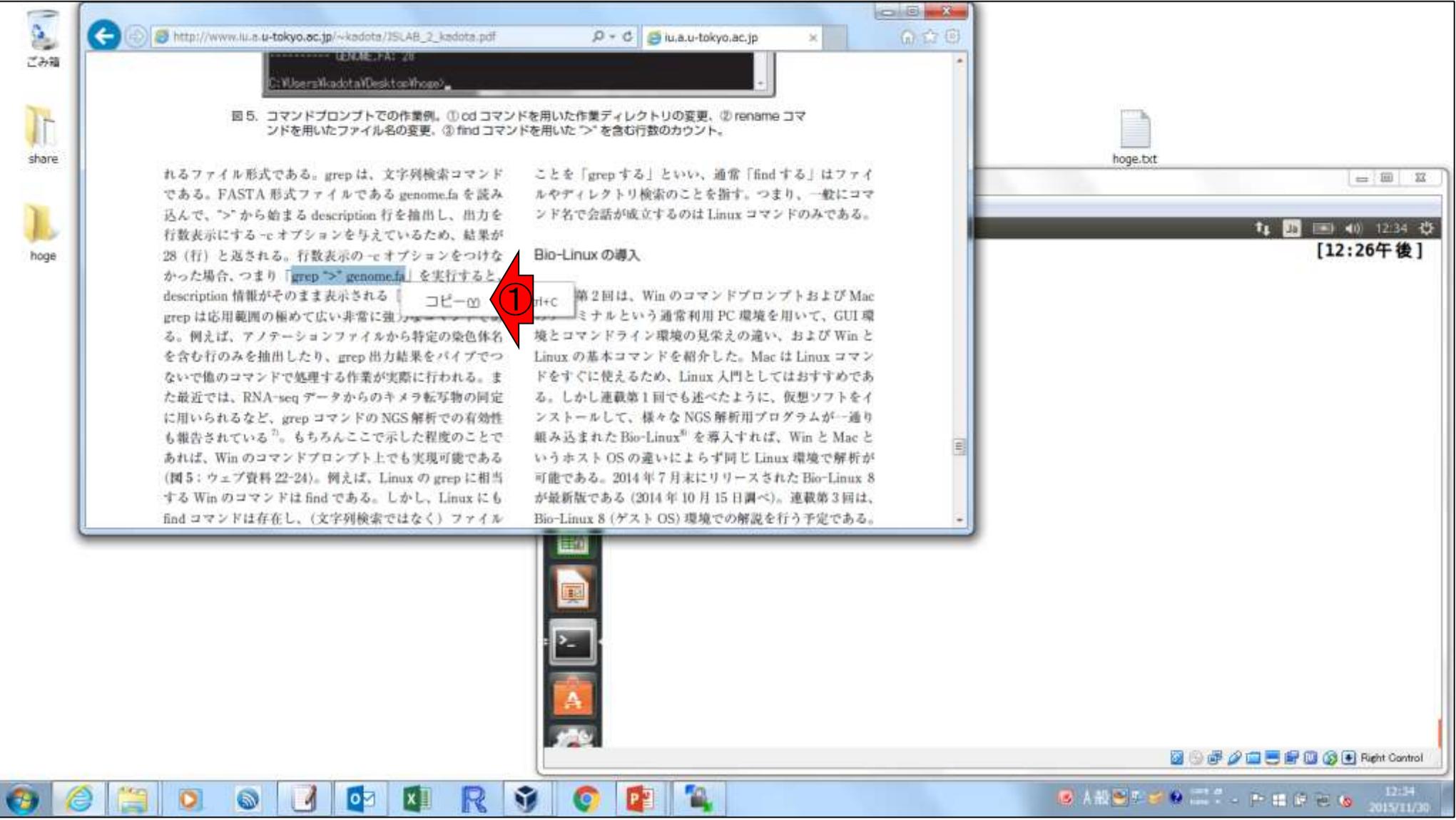


図5. コマンドプロンプトでの作業例。① cd コマンドを用いた作業ディレクトリの変更、② rename コマンドを用いたファイル名の変更、③ find コマンドを用いた ">" を含む行数のカウント。

れるファイル形式である。grep は、文字列検索コマンドである。FASTA 形式ファイルである genome.fa を読み込んで、">" から始まる description 行を抽出し、出力を行数表示にする -c オプションを与えているため、結果が 28 (行) と返される。行数表示の -c オプションをつけなかった場合、つまり `grep > genome.fa` を実行すると、description 情報がそのまま表示される。grep は応用範囲の極めて広い非常に強力な検索ツールである。例えば、アノテーションファイルから特定の染色体名を含む行のみを抽出したり、grep 出力結果をパイプでつないで他のコマンドで処理する作業が実際に行われる。また最近では、RNA-seq データからのキメラ転写物の同定に用いられるなど、grep コマンドの NGS 解析での有効性も報告されている²⁾。もちろんここで示した程度のことであれば、Win のコマンドプロンプト上でも実現可能である (図5: ウェブ資料 22-24)。例えば、Linux の grep に相当する Win のコマンドは find である。しかし、Linux にも find コマンドは存在し、(文字列検索ではなく) ファイル

ことを「grep する」といい、通常「find する」はファイルやディレクトリ検索のことを指す。つまり、一般にコマンド名で会話が成立するのは Linux コマンドのみである。

Bio-Linux の導入

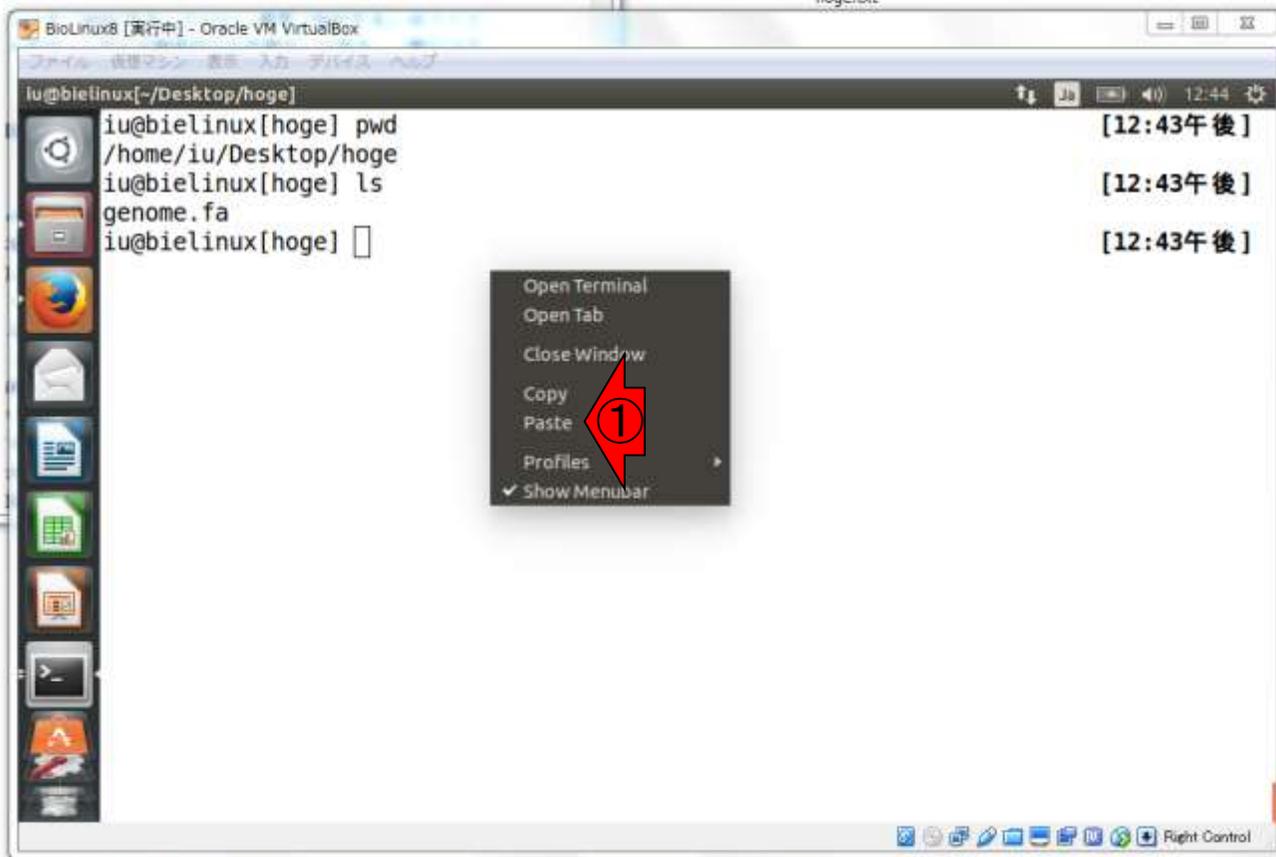
第2回は、Win のコマンドプロンプトおよび Mac ミナルという通常利用 PC 環境を用いて、GUI 環境とコマンドライン環境の見栄えの違い、および Win と Linux の基本コマンドを紹介した。Mac は Linux コマンドをすぐに使えるため、Linux 入門としてはおすすめである。しかし連載第1回でも述べたように、仮想ソフトをインストールして、様々な NGS 解析用プログラムが一通り組み込まれた Bio-Linux⁴⁾ を導入すれば、Win と Mac というホスト OS の違いによらず同じ Linux 環境で解析が可能である。2014 年 7 月末にリリースされた Bio-Linux 8 が最新版である (2014 年 10 月 15 日調べ)。連載第3回は、Bio-Linux 8 (ゲスト OS) 環境での解説を行う予定である。

W16-1: "" に注意!

①右クリックでペースト。手打ちはうまくいってしまうからダメ!

図5. コマンドプロンプトでの作業例。① cd コマンドを用いた作業ディレクトリの変更、② rename コマンドを用いたファイル名の変更、③ find コマンドを用いた ">" を含む行数のカウント。

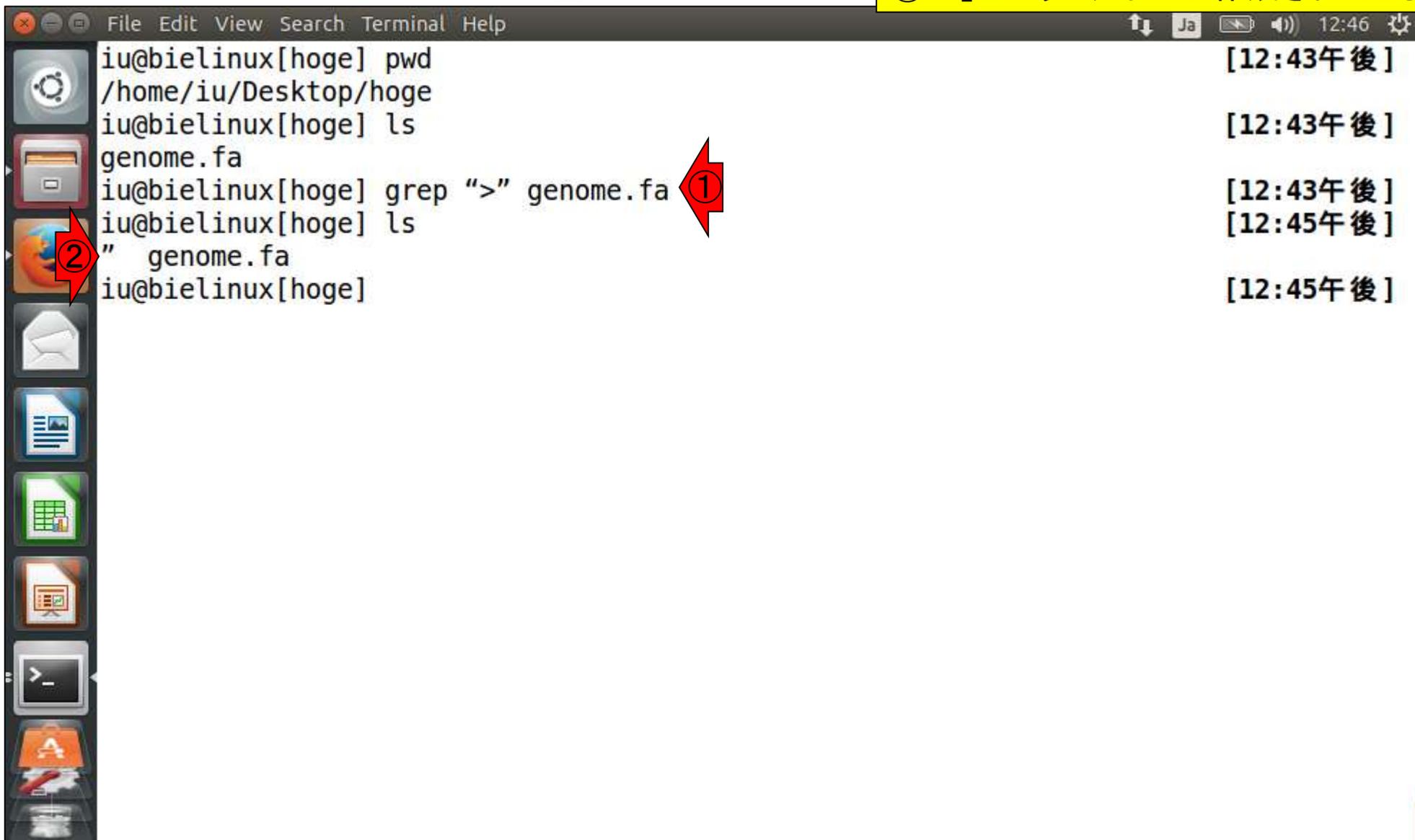
れるファイル形式である。grep は、文字列検索コマンドである。FASTA 形式ファイルである genome.fa を読み込んで、">" から始まる description 行を抽出し、出力を行数表示にする -c オプションを与えているため、結果が 28 (行) と返される。行数表示の -c オプションをつけなかった場合、つまり `grep > genome.fa` を実行すると、description 情報がそのまま表示される [ウェブ資料 21]。grep は応用範囲の極めて広い非常に強力なコマンドである。例えば、アノテーションファイルから特定の染色体名を含む行のみを抽出したり、grep 出力結果をパイプでつないで他のコマンドで処理する作業が実際に行われる。また最近では、RNA-seq データからのキメラ転写物の同定に用いられるなど、grep コマンドの NGS 解析での有効性も報告されている²⁾。もちろんここで示した程度のことであれば、Win のコマンドプロンプト上でも実現可能である (図 5: ウェブ資料 22-24)。例えば、Linux の grep に相当する Win のコマンドは find である。しかし、Linux にも find コマンドは存在し、(文字列検索ではなく) ファイル



W16-1: "" に注意！

①ペーストしても、「第2回のW21」と同じ結果にならず、何も表示されない。また、
②「”」というファイルが作成されてしまう

```
File Edit View Search Terminal Help
iu@bielinux[hoge] pwd [12:43午後]
/home/iu/Desktop/hoge
iu@bielinux[hoge] ls [12:43午後]
genome.fa
iu@bielinux[hoge] grep ">" genome.fa [12:43午後]
iu@bielinux[hoge] ls [12:45午後]
" genome.fa
iu@bielinux[hoge] [12:45午後]
```



W16-1: "" に注意!

初心者はコマンドライン上での削除は難しい。①ターミナル画面上で通常のダブルクォーテーションで消そうとしても「dquote>」という変なプロンプトが出てしまう。②リターンキーを押しても、③「qキー」を押してもだめだが、④「CTRLキー + Cキー」で復旧できる

```
iu@bielinux[hoge] pwd
/home/iu/Desktop/hoge
iu@bielinux[hoge] ls
genome.fa
iu@bielinux[hoge] grep ">" genome.fa
iu@bielinux[hoge] ls
" genome.fa
① iu@bielinux[hoge] rm "
② dquote>
③ dquote> q
④ dquote>
iu@bielinux[hoge] █
```

[12:43午後]

[12:45午後]

[12:45午後]

[12:49午後]

- ①「rm」のあとに、消したいファイルを反転させて
- ②コピーでうまく消すことができる。

W16-1: "" に注意!

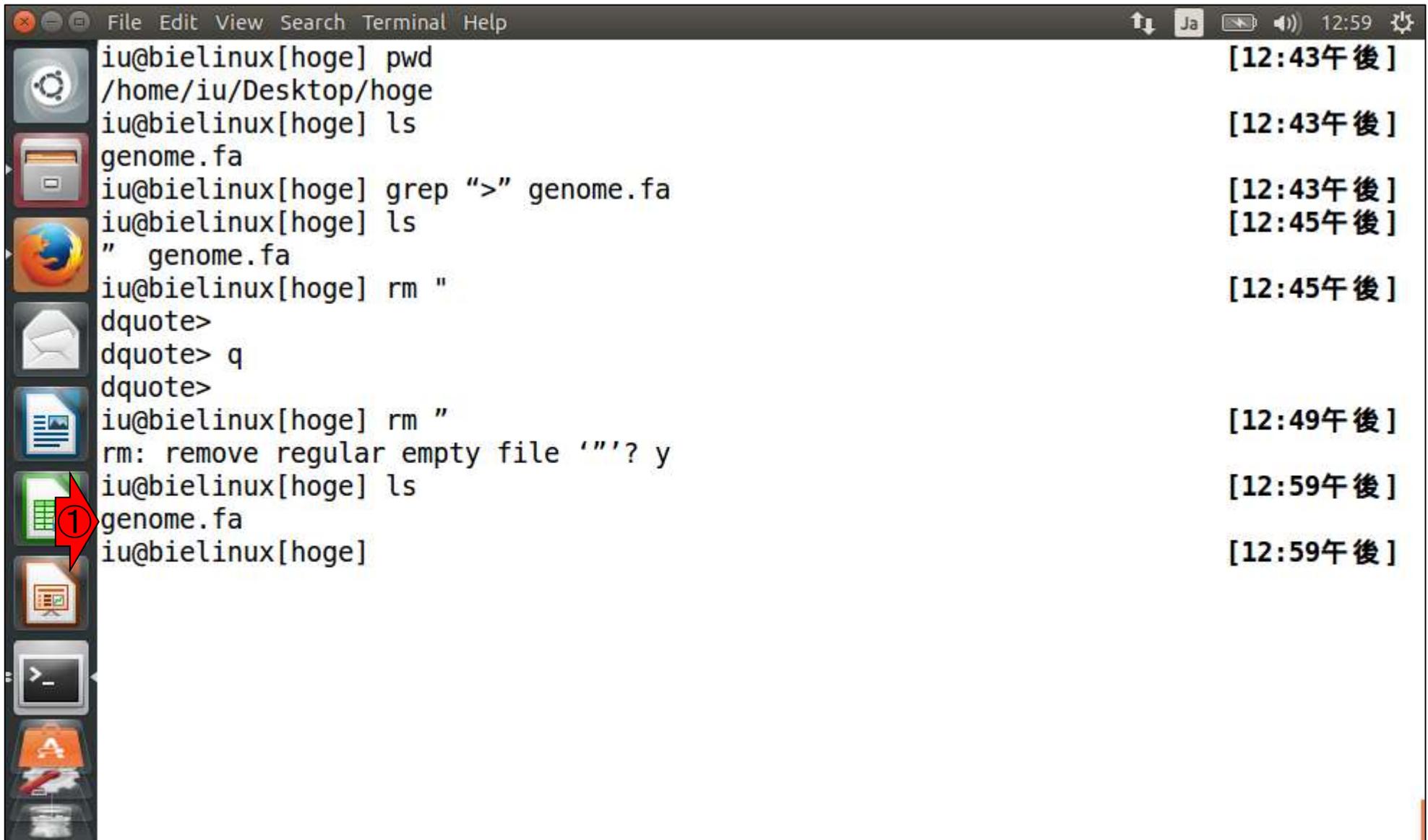
```
iu@bielinux[~/Desktop/hoge]
iu@bielinux[hoge] pwd [12:43午後]
/home/iu/Desktop/hoge
iu@bielinux[hoge] ls [12:43午後]
genome.fa
iu@bielinux[hoge] grep ">" genome.fa [12:43午後]
iu@bielinux[hoge] ls [12:45午後]
genome fa
iu@bielinux[hoge] rm " [12:45午後]
iu@bielinux[hoge] rm [12:49午後]
```

W16-1: "" に注意!

```
File Edit View Search Terminal Help 12:57
iu@bielinux[hoge] pwd [12:43午後]
/home/iu/Desktop/hoge
iu@bielinux[hoge] ls [12:43午後]
genome.fa
iu@bielinux[hoge] grep ">" genome.fa [12:43午後]
iu@bielinux[hoge] ls [12:45午後]
" genome.fa
iu@bielinux[hoge] rm " [12:45午後]
dquote>
dquote> q
dquote>
iu@bielinux[hoge] rm " [12:49午後]
rm: remove regular empty file ""? y
```



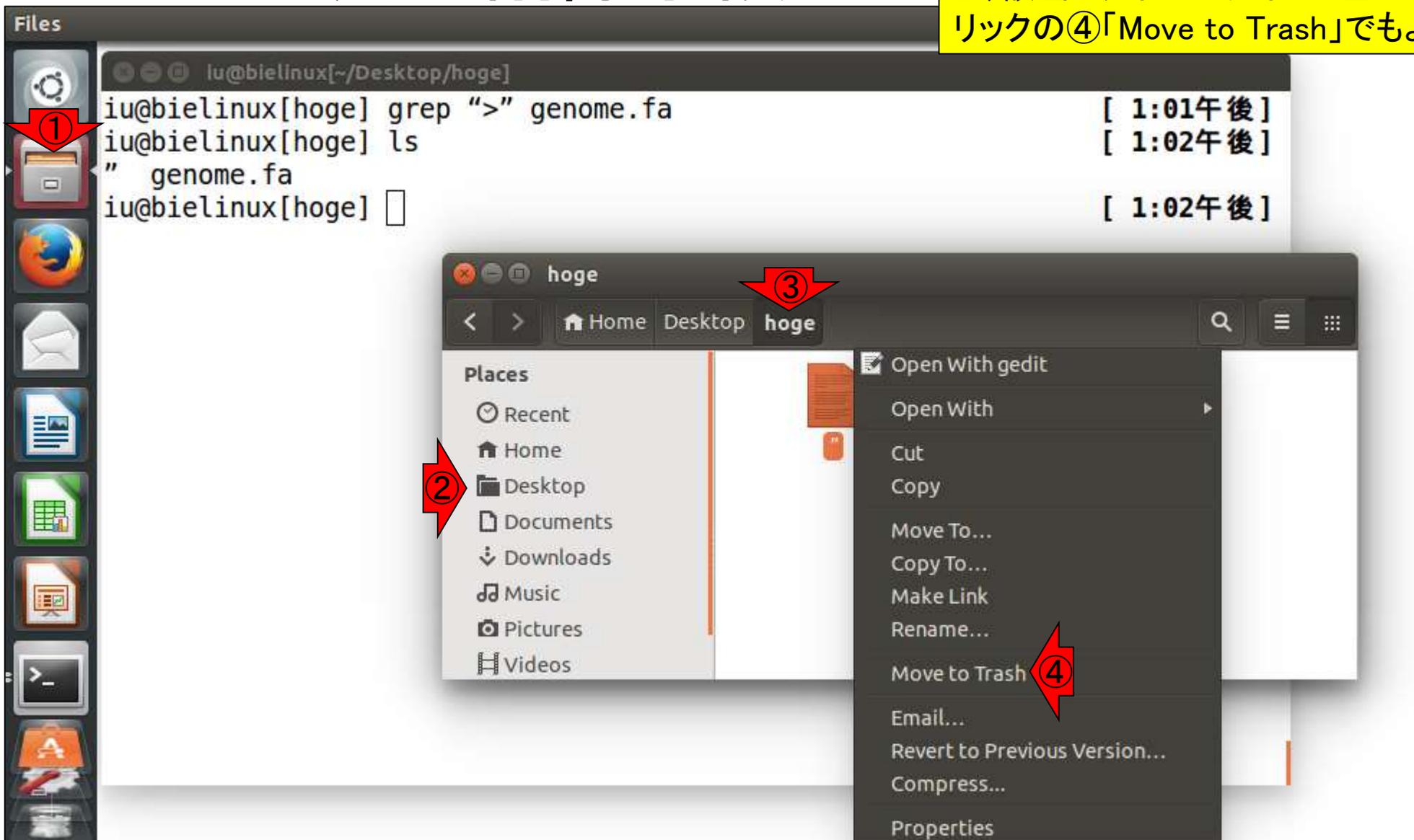
W16-1: "" に注意!



```
iu@bielinux[hoge] pwd [12:43午後]
/home/iu/Desktop/hoge
iu@bielinux[hoge] ls [12:43午後]
genome.fa
iu@bielinux[hoge] grep ">" genome.fa [12:43午後]
iu@bielinux[hoge] ls [12:45午後]
" genome.fa
iu@bielinux[hoge] rm " [12:45午後]
dquote>
dquote> q
dquote>
iu@bielinux[hoge] rm " [12:49午後]
rm: remove regular empty file ""? y
iu@bielinux[hoge] ls [12:59午後]
genome.fa
iu@bielinux[hoge] [12:59午後]
```

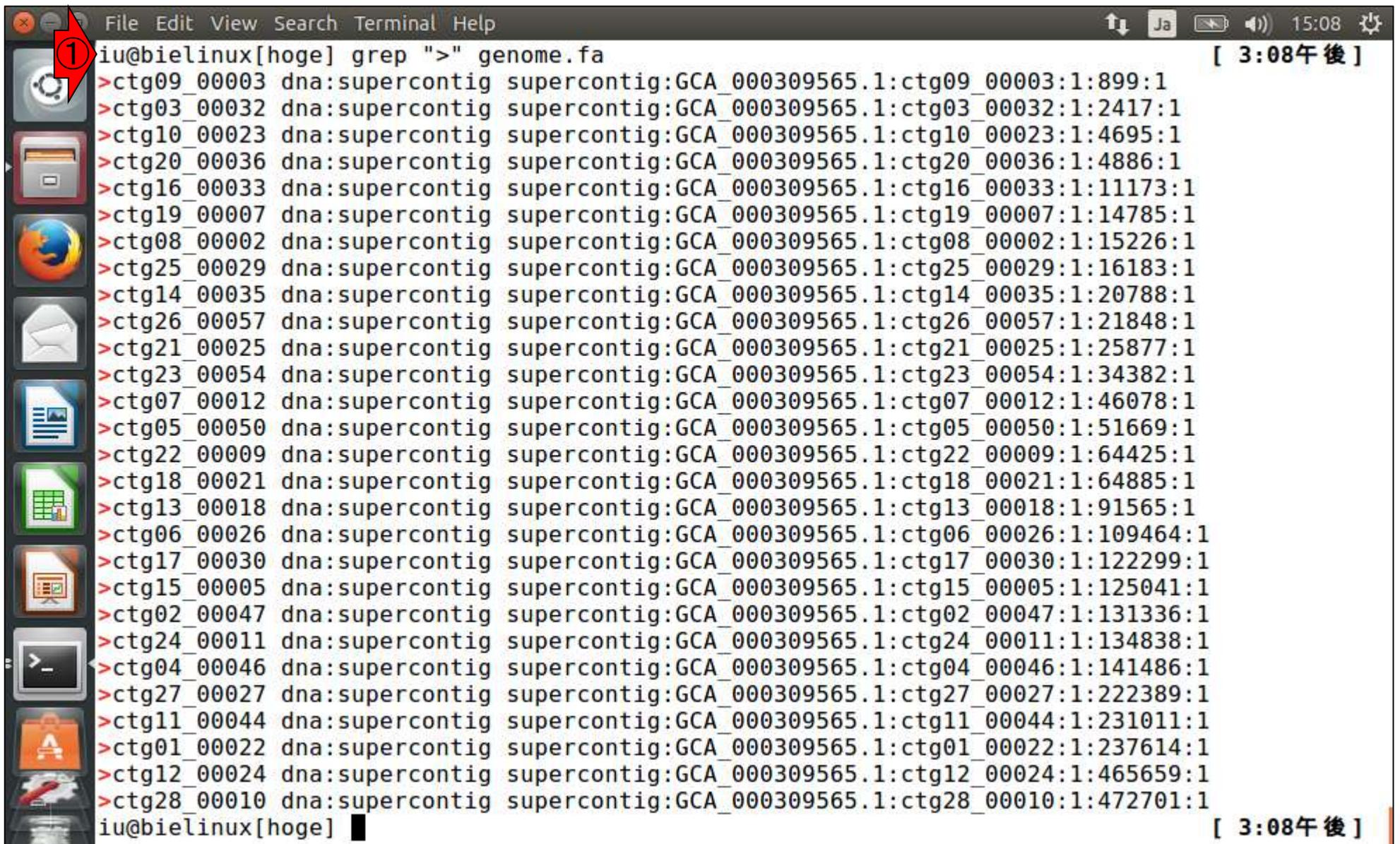
W16-1: 別の削除手段

別の削除手段。①引き出しみたいないなアイコンから③hogeフォルダに移動し、該当ファイルのアイコン上で右クリックの④「Move to Trash」でもよい



①Linux上での"正しい"ダブルクォーテーションマークは、開始記号と終了記号が同じほう

W16-2: 正しいのは...



```
iu@bielinux[hoge] grep ">" genome.fa [ 3:08午後 ]
>ctg09_00003 dna:supercontig supercontig:GCA_000309565.1:ctg09_00003:1:899:1
>ctg03_00032 dna:supercontig supercontig:GCA_000309565.1:ctg03_00032:1:2417:1
>ctg10_00023 dna:supercontig supercontig:GCA_000309565.1:ctg10_00023:1:4695:1
>ctg20_00036 dna:supercontig supercontig:GCA_000309565.1:ctg20_00036:1:4886:1
>ctg16_00033 dna:supercontig supercontig:GCA_000309565.1:ctg16_00033:1:11173:1
>ctg19_00007 dna:supercontig supercontig:GCA_000309565.1:ctg19_00007:1:14785:1
>ctg08_00002 dna:supercontig supercontig:GCA_000309565.1:ctg08_00002:1:15226:1
>ctg25_00029 dna:supercontig supercontig:GCA_000309565.1:ctg25_00029:1:16183:1
>ctg14_00035 dna:supercontig supercontig:GCA_000309565.1:ctg14_00035:1:20788:1
>ctg26_00057 dna:supercontig supercontig:GCA_000309565.1:ctg26_00057:1:21848:1
>ctg21_00025 dna:supercontig supercontig:GCA_000309565.1:ctg21_00025:1:25877:1
>ctg23_00054 dna:supercontig supercontig:GCA_000309565.1:ctg23_00054:1:34382:1
>ctg07_00012 dna:supercontig supercontig:GCA_000309565.1:ctg07_00012:1:46078:1
>ctg05_00050 dna:supercontig supercontig:GCA_000309565.1:ctg05_00050:1:51669:1
>ctg22_00009 dna:supercontig supercontig:GCA_000309565.1:ctg22_00009:1:64425:1
>ctg18_00021 dna:supercontig supercontig:GCA_000309565.1:ctg18_00021:1:64885:1
>ctg13_00018 dna:supercontig supercontig:GCA_000309565.1:ctg13_00018:1:91565:1
>ctg06_00026 dna:supercontig supercontig:GCA_000309565.1:ctg06_00026:1:109464:1
>ctg17_00030 dna:supercontig supercontig:GCA_000309565.1:ctg17_00030:1:122299:1
>ctg15_00005 dna:supercontig supercontig:GCA_000309565.1:ctg15_00005:1:125041:1
>ctg02_00047 dna:supercontig supercontig:GCA_000309565.1:ctg02_00047:1:131336:1
>ctg24_00011 dna:supercontig supercontig:GCA_000309565.1:ctg24_00011:1:134838:1
>ctg04_00046 dna:supercontig supercontig:GCA_000309565.1:ctg04_00046:1:141486:1
>ctg27_00027 dna:supercontig supercontig:GCA_000309565.1:ctg27_00027:1:222389:1
>ctg11_00044 dna:supercontig supercontig:GCA_000309565.1:ctg11_00044:1:231011:1
>ctg01_00022 dna:supercontig supercontig:GCA_000309565.1:ctg01_00022:1:237614:1
>ctg12_00024 dna:supercontig supercontig:GCA_000309565.1:ctg12_00024:1:465659:1
>ctg28_00010 dna:supercontig supercontig:GCA_000309565.1:ctg28_00010:1:472701:1
iu@bielinux[hoge] [ 3:08午後 ]
```

W16-3: 横棒にも注意

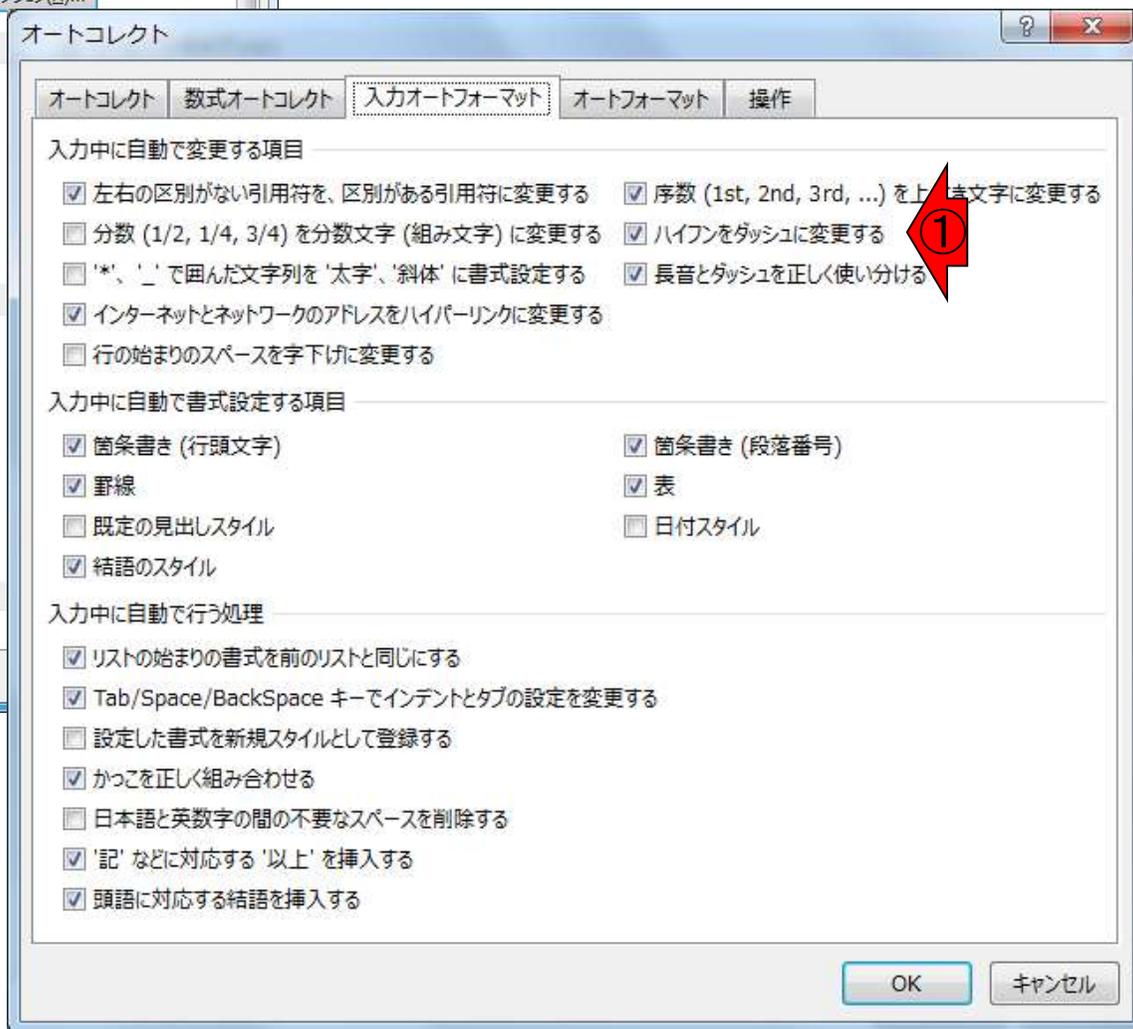
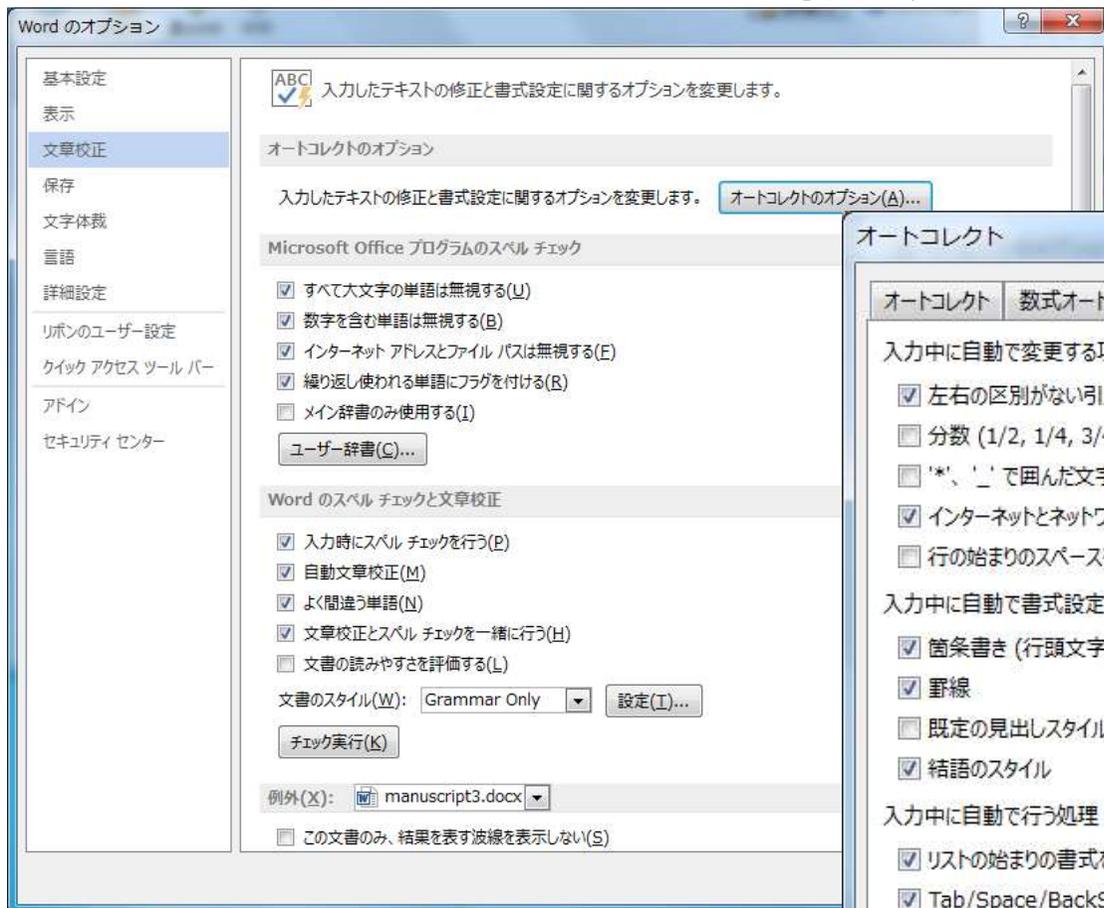
原因既知状態で眺めると①「-(ハイフン)」と②「-(ダッシュ)」の違いは明白。但し、③「-(全角ハイフン)」はほとんど見分けがつかない。Bio-Linuxで用いられているフォントは④いち、と⑤える、の区別が付きやすい。「全角」は利用しないが基本中の基本。

```
File Edit View Search Terminal Help
iu@bielinux[hoge] pwd
/home/iu/Desktop/hoge
iu@bielinux[hoge] ls
genome.fa
① iu@bielinux[hoge] grep -c ">" genome.fa
28
iu@bielinux[hoge] grep -c ">" genome.fa ②
grep: >: No such file or directory
③ iu@bielinux[hoge] grep -c ">" genome.fa
grep: >: No such file or directory
④ iu@bielinux[hoge] l
zsh: command not found: l
⑤ iu@bielinux[hoge] l
zsh: command not found: l
iu@bielinux[hoge] █
```

[3:22午後]

W16-4: wordの設定

著者のデフォルト環境では、Microsoft Wordの「オプション」-「オートコレクトのオプション」-「入力オートフォーマット」で①「ハイフンをダッシュに変更する」のチェックが入っている



W17-1: エディタ

Bio-Linux上で利用可能なテキストエディタは、vi, emacs, geditなど。「vi genome.fa」、「emacs genome.fa」、①「gedit genome.fa」などでそれぞれのエディタを起動できる。

The screenshot displays a Linux desktop environment. On the left is a vertical dock with icons for applications like a file manager, Firefox (with a red circle and arrow labeled '1'), a mail client, a document viewer, a spreadsheet, a presentation, and a terminal. The main window is a terminal titled 'iu@bielinux[~/Desktop/hoge]'. It shows the following commands and their outputs:

```
iu@bielinux[hoge] pwd [ 3:34午後 ]
/home/iu/Desktop/hoge
iu@bielinux[hoge] ls [ 3:34午後 ]
genome.fa
iu@bielinux[hoge] gedit genome.fa [ 3:34午後 ]
█
```

An inset window titled 'genome.fa (~/Desktop/hoge) - gedit' shows the contents of the file genome.fa. The text is as follows:

```
genome.fa x
>ctg09_00003 dna:supercontig
supercontig:GCA_000309565.1:ctg09_00003:1:899:1
AGTAAGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCT
TATGACCTGGGCTACACACGTGCTACAATGGATGGTACAACGAGTTGCGAGACCGCGAGG
TYAAGCTAATCTCTTAAAGCCATTCTCAGTTCGGACTGTAGGCTGCAACTCGCCTACACG
AAGTCGGAATCGCTAGTAATCGCGGATCAGCACGCCGCGGTGAATACGTTCCCCGGGCCTT
GTACACACCGCCCGTCACACCATGAGAGTTTGTAAACCCGAAGCCGGTGGCGTAACCCT
TTTAGGGAGCGAGCCGTCTAAGGTGGGACAAATGATTAGGGTGAAGTCGTAACAAGGTAG
CCGTAGGAGAACCTGCGGCTGGATCACCTCCTTTCTAAGGAAACAGACTRAAAGTCTGAC
GGAAACCTGCACACACGAACTTTGTTTAGTTTTGAGGGGATCACCTCAAGCACCCCTAA
CGGGTGCGACTTTGTTCTTTGAAACTGGATATCATTGTATTAATTGTTTTAAATTGCCG
AGAACACAGCGTATTTGTATGAGTTTCTGAAAAAGAAATTCGCATCGCATAACCGCTGAC
GCAAGTCAGTACAGGTTAAGTTRCAAAGGGCGCACGGTGGATGCCTTGGCACTAGGAGCC
GATGAAGGACGGAACATAACCGATATGCTTCGGGGAGCTATAAGTAAGCTTTGATCCGG
```

The status bar at the bottom of the gedit window shows 'Plain Text', 'Tab Width: 8', 'Ln 1, Col 1', and 'INS'.

W17-1: エディタ

gedit起動中は、geditというコマンド実行中と同義である。そのため、①このターミナル上では何か新たなコマンドを打ち込むことはできない。②赤矢印の×ボタンを押すとコマンド入力待ち状態になる。×を押してみる。

Terminal output:

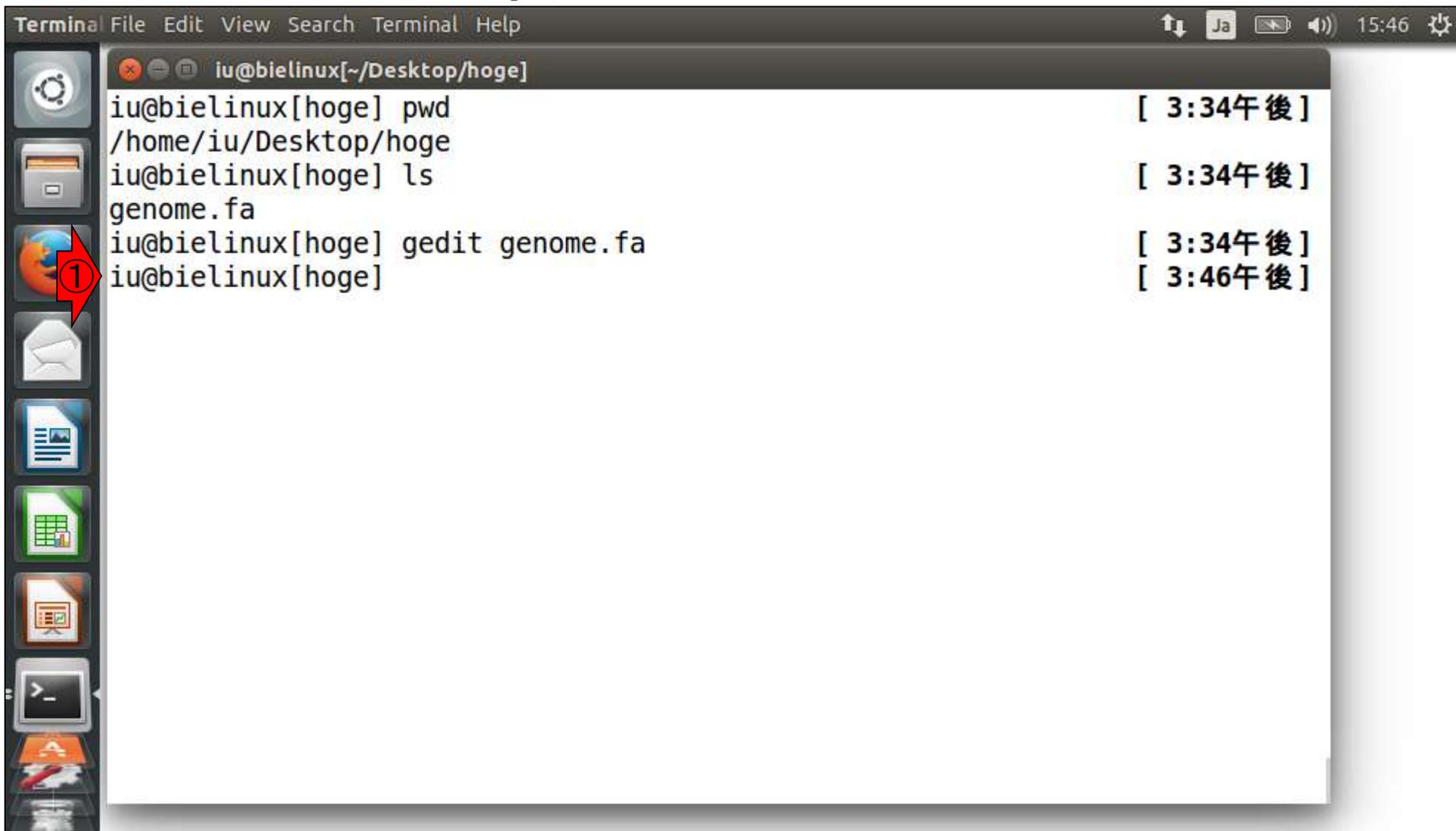
```
iu@bielinux[~/Desktop/hoge]
iu@bielinux[hoge] pwd
/home/iu/Desktop/hoge
iu@bielinux[hoge] ls
genome.fa
iu@bielinux[hoge] gedit genome.fa
```

gedit window content:

```
genome.fa x
>ctg09_00003 dna:supercontig
supercontig:GCA_000309565.1:ctg09_00003:1:899:1
AGTAAGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCT
TATGACCTGGGCTACACACGTGCTACAATGGATGGTACAACGAGTTGCGAGACCGCGAGG
TYAAGCTAATCTCTTAAAGCCATTCTCAGTTCGGACTGTAGGCTGCAACTCGCCTACACG
AAGTCGGAATCGCTAGTAATCGCGGATCAGCACGCCGCGGTGAATACGTTCCCGGGCCTT
GTACACACCGCCCGTCACACCATGAGAGTTTGTAAACCCGAAGCCGGTGGCGTAACCCT
TTTAGGGAGCGAGCCGTCTAAGGTGGGACAAATGATTAGGGTGAAGTCGTAACAAGGTAG
CCGTAGGAGAACCTGCGGCTGGATCACCTCCTTTCTAAGGAAACAGACTRAAAGTCTGAC
GGAAACCTGCACACACGAACTTTGTTTAGTTTTGAGGGGATCACCTCAAGCACCCCTAA
CGGGTGCGACTTTGTTCTTTGAAAACCTGGATATCATTGTATTAATTGTTTTAAATTGCCG
AGAACACAGCGTATTTGTATGAGTTTCTGAAAAAGAAATTCGCATCGCATAACCGCTGAC
GCAAGTCAGTACAGGTTAAGTTRCAAAGGGCGCACGGTGGATGCCTTGGCACTAGGAGCC
GATGAAGGACGGAACCTAATACCGATATGCTTCGGGGAGCTATAAGTAAGCTTTGATCCGG
```

W17-1: エディタ

×を押して、gedit終了後の状態。①コマンド入力待ち状態に変わったことがわかる。



The image shows a terminal window with a menu bar (Terminal, File, Edit, View, Search, Terminal, Help) and a system tray (Ja, 15:46). The terminal output is as follows:

```
iu@bielinux[~/Desktop/hoge]
iu@bielinux[hoge] pwd [ 3:34午後]
/home/iu/Desktop/hoge
iu@bielinux[hoge] ls [ 3:34午後]
genome.fa
iu@bielinux[hoge] gedit genome.fa [ 3:34午後]
iu@bielinux[hoge] [ 3:46午後]
```

A red arrow with the number 1 points to the prompt `iu@bielinux[hoge]` on the line following the `gedit` command, indicating the state after the editor has closed.

W17-2: コマンドファイル

一連のコマンドを記したスクリプトファイル(JSLAB3_code.txt)を取得。ここでは①wgetを利用。(記述内容量が少ないことが分かっているので)ダウンロード後に②moreコマンドで眺める。

```
iu@bielinux[hoge] ls
genome.fa
① iu@bielinux[hoge] wget -c http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB3_code.txt
--2015-11-30 15:56:19-- http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB3_code.txt
Resolving www.iu.a.u-tokyo.ac.jp (www.iu.a.u-tokyo.ac.jp)... 133.11.224.26
Connecting to www.iu.a.u-tokyo.ac.jp (www.iu.a.u-tokyo.ac.jp)|133.11.224.26|:80...
connected.
HTTP request sent, awaiting response... 200 OK
Length: 579 [text/plain]
Saving to: 'JSLAB3_code.txt'

100%[=====] 579      ---K/s   in 0s

2015-11-30 15:56:19 (79.3 MB/s) - 'JSLAB3_code.txt' saved [579/579]

iu@bielinux[hoge] ls
genome.fa JSLAB3_code.txt
② iu@bielinux[hoge] more JSLAB3_code.txt
```

http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB3_code.txt

W17-2: コマンドファイル

②moreコマンド実行結果。赤枠部分がJSLAB3_code.txtの中身。

```
File Edit View Search Terminal Help
2015-11-30 15:56:19 (79.3 MB/s) - 'JSLAB3_code.txt' saved [579/579]
iu@bielinux[hoge] ls [ 3:56午後]
genome.fa JSLAB3_code.txt
iu@bielinux[hoge] more JSLAB3_code.txt [ 3:56午後]
#####
### Download file
#####
wget -c ftp://ftp.ensemblgenomes.org/pub/bacteria/release-22/fasta/bacteria_15_coll
ection/lactobacillus_casei_12a/dna/Lactobacillus_casei_12a.GCA_000309565.1.22.dna.t
oplevel.fa.gz
#####
### Expand gzipped file
#####
gunzip Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa.gz
#####
### Rename
#####
mv Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa genome.fa
iu@bielinux[hoge] [ 4:12午後]
```

W17-2: rm -f

別のやり方を示すために一旦削除。rmコマンド実行時に、「-f」オプションをつけると「本当に消していいか？」という警告メッセージが表示されない。

```
iu@bielinux[hoge] more JSLAB3_code.txt
#####
### Download file
#####
wget -c ftp://ftp.ensemblgenomes.org/pub/bacteria/release-22/fasta/bacteria_15_coll
ection/lactobacillus_casei_12a/dna/Lactobacillus_casei_12a.GCA_000309565.1.22.dna.t
oplevel.fa.gz
#####
### Expand gzipped file
#####
gunzip Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa.gz
#####
### Rename
#####
mv Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa genome.fa
iu@bielinux[hoge] ls
genome.fa JSLAB3_code.txt
① iu@bielinux[hoge] rm -f JSLAB3_code.txt
iu@bielinux[hoge] ls
genome.fa
iu@bielinux[hoge]
```

[3:50午後]
[4:12午後]
[4:19午後]
[4:20午後]
[4:20午後]

W17-3: gedit上でコピー

別のやり方。①geditを起動して、ウェブ上の情報をコピーし、任意のファイル名で保存するのもよい。

The screenshot shows a Linux desktop environment. In the foreground, a terminal window displays a series of commands and their outputs. A red arrow with the number '1' points to the command `iu@bielinux[hoge] gedit`. In the background, a Mozilla Firefox browser window is open, displaying a webpage with text that has been copied into the gedit editor window. The gedit window is titled "Untitled Document 1 - gedit" and shows the copied text. The terminal window also shows the output of the `gedit` command, which is an empty file.

```
#####  
### Download file  
#####  
##### iu@bielinux[~/Desktop/  
wget -c### Expand gzipped fi  
/bacter#####  
/dna/Li gunzip Lactobacillus_c  
#####  
#####  
### E#####  
#####  
#####  
gunzip mv Lactobacillus_casei  
#####  
iu@bielinux[hoge] ls  
##### genome.fa JSLAB3_code.txt  
### R#####  
##### iu@bielinux[hoge] rm -f JSLAB3_code.txt [ 4:  
##### iu@bielinux[hoge] ls [ 4:  
mv lac genome.fa  
#####  
iu@bielinux[hoge] gedit [ 4:27午後 ]
```

W17-3: gedit上でコピー

別のやり方。①geditを起動して、ウェブ上の情報をコピーし、任意のファイル名で保存するのもよい。②右クリックのコピーが基本です。

Firefox Web Browser

Mozilla Firefox

http://www...3_code.txt x +

www.iu.a.u-tokyo.ac.jp/~kadota/b Search

```
### Download file  
#####  
wget -c ftp://ftp.ensemblgenom  
/bacteria_15_collection/lactoba  
/dna/Lactobacillus_casei_12a.G  
#####  
### Expand gzipped file  
#####  
gunzip Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa.gz  
#####  
### Rename  
#####  
mv Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa genome.fa  
genome.fa  
iu@bielinux[hoge] gedit
```

Undo

Copy ②

Select A

Search Google for "#####..."

View Selection Source

Inspect Element (Q)

Width: 8 Ln 1, Col 1 INS

[4:

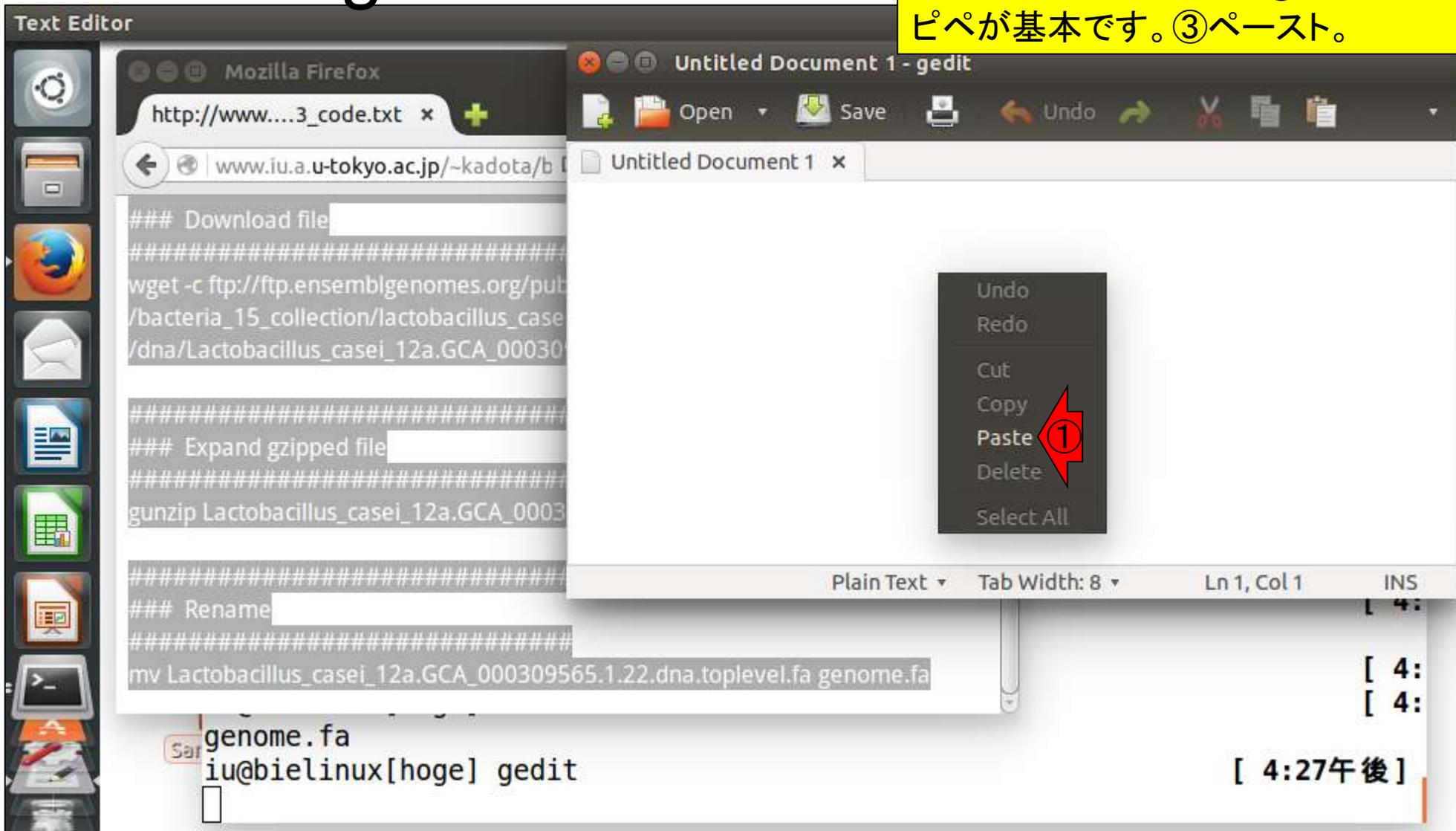
[4:

[4:

[4:27午後]

W17-3: gedit上でコピー

別のやり方。①geditを起動して、ウェブ上の情報をコピーし、任意のファイル名で保存するのもよい。②右クリックのコピーが基本です。③ペースト。



W17-3: gedit上でコピー

The screenshot shows a terminal window with the following content:

```
Sar genome.fa
iu@bielinux[hoge] gedit
```

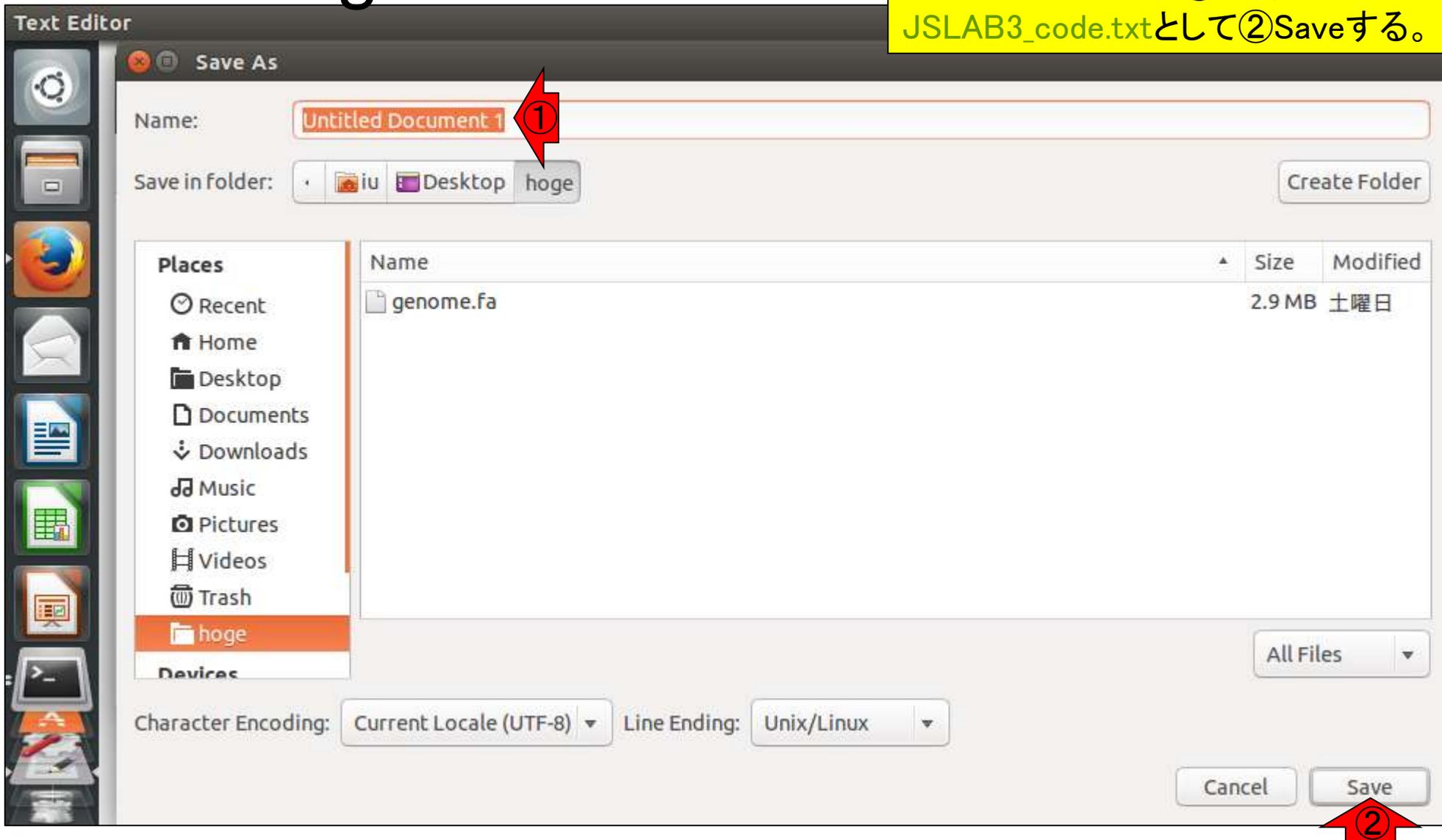
The gedit window shows the following content:

```
#####
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa.gz
#####
### Expand gzipped file
#####
gunzip
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa.gz
#####
### Rename
#####
mv Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa
genome.fa
```

The gedit window also shows the status bar: Plain Text, Tab Width: 8, Ln 15, Col 1, INS.

W17-3: gedit上でコピー

デフォルトの保存先は、geditを起動した「カレントディレクトリ」になるようですね。ここではファイル名を①の部分を JSLAB3_code.txtとして②Saveする。



W17-3: gedit上でコピー

Saveボタンを押した後の状態。Linux環境とはいえ、通常のテキストエディタと同じような感覚で取り扱えることがわかる。
①×を押してgeditを終了。

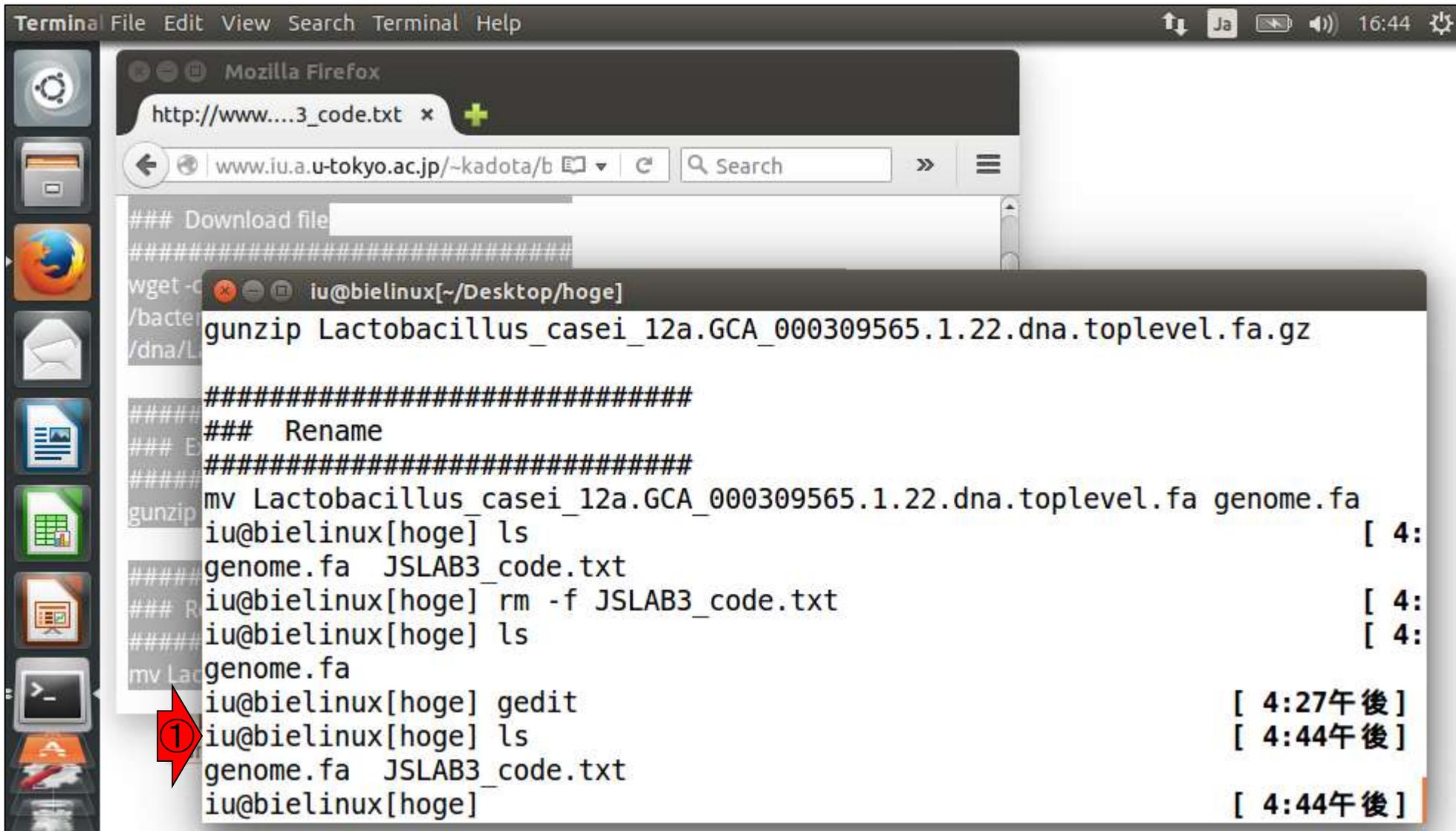
The screenshot shows a Linux desktop environment. In the foreground, a terminal window is open, displaying a series of commands and their outputs. The commands include `wget`, `gunzip`, and `mv`. The outputs show the download of a file, its expansion, and its renaming to `genome.fa`. In the background, a gedit editor window is open, showing the same text being edited. A red arrow points to the close button (X) in the gedit window's title bar, indicating that the user should press it to exit gedit.

```
### Download file [redacted]
#####
wget -c ftp://ftp.ensemblgenomes.org/pub/bacteria_15_collection/lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa.gz
#####
### Expand gzipped file
#####
gunzip
Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa.gz
#####
### Rename
#####
mv Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa genome.fa
#####
### Rename [redacted]
#####
mv Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa genome.fa

genome.fa
iu@bielinux[hoge] gedit
```

ターミナル画面上で①lsすると、確かにgedit上で作成したJSLAB3_code.txtが存在することがわかる。

W17-3: 確認



```
Terminal File Edit View Search Terminal Help
Mozilla Firefox
http://www...3_code.txt x
www.iu.a.u-tokyo.ac.jp/~kadota/b Search
### Download file
#####
wget -c
/bacter
/dna/L
iu@bielinux[~/Desktop/hoge]
gunzip Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa.gz
#####
### Rename
### E
#####
mv Lactobacillus_casei_12a.GCA_000309565.1.22.dna.toplevel.fa genome.fa
gunzip
#####
### R
#####
mv Lac
iu@bielinux[hoge] ls [ 4:
genome.fa JSLAB3_code.txt
iu@bielinux[hoge] rm -f JSLAB3_code.txt [ 4:
iu@bielinux[hoge] ls [ 4:
genome.fa
iu@bielinux[hoge] gedit [ 4:27午後]
iu@bielinux[hoge] ls [ 4:44午後]
genome.fa JSLAB3_code.txt
iu@bielinux[hoge] [ 4:44午後]
```

W18-1:リダイレクト

①リダイレクトで実行結果をファイルに保存。赤下線で示すように、出力ファイル名で指定したheaders.txtが確かに生成されていることがわかる。②headers.txtの最初の11行分表示して確認している。

```
File Edit View Search Terminal Help
iu@bielinux[hoge] pwd
/home/iu/Desktop/hoge
iu@bielinux[hoge] ls
genome.fa  JSLAB3_code.txt
iu@bielinux[hoge] grep ">" genome.fa > headers.txt
iu@bielinux[hoge] ls
genome.fa  headers.txt  JSLAB3_code.txt
iu@bielinux[hoge] head -n 11 headers.txt
>ctg09_00003 dna:supercontig supercontig:GCA_000309565.1:ctg09_00003:1:899:1
>ctg03_00032 dna:supercontig supercontig:GCA_000309565.1:ctg03_00032:1:2417:1
>ctg10_00023 dna:supercontig supercontig:GCA_000309565.1:ctg10_00023:1:4695:1
>ctg20_00036 dna:supercontig supercontig:GCA_000309565.1:ctg20_00036:1:4886:1
>ctg16_00033 dna:supercontig supercontig:GCA_000309565.1:ctg16_00033:1:11173:1
>ctg19_00007 dna:supercontig supercontig:GCA_000309565.1:ctg19_00007:1:14785:1
>ctg08_00002 dna:supercontig supercontig:GCA_000309565.1:ctg08_00002:1:15226:1
>ctg25_00029 dna:supercontig supercontig:GCA_000309565.1:ctg25_00029:1:16183:1
>ctg14_00035 dna:supercontig supercontig:GCA_000309565.1:ctg14_00035:1:20788:1
>ctg26_00057 dna:supercontig supercontig:GCA_000309565.1:ctg26_00057:1:21848:1
>ctg21_00025 dna:supercontig supercontig:GCA_000309565.1:ctg21_00025:1:25877:1
iu@bielinux[hoge]
```

[4:55午後]
[4:55午後]
[4:55午後]
[4:55午後]
[4:55午後]

W18-1: wc, grep

①「grep ">" genome.fa > headers.txt」に引き続いて③「wc -l headers.txt」とすることでコンティグ数を調べることができる。

```
iu@bielinux[hoge] grep ">" genome.fa > headers.txt [ 4:55午後]
iu@bielinux[hoge] ls [ 4:55午後]
genome.fa headers.txt JSLAB3_code.txt
iu@bielinux[hoge] head -n 11 headers.txt [ 4:55午後]
>ctg09_00003 dna:supercontig supercontig:GCA_000309565.1:ctg09_00003:1:899:1
>ctg03_00032 dna:supercontig supercontig:GCA_000309565.1:ctg03_00032:1:2417:1
>ctg10_00023 dna:supercontig supercontig:GCA_000309565.1:ctg10_00023:1:4695:1
>ctg20_00036 dna:supercontig supercontig:GCA_000309565.1:ctg20_00036:1:4886:1
>ctg16_00033 dna:supercontig supercontig:GCA_000309565.1:ctg16_00033:1:11173:1
>ctg19_00007 dna:supercontig supercontig:GCA_000309565.1:ctg19_00007:1:14785:1
>ctg08_00002 dna:supercontig supercontig:GCA_000309565.1:ctg08_00002:1:15226:1
>ctg25_00029 dna:supercontig supercontig:GCA_000309565.1:ctg25_00029:1:16183:1
>ctg14_00035 dna:supercontig supercontig:GCA_000309565.1:ctg14_00035:1:20788:1
>ctg26_00057 dna:supercontig supercontig:GCA_000309565.1:ctg26_00057:1:21848:1
>ctg21_00025 dna:supercontig supercontig:GCA_000309565.1:ctg21_00025:1:25877:1
iu@bielinux[hoge] wc -l headers.txt [ 4:55午後]
28 headers.txt
iu@bielinux[hoge] wc headers.txt [ 5:08午後]
 28  84 2218 headers.txt
iu@bielinux[hoge] grep -c ">" genome.fa [ 5:08午後]
28
iu@bielinux[hoge] grep -c ">" genome.fa > tmp.txt [ 5:08午後]
iu@bielinux[hoge] [ 5:08午後]
```

W18-2: df -h

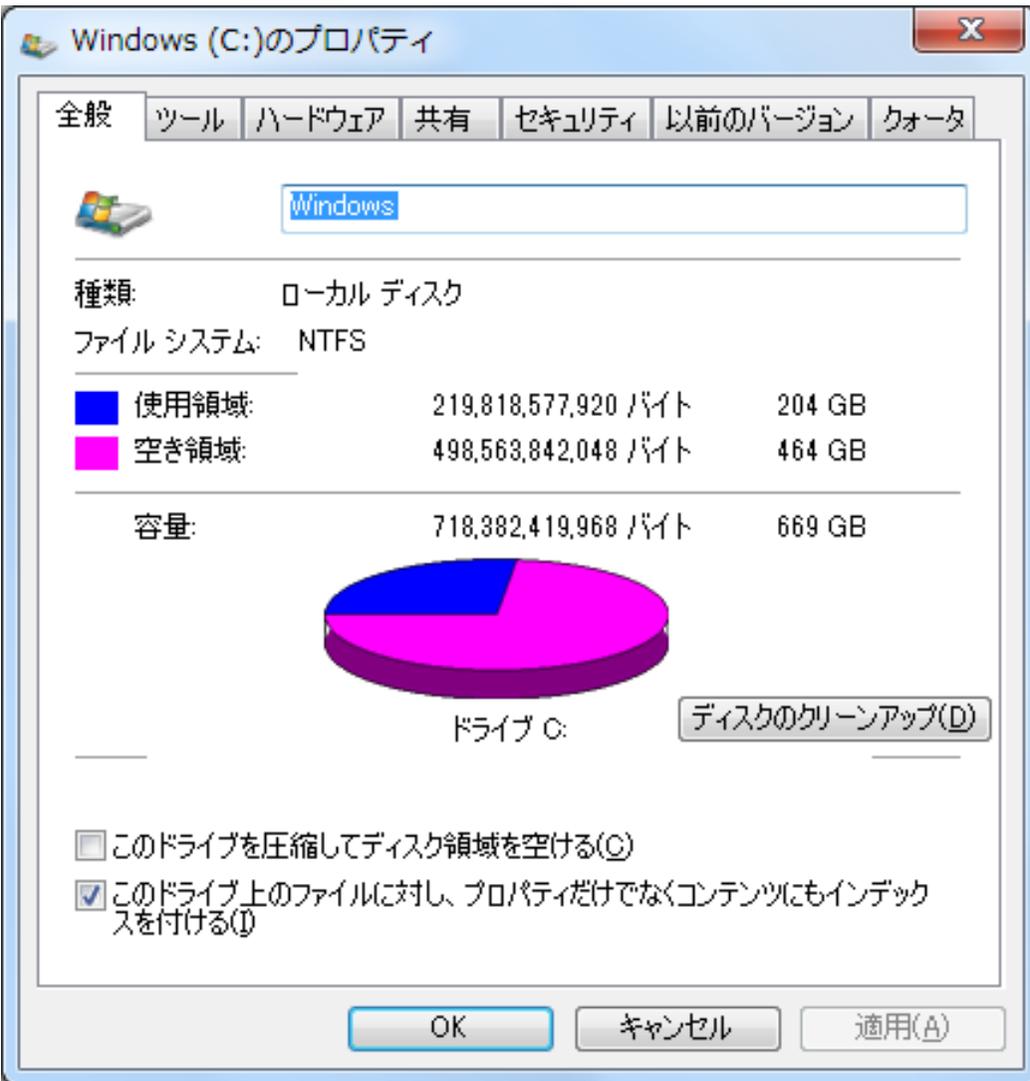
どのディレクトリ上でもいいので①「df -h」を実行。
②Sizeのところが97Gとなっているのは、仮想マシン設定時に割り当てを100GBとしたから。Use%のところが全体の何%を利用したかを表している。ここでは11%となっているが、90%を超えない程度に適宜不要なファイルを削除したほうがいいだろう。

```
iu@bielinux[iu] pwd
/home/iu
iu@bielinux[iu] df -h
df: 'mac_share': No such file or directory
Filesystem      Size  Used Avail Use% Mounted on
/dev/sda1       97G   9.7G   82G   11% /
none            4.0K    0   4.0K    0% /sys/fs/cgroup
udev           991M   4.0K  991M    1% /dev
tmpfs          201M  932K  200M    1% /run
none           5.0M    0   5.0M    0% /run/lock
none          1001M  156K  1001M    1% /run/shm
none           100M   64K   100M    1% /run/user
iu@bielinux[iu] cd
iu@bielinux[iu] pwd
/home/iu
iu@bielinux[iu] df -h
df: 'mac_share': No such file or directory
Filesystem      Size  Used Avail Use% Mounted on
/dev/sda1       97G   9.7G   82G   11% /
none            4.0K    0   4.0K    0% /sys/fs/cgroup
udev           991M   4.0K  991M    1% /dev
tmpfs          201M  932K  200M    1% /run
none           5.0M    0   5.0M    0% /run/lock
none          1001M  156K  1001M    1% /run/shm
none           100M   64K   100M    1% /run/user
iu@bielinux[iu]
```

[5:20午後]
[5:20午後]
[5:20午後]
[5:20午後]

「df -h」結果で眺めるUse%は、ホストOS (Windows)上で使用領域と空き領域の割合を眺めているのと同じようなもの

W18-2: df -h



W18-3: du -sh

カレントディレクトリ(/home/iu/Desktop/hoge)のディスク使用量(2.9 MB)は、①lsコマンド実行結果(赤下線)からも一応分かる。が、ディレクトリ内の全容量を調べるのは②duコマンドを利用するのが一般的。

```
iu@bielinux[hoge] pwd [ 5:34午後 ]
/home/iu/Desktop/hoge
① iu@bielinux[hoge] ls -lh [ 5:34午後 ]
total 2.9M
-rw-rw-r-- 1 iu iu 2.8M 11月 28 12:23 genome.fa
-rw-rw-r-- 1 iu iu 2.2K 11月 30 16:55 headers.txt
-rw-rw-r-- 1 iu iu 566 11月 30 16:42 JSLAB3_code.txt
-rw-rw-r-- 1 iu iu 3 11月 30 17:08 tmp.txt
② iu@bielinux[hoge] du -sh [ 5:34午後 ]
2.9M .
iu@bielinux[hoge] cd .. [ 5:34午後 ]
iu@bielinux[Desktop] pwd [ 5:34午後 ]
/home/iu/Desktop
iu@bielinux[Desktop] ls -lh [ 5:34午後 ]
total 12K
drwxr-xr-x 2 iu iu 4.0K 11月 17 13:44 Bio-Linux Documentation
drwxrwxr-x 2 iu iu 4.0K 11月 30 17:08 hoge
drwxrwxrwx 1 root root 0 11月 28 11:13 mac_share
drwxrwxr-x 2 iu iu 4.0K 11月 26 19:15 mac_share2
lrwxrwxrwx 1 iu iu 29 11月 17 13:44 Sample Data -> /usr/local/bioinf/sampledata/
iu@bielinux[Desktop] du -sh [ 5:34午後 ]
5.8M .
iu@bielinux[Desktop] [ 5:34午後 ]
```

①「どっと(.)」はカレントディレクトリという意味。つまり、「どっと(.)」以下の全ファイルの容量は2.9Mですよという出力結果

W18-3: du -sh

```
iu@bielinux[hoge] pwd [ 5:34午後 ]
/home/iu/Desktop/hoge
iu@bielinux[hoge] ls -lh [ 5:34午後 ]
total 2.9M
-rw-rw-r-- 1 iu iu 2.8M 11月 28 12:23 genome.fa
-rw-rw-r-- 1 iu iu 2.2K 11月 30 16:55 headers.txt
-rw-rw-r-- 1 iu iu 566 11月 30 16:42 JSLAB3_code.txt
-rw-rw-r-- 1 iu iu 3 11月 30 17:08 tmp.txt
iu@bielinux[hoge] du -sh [ 5:34午後 ]
2.9M .
iu@bielinux[hoge] cd .. [ 5:34午後 ]
iu@bielinux[Desktop] pwd [ 5:34午後 ]
/home/iu/Desktop
iu@bielinux[Desktop] ls -lh [ 5:34午後 ]
total 12K
drwxr-xr-x 2 iu iu 4.0K 11月 17 13:44 Bio-Linux Documentation
drwxrwxr-x 2 iu iu 4.0K 11月 30 17:08 hoge
drwxrwxrwx 1 root root 0 11月 28 11:13 mac_share
drwxrwxr-x 2 iu iu 4.0K 11月 26 19:15 mac_share2
lrwxrwxrwx 1 iu iu 29 11月 17 13:44 Sample Data -> /usr/local/bioinf/sampledata/
iu@bielinux[Desktop] du -sh [ 5:34午後 ]
5.8M .
iu@bielinux[Desktop] [ 5:34午後 ]
```



①

①「cd ..」は1つ上の階層のディレクトリに移動するときのお約束的な指令。

W18-3: du -sh

```
iu@bielinux[hoge] pwd [ 5:34午後 ]
/home/iu/Desktop/hoge
iu@bielinux[hoge] ls -lh [ 5:34午後 ]
total 2.9M
-rw-rw-r-- 1 iu iu 2.8M 11月 28 12:23 genome.fa
-rw-rw-r-- 1 iu iu 2.2K 11月 30 16:55 headers.txt
-rw-rw-r-- 1 iu iu 566 11月 30 16:42 JSLAB3_code.txt
-rw-rw-r-- 1 iu iu 3 11月 30 17:08 tmp.txt
iu@bielinux[hoge] du -sh [ 5:34午後 ]
2.9M .
iu@bielinux[hoge] cd .. [ 5:34午後 ]
iu@bielinux[Desktop] pwd [ 5:34午後 ]
/home/iu/Desktop
iu@bielinux[Desktop] ls -lh [ 5:34午後 ]
total 12K
drwxr-xr-x 2 iu iu 4.0K 11月 17 13:44 Bio-Linux Documentation
drwxrwxr-x 2 iu iu 4.0K 11月 30 17:08 hoge
drwxrwxrwx 1 root root 0 11月 28 11:13 mac_share
drwxrwxr-x 2 iu iu 4.0K 11月 26 19:15 mac_share2
lrwxrwxrwx 1 iu iu 29 11月 17 13:44 Sample Data -> /usr/local/bioinf/sampledata/
iu@bielinux[Desktop] du -sh [ 5:34午後 ]
5.8M .
iu@bielinux[Desktop] [ 5:34午後 ]
```



W18-3: du -sh

カレントディレクトリが/home/iu/Desktopの状態、③ls結果と④du結果を比較するとduコマンドを使う意義が分かる。(lsはディレクトリ内のファイルサイズは計算していない)

```
iu@bielinux[hoge] pwd [ 5:34午後 ]
/home/iu/Desktop/hoge
iu@bielinux[hoge] ls -lh [ 5:34午後 ]
total 2.9M
-rw-rw-r-- 1 iu iu 2.8M 11月 28 12:23 genome.fa
-rw-rw-r-- 1 iu iu 2.2K 11月 30 16:55 headers.txt
-rw-rw-r-- 1 iu iu 566 11月 30 16:42 JSLAB3_code.txt
-rw-rw-r-- 1 iu iu 3 11月 30 17:08 tmp.txt
iu@bielinux[hoge] du -sh [ 5:34午後 ]
2.9M .
iu@bielinux[hoge] cd .. [ 5:34午後 ]
iu@bielinux[Desktop] pwd [ 5:34午後 ]
/home/iu/Desktop
iu@bielinux[Desktop] ls -lh [ 5:34午後 ]
total 12K
drwxr-xr-x 2 iu iu 4.0K 11月 17 13:44 Bio-Linux Documentation
drwxrwxr-x 2 iu iu 4.0K 11月 30 17:08 hoge
drwxrwxrwx 1 root root 0 11月 28 11:13 mac_share
drwxrwxr-x 2 iu iu 4.0K 11月 26 19:15 mac_share2
lrwxrwxrwx 1 iu iu 29 11月 17 13:44 Sample Data -> /usr/local/bioinf/sampledata/
iu@bielinux[Desktop] du -sh [ 5:34午後 ]
5.8M .
iu@bielinux[Desktop] [ 5:34午後 ]
```

W18-3: du -sh

①cdでカレントディレクトリをホームディレクトリとして②「du -sh」を実行した結果。ユーザiuは、現在のところトータルで56MBディスクを使用していることがわかる。見えている数値はヒトによって若干異なる。③のようにディレクトリ名を指定してディスク使用量を調べることができる

```
iu@bielinux[Desktop] pwd
/home/iu/Desktop
iu@bielinux[Desktop] ls
Bio-Linux Documentation hoge mac_share mac_share2 Sample Data
iu@bielinux[Desktop] cd
iu@bielinux[iu] pwd
/home/iu
iu@bielinux[iu] ls
Desktop Documents Downloads Music Pictures Public Templates Videos
iu@bielinux[iu] du -sh
56M .
iu@bielinux[iu] du -sh Desktop
5.8M Desktop
iu@bielinux[iu] du -sh Documents
4.0K Documents
iu@bielinux[iu] du -sh Downloads
4.0K Downloads
iu@bielinux[iu]
```

W18-4: rm -r

ディレクトリ削除例。①rmコマンド実行時に-rをつけると「ディレクトリ内のファイルも削除」となる。fオプションも同時につけることによって、ファイル数分の確認メッセージへの対処を回避している。ごみ箱行きではなく消滅なので実行には十分に注意。

```
iu@bielinux[iu] cd Desktop [ 6:04午後]
iu@bielinux[Desktop] pwd [ 6:04午後]
/home/iu/Desktop
iu@bielinux[Desktop] ls [ 6:04午後]
Bio-Linux Documentation hoge mac_share mac_share2 Sample Data
iu@bielinux[Desktop] du -sh [ 6:04午後]
5.8M .
iu@bielinux[Desktop] du -sh mac_share2 [ 6:04午後]
4.0K mac_share2
iu@bielinux[Desktop] ls mac_share2 [ 6:04午後]
iu@bielinux[Desktop] rm -rf mac_share2 [ 6:04午後]
iu@bielinux[Desktop] ls [ 6:04午後]
Bio-Linux Documentation hoge mac_share Sample Data
iu@bielinux[Desktop] du -sh [ 6:04午後]
5.7M .
iu@bielinux[Desktop] [ 6:04午後]
```

①

W19-1:パイプ

```
iu@bielinux[Desktop] cd hoge [10:07午前]
iu@bielinux[hoge] pwd [10:07午前]
/home/iu/Desktop/hoge
iu@bielinux[hoge] ls [10:07午前]
genome.fa headers.txt JSLAB3_code.txt tmp.txt
① iu@bielinux[hoge] grep ">" genome.fa | wc [10:07午前]
    28     84    2218
② iu@bielinux[hoge] grep ">" genome.fa | wc -l [10:07午前]
28
③ iu@bielinux[hoge] grep ">" genome.fa | head -n 3 [10:07午前]
>ctg09_00003 dna:supercontig supercontig:GCA_000309565.1:ctg09_00003:1:899:1
>ctg03_00032 dna:supercontig supercontig:GCA_000309565.1:ctg03_00032:1:2417:1
>ctg10_00023 dna:supercontig supercontig:GCA_000309565.1:ctg10_00023:1:4695:1
iu@bielinux[hoge] grep ">" genome.fa | tail -n 2 [10:07午前]
④ >ctg12_00024 dna:supercontig supercontig:GCA_000309565.1:ctg12_00024:1:465659:1
>ctg28_00010 dna:supercontig supercontig:GCA_000309565.1:ctg28_00010:1:472701:1
iu@bielinux[hoge] █ [10:07午前]
```

W19-2: パイプ

パイプ利用例2。①赤枠の結果が示すように、パイプ(!)の両側のスペースはあってもなくてもいいようだ。②grepの結果をgrepすることもできる。右側の2回目のgrepでは“_0002”を含む行を出力している。

```
iu@bielinux[hoge] pwd
/home/iu/Desktop/hoge
iu@bielinux[hoge] ls
genome.fa headers.txt JSLAB3 code.txt tmp.txt
iu@bielinux[hoge] grep ">" genome.fa | wc -l
28
iu@bielinux[hoge] grep ">" genome.fa |wc -l
28
iu@bielinux[hoge] grep ">" genome.fa|wc -l
28
iu@bielinux[hoge] grep ">" genome.fa | grep "_0002"
>ctg10_00023 dna:supercontig supercontig:GCA_000309565.1:ctg10_00023:1:4695:1
>ctg25_00029 dna:supercontig supercontig:GCA_000309565.1:ctg25_00029:1:16183:1
>ctg21_00025 dna:supercontig supercontig:GCA_000309565.1:ctg21_00025:1:25877:1
>ctg18_00021 dna:supercontig supercontig:GCA_000309565.1:ctg18_00021:1:64885:1
>ctg06_00026 dna:supercontig supercontig:GCA_000309565.1:ctg06_00026:1:109464:1
>ctg27_00027 dna:supercontig supercontig:GCA_000309565.1:ctg27_00027:1:222389:1
>ctg01_00022 dna:supercontig supercontig:GCA_000309565.1:ctg01_00022:1:237614:1
>ctg12_00024 dna:supercontig supercontig:GCA_000309565.1:ctg12_00024:1:465659:1
iu@bielinux[hoge]
```



W19-3: パイプ

パイプ利用例3。①3つ以上のコマンド同士をパイプで連結することも可能。②headとtailを連結することで、コマンド単体では概観しづらい箇所も眺めることができる。

```
iu@bielinux[hoge] grep ">" genome.fa | grep "_0002" [10:25午前]
>ctg10_00023 dna:supercontig supercontig:GCA_000309565.1:ctg10_00023:1:4695:1
>ctg25_00029 dna:supercontig supercontig:GCA_000309565.1:ctg25_00029:1:16183:1
>ctg21_00025 dna:supercontig supercontig:GCA_000309565.1:ctg21_00025:1:25877:1
>ctg18_00021 dna:supercontig supercontig:GCA_000309565.1:ctg18_00021:1:64885:1
>ctg06_00026 dna:supercontig supercontig:GCA_000309565.1:ctg06_00026:1:109464:1
>ctg27_00027 dna:supercontig supercontig:GCA_000309565.1:ctg27_00027:1:222389:1
>ctg01_00022 dna:supercontig supercontig:GCA_000309565.1:ctg01_00022:1:237614:1
>ctg12_00024 dna:supercontig supercontig:GCA_000309565.1:ctg12_00024:1:465659:1
iu@bielinux[hoge] grep ">" genome.fa | grep -c "_0002" [10:25午前]
8
iu@bielinux[hoge] grep ">" genome.fa | grep "_0002" | wc [10:25午前]
      8      24     635
iu@bielinux[hoge] grep ">" genome.fa | grep "_0002" | wc -l [10:25午前]
8
iu@bielinux[hoge] head -n 10 genome.fa | tail -n 10 [10:25午前]
GTTTTGTATGCTTTGACCCAAATTCACCGCTGACAGCAACAGTGCAAAGACTACCATCAC
CCGCAACAAAAC TTTATGTCGCGATCTTACCAGTTTCAATAAACGATCTTCACCCATTTA
TTTCACCTCCGCGTTTTTGGCGGAACTCAACTTTAAATAAATCAGATAGCGATAACGG
CAAGCTCTCATT CAGCACGGGTTCTTCGGCGGCAATCGCGGCCTTGACTTCAGGTGTGAA
ATGCGGGAAAAC CGCTTCGATCACGCGTCCACGAATGGTCAGAATCCGACCTTGTTCTT
GATCACATCAGGAATCTCCTTGGTTTTGAAAATAACTGCAGCTTGATCGCCTGCTGTCT
GACGTCCTCTAGCGTCAAGTCCTGACTGATCCGACCGGCTTTAACGAACAAAACGCGATC
TGACAAGCCGTCTAATTCGACTAGATTATGTGAGGCGATGAGAAAGGTTGTTTTACCATC
AGCAACAGCATCTACCACCAAATTGACAATCCGCTCGCGAATCAGCAAATCCAACCCATC
AAATGGTTCATCCAGAAAAACATACGGTGCCCCGCTCGCCAGCGCCAACGCCACCATCAC
iu@bielinux[hoge]
```

W20-1 : SRA

The screenshot shows the NCBI SRA website. The search bar is at the top, with a dropdown menu set to 'SRA' and the search term 'Lactobacillus casei' entered. A red arrow labeled ① points to the search input field. To the right of the input field is a blue 'Search' button, with a red arrow labeled ② pointing to it. Below the search bar, there is a banner for SRA with the text: 'Sequence Read Archive (SRA) makes biological sequence data available to the research community to enhance reproducibility and allow for new discoveries by comparing data sets. The SRA stores raw sequencing data and alignment information from high-throughput sequencing platforms, including Roche 454 GS System®, Illumina Genome Analyzer®, Applied Biosystems SOLiD System®, Helicos Heliscope®, Complete Genomics®, and Pacific Biosciences SMRT®'. Below the banner, there are three columns of links: 'Getting Started' (How to Submit, Login to SRA, Login to Submission Portal, SRA Handbook, Download Guide, SRA Fact Sheet (.pdf)), 'Tools and Software' (Download SRA Toolkit, SRA Toolkit Documentation, SRA-BLAST, SRA Run Browser, SRA Run Selector), and 'Related Resources' (Submission Portal, Trace Archive, dbGaP Home, BioProject, BioSample).

W20-1 : SRA

①33件ヒットしたようだ(2015年12月1日現在)。②ざっと眺める

NCBI Resources How To Sign in to NCBI

SRA SRA Lactobacillus casei Search

Summary 20 per page Send to: Filters: Manage Filters

Public (33)

Source DNA (29) RNA (4) metagenomic (4)

Type genome (25)

Other aligned data (19)

Clear all Show additional filters

Search results

Items: 1 to 20 of 33

1. [Illumina HiSeq 2000 paired end sequencing](#)
1 ILLUMINA (Illumina HiSeq 2000) run: 2.5M spots, 507.9M bases, 263Mb downloads
Accession: ERX529187

2. [Illumina HiSeq 2000 paired end sequencing](#)
1 ILLUMINA (Illumina HiSeq 2000) run: 1.5M spots, 301.1M bases, 149.7Mb downloads
Accession: ERX450909

3. [Illumina HiSeq 2000 paired end sequencing](#)
1 ILLUMINA (Illumina HiSeq 2000) run: 1.7M spots, 348.6M bases, 171.5Mb downloads
Accession: ERX450901

4. [Illumina HiSeq 2000 paired end sequencing](#)
1 ILLUMINA (Illumina HiSeq 2000) run: 1.7M spots, 340M bases, 168.6Mb downloads
Accession: ERX450893

Results by taxon

Top Organisms [Tree]
Lactobacillus casei (29)
synthetic metagenome (4)

Search in related databases

Database	Access		all
	public	controlled	
BioSample	25		25
BioProject	11		11
dbGaP			
GEO Datasets			

Find related data

Database: Select

W20-1 : SRA

①ERXや②SRX単位での出力であることがわかる。ERPやDRP単位ではないのでわかりづらいが、タイトルの文字列から「赤枠の2つは同じデータセットに含まれるのだろう」という程度の予想はつく

http://www.ncbi.nlm.nih.gov/sra/?term=Lactobacillus+casei

Lactobacillus casei - SR...

Search details

"Lactobacillus casei" [Organism] OR Lactobacillus casei [All Fields]

Search See more...

Recent activity

Turn Off Clear

Lactobacillus casei (33) SRA

See more...

- [Illumina HiSeq 2000 paired end sequencing](#)
6. 1 ILLUMINA (Illumina HiSeq 2000) run: 1.7M spots, 333.9M bases, 170.7Mb downloads
Accession: ERX450875
- [Illumina HiSeq 2000 paired end sequencing](#)
7. 1 ILLUMINA (Illumina HiSeq 2000) run: 1.7M spots, 332.8M bases, 164.3Mb downloads
Accession: ERX450870
- [Whole genome seq of **Lactobacillus casei** clinical isolate: Sample](#)
8. [867 LCAS](#)
1 ILLUMINA (Illumina HiSeq 2000) run: 632,555 spots, 119.5M bases, 80Mb downloads
Accession: SRX762773
- [Whole genome seq of **Lactobacillus casei** clinical isolate: Sample](#)
9. [844 LCAS](#)
1 ILLUMINA (Illumina HiSeq 2000) run: 450,027 spots, 83.8M bases, 59.4Mb downloads
Accession: SRX762750
- [Population Genomics of **Lactobacillus casei**](#)
10. 1 ILLUMINA (Illumina HiSeq 2000) run: 5.5M spots, 1.1G bases, 739.1Mb downloads
Accession: ERX358287
- [Population Genomics of **Lactobacillus casei**](#)
11. 1 ILLUMINA (Illumina HiSeq 2000) run: 5.8M spots, 1.2G bases, 761.4Mb downloads
Accession: ERX358286
- [Whole genome analysis of **Lactobacillus casei** JCM 1134](#)

W20-2: ENA

The screenshot shows the ENA website in a browser window. The address bar displays 'http://www.ebi.ac.uk/ena'. A cookie consent banner is visible at the top. The main navigation bar includes 'EMBL-EBI', 'Services', 'Research', 'Training', and 'About us'. The ENA logo and 'European Nucleotide Archive' text are prominently displayed. A search bar contains the text 'Lactobacillus casei', with a red arrow labeled '1' pointing to it. Below the search bar, examples 'BN000065, histone' are listed. To the right of the search bar is a 'Search' button, with a red arrow labeled '2' pointing to it. Below the search bar, there are links for 'Advanced' and 'Sequence'. A secondary navigation bar includes 'Home', 'Search & Browse', 'Submit & Update', 'Software', 'About ENA', and 'Support'. The main content area features the title 'European Nucleotide Archive' and a descriptive paragraph: 'The European Nucleotide Archive (ENA) provides a comprehensive record of the world's nucleotide sequencing information, covering raw sequencing data, sequence assembly information and functional annotation. More about ENA'. Below this, it states: 'Access to ENA data is provided through the browser, through search tools, large scale file download and through the API.' There is a 'Text Search' section with an input field and examples 'BN000065, histone'. On the right side, there is a 'Popular' section with a list of links: 'Submit and update', 'Sequence submissions', 'Genome assembly submissions', 'Submitting environmental sequences', 'Citing ENA data', 'Rest URLs for data retrieval', and 'Rest URLs to search ENA'. At the bottom right, there is a 'Latest ENA news' section.

W20-2: ENA

The screenshot shows a web browser window displaying the ENA search results for 'Lactobacillus casei'. The browser address bar shows the URL: <http://www.ebi.ac.uk/ena/data/search?query=Lactobacillus+casei>. The ENA logo and 'European Nucleotide Archive' text are visible. The search bar contains 'Lactobacillus casei' and a 'Search' button. Below the search bar, there are navigation links: Home, Search & Browse, Submit & Update, About ENA, and Support. A blue banner contains a message: 'Please subscribe to ena-announce mailing list here: listserv.ebi.ac.uk/mailman/listin... to receive alerts about ENA services.' The search results are displayed under the heading 'Search results for *Lactobacillus casei*'. There are two main result sections: 'Assembly (40 results found)' and 'Sequence (Update) (1,663 results found)'. The first result in the Assembly section is 'GCA_000026485.1 ASM2648v1 assembly for Lactobacillus casei BL23'. The first result in the Sequence section is 'JN315659 Lactobacillus casei strain Lac3 16S ribosomal RNA gene, partial sequence.' A red arrow labeled '1' points to the bottom of the page, indicating the location of the search results.

W20-2: ENA

DRP, ERP, SRPのように、3文字目がPになっているIDをクリックすると全体像が分かりやすいのでよい。例えば①DRP000852をクリックして眺めると、②DRXや③DRRは、①DRPの下の階層のIDであることがわかる。

The screenshot shows a web browser window with the URL <http://www.ebi.ac.uk/ena/data/search?query=Lactobacillus+casei>. The search results are categorized into several sections:

- Experiment (12 results found)**:
 - ② DRX002585 Ion Torrent PGM sequencing; Whole genome analysis of *Lactobacillus casei* JCM 1134
View all 12 results
- Run (8 results found)**:
 - ③ DRR003254 Ion Torrent PGM sequencing; Whole genome analysis of *Lactobacillus casei* JCM 1134
View all 8 results
- Study (12 results found)**:
 - ① DRP000852 Genome sequencing of *Lactobacillus casei* JCM 1134
View all 12 results
- Study (Sequence) (59 results found)**:
 - PRJNA165565 *Lactobacillus casei* UW4 Genome sequencing
View all 59 results
- Taxon (37 results found)**

W20-2: ENA

DRP, ERP, SRPのように、3文字目がPになっているIDをクリックすると全体像が分かりやすいのでよい。例えば①DRP000852をクリックして眺めると、②DRXや③DRRは、①DRPの下の階層のIDであることがわかる。

Navigation Read Files

Download files

Download: 1 - 1 of 1 results in TEXT

Select columns

Showing results 1 - 1 of 1 results

Study accession	Secondary study accession	Sample accession	Secondary sample accession	Experiment accession	Run accession	Tax ID	Scientific name	Instrument model	Library layout	Fastq files (ftp)	Fastq files (galaxy)
PRJDB638	DRP000852	SAMD00015724	DRS002689	DRX002585	DRR003254	1423732	Lactobacillus casei DSM 20011 = JCM 1134	Ion Torrent PGM	SINGLE	File 1	File 1

Prev Next Go to page: 1 Go

For Aspera download, please download and install Aspera Connect

W20-3: DRA

http://trace.ddbj.nig.ac.jp/dra/index.html

DDBJ
DNA Data Bank of Japan

Login & Submit | Databases ▾ | English | Contact

Google™ カスタム検索

Home | Handbook | FAQ | Search | Download ▾ | Pipeline | About DRA

News

2015年12月01日 **New:** スパコン停止に伴う各種サービスの停止 more...

DDBJ Sequence Read Archive (DRA) は Roche 454 GS System®, Illumina Genome Analyzer®, Applied Biosystems SOLiD® System などの次世代シーケンサからの出力データのためのデータベースです。DRA は International Nucleotide Sequence Database Collaboration (INSDC) のメンバーであり、NCBI Sequence Read Archive (SRA) と EBI Sequence Read Archive (ERA) との国際協力のもと、運営されています。従来のキャピラリー式シーケンサからの出力データは DDBJ Trace Archive にご登録ください。

検索

登録

W20-3: DRA

DRASearch Send Feedback [Search Home](#) [DRA Home](#)

Accession :

Organism : StudyType :

CenterName : Platform :

Keyword :

Show records Sort by

Data Last Update 2015-12-01

Statistics

Released Entries

Type	Count
Submission	458071
Study	61286
Experiment	1402638
Sample	1313037
Run	1682069

Organism			Study Type		
#	Organism Name	Study	#	Study Type	Study

W20-3: DRA

DRASearch Send Feedback [Search Home](#) [DRA Home](#)

Accession :

Organism : StudyType :

CenterName : Platform :

Keyword :

Show records Sort by

Search Results (3 studies) ① << < 1 / 1 Page > >>

#	STUDY	SUBMISSION	STUDY_TITLE	STUDY_TYPE	ORGANISM	BAS
					Atopobium minutum	
					Atopobium rimae	
					Carnobacterium divergens	
					Carnobacterium maltaromaticum	
					Fructobacillus fructosus	
					Kandleria vitulina	
					Lactobacillus acetotolerans	
					Lactobacillus acidifarinae	
					Lactobacillus	

W20-4: 検索場所に注意

①Organismではなく②Keywordのところ
で③検索すると、④ヒット数は1251件に
激増。→検索する場所に気を付けよう。

DRASearch

Accession :

Organism : StudyType :

CenterName : Platform :

Keyword :

Show records Sort by Search Clear

Search Results (1251 records) << < 1 / 63 Page > >>

Filtered by
document type:sample(675) study(291) experiment(185) submission(78) run(22)
organism:Lactobacillus gasseri(72) Lactobacillus fermentum(65) Lactobacillus rhamnosus(58)
Lactobacillus sp.(49) Lactobacillus plantarum(32) Lactobacillus salivarius(28)

#	META_FILE	ACCESSION	STUDY	STUDY_TITLE	STUDY_TYPE	ORGANISM	BASES	SUBMITTED	CENTER_NAME
1	SRA311747.sample.xml org/2001/XMLSchema-instance"> <SAMPLE alias="Lactobacillus casei BM- LC14617" accession="SRS1164278"> <IDENTIFIERS> <PRIMARY	SRS1164278							
2	SRA148846.submission.xml <?xml version="1.0" encoding="UTF-8"? ><SUBMISSION alias="The	SRA148846	SRP040478	16S rRNA gene sequencing (human GM)	Other	Homo sapiens	898.1M		BioProject

W21-1 : L. casei 12A

Organismのところすでに気づいているかもしれないが、①「Lactobacillus casei」以降の候補が表示されるので、そこで目的の②L. casei 12Aを選択するのがより直接的。

DRASearch

Accession :

Organism : **①** Lactobacillus casei x StudyType :

CenterName : Lactobacillus casei

Keyword : Lactobacillus casei 12A **②**

Lactobacillus casei A2-362

Lactobacillus casei DSM 20011 = JCM 1134

Clear

Data Last Update 2015-12-01

Statistics

Released Entries

Type	Count
Submission	458071
Study	61286
Experiment	1402638
Sample	1313037
Run	1682069

Organism		Study Type		Center Name				
#	Organism Name	Study	#	Study Type	Study	#	Center Name	Study

W21-1 : L. casei 12A

①検索結果は、②1件のヒット。
③SRP017156をクリック。

The screenshot shows the DRA Search interface. The search criteria are: Organism: Lactobacillus casei 12A, StudyType: (empty), CenterName: (empty), Platform: (empty), Keyword: (empty). The search results show 1 study. The search button is labeled ①, the search results count is labeled ②, and the study ID SRP017156 is labeled ③.

DRASearch Send Feedback [Search Home](#) [DRA Home](#)

Accession :

Organism : StudyType :

CenterName : Platform :

Keyword :

Show records Sort by

Search Results (1 studies) / 1 Page / 1 Page / 1 Page

#	STUDY	SUBMISSION	STUDY_TITLE	STUDY_TYPE	ORGANISM	BASES	SUBMITTED	CENTER_NAME
1	SRP017156	SRA061483	Lactobacillus casei 12A Project	Transcriptome Analysis	Lactobacillus casei 12A	53.9G	2012-11-12	JGI

Website policy | © DNA Data Bank of Japan

W21-1:L. casei 12A

実際にダウンロードするのは①SRRから始まるIDのファイル。DRAは、FASTQ形式のbzip2圧縮ファイルとsra形式ファイルの両方を提供している。②SRXから始まるIDを眺めることで、用いた実験機器やpaired or single-endなどの情報が分かる。②をクリック。

http://trace.ddbj.nig.ac.jp/DRAsearch/study?acc=SRP017156

DRASearch

SRP017156

Study Detail

Title	Lactobacillus casei 12A Project
Study Type	Transcriptome Analysis
Abstract	none provided
Description	none provided
Center Name	JGI

Navigation

Submission SRA061483	FTP
Experiment SRX204226	FASTQ S
SRP017156	FASTQ S

Website policy | © DNA Data Bank of Japan

W21-1:L. casei 12A

SRX204226の実験デザインで得られたデータ (Run)は、①SRR616268であり、②pairedデータであることがわかる。FASTQ形式でダウンロードする際に、*_1.fastq.bz2と*_2.fastq.bz2の2つに分かれている事実からも認識可能。公共データをダウンロードする際には③のところの記述内容などを眺めて、RNA鎖の方向性を考慮(stranded)しているかどうかをちゃんと認識しておく必要がある。④をクリック。

DRASearch

SRX204226 FASTQ SRA

Experiment Detail

Title	Lactobacillus casei 12A library
Design Description	Illumina Stranded Std PE (amplified) ③
Organism	Lactobacillus casei 12A

Library Description

Name	COHG
Strategy	EST
Source	TRANSCRIPTOMIC
Selection	cDNA
Layout	PAIRED ②
Orientation	
Nominal Length	300
Nominal Sdev	
Construction Protocol	

Platform

Submission	SRA061483	FTP
Study	SRP017156	
Sample	SRS375150	
Run	SRR616268	FASTQ SRA

① ↑ ④ ↑

W21-2: URL情報取得

ダウンロードしたいFASTQ形式のbzip2圧縮ファイルは、赤枠の2つ。それぞれのファイル上で、右クリックの「ショートカットのコピー」でURL情報を取得可能。W22-2でペーストする

FTP ディレクトリ /ddbj_database/dra/fastq/SRA061/SRA061483/SRX204226 /
ftp.ddbj.nig.ac.jp

FTP ディレクトリ /ddbj_database/dra/fastq/SRA061/SRA061483/SRX204226 /
ftp.ddbj.nig.ac.jp

エクスプローラーでこの FTP サイトを表示するには、Alt キーを押して、[表示]をクリックし、[エクスプローラーで FTP サイトを開く]をクリックしてください。

-Welcome to DDBJ FTP Archive, running on ftp.ddbj.nig.ac.jp!
-
-Please contact ddbj@ddbj.nig.ac.jp when you have any problem for getting
-access to this archive, downloading the data, and etc.
-
-For details on the directory structure and file contents, please refer
-to the README.TXT placed in the "ddbj_database".

[1 階層上のディレクトリへ](#)

08/06/2015 02:58午前	14,575,885	SRR616268.fastq.bz2
08/06/2015 02:58午前	7,662,128,101	SRR616268.1.fastq.bz2
08/06/2015 02:58午前	7,017,031,734	SRR616268.2.fastq.bz2

- 開く(O)
- 新しいタブで開く(W)
- 新しいウィンドウで開く(N)
- 対象をファイルに保存(A)...
- 対象を印刷(P)
- 切り取り
- コピー(C)
- ショートカットのコピー(T)
- 貼り付け(P)

W21-3: ENA

ENAにて①「Lactobacillus casei 12A」
のデータを②検索。

The screenshot shows the ENA website interface. At the top, there is a navigation bar with 'EMBL-EBI' logo and links for 'Services', 'Research', 'Training', and 'About us'. Below this is a large teal banner with the ENA logo and a search bar. The search bar contains the text 'Lactobacillus casei 12A' and has a red arrow with the number '1' pointing to it. To the right of the search bar is a 'Search' button with a red arrow and the number '2' pointing to it. Below the search bar, there are links for 'Advanced' and 'Sequence'. The main content area has a 'Home' button and a 'Text Search' input field. On the right side, there is a 'Popular' section with a list of links: 'Submit and update', 'Sequence submissions', 'Genome assembly submissions', 'Submitting environmental sequences', 'Citing ENA data', 'Rest URLs for data retrieval', and 'Rest URLs to search ENA'. At the bottom right, there is a 'Latest ENA news' section.

W21-3: ENA

「Lactobacillus casei 12A」検索結果。
①ページ下部に移動するとDRAで見
たものと同じIDを発見できる。

The screenshot shows a web browser window with the URL <http://www.ebi.ac.uk/ena/data/search?query=Lactobacillus+casei+12A>. The page header includes the EMBL-EBI logo and navigation links for Services, Research, Training, and About us. The main content area features the ENA logo and a search bar containing "Lactobacillus casei 12A". Below the search bar, there are links for "Advanced Sequence" and "Search". A red arrow with the number 1 points to the bottom of the page, indicating where to scroll to find the same IDs as in the DRA.

Search results for *Lactobacillus casei* 12A

Show more data from EMBL-EBI

Assembly
Assembly (3)

Sequence
Sequence (Release) (406)
Assembly contig set (1)

Coding

Assembly (3 results found)

GCA_000309565.1 ASM30956v1 assembly for Lactobacillus casei 12A
View all 3 results

Sequence (Release) (406 results found)

W21-3: ENA

SRP, SRA, SRX, SRRなど沢山のIDが存在するが、①SRPというのが一番の大元に近いID。

The screenshot shows a web browser window with the URL <http://www.ebi.ac.uk/ena/data/search?query=Lactobacillus+casei+12A>. The search results are categorized into several sections:

- Experiment (2 results found)**
 - SRX204226 Illumina HiSeq 2000 paired end sequencing; Lactobacillus casei 12A library
 - View all 2 results
- Run (2 results found)**
 - SRR616268 Illumina HiSeq 2000 paired end sequencing; Lactobacillus casei 12A library
 - View all 2 results
- Study (1 results found)**
 - ①** SRP017156 Lactobacillus casei 12A Project
 - View all 1 results
- Study (Sequence) (2 results found)**
 - PRJNA217366 Lactobacillus casei 12A Genome sequencing
 - View all 2 results
- Taxon (1 results found)**
 - 1051650 Lactobacillus casei 12A

W21-3: ENA

The screenshot shows a web browser window with the URL <http://www.ebi.ac.uk/ena/data/view/SRP017156>. The page features a search bar with a 'Search' button and a red arrow labeled '1' pointing to it. The search bar contains the text 'Examples: BN000085, histone'. Below the search bar is a navigation menu with 'Home', 'Search & Browse', 'Submit & Update', 'About ENA', and 'Support'. A blue banner below the navigation menu contains the text: 'Please subscribe to ena-announce mailing list here: listserv.ebi.ac.uk/mailman/listin... to receive alerts about ENA services.'

Study: SRP017156
Lactobacillus casei 12A Project

View: [XML](#) [Send Feedback](#)  [Download: XML](#)

Submitting Centre	Study Type	Read Count	Base Count
JGI	Transcriptome Analysis	269,298,092	53,859,618,400

Broker Name

W21-3: ENA

赤枠で示すように、SRP, SRA, SRX, SRRなど様々な種類のIDの対応関係を俯瞰できるのがENAの長所。

Navigation Read Files

Download files

Download: 1 - 2 of 2 results in TEXT

Select columns

Showing results 1 - 2 of 2 results

Study accession	Secondary study accession	Sample accession	Secondary sample accession	Experiment accession	Run accession	Tax ID	Scientific name	Instrument model	Library layout	Fastq files (ftp)
SRP017156	SRP017156	SAMN01813467	SRS375150	SRX204226	SRR616268	1051650	Lactobacillus casei 12A	Illumina HiSeq 2000	PAIRED	File 1 File 2
SRP017156	SRP017156	SAMN01813467	SRS375150	SRX204227	SRR616269	1051650	Lactobacillus casei 12A	Illumina HiSeq 2000	PAIRED	File 1 File 2

Prev Next Go to page: 1 Go

For Aspera download, please download and install Aspera Connect

W21-3: ENA

ENAは、①FASTQ形式の②gzip圧縮ファイルとして提供している。

Navigation Read Files

Download files

Download: 1 - 2 of 2 results in TEXT

Select columns

Showing results 1 - 2 of 2 results

Study accession	Secondary study accession	Sample accession	Secondary sample accession	Experiment accession	Run accession	Tax ID	Scientific name	Instrument model	Library layout	Fastq files (ftp)
SRP017156	SRP017156	SAMN01813467	SRS375150	SRX204226	SRR616268	1051650	Lactobacillus casei 12A	ILLUMINA HiSeq 2000	PAIRED	File 1 File 2
SRP017156	SRP017156	SAMN01813467	SRS375150	SRX204227	SRR616269	1051650	Lactobacillus casei 12A	ILLUMINA HiSeq 2000	PAIRED	File 1 File 2

Prev Next Go to page: 1 Go

For Aspera download, please download and install Aspera Connect

ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR616/SRR616268/SRR616268_1.fastq.gz

W22-1:ダウンロード準備

①srp017156ディレクトリの作成。ここでは②/home/iu/Documents/srp017156で作業を行う。

```
iu@bielinux[hoge] cd [10:10午後]
iu@bielinux[iu] pwd [10:10午後]
/home/iu
iu@bielinux[iu] ls [10:10午後]
Desktop Documents Downloads Music Pictures Public Templates Videos
iu@bielinux[iu] cd Documents [10:10午後]
iu@bielinux[Documents] ls [10:10午後]
iu@bielinux[Documents] pwd [10:10午後]
/home/iu/Documents
① iu@bielinux[Documents] mkdir srp017156 [10:10午後]
iu@bielinux[Documents] ls -la [10:10午後]
total 12
drwxr-xr-x 3 iu iu 4096 12月 1 22:10 .
drwxr-xr-x 19 iu iu 4096 11月 30 17:34 ..
drwxrwxr-x 2 iu iu 4096 12月 1 22:10 srp017156
iu@bielinux[Documents] cd srp017156 [10:10午後]
② iu@bielinux[srp017156] pwd [10:10午後]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] [10:10午後]
```

W22-2: wget

オリジナル版は、計約15GBのpaired-endファイルでしたが、デカすぎるので全リード(約1.35億)ではなく、ダウンロードの段階から最初の150万リードのみにする方法に変更。①赤下線部分が実際のコマンド。1リード4行分なので、実際に行うのは6000000行分の抽出。②etaは到着予定時刻(ここではダウンロード終了予定時刻)の意味

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls
iu@bielinux[srp017156] wget -c ftp://ftp.ddbj.nig.ac.jp/ddbj_database/dra/fastq/SRA061/SRA061483/SRX204226/SRR616268_1.fastq.bz2 -O - | bzip2 -dc | head -n 600000
0 > SRR616268_1.fastq
--2015-12-03 14:55:50-- ftp://ftp.ddbj.nig.ac.jp/ddbj_database/dra/fastq/SRA061/SRA061483/SRX204226/SRR616268_1.fastq.bz2
=> '-'
Resolving ftp.ddbj.nig.ac.jp (ftp.ddbj.nig.ac.jp)... 133.39.224.12
Connecting to ftp.ddbj.nig.ac.jp (ftp.ddbj.nig.ac.jp)|133.39.224.12|:21... connected.
Logging in as anonymous ... Logged in!
==> SYST ... done.      ==> PWD ... done.
==> TYPE I ... done.   ==> CWD (1) /ddbj_database/dra/fastq/SRA061/SRA061483/SRX204226 ... done.
==> SIZE SRR616268_1.fastq.bz2 ... 7662128101
==> PASV ... done.     ==> RETR SRR616268_1.fastq.bz2 ... done.
Length: 7662128101 (7.1G) (unauthoritative)

1% [                               ] 81,536,852  3.33MB/s  eta 44m 45s
```



W22-2: wget

①私の(東大有線LAN)環境では31秒で終了。
②「Broken pipe, closing control connection」というヤバそう?!なメッセージが出ているが、結論としては問題ない。高度ではあるが、一応③のwgetコマンドで行っていることを説明すると...

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls
iu@bielinux[srp017156] wget -c ftp://ftp.ddbj.nig.ac.jp/ddbj_database/dra/fastq/SRA061/SRA061483/SRX204226/SRR616268_1.fastq.bz2 -O - | bzip2 -dc | head -n 600000
0 > SRR616268_1.fastq
--2015-12-03 14:55:50-- ftp://ftp.ddbj.nig.ac.jp/ddbj_database/dra/fastq/SRA061/SRA061483/SRX204226/SRR616268_1.fastq.bz2
=> '-'
Resolving ftp.ddbj.nig.ac.jp (ftp.ddbj.nig.ac.jp)... 133.39.224.12
Connecting to ftp.ddbj.nig.ac.jp (ftp.ddbj.nig.ac.jp)|133.39.224.12|:21... connected.
Logging in as anonymous ... Logged in!
==> SYST ... done.      ==> PWD ... done.
==> TYPE I ... done.    ==> CWD (1) /ddbj_database/dra/fastq/SRA061/SRA061483/SRX204226 ... done.
==> SIZE SRR616268_1.fastq.bz2 ... 7662128101
==> PASV ... done.     ==> RETR SRR616268_1.fastq.bz2 ... done.
Length: 7662128101 (7.1G) (unauthoritative)

1% [          ] 87,107,412  3.09MB/s  in 31s
SRR616268_1.fastq.bz2: Broken pipe, closing control connection.
iu@bielinux[srp017156]
```

[2:55午後]

[2:56午後]



W22-2: wget

① wgetで取得するデータをファイルに保存せずに(「-O -」に相当)そのままパイプで流し、②元ファイルを残した状態(-c)で解凍(-d)し、③最初の6000000行分(150万リード分)に達した段階で、④SRR616268_1.fastqというファイル名で保存。(bzip2圧縮はされていない状態なのでbz2はつけない)

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls
iu@bielinux[srp017156] wget -c ftp://ftp.ddbj.nig.ac.jp/ddbj_database/dra/fastq/S
RA061/SRA061483/SRX204226/SRR616268_1.fastq.bz2 -O - | bzip2 -dc | head -n 600000
0 > SRR616268_1.fastq
--2015-12-09 14:55:50-- ftp://ftp.ddbj.nig.ac.jp/ddbj_database/dra/fastq/SRA061/
SRA061483/SRX204226/SRR616268_1.fastq.bz2
=> '-'
Resolving ftp.ddbj.nig.ac.jp (ftp.ddbj.nig.ac.jp)... 133.39.224.12
Connecting to ftp.ddbj.nig.ac.jp (ftp.ddbj.nig.ac.jp)|133.39.224.12|:21... connec
ted.
Logging in as anonymous ... Logged in!
==> SYST ... done.      ==> PWD ... done.
==> TYPE I ... done.    ==> CWD (1) /ddbj_database/dra/fastq/SRA061/SRA061483/SRX20
4226 ... done.
==> SIZE SRR616268_1.fastq.bz2 ... 7662128101
==> PASV ... done.     ==> RETR SRR616268_1.fastq.bz2 ... done.
Length: 7662128101 (7.1G) (unauthoritative)

1% [                               ] 87,107,412  3.09MB/s  in 31s

SRR616268_1.fastq.bz2: Broken pipe, closing control connection.
iu@bielinux[srp017156]
```

[2:55午後]

[2:56午後]



W22-2: 確認

①lsで確認。確かにファイルは存在する。②非圧縮状態でも500MB以下なのでお手頃。最初の6000000行分のみダウンロードしているので、ファイルサイズを頼りに正しくダウンロードできているかを確認することはできないため、③wcコマンドでファイルの行数が正しく6000000行になっているかどうかは確認する。

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls
SRR616268_1.fastq
iu@bielinux[srp017156] ls -l
total 471908
-rw-rw-r-- 1 iu iu 483233202 12月  3 14:56 SRR616268_1.fastq
iu@bielinux[srp017156] ls -lh
total 461M
-rw-rw-r-- 1 iu iu 461M 12月  3 14:56 SRR616268_1.fastq
iu@bielinux[srp017156] wc SRR616268_1.fastq
 6000000 12000000 483233202 SRR616268_1.fastq
iu@bielinux[srp017156] █
```

[3:35午後]
[3:35午後]
[3:35午後]
[3:35午後]

(wgetを除く)他の一般的なコマンドは実行時間は表示されない。実行時間を計測したい場合は、①コマンドの左側(一番最初)にtimeをつける。②実行時間は約24.2秒

W22-2: もう片方も

```
iu@bielinux[srp017156] time wget -c ftp://ftp.ddbj.nig.ac.jp/ddbj_database/dra/fastq/SRA061/SRA061483/SRX204226/SRR616268_2.fastq.bz2 -O - | bzip2 -dc | head -n 6000000 > SRR616268_2.fastq
--2015-12-03 15:50:55-- ftp://ftp.ddbj.nig.ac.jp/ddbj_database/dra/fastq/SRA061/SRA061483/SRX204226/SRR616268_2.fastq.bz2
=> '-'
Resolving ftp.ddbj.nig.ac.jp (ftp.ddbj.nig.ac.jp)... 133.39.224.12
Connecting to ftp.ddbj.nig.ac.jp (ftp.ddbj.nig.ac.jp)|133.39.224.12|:21... connected.
Logging in as anonymous ... Logged in!
==> SYST ... done.      ==> PWD ... done.
==> TYPE I ... done.   ==> CWD (1) /ddbj_database/dra/fastq/SRA061/SRA061483/SRX204226 ... done.
==> SIZE SRR616268_2.fastq.bz2 ... 7017031734
==> PASV ... done.     ==> RETR SRR616268_2.fastq.bz2 ... done.
Length: 7017031734 (6.5G) (unauthoritative)

1% [          ] 76,952,004  3.47MB/s   in 24s

SRR616268 2.fastq.bz2: Broken pipe, closing control connection.
wget -c -O - 0.01s user 0.59s system 2% cpu 24.222 total
bzip2 -dc 19.13s user 0.62s system 81% cpu 24.222 total
head -n 6000000 > SRR616268_2.fastq 0.38s user 0.62s system 4% cpu 24.219 total
iu@bielinux[srp017156] [ 3:51午後]
```

W22-3: df -h

最初の150万リード(オリジナルの1/90程度)のみにしているおかげで、非圧縮状態でも2つのファイル合わせて1GB未満のディスク使用量であることがわかる。1.35億リードのオリジナルデータの場合は80GB程度になってしまうため、一つ一つのコマンド実行時間(やディスク使用量)が90倍弱かかることを想像するとよい。

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -l
total 899872
-rw-rw-r-- 1 iu iu 483233202 12月  3 14:56 SRR616268_1.fastq
-rw-rw-r-- 1 iu iu 438233202 12月  3 15:51 SRR616268_2.fastq
iu@bielinux[srp017156] ls -lh
total 879M
-rw-rw-r-- 1 iu iu 461M 12月  3 14:56 SRR616268_1.fastq
-rw-rw-r-- 1 iu iu 418M 12月  3 15:51 SRR616268_2.fastq
iu@bielinux[srp017156] df -h
df: 'mac_share': No such file or directory
Filesystem      Size  Used Avail Use% Mounted on
/dev/sda1        97G   11G   81G  12% /
none             4.0K    0 4.0K   0% /sys/fs/cgroup
udev            991M   4.0K 991M   1% /dev
tmpfs            201M  932K 200M   1% /run
none             5.0M    0 5.0M   0% /run/lock
none            1001M  152K 1001M   1% /run/shm
none             100M   44K  100M   1% /run/user
iu@bielinux[srp017156] █
```

[3:58午後]

[3:58午後]

[3:59午後]

W22-4: ディスク使用量

つまり、このpaired-endデータ(SRR616268)の場合、1.35億リードのオリジナルデータで2つのファイルともに解凍するには、①のところが100GB程度ないと厳しいということ。②もしオリジナルデータで解凍すると、ここが87GB程度まで膨れ上がる。

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -l
total 899872
-rw-rw-r-- 1 iu iu 483233202 12月  3 14:56 SRR616268_1.fastq
-rw-rw-r-- 1 iu iu 438233202 12月  3 15:51 SRR616268_2.fastq
iu@bielinux[srp017156] ls -lh
total 879M
-rw-rw-r-- 1 iu iu 461M 12月  3 14:56 SRR616268_1.fastq
-rw-rw-r-- 1 iu iu 418M 12月  3 15:51 SRR616268_2.fastq
iu@bielinux[srp017156] df -h
df: 'mac_share': No such file or directory
Filesystem      Size  Used Avail Use% Mounted on
/dev/sda1       97G   11G   81G  12% /
none            4.0K   0     4.0K   0% /sys/fs/cgroup
udev            991M   4.0K  991M   1% /dev
tmpfs           201M   932K  200M   1% /run
none            5.0M   0     5.0M   0% /run/lock
none           1001M  152K  1001M   1% /run/shm
none            100M   44K   100M   1% /run/user
iu@bielinux[srp017156] █
```

[3:58午後]

[3:58午後]

[3:59午後]

W22-5: bzip2の性能

bzip2の圧縮パフォーマンス。①461MBが83MB、②418MBが74MB。③処理時間は、それぞれ46.364秒と41.399秒。

```
iu@bielinux[srp017156] pwd [ 4:36午後 ]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh [ 4:37午後 ]
total 879M
-rw-rw-r-- 1 iu iu 461M 12月 3 14:56 SRR616268_1.fastq
-rw-rw-r-- 1 iu iu 418M 12月 3 15:51 SRR616268_2.fastq
① iu@bielinux[srp017156] time bzip2 SRR616268_1.fastq [ 4:37午後 ]
bzip2 SRR616268_1.fastq 45.95s user 0.29s system 99% cpu 46.364 total
② iu@bielinux[srp017156] time bzip2 SRR616268_2.fastq [ 4:38午後 ]
bzip2 SRR616268_2.fastq 40.22s user 0.33s system 97% cpu 41.399 total
iu@bielinux[srp017156] ls -lh [ 4:39午後 ]
total 157M
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] [ 5:06午後 ]
```

W22-5: bzip2の性能

bzip2の解凍パフォーマンス。ここではbunzip2コマンドを利用しているが、bzip2 -dでもよい。83MBが解凍後に461MBに、②74MBが解凍後に418MBになるのは自明。③処理時間は、それぞれ23.338秒と22.056秒。解凍にかかる時間は圧縮にかかる時間の約半分であることがわかる。

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh
total 157M
-rw-rw-r-- 1 iu iu 83M 12月  3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月  3 15:51 SRR616268_2.fastq.bz2
① iu@bielinux[srp017156] time bunzip2 SRR616268_1.fastq.bz2
bunzip2 SRR616268_1.fastq.bz2 20.13s user 0.60s system 88% cpu 23.338 total [ 5:19午後 ]
② iu@bielinux[srp017156] time bunzip2 SRR616268_2.fastq.bz2
bunzip2 SRR616268_2.fastq.bz2 17.08s user 0.56s system 79% cpu 22.056 total [ 5:20午後 ]
iu@bielinux[srp017156] ls -lh
total 879M
-rw-rw-r-- 1 iu iu 461M 12月  3 14:56 SRR616268_1.fastq
-rw-rw-r-- 1 iu iu 418M 12月  3 15:51 SRR616268_2.fastq
iu@bielinux[srp017156] [ 5:21午後 ]
```



W22-6: gzipの性能

gzipの圧縮パフォーマンス。①461MBが113MB、②418MBが100MB。③処理時間は、それぞれ29.836秒と32.248秒。gzipはbzip2に比べて圧縮率は劣るが処理時間は速いことがわかる。

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh
total 879M
-rw-rw-r-- 1 iu iu 461M 12月  3 14:56 SRR616268_1.fastq
-rw-rw-r-- 1 iu iu 418M 12月  3 15:51 SRR616268_2.fastq
① iu@bielinux[srp017156] time gzip SRR616268_1.fastq
gzip SRR616268_1.fastq 27.86s user 1.35s system 97% cpu 29.836 total
② iu@bielinux[srp017156] time gzip SRR616268_2.fastq
gzip SRR616268_2.fastq 27.44s user 0.43s system 86% cpu 32.248 total
iu@bielinux[srp017156] ls -lh
total 213M
-rw-rw-r-- 1 iu iu 113M 12月  3 14:56 SRR616268_1.fastq.gz
-rw-rw-r-- 1 iu iu 100M 12月  3 15:51 SRR616268_2.fastq.gz
iu@bielinux[srp017156] █
```

[5:39午後]

[5:39午後]

[5:40午後]

[5:40午後]

[5:40午後]



W22-6: gzipの性能

gzipの解凍パフォーマンス。ここではgunzipコマンドを利用しているが、gzip -dでもよい。113MBが解凍後に461MBに、②100MBが解凍後に418MBになるのは自明。③処理時間は、それぞれ8.569秒と17.029秒。Bzip2同様、解凍にかかる時間は短い、ファイルサイズと比例しないのは個人的には気持ち悪い。

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh
total 213M
-rw-rw-r-- 1 iu iu 113M 12月  3 14:56 SRR616268_1.fastq.gz
-rw-rw-r-- 1 iu iu 100M 12月  3 15:51 SRR616268_2.fastq.gz
① iu@bielinux[srp017156] time gunzip SRR616268_1.fastq.gz [ 5:52午後 ]
gunzip SRR616268_1.fastq.gz 3.00s user 0.38s system 39% cpu 8.569 total
② iu@bielinux[srp017156] time gunzip SRR616268_2.fastq.gz [ 5:52午後 ]
gunzip SRR616268_2.fastq.gz 2.99s user 0.42s system 20% cpu 17.029 total
iu@bielinux[srp017156] ls -lh [ 5:53午後 ]
total 879M
-rw-rw-r-- 1 iu iu 461M 12月  3 14:56 SRR616268_1.fastq
-rw-rw-r-- 1 iu iu 418M 12月  3 15:51 SRR616268_2.fastq
iu@bielinux[srp017156] [ 5:54午後 ]
```

W23-1: wcで確認

最初の6000000行を取ったデータなので、行数は同じはず。一般にはペアエンドの場合は行数が同じになることを確認すべし。①と②でそれぞれ独立にwcコマンドで行数を調べている。数が違っているときは、ちゃんとダウンロードできているのかななどを様々な角度から検証。

```
File Edit View Search Terminal Help
iu@bielinux[srp017156] date
2015年 12月 3日 木曜日 18:07:35 JST
① iu@bielinux[srp017156] wc SRR616268_1.fastq
6000000 12000000 483233202 SRR616268_1.fastq
② iu@bielinux[srp017156] wc SRR616268_2.fastq
6000000 12000000 438233202 SRR616268_2.fastq
iu@bielinux[srp017156]
```

[6:07午後]

[6:07午後]

W23-1: wcで確認

①wcコマンドのあとに複数のファイル名を与えるやり方。②ファイル名の最後が「.fastq」で終わるファイル全てに対してwcコマンドを実行せよという命令。この場合、ペアエンドの2つのファイルしかないので実質的には①と同じ結果が得られるが、カレントディレクトリ中に10個の.fastqで終わるファイルが存在するときは10個分実行される。

```
iu@bielinux[srp017156] date
2015年 12月 3日 木曜日 18:07:35 JST
iu@bielinux[srp017156] wc SRR616268_1.fastq
6000000 12000000 483233202 SRR616268_1.fastq
iu@bielinux[srp017156] wc SRR616268_2.fastq
6000000 12000000 438233202 SRR616268_2.fastq
① iu@bielinux[srp017156] wc SRR616268_1.fastq SRR616268_2.fastq
6000000 12000000 483233202 SRR616268_1.fastq
6000000 12000000 438233202 SRR616268_2.fastq
12000000 24000000 921466404 total
② iu@bielinux[srp017156] wc *.fastq
6000000 12000000 483233202 SRR616268_1.fastq
6000000 12000000 438233202 SRR616268_2.fastq
12000000 24000000 921466404 total
iu@bielinux[srp017156]
```

[6:07午後]

[6:07午後]

[6:09午後]

[6:09午後]

W23-2: time

①150万リード×2ファイルなので、②11.035秒で済んでいるが、約1.35億リード×2ファイルだとwc実行にも一苦労(15分以上かかる見積もり)。

```
iu@bielinux[srp017156] date [ 6:07午後 ]
2015年 12月 3日 木曜日 18:07:35 JST
iu@bielinux[srp017156] wc SRR616268_1.fastq [ 6:07午後 ]
6000000 12000000 483233202 SRR616268_1.fastq
iu@bielinux[srp017156] wc SRR616268_2.fastq [ 6:07午後 ]
6000000 12000000 438233202 SRR616268_2.fastq
iu@bielinux[srp017156] wc SRR616268_1.fastq SRR616268_2.fastq [ 6:07午後 ]
6000000 12000000 483233202 SRR616268_1.fastq
6000000 12000000 438233202 SRR616268_2.fastq
12000000 24000000 921466404 total
iu@bielinux[srp017156] wc *.fastq [ 6:09午後 ]
6000000 12000000 483233202 SRR616268_1.fastq
6000000 12000000 438233202 SRR616268_2.fastq
12000000 24000000 921466404 total
① iu@bielinux[srp017156] time wc *.fastq [ 6:09午後 ]
6000000 12000000 483233202 SRR616268_1.fastq
6000000 12000000 438233202 SRR616268_2.fastq
12000000 24000000 921466404 total
wc *.fastq 10.84s user 0.12s system 99% cpu 11.035 total
iu@bielinux[srp017156] date [ 6:11午後 ]
2015年 12月 3日 木曜日 18:11:30 JST
iu@bielinux[srp017156] [ 6:11午後 ]
```

W23-2: time

NGSデータ解析は、このように各コマンドの実行が十数分から数時間にも及ぶ。そのため、1つのコマンド実行が終わるのを待ち構えて、すぐに次のコマンドを打ち込んでいる...わけではない。ここでは、ホストOS(Windows)のブラウザ上のスクリプトをコピーして実行している。尚、コピーは「CTRL + CおよびV」ではなく右クリックでやるべし。

```
iu@bielinux[srp017156] date
2015年 12月 3日 木曜日 18:07:35
iu@bielinux[srp017156] wc SRR616268_1.fastq
6000000 12000000 483233202 SRR616268_1.fastq
iu@bielinux[srp017156] wc SRR616268_2.fastq
6000000 12000000 438233202 SRR616268_2.fastq
iu@bielinux[srp017156] wc SRR616268_1.fastq SRR616268_2.fastq
6000000 12000000 483233202 SRR616268_1.fastq
6000000 12000000 438233202 SRR616268_2.fastq
12000000 24000000 921466404 total
iu@bielinux[srp017156] wc *.fastq
6000000 12000000 483233202 SRR616268_1.fastq
6000000 12000000 438233202 SRR616268_2.fastq
12000000 24000000 921466404 total
iu@bielinux[srp017156] time wc *.fastq
wc *.fastq 10.84s user 0.12s system 9
iu@bielinux[srp017156] date
2015年 12月 3日 木曜日 18:11:30 JST
iu@bielinux[srp017156]
```

• dateとwc[W23-1]

- date
- wc SRR616268_1.fastq
- wc SRR616268_2.fastq
- wc SRR616268_1.fastq SRR616268_2.fastq
- wc *.fastq
- time wc *.fastq
- date

切り取り(T)
コピー(C) ①
貼り付け
すべて選択(A)

W24-3: DRA特有の...

DRAは、生データのsraファイルからFASTQファイルを生成する際、クオリティの低いリードなどを除去しているため

DRA で公開されている fastq のリード数が生データのそれよりも少ないのは何故でしょうか?

DRA では NCBI SRA Toolkit に含まれている fastq-dump を使い、以下のオプションで生データである SRA ファイルから fastq ファイルを作成しています。

```
fastq-dump -M 25 -E --skip-technical --split-3 -W <SRA file>
```

- -M 25: 25 塩基以上の配列のみを含める。デフォルトは 25。
- -E: リードの開始、もしくは終わりに 10 以上の N が存在しない
- --skip-technical: technical read を除き biological read のみを出力
- --split-3: ペアリードで最初と二番目の biological read をそれぞれ *_1.fastq と *_2.fastq として出力する。一つしか biological read が存在しない場合、*.fastq として出力する。
- -W: 指定されていた場合、left と right を clip する

上記の出力条件でリードがフィルタリング、トリミングされるため、一般的に fastq のリード数は SRA ファイルのそれよりも少なくなっています。フィルタリング、トリミングされていない fastq ファイルを得るには以下のコマンドで fastq を生成します。

```
fastq-dump -M 1 --split-3 <SRA file>
```

作成日: 2013年10月8日; 最終更新日: 2014年6月6日

<http://trace.ddbj.nig.ac.jp/dra/faq.html#read-number-fastq>

W25-1 : time

①bzip2とwcをパイプで連結した場合は、赤下線で示すコマンドごとに②処理時間が出力されるものの、同じ時間(33.632秒)が表示されている。つまり、①で入力したコマンドトータル時間なのだろう。

```
iu@bielinux[srp017156] pwd [ 3:30午後 ]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh [ 3:37午後 ]
total 157M
-rw-rw-r-- 1 iu iu 83M 12月  3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月  3 15:51 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] time bzip2 -d -c SRR616268_1.fastq.bz2 | wc [ 3:37午後 ]
60000000 120000000 483233202
bzip2 -d -c SRR616268_1.fastq.bz2 24.48s user 1.41s system 76% cpu 33.632 total
wc 6.52s user 0.21s system 20% cpu 33.632 total
iu@bielinux[srp017156] [ 3:37午後 ]
```



W25-1 : time

①で-cオプションを入れているので、*.fastq.bz2ファイルは残る。それゆえ、②でもう一度同じ*.fastq.bz2ファイルに対してbzip2コマンドを実行可能。29.081秒かかった②と7.373秒かかった③を合わせて①と同じことになるが、パイプで連結する①の処理時間(33.632秒)のほうが、中間ファイルhoge_1.fastqを作成しない分だけ早いのだろう

```
iu@bielinux[srp017156] pwd
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh
total 157M
-rw-rw-r-- 1 iu iu 83M 12月  3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月  3 15:51 SRR616268_2.fastq.bz2
① iu@bielinux[srp017156] time bzip2 -d -c SRR616268_1.fastq.bz2 | wc [ 3:37午後 ]
6000000 12000000 483233202
bzip2 -d -c SRR616268_1.fastq.bz2 24.48s user 1.41s system 76% cpu 33.632 total
wc 6.52s user 0.21s system 20% cpu 33.632 total
② iu@bielinux[srp017156] time bzip2 -d -c SRR616268_1.fastq.bz2 > hoge_1.fastq
bzip2 -d -c SRR616268_1.fastq.bz2 > hoge_1.fastq 20.24s user 1.45s system 74% c
pu 29.081 total
③ iu@bielinux[srp017156] time wc hoge_1.fastq [ 3:42午後 ]
6000000 12000000 483233202 hoge_1.fastq
wc hoge_1.fastq 5.63s user 0.29s system 80% cpu 7.373 total
iu@bielinux[srp017156] ls -lh [ 3:42午後 ]
total 618M
-rw-rw-r-- 1 iu iu 461M 12月  7 15:42 hoge_1.fastq
-rw-rw-r-- 1 iu iu  83M 12月  3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu  74M 12月  3 15:51 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] [ 3:42午後 ]
```

W25-2: サブセット

①と②は、パイプで連結して最初の400万行分をリダイレクトでファイル保存しているので、十数秒で処理が終わっていることがわかる。

```
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] ls -lh
total 618M
-rw-rw-r-- 1 iu iu 461M 12月  7 15:42 hoge_1.fastq
-rw-rw-r-- 1 iu iu  83M 12月  3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu  74M 12月  3 15:51 SRR616268_2.fastq.bz2
① iu@bielinux[srp017156] time bzip2 -d -c SRR616268_1.fastq.bz2 | head -n 4000000
> subset_1.fastq
bzip2 -d -c SRR616268_1.fastq.bz2 14.08s user 0.45s system 77% cpu 18.775 total
head -n 4000000 > subset_1.fastq 0.32s user 0.42s system 3% cpu 18.774 total
② iu@bielinux[srp017156] time bzip2 -d -c SRR616268_2.fastq.bz2 | head -n 4000000
> subset_2.fastq
bzip2 -d -c SRR616268_2.fastq.bz2 11.41s user 0.84s system 74% cpu 16.443 total
head -n 4000000 > subset_2.fastq 0.33s user 0.35s system 4% cpu 16.443 total
iu@bielinux[srp017156] █
[ 3:47午後]
```

W25-2: サブセット

①一旦461MBのhoge_1.fastqという解凍後のファイル(600万行;150万リード)を作成したのち、最初の400万行分(100万リード分)をリダイレクトでファイル保存したのが306MBのhoge_subset_1.fastq。

```
iu@bielinux[srp017156] ls -lh
total 618M
-rw-rw-r-- 1 iu iu 461M 12月  7 15:42 hoge_1.fastq
-rw-rw-r-- 1 iu iu  83M 12月  3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu  74M 12月  3 15:51 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] time bzip2 -d -c SRR616268_1.fastq.bz2 | head -n 4000000
> subset_1.fastq
bzip2 -d -c SRR616268_1.fastq.bz2 14.08s user 0.45s system 77% cpu 18.775 total
head -n 4000000 > subset_1.fastq 0.32s user 0.42s system 3% cpu 18.774 total
iu@bielinux[srp017156] time bzip2 -d -c SRR616268_2.fastq.bz2 | head -n 4000000
> subset_2.fastq
bzip2 -d -c SRR616268_2.fastq.bz2 11.41s user 0.84s system 74% cpu 16.443 total
head -n 4000000 > subset_2.fastq 0.33s user 0.35s system 4% cpu 16.443 total
iu@bielinux[srp017156] head -n 4000000 hoge_1.fastq > hoge_subset_1.fastq
iu@bielinux[srp017156] ls -lh
total 1.5G
-rw-rw-r-- 1 iu iu 461M 12月  7 15:42 hoge_1.fastq
-rw-rw-r-- 1 iu iu 306M 12月  7 15:53 hoge_subset_1.fastq
-rw-rw-r-- 1 iu iu  83M 12月  3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu  74M 12月  3 15:51 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 12月  7 15:47 subset_1.fastq
-rw-rw-r-- 1 iu iu 278M 12月  7 15:47 subset_2.fastq
iu@bielinux[srp017156]
```

①

W25-2: diff

①diffコマンドで2つのファイルの同一性を確認している。ファイルの中身が全く同じ場合、ここで示す結果のように何も表示されない

```
iu@bielinux[srp017156] ls -lh
total 618M
-rw-rw-r-- 1 iu iu 461M 12月 7 15:42 hoge_1.fastq
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
iu@bielinux[srp017156] time bzip2 -d -c SRR616268_1.fastq.bz2 | head -n 4000000
> subset_1.fastq
bzip2 -d -c SRR616268_1.fastq.bz2 14.08s user 0.45s system 77% cpu 18.775 total
head -n 4000000 > subset_1.fastq 0.32s user 0.42s system 3% cpu 18.774 total
iu@bielinux[srp017156] time bzip2 -d -c SRR616268_2.fastq.bz2 | head -n 4000000
> subset_2.fastq
bzip2 -d -c SRR616268_2.fastq.bz2 11.41s user 0.84s system 74% cpu 16.443 total
head -n 4000000 > subset_2.fastq 0.33s user 0.35s system 4% cpu 16.443 total
iu@bielinux[srp017156] head -n 4000000 hoge_1.fastq > hoge_subset_1.fastq
iu@bielinux[srp017156] ls -lh
total 1.5G
-rw-rw-r-- 1 iu iu 461M 12月 7 15:42 hoge_1.fastq
-rw-rw-r-- 1 iu iu 306M 12月 7 15:53 hoge_subset_1.fastq
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 12月 7 15:47 subset_1.fastq
-rw-rw-r-- 1 iu iu 278M 12月 7 15:47 subset_2.fastq
iu@bielinux[srp017156] diff hoge_subset_1.fastq subset_1.fastq
iu@bielinux[srp017156]
```



W25-3: 応用編

①subsetから始まるファイル群を一気にgzip圧縮することが可能。

```
iu@bielinux[srp017156] pwd [ 4:16午後 ]
/home/iu/Documents/srp017156
iu@bielinux[srp017156] ls -lh [ 4:17午後 ]
total 1.5G
-rw-rw-r-- 1 iu iu 461M 12月 7 15:42 hoge_1.fastq
-rw-rw-r-- 1 iu iu 306M 12月 7 15:53 hoge_subset_1.fastq
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 306M 12月 7 15:47 subset_1.fastq
-rw-rw-r-- 1 iu iu 278M 12月 7 15:47 subset_2.fastq
iu@bielinux[srp017156] gzip subset* [ 4:17午後 ]
iu@bielinux[srp017156] ls -lh [ 4:17午後 ]
total 1.1G
-rw-rw-r-- 1 iu iu 461M 12月 7 15:42 hoge_1.fastq
-rw-rw-r-- 1 iu iu 306M 12月 7 15:53 hoge_subset_1.fastq
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 7 15:47 subset_1.fastq.gz
-rw-rw-r-- 1 iu iu 66M 12月 7 15:47 subset_2.fastq.gz
iu@bielinux[srp017156] [ 4:17午後 ]
```

①

W25-3: 応用編

①と②で示すように、パイプで一気にgzip圧縮まで行うことも可能。③と④は全く同じファイルサイズになることを想定していたが若干異なる(76659480 bytesと76659495 bytes)のが気になる。

```
iu@bielinux[srp017156] ls -lh
total 1.1G
-rw-rw-r-- 1 iu iu 461M 12月 7 15:42 hoge_1.fastq
-rw-rw-r-- 1 iu iu 306M 12月 7 15:53 hoge_subset_1.fastq
-rw-rw-r-- 1 iu iu 83M 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 3 15:51 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 74M 12月 7 15:47 subset_1.fastq.gz
-rw-rw-r-- 1 iu iu 66M 12月 7 15:47 subset_2.fastq.gz
① iu@bielinux[srp017156] bzip2 -d -c SRR616268_1.fastq.bz2 | head -n 4000000 | gzi
p > SRR616268sub_1.fastq.gz
② iu@bielinux[srp017156] bzip2 -d -c SRR616268_2.fastq.bz2 | head -n 4000000 | gzi
p > SRR616268sub_2.fastq.gz
iu@bielinux[srp017156] ls -l
total 1229020
-rw-rw-r-- 1 iu iu 483233202 12月 7 15:42 hoge_1.fastq
-rw-rw-r-- 1 iu iu 320760716 12月 7 15:53 hoge_subset_1.fastq
-rw-rw-r-- 1 iu iu 87005342 12月 3 14:56 SRR616268_1.fastq.bz2
-rw-rw-r-- 1 iu iu 76813406 12月 3 15:51 SRR616268_2.fastq.bz2
-rw-rw-r-- 1 iu iu 76659480 12月 7 16:27 SRR616268sub_1.fastq.gz ③
-rw-rw-r-- 1 iu iu 68682938 12月 7 16:27 SRR616268sub_2.fastq.gz
-rw-rw-r-- 1 iu iu 76659495 12月 7 15:47 subset_1.fastq.gz ④
-rw-rw-r-- 1 iu iu 68682953 12月 7 15:47 subset_2.fastq.gz
iu@bielinux[srp017156] █
```

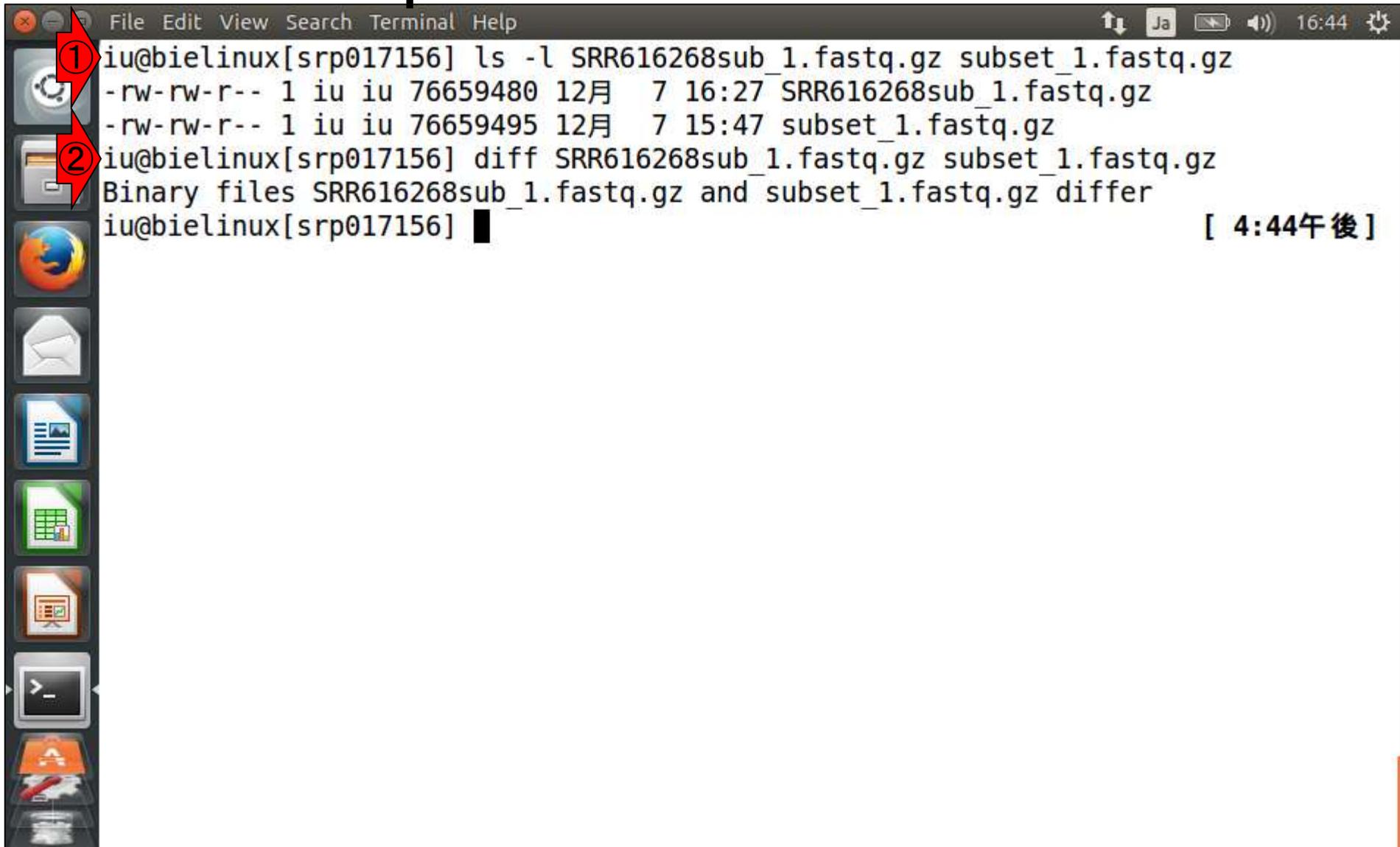
[4:20午後]

[4:27午後]

[4:28午後]

①ls、②diff。結果として、確かにバイナリファイル状態で比較する2つのファイルに違いがあることがわかる。

W25-4: Tips

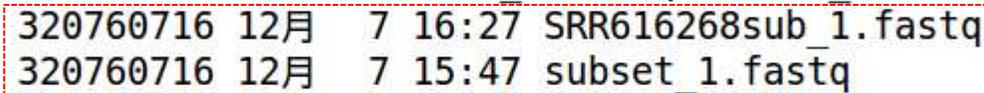


```
File Edit View Search Terminal Help 16:44
iu@bielinux[srp017156] ls -l SRR616268sub_1.fastq.gz subset_1.fastq.gz
-rw-rw-r-- 1 iu iu 76659480 12月  7 16:27 SRR616268sub_1.fastq.gz
-rw-rw-r-- 1 iu iu 76659495 12月  7 15:47 subset_1.fastq.gz
iu@bielinux[srp017156] diff SRR616268sub_1.fastq.gz subset_1.fastq.gz
Binary files SRR616268sub_1.fastq.gz and subset_1.fastq.gz differ
iu@bielinux[srp017156] [ 4:44午後]
```

W25-4: Tips

①と②でgzip圧縮ファイルを解凍し、テキストファイル状態でどうなっているかを調べる。解凍後に③lsすると、赤枠で示すようになぜか同じファイルサイズ(320760716 bytes)になっている

```
iu@bielinux[~/Documents/srp017156]
iu@bielinux[srp017156] ls -l SRR616268sub_1.fastq.gz subset_1.fastq.gz
-rw-rw-r-- 1 iu iu 76659480 12月  7 16:27 SRR616268sub_1.fastq.gz
-rw-rw-r-- 1 iu iu 76659495 12月  7 15:47 subset_1.fastq.gz
iu@bielinux[srp017156] diff SRR616268sub_1.fastq.gz subset_1.fastq.gz
Binary files SRR616268sub_1.fastq.gz and subset_1.fastq.gz differ
iu@bielinux[srp017156] gunzip SRR616268sub_1.fastq.gz
iu@bielinux[srp017156] gunzip subset_1.fastq.gz
iu@bielinux[srp017156] ls -l SRR616268sub 1.fastq subset 1.fastq
-rw-rw-r-- 1 iu iu 320760716 12月  7 16:27 SRR616268sub 1.fastq
-rw-rw-r-- 1 iu iu 320760716 12月  7 15:47 subset 1.fastq
iu@bielinux[srp017156]
```



W25-4: Tips

```
iu@bielinux[srp017156] ls -l SRR616268sub_1.fastq.gz subset_1.fastq.gz
-rw-rw-r-- 1 iu iu 76659480 12月  7 16:27 SRR616268sub_1.fastq.gz
-rw-rw-r-- 1 iu iu 76659495 12月  7 15:47 subset_1.fastq.gz
iu@bielinux[srp017156] diff SRR616268sub_1.fastq.gz subset_1.fastq.gz
Binary files SRR616268sub_1.fastq.gz and subset_1.fastq.gz differ
iu@bielinux[srp017156] gunzip SRR616268sub_1.fastq.gz           [ 4:44午後 ]
iu@bielinux[srp017156] gunzip subset_1.fastq.gz                 [ 4:46午後 ]
iu@bielinux[srp017156] ls -l SRR616268sub_1.fastq subset_1.fastq [ 4:47午後 ]
-rw-rw-r-- 1 iu iu 320760716 12月  7 16:27 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 320760716 12月  7 15:47 subset_1.fastq
iu@bielinux[srp017156] diff SRR616268sub_1.fastq subset_1.fastq [ 4:47午後 ]
iu@bielinux[srp017156] █                                         [ 4:50午後 ]
```



W25-4: Tips

再度①gzip圧縮して②diffして、③ls。上下の④赤枠が示すように、圧縮ごとにファイルサイズは異なる。結論としては、気にしなくてよい。

```
iu@bielinux[srp017156] ls -l SRR616268sub_1.fastq.gz subset_1.fastq.gz
-rw-rw-r-- 1 iu iu 76659480 12月 7 16:27 SRR616268sub_1.fastq.gz
-rw-rw-r-- 1 iu iu 76659495 12月 7 15:47 subset_1.fastq.gz
iu@bielinux[srp017156] diff SRR616268sub_1.fastq.gz subset_1.fastq.gz
Binary files SRR616268sub_1.fastq.gz and subset_1.fastq.gz differ
iu@bielinux[srp017156] gunzip SRR616268sub_1.fastq.gz [ 4:44午後]
iu@bielinux[srp017156] gunzip subset_1.fastq.gz [ 4:46午後]
iu@bielinux[srp017156] ls -l SRR616268sub_1.fastq subset_1.fastq [ 4:47午後]
-rw-rw-r-- 1 iu iu 320760716 12月 7 16:27 SRR616268sub_1.fastq
-rw-rw-r-- 1 iu iu 320760716 12月 7 15:47 subset_1.fastq
iu@bielinux[srp017156] diff SRR616268sub_1.fastq subset_1.fastq [ 4:47午後]
① iu@bielinux[srp017156] gzip SRR616268sub_1.fastq [ 4:50午後]
iu@bielinux[srp017156] gzip subset_1.fastq [ 4:57午後]
② iu@bielinux[srp017156] diff SRR616268sub_1.fastq.gz subset_1.fastq.gz
Binary files SRR616268sub_1.fastq.gz and subset_1.fastq.gz differ
③ iu@bielinux[srp017156] ls -l SRR616268sub_1.fastq.gz subset_1.fastq.gz
-rw-rw-r-- 1 iu iu 76659501 12月 7 16:27 SRR616268sub_1.fastq.gz
-rw-rw-r-- 1 iu iu 76659495 12月 7 15:47 subset_1.fastq.gz
iu@bielinux[srp017156] [ 4:58午後]
```