

次世代シーケンサーデータの解析手法 第8回アセンブリ後の解析: ウェブ資料

谷澤 靖洋、神沼英里*、中村保一、遠野雅徳、
寺田朋子、清水謙多郎、門田 幸二*

*東京大学・大学院農学生命科学研究科
kadota@bi.a.u-tokyo.ac.jp

<http://www.iu.a.u-tokyo.ac.jp/~kadota/>

W1: 初心者向け教材

The screenshot shows a web browser window with the address bar containing the URL http://www.iu.a.u-tokyo.ac.jp/~kadota/r_seq.html. The browser's menu bar includes options like 'ファイル(F)', '編集(E)', '表示(V)', 'お気に入り(A)', 'ツール(T)', and 'ヘルプ(H)'. The main content area features the title '(Rで)塩基配列解析' with a sub-note '(last modified 2016/08/27, since 2011)'. Below the title is a paragraph of introductory text. A 'What's new?' section follows, containing a list of updates. A red diamond icon with the number '1' is placed over the link '講習会、講義、講演資料' in the list. At the bottom right of the page, there is a link 'トップページへ'.

(Rで)塩基配列解析

(last modified 2016/08/27, since 2011)

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

What's new?

- バイオインフォマティクスやNGSをキーワード程度は知っているヒト(学部3年生程度)向けの講義資料を作成しました。「参考資料 | [講習会](#)、[講義](#)、[講演資料](#)」の2016.09.12-16の日付のものです。[日本乳酸菌学会誌のNGS連載](#)第1-4回あたりのダイジェスト版のようなものです。[NGSハンズオン講習会2016](#)の予習事項が厳しすぎて受講を断念したヒトも、ここからだとスムーズに頭に入っていきかもしれません。(2016/08/26) **NEW**
- 参考資料の項目の情報量が多くなってきたので、「[書籍](#)、[学会誌](#)」と「[講習会](#)、[講義](#)、[講演資料](#)」の2つに分割しました。(2016/08/23) **NEW**
- [NGSハンズオン講習会2016](#)の「昨年度との違い、対応関係、想定受講者、および予習事項」のPDFに注釈を入れました。8/4のcuffdiffの入力ファイル周辺の話です。経緯などの記憶は定かではありませんが、ここで誤解を与えていたかもしれないと思いましたので、念のため追記しました。これ以外にも間違いや私の誤解などありましたら、よりよい情報共有のために遠慮なくご指摘よろしくお願いたしますm(_ _)m(2016/08/19) **NEW**
- [NGSハンズオン講習会2016](#)の第3部(8/4)の[講義資料PDF](#)を修正しました。スライド107 (Trinityのパスを通すところのミス)、スライド140 (Bridgerに必要のboostパッケージをapt-getでやるとサンプルデータの実行もうまくいく)あたりです。(2016/08/12) **NEW**
- [NGSハンズオン講習会2016](#)の第2部(7/25-28)の講義資料修正版を公開しました。(2016/08/03) **NEW**

- [門田からメール返信をもらえない場合は](#) (last modified 2016/08/23) **NEW**
- [はじめに](#) (last modified 2015/03/31)
- 参考資料 | [書籍](#)、[学会誌](#) (last modified 2016/08/23) **NEW**
- 参考資料 | [講習会](#)、[講義](#)、[講演資料](#) **①** (last modified 2016/08/26) **NEW**
- [過去のお知らせ](#) (last modified 2016/08/24) **NEW**

[トップページへ](#)

W1: 初心者向け教材

新しいものから順にリストアップされている。①
今回の目的物は、②2016年9月12-16日の講義
資料なので、たまたま一番上に見えているだけ

参考資料 | 講習会、講義、講演資料 NEW

基本的に私門田の個人ページに記載してあるものです。かなり古い講演資料などの情報をもとに勉強されている方もいらっしゃるようですので、ここでは2013年秋以降の情報のみです。大まかな内容についても述べています。講義資料としての利用などは事前連絡や私個人への謝辞も気にせずご自由にお使いください。

- 門田幸二、「[講義資料](#)(2016.08.26版; 約5.3MB)」, 東京大学農学部展開科目: バイオインフォマティクス, 東京大学(東京), 2016.09.12-16
内容: バイオインフォマティクスやNGSをキーワード程度は知っているヒト(学部3年生程度)向けのアクティブ・ラーニング系講義。バイオインフォマティクスとLinuxスキル習得の意義。バイオインフォマティクス分野で需要が多いのはNGSデータの解析(JST-NBDCIによる[アンケート結果](#))。バイオインフォマティクス人材育成のための講習会の一環として行われているNGS講習会資料を有効活用するために必要な基本スキルの習得。具体的には、日本乳酸菌学会誌のNGS連載内容を理解するための基礎として、[Bio-Linux](#)のノリに慣れるのが最低限の到達目標。何かやったという達成感を味わってもらうため、NGSデータのde novoゲノムアセンブリを行い、元のリード長(塩基配列の長さ)よりも伸びたことを実感してもらう。また、アセンブリ結果の一部をBLAST検索し、入力データ([hoge.fasta](#); 約6.4MB)がどんな生物種であったかを調べるあたりまで。課題は[kadai1.fasta](#), [kadai2.fasta](#), [kadai3.fasta](#), [kadai4.fasta](#)のいずれかを選択して実行。下記と同様の手順で「データのダウンロード → de novoアセンブリ → BLAST検索」し、得られた結果を発表。約2日分。

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

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

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

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

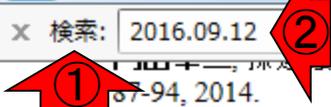
#de novoアセンブリを実行し、結果をugeディレクトリに保存
velveth uge 31 -short -fasta hoge.fasta
velvetg uge

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

[トップページへ](#)

W1: 初心者向け教材

Windowsの場合は「CTRL + F」を押して、①ページ内検索で、目的とする講義資料の②開催日時(この場合2016.09.12)で検索してもよい。③のような感じで引っ掛けられます



- ① 07-94, 2014.
- 内容: 日本乳酸菌学会誌のNGS関連連載の第1回分です。NGS解析関連の情報収集先、NGS用教育カリキュラム、Windows上でのLinux環境構築、Bio-Linux、R環境構築、Rで一通り解析ができたことを知っているマイクロアレイ解析出身のヒトがRNA-seq解析でLinuxを強く勧められる理由など。Perl, Python, Rubyなどの他のプログラミング言語の概観。DDBJパイプラインなどのウェブツール、Galaxyの簡単な説明。Expression Atlas中のbaselineやcontrastの説明を通じた解析目的別留意点、Expression Atlasと似た思想のRefEx紹介。Rで乳酸菌ゲノム配列を読み込んでGC含量をコピペで得られることなど。書籍中のリンク先やRコードは「書籍 | 日本乳酸菌学会誌 | [第1回イントロダクション](#)」の項目をご覧ください。
- 門田幸二著(金明哲 編), シリーズ Useful R 第7巻トランスクリプトーム解析, 共立出版, 2014. ISBN: 978-4-320-12370-0
内容: マイクロアレイとRNA-seq解析を例としてRを用いてトランスクリプトーム解析を行うための体系的な本としてまとめました。数式が苦手なヒト向けに、重みつき平均の具体的な計算例などを挙げてオプションの意味などがわかるような中身の理解に重点を置いた構成にしています。書籍中のRコードは「書籍 | トランスクリプトーム解析 | ...」をご覧ください。
 - 門田幸二, 「トランスクリプトミクスの推奨データ解析ガイドライン」, ニュートリゲノミクスを基盤としたバイオマーカーの開発, シーエムシー出版, 45-52, 2013. ISBN: 978-4-7813-0820-3
内容: マイクロアレイ解析の話がメインです。実験デザインの重要性を述べています。Affymetrix GeneChipデータの数値化と発現変動遺伝子(DEG)検出法の組合せの重要性の話や、サンプル間クラスタリングである程度DEGに関する情報がわかることを述べています。MAS5データを用いる場合は特に倍率変化で議論することも無意味であること、RMAのようなマルチアレイ正規化法を用いて得られたマイクロアレイデータの場合にはなぜ倍率変化でうまくいく傾向にあるかなどの理由をM-A plotを用いて説明しています。

参考資料 | 講習会、講義、講演資料 NEW

基本的に私門田の個人ページに記載してあるものです。かなり古い講演資料などの情報をもとに勉強されている方もいらっしゃるようですので、ここでは2013年秋以降の情報のみです。大まかな内容についても述べています。講義資料としての利用などは事前連絡や私個人への謝辞も気にせずご自由にお使いください。

- 門田幸二, 「[講義資料\(2016.08.26版; 約5.3MB\)](#)」, 東京大学農学部展開科目: バイオインフォマティクス, 東京大学(東京), 2016.09.12-16
内容: バイオインフォマティクスやNGSをキーワード程度は知っているヒト(学部3年生程度)向けのアクティブ・ラーニング系講義。[トップページへ](#)
フォマティクスとLinuxスキル習得の意義。バイオインフォマティクス分野で需要が多いのはNGSデータの解析 (JST-NBDCIによる[アンケート結](#)

W2-1 : result.zipダウンロード

①HGAP実行結果ファイル(result.zip)のダウンロード。第6回W9-3

```
iu@bielinux[iu] cd ~/Desktop/ [11:46午前]
iu@bielinux[Desktop] pwd [11:46午前]
/home/iu/Desktop
iu@bielinux[Desktop] ls [11:46午前]
Bio-Linux Documentation hoge mac share Sample Data
① iu@bielinux[Desktop] wget -cq http://www.iu.a.u-tokyo.ac.jp/~kadota/bo [11:47午前]
ok/DRR054113/result.zip
iu@bielinux[Desktop] ls [11:47午前]
Bio-Linux Documentation hoge mac share result.zip Sample Data
iu@bielinux[Desktop] █ [11:47午前]
```

①unzipで解凍(第6回W9-4)
。②resultディレクトリに移動

W2-2: result.zipの解凍

```
iu@bielinux[iu] cd ~/Desktop/ [11:46午前]
iu@bielinux[Desktop] pwd [11:46午前]
/home/iu/Desktop
iu@bielinux[Desktop] ls [11:46午前]
Bio-Linux Documentation hoge mac share Sample Data
iu@bielinux[Desktop] wget -cq http://www.iu.a.u-tokyo.ac.jp/~kadota/bo [11:47午前]
ok/DRR054113/result.zip
iu@bielinux[Desktop] ls [11:47午前]
Bio-Linux Documentation hoge mac share result.zip Sample Data
① iu@bielinux[Desktop] unzip result.zip [11:47午前]
Archive: result.zip
  creating: result/
  inflating: result/corrected.fastq
  inflating: result/smrtpipe.log
  inflating: result/polished_assembly.fastq
  inflating: result/polished_assembly.fasta
② iu@bielinux[Desktop] cd result [12:58午後]
iu@bielinux[result] pwd [12:58午後]
/home/iu/Desktop/result
iu@bielinux[result] █ [12:58午後]
```

W2-3: FASTAファイル

①アセンブリ結果のFASTAファイル (polished_assembly.fasta)は、塩基配列部分が複数行に分かれている

```
iu@bielinux[result] pwd [ 1:06午後 ]
/home/iu/Desktop/result
iu@bielinux[result] ls [ 1:06午後 ]
corrected.fastq polished_assembly.fastq
polished_assembly.fasta smrtpipe.log
iu@bielinux[result] head polished_assembly.fasta [ 1:06午後 ]
>unitig_0|quiver
tgttgggctgctgaatgaacatagcgaatttgcccggaaactactttttggcgggtggcaa
tcgcgctgacagatttacgctcaaaggaaacctgatgatggttagtggcaggattctgca
atatatcgaaggggattacgactacattaatgcgcgatggggctttaggctattcggct
gcaattcctaccgtgatttagtgacgaccaattaagcaactttatcatactggcgatttg
cttttaatgagcatcgggttgacagatagcctgtatttttacaaggactatccggataaa
aagggtgcggtgattgggcttgattcctaatgacggaccggaagtgctagatgacaaaggc
ttacccaaaatatttttgctattaagcacatgggggtatcagcaagcccagattgattggct
ggccaacacggcactcacgggggtacctgacgactatgtggtaatggttagttggccatat
tccagccgcggcagtgggcagtgatcagaacaaccagacgcttatcaaccaaattctaaa
iu@bielinux[result] [ 1:06午後 ]
```

W2-4: 配列長でソート

配列長(降順)でソートし、コンティグの塩基配列部分を1行にすることを目的としてfastaLengthFilter.pyを利用。

①whereでパスの確認。第6回W12

```
iu@bielinux[result] pwd
/home/iu/Desktop/result
iu@bielinux[result] ls
corrected.fastq          polished_assembly.fastq
polished_assembly.fasta smrtpipe.log
iu@bielinux[result] head polished_assembly.fasta
>unitig_0|quiver
tgttgggctgctgaatgaacatagcgaatttggcccggaaactactttttggcgggtggcaa
tcgcgctgacagatttacgctcaaaggaaacctgatgatggttagtggcaggattctgca
atatatcgaaggggattacgactacattaaatgcgcgatggggctttaggctattcggct
gcaattcctaccgtgatttagtgacgaccaattaaagcaactttatcatactggcgatttg
ctttttaatgagcatcgggttgacagatagcctgtatttttacaaggactatccggataaa
aagggtgcggtgattgggcttgattcctaatgacggaccggaagtgctagatgacaaaggc
ttacccaaaatatttttgctattaagcacatgggggtatcagcaagcccagattgattggct
ggccaacacggcactcacgggggtacctgacgactatgtggtaatggttagttggccatat
tccagccgcggcagtgggcagtgatcagaacaaccagacgcttatcaaccaaattctaaa
iu@bielinux[result] where fastaLengthFilter.py
/home/iu/bin/fastaLengthFilter.py
/home/iu/bin/fastaLengthFilter.py
iu@bielinux[result] █
```

[1:06午後]

[1:06午後]

[1:06午後]

[1:06午後]

[1:12午後]



W2-5: 実行

①polished_assembly.fastaを入力として、0塩基以上の配列を抽出した結果をLH_hgap.faというファイル名で保存する
fastaLengthFilter.pyを実行。オリジナルは4コンティグのFASTAファイルなので、②wc実行結果が③8になるのは妥当

```
iu@bielinux[result] pwd [ 1:28午後 ]
/home/iu/Desktop/result
iu@bielinux[result] ls [ 1:28午後 ]
corrected.fastq polished_assembly.fastq
polished_assembly.fasta smrtpipe.log
iu@bielinux[result] fastaLengthFilter.py polished_assembly.fasta 0 > L
H_hgap.fa
iu@bielinux[result] ls [ 1:28午後 ]
corrected.fastq polished_assembly.fasta smrtpipe.log
LH_hgap.fa polished_assembly.fastq
iu@bielinux[result] wc LH_hgap.fa [ 1:28午後 ]
      8      8 2433662 LH_hgap.fa
iu@bielinux[result] [ 1:28午後 ]
```



W2-6: ファイル分割

①headとtailコマンドを組み合わせて、multi-FASTAファイルからsingle-FASTAファイルに分割。第7回W10-3と10-4

```
iu@bielinux[result] pwd [ 2:06午後 ]
/home/iu/Desktop/result
iu@bielinux[result] ls [ 2:06午後 ]
corrected.fastq polished_assembly.fasta smrtpipe.log
LH_hgap.fa polished_assembly.fastq
iu@bielinux[result] head -n 2 LH_hgap.fa | tail -n 2 > sequence1.fa
iu@bielinux[result] head -n 4 LH_hgap.fa | tail -n 2 > sequence2.fa
iu@bielinux[result] head -n 6 LH_hgap.fa | tail -n 2 > sequence3.fa
iu@bielinux[result] head -n 8 LH_hgap.fa | tail -n 2 > sequence4.fa
iu@bielinux[result] ls -l *.fa [ 2:06午後 ]
-rw-rw-r-- 1 iu iu 2433662 8月 29 13:28 LH_hgap.fa
-rw-rw-r-- 1 iu iu 2289509 8月 29 14:06 sequence1.fa
-rw-rw-r-- 1 iu iu 86904 8月 29 14:06 sequence2.fa
-rw-rw-r-- 1 iu iu 45865 8月 29 14:06 sequence3.fa
-rw-rw-r-- 1 iu iu 11384 8月 29 14:06 sequence4.fa
iu@bielinux[result] █ [ 2:06午後 ]
```



W3-1: シェルスクリプト

第7回W10-3やW10-4とは異なり、① sequence[0-9].fa作成のためのスクリプトを統一的に記述している。このような体系的なコマンドにすることで、自動処理が容易になる。シェルスクリプトだと、どのように記述できるかを示す(示したかったから①のように記述しているのです)

```
iu@bielinux[result] pwd
/home/iu/Desktop/result
iu@bielinux[result] ls
corrected.fastq  polished_assembly.fasta  smrtpip
LH_hgap.fa      polished_assembly.fastq

iu@bielinux[result] head -n 2 LH_hgap.fa | tail -n 2 > sequence1.fa
iu@bielinux[result] head -n 4 LH_hgap.fa | tail -n 2 > sequence2.fa
iu@bielinux[result] head -n 6 LH_hgap.fa | tail -n 2 > sequence3.fa
iu@bielinux[result] head -n 8 LH_hgap.fa | tail -n 2 > sequence4.fa

iu@bielinux[result] ls -l *.fa
-rw-rw-r-- 1 iu iu 2433662  8月 29 13:28 LH_hgap.fa
-rw-rw-r-- 1 iu iu 2289509  8月 29 14:06 sequence1.fa
-rw-rw-r-- 1 iu iu  86904   8月 29 14:06 sequence2.fa
-rw-rw-r-- 1 iu iu  45865   8月 29 14:06 sequence3.fa
-rw-rw-r-- 1 iu iu  11384   8月 29 14:06 sequence4.fa
iu@bielinux[result] █
```

[2:06午後]

[2:06午後]



①

W3-2: 参考資料

平成27年度NGSハンズオン講習会

①

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

H27年度概要

H27年度講義日程・参考資料

H26年度講習会の情報については[こちら](#)をご覧ください。

H27年度実施報告書・講義資料・動画等

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

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

実施日	実施時間	大項目	項目	レベル	習得技術	担当講師(敬称略)	講義資料・動画(統合TV)				
7月22日 (水)	10:30-12:00	PC環境の構築	Bio-Linux8とRのインストール状況確認		<ul style="list-style-type: none"> Linux導入 R導入 NGS解析に必要な環境構築技術 	門田 幸二 (東京大学)	事前予習資料一覧 (PDF:52KB)				
	13:15-14:45										
	15:00-16:30										
	16:45-18:15										講義資料一覧 (PDF:108KB)
7月23日 (木)	10:30-12:00	UNIX/Linuxとスクリプト言語	Linux基礎	初級	UNIXの基礎の理解	門田 幸二 (東京大学)	講義資料一覧 (PDF:32KB)				
	13:15-14:45			中級							
	15:00-16:30								統合TV		
	16:45-18:15										
7月24日 (金)	10:30-12:00		スクリプト言語	中級	シェルスクリプト	服部 恵美 (アメリエフ)	講義資料 (PDF:1.8MB)				
	13:15-14:45										
	15:00-16:30									統合TV	

②

W3-3: 戦略を練る

まず、①の部分は2で固定。理由は、出力ファイルがsingle-FASTA形式だから

```
iu@bielinux[result] pwd [ 2:06午後 ]
/home/iu/Desktop/result
iu@bielinux[result] ls [ 2:06午後 ]
corrected.fastq polished_assembly.fasta smrtpipe_log
LH_hgap.fa polished_assembly.fastq
iu@bielinux[result] head -n 2 LH_hgap.fa | tail -n 2 > sequence1.fa
iu@bielinux[result] head -n 4 LH_hgap.fa | tail -n 2 > sequence2.fa
iu@bielinux[result] head -n 6 LH_hgap.fa | tail -n 2 > sequence3.fa
iu@bielinux[result] head -n 8 LH_hgap.fa | tail -n 2 > sequence4.fa
iu@bielinux[result] ls -l *.fa [ 2:06午後 ]
-rw-rw-r-- 1 iu iu 2433662 8月 29 13:28 LH_hgap.fa
-rw-rw-r-- 1 iu iu 2289509 8月 29 14:06 sequence1.fa
-rw-rw-r-- 1 iu iu 86904 8月 29 14:06 sequence2.fa
-rw-rw-r-- 1 iu iu 45865 8月 29 14:06 sequence3.fa
-rw-rw-r-- 1 iu iu 11384 8月 29 14:06 sequence4.fa
iu@bielinux[result] [ 2:06午後 ]
```



W3-3: 戦略を練る

変わっているのは①と②の部分。①1-4までのループを回して、その2倍の値として②を表現できるだろうと妄想する

```
iu@bielinux[result] pwd [ 2:06午後 ]
/home/iu/Desktop/result
iu@bielinux[result] ls [ 2:06午後 ]
corrected.fastq polished assembly.fasta smrtpipe.log
LH_hgap.fa polished assembly.fasta
iu@bielinux[result] head -n 2 LH_hgap.fa | tail -n 2 > sequence1.fa
iu@bielinux[result] head -n 4 LH_hgap.fa | tail -n 2 > sequence2.fa
iu@bielinux[result] head -n 6 LH_hgap.fa | tail -n 2 > sequence3.fa
iu@bielinux[result] head -n 8 LH_hgap.fa | tail -n 2 > sequence4.fa
iu@bielinux[result] ls -l *.fa [ 2:06午後 ]
-rw-rw-r-- 1 iu iu 2433662 8月 29 13:28 LH_hgap.fa
-rw-rw-r-- 1 iu iu 2289509 8月 29 14:06 sequence1.fa
-rw-rw-r-- 1 iu iu 86904 8月 29 14:06 sequence2.fa
-rw-rw-r-- 1 iu iu 45865 8月 29 14:06 sequence3.fa
-rw-rw-r-- 1 iu iu 11384 8月 29 14:06 sequence4.fa
iu@bielinux[result] [ 2:06午後 ]
```

W3-4: 基本形

①シェルスクリプトの基本形ファイルJSLAB8_1.shを、②wgetし、③moreで確認

```
iu@bielinux[result] pwd [10:23午前]
/home/iu/Desktop/result
② iu@bielinux[result] wget -cq http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB8_1.sh
iu@bielinux[result] ls [10:23午前]
corrected.fastq polished_assembly.fasta sequence2.fa smrtpipe.log
JSLAB8_1.sh polished_assembly.fastq sequence3.fa
LH_hgap.fa sequence1.fa sequence4.fa
③ iu@bielinux[result] more JSLAB8_1.sh [10:23午前]
#!/bin/sh
for i in `seq 1 4`
do
    echo "head -n $i*2 LH_hgap.fa | tail -n 2 > test$i.fa"
done
① iu@bielinux[result] [10:23午前]
```

W3-5: 解説

①1から②4までの1刻みのループを回し、③iという変数で取り扱う。別にiでなくてもよいが、変数名として、i, j, kといった順番で使うヒトも一定数存在する

```
iu@bielinux[result] pwd [10:23午前]
/home/iu/Desktop/result
iu@bielinux[result] wget -cq http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB8_1.sh
iu@bielinux[result] ls [10:23午前]
corrected.fastq polished_assembly.fasta sequence2.fa smrtpipe.log
JSLAB8_1.sh polished_assembly.fastq sequence3.fa
LH_hgap.fa sequence1.fa sequence4.fa
iu@bielinux[result] more JSLAB8_1.sh [10:23午前]
#!/bin/sh
for i in `seq 1 4`
do
    echo "head -n $i*2 LH_hgap.fa | tail -n 2 > test$i.fa"
done
iu@bielinux[result] [10:23午前]
```

①

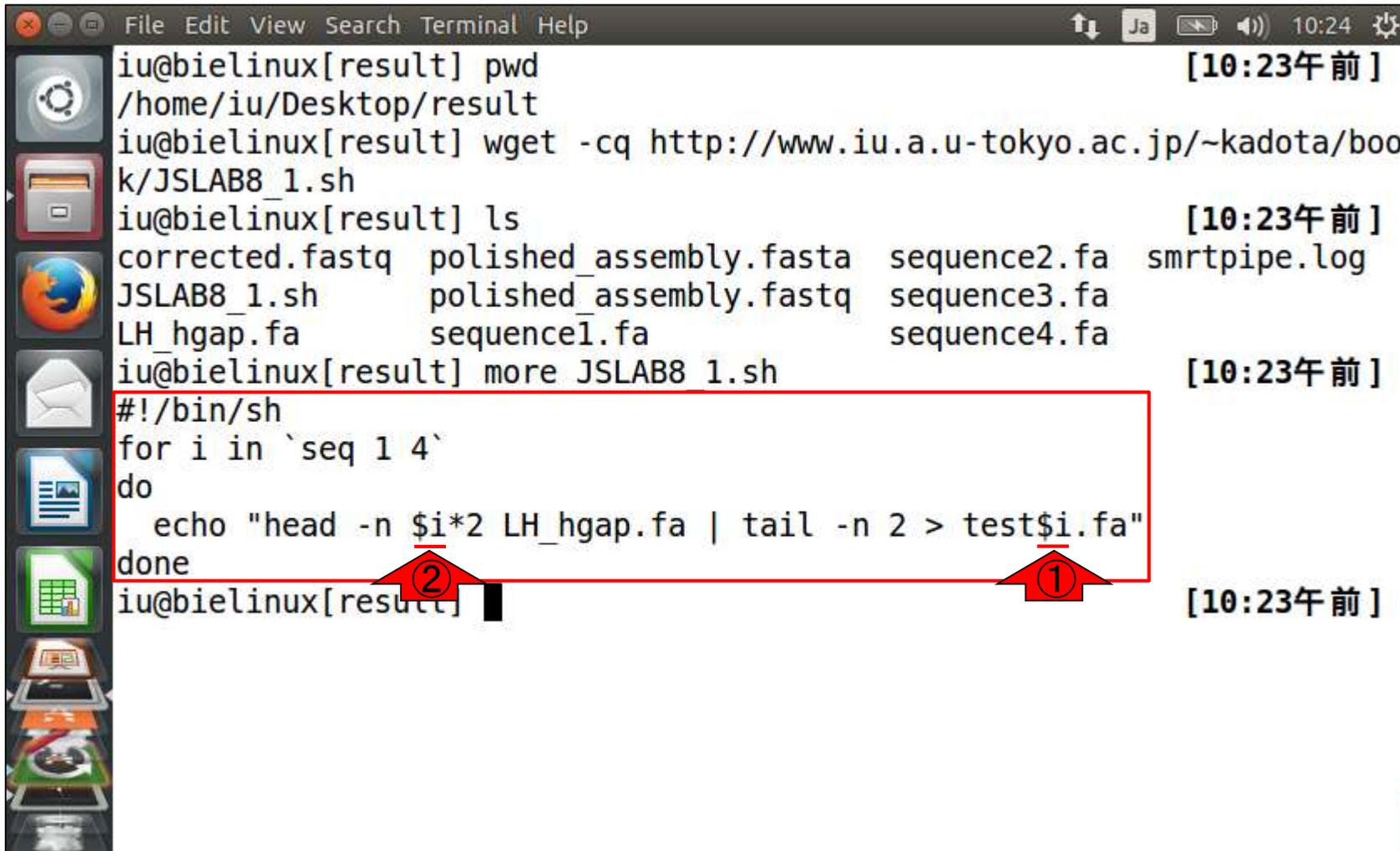
③

②

変数*i*は、*\$i*として取り扱う。つまり\$を追加する。
なぜ?という類のものではなく、お約束です

W3-5: 解説

```
iu@bielinux[result] pwd [10:23午前]
/home/iu/Desktop/result
iu@bielinux[result] wget -cq http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB8_1.sh
iu@bielinux[result] ls [10:23午前]
corrected.fastq polished_assembly.fasta sequence2.fa smrtpipe.log
JSLAB8_1.sh polished_assembly.fastq sequence3.fa
LH_hgap.fa sequence1.fa sequence4.fa
iu@bielinux[result] more JSLAB8_1.sh [10:23午前]
#!/bin/sh
for i in `seq 1 4`
do
    echo "head -n $i*2 LH_hgap.fa | tail -n 2 > test$i.fa"
done
iu@bielinux[result] [10:23午前]
```



W3-5: 解説

①「echo “…”」で困っていますが、これはLinuxコマンドのechoであり、“”内部の実際に実行したいコマンドがどんな感じになっているかを実行前に確認するためにつけています

```
iu@bielinux[result] pwd [10:23午前]
/home/iu/Desktop/result
iu@bielinux[result] wget -cq http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB8_1.sh
iu@bielinux[result] ls [10:23午前]
corrected.fastq polished_assembly.fasta sequence2.fa smrtpipe.log
JSLAB8_1.sh polished_assembly.fastq sequence3.fa
LH_hgap.fa sequence1.fa sequence4.fa
iu@bielinux[result] more JSLAB8_1.sh [10:23午前]
#!/bin/sh
for i in `seq 1 4`
do
  echo "head -n $i*2 LH_hgap.fa | tail -n 2 > test$i.fa"
done
iu@bielinux[result] [10:23午前]
```

W3-6: echoで確認

①shコマンドでJSLAB8_1.shを実行。②echoで囲った中身に相当する、赤下線部分が表示されていることがわかる

```
iu@bielinux[result] pwd [10:23午前]
/home/iu/Desktop/result
iu@bielinux[result] wget -cq http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB8_1.sh
iu@bielinux[result] ls [10:23午前]
corrected.fastq polished_assembly.fasta sequence2.fa smrtpipe.log
JSLAB8_1.sh polished_assembly.fastq sequence3.fa
LH_hgap.fa sequence1.fa sequence4.fa
iu@bielinux[result] more JSLAB8_1.sh [10:23午前]
#!/bin/sh
for i in `seq 1 4`
do
  echo "head -n $i*2 LH_hgap.fa | tail -n 2 > test$i.fa"
done
iu@bielinux[result] sh JSLAB8_1.sh [10:23午前]
head -n 1*2 LH_hgap.fa | tail -n 2 > test1.fa
head -n 2*2 LH_hgap.fa | tail -n 2 > test2.fa
head -n 3*2 LH_hgap.fa | tail -n 2 > test3.fa
head -n 4*2 LH_hgap.fa | tail -n 2 > test4.fa
iu@bielinux[result] [10:57午前]
```



W3-7: head部分

実際に全体を実行してエラーに遭遇してもよいが一応説明。headコマンド部分の①では最初のy行という数値を指定するところだが、1*2とか4*2という掛け算の*が含まれている。このような指定法はダメ(だということをエラーに遭遇して学習する)

```
iu@bielinux[result] pwd
/home/iu/Desktop/result
iu@bielinux[result] wget -cq http://www.iu.a.u-tokyo.ac.jp/~kadota/boo
k/JSLAB8_1.sh
iu@bielinux[result] ls
corrected.fastq polished_assembly.fasta sequence2.fa smrtpipe.log
JSLAB8_1.sh polished_assembly.fastq sequence3.fa
LH_hgap.fa sequence1.fa sequence4.fa
iu@bielinux[result] more JSLAB8_1.sh
#!/bin/sh
for i in `seq 1 4`
do
  echo "head -n $i*2 LH_hgap.fa | tail -n 2 > test$i.fa"
done
iu@bielinux[result] sh JSLAB8_1.sh
head -n 1*2 LH_hgap.fa | tail -n 2 > test1.fa
head -n 2*2 LH_hgap.fa | tail -n 2 > test2.fa
head -n 3*2 LH_hgap.fa | tail -n 2 > test3.fa
head -n 4*2 LH_hgap.fa | tail -n 2 > test4.fa
iu@bielinux[result] █
```



W3-7: head部分

①や②を実際に実行してみても、赤下線部分が原因でエラーが出ます。エラーが出ないようにするには、③のように4*2の結果である8を与える必要があります

```
iu@bielinux[result] ls [10:23午前]
corrected.fastq polished_assembly.fasta sequence2.fa smrtpipe.log
JSLAB8_1.sh polished_assembly.fastq sequence3.fa
LH_hgap.fa sequence1.fa sequence4.fa
iu@bielinux[result] more JSLAB8_1.sh [10:23午前]
#!/bin/sh
for i in `seq 1 4`
do
    echo "head -n $i*2 LH_hgap.fa | tail -n 2 > test$i.fa"
done
iu@bielinux[result] sh JSLAB8_1.sh [10:23午前]
head -n 1*2 LH_hgap.fa | tail -n 2 > test1.fa
head -n 2*2 LH_hgap.fa | tail -n 2 > test2.fa
head -n 3*2 LH_hgap.fa | tail -n 2 > test3.fa
head -n 4*2 LH_hgap.fa | tail -n 2 > test4.fa
iu@bielinux[result] [10:57午前]
iu@bielinux[result] head -n 4*2 LH_hgap.fa [11:29午前]
zsh: no matches found: 4*2
iu@bielinux[result] head -n 4*2 LH_hgap.fa | tail -n 2 [11:31午前]
zsh: no matches found: 4*2
iu@bielinux[result] head -n 8 LH_hgap.fa | tail -n 2 [11:31午前]
```



①シェルスクリプトの発展形ファイルJSLAB8_2.sh
を、②wgetし、③moreで確認

W3-8: 発展形

```
iu@bielinux[result] pwd [11:53午前]
/home/iu/Desktop/result
② iu@bielinux[result] wget -cq http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB8_2.sh
iu@bielinux[result] ls [11:53午前]
corrected.fastq LH hgap.fa sequence1.fa sequence4.fa
JSLAB8_1.sh polished_assembly.fasta sequence2.fa smrtpipe.log
JSLAB8_2.sh polished_assembly.fastq sequence3.fa
③ iu@bielinux[result] more JSLAB8_2.sh [11:53午前]
#!/bin/sh
for i in `seq 1 4`
do
    j=`expr $i \* 2`
    echo "head -n $j LH hgap.fa | tail -n 2 > test$i.fa"
done
iu@bielinux[result] [11:53午前]
```



W3-8: 発展形

JSLAB8_1.shとの違いは、赤下線部分のみ。jという別の変数を用いて*\$i*2*を表現したいただけだが、シェルスクリプトの掟に従うと、このようになります

```
iu@bielinux[result] pwd [11:53午前]
/home/iu/Desktop/result
iu@bielinux[result] wget -cq http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB8_2.sh
iu@bielinux[result] ls [11:53午前]
corrected.fastq LH_hgap.fa sequence1.fa sequence4.fa
JSLAB8_1.sh polished_assembly.fasta sequence2.fa smrtpipe.log
JSLAB8_2.sh polished_assembly.fastq sequence3.fa
iu@bielinux[result] more JSLAB8_2.sh [11:53午前]
#!/bin/sh
for i in `seq 1 4`
do
  j=`expr $i \* 2`
  echo "head -n $j LH_hgap.fa | tail -n 2 > test$i.fa"
done
iu@bielinux[result] [11:53午前]
```

W3-8: 発展形

①は $\$i*2$ の*に相当する部分ですが、「なんでもよい、というワイルドカードの*」と「掛け算の*」の意味を区別する必要があります。後者の意味として用いたい場合に、②¥(入力できませんが、バックスラッシュです)を添えます

```
iu@bielinux[result] pwd
/home/iu/Desktop/result
iu@bielinux[result] wget -cq http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB8_2.sh
iu@bielinux[result] ls
corrected.fastq  LH_hgap.fa          sequence1.fa  sequence4.fa
JSLAB8_1.sh     polished_assembly.fasta  sequence2.fa  smrtpipe.log
JSLAB8_2.sh     polished_assembly.fastq  sequence3.fa
iu@bielinux[result] more JSLAB8_2.sh
#!/bin/sh
for i in `seq 1 4`
do
  j=`expr $i \* 2`
  echo "head -n 2 $j LH_hgap.fa | tail -n 2 > test$i.fa"
done
iu@bielinux[result]
```

[11:53午前]

[11:53午前]

[11:53午前]

[11:53午前]

W3-8: 発展形

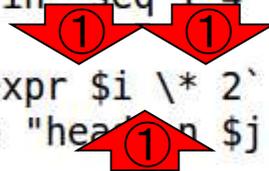
```
iu@bielinux[result] pwd [11:53午前]
/home/iu/Desktop/result
iu@bielinux[result] wget -cq http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB8_2.sh
iu@bielinux[result] ls [11:53午前]
corrected.fastq LH_hgap.fa sequence1.fa sequence4.fa
JSLAB8_1.sh polished_assembly.fasta sequence2.fa smrtpipe.log
JSLAB8_2.sh polished_assembly.fastq sequence3.fa
iu@bielinux[result] more JSLAB8_2.sh [11:53午前]
#!/bin/sh
for i in `seq 1 4`
do
    j=`expr $i \* 2`
    "head -n $j LH_hgap.fa | tail -n 2 > test$i.fa"
done
iu@bielinux[result] [11:53午前]
```



掟の続き。①のところにはスペースを入れないといけません

W3-8: 発展形

```
iu@bielinux[result] pwd [11:53午前]
/home/iu/Desktop/result
iu@bielinux[result] wget -cq http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB8_2.sh
iu@bielinux[result] ls [11:53午前]
corrected.fastq LH_hgap.fa sequence1.fa sequence4.fa
JSLAB8_1.sh polished_assembly.fasta sequence2.fa smrtpipe.log
JSLAB8_2.sh polished_assembly.fastq sequence3.fa
iu@bielinux[result] more JSLAB8_2.sh [11:53午前]
#!/bin/sh
for i in `seq 1 4`
do
    j=`expr $i \* 2`
    echo "head -n $j LH_hgap.fa | tail -n 2 > test$i.fa"
done
iu@bielinux[result] [11:53午前]
```



W3-8: 発展形

①の記号は、②Shiftキーを押しながら、③@のキーを押すと出ます

```
iu@bielinux[result] pwd
/home/iu/Desktop/result
iu@bielinux[result] wget -cq http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB8_2.sh
iu@bielinux[result] ls
corrected.fastq  LH_hgap.fa          sequence1.fa  sequence4.fa
JSLAB8_1.sh      polished_assembly.fa
JSLAB8_2.sh      polished_assembly.fa
iu@bielinux[result] more JSLAB8_2.sh
#!/bin/sh
for i in `seq 1 4`
do
  j=`expr $i \* 2`
  echo "head -n $j LH_hgap.fa | tail
done
iu@bielinux[result]
```



W3-9: echoで確認

- ①shコマンドでJSLAB8_2.shを実行。
- ②意図通りになっていることがわかる

```
iu@bielinux[result] pwd [ 1:54午後 ]
/home/iu/Desktop/result
iu@bielinux[result] ls [ 1:56午後 ]
corrected.fastq LH_hgap.fa sequence1.fa sequence4.fa
JSLAB8_1.sh polished_assembly.fasta sequence2.fa smrtpipe.log
JSLAB8_2.sh polished_assembly.fastq sequence3.fa
iu@bielinux[result] sh JSLAB8_2.sh [ 1:56午後 ]
head -n 2 LH_hgap.fa | tail -n 2 > test1.fa
head -n 4 LH_hgap.fa | tail -n 2 > test2.fa
head -n 6 LH_hgap.fa | tail -n 2 > test3.fa
head -n 8 LH_hgap.fa | tail -n 2 > test4.fa
iu@bielinux[result] [ 1:56午後 ]
```



W3-10: 発展形2

```
iu@bielinux[result] pwd [ 2:10午後 ]
/home/iu/Desktop/result
② iu@bielinux[result] wget -cq http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB8_3.sh
iu@bielinux[result] ls [ 2:10午後 ]
corrected.fastq LH_hgap.fa sequence2.fa
JSLAB8_1.sh polished_assembly.fasta sequence3.fa
JSLAB8_2.sh polished_assembly.fastq sequence4.fa
JSLAB8_3.sh sequence1.fa smrtpipe.log
③ iu@bielinux[result] more JSLAB8_3.sh [ 2:10午後 ]
#!/bin/sh
for i in `seq 1 4`
do
  j=`expr $i \* 2`
  #echo "head -n $j LH_hgap.fa | tail -n 2 > test$i.fa"
  head -n $j LH_hgap.fa | tail -n 2 > test$i.fa
done
iu@bielinux[result] [ 2:10午後 ]
```



W3-10: 発展形2

JSLAB8_2.shとの違いは、①と②の行の部分。①のecho行頭に#を入れてコメントアウト(実行されないように)している。削除するのと同じだが、エラーなど問題が起こったときの対処などの目的で#をつけたまま残しておくことはよくやる。②は①の赤下線部分と同じもの。ここが実際に実行される部分

```
iu@bielinux[result] pwd
/home/iu/Desktop/result
iu@bielinux[result] wget -cq http://www.iu.a.u-tokyo.ac.jp/~kadota/boo
k/JSLAB8_3.sh
iu@bielinux[result] ls
corrected.fastq  LH_hgap.fa          sequence2.fa
JSLAB8_1.sh      polished_assembly.fasta sequence3.fa
JSLAB8_2.sh      polished_assembly.fastq sequence4.fa
JSLAB8_3.sh      sequence1.fa         smrtpipe.log
iu@bielinux[result] more JSLAB8_3.sh
#!/bin/sh
for i in `seq 1 4`
do
  j=`expr $i \* 2`
  #echo "head -n $j LH_hgap.fa | tail -n 2 > test$i.fa"
  head -n $j LH_hgap.fa | tail -n 2 > test$i.fa
done
iu@bielinux[result] █
```

[2:10午後]

[2:10午後]

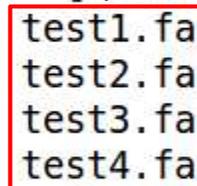
[2:10午後]



W3-11: 実行

①「sh JSLAB8_3.sh」を実行。echoをコメントアウトしているので、意図通り何も表示されない。また、②イメージ通りの出力ファイルtest[0-9].faが作成されている

```
iu@bielinux[result] pwd [ 2:43午後 ]
/home/iu/Desktop/result
iu@bielinux[result] ls [ 2:44午後 ]
corrected.fastq LH_hgap.fa sequence2.fa
JSLAB8_1.sh polished_assembly.fasta sequence3.fa
JSLAB8_2.sh polished_assembly.fastq sequence4.fa
JSLAB8_3.sh sequence1.fa smrtpipe.log
iu@bielinux[result] sh JSLAB8_3.sh [ 2:44午後 ]
iu@bielinux[result] ls [ 2:44午後 ]
corrected.fastq LH_hgap.fa sequence2.fa test1.fa
JSLAB8_1.sh polished_assembly.fasta sequence3.fa test2.fa
JSLAB8_2.sh polished_assembly.fastq sequence4.fa test3.fa
JSLAB8_3.sh sequence1.fa smrtpipe.log test4.fa
iu@bielinux[result] [ 2:44午後 ]
```



W3-12: 確認

①「ls -l」で詳細情報を表示。ファイルサイズの観点からは、sequence[0-9].faとtest[0-9].faは全く同じ。②念のためdiffコマンドでも確認。何も表示されていないので中身も同一であることを確認したことになる。③違うもの同士だと、違いのある部分がババッと表示されます

```
File Edit View Search Terminal Help
iu@bielinux[result] pwd
/home/iu/Desktop/result
iu@bielinux[result] ls -l sequence* test*
-rw-rw-r-- 1 iu iu 2289509  8月 29 14:06 sequence1.fa
-rw-rw-r-- 1 iu iu   86904  8月 29 14:06 sequence2.fa
-rw-rw-r-- 1 iu iu   45865  8月 29 14:06 sequence3.fa
-rw-rw-r-- 1 iu iu   11384  8月 29 14:06 sequence4.fa
-rw-rw-r-- 1 iu iu 2289509  8月 30 14:44 test1.fa
-rw-rw-r-- 1 iu iu   86904  8月 30 14:44 test2.fa
-rw-rw-r-- 1 iu iu   45865  8月 30 14:44 test3.fa
-rw-rw-r-- 1 iu iu   11384  8月 30 14:44 test4.fa
iu@bielinux[result] diff sequence1.fa test1.fa
iu@bielinux[result] diff sequence2.fa test2.fa
iu@bielinux[result] diff sequence3.fa test3.fa
iu@bielinux[result] diff sequence4.fa test4.fa
iu@bielinux[result] diff sequence4.fa sequence3.fa
```

[3:24午後]
[3:24午後]
[3:24午後]
[3:24午後]
[3:24午後]

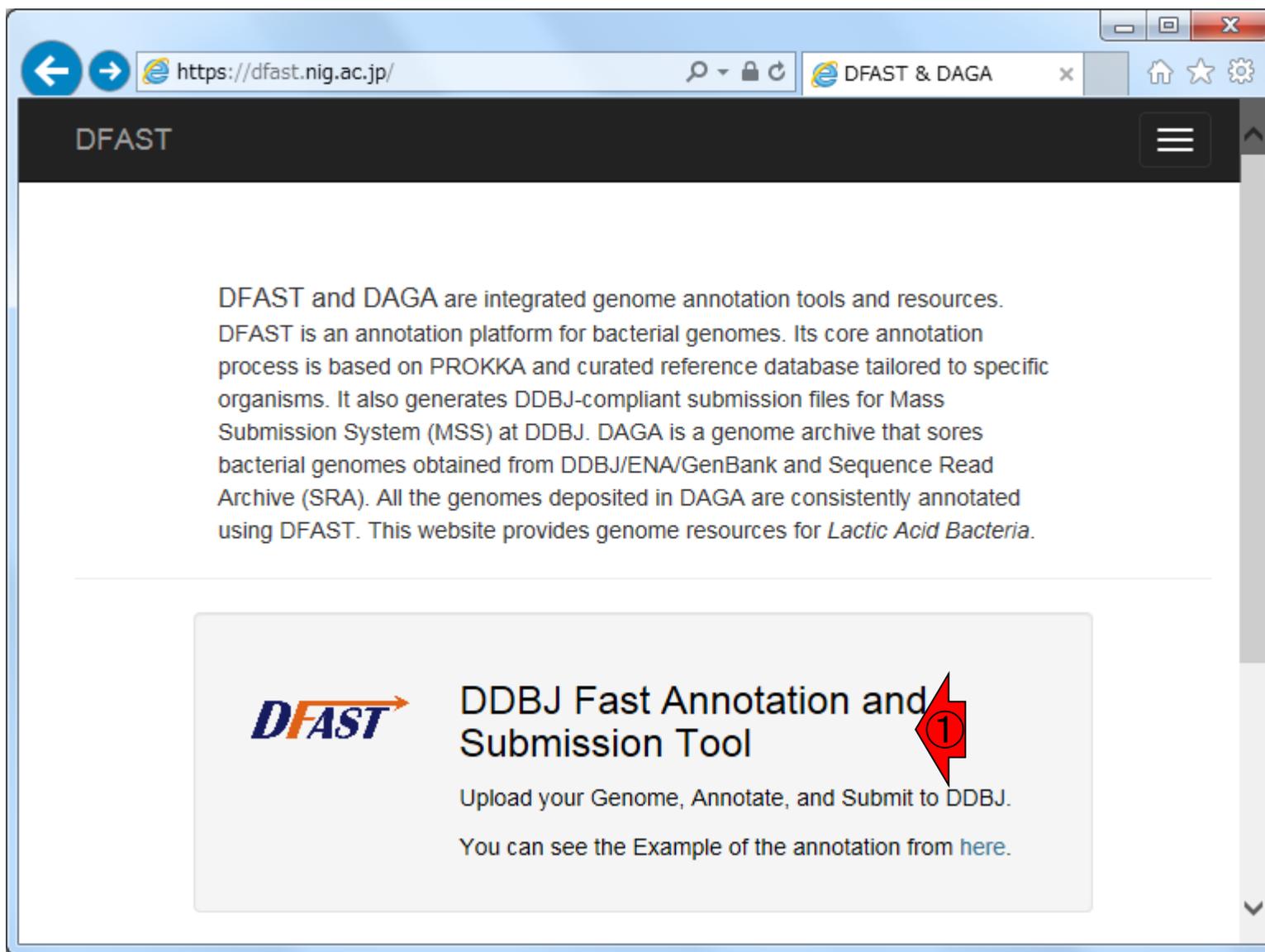
W3-13: 発展形3

「expr シェル 遅い」などでググればわかりますが、JSLAB8_3.sh内で使われているexprは遅いらしいです。赤下線のように書くといいらしい。①高速版のJSLAB8_4.shを②wgetし、③more。後は省略

```
iu@bielinux[result] pwd
/home/iu/Desktop/result
iu@bielinux[result] wget -cq http://www.iu.a.u-tokyo.ac.jp/~kadota/book/JSLAB8_4.sh
iu@bielinux[result] more JSLAB8_4.sh
#!/bin/sh
for i in `seq 1 4`
do
  j=$((i \* 2))
  #echo "head -n $j LH_hgap.fa | tail -n 2 > test$i.fa"
  head -n $j LH_hgap.fa | tail -n 2 > test$i.fa
done
iu@bielinux[result] █
```



W4-1: DFAST



DFAST and DAGA are integrated genome annotation tools and resources. DFAST is an annotation platform for bacterial genomes. Its core annotation process is based on PROKKA and curated reference database tailored to specific organisms. It also generates DDBJ-compliant submission files for Mass Submission System (MSS) at DDBJ. DAGA is a genome archive that stores bacterial genomes obtained from DDBJ/ENA/GenBank and Sequence Read Archive (SRA). All the genomes deposited in DAGA are consistently annotated using DFAST. This website provides genome resources for *Lactic Acid Bacteria*.

DFAST **DDBJ Fast Annotation and Submission Tool**

Upload your Genome, Annotate, and Submit to DDBJ.

You can see the Example of the annotation from [here](#).

W4-1: DFAST

①のところで、参照ボタンを押して、アノテーションしたいmulti-FASTAファイル(LH_hgap.fa)を指定

The screenshot shows the DFAST web interface. The browser address bar displays `https://dfast.nig.ac.jp/analysis/annotation/`. The page title is "DFAST: DDBJ Fast ...". The main content area includes a description of DFAST as an annotation platform for bacterial genomes. Below the description, there are two input fields: "Query File (Fasta format)" and "Job Title". The "Query File" field has a "参照..." button next to it, which is highlighted with a red arrow and a circled "1". The "Job Title" field contains the text "(optional)". Below these fields is a "Mail Address" field with the placeholder text "E-mail notification will be sent to this address when the job is completed. (opti...". At the bottom, there is a section titled "Specify metadata and parameters." with a note that data other than minimum contig length can be altered later and that reference databases for genera other than Lactobacillus and Pediococcus are not fully supported.

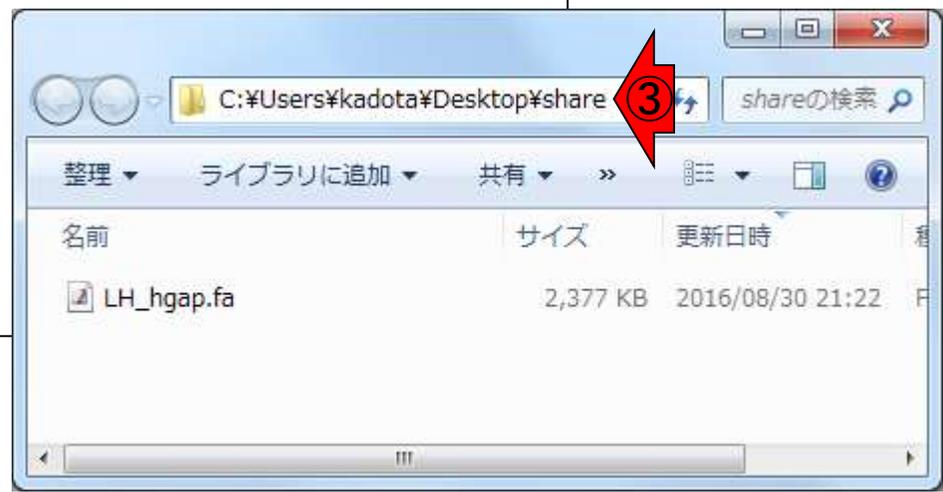
W4-2:LH_hgap.fa

①アノテーションしたいmulti-FASTAファイル(LH_hgap.fa)を共有フォルダにコピー。②共有フォルダ内の状況はヒトそれぞれだが、最低限LH_hgap.faが見えていけばよい。(私のWindows環境では)ホストOS上の絶対パスとして③のところにLH_hgap.faが見えるので、これをアップロードします

```
iu@bielinux[~/Desktop/result]
iu@bielinux[result] pwd
/home/iu/Desktop/result
iu@bielinux[result] ls
corrected.fastq  LH_hgap.fa          sequence3.fa  test3.fa
JSLAB8_1.sh     polished_assembly.fasta  sequence4.fa  test4.fa
JSLAB8_2.sh     polished_assembly.fastq  smrtpipe.log
JSLAB8_3.sh     sequence1.fa          test1.fa
JSLAB8_4.sh     sequence2.fa          test2.fa
① iu@bielinux[result] cp LH_hgap.fa ../mac_share
iu@bielinux[result] ls -l ../mac share
total 2377
-rwxrwxrwx 1 iu iu 2433662  8月 30 21:22 LH_hgap.fa
iu@bielinux[result] █
```

[9:21午後]
[9:22午後]

[9:22午後]



W4-3: アップロードと実行

①参照ボタン、②に移動して、③目的のファイル(LH_hgap.fa)を指定して、④開く

The image shows a web browser window displaying the DFAST website. The browser address bar shows the URL `https://dfast.nig.ac.jp/analysis/annotation/`. The page content includes a description of DFAST and a form for uploading a query file. A red arrow labeled ① points to the '参照...' button in the 'Query File (Fasta format)' section. A file selection dialog box is overlaid on the page, showing the path `C:\Users\kadota\Desktop\share` (2) and a list of files. The file `LH_hgap.fa` is selected (3). A red arrow labeled ④ points to the '開く(O)' button at the bottom of the dialog.

Query File (Fasta format)

参照... ①

Job Title

(optional)

Mail Address

E-mail notification will be sent to this address when the

Specify metadata and parameters.

These data other than minimum contig length can be altered. Databases for genera other than Lactobacillus and Pedicoccus are supported.

アップロードするファイルの選択

C:\Users\kadota\Desktop\share ②

名前	更新日時	種類	サイズ
LH_hgap.fa ③	2016/08/30 21:22	FA ファイル	2,377 KB

ファイル名(N): LH_hgap.fa

開く(O) ④ キャンセル

W4-3: アップロードと実行

①ジョブタイトル(hogegeee)をテキストにつけて、②計算が終わったらメールでお知らせしてくれるので入力して、③ページ下部に移動

DFAST is an annotation platform for bacterial genomes. Its core annotation process is based on PROKKA and curated reference database tailored to specific organisms. It also generates DDBJ-compliant submission files for Mass Submission System (MSS) at DDBJ.

Query File (Fasta format)
C:\User: 参照...

Job Title
hogegeee ①

Mail Address
kadota@iu.a.u-tokyo.ac.jp ②

Specify metadata and parameters.
These data other than minimum contig length can be altered later. Reference Databases for genera other than Lactobacillus and Pediococcus are not fully supported.

Genus Lactob
Species sp
Strain unkown



W4-3: アップロードと実行

①いろいろオプション指定できるがとりあえずここは無視して、②Run。③は設定値(デフォルトは200 bp)以下の短い配列を除くため。それ以外の赤枠の属・種名等のオプションはアノテーション結果に影響を与えることはなく、後で変更することも可能であるため、規定値のままでOK

Specify metadata and parameters.
These data other than minimum contig length can be altered later. Reference Databases for genera other than Lactobacillus and Pediococcus are not fully supported.

Genus	Species	Strain
Lactob <input type="button" value="v"/>	sp. ex) plantarum, delbrueckii subsp. bulgaricus	unkown

Locus Tag Prefix **Minimum Contig Length**

LOCUS 200

Perform Genome Assessment (optional)

W4-3: アップロードと実行

計算が始まったようだ。とりあえず計算終了メールが来るまで思考停止

DFAST

Remember the [current URL](#) to access this page. The result will be deleted 30 days after your last visit.
Delete this job now. => This procedure cannot be undone.

Title : hogegeee

JobID : f341d803-072b-48db-a363-1de4c3a686a5

Status : RUNNING

[Features](#) [DDBJ Submission](#) [Log](#)

[2016-08-31 15:18:55.909400] Job submitted.
[2016-08-31 15:18:55.932959] Job started.

計算終了メールが届く。このときは、確かに5分以内に計算が終了。①にアクセス

W4-4: 計算終了

 2016/08/31 (水) 15:22
DFAST Report <dfast@nig.ac.jp>
[DFAST] Your requested job completed.

宛先 kadota@iu.a.u-tokyo.ac.jp

Your requested job has completed.

Job Title : hogegeee
Job ID : f341d803-072b-48db-a363-1de4c3a686a5
Submitted at :2016-08-31 15:18:55.909400.

Please visit the link below to check the result.
<https://dfast.nig.ac.jp/analysis/annotation/f341d803-072b-48db-a363-1de4c3a686a5>

DFAST <dfast@nig.ac.jp>



計算結果が見えている。
①ページ下部に移動

W4-5: 結果を眺める

DFAST

Remember the [current URL](#) to access this page. The result will be deleted 30 days after your last visit.
Delete this job now. => This procedure cannot be undone.

Title : hogegeee

JobID : f341d803-072b-48db-a363-1de4c3a686a5

Status : COMPLETE

```
[2016-08-31 15:18:55.909400] Job submitted.  
[2016-08-31 15:18:55.932959] Job started.  
[2016-08-31 15:21:52.789970] Job completed.
```

Result | Features | DDBJ Submission | Log

Genome Statistics

Total Length (bp)	2,433,614
No. of Sequences	4

Download Files

Genbank Flat
File : [annotation.gbk](#)

GFF3-formated

W4-5: 結果を眺める

これがDFASTによるアノテーション結果の全体像。①このあたりの数値は、DDBJ Pipeline上でHGAPを実行したときの結果(第7回W9-2)と全く同じで妥当

The screenshot shows the DFAST Job Result page. The browser address bar displays <https://dfast.nig.ac.jp/analysis/annotation/f341d803>. The page has tabs for 'Result', 'Features', 'DDBJ Submission', and 'Log'. The 'Result' tab is active, showing 'Genome Statistics' and 'Download Files'.

Genome Statistics	
Total Length (bp)	2,433,614
No. of Sequences	4
GC Content (%)	38.2%
N50	2,289,497
Gap Ratio (%)	0.0%
No. of CDSs	2,389
No. of rRNA	12
No. of tRNA	56
No. of CRISPRS	1
Coding Ratio (%)	86.7%

The 'Download Files' section lists the following files:

- Genbank Flat File : [annotation.gbk](#)
- GFF3-formated File : [annotation.gff](#)
- Genome Fasta File : [genome.fna](#)
- Protein Fasta File : [protein.faa](#)
- CDS Fasta File : [cds.fna](#)
- RNA Fasta File : [rna.fna](#)
- Feature Table : [features.tsv](#)
- Genome Statistics :

A red box highlights the 'Total Length (bp)' and 'No. of Sequences' rows, with a red arrow pointing to the value '4' in the 'No. of Sequences' row, labeled with a circled '1'.

W4-5: 結果を眺める

①アノテーションファイルの一般的な形式であるGFFファイルをダウンロードしてエクセルなどで眺めてもよいが、ここでは②Featuresタブをクリックして、ウェブ上でアノテーション結果を眺める

Result **Features** DDBJ Submission Log

Genome Statistics

Total Length (bp)	2,433,614
No. of Sequences	4
GC Content (%)	38.2%
N50	2,289,497
Gap Ratio (%)	0.0%
No. of CDSs	2,389
No. of rRNA	12
No. of tRNA	56
No. of CRISPRS	1
Coding Ratio (%)	86.7%

Download Files

Genbank Flat
File : [annotation.gbk](#)

GFF3-formated
File : [annotation.gff](#) ①

Genome Fasta
File : [genome.fna](#)

Protein Fasta
File : [protein.faa](#)

CDS Fasta File : [cds.fna](#)

RNA Fasta File : [ma.fna](#)

Feature Table : [features.tsv](#)

Genome Statistics :

W4-5: 結果を眺める

こんな感じになります。①デフォルトはページあたり25エントリーになっているので、②の部分が25行分ずつ表示されます

Result Features DDBJ Submission Log

Annotated Features

Show 25 entries Search:

	LocusTag	Seq. ID	Location	Feature Type	Product	Gene	Nucleotide	Translation	Edit
1	LOCUS_00001	sequence1	151..384	CDS	hypothetical protein		View	View	Edit
2	LOCUS_00002	sequence1	350..886	CDS	hypothetical protein		View	View	Edit
3	LOCUS_00003	sequence1	883..1311	CDS	hypothetical protein		View	View	Edit
4	LOCUS_00004	sequence1	1637..1849	CDS	hypothetical protein		View	View	Edit
5	LOCUS_00005	sequence1	1968..2165	CDS	hypothetical protein		View	View	Edit
6	LOCUS_00006	sequence1	2355..2732	CDS	prophage protein		View	View	Edit
7	LOCUS_00007	sequence1	2725..2940	CDS	hypothetical protein		View	View	Edit

全エントリーを一気に表示させて説明したいので、①Allにする

W4-5: 結果を眺める

Result Features DDBJ Submission Log

Annotated Features

Show **25** entries Search:

50
100
All

No.	Locu Tag	Seq. ID	Location	Feature Type	Product	Gene	Nucleotide	Translation	Edit
1	LOCUS_00001	sequence1	151..384	CDS	hypothetical protein		View	View	Edit
2	LOCUS_00002	sequence1	350..886	CDS	hypothetical protein		View	View	Edit
3	LOCUS_00003	sequence1	883..1311	CDS	hypothetical protein		View	View	Edit
4	LOCUS_00004	sequence1	1637..1849	CDS	hypothetical protein		View	View	Edit
5	LOCUS_00005	sequence1	1968..2165	CDS	hypothetical protein		View	View	Edit
6	LOCUS_00006	sequence1	2355..2732	CDS	prophage protein		View	View	Edit
7	LOCUS_00007	sequence1	2725..2940	CDS	hypothetical protein		View	View	Edit

W4-5: 結果を眺める

こんな感じになります。横幅が広がるが細かいことは気にしない。①のあたりを見ることで、入力ファイル(LH_hgap.fa)の配列および座標順にアノテーションされた結果が表示されていることがわかる

Result Features DDBJ Submission Log

Annotated Features

Show entries Search:

No.	Locus Tag	Seq. ID	Position	Feature Type	Product	Gene	Nucleotide	Transcript
1	LOCUS_00001	sequence1	151..384	CDS	hypothetical protein		<input type="button" value="View"/>	<input type="button" value="View"/>
2	LOCUS_00002	sequence1	350..886	CDS	hypothetical protein		<input type="button" value="View"/>	<input type="button" value="View"/>
3	LOCUS_00003	sequence1	883..1311	CDS	hypothetical protein		<input type="button" value="View"/>	<input type="button" value="View"/>
4	LOCUS_00004	sequence1	1637..1849	CDS	hypothetical protein		<input type="button" value="View"/>	<input type="button" value="View"/>
5	LOCUS_00005	sequence1	1968..2165	CDS	hypothetical protein		<input type="button" value="View"/>	<input type="button" value="View"/>
6	LOCUS_00006	sequence1	2355..2732	CDS	prophage protein		<input type="button" value="View"/>	<input type="button" value="View"/>
7	LOCUS_00007	sequence1	2725..2940	CDS	hypothetical protein		<input type="button" value="View"/>	<input type="button" value="View"/>
8	LOCUS_00008	sequence1	2930..3031	CDS	hypothetical protein		<input type="button" value="View"/>	<input type="button" value="View"/>
9	LOCUS_00009	sequence1	3067..3534	CDS	holin		<input type="button" value="View"/>	<input type="button" value="View"/>
10	LOCUS_00010	sequence1	3547..4578	CDS	1,4-beta-N-acetylmuramidase		<input type="button" value="View"/>	<input type="button" value="View"/>

W4-6: sequence1 概観

①と②を関連づけながら、アノテーション結果の全体像を把握する

Result Features DDBJ Submission Log

Annotated Features

Show **All** entries Search:

No.	Locus Tag	Seq. ID	Position	Feature Type	Product	Gene	Nucleotide	Translation
1	LOCUS_00001	sequence1	151..384	CDS	hypothetical protein		View	View
2	LOCUS_00002	sequence1	350..886	CDS	hypothetical protein		View	View
3	LOCUS_00003	sequence1	883..1311	CDS	hypothetical protein		View	View
4	LOCUS_00004	sequence1	1637..1849	CDS	hypothetical protein		View	View
5	LOCUS_00005	sequence1	1968..2165	CDS	hypothetical protein		View	View
6	LOCUS_00006	sequence1	2355..2732	CDS	prophage protein		View	View
7	LOCUS_00007	sequence1	2725..2940	CDS	hypothetical protein		View	View
8	LOCUS_00008	sequence1	2930..3031	CDS	hypothetical protein		View	View
9	LOCUS_00009	sequence1	3067..3534	CDS	holin		View	View
10	LOCUS_00010	sequence1	3547..4578	CDS	1,4-beta-N-acetylmuramidase		View	View

W4-6: sequence1 概観

① sequence1の左端(最初の3031 bpまで)は「hypothetical proteinが多いなあ…」とか、②「prophage proteinがある」とか…

Result Features DDBJ Submission Log

Annotated Features

Show entries Search:

No.	Locus Tag	Seq. ID	Location	Feature Type	Product	Gene	Nucleotide	Transcript
1	LOCUS_00001	sequence1	151..384	CDS	hypothetical protein		<input type="button" value="View"/>	<input type="button" value="View"/>
2	LOCUS_00002	sequence1	350..886	CDS	hypothetical protein		<input type="button" value="View"/>	<input type="button" value="View"/>
3	LOCUS_00003	sequence1	883..1311	CDS	hypothetical protein		<input type="button" value="View"/>	<input type="button" value="View"/>
4	LOCUS_00004	sequence1	1637..1849	CDS	hypothetical protein		<input type="button" value="View"/>	<input type="button" value="View"/>
5	LOCUS_00005	sequence1	1968..2165	CDS	hypothetical protein		<input type="button" value="View"/>	<input type="button" value="View"/>
6	LOCUS_00006	sequence1	2355..2732	CDS	prophage protein		<input type="button" value="View"/>	<input type="button" value="View"/>
7	LOCUS_00007	sequence1	2725..2940	CDS	hypothetical protein		<input type="button" value="View"/>	<input type="button" value="View"/>
8	LOCUS_00008	sequence1	2930..3031	CDS	hypothetical protein		<input type="button" value="View"/>	<input type="button" value="View"/>
9	LOCUS_00009	sequence1	3067..3534	CDS	holin		<input type="button" value="View"/>	<input type="button" value="View"/>
10	LOCUS_00010	sequence1	3547..4578	CDS	1,4-beta-N-acetylmuramidase		<input type="button" value="View"/>	<input type="button" value="View"/>

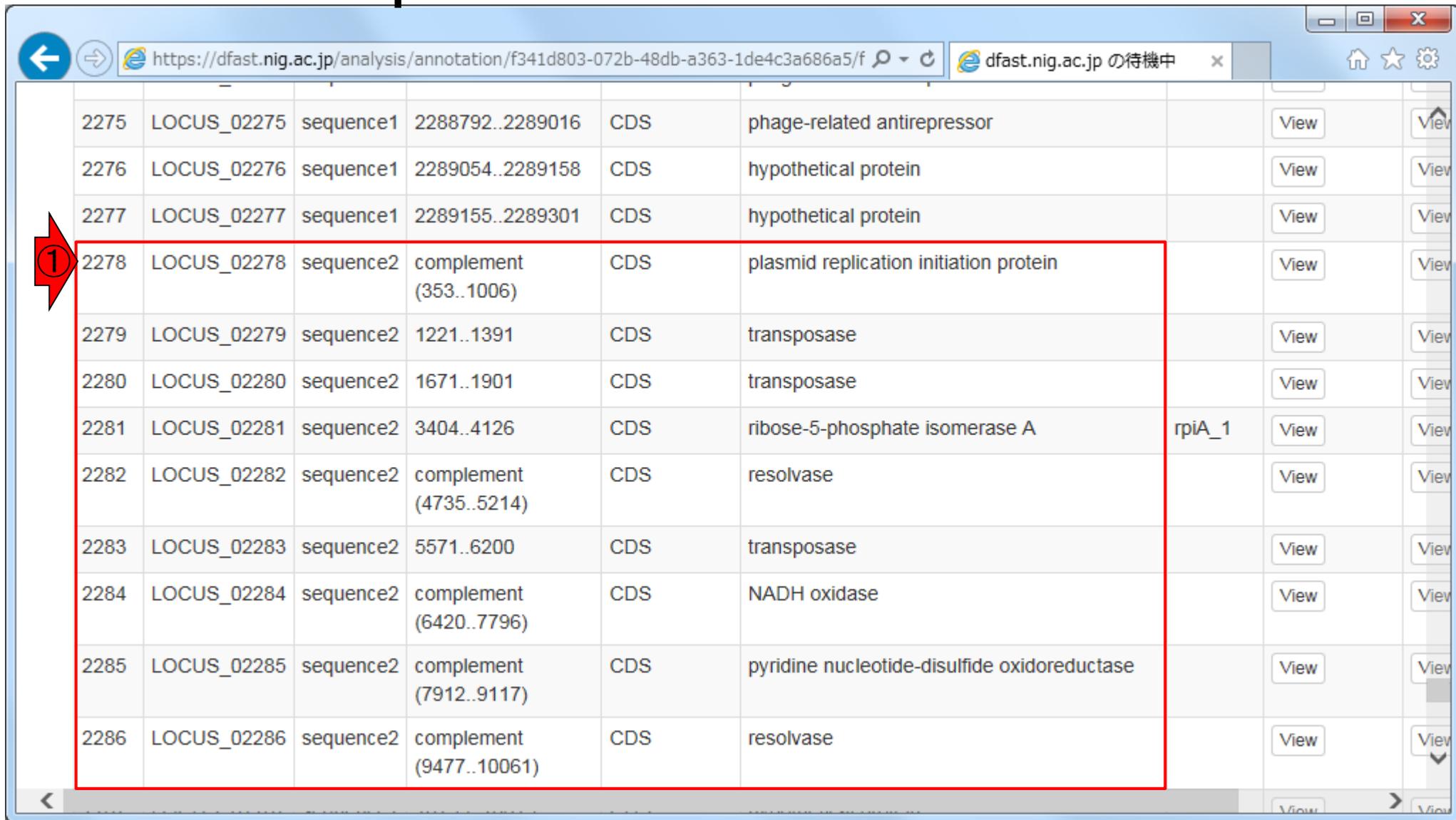
W4-6: sequence1 概観

赤枠はsequence1の①19649から②34418 bpの範囲(全部で2,289,497 bpの長さがあるので、このあたりもまだ左端といえる)。赤下線で示すように、このあたりにも**ファージ(phage)**関連のものがちらほら存在

LOCUS	sequence1	Start	End	Type	Protein Name	Gene Name	Action
41	sequence1	19314	19339	CDS	hypothetical protein		View
42	sequence1	19649	20095	CDS	<u>phage protein</u>		View
43	sequence1	20310	20585	CDS	hypothetical protein		View
44	sequence1	20578	21057	CDS	hypothetical protein		View
45	sequence1	21180	21698	CDS	<u>phage terminase small subunit</u>	xtmA	View
46	sequence1	21698	23599	CDS	<u>phage terminase large subunit</u>	xtmB	View
47	sequence1	23609	23785	CDS	hypothetical protein		View
48	sequence1	23786	24961	CDS	<u>phage portal protein</u>		View
49	sequence1	24942	26834	CDS	<u>phage capsid protein</u>		View
50	sequence1	27026	27313	CDS	hypothetical protein		View
51	sequence1	27294	27671	CDS	hypothetical protein		View
52	sequence1	27674	28096	CDS	hypothetical protein		View
53	sequence1	28096	28476	CDS	hypothetical protein		View
54	sequence1	28480	29088	CDS	<u>phage tail protein</u>		View
55	sequence1	29239	29589	CDS	tail protein		View
56	sequence1	29793	34418	CDS	phage tail tape measure protein		View

①ここからが、sequence2
のアノテーション結果

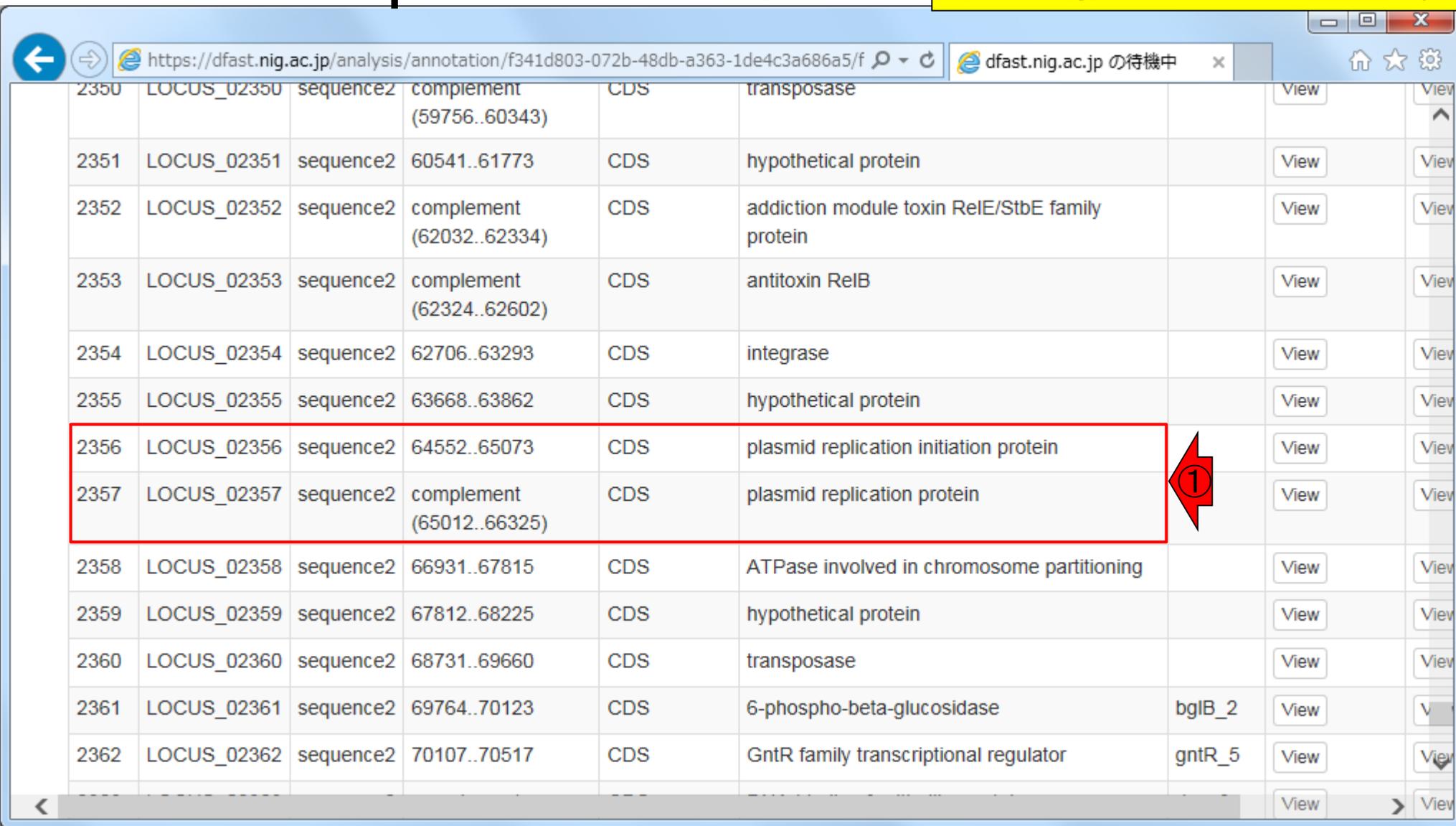
W4-7: sequence2概観



2275	LOCUS_02275	sequence1	2288792..2289016	CDS	phage-related antirepressor		View	View	
2276	LOCUS_02276	sequence1	2289054..2289158	CDS	hypothetical protein		View	View	
2277	LOCUS_02277	sequence1	2289155..2289301	CDS	hypothetical protein		View	View	
1	2278	LOCUS_02278	sequence2	complement (353..1006)	CDS	plasmid replication initiation protein		View	View
	2279	LOCUS_02279	sequence2	1221..1391	CDS	transposase		View	View
	2280	LOCUS_02280	sequence2	1671..1901	CDS	transposase		View	View
	2281	LOCUS_02281	sequence2	3404..4126	CDS	ribose-5-phosphate isomerase A	rpiA_1	View	View
	2282	LOCUS_02282	sequence2	complement (4735..5214)	CDS	resolvase		View	View
	2283	LOCUS_02283	sequence2	5571..6200	CDS	transposase		View	View
	2284	LOCUS_02284	sequence2	complement (6420..7796)	CDS	NADH oxidase		View	View
	2285	LOCUS_02285	sequence2	complement (7912..9117)	CDS	pyridine nucleotide-disulfide oxidoreductase		View	View
	2286	LOCUS_02286	sequence2	complement (9477..10061)	CDS	resolvase		View	View

W4-7: sequence2概観

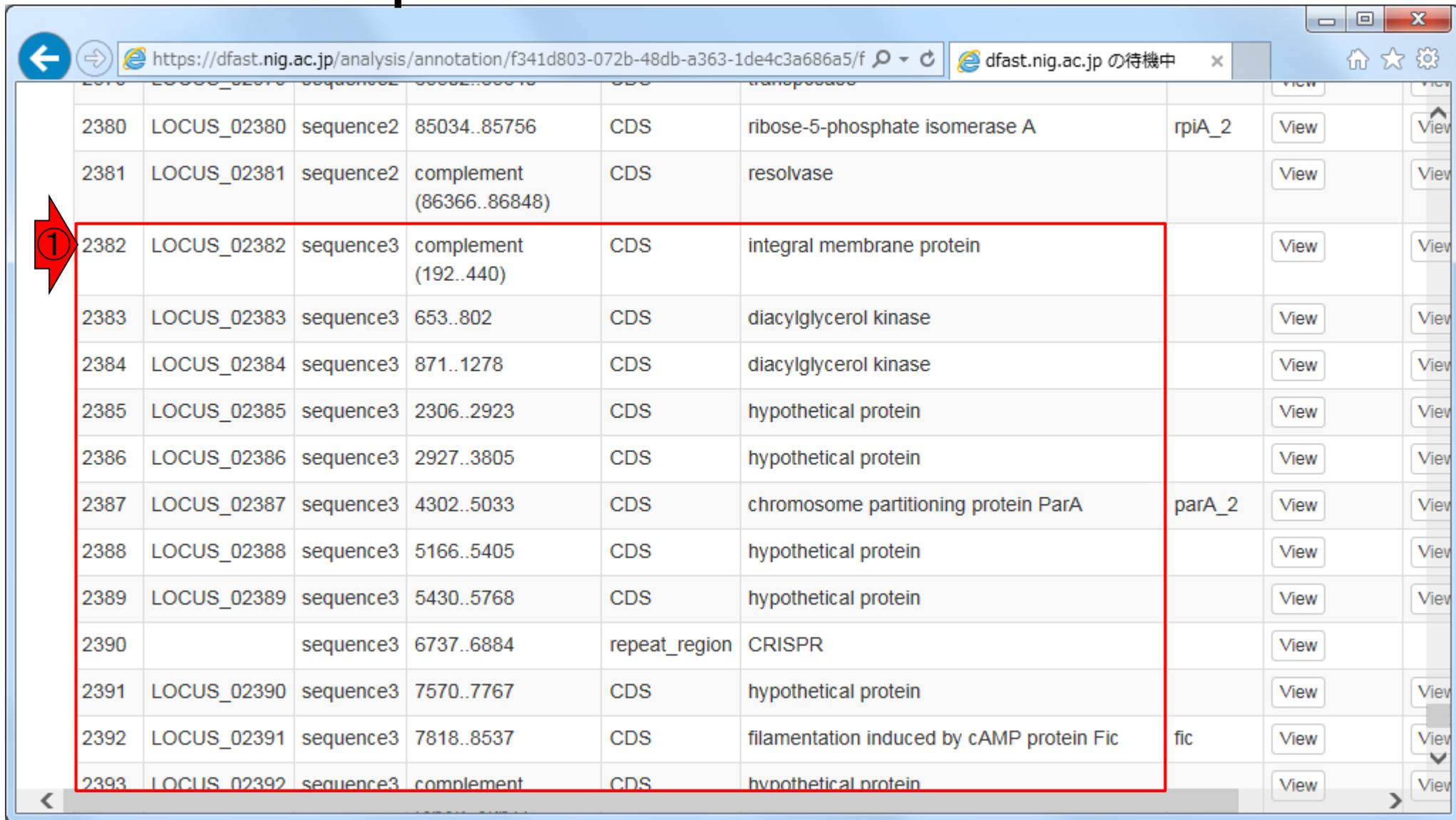
第7回でsequence2は環状のプラスミド配列だろうと予想していたが、①それを補強するアノテーション結果



LOCUS	sequence2	coordinates	CDS	protein	View
2350	LOCUS_02350	sequence2	complement (59756..60343)	transposase	View
2351	LOCUS_02351	sequence2	60541..61773	hypothetical protein	View
2352	LOCUS_02352	sequence2	complement (62032..62334)	addiction module toxin RelE/StbE family protein	View
2353	LOCUS_02353	sequence2	complement (62324..62602)	antitoxin RelB	View
2354	LOCUS_02354	sequence2	62706..63293	integrase	View
2355	LOCUS_02355	sequence2	63668..63862	hypothetical protein	View
2356	LOCUS_02356	sequence2	64552..65073	plasmid replication initiation protein	View
2357	LOCUS_02357	sequence2	complement (65012..66325)	plasmid replication protein	View
2358	LOCUS_02358	sequence2	66931..67815	ATPase involved in chromosome partitioning	View
2359	LOCUS_02359	sequence2	67812..68225	hypothetical protein	View
2360	LOCUS_02360	sequence2	68731..69660	transposase	View
2361	LOCUS_02361	sequence2	69764..70123	6-phospho-beta-glucosidase	bgIB_2 View
2362	LOCUS_02362	sequence2	70107..70517	GntR family transcriptional regulator	gntR_5 View

①ここからが、sequence3
のアノテーション結果

W4-8: sequence3概観



2380	LOCUS_02380	sequence2	85034..85756	CDS	ribose-5-phosphate isomerase A	rpiA_2	View	View
2381	LOCUS_02381	sequence2	complement (86366..86848)	CDS	resolvase		View	View
2382	LOCUS_02382	sequence3	complement (192..440)	CDS	integral membrane protein		View	View
2383	LOCUS_02383	sequence3	653..802	CDS	diacylglycerol kinase		View	View
2384	LOCUS_02384	sequence3	871..1278	CDS	diacylglycerol kinase		View	View
2385	LOCUS_02385	sequence3	2306..2923	CDS	hypothetical protein		View	View
2386	LOCUS_02386	sequence3	2927..3805	CDS	hypothetical protein		View	View
2387	LOCUS_02387	sequence3	4302..5033	CDS	chromosome partitioning protein ParA	parA_2	View	View
2388	LOCUS_02388	sequence3	5166..5405	CDS	hypothetical protein		View	View
2389	LOCUS_02389	sequence3	5430..5768	CDS	hypothetical protein		View	View
2390		sequence3	6737..6884	repeat_region	CRISPR		View	
2391	LOCUS_02390	sequence3	7570..7767	CDS	hypothetical protein		View	View
2392	LOCUS_02391	sequence3	7818..8537	CDS	filamentation induced by cAMP protein Fic	fic	View	View
2393	LOCUS_02392	sequence3	complement	CDS	hypothetical protein		View	View

W4-8: sequence3概観

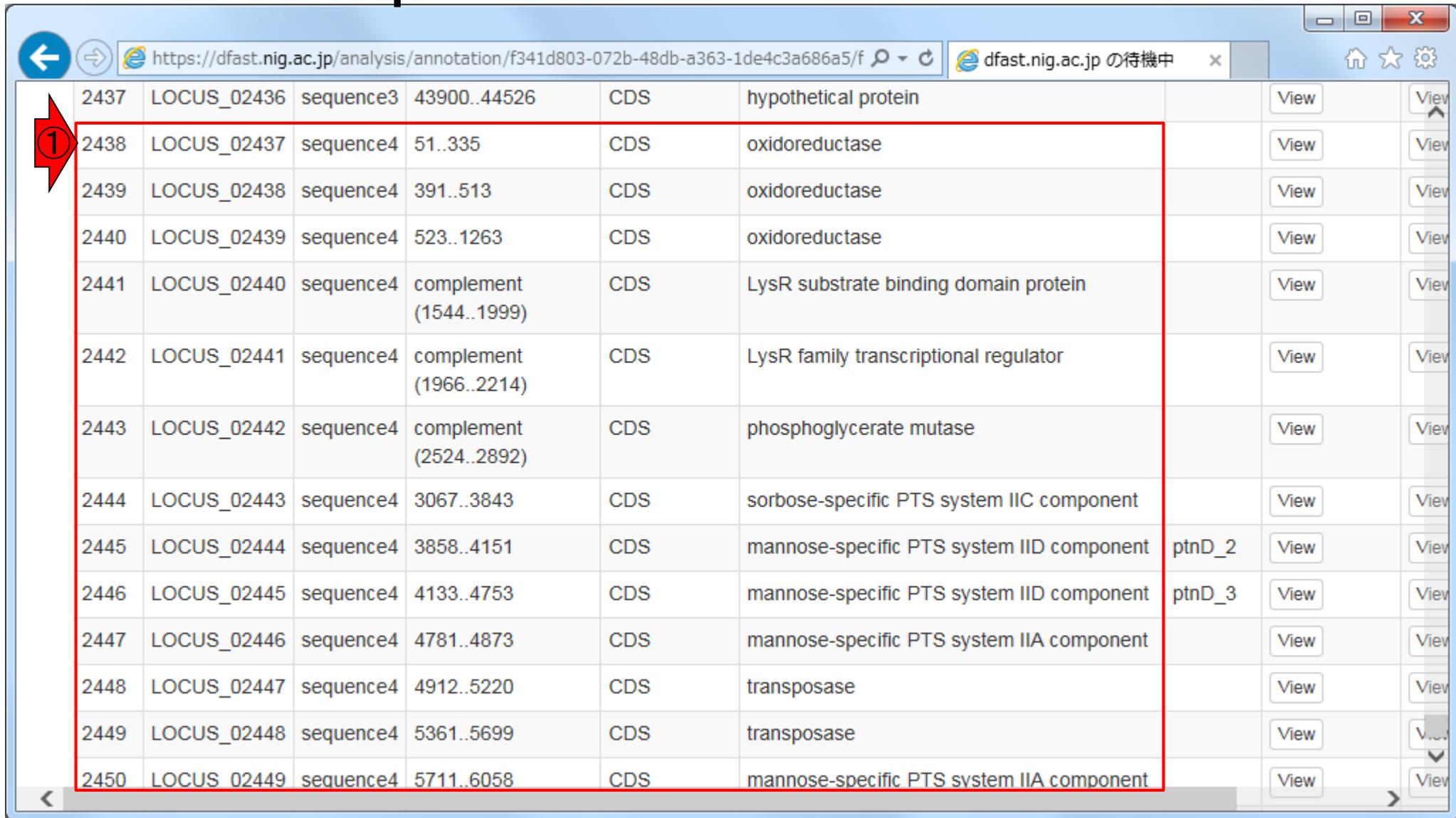
①赤下線で示す接合伝達(conjugal transfer)関連遺伝子が多く見られることから、sequence3がプラスミドであることを裏付けている

2396	LOCUS_02395	sequence3	11613..11924	CDS	hypothetical protein	View	View
2397	LOCUS_02396	sequence3	11963..12577	CDS	hypothetical protein	View	View
2398	LOCUS_02397	sequence3	12579..12914	CDS	<u>conjugal transfer protein</u>	View	View
2399	LOCUS_02398	sequence3	12935..13297	CDS	<u>conjugal transfer protein</u>	View	View
2400	LOCUS_02399	sequence3	13266..13925	CDS	<u>conjugal transfer protein</u>	View	View
2401	LOCUS_02400	sequence3	13937..15955	CDS	<u>conjugal transfer protein</u>	View	View
2402	LOCUS_02401	sequence3	15948..17366	CDS	<u>conjugal transfer protein</u>	View	View
2403	LOCUS_02402	sequence3	17367..18521	CDS	hypothetical protein	View	View
2404	LOCUS_02403	sequence3	18535..19152	CDS	hypothetical protein	View	View
2405	LOCUS_02404	sequence3	19106..19507	CDS	hypothetical protein	View	View
2406	LOCUS_02405	sequence3	19508..19978	CDS	<u>conjugal transfer protein</u>	View	View
2407	LOCUS_02406	sequence3	19980..21491	CDS	<u>conjugal transfer protein</u>	View	View
2408	LOCUS_02407	sequence3	21506..21895	CDS	hypothetical protein	View	View
2409	LOCUS_02408	sequence3	21914..22753	CDS	<u>conjugal transfer protein</u>	View	View
2410	LOCUS_02409	sequence3	22769..23179	CDS	hypothetical protein	View	View



①ここからが、sequence4
のアノテーション結果

W4-9: sequence4概観



LOCUS	LOCUS	sequence	coordinates	feature	description	accession	actions
2437	LOCUS_02436	sequence3	43900..44526	CDS	hypothetical protein		View
2438	LOCUS_02437	sequence4	51..335	CDS	oxidoreductase		View
2439	LOCUS_02438	sequence4	391..513	CDS	oxidoreductase		View
2440	LOCUS_02439	sequence4	523..1263	CDS	oxidoreductase		View
2441	LOCUS_02440	sequence4	complement (1544..1999)	CDS	LysR substrate binding domain protein		View
2442	LOCUS_02441	sequence4	complement (1966..2214)	CDS	LysR family transcriptional regulator		View
2443	LOCUS_02442	sequence4	complement (2524..2892)	CDS	phosphoglycerate mutase		View
2444	LOCUS_02443	sequence4	3067..3843	CDS	sorbose-specific PTS system IIC component		View
2445	LOCUS_02444	sequence4	3858..4151	CDS	mannose-specific PTS system IID component	ptnD_2	View
2446	LOCUS_02445	sequence4	4133..4753	CDS	mannose-specific PTS system IID component	ptnD_3	View
2447	LOCUS_02446	sequence4	4781..4873	CDS	mannose-specific PTS system IIA component		View
2448	LOCUS_02447	sequence4	4912..5220	CDS	transposase		View
2449	LOCUS_02448	sequence4	5361..5699	CDS	transposase		View
2450	LOCUS_02449	sequence4	5711..6058	CDS	mannose-specific PTS system IIA component		View

W4-9: sequence4概観

LOCUS	sequence4	coordinates	CDS	protein names	view
2446	LOCUS_02446	sequence4 4755..4755	CDS	mannose-specific PTS system IID component	view
2447	LOCUS_02446	sequence4 4781..4873	CDS	mannose-specific PTS system IIA component	View
2448	LOCUS_02447	sequence4 4912..5220	CDS	transposase	View
2449	LOCUS_02448	sequence4 5361..5699	CDS	transposase	View
2450	LOCUS_02449	sequence4 5711..6058	CDS	mannose-specific PTS system IIA component	View
2451	LOCUS_02450	sequence4 6114..6596	CDS	mannose/fructose/sorbose-specific PTS system IID component	View
2452	LOCUS_02451	sequence4 6738..7010	CDS	hypothetical protein	View
2453	LOCUS_02452	sequence4 7028..7204	CDS	hypothetical protein	View
2454	LOCUS_02453	sequence4 7515..7844	CDS	excinuclease ABC subunit A	uvrA_3 View
2455	LOCUS_02454	sequence4 8068..9012	CDS	excinuclease ABC subunit A	uvrA_4 View
2456	LOCUS_02455	sequence4 9012..9446	CDS	excinuclease ABC subunit A	uvrA_5 View
2457	LOCUS_02456	sequence4 9552..10025	CDS	excinuclease ABC subunit A	uvrA_6 View
2458	LOCUS_02457	sequence4 10378..10611	CDS	penicillin-binding protein 2A	View
2459	LOCUS_02458	sequence4 10608..11261	CDS	penicillin-binding protein 2A	View

Showing 1 to 2,459 of 2,459 entries

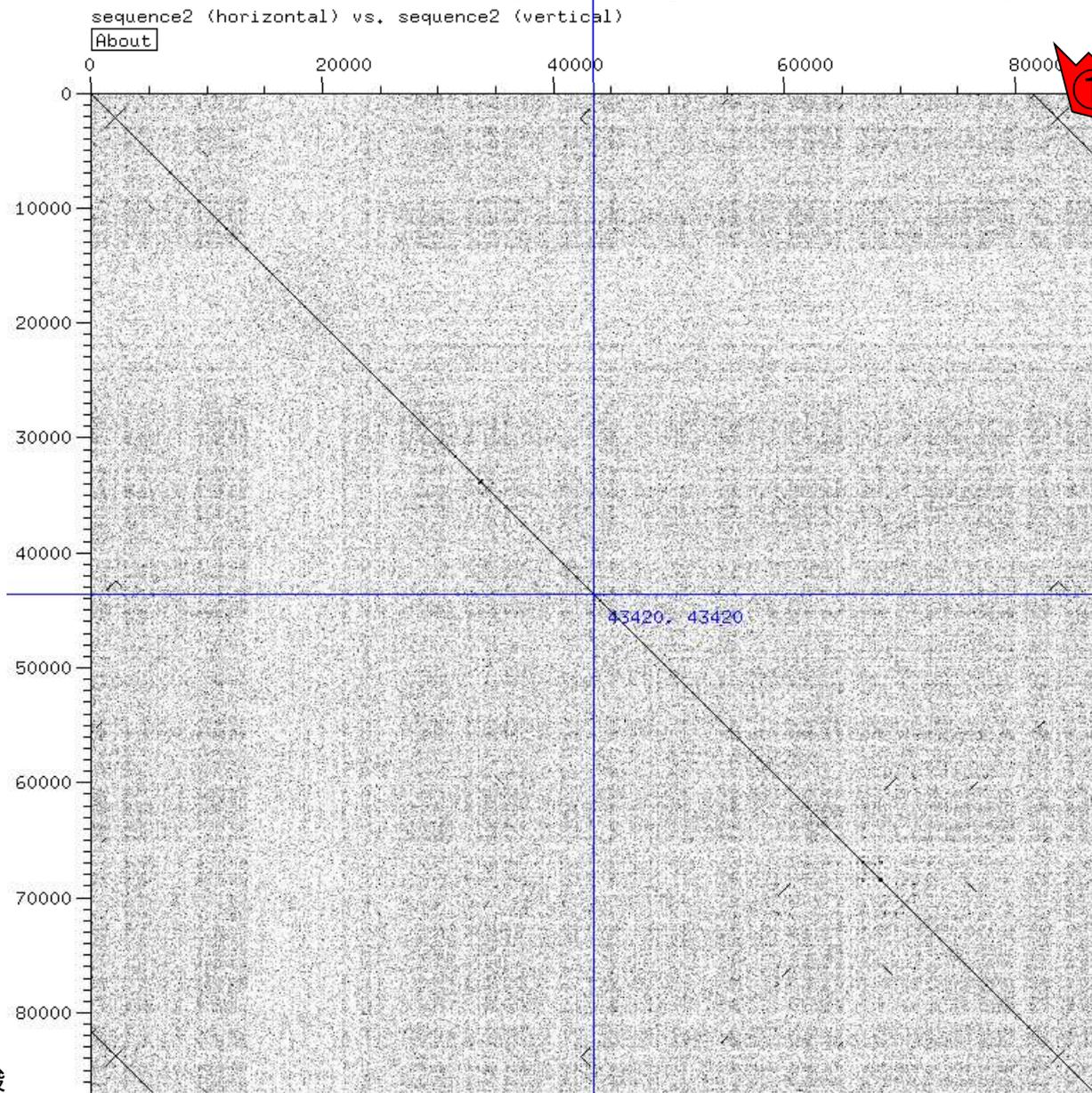
Previous 1 Next

W5-1 : dotter (sequence2)

```
iu@bielinux[result] pwd [12:03午後]
/home/iu/Desktop/result
iu@bielinux[result] ls [12:03午後]
corrected.fastq LH_hgap.fa sequence3.fa test3.fa
JSLAB8_1.sh polished_assembly.fasta sequence4.fa test4.fa
JSLAB8_2.sh polished_assembly.fastq smrtpipe.log
JSLAB8_3.sh sequence1.fa test1.fa
JSLAB8_4.sh sequence2.fa test2.fa
iu@bielinux[result] dotter sequence2.fa sequence2.fa [12:03午後]
```

W5-1 : dotter (sequence2)

実行結果。①このあたりで5,000 bpほど重複していることがわかる。両末端の5,000 bpあたりはアノテーション結果もほぼ同じなのだろうと妄想する



W5-2: dotter (sequence4)

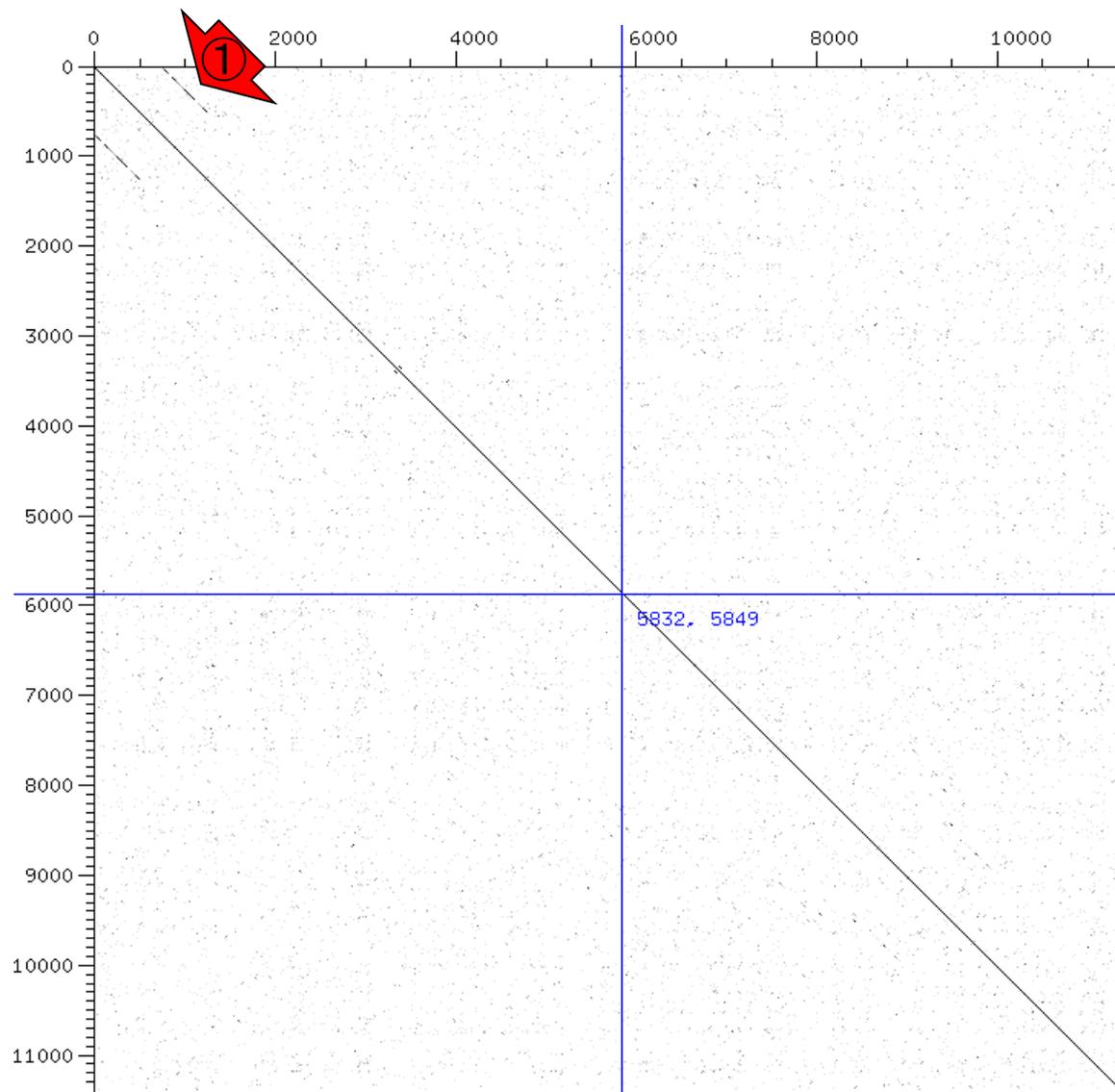
```
iu@bielinux[result] pwd [12:03午後]
/home/iu/Desktop/result
iu@bielinux[result] ls [12:03午後]
corrected.fastq LH_hgap.fa sequence3.fa test3.fa
JSLAB8_1.sh polished_assembly.fasta sequence4.fa test4.fa
JSLAB8_2.sh polished_assembly.fastq smrtpipe.log
JSLAB8_3.sh sequence1.fa test1.fa
JSLAB8_4.sh sequence2.fa test2.fa
iu@bielinux[result] dotter sequence4.fa sequence4.fa [12:03午後]
```



①

W5-2: dotter (sequence4)

①この結果は、(1 - 500)番目と(750 - 1,350)番目付近の領域が似ていることを意味する。しかし両末端ではないので、sequence4は環状(プラスミド)ではないのだろうと判断



W5-3: dotter (sequence1)

```
iu@bielinux[~/Desktop/result]                               [12:03午後]
iu@bielinux[result] pwd                                     [12:03午後]
/home/iu/Desktop/result
iu@bielinux[result] ls                                       [12:03午後]
corrected.fastq  LH_hgap.fa          sequence3.fa  test3.fa
JSLAB8_1.sh      polished_assembly.fasta sequence4.fa  test4.fa
JSLAB8_2.sh      polished_assembly.fastq smrtpipe.log
JSLAB8_3.sh      sequence1.fa        test1.fa
JSLAB8_4.sh      sequence2.fa        test2.fa
iu@bielinux[result] dotter sequence1.fa sequence1.fa [12:03午後]
```



この状態からいつまでも変わらないので、「①CTRL + ②C」で脱出

W5-3: dotter (sequence1)

```
File Edit View Search Terminal Help
/home/iu/Desktop/result
iu@bielinux[result] ls
corrected.fastq  LH_hgap.fa          sequence3.fa  test3.fa
JSLAB8_1.sh     polished_assembly.fasta sequence4.fa  test4.fa
JSLAB8_2.sh     polished_assembly.fastq smrtpipe.log
JSLAB8_3.sh     sequence1.fa        test1.fa
JSLAB8_4.sh     sequence2.fa        test2.fa
iu@bielinux[result] dotter sequence1.fa sequence1.fa
```

[4:00午後]

[4:00午後]

```
Detected sequence types: DNA vs.
Karlin/Altschul statistics for th
K      = 0.161
Lambda = 0.176
=> Expected MSP score in a 100
Expected residue score in MSP
=> Expected MSP length = 24
2289497 vs. 2289497 residues => 5
minutes on an SGI MIPS R10000)
```



W5-3: dotter (sequence1)

①「CTRL + C」を押しても変化がない場合は、リターンキーを押したりする。それでもコマンド入力待ち状態に戻らない場合は、「CTRL + Z」

```
iu@bielinux[result] ls
corrected.fastq  LH_hgap.fa          sequence3.fa  test3.fa
JSLAB8_1.sh     polished_assembly.fasta  sequence4.fa  test4.fa
JSLAB8_2.sh     polished_assembly.fastq  smrtpipe.log
JSLAB8_3.sh     sequence1.fa         test1.fa
JSLAB8_4.sh     sequence2.fa         test2.fa
iu@bielinux[result] dotter sequence1.fa sequence1.fa [ 4:00午後 ]

Detected sequence types: DNA vs. DNA
Karlin/Altschul statistics for these sequences and score matrix:
K      = 0.161
Lambda = 0.176
=> Expected MSP score in a 100x100 matrix = 42.067
Expected residue score in MSP = 1.719
=> Expected MSP length = 24
2289497 vs. 2289497 residues => 5241796.50 million dots. (Takes 5079:15
minutes on an SGI MIPS R10000)
^C
```



①「CTRL + Z」だと、コマンド入力待ち状態に戻ることができます

W5-3: dotter (sequence1)

```
File Edit View Search Terminal Help 16:30
JSLAB8_1.sh polished_assembly.fasta sequence4.fa test4.fa
JSLAB8_2.sh polished_assembly.fastq smrtpipe.log
JSLAB8_3.sh sequence1.fa test1.fa
JSLAB8_4.sh sequence2.fa test2.fa
iu@bielinux[result] dotter sequence1.fa sequence1.fa [ 4:00午後 ]

Detected sequence types: DNA vs. DNA
Karlin/Altschul statistics for these sequences and score matrix:
K = 0.161
Lambda = 0.176
=> Expected MSP score in a 100x100 matrix = 42.067
Expected residue score in MSP = 1.719
=> Expected MSP length = 24
2289497 vs. 2289497 residues => 5241796.50 million dots. (Takes 5079:15
minutes on an SGI MIPS R10000)
^C
^Z
zsh: suspended dotter sequence1.fa sequence1.fa
iu@bielinux[result] [ 4:29午後 ]
```



W5-4: dotterのオプション

ちなみに①実行時間は5079分と書かれています。84.65時間なので、約3.5日ですねw。数分で終了しないわけだ。このように大まかな実行時間が書かれていることもありますので、なかなか計算がおわらないときは、①のような時間に関する情報がどこかに書かれていないかチェックしてもいいでしょう

```
File Edit View Search Terminal Help
JSLAB8_1.sh polished_assembly.fasta sequence4
JSLAB8_2.sh polished_assembly.fastq smrtpipe.
JSLAB8_3.sh sequence1.fa test1.fa
JSLAB8_4.sh sequence2.fa test2.fa
iu@bielinux[result] dotter sequence1.fa sequence1.f

Detected sequence types: DNA vs. DNA
Karlin/Altschul statistics for these sequences and score matrix:
K = 0.161
Lambda = 0.176
=> Expected MSP score in a 100x100 matrix = 42.067
Expected residue score in MSP = 1.719
=> Expected MSP length = 24
2289497 vs. 2289497 residues => 5241796.50 million dots. (Takes 5079:15
minutes on an SGI MIPS R10000)
^C
^Z
zsh: suspended dotter sequence1.fa sequence1.fa
iu@bielinux[result]
```



[4:29午後]

W5-4: dotterのオプション

①でdotterのオプションを眺めると、②デフォルトの使用メモリはたった0.5Mb!これを1000とかにすれば10時間程度で計算が終了するかもしれません。

```
File Edit View Search Terminal Help
iu@bielinux[result] dotter [ 1:49午後 ]

Dotter - Sequence dotplots with image enhancement tools.

Reference: Sonnhammer ELL & Durbin R (1995). A dot-matrix program
with dynamic threshold control suited for genomic DNA and protein
sequence analysis. Gene 167(2):GC1-10.

Usage: dotter [options] <horizontal_sequence> <vertical_sequence> [X
options]

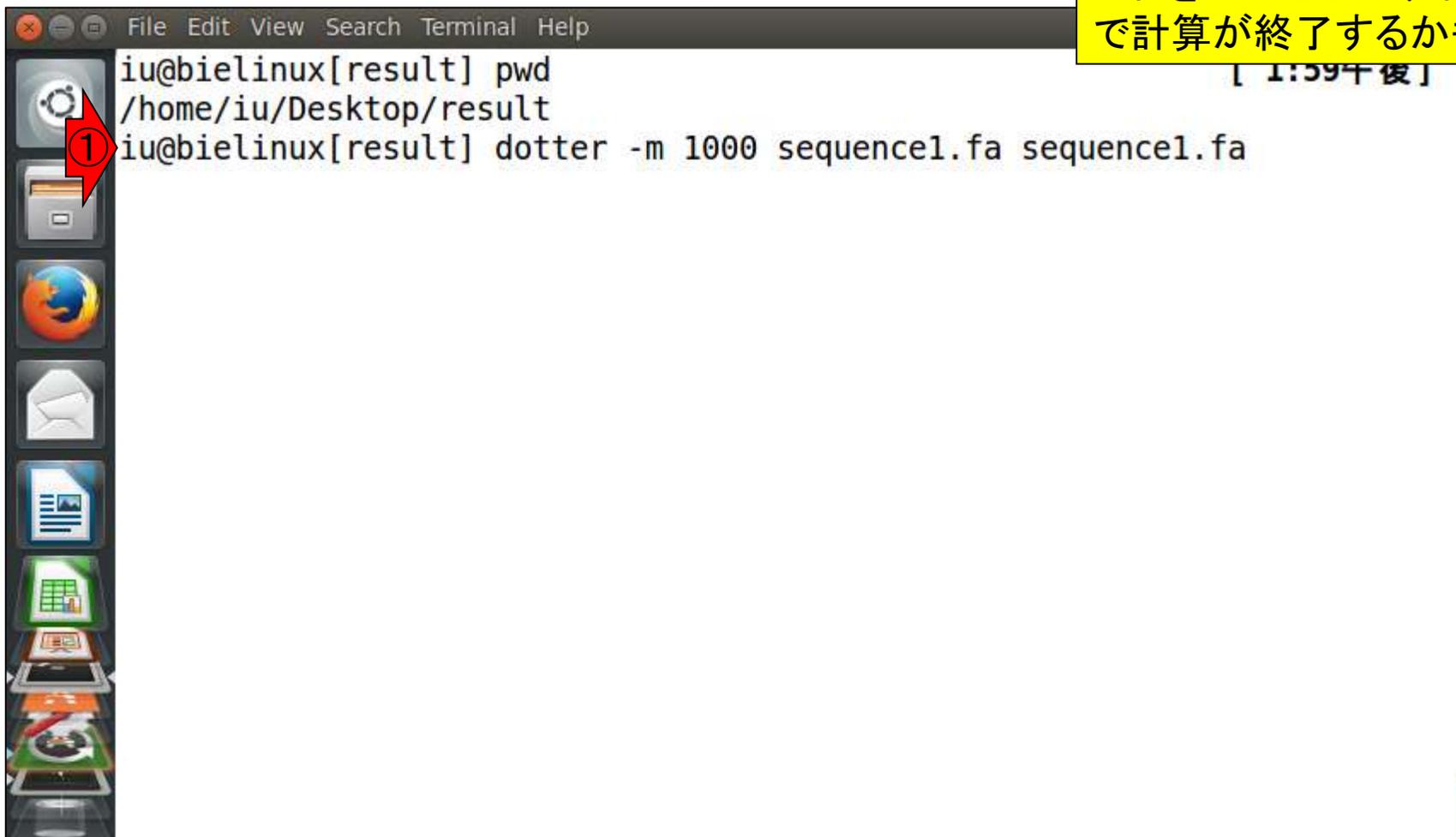
Allowed types:
                Protein      -      Protein
                DNA          -      DNA
                DNA          -      Protein

Options:
-b <file>      Batch mode, write dotplot to <file>
-l <file>      Load dotplot from <file>
-m <float>     Memory usage limit in Mb (default 0.5)
```



W5-4: dotterのオプション

①でdotterのオプションを眺めると、②デフォルトの使用メモリはたった0.5Mb!これを1000とかにすれば10時間程度で計算が終了するかもしれません。



A terminal window with a dark background and a light-colored text. The window title bar shows 'File Edit View Search Terminal Help'. The terminal content is as follows:

```
iu@bielinux[result] pwd
/home/iu/Desktop/result
iu@bielinux[result] dotter -m 1000 sequence1.fa sequence1.fa
```

A red arrow with the number '1' points to the second line of the terminal output. The terminal window is part of a desktop environment with a vertical dock on the left containing icons for a gear, a folder, Firefox, an envelope, a document, a spreadsheet, and a presentation.

[1:59午後]

W6-1 : makeblastdb

```
iu@bielinux[result] pwd [ 9:40午前 ]
/home/iu/Desktop/result
iu@bielinux[result] ls LH_hgap.fa* [ 9:40午前 ]
LH_hgap.fa
① iu@bielinux[result] makeblastdb -in LH_hgap.fa -dbtype nucl -hash_index

Building a new DB, current time: 09/04/2016 09:40:32
New DB name: LH_hgap.fa
New DB title: LH_hgap.fa
Sequence type: Nucleotide
Keep Linkouts: T
Keep MBits: T
Maximum file size: 1000000000B
Adding sequences from FASTA; added 4 sequences in 0.154485 seconds.
iu@bielinux[result] ls LH_hgap.fa* [ 9:40午前 ]
LH_hgap.fa LH_hgap.fa.nhr LH_hgap.fa.nsd
LH_hgap.fa.nhd LH_hgap.fa.nin LH_hgap.fa.nsi
LH_hgap.fa.nhi LH_hgap.fa.nog LH_hgap.fa.nsq
iu@bielinux[result] [ 9:40午前 ]
```

W6-2: blastn

①DB側をLH_hgap.fa、query側をsequence1.fa、出力ファイルをsequence1_blast.txtとしてblastnを実行。
②出力ファイルは12MB。sequence1.fa同士のアラインメント結果を丸々含んでいるためであろう

```
iu@bielinux[result] pwd [10:18午前]
/home/iu/Desktop/result
iu@bielinux[result] ls -l sequence1* [10:18午前]
-rw-rw-r-- 1 iu iu 2289509  8月 29 14:06 sequence1.fa
① iu@bielinux[result] blastn -db LH_hgap.fa -query sequence1.fa \
> -out sequence1_blast.txt
iu@bielinux[result] ls -l sequence1* [10:19午前]
-rw-rw-r-- 1 iu iu 12473981  9月  4 10:19 sequence1_blast.txt
-rw-rw-r-- 1 iu iu  2289509  8月 29 14:06 sequence1.fa
② iu@bielinux[result] ls -lh sequence1* [10:19午前]
-rw-rw-r-- 1 iu iu  12M  9月  4 10:19 sequence1_blast.txt
-rw-rw-r-- 1 iu iu 2.2M  8月 29 14:06 sequence1.fa
iu@bielinux[result] [10:19午前]
```

W7-1 : BLASTGrabber

①BLASTGrabberのウェブページ。②BLASTGrabber 2.0からzip圧縮ファイルをダウンロード。右クリックでショートカットのコピー、でURL情報を取得してwgetするやり方が次のスライド

Resources

BLASTGrabber

The BLASTGrabber program can visualize and analyse the output of the BLAST sequence search algorithm. It can handle tens of thousands of queries even on a laptop computer. NCBI taxonomy can be inferred and visualized for BLAST hits.

The program is a Java application with a graphical user interface.

- To install it, merely download and unzip the distribution file.
- To use it, download the user guide or the tutorial video.

BLASTGrabber supports the addition of 3rd-party plug-in components. Download the plug-in description and source code examples for details.

[BLASTGrabber 2.0](#)

[New in BLASTGrabber 2.0](#)

[BLASTGrabber 1.7](#)

Neumann et al., *BMC Bioinformatics*, **15**: 128, 2014

①~/Downloadsに移動し
、②wget、③unzipで解凍

W7-1 : BLASTGrabber

```
File Edit View Search Terminal Help [ 3:42午後 ]
iu@bielinux[result] cd ~/Downloads [ 3:42午後 ]
iu@bielinux[Downloads] pwd [ 3:42午後 ]
/home/iu/Downloads
iu@bielinux[Downloads] wget -cq http://folk.uio.no/ralfne/BLASTGrabber_
2_0.zip
iu@bielinux[Downloads] unzip BLASTGrabber_2_0.zip [ 3:48午後 ]
Archive:  BLASTGrabber_2_0.zip
  inflating: BLASTGrabber/BLASTGrabber
  inflating: BLASTGrabber/BLASTGrabber.bat
  inflating: BLASTGrabber/BLASTGrabber.jar
  inflating: BLASTGrabber/BLASTGrabberUserGuide.pdf
   creating: BLASTGrabber/lib/
  inflating: BLASTGrabber/lib/appframework-1.0.3.jar
  inflating: BLASTGrabber/lib/biojava-1.7.1.jar
  inflating: BLASTGrabber/lib/jcommon-1.0.16.jar
  inflating: BLASTGrabber/lib/jfreechart-1.0.13.jar
  inflating: BLASTGrabber/lib/jlfr-1_0.jar
  inflating: BLASTGrabber/lib/PDFRenderer.jar
  inflating: BLASTGrabber/lib/swing-worker-1.1.jar
   creating: BLASTGrabber/plugins/
  inflating: BLASTGrabber/plugins/plugins.sys
```

W7-1 : BLASTGrabber

①解凍後に作成されたBLASTGrabberディレクトリに移動し、③ls

```
File Edit View Search Terminal Help 15:53
inflating: BLASTGrabber/lib/jfreechart-1.0.13.jar
inflating: BLASTGrabber/lib/jlfr-1_0.jar
inflating: BLASTGrabber/lib/PDFRenderer.jar
inflating: BLASTGrabber/lib/swing-worker-1.1.jar
  creating: BLASTGrabber/plugins/
inflating: BLASTGrabber/plugins/plugins.sys
inflating: BLASTGrabber/README.TXT
  creating: BLASTGrabber/samples/
inflating: BLASTGrabber/samples/MC-Domain_blastout.bgr
inflating: BLASTGrabber/samples/MC-Domain_blastout.txt
inflating: BLASTGrabber/samples/MC-domain_queries.fasta
inflating: BLASTGrabber/taxnodes.bin
iu@bielinux[Downloads] [ 3:48午後 ]
① iu@bielinux[Downloads] cd BLASTGrabber [ 3:51午後 ]
iu@bielinux[BLASTGrabber] pwd [ 3:52午後 ]
/home/iu/Downloads/BLASTGrabber
② iu@bielinux[BLASTGrabber] ls [ 3:52午後 ]
BLASTGrabber      BLASTGrabberUserGuide.pdf  README.TXT
BLASTGrabber.bat  lib                          samples
BLASTGrabber.jar  plugins                       taxnodes.bin
iu@bielinux[BLASTGrabber] [ 3:53午後 ]
```

W7-1 : BLASTGrabber

①BLASTGrabberの起動はこんな感じ。いろいろ試してみたが、直感的に使いづらく断念

```
File Edit View Search Terminal Help 11:23
inflating: BLASTGrabber/lib/swing-worker-1.1.jar
creating: BLASTGrabber/plugins/
inflating: BLASTGrabber/plugins/plugins.sys
inflating: BLASTGrabber/README.TXT
creating: BLASTGrabber/samples/
inflating: BLASTGrabber/samples/MC-Domain_blastout.bgr
inflating: BLASTGrabber/samples/MC-Domain_blastout.txt
inflating: BLASTGrabber/samples/MC-domain_queries.fasta
inflating: BLASTGrabber/taxnodes.bin
iu@bielinux[Downloads] [ 3:48午後]
iu@bielinux[Downloads] cd BLASTGrabber [ 3:51午後]
iu@bielinux[BLASTGrabber] pwd [ 3:52午後]
/home/iu/Downloads/BLASTGrabber
iu@bielinux[BLASTGrabber] ls [ 3:52午後]
BLASTGrabber      BLASTGrabberUserGuide.pdf  README.TXT
BLASTGrabber.bat  lib                          samples
BLASTGrabber.jar  plugins                       taxnodes.bin
iu@bielinux[BLASTGrabber] java -Xmx512m -jar ~/Downloads/BLASTGrabber/
LASTGrabber.jar
```



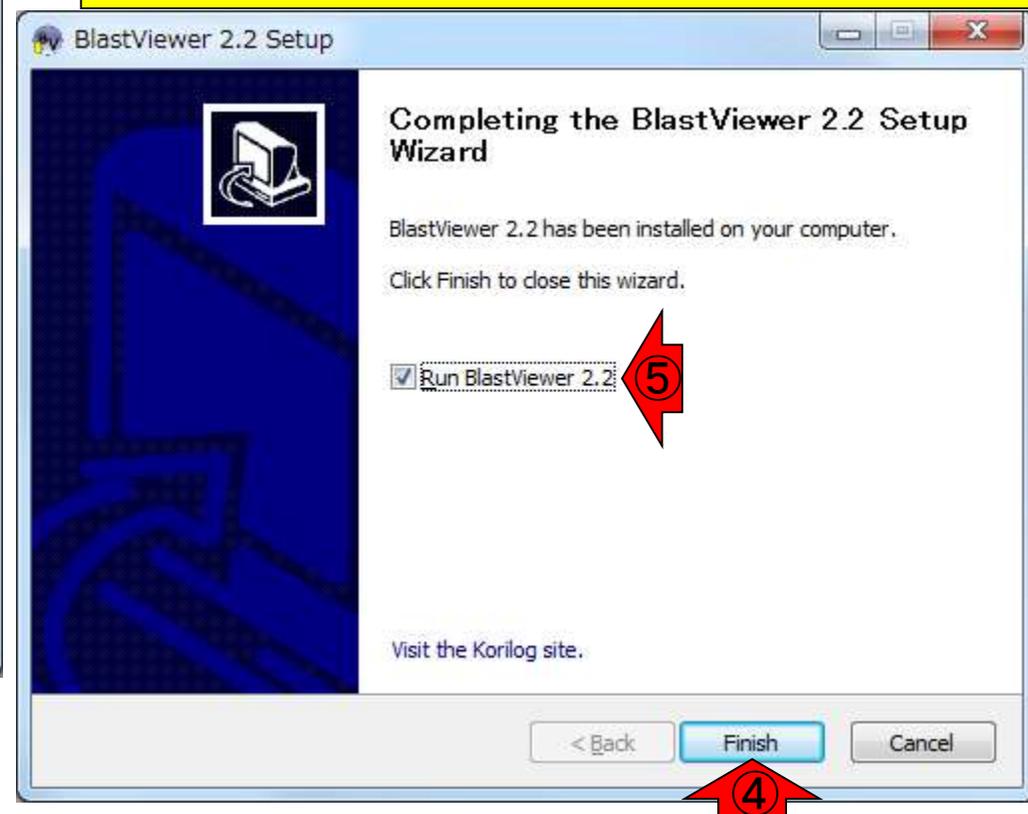
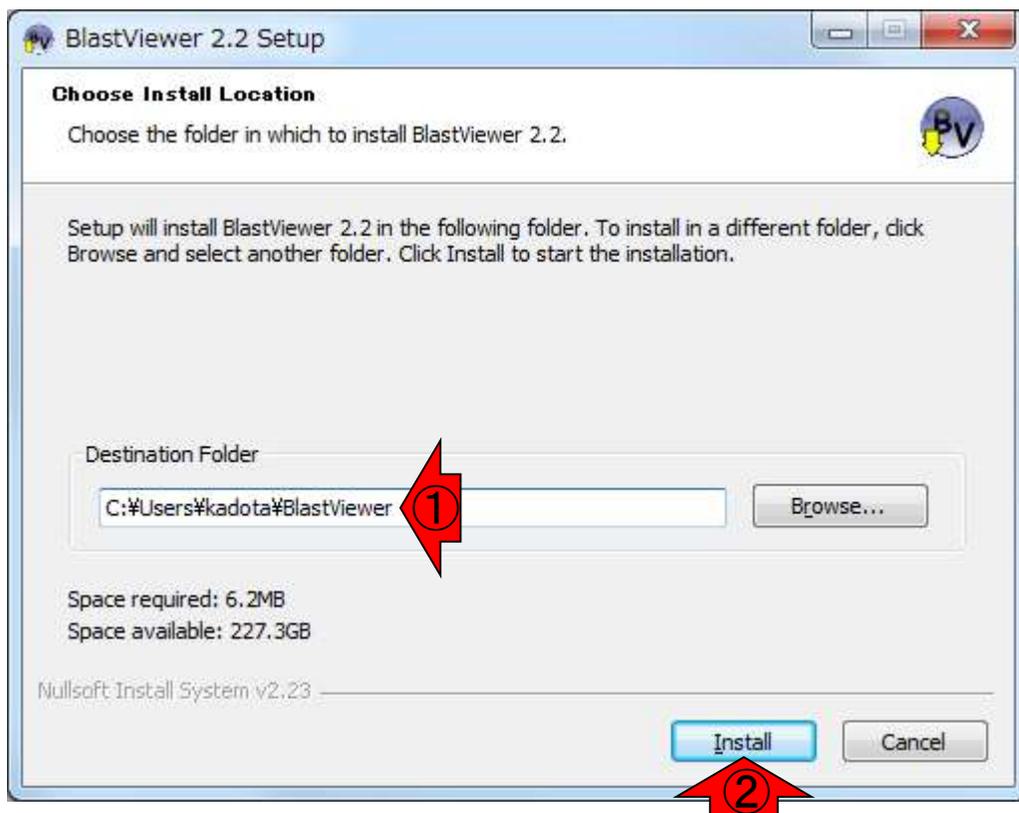
W7-2: BLASTViewer

①Korilogというところが提供しているBlastViewerのウェブページ。②Windows版と③Macintosh版。次のスライド以降はWindows版で説明

The screenshot shows the Korilog website interface. At the top, there is a navigation bar with links for CNET, REVIEWS, NEWS, DOWNLOAD, VIDEO, and HOW TO. Below this is a search bar and icons for Windows, Apple, and Android. A large green button labeled 'Start Download' is prominent, with a sub-button 'Start Download' and a list of steps: '1. Click to Start Download', '2. Run and Install', and '3. Scan for Issues'. The main content area features the 'Korilog' logo and a description of the company. On the left, there are filters for 'Narrow Results' including 'By Price', 'By Category', and 'By Operating System'. The main list shows two entries for 'BlastViewer'. The first entry is for the Windows version, with a 'Download' button highlighted by a red arrow labeled '2'. The second entry is for the Mac version, with a 'Download' button highlighted by a red arrow labeled '3'. A red arrow labeled '1' points to the 'Korilog' header.

W7-2: BLASTViewer

Windows 7上でのインストール画面。①ここに作成されるようだ。②Install。③デスクトップにBlastViewerのアイコンが作成されるはず。④Finish。デフォルトでは⑤にチェックが入っているのでBlastViewer (ver. 2.2)が起動する



W7-2: BLASTViewer

①を見ることで、BLAST結果ファイルを開く際には、②を押せばよいことがわかる。③をよく見るとBLAST出力結果ファイルはXML形式しか受け付けていないようだ

The screenshot shows the BlastViewer application window. The interface is divided into several sections:

- Left Panel:** Contains a "BLAST results" section with an "Open" icon (circled 2) and a text box (circled 3) explaining how to open BLAST files. A red arrow (circled 1) points to the text box.
- Summary Panel:** Shows a "Hits" table with columns for "#", "Accession", "Definition", "Quality", and "# HSPs".
- Alignment Panel:** Shows tabs for "HSP Map", "Definition", "Statistics", and "Alignment".
- Bottom Panel:** Includes a "Discover KoriBlast" link, a "Welcome to BlastViewer" message, a URL "www.korilog.com", and a date "14Mo/123Mo".

W7-2: BLASTViewer

①ためしにW6-2で作成したテキスト形式のBLAST結果ファイル(sequence1_blast.txt)を読み込もうとしたらダメでした。②OK

The screenshot shows the BlastViewer application window. The main area displays a table with columns for '#', 'Accession', 'Definition', 'Quality', and '# HSPs'. A modal dialog box is overlaid on the table, containing a warning icon and the text: "BlastViewer" followed by "BLAST file sequence1_blast.txt does not contain any readable data." Below the text is an "OK" button. Red arrows with circled numbers 1 and 2 point to the dialog box and the "OK" button, respectively. The left sidebar contains instructions for opening BLAST files. The bottom status bar shows "Welcome to BlastViewer", a URL "www.korilog.com", and a file size "14Mo/123Mo".

File Help

BLAST results

Summary

Hits

#	Accession	Definition	Quality	# HSPs
---	-----------	------------	---------	--------

BlastViewer

! BLAST file sequence1_blast.txt does not contain any readable data.

OK

Alignment

HSP Map Definition Statistics Alignment

HSP: < >

Discover KoriBlast to go beyond the viewer

Welcome to BlastViewer

www.korilog.com 14Mo/123Mo

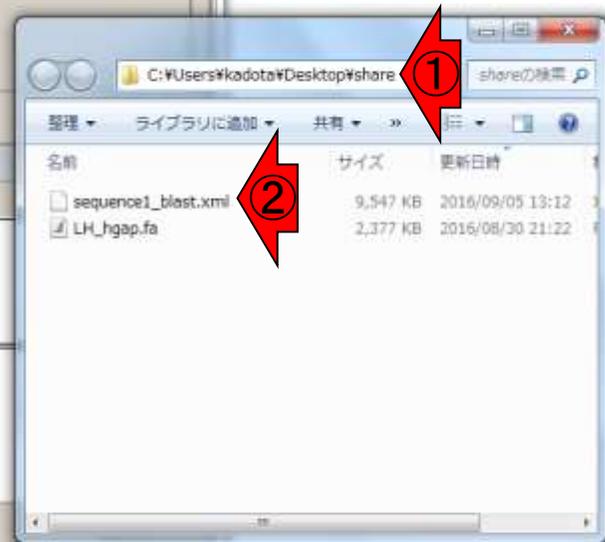
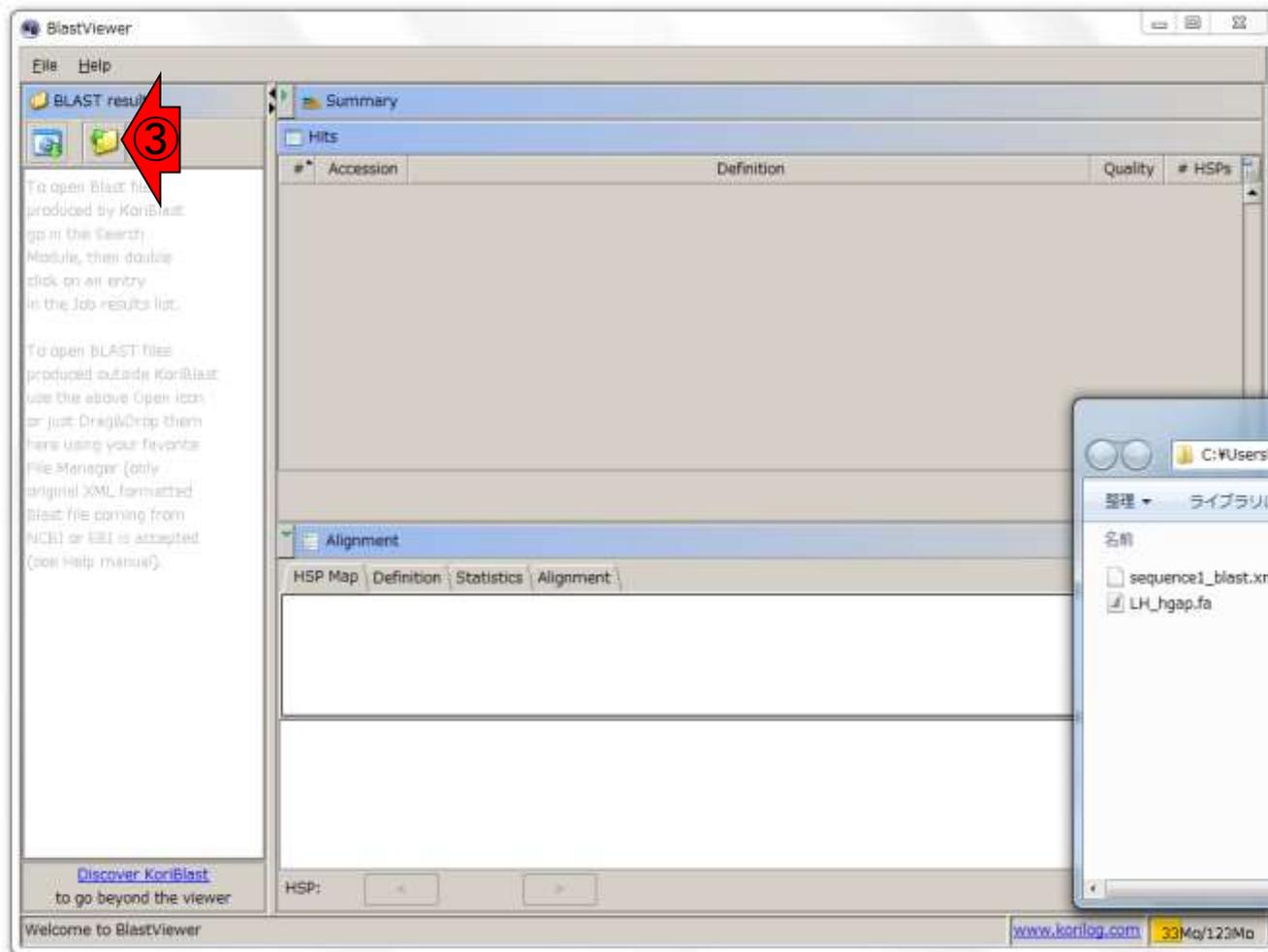
W7-3: BLAST再実行

①XML形式のBLAST結果ファイル(sequence1_blast.xml)を作成。②確かにあります。③共有フォルダにコピー

```
iu@bielinux[result] pwd [ 1:11午後 ]
/home/iu/Desktop/result
iu@bielinux[result] ls -l sequence1* [ 1:11午後 ]
-rw-rw-r-- 1 iu iu 132822 9月 4 16:22 sequence1_blast.bgr
-rw-rw-r-- 1 iu iu 12473981 9月 4 10:19 sequence1_blast.txt
-rw-rw-r-- 1 iu iu 2289509 8月 29 14:06 sequence1.fa
① iu@bielinux[result] blastn -db LH_hgap.fa -query sequence1.fa \
> -out sequence1_blast.xml -outfmt 5
iu@bielinux[result] ls -lh sequence1* [ 1:11午後 ]
-rw-rw-r-- 1 iu iu 130K 9月 4 16:22 sequence1_blast.bgr
-rw-rw-r-- 1 iu iu 12M 9月 4 10:19 sequence1_blast.txt
-rw-rw-r-- 1 iu iu 9.4M 9月 5 13:11 sequence1_blast.xml ②
-rw-rw-r-- 1 iu iu 2.2M 8月 29 14:06 sequence1.fa
③ iu@bielinux[result] cp sequence1_blast.xml ~/Desktop/mac_share
iu@bielinux[result] [ 1:12午後 ]
```

W7-4: BLASTViewer

①共有フォルダ上の、②XML形式のBLAST結果ファイル(sequence1_blast.xml)を、③開く



W7-4: BLASTViewer

①共有フォルダ上の、②XML形式のBLAST結果ファイル(sequence1_blast.xml)を、③開く

The screenshot shows the BLASTViewer application window. The main window has a menu bar (File, Help) and a toolbar. The left sidebar contains instructions for opening BLAST files. The main area is divided into sections: Summary, Hits (with a table header: #, Accession, Definition, Quality, # HSPs), and Alignment (with sub-tabs: HSP Map, Definition, Statistics, Alignment). An 'Open' dialog box is overlaid on the main window. The dialog box shows the file location as 'share' (indicated by a red arrow labeled '1') and the selected file as 'sequence1_blast.xml' (indicated by a red arrow labeled '2'). The file name field contains 'sequence1_blast.xml' and the file type is set to 'Blast file (NCBI's XML Format) (.xml)'. The '開く' (Open) button is highlighted with a red arrow labeled '3'. The '取消' (Cancel) button is also visible. The status bar at the bottom of the main window shows 'Welcome to BlastViewer', 'www.korilog.com', and '14Mo/123Mo'.

W7-4: BLASTViewer

読み込み後の状態。①query側が sequence1.fa、②DB側がsequence1-4からなるLH_hgap.faだったことを思い出そう

The screenshot shows the BlastViewer application window. On the left, a file explorer shows 'sequence1_blast.xml' and 'blastn vs. LH_hgap.fa'. A red arrow labeled '1' points to 'sequence1_blast.xml' and another red arrow labeled '2' points to 'blastn vs. LH_hgap.fa'. The main window displays the 'Summary' tab with a table of hits:

#	Accession	Definition	Quality	# HSPs
1	0	sequence1	😊	1347
2	3	sequence4	😊	2
3	2	sequence3	😊	1
4	1	sequence2	😊	1

Below the table, the 'Alignment' tab is active, showing 'Alignment: Query (2289497 nuc) vs. 0 (2,289,497 nuc)'. It includes a sequence alignment view with nucleotide letters (A, C, G, T) and a progress indicator at the bottom showing 'HSP: 1/1347'.

W7-5: 解釈

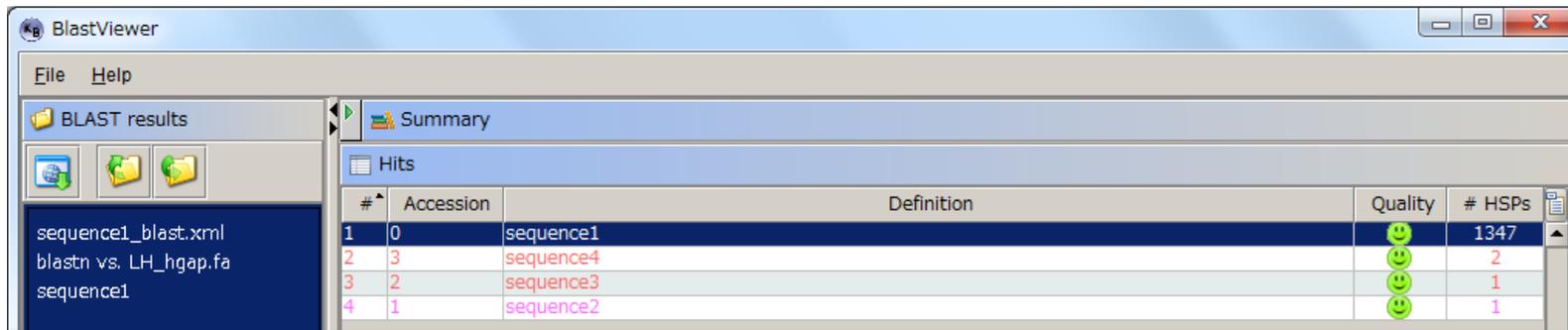
①は、DB側のヒット数(配列類似領域の数;HSPの数)が、sequence1中に1347個、sequence4中に2個、sequence3中に1個、sequence2中に1個あったことを示す。sequence1-4の並びではないので、HSP数がスコアの高い順でソートされているのだろうと妄想

The screenshot shows the BlastViewer interface. On the left, there's a sidebar with 'BLAST results' and a list of files: 'sequence1_blast.xml', 'blastn vs. LH_hgap.fa', and 'sequence1'. The main area is titled 'Summary' and contains a 'Hits' table. The table has five columns: '#', 'Accession', 'Definition', 'Quality', and '# HSPs'. The first row is highlighted in blue and has a red box around it, with a red arrow pointing to it from the right. This row is labeled with a circled '1'. Below the table, there's an 'Alignment' section showing 'Query (2289497 nuc) vs. 0 (2,289,497 nuc)' with tabs for 'HSP Map', 'Definition', 'Statistics', and 'Alignment'. The alignment view shows the query sequence and the HSP(s) with a sequence alignment below it. At the bottom, there's a footer with 'Welcome to BlastViewer', 'www.korilog.com', and '90Mo/123Mo'.

#	Accession	Definition	Quality	# HSPs
1	0	sequence1	😊	1347
2	3	sequence4	😊	2
3	2	sequence3	😊	1
4	1	sequence2	😊	1

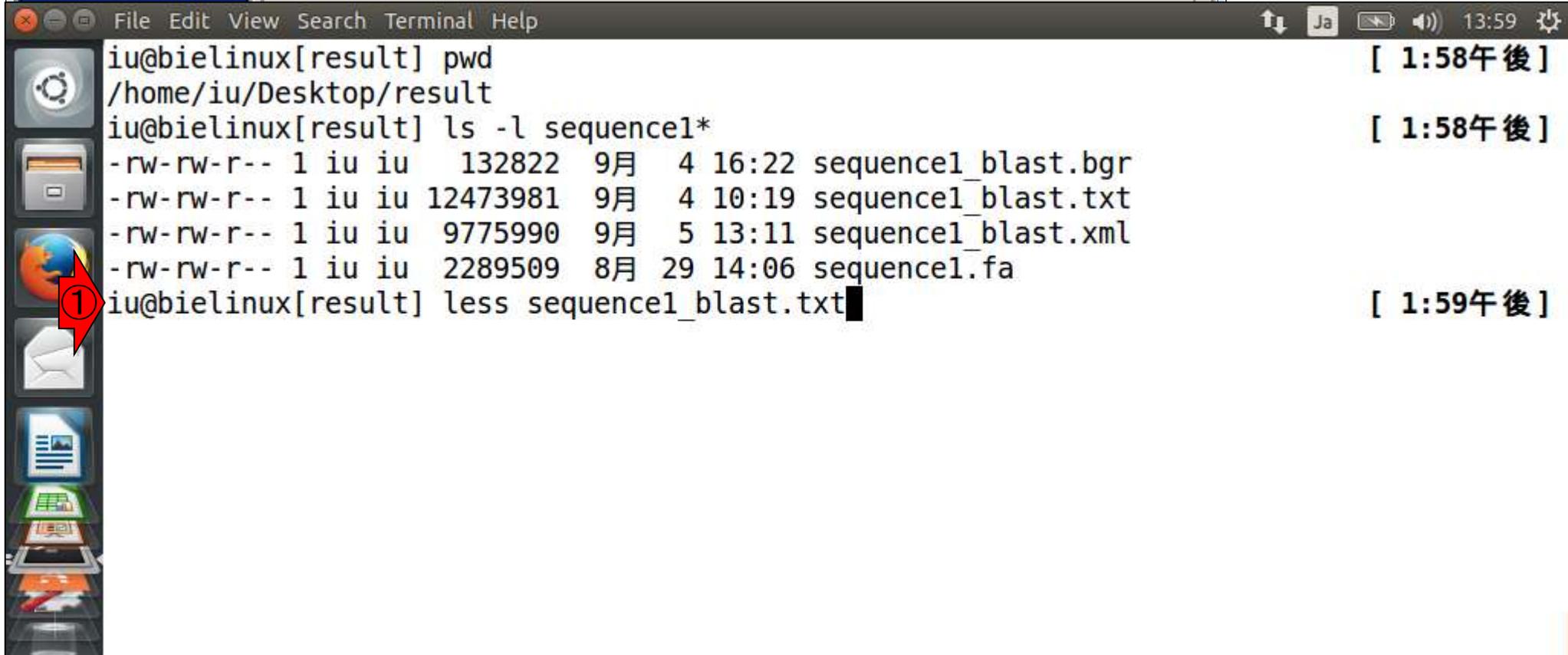
①テキスト形式のBLAST結果ファイル(sequence1_blast.txt)をlessで眺めて全体像を大まかに把握

W7-5: 解釈



The screenshot shows the BlastViewer application window. The 'Summary' tab is active, displaying a table of BLAST hits. The table has columns for '#', 'Accession', 'Definition', 'Quality', and '# HSPs'. The first hit (row 1) is 'sequence1' with an accession of '0', a quality score of 1347, and 1347 HSPs. The other three hits are 'sequence4', 'sequence3', and 'sequence2' with lower quality scores and fewer HSPs.

#	Accession	Definition	Quality	# HSPs
1	0	sequence1	1347	1347
2	3	sequence4	2	2
3	2	sequence3	1	1
4	1	sequence2	1	1



The screenshot shows a Linux terminal window with the following commands and output:

```
iu@bielinux[result] pwd [ 1:58午後 ]
/home/iu/Desktop/result
iu@bielinux[result] ls -l sequence1* [ 1:58午後 ]
-rw-rw-r-- 1 iu iu 132822 9月 4 16:22 sequence1_blast.bgr
-rw-rw-r-- 1 iu iu 12473981 9月 4 10:19 sequence1_blast.txt
-rw-rw-r-- 1 iu iu 9775990 9月 5 13:11 sequence1_blast.xml
-rw-rw-r-- 1 iu iu 2289509 8月 29 14:06 sequence1.fa
iu@bielinux[result] less sequence1_blast.txt [ 1:59午後 ]
```

A red arrow with the number '1' points to the 'less sequence1_blast.txt' command.

W7-5: 解釈

lessで開いた直後の状態。①用いたBLASTNのバージョン、②DB側、③query側の情報

BlastViewer interface showing BLAST results. The 'Summary' tab is active, displaying a table of hits. The table has columns for Hit number, Accession, Definition, Quality, and # HSPs.

#	Accession	Definition	Quality	# HSPs
1	0	sequence1	☺	1347
2	3	sequence4	☺	2
3	2	sequence3	☺	1
4	1	sequence2	☺	1

```
File Edit View Search Terminal Help
BLASTN 2.2.28+
Reference: Zheng Zhang, Scott Schwartz, Lukas Wagner, and Webb Miller (2000), "A greedy algorithm for aligning DNA sequences", J Comput Biol 2000; 7(1-2):203-14.
Database: LH_hgap.fa
         4 sequences; 2,433,614 total letters
Query= sequence1
█
```

Red arrows point to the following information in the terminal output:

- ① BLASTN 2.2.28+
- ② Database: LH_hgap.fa
- ③ Query= sequence1

W7-5: 解釈

1ページ分ほど下に移動。①確かにBLASTViewerと同じ並び(sequence1, 4, 3, 2)になっている

The screenshot shows the BlastViewer application window. The 'Summary' tab is active, displaying a table of hits. The table has columns for '#', 'Accession', 'Definition', 'Quality', and '# HSPs'. The hits are as follows:

#	Accession	Definition	Quality	# HSPs
1	0	sequence1	😊	1347
2	3	sequence4	😊	2
3	2	sequence3	😊	1
4	1	sequence2	😊	1

The screenshot shows a terminal window with the following content:

```
Sequences producing significant alignments:
```

	Score (Bits)	E Value
sequence1	4.228e+06	0.0
sequence4	1.032e+04	0.0
sequence3	2113	0.0
sequence2	124	1e-25

```
> sequence1  
Length=2289497  
  
Score = 4.228e+06 bits (2289497), Expect = 0.0  
Identities = 2289497/2289497 (100%), Gaps = 0/2289497 (0%)  
Strand=Plus/Plus  
:  
█
```

A red arrow with the number '1' points to the 'sequence1' entry in the table above.

W7-5: 解釈

①これがトップヒットの基本情報。query (sequence1)とDB (sequence1-4)の関係が分かれば、これがsequence1 vs. sequence1の100%一致の結果であることがわかる。スコアの計算方法がよくわかっていなくても、大まかに配列長 (2,289,497 bp)の2倍程度の値がスコアっぽいなどと学習する

The image shows a screenshot of the BlastViewer application. The top window displays the BLAST results summary, and the bottom window shows a terminal with detailed alignment information for the top hit.

#	Accession	Definition	Quality	# HSPs
1	0	sequence1	😊	1347
2	3	sequence4	😊	2
3	2	sequence3	😊	1
4	1	sequence2	😊	1

	Score (Bits)	E Value
sequence1	4.228e+06	0.0
sequence4	1.032e+04	0.0
sequence3	2113	0.0
sequence2	124	1e-25


```
> sequence1
Length=2289497

Score = 4.228e+06 bits (2289497), Expect = 0.0
Identities = 2289497/2289497 (100%), Gaps = 0/2289497 (0%)
Strand=Plus/Plus
```



①

W7-5: 解釈

となると、いくら①E-valueがそこそこ低くても、② sequence2にヒットしているsequence1の断片配列の領域は、スコアが124なので60塩基程度だろうと予想。また、③sequence3にヒットしているsequence1の断片配列の領域は、スコアが2113なので1000塩基程度だろうと予想

BlastViewer

File Help

BLAST results Summary

#	Accession	Definition	Quality	# HSPs
1	0	sequence1	😊	1347
2	3	sequence4	😊	2
3	2	sequence3	😊	1
4	1	sequence2	😊	1

Sequences producing significant alignments:

	Score (Bits)	E Value
sequence1	4.228e+06	0.0
sequence4	1.032e+04	0.0
sequence3	2113	0.0
sequence2	124	1e-25

```
> sequence1
Length=2289497

Score = 4.228e+06 bits (2289497), Expect = 0.0
Identities = 2289497/2289497 (100%), Gaps = 0/2289497 (0%)
Strand=Plus/Plus
```

W7-5: 解釈

今は、テキスト形式のBLAST結果ファイル(sequence1_blast.txt)をlessで眺めている。①sequence2でキーワード検索(nで順方向に検索、Nで逆方向に検索)し、アラインメントを表示させているところ。②ヒットしている領域としては76塩基

The image shows a computer screen with two windows. The top window is BlastViewer, displaying a table of BLAST hits. The bottom window is a terminal running the 'less' command to view the BLAST results file.

#	Accession	Definition	Quality	# HSPs
1	0	sequence1	😊	1347
2	3	sequence4	😊	2
3	2	sequence3	😊	1
4	1	sequence2	😊	1

```
> sequence2
Length=86892

Score = 124 bits (67), Expect = 1e-25
Identities = 73/76 (96%), Gaps = 0/76 (0%)
Strand=Plus/Minus

Query 1203803 TTGGATGTAAGCTGATTCCTGAGACAACTTTTAAGAGAGCGTATAATGAATAAATCGTCT 1203862
          |||
Sbjct 31521   TTGGATGTAAGCTGATTCCTGAGACAACTTTTAAGAGAGGGTATGATGAATAAATCATCT 31462

Query 1203863 CTCAAAAGGAAGGAAT 1203878
          |||
Sbjct 31461   CTCAAAAGGAAGGAAT 31446

:
```

W7-5: 解釈

今は、テキスト形式のBLAST結果ファイル(sequence1_blast.txt)をlessで眺めている。①sequence3でキーワード検索(nで順方向に検索、Nで逆方向に検索)し、アラインメントを表示させているところ。②ヒットしている領域としては1276塩基

The image shows a screenshot of a computer interface. At the top, there is a yellow box with Japanese text explaining the context: viewing a BLAST result file (sequence1_blast.txt) using less, searching for sequence3, and highlighting a hit region of 1276 bases. Below this, the BlastViewer application is open, showing a 'Summary' tab with a 'Hits' table. The table lists four hits: sequence1 (Quality: 1347), sequence4 (Quality: 2), sequence3 (Quality: 1), and sequence2 (Quality: 1). Below the BlastViewer, a terminal window is open, displaying the command 'sequence3' and its output. The output shows search parameters (Score = 2113 bits, Expect = 0.0, Identities = 1232/1276 (97%), Gaps = 0/1276 (0%), Strand=Plus/Minus) and three alignment blocks. Red arrows point to 'sequence3' in the terminal and the 'Identities' line.

#	Accession	Definition	Quality	# HSPs
1	0	sequence1	1347	1
2	3	sequence4	2	1
3	2	sequence3	1	1
4	1	sequence2	1	1

```
> sequence3
Length=45853

Score = 2113 bits (1144), Expect = 0.0
Identities = 1232/1276 (97%), Gaps = 0/1276 (0%)
Strand=Plus/Minus

Query 1203879 TTTTACATGCCAACTCGTTACGACAAAGAATTCAAACAAAACATCATCAACCTATATAAA 1203938
          |||
Sbjct 36875 TTTTACATGCCAACTCGTTACGACAAAGAATTCAAACAAAACATTATCAACCTATATAAG 36816

Query 1203939 CAAGGCGAATCAGCCGCCCAACTGGCCAGAGAATATGGCATTGGCTATTCAACCGTTCAT 1203998
          |||
Sbjct 36815 CAAGGCGAATCAGCTGCCCAACTGGCCAGAGAATATGGCATTGGCTATTCAACCGTTCAT 36756

Query 1203999 AAGTGGATCCAGGGCCAAGCCAAAACCTCAATCCGGTAAATCGCCAGACGAAATTAAGCG 1204058
:
```

W7-6 : grep

トータルのヒット数が(1347 + 2 + 1 + 1) = 1351個だったということは、①「Score = 」を含む行数も1351個あるのだろうと予想。grepで確認すべく、qを押してlessから一旦抜ける

The screenshot shows the BlastViewer application window. The top part displays the BLAST results summary with a table of hits:

#	Accession	Definition	Quality	# HSPs
1	0	sequence1	😊	1347
2	3	sequence4	😊	2
3	2	sequence3	😊	1
4	1	sequence2	😊	1

Below the summary, a terminal window is open with the following content:

```
> sequence3
Length=45853

Score = 2113 bits (1144), Expect = 0.0
Identities = 1232/1276 (97%), Gaps = 0/1276 (0%)
Strand=Plus/Minus

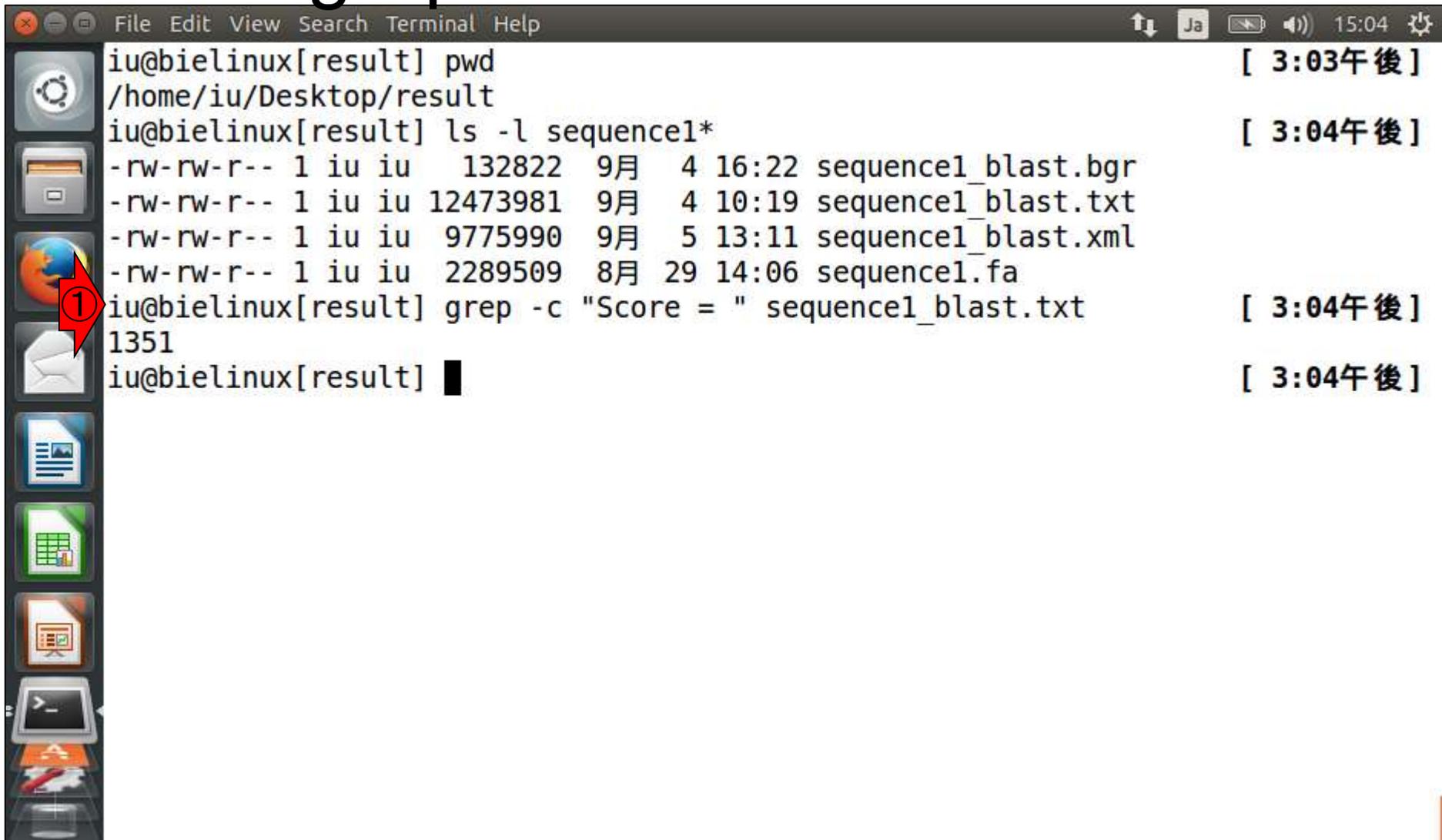
Query 1203879 TTTTACATGCCAACTCGTTACGACAAAGAATTCAAACAAAACATCATCAACCTATATAAA 1203938
          |||
Sbjct 36875 TTTTACATGCCAACTCGTTACGACAAAGAATTCAAACAAAACATTATCAACCTATATAAG 36816

Query 1203939 CAAGGCGAATCAGCCGCCCAACTGGCCAGAGAATATGGCATTGGCTATTCAACCGTTCAT 1203998
          |||
Sbjct 36815 CAAGGCGAATCAGCTGCCCAACTGGCCAGAGAATATGGCATTGGCTATTCAACCGTTCAT 36756

Query 1203999 AAGTGGATCCAGGGCCAAGCCAAAACCTCAATCCGGTAAATCGCCAGACGAAATTAAGCG 1204058
          :
```

A red arrow with the number 1 points to the "Score =" line in the terminal output.

W7-6: grep



```
iu@bielinux[result] pwd [ 3:03午後 ]
/home/iu/Desktop/result
iu@bielinux[result] ls -l sequence1* [ 3:04午後 ]
-rw-rw-r-- 1 iu iu 132822 9月 4 16:22 sequence1_blast.bgr
-rw-rw-r-- 1 iu iu 12473981 9月 4 10:19 sequence1_blast.txt
-rw-rw-r-- 1 iu iu 9775990 9月 5 13:11 sequence1_blast.xml
-rw-rw-r-- 1 iu iu 2289509 8月 29 14:06 sequence1.fa
① iu@bielinux[result] grep -c "Score = " sequence1_blast.txt [ 3:04午後 ]
1351
iu@bielinux[result] █ [ 3:04午後 ]
```

W7-6: grep

①「Score = 」を含む最初の10行分を表示。②のように同じスコアのものが2個ずつ表示されている状態を見て、「sequence1も両末端に重複領域がある環状コンティグなのだろう」と予想する

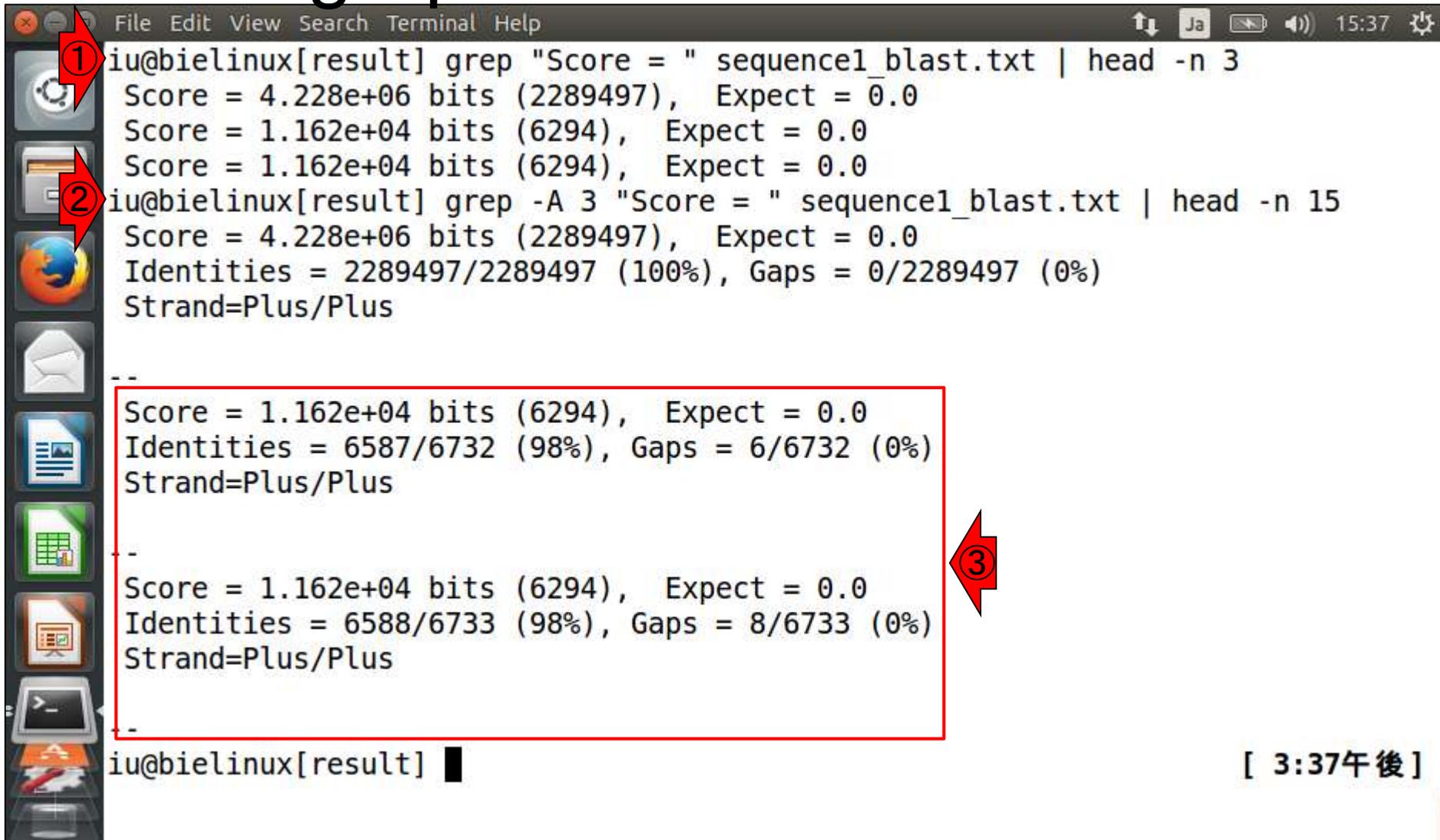
```
iu@bielinux[result] pwd
/home/iu/Desktop/result
iu@bielinux[result] ls -l sequence1*
-rw-rw-r-- 1 iu iu 132822 9月 4 16:22 sequence1_blast.bgr
-rw-rw-r-- 1 iu iu 12473981 9月 4 10:19 sequence1_blast.txt
-rw-rw-r-- 1 iu iu 9775990 9月 5 13:11 sequence1_blast.xml
-rw-rw-r-- 1 iu iu 2289509 8月 29 14:06 sequence1.fa
iu@bielinux[result] grep -c "Score = " sequence1_blast.txt
1351
iu@bielinux[result] grep "Score = " sequence1_blast.txt | head
Score = 4.228e+06 bits (2289497), Expect = 0.0
Score = 1.162e+04 bits (6294), Expect = 0.0 }
Score = 1.162e+04 bits (6294), Expect = 0.0 }
Score = 1.075e+04 bits (5823), Expect = 0.0 }
Score = 1.075e+04 bits (5823), Expect = 0.0 }
Score = 1.052e+04 bits (5694), Expect = 0.0 }
Score = 1.052e+04 bits (5694), Expect = 0.0 }
Score = 1.043e+04 bits (5647), Expect = 0.0 }
Score = 1.043e+04 bits (5647), Expect = 0.0 }
Score = 9679 bits (5241), Expect = 0.0
iu@bielinux[result]
```



[3:03午後]
[3:04午後]
[3:04午後]
[3:04午後]
[3:07午後]

①「Score = 」を含む最初の3行分を表示。②grep -Aオプションで一致した行を含め後ろの3行分を表示。③が今注目しているところ

W7-6: grep



```
iu@bielinux[result] grep "Score = " sequence1_blast.txt | head -n 3
Score = 4.228e+06 bits (2289497), Expect = 0.0
Score = 1.162e+04 bits (6294), Expect = 0.0
Score = 1.162e+04 bits (6294), Expect = 0.0
iu@bielinux[result] grep -A 3 "Score = " sequence1_blast.txt | head -n 15
Score = 4.228e+06 bits (2289497), Expect = 0.0
Identities = 2289497/2289497 (100%), Gaps = 0/2289497 (0%)
Strand=Plus/Plus
--
Score = 1.162e+04 bits (6294), Expect = 0.0
Identities = 6587/6732 (98%), Gaps = 6/6732 (0%)
Strand=Plus/Plus
-
Score = 1.162e+04 bits (6294), Expect = 0.0
Identities = 6588/6733 (98%), Gaps = 8/6733 (0%)
Strand=Plus/Plus
-
iu@bielinux[result] █
```

[3:37午後]

W7-6: grep -A

①Plus/Plusになっているので、おそらく両端に重複領域があるのだろう。sequence3同士のBLAST実行結果(第7回W15-6)も両端でPlus/Plusになっていたことを思い出せば納得できるだろう。本当に両端かどうかはBLASTViewerなり、このテキスト形式のBLAST結果ファイル(sequence1_blast.txt)を詳細に眺めて確認する

```
iu@bielinux[result] grep "Score = "  
Score = 4.228e+06 bits (2289497),  
Score = 1.162e+04 bits (6294), Expect = 0.0  
Score = 1.162e+04 bits (6294), Expect = 0.0  
iu@bielinux[result] grep -A 3 "Score = " sequence1_blast.txt | head -n 15  
Score = 4.228e+06 bits (2289497), Expect = 0.0  
Identities = 2289497/2289497 (100%), Gaps = 0/2289497 (0%)  
Strand=Plus/Plus  
--  
Score = 1.162e+04 bits (6294), Expect = 0.0  
Identities = 6587/6732 (98%), Gaps = 6/6732 (0%)  
Strand=Plus/Plus  
--  
Score = 1.162e+04 bits (6294), Expect = 0.0  
Identities = 6588/6733 (98%), Gaps = 8/6733 (0%)  
Strand=Plus/Plus  
--  
iu@bielinux[result] █
```

[3:37午後]

W7-6: grep -n

①grep -nで検索文字列(この場合"> sequence")を含む行番号を表示。例えば②sequence4とヒットしたものは、入力ファイル(sequence1_blast.txt)の205785行目からスタートしていることがわかる。第7回W15-5

```
iu@bielinux[result] pwd [11:45午前]
/home/iu/Desktop/result
iu@bielinux[result] ls -l sequence1* [11:45午前]
-rw-rw-r-- 1 iu iu 132822 9月 4 16:22 sequence1_blast.bgr
-rw-rw-r-- 1 iu iu 12473981 9月 4 10:19 sequence1_blast.txt
-rw-rw-r-- 1 iu iu 9775990 9月 5 13:11 sequence1_blast.xml
-rw-rw-r-- 1 iu iu 2289509 8月 29 14:06 sequence1.fa
① iu@bielinux[result] grep -n "> sequence" sequence1_blast.txt
27:> sequence1
205785:> sequence4 ②
206506:> sequence3
206602:> sequence2
iu@bielinux[result] █ [11:45午前]
```

W7-6: grep -n

①grep -nで検索文字列(この場合"Score =")を含む行番号も表示。②赤枠の上下関係(sequence1の最後の205776とsequence3の206509)から、sequence4とヒットした領域は2か所あり、いずれもスコア(約10320と8907)がそこそこ高い。スコアの合計(2万弱)とsequence4の長さ(11,372 bp)の関係から、sequence4の大部分の領域がsequence1と類似していると判断できる

```
iu@bielinux[result] pwd
/home/iu/Desktop/result
iu@bielinux[result] ls -l s
-rw-rw-r-- 1 iu iu 132822 9月 4 16:22 sequence1_blast.bgr
-rw-rw-r-- 1 iu iu 12473981 9月 4 10:19 sequence1_blast.txt
-rw-rw-r-- 1 iu iu 9775990 9月 5 13:11 sequence1_blast.xml
-rw-rw-r-- 1 iu iu 2289509 8月 29 14:06 sequence1.fa
iu@bielinux[result] grep -n "> sequence" sequence1_blast.txt
27:> sequence1
205785:> sequence4
206506:> sequence3
206602:> sequence2
iu@bielinux[result] grep -n "Score =" sequence1_blast.txt | tail -n 6
205767: Score = 54.7 bits (29), Expect = 2e-04
205776: Score = 54.7 bits (29), Expect = 2e-04
205788: Score = 1.032e+04 bits (5588), Expect = 0.0
206173: Score = 8907 bits (4823), Expect = 0.0
206509: Score = 2113 bits (1144), Expect = 0.0
206605: Score = 124 bits (67), Expect = 1e-25
iu@bielinux[result] █
```

[11:52午前]

W8-1 : seq1 vs. seq4

BLASTViewerに戻り、①sequence4
とのヒット領域を眺める

BLAST results

- sequence1_blast.xml
- blastn vs. LH_hgap.fa
- sequence1

#	Accession	Definition	Quality	# HSPs
1	0	sequence1	☺	1347
2	3	sequence4	☺	2
3	2	sequence3	☺	1
4	1	sequence2	☺	1

Alignment: Query (2289497 nuc) vs. 3 (11,372 nuc)

HSP Map Definition Statistics Alignment

Query

HSP(s)

3

549500 549510 549520 549530 549540

T T T C C A G T G C T G G T G T A T A A A C G G C A A T C G G T T T A A T C G C T G A T C C T G G T T

T T T C C A G T G C T G G T G T A T A A A C G G C A A T C G G T T T A A T C G C T G A T C C T G G T T

11370 11360 11350 11340 11330

HSP: < 1/2 >

Discover KoriBlast
to go beyond the viewer

Welcome to BlastViewer

www.korilog.com 71Mo/123Mo

W8-1 : seq1 vs. seq4

大まかな見方を説明。①3という数字は、②Accessionのところの数字と同じものであり、3'末端などという意味ではない。③DB側配列(sequence4; 11,372 bp)、④query側配列(sequence1; 2,289,497 bp)

The screenshot shows the BlastViewer interface. On the left, the 'BLAST results' panel lists 'sequence1_blast.xml', 'blastn vs. LH_hgap.fa', and 'sequence1'. The main 'Summary' panel displays a table of hits:

#	Accession	Definition	Quality	# HSPs
1	0	sequence1	☺	1347
2	3	sequence4	☺	2
3	2	sequence3	☺	1
4	1	sequence2	☺	1

Below the table, the 'Alignment: Query (2289497 nuc) vs. 3 (11,372 nuc)' section is shown. It includes tabs for 'HSP Map', 'Definition', 'Statistics', and 'Alignment'. The alignment view shows the query sequence (top) and HSP(s) (middle) with red bars indicating high-scoring pairs. The sequence alignment (bottom) shows the query sequence (T T T C C A G T G C T G G T G T A T A A A C G G C A A T C G G T T T A A T C G C T G A T C C T G G T T) and the HSP(s) sequence (T T T C C A G T G C T G G T G T A T A A A C G G C A A T C G G T T T A A T C G C T G A T C C T G G T T) with vertical lines indicating matches. The alignment is viewed from position 11370 to 11330. Red callouts 1, 2, 3, and 4 point to specific elements: 1 points to the HSP(s) bar, 2 points to the Accession '3' in the hits table, 3 points to the HSP(s) bar, and 4 points to the HSP(s) bar.

W8-1 : seq1 vs. seq4

①DB側配列(sequence4; 11,372 bp)と、②query側配列(sequence1; 2,289,497 bp)は実際の長さは異なるが、長さを揃えて表示している

BlastViewer

File Help

BLAST results

sequence1_blast.xml
blastn vs. LH_hgap.fa
sequence1

Summary

Hits

#	Accession	Definition	Quality	# HSPs
1	0	sequence1	☺	1347
2	3	sequence4	☺	2
3	2	sequence3	☺	1
4	1	sequence2	☺	1

Alignment: Query (2289497 nuc) vs. 3 (11,372 nuc)

HSP Map Definition Statistics Alignment

Query

HSP(s)

3

549500 549510 549520 549530 549540

T T T C C A G T G C T G G T G T A T A A A C G G C A A T C G G T T T A A T C G C T G A T C C T G G T T
T T T C C A G T G C T G G T G T A T A A A C G G C A A T C G G T T T A A T C G C T G A T C C T G G T T

11370 11360 11350 11340 11330

Discover KoriBlast
to go beyond the viewer

Welcome to BlastViewer

HSP: < 1/2 >

www.korilog.com 71Mo/123Mo

99

W8-1 : seq1 vs. seq4

- ①(sequence4; 11,372 bp)中には、配列類似領域が2つ(HSPが2個)あったことを思い出そう。
- ②をクリックして最初に見ているのは、③スコアの大きい「Score = 1.032e+04 bits (5588)」のHSP。
- ③スコアの大きい「Score = 1.032e+04 bits (5588)」のHSP。
- ④対応するquery側のアラインメント領域

BLAST results

sequence1_blast.xml
blastn vs. LH_hgap.fa
sequence1

Summary

#	Accession	Definition	Quality	# HSPs
1	0	sequence1	☺	1347
2	3	sequence4	☺	2
3	2	sequence3	☺	1
4	1	sequence2	☺	1

Alignment: Query (2289497 nuc) vs. 3 (11,372 nuc)

HSP Map Definition Statistics Alignment

Query

HSP(s)

3

549500 549510 549520 549530 549540

T T T C C A G T G C T G G T G T A T A A A C G G C A A T C G G T T T A A T C G C T G A T C C T G G T T

T T T C C A G T G C T G G T G T A T A A A C G G C A A T C G G T T T A A T C G C T G A T C C T G G T T

11370 11360 11350 11340 11330

HSP: < 1/2 >

Discover KoriBlast
to go beyond the viewer

Welcome to BlastViewer

www.korilog.com 71Mo/123Mo

W8-1 : seq1 vs. seq4

①2つあるHSPのうち、②最初の1個目という意味。③次のHSPの Alignmentが見られる

BLAST results

sequence1_blast.xml
blastn vs. LH_hgap.fa
sequence1

Summary

Hits

#	Accession	Definition	Quality	# HSPs
1	0	sequence1	☺	1347
2	3	sequence4	☺	2
3	2	sequence3	☺	1
4	1	sequence2	☺	1

Alignment: Query (2289497 nuc) vs. 3 (11,372 nuc)

HSP Map Definition Statistics Alignment

Query

HSP(s)

3

549500 549510 549520 549530 549540

T T T C C A G T G C T G G T G T A T A A A C G G C A A T C G G T T T A A T C G C T G A T C C T G G T T

T T T C C A G T G C T G G T G T A T A A A C G G C A A T C G G T T T A A T C G C T G A T C C T G G T T

11370 11360 11350 11340 11330

HSP: < 1/2 >

Discover KoriBlast
to go beyond the viewer

Welcome to BlastViewer

www.korilog.com 71Mo/123Mo

W8-1 : seq1 vs. seq4

次のHSP(「Score = 8907 bits (4823)」)の ALIGNMENT。①赤枠の位置が変わっているのがわかる。②対応するquery側の ALIGNMENT 領域もわずかに右側にシフトしていることがわかる

The screenshot shows the BlastViewer interface. On the left, the 'BLAST results' panel lists 'sequence1_blast.xml', 'blastn vs. LH_hgap.fa', and 'sequence1'. The main area shows a 'Summary' tab with a 'Hits' table:

#	Accession	Definition	Quality	# HSPs
1	0	sequence1	😊	1347
2	3	sequence4	😊	2
3	2	sequence3	😊	1
4	1	sequence2	😊	1

Below the table is the 'Alignment: Query (2289497 nuc) vs. 3 (11,372 nuc)' section. It has tabs for 'HSP Map', 'Definition', 'Statistics', and 'Alignment'. The 'Alignment' tab is active, showing a query sequence and two HSPs. Red arrows point to a mismatch in the HSP alignment (1) and a shift in the query sequence alignment (2).

Query: T A A G A G T T C C A T A C T T T T G A C T G T T T C A G T A C C C A T G T C A C C A T G T G T A A T
HSP(s): T A A G A G T T C C A T A C T T T T G A C T G T T T C A G T A C C C A T G T C A C C A T G T G T A A T
3: T A A G A G T T C C A T A C T T T T G A C T G T T T C A G T A C C C A T G T C A C C A T G T G T A A T

Coordinates: 555170, 555180, 555190, 555200, 555210 (top); 4850, 4840, 4830, 4820, 4810 (bottom)

W8-1 : seq1 vs. seq4

これは、①query側配列(sequence1)が「左から右の方向(Plus鎖)」に並んでいるのに対して、②DB側配列(sequence4)が「右から左の方向(Minus鎖)」に並んでいることから納得できる。テキスト形式のBLAST結果ファイル(sequence1_blast.txt)を詳細に眺めて確認してもいいだろう

BlastViewer

File Help

BLAST results

sequence1_blast.xml
blastn vs. LH_hgap.fa
sequence1

Summary

Hits

#	Accession	Definition	Quality	# HSPs
1	0	sequence1	😊	1347
2	3	sequence4	😊	2
3	2	sequence3	😊	1
4	1	sequence2	😊	1

Alignment: Query (2289497 nuc) vs. 3 (11,372 nuc)

HSP Map Definition Statistics Alignment

Query

HSP(s)

3

555170 555180 555190 555200 555210

T A A G A G T T C C A T A C T T T T G A C T G T T T C A G T A C C C A T G T C A C C A T G T G T A A T

4850 4840 4830 4820 4810

T A A G A G T T C C A T A C T T T T G A C T G T T T C A G T A C C C A T G T C A C C A T G T G T A A T

Discover KoriBlast
to go beyond the viewer

Welcome to BlastViewer

HSP: < 2/2 >

www.korilog.com 82Mo/123Mo

①
②
①
②

①grep実行時に②-nと③-A 3を組み合わせて、
"Score ="と一致した行を含めた後ろの3行分を行
番号とともに表示させてもう少し詳細情報を表示

W8-2: seq1 vs. seq4

```
iu@bielinux[result] grep -n -A 3 "Score =" sequence1_blast.txt | tail -n 19
205788: Score = 1.032e+04 bits (5588), Expect = 0.0
205789- Identities = 5656/5684 (99%), Gaps = 23/5684 (0%)
205790- Strand=Plus/Minus
205791-
--
206173: Score = 8907 bits (4823), Expect = 0.0
206174- Identities = 4851/4862 (99%), Gaps = 11/4862 (0%)
206175- Strand=Plus/Minus
206176-
--
206509: Score = 2113 bits (1144), Expect = 0.0
206510- Identities = 1232/1276 (97%), Gaps = 0/1276 (0%)
206511- Strand=Plus/Minus
206512-
--
206605: Score = 124 bits (67), Expect = 1e-25
206606- Identities = 73/76 (96%), Gaps = 0/76 (0%)
206607- Strand=Plus/Minus
206608-
iu@bielinux[result] █
```

[3:25午後]

sequence4上の配列類似領域の①1つめと②2つめ。
③確かにストランド情報として、query側(sequence1)がPlus鎖、DB側(sequence4)がMinus鎖となっている

W8-2: seq1 vs. seq4

```
iu@bielinux[result] grep -n -A 3 "Score = " sequence1_blast.txt | tail -n 19
205788: Score = 1.032e+04 bits (5588), Expect = 0.0
205789- Identities = 5656/5684 (99%), Gaps = 23/5684 (0%)
205790- Strand=Plus/Minus
205791-
--
206173: Score = 8907 bits (4823), Expect = 0.0
206174- Identities = 4851/4862 (99%), Gaps = 11/4862 (0%)
206175- Strand=Plus/Minus
206176-
--
206509: Score = 2113 bits (1144), Expect = 0.0
206510- Identities = 1232/1276 (97%), Gaps = 0/1276 (0%)
206511- Strand=Plus/Minus
206512-
--
206605: Score = 124 bits (67), Expect = 1e-25
206606- Identities = 73/76 (96%), Gaps = 0/76 (0%)
206607- Strand=Plus/Minus
206608-
iu@bielinux[result] █
```



[3:25午後]

W8-3: HSP1

①1つめのHSP (以下、HSP1)に切り替えて、②Alignmentタブをクリック。ここでアラインメント結果の詳細情報がわかる

The screenshot shows the BlastViewer interface. On the left, the 'BLAST results' panel lists 'sequence1_blast.xml', 'blastn vs. LH_hgap.fa', and 'sequence1'. The main area is divided into 'Summary' and 'Alignment' sections. The 'Summary' section contains a table of hits:

#	Accession	Definition	Quality	# HSPs
1	0	sequence1	☺	1347
2	3	sequence4	☺	2
3	2	sequence3	☺	1
4	1	sequence2	☺	1

The 'Alignment' section is active, showing 'Alignment: Query (2289497 nuc) vs. 3 (11,372 nuc)'. It includes tabs for 'HSP Map', 'Definition', 'Statistics', and 'Alignment'. The 'Alignment' tab displays the following information:

Query: from 549,494 to 555,171, strand '+'
3: from 11,372 to 5,706, strand '-'
length: 5,684

The alignment is visualized as two rows of nucleotides (A, C, G, T) with vertical lines indicating matches. The top row is labeled with coordinates 549500, 549510, 549520, 549530, and 549540. The bottom row is labeled with coordinates 11370, 11360, 11350, 11340, and 11330. At the bottom of the alignment view, there are navigation controls for HSPs, including a left arrow, a '1/2' indicator, and a right arrow. A red arrow labeled '1' points to these controls. Another red arrow labeled '2' points to the 'Alignment' tab.

Discover KoriBlast
to go beyond the viewer

Welcome to BlastViewer

www.korilog.com 99Mo/123Mo

W8-3: HSP1

①query側(sequence1)の549,494番目と、②DB側(sequence4)の11,372番目から…③を一番右まで移動

The screenshot shows the BlastViewer interface. On the left, the 'BLAST results' panel lists 'sequence1_blast.xml', 'blastn vs. LH_hgap.fa', and 'sequence1'. The main window displays a 'Summary' tab with a 'Hits' table:

#	Accession	Definition	Quality	# HSPs
1	0	sequence1	☺	1347
2	3	sequence4	☺	2
3	2	sequence3	☺	1
4	1	sequence2	☺	1

Below the table is the 'Alignment: Query (2289497 nuc) vs. 3 (11,372 nuc)' section. It has tabs for 'HSP Map', 'Def', 'Statistics', and 'Alignment'. The 'Def' tab is active, showing:

Query: from 549,494 to 555,171, strand '+'
3: from 11,372 to 5,706, strand '-'
length: 5,684

The alignment view shows two DNA sequences with vertical lines indicating matches. Red arrows point to specific positions: arrow ① points to the 'Def' tab; arrow ② points to the '11370' position on the bottom sequence; arrow ③ points to the '11370' position on the top sequence.

At the bottom, there is a 'Discover KoriBlast' link, a 'Welcome to BlastViewer' message, a 'www.korilog.com' link, and a '99Mo/123Mo' indicator. The page number '107' is in the bottom right corner.

W8-3:HSP1

①query側(sequence1)の555,171番目と、②DB側(sequence4)の5,706番目の領域がHSP1。③を一番右まで移動させた結果で見えています

The screenshot displays the BlastViewer interface. On the left, the 'BLAST results' sidebar shows 'sequence1_blast.xml', 'blastn vs. LH_hgap.fa', and 'sequence1'. The main window shows a 'Summary' tab with a 'Hits' table:

#	Accession	Definition	Quality	# HSPs
1	0	sequence1	☺	1347
2	3	sequence4	☺	2
3	2	sequence3	☺	1
4	1	sequence2	☺	1

Below the table, the 'Alignment: Query (2289497 nuc) vs. 3 (11,372 nuc)' section is visible. The 'Stable' tab is selected, indicated by arrow ①. The query and subject sequences are shown:

Query: from 549,494 to 555,171, strand '+'
3: from 11,372 to 5,706, strand '-'
length: 5,684 (indicated by arrow ②)

The alignment sequence is displayed with nucleotide bases (C, G, C, C, T, G, G, G, T, G, T, A, A, A, C, C, C, A, A, C, G, C, A, G, T, T, A, C, A, T, T, T, T, C, T, T, G, G, G, G, A, C, C, C, A, T, T, A, A, G, A) and positions (120, 555130, 555140, 555150, 555160, 555170) above and below. Arrow ③ points to the alignment sequence.

At the bottom, the 'HSP:' section shows navigation buttons and '1/2'. The footer includes 'Welcome to BlastViewer', 'www.korilog.com', and '87Mo/123Mo'.

W8-4:HSP2

BlastViewer

File Help

BLAST results

sequence1_blast.xml
blastn vs. LH_hgap.fa
sequence1

Summary

Hits

#	Accession	Definition	Quality	# HSPs
1	0	sequence1	☺	1347
2	3	sequence4	☺	2
3	2	sequence3	☺	1
4	1	sequence2	☺	1

Alignment: Query (2289497 nuc) vs. 3 (11,372 nuc)

HSP Map Definition Statistics Alignment

Query: from 555,167 to 560,027, strand '+'
3: from 4,852 to 1, strand '-'
length: 4,862

555170 555180 555190 555200 555210

T A A G A G T T C C A T A C T T T T G A C T G T T T C A G T A C C C A T G T C A C C A T G T G T A A T
T A A G A G T T C C A T A C T T T T G A C T G T T T C A G T A C C C A T G T C A C C A T G T G T A A T

4850 4840 4830 4820 4810

Discover KoriBlast
to go beyond the viewer

Welcome to BlastViewer

HSP: < 2/2 >

www.korilog.com 68Mo/123Mo

W8-4: HSP2

①query側(sequence1)の555,167番目と、②DB側(sequence4)の4,852番目から…③を一番右まで移動

BLAST results

#	Accession	Definition	Quality	# HSPs
1	0	sequence1	☺	1347
2	3	sequence4	☺	2
3	2	sequence3	☺	1
4	1	sequence2	☺	1

Alignment: Query (2289497 nuc) vs. 3 (11,372 nuc)

HSP Map Def **①** Statistics Alignment

Query: from 555,167 to 560,027, strand '+'
3: from 4,852 to 1, strand '-'
length: 4,862 **②**

①

555170 555180 555190 555200 555210

T A A G A G T T C C A T A C T T T T G A C T G T T T C A G T A C C C A T G T C A C C A T G T G T A A T

T A A G A G T T C C A T A C T T T T G A C T G T T T C A G T A C C C A T G T C A C C A T G T G T A A T

4850 4840 4830 4820 4810

② **③**

HSP: < 2/2 >

Discover KoriBlast
to go beyond the viewer

Welcome to BlastViewer www.korilog.com 68Mo/123Mo

W8-4: HSP2

①query側(sequence1)の560,027番目と、②DB側(sequence4)の1番目の領域がHSP2。③を一番右まで移動させた結果で見えています

BLAST results

#	Accession	Definition	Quality	# HSPs
1	0	sequence1	☺	1347
2	3	sequence4	☺	2
3	2	sequence3	☺	1
4	1	sequence2	☺	1

Alignment: Query (2289497 nuc) vs. 3 (11,372 nuc)

HSP Map Definition Start **①** Alignment

Query: from 555,167 to 560,027, strand '+'
3: from 4,852 to 1, strand '-'
length: 4,862 **②**

559980 559990 560000 560010 560020

T C A C A C C A G C A T T G C A A T A T A A T A C G T C A A C C C G G C C A A A T A C T T T C T T A G
T C A C A C C A G C A T T G C A A T A T A A T A C G T C A A C C C G G C C A A A T A C T T T C T T A G

50 40 30 20 10

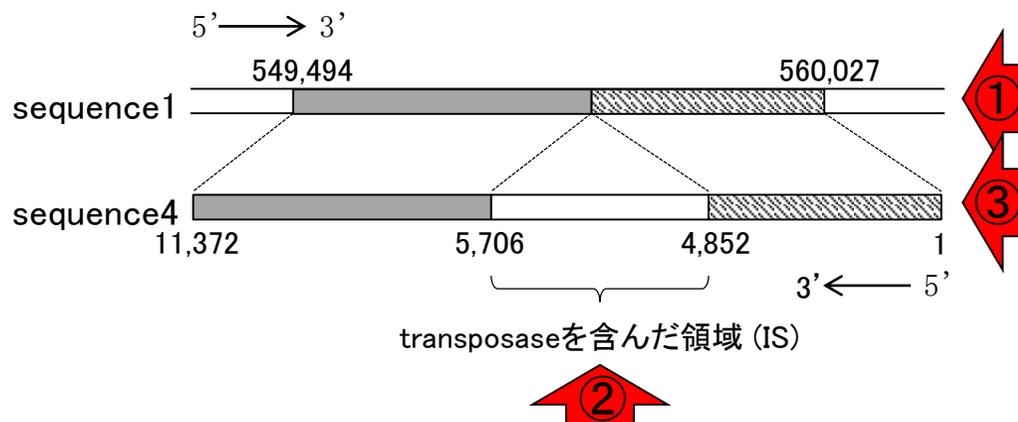
HSP: < 2/2 >

Discover KoriBlast
to go beyond the viewer

Welcome to BlastViewer www.korilog.com 90Mo/123Mo

W8-5: 模式図

sequence1とsequence4のアラインメント模式図。① sequence1の一続きの領域[549494, 560027 bp]と、②領域[4853, 5705 bp]を除く、③sequence4の全長がほぼ一致。そして、④アラインメントされなかった領域[4853, 5705 bp]には、transposaseがコードされていた(W4-9)。第8回の図1と同じ



sequence4	4133..4753	CDS	mannose-specific PTS system IID component
sequence4	4781..4873	CDS	mannose-specific PTS system IIA component
sequence4	4912..5220	CDS	transposase
sequence4	5361..5699	CDS	transposase
sequence4	5711..6058	CDS	mannose-specific PTS system IIA component
sequence4	6114..6596	CDS	mannose/fructose/sorbose-specific PTS system IID component



W9-1 : seq1 vs. seq1

query側(sequence1)は固定で、①DB側が sequence1のBLAST結果を表示。②配列類似領域(HSP)は1,347個あったことを思い出そう。③スコアトップのHSP (HSP1)が、④seq1 vs. seq1の全長が100%一致のHSPとなるのは当たり前

BLAST results

- sequence1_blast.xml
- blastn vs. LH_hgap.fa
- sequence1

Summary

Hits

#	Accession	Definition	Quality	# HSPs
1	0	sequence1	☺	1347
2	3	sequence4	☺	2
3	2	sequence3	☺	1
4	1	sequence2	☺	1

Alignment: Query (2289497 nuc) vs. 0 (2,289,497 nuc)

HSP Map Definition Statistics Alignment

Query: from 1 to 2,289,497, strand '+'
0: from 1 to 2,289,497, strand '+'
length: 2,289,497

60 70 80 90 100
G C G G T G G C A A T C G C G C T G A C A G A T T T A C G C T C A A A G G A A A C C A T G A T G A T G C
G C G G T G G C A A T C G C G C T G A C A G A T T T A C G C T C A A A G G A A A C C A T G A T G A T G C
60 70 80 90 100

HSP: < 1/1347 >

Discover KoriBlast
to go beyond the viewer

Welcome to BlastViewer

www.korilog.com 77Mo/123Mo

W9-1 : seq1 vs. seq1

①や②や③を押して、スコアの上位33位 (HSP1-33)までの全体像をまとめたのが…

BLAST results

#	Accession	Definition	Quality	# HSPs
1	0	sequence1	☺	1347
2	3	sequence4	☺	2
3	2	sequence3	☺	1
4	1	sequence2	☺	1

Alignment: Query (2,289,497 nuc)

HSP Map | Definition | Statistics | Alignment

Query: from 1 to 2,289,497, strand '+'
0: from 1 to 2,289,497, strand '+'
length: 2,289,497

```
      60      70      80      90     100
G C G G T G G C A A T C G C G C T G A C A G A T T T A C G C T C A A A G G A A A C C A T G A T G A T G C
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
G C G G T G G C A A T C G C G C T G A C A G A T T T A C G C T C A A A G G A A A C C A T G A T G A T G C
      60      70      80      90     100
```

HSP: < 1/1347 >

Discover KoriBlast
to go beyond the viewer

Welcome to BlastViewer www.korilog.com 77Mo/123Mo

W9-2:HSP1-33

これです。①括弧内のHSPは、カッコ外のHSPとqueryとDB側の一致領域が入れ替わっているだけで、実質的に同じものです

HSP番号	スコア	一致領域(query側)			一致領域(DB側)			Length	query側の領域	DB側の領域
		start	end	strand	start	end	strand			
HSP1	4,227,900	1	2,289,497	plus	1	2,289,497	plus	2,289,497		
HSP2 (HSP3)	11,624	2,058,465	2,065,193	plus	2,057,850	2,064,578	plus	6,732		
HSP4 (HSP5)	10,754	37,329	43,187	plus	1	5,860	plus	5,865	①'	①
HSP6 (HSP7)	10,516	2,059,080	2,065,193	plus	2,057,850	2,063,963	plus	6,117		
HSP8 (HSP9)	10,429	2,283,820	2,289,497	plus	5,839	11,509	plus	5,679	②'	②
HSP10 (HSP11)	9,679	2,273,804	2,279,100	plus	717,190	711,892	minus	5,301		
HSP12 (HSP13)	9,524	2,059,672	2,065,193	plus	2,057,830	2,063,348	plus	5,525		
HSP14 (HSP15)	9,362	1,088,170	1,093,274	plus	999,673	1,004,778	plus	5,106		
HSP16 (HSP17)	8,385	2,060,310	2,065,194	plus	2,057,850	2,062,735	plus	4,887		
HSP18 (HSP19)	7,301	2,060,901	2,065,194	plus	2,057,830	2,062,120	plus	4,296		
HSP20 (HSP21)	6,186	2,061,516	2,065,193	plus	2,057,830	2,061,504	plus	3,680		
HSP22 (HSP23)	6,047	2,275,808	2,279,168	plus	1,002,969	999,608	minus	3,365		
HSP24 (HSP26)	6,043	999,670	1,002,969	plus	711,886	715,184	plus	3,300		
HSP25 (HSP27)	6,043	1,088,173	1,091,466	plus	711,892	715,184	plus	3,294		
HSP28 (HSP29)	6,038	2,275,808	2,279,100	plus	1,091,466	1,088,173	minus	3,295		
HSP30 (HSP31)	5,050	2,062,154	2,065,193	plus	2,057,850	2,060,889	plus	3,042		
HSP32 (HSP33)	4,021	2,062,769	2,065,197	plus	2,057,850	2,060,279	plus	2,431		



W9-2:HSP1-33

HSP番号	スコア	一致領域(query側)			一致領域(DB側)			Length	query側の領域	DB側の領域
		start	end	strand	start	end	strand			
HSP1	4,227,900	1	2,289,497	plus	1	2,289,497	plus	2,289,497		
HSP4 (HSP5)	10,754	37,329	43,187	plus	1	5,860	plus	5,865	①'	①
HSP24 (HSP26)	6,043	999,670	1,002,969	plus	711,886	715,184	plus	3,300		
HSP14 (HSP15)	9,362	1,088,170	1,093,274	plus	999,673	1,004,778	plus	5,106		
HSP25 (HSP27)	6,043	1,088,173	1,091,466	plus	711,892	715,184	plus	3,294		
HSP2 (HSP3)	11,624	2,058,465	2,065,193	plus	2,057,850	2,064,578	plus	6,732		
HSP6 (HSP7)	10,516	2,059,080	2,065,193	plus	2,057,850	2,063,963	plus	6,117		
HSP12 (HSP13)	9,524	2,059,672	2,065,193	plus	2,057,830	2,063,348	plus	5,525		
HSP16 (HSP17)	8,385	2,060,310	2,065,194	plus	2,057,850	2,062,735	plus	4,887		
HSP18 (HSP19)	7,301	2,060,901	2,065,194	plus	2,057,830	2,062,120	plus	4,296		
HSP20 (HSP21)	6,186	2,061,516	2,065,193	plus	2,057,830	2,061,504	plus	3,680		
HSP30 (HSP31)	5,050	2,062,154	2,065,193	plus	2,057,850	2,060,889	plus	3,042		
HSP32 (HSP33)	4,021	2,062,769	2,065,197	plus	2,057,850	2,060,279	plus	2,431		
HSP10 (HSP11)	9,679	2,273,804	2,279,100	plus	717,190	711,892	minus	5,301		
HSP22 (HSP23)	6,047	2,275,808	2,279,168	plus	1,002,969	999,608	minus	3,365		
HSP28 (HSP29)	6,038	2,275,808	2,279,100	plus	1,091,466	1,088,173	minus	3,295		
HSP8 (HSP9)	10,429	2,283,820	2,289,497	plus	5,839	11,509	plus	5,679	②'	②



①

W9-2:HSP1-33

HSP番号	スコア	一致領域(query側)			一致領域(DB側)			Length	query側の領域	DB側の領域
		start	end	strand	start	end	strand			
HSP1	4,227,901	1	2,289,497	plus	1	2,289,497	plus	2,289,497		
HSP4 (HSP5)	10,741	37,329	43,187	plus	1	5,860	plus	5,865	①'	①
HSP24 (HSP26)	6,045	999,670	1,002,969	plus	711,886	715,184	plus	3,300		
HSP14 (HSP15)	9,362	1,088,170	1,093,274	plus	999,673	1,004,778	plus	5,106		
HSP25 (HSP27)	6,043	1,088,173	1,091,466	plus	711,892	715,184	plus	3,294		
HSP2 (HSP3)	11,624	2,058,465	2,065,193	plus	2,057,850	2,064,578	plus	6,732		
HSP6 (HSP7)	10,516	2,059,080	2,065,193	plus	2,057,850	2,063,963	plus	6,117		
HSP12 (HSP13)	9,524	2,059,672	2,065,193	plus	2,057,830	2,063,348	plus	5,525		
HSP16 (HSP17)	8,385	2,060,310	2,065,194	plus	2,057,850	2,062,735	plus	4,887		
HSP18 (HSP19)	7,301	2,060,901	2,065,194	plus	2,057,830	2,062,120	plus	4,296		
HSP20 (HSP21)	6,186	2,061,516	2,065,193	plus	2,057,830	2,061,504	plus	3,680		
HSP30 (HSP31)	5,050	2,062,154	2,065,193	plus	2,057,850	2,060,889	plus	3,042		
HSP32 (HSP33)	4,021	2,062,769	2,065,197	plus	2,057,850	2,060,279	plus	2,431		
HSP10 (HSP11)	9,679	2,273,804	2,279,100	plus	717,190	711,892	minus	5,301		
HSP22 (HSP23)	6,047	2,275,808	2,279,168	plus	1,002,969	999,608	minus	3,365		
HSP28 (HSP29)	6,031	2,275,808	2,279,100	plus	1,091,466	1,088,173	minus	3,295		
HSP8 (HSP9)	10,412	2,283,820	2,289,497	plus	5,839	11,509	plus	5,679	②'	②

W9-2:HSP1-33

右側の①' が領域[37329, 43187 bp]、①が[1, 5860 bp]、
②' が [2283820, 2289497 bp]、②が[5839, 11509 bp]に相
当。この「①' , ①, ②' , ②」が次のスライドで使われる

HSP番号	スコア	一致領域(query側)			一致領域(DB側)			Length	query側 の領域	DB側の 領域
		start	end	strand	start	end	strand			
HSP1	4,227,900	1	2,289,497	plus	1	2,289,497	plus	2,289,497		
HSP4 (HSP5)	10,754	37,329	43,187	plus	1	5,860	plus	5,865	①'	①
HSP24 (HSP26)	6,043	999,670	1,002,969	plus	711,886	715,184	plus	3,300		
HSP14 (HSP15)	9,362	1,088,170	1,093,274	plus	999,673	1,004,778	plus	5,106		
HSP25 (HSP27)	6,043	1,088,173	1,091,466	plus	711,892	715,184	plus	3,294		
HSP2 (HSP3)	11,624	2,058,465	2,065,193	plus	2,057,850	2,064,578	plus	6,732		
HSP6 (HSP7)	10,516	2,059,080	2,065,193	plus	2,057,850	2,063,963	plus	6,117		
HSP12 (HSP13)	9,524	2,059,672	2,065,193	plus	2,057,830	2,063,348	plus	5,525		
HSP16 (HSP17)	8,385	2,060,310	2,065,194	plus	2,057,850	2,062,735	plus	4,887		
HSP18 (HSP19)	7,301	2,060,901	2,065,194	plus	2,057,830	2,062,120	plus	4,296		
HSP20 (HSP21)	6,186	2,061,516	2,065,193	plus	2,057,830	2,061,504	plus	3,680		
HSP30 (HSP31)	5,050	2,062,154	2,065,193	plus	2,057,850	2,060,889	plus	3,042		
HSP32 (HSP33)	4,021	2,062,769	2,065,197	plus	2,057,850	2,060,279	plus	2,431		
HSP10 (HSP11)	9,679	2,273,804	2,279,100	plus	717,190	711,892	minus	5,301		
HSP22 (HSP23)	6,047	2,275,808	2,279,168	plus	1,002,969	999,608	minus	3,365		
HSP28 (HSP29)	6,038	2,275,808	2,279,100	plus	1,091,466	1,088,173	minus	3,295		
HSP8 (HSP9)	10,429	2,283,820	2,289,497	plus	5,839	11,509	plus	5,679	②'	②

W9-3: アノテーション

HSP番号	スコア	一致領域(query側)			一致領域(DB側)			Length	query側の領域	DB側の領域
		start	end	strand	start	end	strand			
HSP1	4,227,900	1	2,289,497	plus	1	2,289,497	plus	2,289,497		
HSP4 (HSP5)	10,754	37,329	43,187	plus	1	5,860	plus	5,865	①'	①
HSP24 (HSP26)	6,041	999,670	1,002,969	plus	711,886	715,184	plus	3,300	ribosomal RNA	
HSP14 (HSP15)	9,303	1,088,170	1,093,274	plus	999,673	1,004,778	plus	5,106	ribosomal RNA	
HSP25 (HSP27)	6,041	1,088,173	1,091,466	plus	711,892	715,184	plus	3,294	ribosomal RNA	
HSP2 (HSP3)	11,624	2,058,465	2,065,193	plus	2,057,850	2,064,578	plus	6,732		
HSP6 (HSP7)	10,516	2,059,080	2,065,193	plus	2,057,850	2,063,963	plus	6,117		
HSP12 (HSP13)	9,524	2,059,672	2,065,193	plus	2,057,830	2,063,348	plus	5,525		
HSP16 (HSP17)	8,385	2,060,310	2,065,194	plus	2,057,850	2,062,735	plus	4,887		
HSP18 (HSP19)	7,301	2,060,901	2,065,194	plus	2,057,830	2,062,120	plus	4,296		
HSP20 (HSP21)	6,186	2,061,516	2,065,193	plus	2,057,830	2,061,504	plus	3,680		
HSP30 (HSP31)	5,050	2,062,154	2,065,193	plus	2,057,850	2,060,889	plus	3,042		
HSP32 (HSP33)	4,021	2,062,769	2,065,197	plus	2,057,850	2,060,279	plus	2,431		
HSP10 (HSP11)	9,671	2,273,804	2,279,100	plus	717,190	711,892	minus	5,301	ribosomal RNA	
HSP22 (HSP23)	6,035	2,275,808	2,279,168	plus	1,002,969	999,608	minus	3,365	ribosomal RNA	
HSP28 (HSP29)	6,035	2,275,808	2,279,100	plus	1,091,466	1,088,173	minus	3,295	ribosomal RNA	
HSP8 (HSP9)	10,429	2,283,820	2,289,497	plus	5,839	11,509	plus	5,679	②'	②

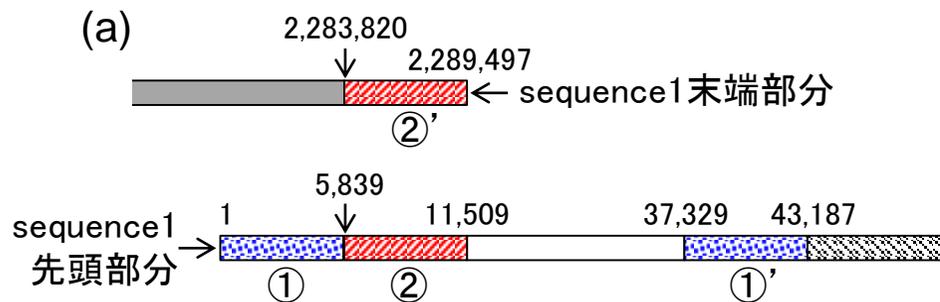
W9-3: アノテーション

④のあたりは、「adhesion exoprotein、mucus-binding protein、hypothetical protein」となっており、はっきり言ってよくわからない

HSP番号	スコア	一致領域(query側)			一致領域(DB側)			Length	query側の領域	DB側の領域
		start	end	strand	start	end	strand			
HSP1	4,227,900	1	2,289,497	plus	1	2,289,497	plus	2,289,497		
HSP4 (HSP5)	10,754	37,329	43,187	plus	1	5,860	plus	5,865	①'	①
HSP24 (HSP26)	6,043	999,670	1,002,969	plus	711,886	715,184	plus	3,300	ribosomal RNA	
HSP14 (HSP15)	9,362	1,088,170	1,093,274	plus	999,673	1,004,778	plus	5,106	ribosomal RNA	
HSP25 (HSP27)	6,043	1,088,173	1,091,466	plus	711,892	715,184	plus	3,294	ribosomal RNA	
HSP2 (HSP3)	11,624	2,058,465	2,065,193	plus	2,057,850	2,064,578	plus	6,732		
HSP6 (HSP7)	10,516	2,059,080	2,065,193	plus	2,057,850	2,063,963	plus	6,117		
HSP12 (HSP13)	9,524	2,059,672	2,065,193	plus	2,057,830	2,063,348	plus	5,525		
HSP16 (HSP17)	8,388	2,060,310	2,065,194	plus	2,057,850	2,062,735	plus	4,887		
HSP18 (HSP19)	7,360	2,060,901	2,065,194	plus	2,057,830	2,062,120	plus	4,296		
HSP20 (HSP21)	6,186	2,061,516	2,065,193	plus	2,057,830	2,061,504	plus	3,680		
HSP30 (HSP31)	5,050	2,062,154	2,065,193	plus	2,057,850	2,060,889	plus	3,042		
HSP32 (HSP33)	4,021	2,062,769	2,065,197	plus	2,057,850	2,060,279	plus	2,431		
HSP10 (HSP11)	9,679	2,273,804	2,279,100	plus	717,190	711,892	minus	5,301	ribosomal RNA	
HSP22 (HSP23)	6,047	2,275,808	2,279,168	plus	1,002,969	999,608	minus	3,365	ribosomal RNA	
HSP28 (HSP29)	6,038	2,275,808		sequence1	2055735..2060864	CDS		adhesion exoprotein		NA
HSP8 (HSP9)	10,429	2,283,820		sequence1	2060942..2063323	CDS		mucus-binding protein		②
				sequence1	2063401..2066928	CDS		hypothetical protein		



W10-1 : sequence1

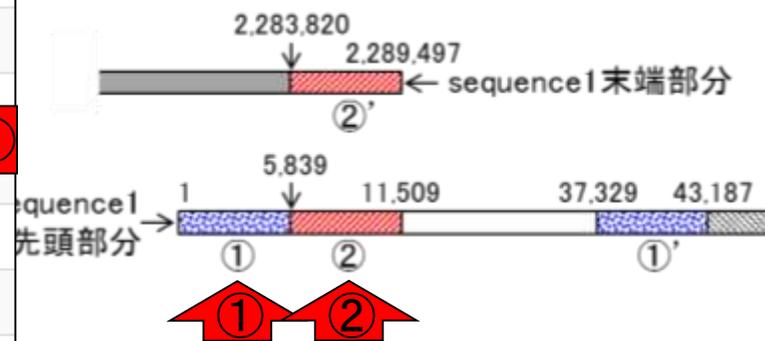


sequence1の末端付近の重複は、HSP8の一致領域(②と②')に相当。これにHSP4の一致領域(①と①')を含めた模式図。第8回の図2aと同じ。後の検証により、①の領域は染色体上に存在しないことが確認された

W10-2: アノテーション

HSP4の一致領域①[1, 5860 bp]のDFAST
アノテーション。HSP8の一致領域②[5839,
11509 bp]のアノテーションの一部

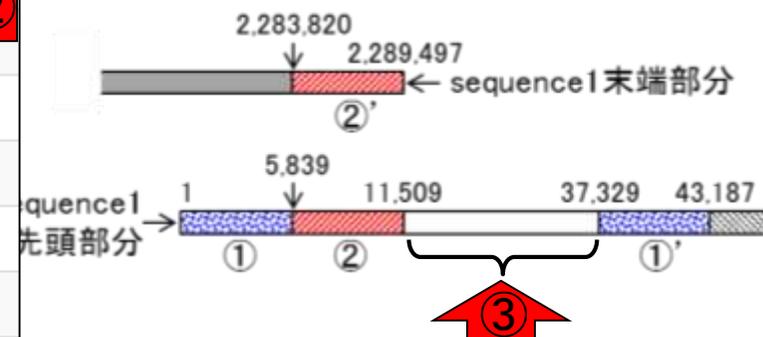
LOCUS_00001	sequence1	151..384	CDS	hypothetical protein	
LOCUS_00002	sequence1	350..886	CDS	hypothetical protein	
LOCUS_00003	sequence1	883..1311	CDS	hypothetical protein	
LOCUS_00004	sequence1	1637..1849	CDS	hypothetical protein	
LOCUS_00005	sequence1	1968..2165	CDS	hypothetical protein	
LOCUS_00006	sequence1	2355..2732	CDS	prophage protein	①
LOCUS_00007	sequence1	2725..2940	CDS	hypothetical protein	
LOCUS_00008	sequence1	2930..3031	CDS	hypothetical protein	
LOCUS_00009	sequence1	3067..3534	CDS	holin	
LOCUS_00010	sequence1	3547..4578	CDS	1,4-beta-N-acetylmuramidase	
LOCUS_00011	sequence1	4765..4935	CDS	hypothetical protein	
LOCUS_00012	sequence1	4936..5517	CDS	hypothetical protein	
LOCUS_00013	sequence1	complement (6059..7228)	CDS	integrase	
LOCUS_00014	sequence1	complement (7407..8267)	CDS	hypothetical protein	②
LOCUS_00015	sequence1	complement (8301..8813)	CDS	hypothetical protein	



W10-2: アノテーション

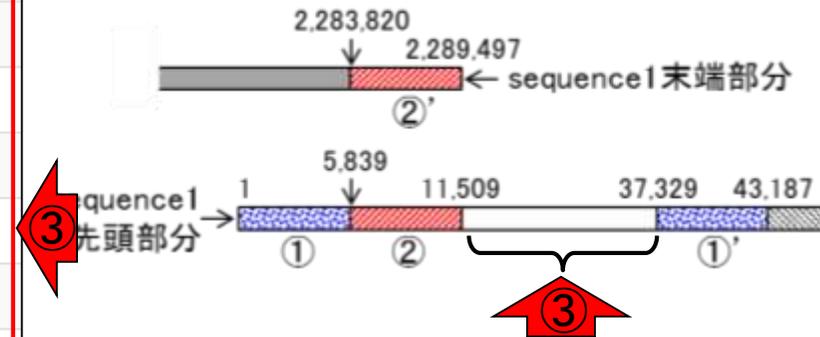
②[5839, 11509 bp]のアノテーションの残り。
 ③[11510, 37328 bp]のアノテーションの一部

LOCUS_00016	sequence1	complement (8911..9315)	CDS	hypothetical protein	
LOCUS_00017	sequence1	complement (9317..9640)	CDS	XRE family transcriptional regulator	
LOCUS_00018	sequence1	9916..10098	CDS	hypothetical protein	②
LOCUS_00019	sequence1	10104..10265	CDS	hypothetical protein	
LOCUS_00020	sequence1	10317..11030	CDS	phage-related antirepressor	
LOCUS_00021	sequence1	11068..11355	CDS	hypothetical protein	
LOCUS_00022	sequence1	11395..11490	CDS	hypothetical protein	
LOCUS_00023	sequence1	11503..11736	CDS	hypothetical protein	
LOCUS_00024	sequence1	complement (11714..12211)	CDS	hypothetical protein	
LOCUS_00025	sequence1	12396..12629	CDS	hypothetical protein	
LOCUS_00026	sequence1	12616..13254	CDS	hypothetical protein	③
LOCUS_00027	sequence1	13258..13974	CDS	phage NTP-binding protein	
LOCUS_00028	sequence1	13977..14576	CDS	hypothetical protein	
LOCUS_00029	sequence1	14593..14688	CDS	hypothetical protein	
LOCUS_00030	sequence1	14689..15522	CDS	hypothetical protein	



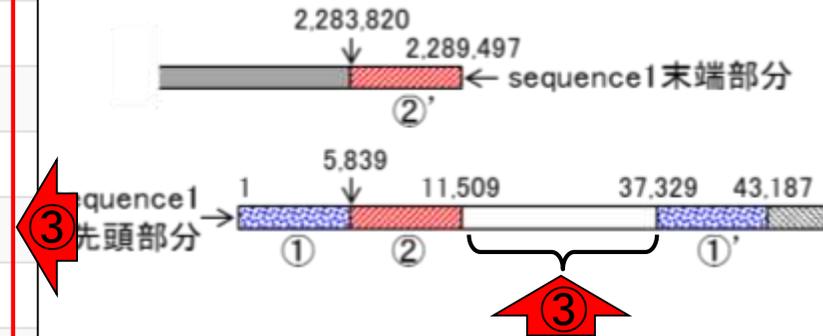
W10-2: アノテーション

LOCUS_00031	sequence1	15519..16757	CDS	replicative DNA helicase
LOCUS_00032	sequence1	16760..17158	CDS	hypothetical protein
LOCUS_00033	sequence1	17162..17302	CDS	hypothetical protein
LOCUS_00034	sequence1	17306..17671	CDS	hypothetical protein
LOCUS_00035	sequence1	17668..17859	CDS	hypothetical protein
LOCUS_00036	sequence1	17908..18282	CDS	hypothetical protein
LOCUS_00037	sequence1	18279..18482	CDS	hypothetical protein
LOCUS_00038	sequence1	18533..18700	CDS	hypothetical protein
LOCUS_00039	sequence1	18684..18857	CDS	hypothetical protein
LOCUS_00040	sequence1	18859..19314	CDS	Holliday junction resolvase
LOCUS_00041	sequence1	19314..19559	CDS	hypothetical protein
LOCUS_00042	sequence1	19649..20095	CDS	phage protein
LOCUS_00043	sequence1	20310..20585	CDS	hypothetical protein
LOCUS_00044	sequence1	20578..21057	CDS	hypothetical protein
LOCUS_00045	sequence1	21180..21698	CDS	phage terminase small subunit



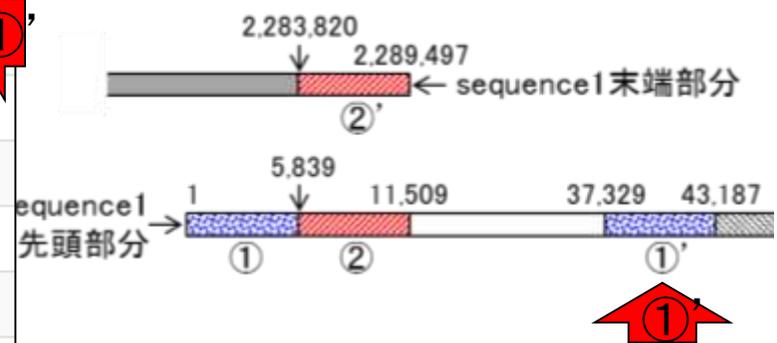
W10-2: アノテーション

LOCUS_00046	sequence1	21698..23599	CDS	phage terminase large subunit
LOCUS_00047	sequence1	23609..23785	CDS	hypothetical protein
LOCUS_00048	sequence1	23786..24961	CDS	phage portal protein
LOCUS_00049	sequence1	24942..26834	CDS	phage capsid protein
LOCUS_00050	sequence1	27026..27313	CDS	hypothetical protein
LOCUS_00051	sequence1	27294..27671	CDS	hypothetical protein
LOCUS_00052	sequence1	27674..28096	CDS	hypothetical protein
LOCUS_00053	sequence1	28096..28476	CDS	hypothetical protein
LOCUS_00054	sequence1	28480..29088	CDS	phage tail protein
LOCUS_00055	sequence1	29239..29589	CDS	tail protein
LOCUS_00056	sequence1	29793..34418	CDS	phage tail tape measure protein
LOCUS_00057	sequence1	34434..35273	CDS	phage protein
LOCUS_00058	sequence1	35230..36441	CDS	prophage protein
LOCUS_00059	sequence1	36425..36751	CDS	hypothetical protein
LOCUS_00060	sequence1	36751..38214	CDS	hypothetical protein



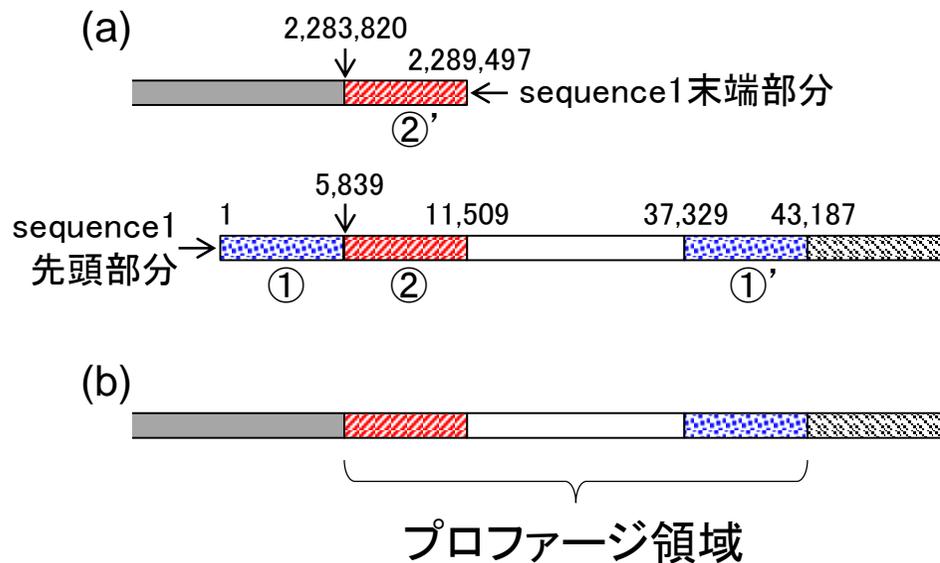
W10-2: アノテーション

LOCUS_00061	sequence1	38211..38966	CDS	hypothetical protein
LOCUS_00062	sequence1	38966..39298	CDS	hypothetical protein
LOCUS_00063	sequence1	39295..39492	CDS	hypothetical protein
LOCUS_00064	sequence1	39683..40060	CDS	prophage protein
LOCUS_00065	sequence1	40053..40268	CDS	hypothetical protein
LOCUS_00066	sequence1	40394..40861	CDS	holin
LOCUS_00067	sequence1	40874..41905	CDS	1,4-beta-N-acetylmuramidase
LOCUS_00068	sequence1	42092..42262	CDS	hypothetical protein
LOCUS_00069	sequence1	42263..42844	CDS	hypothetical protein
LOCUS_00070	sequence1	43586..44437	CDS	homoserine O-succinyltransferase
LOCUS_00071	sequence1	44450..45724	CDS	O-acetylhomoserine aminocarboxypropyltransferase
LOCUS_00072	sequence1	45808..46845	CDS	competence protein CoiA
LOCUS_00073	sequence1	complement (46909..47541)	CDS	dithiol-disulfide isomerase
LOCUS_00074	sequence1	47818..48507	CDS	GTP pyrophosphokinase
LOCUS_00075	sequence1	48519..49328	CDS	ATP-NAD kinase



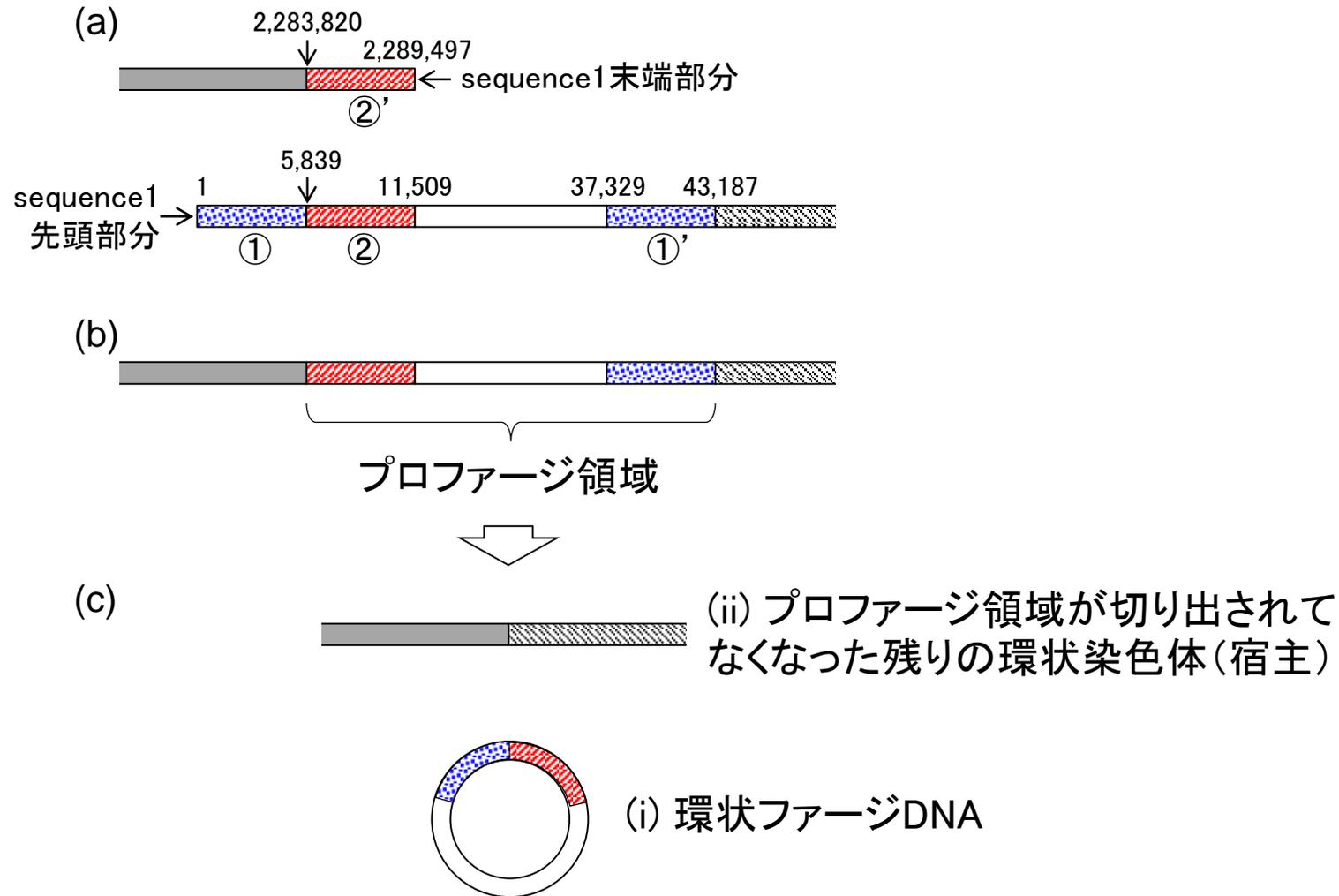
W10-3: 染色体構造

(b) sequence1の両末端である①と②'の領域をトリム。得られた領域[5839, 2283819 bp]の両末端を結合した環状コンティグが実際の染色体構造であると予想した



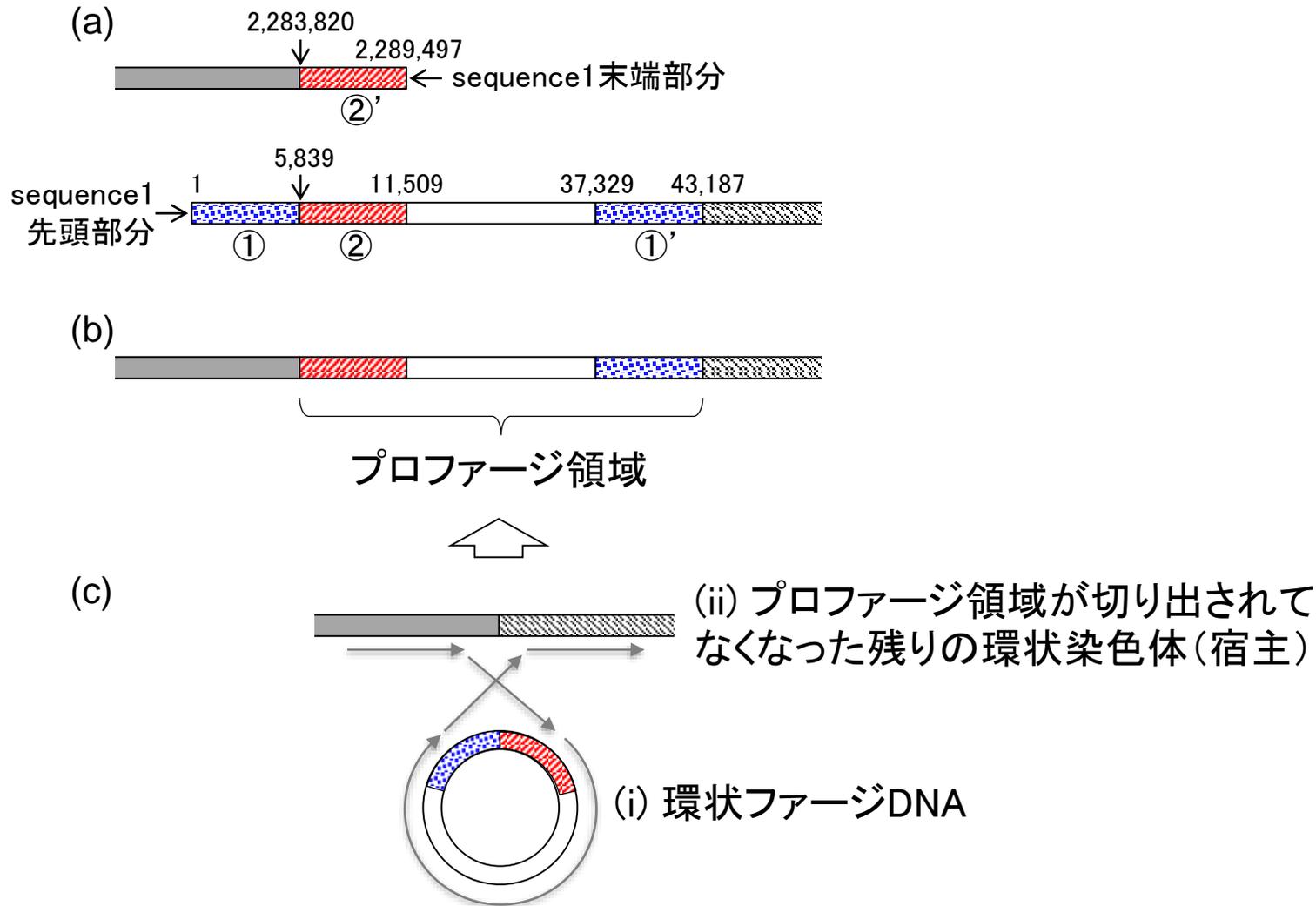
(b) → (c)プロファージ領域が染色体から切り出されて環状化した状態(excision)

W10-4: ファージの機構



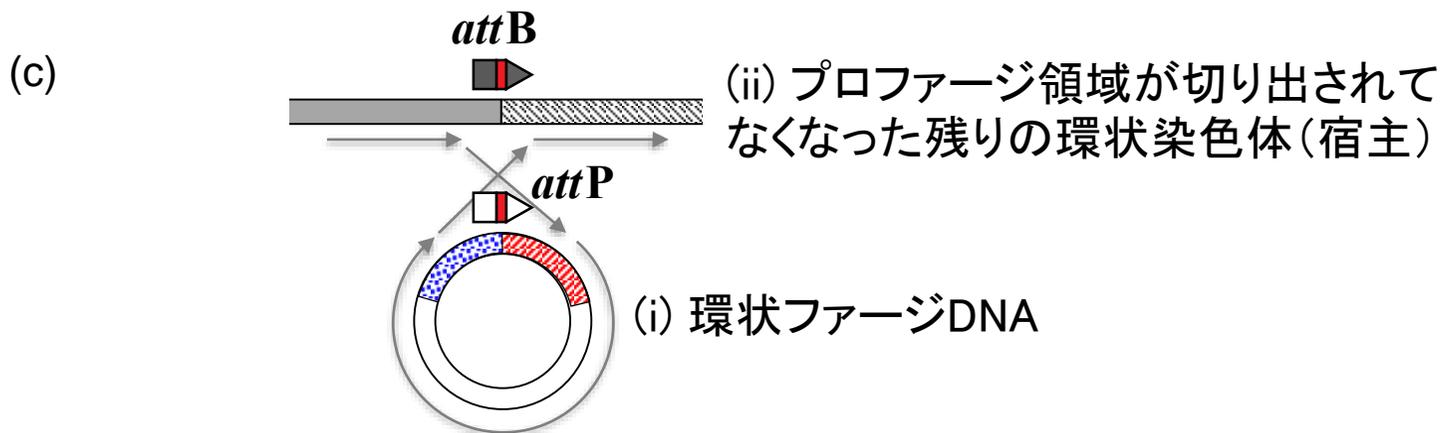
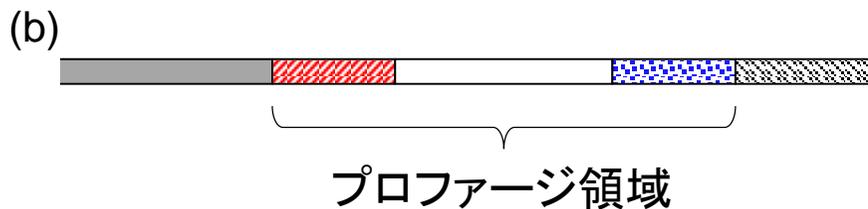
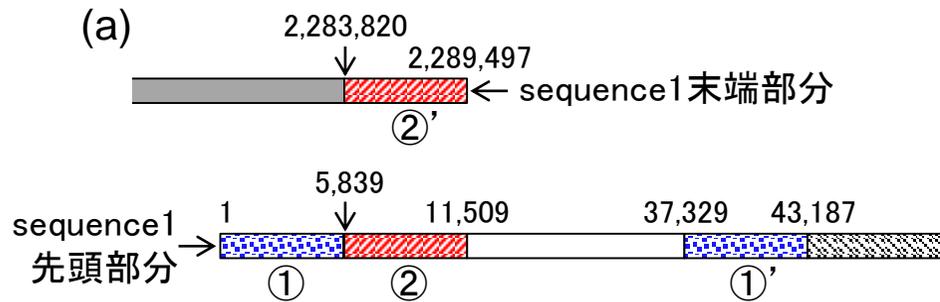
W10-4: ファージの機構

(c) → (b) ファージが染色体に組み込まれる(integration)場合は、灰色矢印の方向に沿って行われる



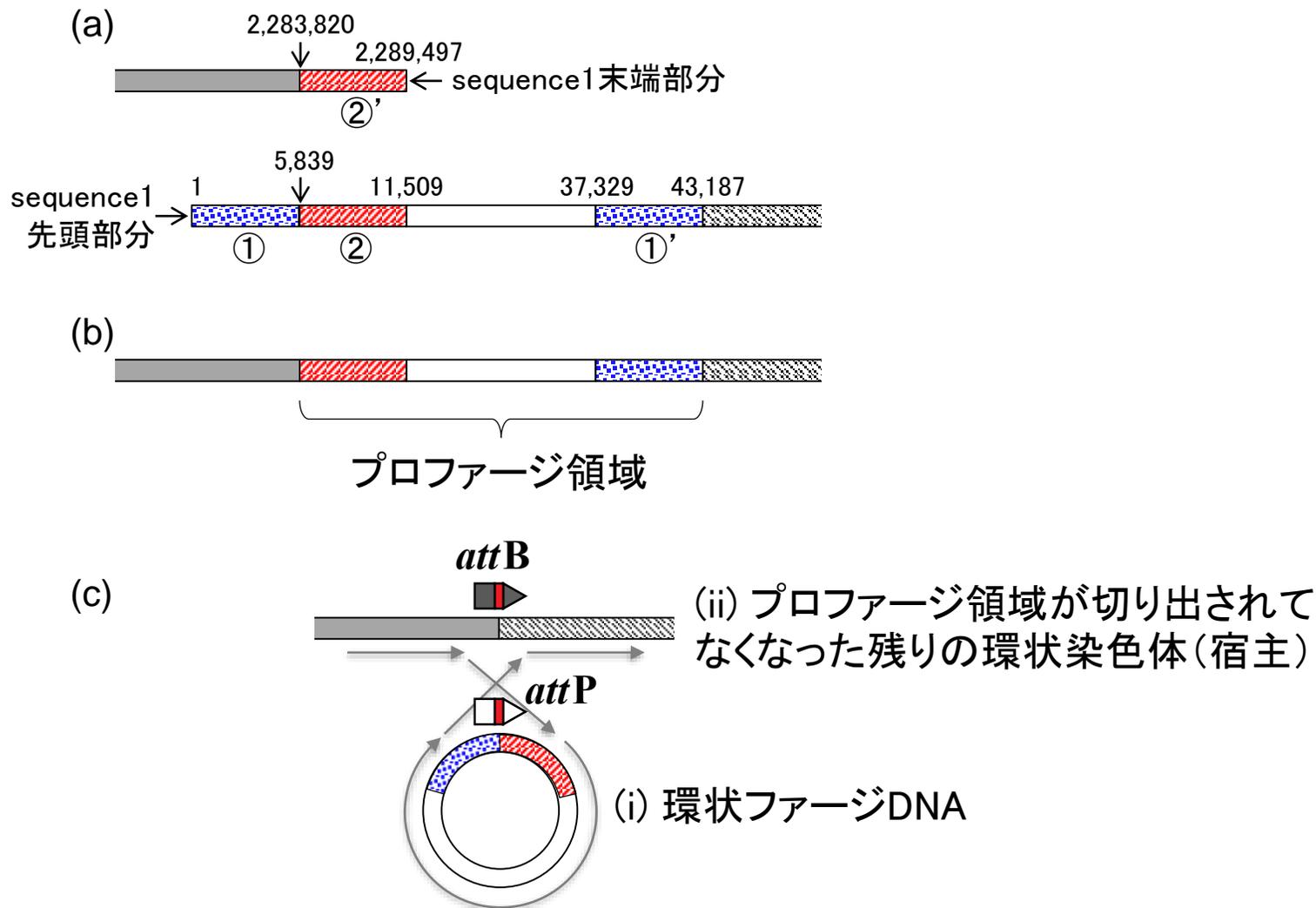
W10-4: ファージの機構

excisionやintegrationが行われるためには、まず染色体上にファージが付着しなければならない。*att*がファージ付着部位(attachment site)であり、**赤四角部分**が付着部位のコア領域



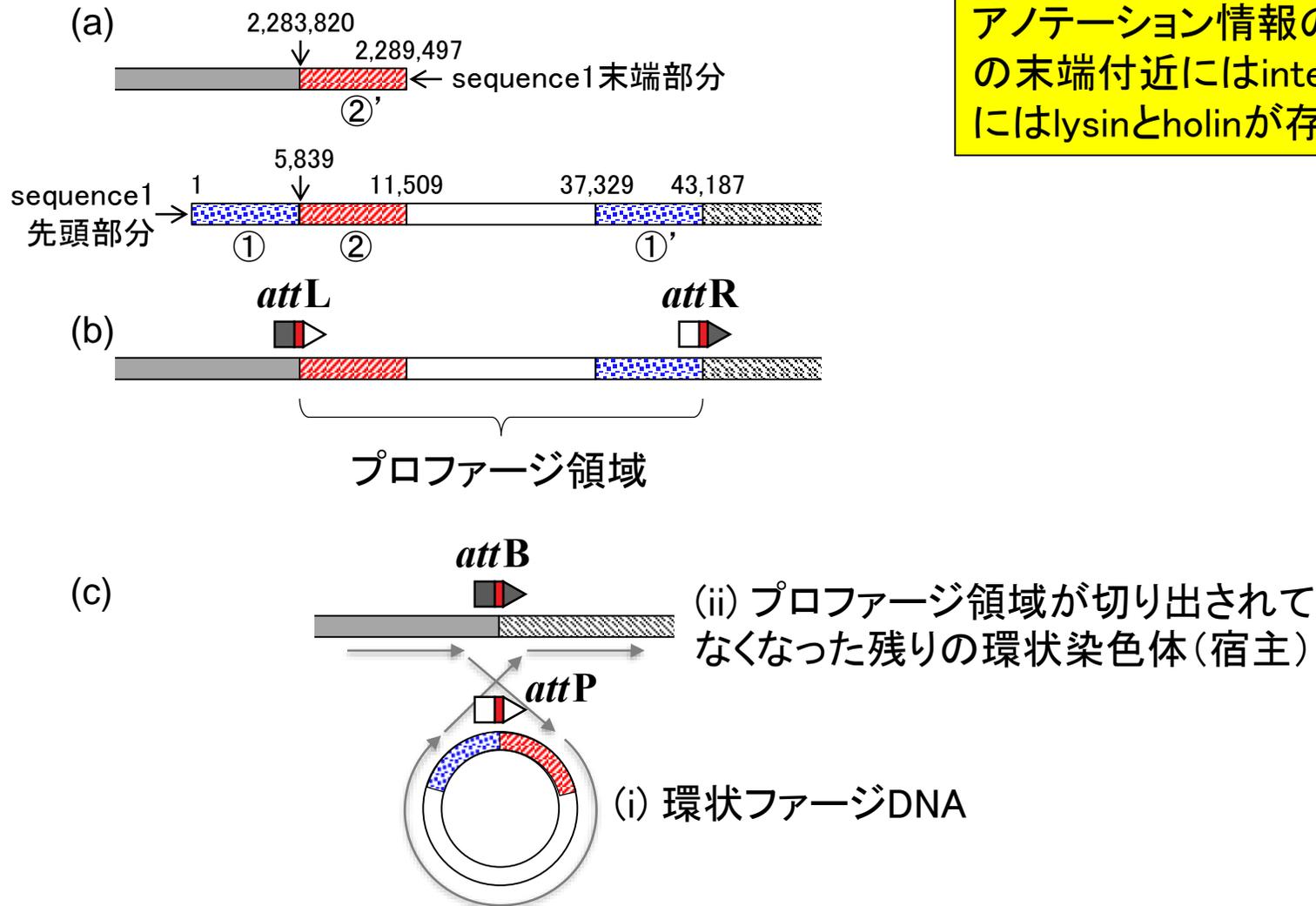
プロファージの特徴1。attBの近傍には、tRNAがコードされていることが多い(宿主側)

W10-4: ファージの機構



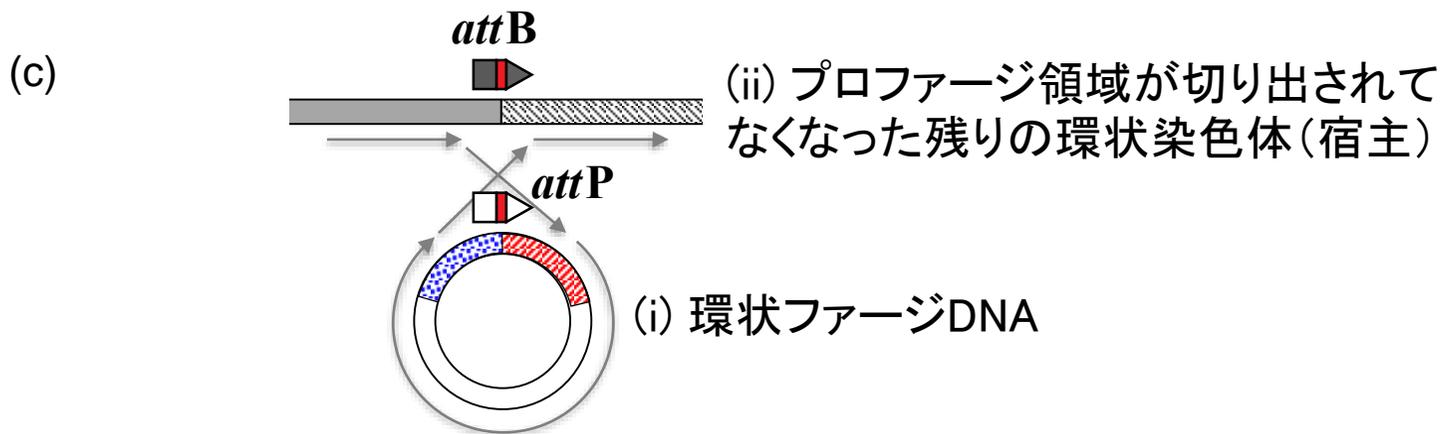
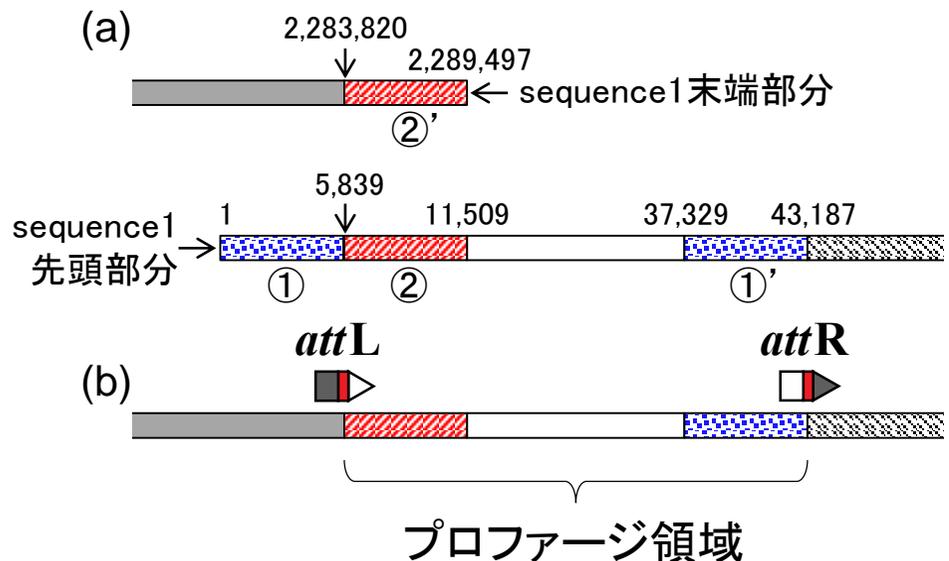
W10-4: ファージの機構

プロファージの特徴2。プロファージ領域の両末端(*attL*と*attR*)には、10-20塩基の相同な配列が存在する。付着部位のコア領域のことを指し、赤四角部分に相当する。アノテーション情報の特徴としては、一方の末端付近には*integrase*、逆の末端付近には*lysine*と*holin*が存在する



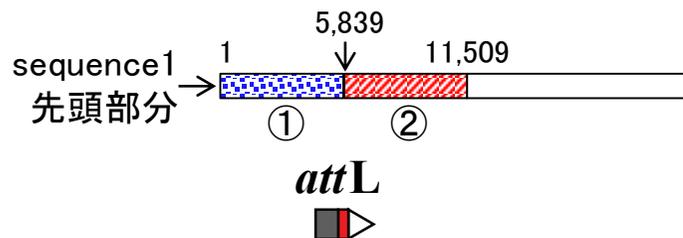
W10-4: ファージの機構

プロファージ領域の右末端(*attR*)は、領域①'の右末端ということ。それは領域①[1, 5860 bp]の右末端と同じ。つまり赤四角部分は、領域①[1, 5860 bp]の右末端と、領域②[5839, 11509 bp]の左末端あたりにあるということです



W10-4: ファージの機構

今は左図あたりの議論をしています。領域①[1, 5860 bp]の右末端と、領域②[5839, 11509 bp]の左末端には**重複領域**[5839, 5860 bp]があることに気づくでしょう。これが**赤四角部分**に相当



W10-4: ファージの機構

HSP4の ALIGNMENT 結果の右端部分を表示。領域①[1, 5860 bp]の右末端に相当。赤枠部分が赤四角部分に相当する

The screenshot shows the BlastViewer interface. On the left, the 'BLAST results' pane lists 'sequence1_blast.xml', 'blastn vs. LH_hgap.fa', and 'sequence1'. The main area shows a 'Summary' tab with a 'Hits' table:

#	Accession	Definition	Quality	# HSPs
1	0	sequence1	☺	1347
2	3	sequence4	☺	2
3	2	sequence3	☺	1
4	1	sequence2	☺	1

Below the table is the 'Alignment: Query (2289497 nuc) vs. 0 (2,289,497 nuc)' section. It includes a 'HSP Map' and a sequence alignment view. The alignment view shows the query sequence (top) and the HSP(s) (middle) aligned to the reference sequence (bottom). The reference sequence is shown in a grid format with positions 5810, 5820, 5830, 5840, 5850, and 5860 marked. A red box highlights the sequence 'A C G C G T T T T T A T T T T T C T A A G T' at the end of the alignment. At the bottom, the 'HSP:' section shows '4/1347' with navigation buttons. A red arrow labeled '1' points to the '4/1347' text, and another red arrow labeled '2' points to the right navigation button.

W10-4: ファージの機構

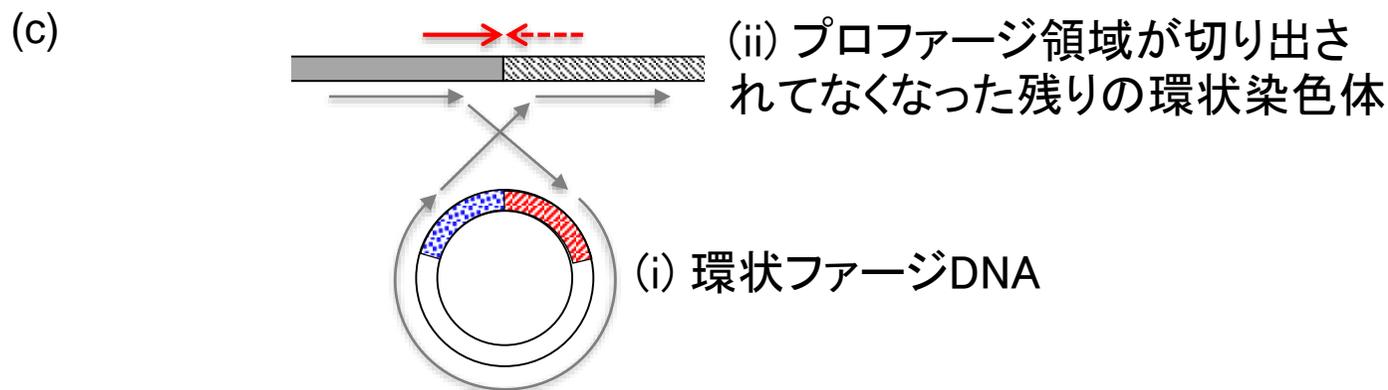
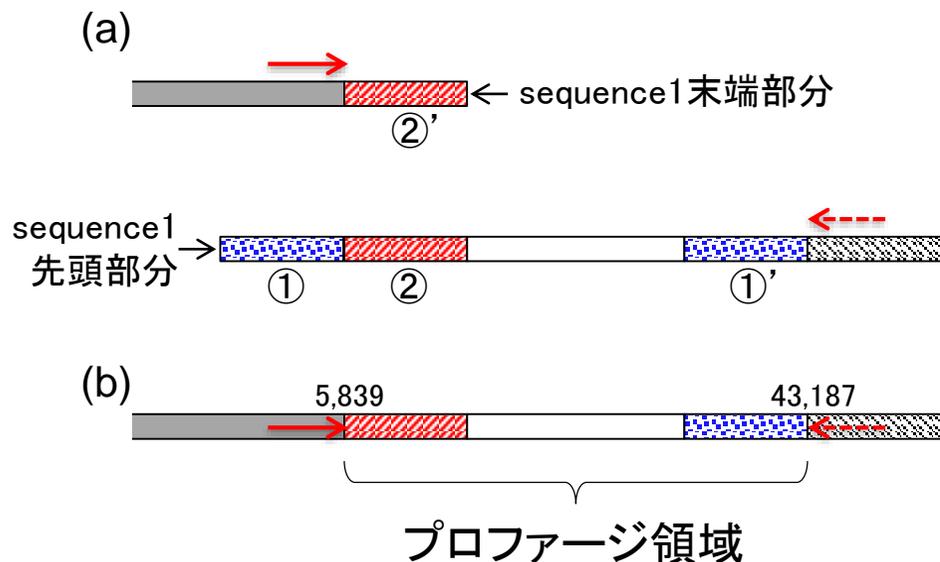
HSP8のALIGNMENT結果の右端部分を表示。領域②[5839, 11509 bp]の左末端に相当。赤枠部分が赤四角部分に相当する

The screenshot shows the BlastViewer interface. On the left, there is a sidebar with 'BLAST results' and a list of files: 'sequence1_blast.xml', 'blastn vs. LH_hgap.fa', and 'sequence1'. The main window is titled 'BlastViewer' and has a menu bar with 'File' and 'Help'. Below the menu bar, there is a 'Summary' tab and a 'Hits' table. The 'Hits' table has the following data:

#	Accession	Definition	Quality	# HSPs
1	0	sequence1	☺	1347
2	3	sequence4	☺	2
3	2	sequence3	☺	1
4	1	sequence2	☺	1

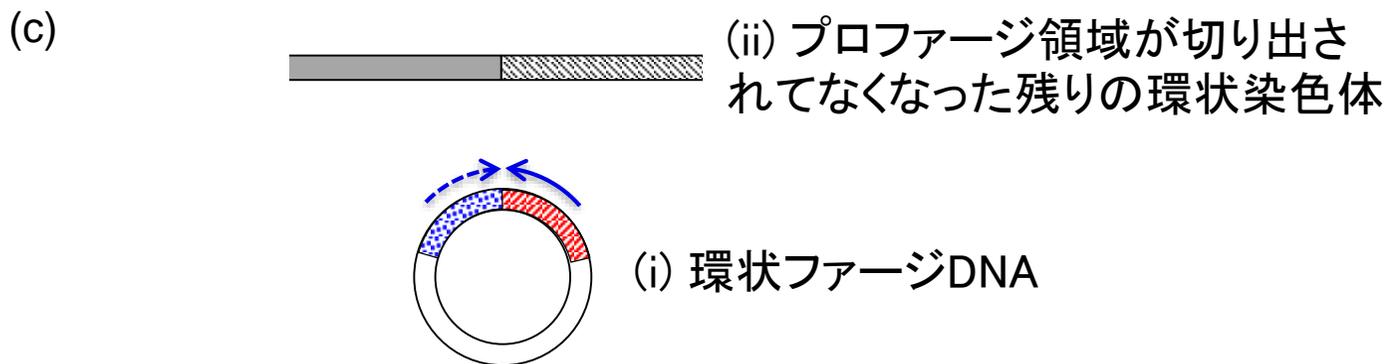
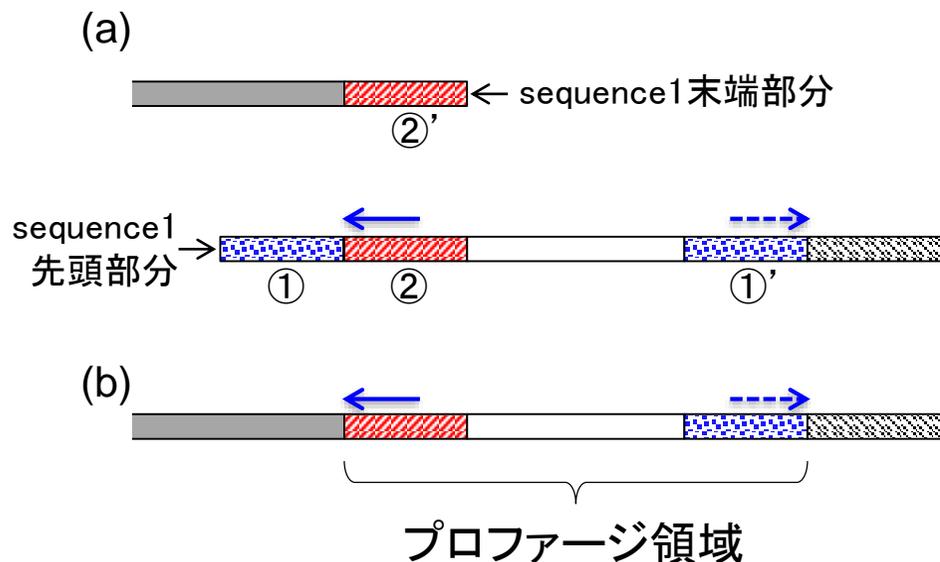
Below the table, there is an 'Alignment: Query (2289497 nuc) vs. 0 (2,289,497 nuc)' section. It has tabs for 'HSP Map', 'Definition', 'Statistics', and 'Alignment'. The 'Alignment' tab is selected, showing a sequence alignment. The query sequence is shown as a black bar at the top. Below it, the HSP(s) sequence is shown as a red bar. The alignment is displayed as a sequence of nucleotides (A, C, G, T) with a red box highlighting a specific region. The sequence is: A C G C G T T T T T A T T T T T C T A A G T G A T T T G G G A A T T A T C T G G A G T C A C A A G G A T. The red box highlights the sequence: A C G C G T T T T T A T T T T T C T A A G T. Below the sequence, there are coordinates: 5840, 5850, 5860, 5870, 5880, 5890. At the bottom of the window, there is a navigation bar with a left arrow, a right arrow, and the text '8/1347'. Red arrows point to the left arrow (labeled 2) and the right arrow (labeled 1). The bottom right corner of the window shows the URL 'www.korilog.com' and the text '85Mo/123Mo'.

W10-5: PCR検証1



(a) 赤い矢印の位置でPCRのプライマーを設計。(b) プロファージ領域の長さは40kbp程度あるため、通常はPCRで増幅されない。もしシーケンス対象となった細胞集団中に“プロファージ領域を含む環状染色体”しか存在しなければPCR増幅されるものはないはず。(c) シーケンス対象となった細胞集団中には、(b)のプロファージ領域を含む環状染色体以外に、(i) プロファージ領域が切り出されてできた環状ファージDNA、および(ii) プロファージ領域が切り出されてなくなった残りの環状染色体が含まれるとすれば、(ii)の存在比に応じてPCR増幅されるはず。→実際に増幅された

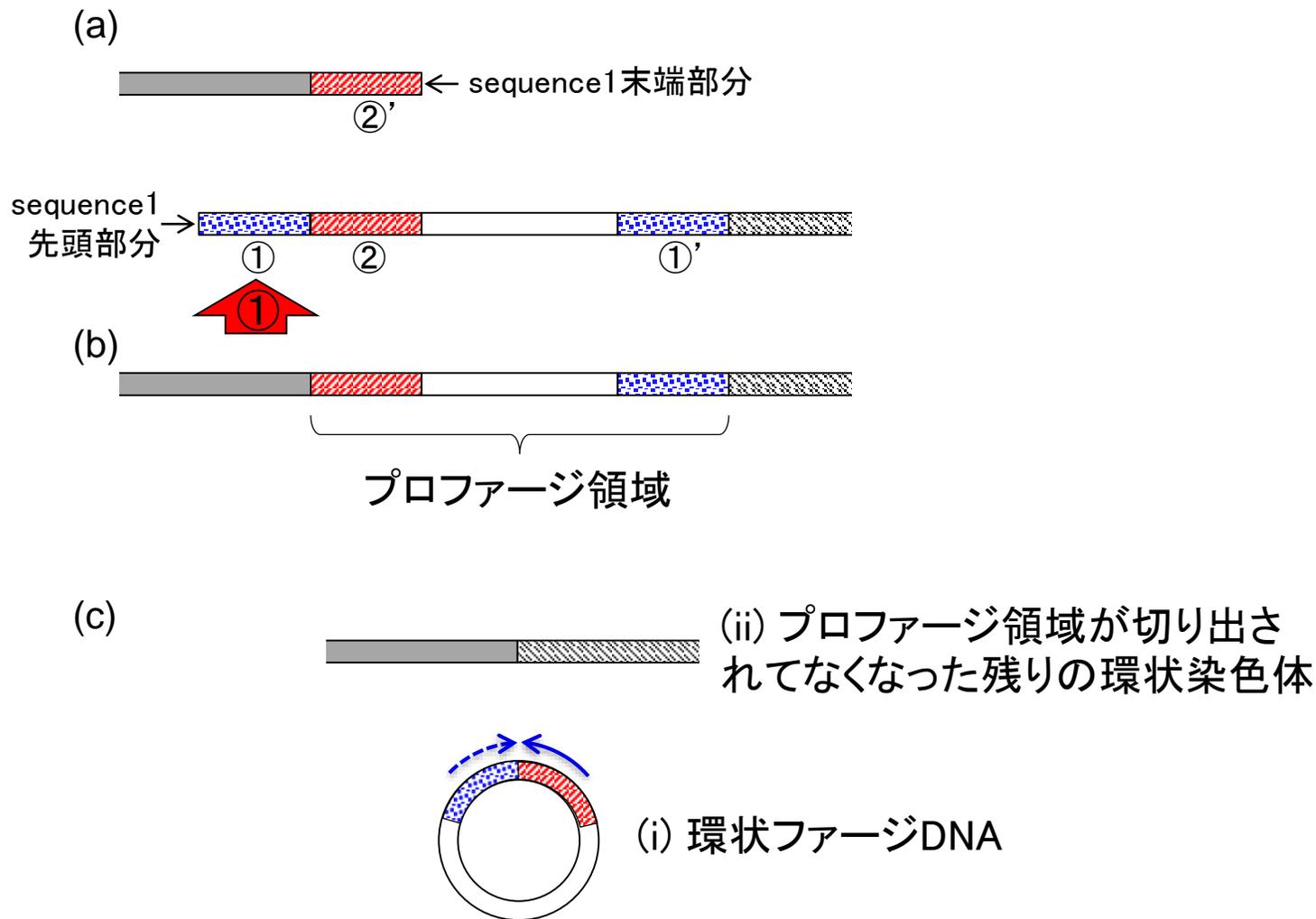
W10-5: PCR検証2



(a) 青い矢印の位置でPCRのプライマーを設計。(b) 逆向きにプライマーを設計しているため、通常はPCRで増幅されない。もしシーケンス対象となった細胞集団中に“プロファージ領域を含む環状染色体”しか存在しなければPCR増幅されるものはないはず。(c) シーケンス対象となった細胞集団中には、(b)のプロファージ領域を含む環状染色体以外に、(i) プロファージ領域が切り出されてできた環状ファージDNA、および(ii) プロファージ領域が切り出されてなくなった残りの環状染色体が含まれるとすれば、(i)の存在比に応じてPCR増幅されるはず。→実際に増幅された

W10-6: ①の領域

①の領域は、(i)のように環状化したファージDNAがシーケンスされた結果として生じたものであり、実際の染色体上には存在しないと結論づけた



sequence2のBLASTを実行。①makeblastdb
、②blastn。③lsで確認して、④lessで開く

W11-1: BLAST

```
iu@bielinux[result] makeblastdb -in sequence2.fa -dbtype nucl -hash_index
Building a new DB, current time: 09/10/2016 10:16:01
New DB name: sequence2.fa
New DB title: sequence2.fa
Sequence type: Nucleotide
Keep Linkouts: T
Keep MBits: T
Maximum file size: 10000000000B
Adding sequences from FASTA; added 1 sequences in 0.119973 seconds.
iu@bielinux[result] blastn -db sequence2.fa -query sequence2.fa -out sequence2_blast.txt
iu@bielinux[result] pwd [10:16午前]
/home/iu/Desktop/result
iu@bielinux[result] ls sequence2* [10:16午前]
sequence2_blast.txt sequence2.fa.nhr sequence2.fa.nsi
sequence2.fa sequence2.fa.nin sequence2.fa.nsq
sequence2.fa.nhd sequence2.fa.nog
sequence2.fa.nhi sequence2.fa.nsd
iu@bielinux[result] less sequence2_blast.txt [10:16午前]
```

W11-1: BLAST

①Scoreで文字列検索して、2つめのHSPに辿り着いたところ(第7回W15-7)。②長さが5264塩基なので、概ねその半分程度の長さのところのアラインメントを眺め、重複除去する場所を決定する。③1塩基目から始まっているので、下側(Sbjct側)が2600番目あたりを探すべく、ページ下部に移動(第7回W16-1)

```
File Edit View Search Terminal Help
Score = 9651 bits (5226), Expect = 0
Identities = 5253/5264 (99%), Gaps =
Strand=Plus/Plus

Query 81634 AGCAGTTGTTTGTCTGTAAGGCCAATCTTTGTTGCTGATCTAGTAACTTATGTA ACTCT 81693
      |||
Sbjct 1 AGCAGTTGTTTGTCTGTAAGGCCAATCTTTGTTGCTGATCTAGTAACTTATGTA ACTCT 60

Query 81694 TTCTTTTCAGATTGAATTTCTGATACTAAATTTCGAAGGAATTTCACTTCATCTTG TCCA 81753
      |||
Sbjct 61 TTCTTTTCAGATTGAATTTCTGATACTAAATTTCGAAGGAATTTCACTTCATCTTG TCCA 120

Query 81754 TCAACCGCATCTTTTTTGTGTCGTCAACATCATCTTTTTGTTGACGGTATCTTG TTGAC 81813
      |||
Sbjct 121 TCAACCGCATCTTTTTTGTGTCGTCAACATCATCTTTTTGTTGACGGTATCTTG TTGAC 180

Query 81814 GAATTGTTTGAGAACATTGCTTTAATAGCTTTTTGCCATCAACATCAACGTAAACAGTA 81873
      |||
Sbjct 181 GAATTGTTTGAGAACATTGCTTTAATAGCTTTTTGCCATCAACATCAACGTAAACAGTA 240

Query 81874 TTACCTTTTGTTGATGTAAATTGACGTAAATCGATTGACGCATCTC-TTTTTATCTTTTG 81932
      :

```

W11-1: BLAST

①の赤枠分くらいを眺め、どこにもミスマッチやGapがないことを確認。②のところでトリムすることにする。左端にする理由は、上が84271番目、下が2641番目の塩基だとすぐにわかるから

```
File Edit View Search Terminal Help
Query 84151 GCCCTTGCCTTAATCTTCGATCAGCGACAATTTTTTAAAACCTGAGTTAATAAAGCTGGT 84210
      |||
Sbjct 2521 GCCCTTGCCTTAATCTTCGATCAGCGACAATTTTTTAAAACCTGAGTTAATAAAGCTGGT 2580

Query 84211 TTATCAGGGGTGGGGGCTTTCGACCAATTCATCCAATAGTAGTAAACAGTGGACCACTTT 84270
      |||
Sbjct 2581 TTATCAGGGGTGGGGGCTTTCGACCAATTCATCCAATAGTAGTAAACAGTGGACCACTTT 2640

Query 84271 GGAAAATCAGCCGGCAAATCACGCCAAGTACAACCATTTTTAAGGACATAGAGCATGGCA 84330
      |||
Sbjct 2641 GGAAAATCAGCCGGCAAATCACGCCAAGTACAACCATTTTTAAGGACATAGAGCATGGCA 2700

Query 84331 CAGAAGATATCATAAGGATCAATGTGTTTTGGATGAGTTCGTTTTCTTGAAGCTTCCAGA 84390
      |||
Sbjct 2701 CAGAAGATATCATAAGGATCAATGTGTTTTGGATGAGTTCGTTTTCTTGAAGCTTCCAGA 2760

Query 84391 TCCTGGCGGATCAAATCGAATTGTTGCCGAGAAATATCGCTTTGGTAATGATGACTAAAA 84450
      |||
Sbjct 2761 TCCTGGCGGATCAAATCGAATTGTTGCCGAGAAATATCGCTTTGGTAATGATGACTAAAA 2820
:
```



W11-1: BLAST

①2641番目の塩基をトリム後の1塩基目にする場合は、[2641, 84270 bp]を残せばよい。こうすることで、トリム後の塩基配列の最初のほうは①の赤枠のようになり、最後のほうは②のようになるはずである。③qで終了

```
File Edit View Search Terminal Help
Query 84151 GCCCTTGCCTTAATCTTCGATCAGCGACAATTTTTTAAAACCTGAGTTAATAAAGCTGGT 84210
      |||
Sbjct 2521  GCCCTTGCCTTAATCTTCGATCAGCGACAATTTTTTAAAACCTGAGTTAATAAAGCTGGT 2580

Query 84211 TTATCAGGGGTGGGGGCTTTCGACCAATTCATCCAATAGTAGTAAACAGTGGACCACTTT 4270
      |||
Sbjct 2581  TTATCAGGGGTGGGGGCTTTCGACCAATTCATCCAATAGTAGTAAACAGTGGACCACTTT 2640

Query 84271 GGAAAATCAGCCGGCAAATCACGCCAAGTACAACCATTTTTAAGGACATAGAGCATGGCA 84330
      |||
Sbjct 2641  GGAAAATCAGCCGGCAAATCACGCCAAGTACAACCATTTTTAAGGACATAGAGCATGGCA 2700

Query 84331 CAGAAGATATCATAAGGATCAATGTGTTTTGGATGAGTTCGTTTTCTTGAAGCTTCCAGA 84390
      |||
Sbjct 2701  CAGAAGATATCATAAGGATCAATGTGTTTTGGATGAGTTCGTTTTCTTGAAGCTTCCAGA 2760

Query 84391 TCCTGGCGGATCAAATCGAATTGTTGCCGAGAAATATCGCTTTGGTAATGATGACTAAAA 84450
      |||
Sbjct 2761  TCCTGGCGGATCAAATCGAATTGTTGCCGAGAAATATCGCTTTGGTAATGATGACTAAAA 2820

:
```



W11-2:ドラフト配列作成

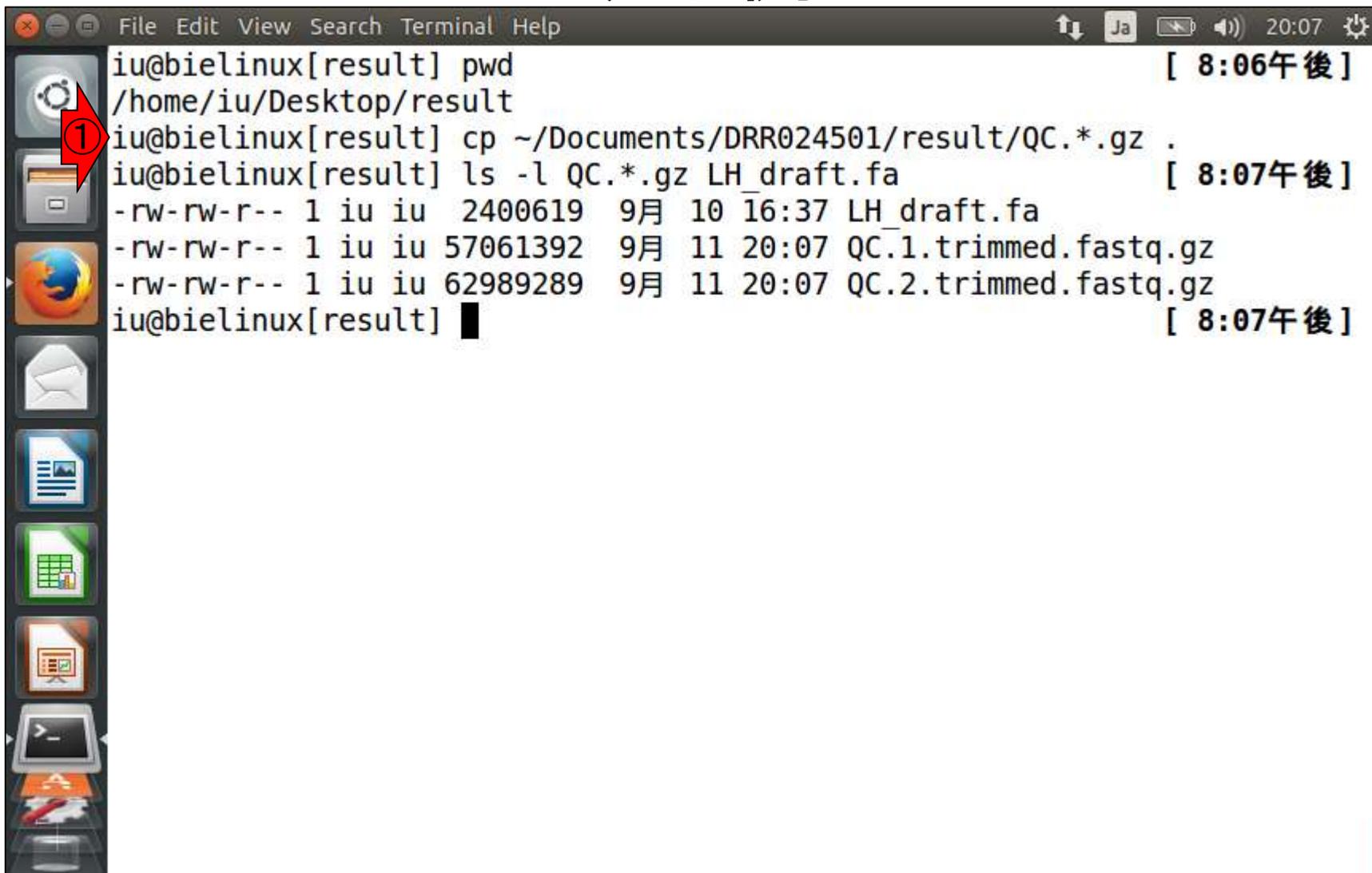
①アセンブリ結果ファイル(LH_hgap.fa)を入力として、重複除去を行う一連のコマンド。出力ファイルはLH_draft.fa。②ファイルサイズの減少度合い的に妥当。ここでは示さないがlessで配列の最初と最後を思い通りに抽出できているか確認しておこう

```
iu@bielinux[result] pwd
/home/iu/Desktop/result
iu@bielinux[result] ls -l LH*.fa
-rw-rw-r-- 1 iu iu 2433662  8月 29 13:28 LH_hgap.fa
iu@bielinux[result] echo ">chromosome" > LH_draft.fa [ 4:37午後]
iu@bielinux[result] head -2 LH_hgap.fa | tail -1 | cut -c 5839-2283819 >> LH_draft.fa [ 4:37午後]
iu@bielinux[result] echo ">plasmid1" >> LH_draft.fa [ 4:37午後]
iu@bielinux[result] head -4 LH_hgap.fa | tail -1 | cut -c 2641-84270 >> LH_draft.fa [ 4:37午後]
iu@bielinux[result] echo ">plasmid2" >> LH_draft.fa [ 4:37午後]
iu@bielinux[result] head -6 LH_hgap.fa | tail -1 | cut -c 2450-43422 >> LH_draft.fa [ 4:37午後]
iu@bielinux[result] ls -l LH*.fa [ 4:37午後]
-rw-rw-r-- 1 iu iu 2400619  9月 10 16:37 LH_draft.fa [ 4:37午後]
-rw-rw-r-- 1 iu iu 2433662  8月 29 13:28 LH_hgap.fa [ 4:37午後]
iu@bielinux[result]
```



W12-1: マップする側

①~/Documents/DRR024501/result/上にあるQC.*.gzをカレントディレクトリにコピー



The image shows a terminal window with a dark background and a light-colored text. The window title bar includes 'File Edit View Search Terminal Help' and system icons for volume, network, and battery. The terminal output shows the following commands and results:

```
iu@bielinux[result] pwd [ 8:06午後]
/home/iu/Desktop/result
iu@bielinux[result] cp ~/Documents/DRR024501/result/QC.*.gz .
iu@bielinux[result] ls -l QC.*.gz LH_draft.fa [ 8:07午後]
-rw-rw-r-- 1 iu iu 2400619 9月 10 16:37 LH_draft.fa
-rw-rw-r-- 1 iu iu 57061392 9月 11 20:07 QC.1.trimmed.fastq.gz
-rw-rw-r-- 1 iu iu 62989289 9月 11 20:07 QC.2.trimmed.fastq.gz
iu@bielinux[result] █ [ 8:07午後]
```

A red arrow with the number '1' points to the second line of the terminal output, which is the 'cp' command.

W12-2: bwa

①Bio-Linuxにはbwaがプリインストールされているので、bwaがすぐに利用可能な状態。②バージョンは0.7.12-r1039。BWAの使用法については、平成28年度NGSハンズオン講習会(2016.07.26)にもあり

```
File Edit View Search Terminal Help
iu@bielinux[result] bwa [ 8:19午後 ]
Program: bwa (alignment via Burrows-Wheeler transformation)
Version: 0.7.12-r1039
Contact: Heng Li <lh3@sanger.ac.uk>

Usage: bwa <command> [options]

Command: index          index sequences in the FASTA format
         mem            BWA-MEM algorithm
         fastmap        identify super-maximal exact matches
         pmerge         merge overlapping paired ends (EXPERIMENTAL)
         aln            gapped/ungapped alignment
         samse          generate alignment (single ended)
         sampe          generate alignment (paired ended)
         bwasw          BWA-SW for long queries

         shm           manage indices in shared memory
         fa2pac         convert FASTA to PAC format
         pac2bwt        generate BWT from PAC
         pac2bwtgen     alternative algorithm for generating BWT
         bwtupdate      update .bwt to the new format
```

W12-2:bwa

```
File Edit View Search Terminal Help 21:40
mem      BWA-MEM algorithm
fastmap  identify super-maximal exact matches
pmerge  merge overlapping paired ends (EXPERIMENTAL)
aln      gapped/ungapped alignment
samse    generate alignment (single ended)
sampe    generate alignment (paired ended)
bwasw    BWA-SW for long queries

shm      manage indices in shared memory
fa2pac   convert FASTA to PAC format
pac2bwt  generate BWT from PAC
pac2bwtgen alternative algorithm for generating BWT
bwtupdate update .bwt to the new format
bwt2sa   generate SA from BWT and Occ

Note: To use BWA, you need to first index the genome with `bwa index`.
There are three alignment algorithms in BWA: `mem`, `bwasw`, and
`aln/samse/sampe`. If you are not sure which to use, try `bwa mem`
first. Please `man ./bwa.1` for the manual.

iu@bielinux[result] [ 8:19午後 ]
iu@bielinux[result] bwa index LH_draft.fa [ 9:39午後 ]
```



W12-2:bwa

①bwa index実行結果。②5つのインデックスファイルが作成されたようだ。③BWAには3つのアルゴリズム(mem, bwasw, aln/samse/sampe)があり、よくわからなければbwa memを使えと書いている

```
File Edit View Search Terminal Help
bwt2sa          generate SA from BWT and OCC

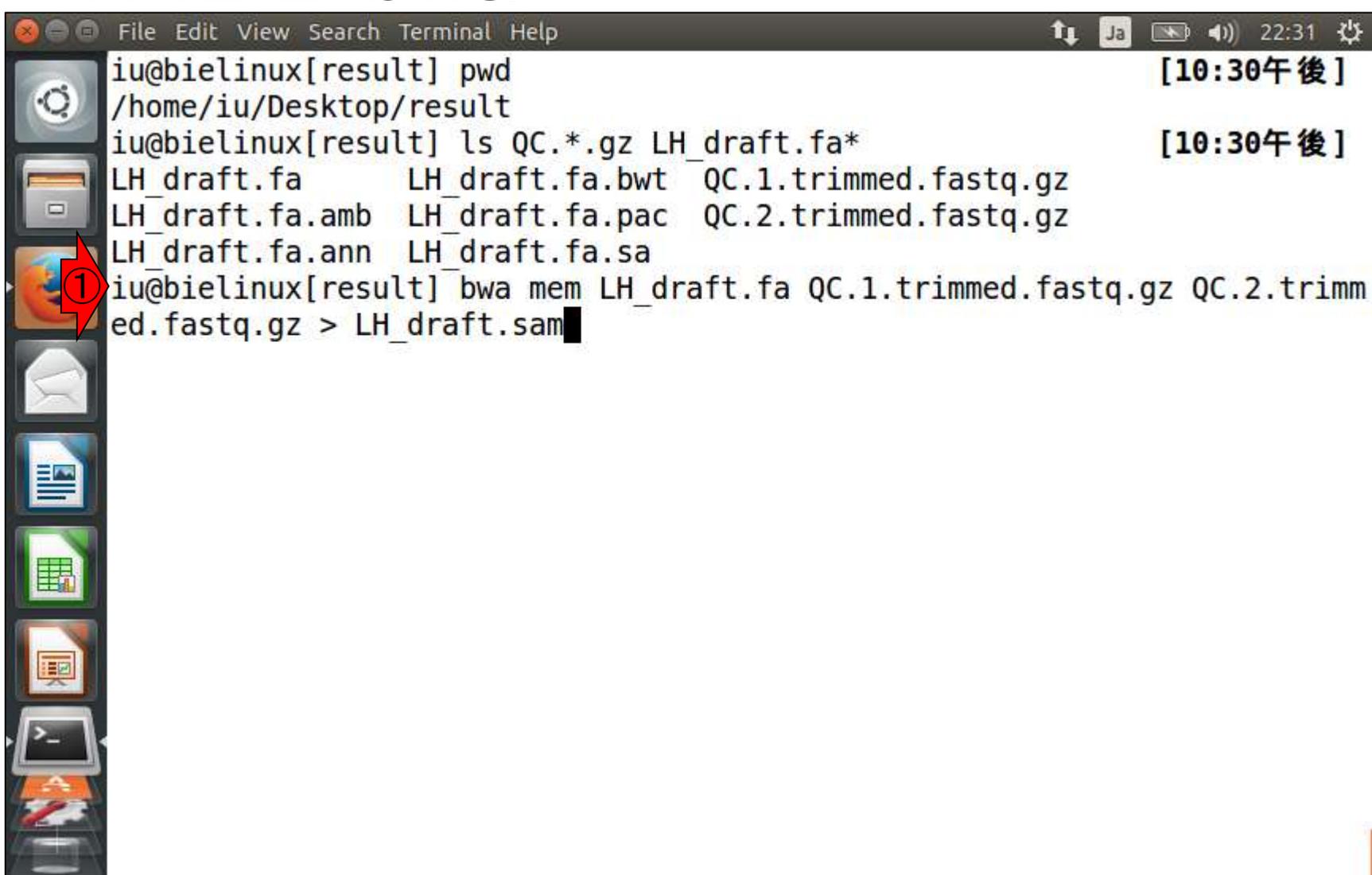
Note: To use BWA, you need to first index the genome with `bwa index`.
There are three alignment algorithms in BWA: `mem`, `bwasw`, and
`aln/samse/sampe`. If you are not sure which to use, try `bwa mem`
first. Please `man ./bwa.1` for the manual.

iu@bielinux[result] [ 8:19午後]
iu@bielinux[result] bwa index LH_draft.fa [ 9:39午後]
[bwa_index] Pack FASTA... 0.06 sec
[bwa_index] Construct BWT for the packed sequence...
[bwa_index] 0.78 seconds elapse.
[bwa_index] Update BWT... 0.04 sec
[bwa_index] Pack forward-only FASTA... 0.04 sec
[bwa_index] Construct SA from BWT and Occ... 0.33 sec
[main] Version: 0.7.12-r1039
[main] CMD: bwa index LH_draft.fa
[main] Real time: 2.842 sec; CPU: 1.242 sec
iu@bielinux[result] ls LH_draft.fa* [ 9:42午後]
LH_draft.fa      LH_draft.fa.ann LH_draft.fa.pac
LH_draft.fa.amb LH_draft.fa.bwt LH_draft.fa.sa
iu@bielinux[result] [ 9:43午後]
```



W12-2:bwa

①bwa memの実行。SAM形式のマッピング結果ファイル(LH_draft.sam)を得るところまで。約2分



A terminal window titled "Terminal" with a menu bar (File, Edit, View, Search, Terminal, Help) and a system tray (Ja, 22:31). The terminal shows the following commands and output:

```
iu@bielinux[result] pwd [10:30午後]
/home/iu/Desktop/result
iu@bielinux[result] ls QC.*.gz LH_draft.fa* [10:30午後]
LH_draft.fa      LH_draft.fa.bwt  QC.1.trimmed.fastq.gz
LH_draft.fa.amb  LH_draft.fa.pac  QC.2.trimmed.fastq.gz
LH_draft.fa.ann  LH_draft.fa.sa
iu@bielinux[result] bwa mem LH_draft.fa QC.1.trimmed.fastq.gz QC.2.trimmed.fastq.gz > LH_draft.sam
```

A red arrow with the number "1" points to the `bwa mem` command line.

無事計算終了。①実際にかかった時間は約82秒。②lsで確認。③ファイルサイズは約335MB

W12-2:bwa

```
File Edit View Search Terminal Help 22:35
(238, 488)
[M::mem_pestat] mean and std.dev: (365.05, 38.12)
[M::mem_pestat] low and high boundaries for proper pairs: (188, 538)
[M::mem_pestat] skip orientation RF as there are not enough pairs
[M::mem_pestat] skip orientation RR as there are not enough pairs
[M::mem_process_seqs] Processed 22156 reads in 2.492 CPU sec, 2.529 real
sec
[main] Version: 0.7.12-r1039
[main] CMD: bwa mem LH_draft.fa QC.1.trimmed.fastq.gz QC.2.trimmed.fastq
.gz
[main] Real time: 81.798 sec; CPU: 76.086 sec
iu@bielinux[result] ls -l LH_draft.*           [10:34午後]
-rw-rw-r-- 1 iu iu 2400619  9月 10 16:37 LH_draft.fa
-rw-rw-r-- 1 iu iu      12  9月 11 21:42 LH_draft.fa.amb
-rw-rw-r-- 1 iu iu    113  9月 11 21:42 LH_draft.fa.ann
-rw-rw-r-- 1 iu iu 2400684  9月 11 21:42 LH_draft.fa.bwt
-rw-rw-r-- 1 iu iu  600148  9月 11 21:42 LH_draft.fa.pac
-rw-rw-r-- 1 iu iu 1200344  9月 11 21:42 LH_draft.fa.sa
-rw-rw-r-- 1 iu iu 350505158 9月 11 22:34 LH_draft.sam
iu@bielinux[result] ls -lh LH_draft.sam       [10:35午後]
-rw-rw-r-- 1 iu iu 335M  9月 11 22:34 LH_draft.sam
iu@bielinux[result] █                         [10:35午後]
```

W13-1: Cyberduck

Cyberduckのインストール。①Windows用をインストールしたい場合はこちら。②実行



[Changelog](#) | [Blog](#) | [Development](#) | [CLI](#) | [Mountain Duck](#) | [Help](#) | [Donate](#)

 **chrome** Switch to Google Chrome for a better browsing experience. [Download Now](#)

Cyberduck is a libre FTP, SFTP, WebDAV, S3, Backblaze B2, Azure & OpenStack Swift browser for Mac and Windows.

[Download Cyberduck for Windows.](#) ①

[Download Cyberduck for Mac.](#) X

[Mac App Store](#)

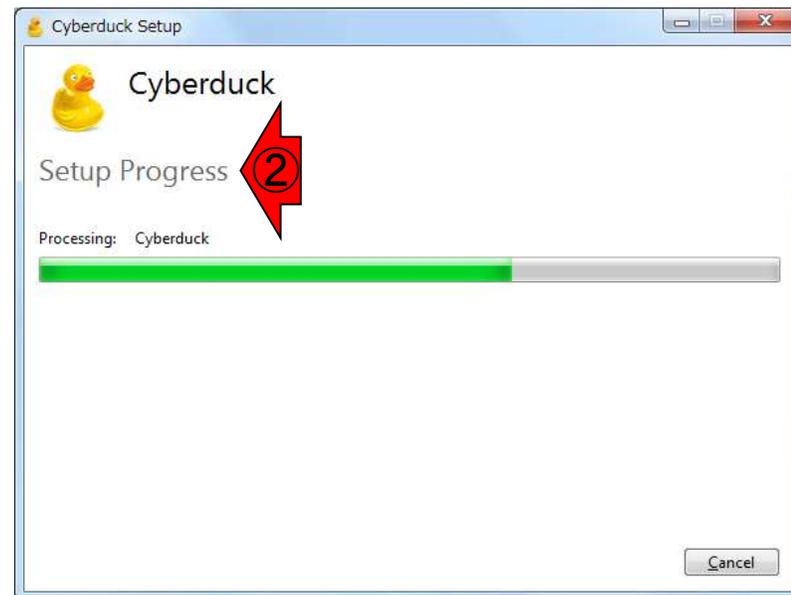
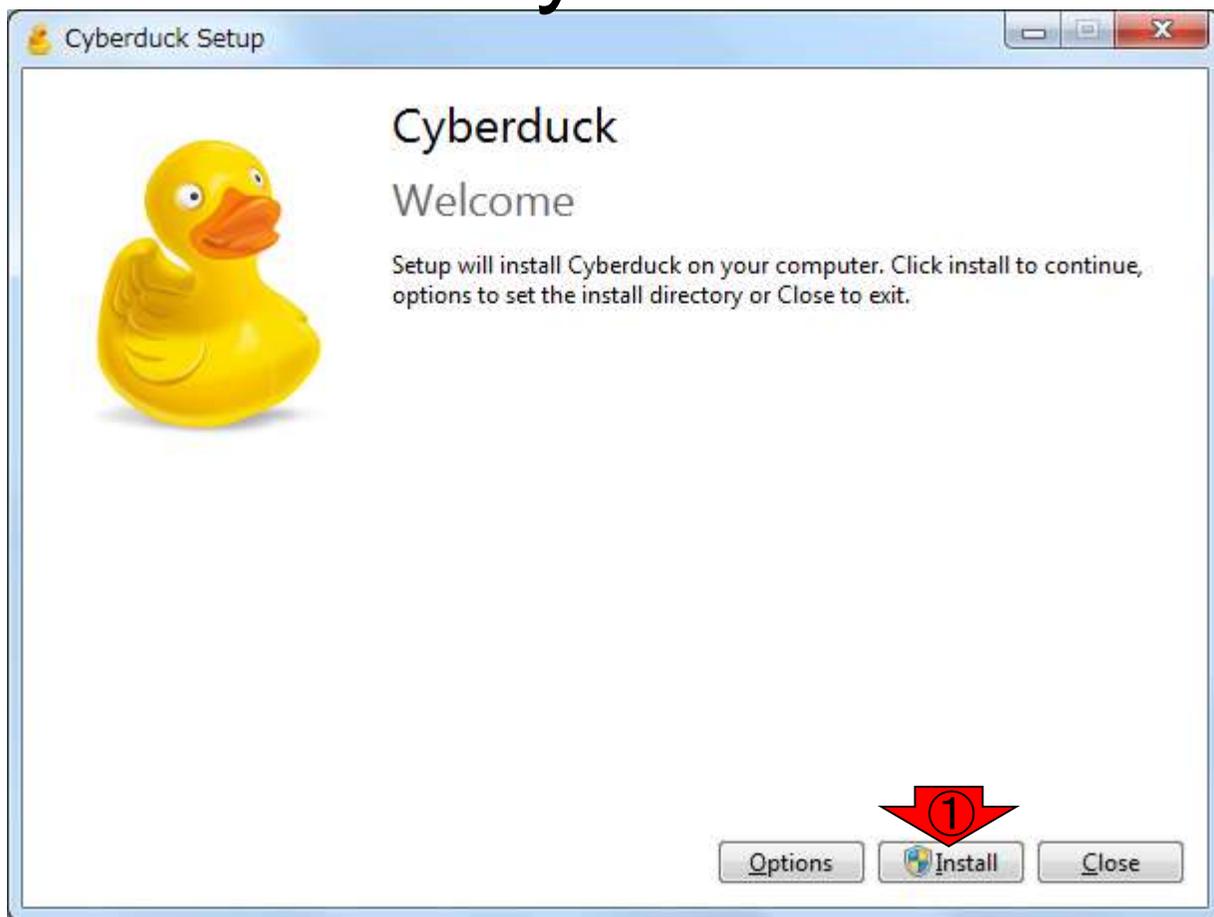
AWS - iterate GmbH
Quick Connect
Action Get Info Edit Transfers Disconnect

- Backblaze B2
api.backblaze.com
https://ad96b6ea5ed0@api.backblaze.com
- AWS - iterate GmbH
s3.amazonaws.com
https://AKIAJ5AERZY4GU3I3A@s3.amazonaws.com
- Oracle Storage Cloud Service
cyduck.storage.oraclecloud.com
https://Storage-cyduck:dkocher@cyduck.storage.oraclecloud.com

...23c7ec3cf4539564-60c534a1284a12ce74ef84032e9b4e46.ssl.cf1.rackcdn.com から Cyberduck-... (45.8 MB) を実行または保存しますか? X

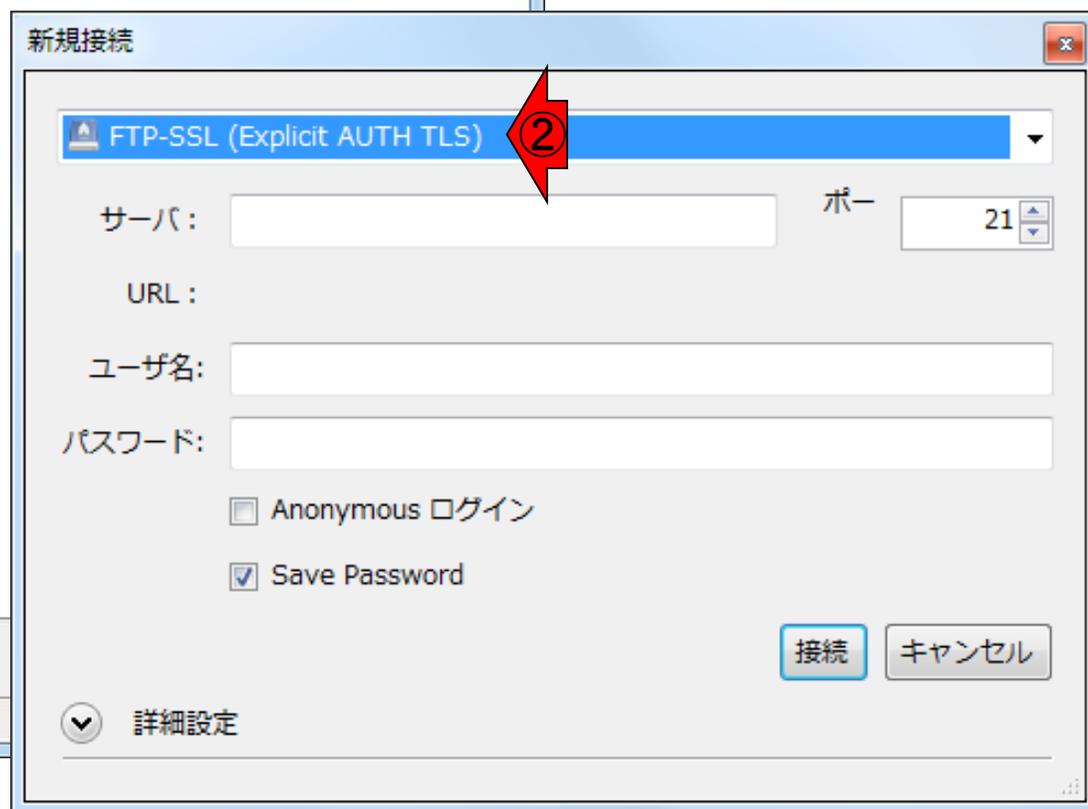
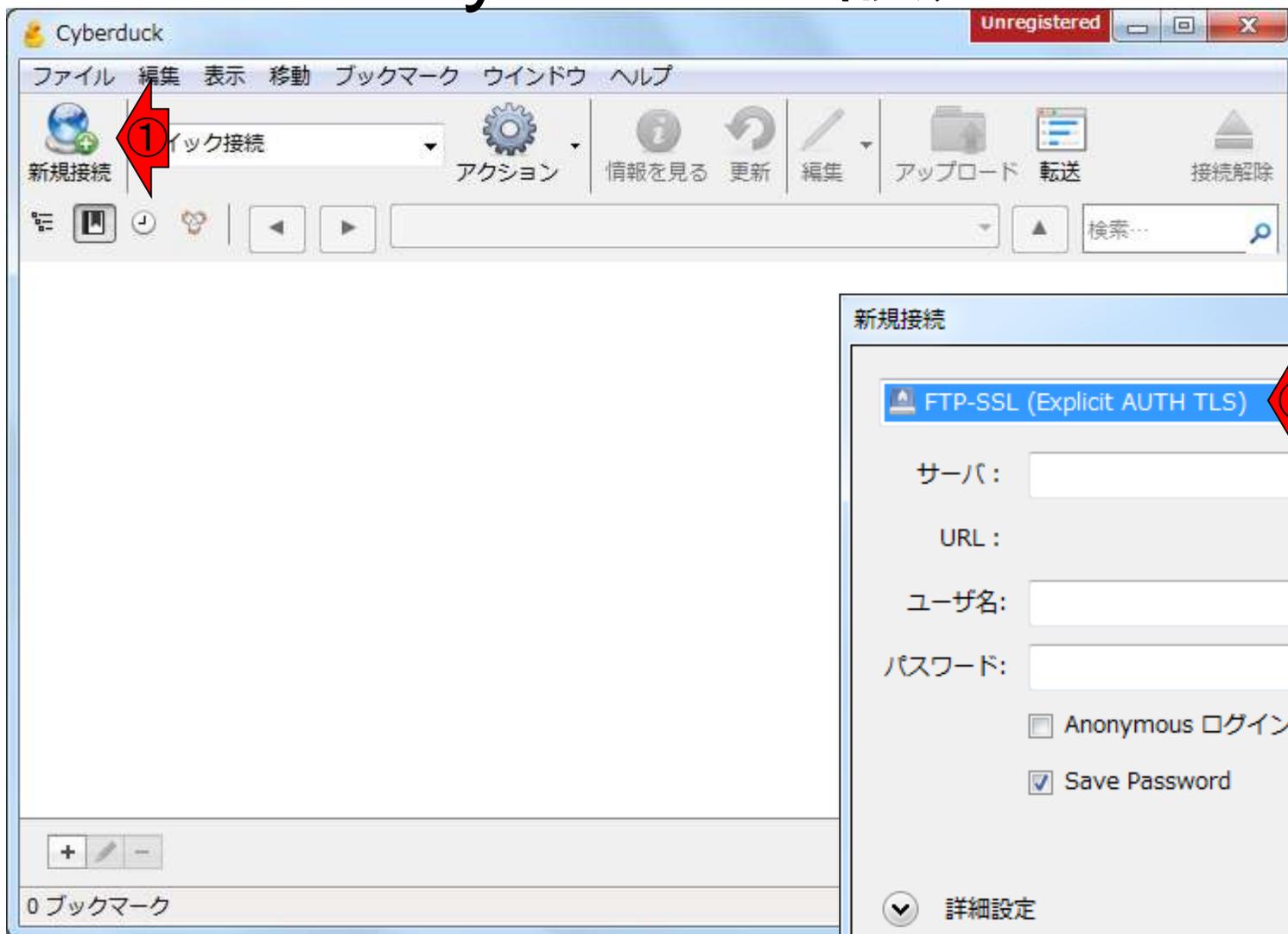
W13-1 : Cyberduck

①Install。②途中経過。③無事終了すると右下のような感じになる。④Close

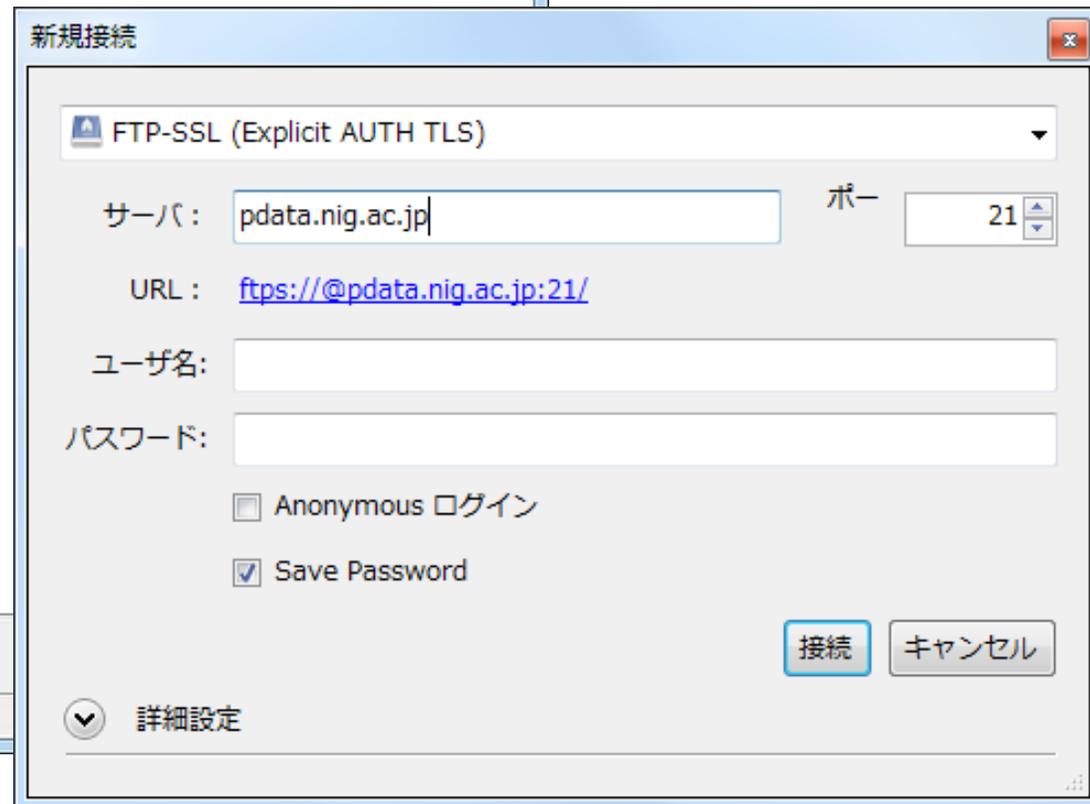
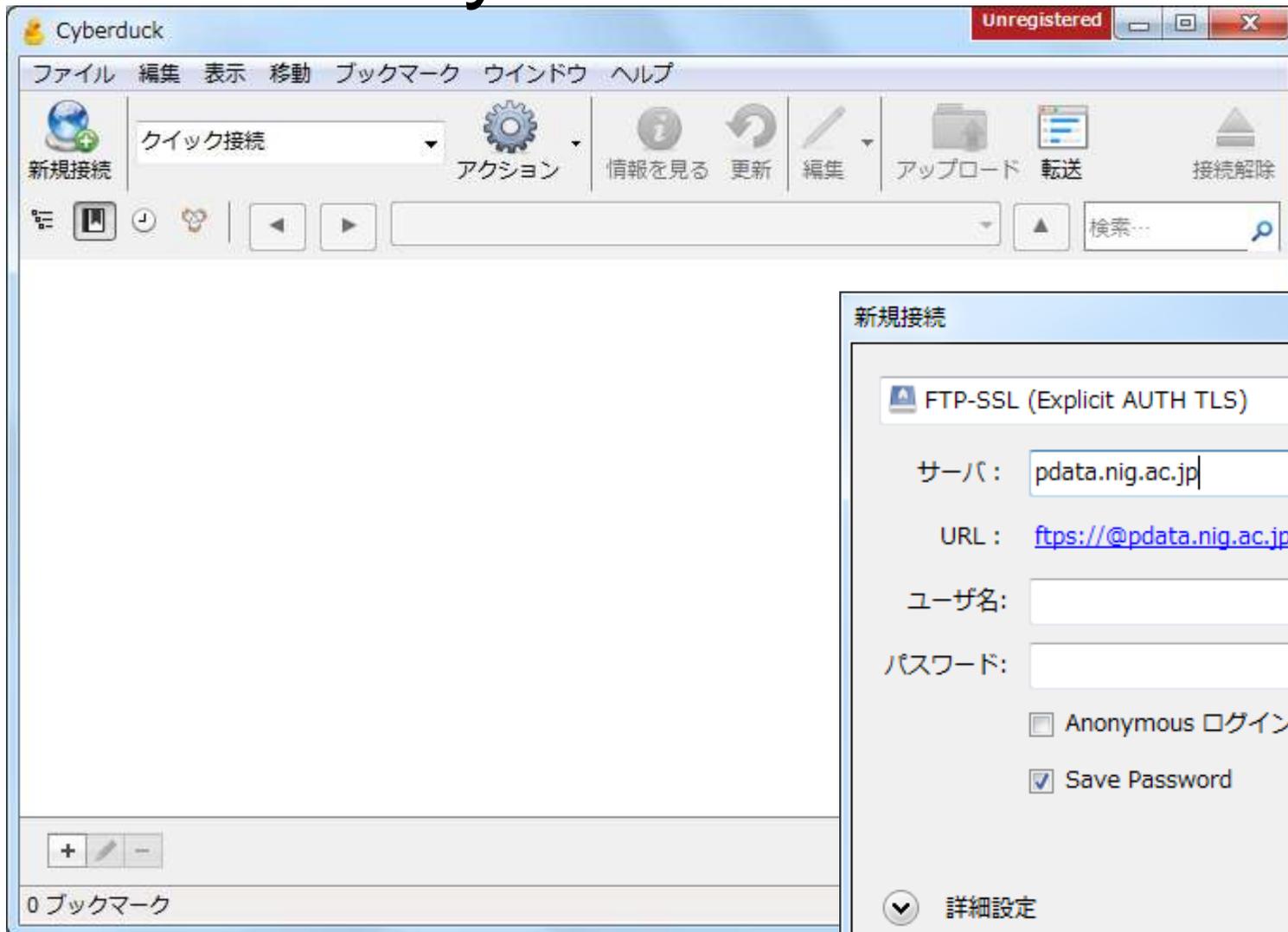


W13-2: Cyberduck設定

Cyberduckを起動し、①新規接続。②FTP-SSLを選択

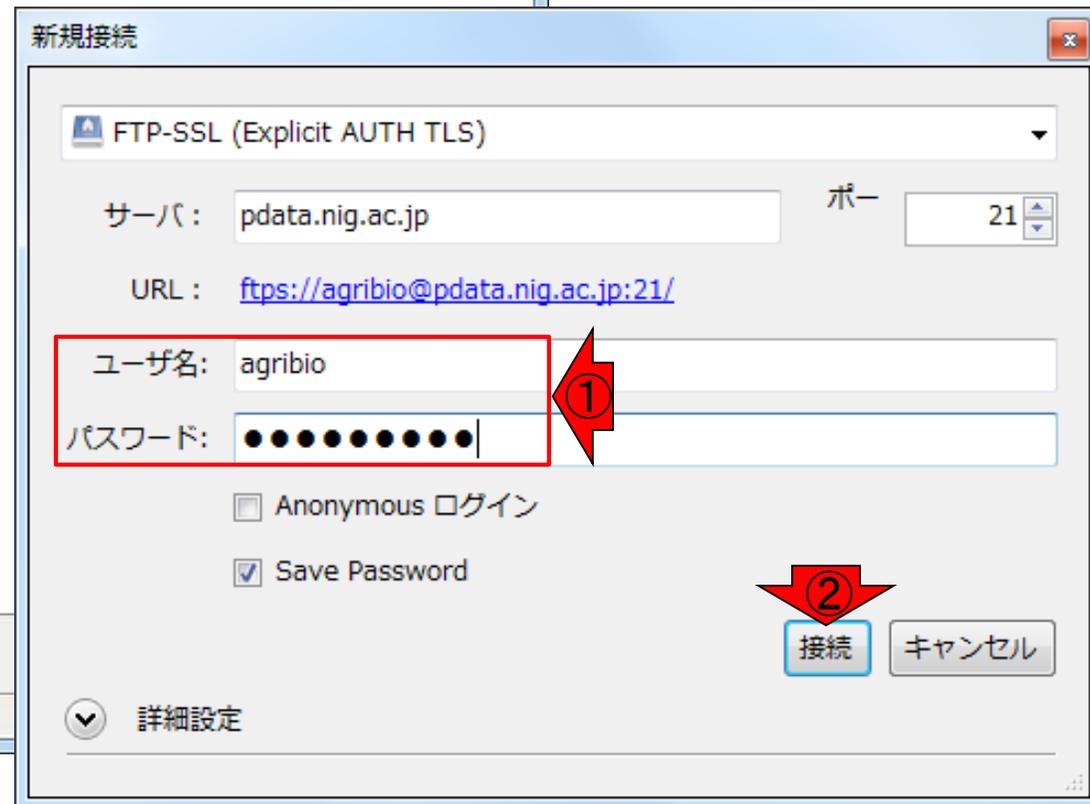
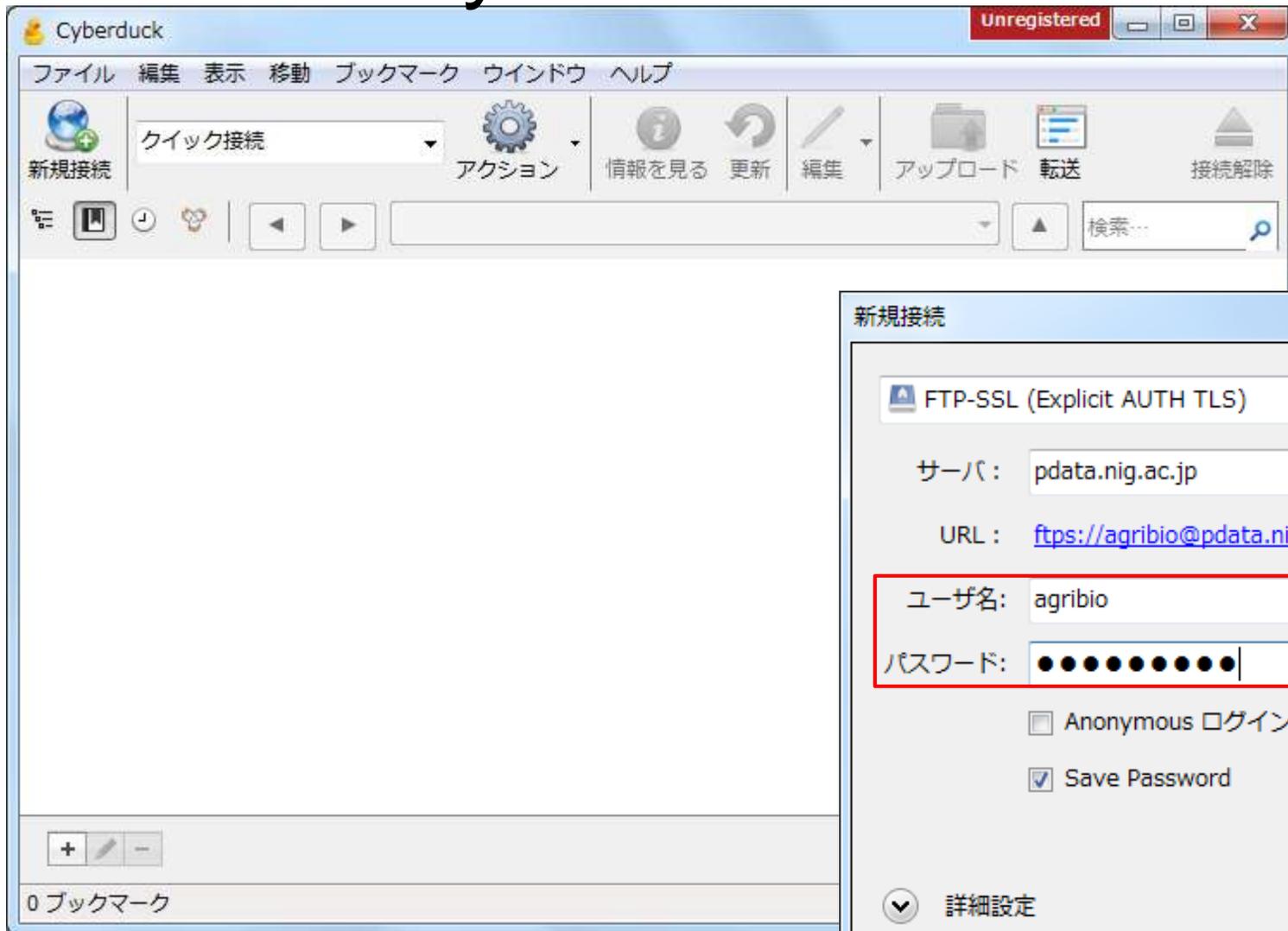


W13-2: Cyberduck設定



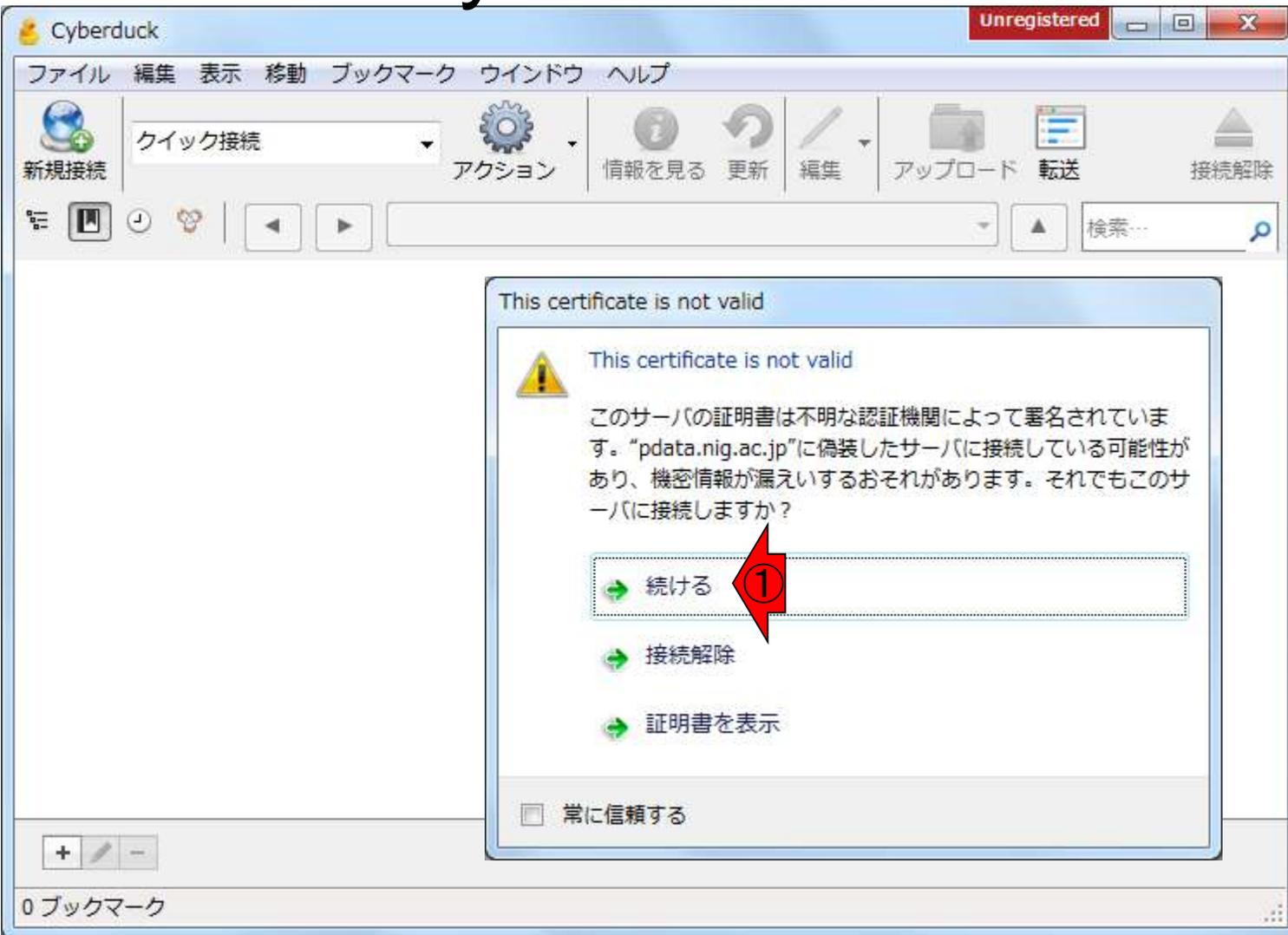
W13-2: Cyberduck設定

①ユーザ名とパスワードはDDBJ Pipelineにログインするときに使うものを入力。②接続



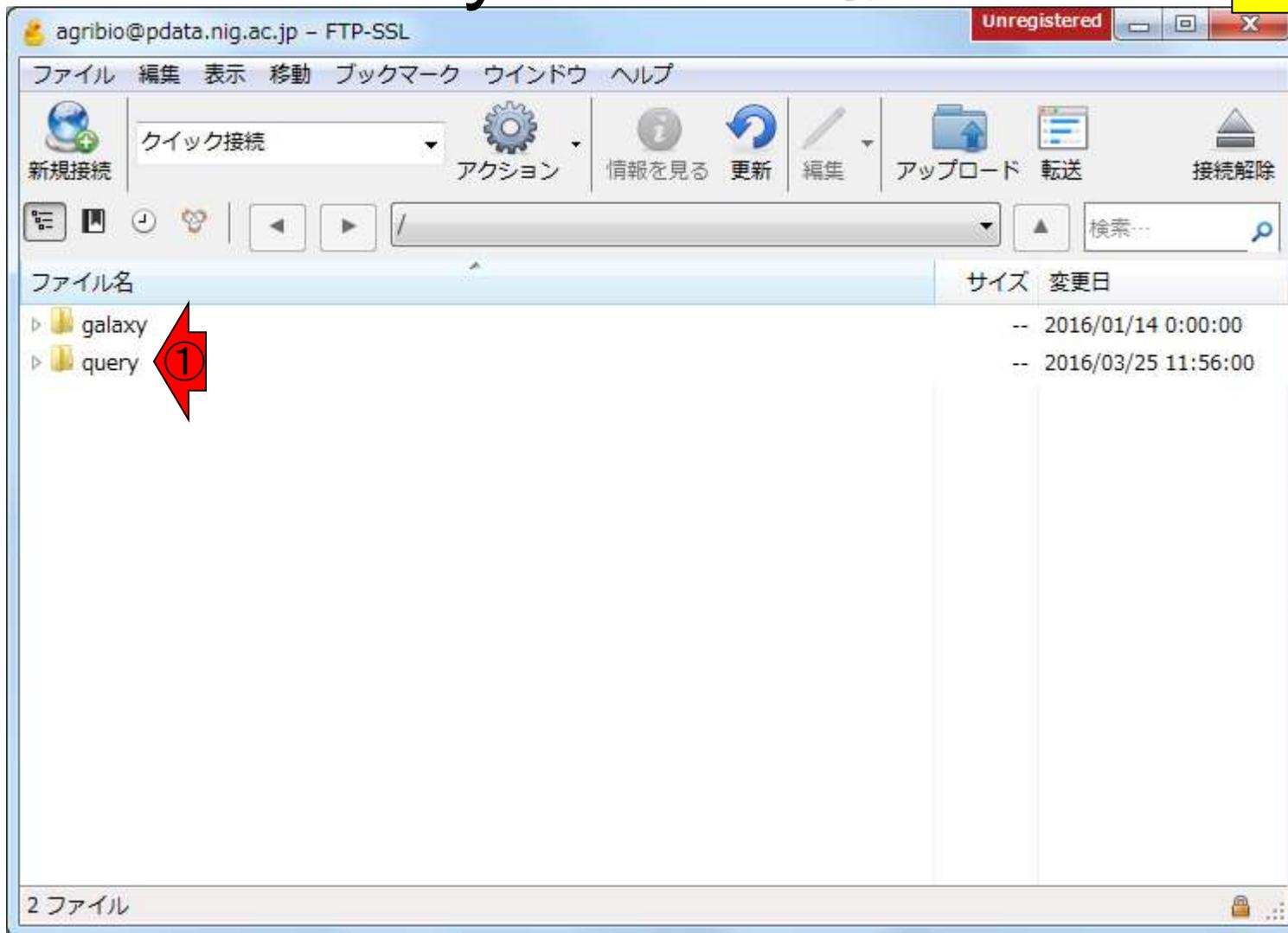
W13-2: Cyberduck設定

①ユーザ名とパスワードはDDBJ Pipelineにログインするときに入力



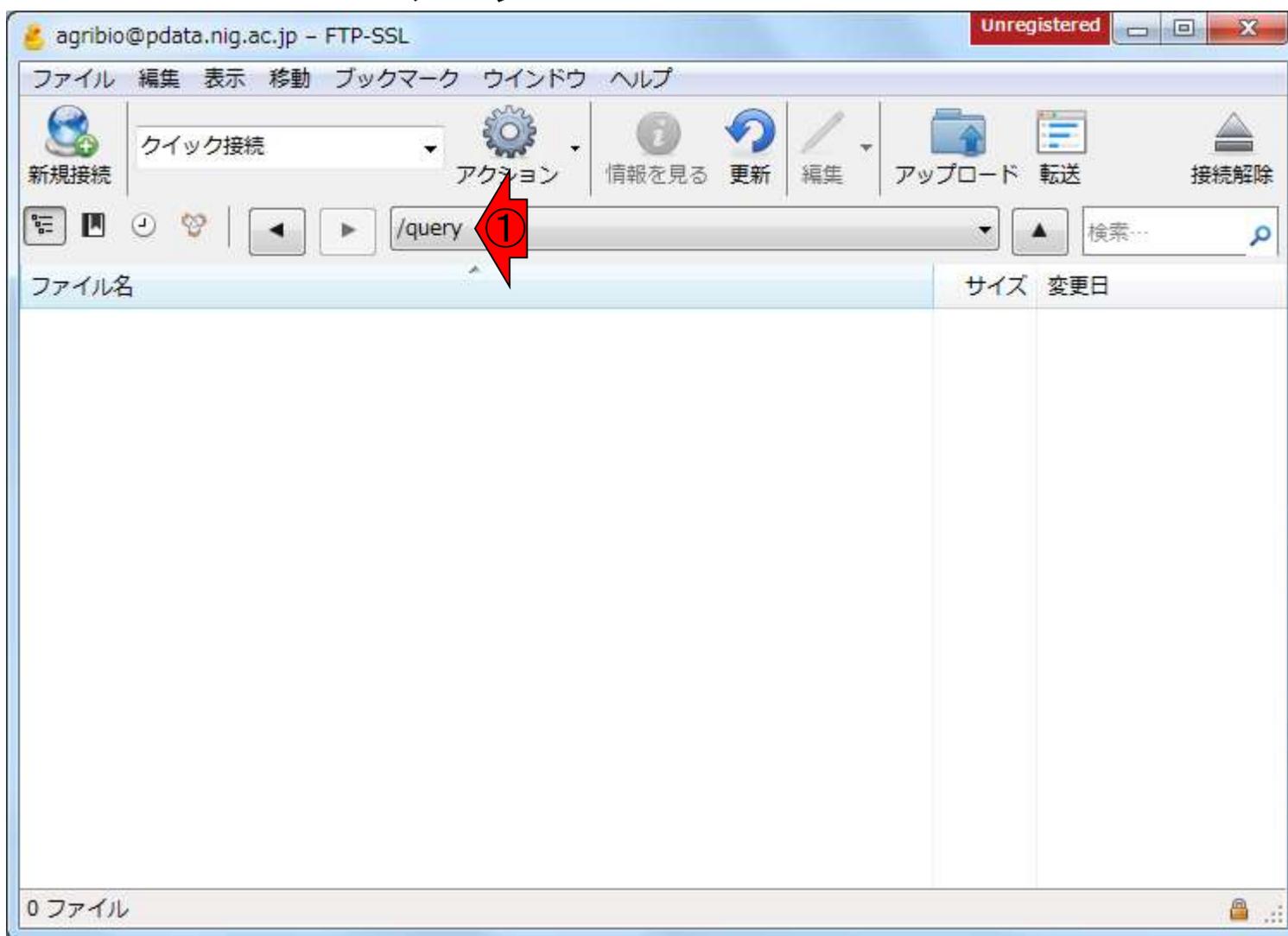
W13-3: Cyberduck接続

第6回W14-4とほぼ同じ状態。
DDBJ Pipelineにアップロード
したいので、①queryをクリック



W13-4: アップロード

①queryに移動できていることがわかる。②アップロードしたいファイルをドラッグ&ドロップで移動



W14-1: DDBJ Pipeline

DDBJ Pipelineの初期画面。①自分のIDとパスワードを入力して、②Login

http://p.ddbj.nig.ac.jp/pipeline/Login.do

DDBJ Read Annotation Pipeline

English Japanese

DDBJ Read Annotation Pipelineは、次世代シーケンサ配列のクラウド型データ解析プラットフォームです。

LOGIN

新規アカウント作成 ゲストとしてログイン

User ID:

Password:

Login

動作中JOBの確認

PipelineのIDをお持ちでない場合、[ゲストとしてログイン](#)することができます。

■ マニュアルおよびチュートリアル

- 日本語チュートリアル (FAQ)
- 英語マニュアル
- DBCLS 統合TVチュートリアル1 - 今日からはじめるDDBJ Read Annotation Pipeline
- DBCLS 統合TVチュートリアル2 - DDBJ Read Annotation Pipelineによるde novo Assembly解析
- チュートリアル: FTPでファイルをアップロードしDDBJ Pipelineへ登録する方法
- チュートリアル: DDBJ PipelineでHGAP法でPacBioリードの解析

Pipelineフローチャート

ユーザー → File Upload → Reads, metadata → DDBJ Read Archive → Reads, metadata → map positions → DDBJ → metadata, annotation, Manual curation, WGS/CON

基礎処理部: Mapping, de novo assembly

解析目的別ワークフロー (開発中): ゲノム解析, RNA-seq, ChIP-seq

①FTP upload。マップする側の
ファイルを指定するのが目的

W14-1 : DDBJ Pipeline

The screenshot shows the DDBJ Pipeline web interface. At the top, a navigation bar contains a sequence of steps: 'Select Query Files' (highlighted in orange), 'Select Tools', 'Set QuerySet', 'Set GenomeSet', 'Set Map Options', and 'Confirmation'. Below this is a 'Running Status' button. The main content area is titled 'Selecting Query Files' and features a horizontal menu with options: 'FTP upload', 'Private DRA entry' (highlighted in yellow), 'Import public DRA', 'Preprocessing', and 'HTTP upload'. A red arrow with the number '1' points to the 'FTP upload' option. Below the menu is a text area labeled 'Metadata of the DRA entry.' containing the text 'No DRA data'. On the right side, there are 'NEXT' buttons at the top and bottom right. The left sidebar contains navigation menus for 'ACCOUNT', 'ANALYSIS', and 'JOB STATUS'. The browser address bar shows 'http://p.ddbj.nig.ac.jp/pipeline/User.do' and the page title is 'Selecting Query Files'.

①赤枠部分の見え方はヒトそれぞれだが、連載第6,7回をやったヒトはこんな感じになっているはず

W14-2: マップする側

The screenshot shows the DDBJ pipeline web interface. At the top, a progress bar indicates the current step: 'Select Query Files' (highlighted in orange), followed by 'Select Tools', 'Set QuerySet', 'Set GenomeSet', 'Set Map Options', and 'Confirmation'. Below the progress bar, there are buttons for 'Running Status' and 'NEXT'. The main heading is 'Selecting Query Files'. Underneath, there are tabs for different upload methods: 'FTP upload' (highlighted in yellow), 'Private DRA entry', 'Import public DRA', 'Preprocessing', and 'HTTP upload'. A text box contains the text 'List of your uploaded files by FTP client. [Add new files]'. Below this are 'Select All' and 'Clear All' buttons. A table lists the uploaded files with columns for 'Filename', 'Description', 'Layout', and 'Instrmc'. The table is highlighted with a red border. At the bottom right of the table area, there are 'DELETE' and 'NEXT' buttons. On the right side of the page, there is a red arrow pointing to a circled number '1'.

	Filename	Description	Layout	Instrmc
<input type="checkbox"/>	m130821_065825_42195_c100539522550000001823089611241356_s1_p0.3.bax.h5	L.hokkaidonensis.PacBio3	single	PacBio
<input type="checkbox"/>	m130821_065825_42195_c100539522550000001823089611241356_s1_p0.2.bax.h5	L.hokkaidonensis.PacBio2	single	PacBio
<input type="checkbox"/>	m130821_065825_42195_c100539522550000001823089611241356_s1_p0.1.bax.h5	L.hokkaidonensis.PacBio1	single	PacBio
<input type="checkbox"/>	QC.1.trimmed.fastq.gz (more 1 files)	L.hokkaidonensis_MiSeq_denovo	paired	ILLUM

W14-2: マップする側

第6回W14-5あたりまでで登録 (registration)した①が残っているヒトはこれにチェックを入れて、②NEXT

ACCOUNT
login ID [agribio]
Logout
Change password

ANALYSIS
Data setup
DRA Start
FTP upload
HTTP upload
DRA Import
Preprocessing Start

step-1
Preprocessing
Mapping / de novo Assembly

step-2
Workflow
Genome (SNP/Short Indel)
RNA-seq (Tag count)
ChIP-seq

JOB STATUS
step1.

Select Query Files → Select Tools → Set QuerySet → Set GenomeSet → Set Map Options → Confirmation

Running Status

Selecting Query Files

NEXT

FTP upload Private DRA entry Import public DRA Preprocessing HTTP upload

List of your uploaded files by FTP client. [\[Add new files\]](#)

Select All Clear All

	Filename	Description	Layout	Instrument
<input type="checkbox"/>	m130821_065825_42195_c100539522550000001823089611241356_s1_p0.3.bax.h5	L.hokkaidonensis.PacBio3	single	PacBio
<input type="checkbox"/>	m130821_065825_42195_c100539522550000001823089611241356_s1_p0.2.bax.h5	L.hokkaidonensis.PacBio2	single	PacBio
<input type="checkbox"/>	m130821_065825_42195_c100539522550000001823089611241356_s1_p0.1.bax.h5	L.hokkaidonensis.PacBio1	single	PacBio
<input checked="" type="checkbox"/>	QC.1.trimmed.fastq.gz (more 1 files)	L.hokkaidonensis_MiSeq_denovo	paired	ILLUM

DELETE NEXT

W14-3: bwa

http://p.ddbj.nig.ac.jp/pipeline/SelectTool.do

Selecting Tools for Basic...

ACCOUNT

login ID [agribio]

Logout

Change password

ANALYSIS

Data setup

DRA Start

FTP upload

HTTP upload

DRA Import

Preprocessing Start

step-1

Preprocessing

Mapping / *de novo* Assembly

step-2

Workflow

Genome (SNP/Short Indel)

RNA-seq (Tag count)

ChIP-seq

JOB STATUS

step1. Preprocessing

Select Query Files → **Select Tools** → Set QuerySet → Set GenomeSet → Set Map Options → Confirmation

Running Status

Selecting Tools for Basic Analysis of DDBJ ANNOTATION PIPELINE

BACK NEXT

Reference Genome Mapping

	Tool	Help	Version	Input data			Evaluation			Analysis			Output format			Comment
				Base space	Color space	Paired end	Depth	Coverage	Error rate	SNP	Indel	.gff	.bed	SAM		
<input type="checkbox"/>	BLAT		34	✓						✓						Single-end analysis only
<input checked="" type="checkbox"/>	bwa		0.6.1	✓		✓	✓	✓	✓	✓				✓		
<input type="checkbox"/>	Bowtie		0.12.7	✓	✓	✓	✓	✓	✓	✓				✓		
<input type="checkbox"/>	TopHat		1.0.11	✓		✓	✓	✓	✓					✓		
<input type="checkbox"/>	Bowtie2		2.2.6	✓	✓	✓	✓	✓	✓	✓				✓	For reads longer than about 50 bp, Bowtie2 is generally faster, more sensitive, and uses less memory than Bowtie1.	
<input type="checkbox"/>	TopHat2		2.1.0	✓		✓	✓	✓	✓					✓		

de novo Assembly

W14-3:bwa

①解析したいデータにチェックをいれて、②Set as Pair-End

ACCOUNT

login ID [agribio]

Logout

Change password

ANALYSIS

Data setup

DRA Start

FTP upload

HTTP upload

DRA Import

Preprocessing Start

step-1

Preprocessing

Mapping / de novo Assembly

step-2

Workflow

Genome (SNP/Short Indel)

RNA-seq (Tag count)

ChIP-seq

JOB STATUS

step1. Preprocessing

Select Query Files → Select Tools → **Set QuerySet** → Set GenomeSet → Set Map Options → Confirmation

Running Status

Generating Query Sets from Query Read Files

RESET BACK NEXT

Paired-end analysis

Layout of paired sequence. 5'-3' 3'-5'

5' 3' 3' 5'

Linker(1)	Target	Linker(2)	Linker(3)	Target	Linker(4)
-----------	--------	-----------	-----------	--------	-----------

	Run ACCESSION	Read length	Quality Score
①	<input checked="" type="checkbox"/> QC.1.trimmed.fastq.gz	bp	

②

Set as Pair-End

RESET BACK NEXT

QUERY SET

RESET BACK NEXT

W14-3:bwa

http://p.ddbj.nig.ac.jp/pipeline/Confirm2ToSelectQuery.do

Generating Query Sets f... x

DDBJ
DNA Data Bank of Japan

ACCOUNT

login ID [agribio]

Logout

Change password

ANALYSIS

Data setup

DRA Start

FTP upload

HTTP upload

DRA Import

Preprocessing Start

step-1

Preprocessing

Mapping /
de novo Assembly

step-2

Workflow

Genome (SNP/Short Indel)

RNA-seq (Tag count)

ChIP-seq

JOB STATUS

step1.
Preprocessing

Select Query Files → Select Tools → **Set QuerySet** → Set GenomeSet → Set Map Options → Confirmation

Running Status

Generating Query Sets from Query Read Files

RESET BACK **NEXT**

Paired-end analysis

Layout of paired sequence. 5'-3' 3'-5'

5' 3' 3' 5'

Linker(1)	Target	Linker(2)	Linker(3)	Target	Linker(4)
-----------	--------	-----------	-----------	--------	-----------

Run ACCESSION Read length Quality Score

Set as Pair-End

QUERY SET

Query set1

PairedOrientation	RunAccession	RunAlias	RowLength	QualityScore1	QualityScore2
paired	20392	L.hokkaidonensis_MiSeq_denovo			

RESET BACK NEXT

W14-4: マップされる側

マップされる側のリファレンス配列ファイル(LH_draft.fa)を指定。①デフォルトはMajor genome setsになっているが、ここでは②ページ下部に移動して…

The screenshot shows the DDBJ pipeline web interface. At the top, a progress bar indicates the current step is 'Set GenomeSet'. The main heading is 'Specifying Database of Reference Genome'. On the left, there are navigation menus for 'ACCOUNT', 'ANALYSIS', and 'JOB STATUS'. The 'ANALYSIS' menu is expanded, showing 'Data setup' as the current step. A red arrow labeled '1' points to the 'Major genome sets' radio button. Below this, there are dropdown menus for 'Organisms' (set to 'Arabidopsis thaliana') and 'Genome sets' (set to 'TAIR8'). There are 'all check' and 'all clear' buttons. A list of genome files is shown with checkboxes: 'all.fa' (checked), 'chr01.fa', 'chr02.fa', 'chr03.fa', 'chr04.fa', 'chr05.fa', 'chrC.fa', and 'chrM.fa'. At the bottom, there is a radio button for 'User original sets'. A red arrow labeled '2' points to the bottom right corner of the page.

W14-4: マップされる側

Preprocessing Start

step-1

Preprocessing

Mapping /
de novo Assembly

step-2

Workflow

Genome (SNP/Short
Indel)
RNA-seq (Tag count)
ChIP-seq

JOB STATUS

step1.
Preprocessing

step1.
Mapping

step1.
***de novo* Assembly**

step2-All status

HELP

HELP

TUTORIAL

Contact Us.
DOBJ Read Annotation
Pipeline.
Development Team.

all check all clear

all.fa
 chr01.fa
 chr02.fa
 chr03.fa
 chr04.fa
 chr05.fa
 chrC.fa
 chrM.fa

User original sets

Download or upload reference

RESET BACK NEXT

W14-4: マップされる側

①でマップされる側のリファレンス配列ファイル(LH_draft.fa)を指定

The screenshot shows the DDBJ pipeline web interface. The browser address bar displays `http://p.ddbj.nig.ac.jp/pipeline/ChangeGenomeSetTypeToSelectGenome.do`. The page title is "Specifying Database of Reference Genome". A progress bar at the top indicates the current step: "Set GenomeSet".

On the left sidebar, there are sections for "ACCOUNT" (login ID: agribio, Logout, Change password), "ANALYSIS" (Data setup, DRA Start, FTP upload, HTTP upload, DRA Import, Preprocessing Start), "step-1" (Preprocessing, Mapping / de novo Assembly), "step-2" (Workflow: Genome (SNP/Short Indel), RNA-seq (Tag count), ChIP-seq), and "JOB STATUS" (step1. Preprocessing).

The main content area has a "Running Status" button and a "Specifying Database of Reference Genome" header. Below the header are "RESET", "BACK", and "NEXT" buttons. There are three radio button options:

- Major genome sets
- User original sets
- Download or upload reference

Below these options is a diagram titled "Retrieving a chromosome from DDBJ-DB by using HTTP REST". The diagram shows a user on the "INTERNET" sending a "Request" (HTTP REST *) to a "PIPELINE" server, which then sends "Data (fasta)" to the "DDBJ-DB" server. A note below the diagram states: "* Representational State Transfer (REST)".

Below the diagram is a form for "Input Accession Number (INSD) or (RefseqID)" with the value "CP000075" and a "LOAD" button.

At the bottom, there is a section for "Uploading reference from local" with a red arrow pointing to a "1" in a circle. Below this are "FASTA only" and "参照..." buttons, and an "UPLOAD" button.

W14-4: マップされる側

マップされる側のリファレンス配列ファイル (LH_draft.fa)を指定し終わったら、①UPLOAD

The screenshot shows the DDBJ pipeline web interface. The browser address bar displays `http://p.ddbj.nig.ac.jp/pipeline/ChangeGenomeSetTypeToSelectGenome.do`. The page title is "Specifying Database of Reference Genome". A progress bar at the top indicates the current step is "Set GenomeSet".

On the left sidebar, the "ACCOUNT" section shows the login ID "agribio" and options for "Logout" and "Change password". The "ANALYSIS" section lists various data setup options, with "Preprocessing Start" selected. Below this, "step-1" includes "Preprocessing" and "Mapping / de novo Assembly". "step-2" includes a "Workflow" section with options for "Genome (SNP/Short Indel)", "RNA-seq (Tag count)", and "ChIP-seq". The "JOB STATUS" section shows "step1. Preprocessing".

The main content area is titled "Specifying Database of Reference Genome" and includes "RESET", "BACK", and "NEXT" buttons. Three radio button options are present:

- Major genome sets
- User original sets
- Download or upload reference

Below the options is a diagram titled "Retrieving a chromosome from DDBJ-DB by using HTTP REST". It shows a user on the "INTERNET" sending a "Request" (HTTP REST) to the "PIPELINE", which then sends "Data (fasta)" to "DDBJ-DB". A note below the diagram states: "* Representational State Transfer (REST)".

At the bottom, there is a section for "Uploading reference from local drive." with a text input field containing "FASTA only C:\Users\kadota\Desktop\参照..." and a "参照..." button. A red arrow with the number "1" points to the "UPLOAD" button.

特に変化がないが、①ページ下部に移動すると…

W14-4: マップされる側

The screenshot shows a web browser window with the URL <http://p.ddbj.nig.ac.jp/pipeline/GenomeUpload.do>. The page title is "Specifying Database of Reference Genome". A progress bar at the top indicates the current step: "Set GenomeSet".

On the left sidebar, there are sections for "ACCOUNT" (login ID [agribio], Logout, Change password), "ANALYSIS" (Data setup, DRA Start, FTP upload, HTTP upload, DRA Import, Preprocessing Start), "step-1" (Preprocessing, Mapping / de novo Assembly), "step-2" (Workflow: Genome (SNP/Short Indel), RNA-seq (Tag count), ChIP-seq), and "JOB STATUS" (step1. Preprocessing).

The main content area has a "Running Status" button and a "Specifying Database of Reference Genome" header. Below the header are "RESET", "BACK", and "NEXT" buttons. There are three radio button options:

- Major genome sets
- User original sets
- Download or upload reference

A red arrow with the number "1" points to the "Download or upload reference" option.

Below the options is a diagram titled "Retrieving a chromosome from DDBJ-DB by using HTTP REST". The diagram shows a user on the "INTERNET" sending a "Request" (HTTP REST *) to a "PIPELINE" server, which then sends "Data (fasta)" to the "DDBJ-DB" server. A note below the diagram states: "* Representational State Transfer (REST)".

Below the diagram is a form for "Input Accession Number (INSD) or (RefseqID)" with the value "CP000075" and a "LOAD" button.

At the bottom, there is a section for "Uploading reference from local drive." with a "FASTA only" input field, a "参照..." button, and an "UPLOAD" button.

W14-4: マップされる側

①こんなのが見える。②「>chromosome」という文字は、アップロードしたファイルの最初の1行目と同じものなので、うまくアップロードできたのだらうと判断。③NEXT。④CREATE DATASETをしておくと、毎回リファレンス配列をアップロードし直さなくても済むが、ここではやらない

The screenshot shows the 'Download or upload reference' section of the GenomeUpload pipeline. The URL is <http://p.ddbj.nig.ac.jp/pipeline/GenomeUpload.do>. The interface includes a sidebar with navigation options like 'Preprocessing Start', 'Mapping / de novo Assembly', and 'JOB STATUS'. The main content area has two sections: 'Retrieving a chromosome from DDBJ-DB by using HTTP REST' and 'Uploading reference from local drive.'.

The 'Retrieving a chromosome from DDBJ-DB by using HTTP REST' section contains a form with 'Input Accession Number (INSD) or (RefseqID)' set to 'CP000075' and a 'LOAD' button. Below this is a diagram showing the flow from 'INTERNET' to 'PIPELINE' and then to 'DDBJ-DB'. A 'Request' (HTTP REST) is sent to DDBJ-DB, and 'Data (fasta)' is returned. A note indicates '* Representational State Transfer (REST)'.

The 'Uploading reference from local drive.' section has a 'FASTA only' field, a '参照...' button, and an 'UPLOAD' button. Below this is a table with one row:

<input checked="" type="checkbox"/>	>chromosome	DELETE
-------------------------------------	-------------	--------

At the bottom right, there are buttons for 'RESET', 'BACK', and 'NEXT'. A 'CREATE DATASET' button is also present. Red arrows with numbers 1 through 4 point to specific elements: 1 points to the 'CREATE DATASET' button, 2 points to the '>chromosome' text in the table, 3 points to the 'NEXT' button, and 4 points to the 'CREATE DATASET' button.

W14-5: オプション指定

基本はデフォルトのままでよい。① Step1) リファレンス配列のインデックス化、② Step2) BWAマッピング本番

The screenshot shows the DDBJ pipeline web interface. At the top, a navigation bar contains a sequence of steps: 'Select Query Files', 'Select Tools', 'Set QuerySet', 'Set GenomeSet', 'Set Map Options' (highlighted in orange), and 'Confirmation'. Below this, a sidebar on the left lists 'ACCOUNT' (login ID, Logout, Change password) and 'ANALYSIS' (Data setup, DRA Start, FTP upload, HTTP upload, DRA Import, Preprocessing Start, step-1, step-2, Workflow, JOB STATUS). The main content area is titled 'Setting for Reference Genome Mapping' and includes 'bwa' and 'Set optional parameters of the paired-end analysis' sections. Two red arrows with circled numbers 1 and 2 point to the 'Step1) Convert reference sequence' and 'Step2) Map' sections respectively. The 'Step1' section shows a dropdown menu for 'bwa index' set to '-a is (for small-size reference)' and a text field for 'refgenome.fasta'. The 'Step2' section shows three command-line examples for 'bwa aln -t 4' and 'bwa sampe'. Below, there is a section for 'Step3)uniq' with radio button options for removing multiple hits.

W14-5: オプション指

①Step3) ユニーク化処理。②デフォルトは、ペアのリードの両方がユニークにマップされたものを残し、それ以外のリードは除去する処理を行っている。反復配列のような領域が存在すると、1つのリードが複数箇所にマップされてしまい、解析結果の解釈を難しくすることがあるため、変異解析などではこれらのリードを除外する処理を行うことがある

The screenshot shows the DDBJ pipeline web interface. On the left, there is a sidebar with navigation links: 'ChIP-seq', 'JOB STATUS' (with sub-items: 'step1. Preprocessing', 'step1. Mapping', 'step1. de novo Assembly', 'step2-All status'), 'HELP' (with sub-items: 'HELP', 'TUTORIAL'), and 'Contact Us. DDBJ Read Annotation Pipeline. Development Team.'.

The main content area is titled 'Step3) 'uniq': Remove multiple hits on the genome'. It includes the instruction 'Please choose uniq mode.' and four radio button options:

- Do not remove any read.
- Retain pairs when both reads mapped uniquely or one of reads mapped uniquely, and Discard other pairs.
- Retain pairs when both reads mapped uniquely, and Discard other pairs.
- Retain uniquely mapped reads and discard multiply mapped reads.

Below this, there are sections for 'Step4) Convert the read alignment to .BAM format' (with command: `samtools view -bS -o out.bam out.sam`), 'Step5) Detect DNA polymorphism' (with instruction 'Please choose one of the following.' and three radio button options, the second of which is selected and highlighted in yellow: `out.var.raw.bcf`), and 'Step6) Analysis for Depth, Coverage' (with commands: `samtools sort -o out.bam out_sorted.bam`, `samtools pileup -c -f reference.fa out_sorted.bam > out.pileup`, and `perl pileup_for_CoverageDepth.pl out.pileup reference.fa`).

At the bottom, there is a link for 'Step7) Create assembled sequences in FASTA file from pileupped reads to submit WGS division of DDBJ.'

Red arrows with numbers 1 and 2 point to the Step3 title and the selected Step5 option, respectively.

W14-5: オプション指定

①Step4) BWAの出力であるSAMファイルを入力として、BAMファイルを作成。②Step5) DNA多型の検出(変異解析のこと)。③デフォルトの手順の、④最後の出力がリファレンス配列と異なる部分のみを抽出したVCF形式ファイル

http://p.ddbj.nig.ac.jp/pipeline/SelectGenomeNext.do

ChIP-seq

JOB STATUS

- step1. Preprocessing
- step1. Mapping
- step1. **de novo Assembly**
- step2-All status

HELP

- HELP
- TUTORIAL
- Contact Us. DDBJ Read Annotation Pipeline. Development Team.

Step3)'uniq': Remove multiple hits on the genome from out.sam.

Please choose uniq mode.

- Do not remove any read.
- Retain pairs when both reads mapped uniquely or one of reads mapped uniquely, and Discard other pairs.
- Retain pairs when both reads mapped uniquely, and Discard other pairs.
- Retain uniquely mapped reads and discard multiply mapped reads.

Step4) Convert the read alignment to .BAM format

```
samtools view -bS -o out.bam out.sam
```

Step5) Detect DNA polymorphism

Please choose one of the following.

- samtools pileup -f refgenome.fasta out.bam | bcftools view
- samtools mpileup -f refgenome.fasta out.bam | bcftools view - >
- out.var.raw.bcf
- bcftools view out.var.raw.bcf | vcftutils.pl varFilter > out.var.fit.vcf

Step6) Analysis for Depth, Coverage

```
samtools sort -o out.bam out_sorted.bam
samtools pileup -c -f reference.fa out_sorted.bam > out.pileup
perl pileup_for_CoverageDepth.pl out.pileup reference.fa
* This command does not appear in the list.
```

Step7) Create assembled sequences in FASTA file from pileupped reads to submit WGS division of DDBJ.

①Step6)や、②Step7)は、第8回
内容とは関係が薄いので省略

W14-5: オプション指定

Mapping

step1. **de novo Assembly**

step2-All status

HELP

HELP ↗

TUTORIAL

Contact Us.
DDBJ Read Annotation Pipeline.
Development Team.

Retain pairs when both reads mapped uniquely, and Discard other pairs.

Retain uniquely mapped reads and discard multiply mapped reads.

Step4) Convert the read alignment to .BAM format

```
samtools view -bS -o out.bam out.sam
```

Step5) Detect DNA polymorphism

Please choose one of the following.

samtools pileup -f refgenome.fasta out.bam | bcftools view

samtools mpileup -f refgenome.fasta out.bam | bcftools view ->

out.var.raw.bcf

```
bcftools view out.var.raw.bcf | vcutils.pl varFilter  > out.var.fit.vcf
```

① Step6) Analysis for Depth, Coverage

```
samtools sort -o out.bam out_sorted.bam  
samtools pileup -c -f reference.fa out_sorted.bam > out.pileup  
perl pileup_for_CoverageDepth.pl out.pileup reference.fa  
* This command does not appear in the list.
```

② Step7) Create assembled sequences in FASTA file from pileupped reads to [submit WGS division of DDBJ](#).

perl getConsGeno_4pipeline.pl pileupFile out_WGS.txt

** Threshold of insertion of pileupped reads: the quality threshold for indels <= 50 and allele constitutes 80% of pileupped reads.*

BACK NEXT

W14-6: 確認して実行

①ジョブ完了後のメール送信先を指定して、②一応ページ下部までざっと眺める

http://p.ddbj.nig.ac.jp/pipeline/Confirm.do

Run Confirmation

Select Query Files → Select Tools → Set QuerySet → Set GenomeSet → Set Map Options → Confirmation

Running Status

Run Confirmation

BACK RUN

Destination of mail

When the request is completed, the system sends an email to this address.

* Required

Result files will be deleted 60 days after submission.

Reference Genome Map [bwa]

Query sets

Query set1	PairedOrientation	RunAccession	RunAlias	RowLength	Quality Score1	Quality Score2
paired		20392	L.hokkaidonensis_MiSeq_denovo			

genome sets

Upload file

- >chromosome

Command Options

bwa

Set optional parameters of the paired-end analysis

W14-6: 確認して実行

①一応ページ下部までざっと眺めて問題ないようなら、②RUN、③OK

The screenshot shows a web browser window with the URL `http://p.ddbj.nig.ac.jp/pipeline/Confirm.do`. The page contains several steps of a pipeline configuration:

- Step 4: Convert the read alignment to .BAM format. Command: `samtools view -bS -o out.bam out.sam`
- Step 5: Detect DNA polymorphism. Please choose one of the following:
 - `samtools pileup -c`
 - `samtools mpileup -u -CS`
 - `out.var.raw.bcf`
 - `bcftools view out.var.raw.bcf`
- Step 6: Analysis for Depth, Coverage, etc. Commands: `samtools sort -o out.bam out.sam`, `samtools pileup -c -f reference.fasta`, `perl pileup_for_CoverageDepth.pl`. Note: * This command does not appear to work.
- Step 7: Create assembled sequences in FASTA file from pileupped reads to submit WGS division of DDBJ. Command: `perl getConsGeno_4pipeline.pl pileupFile [Not to include insertion of pileupped reads.] out_WGS.txt`. Note: * Threshold of insertion of pileupped reads: the quality threshold for indels <= 50 and allele constitutes 80% of pileupped reads.

A modal dialog box titled "Web ページからのメッセージ" (Message from the Web Page) is displayed in the center, asking "Do you really want to execute pipeline programs?". It has "OK" and "キャンセル" (Cancel) buttons. A red arrow labeled "③" points to the "OK" button.

At the bottom right of the page, there are "BACK" and "RUN" buttons. A red arrow labeled "②" points to the "RUN" button.

On the right side of the page, there is a vertical scrollbar. A red arrow labeled "①" points to the scrollbar area, indicating that the user should scroll down to check the bottom of the page.

ジョブ投入完了。メール
通知があるまで思考停止

W14-6: 確認して実行

The screenshot shows a web browser window with the URL <http://p.ddbj.nig.ac.jp/pipeline/ConfirmRun.do>. The page features the DDBJ logo and a navigation menu on the left with sections for ACCOUNT, ANALYSIS, and JOB STATUS. At the top, a workflow diagram shows steps: Select Query Files, Select Tools, Set QuerySet, Set GenomeSet, Set Map Options, and Confirmation. A large blue banner in the center reads "The reservation was completed." Below this banner are buttons for "STATUS" and "NEXT JOB".

W15-1: 計算終了

2016/09/12 (月) 17:38

pipeline_team@g.nig.ac.jp

Job finished : DDBJ Read Annotation Pipeline

宛先 kadota@bi.au-tokyo.ac.jp

C C pipeline_report@g.nig.ac.jp

 このメッセージから余分な改行を削除しました。

Dear agribio,

Your request to DDBJ pipeline service has finished.

Please visit the web site to obtain analytical results.

Request ID: 24152

URL: <https://p.ddbj.nig.ac.jp/>

If you have troubles in this service, please write to pipeline_dev@ddbj.nig.ac.jp Thank you for trying our analytical service.

Regards,

DDBJ

①用いたプログラムはbwa (ver. 0.6.1)。②ページ下部に移動

W15-2: 結果の要約

http://p.ddbj.nig.ac.jp/pipeline/DetailView.do?query_set_id=24152

Detail view

ACCOUNT

login ID [agribio]

Logout

Change password

ANALYSIS

Data setup

DRA Start

FTP upload

HTTP upload

DRA Import

Preprocessing Start

step-1

Preprocessing

Mapping / de novo Assembly

step-2

Workflow

Genome (SNP/Short Indel)

RNA-seq (Tag count)

ChIP-seq

JOB STATUS

step1. Preprocessing

Select Query Files → Select Tools → Set QuerySet → Set GenomeSet → Set Map Options → Confirmation

Running Status

Detail view

BACK

Job info

ID: 24152

Tool (Version): bwa (0.6.1)

RunAccession	Filename	Download	Read length	Alias
QC.1.trimmed.fastq.gz		QC.1.trimmed.fastq.gz	N.A. bp	L.hokkaidonensis_MiSeq_denovo

Genome set: ---

Chromosome: [LH_draft.fa](#)

Download modified queries

- [QC.1.trimmed.fastq.gz](#) (Original size 189.4 MB)
- [QC.2.trimmed.fastq.gz](#) (Original size 189.6 MB)

Download merged pileup file

Uniq.sam files have been merged if you specified 'uniq' option.

Top of page

W15-2: 結果の要約

①Map ratioのところの解説。全297,633リード中、281,303リードがペアでマップされ、マップ率は281,303/297,633 = 94.513%であった

http://p.ddbj.nig.ac.jp/pipeline/DetailView.do?query_set_id=24152

Detail view

JOB STATUS

- step1. Preprocessing
- step1. Mapping
- step1. *de novo* Assembly
- step2-All status

HELP

- HELP
- TUTORIAL
- Contact Us. DDBJ Read Annotation Pipeline. Development Team.

Download merged pileup file

Uniq.sam files have been merged if you specified 'uniq' option.

- [merged.var.fit.vcf.gz \(Original size 6.2 KB\)](#)
- [merged.sam.gz \(Original size 320.9 MB\)](#)

Download wgs file

- [out_WGS.fasta.gz \(Original size 197 byte\)](#)

Position errors	Map ratio	Depth, Coverage
PDF download	total query # : 297,633 mapped query # : 281,303 map ratio : 94.513 %	coverage : 2400552 / 2400584 * 100 = 99.999 depth : 130565653 / 2400552 = 54.390

Time

Wait time	Start time	End time
0: 0:18	2016-09-12 17:16:14	2016-09-12 17:37:57

LH_draft.fa

Command	Start time	End time	Log1	Log2	Result	MD5
Create BWA Index File bwa index [-a is] LH_draft.fa	2016-09-12 17:16:15	2016-09-12 17:16:48		View		
BWA : Alignment bwa aln LH_draft.fa QC.1.trimmed.fastq > 1.sai	2016-09-12 17:16:48	2016-09-12 17:17:10		View		
BWA : Alignment bwa aln LH_draft.fa QC.2.trimmed.fastq > 2.sai	2016-09-12 17:17:11	2016-09-12 17:17:33		View		
BWA : SAMPE bwa sampe LH_draft.fa 1.sai 2.sai QC.1.trimmed.fastq QC.2.trimmed.fastq > out.sam	2016-09-12 17:17:34	2016-09-12 17:19:43		View	Download(116.3 MB)	
Extract Unmapped Reads	2016-09-12	2016-09-12				

Top of page

W15-2: 結果の要約

①Depth, Coverageのところの解説で、まずは② Coverage。ゲノムサイズ(全長)が2,400,584 bp。そのうち2,400,552 bp分がマップされたリードで覆われている。被覆率(coverage)は、 $2,400,552 / 2,400,584 = 99.99867\%$

http://p.ddbj.nig.ac.jp/pipeline/DetailView.do?query_set_id=24152

Detail view

JOB STATUS

step1. **Preprocessing**

step1. **Mapping**

step1. **de novo Assembly**

step2-All status

HELP

HELP [↗](#)

TUTORIAL

Contact Us.
DDBJ Read Annotation Pipeline. Development Team.

Download merged pileup file

Uniq.sam files have been merged if you specified 'uniq' option.

- [merged.var.fit.vcf.gz \(Original size 6.2 KB\)](#)
- [merged.sam.gz \(Original size 320.9 MB\)](#)

Download wgs file

- [out_WGS.fasta.gz \(Original size 197 byte\)](#)

Position errors	Map ratio	Depth, Coverage
PDF download	total query # : 297,633 mapped query # : 281,303 map ratio : 94.513 %	coverage : $2400552 / 2400584 * 100 = 99.999$ depth : $130565653 / 2400552 = 54.390$

Time

Wait time	Start time	End time
0: 0:18	2016-09-12 17:16:14	2016-09-12 17:37:57

LH_draft.fa

Command	Start time	End time	Log1	Log2	Result	MD5
Create BWA Index File bwa index [-a is] LH_draft.fa	2016-09-12 17:16:15	2016-09-12 17:16:48		View		
BWA : Alignment bwa aln LH_draft.fa QC.1.trimmed.fastq > 1.sai	2016-09-12 17:16:48	2016-09-12 17:17:10		View		
BWA : Alignment bwa aln LH_draft.fa QC.2.trimmed.fastq > 2.sai	2016-09-12 17:17:11	2016-09-12 17:17:33		View		
BWA : SAMPE bwa sampe LH_draft.fa 1.sai 2.sai QC.1.trimmed.fastq QC.2.trimmed.fastq > out.sam	2016-09-12 17:17:34	2016-09-12 17:19:43		View	Download(116.3 MB)	
Extract Unmapped Reads	2016-09-12	2016-09-12				



W15-2: 結果の要

①Depth, Coverageのところの解説で、次は②depth。③マップされたリードで覆われている領域(2,400,552 bp)が平均してどれだけの厚み(depth)でマップされているかを示す。
 ④130,565,653は、マップされたリードの総塩基数。リード長が251 bp、paired-endなので×2、マップされたリード数が281,303なので、 $251 \times 2 \times 281,303 = 141,214,106$ 。この値は、リード長が全て251 bpであった場合であり、実際にはFaQCsでアダプタートリムやクオリティフィルタリングがかけられているので、ちょっと小さ目の値になるのは妥当

JOB STATUS

step1. Preprocessing

step1. Mapping

step1. *de novo* Assembly

step2-All status

HELP

HELP [↗](#)

TUTORIAL

Contact Us.
 DDBJ Read Annotation Pipeline.
 Development Team.

http://p.ddbj.nig.ac.jp/pipeline/DetailView.do?query_set

Download merged pileup file

Uniq.sam files have been merged if you spe

- [merged.var.fit.vcf.gz](#) (Original size 6
- [merged.sam.gz](#) (Original size 320.9

Download wgs file

- [out_WGS.fasta.gz](#) (Original size 197 byte)

Position errors	Map ratio	Depth, Coverage
PDF download	total query # : 297,633 mapped query # : 281,303 map ratio : 94.513 %	average : $\frac{2400552}{2400584} \times 100 = 99.999$ depth : $\frac{130565653}{2400552} = 54.390$

Time

Wait time	Start time	End time
0: 0:18	2016-09-12 17:16:14	2016-09-12 17:37:57

LH_draft.fa

Command	Start time	End time	Log1	Log2	Result	MD5
Create BWA Index File bwa index [-a is] LH_draft.fa	2016-09-12 17:16:15	2016-09-12 17:16:48		View		
BWA : Alignment bwa aln LH_draft.fa QC.1.trimmed.fastq > 1.sai	2016-09-12 17:16:48	2016-09-12 17:17:10		View		
BWA : Alignment bwa aln LH_draft.fa QC.2.trimmed.fastq > 2.sai	2016-09-12 17:17:11	2016-09-12 17:17:33		View		
BWA : SAMPE bwa sampe LH_draft.fa 1.sai 2.sai QC.1.trimmed.fastq QC.2.trimmed.fastq > out.sam	2016-09-12 17:17:34	2016-09-12 17:19:43		View	Download(116.3 MB)	
Extract Unmapped Reads	2016-09-12	2016-09-12				



W15-2: 結果の要約

← → 🌐 http://p.ddbj.nig.ac.jp/pipeline/DetailView.do?query_set_id=24152
🔍
🔄
🌐 Detail view
🏠
★
⚙️

JOB STATUS

step1. **Preprocessing**

step1. **Mapping**

step1. **de novo Assembly**

step2-All status

HELP

HELP 🔗

TUTORIAL

Contact Us.
DDBJ Read Annotation Pipeline.
Development Team.

Download merged pileup file
 Uniq.sam files have been merged if you specified 'uniq' option.

- [merged.var.fit.vcf.gz \(Original size 6.2 KB\)](#)
- [merged.sam.gz \(Original size 320.9 MB\)](#)

Download wgs file

- [out_WGS.fasta.gz \(Original size 197 byte\)](#)

Position errors	Map ratio	Depth, Coverage
PDF download	total query # : 297,633 mapped query # : 281,303 map ratio : 94.513 %	coverage : 2400552 / 2400584 * 100 = 99.999 depth : 130565653 / 2400552 = 54.390

Time

Wait time	Start time	End time
0: 0:18	2016-09-12 17:16:14	2016-09-12 17:37:57

LH_draft.fa							
Command	Start time	End time	Log1	Log2	Result	MD5	
Create BWA Index File bwa index [-a is] LH_draft.fa	2016-09-12 17:16:15	2016-09-12 17:16:48		View			
BWA : Alignment bwa aln LH_draft.fa QC.1.trimmed.fastq > 1.sai	2016-09-12 17:16:48	2016-09-12 17:17:10		View			
BWA : Alignment bwa aln LH_draft.fa QC.2.trimmed.fastq > 2.sai	2016-09-12 17:17:11	2016-09-12 17:17:33		View			
BWA : SAMPE bwa sampe LH_draft.fa 1.sai 2.sai QC.1.trimmed.fastq QC.2.trimmed.fastq > out.sam	2016-09-12 17:17:34	2016-09-12 17:19:43		View	Download(116.3 MB)		Top of page
Extract Unmapped Reads	2016-09-12	2016-09-12					

W15-3:ダウンロード

①DDBJ Pipelineでは、いくつかのBWA実行結果ファイルをダウンロード可能。まずは、②ユニーク化処理前のSAM形式ファイル(out.sam.zip; 116.3MB)をダウンロード

The screenshot shows the 'Detail view' page for a pipeline job. The left sidebar contains navigation links for 'JOB STATUS' (Preprocessing, Mapping, de novo Assembly, All status), 'HELP', 'TUTORIAL', and 'Contact Us'. The main content area is divided into several sections:

- Download merged pileup file:** Uniq.sam files have been merged if you specified 'uniq' option. Links for [merged.var.fit.vcf.gz](#) (6.2 KB) and [merged.sam.gz](#) (320.9 MB).
- Download wgs file:** Link for [out_WGS.fasta.gz](#) (197 byte).
- Position errors, Map ratio, Depth, Coverage:** A summary table with a 'PDF download' button. Map ratio: total query # : 297,633, mapped query # : 281,303, map ratio : 94.513 %. Depth, Coverage: coverage : 2400552 / 2400584 * 100 = 99.999, depth : 130565653 / 2400552 = 54.390.
- Time:** A table showing Wait time (0: 0:18), Start time (2016-09-12 17:16:14), and End time (2016-09-12 17:37:57).
- LH_draft.fa:** A table of job steps with columns for Command, Start time, End time, Log1, Log2, Result, and MD5. A red arrow labeled '1' points to the 'Result' column. The 'BWA : SAMPE' row has a 'Download(116.3 MB)' link in the Result column, which is highlighted by a red arrow labeled '2'.

W15-3: ダウンロード

- ③ユニーク化処理後のSAM形式ファイル(uniqout.sam.zip)、
- ④ユニーク化処理後のソート済みのBAM形式ファイル(uniqout.bam.zip)、
- ⑤BAMインデックスファイル(out2.bam.bai.zip)、
- ⑥VCFファイル(out-unique.var.flt.vcf.zip)

http://p.ddbj.nig.ac.jp/pipeline/DetailView.do?query_set_id=24152 Detail view

LH_draft.fa	Command	Start time	End time	Log1	Log2	Result	MD5
	Create BWA Index File bwa index [-a is] LH_draft.fa	2016-09-12 17:16:15	2016-09-12 17:16:48		View		
	BWA : Alignment bwa aln LH_draft.fa QC.1.trimmed.fastq > 1.sai	2016-09-12 17:16:48	2016-09-12 17:17:10		View		
	BWA : Alignment bwa aln LH_draft.fa QC.2.trimmed.fastq > 2.sai	2016-09-12 17:17:11	2016-09-12 17:17:33		View		
	BWA : SAMPE bwa sampe LH_draft.fa 1.sai 2.sai QC.1.trimmed.fastq QC.2.trimmed.fastq > out.sam	2016-09-12 17:17:34	2016-09-12 17:19:43		View	Download(116.3 MB)	MD5
	Extract Unmapped Reads python extractUnmappedFASTQ.py QC.1.trimmed.fastq QC.2.trimmed.fastq out.sam	2016-09-12 17:20:38	2016-09-12 17:21:00			Download(4.2 MB)	MD5
	Convert SAM to BAM samtools view -bS -o out.bam out.sam	2016-09-12 17:21:23	2016-09-12 17:21:56		View	Download(124.2 MB)	MD5
	Sort BAM File samtools sort out.bam out2	2016-09-12 17:22:08	2016-09-12 17:22:50			Download(87.9 MB)	MD5
	Create BAM Index File samtools index out2.bam	2016-09-12 17:23:02	2016-09-12 17:23:12			Download(3.6 KB)	MD5
	Uniquify SAM (Remove Multiple Hits) perl sam2uniq.pl out.sam UBE > uniqout.sam	2016-09-12 17:23:24	2016-09-12 17:23:36			Download(96.3 MB)	MD5
	Convert SAM to BAM [For Unique SAM] samtools view -bS -o uniqout.bam uniqout.sam	2016-09-12 17:24:31	2016-09-12 17:25:04		View	Download(199.9 MB)	MD5
	Sort BAM File [For Unique SAM] samtools sort uniqout.bam out2	2016-09-12 17:25:59	2016-09-12 17:26:20			Download(71.3 MB)	MD5
	Create BAM Index File [For Unique SAM] samtools index out2.bam	2016-09-12 17:26:32	2016-09-12 17:26:43			Download(3.2 KB)	MD5
	Mpileup and Create BCF File [For Unique SAM] samtools mpileup -u -C50 -BQ0 -d10000000 -f LH_draft.fa out2.bam bcftools view -bvcg - > uniq.var.bcf	2016-09-12 17:26:55	2016-09-12 17:28:42		View		
	Filter BCF and Convert to VCF File [For Unique SAM] bcftools view uniq.var.bcf perl vcfutils.pl varFilter -D10000 > out-unique.var.flt.vcf	2016-09-12 17:28:44	2016-09-12 17:28:55			Download(2.2 KB)	MD5
	Mpileup and Create BCF File samtools mpileup -u -C50 -BQ0 -d10000000 -f LH_draft.fa	2016-09-12 17:29:06	2016-09-12 17:30:58		View		



W15-3:ダウンロード

md5sumコマンドで確認。②から⑥で提示されている文字列と同じであればOK。第3回W12、第7回W9-3にもあり

The screenshot shows a web browser window displaying a pipeline detail view for 'LH_draft.fa'. The browser address bar shows 'http://p.ddbj.nig.ac.jp/pipeline/DetailView.do?query_set_id=24152'. The main content area contains a table with columns: Command, Start time, End time, Log1, Log2, Result, and MD5. A terminal window is overlaid on the browser, showing the execution of 'md5sum *.zip' and its output. The output is highlighted in a red box. Red arrows and numbers 1-6 point to specific elements in both the browser and terminal.

Command	Start time	End time	Log1	Log2	Result	MD5
Create BWA Index File bwa index [-a is] LH_draft.fa	2016-09-12 17:16:15	2016-09-12 17:16:48		View		
Filter BCF and Convert to VCF File [For Unique SAM] bcftools view uniq.var.bcf perl vcfutils.pl varFilter -D10000 > out-unique.var.fit.vcf	2016-09-12 17:28:44	2016-09-12 17:28:55			Download(2.2 KB)	MD5
Mpileup and Create BCF File samtools mpileup -u -C50 -BQ0 -d10000000 -f LH_draft.fa	2016-09-12 17:29:06	2016-09-12 17:30:58		View		

```
iu@bielinux[mac_share] pwd [ 7:51午後 ]
/home/iu/Desktop/mac_share
iu@bielinux[mac_share] ls [ 7:51午後 ]
LH_draft.fa out.sam.zip
LH_hgap.fa out-unique.var.flt.vcf.zip
out2.bam.bai.zip sequence1_blast.xml
out2.bam.zip uniqout.sam.zip
iu@bielinux[mac_share] md5sum *.zip [ 7:51午後 ]
83c0de09f4d1a32237058c1a0631b11b out2.bam.bai.zip
c71e4beb6c545d0b7c9d8ab5fc03496a out2.bam.zip
90052b9dced3bfd8393cf30f82f5b2a7 out.sam.zip
8f5c3c8877ecc7d15557bd9dd3581acd out-unique.var.flt.vcf.zip
2a18d77b4177bd92d0acdb584978bd37 uniqout.sam.zip
iu@bielinux[mac_share] [ 7:51午後 ]
```

W16-1: 解凍

```
iu@bielinux[mac_share] pwd [ 7:56午後 ]
/home/iu/Desktop/mac_share
iu@bielinux[mac_share] ls -l *.zip [ 7:56午後 ]
-rwxrwxrwx 1 iu iu 3267 9月 13 20:03 out2.bam.bai.zip
-rwxrwxrwx 1 iu iu 71322183 9月 13 20:02 out2.bam.zip
-rwxrwxrwx 1 iu iu 116389170 9月 13 19:42 out.sam.zip
-rwxrwxrwx 1 iu iu 2242 9月 13 20:04 out-unique.var.flt.vcf.zip
-rwxrwxrwx 1 iu iu 96393307 9月 13 20:02 uniqout.sam.zip
① iu@bielinux[mac_share] unzip out.sam.zip [ 7:56午後 ]
Archive: out.sam.zip
inflating: out.sam
② iu@bielinux[mac_share] unzip uniqout.sam.zip [ 7:56午後 ]
Archive: uniqout.sam.zip
inflating: uniqout.sam
③ iu@bielinux[mac_share] unzip out2.bam.zip [ 7:56午後 ]
Archive: out2.bam.zip
inflating: out2.bam
④ iu@bielinux[mac_share] unzip out2.bam.bai.zip [ 7:56午後 ]
Archive: out2.bam.bai.zip
inflating: out2.bam.bai
⑤ iu@bielinux[mac_share] unzip out-unique.var.flt.vcf.zip [ 7:56午後 ]
Archive: out-unique.var.flt.vcf.zip
inflating: out-unique.var.flt.vcf
iu@bielinux[mac_share] [ 7:57午後 ]
```

W16-1: 解凍

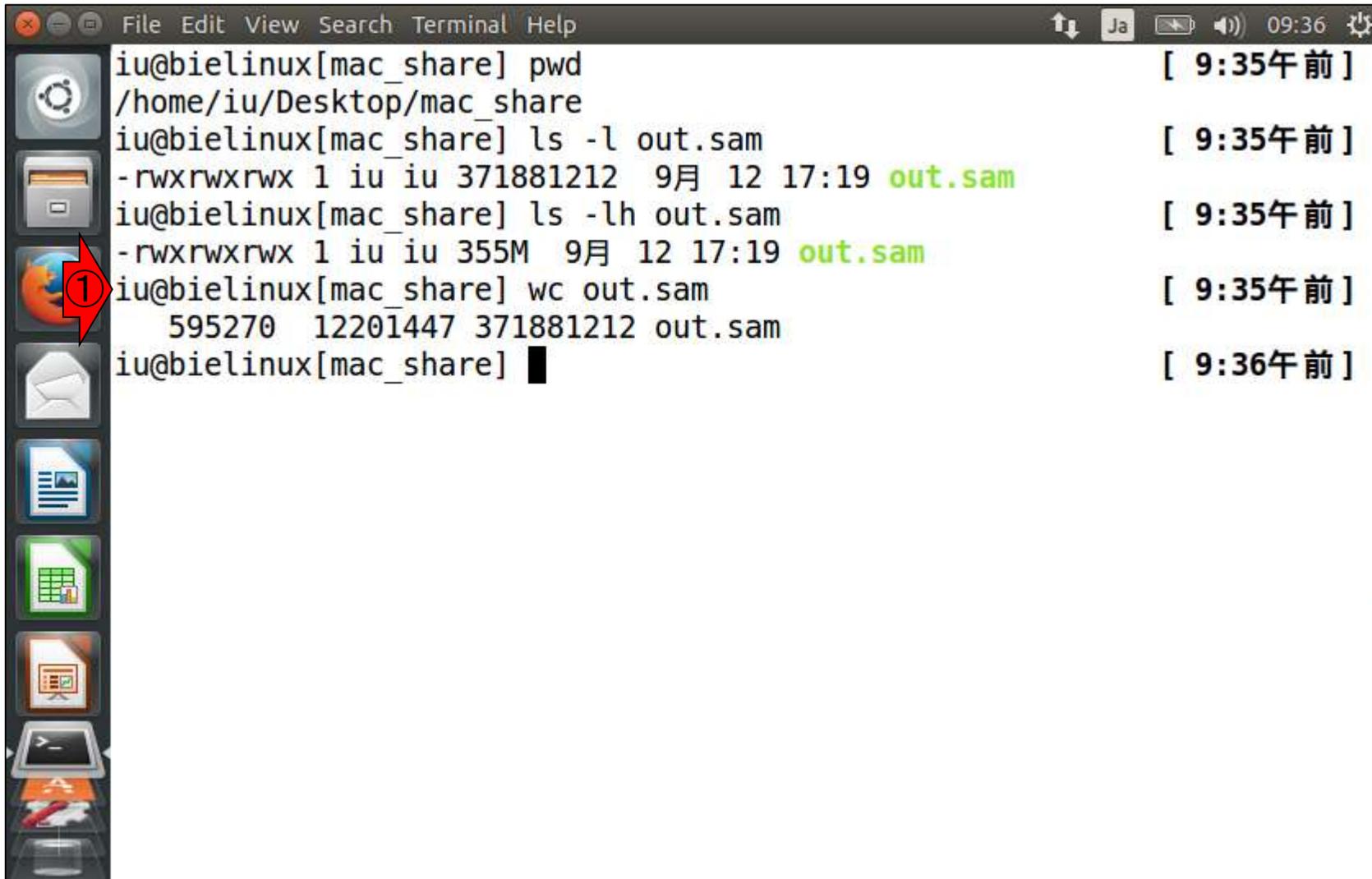
①解凍後にファイルサイズを確認。全体的に圧縮後に2-3倍程度になっている。②out2.bamのみ、③解凍前のout2.bam.zipとサイズがほとんど変わっていないが、out2.bam.zipのチェックサム(md5sum)値が提供元と同じであることを確認済みなので事実を素直に受け止めるのみ

```
iu@bielinux[mac_share] unzip uniqout.sam.zip
Archive:  uniqout.sam.zip
  inflating:  uniqout.sam
iu@bielinux[mac_share] unzip out2.bam.zip
Archive:  out2.bam.zip
  inflating:  out2.bam
iu@bielinux[mac_share] unzip out2.bam.bai.zip
Archive:  out2.bam.bai.zip
  inflating:  out2.bam.bai
iu@bielinux[mac_share] unzip out-unique.var.flt.vcf.zip
Archive:  out-unique.var.flt.vcf.zip
  inflating:  out-unique.var.flt.vcf
iu@bielinux[mac_share] ls -l out* uniq*
-rwxrwxrwx 1 iu iu 71310955  9月 12 17:26 out2.bam
-rwxrwxrwx 1 iu iu  7456  9月 12 17:26 out2.bam.bai
-rwxrwxrwx 1 iu iu  3267  9月 13 20:03 out2.bam.bai.zip
-rwxrwxrwx 1 iu iu 71322183  9月 13 20:02 out2.bam.zip
-rwxrwxrwx 1 iu iu 371881212  9月 12 17:19 out.sam
-rwxrwxrwx 1 iu iu 116389170  9月 13 19:42 out.sam.zip
-rwxrwxrwx 1 iu iu  6270  9月 12 17:28 out-unique.var.flt.vcf
-rwxrwxrwx 1 iu iu  2242  9月 13 20:04 out-unique.var.flt.vcf.zip
-rwxrwxrwx 1 iu iu 320957638  9月 12 17:23 uniqout.sam
-rwxrwxrwx 1 iu iu 96393307  9月 13 20:02 uniqout.sam.zip
iu@bielinux[mac_share]
```



①out.samファイル(371,881,212 bytes; 355MB)は、595,270行

W16-2: out.sam



A terminal window titled "Terminal" with a menu bar (File, Edit, View, Search, Terminal, Help) and a system tray (Ja, 09:36). The terminal shows the following commands and output:

```
iu@bielinux[mac_share] pwd [ 9:35午前 ]
/home/iu/Desktop/mac_share
iu@bielinux[mac_share] ls -l out.sam [ 9:35午前 ]
-rwxrwxrwx 1 iu iu 371881212  9月 12 17:19 out.sam
iu@bielinux[mac_share] ls -lh out.sam [ 9:35午前 ]
-rwxrwxrwx 1 iu iu 355M  9月 12 17:19 out.sam
iu@bielinux[mac_share] wc out.sam [ 9:35午前 ]
 595270 12201447 371881212 out.sam
iu@bielinux[mac_share] [ 9:36午前 ]
```

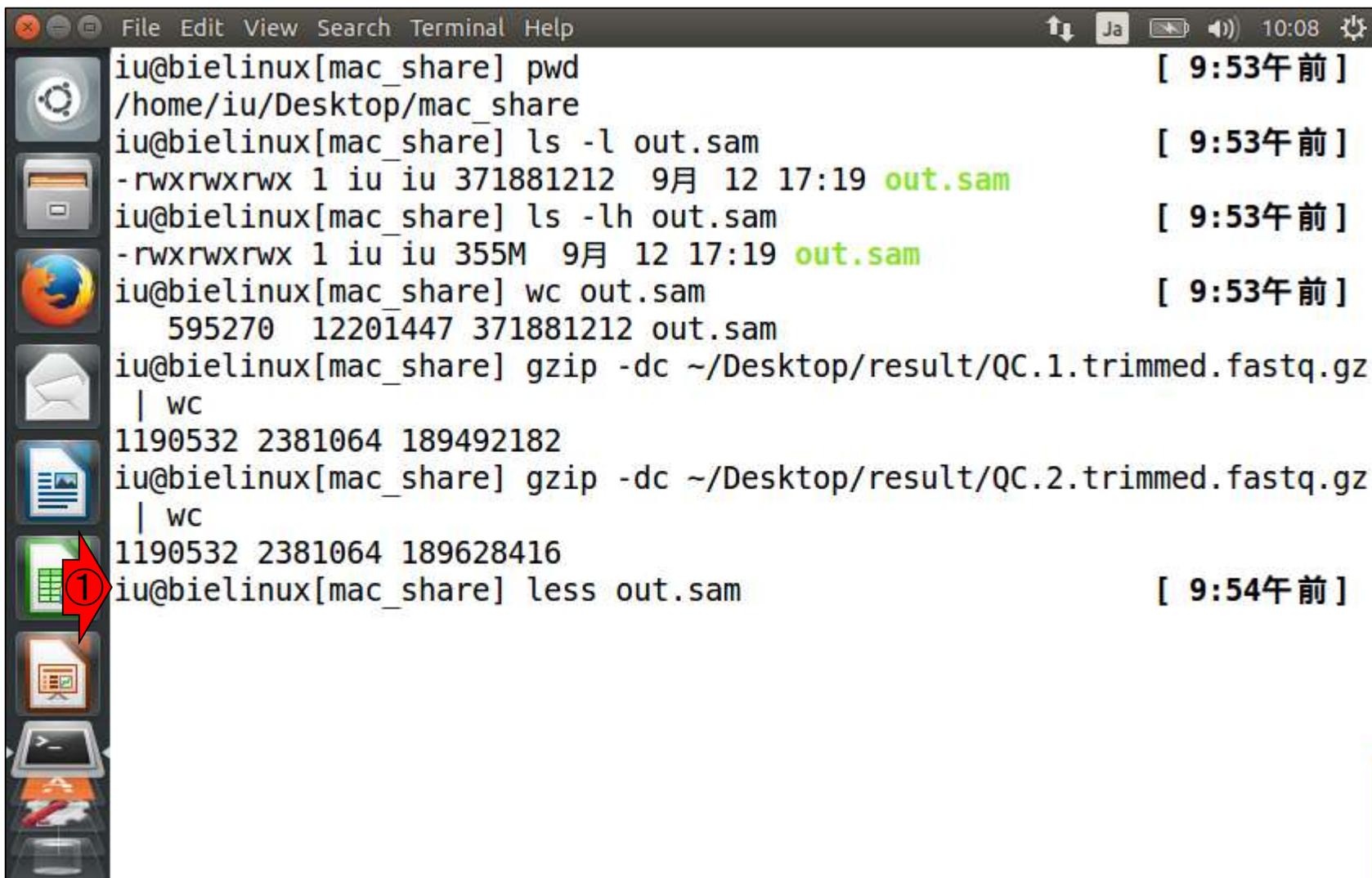
A red arrow with the number "1" points to the "wc out.sam" command line.

W16-3: マップする側

マップする側のファイル(QC.*.trimmed.fastq.gz)の行数をおさらい。1,190,532行。FASTQ形式なので $1,190,532 / 4 = 297,633$ リード。①②gzipのdオプションが解凍、cオプションが元ファイルを残す指定

```
iu@bielinux[mac_share] pwd [ 9:53午前 ]
/home/iu/Desktop/mac_share
iu@bielinux[mac_share] ls -l out.sam [ 9:53午前 ]
-rwxrwxrwx 1 iu iu 371881212  9月 12 17:19 out.sam
iu@bielinux[mac_share] ls -lh out.sam [ 9:53午前 ]
-rwxrwxrwx 1 iu iu 355M  9月 12 17:19 out.sam
iu@bielinux[mac_share] wc out.sam [ 9:53午前 ]
 595270 12201447 371881212 out.sam
① iu@bielinux[mac_share] gzip -dc ~/Desktop/result/QC.1.trimmed.fastq.gz
 | wc
1190532 2381064 189492182
② iu@bielinux[mac_share] gzip -dc ~/Desktop/result/QC.2.trimmed.fastq.gz
 | wc
1190532 2381064 189628416
iu@bielinux[mac_share] █ [ 9:54午前 ]
```

W16-4: less out.sam



```
iu@bielinux[mac_share] pwd [ 9:53午前 ]
/home/iu/Desktop/mac_share
iu@bielinux[mac_share] ls -l out.sam [ 9:53午前 ]
-rwxrwxrwx 1 iu iu 371881212  9月 12 17:19 out.sam
iu@bielinux[mac_share] ls -lh out.sam [ 9:53午前 ]
-rwxrwxrwx 1 iu iu 355M  9月 12 17:19 out.sam
iu@bielinux[mac_share] wc out.sam [ 9:53午前 ]
 595270 12201447 371881212 out.sam
iu@bielinux[mac_share] gzip -dc ~/Desktop/result/QC.1.trimmed.fastq.gz
| wc
1190532 2381064 189492182
iu@bielinux[mac_share] gzip -dc ~/Desktop/result/QC.2.trimmed.fastq.gz
| wc
1190532 2381064 189628416
① iu@bielinux[mac_share] less out.sam [ 9:54午前 ]
```


行番号も表示させたいときは、①
less実行時にNオプションをつける

W16-4: lessで眺める

```
iu@bielinux[mac_share] pwd [10:37午前]
/home/iu/Desktop/mac_share
iu@bielinux[mac_share] ls -l out.sam [10:37午前]
-rwxrwxrwx 1 iu iu 371881212  9月 12 17:19 out.sam
iu@bielinux[mac_share] ls -lh out.sam [10:37午前]
-rwxrwxrwx 1 iu iu 355M  9月 12 17:19 out.sam
iu@bielinux[mac_share] wc out.sam [10:37午前]
 595270 12201447 371881212 out.sam
iu@bielinux[mac_share] gzip -dc ~/Desktop/result/QC.1.trimmed.fastq.gz
| wc
1190532 2381064 189492182
iu@bielinux[mac_share] gzip -dc ~/Desktop/result/QC.2.trimmed.fastq.gz
| wc
1190532 2381064 189628416
iu@bielinux[mac_share] less out.sam [10:39午前]
iu@bielinux[mac_share] less -N out.sam [10:39午前]
```



W16-4: lessで眺める

```

iu@bielinux[~/Desktop/mac_share]
1 @SQ      SN:chromosome   LN:2277981
2 @SQ      SN:plasmid1     LN:81630
3 @SQ      SN:plasmid2     LN:40973
4 @PG      ID:bwa   PN:bwa   VN:0.6.1-r104
5 DRR024501.1      83      chromosome      565592  60      251M
5      =      565460  -383      CGCCACTCCAAGCGTCTGCATCCCATTACTAAAC
5 AATAAATGGCGAGGCTCATAACTTCGGAAGTGCCGGGCCATCCAAATATAATGGCTCCCCAC
5 ATCAACCGAAACGGTCATTTTATCATCAACTCGTTGCTGCAAAGCTTCAATAACAGCTAATG
5 GGTGGTTTAGATTACCCTTGCCGGCTGGAATGGTTTCGTGTTGATCAGCTTCAATTTAGTA
5 TGCAAATCTTGTAATACTGTTTCGATNCAT      888EEA??80);?EA*/?)AA8:AEE
5 ?/*A?00**/*?:8E??EEEEC::*CECC?EEE:*AE?;8;48*CCAACC:ECACC:A:*??
5 ;?;8?:?:EE?E?EEEEEEEEEB?;;;EEEEEEEEEEEEEEEEEEEEFFHHFFHHFFHH
5 HFHFHHHHHDDFHGHHHHHHHFC5HFFHHHHHHHHHHHHHHHHHHHEHFF@HHHHHHHHHHHHHHF
5 FHHHHGHGDFHHHHHHHFFACFB?BBBA<B?<55#???      XT:A:U  NM:i:2  SM
5 :i:37 AM:i:37 X0:i:1  X1:i:0  XM:i:2  X0:i:0  XG:i:0  MD:Z:180
5 A66T3
6 DRR024501.1      163     chromosome      565460  60      251M
6      =      565592  383      ATGAACCGTATTAAGGCCTAAACGAACGGCTGTC
6 TCCAGTTCTTGTCCAGTAAATAAGAATCCGGCATCCCCAGAAACAGAGACTGATTTAGCATT
6 GGGCCGAACACTAACGCAGCCGAAATTGACCAAGGTAGCGCCACTCCAAGCGTCTGCATCCCAT
6 TACTAAACAATAAATGGCGAGGCTCATAACTTCGGAAGTGCCGGGCCATCCAAATATAATGG
out.sam
  
```

W16-4: lessで眺める

例えば①の赤枠内が5行目の情報。
この画面上では、おそらく6行目の途中までの情報が表示されています

```
iu@bielinux[~/Desktop/mac_share]
1 @SQ      SN:chromosome  LN:2277981
2 @SQ      SN:plasmid1    LN:81630
3 @SQ      SN:plasmid2    LN:40973
4 @PG      ID:bwa  PN:bwa  VN:0.6.1-r104
5 DRR024501.1  83      chromosome  565592  60      251M
5      =      565460  -383      CGCCACTCCAAGCGTCTGCATCCCATTACTAAAC
5 AATAAATGGCGAGGCTCATAACTTCGGAAGTGCCGGGCCATCCAAATATAATGGCTCCCCAC
5 ATCAACCGAAACGGTCATTTTATCATCAACTCGTTGCTGCAAAGCTTCAATAACAGCTAATG
5 GGTGGTTTAGATTACCCTTGCCGGCTGGAATGGTTTCGTGTTGATCAGCTTCAATTTAGTA
5 TGCAAATCTTGTAATACTGTTTCGATNCAT      888EEA??80);?EA*:/)AA8:AEE
5 ?/*A?00**/*?:8E??EEEEC::*CECC?EEE:*AE?;8;48*CCAACC:ECACC:A:*??
5 ;?;8?:?:EE?E?EEEEEEEEEB?;;;EEEEEEEEEEEEEEEEEEEEFFFFFFFFFFFHFFHHFFHH
5 HFHFHHHHHDDFHGHHHHHHHFC5HFFHHHHHHHHHHHHHHHHHHHEHFF@HHHHHHHHHHHHHHHF
5 FHHHHGHGDFHHHHHHHFFFACFB?BBBA<B?<55#???      XT:A:U  NM:i:2  SM
5 :i:37 AM:i:37 X0:i:1  X1:i:0  XM:i:2  X0:i:0  XG:i:0  MD:Z:180
5 A66T3
6 DRR024501.1  163     chromosome  565460  60      251M
6      =      565592  383      ATGAACCGTATTAAGGCCTAAACGAACGGCTGTC
6 TCCAGTTCTTGTCCAGTAAATAAGAATCCGGCATCCCCAGAAACAGAGACTGATTTAGCATT
6 GGGCCGAACCTAACGCAGCCGAAATTGACCAAGGTAGCGCCACTCCAAGCGTCTGCATCCCAT
6 TACTAAACAATAAATGGCGAGGCTCATAACTTCGGAAGTGCCGGGCCATCCAAATATAATGG
```



out sam

W16-4: lessで眺める

今はout.samを眺めることで、SAM形式を学んでいます。①@からはじまる行は、ヘッダー行です。out.samの場合は4行ですが、ヘッダーの行数はマップされる側の配列数に依存します

```
iu@bielinux[~/Desktop/mac_share]
1 @SQ      SN:chromosome  LN:2277981
2 @SQ      SN:plasmid1    LN:81630
3 @SQ      SN:plasmid2    LN:40973
4 @PG      ID:bwa  PN:bwa  VN:0.6.1-r104
5 DRR024501.1      83      chromosome      565592  60      251M
5      =      565460  -383      CGCCACTCCAAGCGTCTGCATCCCATTACTAAAC
5 AATAAATGGCGAGGCTCATAACTTCGGAAGTGCCGGGCCATCCAAATATAATGGCTCCCCAC
5 ATCAACCGAAACGGTCATTTTATCATCAACTCGTTGCTGCAAAGCTTCAATAACAGCTAATG
5 GGTGGTTTAGATTACCCTTGCCGGCTGGAATGGTTTCGTGTTGATCAGCTTCAATTTAGTA
5 TGCAAATCTTGTAATACTGTTTCGATNCAT      888EEA??80);?EA*/?)AA8:AEE
5 ?/*A?00**/*?:8E??EEEEC::*CECC?EEE:*AE?;8;48*CCAACC:ECACC:A:*??
5 ;?;8?:?:EE?E?EEEEEEEEEB?;;;EEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE
5 HFHFHHHHHDDFHGHHHHHHHFC5HFFHHHHHHHHHHHHHHHHHHHEHHF@HHHHHHHHHHHHHHHF
5 FHHHHGHGDFHHHHHHHFFACFB?BBBA<B?<55#???      XT:A:U  NM:i:2  SM
5 :i:37 AM:i:37 X0:i:1  X1:i:0  XM:i:2  X0:i:0  XG:i:0  MD:Z:180
5 A66T3
6 DRR024501.1      163     chromosome      565460  60      251M
6      =      565592  383      ATGAACCGTATTAAGGCCTAAACGAACGGCTGTC
6 TCCAGTTCTTGTCCAGTAAATAAGAATCCGGCATCCCCAGAAACAGAGACTGATTTAGCATT
6 GGGCCGAACACTAACGCAGCCGAAATTGACCAAGGTAGCGCCACTCCAAGCGTCTGCATCCCAT
6 TACTAAACAATAAATGGCGAGGCTCATAACTTCGGAAGTGCCGGGCCATCCAAATATAATGG
```



out.sam

W16-4: lessで眺める

配列数依存の意味は、ここを見ればわかる。
。description部分をchromosome, plasmid1, plasmid2としたリファレンス配列(LH_draft.fa)作成時の手順を思い出そう(W11-2)

iu@bielinux[~/Desktop/mac_share]

```
1 @SQ      SN:chromosome  LN:2277981
2 @SQ      SN:plasmid1   LN:81630
3 @SQ      SN:plasmid2   LN:40973
4 @PG      ID:bwa    PN:bwa    VN:0.6.1-r104
5 DRR024501.1      83      chromosome      565592  60      251M
5      =      565460  -383      CGCCACTCCAAGCGTCTGCATCCCATTACTAAAC
5 AATAAATGGCGAGGCTCATAACTTCGGAAGTGCCGGGCCATCCAAATATAATGGCTCCCCAC
5 ATCAACCGAAACGGTCATTTTATCATCAACTCGTTGCTGCAAAGCTTCAATAACAGCTAATG
5 GGTGGTTTAGATTACCCTTGCCGGCTGGAATGGTTTCGTGTTGATCAGCTTCAATTTAGTA
5 TGCAAATCTTGTAATACTGTTTCGATNCAT      888EEA??80);?EA*:/)AA8:AEE
5 ?/*A?00**/*?:8E??EEEEC::*CECC?EEE:*AE?;8;48*CCAACC:ECACC:A:*??
5 ;?;8?:?:EE?E?EEEEEEEEEB?;;;EEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE
5 HFHFHHHHHDDFHGHHHHHHFC5HFFHHHHHHHHHHHHHHHHHHHHHEHFF@HHHHHHHHHHHHHHHF
5 FHHHHGHGDFHHHHHHHFFACFB?BBBA<B?<55#???      XT:A:U  NM:i:2  SM
5 :i:37 AM:i:37 X0:i:1  X1:i:0  XM:i:2  X0:i:0  XG:i:0  MD:Z:180
5 A66T3
6 DRR024501.1      163      chromosome      565460  60      251M
6      =      565592  383      ATGAACCGTATTAAGGCCTAAACGAACGGCTGTC
6 TCCAGTTCTTGTCCAGTAAATAAGAATCCGGCATCCCCAGAAACAGAGACTGATTTAGCATT
6 GGGCCGAACCTAACGCAGCCGAAATTGACCAAGGTAGCGCCACTCCAAGCGTCTGCATCCCAT
6 TACTAAACAATAAATGGCGAGGCTCATAACTTCGGAAGTGCCGGGCCATCCAAATATAATGG
```



out.sam

W16-4: lessで眺める

この数値は配列ごとの塩基数です。
 $2,277,981 + 81,630 + 40,973 = 2,400,584$
がゲノムサイズに相当。この数値は…

```
iu@bielinux[~/Desktop/mac_share]
1 @SQ      SN:chromosome  LN:2277981
2 @SQ      SN:plasmid1    LN:81630
3 @SQ      SN:plasmid2    LN:40973
4 @PG      ID:bwa  PN:bwa  VN:0.6.1-r104
5 DRR024501.1      83      chromosome      565592  60      251M
5      =      565460  -383      CGCCACTCCAAGCGTCTGCATCCCATTACTAAAC
5 AATAAATGGCGAGGCTCATAACTTCGGAAGTGCCGGGCCATCCAAATATAATGGCTCCCCAC
5 ATCAACCGAAACGGTCATTTTATCATCAACTCGTTGCTGCAAAGCTTCAATAACAGCTAATG
5 GGTGGTTTAGATTACCCTTGCCGGCTGGAATGGTTTCGTGTTGATCAGCTTCAATTTAGTA
5 TGCAAATCTTGTAATACTGTTTCGATNCAT      888EEA??80);?EA*/?)AA8:AEE
5 ?/*A?00**/*?:8E??EEEEC::*CECC?EEE:*AE?;8;48*CCAACC:ECACC:A:*??
5 ;?;8?:?:EE?E?EEEEEEEEEB?;;;EEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE
5 HFHFHHHHHDDFHGHHHHHHHFC5HFFHHHHHHHHHHHHHHHHHHHEHFF@HHHHHHHHHHHHHHHF
5 FHHHHGHGDFHHHHHHHFFACFB?BBBA<B?<55#???      XT:A:U  NM:i:2  SM
5 :i:37 AM:i:37 X0:i:1  X1:i:0  XM:i:2  X0:i:0  XG:i:0  MD:Z:180
5 A66T3
6 DRR024501.1      163      chromosome      565460  60      251M
6      =      565592  383      ATGAACCGTATTAAGGCCTAAACGAACGGCTGTC
6 TCCAGTTCTTGTCCAGTAAATAAGAATCCGGCATCCCCAGAAACAGAGACTGATTTAGCATT
6 GGGCCGAACCTAACGCAGCCGAAATTGACCAAGGTAGCGCCACTCCAAGCGTCTGCATCCCAT
6 TACTAAACAATAAATGGCGAGGCTCATAACTTCGGAAGTGCCGGGCCATCCAAATATAATGG
out.sam
```



W16-4: lessで眺める

DDBJ Pipeline実行結果画面上の、
①と同じことから納得できるでしょう

JOB STATUS

- step1. Preprocessing
- step1. Mapping
- step1. *de novo* Assembly
- step2-All status

HELP

- HELP
- TUTORIAL
- Contact Us. DDBJ Read Annotation Pipeline. Development Team.

Download merged pileup file

Uniq.sam files have been merged if you specified 'uniq' option.

- [merged.var.fit.vcf.gz \(Original size 6.2 KB\)](#)
- [merged.sam.gz \(Original size 320.9 MB\)](#)

Download wgs file

- [out_WGS.fasta.gz \(Original size 197 byte\)](#)

Position errors **Map ratio** **Depth, Coverage** ①

PDF download total query # : 297,633 coverage : 2400552 / 2400584 * 100 = 99.999
mapped query # : 281,303 depth : 130565653 / 2400552 = 54.390
map ratio : 94.513 %

Time

Wait time	Start time	End time
0: 0:18	2016-09-12 17:16:14	2016-09-12 17:37:57

LH_draft.fa

Command	Start time	End time	Log1	Log2	Result	MD5
Create BWA Index File bwa index [-a is] LH_draft.fa	2016-09-12 17:16:15	2016-09-12 17:16:48		View		
BWA : Alignment bwa aln LH_draft.fa QC.1.trimmed.fastq > 1.sai	2016-09-12 17:16:48	2016-09-12 17:17:10		View		
BWA : Alignment bwa aln LH_draft.fa QC.2.trimmed.fastq > 2.sai	2016-09-12 17:17:11	2016-09-12 17:17:33		View		
BWA : SAMPE bwa sampe LH_draft.fa 1.sai 2.sai QC.1.trimmed.fastq QC.2.trimmed.fastq > out.sam	2016-09-12 17:17:34	2016-09-12 17:19:43		View	Download(116.3 MB)	
Extract Unmapped Reads	2016-09-12	2016-09-12				

Top of page

W16-4: lessで眺める

①はマッピングに用いたプログラム、およびバージョン情報が記載されています

```
iu@bielinux[~/Desktop/mac_share] 10:41
1 @SQ      SN:chromosome   LN:2277981
2 @SQ      SN:plasmid1     LN:81630
3 @SQ      SN:plasmid2     LN:40973
4 @PG      ID:bwa          PN:bwa          VN:0.6.1-r104
5 DRR024501.1 83 chromosome 565592 60 251M
5 = 565460 -383 CGCCACTCCAAGCGTCTGCATCCCATTACTAAAC
5 AATAAATGGCGAGGCTCATAACTTCGGAAGTGCCGGGCCATCCAAATATAATGGCTCCCCAC
5 ATCAACCGAAACGGTCATTTTATCATCAACTCGTTGCTGCAAAGCTTCAATAACAGCTAATG
5 GGTGGTTTAGATTACCCTTGCCGGCTGGAATGGTTTCGTGTTGATCAGCTTCAATTTAGTA
5 TGCAAATCTTGTAATACTGTTTCGATNCAT 888EEA??80);?EA*/?)AA8:AEE
5 ?/*A?00**/*?:8E??EEEEC::*CECC?EEE:*AE?;8;48*CCAACC:ECACC:A:*??
5 ;?;8?:?:EE?E?EEEEEEEEEB?;;;EEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE
5 HFHFHHHHHDDFHGHHHHHHHFC5HFFHHHHHHHHHHHHHHHHHHHEHFF@HHHHHHHHHHHHHHHF
5 FHHHHGHGDFHHHHHHHFFACFB?BBBA<B?<55#??? XT:A:U NM:i:2 SM
5 :i:37 AM:i:37 X0:i:1 X1:i:0 XM:i:2 X0:i:0 XG:i:0 MD:Z:180
5 A66T3
6 DRR024501.1 163 chromosome 565460 60 251M
6 = 565592 383 ATGAACCGTATTAAGGCCTAAACGAACGGCTGTC
6 TCCAGTTCTTGTCCAGTAAATAAGAATCCGGCATCCCCAGAAACAGAGACTGATTTAGCATT
6 GGGCCGAACACTAACGCAGCCGAAATTGACCAAGGTAGCGCCACTCCAAGCGTCTGCATCCCAT
6 TACTAAACAATAAATGGCGAGGCTCATAACTTCGGAAGTGCCGGGCCATCCAAATATAATGG
```

out.sam



W16-5:5行目以降

①5行目の最初、②6行目の最初をみると、同じリードIDになっていることがわかる。並びは、マップする側のpaired-endファイルのリードの順番。リードIDの「1番目のforward側reverse側」、「2番目のforward側reverse側」、のような感じで、2行で1つのリードペアを表している

```
iu@bielinux[~/Desktop/mac_share]
1 @SQ      SN:chromosome  LN:2277981
2 @SQ      SN:plasmid1    LN:81630
3 @SQ      SN:plasmid2    LN:40973
4 @PG      ID:bwa  PN:bwa  VN:0.6.1-r104
① 5 DRR024501.1      83      chromosome      565592  60      251M
5      =      565460  -383      CGCCACTCCAAGCGTCTGCATCCCATTACTAAAC
5 AATAAATGGCGAGGCTCATAACTTCGGAAGTGCCGGGCCATCCAAATATAATGGCTCCCCAC
5 ATCAACCGAAACGGTCATTTTATCATCAACTCGTTGCTGCAAAGCTTCAATAACAGCTAATG
5 GGTGGTTTAGATTACCCTTGCCGGCTGGAATGGTTTCGTGTTGATCAGCTTCAATTTTCAGTA
5 TGCAAATCTTGTAATACTGTTTCGATNCAT      888EEA??80);?EA*:/)AA8:AEE
5 ?/*A?00**/*?:8E??EEEEC::*CECC?EEE:*AE?;8;48*CCAACC:ECACC:A:*??
5 ;?;8?:?:EE?E?EEEEEEEEEB?;;;EEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE
5 HFHFHHHHHDDFHGHHHHHHHFC5HFFHHHHHHHHHHHHHHHHHHHEHFF@HHHHHHHHHHHHHHHF
5 FHHHHGHGDFHHHHHHHFFACFB?BBBA<B?<55#???      XT:A:U  NM:i:2  SM
5 :i:37 AM:i:37 X0:i:1  X1:i:0  XM:i:2  X0:i:0  XG:i:0  MD:Z:180
5 A66T3
② 6 DRR024501.1      163     chromosome      565460  60      251M
6      =      565592  383      ATGAACCGTATTAAGGCCTAAACGAACGGCTGTC
6 TCCAGTTCTTGTCCAGTAAATAAGAATCCGGCATCCCCAGAAACAGAGACTGATTTAGCATT
6 GGGCCGAACCTAACGCAGCCGAAATTGACCAAGGTAGCGCCACTCCAAGCGTCTGCATCCCAT
6 TACTAAACAATAAATGGCGAGGCTCATAACTTCGGAAGTGCCGGGCCATCCAAATATAATGG
out.sam
```

リードIDの「1番目の①forward側②reverse側」
を表示。確かに同じリードID

W16-5:5行目以降

```
iu@bielinux[mac_share] gzip -dc ~/Desktop/result/QC.1.trimmed.fastq.gz
| head -n 4
@DRR024501.1 M00278:15:000000000-A2RK1:1:1101:18783:2260 length=251
ATGNATCGAAACAGTATTTACAAGATTTGCATACTGAAATTGAAGCTGATCAACACGAAACCATTCCAGC
CGGCAAGGGTAATCTAAACCACCCATTAGCTGTTATTGAAGCTTTGCAGCAACGAGTTGATGATAAAATG
ACCGTTTCGGTTGATGTGGGGAGCCATTATATTTGGATGGCCCGGCACTTCCGAAGTTATGAGCCTCGCC
ATTTATTGTTTAGTAATGGGATGCAGACGCTTGGAGTGGCG
+DRR024501.1 M00278:15:000000000-A2RK1:1:1101:18783:2260 length=251
???.#55<?B<ABBB?BFCAFFFHHHHHHFFDGHGHHHHFFHHHHHHHHHHHHHHH@FHHEHHHHHHHHHHHH
HHHHHHFFH5CFHHHHHHGHFDDHHHHFHFHHHFFHHHFFHHFFHHFFHHFFHHFFHHFFHHFFHHFFHHFF
?BEEEEEEEE?E?EE:?:?8;?;??*:A:CCACE:CCAACC*84;8;?EA*:EEE?CCEC*: :CEEEE??
E8:?*/**00?A*/?EEA:8AA)/:*AE?;)08??AEE888
iu@bielinux[mac_share] gzip -dc ~/Desktop/result/QC.2.trimmed.fastq.gz
| head -n 4
@DRR024501.1 M00278:15:000000000-A2RK1:1:1101:18783:2260 length=251
ATGAACCGTATTAAGGCCTAAACGAACGGCTGTCTCCAGTTCTTGTCCAGTAAATAAGAATCCGGCATCC
CCAGAAACAGAGACTGATTTAGCATTGGGCCGAACAAACGCAGCCGAAATTGACCAAGGTAGCGCCACTC
CAAGCGTCTGCATCCCATTACTAAACAATAAATGGCGAGGCTCATAACTTCGGAAGTGCCGGGCCATCCA
AATATAATGGCTCCCCACATCAAACGCAACGTTTCATATTAT
+DRR024501.1 M00278:15:000000000-A2RK1:1:1101:18783:2260 length=251
?????B?@<ABDDDDD;FFFFFHEBHACCCCCFFGGHBGFHHHBEDGHDACFHHF=?-A5AFC>E>EHH
HHHD+CFDFFHHBFGHG7CB@@DF.@FDDE)@EFFEEEE):);BEB6???:BECE;3:C?A2;;EEEE
```

W16-5:5行目以降

リードIDの「1番目の①forward側②reverse側」の塩基配列の最初の7塩基分を赤矢印で、最後の7塩基分を青矢印で示す

```
iu@bielinux[mac_share] gzip -dc ~/Desktop/result/QC.1.trimmed.fastq.gz
| head -n 4
@DRR024501.1 M00278:15:000000000-A2RK1:1:1101:18783:2260 length=251
ATGNATCGAAACAGTATTTACAAGATTTGCATACTGAAATTGAAGCTGATCAACACGAAACCATTCCAGC
CCGAAAGGGTAATCTAAACCACCCATTAGCTGTTATTGAAGCTTTGCAGCAACGAGTTGATGATAAAATG
ACCGTTTCGGTTGATGTGGGGAGCCATTATATTTGGATGGCCCGGCACTTCCGAAGTTATGAGCCTCGCC
ATTTATTGTTTAGTAATGGGATGCAGACGCTTGGAGTGGCG
+DRR024501.1 M00278:15:000000000-A2RK1:1:1101:18783:2260 length=251
???#55<?B<ABBB?BFCAFFFHHHHHHFFDGHGHHHHFFHHHHHHHHHHHHHHH@FHHEHHHHHHHHHHHH
HHHHHHFFH5CFHHHHHHGHFDDHHHHFHFFHHFFHHFFHHFFHHFFHHFFHHFFHHFFHHFFHHFFHHFF
?BEEEEEEEE?E?EE?:?:?8;?;??:A:CCACE:CCAACC*84;8;?EA*:EEE?CCEC*: :CEEEE??
E8:?*/*00?A*/?EEA:8AA)/:*AE?;)08??AEE888
iu@bielinux[mac_share] gzip -dc ~/Desktop/result/QC.2.trimmed.fastq.gz
| head -n 4
@DRR024501.1 M00278:15:000000000-A2RK1:1:1101:18783:2260 length=251
ATGAACCGTATTAAGGCCTAAACGAACGGCTGTCTCCAGTTCTTGTCCAGTAAATAAGAATCCGGCATCC
CCGAAACAGAGACTGATTTAGCATTGGGCCGAACAACTAACGCAGCCGAAATTGACCAAGGTAGCGCCACTC
CAAGCGTCTGCATCCCATTACTAAACAATAAATGGCGAGGCTCATAACTTCGGAAGTGCCGGGCCATCCA
AATATAATGGCTCCCCACATCAAACGCAACGTTTCATATTAT
+DRR024501.1 M00278:15:000000000-A2RK1:1:1101:18783:2260 length=251
?????B?@<ABDDDDD;FFFFFHEBHACCCCFGGHGBGHFHBBEDGHDACFHHF=?-A5AFC>E>EHH
HHHD+CFDFFHHBFGHG7CB@@DF.@FDDE)@EFFEEEE):);BEB6???:BECE;3:C?A2;;EEEE
```

W16-5:5行目以降

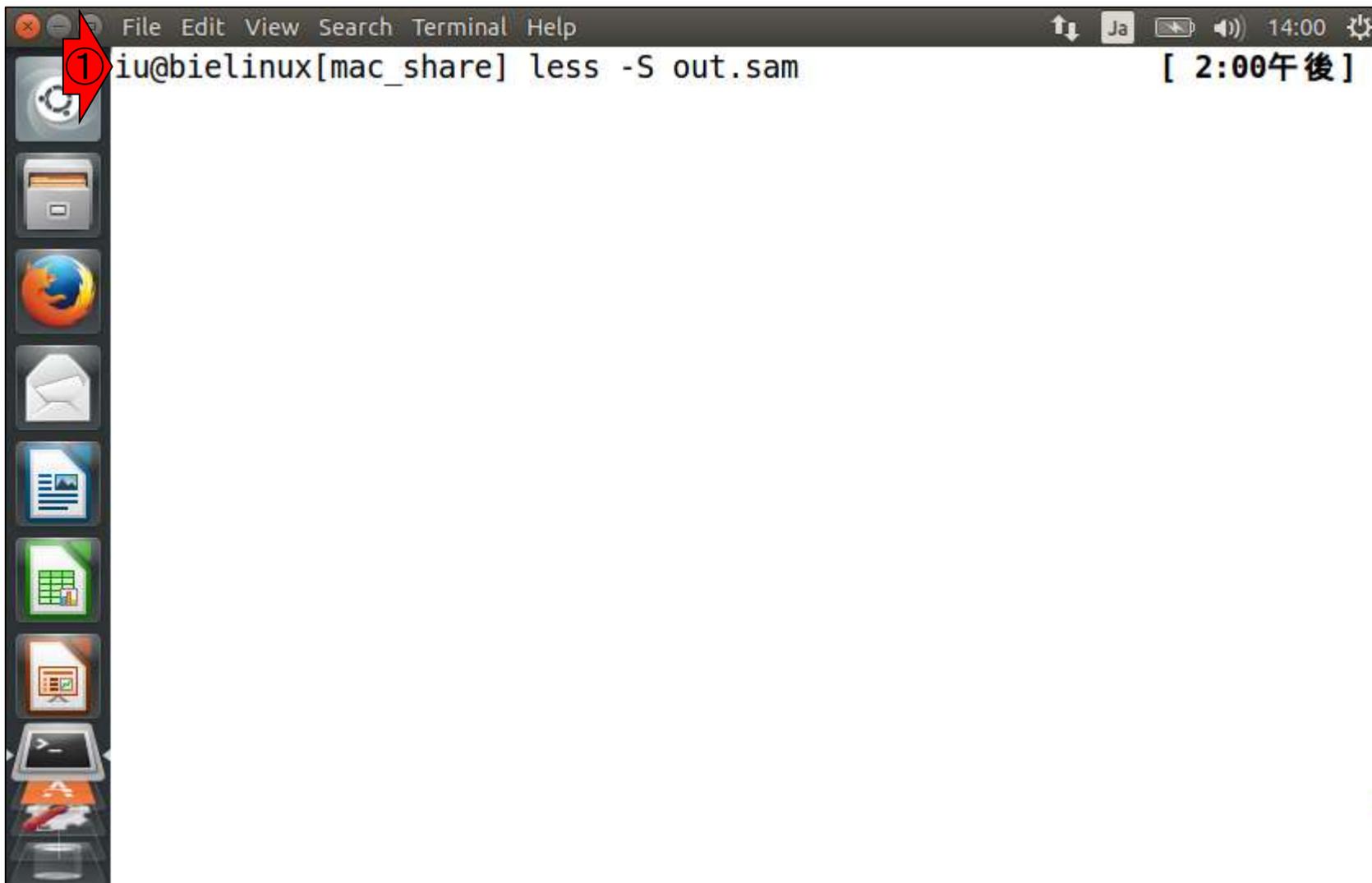
「less -N out.sam」上で、さきほどの赤矢印に対応する部分を示す。①forward側のリードは逆相補鎖(reverse complement)が、②reverse側は同じ並びの塩基配列になっていることがわかる。2本鎖の場合は、逆相補鎖がマップされるのは普通なので妥当

```
iu@bielinux[~/Desktop/mac_share]
1 @SQ      SN:chromosome  LN:2277981
2 @SQ      SN:plasmid1    LN:81630
3 @SQ      SN:plasmid2    LN:40973
4 @PG      ID:bwa  PN:bwa  VN:0.6.1-r104
5 DRR024501.1      83      chromosome      565592  60      251M
5      =      565460  -383      CGCCACTCCAAGCGTCTGCATCCCATTACTAAAC
5 AATAAATGGCGAGGCTCATAACTTCGGAAGTGCCGGGCCATCCAAATATAATGGCTCCCCAC
5 ATCAACCGAAACGGTCATTTTATCATCAACTCGTTGCTGCAAAGCTTCAATAACAGCTAATG
5 GGTGGTTTAGATTACCCTTGCCGGCTGGAATGGTTTCGTGTTGATCAGCTTCAATTTTCAGTA
5 TGCAAATCTTGTAATACTGTTTCGATNCAT      888EEA??80);?EA*/?)AA8:AEE
5 ?/*A?00**/*?:8E??EEEEC::*CECC?EEE:*AE?;8;48*CCAACC:ECACC:A:*??
5 ;?;8?:?:EE?E?EEEEEEEEEB?;;;EEEEEEEEEEEEEEEEEEEEFFFFFFFFFFFHFFHHFFHH
5 HFHFHHHHHDDFHGHHHHHHFC5HFFHHHHHHHHHHHHHHHHHHHEHHF@HHHHHHHHHHHHHHHF
5 FHHHHGHGDFHHHHHHHFFACFB?BBBA<B?<55#???      XT:A:U  NM:i:2  SM
5 :i:37 AM:i:37 X0:i:1  X1:i:0  XM:i:2  X0:i:0  XG:i:0  MD:Z:180
5 A66T3
6 DRR024501.1      163     chromosome      565460  60      251M
6      =      565592  383      ATGAACCGTATTAAGGCCTAAACGAACGGCTGTC
6 TCCAGTTCTTGTCCAGTAAATAAGAATCCGGCATCCCCAGAAACAGAGACTGATTTAGCATT
6 GGGCCGAACCTAACGCAGCCGAAATTGACCAAGGTAGCGCCACTCCAAGCGTCTGCATCCCAT
6 TACTAAACAATAAATGGCGAGGCTCATAACTTCGGAAGTGCCGGGCCATCCAAATATAATGG
out.sam
```



W16-6: less -S

less実行時にSオプションをつけると1行分を折り返さずに表示してくれます



W16-6: less -S

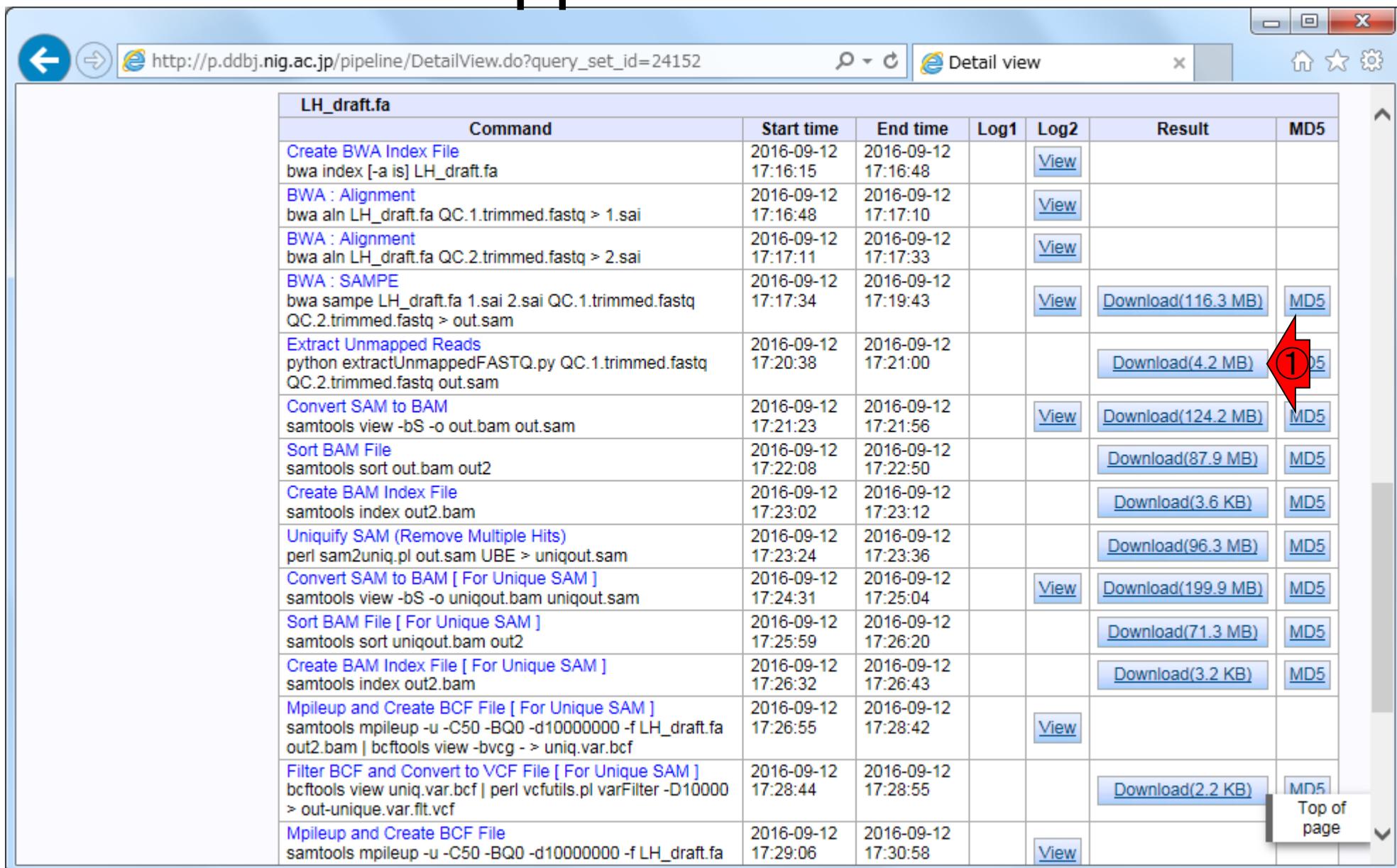
こんな感じ。1行が長いファイルで、各行の最初のほうのみ眺めたい場合に便利。SAMファイルはマップされなかったリード情報も保持している。その例が①5番目のリード。マップされなかったことを示すアスタリスク(*)で認識可能

```
File Edit View Search Terminal Help
@SQ      SN:chromosome  LN:2277981
@SQ      SN:plasmid1   LN:81630
@SQ      SN:plasmid2   LN:40973
@PG      ID:bwa       PN:bwa       VN:0.6.1-r104
DRR024501.1  83  chromosome  565592  60  251M  =
DRR024501.1  163 chromosome  565460  60  251M  =
DRR024501.2  83  chromosome  1188239 60  251M  =
DRR024501.2  163 chromosome  1188115 60  251M  =
DRR024501.3  99  chromosome  721208  29  251M  =
DRR024501.3  147 chromosome  721460  29  136S115M
DRR024501.4  99  chromosome  236084  60  184M  =
DRR024501.4  147 chromosome  236106  60  175M  =
DRR024501.5  77  *          0        0        *        *        0
DRR024501.5  141 *          0        0        *        *        0
DRR024501.6  89  chromosome  1861136 37  188M  =
DRR024501.6  181 chromosome  1861136 0   *      =
DRR024501.7  83  chromosome  2174307 29  251M  =
DRR024501.7  163 chromosome  2174200 29  107M144S
DRR024501.8  83  chromosome  1484623 60  251M  =
DRR024501.8  163 chromosome  1484466 60  251M  =
DRR024501.9  99  chromosome  621882  60  251M  =
out.sam
```



W16-7: unmapped reads

① マップされなかったリード情報のみ FASTQ形式ファイルとして得たい場合



LH_draft.fa							
Command	Start time	End time	Log1	Log2	Result	MD5	
Create BWA Index File bwa index [-a is] LH_draft.fa	2016-09-12 17:16:15	2016-09-12 17:16:48		View			
BWA : Alignment bwa aln LH_draft.fa QC.1.trimmed.fastq > 1.sai	2016-09-12 17:16:48	2016-09-12 17:17:10		View			
BWA : Alignment bwa aln LH_draft.fa QC.2.trimmed.fastq > 2.sai	2016-09-12 17:17:11	2016-09-12 17:17:33		View			
BWA : SAMPE bwa sampe LH_draft.fa 1.sai 2.sai QC.1.trimmed.fastq QC.2.trimmed.fastq > out.sam	2016-09-12 17:17:34	2016-09-12 17:19:43		View	Download(116.3 MB)	MD5	
Extract Unmapped Reads python extractUnmappedFASTQ.py QC.1.trimmed.fastq QC.2.trimmed.fastq out.sam	2016-09-12 17:20:38	2016-09-12 17:21:00			Download(4.2 MB)	MD5	①
Convert SAM to BAM samtools view -bS -o out.bam out.sam	2016-09-12 17:21:23	2016-09-12 17:21:56		View	Download(124.2 MB)	MD5	
Sort BAM File samtools sort out.bam out2	2016-09-12 17:22:08	2016-09-12 17:22:50			Download(87.9 MB)	MD5	
Create BAM Index File samtools index out2.bam	2016-09-12 17:23:02	2016-09-12 17:23:12			Download(3.6 KB)	MD5	
Uniquify SAM (Remove Multiple Hits) perl sam2uniq.pl out.sam UBE > uniqout.sam	2016-09-12 17:23:24	2016-09-12 17:23:36			Download(96.3 MB)	MD5	
Convert SAM to BAM [For Unique SAM] samtools view -bS -o uniqout.bam uniqout.sam	2016-09-12 17:24:31	2016-09-12 17:25:04		View	Download(199.9 MB)	MD5	
Sort BAM File [For Unique SAM] samtools sort uniqout.bam out2	2016-09-12 17:25:59	2016-09-12 17:26:20			Download(71.3 MB)	MD5	
Create BAM Index File [For Unique SAM] samtools index out2.bam	2016-09-12 17:26:32	2016-09-12 17:26:43			Download(3.2 KB)	MD5	
Mpileup and Create BCF File [For Unique SAM] samtools mpileup -u -C50 -BQ0 -d10000000 -f LH_draft.fa out2.bam bcftools view -bvcg - > uniq.var.bcf	2016-09-12 17:26:55	2016-09-12 17:28:42		View			
Filter BCF and Convert to VCF File [For Unique SAM] bcftools view uniq.var.bcf perl vcfutils.pl varFilter -D10000 > out-unique.var.fit.vcf	2016-09-12 17:28:44	2016-09-12 17:28:55			Download(2.2 KB)	MD5	
Mpileup and Create BCF File samtools mpileup -u -C50 -BQ0 -d10000000 -f LH_draft.fa	2016-09-12 17:29:06	2016-09-12 17:30:58		View			

W17-1: Excel

out.samをExcelで開いた状態。①QNAME (Query NAME)。リードIDのこと。②FLAG。フラグ情報(マップされたかどうかなどの情報だが、ややこしいので省略)。③RNAME (Reference NAME)。マップされたリファレンスの配列名情報。マップされなかったら*。④POS (start POSition)。マップされたリードの開始点。⑤MAPQ (MAPping Quality)。マッピングのクオリティ。高いほどよい

	A	B	C								
1	@SQ	SN:chromosome	LN:2277981								
2	@SQ	SN:plasmid1	LN:81630								
3	@SQ	SN:plasmid2	LN:40973								
4	@PG	ID:bwa	PN:bwa	VN:0.1.0	F104						
5	DRR024501.1	83	chromosome	565592	60	251M	=	565460	-383	CGCCAC888EEA???	XT:A:I
6	DRR024501.1	163	chromosome	565460	60	251M	=	565592	383	ATGAAC?????B?@<	XT:A:I
7	DRR024501.2	83	chromosome	1188239	60	251M	=	1188115	-375	TCACTT?0*)'*C?:(XT:A:I
8	DRR024501.2	163	chromosome	1188115	60	251M	=	1188239	375	TGGACC??<??B?9E	XT:A:I
9	DRR024501.3	99	chromosome	721208	29	251M	=	721460	367	CAANGA===#55<=€	XT:A:I
10	DRR024501.3	147	chromosome	721460	29	136S115M	=	721208	-367	ATTACT(//((//(((/(XT:A:I
11	DRR024501.4	99	chromosome	236084	60	184M	=	236106	197	AGANAC???#5<<?D	XT:A:I
12	DRR024501.4	147	chromosome	236106	60	175M	=	236084	-197	TTAATT ECA?.C:*?	XT:A:I
13	DRR024501.5	77	*	0	0	*	*	0	0	AGTNCC?,<#55?,<+<5?<<	
14	DRR024501.5	141	*	0	0	*	*	0	0	GATAAA,55=,<5<--5@@@	
15	DRR024501.6	89	chromosome	1861136	37	188M	=	1861136	0	TTTACC EEEEEEEFI	XT:A:I

W17-2:CIGAR

⑥CIGAR。CIGAR stringと呼ばれるもので、マッピング結果の概要を把握する情報。251Mは、251塩基が一致。136S115MのSは、Soft clippingのこと。136塩基がマップされず、115塩基が(マップされて)一致していたということ。オリジナルのリード長は136+115 = 251塩基の長さ。184Mは、184塩基が一致。入力ファイル中のリード長は(FaQCsでアダプタートリムも行っているので)一定ではなく、このリードは184塩基の長さしかなかったのだろう

	A	B	C	D
1	@SQ	SN:chromosome	LN:2277981	
2	@SQ	SN:plasmid1	LN:81630	
3	@SQ	SN:plasmid2	LN:40973	
4	@PG	ID:bwa	PN:bwa	VN:0.6.1-r104
5	DRR024501.1	83	chromosome	565592 60 251M = 565460 -383 CGCCAC888EEA???
6	DRR024501.1	163	chromosome	565460 60 251M = 565592 383 ATGAAC????B?@<
7	DRR024501.2	83	chromosome	1188239 60 251M = 1188115 -375 TCACTT?0*)?C?:(XT:A:
8	DRR024501.2	163	chromosome	1188115 60 251M = 1188239 375 TGGACC??<??B?9EXT:A:
9	DRR024501.3	99	chromosome	721208 29 251M = 721460 367 CAANGA===#55<=€XT:A:
10	DRR024501.3	147	chromosome	721460 29 136S115M = 721208 -367 ATTACT(//((//(((/(XT:A:
11	DRR024501.4	99	chromosome	236084 60 184M = 236106 197 AGANAC???#5<<?DXT:A:
12	DRR024501.4	147	chromosome	236106 60 175M = 236084 -197 TTAATT ECA?.C:*?:XT:A:
13	DRR024501.5	77	*	0 0 * * 0 0 AGTNCC?,<#55?,<+<5?<<
14	DRR024501.5	141	*	0 0 * * 0 0 GATAAA,55=,<5<--5@@@
15	DRR024501.6	89	chromosome	1861136 37 188M = 1861136 0 TTTACC EEEEEEEFI XT:A:

⑥

W17-3: RNEXTとPNEXT

⑦RNEXT。*は、マップされなかったもの。=は、paired-endでもう片方が同じリファレンス配列にマップされた場合。もし異なっていれば配列名(chromosome, plasmid1, plasmid2)となる。⑧PNEXTは、paired-endのもう片方のマップされた座標情報

	A	B	C	D	E	F					
1	@SQ	SN:chromosome	LN:2277981								
2	@SQ	SN:plasmid1	LN:81630								
3	@SQ	SN:plasmid2	LN:40973								
4	@PG	ID:bwa	PN:bw	VN:0.6.1-r104							
5	DRR024501.1	83	chromosome	565592	60	251M	=	565460	-383	CGCCAC888EEA???	EXT:A:J
6	DRR024501.1	163	chromosome	565460	60	251M	=	565592	383	ATGAAC?????B?@<	XT:A:J
7	DRR024501.2	83	chromosome	1188239	60	251M	=	1188115	-375	TCACTT?0*)'*C?:(XT:A:J
8	DRR024501.2	163	chromosome	1188115	60	251M	=	1188239	375	TGGACC??<??B?9E	XT:A:J
9	DRR024501.3	99	chromosome	721208	29	251M	=	721460	367	CAANGA===#55<=€	XT:A:J
10	DRR024501.3	147	chromosome	721460	29	136S115M	=	721208	-367	ATTACT(//((/((((/(XT:A:J
11	DRR024501.4	99	chromosome	236084	60	184M	=	236106	197	AGANAC???#5<<?D	XT:A:J
12	DRR024501.4	147	chromosome	236106	60	175M	=	236084	-197	TTAATT ECA?.C:*?;	XT:A:J
13	DRR024501.5	77	*	0	0	*	*	0	0	AGTNCC?,<#55?,<+<5?<<	
14	DRR024501.5	141	*	0	0	*	*	0	0	GATAAA,55=,<5<--5@@@	
15	DRR024501.6	89	chromosome	1861136	37	188M	=	1861136	0	TTTACC EEEEEEEFI	XT:A:J



W17-3: RNEXTとPNEXT

	A	B	C	D	E	F	G	H	I	J	K	L
1	@SQ	SN:chromosome	LN:2277981									
2	@SQ	SN:plasmid1	LN:81630									
3	@SQ	SN:plasmid2	LN:40973									
4	@PG	ID:bwa	PN:DRR024501.1	VN:0.1.104								
5	DRR024501.1	83	chromosome	565592	←60 251M		=	565460	-383	CGCCAC(888EEA???)XT:A:1		
6	DRR024501.1	163	chromosome	565460	←60 251M		=	565592	383	ATGAAC?????B?@<XT:A:1		
7	DRR024501.2	83	chromosome	1188239	←60 251M		=	1188115	-375	TCACTT(??*)?C?:(XT:A:1		
8	DRR024501.2	163	chromosome	1188115	←60 251M		=	1188239	375	TGGACC(??<??B?9E)XT:A:1		
9	DRR024501.3	99	chromosome	721208	←29 251M		=	721460	367	CAANGA===#55<=€XT:A:1		
10	DRR024501.3	147	chromosome	721460	←29 136S115M		=	721208	-367	ATTACT(//((//(((/(XT:A:1		
11	DRR024501.4	99	chromosome	236084	←60 184M		=	236106	197	AGANAC???#5<<?D)XT:A:1		
12	DRR024501.4	147	chromosome	236106	←60 175M		=	236084	-197	TTAATT ECA?.C:*?:XT:A:1		
13	DRR024501.5	77	*	0	0 *		*	0	0	AGTNCC?,<#55?,<+<5?<<		
14	DRR024501.5	141	*	0	0 *		*	0	0	GATAAA,55=,<5<--5@@@		
15	DRR024501.6	89	chromosome	1861136	37 188M		=	1861136	0	TTTACC EEEEEEEE)XT:A:1		

W17-3: RNEXTとPNI

⑦のRNEXTが=yや*でないものの例。sequence1をqueryとしてBLAST実行した結果にも、plasmid1 (sequence2)とplasmid2 (sequence3)にヒットしたものがあつたので妥当。W7-4



	A	B	C	D	E	F	G	H	I	J	K	L
8618	DRR024501.4314	163	chromosome	1352183	60	251M	=	1352317	385	CAACG/????<?<?B	XT:A:U	
8619	DRR024501.4315	99	chromosome	87242	60	251M	=	87383	392	CAACT/AAAA?BBE	XT:A:U	
8620	DRR024501.4315	147	chromosome	87383	60	251M	=	87242	-392	GCAATT*C:0'2.*:10	XT:A:U	
8621	DRR024501.4316	83	chromosome	1523909	60	251M	=	1523769	-391	ATCGAC.EGGECCE	XT:A:U	
8622	DRR024501.4316	163	chromosome	1523769	60	251M	=	1523909	391	TCAGGT?????BBB	XT:A:U	
8623	DRR024501.4317	113	plasmid2	5400	37	251M	chr	668425	0	ATAAAT*EE?CEC:C	XT:A:U	
8624	DRR024501.4317	177	chromosome	668425	37	251M	plac	5400	0	AATAGT0:*'?//*0*	XT:A:U	
8625	DRR024501.4318	83	chromosome	1698448	60	251M	=	1698316	-383	GACCGC>?849CGE	XT:A:U	
8626	DRR024501.4318	163	chromosome	1698316	60	251M	=	1698448	383	CTAGTT?????BB?D	XT:A:U	
8627	DRR024501.4319	83	chromosome	1401590	29	251M	=	1401462	-379	ATTGGT*EGEOGEC	XT:A:U	
8628	DRR024501.4319	163	chromosome	1401462	29	128M123S	=	1401590	379	CCACGC?????ABB@	XT:A:M	
8629	DRR024501.4320	69	chromosome	1220366	0	*	=	1220366	0	GTATCT13;?*0:A.8?CCC?		
8630	DRR024501.4320	137	chromosome	1220366	37	109M	=	1220366	0	TGATTA?????BBB	XT:A:U	
8631	DRR024501.4321	83	chromosome	873625	60	251M	=	873528	-348	GGCTTC088A8>CE	XT:A:U	
8632	DRR024501.4321	163	chromosome	873528	60	251M	=	873625	348	CTAAAT?????BB?E	XT:A:U	
8633	DRR024501.4322	77	*	0	0	*	*	0	0	GTATCTFBFB5BFF@DEEB		
8634	DRR024501.4322	141	*	0	0	*	*	0	0	TCTCGC++4.@@99@1*109		

W17-3: RNEXTとPNEXT

⑦のRNEXTが=の例ですが、なんだかよくわからないものもありますね。これは、、、ペアのリードが同じ場所にマップされたということでしょうか。。。



	A	B	C	D	E	F	G	H	I	J	K	L
8618	DRR024501.4314	163	chromosome	1352183	60	251M	=	1352317	385	CAACG/????<?<?B	XT:A:U	
8619	DRR024501.4315	99	chromosome	87242	60	251M	=	87383	392	CAACT/AAAA?BBE	XT:A:U	
8620	DRR024501.4315	147	chromosome	87383	60	251M	=	87242	-392	GCAATT*C:0'2.*:10	XT:A:U	
8621	DRR024501.4316	83	chromosome	1523909	60	251M	=	1523769	-391	ATCGAC.EGGECCE	XT:A:U	
8622	DRR024501.4316	163	chromosome	1523769	60	251M	=	1523909	391	TCAGGT?????BBB	XT:A:U	
8623	DRR024501.4317	113	plasmid2	5400	37	251M	chr	668425	0	ATAAAT*EE?CEC:C	XT:A:U	
8624	DRR024501.4317	177	chromosome	668425	37	251M	plas	5400	0	AATAGT0:*'?//*0*	XT:A:U	
8625	DRR024501.4318	83	chromosome	1698448	60	251M	=	1698316	-383	GACCGC>?849CGE	XT:A:U	
8626	DRR024501.4318	163	chromosome	1698316	60	251M	=	1698448	383	CTAGTT?????BB?D	XT:A:U	
8627	DRR024501.4319	83	chromosome	1401590	29	251M	=	1401462	-379	ATTGGT*EGEOGEC	XT:A:U	
8628	DRR024501.4319	163	chromosome	1401462	29	128M123S	=	1401590	379	CCACGC(????ABB@	XT:A:M	
8629	DRR024501.4320	69	chromosome	1220366	0	*	=	1220366	0	GTATCT13;?*0:A.8?CCC?		
8630	DRR024501.4320	137	chromosome	1220366	37	109M	=	1220366	0	TGATTA?????BBB	XT:A:U	
8631	DRR024501.4321	83	chromosome	873625	60	251M	=	873528	-348	GGCTTC088A8>CE	XT:A:U	
8632	DRR024501.4321	163	chromosome	873528	60	251M	=	873625	348	CTAAAT?????BB?E	XT:A:U	
8633	DRR024501.4322	77	*	0	0	*	*	0	0	GTATCTFBFB5BFF@DEEB		
8634	DRR024501.4322	141	*	0	0	*	*	0	0	TCTCGC++4.@@99@1*109		

W17-4: TLEN

⑨TLEN (observed Template LENgth)。paired-endでシークエンスする前のテンプレート断片配列の長さ。概ね、「(⑧-④)の絶対値 + ⑥」のような感じで計算されます

	A	B	C	D	E	F	G	H	I	J	K	L
1	@SQ	SN:chromosome	LN:2277981									
2	@SQ	SN:plasmid1	LN:81630									
3	@SQ	SN:plasmid2	LN:40973									
4	@PG	ID:bwa	PN:bwa	VN:0.1.104								
5	DRR024501.1	83	chromosome	565592	60	251M	=	565460	-383	CGCCAC(888EEA???)XT:A:1		
6	DRR024501.1	163	chromosome	565460	60	251M	=	565592	383	ATGAAC?????B?@<XT:A:1		
7	DRR024501.2	83	chromosome	1188239	60	251M	=	1188115	-375	TCACTT(??*)?C?:(XT:A:1		
8	DRR024501.2	163	chromosome	1188115	60	251M	=	1188239	375	TGGACC??<??B?9EXT:A:1		
9	DRR024501.3	99	chromosome	721208	29	251M	=	721460	367	CAANGA===#55<=€XT:A:1		
10	DRR024501.3	147	chromosome	721460	29	136S115M	=	721208	-367	ATTACT(//((//(((//((XT:A:1		
11	DRR024501.4	99	chromosome	236084	60	184M	=	236106	197	AGANAC???#5<<?DXT:A:1		
12	DRR024501.4	147	chromosome	236106	60	175M	=	236084	-197	TTAATT ECA?.C:*?:XT:A:1		
13	DRR024501.5	77	*	0	0	*	*	0	0	AGTNCC?,<#55?,<+<5?<<		
14	DRR024501.5	141	*	0	0	*	*	0	0	GATAAA,55=,<5<--5@@@		
15	DRR024501.6	89	chromosome	1861136	37	188M	=	1861136	0	TTTACC EEEEEEEI XT:A:1		

forward側



reverse側

W17-4:TLEN

例えば、⑨ = $|565460 - 565592| + 251 = 383$ のように計算します。④と⑧は入れ替わってもいいので、⑨の値がプラスかマイナスかは本質的な問題ではありません。⑥はリードの長さです。アダプタートリムされているものもあるので一定ではありません

	A	B	C	D	E	F	G	H	I	J	K	L
1	@SQ	SN:chromosome	LN:2277981									
2	@SQ	SN:plasmid1	LN:81630									
3	@SQ	SN:plasmid2	LN:40973									
4	@PG	ID:bwa	PN:bwa	VN:0.1.104								
5	DRR024501.1	83	chromosome	565592	60	251M	=	565460	-383	CGCCAC888EEA???	EXT:A:1	
6	DRR024501.1	163	chromosome	565460	60	251M	=	565592	383	ATGAAC?????	B?@<XT:A:1	
7	DRR024501.2	83	chromosome	1188239	60	251M	=	1188115	-375	TCACTT?0*)?*	C?:(XT:A:1	
8	DRR024501.2	163	chromosome	1188115	60	251M	=	1188239	375	TGGACC??<??B?9E	XT:A:1	
9	DRR024501.3	99	chromosome	721208	29	251M	=	721460	367	CAANGA===#55<=	EXT:A:1	
10	DRR024501.3	147	chromosome	721460	29	136S115M	=	721208	-367	ATTACT(//((/((((/(XT:A:1	
11	DRR024501.4	99	chromosome	236084	60	184M	=	236106	197	AGANAC???#5<<?	DXT:A:1	
12	DRR024501.4	147	chromosome	236106	60	175M	=	236084	-197	TTAATT ECA?.C:*	?XT:A:1	
13	DRR024501.5	77	*	0	0	*	*	0	0	AGTNCC?,<#55?,<+<5?<<		
14	DRR024501.5	141	*	0	0	*	*	0	0	GATAAA,55=,<5<--5@@@		
15	DRR024501.6	89	chromosome	1861136	37	189M	=	1861136	0	TTTACC EEEEEEEFI	XT:A:1	

forward側



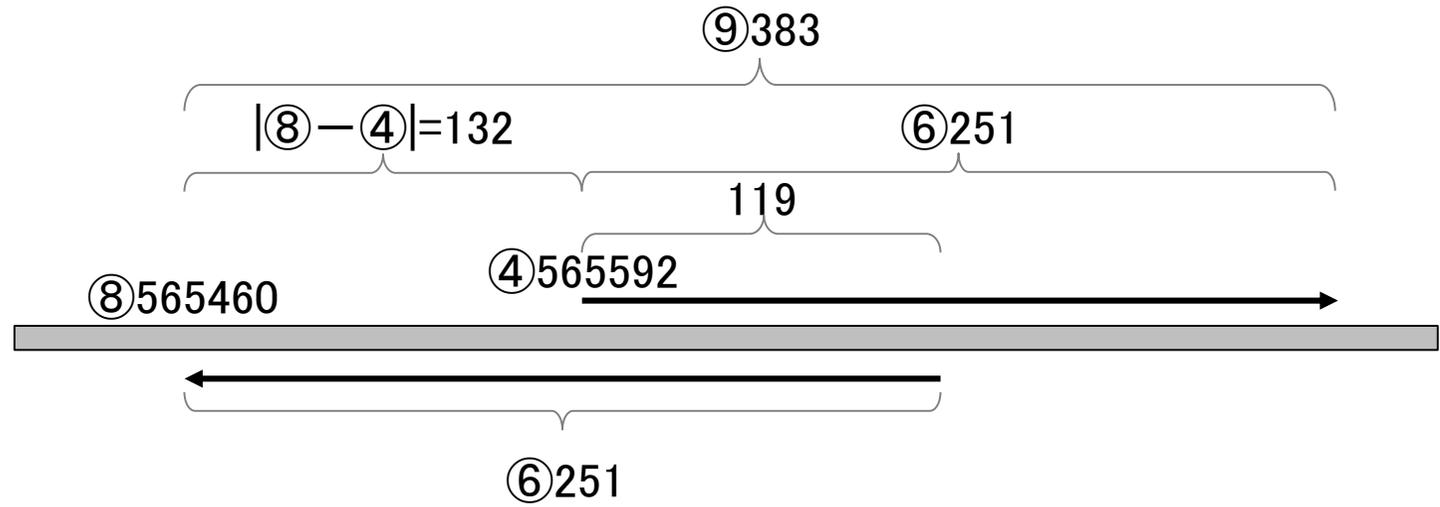
reverse側



W17-4:TLEN

- ④ POS (start POSition)
- ⑥ CIGAR (CIGAR string)
- ⑧ PNEXT (paired-endのもう片方のマップされた座標情報)
- ⑨ TLEN (observed Template LENgth) = | ⑧ - ④ | + ⑥

	A	B	C	D	E	F	G	H	I	J	K	L
1	@SQ	SN:chromosome	LN:2277981									
2	@SQ	SN:plasmid1	LN:81630									
3	@SQ	SN:plasmid2	LN:40973									
4	@PG	ID:bwa	PN:bwa	VN:0.1.1	r104							
5	DRR024501.1	83	chromosome	565592	60	251M	=	565460	-383	CGCCA(888EEA???)XT:A:1		
6	DRR024501.1	163	chromosome	565460	60	251M	=	565592	383	ATGAAC?????B?@<XT:A:1		



W17-5: SEQとQUAL

⑩SEQ (SEQUence)。リード塩基配列情報。⑪QUAL (QUALity)。1文字表記のクオリティスコア

	A	B	C	D	E	F	G	H	I	J	K	L
1	@SQ	SN:chromosome	LN:2277981									
2	@SQ	SN:plasmid1	LN:81630									
3	@SQ	SN:plasmid2	LN:40973									
4	@PG	ID:bwa	PN:bwa	VN:0.6.1-r104								
5	DRR024501.1	83	chromosome	565592	60	251M	=	565460	-383	CGCCA(888EEA???	EXT:A:	
6	DRR024501.1	163	chromosome	565460	60	251M	=	565592	383	ATGAAC?????B?@<	XT:A:	
7	DRR024501.2	83	chromosome	1188239	60	251M	=	1188115	-375	TCACTT(??*)?C?:(XT:A:	
8	DRR024501.2	163	chromosome	1188115	60	251M	=	1188239	375	TGGACC??<??B?9E	XT:A:	
9	DRR024501.3	99	chromosome	721208	29	251M	=	721460	367	CAANGA===#55<=€	XT:A:	
10	DRR024501.3	147	chromosome	721460	29	136S115M	=	721208	-367	ATTACT(//((/((((/(XT:A:	
11	DRR024501.4	99	chromosome	236084	60	184M	=	236106	197	AGANAC???#5<<?D	XT:A:	
12	DRR024501.4	147	chromosome	236106	60	175M	=	236084	-197	TTAATT ECA?.C:*?	XT:A:	
13	DRR024501.5	77	*	0	0	*	*	0	0	AGTNCC?,<#55?,<+<5?<<		
14	DRR024501.5	141	*	0	0	*	*	0	0	GATAAA,55=,<5<--5@@@		
15	DRR024501.6	89	chromosome	1861136	37	188M	=	1861136	0	TTTACC EEEEEEEF	XT:A:	



W17-5: optional field

⑫12列目以降は付加的な情報を含む場所 (optional field)。ここが「XT:A:U」となっているものがユニークヒットに相当するもの。uniqout.sam は、この条件を満たすリードペアのみからなる

	A	B	C	D	E	F	G	H	I	J	K	L
1	@SQ	SN:chromosome	LN:2277981									
2	@SQ	SN:plasmid1	LN:81630									
3	@SQ	SN:plasmid2	LN:40973									
4	@PG	ID:bwa	PN:bwa	VN:0.6.1-r104								
5	DRR024501.1	83	chromosome	565592	60	251M	=	565460	-383	CGCCAC888EEA???	XT:A:U	
6	DRR024501.1	163	chromosome	565460	60	251M	=	565592	383	ATGAAC?????B?@<	XT:A:U	
7	DRR024501.2	83	chromosome	1188239	60	251M	=	1188115	-375	TCACTT?0*)'*C?:(XT:A:U	
8	DRR024501.2	163	chromosome	1188115	60	251M	=	1188239	375	TGGACC??<??B?9E	XT:A:U	
9	DRR024501.3	99	chromosome	721208	29	251M	=	721460	367	CAANGA===#55<=€	XT:A:U	
10	DRR024501.3	147	chromosome	721460	29	136S115M	=	721208	-367	ATTACT(//((/((((/(XT:A:U	
11	DRR024501.4	99	chromosome	236084	60	184M	=	236106	197	AGANAC???#5<<?D	XT:A:U	
12	DRR024501.4	147	chromosome	236106	60	175M	=	236084	-197	TTAATT ECA?.C:*?	XT:A:U	
13	DRR024501.5	77	*	0	0	*	*	0	0	AGTNCC?,<#55?,<+<5?<<		
14	DRR024501.5	141	*	0	0	*	*	0	0	GATAAA,55=,<5<--5@@@		
15	DRR024501.6	89	chromosome	1861136	37	188M	=	1861136	0	TTTACC EEEEEEEFI	XT:A:U	



W17-6: out2.bam

①out2.bamは、out.samからユニーク化処理後のSAMファイルである②uniqout.samをベースに作成。
③out2.bamはバイナリファイル。直接眺めることはできないため、samtools viewコマンドで見ている。
less -Sで開き、1行分を折り返さずに表示(W16-6)

```
iu@bielinux[mac_share] pwd
/home/iu/Desktop/mac_share
iu@bielinux[mac_share] ls -l *.sam *.bam
-rwxrwxrwx 1 iu iu 71310955  9月 12 17:26 out2.bam
-rwxrwxrwx 1 iu iu 371881212  9月 12 17:19 out.sam
-rwxrwxrwx 1 iu iu 320957638  9月 12 17:23 uniqout.sam
iu@bielinux[mac_share] samtools view out2.bam | less -S
```

[2:20午後]

[2:20午後]



W17-6: out2.bam

こんな感じ。若干ずれていますが、chromosomeの右隣の数値が昇順にソートされていることがわかる。Gキーを押してページの最終行に移動

```
File Edit View Search Terminal Help 14:33
DRR024501.86548 163 chromosome 4 60 251M =
DRR024501.295594 163 chromosome 6 60 251M
DRR024501.296978 99 chromosome 8 60 251M
DRR024501.292258 99 chromosome 9 60 251M
DRR024501.275118 99 chromosome 24 60 251M
DRR024501.295594 83 chromosome 32 60 251M
DRR024501.122812 99 chromosome 53 60 251M
DRR024501.288276 163 chromosome 53 60 251M
DRR024501.122812 147 chromosome 57 60 251M
DRR024501.171571 99 chromosome 63 60 251M
DRR024501.44079 99 chromosome 72 60 251M =
DRR024501.181284 99 chromosome 72 60 251M
DRR024501.266245 163 chromosome 76 60 251M
DRR024501.106304 163 chromosome 84 60 251M
DRR024501.106304 83 chromosome 89 60 251M
DRR024501.28563 163 chromosome 92 60 251M =
DRR024501.58849 163 chromosome 94 60 251M =
DRR024501.79179 99 chromosome 102 60 251M =
DRR024501.188616 99 chromosome 106 60 251M
DRR024501.28563 83 chromosome 107 60 251M =
DRR024501.216324 163 chromosome 107 60 251M
:
```

W17-6: out2.bam

out2.bamの最終行はこんな感じ。plasmid2は40,973塩基なので(W16-6)、①最後のほうが40786で終わっているのは妥当といえる。尚、out2.bam作成過程でマップされなかったリードは除外されているので、このファイル中には存在しない

RRR024501.228401	83	plasmid2			
RRR024501.178522	147	plasmid2			
RRR024501.240655	83	plasmid2	40680	60	251M
RRR024501.73398 147		plasmid2	40685	60	251M =
RRR024501.267503	147	plasmid2	40688	60	251M
RRR024501.254114	145	plasmid2	40694	37	137M
RRR024501.254114	97	plasmid2	40695	37	132M
RRR024501.264041	147	plasmid2	40695	60	251M
RRR024501.147571	83	plasmid2	40706	60	251M
RRR024501.141761	147	plasmid2	40707	60	251M
RRR024501.218771	83	plasmid2	40708	60	251M
RRR024501.217892	147	plasmid2	40710	60	251M
RRR024501.165704	83	plasmid2	40714	60	251M
RRR024501.13639 83		plasmid2	40717	60	251M =
RRR024501.127223	147	plasmid2	40718	60	251M
RRR024501.176446	99	plasmid2	40723	60	130M
RRR024501.176446	147	plasmid2	40741	60	113M
RRR024501.103552	99	plasmid2	40772	60	130M
RRR024501.103552	147	plasmid2	40781	60	128M
RRR024501.271587	145	plasmid2	40785	37	138M
RRR024501.271587	97	plasmid2	40786	37	123M
(END)					



W18-1: VCFファイル

VCF (Variant Call Format)形式ファイルの説明。
DDBJ Pipelineからダウンロードして解凍した、①
out-unique.var.flt.vcfは、55行。②less -Nで確認

```
iu@bielinux[mac_share] pwd [ 3:59午後 ]
/home/iu/Desktop/mac_share
iu@bielinux[mac_share] ls -l *.vcf [ 3:59午後 ]
-rwxrwxrwx 1 iu iu 6270  9月 12 17:28 out-unique.var.flt.vcf
① iu@bielinux[mac_share] wc *.vcf [ 3:59午後 ]
  55  461 6270 out-unique.var.flt.vcf
② iu@bielinux[mac_share] less -N *.vcf [ 3:59午後 ]
```

W18-1: VCFファイル

less -Nで確認しているところ。ここで見えているのは、行頭が#から始まるヘッダー行(メタデータ情報部分)。全部で55行しかないのに、①少なくとも最初の11行はヘッダー部分であることがわかる

```
File Edit View Search Terminal Help
1 ##fileformat=VCFv4.1
2 ##samtoolsVersion=0.1.18 (r982:295)
3 ##INFO=<ID=DP,Number=1,Type=Integer,Description="Raw read dept
3 h">
4 ##INFO=<ID=DP4,Number=4,Type=Integer,Description="# high-quali
4 ty ref-forward bases, ref-reverse, alt-forward and alt-reverse
4 bases">
5 ##INFO=<ID=MQ,Number=1,Type=Integer,Description="Root-mean-squ
5 are mapping quality of covering reads">
6 ##INFO=<ID=FQ,Number=1,Type=Float,Description="Phred probabili
6 ty of all samples being the same">
7 ##INFO=<ID=AF1,Number=1,Type=Float,Description="Max-likelihood
7 estimate of the first ALT allele frequency (assuming HWE)">
8 ##INFO=<ID=AC1,Number=1,Type=Float,Description="Max-likelihood
8 estimate of the first ALT allele count (no HWE assumption)">
9 ##INFO=<ID=G3,Number=3,Type=Float,Description="ML estimate of
9 genotype frequencies">
10 ##INFO=<ID=HWE,Number=1,Type=Float,Description="Chi^2 based HW
10 E test P-value based on G3">
11 ##INFO=<ID=CLR,Number=1,Type=Integer,Description="Log ratio of
11 genotype likelihoods with and without the constraint">
out-unique.var.flt.vcf
```



W18-1: VCFファイル

下矢印キーを押してページ下部に移動。①27行目がヘッダー行の最後。この情報が28行目以降の情報の列名に相当するのだろう

```
File Edit View Search Terminal Help
22 ality">
23 ##FORMAT=<ID=GL,Number=3,Type=Float,Description="Likelihoods f
23 or RR,RA,AA genotypes (R=ref,A=alt)">
24 ##FORMAT=<ID=DP,Number=1,Type=Integer,Description="# high-qual
24 ity bases">
25 ##FORMAT=<ID=SP,Number=1,Type=Integer,Description="Phred-scale
25 d strand bias P-value">
26 ##FORMAT=<ID=PL,Number=G,Type=Integer,Description="List of Phr
26 ed-scaled genotype likelihoods">
27 #CHROM POS ID REF ALT QUAL FILTER INFO
27 FORMAT /home/w3pipeline/refdata/result/agribio/24152/2395
27 9/37608/uniqsam/out2.bam
28 chromosome 141598 . A G 17.1 .
28 DP=7;VDB=0.0077;AF1=0.5;AC1=1;DP4=4,1,1,1;MQ=39;FQ=20.1
28 ;PV4=1,1,0,1 GT:PL:GQ 0/1:47,0,138:50
29 chromosome 242819 . T C 20 .
29 DP=9;VDB=0.0111;AF1=0.5;AC1=1;DP4=5,2,1,1;MQ=49;FQ=23;P
29 V4=1,1,0,1 GT:PL:GQ 0/1:50,0,139:53
30 chromosome 359141 . T C 4.13 .
30 DP=14;VDB=0.0037;AF1=0.4998;AC1=1;DP4=11,1,1,1;MQ=49;FQ
30 =6.2;PV4=0.27,1,0.021,1 GT:PL:GQ 0/1:32,0,187:31
```



Excelで眺めるとこんな感じ。1行目はVCFのバージョン。これはver. 4.1のようだ

W18-2: Excel

	A	B	C	D	E	F	G	H	I	J	K	L
1	##	fileformat=	VCFv4.1									
2	##	samtoolsVersion=	0.1.18	(r982:295)								
3	##	INFO=<ID=DP,Number=1,Type=Integer,Description="Raw read depth">										
4	##	INFO=<ID=DP4,Number=4,Type=Integer,Description="# high-quality ref-forward bases, ref-reverse, alt-										
5	##	INFO=<ID=MQ,Number=1,Type=Integer,Description="Root-mean-square mapping quality of covering re-										
6	##	INFO=<ID=FQ,Number=1,Type=Float,Description="Phred probability of all samples being the same">										
7	##	INFO=<ID=AF1,Number=1,Type=Float,Description="Max-likelihood estimate of the first ALT allele frequ										
8	##	INFO=<ID=AC1,Number=1,Type=Float,Description="Max-likelihood estimate of the first ALT allele cour										
9	##	INFO=<ID=G3,Number=3,Type=Float,Description="ML estimate of genotype frequencies">										
10	##	INFO=<ID=HWE,Number=1,Type=Float,Description="Chi^2 based HWE test P-value based on G3">										
11	##	INFO=<ID=CLR,Number=1,Type=Integer,Description="Log ratio of genotype likelihoods with and without										
12	##	INFO=<ID=UGT,Number=1,Type=String,Description="The most probable unconstrained genotype config										
13	##	INFO=<ID=CGT,Number=1,Type=String,Description="The most probable constrained genotype configura										
14	##	INFO=<ID=PV4,Number=4,Type=Float,Description="P-values for strand bias, baseQ bias, mapQ bias and										

W18-2: Excel

27行目以降の主要な結果の全体像。①REFは、概要配列側の塩基。②ALTは、変異塩基(ALTnative base)。③QUALは、クオリティスコア(高いほどよい)

	A	B	C	①	②	③	G	H	I	J
27	#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	/home/w3pipeline/refdata/
28	chromosome	141598	.	A	G	17.1	.	DP=7;VDB=0.0077;AF1=0.5;AC1	GT:PL:GQ	0/1:47,0,138:50
29	chromosome	242819	.	T	C	20	.	DP=9;VDB=0.0111;AF1=0.5;AC1	GT:PL:GQ	0/1:50,0,139:53
30	chromosome	359141	.	T	C	4.13	.	DP=14;VDB=0.0037;AF1=0.4998	GT:PL:GQ	0/1:32,0,187:31
31	chromosome	463882	.	T	G	6.2	.	DP=13;VDB=0.0074;AF1=0.4999	GT:PL:GQ	0/1:35,0,216:36
32	chromosome	660031	.	C	T	8.64	.	DP=11;VDB=0.0063;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,197:40
33	chromosome	792963	.	T	C	4.13	.	DP=14;VDB=0.0153;AF1=0.4998	GT:PL:GQ	0/1:32,0,200:31
34	chromosome	887422	.	G	A	10.4	.	DP=11;VDB=0.0070;AF1=0.5;AC	GT:PL:GQ	0/1:40,0,188:42
35	chromosome	913470	.	T	C	14.2	.	DP=9;VDB=0.0056;AF1=0.5;AC1	GT:PL:GQ	0/1:44,0,144:47
36	chromosome	965171	.	T	C	17.1	.	DP=20;VDB=0.0071;AF1=0.5;AC	GT:PL:GQ	0/1:47,0,249:50
37	chromosome	1134950	.	A	G	7.8	.	DP=23;VDB=0.0076;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,255:39
38	chromosome	1160048	.	A	G	29	.	DP=17;VDB=0.0054;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
39	chromosome	1160095	.	A	G	19.1	.	DP=10;VDB=0.0033;AF1=0.5;AC	GT:PL:GQ	0/1:49,0,191:52
40	chromosome	1240248	.	GCCCC	GCCCCC	214	.	INDEL;DP=48;VDB=0.0137;AF1=	GT:PL:GQ	1/1:255,128,0:99
41	chromosome	1455552	.	CAAAA	CAAAAA	214	.	INDEL;DP=58;VDB=0.0164;AF1=	GT:PL:GQ	1/1:255,130,0:99
42	chromosome	1475721	.	T	C	52	.	DP=19;VDB=0.0021;AF1=0.5;AC	GT:PL:GQ	0/1:82,0,246:85
43	chromosome	1752680	.	G	A	29	.	DP=24;VDB=0.0129;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
44	chromosome	1941380	.	T	C	7.8	.	DP=13;VDB=0.0046;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,214:39
45	chromosome	2052955	.	A	G	12.3	.	DP=8;VDB=0.0084;AF1=0.5;AC1	GT:PL:GQ	0/1:42,0,103:44
46	chromosome	2054084	.	TC	T	13.7	.	INDEL;DP=8;VDB=0.0009;AF1=0	GT:PL:GQ	0/1:51,0,174:54
47	chromosome	2056279	.	A	G	24	.	DP=6;VDB=0.0057;AF1=0.5;AC1	GT:PL:GQ	0/1:54,0,113:57
48	chromosome	2057564	.	T	C	8.64	.	DP=10;VDB=0.0041;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,129:40
49	chromosome	2058145	.	C	T,G	108	.	DP=14;VDB=0.0147;AF1=1;AC1=	GT:PL:GQ	1/1:141,39,0,141,28,138:75
50	plasmid1	37177	.	G	T	54	.	DP=49;VDB=0.0087;AF1=0.5;AC	GT:PL:GQ	0/1:84,0,255:87
51	plasmid1	66609	.	A	G	225	.	DP=55;VDB=0.0060;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
52	plasmid1	66637	.	C	T	225	.	DP=71;VDB=0.0141;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,249:99
53	plasmid1	66668	.	C	T	225	.	DP=80;VDB=0.0153;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
54	plasmid2	2569	.	CAAAA	CAAAA	177	.	INDEL;DP=110;VDB=0.0166;AF1	GT:PL:GQ	1/1:218,255,0:99
55	plasmid2	37805	.	TAAAAA	TAAAA	104	.	INDEL;DP=31;VDB=0.0058;AF1=	GT:PL:GQ	1/1:145,93,0:99

④INFOは、変異情報。⑤FORMATは、genotype (GT)のフォーマット、⑥は⑤の形式でgenotypeの具体的な情報が記載されている

W18-2: Excel

	A	B	C	D	E	F	G	④	⑤	⑥
27	#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	/home/w3pipeline/refdata/
28	chromosome	141598	.	A	G	17.1	.	DP=7;VDB=0.0077;AF1=0.5;AC1	GT:PL:GQ	0/1:47,0,138:50
29	chromosome	242819	.	T	C	20	.	DP=9;VDB=0.0111;AF1=0.5;AC1	GT:PL:GQ	0/1:50,0,139:53
30	chromosome	359141	.	T	C	4.13	.	DP=14;VDB=0.0037;AF1=0.4998	GT:PL:GQ	0/1:32,0,187:31
31	chromosome	463882	.	T	G	6.2	.	DP=13;VDB=0.0074;AF1=0.4999	GT:PL:GQ	0/1:35,0,216:36
32	chromosome	660031	.	C	T	8.64	.	DP=11;VDB=0.0063;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,197:40
33	chromosome	792963	.	T	C	4.13	.	DP=14;VDB=0.0153;AF1=0.4998	GT:PL:GQ	0/1:32,0,200:31
34	chromosome	887422	.	G	A	10.4	.	DP=11;VDB=0.0070;AF1=0.5;AC	GT:PL:GQ	0/1:40,0,188:42
35	chromosome	913470	.	T	C	14.2	.	DP=9;VDB=0.0056;AF1=0.5;AC1	GT:PL:GQ	0/1:44,0,144:47
36	chromosome	965171	.	T	C	17.1	.	DP=20;VDB=0.0071;AF1=0.5;AC	GT:PL:GQ	0/1:47,0,249:50
37	chromosome	1134950	.	A	G	7.8	.	DP=23;VDB=0.0076;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,255:39
38	chromosome	1160048	.	A	G	29	.	DP=17;VDB=0.0054;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
39	chromosome	1160095	.	A	G	19.1	.	DP=10;VDB=0.0033;AF1=0.5;AC	GT:PL:GQ	0/1:49,0,191:52
40	chromosome	1240248	.	GCCCC	GCCCCC	214	.	INDEL;DP=48;VDB=0.0137;AF1=	GT:PL:GQ	1/1:255,128,0:99
41	chromosome	1455552	.	CAAAA	CAAAAA	214	.	INDEL;DP=58;VDB=0.0164;AF1=	GT:PL:GQ	1/1:255,130,0:99
42	chromosome	1475721	.	T	C	52	.	DP=19;VDB=0.0021;AF1=0.5;AC	GT:PL:GQ	0/1:82,0,246:85
43	chromosome	1752680	.	G	A	29	.	DP=24;VDB=0.0129;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
44	chromosome	1941380	.	T	C	7.8	.	DP=13;VDB=0.0046;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,214:39
45	chromosome	2052955	.	A	G	12.3	.	DP=8;VDB=0.0084;AF1=0.5;AC1	GT:PL:GQ	0/1:42,0,103:44
46	chromosome	2054084	.	TC	T	13.7	.	INDEL;DP=8;VDB=0.0009;AF1=0	GT:PL:GQ	0/1:51,0,174:54
47	chromosome	2056279	.	A	G	24	.	DP=6;VDB=0.0057;AF1=0.5;AC1	GT:PL:GQ	0/1:54,0,113:57
48	chromosome	2057564	.	T	C	8.64	.	DP=10;VDB=0.0041;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,129:40
49	chromosome	2058145	.	C	T,G	108	.	DP=14;VDB=0.0147;AF1=1;AC1=	GT:PL:GQ	1/1:141,39,0,141,28,138:75
50	plasmid1	37177	.	G	T	54	.	DP=49;VDB=0.0087;AF1=0.5;AC	GT:PL:GQ	0/1:84,0,255:87
51	plasmid1	66609	.	A	G	225	.	DP=55;VDB=0.0060;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
52	plasmid1	66637	.	C	T	225	.	DP=71;VDB=0.0141;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,249:99
53	plasmid1	66668	.	C	T	225	.	DP=80;VDB=0.0153;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
54	plasmid2	2569	.	CAAAA	CAAAA	177	.	INDEL;DP=110;VDB=0.0166;AF1	GT:PL:GQ	1/1:218,255,0:99
55	plasmid2	37805	.	TAAAAA	TAAAA	104	.	INDEL;DP=31;VDB=0.0058;AF1=	GT:PL:GQ	1/1:145,93,0:99

今回はpaired-endの1サンプルのみを入力としたので、⑥の部分が一列分のみのgenotype情報からなる

W18-2: Excel



	A	B	C	D	E	F	G	H	I	J
27	#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	/home/w3pipeline/refdata/
28	chromosome	141598	.	A	G	17.1	.	DP=7;VDB=0.0077;AF1=0.5;AC1	GT:PL:GQ	0/1:47,0,138:50
29	chromosome	242819	.	T	C	20	.	DP=9;VDB=0.0111;AF1=0.5;AC1	GT:PL:GQ	0/1:50,0,139:53
30	chromosome	359141	.	T	C	4.13	.	DP=14;VDB=0.0037;AF1=0.4998	GT:PL:GQ	0/1:32,0,187:31
31	chromosome	463882	.	T	G	6.2	.	DP=13;VDB=0.0074;AF1=0.4999	GT:PL:GQ	0/1:35,0,216:36
32	chromosome	660031	.	C	T	8.64	.	DP=11;VDB=0.0063;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,197:40
33	chromosome	792963	.	T	C	4.13	.	DP=14;VDB=0.0153;AF1=0.4998	GT:PL:GQ	0/1:32,0,200:31
34	chromosome	887422	.	G	A	10.4	.	DP=11;VDB=0.0070;AF1=0.5;AC	GT:PL:GQ	0/1:40,0,188:42
35	chromosome	913470	.	T	C	14.2	.	DP=9;VDB=0.0056;AF1=0.5;AC1	GT:PL:GQ	0/1:44,0,144:47
36	chromosome	965171	.	T	C	17.1	.	DP=20;VDB=0.0071;AF1=0.5;AC	GT:PL:GQ	0/1:47,0,249:50
37	chromosome	1134950	.	A	G	7.8	.	DP=23;VDB=0.0076;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,255:39
38	chromosome	1160048	.	A	G	29	.	DP=17;VDB=0.0054;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
39	chromosome	1160095	.	A	G	19.1	.	DP=10;VDB=0.0033;AF1=0.5;AC	GT:PL:GQ	0/1:49,0,191:52
40	chromosome	1240248	.	GCCCC	GCCCCC	214	.	INDEL;DP=48;VDB=0.0137;AF1=	GT:PL:GQ	1/1:255,128,0:99
41	chromosome	1455552	.	CAAAA	CAAAAA	214	.	INDEL;DP=58;VDB=0.0164;AF1=	GT:PL:GQ	1/1:255,130,0:99
42	chromosome	1475721	.	T	C	52	.	DP=19;VDB=0.0021;AF1=0.5;AC	GT:PL:GQ	0/1:82,0,246:85
43	chromosome	1752680	.	G	A	29	.	DP=24;VDB=0.0129;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
44	chromosome	1941380	.	T	C	7.8	.	DP=13;VDB=0.0046;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,214:39
45	chromosome	2052955	.	A	G	12.3	.	DP=8;VDB=0.0084;AF1=0.5;AC1	GT:PL:GQ	0/1:42,0,103:44
46	chromosome	2054084	.	TC	T	13.7	.	INDEL;DP=8;VDB=0.0009;AF1=0	GT:PL:GQ	0/1:51,0,174:54
47	chromosome	2056279	.	A	G	24	.	DP=6;VDB=0.0057;AF1=0.5;AC1	GT:PL:GQ	0/1:54,0,113:57
48	chromosome	2057564	.	T	C	8.64	.	DP=10;VDB=0.0041;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,129:40
49	chromosome	2058145	.	C	T,G	108	.	DP=14;VDB=0.0147;AF1=1;AC1=	GT:PL:GQ	1/1:141,39,0,141,28,138:75
50	plasmid1	37177	.	G	T	54	.	DP=49;VDB=0.0087;AF1=0.5;AC	GT:PL:GQ	0/1:84,0,255:87
51	plasmid1	66609	.	A	G	225	.	DP=55;VDB=0.0060;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
52	plasmid1	66637	.	C	T	225	.	DP=71;VDB=0.0141;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,249:99
53	plasmid1	66668	.	C	T	225	.	DP=80;VDB=0.0153;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
54	plasmid2	2569	.	CAAAA	CAAAA	177	.	INDEL;DP=110;VDB=0.0166;AF1	GT:PL:GQ	1/1:218,255,0:99
55	plasmid2	37805	.	TAAAAA	TAAAA	104	.	INDEL;DP=31;VDB=0.0058;AF1=	GT:PL:GQ	1/1:145,93,0:99

W18-2: Excel

もしBWAマッピング時の入力が3サンプル分だったら、⑥の部分
が3列分の情報になる。具体的には、ExcelでいうところのK列と
L列に2サンプル目と3サンプル目のgenotype情報が追加される

	A	B	C	D	E	F	G	H	I	J
27	#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	/home/w3pipeline/refdata/
28	chromosome	141598	.	A	G	17.1	.	DP=7;VDB=0.0077;AF1=0.5;AC1	GT:PL:GQ	0/1:47,0,138:50
29	chromosome	242819	.	T	C	20	.	DP=9;VDB=0.0111;AF1=0.5;AC1	GT:PL:GQ	0/1:50,0,139:53
30	chromosome	359141	.	T	C	4.13	.	DP=14;VDB=0.0037;AF1=0.4998	GT:PL:GQ	0/1:32,0,187:31
31	chromosome	463882	.	T	G	6.2	.	DP=13;VDB=0.0074;AF1=0.4999	GT:PL:GQ	0/1:35,0,216:36
32	chromosome	660031	.	C	T	8.64	.	DP=11;VDB=0.0063;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,197:40
33	chromosome	792963	.	T	C	4.13	.	DP=14;VDB=0.0153;AF1=0.4998	GT:PL:GQ	0/1:32,0,200:31
34	chromosome	887422	.	G	A	10.4	.	DP=11;VDB=0.0070;AF1=0.5;AC	GT:PL:GQ	0/1:40,0,188:42
35	chromosome	913470	.	T	C	14.2	.	DP=9;VDB=0.0056;AF1=0.5;AC1	GT:PL:GQ	0/1:44,0,144:47
36	chromosome	965171	.	T	C	17.1	.	DP=20;VDB=0.0071;AF1=0.5;AC	GT:PL:GQ	0/1:47,0,249:50
37	chromosome	1134950	.	A	G	7.8	.	DP=23;VDB=0.0076;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,255:39
38	chromosome	1160048	.	A	G	29	.	DP=17;VDB=0.0054;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
39	chromosome	1160095	.	A	G	19.1	.	DP=10;VDB=0.0033;AF1=0.5;AC	GT:PL:GQ	0/1:49,0,191:52
40	chromosome	1240248	.	GCCCC	GCCCCC	214	.	INDEL;DP=48;VDB=0.0137;AF1=	GT:PL:GQ	1/1:255,128,0:99
41	chromosome	1455552	.	CAAAA	CAAAAA	214	.	INDEL;DP=58;VDB=0.0164;AF1=	GT:PL:GQ	1/1:255,130,0:99
42	chromosome	1475721	.	T	C	52	.	DP=19;VDB=0.0021;AF1=0.5;AC	GT:PL:GQ	0/1:82,0,246:85
43	chromosome	1752680	.	G	A	29	.	DP=24;VDB=0.0129;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
44	chromosome	1941380	.	T	C	7.8	.	DP=13;VDB=0.0046;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,214:39
45	chromosome	2052955	.	A	G	12.3	.	DP=8;VDB=0.0084;AF1=0.5;AC1	GT:PL:GQ	0/1:42,0,103:44
46	chromosome	2054084	.	TC	T	13.7	.	INDEL;DP=8;VDB=0.0009;AF1=0	GT:PL:GQ	0/1:51,0,174:54
47	chromosome	2056279	.	A	G	24	.	DP=6;VDB=0.0057;AF1=0.5;AC1	GT:PL:GQ	0/1:54,0,113:57
48	chromosome	2057564	.	T	C	8.64	.	DP=10;VDB=0.0041;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,129:40
49	chromosome	2058145	.	C	T,G	108	.	DP=14;VDB=0.0147;AF1=1;AC1=	GT:PL:GQ	1/1:141,39,0,141,28,138:75
50	plasmid1	37177	.	G	T	54	.	DP=49;VDB=0.0087;AF1=0.5;AC	GT:PL:GQ	0/1:84,0,255:87
51	plasmid1	66609	.	A	G	225	.	DP=55;VDB=0.0060;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
52	plasmid1	66637	.	C	T	225	.	DP=71;VDB=0.0141;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,249:99
53	plasmid1	66668	.	C	T	225	.	DP=80;VDB=0.0153;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
54	plasmid2	2569	.	CAAAA	CAAAA	177	.	INDEL;DP=110;VDB=0.0166;AF1	GT:PL:GQ	1/1:218,255,0:99
55	plasmid2	37805	.	TAAAAA	TAAAA	104	.	INDEL;DP=31;VDB=0.0058;AF1=	GT:PL:GQ	1/1:145,93,0:99

⑥

W18-3: INFO

	A	B	C	D	E	F	G	H
	#CHROM	POS	ID	REF	ALT	QUAL	FI	INFO
27	chromosome	141598	A	G	17	DP=7;VDB=0.0077;AF1=0.5;AC1=1;DP4=4,1,1,1;MQ=39;FQ=20.1;PV4=1,1,0,1		
29	chromosome	242819	T	C	20	DP=9;VDB=0.0111;AF1=0.5;AC1=1;DP4=5,2,1,1;MQ=49;FQ=23;PV4=1,1,0,1		
30	chromosome	359141	T	C	4.1	DP=14;VDB=0.0037;AF1=0.4998;AC1=1;DP4=11,1,1,1;MQ=49;FQ=6.2;PV4=0.27		
31	chromosome	463882	T	G	6.2	DP=13;VDB=0.0074;AF1=0.4999;AC1=1;DP4=2,9,1,1;MQ=48;FQ=8.65;PV4=0.42		
32	chromosome	660031	C	T	8.6	DP=11;VDB=0.0063;AF1=0.5;AC1=1;DP4=7,2,1,1;MQ=49;FQ=11.3;PV4=0.49,0.2		
33	chromosome	792963	T	C	4.1	DP=14;VDB=0.0153;AF1=0.4998;AC1=1;DP4=2,10,1,1;MQ=48;FQ=6.2;PV4=0.4,1		
34	chromosome	887422	G	A	10	DP=11;VDB=0.0070;AF1=0.5;AC1=1;DP4=7,2,1,1;MQ=49;FQ=13.2;PV4=0.49,1,0		
35	chromosome	913470	T	C	14	DP=9;VDB=0.0056;AF1=0.5;AC1=1;DP4=1,6,1,1;MQ=49;FQ=17.1;PV4=0.42,1,0.2		
36	chromosome	965171	T	C	17	DP=20;VDB=0.0071;AF1=0.5;AC1=1;DP4=14,3,2,1;MQ=49;FQ=20.1;PV4=0.51,1,		
37	chromosome	1134950	A	G	7.8	DP=23;VDB=0.0076;AF1=0.5;AC1=1;DP4=5,15,1,2;MQ=49;FQ=10.4;PV4=1,1,0,0		
38	chromosome	1160048	A	G	29	DP=17;VDB=0.0054;AF1=0.5;AC1=1;DP4=6,8,1,2;MQ=48;FQ=32;PV4=1,1,0.041,		
39	chromosome	1160095	A	G	19	DP=10;VDB=0.0033;AF1=0.5;AC1=1;DP4=6,2,1,1;MQ=48;FQ=22;PV4=1,1,0.007,		
40	chromosome	1240248	GCCCC	GCCCCC	214	INDEL;DP=48;VDB=0.0137;AF1=1;AC1=2;DP4=1,0,21,26;MQ=50;FQ=-163;PV4=		
41	chromosome	1455552	CAAAA	CAAAAA	214	INDEL;DP=58;VDB=0.0164;AF1=1;AC1=2;DP4=2,1,28,27;MQ=49;FQ=-165;PV4=		
42	chromosome	1475721	T	C	52	DP=19;VDB=0.0021;AF1=0.5;AC1=1;DP4=11,4,2,2;MQ=48;FQ=55;PV4=0.56,1,2,		
43	chromosome	1752680	G	A	29	DP=24;VDB=0.0129;AF1=0.5;AC1=1;DP4=14,6,2,2;MQ=48;FQ=32;PV4=0.58,1,0,		
44	chromosome	1941380	T	C	7.8	DP=13;VDB=0.0046;AF1=0.5;AC1=1;DP4=9,2,1,1;MQ=50;FQ=10.4;PV4=0.42,1,0		
45	chromosome	2052955	A	G	12	DP=8;VDB=0.0084;AF1=0.5;AC1=1;DP4=2,3,2,1;MQ=50;FQ=15.1;PV4=1,0.34,0.1		
46	chromosome	2054084	TC	T	14	INDEL;DP=8;VDB=0.0009;AF1=0.5;AC1=1;DP4=3,3,1,1;MQ=47;FQ=16.6;PV4=1,4		
47	chromosome	2056279	A	G	24	DP=6;VDB=0.0057;AF1=0.5;AC1=1;DP4=2,2,1,1;MQ=46;FQ=27;PV4=1,1,1,1		
48	chromosome	2057564	T	C	8.6	DP=10;VDB=0.0041;AF1=0.5;AC1=1;DP4=4,4,1,1;MQ=33;FQ=11.3;PV4=1,1,1,1		
49	chromosome	2058145	C	T,G	108	DP=14;VDB=0.0147;AF1=1;AC1=2;DP4=0,0,7,7;MQ=23;FQ=-66		
50	plasmid1	37177	G	T	54	DP=49;VDB=0.0087;AF1=0.5;AC1=1;DP4=24,16,5,4;MQ=49;FQ=57;PV4=1,0.1,0,		
51	plasmid1	66609	A	G	225	DP=55;VDB=0.0060;AF1=0.5;AC1=1;DP4=16,8,22,9;MQ=43;FQ=225;PV4=0.78,1		
52	plasmid1	66637	C	T	225	DP=71;VDB=0.0141;AF1=0.5;AC1=1;DP4=17,11,23,20;MQ=43;FQ=221;PV4=0.6,		
53	plasmid1	66668	C	T	225	DP=80;VDB=0.0153;AF1=0.5;AC1=1;DP4=17,17,23,23;MQ=43;FQ=225;PV4=1,0,		
54	plasmid2	2569	CAAAA	CAAAA	177	INDEL;DP=110;VDB=0.0166;AF1=1;AC1=2;DP4=0,0,51,59;MQ=49;FQ=-290		
55	plasmid2	37805	TAAAA	TAAAA	104	INDEL;DP=31;VDB=0.0058;AF1=1;AC1=2;DP4=0,0,5,26;MQ=31;FQ=-128		

W18-3: INFO

①DPはdepthのこと。数が多いほど多くのリードがマップされたことを意味し、信頼度が高い

	A	B	C	D	E	F	G	H
27	#CHROM	POS	ID	REF	ALT	QUAL	FI	INFO
28	chromosome	141598	A	G	G	17		DP=7;VDB=0.0077;AF1=0.5;AC1=1;DP4=4,1,1,1;MQ=39;FQ=20.1;PV4=1,1,0,1
29	chromosome	242819	T	C	C	20		DP=9;VDB=0.0111;AF1=0.5;AC1=1;DP4=5,2,1,1;MQ=49;FQ=23;PV4=1,1,0,1
30	chromosome	359141	T	C	C	4.1		DP=14;VDB=0.0037;AF1=0.4998;AC1=1;DP4=11,1,1,1;MQ=49;FQ=6.2;PV4=0.27
31	chromosome	463882	T	G	G	6.2		DP=13;VDB=0.0074;AF1=0.4999;AC1=1;DP4=2,9,1,1;MQ=48;FQ=8.65;PV4=0.42
32	chromosome	660031	C	T	T	8.6		DP=11;VDB=0.0063;AF1=0.5;AC1=1;DP4=7,2,1,1;MQ=49;FQ=11.3;PV4=0.49,0.2
33	chromosome	792963	T	C	C	4.1		DP=14;VDB=0.0153;AF1=0.4998;AC1=1;DP4=2,10,1,1;MQ=48;FQ=6.2;PV4=0.4,1
34	chromosome	887422	G	A	A	10		DP=11;VDB=0.0070;AF1=0.5;AC1=1;DP4=7,2,1,1;MQ=49;FQ=13.2;PV4=0.49,1,0
35	chromosome	913470	T	C	C	14		DP=9;VDB=0.0056;AF1=0.5;AC1=1;DP4=1,6,1,1;MQ=49;FQ=17.1;PV4=0.42,1,0.2
36	chromosome	965171	T	C	C	17		DP=20;VDB=0.0071;AF1=0.5;AC1=1;DP4=14,3,2,1;MQ=49;FQ=20.1;PV4=0.51,1,
37	chromosome	1134950	A	G	G	7.8		DP=23;VDB=0.0076;AF1=0.5;AC1=1;DP4=5,15,1,2;MQ=49;FQ=10.4;PV4=1,1,0,0
38	chromosome	1160048	A	G	G	29		DP=17;VDB=0.0054;AF1=0.5;AC1=1;DP4=6,8,1,2;MQ=48;FQ=32;PV4=1,1,0.041,
39	chromosome	1160095	A	G	G	19		DP=10;VDB=0.0033;AF1=0.5;AC1=1;DP4=6,2,1,1;MQ=48;FQ=22;PV4=1,1,0.007,
40	chromosome	1240248	GCCCC	GCCCC	GCCCC	214		INDEL;DP=48;VDB=0.0137;AF1=1;AC1=2;DP4=1,0,21,26;MQ=50;FQ=-163;PV4=
41	chromosome	1455552	CAAAA	CAAAA	CAAAA	214		INDEL;DP=58;VDB=0.0164;AF1=1;AC1=2;DP4=2,1,28,27;MQ=49;FQ=-165;PV4=
42	chromosome	1475721	T	C	C	52		DP=19;VDB=0.0021;AF1=0.5;AC1=1;DP4=11,4,2,2;MQ=48;FQ=55;PV4=0.56,1,2,
43	chromosome	1752680	G	A	A	29		DP=24;VDB=0.0129;AF1=0.5;AC1=1;DP4=14,6,2,2;MQ=48;FQ=32;PV4=0.58,1,0,
44	chromosome	1941380	T	C	C	7.8		DP=13;VDB=0.0046;AF1=0.5;AC1=1;DP4=9,2,1,1;MQ=50;FQ=10.4;PV4=0.42,1,0
45	chromosome	2052955	A	G	G	12		DP=8;VDB=0.0084;AF1=0.5;AC1=1;DP4=2,3,2,1;MQ=50;FQ=15.1;PV4=1,0.34,0.1
46	chromosome	2054084	TC	T	T	14		INDEL;DP=8;VDB=0.0009;AF1=0.5;AC1=1;DP4=3,3,1,1;MQ=47;FQ=16.6;PV4=1,4
47	chromosome	2056279	A	G	G	24		DP=6;VDB=0.0057;AF1=0.5;AC1=1;DP4=2,2,1,1;MQ=46;FQ=27;PV4=1,1,1,1
48	chromosome	2057564	T	C	C	8.6		DP=10;VDB=0.0041;AF1=0.5;AC1=1;DP4=4,4,1,1;MQ=33;FQ=11.3;PV4=1,1,1,1
49	chromosome	2058145	C	T,G	T,G	108		DP=14;VDB=0.0147;AF1=1;AC1=2;DP4=0,0,7,7;MQ=23;FQ=-66
50	plasmid1	37177	G	T	T	54		DP=49;VDB=0.0087;AF1=0.5;AC1=1;DP4=24,16,5,4;MQ=49;FQ=57;PV4=1,0,1,0,
51	plasmid1	66609	A	G	G	225		DP=55;VDB=0.0060;AF1=0.5;AC1=1;DP4=16,8,22,9;MQ=43;FQ=225;PV4=0.78,1
52	plasmid1	66637	C	T	T	225		DP=71;VDB=0.0141;AF1=0.5;AC1=1;DP4=17,11,23,20;MQ=43;FQ=221;PV4=0.6,
53	plasmid1	66668	C	T	T	225		DP=80;VDB=0.0153;AF1=0.5;AC1=1;DP4=17,17,23,23;MQ=43;FQ=225;PV4=1,0,
54	plasmid2	2569	CAAAA	CAAAA	CAAAA	177		INDEL;DP=110;VDB=0.0166;AF1=1;AC1=2;DP4=0,0,51,59;MQ=49;FQ=-290
55	plasmid2	37805	TAAAA	TAAAA	TAAAA	104		INDEL;DP=31;VDB=0.0058;AF1=1;AC1=2;DP4=0,0,5,26;MQ=31;FQ=-128

W18-3: INFO

①VDBは、Variant Distance Biasの略。Higher values indicate higher likelihoods that the variant is distributed within the reads randomly.だそうです。(Biostars情報)。
 実用上、ここを見て判断することはほぼない

	A	B	C	D	E	F	G
27	#CHROM	POS	ID REF	ALT	QUAL	FI	INFO
28	chromosome	141598	A	G	17		DP=7;VDB=0.0077;AF1=0.5;AC1=1;DP4=4,1,1,1;MQ=39;FQ=20.1;PV4=1,1,0,1
29	chromosome	242819	T	C	20		DP=9;VDB=0.0111;AF1=0.5;AC1=1;DP4=5,2,1,1;MQ=49;FQ=23;PV4=1,1,0,1
30	chromosome	359141	T	C	4.1		DP=14;VDB=0.0037;AF1=0.4998;AC1=1;DP4=11,1,1,1;MQ=49;FQ=6.2;PV4=0.27
31	chromosome	463882	T	G	6.2		DP=13;VDB=0.0074;AF1=0.4999;AC1=1;DP4=2,9,1,1;MQ=48;FQ=8.65;PV4=0.42
32	chromosome	660031	C	T	8.6		DP=11;VDB=0.0063;AF1=0.5;AC1=1;DP4=7,2,1,1;MQ=49;FQ=11.3;PV4=0.49,0.2
33	chromosome	792963	T	C	4.1		DP=14;VDB=0.0153;AF1=0.4998;AC1=1;DP4=2,10,1,1;MQ=48;FQ=6.2;PV4=0.4,1
34	chromosome	887422	G	A	10		DP=11;VDB=0.0070;AF1=0.5;AC1=1;DP4=7,2,1,1;MQ=49;FQ=13.2;PV4=0.49,1,0
35	chromosome	913470	T	C	14		DP=9;VDB=0.0056;AF1=0.5;AC1=1;DP4=1,6,1,1;MQ=49;FQ=17.1;PV4=0.42,1,0.2
36	chromosome	965171	T	C	17		DP=20;VDB=0.0071;AF1=0.5;AC1=1;DP4=14,3,2,1;MQ=49;FQ=20.1;PV4=0.51,1,
37	chromosome	1134950	A	G	7.8		DP=23;VDB=0.0076;AF1=0.5;AC1=1;DP4=5,15,1,2;MQ=49;FQ=10.4;PV4=1,1,0,0
38	chromosome	1160048	A	G	29		DP=17;VDB=0.0054;AF1=0.5;AC1=1;DP4=6,8,1,2;MQ=48;FQ=32;PV4=1,1,0.041,
39	chromosome	1160095	A	G	19		DP=10;VDB=0.0033;AF1=0.5;AC1=1;DP4=6,2,1,1;MQ=48;FQ=22;PV4=1,1,0.007,
40	chromosome	1240248	GCCCC	GCCCCC	214		INDEL;DP=48;VDB=0.0137;AF1=1;AC1=2;DP4=1,0,21,26;MQ=50;FQ=-163;PV4=
41	chromosome	1455552	CAAAA	CAAAAA	214		INDEL;DP=58;VDB=0.0164;AF1=1;AC1=2;DP4=2,1,28,27;MQ=49;FQ=-165;PV4=
42	chromosome	1475721	T	C	52		DP=19;VDB=0.0021;AF1=0.5;AC1=1;DP4=11,4,2,2;MQ=48;FQ=55;PV4=0.56,1,2.
43	chromosome	1752680	G	A	29		DP=24;VDB=0.0129;AF1=0.5;AC1=1;DP4=14,6,2,2;MQ=48;FQ=32;PV4=0.58,1,0.
44	chromosome	1941380	T	C	7.8		DP=13;VDB=0.0046;AF1=0.5;AC1=1;DP4=9,2,1,1;MQ=50;FQ=10.4;PV4=0.42,1,0
45	chromosome	2052955	A	G	12		DP=8;VDB=0.0084;AF1=0.5;AC1=1;DP4=2,3,2,1;MQ=50;FQ=15.1;PV4=1,0.34,0.1
46	chromosome	2054084	TC	T	14		INDEL;DP=8;VDB=0.0009;AF1=0.5;AC1=1;DP4=3,3,1,1;MQ=47;FQ=16.6;PV4=1,4
47	chromosome	2056279	A	G	24		DP=6;VDB=0.0057;AF1=0.5;AC1=1;DP4=2,2,1,1;MQ=46;FQ=27;PV4=1,1,1,1
48	chromosome	2057564	T	C	8.6		DP=10;VDB=0.0041;AF1=0.5;AC1=1;DP4=4,4,1,1;MQ=33;FQ=11.3;PV4=1,1,1,1
49	chromosome	2058145	C	T,G	108		DP=14;VDB=0.0147;AF1=1;AC1=2;DP4=0,0,7,7;MQ=23;FQ=-66
50	plasmid1	37177	G	T	54		DP=49;VDB=0.0087;AF1=0.5;AC1=1;DP4=24,16,5,4;MQ=49;FQ=57;PV4=1,0,1,0.
51	plasmid1	66609	A	G	225		DP=55;VDB=0.0060;AF1=0.5;AC1=1;DP4=16,8,22,9;MQ=43;FQ=225;PV4=0.78,1
52	plasmid1	66637	C	T	225		DP=71;VDB=0.0141;AF1=0.5;AC1=1;DP4=17,11,23,20;MQ=43;FQ=221;PV4=0.6,
53	plasmid1	66668	C	T	225		DP=80;VDB=0.0153;AF1=0.5;AC1=1;DP4=17,17,23,23;MQ=43;FQ=225;PV4=1,0.
54	plasmid2	2569	CAAAA	CAAAA	177		INDEL;DP=110;VDB=0.0166;AF1=1;AC1=2;DP4=0,0,51,59;MQ=49;FQ=-290
55	plasmid2	37805	TAAAAA	TAAAA	104		INDEL;DP=31;VDB=0.0058;AF1=1;AC1=2;DP4=0,0,5,26;MQ=31;FQ=-128



①

W18-3: INFO

①AF1は、allele frequency(対立遺伝子頻度)のこと。よく見ると、最低でも0.4998よりも大きい値になっている。この値が約0.5以上のものがリストアップされているのだろう

	A	B	C	D	E	F	G	H
27	#CHROM	POS	ID	REF	ALT	QUAL	FI	INFO
28	chromosome	141598	A	G	G	17		DP=7;VDB=0.0077;AF1=0.5;AC1=1;DP4=4,1,1,1;MQ=39;FQ=20.1;PV4=1,1,0,1
29	chromosome	242819	T	C	C	20		DP=9;VDB=0.0111;AF1=0.5;AC1=1;DP4=5,2,1,1;MQ=49;FQ=23;PV4=1,1,0,1
30	chromosome	359141	T	C	C	4.1		DP=14;VDB=0.0037;AF1=0.4998;AC1=1;DP4=11,1,1,1;MQ=49;FQ=6.2;PV4=0.27
31	chromosome	463882	T	G	G	6.2		DP=13;VDB=0.0074;AF1=0.4999;AC1=1;DP4=2,9,1,1;MQ=48;FQ=8.65;PV4=0.42
32	chromosome	660031	C	T	T	8.6		DP=11;VDB=0.0063;AF1=0.5;AC1=1;DP4=7,2,1,1;MQ=49;FQ=11.3;PV4=0.49,0.2
33	chromosome	792963	T	C	C	4.1		DP=14;VDB=0.0153;AF1=0.4998;AC1=1;DP4=2,10,1,1;MQ=48;FQ=6.2;PV4=0.4,1
34	chromosome	887422	G	A	A	10		DP=11;VDB=0.0070;AF1=0.5;AC1=1;DP4=7,2,1,1;MQ=49;FQ=13.2;PV4=0.49,1,0
35	chromosome	913470	T	C	C	14		DP=9;VDB=0.0056;AF1=0.5;AC1=1;DP4=1,6,1,1;MQ=49;FQ=17.1;PV4=0.42,1,0.2
36	chromosome	965171	T	C	C	17		DP=20;VDB=0.0071;AF1=0.5;AC1=1;DP4=14,3,2,1;MQ=49;FQ=20.1;PV4=0.51,1,
37	chromosome	1134950	A	G	G	7.8		DP=23;VDB=0.0076;AF1=0.5;AC1=1;DP4=5,15,1,2;MQ=49;FQ=10.4;PV4=1,1,0,0
38	chromosome	1160048	A	G	G	29		DP=17;VDB=0.0054;AF1=0.5;AC1=1;DP4=6,8,1,2;MQ=48;FQ=32;PV4=1,1,0.041,
39	chromosome	1160095	A	G	G	19		DP=10;VDB=0.0033;AF1=0.5;AC1=1;DP4=6,2,1,1;MQ=48;FQ=22;PV4=1,1,0.007,
40	chromosome	1240248	GCCCC	GCCCC	GCCCC	214		INDEL;DP=48;VDB=0.0137;AF1=1;AC1=2;DP4=1,0,21,26;MQ=50;FQ=-163;PV4=
41	chromosome	1455552	CAAAA	CAAAA	CAAAA	214		INDEL;DP=58;VDB=0.0164;AF1=1;AC1=2;DP4=2,1,28,27;MQ=49;FQ=-165;PV4=
42	chromosome	1475721	T	C	C	52		DP=19;VDB=0.0021;AF1=0.5;AC1=1;DP4=11,4,2,2;MQ=48;FQ=55;PV4=0.56,1,2,
43	chromosome	1752680	G	A	A	29		DP=24;VDB=0.0129;AF1=0.5;AC1=1;DP4=14,6,2,2;MQ=48;FQ=32;PV4=0.58,1,0,
44	chromosome	1941380	T	C	C	7.8		DP=13;VDB=0.0046;AF1=0.5;AC1=1;DP4=9,2,1,1;MQ=50;FQ=10.4;PV4=0.42,1,0
45	chromosome	2052955	A	G	G	12		DP=8;VDB=0.0084;AF1=0.5;AC1=1;DP4=2,3,2,1;MQ=50;FQ=15.1;PV4=1,0.34,0.1
46	chromosome	2054084	TC	T	T	14		INDEL;DP=8;VDB=0.0009;AF1=0.5;AC1=1;DP4=3,3,1,1;MQ=47;FQ=16.6;PV4=1,4
47	chromosome	2056279	A	G	G	24		DP=6;VDB=0.0057;AF1=0.5;AC1=1;DP4=2,2,1,1;MQ=46;FQ=27;PV4=1,1,1,1
48	chromosome	2057564	T	C	C	8.6		DP=10;VDB=0.0041;AF1=0.5;AC1=1;DP4=4,4,1,1;MQ=33;FQ=11.3;PV4=1,1,1,1
49	chromosome	2058145	C	T,G	T,G	108		DP=14;VDB=0.0147;AF1=1;AC1=2;DP4=0,0,7,7;MQ=23;FQ=-66
50	plasmid1	37177	G	T	T	54		DP=49;VDB=0.0087;AF1=0.5;AC1=1;DP4=24,16,5,4;MQ=49;FQ=57;PV4=1,0,1,0,
51	plasmid1	66609	A	G	G	225		DP=55;VDB=0.0060;AF1=0.5;AC1=1;DP4=16,8,22,9;MQ=43;FQ=225;PV4=0.78,1
52	plasmid1	66637	C	T	T	225		DP=71;VDB=0.0141;AF1=0.5;AC1=1;DP4=17,11,23,20;MQ=43;FQ=221;PV4=0.6,
53	plasmid1	66668	C	T	T	225		DP=80;VDB=0.0153;AF1=0.5;AC1=1;DP4=17,17,23,23;MQ=43;FQ=225;PV4=1,0,
54	plasmid2	2569	CAAAA	CAAAA	CAAAA	177		INDEL;DP=110;VDB=0.0166;AF1=1;AC1=2;DP4=0,0,51,59;MQ=49;FQ=-290
55	plasmid2	37805	TAAAAA	TAAAAA	TAAAAA	104		INDEL;DP=31;VDB=0.0058;AF1=1;AC1=2;DP4=0,0,5,26;MQ=31;FQ=-128

W18-3: INFO

①AC1は、genotype中のallele count (対立遺伝子数)のこと。ほとんどがAC1=1だが、赤矢印で示す5個がAC1=2になっていることがわかる

	A	B	C	D	E	F	G	H
27	#CHROM	POS	ID	REF	ALT	QUAL	FI	INFO
28	chromosome	141598	A	G	G	17		DP=7;VDB=0.0077;AF1=0.5;AC1=1;DP4=4,1,1,1;MQ=39;FQ=20.1;PV4=1,1,0,1
29	chromosome	242819	T	C	C	20		DP=9;VDB=0.0111;AF1=0.5;AC1=1;DP4=5,2,1,1;MQ=49;FQ=23;PV4=1,1,0,1
30	chromosome	359141	T	C	C	4.1		DP=14;VDB=0.0037;AF1=0.4998;AC1=1;DP4=11,1,1,1;MQ=49;FQ=6.2;PV4=0.27
31	chromosome	463882	T	G	G	6.2		DP=13;VDB=0.0074;AF1=0.4999;AC1=1;DP4=2,9,1,1;MQ=48;FQ=8.65;PV4=0.42
32	chromosome	660031	C	T	T	8.6		DP=11;VDB=0.0063;AF1=0.5;AC1=1;DP4=7,2,1,1;MQ=49;FQ=11.3;PV4=0.49,0.2
33	chromosome	792963	T	C	C	4.1		DP=14;VDB=0.0153;AF1=0.4998;AC1=1;DP4=2,10,1,1;MQ=48;FQ=6.2;PV4=0.4,1
34	chromosome	887422	G	A	A	10		DP=11;VDB=0.0070;AF1=0.5;AC1=1;DP4=7,2,1,1;MQ=49;FQ=13.2;PV4=0.49,1,0
35	chromosome	913470	T	C	C	14		DP=9;VDB=0.0056;AF1=0.5;AC1=1;DP4=1,6,1,1;MQ=49;FQ=17.1;PV4=0.42,1,0.2
36	chromosome	965171	T	C	C	17		DP=20;VDB=0.0071;AF1=0.5;AC1=1;DP4=14,3,2,1;MQ=49;FQ=20.1;PV4=0.51,1,
37	chromosome	1134950	A	G	G	7.8		DP=23;VDB=0.0076;AF1=0.5;AC1=1;DP4=5,15,1,2;MQ=49;FQ=10.4;PV4=1,1,0,0
38	chromosome	1160048	A	G	G	29		DP=17;VDB=0.0054;AF1=0.5;AC1=1;DP4=6,8,1,2;MQ=48;FQ=32;PV4=1,1,0.041,
39	chromosome	1160095	A	G	G	19		DP=10;VDB=0.0033;AF1=0.5;AC1=1;DP4=6,2,1,1;MQ=48;FQ=22;PV4=1,1,0.007,
40	chromosome	1240248	GCCCC	GCCCC	GCCCC	4	←	INDEL;DP=48;VDB=0.0137;AF1=1;AC1=2;DP4=1,0,21,26;MQ=50;FQ=-163;PV4=
41	chromosome	1455552	CAAAA	CAAAA	CAAAA	4	←	INDEL;DP=58;VDB=0.0164;AF1=1;AC1=2;DP4=2,1,28,27;MQ=49;FQ=-165;PV4=
42	chromosome	1475721	T	C	C	52		DP=19;VDB=0.0021;AF1=0.5;AC1=1;DP4=11,4,2,2;MQ=48;FQ=55;PV4=0.56,1,2,
43	chromosome	1752680	G	A	A	29		DP=24;VDB=0.0129;AF1=0.5;AC1=1;DP4=14,6,2,2;MQ=48;FQ=32;PV4=0.58,1,0,
44	chromosome	1941380	T	C	C	7.8		DP=13;VDB=0.0046;AF1=0.5;AC1=1;DP4=9,2,1,1;MQ=50;FQ=10.4;PV4=0.42,1,0
45	chromosome	2052955	A	G	G	12		DP=8;VDB=0.0084;AF1=0.5;AC1=1;DP4=2,3,2,1;MQ=50;FQ=15.1;PV4=1,0.34,0.1
46	chromosome	2054084	TC	T	T	14		INDEL;DP=8;VDB=0.0009;AF1=0.5;AC1=1;DP4=3,3,1,1;MQ=47;FQ=16.6;PV4=1,4
47	chromosome	2056279	A	G	G	24		DP=6;VDB=0.0057;AF1=0.5;AC1=1;DP4=2,2,1,1;MQ=46;FQ=27;PV4=1,1,1,1
48	chromosome	2057564	T	C	C	8.6		DP=10;VDB=0.0041;AF1=0.5;AC1=1;DP4=4,4,1,1;MQ=33;FQ=11.3;PV4=1,1,1,1
49	chromosome	2058145	C	T,G	T,G	08	←	DP=14;VDB=0.0147;AF1=1;AC1=2;DP4=0,0,7,7;MQ=23;FQ=-66
50	plasmid1	37177	G	T	T	54		DP=49;VDB=0.0087;AF1=0.5;AC1=1;DP4=24,16,5,4;MQ=49;FQ=57;PV4=1,0,1,0,
51	plasmid1	66609	A	G	G	225		DP=55;VDB=0.0060;AF1=0.5;AC1=1;DP4=16,8,22,9;MQ=43;FQ=225;PV4=0.78,1
52	plasmid1	66637	C	T	T	225		DP=71;VDB=0.0141;AF1=0.5;AC1=1;DP4=17,11,23,20;MQ=43;FQ=221;PV4=0.6,
53	plasmid1	66668	C	T	T	225		DP=80;VDB=0.0153;AF1=0.5;AC1=1;DP4=17,17,23,23;MQ=43;FQ=225;PV4=1,0,
54	plasmid2	2569	CAAAA	CAAAA	CAAAA	77	←	INDEL;DP=110;VDB=0.0166;AF1=1;AC1=2;DP4=0,0,51,59;MQ=49;FQ=-290
55	plasmid2	37805	TAAAAA	TAAAAA	TAAAAA	04	←	INDEL;DP=31;VDB=0.0058;AF1=1;AC1=2;DP4=0,0,5,26;MQ=31;FQ=-128

W18-3: INFO

①DP4は、1番左のDPを割り振ったもの。DP4=REFのforward,REFのreverse,ALTのforward,ALTのreverseの順番

	A	B	C	D	E	F	G	H
27	#CHROM	POS	ID	REF	ALT	QUAL	FI	INFO
28	chromosome	141598	A	G	17	DP=7;VDB=0.0077;AF1=0.5;AC1=1;DP4=4,1,1,1;MQ=39;FQ=20.1;PV4=1,1,0,1		
29	chromosome	242819	T	C	20	DP=9;VDB=0.0111;AF1=0.5;AC1=1;DP4=5,2,1,1;MQ=49;FQ=23;PV4=1,1,0,1		
30	chromosome	359141	T	C	4.1	DP=14;VDB=0.0037;AF1=0.4998;AC1=1;DP4=11,1,1,1;MQ=49;FQ=6.2;PV4=0.27		
31	chromosome	463882	T	G	6.2	DP=13;VDB=0.0074;AF1=0.4999;AC1=1;DP4=2,9,1,1;MQ=48;FQ=8.65;PV4=0.42		
32	chromosome	660031	C	T	8.6	DP=11;VDB=0.0063;AF1=0.5;AC1=1;DP4=7,2,1,1;MQ=49;FQ=11.3;PV4=0.49,0.2		
33	chromosome	792963	T	C	4.1	DP=14;VDB=0.0153;AF1=0.4998;AC1=1;DP4=2,10,1,1;MQ=48;FQ=6.2;PV4=0.4,1		
34	chromosome	887422	G	A	10	DP=11;VDB=0.0070;AF1=0.5;AC1=1;DP4=7,2,1,1;MQ=49;FQ=13.2;PV4=0.49,1,0		
35	chromosome	913470	T	C	14	DP=9;VDB=0.0056;AF1=0.5;AC1=1;DP4=1,6,1,1;MQ=49;FQ=17.1;PV4=0.42,1,0,2		
36	chromosome	965171	T	C	17	DP=20;VDB=0.0071;AF1=0.5;AC1=1;DP4=14,3,2,1;MQ=49;FQ=20.1;PV4=0.51,1,1		
37	chromosome	1134950	A	G	7.8	DP=23;VDB=0.0076;AF1=0.5;AC1=1;DP4=5,15,1,2;MQ=49;FQ=10.4;PV4=1,1,0,0		
38	chromosome	1160048	A	G	29	DP=17;VDB=0.0054;AF1=0.5;AC1=1;DP4=6,8,1,2;MQ=48;FQ=32;PV4=1,1,0,041		
39	chromosome	1160095	A	G	19	DP=10;VDB=0.0033;AF1=0.5;AC1=1;DP4=6,2,1,1;MQ=48;FQ=22;PV4=1,1,0,007		
40	chromosome	1240248	GCCCC	GCCCCC	214	INDEL;DP=48;VDB=0.0137;AF1=1;AC1=2;DP4=1,0,21,26;MQ=50;FQ=-163;PV4=		
41	chromosome	1455552	CAAAA	CAAAAA	214	INDEL;DP=58;VDB=0.0164;AF1=1;AC1=2;DP4=2,1,28,27;MQ=49;FQ=-165;PV4=		
42	chromosome	1475721	T	C	52	DP=19;VDB=0.0021;AF1=0.5;AC1=1;DP4=11,4,2,2;MQ=48;FQ=55;PV4=0.56,1,2		
43	chromosome	1752680	G	A	29	DP=24;VDB=0.0129;AF1=0.5;AC1=1;DP4=14,6,2,2;MQ=48;FQ=32;PV4=0.58,1,0		
44	chromosome	1941380	T	C	7.8	DP=13;VDB=0.0046;AF1=0.5;AC1=1;DP4=9,2,1,1;MQ=50;FQ=10.4;PV4=0.42,1,0		
45	chromosome	2052955	A	G	12	DP=8;VDB=0.0084;AF1=0.5;AC1=1;DP4=2,3,2,1;MQ=50;FQ=15.1;PV4=1,0,34,0,1		
46	chromosome	2054084	TC	T	14	INDEL;DP=8;VDB=0.0009;AF1=0.5;AC1=1;DP4=3,3,1,1;MQ=47;FQ=16.6;PV4=1,4		
47	chromosome	2056279	A	G	24	DP=6;VDB=0.0057;AF1=0.5;AC1=1;DP4=2,2,1,1;MQ=46;FQ=27;PV4=1,1,1,1		
48	chromosome	2057564	T	C	8.6	DP=10;VDB=0.0041;AF1=0.5;AC1=1;DP4=4,4,1,1;MQ=33;FQ=11.3;PV4=1,1,1,1		
49	chromosome	2058145	C	T,G	108	DP=14;VDB=0.0147;AF1=1;AC1=2;DP4=0,0,7,7;MQ=23;FQ=-66		
50	plasmid1	37177	G	T	54	DP=49;VDB=0.0087;AF1=0.5;AC1=1;DP4=24,16,5,4;MQ=49;FQ=57;PV4=1,0,1,0		
51	plasmid1	66609	A	G	225	DP=55;VDB=0.0060;AF1=0.5;AC1=1;DP4=16,8,22,9;MQ=43;FQ=225;PV4=0.78,1		
52	plasmid1	66637	C	T	225	DP=71;VDB=0.0141;AF1=0.5;AC1=1;DP4=17,11,23,20;MQ=43;FQ=221;PV4=0.6		
53	plasmid1	66668	C	T	225	DP=80;VDB=0.0153;AF1=0.5;AC1=1;DP4=17,17,23,23;MQ=43;FQ=225;PV4=1,0		
54	plasmid2	2569	CAAAA	CAAAA	177	INDEL;DP=110;VDB=0.0166;AF1=1;AC1=2;DP4=0,0,51,59;MQ=49;FQ=-290		
55	plasmid2	37805	TAAAA	TAAAA	104	INDEL;DP=31;VDB=0.0058;AF1=1;AC1=2;DP4=0,0,5,26;MQ=31;FQ=-128		

W18-3: INF

例えば①「DP4=4,1,1,1」は②DP=7を割り振ったもの。概要配列と同じ塩基(REF;つまり③A)のforward鎖側でマップされたのが4リード(depth=4)、REFのreverse側がdepth=1、変異塩基(ALT;つまり④G)でforward側がdepth=1、変異塩基(ALT;つまり④G)でreverse側がdepth=1

	A	B	C	D				
#	CHROM	POS	ID	REF	ALT	QUAL	FI	INFO
27	chromosome	141598	A	G		17		DP=7;VDB=0.0077;AF1=0.5;AC1=1;DP4=4,1,1,1;MQ=39;FQ=20.1;PV4=1,1,0,1
28	chromosome	242819	A	A		20		DP=9;VDB=0.0111;AF1=0.5;AC1=1;DP4=9,1,1,1;MQ=49;FQ=23;PV4=1,1,0,1
29	chromosome	359141	T	C		4.1		DP=14;VDB=0.0037;AF1=0.4998;AC1=1;DP4=11,1,1,1;MQ=49;FQ=6.2;PV4=0.27
30	chromosome	463882	T	G		6.2		DP=13;VDB=0.0074;AF1=0.4999;AC1=1;DP4=2,9,1,1;MQ=48;FQ=8.65;PV4=0.42
31	chromosome	660031	C	T		8.6		DP=11;VDB=0.0063;AF1=0.5;AC1=1;DP4=7,2,1,1;MQ=49;FQ=11.3;PV4=0.49,0.2
32	chromosome	792963	T	C		4.1		DP=14;VDB=0.0153;AF1=0.4998;AC1=1;DP4=2,10,1,1;MQ=48;FQ=6.2;PV4=0.4,1
33	chromosome	887422	G	A		10		DP=11;VDB=0.0070;AF1=0.5;AC1=1;DP4=7,2,1,1;MQ=49;FQ=13.2;PV4=0.49,1,0
34	chromosome	913470	T	C		14		DP=9;VDB=0.0056;AF1=0.5;AC1=1;DP4=1,6,1,1;MQ=49;FQ=17.1;PV4=0.42,1,0.2
35	chromosome	965171	T	C		17		DP=20;VDB=0.0071;AF1=0.5;AC1=1;DP4=14,3,2,1;MQ=49;FQ=20.1;PV4=0.51,1,1
36	chromosome	1134950	A	G		7.8		DP=23;VDB=0.0076;AF1=0.5;AC1=1;DP4=5,15,1,2;MQ=49;FQ=10.4;PV4=1,1,0,0
37	chromosome	1160048	A	G		29		DP=17;VDB=0.0054;AF1=0.5;AC1=1;DP4=6,8,1,2;MQ=48;FQ=32;PV4=1,1,0,041
38	chromosome	1160095	A	G		19		DP=10;VDB=0.0033;AF1=0.5;AC1=1;DP4=6,2,1,1;MQ=48;FQ=22;PV4=1,1,0,007
39	chromosome	1240248	GCCCC	GCCCC		214		INDEL;DP=48;VDB=0.0137;AF1=1;AC1=2;DP4=1,0,21,26;MQ=50;FQ=-163;PV4=
40	chromosome	1455552	CAAAA	CAAAA		214		INDEL;DP=58;VDB=0.0164;AF1=1;AC1=2;DP4=2,1,28,27;MQ=49;FQ=-165;PV4=
41	chromosome	1475721	T	C		52		DP=19;VDB=0.0021;AF1=0.5;AC1=1;DP4=11,4,2,2;MQ=48;FQ=55;PV4=0.56,1,2
42	chromosome	1752680	G	A		29		DP=24;VDB=0.0129;AF1=0.5;AC1=1;DP4=14,6,2,2;MQ=48;FQ=32;PV4=0.58,1,0
43	chromosome	1941380	T	C		7.8		DP=13;VDB=0.0046;AF1=0.5;AC1=1;DP4=9,2,1,1;MQ=50;FQ=10.4;PV4=0.42,1,0
44	chromosome	2052955	A	G		12		DP=8;VDB=0.0084;AF1=0.5;AC1=1;DP4=2,3,2,1;MQ=50;FQ=15.1;PV4=1,0,34,0.1
45	chromosome	2054084	TC	T		14		INDEL;DP=8;VDB=0.0009;AF1=0.5;AC1=1;DP4=3,3,1,1;MQ=47;FQ=16.6;PV4=1,4
46	chromosome	2056279	A	G		24		DP=6;VDB=0.0057;AF1=0.5;AC1=1;DP4=2,2,1,1;MQ=46;FQ=27;PV4=1,1,1,1
47	chromosome	2057564	T	C		8.6		DP=10;VDB=0.0041;AF1=0.5;AC1=1;DP4=4,4,1,1;MQ=33;FQ=11.3;PV4=1,1,1,1
48	chromosome	2058145	C	T,G		108		DP=14;VDB=0.0147;AF1=1;AC1=2;DP4=0,0,7,7;MQ=23;FQ=-66
49	plasmid1	37177	G	T		54		DP=49;VDB=0.0087;AF1=0.5;AC1=1;DP4=24,16,5,4;MQ=49;FQ=57;PV4=1,0,1,0
50	plasmid1	66609	A	G		225		DP=55;VDB=0.0060;AF1=0.5;AC1=1;DP4=16,8,22,9;MQ=43;FQ=225;PV4=0.78,1
51	plasmid1	66637	C	T		225		DP=71;VDB=0.0141;AF1=0.5;AC1=1;DP4=17,11,23,20;MQ=43;FQ=221;PV4=0.6
52	plasmid1	66668	C	T		225		DP=80;VDB=0.0153;AF1=0.5;AC1=1;DP4=17,17,23,23;MQ=43;FQ=225;PV4=1,0
53	plasmid2	2569	CAAAA	CAAAA		177		INDEL;DP=110;VDB=0.0166;AF1=1;AC1=2;DP4=0,0,51,59;MQ=49;FQ=-290
54	plasmid2	37805	TAAAA	TAAAA		104		INDEL;DP=31;VDB=0.0058;AF1=1;AC1=2;DP4=0,0,5,26;MQ=31;FQ=-128

W18-3: INFO

概要配列を変更するかどうかの判断基準の1つとして、①を眺める。この場合は、②REFのAを支持するリード数が4+1=5、③ALTのGを支持するリード数が1+1=2。したがって、概要配列であるREF側のAをGには変更しないという結論を下す

	A	B	C	D	E	F
27	#CHROM	POS	ID	REF	ALT	QUAL
28	chromosome	141598	A	G	17	DP=7;VDB=0.0077;AF1=0.5;AC1=1;DP4=4,1,1,1;MQ=39;FQ=20.1;PV4=1,1,0,1
29	chromosome	242819	A	G	20	DP=9;VDB=0.0111;AF1=0.5;AC1=1;DP4=2,2,1,1;MQ=49;FQ=23;PV4=1,1,0,1
30	chromosome	359141	T	C	4.1	DP=14;VDB=0.0037;AF1=0.4998;AC1=1;DP4=11,1,1,1;MQ=49;FQ=6.2;PV4=0.27
31	chromosome	463882	T	G	6.2	DP=13;VDB=0.0074;AF1=0.4999;AC1=1;DP4=2,9,1,1;MQ=48;FQ=8.65;PV4=0.42
32	chromosome	660031	C	T	8.6	DP=11;VDB=0.0063;AF1=0.5;AC1=1;DP4=7,2,1,1;MQ=49;FQ=11.3;PV4=0.49,0.2
33	chromosome	792963	T	C	4.1	DP=14;VDB=0.0153;AF1=0.4998;AC1=1;DP4=2,10,1,1;MQ=48;FQ=6.2;PV4=0.4,1
34	chromosome	887422	G	A	10	DP=11;VDB=0.0070;AF1=0.5;AC1=1;DP4=7,2,1,1;MQ=49;FQ=13.2;PV4=0.49,1,0
35	chromosome	913470	T	C	14	DP=9;VDB=0.0056;AF1=0.5;AC1=1;DP4=1,6,1,1;MQ=49;FQ=17.1;PV4=0.42,1,0,2
36	chromosome	965171	T	C	17	DP=20;VDB=0.0071;AF1=0.5;AC1=1;DP4=14,3,2,1;MQ=49;FQ=20.1;PV4=0.51,1,1
37	chromosome	1134950	A	G	7.8	DP=23;VDB=0.0076;AF1=0.5;AC1=1;DP4=5,15,1,2;MQ=49;FQ=10.4;PV4=1,1,0,0
38	chromosome	1160048	A	G	29	DP=17;VDB=0.0054;AF1=0.5;AC1=1;DP4=6,8,1,2;MQ=48;FQ=32;PV4=1,1,0,041
39	chromosome	1160095	A	G	19	DP=10;VDB=0.0033;AF1=0.5;AC1=1;DP4=6,2,1,1;MQ=48;FQ=22;PV4=1,1,0,007
40	chromosome	1240248	GCCCC	GCCCCC	214	INDEL;DP=48;VDB=0.0137;AF1=1;AC1=2;DP4=1,0,21,26;MQ=50;FQ=-163;PV4=
41	chromosome	1455552	CAAAA	CAAAAA	214	INDEL;DP=58;VDB=0.0164;AF1=1;AC1=2;DP4=2,1,28,27;MQ=49;FQ=-165;PV4=
42	chromosome	1475721	T	C	52	DP=19;VDB=0.0021;AF1=0.5;AC1=1;DP4=11,4,2,2;MQ=48;FQ=55;PV4=0.56,1,2
43	chromosome	1752680	G	A	29	DP=24;VDB=0.0129;AF1=0.5;AC1=1;DP4=14,6,2,2;MQ=48;FQ=32;PV4=0.58,1,0
44	chromosome	1941380	T	C	7.8	DP=13;VDB=0.0046;AF1=0.5;AC1=1;DP4=9,2,1,1;MQ=50;FQ=10.4;PV4=0.42,1,0
45	chromosome	2052955	A	G	12	DP=8;VDB=0.0084;AF1=0.5;AC1=1;DP4=2,3,2,1;MQ=50;FQ=15.1;PV4=1,0,34,0,1
46	chromosome	2054084	TC	T	14	INDEL;DP=8;VDB=0.0009;AF1=0.5;AC1=1;DP4=3,3,1,1;MQ=47;FQ=16.6;PV4=1,4
47	chromosome	2056279	A	G	24	DP=6;VDB=0.0057;AF1=0.5;AC1=1;DP4=2,2,1,1;MQ=46;FQ=27;PV4=1,1,1,1
48	chromosome	2057564	T	C	8.6	DP=10;VDB=0.0041;AF1=0.5;AC1=1;DP4=4,4,1,1;MQ=33;FQ=11.3;PV4=1,1,1,1
49	chromosome	2058145	C	T,G	108	DP=14;VDB=0.0147;AF1=1;AC1=2;DP4=0,0,7,7;MQ=23;FQ=-66
50	plasmid1	37177	G	T	54	DP=49;VDB=0.0087;AF1=0.5;AC1=1;DP4=24,16,5,4;MQ=49;FQ=57;PV4=1,0,1,0
51	plasmid1	66609	A	G	225	DP=55;VDB=0.0060;AF1=0.5;AC1=1;DP4=16,8,22,9;MQ=43;FQ=225;PV4=0.78,1
52	plasmid1	66637	C	T	225	DP=71;VDB=0.0141;AF1=0.5;AC1=1;DP4=17,11,23,20;MQ=43;FQ=221;PV4=0.6
53	plasmid1	66668	C	T	225	DP=80;VDB=0.0153;AF1=0.5;AC1=1;DP4=17,17,23,23;MQ=43;FQ=225;PV4=1,0
54	plasmid2	2569	CAAAA	CAAAA	177	INDEL;DP=110;VDB=0.0166;AF1=1;AC1=2;DP4=0,0,51,59;MQ=49;FQ=-290
55	plasmid2	37805	TAAAA	TAAAA	104	INDEL;DP=31;VDB=0.0058;AF1=1;AC1=2;DP4=0,0,5,26;MQ=31;FQ=-128

W18-3: INFO

概要配列を変更するかどうかの判断基準の1つとして、①を眺める。
 この場合は、②REFのGCCCCを支持するリード数が1+0=1、③ALT
 のGCCCCを支持するリード数が21+26=47。したがって、概要配列
 であるREF側のGCCCCをGCCCCCに変更したほうがよいと判断する

	A	B	C	D		
#	CHROM	POS	ID	REF	ALT	QUALFITINFO
27	chromosome	141598	A	G		17 DP=7;VDB=0.0077;AF1=0.5;AC1=1;DP4=4,1,1,1;MQ=39;FQ=20.1;PV4=1,1,0,1
29	chromosome	242819	T	C		20 DP=9;VDB=0.0111;AF1=0.5;AC1=1;DP4=5,2,1,1;MQ=49;FQ=23;PV4=1,1,0,1
30	chromosome	359141	T	C		4.1 DP=14;VDB=0.0037;AF1=0.4998;AC1=1;DP4=11,1,1,1;MQ=49;FQ=6.2;PV4=0.27
31	chromosome	463882	T	G		6.2 DP=13;VDB=0.0074;AF1=0.4999;AC1=1;DP4=2,9,1,1;MQ=48;FQ=8.65;PV4=0.42
32	chromosome	660031	C	T		8.6 DP=11;VDB=0.0063;AF1=0.5;AC1=1;DP4=7,2,1,1;MQ=49;FQ=11.3;PV4=0.49,0.2
33	chromosome	792963	T	C		4.1 DP=14;VDB=0.0153;AF1=0.4998;AC1=1;DP4=2,10,1,1;MQ=48;FQ=6.2;PV4=0.4,1
34	chromosome	887422	G	A		10 DP=11;VDB=0.0070;AF1=0.5;AC1=1;DP4=7,2,1,1;MQ=49;FQ=13.2;PV4=0.49,1,0
35	chromosome	913470	T	C		14 DP=9;VDB=0.0056;AF1=0.5;AC1=1;DP4=1,6,1,1;MQ=49;FQ=17.1;PV4=0.42,1,0,2
36	chromosome	965171	T	C		17 DP=20;VDB=0.0071;AF1=0.5;AC1=1;DP4=1,4,3,2,1;MQ=49;FQ=20.1;PV4=0.51,1,1
37	chromosome	1134950	A	G		7.8 DP=23;VDB=0.0076;AF1=0.5;AC1=1;DP4=5,15,1,2;MQ=49;FQ=10.4;PV4=1,1,0,0
38	chromosome	1160048	A	G		29 DP=17;VDB=0.0054;AF1=0.5;AC1=1;DP4=6,8,1,2;MQ=48;FQ=32;PV4=1,1,0,0,41
39	chromosome	1160095	A	G		19 DP=10;VDB=0.0033;AF1=0.5;AC1=1;DP4=6,2,1,1;MQ=48;FQ=22;PV4=1,1,0,0,07
40	chromosome	1240248	GCCCC	GCCCCC		214 INDEL;DP=48;VDB=0.0137;AF1=1;AC1=2;DP4=1,0,21,26;MQ=50;FQ=-163;PV4=
41	chromosome	1455552	AAAA	AAAA		214 INDEL;DP=58;VDB=0.0164;AF1=1;AC1=2;DP4=1,0,28,27;MQ=49;FQ=-165;PV4=
42	chromosome	1475721	T	C		52 DP=19;VDB=0.0021;AF1=0.5;AC1=1;DP4=11,4,2,2;MQ=48;FQ=55;PV4=0.56,1,2
43	chromosome	1752680	G	A		29 DP=24;VDB=0.0129;AF1=0.5;AC1=1;DP4=1,4,6,2,2;MQ=48;FQ=32;PV4=0.58,1,0
44	chromosome	1941380	T	C		7.8 DP=13;VDB=0.0046;AF1=0.5;AC1=1;DP4=9,2,1,1;MQ=50;FQ=10.4;PV4=0.42,1,0
45	chromosome	2052955	A	G		12 DP=8;VDB=0.0084;AF1=0.5;AC1=1;DP4=2,3,2,1;MQ=50;FQ=15.1;PV4=1,0,34,0,1
46	chromosome	2054084	TC	T		14 INDEL;DP=8;VDB=0.0009;AF1=0.5;AC1=1;DP4=3,3,1,1;MQ=47;FQ=16.6;PV4=1,4
47	chromosome	2056279	A	G		24 DP=6;VDB=0.0057;AF1=0.5;AC1=1;DP4=2,2,1,1;MQ=46;FQ=27;PV4=1,1,1,1
48	chromosome	2057564	T	C		8.6 DP=10;VDB=0.0041;AF1=0.5;AC1=1;DP4=4,4,1,1;MQ=33;FQ=11.3;PV4=1,1,1,1
49	chromosome	2058145	C	T,G		108 DP=14;VDB=0.0147;AF1=1;AC1=2;DP4=0,0,7,7;MQ=23;FQ=-66
50	plasmid1	37177	G	T		54 DP=49;VDB=0.0087;AF1=0.5;AC1=1;DP4=2,4,16,5,4;MQ=49;FQ=57;PV4=1,0,1,0
51	plasmid1	66609	A	G		225 DP=55;VDB=0.0060;AF1=0.5;AC1=1;DP4=1,6,8,22,9;MQ=43;FQ=225;PV4=0.78,1
52	plasmid1	66637	C	T		225 DP=71;VDB=0.0141;AF1=0.5;AC1=1;DP4=1,7,11,23,20;MQ=43;FQ=221;PV4=0.6
53	plasmid1	66668	C	T		225 DP=80;VDB=0.0153;AF1=0.5;AC1=1;DP4=1,7,17,23,23;MQ=43;FQ=225;PV4=1,0
54	plasmid2	2569	CAAAA	CAAAA		177 INDEL;DP=110;VDB=0.0166;AF1=1;AC1=2;DP4=0,0,51,59;MQ=49;FQ=-290
55	plasmid2	37805	TAAAA	TAAAA		104 INDEL;DP=31;VDB=0.0058;AF1=1;AC1=2;DP4=0,0,5,26;MQ=31;FQ=-128



W18-4: 全体

もう一度全体を眺め直す。VCFファイル中にリストアップされているのは、概要配列と異なる部分のみ。概要配列と同じ塩基がマップされた箇所は、①がおそらく0/0と表現されるが、ないので妥当

	A	B	C	D	E	F	G	H	I	J
27	#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	①
28	chromosome	141598		A	G	17.1		DP=7;VDB=0.0077;AF1=0.5;AC1	GT:PL:GQ	0/1:47,0,138:50
29	chromosome	242819		T	C	20		DP=9;VDB=0.0111;AF1=0.5;AC1	GT:PL:GQ	0/1:50,0,139:53
30	chromosome	359141		T	C	4.13		DP=14;VDB=0.0037;AF1=0.4998	GT:PL:GQ	0/1:32,0,187:31
31	chromosome	463882		T	G	6.2		DP=13;VDB=0.0074;AF1=0.4999	GT:PL:GQ	0/1:35,0,216:36
32	chromosome	660031		C	T	8.64		DP=11;VDB=0.0063;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,197:40
33	chromosome	792963		T	C	4.13		DP=14;VDB=0.0153;AF1=0.4998	GT:PL:GQ	0/1:32,0,200:31
34	chromosome	887422		G	A	10.4		DP=11;VDB=0.0070;AF1=0.5;AC	GT:PL:GQ	0/1:40,0,188:42
35	chromosome	913470		T	C	14.2		DP=9;VDB=0.0056;AF1=0.5;AC1	GT:PL:GQ	0/1:44,0,144:47
36	chromosome	965171		T	C	17.1		DP=20;VDB=0.0071;AF1=0.5;AC	GT:PL:GQ	0/1:47,0,249:50
37	chromosome	1134950		A	G	7.8		DP=23;VDB=0.0076;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,255:39
38	chromosome	1160048		A	G	29		DP=17;VDB=0.0054;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
39	chromosome	1160095		A	G	19.1		DP=10;VDB=0.0033;AF1=0.5;AC	GT:PL:GQ	0/1:49,0,191:52
40	chromosome	1240248		GCCCC	GCCCCC	214		INDEL;DP=48;VDB=0.0137;AF1=	GT:PL:GQ	1/1:255,128,0:99
41	chromosome	1455552		CAAAA	CAAAAA	214		INDEL;DP=58;VDB=0.0164;AF1=	GT:PL:GQ	1/1:255,130,0:99
42	chromosome	1475721		T	C	52		DP=19;VDB=0.0021;AF1=0.5;AC	GT:PL:GQ	0/1:82,0,246:85
43	chromosome	1752680		G	A	29		DP=24;VDB=0.0129;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
44	chromosome	1941380		T	C	7.8		DP=13;VDB=0.0046;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,214:39
45	chromosome	2052955		A	G	12.3		DP=8;VDB=0.0084;AF1=0.5;AC1	GT:PL:GQ	0/1:42,0,103:44
46	chromosome	2054084		TC	T	13.7		INDEL;DP=8;VDB=0.0009;AF1=0	GT:PL:GQ	0/1:51,0,174:54
47	chromosome	2056279		A	G	24		DP=6;VDB=0.0057;AF1=0.5;AC1	GT:PL:GQ	0/1:54,0,113:57
48	chromosome	2057564		T	C	8.64		DP=10;VDB=0.0041;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,129:40
49	chromosome	2058145		C	T,G	108		DP=14;VDB=0.0147;AF1=1;AC1=	GT:PL:GQ	1/1:141,39,0,141,28,138:75
50	plasmid1	37177		G	T	54		DP=49;VDB=0.0087;AF1=0.5;AC	GT:PL:GQ	0/1:84,0,255:87
51	plasmid1	66609		A	G	225		DP=55;VDB=0.0060;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
52	plasmid1	66637		C	T	225		DP=71;VDB=0.0141;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,249:99
53	plasmid1	66668		C	T	225		DP=80;VDB=0.0153;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
54	plasmid2	2569		CAAAA	CAAAA	177		INDEL;DP=110;VDB=0.0166;AF1	GT:PL:GQ	1/1:218,255,0:99
55	plasmid2	37805		TAAAAA	TAAAA	104		INDEL;DP=31;VDB=0.0058;AF1=	GT:PL:GQ	1/1:145,93,0:99

W18-4: 全体

①は、genotype情報。0はREFのアリル、1はALTのアリルを意味する。0/1はREFとALTの両方が存在するもの(ヘテロ接合体; heterozygote)、1/1はALTのみからなるホモ接合体(homozygote)

	A	B	C	D	E	F	G	H	I	J
27	#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	0/1:47,0,138:50
28	chromosome	141598	.	A	G	17.1	.	DP=7;VDB=0.0077;AF1=0.5;AC1	GT:PL:GQ	0/1:47,0,138:50
29	chromosome	242819	.	T	C	20	.	DP=9;VDB=0.0111;AF1=0.5;AC1	GT:PL:GQ	0/1:50,0,139:53
30	chromosome	359141	.	T	C	4.13	.	DP=14;VDB=0.0037;AF1=0.4998	GT:PL:GQ	0/1:32,0,187:31
31	chromosome	463882	.	T	G	6.2	.	DP=13;VDB=0.0074;AF1=0.4999	GT:PL:GQ	0/1:35,0,216:36
32	chromosome	660031	.	C	T	8.64	.	DP=11;VDB=0.0063;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,197:40
33	chromosome	792963	.	T	C	4.13	.	DP=14;VDB=0.0153;AF1=0.4998	GT:PL:GQ	0/1:32,0,200:31
34	chromosome	887422	.	G	A	10.4	.	DP=11;VDB=0.0070;AF1=0.5;AC	GT:PL:GQ	0/1:40,0,188:42
35	chromosome	913470	.	T	C	14.2	.	DP=9;VDB=0.0056;AF1=0.5;AC1	GT:PL:GQ	0/1:44,0,144:47
36	chromosome	965171	.	T	C	17.1	.	DP=20;VDB=0.0071;AF1=0.5;AC	GT:PL:GQ	0/1:47,0,249:50
37	chromosome	1134950	.	A	G	7.8	.	DP=23;VDB=0.0076;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,255:39
38	chromosome	1160048	.	A	G	29	.	DP=17;VDB=0.0054;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
39	chromosome	1160095	.	A	G	19.1	.	DP=10;VDB=0.0033;AF1=0.5;AC	GT:PL:GQ	0/1:49,0,191:52
40	chromosome	1240248	.	GCCCC	GCCCCC	214	.	INDEL;DP=48;VDB=0.0137;AF1=	GT:PL:GQ	1/1:255,128,0:99
41	chromosome	1455552	.	CAAAA	CAAAAA	214	.	INDEL;DP=58;VDB=0.0164;AF1=	GT:PL:GQ	1/1:255,130,0:99
42	chromosome	1475721	.	T	C	52	.	DP=19;VDB=0.0021;AF1=0.5;AC	GT:PL:GQ	0/1:82,0,246:85
43	chromosome	1752680	.	G	A	29	.	DP=24;VDB=0.0129;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
44	chromosome	1941380	.	T	C	7.8	.	DP=13;VDB=0.0046;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,214:39
45	chromosome	2052955	.	A	G	12.3	.	DP=8;VDB=0.0084;AF1=0.5;AC1	GT:PL:GQ	0/1:42,0,103:44
46	chromosome	2054084	.	TC	T	13.7	.	INDEL;DP=8;VDB=0.0009;AF1=0	GT:PL:GQ	0/1:51,0,174:54
47	chromosome	2056279	.	A	G	24	.	DP=6;VDB=0.0057;AF1=0.5;AC1	GT:PL:GQ	0/1:54,0,113:57
48	chromosome	2057564	.	T	C	8.64	.	DP=10;VDB=0.0041;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,129:40
49	chromosome	2058145	.	C	T,G	108	.	DP=14;VDB=0.0147;AF1=1;AC1=	GT:PL:GQ	1/1:141,39,0,141,28,138:75
50	plasmid1	37177	.	G	T	54	.	DP=49;VDB=0.0087;AF1=0.5;AC	GT:PL:GQ	0/1:84,0,255:87
51	plasmid1	66609	.	A	G	225	.	DP=55;VDB=0.0060;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
52	plasmid1	66637	.	C	T	225	.	DP=71;VDB=0.0141;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,249:99
53	plasmid1	66668	.	C	T	225	.	DP=80;VDB=0.0153;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
54	plasmid2	2569	.	CAAAA	CAAAA	177	.	INDEL;DP=110;VDB=0.0166;AF1	GT:PL:GQ	1/1:218,255,0:99
55	plasmid2	37805	.	TAAAAA	TAAAA	104	.	INDEL;DP=31;VDB=0.0058;AF1=	GT:PL:GQ	1/1:145,93,0:99



我々の目的である概要配列の修正箇所
の有力候補は赤矢印部分だということ!

W18-4: 全体

	A	B	C	D	E	F	G	H	I	J
27	#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	/home/w3pipeline/refdata/
28	chromosome	141598	.	A	G	17.1	.	DP=7;VDB=0.0077;AF1=0.5;AC1	GT:PL:GQ	0/1:47,0,138:50
29	chromosome	242819	.	T	C	20	.	DP=9;VDB=0.0111;AF1=0.5;AC1	GT:PL:GQ	0/1:50,0,139:53
30	chromosome	359141	.	T	C	4.13	.	DP=14;VDB=0.0037;AF1=0.4998	GT:PL:GQ	0/1:32,0,187:31
31	chromosome	463882	.	T	G	6.2	.	DP=13;VDB=0.0074;AF1=0.4999	GT:PL:GQ	0/1:35,0,216:36
32	chromosome	660031	.	C	T	8.64	.	DP=11;VDB=0.0063;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,197:40
33	chromosome	792963	.	T	C	4.13	.	DP=14;VDB=0.0153;AF1=0.4998	GT:PL:GQ	0/1:32,0,200:31
34	chromosome	887422	.	G	A	10.4	.	DP=11;VDB=0.0070;AF1=0.5;AC	GT:PL:GQ	0/1:40,0,188:42
35	chromosome	913470	.	T	C	14.2	.	DP=9;VDB=0.0056;AF1=0.5;AC1	GT:PL:GQ	0/1:44,0,144:47
36	chromosome	965171	.	T	C	17.1	.	DP=20;VDB=0.0071;AF1=0.5;AC	GT:PL:GQ	0/1:47,0,249:50
37	chromosome	1134950	.	A	G	7.8	.	DP=23;VDB=0.0076;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,255:39
38	chromosome	1160048	.	A	G	29	.	DP=17;VDB=0.0054;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
39	chromosome	1160095	.	A	G	19.1	.	DP=10;VDB=0.0033;AF1=0.5;AC	GT:PL:GQ	0/1:49,0,191:52
40	chromosome	1240248	.	GCCCC	GCCCCC	214	.	INDEL;DP=48;VDB=0.0137;AF1=	GT:PL:GQ	1/1:←55,128,0:99
41	chromosome	1455552	.	CAAAA	CAAAAA	214	.	INDEL;DP=58;VDB=0.0164;AF1=	GT:PL:GQ	1/1:←55,130,0:99
42	chromosome	1475721	.	T	C	52	.	DP=19;VDB=0.0021;AF1=0.5;AC	GT:PL:GQ	0/1:82,0,246:85
43	chromosome	1752680	.	G	A	29	.	DP=24;VDB=0.0129;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
44	chromosome	1941380	.	T	C	7.8	.	DP=13;VDB=0.0046;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,214:39
45	chromosome	2052955	.	A	G	12.3	.	DP=8;VDB=0.0084;AF1=0.5;AC1	GT:PL:GQ	0/1:42,0,103:44
46	chromosome	2054084	.	TC	T	13.7	.	INDEL;DP=8;VDB=0.0009;AF1=0	GT:PL:GQ	0/1:51,0,174:54
47	chromosome	2056279	.	A	G	24	.	DP=6;VDB=0.0057;AF1=0.5;AC1	GT:PL:GQ	0/1:54,0,113:57
48	chromosome	2057564	.	T	C	8.64	.	DP=10;VDB=0.0041;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,129:40
49	chromosome	2058145	.	C	T,G	108	.	DP=14;VDB=0.0147;AF1=1;AC1=	GT:PL:GQ	1/1:←41,39,0,141,28,138:75
50	plasmid1	37177	.	G	T	54	.	DP=49;VDB=0.0087;AF1=0.5;AC	GT:PL:GQ	0/1:84,0,255:87
51	plasmid1	66609	.	A	G	225	.	DP=55;VDB=0.0060;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
52	plasmid1	66637	.	C	T	225	.	DP=71;VDB=0.0141;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,249:99
53	plasmid1	66668	.	C	T	225	.	DP=80;VDB=0.0153;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
54	plasmid2	2569	.	CAAAA	CAAAA	177	.	INDEL;DP=110;VDB=0.0166;AF1	GT:PL:GQ	1/1:←48,255,0:99
55	plasmid2	37805	.	TAAAAA	TAAAA	104	.	INDEL;DP=31;VDB=0.0058;AF1=	GT:PL:GQ	1/1:←45,93,0:99

W19-1 : Tablet

https://ics.hutton.ac.uk/tablet/ Tablet : Information & ... x

The James Hutton Institute

Home Research Software Blogs Staff Publications Events Contact Us Wiki

Information & Computational Sciences

Tablet

Tablet is a lightweight, high-performance graphical viewer for next generation sequence assemblies and alignments.

enter search terms

search

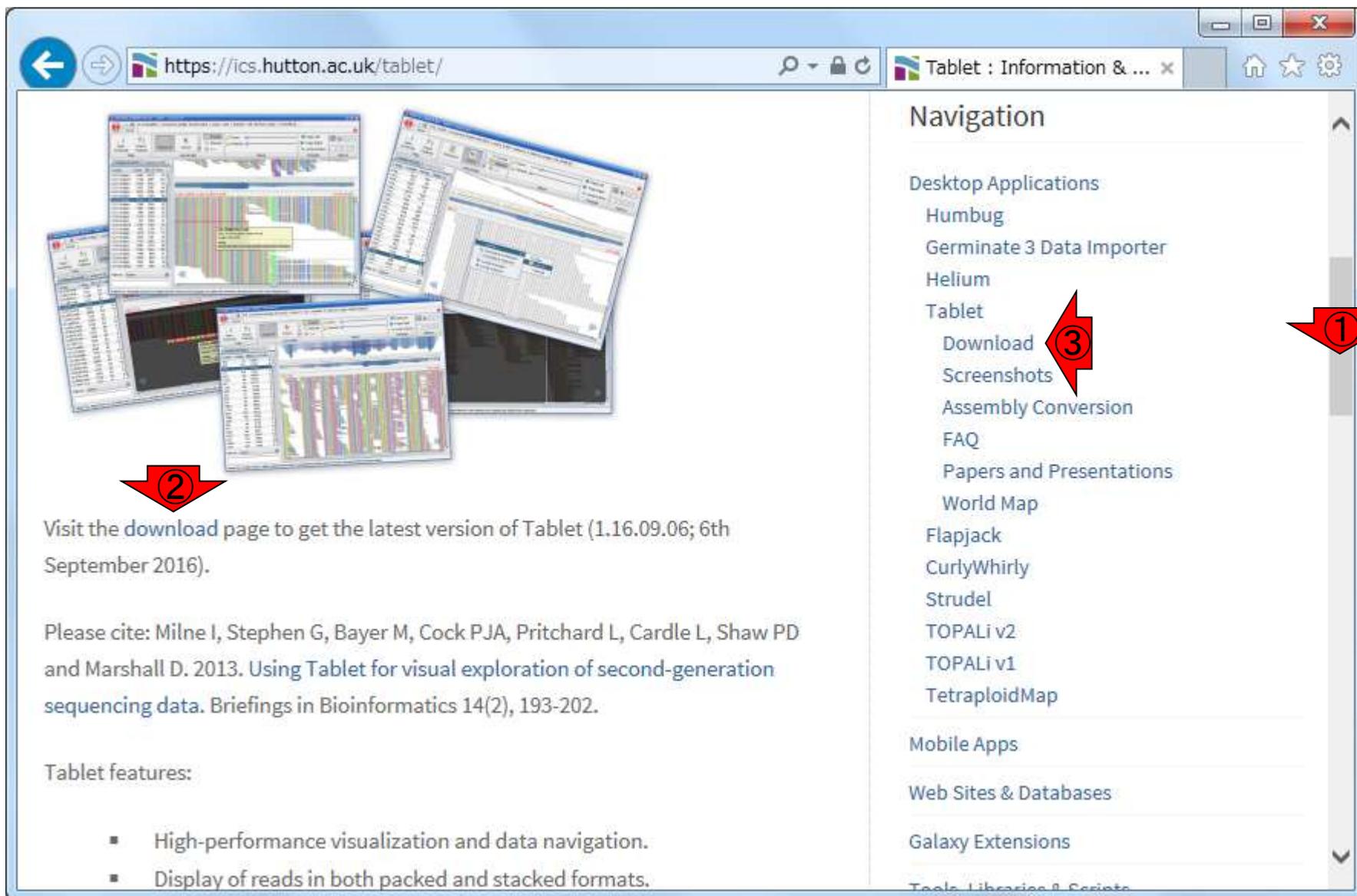
Navigation

Desktop Applications

- Humbug
- Germinate 3 Data Importer
- Helium
- Tablet

W19-1: Tablet

①このあたり。②または③を
押してダウンロードページへ



The screenshot shows a web browser window with the URL <https://ics.hutton.ac.uk/tablet/>. The main content area features a collage of Tablet software interface windows. A red arrow labeled '2' points to the text below the collage. A red arrow labeled '1' points to the 'Download' link in the navigation menu. A red arrow labeled '3' points to the 'Screenshots' link in the navigation menu.

Navigation

- Desktop Applications
 - Humbug
 - Germinate 3 Data Importer
 - Helium
 - Tablet
 - Download
 - Screenshots
 - Assembly Conversion
 - FAQ
 - Papers and Presentations
 - World Map
 - Flapjack
 - CurlyWhirly
 - Strudel
 - TOPALI v2
 - TOPALI v1
 - TetraploidMap
- Mobile Apps
- Web Sites & Databases
- Galaxy Extensions
- Tools, Libraries & Scripts

Visit the download page to get the latest version of Tablet (1.16.09.06; 6th September 2016).

Please cite: Milne I, Stephen G, Bayer M, Cock PJA, Pritchard L, Cardle L, Shaw PD and Marshall D. 2013. Using Tablet for visual exploration of second-generation sequencing data. *Briefings in Bioinformatics* 14(2), 193-202.

Tablet features:

- High-performance visualization and data navigation.
- Display of reads in both packed and stacked formats.

W19-1: Tablet

ダウンロードページ移動後に、①ちょっとページ下部に移動したところ。②赤枠で示すようにWindows、Linux、MacintoshのどれでもOK。ここでは③Windows (64 bit)版のインストール例を示す

The screenshot shows a web browser window with the URL <https://ics.hutton.ac.uk/tablet/download-tablet/>. The page title is "Download Tablet". The main content area contains the following text:

The most recent release of Tablet is **1.16.09.06** (6th September 2016). View the [release notes](#) to see what's new.

Please use the links below to download the Tablet installer most suitable for your operating system. Tablet is currently available for:

-  [Windows \(32 bit\) or Windows \(64 bit\)](#)
-  [Linux \(32 bit\) or Linux \(64 bit\)](#)
-  [OS X \(note requirements below\)](#)

Our installers are built using the multi-platform installer builder *install4j*.
[Checksums can be viewed here.](#)

A **Java Web Start** version of Tablet also exists. Instructions for using it can be found in the [Tablet FAQ](#).

The right sidebar contains a search box with the placeholder text "enter search terms" and a "search" button. Below the search box is a "Navigation" section with a list of links: Desktop Applications, Humbug, Germinate 3 Data Importer, Helium, Tablet, Download, Licence, Release Notes, Screenshots, Assembly Conversion, FAQ, Papers and Presentations, and World Map.

Annotations: A red arrow labeled "1" points to the search box. A red arrow labeled "2" points to the list of operating systems. A red arrow labeled "3" points to the Windows (32 bit) or Windows (64 bit) link.

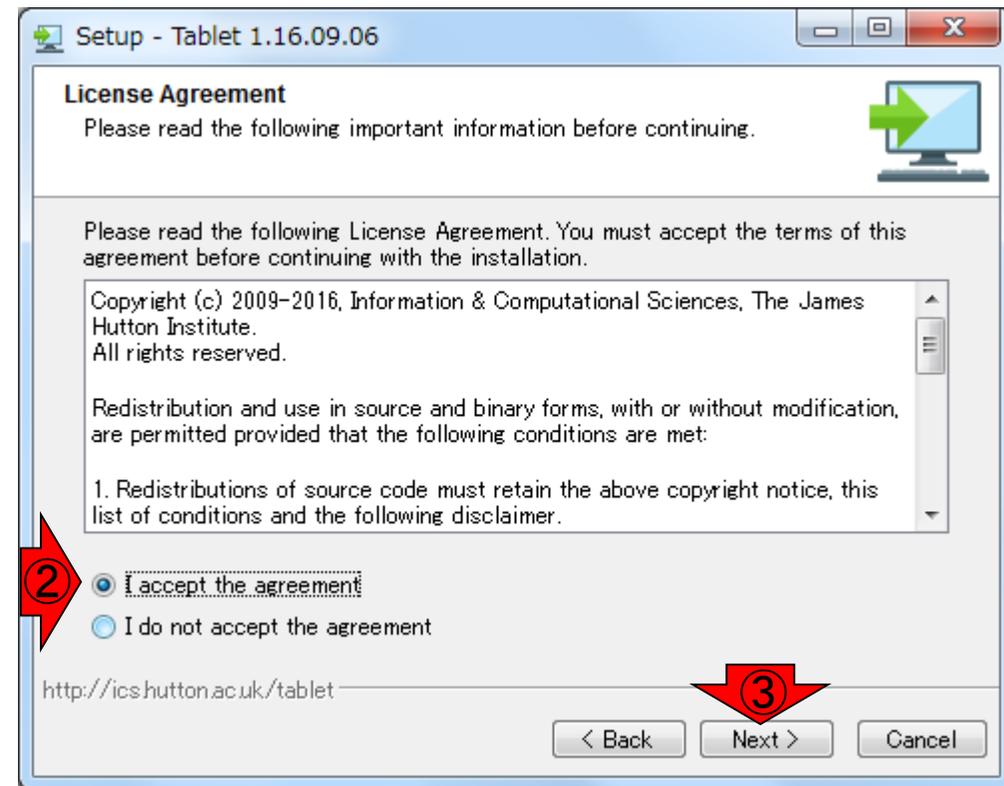
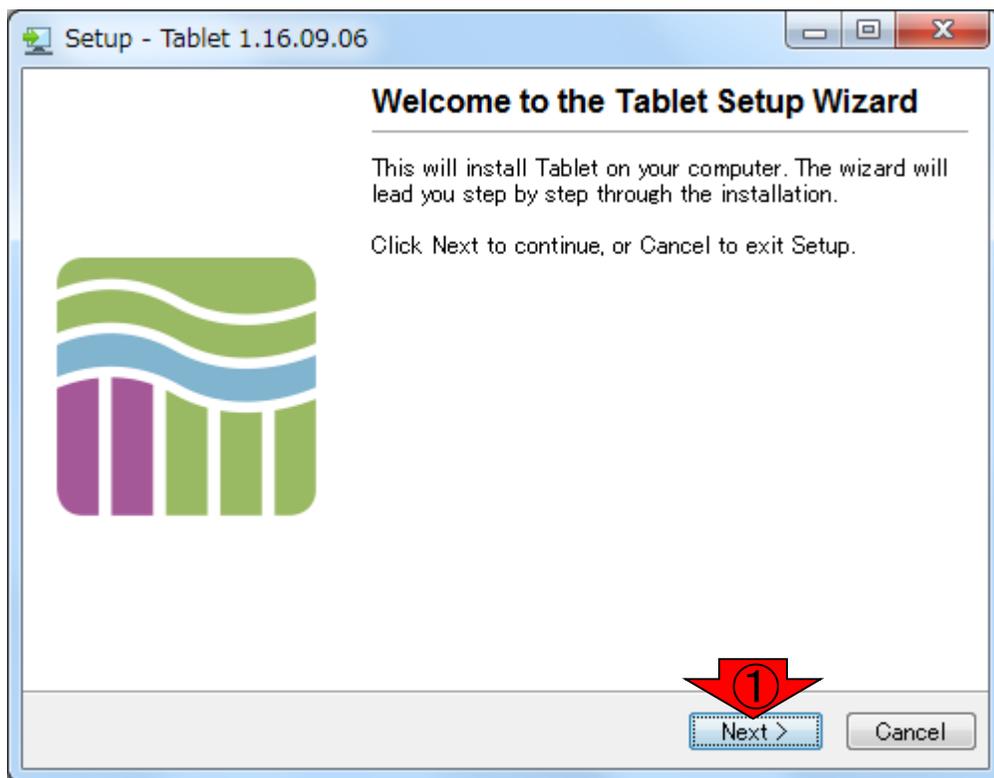
W19-1: Tablet

The screenshot shows a web browser window with the URL <https://ics.hutton.ac.uk/tablet/download-tablet/>. The page title is "Download Tablet" and it is part of the "Information & Computational Sciences" website. The main content area says "The most recent release of Tablet is 1.16.09.06 (6th September 2016). View the release notes to see what's new." Below this, it says "Please use the links below to download the Tablet installer most suitable for your operating system. Tablet is currently available for:" followed by three options: "Windows (32 bit) or Windows (64 bit)", "Linux (32 bit) or Linux (64 bit)", and "OS X (note requirements below)".

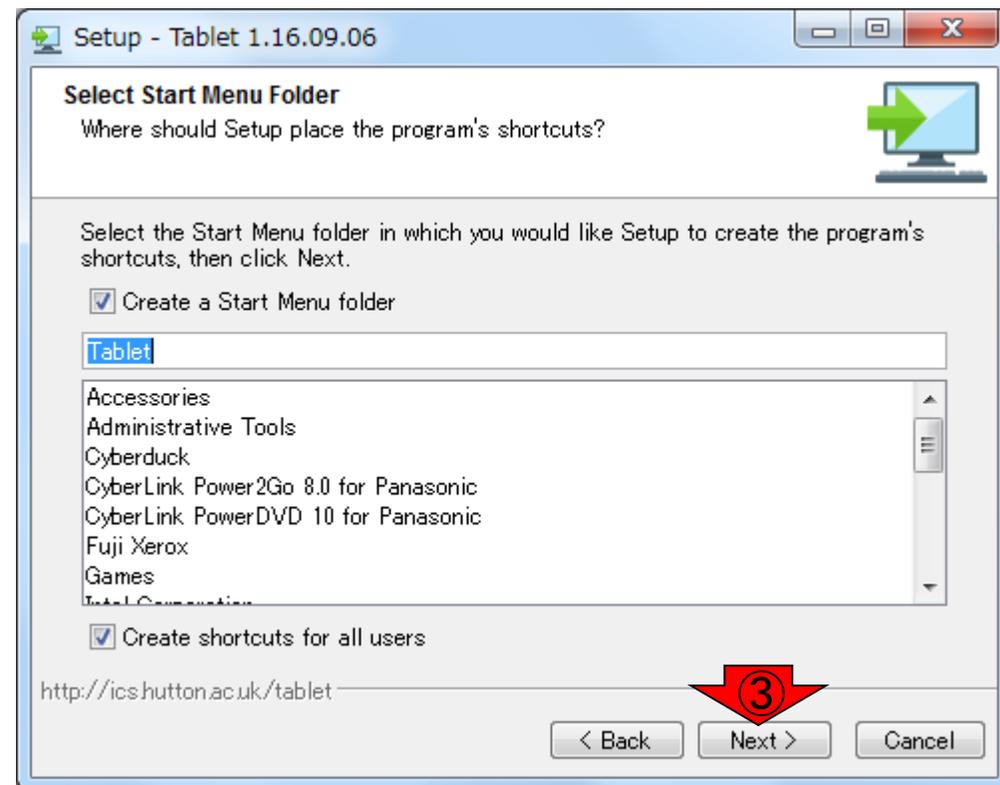
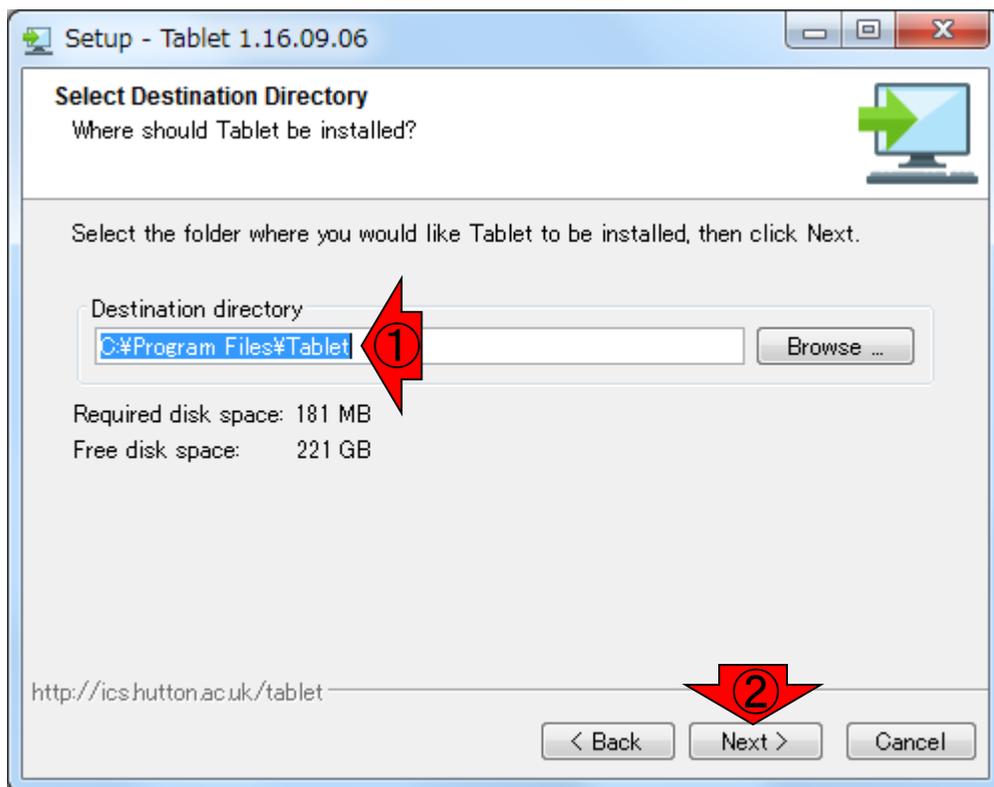
Overlaid on the bottom of the browser window is a yellow dialog box with the text: "bioinf.hutton.ac.uk から tablet_windows_x64_1_16_09_06.exe (45.8 MB) を実行または保存しますか?". At the bottom of this dialog are three buttons: "実行(R)", "保存(S)", and "キャンセル(C)". A red arrow labeled "1" points to the "実行(R)" button.

Overlaid on the right side of the browser window is a red-bordered dialog box titled "install4j Wizard". It contains a green progress bar and a "Cancel" button. The text inside says: "Tablet is preparing the install4j Wizard which will guide you through the rest of the setup process." A red arrow labeled "2" points to the search bar area of the background page.

W19-1 : Tablet

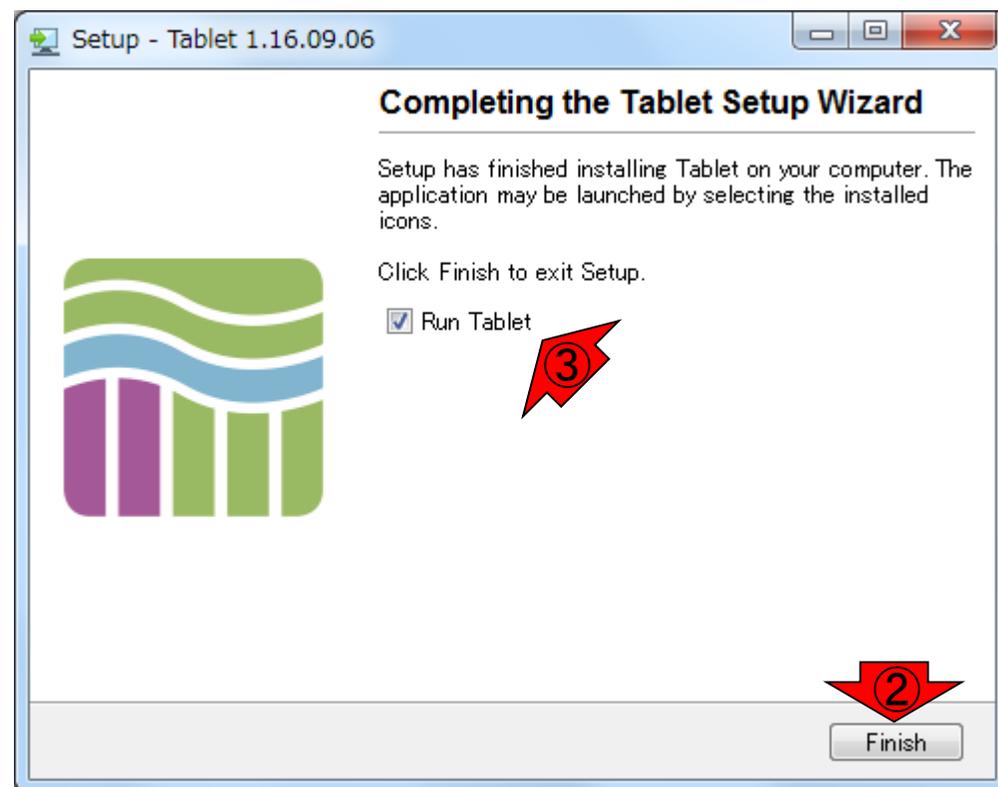
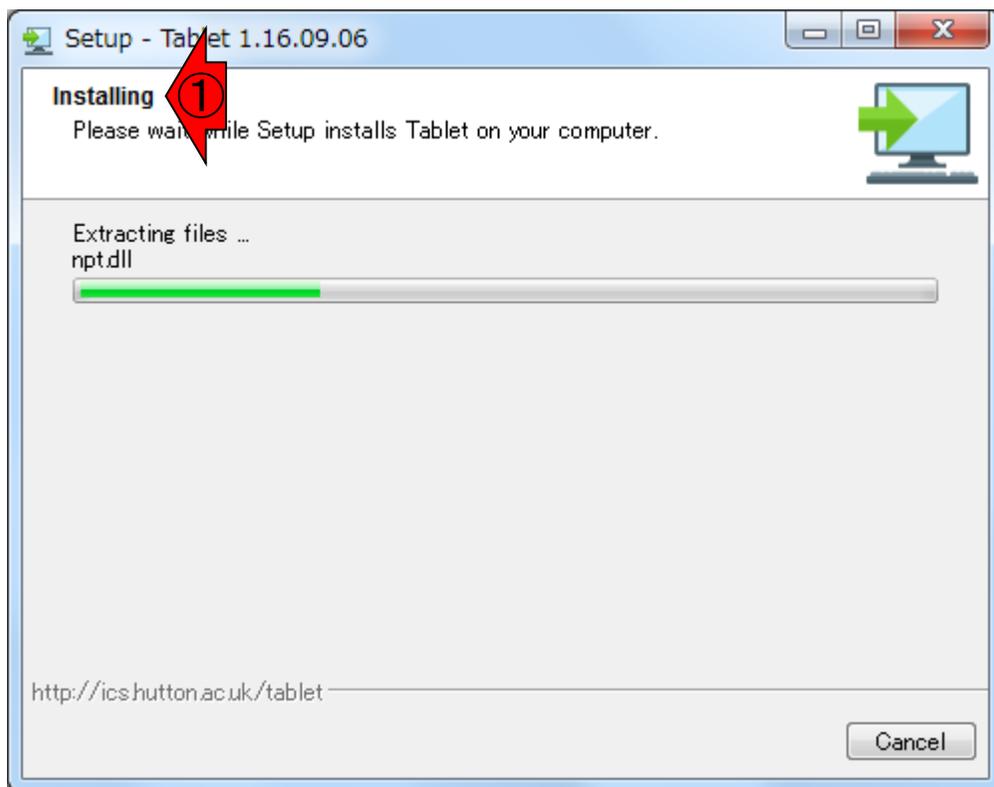


W19-1: Tablet

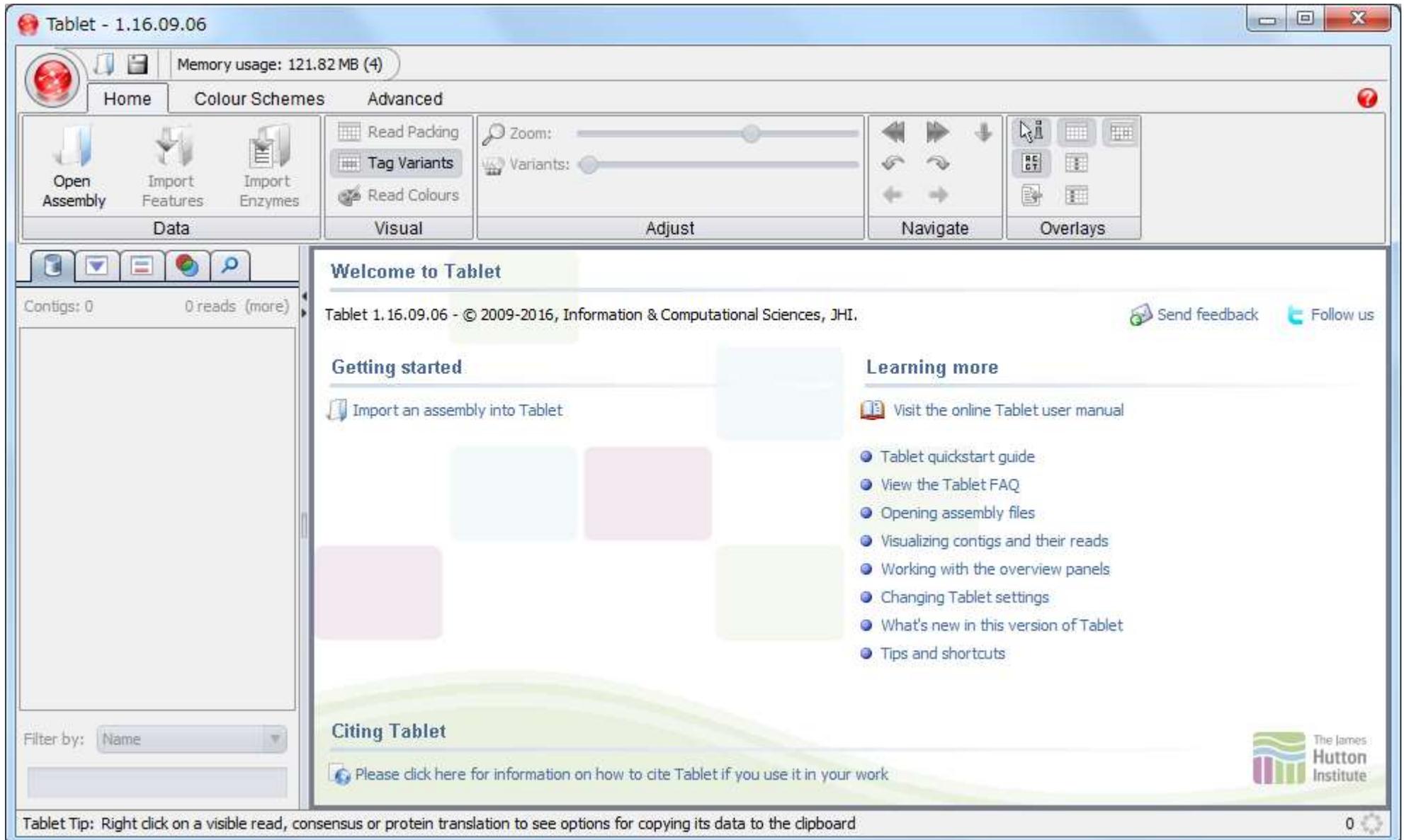


W19-1: Tablet

①インストール中。無事インストール完了したので、②Finish。③Run Tabletにチェックが入っているので自動的に起動する



W19-2: 起動直後



W19-3: 読み込み

①Open Assemblyボタンを押して、②Browse。ここではソート済みBAMファイル(out2.bam)を読み込む

Tablet - 1.16.09.06

Memory usage: 121.82 MB (4)

Home Colour Schemes Advanced

Open Assembly Import Features Import Enzymes

Data

Read Packing Zoom: [Slider]

Open Assembly

Select assembly files:

Primary assembly file or URL: [Text Field] Browse... ②

Reference/consensus file or URL: [Text Field] Browse...

Current status: Assembly - Unknown

Notes:

- Tablet currently supports ACE, AFG, MAQ (text), SOAP, SAM, and (indexed) BAM assemblies.
- Reference files (if needed for MAQ, SOAP, SAM and BAM) can be in FASTA or FASTQ format.
- Unsure how to get started? Click here to open an example assembly.

Open Cancel Help

Tablet Tip: Right click on a visible read, consensus or protein translation to see options for copying its data to the clipboard

W19-3: 読み込み

文字化けしているように見づらいが、どうかして①(ここでは)共有フォルダ上にある、②ソート済みBAMファイル(out2.bam)を読み込む。③ここ

Tablet - 1.16.09.06

Memory usage: 121.82 MB (4)

Home Colour Schemes Advanced

Open Assembly Import Features Import Enzymes

Read Packing Zoom: [Slider]

Open Assembly

Select assembly files:

C:\Users\kadota\Desktop\share

名前	サイズ	更新日時	種類
out-unique.var.flt.vcf.zip	3 KB	2016/09/13 20:04	ZIP
out2.bam.bai.zip	4 KB	2016/09/13 20:03	ZIP
out2.bam.zip	69,651 KB	2016/09/13 20:02	ZIP
uniqout.sam.zip	94,135 KB	2016/09/13 20:02	ZIP
out.sam.zip	113,662 KB	2016/09/13 19:42	ZIP
out-unique.var.flt.vcf	7 KB	2016/09/12 17:28	vcf
out2.bam.bai	8 KB	2016/09/12 17:26	BAM
out2.bam	69,640 KB	2016/09/12 17:26	BAM
uniqout.sam	313,436 KB	2016/09/12 17:23	SAM
out.sam	363,166 KB	2016/09/12 17:19	SAM
LH_draft.fa	2,345 KB	2016/09/10 16:52	FASTA
sequence1_blast.xml	9,547 KB	2016/09/05 13:12	XML
LH_hgap.fa	2,377 KB	2016/08/30 21:22	FASTA

share

LH_draft.fa out2.bam.bai.zip uniqout.sam.zip

LH_hgap.fa out2.bam.zip

out.sam out-unique.var.flt.vcf

out.sam.zip out-unique.var.flt.vcf.zip

out2.bam

out2.bam.bai sequence1_blast.xml

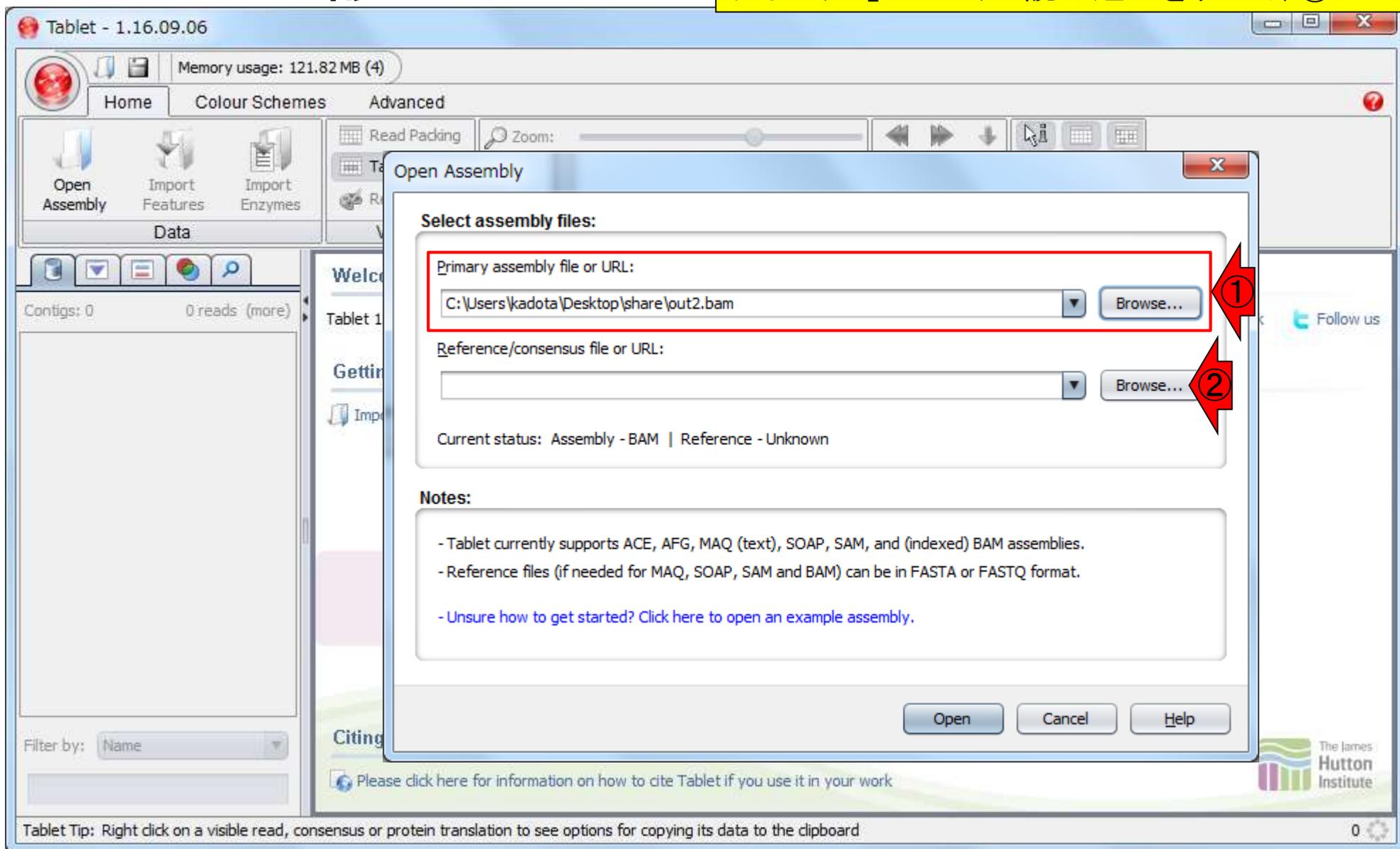
uniqout.sam

out2.bam

③

W19-3: 読み込み

①ソート済みBAMファイル(out2.bam)の読み込みは成功。次はマッピングに用いたリファレンス配列ファイル(LH_draft.fa)の読み込みをすべく、②Browse



W19-3: 読み込み

文字化けしているようで見づらいが、どうにかして①(ここでは)共有フォルダ上にある、②リファレンス配列ファイル(LH_draft.fa)を読み込む。③ここ

The screenshot shows the Tablet software interface with the following elements:

- Open Assembly Dialog:** "Select assembly files:"
- File Explorer:** Shows the path `C:\Users\kadota\Desktop\share`. The file list includes:

名前	サイズ	更新日時	種類
out-unique.var.flit.vcf.zip	3 KB	2016/09/13 20:04	ZIP
out2.bam.bai.zip	4 KB	2016/09/13 20:03	ZIP
out2.bam.zip	69,651 KB	2016/09/13 20:02	ZIP
uniqout.sam.zip	94,135 KB	2016/09/13 20:02	ZIP
out.sam.zip	113,662 KB	2016/09/13 19:42	ZIP
out-unique.var.flit.vcf	7 KB	2016/09/12 17:28	vcf
out2.bam.bai	8 KB	2016/09/12 17:26	BAI
out2.bam	69,640 KB	2016/09/12 17:26	BAI
uniqout.sam	313,436 KB	2016/09/12 17:23	SAM
out.sam	363,166 KB	2016/09/12 17:19	SAM
LH_draft.fa	2,345 KB	2016/09/10 16:52	FA
sequence1_blast.xml	9,547 KB	2016/09/05 13:12	XML
LH_hgap.fa	2,377 KB	2016/08/30 21:22	FA
- Browse For File Dialog:** Shows the "share" folder selected. The file list includes:

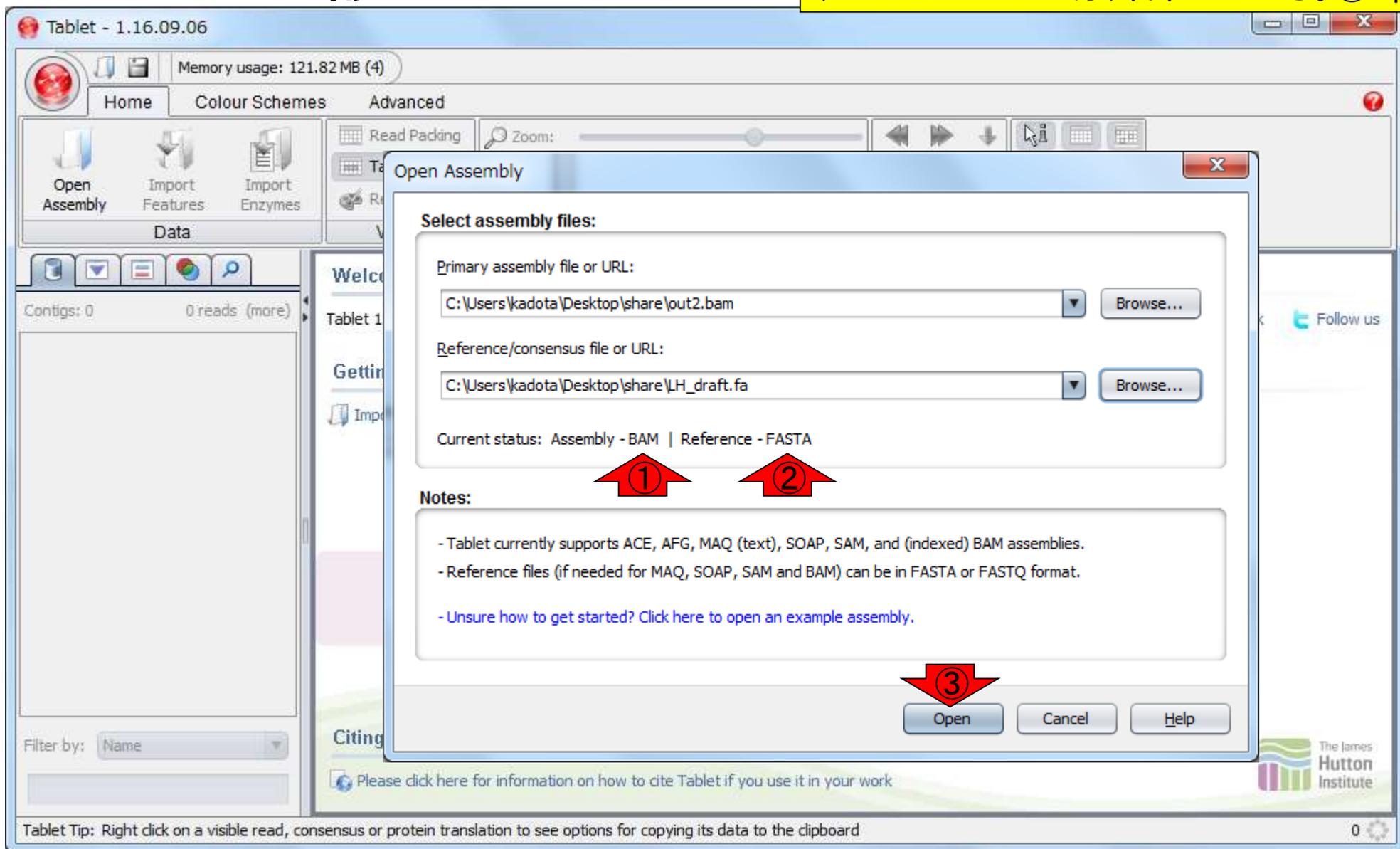
LH_draft.fa	out2.bam.bai.zip	uniqout.sam.zip
LH_hgap.fa	out2.bam.zip	
out.sam	out-unique.var.flit.vcf	
out.sam.zip	out-unique.var.flit.vcf.zip	
out2.bam	sequence1_blast.xml	
out2.bam.bai	uniqout.sam	

Red arrows indicate the following steps:

- ① Select the share folder in the file explorer.
- ② Select the LH_draft.fa file in the file explorer.
- ③ Enter the file name LH_draft.fa in the "Browse For File" dialog.

W19-3: 読み込み

2つのファイルともに無事読み込み成功。よく見ると、①と②のところもunknownだったのが、BAMとFASTAに切り替わっている。③Open



W19-3: 読み込み

読み込み直後の状態。赤枠あたりが変わった。ここで見えているのは概要配列中の3つ(chromosome, plasmid1, plasmid2)

out2.bam - Tablet - 1.16.09.06

Memory usage: 124.09 MB (3)

Home Colour Schemes Advanced

Open Assembly Import Features Import Enzymes

Read Packing Tag Variants Read Colours

Zoom: Variants:

Adjust Navigate Overlays

Contigs: 3 509.1 k reads (more)

Co...	Len...	Re...	Fe...	Mis...
chr...	2,2...	47...	0	?
pla...	81,...	14,...	0	?
pla...	40,...	15,...	0	?

Filter by: Name

- select a contig to begin visualization -

Tablet Tip: Load data more quickly by simply dragging and dropping the assembly (and reference file if needed) directly into Tablet

W19-4: 概観

①のあたりにマウス移動させると、plasmid2の配列長(40,973 bp)とマップされたリード数(15,434)が表示される。クリックすると…

The screenshot shows the Tablet software interface. The top bar displays 'out2.bam - Tablet - 1.16.09.06' and 'Memory usage: 65.09 MB (3)'. The main interface is divided into several sections: 'Data' (Open Assembly, Import Features, Import Enzymes), 'Visual' (Read Packing, Tag Variants, Read Colours), 'Adjust' (Zoom and Variants sliders), 'Navigate' (directional arrows), and 'Overlays' (grid and zoom icons). On the left, a 'Contigs' table lists three contigs. A red arrow labeled '1' points to the second row, 'plasmid2'. A tooltip for 'plasmid2' is displayed, showing its length (40,973 bp) and the number of mapped reads (15,434). The main visualization area is currently empty, with the text '- select a contig to begin visualization -' centered on the screen. A 'Tablet Tip' at the bottom explains that right-clicking on a visible read, consensus, or protein translation will show options for copying its data to the clipboard.

Co...	Len...	Re...	Fe...	Mis...
chr...	2,2...	47...	0	?
pla...	81,...	14,...	0	?
pla...	40,...	15,...	0	?

Contigs: 3 509.1 k reads (more)

plasmid2
Length: 40,973
Reads: 15,434
Features: 0
Mismatch: Unknown

Filter by: Name

Tablet Tip: Right click on a visible read, consensus or protein translation to see options for copying its data to the clipboard

W19-4: 概観

out2.bam - Tablet - 1.16.09.06

plasmid2 | consensus length: 40,973 (0) | reads: 9,503 | features: 1 | Memory usage: 102.23 MB (10)

Home Colour Schemes Advanced

Open Assembly Import Features Import Enzymes

Read Packing Tag Variants Read Colours

Zoom: Variants:

Adjust Navigate Overlays

Contigs: 3 509.1 k reads (more)

Co...	Len...	Re...	Fe...	Mis...
chr...	2,2...	47...	0	?
pla...	81,...	14,...	0	?
pla...	40,...	15,...	1	0.5

1 to 25,000 (25 Kb) 1 to 47 (47 bp)

E Q G Y Y Y I I P N G N G T

G A G C A G G G G T A T T A C T A C T A C A T C A T T C C A A A T G G G A A C G G C A C T T

CIGAR:0

1 U1 46 U46

Tablet Tip: Mousing over CIGAR "I" features on the features track highlights the reads - and locations - the insertion relates to

W19-4: 概観

①の部分が概要配列(LH_draft.fa)。②の部分がマップされたリードの情報です。ソート済みBAMファイル(out2.bam)を読み込ませているので(本当はソートしてなくてもかもしれませんが)、概要配列の位置順にリードが張り付いている様子がよくわかります

out2.bam - Tablet - 1.16.09.06

plasmid2 | consensus length: 40,973 (0) | reads: 9,503 | features: 1 | Memory: 100 MB

Home Colour Schemes Advanced

Open Assembly Import Features Import Enzymes

Data Visual Adjust Navigate Overlays

Contigs: 3 509.1 k reads (more)

Co...	Len...	Re...	Fe...	Mis...
chr...	2,2...	47...	0	?
pla...	81,...	14,...	0	?
pla...	40,...	15,...	1	0.5

1 to 25,000 (25 Kb) 1 to 47 (47 bp)

E Q G Y Y Y I I P N G N G T

G A G C A G G G G T A T T A C T A C T A C A T C A T T C C A A A T G G G A A C G G C A C T T

CIGAR:0

1 U1 46 U46

A G G G G T A T T A C T A C T A C A T C A T T C C A A A T G G G A A C G G G C A C T T T

T T A T T A C T A C T A C A T C A T T C C A A A T G G G A A C G G G C A C T T T

T A T T A C T A C T A C A T C A T T C C A A A T G G G A A C G G G C A C T T T

T A C T A C T A C A T C A T T C C A A A T G G G A A C G G G C A C T T T

C T A C T A C A T C A T T C C A A A T G G G A A C G G G C A C T T T

C A A A T G G G A A C G G G C A C T T T

G G A A C G G G C A C T T T

A C

Tablet Tip: Mousing over CIGAR "I" features on the features track highlights the reads - and locations - the insertion relates to

W19-5: 目的おさらい

他の配列 (chromosomeとplasmid1) についても同様。我々が今やりたいのは、ミスマッチやindel(挿入や欠失)を許容してマップした結果から、リファレンスである概要配列と異なる領域を重点的に眺め、概要側を修正すること

out2.bam - Tablet - 1.16.09.06

plasmid2 | consensus length: 40,973 (0) | reads: 9,503 | features: 1 | Memory usage: 10...

Home Colour Schemes Advanced

Open Assembly Import Features Import Enzymes

Read Packing Tag Variants Read Colours

Zoom: Variants:

Adjust Navigate Overlays

Contigs: 3 509.1 k reads (more)

Co...	Len...	Re...	Fe...	Mis...
chr...	2,2...	47...	0	?
pla...	81,...	14,...	0	?
pla...	40,...	15,...	1	0.5

to 25,000 (25 Kb) 1 to 47 (47 bp)

E Q G Y Y Y Y I I P N G N G T

A G C A G G G G T A T T A C T A C A T C A T T C C A A A T G G G A A C G G C A C T T

1 U1 46 U46

A G G G G T T A T T A C T A C T A C A T C A T T C C A A A T G G G A A C G G C A C T T

T T A T T A C T A C T A C A T C A T T C C A A A T G G G A A C G G C A C T T

T A T T A C T A C T A C A T C A T T C C A A A T G G G A A C G G C A C T T

T A C T A C T A C A T C A T T C C A A A T G G G A A C G G C A C T T

C T A C T A C A T C A T T C C A A A T G G G A A C G G C A C T T

C A A A T G G G A A C G G C A C T T

G G A A C G G C A C T T

A C T T

Tablet Tip: Mousing over CIGAR "I" features on the features track highlights the reads - and locations - the insertion relates to

W19-5: 目的おさらい

つまり、概要配列と同じ塩基の領域なんて興味ないんです。興味ある領域は、この画面上であえて言えば①になります。概要側はGで、マップされたリードは2本(depth = 2)。1つはGで概要側と同じ、そしてもう1つはTで異なるからです

The screenshot shows the Tablet software interface. The top bar displays 'out2.bam - Tablet - 1.16.09.06'. Below the navigation tabs (Home, Colour Schemes, Advanced), there are sections for Data (Open Assembly, Import Features, Import Enzymes), Visual (Read Packing, Tag Variants, Read Colours), Adjust (Zoom, Variants), Navigate, and Overlays. The main view shows a genomic track with a consensus sequence and a CIGAR track. A red arrow points to a 'T' in the CIGAR track, which is highlighted with a red circle containing the number 1. The CIGAR track shows '1 U1' and '46 U46'. The consensus sequence is 'E Q G Y Y Y I I P N G N G T'. The CIGAR sequence is 'G A G C A G G G G T A T T A C T A C T A C A T C A T T C C A A A T G G G A A C G G C A C T T'. The CIGAR track shows '1 U1' and '46 U46'. The CIGAR track shows '1 U1' and '46 U46'.

Co...	Len...	Re...	Fe...	Mis...
chr...	2,2...	47...	0	?
pla...	81,...	14,...	0	?
pla...	40,...	15,...	1	0.5

Tablet Tip: Mousing over CIGAR 'T' features on the features track highlights the reads - and locations - the insertion relates to

W19-5: 目的おさら

しかし、この領域(塩基)の場合はdepth = 2なので、本当はTかもしれないけどGでいいのかもしれないと悩みます。結論としては概要側のGをTに修正しません。理由はdepth = 2で低い(つまり信頼性が低い)からです。例えばこの領域に沢山のリードがマップされていて、そのほとんどがTなら、概要側のGをTに修正します

out2.bam - Tablet - 1.16.09.06

plasmid2 | consensus length: 40,973 (0) | reads: 9,503 | features: 1

Home Colour Schemes Advanced

Open Assembly Import Features Import Enzymes

Data Visual Adjust Navigate Overlays

Contigs: 3 509.1 k reads (more)

Co...	Len...	Re...	Fe...	Mis...
chr...	2,2...	47...	0	?
pla...	81,...	14,...	0	?
pla...	40,...	15,...	1	0.5

1 to 25,000 (25 Kb) 1 to 47 (47 bp)

E Q G Y Y Y I I P N G N G T

G A G C A G G G G T A T T A C T A C A T C A T T C C A A A T G G G A A C G G C A C T T

CIGAR-D

1 U1 46 U46

A G G G G T A T T A C T A C T A C A T C A T T C C A A A T G G G A A C G G C A C T T

Tablet Tip: Mousing over CIGAR "T" features on the features track highlights the reads - and locations - the insertion relates to

W19-6: VCF

次に読み込むVCFファイル(out-unique.var.flt.vcf)は、一定の信頼性のある概要配列と異なる部分の情報を保持している。①Import Features

The screenshot shows the Tablet software interface for analyzing sequencing data. The main window displays a sequence alignment track with a consensus sequence and individual reads. The control panel at the top includes buttons for 'Open Assembly', 'Import Features', and 'Import Enzymes'. The 'Import Features' button is highlighted with a red circle and the number '1'. The interface also shows a 'Contigs' table and a 'Filter by' dropdown menu.

Co...	Len...	Re...	Fe...	Mis...
chr...	2,2...	47...	0	?
pla...	81,...	14,...	0	?
pla...	40,...	15,...	1	0.5

Tablet Tip: Mousing over CIGAR "I" features on the features track highlights the reads - and locations - the insertion relates to

W19-6: VCF

文字化けしているようで見づらいが、どうにかして
①(ここでは)共有フォルダ上にある、②VCFファイル(out-unique.var.ft.vcf)を読み込む。③ここ

out2.bam - Tablet - 1.16.09.06

plasmid2 | consensus length: 40,973 (0) | reads: 9,503 | features: 1 | Memory usage: 102.23 MB (10)

Home Colour Schemes Advanced

Open Assembly Import Features Import Enzymes

Read Packing Tag Variants Read Colours

Zoom: Variants:

Navigate Overlays

C:\Users\kadota\Desktop\share

名前	サイズ	更新日時	種類
out-unique.var.ft.vcf.zip	3 KB	2016/09/13 20:04	ZIP
out2.bam.bai.zip	4 KB	2016/09/13 20:03	ZIP
out2.bam.zip	69,651 KB	2016/09/13 20:02	ZIP
uniqout.sam.zip	94,135 KB	2016/09/13 20:02	ZIP
out.sam.zip	113,662 KB	2016/09/13 19:42	ZIP
out-unique.var.ft.vcf	7 KB	2016/09/12 17:28	vCa
out2.bam.bai	8 KB	2016/09/12 17:26	BAI
out2.bam	69,640 KB	2016/09/12 17:26	BAI
uniqout.sam	313,436 KB	2016/09/12 17:23	SAM
out.sam	363,166 KB	2016/09/12 17:19	SAM
LH_draft.fa	2,345 KB	2016/09/10 16:52	FA
sequence1_blast.xml	9,547 KB	2016/09/05 13:12	XML
LH_hgap.fa	2,377 KB	2016/08/30 21:22	FA

Import Features

share

- LH_draft.fa
- LH_hgap.fa
- out.sam
- out2.bam
- out2.bam.bai
- out2.bam.bai.zip
- out2.bam.zip
- out-unique.var.ft.vcf
- out-unique.var.ft.vcf.zip
- sequence1_blast.xml
- uniqout.sam
- uniqout.sam.zip

out-unique.var.ft.vcf

0

W19-6: VCF

①28個のfeatureがあるようだ。これは変異数に相当(W18-2)。②Close

The screenshot shows the Tablet software interface with a dialog box titled "Tablet" open. The dialog box contains the following text:

Tablet has finished importing features. A summary is given below:

- Number of features read from the file: 28
- Number of features added to Tablet: 28

The number added may be less than the number read if duplicates were detected or the contig name assigned to a feature does not match the name of any contigs currently in memory.

The dialog box has a "Close" button at the bottom right. Red arrows with numbers 1 and 2 point to the summary text and the "Close" button, respectively.

The background interface shows a contig table on the left with the following data:

Co...	Len...	Re...	Fe...	Mis...
chr...	2,2...	47...	22	?
pla...	81,...	14,...	4	?
pla...	40,...	15,...	3	0.5

The main display area shows a genomic overview with a sequence alignment below it.

Tablet Tip: CTRL / CMD drag the mouse on the overview display to subset the overview

W19-6: VCF

①この部分にVCFの情報が追加されていることがわかります

The screenshot shows the Tablet software interface. The top bar displays 'out2.bam - Tablet - 1.16.09.06'. Below the bar, there are tabs for 'Home', 'Colour Schemes', and 'Advanced'. The 'Advanced' tab is active, showing various controls like 'Read Packing', 'Tag Variants', 'Read Colours', 'Visual', 'Adjust', 'Navigate', and 'Overlays'. On the left, there is a 'Contigs' table with columns for 'Co...', 'Len...', 'Re...', 'Fe...', and 'Mis...'. The table shows three contigs, with the third one highlighted. A red arrow with the number '1' points to the 'VCF' track in the main display area. The VCF track shows a list of variants for the selected contig. The main display area shows a genomic overview with a zoomed-in view of a region from 1 to 47 (47 bp). The zoomed-in view shows the reference sequence and the aligned reads. The reference sequence is: E Q G Y Y Y Y I I P N G N G T. The aligned reads are: G A G C A G G G G T A T T A C T A C A T C A T T C C A A A T G G G A A C G G C A C T T. The VCF track below the reference sequence shows the following variants: 1 U1, 46 U46.

Co...	Len...	Re...	Fe...	Mis...
chr...	2,2...	47...	22	?
pla...	81,...	14,...	4	?
pla...	40,...	15,...	3	0.

Contigs: 3 509.1 k reads (more)

1 to 25,000 (25 Kb) 1 to 47 (47 bp)

E Q G Y Y Y Y I I P N G N G T

G A G C A G G G G T A T T A C T A C A T C A T T C C A A A T G G G A A C G G C A C T T

CIGAR:0

VCF

1 U1 46 U46

A G G G G T A T T A C T A C T A C A T C A T T C C A A A T G G G A A C G G G C A C T T

T T A T T A C T A C T A C A T C A T T C C A A A T G G G A A C G G G C A C T T

T A T T A C T A C T A C A T C A T T C C A A A T G G G A A C G G G C A C T T

T A C T A C T A C A T C A T T C C A A A T G G G A A C G G G C A C T T

C T A C A T C A T T C C A A A T G G G A A C G G G C A C T T

C A A A T G G G A A C G G G C A C T T

A A T G G G A A C G G G C A C T T

G G A A C G G G C A C T T

A C

Tablet Tip: CTRL / CMD drag the mouse on the overview display to subset the overview

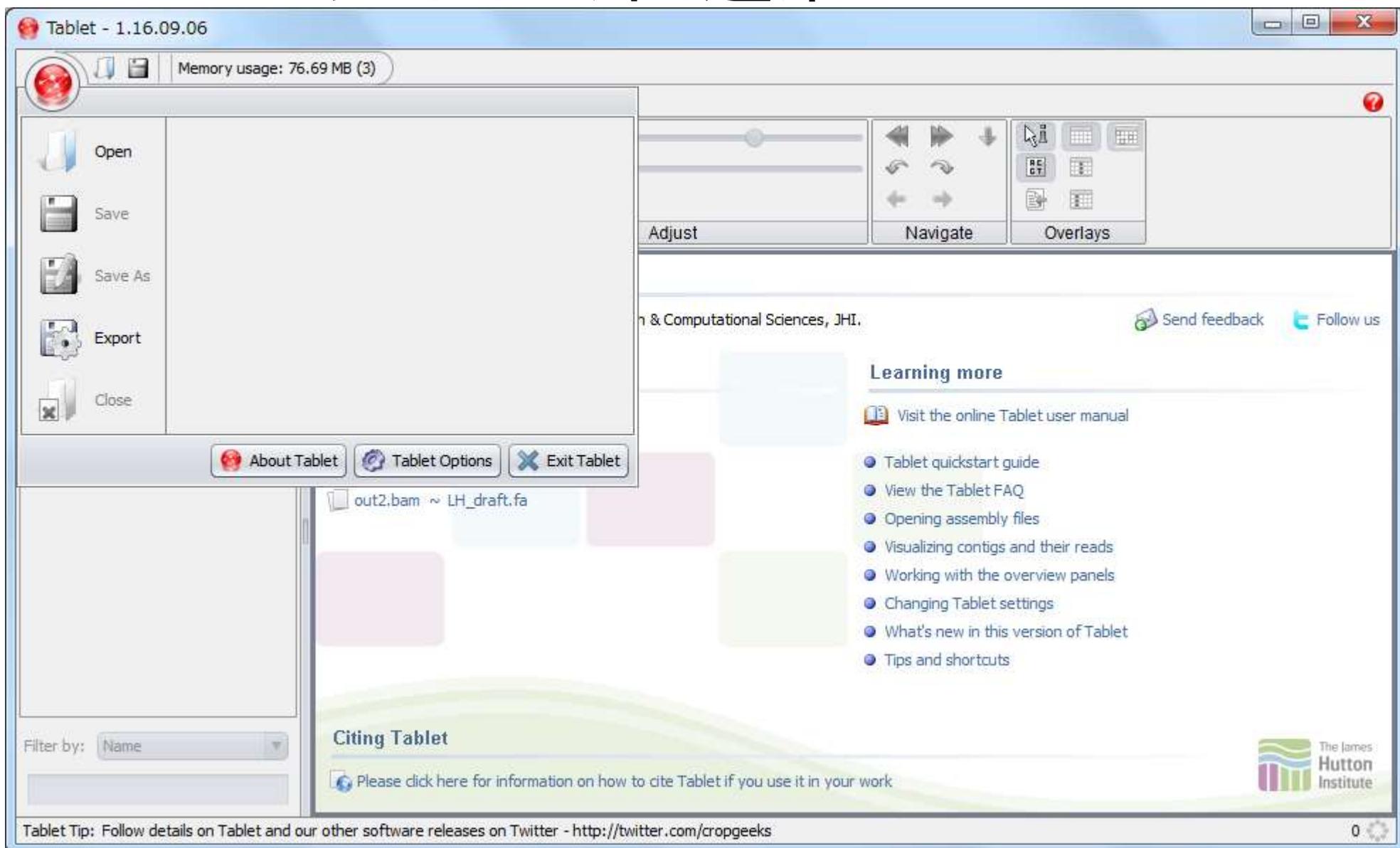
W19-7: 文字化け回避策

①この部分にVCFの情報が追加されていることがわかります

The screenshot shows the Tablet software interface. The window title is "Tablet - 1.16.09.06". The top bar includes a memory usage indicator: "Memory usage: 68.82 MB (3)". Below this are tabs for "Home", "Color Schemes", and "Advanced". The main toolbar is divided into several sections: "Data" (Open Assembly, Import Features, Import Enzymes), "Visual" (Read Packing, Tag Variants, Read Colors), "Adjust" (Zoom and Variants sliders), "Navigate" (directional arrows), and "Overlays" (various grid and overlay icons). On the left, there is a sidebar with "Contigs: 0" and "0 reads (more)". The main content area displays a "Welcome to Tablet" message, copyright information for 2009-2016, and a list of links for "Getting started" and "Learning more". A "Citing Tablet" section is also visible at the bottom. The footer contains a "Tablet Tip" and a Twitter link.

W19-7: 文字化け回避策

①この部分にVCFの情報が追加されていることがわかります



W19-7: 文字化け回避策

①この部分にVCFの情報が追加されていることがわかります

The screenshot shows the Tablet software interface. The main window displays a 'Welcome to Tablet' message and a list of assemblies. The 'Tablet Options' dialog box is open, showing the 'General' tab. The 'General options (restart to apply):' section includes 'Interface display language' set to 'Automatic' and 'Check for newer Tablet versions' set to 'At application startup'. The 'Cache options:' section shows the local directory for temporary files as 'C:\Users\kadota\AppData\Local\Temp\scri-tablet-kadota' and a cache size of 1.14 MB. The dialog box has 'OK', 'Cancel', and 'Help' buttons at the bottom.

W19-7: 文字化け回避策

①この部分にVCFの情報が追加されていることがわかります

The screenshot shows the Tablet software interface. The main window displays a 'Welcome to Tablet' message and a list of assemblies. The 'Tablet Options' dialog box is open, showing the 'General' tab. The 'Interface display language' dropdown menu is open, showing 'Automatic', 'English (UK)', and 'English (US)'. The 'Cache options' section shows the local directory for temporary files and the cache size.

Tablet - 1.16.09.06

Memory usage: 68.82 MB (3)

Home Color Schemes Advanced

Open Assembly Import Features Import Enzymes

Data Visual

Read Packing Tag Variants Read Colors

Zoom: Variants:

Tablet Options

General Visualization Importing Warnings

General options (restart to apply):

Interface display language: Automatic

Check for newer Tablet versions: Automatic

Cache options:

Tablet requires a local directory for storing temporary files:

C:\Users\kadota\AppData\Local\Temp\scri-tablet-kadota

Cache size: 1.14 MB (Clear cache)

OK Cancel Help

Tablet Tip: Follow details on Tablet and our other software releases on Twitter - <http://twitter.com/cropgeeks>

The James Hutton Institute

W20-1 : plasmid2

VCFファイルの27行目以降を、Excelで眺めたものを再掲(W18-2)。①plasmid2上には、2か所変異があるようなので2,569番目と37,805番目をTabletで眺める

	A	B	C	D	E	F	G	H	I	J
27	#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	/home/w3pipeline/refdata/
28	chromosome	141598	.	A	G	17.1	.	DP=7;VDB=0.0077;AF1=0.5;AC1	GT:PL:GQ	0/1:47,0,138:50
29	chromosome	242819	.	T	C	20	.	DP=9;VDB=0.0111;AF1=0.5;AC1	GT:PL:GQ	0/1:50,0,139:53
30	chromosome	359141	.	T	C	4.13	.	DP=14;VDB=0.0037;AF1=0.4998	GT:PL:GQ	0/1:32,0,187:31
31	chromosome	463882	.	T	G	6.2	.	DP=13;VDB=0.0074;AF1=0.4999	GT:PL:GQ	0/1:35,0,216:36
32	chromosome	660031	.	C	T	8.64	.	DP=11;VDB=0.0063;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,197:40
33	chromosome	792963	.	T	C	4.13	.	DP=14;VDB=0.0153;AF1=0.4998	GT:PL:GQ	0/1:32,0,200:31
34	chromosome	887422	.	G	A	10.4	.	DP=11;VDB=0.0070;AF1=0.5;AC	GT:PL:GQ	0/1:40,0,188:42
35	chromosome	913470	.	T	C	14.2	.	DP=9;VDB=0.0056;AF1=0.5;AC1	GT:PL:GQ	0/1:44,0,144:47
36	chromosome	965171	.	T	C	17.1	.	DP=20;VDB=0.0071;AF1=0.5;AC	GT:PL:GQ	0/1:47,0,249:50
37	chromosome	1134950	.	A	G	7.8	.	DP=23;VDB=0.0076;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,255:39
38	chromosome	1160048	.	A	G	29	.	DP=17;VDB=0.0054;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
39	chromosome	1160095	.	A	G	19.1	.	DP=10;VDB=0.0033;AF1=0.5;AC	GT:PL:GQ	0/1:49,0,191:52
40	chromosome	1240248	.	GCCCC	GCCCCC	214	.	INDEL;DP=48;VDB=0.0137;AF1 =	GT:PL:GQ	1/1:255,128,0:99
41	chromosome	1455552	.	CAAAA	CAAAAA	214	.	INDEL;DP=58;VDB=0.0164;AF1 =	GT:PL:GQ	1/1:255,130,0:99
42	chromosome	1475721	.	T	C	52	.	DP=19;VDB=0.0021;AF1=0.5;AC	GT:PL:GQ	0/1:82,0,246:85
43	chromosome	1752680	.	G	A	29	.	DP=24;VDB=0.0129;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
44	chromosome	1941380	.	T	C	7.8	.	DP=13;VDB=0.0046;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,214:39
45	chromosome	2052955	.	A	G	12.3	.	DP=8;VDB=0.0084;AF1=0.5;AC1	GT:PL:GQ	0/1:42,0,103:44
46	chromosome	2054084	.	TC	T	13.7	.	INDEL;DP=8;VDB=0.0009;AF1=0	GT:PL:GQ	0/1:51,0,174:54
47	chromosome	2056279	.	A	G	24	.	DP=6;VDB=0.0057;AF1=0.5;AC1	GT:PL:GQ	0/1:54,0,113:57
48	chromosome	2057564	.	T	C	8.64	.	DP=10;VDB=0.0041;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,129:40
49	chromosome	2058145	.	C	T,G	108	.	DP=14;VDB=0.0147;AF1=1;AC1=	GT:PL:GQ	1/1:141,39,0,141,28,138:75
50	plasmid1	37177	.	G	T	54	.	DP=49;VDB=0.0087;AF1=0.5;AC	GT:PL:GQ	0/1:84,0,255:87
51	plasmid1	66609	.	A	G	225	.	DP=55;VDB=0.0060;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
52	plasmid1	66637	.	C	T	225	.	DP=71;VDB=0.0141;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,249:99
53	plasmid1	66668	.	C	T	225	.	DP=80;VDB=0.0153;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
54	plasmid2	2569	.	CAAAA	CAAAA	177	.	INDEL;DP=110;VDB=0.0166;AF1	GT:PL:GQ	1/1:218,255,0:99
55	plasmid2	37805	.	TAAAAA	TAAAA	104	.	INDEL;DP=31;VDB=0.0058;AF1 =	GT:PL:GQ	1/1:145,93,0:99



W20-1 : plasmid2

①VCFファイル読み込み時に28 featuresと書いてあったが、ここでは(22+4+3) = 29個となっている。plasmid2が2個ではなく3個となっているところが変だと思われる。VCFファイル読み込み前を見直してみると、0,0,1となっていてplasmid2のところに1つのfeatureがあるとなっていた(W19-4)。そのせいかもしれない

out2.bam - Tablet - 1.16.09.06

plasmid2 | consensus length: 40,973 (0) | reads: 9,503 | features: 3

Home Colour Schemes Advanced

Open Assembly Import Features Import Enzymes

Read Packing Tag Variants Read Colours

Zoom: Variants:

Adjust Navigate Overlays

Contigs: 3 509.1 k reads (more)

Co...	Len...	Re...	Fe...	Mis...
chr...	2,2...	47...	22	?
pla...	81,...	14,...	4	?
pla...	40,...	15,...	3	0.5

1 to 25,000 (25 Kb) 1 to 47 (47 bp)

E Q G Y Y Y Y I I P N G N G T

G A G C A G G G G T A T T A C T A C A T C A T T C C A A A T G G G A A C G G C A C T T

CIGAR:0

VCF:

1 U1 46 U46

A G G G G T A T T A C T A C T A C A T C A T T C C A A A T G G G A A C G G G C A C T T T

T T A T T A C T A C T A C A T C A T T C C A A A T G G G A A C G G G C A C T T T

T T A T T A C T A C T A C A T C A T T C C A A A T G G G A A C G G G C A C T T T

T T A T T A C T A C T A C A T C A T T C C A A A T G G G A A C G G G C A C T T T

C T A C A T C A T T C C A A A T G G G A A C G G G C A C T T T

C A A A T G G G A A C G G G C A C T T T

A A T G G G A A C G G G C A C T T T

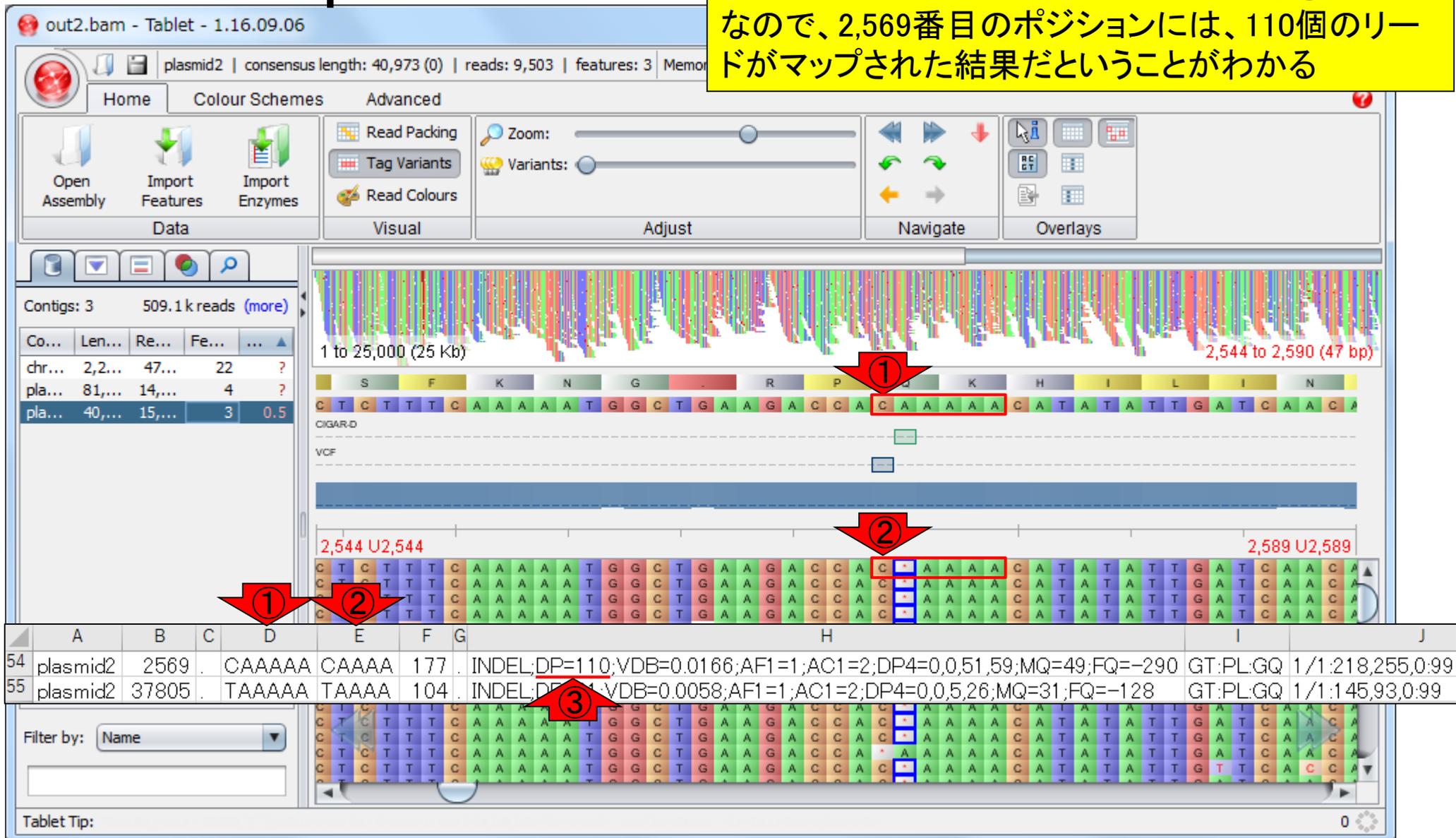
G G A A C G G G C A C T T T

A C

Tablet Tip: CTRL / CMD drag the mouse on the overview display to subset the overview

W20-1 : plasmid2

①リファレンス、CAAAAA。②ALT列の変異側(つまりマップする側のIlluminaリード配列)は、CAAAA。確かにマッピング結果はそうなっている。③DP=110なので、2,569番目のポジションには、110個のリードがマップされた結果だということがわかる



W20-1 : plasmid2

①下のほうに移動させて全体を眺めたところ。確かに概要配列のCAAAAAから、Aを1つ除去してCAAAAとしたほうがよさそうと判断できる

out2.bam - Tablet - 1.16.09.06

plasmid2 | consensus length: 40,973 (0) | reads: 9,503 | features: 3 | Memory usage: 86.52 MB (11)

Home Colour Schemes Advanced

Open Assembly Import Features Import Enzymes

Read Packing Tag Variants Read Colours

Zoom: Variants:

Adjust Navigate Overlays

Contigs: 3 509.1 k reads (more)

Co...	Len...	Re...	Fe...	...
chr...	2,2...	47...	22	?
pla...	81,...	14,...	4	?
pla...	40,...	15,...	3	0.5

1 to 25,000 (25 Kb) 2,544 to 2,590 (47 bp)

S F K N G R P Q K H I L I N

C T C T T T C A A A A A A T G G C T G A A G A C C A C A A A A A C A T A T A T T G A T C A A C A A

CIGAR-D

VCF

2,544 U2,544 2,589 U2,589

C T C T T T C A A A A A A T G G C T G A A G A C C A C A A A A A C A T A T A T T G A T C A A C A A

C T C T T T C A A A A A A T G G C T G A A G A C C A C A A A A A C A T A T A T T G A T C A A C A A

C T C T T T C A A A A A A T G G C T G A A G A C C A C A A A A A C A T A T A T T G A T C A A C A A

C T C T T T C A A A A A A T G G C T G A A G A C C A C A A A A A C A T A T A T T G A T C A A C A A

C T C T T T C A A A A A A T G G C T G A A G A C C A C A A A A A C A T A T A T T G A T C A A C A A

C T C T T T C A A A A A A T G G C T G A A G A C C A C A A A A A C A T A T A T T G A T C A A C A A

C T C T T T C A A A A A A T G G C T G A A G A C C A C A A A A A C A T A T A T T G A T C A A C A A

C T C T T T C A A A A A A T G G C T G A A G A C C A C A A A A A C A T A T A T T G A T C A A C A A

C T C T T T C A A A A A A T G G C T G A A G A C C A C A A A A A C A T A T A T T G A T C A A C A A

C T C T T T C A A A A A A T G G C T G A A G A C C A C A A A A A C A T A T A T T G A T C A A C A A

C T C T T T C A A A A A A T G G C T G A A G A C C A C A A A A A C A T A T A T T G A T C A A C A A

C T C T T T C A A A A A A T G G C T G A A G A C C A C A A A A A C A T A T A T T G A T C A A C A A

Filter by: Name

Tablet Tip: Navigate around an alignment by clicking and dragging on either the overview display area or the main display area

W20-2:37,805番目

plasmid2中の2つめの変異を確認。①リファレンス、TAAAAA。②ALT列の変異側(つまりマップする側のIlluminaリード配列)は、TAAAA。確かにマッピング結果はそうになっている。③DP=31なので、37,805番目のポジションには、31個のリードがマップされた結果だということがわかる

	A	B	C	D	E	F	G	H	I	J
54	plasmid2	2569	CAAAAA	CAAAA	177	INDEL;DP=110;VDB=0.0166;AF1=1;AC1=2;DP4=0,0,51,59;MQ=49;FQ=-290				
55	plasmid2	37805	TAAAAA	TAAAA	104	INDEL;DP=31;VDB=0.0058;AF1=1;AC1=2;DP4=0,0,5,26;MQ=31;FQ=-128				

Tablet Tip: Quickly zoom in on points of interest by double clicking within the main display area

W20-2:37,805番目

①下のほうに移動させて全体を眺めたところ。確かに概要配列のTAAAAAから、Aを1つ除去してTAAAAとしたほうがよさそうと判断できる

out2.bam - Tablet - 1.16.09.06

plasmid2 | consensus length: 40,973 (0) | reads: 9,379 | features: 3 | Memory usage: 93.79 MB (11)

Home Colour Schemes Advanced

Open Assembly Import Features Import Enzymes

Read Packing Tag Variants Read Colours

Zoom: Variants:

Navigate Overlays

Contigs: 3 509.1 k reads (more)

Co...	Len...	Re...	Fe...	...
chr...	2,2...	47...	22	?
pla...	81,...	14,...	4	?
pla...	40,...	15,...	3	0.5

15,974 to 40,973 (25 Kb) 37,783 to 37,829 (47 bp)

R F L W A Y R L K R S D D S D C

G G T T C C T T T G G G C A T A C C G G T T A A A A A G A T C G G A T G A C A G C G A C T G

CIGAR: VCF:

37,783 U37,783 37,828 U37,828

G G T T C C T T T G G G C A T A C C G G T T A A A A A G A T C G G A T G A C A G C G A C T G

G G T T C C T T T G G G C A T A C C G G T T A A A A A G A T C G G A T G A C A G C G A C T G

G G T T C C T T T G G G C A T A C C G G T T A A A A A G A T C G G A T G A C A G C G A C T G

G G T T C C T T T G G G C A T A C C G G T T A A A A A G A T C G G A T G A C A G C G A C T G

G G T T C C T T T G G G C A T A C C G G T T A A A A A G A T C G G A T G A C A G C G A C T G

Filter by: Name

Tablet Tip: Right click on the Contigs table to see options for copying or saving its data to the clipboard or a file

次は、①plasmid1上の、4か所の変異をチェック。まずは②37,177番目

W21-1 : plasmid1

	A	B	C	D	E	F	G	H	I	J
27	#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	/home/w3pipeline/refdata/
28	chromosome	141598	.	A	G	17.1	.	DP=7;VDB=0.0077;AF1=0.5;AC1	GT:PL:GQ	0/1:47,0,138:50
29	chromosome	242819	.	T	C	20	.	DP=9;VDB=0.0111;AF1=0.5;AC1	GT:PL:GQ	0/1:50,0,139:53
30	chromosome	359141	.	T	C	4.13	.	DP=14;VDB=0.0037;AF1=0.4998	GT:PL:GQ	0/1:32,0,187:31
31	chromosome	463882	.	T	G	6.2	.	DP=13;VDB=0.0074;AF1=0.4999	GT:PL:GQ	0/1:35,0,216:36
32	chromosome	660031	.	C	T	8.64	.	DP=11;VDB=0.0063;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,197:40
33	chromosome	792963	.	T	C	4.13	.	DP=14;VDB=0.0153;AF1=0.4998	GT:PL:GQ	0/1:32,0,200:31
34	chromosome	887422	.	G	A	10.4	.	DP=11;VDB=0.0070;AF1=0.5;AC	GT:PL:GQ	0/1:40,0,188:42
35	chromosome	913470	.	T	C	14.2	.	DP=9;VDB=0.0056;AF1=0.5;AC1	GT:PL:GQ	0/1:44,0,144:47
36	chromosome	965171	.	T	C	17.1	.	DP=20;VDB=0.0071;AF1=0.5;AC	GT:PL:GQ	0/1:47,0,249:50
37	chromosome	1134950	.	A	G	7.8	.	DP=23;VDB=0.0076;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,255:39
38	chromosome	1160048	.	A	G	29	.	DP=17;VDB=0.0054;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
39	chromosome	1160095	.	A	G	19.1	.	DP=10;VDB=0.0033;AF1=0.5;AC	GT:PL:GQ	0/1:49,0,191:52
40	chromosome	1240248	.	GCCCC	GCCCCC	214	.	INDEL;DP=48;VDB=0.0137;AF1=	GT:PL:GQ	1/1:255,128,0:99
41	chromosome	1455552	.	CAAAA	CAAAAA	214	.	INDEL;DP=58;VDB=0.0164;AF1=	GT:PL:GQ	1/1:255,130,0:99
42	chromosome	1475721	.	T	C	52	.	DP=19;VDB=0.0021;AF1=0.5;AC	GT:PL:GQ	0/1:82,0,246:85
43	chromosome	1752680	.	G	A	29	.	DP=24;VDB=0.0129;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
44	chromosome	1941380	.	T	C	7.8	.	DP=13;VDB=0.0046;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,214:39
45	chromosome	2052955	.	A	G	12.3	.	DP=8;VDB=0.0084;AF1=0.5;AC1	GT:PL:GQ	0/1:42,0,103:44
46	chromosome	2054084	.	TC	T	13.7	.	INDEL;DP=8;VDB=0.0009;AF1=0	GT:PL:GQ	0/1:51,0,174:54
47	chromosome	2056279	.	A	G	24	.	DP=6;VDB=0.0057;AF1=0.5;AC1	GT:PL:GQ	0/1:54,0,113:57
48	chromosome	2057564	.	T	C	8.64	.	DP=10;VDB=0.0041;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,129:40
49	chromosome	2058145	.	C	T,G	108	.	DP=14;VDB=0.0147;AF1=1;AC1=	GT:PL:GQ	1/1:141,39,0,141,28,138:75
50	plasmid1	37177	.	G	T	54	.	DP=49;VDB=0.0087;AF1=0.5;AC	GT:PL:GQ	0/1:84,0,255:87
51	plasmid1	66609	.	A	G	225	.	DP=55;VDB=0.0060;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
52	plasmid1	66637	.	C	T	225	.	DP=71;VDB=0.0141;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,249:99
53	plasmid1	66668	.	C	T	225	.	DP=80;VDB=0.0153;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
54	plasmid2	2569	.	CAAAA	CAAAA	177	.	INDEL;DP=110;VDB=0.0166;AF1	GT:PL:GQ	1/1:218,255,0:99
55	plasmid2	37805	.	TAAAAA	TAAAA	104	.	INDEL;DP=31;VDB=0.0058;AF1=	GT:PL:GQ	1/1:145,93,0:99



W21-1:37,177番目

①plasmid1に切り替え。②リファレンスはG。③ALT列の変異側(つまりマップする側のIlluminaリード配列)はT。そういうのも確かにあるが、この場合はGのままでいいだろうと判断。④長いほうを見られないときはここを押す

out2.bam - Tablet - 1.16.09.06

plasmid1 | consensus length: 81,630 (0) | reads: 4,722 | features: 4 | Memory usage: ...

Home Colour Schemes Advanced

Open Assembly Import Features Import Enzymes

Read Packing Tag Variants Read Colours

Zoom: Variants:

Navigate Overlays

Contigs: 3 509.1 k reads (more)

Co...	Len...	Re...	Fe...	...
chr...	2,2...	47...	22	
pla...	81,...	14,...	4	
pla...	40,...	15,...	3	

20,001 to 45,000 (25 Kb) 37,155 to 37,201 (47 bp)

K T S V S L A A L R S S Q P

A A A A C A T C G G T A T C G C T A G C G G G C T T A A C T G C G A T C C T C G C A G C C C A

CIGAR-D

VCF

37,155 U37,155 37,200 U37,200

A A A A C A T C G G T A T C G C T A G C G G G C T T A A C T G C G A T C C T C G C A G C C C A

Filter by: Name

Tablet Tip: Right click on the features track to access the option to select which tracks are visible

W21-2: 66,609番目

66,609番目の塩基を表示。①リファレンスはA。
②ALT列の変異側(つまりマップする側のIllumina
リード配列)はG。Gが全体的にAに比べて多い
が、これくらいならAのままでいいだろうと判断

out2.bam - Tablet - 1.16.09.06

plasmid1 | consensus length: 81,630 (0) | reads: 4,797 | features: 4 | Memory usage: 113.00 MB (12)

Home Colour Schemes Advanced

Open Assembly Import Features Import Enzymes

Read Packing Tag Variants Read Colours

Zoom: Variants:

Navigate Overlays

Contigs: 3 509.1 k reads (more)

Co...	Len...	Re...	Fe...	...
chr...	2,2...	47...	22	?
pla...	81,...	14,...	4	0.5
pla...	40,...	15,...	3	?

56,631 to 81,630 (25 Kb) 66,587 to 66,634 (48 bp)

G H S K A V L L T L I D R K S

T G G G C A T A G T A A G G C A G T T T T A T T A A C T T T A A T C G A T C G A A A A T C A

CIGAR-D

VCF

66,587 U66,587 66,633 U66,633

T G G G C A T A G T A A G G C A G T T T T A T T A A C T T T A A T C G A T C G A A A A T C A

T G G G C A T A G T A A G G C A G T T T T A T T A A C T T T A A T C G A T C G A A A A T C A

T G G G C A T A G T A A G G C A G T T T T A T T A A C T T T A A T C G A T C G A A A A T C A

T G G G C A T A G T A A G G C A G T T T T A T T A A C T T T A A T C G A T C G A A A A T C A

T G G G C A T A G T A A G G C A G T T T T A T T A A C T T T A A T C G A T C G A A A A T C A

T G G G C A T A G T A A G G C A G T T T T A T T A A C T T T A A T C G A T C G A A A A T C A

T G G G C A T A G T A A G G C A G T T T T A T T A A C T T T A A T C G A T C G A A A A T C A

T G G G C A T A G T A A G G C A G T T T T A T T A A C T T T A A T C G A T C G A A A A T C A

T G G G C A T A G T A A G G C A G T T T T A T T A A C T T T A A T C G A T C G A A A A T C A

T G G G C A T A G T A A G G C A G T T T T A T T A A C T T T A A T C G A T C G A A A A T C A

Tablet Tip: Mouse over a feature on the feature track to see information including its name and start and end positions

W21-3: 66,637と66,668番目

①66,637番目と②66,668番目の塩基を表示。このあたりも概要側を変えなくてもいいかな…(個人の感想です)

out2.bam - Tablet - 1.16.09.06

plasmid1 | consensus length: 81,630 (0) | reads: 4,797 | features: 4 | Memory usage: 107.55 MB (11)

Home | Colour Schemes | Advanced

Open Assembly | Import Features | Import Enzymes

Read Packing | Tag Variants | Read Colours

Zoom: [Slider] | Variants: [Slider]

Adjust | Navigate | Overlays

Contigs: 3 | 509.1 k reads (more)

Co...	Len...	Re...	Fe...	...
chr...	2,2...	47...	22	?
pla...	81,...	14,...	4	0.5
pla...	40,...	15,...	3	?

56,631 to 81,630 (25,146)

66,629 to 66,676 (48 bp)

S R L W A Y R L K D R T A

A T C A C G G C T C C T T T G G G C A T A C C G G T T A A A A G A T C G G A C G A C A G C G

CIGAR-D

VCF

66,629 U66,629 | 66,675 U66,675

A T C A C G G C T C C T T T G G G C A T A C C C G G C T C A A A G A T C G G

A T C A C G G C T C C T T T G G G C A T A C C G G G T T A A A A A G A T C G G G A C G A

A T C A C G G C T C C T T T G G G C A T A C C G G G T T A A A A A G A T C G G G C C G C C A G C

A T C A C G G T T C C T T T G G G C A T A C C G G T T A A A A A G A T A G G A T T G A C A G C G

A T C A C G G T T C C T T T G G G C A T A C C G G T T A A A A A G A T C G G A T T G A C A G C G

A T C A C G G T T C C T T T G G G C A T A C C G G T T A A A A A G A T C G G A T T G A C A G C G

A T C A C G G T T C C T T T G G G C A T A C C G G T T A A A A A G A T C G G A T T G A C A G C G

A T C A C G G T T C C T T T G G G C A T A C C G G T T A A A A A G A T C G G A T T G A C A G C G

A T C A C G G C T C C T T T G G G C A T A C C G G T T A A A A A G A T C G G A C C G A C A G C G

A T C A C G G C T C C T T T G G G C A T A C C G G T T A A A A A G A T C G G A C C G A C A G C G

A T C A C G G C T C C T T T G G G C A T A C C G G T T A A A A A G A T C G G A C C G A C A G C G

A T C A C G G C T C C T T T G G G C A T A C C G G T T A A A A A G A T C G G A C C G A C A G C G

Filter by: Name

Tablet Tip: Use the Search tab to search for subsequences within the reads, or within the consensus / reference

今度はchromosomeを確認。
とりあえず①2,054,084番目

W22-1 : chromosome

	A	B	C	D	E	F	G	H	I	J
27	#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	/home/w3pipeline/refdata/
28	chromosome	141598	A	G	G	17.1		DP=7;VDB=0.0077;AF1=0.5;AC1	GT:PL:GQ	0/1:47,0,138:50
29	chromosome	242819	T	C	C	20		DP=9;VDB=0.0111;AF1=0.5;AC1	GT:PL:GQ	0/1:50,0,139:53
30	chromosome	359141	T	C	C	4.13		DP=14;VDB=0.0037;AF1=0.4998	GT:PL:GQ	0/1:32,0,187:31
31	chromosome	463882	T	G	G	6.2		DP=13;VDB=0.0074;AF1=0.4999	GT:PL:GQ	0/1:35,0,216:36
32	chromosome	660031	C	T	T	8.64		DP=11;VDB=0.0063;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,197:40
33	chromosome	792963	T	C	C	4.13		DP=14;VDB=0.0153;AF1=0.4998	GT:PL:GQ	0/1:32,0,200:31
34	chromosome	887422	G	A	A	10.4		DP=11;VDB=0.0070;AF1=0.5;AC	GT:PL:GQ	0/1:40,0,188:42
35	chromosome	913470	T	C	C	14.2		DP=9;VDB=0.0056;AF1=0.5;AC1	GT:PL:GQ	0/1:44,0,144:47
36	chromosome	965171	T	C	C	17.1		DP=20;VDB=0.0071;AF1=0.5;AC	GT:PL:GQ	0/1:47,0,249:50
37	chromosome	1134950	A	G	G	7.8		DP=23;VDB=0.0076;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,255:39
38	chromosome	1160048	A	G	G	29		DP=17;VDB=0.0054;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
39	chromosome	1160095	A	G	G	19.1		DP=10;VDB=0.0033;AF1=0.5;AC	GT:PL:GQ	0/1:49,0,191:52
40	chromosome	1240248	GCCCC	GCCCCC	GCCCCC	214		INDEL;DP=48;VDB=0.0137;AF1=	GT:PL:GQ	1/1:255,128,0:99
41	chromosome	1455552	CAAAA	CAAAAA	CAAAAA	214		INDEL;DP=58;VDB=0.0164;AF1=	GT:PL:GQ	1/1:255,130,0:99
42	chromosome	1475721	T	C	C	52		DP=19;VDB=0.0021;AF1=0.5;AC	GT:PL:GQ	0/1:82,0,246:85
43	chromosome	1752680	G	A	A	29		DP=24;VDB=0.0129;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
44	chromosome	1941380	T	C	C	7.8		DP=13;VDB=0.0046;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,214:39
45	chromosome	2052955	A	G	G	12.3		DP=8;VDB=0.0084;AF1=0.5;AC1	GT:PL:GQ	0/1:42,0,103:44
46	chromosome	2054084	TC	T	T	13.7		INDEL;DP=8;VDB=0.0009;AF1=0	GT:PL:GQ	0/1:51,0,174:54
47	chromosome	2056279	A	G	G	24		DP=6;VDB=0.0057;AF1=0.5;AC1	GT:PL:GQ	0/1:54,0,113:57
48	chromosome	2057564	T	C	C	8.64		DP=10;VDB=0.0041;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,129:40
49	chromosome	2058145	C	T,G	T,G	108		DP=14;VDB=0.0147;AF1=1;AC1=	GT:PL:GQ	1/1:141,39,0,141,28,138:75
50	plasmid1	37177	G	T	T	54		DP=49;VDB=0.0087;AF1=0.5;AC	GT:PL:GQ	0/1:84,0,255:87
51	plasmid1	66609	A	G	G	225		DP=55;VDB=0.0060;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
52	plasmid1	66637	C	T	T	225		DP=71;VDB=0.0141;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,249:99
53	plasmid1	66668	C	T	T	225		DP=80;VDB=0.0153;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
54	plasmid2	2569	CAAAA	CAAAA	CAAAA	177		INDEL;DP=110;VDB=0.0166;AF1	GT:PL:GQ	1/1:218,255,0:99
55	plasmid2	37805	TAAAAA	TAAAA	TAAAA	104		INDEL;DP=31;VDB=0.0058;AF1=	GT:PL:GQ	1/1:145,93,0:99



W22-2: 2,054,084番目

①chromosomeに変更して、2,054,085番目に移動したい。長すぎるので②または③を押して移動するのは途方に暮れる

The screenshot shows the Tablet software interface. At the top, the window title is "out2.bam - Tablet - 1.16.09.06". Below the title bar, there are tabs for "Home", "Colour Schemes", and "Advanced". The "Home" tab is active, showing various tool icons for "Data" (Open Assembly, Import Features, Import Enzymes), "Visual" (Read Packing, Tag Variants, Read Colours), "Adjust" (Zoom and Variants sliders), "Navigate" (Navigation arrows), and "Overlays" (Grid, Legend, etc.).

The main display area shows a genomic track. The top part is a read alignment view with a consensus sequence: V A L Q K Q M H V A L N V D. Below this is a VCF track and a CIGAR-D track. The bottom part shows a detailed view of the alignment with read identifiers like "347 U347" and "393 U393".

Annotations include "1" pointing to the "Contigs" table and "2" pointing to the read alignment view. "3" points to the navigation arrows in the "Navigate" panel.

Co...	Len...	Re...	Fe...	M...
chr...	2,2...	47...	22	
pla...	81,...	14,...	4	
pla...	40,...	15,...	3	?

Filter by: Name

Tablet Tip: GFF3 or BED data can be displayed on the features track after importing

W22-2: 2,054,084番目

①が指定した塩基ポジションにジャンプできるボタンなので押す。② 2054084と入力して、③Padded Jump

out2.bam - Tablet - 1.16.09.06

chromosome | consensus length: 2,277,981 (0) | reads: 5,469 | features: 22 | Memory usage: 188.24 MB (11)

Home Colour Schemes Advanced

Open Assembly Import Features Import Enzymes

Read Packing Tag Variants Read Colours

Zoom: Variants:

Jump to Base (Ctrl+J)
Instantly jump to a specified base position.

Jump to Base

Jump to base: 2054084 Padded Jump Unpadded Jump

Padded jumps count pad (*) characters in the consensus when determining the base position. Unpadded jumps do not.

You can jump to a position in another contig by entering: contig-name:position

Close Help

Contigs: 3 509.1 k reads (more)

Co...	Len...	Re...	Fe...	Mis...
chr...	2,2...	47...	22	0.4
pla...	81,...	14,...	4	?
pla...	40,...	15,...	3	?

Filter by: Name

Tablet Tip: Use the Search tab to search for subsequences within the reads, or within the consensus / reference

W22-2: 2,054,084番目

①2,054,084番目に移動完了。ここは「リファレンスはTCだけど、Tだけでいいんじゃないでしょうか」と提案されているところ。TCのままでいいんじゃないかという判断を下す

out2.bam - Tablet - 1.16.09.06

chromosome | consensus length: 2,277,981 (0) | reads: 4,002 | features: 22 | Memory usage: 244.41 MB (11)

Home | Colour Schemes | Advanced

Open Assembly | Import Features | Import Enzymes

Read Packing | Tag Variants | Read Colours

Zoom: [Slider] | Variants: [Slider]

Adjust | Navigate | Overlays

Contigs: 3 | 509.1 k reads (more)

Co...	Len...	Re...	Fe...	Mis...
chr...	2,2...	47...	22	0.5
pla...	81,...	14,...	4	?
pla...	40,...	15,...	3	?

2,049,084 to 2,074,083 (25 Kb)

2,054,061 to 2,054,107 (47 bp)

L I V E L L K M V R I R

T T G A T T T A A G T G G A A T T A C T A G T C A A A A T G T A G G T T C G T A T C A G G T

VCF

CIGAR-D

2,054,061 U2,054,061

2,054,106 U2,054,106

T T G A T T T A A G T G G A A T T A C T A G T C A A A A T G T A G G T T C G T A T C A G G T

T T G A T T T A A G T G G A A T T A C T A G T C A A A A T G T A G G T T C G T A T C A G G T

T T G A T T T A A G T G G A A T T A C T A G T C A A A A T G T A G G T T C G T A T C A G G T

T T G A T T T A A G T G G A A T T A C T A G T C A A A A T G T A G G T T C G T A T C A G G T

T T G A T T T A A G T G G A A T T A C T A G T C A A A A T G T A G G T T C G T A T C A G G T

T T G A T T T A A G T G G A A T T A C T A G T C A A A A T G T A G G T T C G T A T C A G G T

Filter by: Name

Tablet Tip: Use the Search tab to search for subsequences within the reads, or within the consensus / reference

W22-3: 1,240,248番目

次は、①1,240,248番目。これはリファレンス配列にCを1個挿入したほうがよいのでは、というもの

	A	B	C	D	E	F	G	H	I	J
27	#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	/home/w3pipeline/refdata/
28	chromosome	141598	A	G		17.1		DP=7;VDB=0.0077;AF1=0.5;AC1	GT:PL:GQ	0/1:47,0,138:50
29	chromosome	242819	T	C		20		DP=9;VDB=0.0111;AF1=0.5;AC1	GT:PL:GQ	0/1:50,0,139:53
30	chromosome	359141	T	C		4.13		DP=14;VDB=0.0037;AF1=0.4998	GT:PL:GQ	0/1:32,0,187:31
31	chromosome	463882	T	G		6.2		DP=13;VDB=0.0074;AF1=0.4999	GT:PL:GQ	0/1:35,0,216:36
32	chromosome	660031	C	T		8.64		DP=11;VDB=0.0063;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,197:40
33	chromosome	792963	T	C		4.13		DP=14;VDB=0.0153;AF1=0.4998	GT:PL:GQ	0/1:32,0,200:31
34	chromosome	887422	G	A		10.4		DP=11;VDB=0.0070;AF1=0.5;AC	GT:PL:GQ	0/1:40,0,188:42
35	chromosome	913470	T	C		14.2		DP=9;VDB=0.0056;AF1=0.5;AC1	GT:PL:GQ	0/1:44,0,144:47
36	chromosome	965171	T	C		17.1		DP=20;VDB=0.0071;AF1=0.5;AC	GT:PL:GQ	0/1:47,0,249:50
37	chromosome	1134950	A	G		7.8		DP=23;VDB=0.0076;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,255:39
38	chromosome	1160048	A	G		29		DP=17;VDB=0.0054;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
39	chromosome	1160095	A	G		19.1		DP=10;VDB=0.0033;AF1=0.5;AC	GT:PL:GQ	0/1:49,0,191:52
40	chromosome	1240248	GCCCC	GCCCCC		214		INDEL;DP=48;VDB=0.0137;AF1=	GT:PL:GQ	1/1:255,128,0:99
41	chromosome	1455552	CAAAA	CAAAAA		214		INDEL;DP=58;VDB=0.0164;AF1=	GT:PL:GQ	1/1:255,130,0:99
42	chromosome	1475721	T	C		52		DP=19;VDB=0.0021;AF1=0.5;AC	GT:PL:GQ	0/1:82,0,246:85
43	chromosome	1752680	G	A		29		DP=24;VDB=0.0129;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
44	chromosome	1941380	T	C		7.8		DP=13;VDB=0.0046;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,214:39
45	chromosome	2052955	A	G		12.3		DP=8;VDB=0.0084;AF1=0.5;AC1	GT:PL:GQ	0/1:42,0,103:44
46	chromosome	2054084	TC	T		13.7		INDEL;DP=8;VDB=0.0009;AF1=0	GT:PL:GQ	0/1:51,0,174:54
47	chromosome	2056279	A	G		24		DP=6;VDB=0.0057;AF1=0.5;AC1	GT:PL:GQ	0/1:54,0,113:57
48	chromosome	2057564	T	C		8.64		DP=10;VDB=0.0041;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,129:40
49	chromosome	2058145	C	T,G		108		DP=14;VDB=0.0147;AF1=1;AC1=	GT:PL:GQ	1/1:141,39,0,141,28,138:75
50	plasmid1	37177	G	T		54		DP=49;VDB=0.0087;AF1=0.5;AC	GT:PL:GQ	0/1:84,0,255:87
51	plasmid1	66609	A	G		225		DP=55;VDB=0.0060;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
52	plasmid1	66637	C	T		225		DP=71;VDB=0.0141;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,249:99
53	plasmid1	66668	C	T		225		DP=80;VDB=0.0153;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
54	plasmid2	2569	CAAAA	CAAAA		177		INDEL;DP=110;VDB=0.0166;AF1	GT:PL:GQ	1/1:218,255,0:99
55	plasmid2	37805	TAAAAA	TAAAA		104		INDEL;DP=31;VDB=0.0058;AF1=	GT:PL:GQ	1/1:145,93,0:99

①

W22-3: 1,240,248番目

こんな感じに見えます。①
指定したポジションはここ

The screenshot shows the Tablet software interface for viewing a BAM file. The window title is "out2.bam - Tablet - 1.16.09.06". The top bar displays "chromosome | consensus length: 2,277,981 (0) | reads: 5,803 | features: 23 | Memory usage: 123.15 MB (11)". The interface includes a menu bar (Home, Colour Schemes, Advanced) and a toolbar with sections for Data (Open Assembly, Import Features, Import Enzymes), Visual (Read Packing, Tag Variants, Read Colours), Adjust (Zoom, Variants), Navigate, and Overlays.

On the left, a contigs table is visible:

Co...	Len...	Re...	Fe...	Mis...
chr...	2,2...	47...	23	0.5
pla...	81,...	14,...	4	?
pla...	40,...	15,...	2	?

The main view shows a genomic track for the region 1,235,248 to 1,260,247 (25 Kb). A red arrow labeled "①" points to a specific position, 1,240,225 to 1,240,271 (47 bp). Below the track, the consensus sequence is displayed: Q T A G S L P R S T A L K P. The corresponding nucleotide sequence is: A G A C T G C G G G C A G C C T C C C A T T G C C C C G G T C A A C C G C T T T G A A A C C. Below this, the CIGAR string is shown as "U1,240,225" and "U1,240,270". The bottom of the interface includes a "Tablet Tip" and a filter dropdown set to "Name".

W22-3: 1,240,248番

②リファレンス配列のGとCの間に、insertion (挿入)の変異が入っている場合には、こんな感じに見えます。③マップされる側(Illuminaリード)は赤枠で表現されます。このあたりは慣れです

out2.bam - Tablet - 1.16.09.06

chromosome | consensus length: 2,277,981 (0) | reads: 5,803 | features: 23 | Memory usage: 123.15 MB (11)

Home | Colour Schemes | Advanced

Open Assembly | Import Features | Import Enzymes

Read Packing | Tag Variants | Read Colours

Zoom: [Slider] | Variants: [Slider]

Navigation: [Left Arrow] [Right Arrow] [Down Arrow] [Up Arrow] [Home] [End]

Overlays: [Grid] [Grid] [Grid] [Grid]

Contigs: 3 | 509.1 k reads (more)

Co...	Len...	Re...	Fe...	Mis...
chr...	2,2...	47...	23	0.5
pla...	81,...	14,...	4	?
pla...	40,...	15,...	2	?

1,235,248 to 1,260,247 (25 Kb) | 1,240,225 to 1,240,271 (47 bp)

Q T A G S L P P R S T A L K P

A G A C T G C G G G C A G C C T C C C A T T G C C C C G G T C A A C C G C T T T G A A A C C

CIGAR: |

VCF: |

1,240,225 U1,240,225 | 1,240,270 U1,240,270

A G A C T G C G G G C A G C C T C C C A T T G C C C C G G T C A A C C G C T T T G A A A C C

A G A C T G C G G G C A G C C T C C C A T T G C C C C G G T C A A C C G C T T T G A A A C C

A G A C T G C G G G C A G C C T C C C A T T G C C C C G G T C A A C C G C T T T G A A A C C

A G A C T G C G G G C A G C C T C C C A T T G C C C C G G T C A A C C G C T T T G A A A C C

A G A C T G C G G G C A G C C T C C C A T T G C C C C G G T C A A C C G C T T T G A A A C C

A G A C T G C G G G C A G C C T C C C A T T G C C C C G G T C A A C C G C T T T G A A A C C

A G A C T G C G G G C A G C C T C C C A T T G C C C C G G T C A A C C G C T T T G A A A C C

A G A C T G C G G G C A G C C T C C C A T T G C C C C G G T C A A C C G C T T T G A A A C C

A G A C T G C G G G C A G C C T C C C A T T G C C C C G G T C A A C C G C T T T G A A A C C

Tablet Tip: Green (rather than blue) navigation arrows mean you have reached the edge of the current BAM data block

W22-3: 1,240,248番目

①のあたりにカーソル移動させると、Cのinsertionであることがわかります

out2.bam - Tablet - 1.16.09.06

chromosome | consensus length: 2,277,981 (0) | reads: 5,803 | features: 23 | Memory usage: 126.34 MB (11)

Home | Colour Schemes | Advanced

Open Assembly | Import Features | Import Enzymes

Read Packing | Tag Variants | Read Colours

Zoom: [Slider] | Variants: [Slider]

Adjust | Navigate | Overlays

Contigs: 3 | 509.1 k reads (more)

Co...	Len...	Re...	Fe...	Mis...
chr...	2,2...	47...	23	0.5
pla...	81,...	14,...	4	?
pla...	40,...	15,...	2	?

1,235,248 to 1,260,247 (25 Kb) | 1,240,225 to 1,240,271 (47 bp)

Q T A G S L P L P R S T A L K P

A G A C T G C G G G C A G C C T C C C A T T G C C C C G G T C A A C C G C T T T G A A A C C

CIGAR:U

VCF

1,240,225 U1,240,225 | 1,240,248 U1,

A G A C T G C G G G C A G C C T C C C A T T T G C C C C G G T C A A C C G C T T T G A A A C C

A G A C T G C G G G C A G C C T C C C A T T T G C C C C G G T C A A C C G C T T T G A A A C C

A G A C T G C G G G C A G C C T C C C A T T T G C C C C G G T C A A C C G C T T T G A A A C C

A G A C T G C G G G C A G C C T C C C A T T T G C C C C G G T C A A C C G C T T T G A A A C C

VCF

Name:

Padded: 1,240,248 to 1,240,248

Unpadded: 1,240,248 to 1,240,248

Tags: GCCCC
GCCCCC
214

INDEL;DP=48;VDB=0.0137;AF1=1;AC1=2;DP4=1,0,21,26;MG
GT:PL:GQ

Tablet Tip: Mouse over a feature on the feature track to see information including its name and start and end positions

W22-4:2,058,145番目

	A	B	C	D	E	F	G	H	I	J
27	#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	/home/w3pipeline/refdata/
28	chromosome	141598	A	G	G	17.1		DP=7;VDB=0.0077;AF1=0.5;AC1	GT:PL:GQ	0/1:47,0,138:50
29	chromosome	242819	T	C	C	20		DP=9;VDB=0.0111;AF1=0.5;AC1	GT:PL:GQ	0/1:50,0,139:53
30	chromosome	359141	T	C	C	4.13		DP=14;VDB=0.0037;AF1=0.4998	GT:PL:GQ	0/1:32,0,187:31
31	chromosome	463882	T	G	G	6.2		DP=13;VDB=0.0074;AF1=0.4999	GT:PL:GQ	0/1:35,0,216:36
32	chromosome	660031	C	T	T	8.64		DP=11;VDB=0.0063;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,197:40
33	chromosome	792963	T	C	C	4.13		DP=14;VDB=0.0153;AF1=0.4998	GT:PL:GQ	0/1:32,0,200:31
34	chromosome	887422	G	A	A	10.4		DP=11;VDB=0.0070;AF1=0.5;AC	GT:PL:GQ	0/1:40,0,188:42
35	chromosome	913470	T	C	C	14.2		DP=9;VDB=0.0056;AF1=0.5;AC1	GT:PL:GQ	0/1:44,0,144:47
36	chromosome	965171	T	C	C	17.1		DP=20;VDB=0.0071;AF1=0.5;AC	GT:PL:GQ	0/1:47,0,249:50
37	chromosome	1134950	A	G	G	7.8		DP=23;VDB=0.0076;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,255:39
38	chromosome	1160048	A	G	G	29		DP=17;VDB=0.0054;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
39	chromosome	1160095	A	G	G	19.1		DP=10;VDB=0.0033;AF1=0.5;AC	GT:PL:GQ	0/1:49,0,191:52
40	chromosome	1240248	GCCCC	GCCCCC	GCCCCC	214		INDEL;DP=48;VDB=0.0137;AF1=	GT:PL:GQ	1/1:255,128,0:99
41	chromosome	1455552	CAAAA	CAAAAA	CAAAAA	214		INDEL;DP=58;VDB=0.0164;AF1=	GT:PL:GQ	1/1:255,130,0:99
42	chromosome	1475721	T	C	C	52		DP=19;VDB=0.0021;AF1=0.5;AC	GT:PL:GQ	0/1:82,0,246:85
43	chromosome	1752680	G	A	A	29		DP=24;VDB=0.0129;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
44	chromosome	1941380	T	C	C	7.8		DP=13;VDB=0.0046;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,214:39
45	chromosome	2052955	A	G	G	12.3		DP=8;VDB=0.0084;AF1=0.5;AC1	GT:PL:GQ	0/1:42,0,103:44
46	chromosome	2054084	TC	T	T	13.7		INDEL;DP=8;VDB=0.0009;AF1=0	GT:PL:GQ	0/1:51,0,174:54
47	chromosome	2056279	A	G	G	24		DP=6;VDB=0.0057;AF1=0.5;AC1	GT:PL:GQ	0/1:54,0,113:57
48	chromosome	2057564	T	C	C	8.64		DP=10;VDB=0.0041;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,129:40
49	chromosome	2058145	C	T,G	T,G	108		DP=14;VDB=0.0147;AF1=1;AC1=	GT:PL:GQ	1/1:141,39,0,141,28,138:75
50	plasmid1	37177	G	T	T	54		DP=49;VDB=0.0087;AF1=0.5;AC	GT:PL:GQ	0/1:84,0,255:87
51	plasmid1	66609	A	G	G	225		DP=55;VDB=0.0060;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
52	plasmid1	66637	C	T	T	225		DP=71;VDB=0.0141;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,249:99
53	plasmid1	66668	C	T	T	225		DP=80;VDB=0.0153;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
54	plasmid2	2569	CAAAA	CAAAA	CAAAA	177		INDEL;DP=110;VDB=0.0166;AF1	GT:PL:GQ	1/1:218,255,0:99
55	plasmid2	37805	TAAAAA	TAAAAA	TAAAAA	104		INDEL;DP=31;VDB=0.0058;AF1=	GT:PL:GQ	1/1:145,93,0:99



W22-4: 2,058,145番目

①リファレンス中のCではなく、TやGとなっていたが、これはTに置換したほうがよいと判断

out2.bam - Tablet - 1.16.09.06

chromosome | consensus length: 2,277,981 (0) | reads: 4,180 | features: 22 | Memory usage: 116.87 MB (11)

Home | Colour Schemes | Advanced

Open Assembly | Import Features | Import Enzymes

Read Packing | Tag Variants | Read Colours

Zoom: [Slider] | Variants: [Slider]

Adjust | Navigate | Overlays

Contigs: 3 | 509.1 k reads (more)

Co...	Len...	Re...	Fe...	Mis...
chr...	2,2...	47...	22	0.5
pla...	81,...	14,...	4	?
pla...	40,...	15,...	2	?

2,053,145 to 2,078,144 (25 Kb) | 2,058,122 to 2,058,168 (47 bp)

I T A A Q I T P T D P N T P

A A T C A C A G C A G C C C A A A T T A C A C C A C C A A C G G A T C C T A A A T A C A C C A

VCF

2,058,122 U2,058,122 | 2,058,167 U2,058,167

A A T C A C A G C A G C C C A A A T T A C A T C A C C A A C G G A G G A G G A T C C T A A T A C A C C A

Tablet Tip: Any CIGAR "I" (Insertion) tags in SAM/BAM files will be listed in the Features table

W23-1: Tips (リード情報)

①のあたりにポインタを移動させたところ。リードごとの情報を見ることができます

The screenshot shows the Tablet software interface. The top bar displays 'out2.bam - Tablet - 1.16.09.06' and various statistics: 'chromosome | consensus length: 2,277,981 (0) | reads: 4,180 | features: 22 | Memory usage: 154.87 MB (12)'. The main interface is divided into several sections: 'Data' (Open Assembly, Import Features, Import Enzymes), 'Visual' (Read Packing, Tag Variants, Read Colours), 'Adjust' (Zoom, Variants), 'Navigate', and 'Overlays'. The central panel shows a genomic track with a read shadower. A red arrow labeled '①' points to a specific read pair. A tooltip window is open over this read pair, displaying the following information:

- DRR024501.17415 (Mate)**
- From: 2,058,046 U2,058,046 to 2,058,296 U2,058,296
- Length: 251 U251 (2 mismatches)
- Cigar: 251M
- Properly paired (2/2); insert size: 395

The bottom of the interface includes a 'Tablet Tip: When the read shadower is in custom mode, right click on any base position to lock the highlighting' and a gear icon for settings.

W23-1:リード情報

①赤枠の上から順に、(1)リードID、(2)マップされた領域、(3)長さミスマッチ数、(4)CIGAR string、(5)paired-endでマップしているのもう片方(メイト)もマップされたかどうかやインサート長の情報を見られます

The screenshot shows the Tablet software interface. At the top, it displays 'out2.bam - Tablet - 1.16.09.06'. Below this, there are tabs for 'Home', 'Colour Schemes', and 'Advanced'. The main area shows a genomic track with reads aligned to a reference sequence. A detailed popup window is open, showing information for a specific read: 'DRR024501.17415 (Mate)'. The popup includes the following details:

- From: 2,058,046 U2,058,046 to 2,058,296 U2,058,296
- Length: 251 U251 (2 mismatches)
- Cigar: 251M
- Properly paired (2/2); insert size: 395

The popup also shows a visual representation of the read alignment and its mate. A red arrow points to the read ID 'DRR024501.17415' in the popup.

W23-1:リード情報

この位置(リファレンス配列の2,058,145番目の塩基上)にマップされた2つめのリードはこんな感じ

The screenshot shows the Tablet software interface. At the top, the window title is "out2.bam - Tablet - 1.16.09.06". Below the title bar, there are navigation buttons (Home, Colour Schemes, Advanced) and a status bar showing "chromosome | consensus length: 2,277,981 (0) | reads: 4,180 | features: 22 | Memory usage: 189.84 MB (11)".

The main interface is divided into several sections:

- Data:** Includes buttons for "Open Assembly", "Import Features", and "Import Enzymes".
- Visual:** Includes "Read Packing", "Tag Variants", and "Read Colours".
- Adjust:** Includes "Zoom:" and "Variants:" sliders.
- Navigate:** Includes navigation arrows.
- Overlays:** Includes icons for various overlays.

The central panel displays a genomic track for "contigs: 3" with "509.1 k reads (more)". The track shows a consensus sequence from position 2,053,145 to 2,078,144 (25 Kb). A zoomed-in view shows a specific region from 2,058,122 to 2,058,168 (47 bp). The reference sequence is shown as "I T A A Q I T P P T D P N T P" and the corresponding bases are "A A T C A C A G C A G C C C A A A T T A C A C C A A C G G A T C C T A A T A C A C C A".

A red arrow points to a specific read alignment at position 2,058,145. A tooltip for this read is shown, containing the following information:

- DRR024501.34809 (Mate)**
- From: 2,058,002 U2,058,002 to 2,058,252 U2,058,252
- Length: 251 U251 (2 mismatches)
- Cigar: 251M
- Properly paired (1/2), insert size: 348

At the bottom left, there is a "Filter by:" dropdown menu set to "Name". At the bottom right, there is a "Tablet Tip:" and a gear icon.

W23-2: Variants

① Variantsのところを右に移動させることは、リファレンス配列と異なる塩基(variants)以外を暗くさせていくことに相当します。つまりvariantsを相対的にハイライトさせることになります

out2.bam - Tablet - 1.16.09.06

chromosome | consensus length: 2,277,981 (0) | reads: 4,180 | features: 22 | Memory usage: 179.47 MB (11)

Home | Colour Schemes | Advanced

Open Assembly | Import Features | Import Enzymes

Read Packing | Tag Variants | Read Colours

Zoom: [Slider] | Variants: [Slider]

Adjust | Navigate | Overlays

Contigs: 3 | 509.1 k reads (more)

Co...	Len...	Re...	Fe...	Mis...
chr...	2,2...	47...	22	0.5
pla...	81,...	14,...	4	?
pla...	40,...	15,...	2	?

2,053,145 to 2,078,144 (25 Kb) | 2,058,122 to 2,058,168 (47 bp)

T I T A A Q I T P P T D P N T P

C A A T C A C A G C A G C C C A A A T T A C A C C A C C A A C G G A T C C T A A T A C A C C A

VCF

2,058,122 U2,058,122 | 2,058,167 U2,058,167

C A A T C A C A G C A G C C C A A A T T A C A T C A C C A A C C G G A G A C C G G A T C C T A A T A C A C C A

C A A T C A C A G C A G C C C A A A T T A C A T T C A C C A A C C G G A T C C C T A A T A C A C C A

C A A T C A C A G C A G C C C A A A T T A C A T T C A C C A A C C G G A T C C C T A A T A C A C C A

C A A T C A C A G C A G C C C A A A T T A C A T T C A C C A A C C G G A T C C C T A A T A C A C C A

C A A T C A C A G C A G C C C A A A T T A C A T T C A C C A A C C G G A T C C C T A A T A C A C C A

C A A T C A C A G C A G C C C A A A T T A C A T T C A C C A A C C G G A T C C C T A A T A C A C C A

C A A T C A C A G C A G C C C A A A T T A C A T T C A C C A A C C G G A T C C C T A A T A C A C C A

C A A T C A C A G C A G C C C A A A T T A C A T T C A C C A A C C G G A T C C C T A A T A C A C C A

C A A T C A C A G C A G C C C A A A T T A C A T T C A C C A A C C G G A T C C C T A A T A C A C C A

Filter by: Name

Tablet Tip: Any CIGAR "I" (Insertion) tags in SAM/BAM files will be listed in the Features table

W23-2: Variants

out2.bam - Tablet - 1.16.09.06

chromosome | consensus length: 2,277,981 (0) | reads: 4,180 | features: 22 | Memory usage: 119.31 MB (11)

Home | Colour Schemes | Advanced

Open Assembly | Import Features | Import Enzymes

Read Packing | Tag Variants | Read Colours

Zoom: [Slider] | Variants: [Slider]

Adjust | Navigate | Overlays

Contigs: 3 | 509.1 k reads (more)

Co...	Len...	Re...	Fe...	Mis...
chr...	2,2...	47...	22	0.5
pla...	81,...	14,...	4	?
pla...	40,...	15,...	2	?

2,053,145 to 2,078,144 (25 Kb) | 2,058,122 to 2,058,168 (47 bp)

T I T A A Q I T P P T D P N T P

C A A T C A C A G C A G C C C A A A T T A C A C C A C C A A C G G A T C C T A A T A C A C C A

VCF

2,058,122 U2,058,122 | 2,058,167 U2,058,167

Filter by: Name

Tablet Tip: Position data is often supplemented with U (unpadded position) and CV (read coverage at that position) values

全体を眺めて下した方針。灰色で示した計5か所が概要配列に変更を加えるところ

W24-1: 方針のまとめ

#CHROM	POS	REF	ALT	方針	理由
chromosome	141598	A	G	Aのまま	Aが5個、Gが2個だったから
chromosome	242819	T	C	Tのまま	Tが7個、Cが2個だったから
chromosome	359141	T	C	Tのまま	Tが多数派だから
chromosome	463882	T	G	Tのまま	Tが2個、Gが1個だったから
chromosome	660031	C	T	Cのまま	Cが8個、Tが2個だったから
chromosome	792963	T	C	Tのまま	Tが多数派だから
chromosome	887422	G	A	Gのまま	Gが多数派だから
chromosome	913470	T	C	Tのまま	Tが多数派だから
chromosome	965171	T	C	Tのまま	Tが多数派だから
chromosome	1134950	A	G	Aのまま	Aが多数派だから
chromosome	1160048	A	G	Aのまま	Aが多数派だから
chromosome	1160095	A	G	Aのまま	Aが多数派だから
chromosome	1240248	GCCCC	GCCCCC	Cを挿入	そのほうが多数派だから
chromosome	1455552	CAAAA	CAAAAA	Aを挿入	そのほうが多数派だから
chromosome	1475721	T	C	Tのまま	Tが多数派だから
chromosome	1752680	G	A	Gのまま	Gが多数派だから
chromosome	1941380	T	C	Tのまま	Tが多数派だから
chromosome	2052955	A	G	Aのまま	Aが多数派だから
chromosome	2054084	TC	T	TCのまま	Cがあるほうが多数派だから
chromosome	2056279	A	G	Aのまま	Aが多数派だから
chromosome	2057564	T	C	Tのまま	Tが多数派だから
chromosome	2058145	C	T,G	Tに変更	Tが多数派だから
plasmid1	37177	G	T	Gのまま	Gが多数派だから
plasmid1	66609	A	G	Aのまま	Gのほうがちょっと多いが微妙
plasmid1	66637	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid1	66668	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid2	2569	CAAAAA	CAAAA	Aを欠失	そのほうが多数派だから
plasmid2	37805	TAAAAA	TAAAA	Aを欠失	そのほうが多数派だから

当たり前と言えば当たり前ですが、VCFファイルから読み取れる結論とほぼ同じになります(W18-4)

W24-1: VCFと一致

	A	B	C	D	E	F	G	H	I	J
27	#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	/home/w3pipeline/refdata/
28	chromosome	141598	.	A	G	17.1	.	DP=7;VDB=0.0077;AF1=0.5;AC1	GT:PL:GQ	0/1:47,0,138:50
29	chromosome	242819	.	T	C	20	.	DP=9;VDB=0.0111;AF1=0.5;AC1	GT:PL:GQ	0/1:50,0,139:53
30	chromosome	359141	.	T	C	4.13	.	DP=14;VDB=0.0037;AF1=0.4998	GT:PL:GQ	0/1:32,0,187:31
31	chromosome	463882	.	T	G	6.2	.	DP=13;VDB=0.0074;AF1=0.4999	GT:PL:GQ	0/1:35,0,216:36
32	chromosome	660031	.	C	T	8.64	.	DP=11;VDB=0.0063;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,197:40
33	chromosome	792963	.	T	C	4.13	.	DP=14;VDB=0.0153;AF1=0.4998	GT:PL:GQ	0/1:32,0,200:31
34	chromosome	887422	.	G	A	10.4	.	DP=11;VDB=0.0070;AF1=0.5;AC	GT:PL:GQ	0/1:40,0,188:42
35	chromosome	913470	.	T	C	14.2	.	DP=9;VDB=0.0056;AF1=0.5;AC1	GT:PL:GQ	0/1:44,0,144:47
36	chromosome	965171	.	T	C	17.1	.	DP=20;VDB=0.0071;AF1=0.5;AC	GT:PL:GQ	0/1:47,0,249:50
37	chromosome	1134950	.	A	G	7.8	.	DP=23;VDB=0.0076;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,255:39
38	chromosome	1160048	.	A	G	29	.	DP=17;VDB=0.0054;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
39	chromosome	1160095	.	A	G	19.1	.	DP=10;VDB=0.0033;AF1=0.5;AC	GT:PL:GQ	0/1:49,0,191:52
40	chromosome	1240248	.	GCCCC	GCCCCC	214	.	INDEL;DP=48;VDB=0.0137;AF1=	GT:PL:GQ	1/1:←55,128,0:99
41	chromosome	1455552	.	CAAAA	CAAAAA	214	.	INDEL;DP=58;VDB=0.0164;AF1=	GT:PL:GQ	1/1:←55,130,0:99
42	chromosome	1475721	.	T	C	52	.	DP=19;VDB=0.0021;AF1=0.5;AC	GT:PL:GQ	0/1:82,0,246:85
43	chromosome	1752680	.	G	A	29	.	DP=24;VDB=0.0129;AF1=0.5;AC	GT:PL:GQ	0/1:59,0,255:62
44	chromosome	1941380	.	T	C	7.8	.	DP=13;VDB=0.0046;AF1=0.5;AC	GT:PL:GQ	0/1:37,0,214:39
45	chromosome	2052955	.	A	G	12.3	.	DP=8;VDB=0.0084;AF1=0.5;AC1	GT:PL:GQ	0/1:42,0,103:44
46	chromosome	2054084	.	TC	T	13.7	.	INDEL;DP=8;VDB=0.0009;AF1=0	GT:PL:GQ	0/1:51,0,174:54
47	chromosome	2056279	.	A	G	24	.	DP=6;VDB=0.0057;AF1=0.5;AC1	GT:PL:GQ	0/1:54,0,113:57
48	chromosome	2057564	.	T	C	8.64	.	DP=10;VDB=0.0041;AF1=0.5;AC	GT:PL:GQ	0/1:38,0,129:40
49	chromosome	2058145	.	C	T,G	108	.	DP=14;VDB=0.0147;AF1=1;AC1=	GT:PL:GQ	1/1:←41,39,0,141,28,138:75
50	plasmid1	37177	.	G	T	54	.	DP=49;VDB=0.0087;AF1=0.5;AC	GT:PL:GQ	0/1:84,0,255:87
51	plasmid1	66609	.	A	G	225	.	DP=55;VDB=0.0060;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
52	plasmid1	66637	.	C	T	225	.	DP=71;VDB=0.0141;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,249:99
53	plasmid1	66668	.	C	T	225	.	DP=80;VDB=0.0153;AF1=0.5;AC	GT:PL:GQ	0/1:255,0,255:99
54	plasmid2	2569	.	CAAAA	CAAAA	177	.	INDEL;DP=110;VDB=0.0166;AF1	GT:PL:GQ	1/1:←48,255,0:99
55	plasmid2	37805	.	TAAAAA	TAAAA	104	.	INDEL;DP=31;VDB=0.0058;AF1=	GT:PL:GQ	1/1:←45,93,0:99

W26-1: 変異の反映

①変異反映後のファイルをLH_draft2.faにすべく、LH_draft.faからコピー。ここでは、**変異の反映を文字列置換で行う戦略**をとる。②オリジナルのファイルサイズ(2,400,619 bytes)は、大まかにはゲノムサイズ。③任意の10-merでは偶然に複数個所ヒットしそうだが、15-merなら一意に決まりそうだ

```
File Edit View Search Terminal Help
iu@bielinux[mac_share] pwd
/home/iu/Desktop/mac_share
iu@bielinux[mac_share] ls -l LH*
-rwxrwxrwx 1 iu iu 2400619  9月 10 16:52 LH_draft.fa
-rwxrwxrwx 1 iu iu 2433662  8月 30 21:22 LH_hgap.fa
iu@bielinux[mac_share] cp LH_draft.fa LH_draft2.fa
iu@bielinux[mac_share] R -q
> 4^10
[1] 1048576
> 4^12
[1] 16777216
> 4^15
[1] 1073741824
> q(save="no")
iu@bielinux[mac_share] █
```

[3:51午後]
[3:51午後]
[3:51午後]

W26-1: 2,058,145番目

①まずはchromosome上の2,058,145番目のCをTに置換すべく、上流12塩基分(つまり2058133-2058144)の塩基配列を抽出。配列ごとに位置を表す数値が大きいものから順番に行えば、位置のずれに悩まされることはない。だから一番最後からやろうとしている

#CHROM	POS	REF	ALT	方針	
chromosome	141598	A	G	Aのまま	Aが5個
chromosome	242819	T	C	Tのまま	Tが7個
chromosome	359141	T	C	Tのまま	Tが
chromosome	463882	T	G	Tのまま	Tが2個、Gが1個だったから
chromosome	660031	C	T	Cのまま	Cが8個、Tが2個だったから
chromosome	792963	T	C	Tのまま	Tが多数派だから
chromosome	887422	G	A	Gのまま	Gが多数派だから
chromosome	913470	T	C	Tのまま	Tが多数派だから
chromosome	965171	T	C	Tのまま	Tが多数派だから
chromosome	1134950	A	G	Aのまま	Aが多数派だから
chromosome	1160048	A	G	Aのまま	Aが多数派だから
chromosome	1160095	A	G	Aのまま	Aが多数派だから
chromosome	1240248	GCCCC	GCCCCC	Cを挿入	そのほうが多数派だから
chromosome	1455552	CAAAA	CAAAAA	Aを挿入	そのほうが多数派だから
chromosome	1475721	T	C	Tのまま	Tが多数派だから
chromosome	1752680	G	A	Gのまま	Gが多数派だから
chromosome	1941380	T	C	Tのまま	Tが多数派だから
chromosome	2052955	A	G	Aのまま	Aが多数派だから
chromosome	2054084	TC	T	TCのまま	Cがあるほうが多数派だから
chromosome	2056279	A	G	Aのまま	Aが多数派だから
chromosome	2057564	T	C	Tのまま	Tが多数派だから
chromosome	2058145	C	T,G	Tに変更	Tが多数派だから
plasmid1	37177	G	T	Gのまま	Gが多数派だから
plasmid1	66609	A	G	Aのまま	Gのほうがちょっと多いが微妙
plasmid1	66637	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid1	66668	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid2	2569	CAAAAA	CAAAA	Aを欠失	そのほうが多数派だから
plasmid2	37805	TAAAAA	TAAAA	Aを欠失	そのほうが多数派だから



W26-1:2,058,145番目

①chromosome上の(2058133-2058144)番目の塩基配列を抽出。②GCCCAAATTACAという配列が得られた。W11-2と同じテクニック。もしこの文字列で検索して1箇所にしかならなければ...

```
iu@bielinux[mac_share] pwd
/home/iu/Desktop/mac_share
iu@bielinux[mac_share] ls -l LH*
-rwxrwxrwx 1 iu iu 2400619  9月 10 16:52 LH_draft.fa
-rwxrwxrwx 1 iu iu 2433662  8月 30 21:22 LH_hgap.fa
iu@bielinux[mac_share] cp LH_draft.fa LH_draft2.fa
iu@bielinux[mac_share] R -q
> 4^10
[1] 1048576
> 4^12
[1] 16777216
> 4^15
[1] 1073741824
> q(save="no")
iu@bielinux[mac_share] head -2 LH_draft.fa | tail -1 | cut -c 2058133-2058144
GCCCAAATTACA
iu@bielinux[mac_share] █
```

[5:09午後]

[5:09午後]

[5:09午後]

[5:16午後]



W26-1: 2,058,145番目

①chromosome上の(2058133-2058144)番目の塩基配列を抽出。②GCCCAAATTACAという配列が得られた。W11-2と同じテクニック。もしこの文字列で検索して1箇所にしかならなければ…他に影響を及ぼすことなく③「置換前:GCCCAAATTACAC」、「置換後:GCCCAAATTACAT」という文字列置換によって変異の反映が可能(という戦略)

```

iu@bielinux[mac_share] pwd
/home/iu/Desktop/mac_share
iu@bielinux[mac_share] ls -l LH*
-rwxrwxrwx 1 iu iu 2400619  9月 10 16:52 LH_dr
-rwxrwxrwx 1 iu iu 2433662  8月 30 21:22 LH_hg
iu@bielinux[mac_share] cp LH_draft.fa LH_draftz.ta
iu@bielinux[mac_share] R -q
> 4^10
[1] 1048576
> 4^12
[1] 16777216
> 4^15
[1] 1073741824
> q(save="no")
iu@bielinux[mac_share] head -2 LH_draft.fa | tail -1 | cut -c 2058133-2058144
GCCCAAATTACA
iu@bielinux[mac_share]

```

chromosome	2058145	C	T,G	Tに変更	Tが多数派だから
plasmid1	37177	G	T	Gのまま	Gが多数派だから
plasmid1	66609	A	G	Aのまま	Gのほうがちょっと多いが微妙
plasmid1	66637	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid1	66668	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid2	2569	CAAAAA	CAAAA	Aを欠失	そのほうが多数派だから
plasmid2	37805	TAAAAA	TAAAA	Aを欠失	そのほうが多数派だから



W26-2: grep -o

①grepのデフォルトはGCCCAAATTACAという文字列と一致した行全体を返すため、一致箇所の個数は不明。-oオプションをつけることで、一致箇所分だけ一致した文字列(この場合はGCCCAAATTACA)がどんどん次の行に表示される。この場合は6箇所一致領域があることを意味するので、12塩基という長さでは危険だと判断する。よって、もっと長い塩基にする

```
File Edit View Search Terminal Help
-rwxrwxrwx 1 iu iu 2400619  9月 10 16:
-rwxrwxrwx 1 iu iu 2433662  8月 30 21:
iu@bielinux[mac_share] cp LH_draft.fa
iu@bielinux[mac_share] R -q
> 4^10
[1] 1048576
> 4^12
[1] 16777216
> 4^15
[1] 1073741824
> q(save="no")
iu@bielinux[mac_share] head -2 LH_draft.fa | tail -1 | cut -c 2058133-2058144
GCCCAAATTACA
iu@bielinux[mac_share] grep -o GCCCAAATTACA LH_draft.fa [ 5:16午後 ]
GCCCAAATTACA
GCCCAAATTACA
GCCCAAATTACA
GCCCAAATTACA
GCCCAAATTACA
GCCCAAATTACA
iu@bielinux[mac_share] [ 5:33午後 ]
```



①

W26-3: 20塩基で

①chromosome上の(2058125-2058144)番目の塩基配列を抽出。この場合でも6箇所一致領域があることを意味するので、20塩基という長さでもまだ危険だと判断

```
iu@bielinux[mac_share] pwd [ 3:13午後 ]
/home/iu/Desktop/mac_share
iu@bielinux[mac_share] ls -l LH* [ 3:13午後 ]
-rwxrwxrwx 1 iu iu 2400619  9月 22 17:09 LH_draft2.fa
-rwxrwxrwx 1 iu iu 2400619  9月 10 16:52 LH_draft.fa
-rwxrwxrwx 1 iu iu 2433662  8月 30 21:22 LH_hgap.fa
① iu@bielinux[mac_share] head -2 LH_draft.fa | tail -1 | cut -c 2058125-2058144
TCACAGCAGCCCAAATTACA
② iu@bielinux[mac_share] grep -o TCACAGCAGCCCAAATTACA LH_draft.fa
TCACAGCAGCCCAAATTACA
TCACAGCAGCCCAAATTACA
TCACAGCAGCCCAAATTACA
TCACAGCAGCCCAAATTACA
TCACAGCAGCCCAAATTACA
TCACAGCAGCCCAAATTACA
TCACAGCAGCCCAAATTACA
iu@bielinux[mac_share] [ 3:13午後 ]
```

W26-4: 50塩基で

①chromosome上の(2058095-2058144)番目の塩基配列を抽出。この場合でも6箇所一致領域があることを意味するので、50塩基という長さでもまだ危険だと判断

```
iu@bielinux[mac_share] pwd [ 3:42午後 ]
/home/iu/Desktop/mac_share
iu@bielinux[mac_share] ls -l LH* [ 3:42午後 ]
-rwxrwxrwx 1 iu iu 2400619  9月 22 17:09 LH_draft2.fa
-rwxrwxrwx 1 iu iu 2400619  9月 10 16:52 LH_draft.fa
-rwxrwxrwx 1 iu iu 2433662  8月 30 21:22 LH_hgap.fa
① iu@bielinux[mac_share] head -2 LH_draft.fa | tail -1 | cut -c 2058095-
2058144
AGGTAAGTGTTAAACCTGGTCAATTTACAATCACAGCAGCCCAAATTACA
② iu@bielinux[mac_share] grep -o AGGTAAGTGTTAAACCTGGTCAATTTACAATCACAGCAG
CCCAAATTACA LH_draft.fa
AGGTAAGTGTTAAACCTGGTCAATTTACAATCACAGCAGCCCAAATTACA
AGGTAAGTGTTAAACCTGGTCAATTTACAATCACAGCAGCCCAAATTACA
AGGTAAGTGTTAAACCTGGTCAATTTACAATCACAGCAGCCCAAATTACA
AGGTAAGTGTTAAACCTGGTCAATTTACAATCACAGCAGCCCAAATTACA
AGGTAAGTGTTAAACCTGGTCAATTTACAATCACAGCAGCCCAAATTACA
AGGTAAGTGTTAAACCTGGTCAATTTACAATCACAGCAGCCCAAATTACA
iu@bielinux[mac_share] [ 3:42午後 ]
```

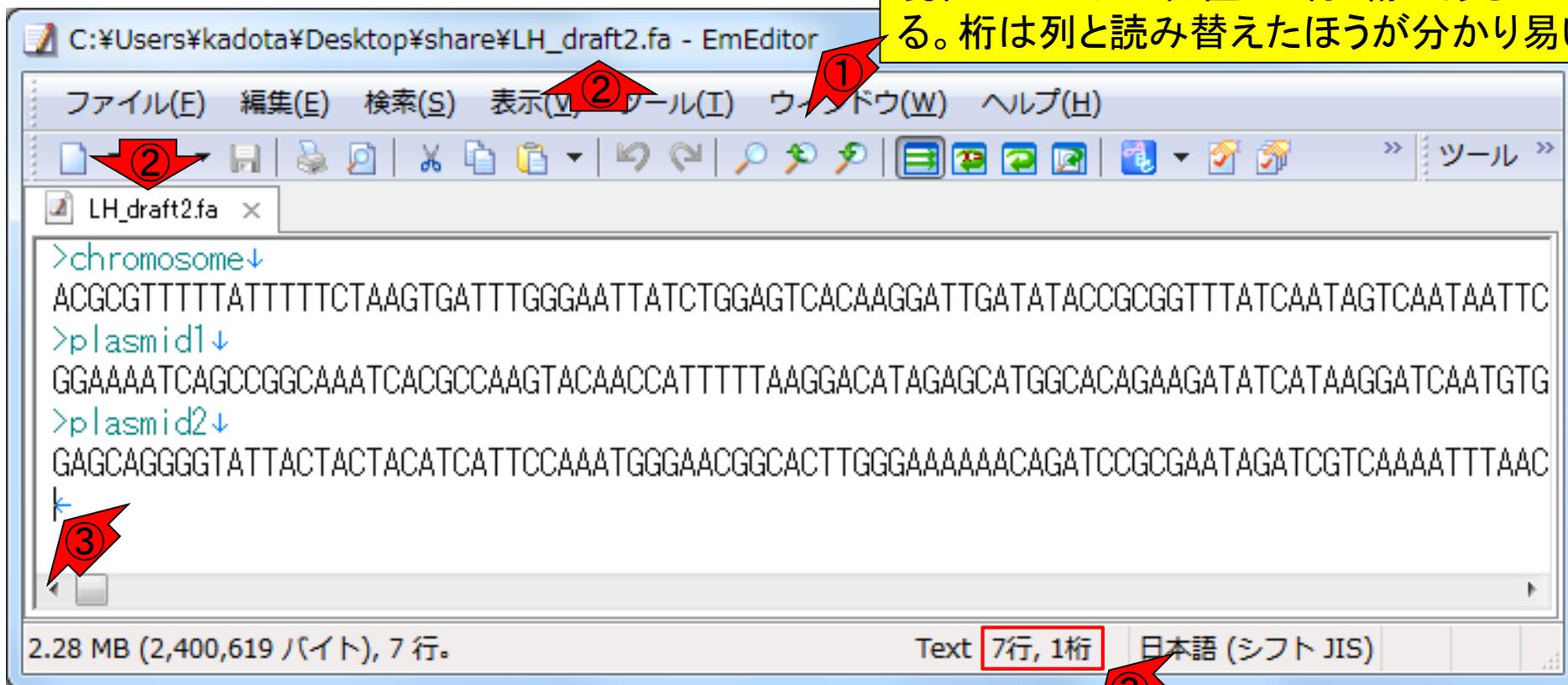
W26-5: 300塩基で

①chromosome上の(2057855-2058144)番目の塩基配列を抽出。200塩基ではだめだったが300塩基の長さでやっと一致領域が一か所のみとなったので、「置換前:…C」、「置換後:…T」という文字列置換によって変異の反映が可能ということになるが、これだけ長くなってしまうとエディタが置換できる最大値を超えているかもしれないのはっきりいってアホラシイ

```
iu@bielinux[mac_share] ls -l LH*
-rwxrwxrwx 1 iu iu 2400619  9月 22 17:
-rwxrwxrwx 1 iu iu 2400619  9月 10 16:
-rwxrwxrwx 1 iu iu 2433662  8月 30 21:
iu@bielinux[mac_share] head -2 LH_draft.fa | tail -1 | cut -c 2057855-2058144
ATGCTAGTGATATTCAAATGCATTGTTTGTGATTACGCCAGCGGCGGTTACAATTACTGCTAATGACAA
TGGTAAAGTGTACGGGACAACAGATCCAATGCTAACCGCACAGGTCAGCGGGTCCCAACAAACGGTGAT
GCGATTAGCGCCAGTGATTACGCAGTTACGCGTGCAGCTGGAGAAAATATTGGCAACTATGACGAAACAG
TTGTTTATACTGGCGGCAATGCCAATTACAAGGTAAGTGTTAAACCTGGTCAATTTACAATCACAGCAGC
CCAAATTACA
iu@bielinux[mac_share] grep -o ATGCTAGTGATATTCAAATGCATTGTTTGTGATTACGC
CAGCGGCGGTTACAATTACTGCTAATGACAATGGTAAAGTGTACGGGACAACAGATCCAATGCTAACCGC
ACAGGTCAGCGGGTCCCAACAAACGGTGATGCGATTAGCGCCAGTGATTACGCAGTTACGCGTGCAGCT
GGAGAAAATATTGGCAACTATGACGAAACAGTTGTTTATACTGGCGGCAATGCCAATTACAAGGTAAGTG
TTAAACCTGGTCAATTTACAATCACAGCAGCCCAAATTACA LH_draft.fa
ATGCTAGTGATATTCAAATGCATTGTTTGTGATTACGCCAGCGGCGGTTACAATTACTGCTAATGACAA
TGGTAAAGTGTACGGGACAACAGATCCAATGCTAACCGCACAGGTCAGCGGGTCCCAACAAACGGTGAT
GCGATTAGCGCCAGTGATTACGCAGTTACGCGTGCAGCTGGAGAAAATATTGGCAACTATGACGAAACAG
TTGTTTATACTGGCGGCAATGCCAATTACAAGGTAAGTGTTAAACCTGGTCAATTTACAATCACAGCAGC
CCAAATTACA
iu@bielinux[mac_share] █ [ 3:57午後 ]
```

W27-1 : EmEditor

①EmEditor (Windows用のみ)で、②変異箇所を変更する予定のLH_draft2.faを開いたところ。③現在のカーソル位置が7行1桁であることがわかる。桁は列と読み替えたほうが分かり易いかも…



W27-2: EmEditor

①W26-2で6箇所的一致領域があることが分かっているGCCCAAATTACAという文字列で検索する。②を何回か押し、目的の「chromosome上の2,058,145番目」あたりになっていることを③で確認する

The screenshot shows the EmEditor interface with a search dialog box open. The search dialog has the following settings:

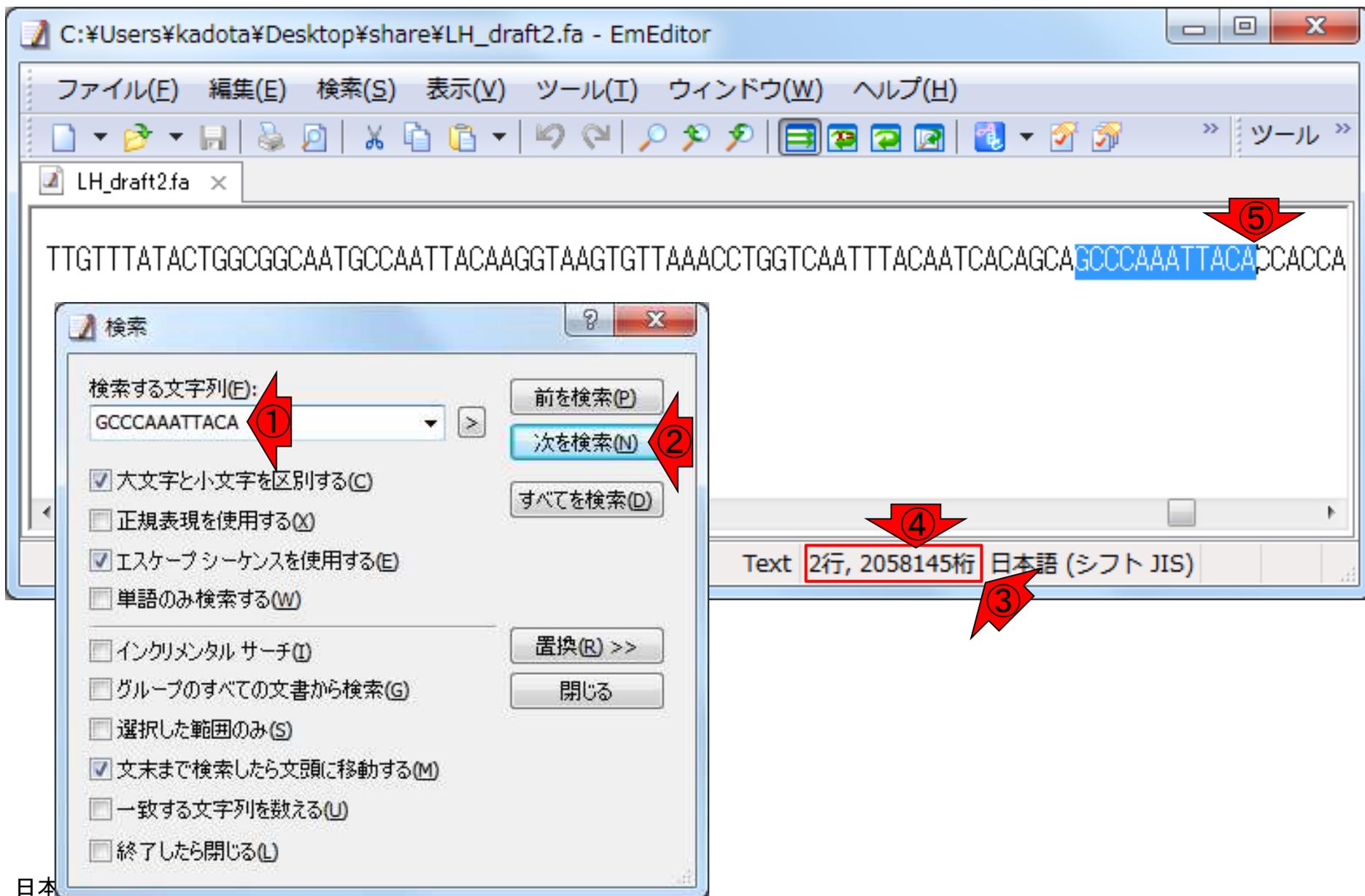
- 検索する文字列(E): GCCCAAATTACA (indicated by red arrow ①)
- 前を検索(P) (indicated by red arrow ②)
- 次を検索(N) (indicated by red arrow ②)
- すべてを検索(D)
- 大文字と小文字を区別する(C)
- 正規表現を使用する(X)
- エスケープシーケンスを使用する(E)
- 単語のみ検索する(W)
- インクリメンタルサーチ(I)
- グループのすべての文書から検索(G)
- 選択した範囲のみ(S)
- 文末まで検索したら文頭に移動する(M)
- 一致する文字列を数える(U)
- 終了したら閉じる(L)

The main text area shows the following DNA sequence: TTGTTTATACTGGCGGCAATGCCAAATTACAAGGTAAGTGTTAAACCTGGTCAATTTACAATCACAGCA GCCCAAATTACA CCACCA. The sequence GCCCAAATTACA is highlighted in blue.

The status bar at the bottom shows: Text 2行, 2058145桁 日本語 (シフト JIS). The text "2行, 2058145桁" is highlighted in red and indicated by red arrow ③.

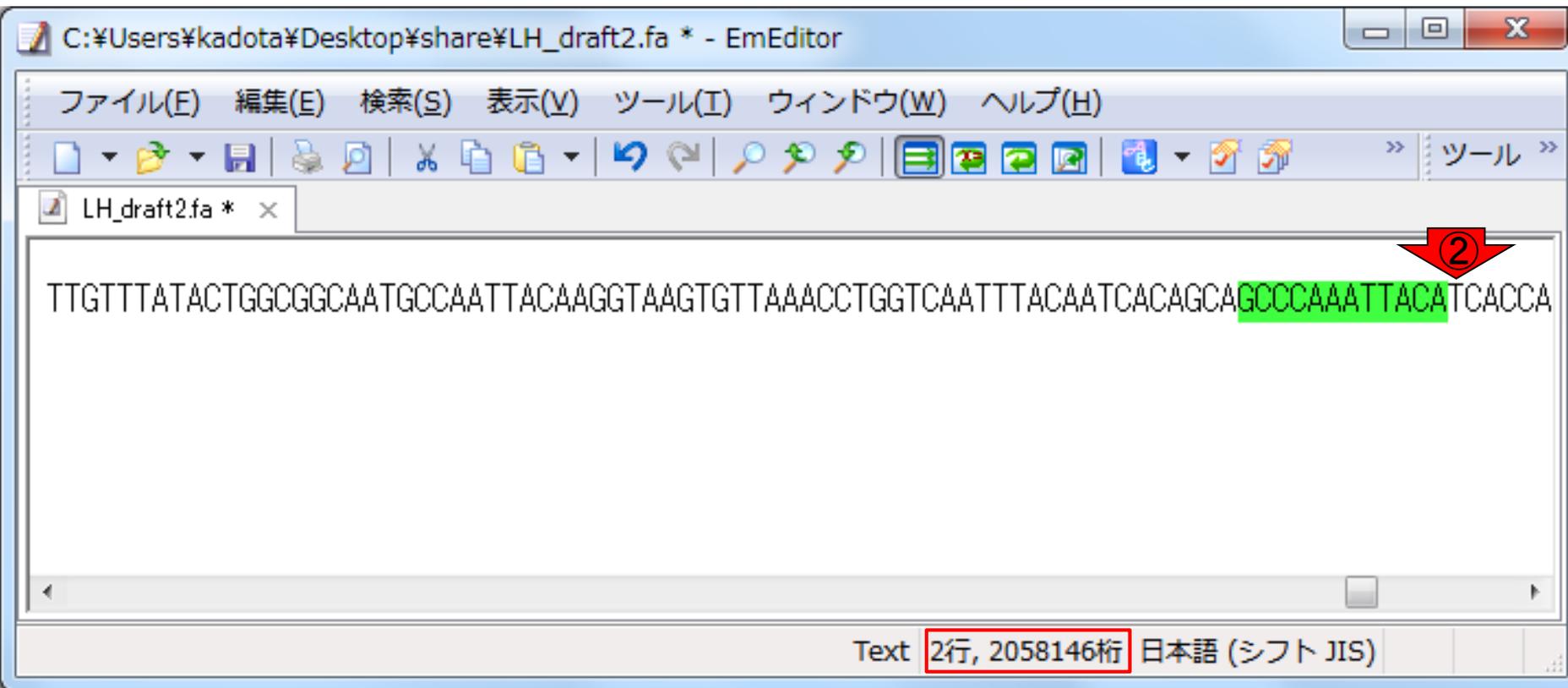
W27-2: EmEditor

④そのものズバリの位置になっているが、⑤の位置にカーソルがあるので妥当



W27-2: EmEditor

①chromosome上の2,058,145番目の、②CをTに変更したところ



chromosome	2056279	A	G	Aのまま	Aが多数派だから
chromosome	2057564	T	C	Tのまま	Tが多数派だから
chromosome	2058145	C	T,G	Tに変更	Tが多数派だから
plasmid1	37177	G	T	Gのまま	Gが多数派だから
plasmid1	66609	A	G	Aのまま	Gのほうがちょっと多いが微妙
plasmid1	66637	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid1	66668	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid2	2569	CAAAAA	CAAAA	Aを欠失	そのほうが多数派だから
plasmid2	37805	TAAAAA	TAAAA	Aを欠失	そのほうが多数派だから



次は欠失の反映例として①plasmid2
上の37,805番目のところを示す

W27-3:37,805番目

#CHROM	POS	REF	ALT	方針	理由
chromosome	141598	A	G	Aのまま	Aが5個、Gが2個だったから
chromosome	242819	T	C	Tのまま	Tが7個、Cが2個だったから
chromosome	359141	T	C	Tのまま	Tが多数派だから
chromosome	463882	T	G	Tのまま	Tが2個、Gが1個だったから
chromosome	660031	C	T	Cのまま	Cが8個、Tが2個だったから
chromosome	792963	T	C	Tのまま	Tが多数派だから
chromosome	887422	G	A	Gのまま	Gが多数派だから
chromosome	913470	T	C	Tのまま	Tが多数派だから
chromosome	965171	T	C	Tのまま	Tが多数派だから
chromosome	1134950	A	G	Aのまま	Aが多数派だから
chromosome	1160048	A	G	Aのまま	Aが多数派だから
chromosome	1160095	A	G	Aのまま	Aが多数派だから
chromosome	1240248	GCCCC	GCCCCC	Cを挿入	そのほうが多数派だから
chromosome	1455552	CAAAA	CAAAAA	Aを挿入	そのほうが多数派だから
chromosome	1475721	T	C	Tのまま	Tが多数派だから
chromosome	1752680	G	A	Gのまま	Gが多数派だから
chromosome	1941380	T	C	Tのまま	Tが多数派だから
chromosome	2052955	A	G	Aのまま	Aが多数派だから
chromosome	2054084	TC	T	TCのまま	Cがあるほうが多数派だから
chromosome	2056279	A	G	Aのまま	Aが多数派だから
chromosome	2057564	T	C	Tのまま	Tが多数派だから
chromosome	2058145	C	T,G	Tに変更	Tが多数派だから
plasmid1	37177	G	T	Gのまま	Gが多数派だから
plasmid1	66609	A	G	Aのまま	Gのほうがちょっと多いが微妙
plasmid1	66637	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid1	66668	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid2	2569	CAAAAA	CAAAA	Aを欠失	そのほうが多数派だから
plasmid2	37805	TAAAAA	TAAAA	Aを欠失	そのほうが多数派だから



W27-3:37,805番

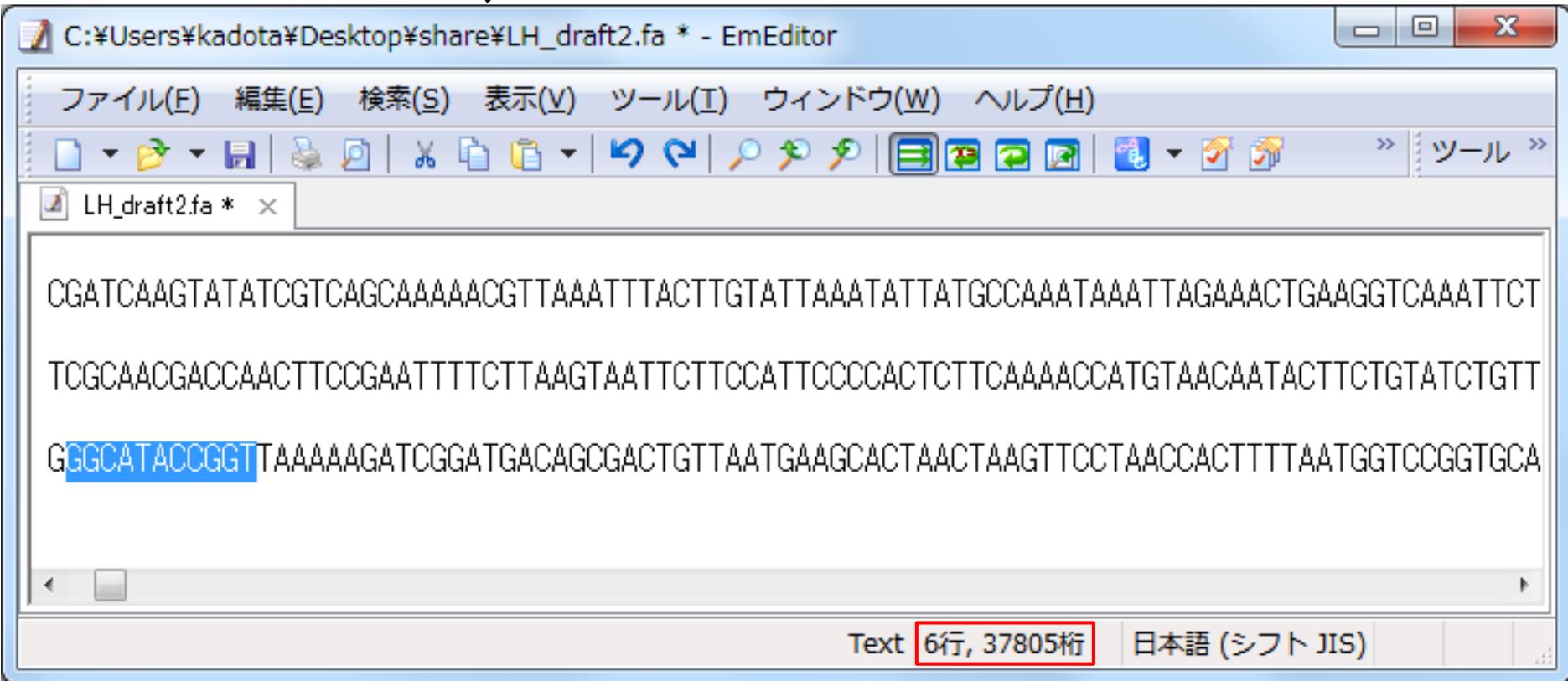
① plasmid2上の37,805番目を合理的に検索すべく、② plasmid2上の(FASTA形式ファイルで3番目の配列なので6)、③37794-37804番目の11塩基分(GGCATACCGGT)を表示させ、④3箇所一致することを把握

```
iu@bielinux[~/Desktop/mac_share]
iu@bielinux[mac_share] pwd [ 6:49午後 ]
/home/iu/Desktop/mac_share
iu@bielinux[mac_share] ls -l LH* [ 6:49午後 ]
-rwxrwxrwx 1 iu iu 2400619 9月 22 17:09 LH_draft2.fa
-rwxrwxrwx 1 iu iu 2400619 9月 10 16:52 LH_draft.fa
-rwxrwxrwx 1 iu iu 2433662 9月 30 21:22 LH_hgap.fa
iu@bielinux[mac_share] head -6 LH_draft.fa | tail -1 | cut -c 37794-37804
GGCATACCGGT
iu@bielinux[mac_share] grep -o GGCATACCGGT LH_draft.fa | wc -l
3
iu@bielinux[mac_share] [ 6:50午後 ]
```

plasmid1	66668	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid2	2569	CAAAAA	CAAAA	Aを欠失	そのほうが多数派だから
plasmid2	37805	TAAAAA	TAAAA	Aを欠失	そのほうが多数派だから



W27-3:37,805番目

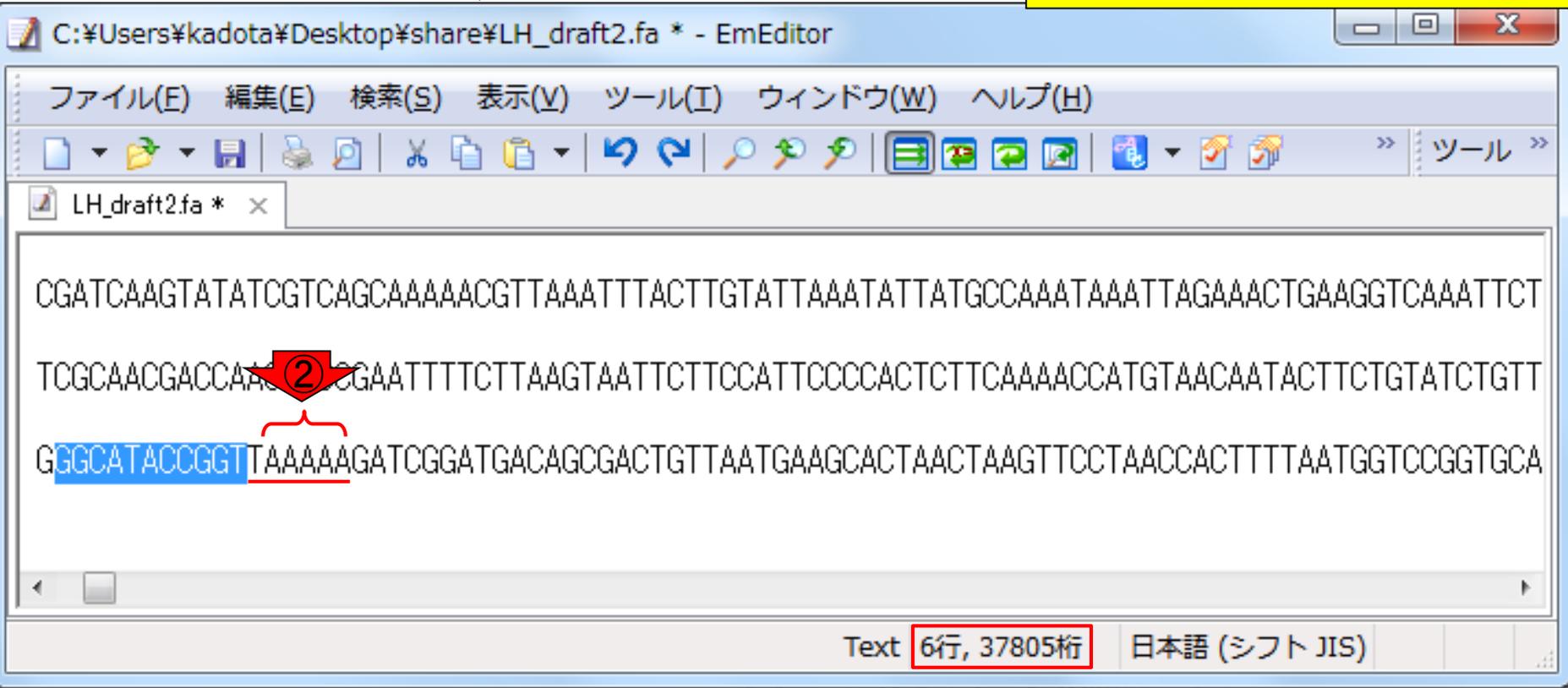


chromosome	2056279	A	G	Aのまま	Aが多数派だから
chromosome	2057564	T	C	Tのまま	Tが多数派だから
chromosome	2058145	C	T,G	Tに変更	Tが多数派だから
plasmid1	37177	G	T	Gのまま	Gが多数派だから
plasmid1	66609	A	G	Aのまま	Gのほうがちょっと多いが微妙
plasmid1	66637	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid1	66668	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid2	2569	CAAAAA	CAAAA	Aを欠失	そのほうが多数派だから
plasmid2	37805	TAAAAA	TAAAA	Aを欠失	そのほうが多数派だから



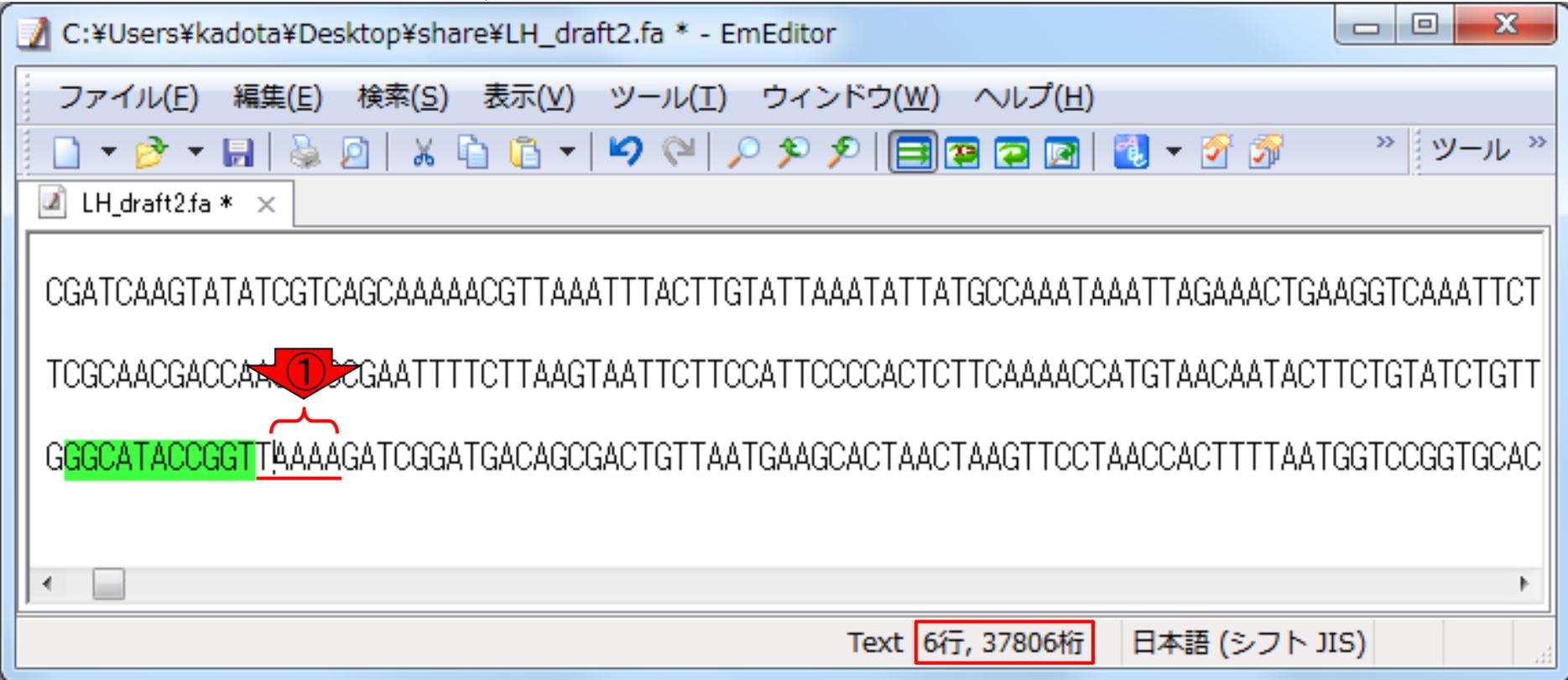
W27-3:37,805番目

実際に変更するのは①37,805番目のTではなく、②37,806-37,810番目の中のいずれか1つのAである点に注意！



chromosome	2056279	A	G	Aのまま	Aが多数派だから
chromosome	2057564	T	C	Tのまま	Tが多数派だから
chromosome	2058145	C	T,G	Tに変更	Tが多数派だから
plasmid1	37177	G	T	Gのまま	Gが多数派だから
plasmid1	66609	A	G	Aのまま	Gのほうがちょっと多いが微妙
plasmid1	66637	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid1	66668	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid2	2569	CAAAAA	CAAAA	Aを欠失	そのほうが多数派だから
plasmid2	37805	TAAAAA	TAAAA	Aを欠失	そのほうが多数派だから

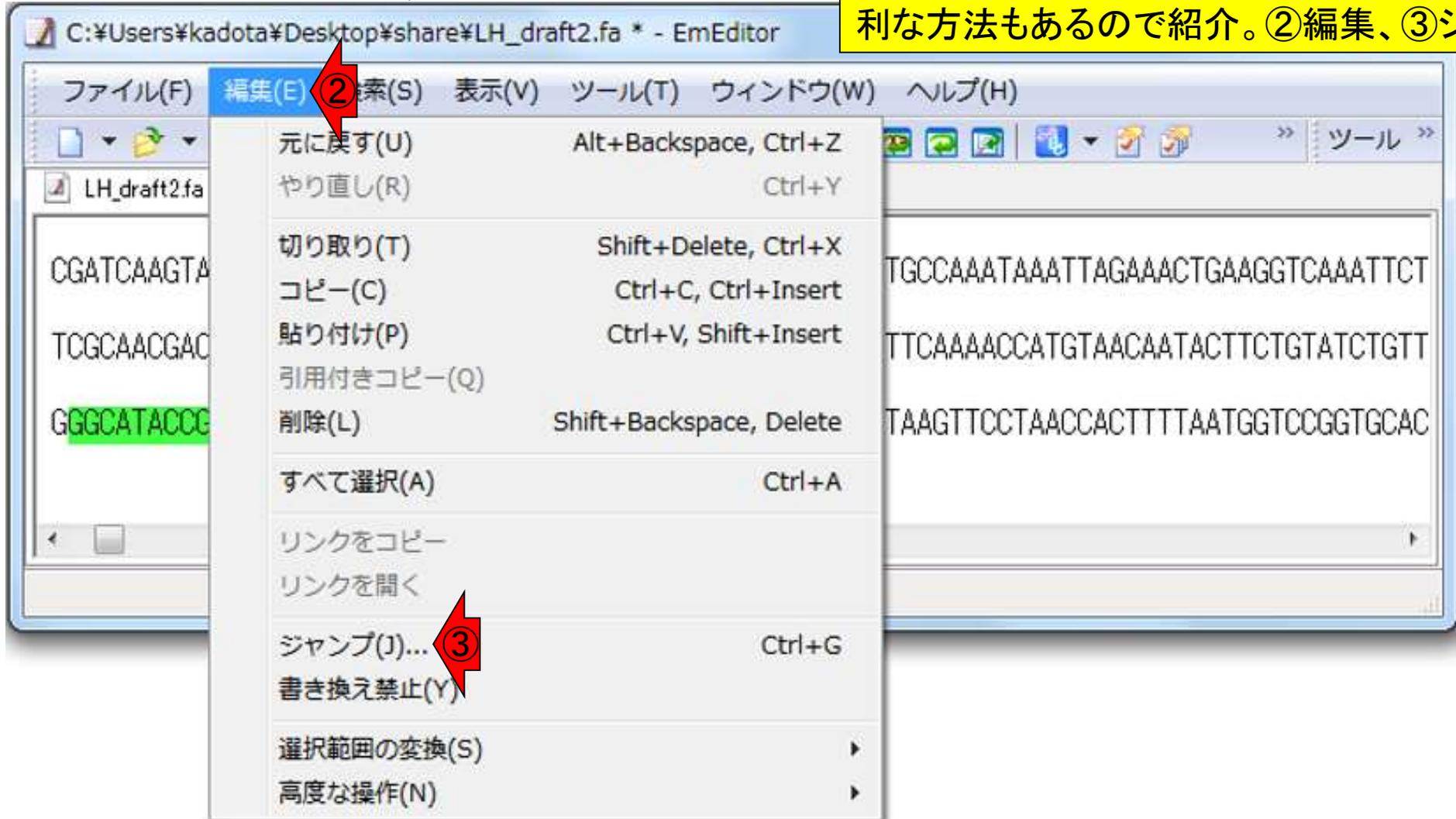
W27-3:37,805番目



chromosome	2056279	A	G	Aのまま	Aが多数派だから
chromosome	2057564	T	C	Tのまま	Tが多数派だから
chromosome	2058145	C	T,G	Tに変更	Tが多数派だから
plasmid1	37177	G	T	Gのまま	Gが多数派だから
plasmid1	66609	A	G	Aのまま	Gのほうがちょっと多いが微妙
plasmid1	66637	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid1	66668	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid2	2569	CAAAAA	CAAAA	Aを欠失	そのほうが多数派だから
plasmid2	37805	TAAAAA	TAAAA	Aを欠失	そのほうが多数派だから

W27-4: 2,569番目

①次は2,569番目。配列数(コンティグ数)が多い場合には意味をなさないが、今回のような3配列(3 contigs)しかない場合には、非常に便利な方法もあるので紹介。②編集、③ジャンプ

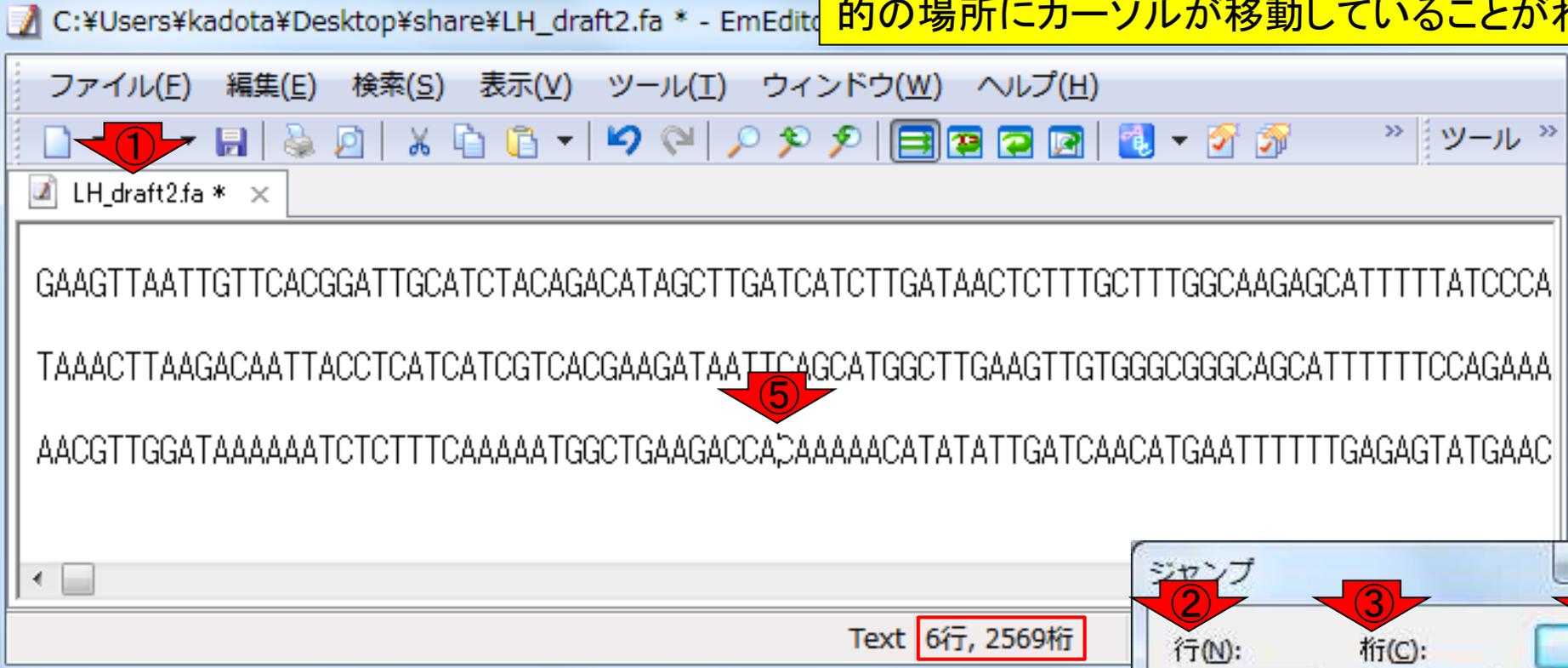


plasmid1	6668	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid2	2569	CAAAAA	CAAAA	Aを欠失	そのほうが多数派だから
plasmid2	37805	TAAAAA	TAAAA	Aを欠失	そのほうが多数派だから



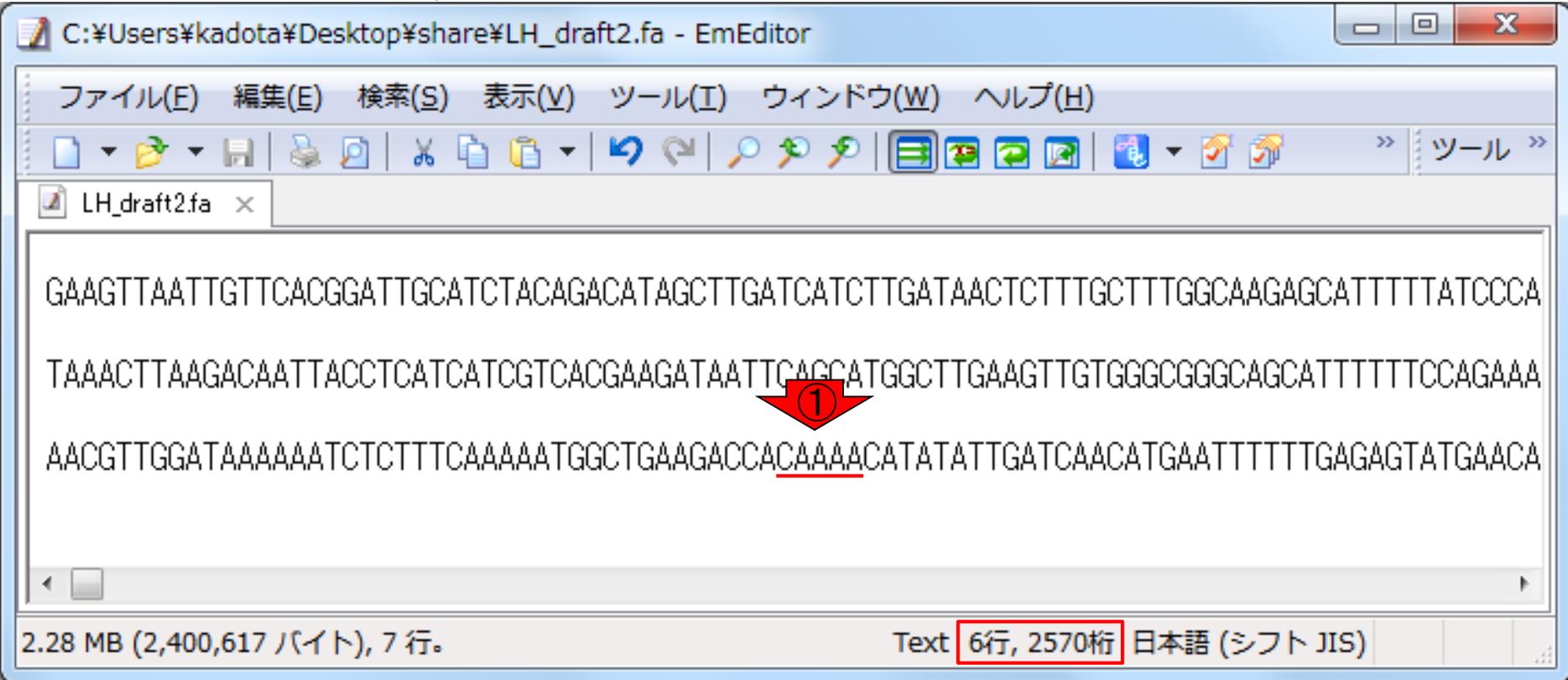
W27-4:2,569番目

今編集しているのは、①multi-FASTA形式のLH_draft2.fa。これの3番目の配列がplasmid2なので、②6行、③2569桁目に④ジャンプすればよい。⑤目的の場所にカーソルが移動していることがわかる



chromosome	2056279	A	G	Aのまま	Aが多数派だから
chromosome	2057564	T	C	Tのまま	Tが多数派だから
chromosome	2058145	C	T,G	Tに変更	Tが多数派だから
plasmid1	37177	G	T	Gのまま	Gが多数派だから
plasmid1	66609	A	G	Aのまま	Gのほうがちょっと多いが微妙
plasmid1	66637	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid1	66668	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid2	2569	CAAAAA	CAAAA	Aを欠失	そのほうが多数派だから
plasmid2	37805	TAAAAA	TAAAA	Aを欠失	そのほうが多数派だから

W27-4:2,569番目



chromosome	2056279	A	G	Aのまま	Aが多数派だから
chromosome	2057564	T	C	Tのまま	Tが多数派だから
chromosome	2058145	C	T,G	Tに変更	Tが多数派だから
plasmid1	37177	G	T	Gのまま	Gが多数派だから
plasmid1	66609	A	G	Aのまま	Gのほうがちょっと多いが微妙
plasmid1	66637	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid1	66668	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid2	2569	CAAAAA	<u>CAAAA</u>	Aを欠失	そのほうが多数派だから
plasmid2	37805	TAAAAA	TAAAA	Aを欠失	そのほうが多数派だから

W27-5: 残りの2箇所も

同じノリで①、②の順番で数値の大きいほうから反映させておく。位置を表す数値が大きいものから順番に行えば、位置のずれに悩まされることはないため

#CHROM	POS	REF	ALT	方針	理
chromosome	141598	A	G	Aのまま	Aが5個、Gが2個だったから
chromosome	242819	T	C	Tのまま	Tが7個、Cが2個だったから
chromosome	359141	T	C	Tのまま	Tが多数派だから
chromosome	463882	T	G	Tのまま	Tが2個、Gが1個だったから
chromosome	660031	C	T	Cのまま	Cが8個、Tが2個だったから
chromosome	792963	T	C	Tのまま	Tが多数派だから
chromosome	887422	G	A	Gのまま	Gが多数派だから
chromosome	913470	T	C	Tのまま	Tが多数派だから
chromosome	965171	T	C	Tのまま	Tが多数派だから
chromosome	1134950	A	G	Aのまま	Aが多数派だから
chromosome	1160048	A	G	Aのまま	Aが多数派だから
chromosome	1160095	A	G	Aのまま	Aが多数派だから
chromosome	② 1240248	GCCCC	GCCCCC	Cを挿入	そのほうが多数派だから
chromosome	1455552	CAAAA	CAAAAA	Aを挿入	そのほうが多数派だから
chromosome	1475721	T	C	Tのまま	Tが多数派だから
chromosome	1752680	G	A	Gのまま	Gが多数派だから
chromosome	1941380	T	C	Tのまま	Tが多数派だから
chromosome	2052955	A	G	Aのまま	Aが多数派だから
chromosome	2054084	TC	T	TCのまま	Cがあるほうが多数派だから
chromosome	2056279	A	G	Aのまま	Aが多数派だから
chromosome	2057564	T	C	Tのまま	Tが多数派だから
chromosome	2058145	C	T,G	Tに変更	Tが多数派だから
plasmid1	37177	G	T	Gのまま	Gが多数派だから
plasmid1	66609	A	G	Aのまま	Gのほうがちょっと多いが微妙
plasmid1	66637	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid1	66668	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid2	2569	CAAAAA	CAAAA	Aを欠失	そのほうが多数派だから
plasmid2	37805	TAAAAA	TAAAA	Aを欠失	そのほうが多数派だから



W27-5: 残りの2箇所も

同じノリで①、②の順番で数値の大きいほうから反映させておく。位置を表す数値が大きいものから順番に行えば、位置のずれに悩まされることはないため

#CHROM	POS	REF	ALT	方針	理
chromosome	141598	A	G	Aのまま	Aが5個、Gが2個だったから
chromosome	242819	T	C	Tのまま	Tが7個、Cが2個だったから
				Tのまま	Tが多数派だから
				Tのまま	Tが2個、Gが1個だったから
				Cのまま	Cが8個、Tが2個だったから
				Tのまま	Tが多数派だから
				Gのまま	Gが多数派だから
				Tのまま	Tが多数派だから
				Tのまま	Tが多数派だから
				Aのまま	Aが多数派だから
				Aのまま	Aが多数派だから
chromosome	1160048	A	G	Aのまま	Aが多数派だから
chromosome	1160095	A	G	Aのまま	Aが多数派だから
chromosome	② 1240248	GCCCC	GCCCCC	Cを挿入	そのほうが多数派だから
chromosome	1455552	CAAAA	CAAAAA	Aを挿入	そのほうが多数派だから
chromosome	1475721	T	C	Tのまま	Tが多数派だから
chromosome	1752680	G	A	Gのまま	Gが多数派だから
chromosome	1941380	T	C	Tのまま	Tが多数派だから
chromosome	2052955	A	G	Aのまま	Aが多数派だから
chromosome	2054084	TC	T	TCのまま	Cがあるほうが多数派だから
chromosome	2056279	A	G	Aのまま	Aが多数派だから
chromosome	2057564	T	C	Tのまま	Tが多数派だから
chromosome	2058145	C	T,G	Tに変更	Tが多数派だから
plasmid1	37177	G	T	Gのまま	Gが多数派だから
plasmid1	66609	A	G	Aのまま	Gのほうがちょっと多い
plasmid1	66637	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid1	66668	C	T	Cのまま	Tのほうがちょっと多いが微妙
plasmid2	2569	CAAAA	CAAAA	Aを欠失	そのほうが多数派だから
plasmid2	37805	TAAAA	TAAAA	Aを欠失	そのほうが多数派だから

ジャンプ

行(N): 2 桁(C): 1240248

論理座標(L)

OK キャンセル

ジャンプ

行(N): 2 桁(C): 1455552

論理座標(L)

OK キャンセル