

ライフサイエンスデータベース統合推進事業
統合化推進プログラム(統合データ解析トライアル)
研究開発課題
「HLA 遺伝子完全配列決定パイプラインの構築」

研究開発終了報告書

研究開発期間:平成26年9月～平成27年2月
研究代表者:細道一善
(国立遺伝学研究所 総合遺伝研究系人類遺
伝研究部門 外来研究員)



§1 研究開発の概要

HLA(ヒト主要組織適合抗原)はヒトの免疫応答の入り口としての役割を果たす重要な分子であり、リウマチや糖尿病など 100 種近くの生活習慣病、自己免疫疾患、がん、造血幹細胞移植に伴うGVH病、薬剤副作用との関連など数多くの医学的興味を有する。将来的に HLA 遺伝子のタイピング法は NGS による手法に進むことが予測されるが、本申請では HLA 遺伝子配列決定法ならびにタイピング法を開発することを目的とした。研究開発の成果として、目的とした一連の解析パイプラインを開発したが、これに加え、構築したパイプラインを p-galaxy に組み込み、ウェブブラウザ上で解析可能なユーザーが使いやすいシステムとして公開した。

§2 研究開発のねらい

ヒトの主要組織適合遺伝子複合体 (Human leukocyte antigen; HLA) 領域はヒト第6染色体短腕部 6p21.31 に位置し、免疫応答をつかさどる HLA 遺伝子を含む 3.8 Mb から構成される。ヒトゲノムの中でも群を抜いて多型性に富むゲノム領域であり、HLA 遺伝子 6 座 (HLA-A, -B, -C, -DR, -DQ, -DP) に計 12,000 種類もの膨大な HLA アレルがこれまでに同定されている。また各遺伝子座における特定の HLA 対立遺伝子が強い連鎖不平衡により、HLA-A24-B52-DR15 など特有の HLA ハプロタイプを形成している。さらにはリウマチや糖尿病など 100 種近くの生活習慣病、自己免疫疾患、がん、造血幹細胞移植に伴う GVH 病、ウイルス感染症における防御と重症化など数多くの医学的興味を有し、さらに HLA 分子による抗原提示を利用したワクチンによる感染防御の有用性とがんの免疫療法の新しい展開が期待されている。特に、スティーブンスン・ジョンソン症候群、中毒性表皮壊死融解症および薬剤性過敏症症候群などの薬剤副作用と HLA アレルはオッズ比が 1,000 を超えるもの、陽性率 100% のものが多々あるなど、極めて強い関連が複数報告されており、HLA アレルが薬剤副作用に大きく寄与していることが報告されている。

現在の HLA タイピングは蛍光ビーズ法による PCR-Sequence specific oligonucleotide (PCR-SSO) およびダイレクトシーケンスによる PCR-Sequencing based typing (PCR-SBT) により行われており、これらの技術は 10 年以上の間置き換わっていない。PCR-SSO は高頻度アレルのみを同定する手法であり、低頻度のものを同定できない。また、塩基配列を直接決定する PCR-SBT は HLA-A, B および C それぞれの遺伝子座で 8,000、18,000 および 5,000 以上の HLA アレルの組み合わせにおいて phase ambiguity に起因する曖昧なタイピング結果が得られるという問題点がある。さらにこの SBT 法は HLA クラス I 遺伝子ではエクソン 2 および 3、クラス II 遺伝子ではエクソン 2 のみで判定されており、上述の曖昧なタイピング結果の問題とあわせて解決すべき問題である。造血幹細胞移植における拒絶反応や重症の移植片対宿主病をなくすことや HLA タイピングを薬剤副作用の診断指針にまで実用するには、将来的に HLA タイピングを“みなしタイピング”を行わない、より高解像度、高精度なものへ改良する必要がある。

NGS による HLA 遺伝子配列決定法ならびにタイピング法の解析パイプラインを構築することが本申請の目的であり、下記の項目にて研究を実施する。

(1) haplotype phasing プログラムの開発

HLA 遺伝子におけるアレル配列決定のための phasing 手法を確立

(2) 解析手法のパイプライン化

haplotype phasing プログラムを含めた HLA 遺伝子配列決定法をパイプライン化

本研究申請は NGS による HLA 遺伝子配列決定法ならびにタイピング法の解析パイプラインを構築することを目的とし、基本となるリソースとしてヒトゲノムバリエーションデータベースの HLA database に登録されている HLA ハプロタイプシーケンス(apd_hap1: 4,622,290 bp、

cox_hap2: 4,795,371 bp、dbb_hap3: 4,610,396 bp、mann_hap4: 4,683,263 bp、mcf_hap5: 4,833,398 bp および qbl_hap6: 4,611,984 bp)を用いる。さらに構築した HLA 遺伝子配列決定法ならびにタイピング法の解析パイプラインを用いて、DRA に登録されているデータ(DRA001289)も追加データとして解析し、HLA 遺伝子配列を決定する。このとき、HLA ハプロタイプ推定の為の必要なデータ量をシミュレーションし、データ量としての推奨値を評価する。これら決定した HLA 遺伝子配列から多様性を精査すると共に、決定した配列を HLA database から参照可能とすることも目指す。また、結果の検証のために市販されている HLA タイピングソフトウェア Omixon との結果の整合性を評価する。Omixon は既知の HLA アレルデータベースに対して相同性を検索しタイピングするアルゴリズムであり、エクソンのみを対象とするため、本申請のデザインとは全く異なるが、タイピング結果を比較することで本申請の評価に用いる。

NGS による HLA 遺伝子配列決定法ならびにタイピング法の解析パイプラインを構築することで、今後研究者が NGS による HLA 遺伝子配列を決定できるようになると共に、より多くのヒトゲノムバリエーションデータベースの HLA database のデータの充実に寄与することが期待される。

§3 研究開発計画

(1) 当初の研究開発計画

本研究開発は haplotype phasing プログラムの作成と解析手法のパイプライン化の 2 つの研究開発項目で構成され、平成 27 年 2 月末までの 6 ヶ月間で完成させるスケジュールで進めた。haplotype phasing プログラムの作成は研究開発時に既に着手しており、平成 26 年内に完成させ、解析手法のパイプライン化は既存のツールで構成される一連の解析手法に新規プログラムを組み込むことで構築し、平成 27 年 2 月末までに完成させる計画で進めた。

(2) 新たに追加・修正など変更した研究開発計画

本研究開発で構築した解析パイプラインの公開を計画に追加した。ユーザーの使いやすさを考慮し、p-galaxy に組み込み、ウェブブラウザ上で解析可能なシステムとして公開した。

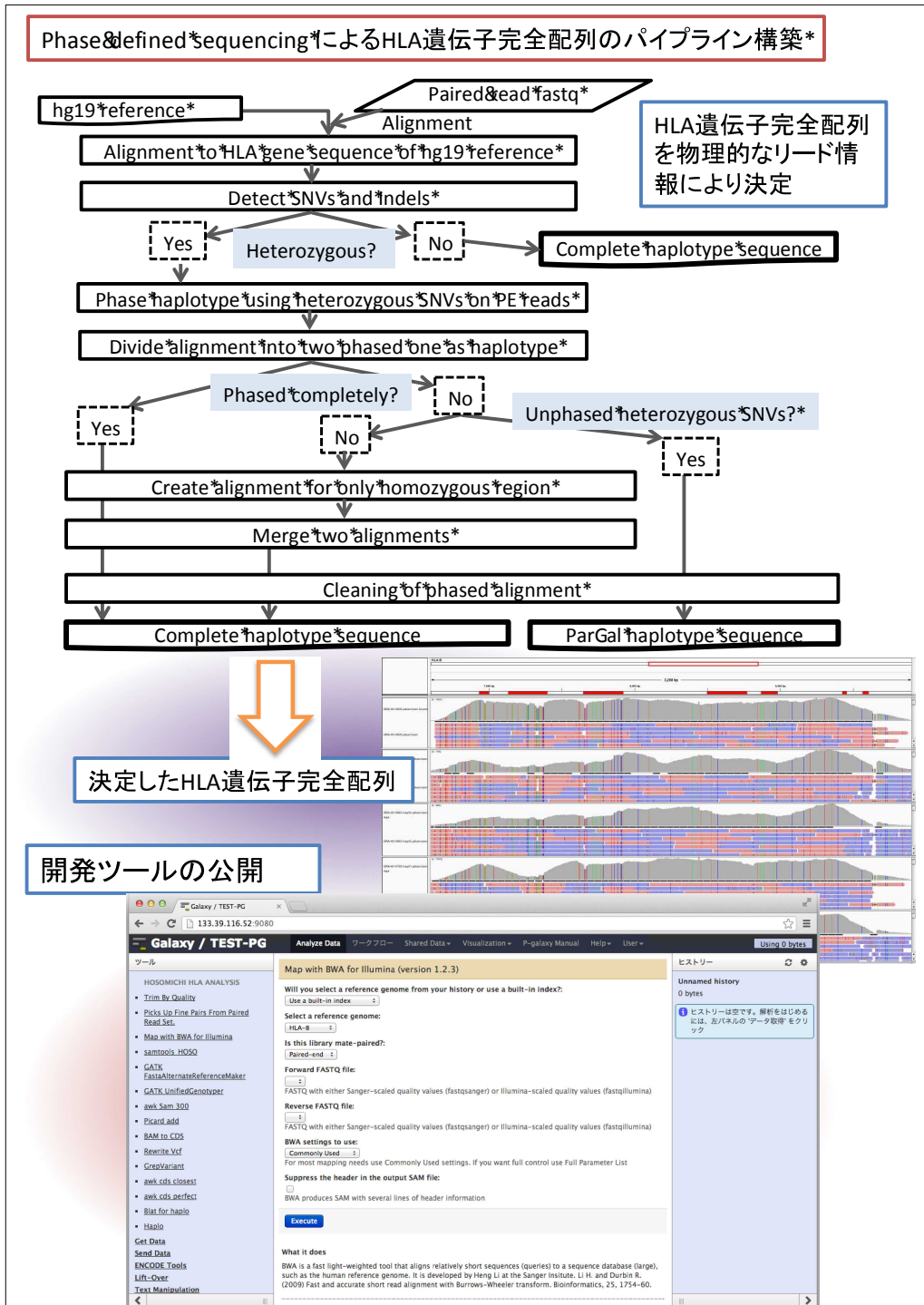
§4 研究開発成果

研究開発のパイプラインの内容は以下のとおりである。まず得られたシーケンスリードを各 PCR 増幅領域のリファレンス配列に bwa および samtools によってマッピングし、ディプロタイプとしての解析結果を得た後、GATK によりリファレンス配列と異なる塩基を検出する。次に検出された一塩基置換(SNV)および挿入欠失(Indel)のうち、SNV のみをシーケンスリードのペア情報を元に相を分類し、この相を分けられたハプロタイプの SNV を元に 2 つの HLA 遺伝子配列を作成する。このとき、1 つのペアエンドのシーケンスリードによって 2 つ以上の SNV 情報を物理的な情報として結合し、相の特定が可能となる。この相の分類をするプログラムを開発したこと、このプログラムを含む一連の解析をパイプラインとしたことが本研究開発の成果である。また、得られた結果を蛍光ビーズ法および市販されている HLA タイピングソフトウェア Omixon の結果と比較したところ、全く矛盾は見られなかった。本研究開発のみが HLA 遺伝子の完全長配列を決定可能であった。作成したプログラムおよびパイプラインは DDBJ の p-galaxy(<https://p-galaxy.ddbj.nig.ac.jp>)のツールの一つとして公開している。ウェブベースのグラフィカルインターフェイスを持つ p-galaxy 上で解析を実行することで、必要項目を画面上で選

択し、実行するのみでバイオインフォマティクスの専門知識を必要とせず誰でも解析結果を得ることが可能となっている。

本成果は統合化推進プログラムで統合化されたデータベースの活用という意味合いは弱いですが、今後、データベースを充実させていくための有用なツールである。

研究開発成果の概要



§5 研究開発計画に対する達成状況と将来展望

(1) 達成状況

研究開発計画は達成された。これに加えてツールとして公開する計画を加え、p-galaxy に組み込み、ウェブブラウザ上で解析可能なシステムとして公開した。

(2) ツール等の将来展望

研究開発のさらなる展開として期待されるのは臨床での診断応用である。2013年9月イルミナ社の次世代シーケンサーMiSeqDx システムが米国 FDA(食品医薬品局)の体外診断用医療機器(IVD)の認可を受け、臨床診断に使用可能となった。このように、次世代シーケンサーが医療機器として認可を受けた意義は極めて大きい。これは臨床シーケンス時代がすでに始まったことを意味し、個々人のゲノム情報を基にした医療(個の医療)が始まっているといえる。HLA はヒトの免疫応答の入り口としての役割を果たす重要な分子であり、HLA タイピングは極めて臨床的意義が高い。HLA が適合しないものはすべて異物と認識して攻撃を始めてしまうことから、造血幹細胞移植や臓器移植では HLA の適合性が重要視される。また、がんワクチン療法に用いるガンに特異的なペプチド・ワクチン開発は特定の HLA を基に開発されている。加えて HLA は 100 以上の疾患ならびに薬剤副作用と関連することから、HLA タイピングが必要となる分野は多岐に渡り、予防医学上の重要性から高精度のタイピングが期待される遺伝子の一つである。とりわけ薬剤の副作用と HLA は強い関連を示し、薬剤投与時の副作用のリスク回避のために HLA タイピングをすることは予防医学的に重要であるとも言える。一例として抗痙攣剤 Carbamazepine は重篤な副作用 Stevens-Johnson 症候群をまれに発症するが、副作用発症者のすべてが HLA-B*15:02 を有し、そのオッズ比は 895.5 と極めて高い。ゲノム情報に基づく個の医療実現において、本研究開発は臨床応用へ展望も期待できる。

§6 研究参加者

氏名	所属	役職	研究開発項目	参加時期
細道一善	国立遺伝学研究所	外来研究員	HLA 遺伝子完全配列決定パイプラインの構築	H26.9-H27.2

§7 成果発表等

(1) 原著論文発表 (国内(和文)誌 0 件、国際(欧文)誌 0 件)

(2) その他の著作物(総説、書籍など)

なし

(3) 国際学会発表及び主要な国内学会発表

① 招待講演 (国内会議 2 件、国際会議 0 件)

1. 細道一善(国立遺伝学研究所総合遺伝研究系人類遺伝研究部門)、臨床シーケンスを目指した HLA 領域のゲノム解析、第20回 NRGIC重点セミナー、長崎大学医学部キャンパス(長崎)、2014年9月12日

2. 細道一善(国立遺伝学研究所総合遺伝研究系人類遺伝研究部門)、臨床シーケンス応用を目指した HLA 領域のゲノム解析、Roche Scientific Symposium 2014 先端医療研究におけるターゲットゲノム解析、JPタワー ホール&カンファレンス(東京)、2014年11月5日

- ② 口頭発表 (国内会議 2件、国際会議 0件)
1. 細道一善 1)、椎名隆 2)、光永滋樹 2)、猪子英俊 2)、井ノ上逸朗 1) (国立遺伝学研究所総合遺伝研究系人類遺伝研究部門、2)東海大学医学部基礎医学系分子生命科学)、HLA-omics : HLA 領域におけるゲノム多様性、メチル化および遺伝子発現の統合的解析、第23回日本組織適合性学会大会、長崎大学医学部キャンパス(長崎)、2014年9月13日
 2. 細道一善 1)、椎名隆 2)、光永滋樹 2)、猪子英俊 2)、井ノ上逸朗 1) (国立遺伝学研究所総合遺伝研究系人類遺伝研究部門、2)東海大学医学部基礎医学系分子生命科学)、HLA-omics : HLA 領域におけるゲノム多様性、メチル化および遺伝子発現の統合的解析、日本人類遺伝学会第59回大会、タワーホール船堀(東京)、2014年11月22日
- ③ ポスター発表 (国内会議 0件、国際会議 1件)
1. Kazuyoshi Hosomichi, Ituro Inoue (Division of Human Genetics, National Institute of Genetics), PCR-based method for complete HLA gene sequencing and capture-based method for entire HLA region sequencing, ASHG 2014, San Diego Convention Center (San Diego, California), October. 18-22, 2014

(4)知財出願

- ①国内出願 (0件)
- ②海外出願 (0件)
- ③その他の知的財産権
なし

(5)受賞・報道等

なし

§8 自己評価

本研究開発期間で得られた研究成果は研究開発のねらい、に対して基盤を確立したと評価している。本研究開発で検証したのは 12,000 を超える HLA アレルの一部のみであり、今後は新規 HLA アレルも含めてファインチューニングしていく必要があるからである。

▼ [home](#)
[howto](#)
[Sitemap](#)

[home](#) > [howto](#) >

HOSOMICHI HLA ANALYSIS TOOLS

Contents

- [1 Introduction](#)
- [2 Preparing User Account in Common with DDBJ Pipeline](#)
- [3 Data Transfer](#)
- [4 Setting Up Workflows](#)
- [5 Running Workflow "HLAwfSeparate1"](#)
- [6 Running Workflow "HLAwfSeparate2"](#)
- [7 NOTICE ABOUT AN ERROR](#)
- [8 Sample FASTQ Files for Trial HOSOMICHI HLA ANALYSIS TOOLS](#)
- [9 K-0328_S66_L001_R1_001.fastq\(forward\) and K-0328_S66_L001_R2_001.fastq\(reverse\)](#)

Introduction

In this section, we would like to introduce p-Galaxy workflows to perform ported "HLA ANALYSIS TOOLS" developed by Dr. Hosomichi of Division of Human Genetics, National Institute of Genetics of Japan.

Human Leukocyte Antigen (HLA) is a group of genes that are extremely polymorphic among individuals and populations. The HLA has been associated with more than 100 different diseases, mostly autoimmune diseases. The workflows are capable to perform the HLA typing for medical research, and population genetics.

References:

Hosomichi *et al.* Phase-defined complete sequencing of the HLA genes by next-generation sequencing. BMC Genomics 2013, 14:355.

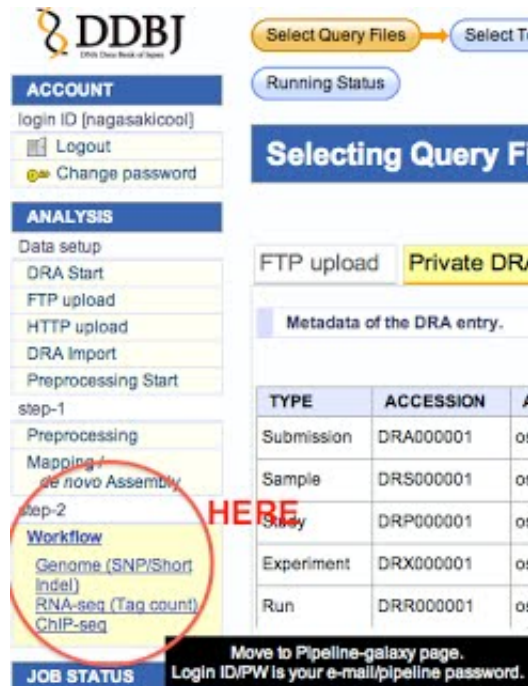
Hosomichi *et al.* A simple and rapid 96-well plate-based complete sequencing method of the HLA-B (Submitted).

Nagasaki *et al.* DDBJ Read Annotation Pipeline: A Cloud Computing-Based Pipeline for High-Throughput Analysis of Next-Generation Sequencing Data. DNA Res. 2013;20(4):383-90.

Preparing User Account in Common with DDBJ Pipeline

Originally, the Galaxy requires your mail address and password. P-Galaxy also requires mail address, which is submitted with user account of [DDBJ Pipeline](#). Therefore if you do not have the pipeline account, you should have the account to use P-Galaxy. To have the account, please read [described part in the manual](#).

When you have user account of the DDBJ Pipeline, start up P-Galaxy from [P-Galaxy website](#) OR from the link ("step-2/Workflow/Genome...") in the left side menu of basic analysis part of DDBJ Pipeline.



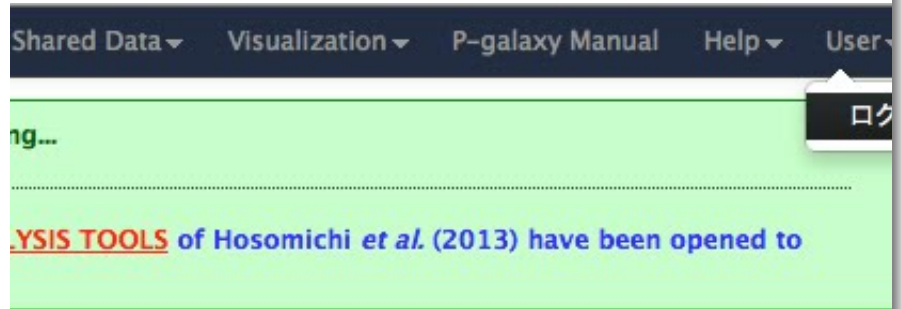
When if you can see [P-Galaxy top page...](#)



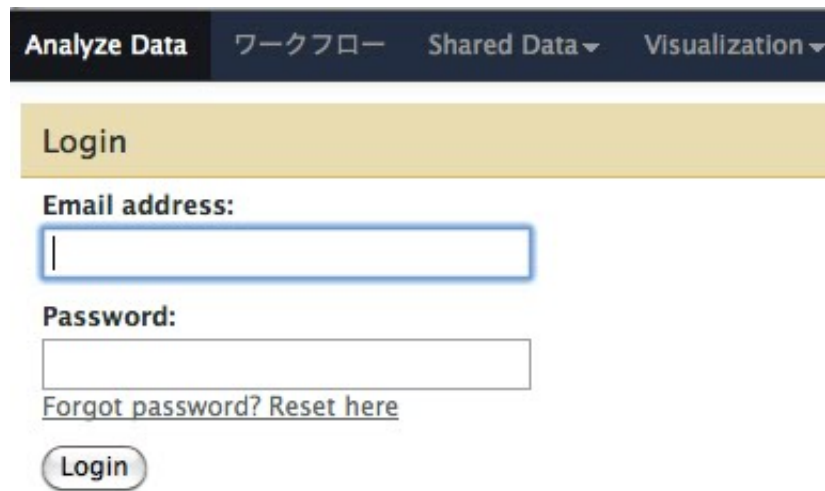
You should check you are already login or not from "top

menu/User".

If you are not login, set e-mail address and password (submitted with the pipeline account).



Login window is appeared.

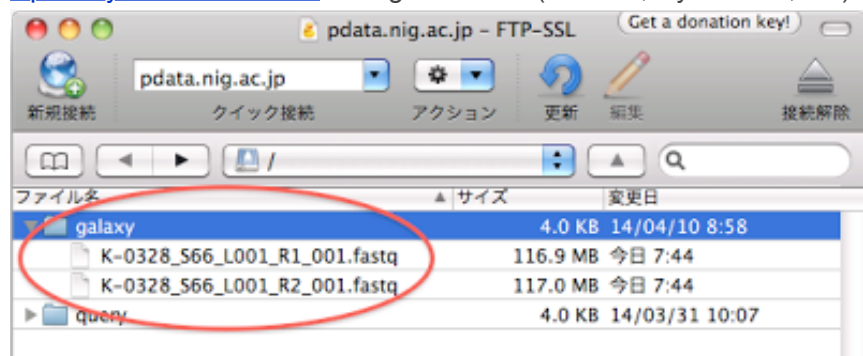


Data Transfer

In this section, how to transfer FASTQ files to HLA typing is described.

Mapping of genotype is performed using BWA software, thus FASTQ files should be included reads by illumina.

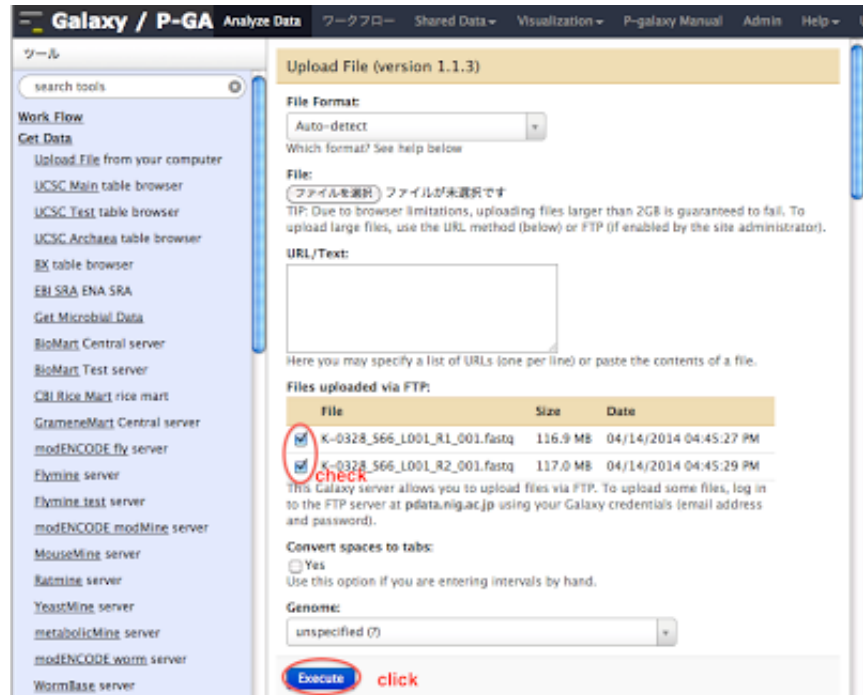
The step of data transfer to "galaxy" directory in the ftp server (pdata.nig.ac.jp) is mostly same as the illustration of section "[5-2-2 Upload your files via FTP](#)" using FTP client(FileZilla, Cyberduck, etc).



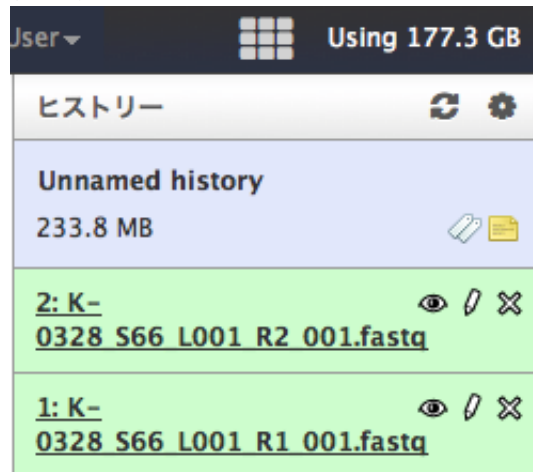
Look to p-galaxy.ddbj.nig.ac.jp, you can find "Get Data/Upload File from your computer" in left side tools of P-Galaxy window.

If the data has been uploaded, you can see the data name will be appeared in center frame of P-Galaxy.

Press "Execute" button to transfer the data to P-Galaxy.

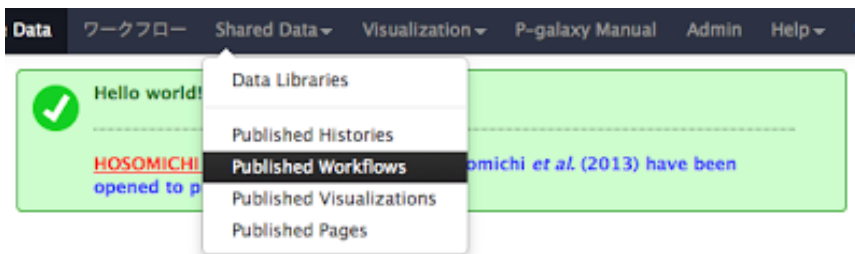


The loaded data will be appeared in "history" area in right side of P-galaxy window.



Setting Up Workflows

Choose the menu of "Published workflows" of "Shared Data" menu in centrally above of the window.



WWFSMD?

The "Published Workflows" flame is opened.

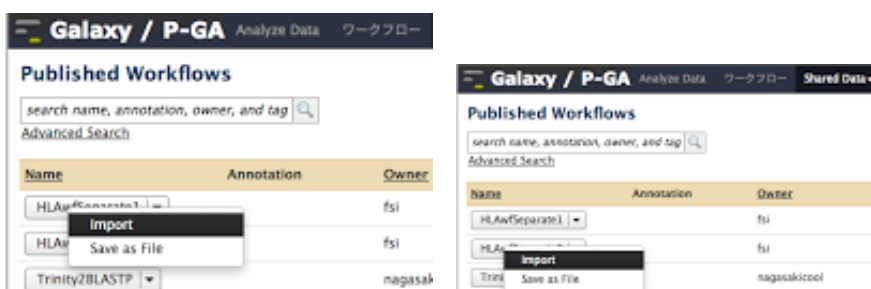


Choose "import" in the menu of "HLAwfSeparate1" and "HLAwfSeparate2".

The workflow "HLAwfSeparate1" performs from trimming reads to building HLA sequences exchanged polymorphisms between mapped reads and HLA genome.

The second,

"HLAwfSeparate2" makes the typing lists from polymorphisms between HLA sequences HLAwfSeparate1 and HLA genome.



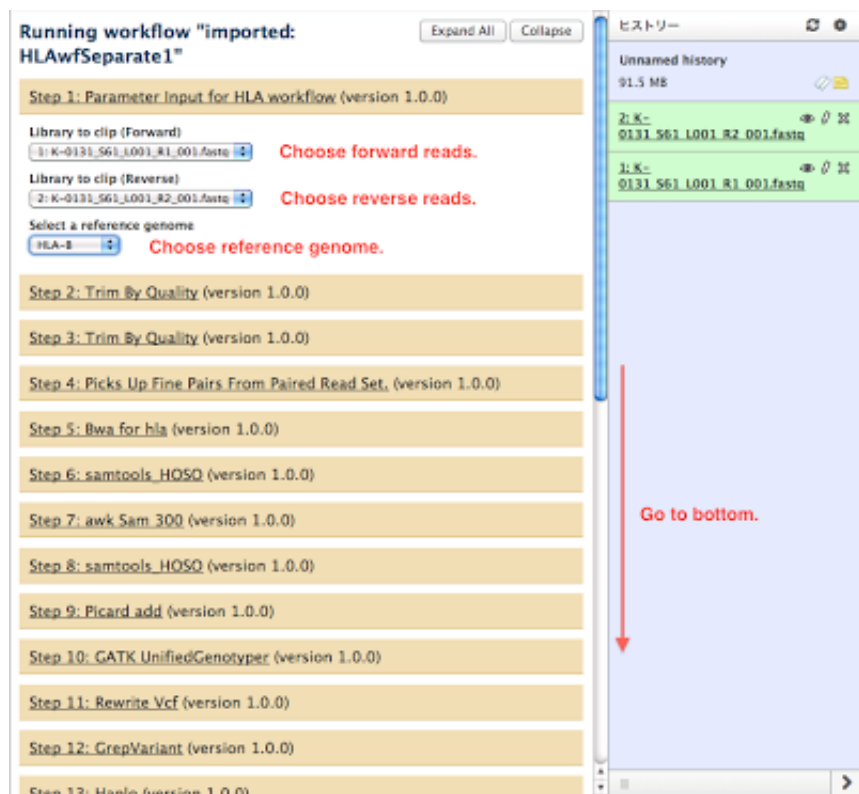
Running Workflow "HLAwfSeparate1"

After import HLAworkflows, click "workflow/ワークフロー" in the center of the upside menu, imported workflows will be appear.

Choose "Run" from "imported: HLAwfSeparate1" menu.



The procedure of the workflow is appeared in the Galaxy window, set the forward/reverse reads and choose HLA (A/B/C/DPB1/DQB1/DRB1) genome. At the end of the procedure, a button of "Run workflow".



Step 24: GATK FastaAlternateReferenceMaker (version 1.0.0)

Step 25: GATK FastaAlternateReferenceMaker (version 1.0.0)

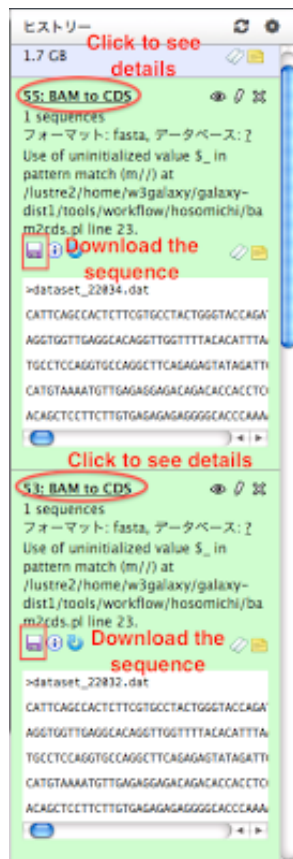
Step 26: BAM to CDS (version 1.0.0)

Step 27: BAM to CDS (version 1.0.0)

Send results to a new history

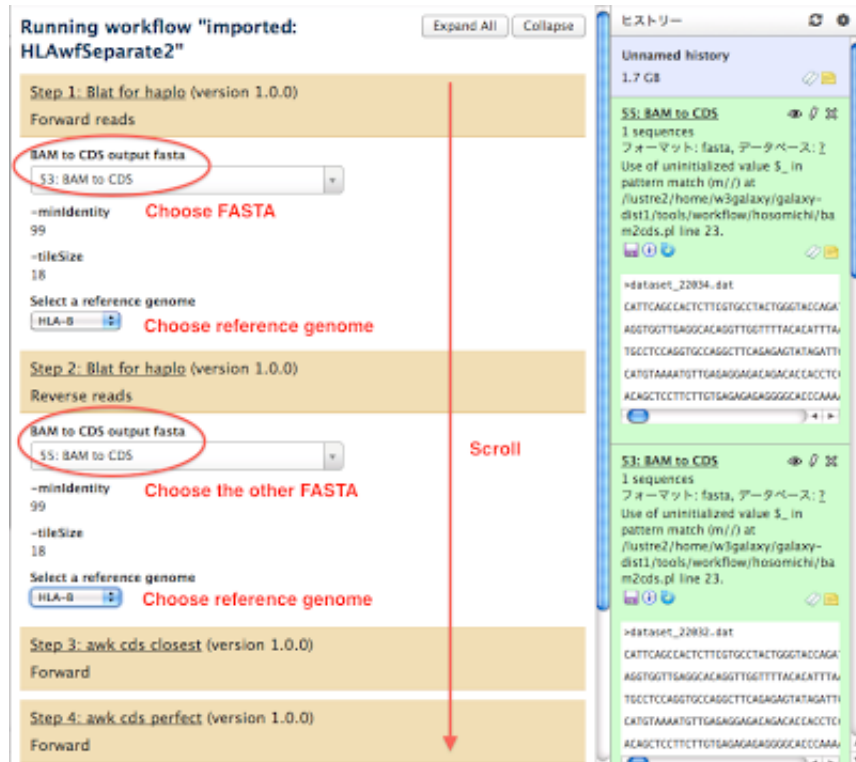
Run workflow Click

HLA sequences will be generated in History flame in the left side of P-Galaxy window.



Running Workflow "HLAwfSeparate2"

Following "HLAwfSeparate1", "HLAwfSeparate2" makes typing results from respective HLA sequence.



The screenshot displays a Galaxy workflow interface. On the left, the workflow steps are listed:

- Step 2: `Blat for haplo` (version 1.0.0) - Reverse reads
- Step 3: `awk cds_closest` (version 1.0.0) - Forward
- Step 4: `awk cds_perfect` (version 1.0.0) - Forward
- Step 5: `awk cds_closest` (version 1.0.0) - Reverse
- Step 6: `awk cds_perfect` (version 1.0.0) - Reverse

Parameters for the workflow include `-minIdentity` (99) and `-tileSize` (18). The reference genome is set to `HLA-B`. A `BAM to CDS` tool is selected. At the bottom, a `Run workflow` button is circled in red, with a `Click` label next to it.

On the right, the history panel shows two instances of the `BAM to CDS` tool. The top instance (ID 55) shows an error message: "Use of uninitialized value \$_ in pattern match (m//) at /lustre2/home/w3galaxy/galaxy-dist1/htdocs/workflow/hosomichi/bam2cds.pl line 23." Below the error is a snippet of DNA sequence from `>dataset_22034.dat`.

NOTICE ABOUT AN ERROR

Some FASTQ files make an error in the step of the finding haplotypes of HLAwfSeparate1 procedure. The cause of this error is that HLA sequences are split by few reads mapping to the reference HLA genome. The mapping methods is BWA MEM, which allows loose mapping, but some read set make that (please check matching to the reference). Recently we are planning to decrease the occurrence of this error.



Sample FASTQ Files for Trial HOSOMICHI HLA ANALYSIS TOOLS

K-0328_S66_L001_R1_001.fastq (forward) and K-0328_S66_L001_R2_001.fastq (reverse)

[K-0328_S66_L001_R1_001.fastq.gz \(30M\)](#)

[K-0328_S66_L001_R2_001.fastq.gz \(30M\)](#)

Choose **HLA-B** as reference genome with above FASTQ data.

コメント

コメントを追加する権限がありません。