

次世代シーケンサーデータの解析手法
第14回RNA-seq解析(その2)
ウェブ資料

寺田朋子、清水謙多郎、門田 幸二*

Contents

- W1: 公共データベースENA
- W2: クラウド解析環境Galaxy
- W3: ENAからGalaxyへのFASTQファイルの送付
- W4: クオリティチェック (FastQC)
- W5: クオリティコントロール (Trimmomatic)
- W6: クオリティチェック (FastQC)
- W7: ゲノム配列へのマッピング (Bowtie2)
- W8: カウント情報取得 (htseq-count)
- W9: カウント情報の連結 (Column join)
- W10: カウントデータの数値行列の行名について
- W11: 公共カウントデータ (GSE107337) を用いて、倍率変化と \log_2 (倍率変化)を算出
- W12: Galaxy上で得られたカウントデータについても同様に計算し、比較
- W13とW14: 原著論文中の記載事項を確認
- W15: 2つのカウントデータから共通遺伝子の情報を用いて類似度を計算

W1: ENA

①ENA上で、②GSE107337を、③検索。
第13回のW18と同じです。

The screenshot shows the ENA website interface. At the top, there is a navigation bar with 'Services', 'Research', 'Training', and 'About us'. Below this is the ENA logo and a search bar. The search bar contains the text 'GSE107337' and has a 'Search' button. Three red arrows with numbers 1, 2, and 3 point to the ENA logo, the search input field, and the search button, respectively. Below the search bar, there are links for 'Advanced' and 'Sequence'. The main content area features a 'European Nucleotide Archive' heading, a description of the archive, and a 'Text Search' section with an input field and examples. On the right side, there is a 'Popular' section with a list of links and a 'Latest ENA news' section with a recent update.

European Nucleotide Archive < E x +

European Bioinformatics Institute [GB] | ebi.ac.uk/ena

This website uses cookies. By continuing to browse this site, you are agreeing to the use of our site cookies. To find out more, see our [Terms of Use.](#) OK

EMBL-EBI Services Research Training About us

ENA ①
European Nucleotide Archive

② Search ③
Examples: [BN000065](#), [histone](#)
[Advanced](#)
[Sequence](#)

[Home](#) [Search & Browse](#) [Submit & Update](#) [Software](#) [About ENA](#) [Support](#)

European Nucleotide Archive

The European Nucleotide Archive (ENA) provides a comprehensive record of the world's nucleotide sequencing information, covering raw sequencing data, sequence assembly information and functional annotation. [More about ENA](#)

Access to ENA data is provided though the browser, through search tools, large scale file download and through the API.

Text Search

Examples: [BN000065](#), [histone](#)

Popular

- [Submit and update](#)
- [Sequence submissions](#)
- [Genome assembly submissions](#)
- [Submitting environmental sequences](#)
- [Citing ENA data](#)
- [Rest URLs for data retrieval](#)
- [Rest URLs to search ENA](#)

Latest ENA news

08 Jul 2019: [Release 140 of ENA's assembled/annotated sequences is now](#)

W1: ENA

GSE107337で検索しているはずですが、検索結果は① PRJNS419802や②SRP125628となっていることがわかります。結論としては問題ない。③ページ下部に移動

The screenshot shows the ENA website interface. At the top, there is a search bar with a 'Search' button and links for 'Advanced' and 'Sequence'. Below the search bar is a navigation menu with 'Home', 'Search & Browse', 'Submit & Update', 'Software', 'About ENA', and 'Support'. The main content area displays the study details for PRJNA419802, including the title 'RNA-seq analysis of Lactobacillus at acidic stress', the submitting center 'Computational and Synthetic Biology Laboratory, Department of Biotechnology, Korea University', and the organism 'Lactobacillus rhamnosus GG'. A secondary accession number SRP125628 is also listed. The description at the bottom explains the study's purpose: 'To understand transcriptional regulation of probiotic bacteria under acidic condition, RNAseq analysis was carried out over different growth conditions Overall design: Comparison of acidic (pH4) and neutral (pH7) conditions by differentially expressed genes'. Three red arrows with circled numbers 1, 2, and 3 point to the study ID, the secondary accession number, and the search bar, respectively.

EMBL-EBI Services Research Training About us

ENA

European Nucleotide Archive

Examples: [BN000065](#), [histone](#) Search

[Advanced](#)
[Sequence](#)

Home **Search & Browse** Submit & Update Software About ENA Support

[Contact Helpdesk](#)

Study: PRJNA419802

RNA-seq analysis of Lactobacillus at acidic stress

View: [Project XML](#) [Study XML](#) Download: [Project XML](#) [Study XML](#)

Name	Submitting Centre	Organism
RNA-seq analysis of Lactobacillus at acidic stress	Computational and Synthetic Biology Laboratory, Department of Biotechnology, Korea University	Lactobacillus rhamnosus GG

Secondary accession(s)
[SRP125628](#)

Description

To understand transcriptional regulation of probiotic bacteria under acidic condition, RNAseq analysis was carried out over different growth conditions Overall design: Comparison of acidic (pH4) and neutral (pH7) conditions by differentially expressed genes

W1: ENA

①このあたりまで移動すると、②全9サンプルに付随する様々なID情報を一覧できる。原著論文には明記されていないが、③このデータがpaired-endであることもわかる。

https://www.ebi.ac.uk/ena/data/

European Bioinformatics Institute [GB] | ebi.ac.uk/ena/data/view/PRJNA419802

Showing results 1 - 9 of 9 results

Study accession	Sample accession	Secondary sample accession	Experiment accession	Run accession	Tax ID	Scientific name	Instrument model	Library layout	FASTQ files (FTP)	FASTQ files (Galaxy)
PRJNA419802	SAMN08098216	SRS2714081	SRX3422361	SRR6322562	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098215	SRS2714083	SRX3422362	SRR6322563	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098214	SRS2714082	SRX3422363	SRR6322564	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098213	SRS2714084	SRX3422364	SRR6322565	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098212	SRS2714085	SRX3422365	SRR6322566	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098211	SRS2714086	SRX3422366	SRR6322567	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2

W1: ENA

①今回はここを利用して、ENAから直接GalaxyにFASTQファイルをアップロードする。ウェブ資料では、冗長な説明を回避すべく、②SRR6322564、③SRR6322567、およびここでは見えていないがSRR6322569の3サンプルに限定して行う。

https://www.ebi.ac.uk/ena/data/ × +
← → ↻ 🏠 🔒 European Bioinformatics Institute [GB] | ebi.

Showing results 1 - 9 of 9 results

Study accession	Sample accession	Secondary sample accession	Experiment accession	Run accession	Tax ID	Scientific name	Instrument model	Library layout	FASTQ files (FTP)	FASTQ files (Galaxy)
PRJNA419802	SAMN08098216	SRS2714081	SRX3422361	SRR6322562	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098215	SRS2714083	SRX3422362	SRR6322563	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098214	SRS2714082	SRX3422363	SRR6322564	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098213	SRS2714084	SRX3422364	SRR6322565	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098212	SRS2714085	SRX3422365	SRR6322566	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098211	SRS2714086	SRX3422366	SRR6322567	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2



Contents

- W1: 公共データベースENA
- W2: クラウド解析環境Galaxy
- W3: ENAからGalaxyへのFASTQファイルの送付
- W4: クオリティチェック (FastQC)
- W5: クオリティコントロール (Trimmomatic)
- W6: クオリティチェック (FastQC)
- W7: ゲノム配列へのマッピング (Bowtie2)
- W8: カウント情報取得 (htseq-count)
- W9: カウント情報の連結 (Column join)
- W10: カウントデータの数値行列の行名について
- W11: 公共カウントデータ (GSE107337) を用いて、倍率変化と \log_2 (倍率変化)を算出
- W12: Galaxy上で得られたカウントデータについても同様に計算し、比較
- W13とW14: 原著論文中の記載事項を確認
- W15: 2つのカウントデータから共通遺伝子の情報を用いて類似度を計算

①新規にタブを開いて、Galaxy上で、②ログインしておく。

W2: Galaxy

The screenshot shows the Galaxy web interface in a browser. A red arrow labeled '1' points to the browser tab titled 'Galaxy'. Another red arrow labeled '2' points to the 'ログイン/登録' (Login/Register) button in the top navigation bar. The interface includes a search bar for tools, a list of tool categories on the left, a central content area with a 'Galaxy 101' tutorial, and a history panel on the right.

Galaxy is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy start here or consult our help resources. You can install your own Galaxy by following the tutorial and choose from thousands of tools from the Tool Shed.

Galaxy 101
an introduction to Galaxy tutorial

Galaxy Training Network

History

search datasets

Unnamed history
(empty)

历史信息は空です。You can load your own data or get data from an external source

Tweets by @galaxyproject

ユーザ名とパスワードを入力して、①Login。

W2: Galaxy

https://www.ebi.ac.uk/ena/data/v x Galaxy x +

usegalaxy.org/login

Galaxy データ解析 ワークフロー 可視化する 共有データ ヘルプ ログイン/登録 Using 0%

Welcome to Galaxy, please log in

Username or Email Address

kadota@iu.a.u-tokyo.ac.jp

Password

Forgot password? Click here to reset your password.

Login ①

Don't have an account? Register here.

W2: Galaxy

私のログイン直後の状態。①予めこれまでの複数のヒストリーを全部消去してあったのでこんな感じになっているが、ヒトによってはここに直前に行ったヒストリーが見えていると思います。例えば、第12回のW11-2。

The screenshot shows the Galaxy web interface. The main content area displays a welcome message: "Galaxy is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy start here or consult our help resources. You can install your own Galaxy by following the tutorial and choose from thousands of tools from the Tool Shed." Below this is a "Galaxy 101 introduction to Galaxy tutorial" banner with a "125" badge. The left sidebar contains navigation menus for "Tools", "Get Data", "Send Data", "Collection Operations", "GENERAL TEXT TOOLS", "GENOMIC FILE MANIPULATION", and "FASTQ Quality Control". The right sidebar shows the "History" panel, which is currently empty and displays the message: "History (empty)". A blue information box in the history panel states: "ヒストリーは空です。 You can load your own data or get data from an external source". A red arrow with the number "1" points to this message box. The browser address bar shows "usegalaxy.org".

W2: Galaxy

The screenshot shows the Galaxy web interface in a browser window. The address bar shows 'https://www.ebi.ac.uk/ena/data/' and 'usegalaxy.org'. The main navigation bar includes 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', and 'ユーザー'. A 'Using 0%' indicator is visible in the top right. The 'Tools' sidebar on the left lists various tool categories like 'Get Data', 'Send Data', 'Collection Operations', 'GENERAL TEXT TOOLS', 'Text Manipulation', 'Filter and Sort', 'Join, Subtract and Group', 'Datamash', 'GENOMIC FILE MANIPULATION', 'FASTA/FASTQ', 'FASTQ Quality Control', 'SAM/BAM', and 'BED'. The main content area displays a welcome message: 'Galaxy is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy start here or consult our help resources. You can install your own Galaxy by following the tutorial and choose from thousands of tools from the Tool Shed.' Below this is a 'Galaxy Help' banner with the text 'Got Questions? Get Answers.' and the URL 'help.galaxyproject.org'. The 'History' panel on the right shows 'Unnamed history (empty)' and a 'Create new history' button, which is highlighted with a red arrow and the number 1. A blue information box below the history panel contains the text: 'i ヒストリーは空です。 You can load your own data or get data from an external source'. The bottom of the page shows a 'Tweets by @galaxyproject' section.

W2: Galaxy

The screenshot shows the Galaxy web interface. The browser address bar displays `https://www.ebi.ac.uk/ena/data/` and `usegalaxy.org`. The main navigation bar includes "Galaxy" and menu items like "データ解析", "ワークフロー", "可視化する", "共有データ", "ヘルプ", and "ユーザー". The "Tools" panel on the left lists various tool categories such as "Get Data", "Send Data", "Collection Operations", "GENERAL TEXT TOOLS", "GENOMIC FILE MANIPULATION", and "FASTA/FASTQ". The central workspace contains a text block about Galaxy and a "Galaxy 101" tutorial banner. The "History" panel on the right shows a search bar and a list of datasets, with the top item being "Unnamed history". A red circle with the number "1" highlights a tooltip that says "ヒストリーの名前を変更するにはクリック" (Click to change the history name). Below the history list, a blue information box states "ヒストリーは空です。You can load your own data or get data from an external source".

W2: Galaxy

とりあえず①ヒストリー名をGSE107337_3samplesと変更。
②変更完了。第12回のW5-2のおさらいでした。

The screenshot shows the Galaxy web interface. The browser address bar displays `https://www.ebi.ac.uk/ena/data/` and `usegalaxy.org`. The main navigation bar includes "Galaxy" and menu items like "データ解析", "ワークフロー", "可視化する", "共有データ", "ヘルプ", and "ユーザー". A status bar on the right indicates "Using 0%".

The left sidebar contains a "Tools" section with a search bar and categories such as "Get Data", "Send Data", "Collection Operations", "GENERAL TEXT TOOLS", "Text Manipulation", "Filter and Sort", "Join, Subtract and Group", "Datamash", "GENOMIC FILE MANIPULATION", "FASTA/FASTQ", "FASTQ Quality Control", "SAM/BAM", and "BED".

The central workspace displays a welcome message: "Galaxy is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy start here or consult our help resources. You can install your own Galaxy by following the tutorial and choose from thousands of tools from the Tool Shed." Below this is a "Galaxy Help" banner with the text "Got Questions? Get Answers." and the URL `help.galaxyproject.org`.

The right sidebar shows the "History" panel with a search bar and a list of datasets. The dataset `GSE107337_3samples` is highlighted with a red arrow and the number 2. Below the list is a blue information box with the text: "历史信息は空です。 You can load your own data or get data from an external source".

Contents

- W1: 公共データベースENA
- W2: クラウド解析環境Galaxy
- W3: ENAからGalaxyへのFASTQファイルの送付
- W4: クオリティチェック (FastQC)
- W5: クオリティコントロール (Trimmomatic)
- W6: クオリティチェック (FastQC)
- W7: ゲノム配列へのマッピング (Bowtie2)
- W8: カウント情報取得 (htseq-count)
- W9: カウント情報の連結 (Column join)
- W10: カウントデータの数値行列の行名について
- W11: 公共カウントデータ (GSE107337) を用いて、倍率変化と \log_2 (倍率変化)を算出
- W12: Galaxy上で得られたカウントデータについても同様に計算し、比較
- W13とW14: 原著論文中の記載事項を確認
- W15: 2つのカウントデータから共通遺伝子の情報を用いて類似度を計算

ブラウザのタブをENAのほうに変更した状態。

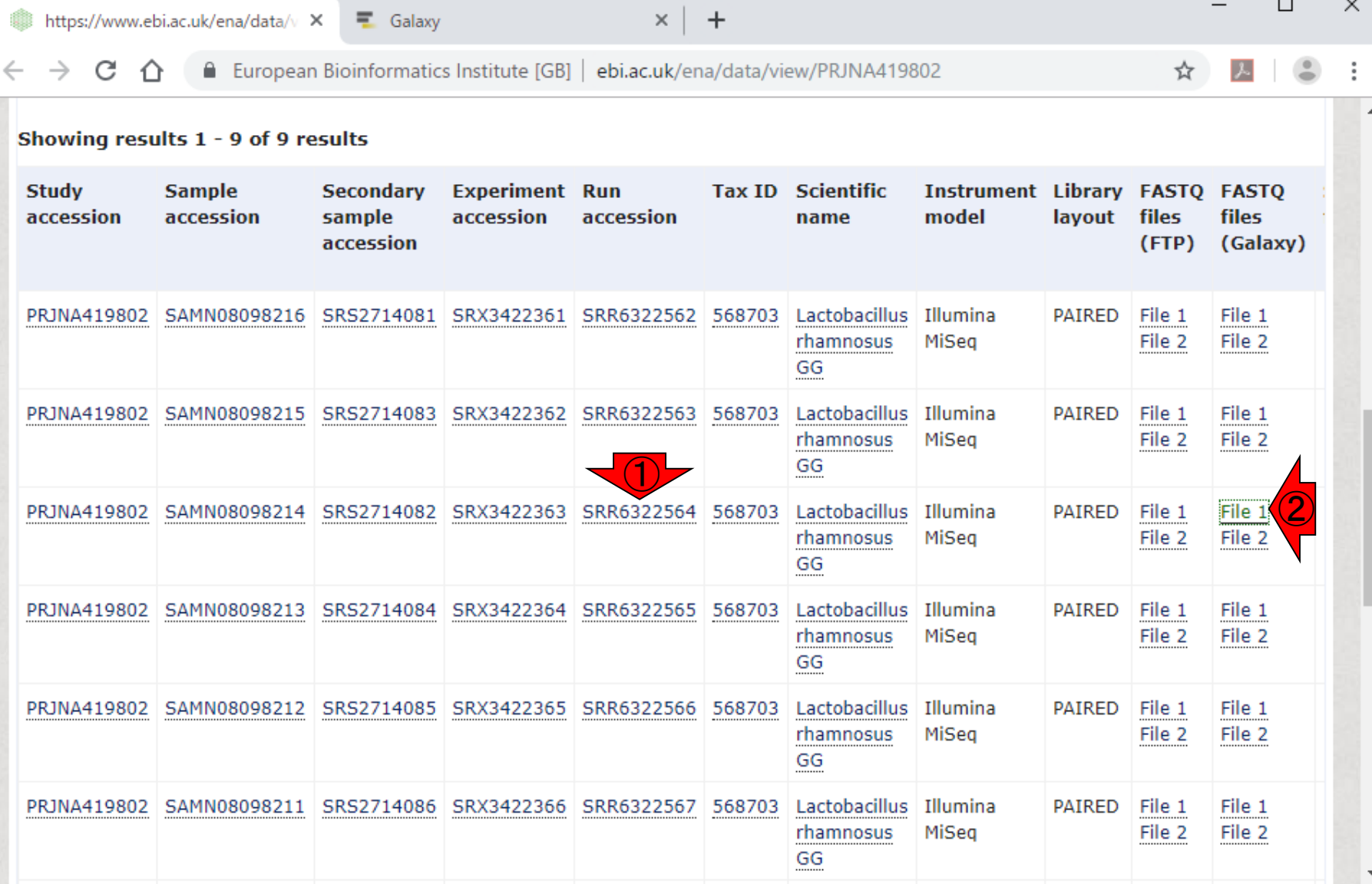
W3-1: ENA → Galaxy

Showing results 1 - 9 of 9 results

Study accession	Sample accession	Secondary sample accession	Experiment accession	Run accession	Tax ID	Scientific name	Instrument model	Library layout	FASTQ files (FTP)	FASTQ files (Galaxy)
PRJNA419802	SAMN08098216	SRS2714081	SRX3422361	SRR6322562	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098215	SRS2714083	SRX3422362	SRR6322563	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098214	SRS2714082	SRX3422363	SRR6322564	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098213	SRS2714084	SRX3422364	SRR6322565	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098212	SRS2714085	SRX3422365	SRR6322566	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098211	SRS2714086	SRX3422366	SRR6322567	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2

W3-1: ENA → Galaxy

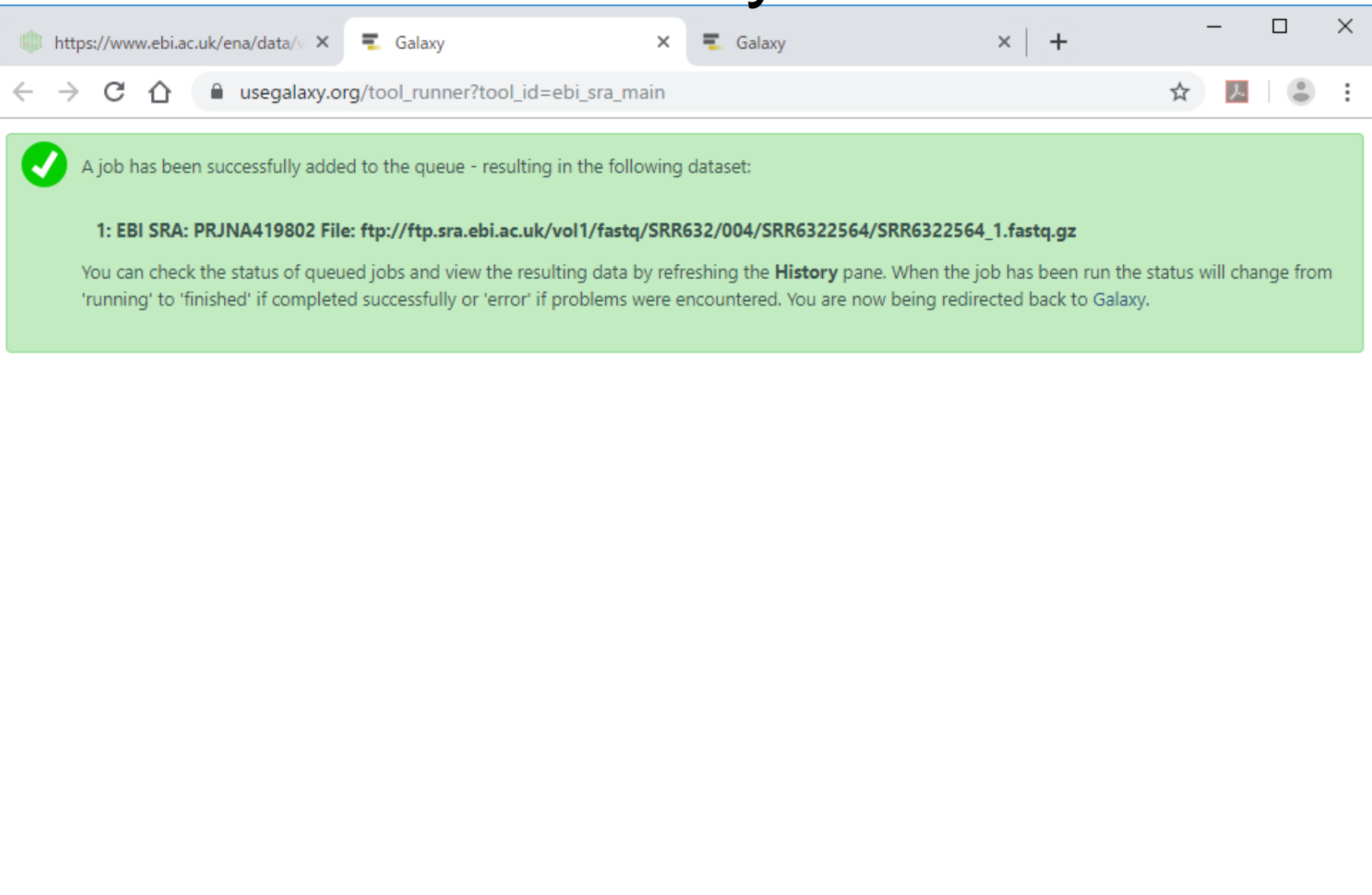
とりあえず、1つめの解析対象サンプルである、① SRR6322564の、② File 1 (paired-endの1つ目; forward側とする)をGalaxyに送る。クリックするのは②のみ。



Showing results 1 - 9 of 9 results

Study accession	Sample accession	Secondary sample accession	Experiment accession	Run accession	Tax ID	Scientific name	Instrument model	Library layout	FASTQ files (FTP)	FASTQ files (Galaxy)
PRJNA419802	SAMN08098216	SRS2714081	SRX3422361	SRR6322562	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098215	SRS2714083	SRX3422362	SRR6322563	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098214	SRS2714082	SRX3422363	SRR6322564	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098213	SRS2714084	SRX3422364	SRR6322565	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098212	SRS2714085	SRX3422365	SRR6322566	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098211	SRS2714086	SRX3422366	SRR6322567	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2

W3-1 : ENA → Galaxy



The screenshot shows a web browser window with two tabs. The active tab is titled 'Galaxy' and shows the URL `usegalaxy.org/tool_runner?tool_id=ebi_sra_main`. A green notification box with a checkmark icon contains the following text:

A job has been successfully added to the queue - resulting in the following dataset:

1: EBI SRA: PRJNA419802 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR632/004/SRR6322564/SRR6322564_1.fastq.gz

You can check the status of queued jobs and view the resulting data by refreshing the **History** pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered. You are now being redirected back to Galaxy.

W3-1: ENA → Galaxy

すぐにこんな感じになる。そしてまたすぐにこんな感じの画面に切り替わり、①灰色の実行待ち状態になる。

The screenshot shows the Galaxy web interface. The main content area displays a tutorial titled "Galaxy 101 an introduction to Galaxy tutorial" from the Galaxy Training Network. The interface includes a left sidebar with tool categories like "Get Data", "Send Data", and "GENERAL TEXT TOOLS". The right sidebar shows the "History" section with a search bar and a list of datasets. The first entry in the history is "1: EBI SRA: PRJNA41980" with a file path "2 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR632/004/SRR6322564/SRR6322564_1.fastq.gz". A red arrow with the number "1" points to this entry.

W3-1: ENA → Galaxy

すぐにこんな感じになる。そしてまたすぐにこんな感じの画面に切り替わり、①灰色の実行待ち状態になる。②さきほど作成したGSE107337_3samplesというヒストリー上で行われています。

The screenshot shows the Galaxy web interface. The main content area displays a tutorial titled "Galaxy 101 an introduction to Galaxy tutorial" from the Galaxy Training Network. On the right side, the "History" panel shows a list of datasets. The first entry is "GSE107337_3samples" with a red arrow labeled "2" pointing to it. Below it, the execution details for "1: EBI SRA: PRJNA41980" and "2 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR632/004/SRR6322564/SRR6322564_1.fastq.gz" are shown, with a red arrow labeled "1" pointing to this entry. The interface includes a navigation menu on the left with categories like "Tools", "Get Data", "Send Data", "Collection Operations", "GENERAL TEXT TOOLS", "GENOMIC FILE MANIPULATION", and "FASTA/FASTQ". The top navigation bar includes "データ解析", "ワークフロー", "可視化する", "共有データ", "ヘルプ", and "ユーザー".

W3-1: ENA → Galaxy

The screenshot shows the Galaxy web interface. The browser address bar displays `https://www.ebi.ac.uk/ena/data/` and `usegalaxy.org`. The main content area features a welcome message and a "Galaxy 101" tutorial banner. On the right, the "History" panel shows a dataset named "GSE107337_3samples" with a file path: `ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR632/004/SRR6322564/SRR6322564_1.fastq.gz`. A red arrow with the number "1" points to this dataset entry.

W3-1: ENA → Galaxy

①取り込み完了。②サイズがemptyから304.99 MBに増加しました。第13回原稿の表1の結果(319.8 MB)と似ており、妥当ですね。

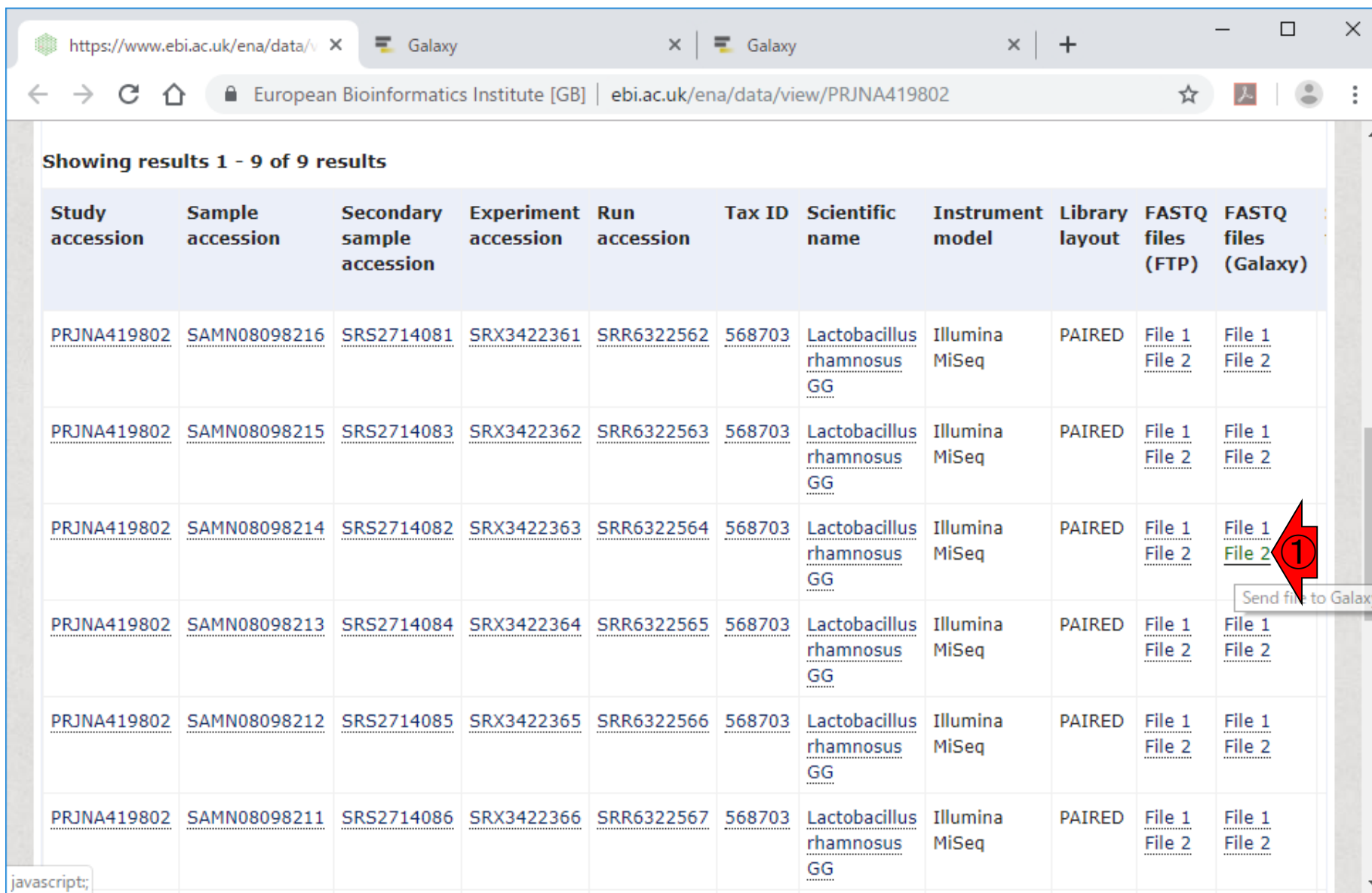
The screenshot shows the Galaxy web interface. The main content area displays a welcome message: "Galaxy is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy start here or consult our help resources. You can install your own Galaxy by following the tutorial and choose from thousands of tools from the Tool Shed." Below this is a large red cross logo for "BASEL JULY 21-25 ISMB ECCB 2019" and a "Tweets by @galaxyproject" section.

The left sidebar contains a "Tools" section with a search bar and categories like "Get Data", "Send Data", "Collection Operations", "GENERAL TEXT TOOLS", "GENOMIC FILE MANIPULATION", and "FASTA/FASTQ".

The right sidebar shows the "History" section with a search bar and a list of datasets. The first dataset is "GSE107337_3samples" with a size of "304.99 MB". A red arrow labeled "2" points to this size. Below it, a green highlighted entry shows the file path: "1: EBI SRA: PRJNA419802 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR632/004/SRR6322564/SRR6322564_1.fastq.gz". A red arrow labeled "1" points to this entry.

①次はSRR6322564のFile 2 (paired-endのreverse側)。

W3-2: SRR6322564のFile2

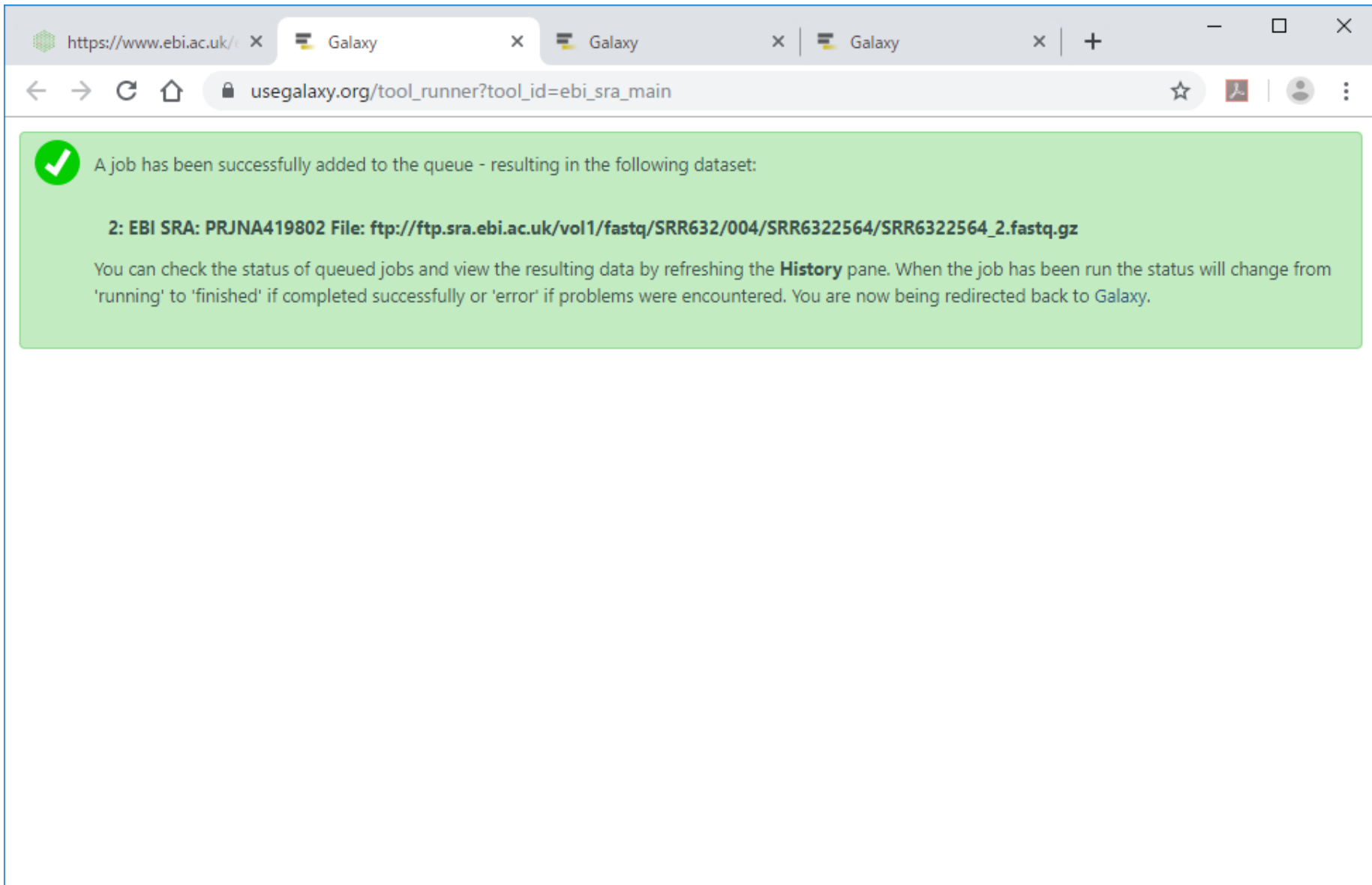


Showing results 1 - 9 of 9 results

Study accession	Sample accession	Secondary sample accession	Experiment accession	Run accession	Tax ID	Scientific name	Instrument model	Library layout	FASTQ files (FTP)	FASTQ files (Galaxy)
PRJNA419802	SAMN08098216	SRS2714081	SRX3422361	SRR6322562	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098215	SRS2714083	SRX3422362	SRR6322563	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098214	SRS2714082	SRX3422363	SRR6322564	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2 ①
PRJNA419802	SAMN08098213	SRS2714084	SRX3422364	SRR6322565	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098212	SRS2714085	SRX3422365	SRR6322566	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098211	SRS2714086	SRX3422366	SRR6322567	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2

さきほどと同様に、すぐにこんな感じになる。ここまでで、どんどんタブが追加されていくのだと学習する。

W3-2: SRR6322564



The screenshot shows a web browser window with the URL `usegalaxy.org/tool_runner?tool_id=ebi_sra_main`. The browser has several tabs open, all labeled "Galaxy". A green notification box with a checkmark icon contains the following text:

A job has been successfully added to the queue - resulting in the following dataset:

2: EBI SRA: PRJNA419802 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR632/004/SRR6322564/SRR6322564_2.fastq.gz

You can check the status of queued jobs and view the resulting data by refreshing the **History** pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered. You are now being redirected back to Galaxy.

W3-2: SRR6322564

ほどなくして、こんな感じになる。実行待ち状態。①や②のタブは不要なので消してよい。



A screenshot of the Galaxy web interface. The top navigation bar includes 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', and 'ユーザー'. The main content area shows a workflow execution status with a central message: 'Galaxy is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy start here or consult our help resources. You can install your own Galaxy by following the tutorial and choose from thousands of tools from the Tool Shed.' Below this is a 'Galaxy 101 an introduction to Galaxy tutorial' banner. On the right, the 'History' panel shows two datasets: '2: EBI SRA: PRJNA41980' and '1: EBI SRA: PRJNA419802 File'. The left sidebar contains various tool categories like 'GENERAL TEXT TOOLS' and 'GENOMIC FILE MANIPULATION'. At the bottom, there is a 'Tweets by @galaxyproject' section.

W3-2: SRR6322564

取り込み(インポート)完了を意味する緑色になった。約3分。ここまでで、SRR6322564のpaired-endの2つのファイルが解析可能状態となった。

The screenshot displays the Galaxy web interface. The top navigation bar includes the Galaxy logo and menu items: データ解析, ワークフロー, 可視化する, 共有データ, ヘルプ, ユーザー. The main content area is divided into three sections:

- Tools:** A search bar and a list of tool categories including 'Get Data', 'Send Data', 'Collection Operations', 'GENERAL TEXT TOOLS', 'Text Manipulation', 'Filter and Sort', 'Join, Subtract and Group', 'Datamash', 'GENOMIC FILE MANIPULATION', 'FASTA/FASTQ', 'FASTQ Quality Control', 'SAM/BAM', and 'BED'.
- History:** A search bar and a list of datasets. The dataset 'GSE107337_3samples' is highlighted in green, indicating it is ready for analysis. It shows 2 datasets shown, with a total size of 643.12 MB. The files listed are:
 - 2: EBI SRA: PRJNA419802 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR632/004/SRR6322564/SRR6322564_2.fastq.gz
 - 1: EBI SRA: PRJNA419802 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR632/004/SRR6322564/SRR6322564_1.fastq.gz
- Workspace:** A central area with a welcome message: 'Galaxy is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy start here or consult our help resources. You can install your own Galaxy by following the tutorial and choose from thousands of tools from the Tool Shed.' Below the message is a logo for 'BASEL JULY 21-25 ISMB/ECCB 2019' and a 'Tweets by @galaxyproject' section.

W3-3: rename

赤枠で示す部分の名前が長いので、①編集ボタンを押して名前を変更する。

The screenshot shows the Galaxy web interface. The main content area displays a tutorial titled "Galaxy 101 an introduction to Galaxy tutorial" from the Galaxy Training Network. On the right, the "History" panel shows a list of datasets. The second dataset, "2: EBI SRA: PRJNA419802 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR632/004/SRR6322564/SRR6322564_2.fastq.gz", is highlighted with a red box. A red arrow with the number "1" points to the edit icon (pencil) next to this dataset. A tooltip with the text "変数を編集する" (Edit variable) is visible over the edit icon. The first dataset, "1: EBI SRA: PRJNA419802 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR632/004/SRR6322564/SRR6322564_1.fastq.gz", is also visible below it.

W3-3: rename

赤枠で示す部分の名前が長いので、①編集ボタンを押して名前を変更する。こんな感じになります。②赤枠で見えているのが、③の部分で編集可能。

The screenshot shows the Galaxy web interface for editing dataset attributes. The main panel is titled "Edit dataset attributes" and contains sections for "Attributes", "Convert", "Datatypes", and "Permissions". The "Name" field is currently set to "EBI SRA: PRJNA419802 File: ftp://ftp.sra.ebi.ac.uk/vol1/1". A red arrow labeled "3" points to the end of this long name. The "History" panel on the right shows a list of datasets, with the second entry highlighted in green. A red arrow labeled "2" points to the edit icon (pencil) next to this entry. A red box highlights the first entry in the history panel, which is partially obscured by the second entry.

W3-3: rename

赤枠で示す部分の名前が長いので、①編集ボタンを押して名前を変更する。こんな感じになります。②赤枠で見えているのが、③の部分で編集可能。ここでは、元の情報を④Infoカラムに保持しつつ、⑤のようにrenameし、⑥Save。

The screenshot shows the Galaxy web interface for editing dataset attributes. The browser address bar shows `https://www.ebi.ac.uk/ena/data/v` and `usegalaxy.org/datasets/edit`. The page title is "Galaxy" and the user is logged in as "ユーザー". The main content area is titled "Edit dataset attributes" and contains several sections:

- Attributes:** Includes a search bar, "Convert" button, and "Datatypes" dropdown.
- Permissions:** Includes a "変数を編集する" button, "Auto-detect" button, and "Save" button (indicated by red arrow 6).
- Name:** A text input field containing "SRR6322564_1" (indicated by red arrow 5).
- Info:** A text area containing the original file path: "EBI SRA: PRJNA419802 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR632/004/SRR6322564/SRR6322564_1.fastq.gz" (indicated by red arrow 4).
- Annotation:** An empty text area.

The right sidebar shows the "History" section with a search bar and a list of datasets. The selected dataset is "GSE107337_3samples" (643.12 MB). The list shows two entries:

- 2: EBI SRA: PRJNA419802 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR632/004/SRR6322564/SRR6322564_2.fastq.gz
- 1: EBI SRA: PRJNA419802 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR632/004/SRR6322564/SRR6322564_1.fastq.gz

W3-3: rename

①forward側ファイルのrename完了。②reverse側についても同様に行う。

The screenshot shows the Galaxy web interface at <https://www.ebi.ac.uk/ena/data/>. The browser address bar shows <https://www.ebi.ac.uk/ena/data/> and the page URL is usegalaxy.org/datasets/edit. The interface is divided into three main sections:

- Tools:** A sidebar on the left with a search bar and categories like "Get Data", "Send Data", "Collection Operations", "GENERAL TEXT TOOLS", "Text Manipulation", "Filter and Sort", "Join, Subtract and Group", "Datamash", "GENOMIC FILE MANIPULATION", "FASTA/FASTQ", "FASTQ Quality Control", "SAM/BAM", and "BED".
- Edit dataset attributes:** The central panel shows a green notification "Attributes updated." and tabs for "Attributes", "Convert", "Datatypes", and "Permissions". Below the tabs are buttons for "変数を編集する", "Auto-detect", and "Save". The "Name" field contains "SRR6322564_1" and the "Info" field contains "EBI SRA: PRJNA419802 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR632/004/SRR6322564/SRR6322564_1.fastq.gz".
- History:** The right panel shows a search bar and a list of datasets. The top entry is "GSE107337_3samples" (2 shown, 643.12 MB). Below it, two dataset entries are visible:
 - Entry 2: "EBI SRA: PRJNA419802 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR632/004/SRR6322564/SRR6322564_2.fastq.gz". A red arrow labeled "2" points to the edit icon (pencil) for this entry.
 - Entry 1: "SRR6322564_1". A red arrow labeled "1" points to the edit icon (pencil) for this entry.

W3-3: rename

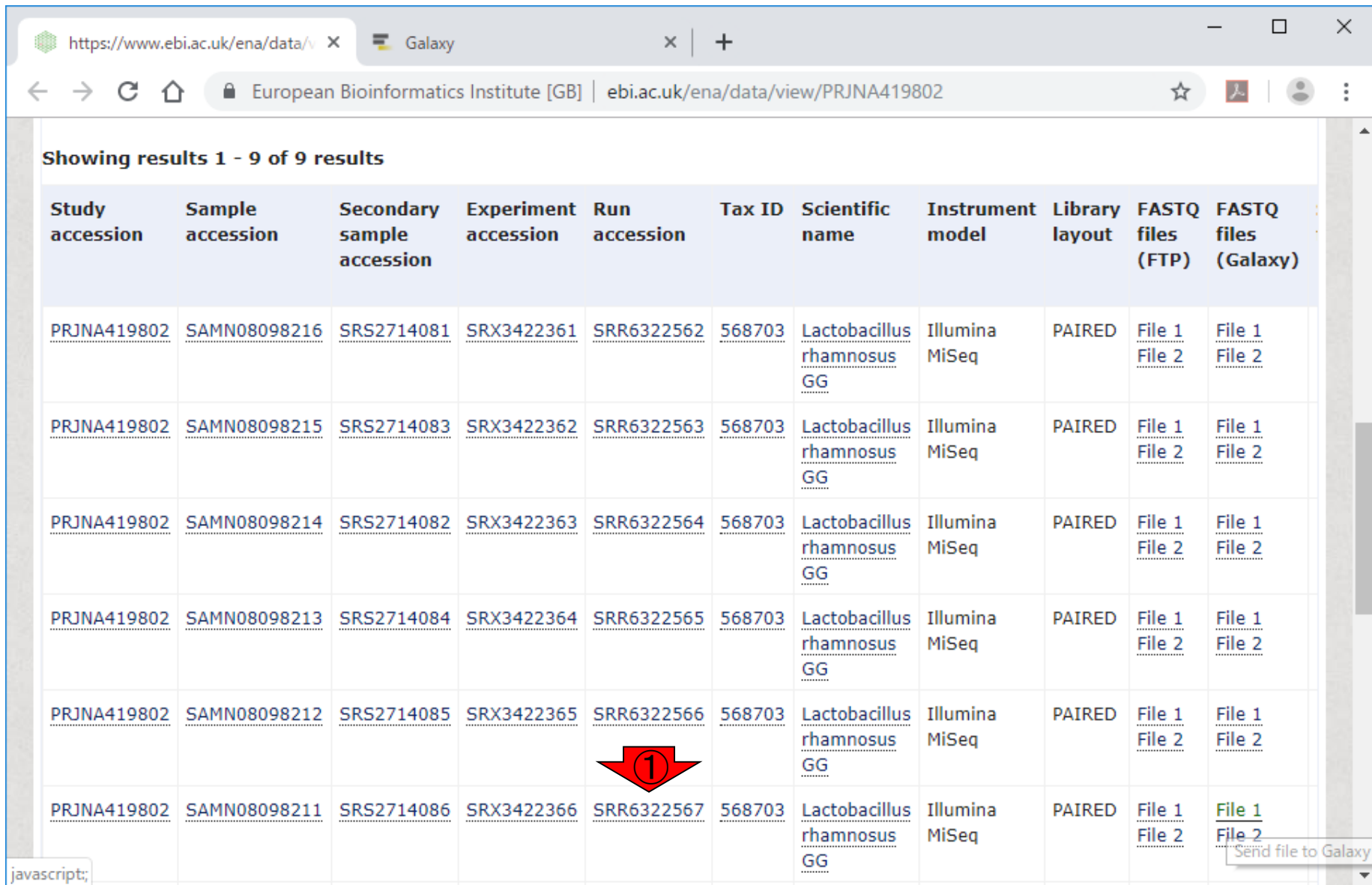
①forward側ファイルのrename完了。②reverse側についても同様に行う。作業完了。

The screenshot displays the Galaxy web interface for editing dataset attributes. The browser address bar shows `https://www.ebi.ac.uk/ena/data/v` and `usegalaxy.org/datasets/edit`. The Galaxy logo and navigation menu are visible at the top. The main content area is divided into three panels:

- Tools:** A sidebar on the left with a search bar and categories like "Get Data", "Send Data", "Collection Operations", "GENERAL TEXT TOOLS", "Text Manipulation", "Filter and Sort", "Join, Subtract and Group", "Datamash", "GENOMIC FILE MANIPULATION", "FASTA/FASTQ", "FASTQ Quality Control", "SAM/BAM", and "BED".
- Edit dataset attributes:** The central panel shows a green notification "Attributes updated." Below it are tabs for "Attributes", "Convert", and "Datatypes". Under "Attributes", there is a "Permissions" section and a button "変数を編集する" (Edit variables). Below that are "Auto-detect" and "Save" buttons. The "Name" field contains "SRR6322564_2". The "Info" field contains the text: "EBI SRA: PRJNA419802 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR632/004/SRR6322564/SRR6322564_2.fastq.gz". The "Annotation" field is currently empty.
- History:** The right panel shows a search bar for datasets. Below it, the dataset "GSE107337_3samples" is listed with "2 shown" and "643.12 MB". Two datasets are visible in the history list:
 - 2: SRR6322564_2 (highlighted in green)
 - 1: SRR6322564_1

再びENAのタブ。今度は①SRR6322567。クリックしない。

W3-4: SRR6322567



Showing results 1 - 9 of 9 results

Study accession	Sample accession	Secondary sample accession	Experiment accession	Run accession	Tax ID	Scientific name	Instrument model	Library layout	FASTQ files (FTP)	FASTQ files (Galaxy)
PRJNA419802	SAMN08098216	SRS2714081	SRX3422361	SRR6322562	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098215	SRS2714083	SRX3422362	SRR6322563	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098214	SRS2714082	SRX3422363	SRR6322564	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098213	SRS2714084	SRX3422364	SRR6322565	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098212	SRS2714085	SRX3422365	SRR6322566	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098211	SRS2714086	SRX3422366	SRR6322567	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2

Send file to Galaxy

W3-4: SRR6322

再びENAのタブ。今度は①SRR6322567。クリックしない。Galaxyへの送付作業は、ファイル1つの送付完了を待つ必要はない。例えば、②File1をクリックして、0.2秒後にFile2をクリックしてもよい。やってみる。以前示したようにすぐに新規タブに飛ばされるが、すぐにENAに戻ってFile2をクリックすると次のスライドのようになる。

Showing results 1 - 9 of 9 results

Study accession	Sample accession	Secondary sample accession	Experiment accession	Run accession	Tax ID	Scientific name	Instrument model	Library layout	FASTQ files (FTP)	FASTQ files (Galaxy)
PRJNA419802	SAMN08098216	SRS2714081	SRX3422361	SRR6322562	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098215	SRS2714083	SRX3422362	SRR6322563	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098214	SRS2714082	SRX3422363	SRR6322564	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098213	SRS2714084	SRX3422364	SRR6322565	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098212	SRS2714085	SRX3422365	SRR6322566	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098211	SRS2714086	SRX3422366	SRR6322567	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2

① ↓ (Red arrow pointing to SRR6322567)

② ↖ (Red arrow pointing to File 2)

Send file to Galaxy

W3-4: SRR6322567

The screenshot shows the Galaxy web interface. The main content area displays a tutorial titled "Running Your Own Understanding how Galaxy works" with the subtitle "An in-depth tutorial". The interface includes a left-hand navigation menu with categories like "Tools", "Get Data", "Send Data", "Collection Operations", "GENERAL TEXT TOOLS", "GENOMIC FILE MANIPULATION", and "FASTA/FASTQ". The right-hand panel shows the "History" section with a search bar and a list of datasets. The top navigation bar includes "データ解析", "ワークフロー", "可視化する", "共有データ", "ヘルプ", and "ユーザー". The browser address bar shows "usegalaxy.org".

History

search datasets

GSE107337_3samples

4 shown

643.12 MB

4: EBI SRA: PRJNA41980
2 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR632/007/SRR6322567/SRR6322567_2.fastq.gz

3: EBI SRA: PRJNA41980
2 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR632/007/SRR6322567/SRR6322567_1.fastq.gz

2: SRR6322564_2

1: SRR6322564_1

Tweets by @galaxyproject

W3-4: SRR6322567

The screenshot shows the Galaxy web interface. The browser address bar displays `https://www.ebi.ac.uk/` and `usegalaxy.org`. The navigation menu on the left includes sections like 'Tools', 'Get Data', 'Send Data', 'Collection Operations', 'GENERAL TEXT TOOLS', 'GENOMIC FILE MANIPULATION', and 'FASTA/FASTQ'. The central content area features a banner for '125+ ways to use Galaxy' and a tweet from @galaxyproject. The right-hand 'History' panel shows a search for 'GSE107337_3samples' with 4 results shown, including file paths like `ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR632/007/SRR6322567/SRR6322567_2.fastq.gz`.

W3-4: SRR6322567

The screenshot shows the Galaxy web interface. The main content area displays a tutorial titled "Galaxy 101: an introduction to Galaxy tutorial" from the Galaxy Training Network. The tutorial text states: "Galaxy is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy start here or consult our help resources. You can install your own Galaxy by following the tutorial and choose from thousands of tools from the Tool Shed." Below the text is a diagram of a Galaxy workflow.

The left sidebar contains a "Tools" section with a search bar and a list of tool categories: "Get Data", "Send Data", "Collection Operations", "GENERAL TEXT TOOLS", "Text Manipulation", "Filter and Sort", "Join, Subtract and Group", "Datamash", "GENOMIC FILE MANIPULATION", "FASTA/FASTQ", "FASTQ Quality Control", "SAM/BAM", and "BED".

The right sidebar shows the "History" section with a search bar and a list of datasets:

- GSE107337_3samples** (4 shown, 643.12 MB)
- 4: EBI SRA: PRJNA41980 2 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR632/007/SRR6322567/SRR6322567_2.fastq.gz
- 3: EBI SRA: PRJNA419802 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR632/007/SRR6322567/SRR6322567_1.fastq.gz
- 2: SRR6322564_2
- 1: SRR6322564_1

At the bottom of the page, there is a "Tweets by @galaxyproject" section.

W3-4: SRR6322567

The screenshot shows the Galaxy web interface. The main content area displays a welcome message: "Galaxy is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy start here or consult our help resources. You can install your own Galaxy by following the tutorial and choose from thousands of tools from the Tool Shed." Below this is a banner for "Try Galaxy on the Cloud" with the text "Now you can have a personal Galaxy within the infinite Universe".

The left sidebar contains a "Tools" section with a search bar and categories like "Get Data", "Send Data", "Collection Operations", "GENERAL TEXT TOOLS", "GENOMIC FILE MANIPULATION", and "FASTQ Quality Control".

The right sidebar shows the "History" section with a search bar and a list of datasets. The dataset list includes:

- GSE107337_3samples** (4 shown, 1.25 GB)
- 4: EBI SRA: PRJNA419802 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR632/007/SRR6322567/SRR6322567_2.fastq.gz
- 3: EBI SRA: PRJNA419802 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR632/007/SRR6322567/SRR6322567_1.fastq.gz
- 2: SRR6322564_2
- 1: SRR6322564_1

At the bottom, there is a "Tweets by @galaxyproject" section.

W3-5: rename

The screenshot shows the Galaxy web interface at <https://www.ebi.ac.uk/>. The browser tabs show multiple instances of Galaxy. The address bar displays usegalaxy.org/datasets/edit. The Galaxy logo and navigation menu are visible at the top, with the user profile set to "ユーザー".

The main content area is divided into three panels:

- Tools:** A sidebar on the left with a search bar and categories like "Get Data", "Send Data", "Collection Operations", "GENERAL TEXT TOOLS", "Text Manipulation", "Filter and Sort", "Join, Subtract and Group", "Datamash", "GENOMIC FILE MANIPULATION", "FASTA/FASTQ", "FASTQ Quality Control", "SAM/BAM", and "BED".
- Edit dataset attributes:** The central panel shows a green notification "Attributes updated." Below it are tabs for "Attributes", "Convert", and "Datatypes". Under "Attributes", there is a "Permissions" section and a button "変数を編集する" (Edit variables). Below this are "Auto-detect" and "Save" buttons. The "Name" field contains "SRR6322567_2". The "Info" field contains "EBI SRA: PRJNA419802 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR632/007/SRR6322567/SRR6322567_2.fastq.gz". The "Annotation" field is currently empty.
- History:** The right panel shows a search bar and a list of datasets under the heading "GSE107337_3samples". It indicates "4 shown" and "1.25 GB". The list contains four entries, each with a checkmark, a trash icon, and a chat icon:
 - 4: SRR6322567_2
 - 3: SRR6322567_1
 - 2: SRR6322564_2
 - 1: SRR6322564_1

W3-6: SRR6322569

PRJNA419802	SAMN08098214	SRS2714082	SRX3422363	SRR6322564	568703	Lactobacillus rhamnosus GG	MiSeq		File 2	File 2
PRJNA419802	SAMN08098214	SRS2714082	SRX3422363	SRR6322564	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098213	SRS2714084	SRX3422364	SRR6322565	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098212	SRS2714085	SRX3422365	SRR6322566	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098211	SRS2714086	SRX3422366	SRR6322567	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098210	SRS2714087	SRX3422367	SRR6322568	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098218	SRS2714088	SRX3422368	SRR6322569	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098217	SRS2714089	SRX3422369	SRR6322570	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2



①

Send to Galaxy

W3-6: SRR6322569

The screenshot shows the Galaxy web interface. The top navigation bar includes 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', and 'ユーザー'. The main content area is divided into three sections:

- Tools:** A sidebar on the left with a search bar and categories like 'Get Data', 'Send Data', 'Collection Operations', 'GENERAL TEXT TOOLS', 'GENOMIC FILE MANIPULATION', and 'FASTA/FASTQ'.
- History:** A sidebar on the right showing a list of datasets. The top entry is 'GSE107337_3samples' (1.25 GB). Below it, a list of SRR datasets is shown:
 - 6: EBI SRA: PRJNA419802 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR632/009/SRR6322569/SRR6322569_2.fastq.gz
 - 5: EBI SRA: PRJNA419802 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR632/009/SRR6322569/SRR6322569_1.fastq.gz
 - 4: SRR6322567_2
 - 3: SRR6322567_1
 - 2: SRR6322564 2
- Central Content:** A text block stating 'Galaxy is an open source, web-based platform for data intensive biomedical research...' followed by a 'Galaxy 101 an introduction to Galaxy tutorial' graphic from the Galaxy Training Network. Below the graphic is a 'Tweets by @galaxyproject' section.

W3-6: SRR6322569

The screenshot shows the Galaxy web interface for editing a dataset. The browser address bar shows `usegalaxy.org/datasets/edit`. The top navigation bar includes the Galaxy logo and menu items like 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', and 'ユーザー'. A status indicator shows 'Using 0%'. The left sidebar lists tool categories: 'Tools', 'Get Data', 'Send Data', 'Collection Operations', 'GENERAL TEXT TOOLS', 'Text Manipulation', 'Filter and Sort', 'Join, Subtract and Group', 'Datamash', 'GENOMIC FILE MANIPULATION', 'FASTA/FASTQ', 'FASTQ Quality Control', 'SAM/BAM', and 'BED'. The central panel, titled 'Edit dataset attributes', displays a green message 'Attributes updated.' and options for 'Attributes', 'Convert', 'Datatypes', and 'Permissions'. Below these is a button '変数を編集する' and 'Auto-detect' and 'Save' buttons. The 'Name' field contains 'SRR6322569_2'. The 'Info' field contains the EBI SRA file path: 'ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR632/009/SRR6322569/SRR6322569_2.fastq.gz'. The right panel, titled 'History', shows a list of datasets under 'GSE107337_3samples' with 6 shown. The list includes: 6: SRR6322569_2, 5: SRR6322569_1, 4: SRR6322567_2, 3: SRR6322567_1, 2: SRR6322564_2, and 1: SRR6322564_1. Each entry has view, edit, and delete icons.

W3-6: SRR6322509

①データ解析のところを押すと、中央パネルをデフォルト画面に戻すことができます。

The screenshot shows the Galaxy web interface. The top navigation bar includes the Galaxy logo, a menu with 'データ解析' (Data Analysis), 'ワークフロー' (Workflow), '可視化する' (Visualize), '共有データ' (Shared Data), and 'ヘルプ' (Help), and a user profile dropdown. A red arrow points to the 'データ解析' menu item. The main content area is divided into three panels: 'Tools' on the left, 'Edit dataset attributes' in the center, and 'History' on the right. The 'Edit dataset attributes' panel shows a green message 'Attributes updated.' and a form for editing dataset attributes. The 'History' panel shows a list of datasets, with the top six highlighted in green.

Dataset ID	View	Edit	Delete
6: SRR6322569_2			
5: SRR6322569_1			
4: SRR6322567_2			
3: SRR6322567_1			
2: SRR6322564_2			
1: SRR6322564_1			

①データ解析のところを押すと、中央パネルをデフォルト画面に戻すことができます。こんな感じ。

W3-7: 解析準備完了

The screenshot shows the Galaxy web interface. The top navigation bar includes 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', and 'ユーザー'. The 'Tools' panel on the left lists various tool categories like 'Get Data', 'Send Data', and 'GENERAL TEXT TOOLS'. The central panel displays a tutorial slide titled 'Galaxy 101 an introduction to Galaxy tutorial' from the 'Galaxy Training Network'. The 'History' panel on the right shows a list of datasets, including 'GSE107337_3samples' and several SRR datasets.

Dataset ID	View	Edit	Delete
6: SRR6322569_2	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5: SRR6322569_1	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4: SRR6322567_2	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3: SRR6322567_1	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2: SRR6322564_2	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1: SRR6322564_1	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Contents

- W1: 公共データベースENA
- W2: クラウド解析環境Galaxy
- W3: ENAからGalaxyへのFASTQファイルの送付
- W4: クオリティチェック(FastQC)
- W5: クオリティコントロール(Trimmomatic)
- W6: クオリティチェック(FastQC)
- W7: ゲノム配列へのマッピング(Bowtie2)
- W8: カウント情報取得(htseq-count)
- W9: カウント情報の連結(Column join)
- W10: カウントデータの数値行列の行名について
- W11: 公共カウントデータ(GSE107337)を用いて、倍率変化と \log_2 (倍率変化)を算出
- W12: Galaxy上で得られたカウントデータについても同様に計算し、比較
- W13とW14: 原著論文中の記載事項を確認
- W15: 2つのカウントデータから共通遺伝子の情報を用いて類似度を計算

次はFastQC。①の中にFastQCがあるので、クリック。

W4-1: FastQC

The screenshot shows the Galaxy web interface. The top navigation bar includes 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', and 'ユーザー'. The 'Tools' sidebar on the left lists various tool categories, with 'FASTQ Quality Control' highlighted by a red arrow and the number 1. The main content area displays a 'Galaxy 101' tutorial banner. The 'History' panel on the right shows a list of datasets, including 'GSE107337_3samples' and several SRR6322569 and SRR6322567 samples.

W4-1 : FastQC

次はFastQC。①の中にFastQCがあるので、クリック。②
FastQC。

The screenshot shows the Galaxy web interface. The browser address bar displays <https://www.ebi.ac.uk/ena/data/> and the page title is "Galaxy". The URL bar shows usegalaxy.org. The main navigation bar includes "Galaxy" and menu items: "データ解析", "ワークフロー", "可視化する", "共有データ", "ヘルプ", "ユーザー", and a "Using 0%" indicator.

The "Tools" panel on the left contains a search bar and a list of tool categories:

- Get Data
- Send Data
- Collection Operations
- GENERAL TEXT TOOLS
 - Text Manipulation
 - Filter and Sort
 - Join, Subtract and Group
 - Datamash
- GENOMIC FILE MANIPULATION
 - FASTA/FASTQ
 - FASTQ Quality Control
 - FastQC Read Quality reports (marked with a red arrow and circled '2')
 - Trimmomatic flexible read trimming

The main content area displays a welcome message: "Galaxy is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy start here or consult our help resources. You can install your own Galaxy by following the tutorial and choose from thousands of tools from the Tool Shed." Below this is a "Galaxy Help" banner with the text "Got Questions? Get Answers." and the URL help.galaxyproject.org.

The "History" panel on the right shows a search bar and a list of datasets under the heading "GSE107337_3samples" (6 shown, 2.15 GB):

Dataset ID	View	Edit	Delete
6: SRR6322569_2			
5: SRR6322569_1			
4: SRR6322567_2			
3: SRR6322567_1			
2: SRR6322564_2			
1: SRR6322564_1			

At the bottom, there is a "Tweets by @galaxyproject" section.

W4-1 : FastQC

次はFastQC。①の中にFastQCがあるので、クリック。② FastQC。こんな感じになる。第11回のW9でもFastQCをやっています。

The screenshot shows the Galaxy web interface. The top navigation bar includes 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', and 'ユーザー'. The main content area is divided into three panels:

- Tools:** A search bar and a list of tool categories including 'GENERAL TEXT TOOLS', 'Text Manipulation', 'Filter and Sort', 'Join, Subtract and Group', 'Datamash', and 'GENOMIC FILE MANIPULATION'. Under 'FASTQ Quality Control', the 'FastQC Read Quality reports' tool is selected.
- FastQC Tool Interface:** Shows the tool name 'FastQC Read Quality reports (Galaxy Version 0.72+galaxy1)' with 'Favorite', 'Versions', and 'Options' buttons. Below are sections for 'Short read data from your current history' (selected: 6: SRR6322569_2), 'Contaminant list' (selected: No tabular dataset), 'Adapter list' (selected: No tabular dataset), and 'Submodule and Limit specifying file' (selected: Nothing selected).
- History:** A list of datasets under the heading 'GSE107337_3samples'. The list shows 6 datasets, with the first one highlighted in green: '6: SRR6322569_2'. Other datasets include '5: SRR6322569_1', '4: SRR6322567_2', '3: SRR6322567_1', '2: SRR6322564_2', and '1: SRR6322564_1'. Each entry has icons for view, edit, and delete.

W4-2: 複数個指定

デフォルトは①Single datasetだが、ここでは②計6ファイル全てについてFastQCを実行したいので、③Multiple datasetsに切り替える。

The screenshot shows the Galaxy web interface with the FastQC tool selected. The 'Content type' dropdown is set to 'Multiple datasets' (indicated by a red arrow and circled '3'). The 'Input' dropdown is set to '6: SRR6322569_2' (indicated by a red arrow and circled '1'). The 'History' panel on the right shows a list of 6 datasets (indicated by a red arrow and circled '2').

Dataset ID	View	Edit	Delete
6: SRR6322569_2	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5: SRR6322569_1	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4: SRR6322567_2	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3: SRR6322567_1	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2: SRR6322564_2	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1: SRR6322564_1	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

W4-2: 複数個指定

こんな感じになる。現状では、まだヒストリーパネル上で実行可能な入力データがリストアップされているだけなので、この中からどれをFastQCの入力として実行するのかを指定せねばならない。①でそれを行う。

The screenshot shows the Galaxy web interface. The main panel displays the configuration for the 'FastQC Read Quality reports' tool. Under the 'Short read data from your current history' section, there is a list of six datasets: 6: SRR6322569_2, 5: SRR6322569_1, 4: SRR6322567_2, 3: SRR6322567_1, 2: SRR6322564_2, and 1: SRR6322564_1. A red arrow with the number '1' points to the 'Multiple datasets' button, indicating that multiple datasets should be selected for input. The right-hand 'History' panel shows a list of datasets from the current history, including the same six datasets, each with a selection icon.

W4-2: 複数個指定

こんな感じになって、ヒストリーパネル上で実行可能なFASTQファイルがリストアップされる。実行したいのは全てなので①全てのファイルをクリックして選択する。

The screenshot shows the Galaxy web interface with a file selection dialog open. The dialog has a search bar and a table of files. The table has three columns: Label, Details, and Time. The files listed are:

Label	Details	Time
6: SRR6322569_2	fastqsanger.gz	2019-08-15 10:25
5: SRR6322569_1	fastqsanger.gz	2019-08-15 10:25
4: SRR6322567_2	fastqsanger.gz	2019-08-15 10:15
3: SRR6322567_1	fastqsanger.gz	2019-08-15 10:15
2: SRR6322564_2	fastqsanger.gz	2019-08-15 09:30
1: SRR6322564_1	fastqsanger.gz	2019-08-15 09:26

A red bracket and arrow labeled '1' point to the right of the table, indicating that all files should be selected. The dialog also has 'Cancel' and 'Ok' buttons at the bottom right.

W4-2: 複数個指定

こんな感じになって、ヒストリーパネル上で実行可能なFASTQファイルがリストアップされる。実行したいのは全てなので①全てのファイルをクリックして選択する。つまりこういう状態にして、②Ok。

The screenshot shows the Galaxy web interface with a file selection dialog open. The dialog has a search bar at the top and a table of files below. The table has three columns: Label, Details, and Time. All files in the table are selected, indicated by a green background. A red arrow labeled '1' points to the selection area. The 'Ok' button is highlighted with a red arrow labeled '2'.

Label	Details	Time
6: SRR6322569_2	fastqsanger.gz	2019-08-15 10:25
5: SRR6322569_1	fastqsanger.gz	2019-08-15 10:25
4: SRR6322567_2	fastqsanger.gz	2019-08-15 10:15
3: SRR6322567_1	fastqsanger.gz	2019-08-15 10:15
2: SRR6322564_2	fastqsanger.gz	2019-08-15 09:30
1: SRR6322564_1	fastqsanger.gz	2019-08-15 09:26

こんな感じになる。①下部に移動して、実行ボタンを探す。

W4-3: FastQC実行

The screenshot shows the Galaxy web interface. The top navigation bar includes 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', and 'ユーザー'. The main content area is divided into three panels:

- Tools:** A search bar and a list of tool categories including 'GENERAL TEXT TOOLS', 'FASTA/FASTQ', and 'FASTQ Quality Control'. The 'FastQC' tool is selected.
- FastQC Configuration:** Shows 'QC Read Quality reports (Galaxy Version 0.72+galaxy1)'. Below this is a 'Short read data from your current history' section with a list of dataset IDs: 6: SRR6322569_2, 5: SRR6322569_1, 4: SRR6322567_2, 3: SRR6322567_1, 2: SRR6322564_2, 1: SRR6322564_1. A red arrow with the number '1' points to the 'Options' button in the top right of this section.
- History:** Shows a list of datasets under 'GSE107337_3samples'. The datasets are listed with their IDs and icons for viewing, editing, and deleting.

こんな感じになる。①下部に移動して、実行ボタンを探す。②Execute。

W4-3: FastQC実行

The screenshot shows the Galaxy web interface. The main content area displays the configuration for the 'Lower limit on the length of the sequence to be shown in the report' tool. The 'length of Kmer to look for' is set to 7. The 'Execute' button is highlighted with a red arrow and the number 2. The 'History' panel on the right shows a list of datasets with a red arrow and the number 1 pointing to the bottom of the list.

Tools

search tools

Get Data

Send Data

Collection Operations

GENERAL TEXT TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

Datamash

GENOMIC FILE MANIPULATION

FASTA/FASTQ

FASTQ Quality Control

[FastQC](#) Read Quality reports

[Trimmomatic](#) flexible read trimming

Lower limit on the length of the sequence to be shown in the report

As long as you set this to a value greater or equal to your longest read length then this will be the sequence length used to create your read groups. This can be useful for making directly comparable statistics from datasets with somewhat variable read lengths. (--min_length)

length of Kmer to look for

7

note: the Kmer test is disabled and needs to be enabled using a custom Submodule and limits file (--kmers)

Execute

Purpose

FastQC aims to provide a simple way to do some quality control checks on raw sequence data coming from high throughput sequencing pipelines. It provides a set of analyses which you can use to get a quick impression of whether your data has any problems of which you should

History

search datasets

GSE107337_3samples

6 shown

2.15 GB

5: SRR6322569_2

5: SRR6322569_1

4: SRR6322567_2

3: SRR6322567_1

2: SRR6322564_2

1: SRR6322564_1

こんな感じになる。①下部に移動して、実行ボタンを探す。②Execute。ボタンを押した直後。

W4-4: 実行中

The screenshot shows the Galaxy web interface. The browser address bar displays 'https://www.ebi.ac.uk/ena/data/v' and 'Galaxy'. The page URL is 'usegalaxy.org'. The navigation bar includes 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', and 'ユーザー'. The main content area is titled 'Lower limit on the length of the sequence to be shown in the report'. It features a search bar for tools, a description of the tool's function, a slider for 'length of Kmer to look for' set to 7, and a 'Sending...' button. The right sidebar shows the 'History' section with a search bar and a list of datasets under 'GSE107337_3samples', including SRR6322569_2, SRR6322569_1, SRR6322567_2, SRR6322567_1, SRR6322564_2, and SRR6322564_1.

W4-4: 実行中

こんな感じになる。①下部に移動して、実行ボタンを探す。②Execute。ボタンを押した直後。すぐにこんな感じに画面が切り替わる。灰色なので実行待ち状態。

The screenshot shows the Galaxy web interface. The main content area is highlighted in green and contains the following text:

...

It produces 12 outputs:

- 18: FastQC on data 6: RawData
- 17: FastQC on data 6: Webpage
- 16: FastQC on data 5: RawData

...

You can check the status of queued jobs and view the resulting data by refreshing the History panel. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

The left sidebar shows the following tool categories:

- Tools
- Get Data
- Send Data
- Collection Operations
- GENERAL TEXT TOOLS
- Text Manipulation
- Filter and Sort
- Join, Subtract and Group
- Datamash
- GENOMIC FILE MANIPULATION
- FASTA/FASTQ
- FASTQ Quality Control
- FastQC Read Quality reports
- Trimmomatic flexible read trimming

The right sidebar shows the History panel with the following jobs:

- 18: FastQC on data 6: RawData
- 17: FastQC on data 6: Webpage
- 16: FastQC on data 5: RawData
- 15: FastQC on data 5: Webpage
- 14: FastQC on data 4: RawData

W4-4: 実行中

こんな感じになる。①下部に移動して、実行ボタンを探す。②Execute。ボタンを押した直後。すぐにこんな感じに画面が切り替わる。灰色なので実行待ち状態。いくつか実行が始まりました。

The screenshot shows the Galaxy web interface. The main content area is highlighted in green and contains the following text:

...

It produces 12 outputs:

- 18: FastQC on data 6: RawData
- 17: FastQC on data 6: Webpage
- 16: FastQC on data 5: RawData

...

You can check the status of queued jobs and view the resulting data by refreshing the History panel. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

The History panel on the right shows a list of jobs:

- 18: FastQC on data 6: RawData
- 17: FastQC on data 6: Webpage
- 16: FastQC on data 5: RawData (highlighted in orange)
- 15: FastQC on data 5: Webpage (highlighted in orange)
- 14: FastQC on data 4: RawData

W4-4: 実行中

こんな感じになる。①下部に移動して、実行ボタンを探す。②Execute。ボタンを押した直後。すぐにこんな感じに画面が切り替わる。灰色なので実行待ち状態。いくつか実行が始まりました。2つ実行完了しました。

The screenshot shows the Galaxy web interface. The top navigation bar includes 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', 'ユーザー', and a 'Using 0%' indicator. The left sidebar contains a 'Tools' section with a search bar and categories like 'Get Data', 'Send Data', 'Collection Operations', 'GENERAL TEXT TOOLS', 'Text Manipulation', 'Filter and Sort', 'Join, Subtract and Group', 'Datamash', 'GENOMIC FILE MANIPULATION', 'FASTA/FASTQ', and 'FASTQ Quality Control'. The main content area is greyed out and overlaid with a green box containing the following text: 'It produces 12 outputs: 18: FastQC on data 6: RawData 17: FastQC on data 6: Webpage 16: FastQC on data 5: RawData ... You can check the status of queued jobs and view the resulting data by refreshing the History panel. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.' The right sidebar shows the 'History' panel with a search bar and a list of jobs. The jobs are: '18: FastQC on data 6: RawData', '17: FastQC on data 6: Webpage', '16: FastQC on data 5: Raw Data', '15: FastQC on data 5: We bpage', and '14: FastQC on data 4: RawData'. The jobs 18 and 17 are highlighted in orange, and jobs 16 and 15 are highlighted in green.

W4-5: 実行完了

こんな感じになる。①下部に移動して、実行ボタンを探す。②Execute。ボタンを押した直後。すぐにこんな感じに画面が切り替わる。灰色なので実行待ち状態。いくつか実行が始まりました。2つ実行完了しました。全部完了したようです。このときは約5分。

The screenshot shows the Galaxy web interface. The main panel displays the output of a job, indicating that it produces 12 outputs. The outputs listed are:

- 18: FastQC on data 6: RawData
- 17: FastQC on data 6: Webpage
- 16: FastQC on data 5: RawData

The History panel on the right shows a list of jobs with their status and size. The jobs listed are:

- 18: FastQC on data 6: Raw Data (2.16 GB)
- 17: FastQC on data 6: Webpage
- 16: FastQC on data 5: Raw Data
- 15: FastQC on data 5: Webpage
- 14: FastQC on data 4: Raw Data

The left sidebar contains navigation menus for Tools, Get Data, Send Data, Collection Operations, GENERAL TEXT TOOLS, Text Manipulation, Filter and Sort, Join, Subtract and Group, Datamash, GENOMIC FILE MANIPULATION, FASTA/FASTQ, and FASTQ Quality Control.

①中央のパネルを一番上まで移動させただけ。

W4-5: 実行完了

The screenshot shows the Galaxy web interface at <https://www.ebi.ac.uk/ena/data/>. The main content area displays a successful execution of the **FastQC** tool, which has added 6 jobs to the queue. The tool uses 6 inputs, including SRR6322569_2, SRR6322569_1, and SRR6322567_2. It produces 12 outputs, including 'FastQC on data 6: RawData' and 'FastQC on data 6: Webpage'. The right-hand panel shows the **History** of datasets, with the top entry being 'GSE107337_3samples' (2.16 GB). A red arrow with the number 1 points to the top of the history panel, indicating the step described in the text above.

W4-6: 解説

①FastQCの入力は、履歴番号1-6のFASTQファイル。②出力は計12個。例えば画面上で見えている履歴17と18は、③data 6(つまり履歴6)の実行結果であることがわかる。④履歴17はhtmlファイル版で、⑤履歴18は生データ版だと読み解く。

The screenshot shows the Galaxy web interface. The main panel displays a job execution summary for FastQC. The summary states: "Executed **FastQC** and successfully added 6 jobs to the queue." It lists 6 inputs: "6: SRR6322569_2", "5: SRR6322569_1", "4: SRR6322567_2", and "...". It also lists 12 outputs: "18: FastQC on data 6: RawData" and "17: FastQC on data 6: Webpage". The history panel on the right shows a list of datasets for "GSE107337_3samples", with 18 shown. The history list includes: "18: FastQC on data 6: Raw Data", "17: FastQC on data 6: Webpage", "16: FastQC on data 5: Raw Data", "15: FastQC on data 5: Webpage", and "14: FastQC on data 4: Raw Data". Red arrows and numbers 1-5 are overlaid on the image to highlight specific elements: 1 points to the input list, 2 points to the output list, 3 points to the input "6", 4 points to job 17, and 5 points to job 18.

①を例として、Webpageのほうを保存。第11回W10-4と同じです。

W4-7: htmlを保存

The screenshot shows the Galaxy web interface. The top navigation bar includes the Galaxy logo and menu items: データ解析, ワークフロー, 可視化する, 共有データ, ヘルプ, ユーザー. The main content area is divided into three panels:

- Tools:** A sidebar on the left with a search bar and categories like Get Data, Send Data, Collection Operations, GENERAL TEXT TOOLS, Text Manipulation, Filter and Sort, Join, Subtract and Group, Datamash, GENOMIC FILE MANIPULATION, FASTA/FASTQ, and FASTQ Quality Control.
- Job Execution:** A central green panel with a checkmark icon. It displays: "Executed **FastQC** and successfully added 6 jobs to the queue." Below this, it lists 6 inputs: "6: SRR6322569_2", "5: SRR6322569_1", "4: SRR6322567_2", and "...". It also lists 12 outputs: "18: FastQC on data 6: RawData" and "17: FastQC on data 6: Webpage".
- History:** A panel on the right showing a list of datasets. The top dataset is "GSE107337_3samples" (18 shown, 2.16 GB). Below it, a list of jobs is shown, including "18: FastQC on data 6: Raw Data", "17: FastQC on data 6: Webpage", "16: FastQC on data 5: Raw Data", "15: FastQC on data 5: Webpage", and "14: FastQC on data 4: Raw Data". A red arrow with the number "1" points to the "17: FastQC on data 6: Webpage" entry.

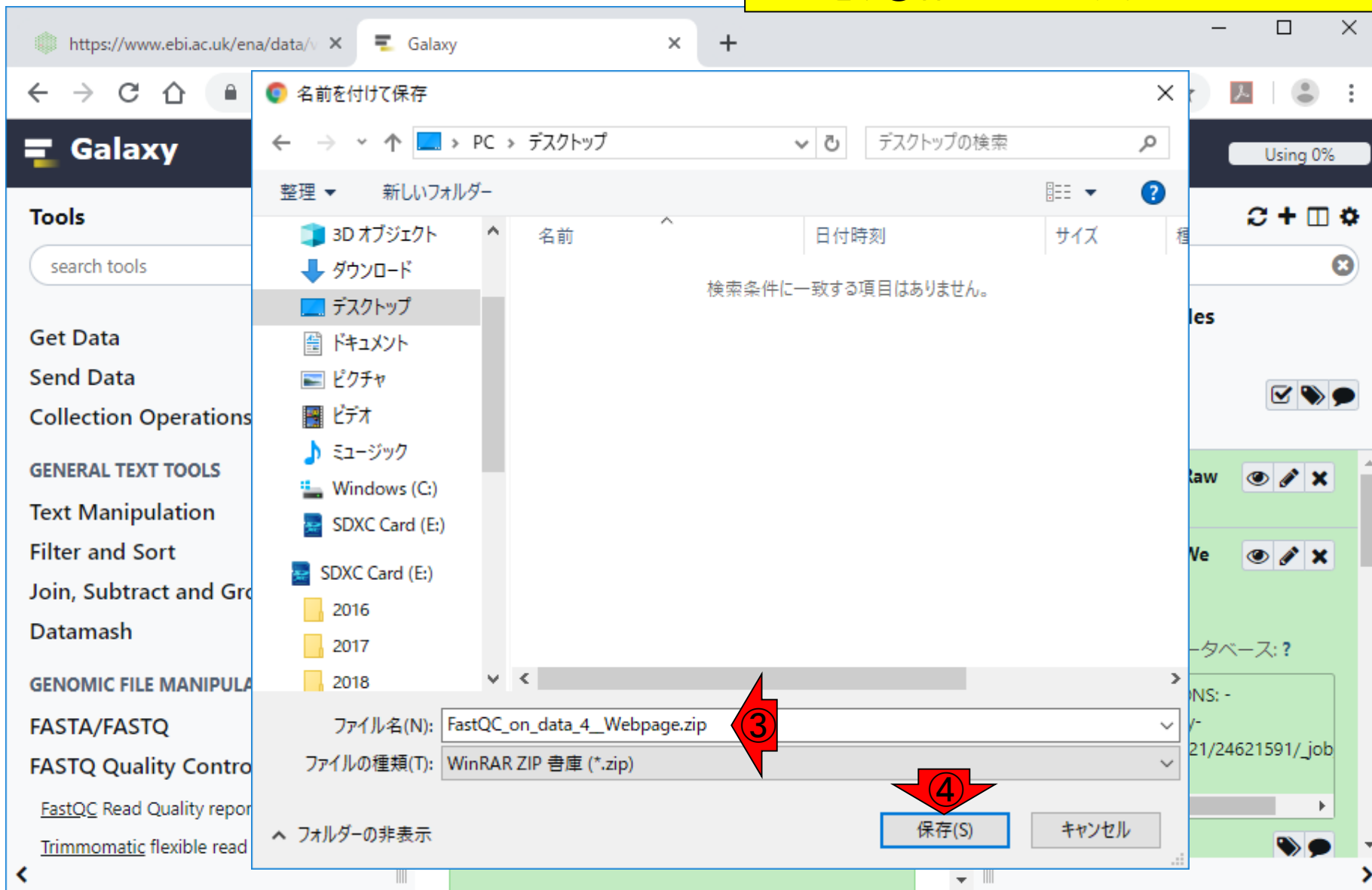
①を例として、Webpageのほうを保存。第11回W10-4と同じです。②を押す。

W4-7: htmlを保存

The screenshot shows the Galaxy web interface. The main content area displays a green notification box with a checkmark, indicating that the **FastQC** tool was executed successfully and 6 jobs were added to the queue. The notification lists the tool's inputs: 6: SRR6322569_2, 5: SRR6322569_1, and 4: SRR6322567_2. It also lists the outputs: 18: FastQC on data 6: RawData and 17: FastQC on data 6: Webpage. The right-hand panel shows the **History** section with a search bar and a list of datasets. The dataset **GSE107337_3samples** is selected, showing 18 datasets shown and 2.16 GB of data. The selected dataset is expanded to show the details of the **17: FastQC on data 6: Webpage** job, which is 621.3 KB in size and has a format of **html**. A red arrow with the number 2 points to the save icon (a floppy disk) in the bottom toolbar of the job details panel.

W4-7:htmlを保存

①を例として、Webpageのほうを保存。第11回W10-4と同じです。②を押す。③のような感じのzip圧縮ファイルを、④保存できます。



①を押す。

W4-8: 元に戻す

The screenshot shows the Galaxy web interface. The top navigation bar includes 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', and 'ユーザー'. The main content area is divided into three panels:

- Tools:** A sidebar on the left with a search bar and categories like 'Get Data', 'Send Data', 'Collection Operations', 'GENERAL TEXT TOOLS', 'Text Manipulation', 'Filter and Sort', 'Join, Subtract and Group', 'Datamash', 'GENOMIC FILE MANIPULATION', 'FASTA/FASTQ', 'FASTQ Quality Control', and 'Trimmomatic flexible read trimming'.
- Job Execution:** A central green panel with a checkmark icon. It displays: 'Executed **FastQC** and successfully added 6 jobs to the queue.' Below this, it lists 6 inputs: '6: SRR6322569_2', '5: SRR6322569_1', '4: SRR6322567_2', and '...'. It also lists 12 outputs: '18: FastQC on data 6: RawData' and '17: FastQC on data 6: Webpage'.
- History:** A panel on the right with a search bar. It shows a dataset 'GSE107337_3samples' (18 shown, 2.16 GB). Below it, a job '17: FastQC on data 6: Webpage' is highlighted in green. A red arrow with the number '1' inside points to the job's icon. Below the job name, it shows '621.3 KB' and 'フォーマット: html, データベース: ?'. A code block contains: 'Picked up _JAVA_OPTIONS: -Djava.io.tmpdir=/galaxy-repl/main/jobdir/024/621/24621591/_job -Xmx7g -Xms256m'.

①を押す。②を押す。

W4-8: 元に戻す

The screenshot shows the Galaxy web interface. The top navigation bar includes 'データ解析' (Data Analysis), 'ワークフロー' (Workflow), '可視化する' (Visualize), '共有データ' (Shared Data), 'ヘルプ' (Help), and 'ユーザー' (User). A red arrow labeled '2' points to the 'データ解析' menu item. The main content area displays a green notification box with a checkmark, stating: 'Executed **FastQC** and successfully added 6 jobs to the queue. The tool uses 6 inputs: 6: SRR6322569_2, 5: SRR6322569_1, 4: SRR6322567_2, ... It produces 12 outputs: 18: FastQC on data 6: RawData, 17: FastQC on data 6: Webpage'. The right sidebar shows the 'History' panel with a search bar and a list of datasets, including 'GSE107337_3samples' and several 'FastQC on data' entries.

①を押す。②を押す。このあたりはただの趣味なので馬鹿正直にトレースしなくてもよいw。

W4-8: 元に戻す



The screenshot shows the Galaxy web interface. The top navigation bar includes 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', and 'ユーザー'. The main content area displays a search for 'FastQC' with a list of history items:

- 18: FastQC on data 6: Raw Data
- 17: FastQC on data 6: We bpage
- 16: FastQC on data 5: Raw Data
- 15: FastQC on data 5: We bpage
- 14: FastQC on data 4: Raw Data

A central banner for 'Galaxy Help' is visible, with the text 'Got Questions? Get Answers.' and the URL 'help.galaxyproject.org'. The bottom of the page shows a tweet by @galaxyproject.

Contents

- W1: 公共データベースENA
- W2: クラウド解析環境Galaxy
- W3: ENAからGalaxyへのFASTQファイルの送付
- W4: クオリティチェック (FastQC)
- W5: クオリティコントロール (Trimmomatic)
- W6: クオリティチェック (FastQC)
- W7: ゲノム配列へのマッピング (Bowtie2)
- W8: カウント情報取得 (htseq-count)
- W9: カウント情報の連結 (Column join)
- W10: カウントデータの数値行列の行名について
- W11: 公共カウントデータ (GSE107337) を用いて、倍率変化と \log_2 (倍率変化)を算出
- W12: Galaxy上で得られたカウントデータについても同様に計算し、比較
- W13とW14: 原著論文中の記載事項を確認
- W15: 2つのカウントデータから共通遺伝子の情報を用いて類似度を計算

W5-1 : Trimmomatic

次は、Trimmomatic。画面左側のツール選択パネル上で、①少し下部に移動したのち、②FASTQ Quality Control。

The screenshot shows the Galaxy web interface. On the left, the 'Tools' panel is visible, listing various genomic tools. A red arrow labeled '2' points to 'FASTQ Quality Control' in the 'FASTQ/FASTQ' category. Another red arrow labeled '1' points to the scroll bar in the Tools panel. The central content area features a banner for '125+ ways to use Galaxy' with a grid of tool thumbnails. The right-hand 'History' panel shows a list of recent jobs, including 'FastQC on data 6: Raw Data' and 'FastQC on data 6: We bpage'. The browser address bar shows 'usegalaxy.org'.

W5-1 : Trimmomatic

次は、Trimmomatic。画面左側のツール選択パネル上で、①少し下部に移動したのち、②FASTQ Quality Control。③Trimmomaticを選択。

The screenshot shows the Galaxy web interface. The browser address bar displays 'usegalaxy.org'. The main navigation bar includes 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', and 'ユーザー'. The 'Tools' panel on the left is expanded to 'FASTQ Quality Control', where 'Trimmomatic flexible read trimming tool for Illumina NGS data' is highlighted with a red arrow and the number 3. The central workspace contains a 'Galaxy 101' tutorial banner. The 'History' panel on the right shows a list of jobs, including 'FastQC on data 6: Raw Data' through 'FastQC on data 4: Raw Data'.

W5-2: paired-end

The screenshot shows the Galaxy web interface with the Trimmomatic tool configuration. The browser address bar shows <https://www.ebi.ac.uk/ena/data/> and the Galaxy logo. The main navigation bar includes "データ解析", "ワークフロー", "可視化する", "共有データ", "ヘルプ", and "ユーザー". The "Tools" sidebar on the left lists categories like "GENOMIC FILE MANIPULATION", "FASTA/FASTQ", "FASTQ Quality Control", "SAM/BAM", and "BED".

The Trimmomatic tool configuration is as follows:

- Trimmomatic flexible read trimming tool for Illumina NGS data (Galaxy Version 0.36.6)**
- Single-end or paired-end reads?**: Single-end
- Input FASTQ file**: 6: SRR6322569_2
- Perform initial ILLUMINACLIP step?**: No
- Trimmomatic Operation**: 1: Trimmomatic Operation
- Select Trimmomatic operation to perform**: Sliding window trimming (SLIDINGWINDOW)

The History panel on the right shows a list of datasets for "GSE107337_3samples" (2.16 GB). The history items are:

- 18: FastQC on data 6: Raw Data
- 17: FastQC on data 6: We bpage
- 16: FastQC on data 5: Raw Data
- 15: FastQC on data 5: We bpage
- 14: FastQC on data 4: Raw Data

わかりづらいが、①Trimmomaticと書いています。
Paired-endなので、②のところを変更する。

W5-2: paired-end

The screenshot shows the Galaxy web interface with the Trimmomatic tool configuration page. The tool is titled "Trimmomatic flexible read trimming tool for Illumina NGS data (Galaxy Version 0.36.6)". The configuration includes a dropdown menu for "Single-end or paired-end reads?" currently set to "Single-end", an "Input FASTQ file" field with a selected dataset "6: SRR6322569_2", and a "Perform initial ILLUMINACLIP step?" section with "Yes" and "No" buttons. The "Trimmomatic Operation" section shows "1: Trimmomatic Operation" and a dropdown for "Select Trimmomatic operation to perform" set to "Sliding window trimming (SLIDINGWINDOW)". The right sidebar shows a "History" panel with a list of datasets, including "18: FastQC on data 6: Raw Data", "17: FastQC on data 6: We bpage", "16: FastQC on data 5: Raw Data", "15: FastQC on data 5: We bpage", and "14: FastQC on data 4: Raw Data".

W5-2: paired-end

わかりづらいが、①Trimmomaticと書いています。Paired-endなので、②のところを変更する。今取り扱っているのはforward側とreverse側に2分割されたものなので、③を選択する。

The screenshot shows the Galaxy web interface with the Trimmomatic tool selected. The tool name is "Trimmomatic flexible read trimming tool for Illumina NGS data (Galaxy Version 0.36.6)". The configuration is as follows:

- Single-end or paired-end reads?**: A dropdown menu is open, showing "Single-end", "Paired-end (two separate input files)" (highlighted in blue), and "Paired-end (as collection)".
- Trimmomatic Operation**: "1: Trimmomatic Operation" is selected.
- Select Trimmomatic operation to perform**: "Sliding window trimming (SLIDINGWINDOW)" is selected.

Red arrows with numbers 1, 2, and 3 point to the tool name, the dropdown menu, and the selected "Paired-end (two separate input files)" option, respectively.

W5-2: paired-end

The screenshot shows the Galaxy web interface with the Trimmomatic tool configuration page. The browser address bar shows <https://www.ebi.ac.uk/ena/data/> and the Galaxy logo. The main navigation bar includes 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', and 'ユーザー'. The 'Tools' sidebar on the left lists categories like 'GENOMIC FILE MANIPULATION', 'FASTA/FASTQ', and 'FASTQ Quality Control'. The main content area shows the Trimmomatic tool configuration for 'mmomatic flexible read trimming tool for Illumina NGS data (Galaxy Version 0.36.6)'. The 'Single-end or paired-end reads?' dropdown is set to 'Paired-end (two separate input files)'. Two 'Input FASTQ file' fields are filled with '6: SRR6322569_2'. The 'Perform initial ILLUMINACLIP step?' section has 'Yes' selected. The 'History' sidebar on the right shows a list of datasets, including '18: FastQC on data 6: Raw Data' and '17: FastQC on data 6: We bpage'.

W5-2: paired-end

こんな感じになる。①の部分でforward側(R1側)のファイル群を、そして②の部分でreverse側(R2側)のファイル群をそれぞれ指定する。作業のノリはW4-2とほぼ同じ。まずは①のforward側から。

The screenshot shows the Galaxy web interface for the Trimmomatic tool. The tool is configured for paired-end reads. The 'Input FASTQ file (R1/first of pair)' field is highlighted with a red arrow and the number 1. The 'Input FASTQ file (R2/second of pair)' field is highlighted with a red arrow and the number 2. The 'History' panel on the right shows a list of datasets, with the top two entries highlighted in green.

Tools

search tools

GENOMIC FILE MANIPULATION

FASTA/FASTQ

FASTQ Quality Control

[FastQC Read Quality reports](#)

[Trimmomatic flexible read trimming tool for Illumina NGS data](#)

[MultiQC aggregate results from bioinformatics analyses into a single report](#)

[FASTQ Summary Statistics by column](#)

[Compute quality statistics](#)

[Draw nucleotides distribution chart](#)

[Draw quality score boxplot](#)

SAM/BAM

BED

Trimmomatic

Trimmomatic flexible read trimming tool for Illumina NGS data (Galaxy Version 0.36.6)

Single-end or paired-end reads?

Paired-end (two separate input files)

Input FASTQ file (R1/first of pair)

6: SRR6322569_2

Input FASTQ file (R2/second of pair)

6: SRR6322569_2

Perform initial ILLUMINACLIP step?

Yes No

Cut adapter and other illumina-specific sequences from the read

History

search datasets

GSE107337_3samples

18 shown

2.16 GB

18: FastQC on data 6: Raw Data

17: FastQC on data 6: We bpage

16: FastQC on data 5: Raw Data

15: FastQC on data 5: We bpage

14: FastQC on data 4: Raw Data

W5-3: forward側

The screenshot displays the Galaxy web interface for configuring the Trimmomatic tool. The main panel shows the tool's configuration options, including input file selection. A red arrow points to the 'Input FASTQ file (R1/first of pair)' field, which is set to '6: SRR6322569_2'. A tooltip labeled 'Multiple datasets' is visible over the input field. The right sidebar shows a history of jobs, including 'FastQC on data 6: Raw Data' and 'FastQC on data 6: We bpage'.

W5-3: forward側

The screenshot shows the Galaxy web interface with the Trimmomatic tool selected. The browser address bar shows <https://www.ebi.ac.uk/ena/data/> and the Galaxy logo. The main navigation bar includes 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', and 'ユーザー'. The 'Tools' sidebar on the left lists various genomic file manipulation tools. The main workspace shows the Trimmomatic tool configuration with the following details:

- Trimmomatic** flexible read trimming tool for Illumina NGS data (Galaxy Version 0.36.6)
- Options: Favorite, Versions, Options
- Single-end or paired-end reads?**: Paired-end (two separate input files)
- Input FASTQ file (R1/first of pair)**:
 - 6: SRR6322569_2
 - 5: SRR6322569_1
 - 4: SRR6322567_2
 - 3: SRR6322567_1
 - 2: SRR6322564_2
 - 1: SRR6322564_1
- A red arrow labeled '2' points to the 'Browse Datasets' button next to the input field.
- Below the input field, it states: "This is a batch mode input field. Separate jobs will be triggered for each dataset selection."
- Input FASTQ file (R2/second of pair)**: 6: SRR6322569_2

The right sidebar shows the **History** panel with a search bar and a list of datasets:

- GSE107337_3samples** (2.16 GB)
- 18 shown
- 18: FastQC on data 6: Raw Data
- 17: FastQC on data 6: We bpage
- 16: FastQC on data 5: Raw Data
- 15: FastQC on data 5: We bpage
- 14: FastQC on data 4: Raw Data

W5-3: forward側

こんな感じになる。履歴の最新のものから表示されるようですね。入力として用いたいのは forward側のFASTQファイルなので、①下部に移動。

The screenshot shows the Galaxy web interface with a search modal open. The modal contains a search bar and a table of results. A red arrow with the number 1 points to the bottom of the table, indicating the user should scroll down to find the correct FASTQ file.

Label	Details	Time
18: FastQC on data 6: RawData	txt	2019-08-16 04:01
17: FastQC on data 6: Webpage	html	2019-08-16 04:01
16: FastQC on data 5: RawData	txt	2019-08-16 04:01
15: FastQC on data 5: Webpage	html	2019-08-16 04:01
14: FastQC on data 4: RawData	txt	2019-08-16 04:02
13: FastQC on data 4: Webpage	html	2019-08-16 04:02
12: FastQC on data 3: RawData	txt	2019-08-16 04:00
11: FastQC on data 3: Webpage	html	2019-08-16 04:00

W5-3: forward側

こんな感じになる。ヒストリーの最新のものから表示されるようですね。入力として用いたいのは forward側のFASTQファイルなので、①下部に移動。

The screenshot shows the Galaxy web interface with a search modal open. The modal contains a search bar and a list of results. A red arrow with the number 1 points to the bottom of the list, indicating the need to scroll down to find the correct file.

File Name	Extension	Date
9: FastQC on data 2: Webpage	html	2019-08-16 04:00
8: FastQC on data 1: RawData	txt	2019-08-16 04:00
7: FastQC on data 1: Webpage	html	2019-08-16 04:00
6: SRR6322569_2	fastqsanger.gz	2019-08-16 03:59
5: SRR6322569_1	fastqsanger.gz	2019-08-16 03:59
4: SRR6322567_2	fastqsanger.gz	2019-08-16 03:59
3: SRR6322567_1	fastqsanger.gz	2019-08-16 03:59
2: SRR6322564_2	fastqsanger.gz	2019-08-16 03:59
1: SRR6322564_1	fastqsanger.gz	2019-08-16 03:59

W5-3: forward側

こんな感じになる。履歴の最新のものから表示されるようです。入力として用いたいのはforward側のFASTQファイルなので、①下部に移動。forward側に相当するのは、②履歴名の最後が_1となっているものたちなので、それをクリックして、③Ok。

The screenshot shows the Galaxy web interface with a search results dialog box open. The dialog box has a search bar at the top and a list of search results below. The results are sorted by date, with the most recent at the top. The file names in the list are: 9: FastQC on data 2: Webpage (html), 8: FastQC on data 1: RawData (txt), 7: FastQC on data 1: Webpage (html), 6: SRR6322569_2 (fastqsanger.gz), 5: SRR6322569_1 (fastqsanger.gz), 4: SRR6322567_2 (fastqsanger.gz), 3: SRR6322567_1 (fastqsanger.gz), 2: SRR6322564_2 (fastqsanger.gz), and 1: SRR6322564_1 (fastqsanger.gz). Red arrows labeled ①, ②, and ③ point to the file names ending in '_1', the 'Ok' button, and the 'Ok' button respectively.

File Name	Extension	Date
9: FastQC on data 2: Webpage	html	2019-08-16 04:00
8: FastQC on data 1: RawData	txt	2019-08-16 04:00
7: FastQC on data 1: Webpage	html	2019-08-16 04:00
6: SRR6322569_2	fastqsanger.gz	2019-08-16 03:59
5: SRR6322569_1	fastqsanger.gz	2019-08-16 03:59
4: SRR6322567_2	fastqsanger.gz	2019-08-16 03:59
3: SRR6322567_1	fastqsanger.gz	2019-08-16 03:59
2: SRR6322564_2	fastqsanger.gz	2019-08-16 03:59
1: SRR6322564_1	fastqsanger.gz	2019-08-16 03:59

W5-3: forward側

こんな感じになる。次はreverse側なので、①中央パネルを少し下部に移動。

The screenshot shows the Galaxy web interface. The main panel displays the configuration for the **mmomatic** tool, which is used for flexible read trimming of Illumina NGS data. The tool version is 0.36.6. The configuration includes a dropdown menu for "Single-end or paired-end reads?" set to "Paired-end (two separate input files)". Below this, there are two input fields for FASTQ files. The first field, "Input FASTQ file (R1/first of pair)", contains a list of six dataset IDs: 6: SRR6322569_2, 5: SRR6322569_1, 4: SRR6322567_2, 3: SRR6322567_1, 2: SRR6322564_2, and 1: SRR6322564_1. A note below the list states: "This is a batch mode input field. Separate jobs will be triggered for each dataset selection." The second input field, "Input FASTQ file (R2/second of pair)", is currently empty. On the right side, the "History" panel is visible, showing a list of recent jobs. A red arrow with the number "1" points to the "search datasets" input field in the History panel. The History panel lists several jobs, including "18: FastQC on data 6: Raw Data", "17: FastQC on data 6: We bpage", "16: FastQC on data 5: Raw Data", "15: FastQC on data 5: We bpage", and "14: FastQC on data 4: Raw Data".

こんな感じになる。次はreverse側なので、①中央パネルを少し下部に移動。こんな感じ。

W5-4: reverse側

The screenshot shows the Galaxy web interface. The top navigation bar includes 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', and 'ユーザー'. The main content area is divided into three panels:

- Tools Panel (Left):** Contains a search bar and a list of tools under categories like 'GENOMIC FILE MANIPULATION', 'FASTA/FASTQ', and 'FASTQ Quality Control'. Tools listed include 'FastQC Read Quality reports', 'Trimmomatic flexible read trimming tool for Illumina NGS data', 'MultiQC aggregate results from bioinformatics analyses into a single report', 'FASTQ Summary Statistics by column', 'Compute quality statistics', 'Draw nucleotides distribution chart', and 'Draw quality score boxplot'.
- Input Panel (Center):** Shows the configuration for the Trimmomatic tool. It includes an 'Input FASTQ file (R2/second of pair)' field with a dropdown menu set to '6: SRR6322569_2'. Below this is a 'Perform initial ILLUMINACLIP step?' section with 'Yes' and 'No' buttons. The 'Cut adapter and other illumina-specific sequences from the read' checkbox is checked. The 'Trimmomatic Operation' section includes a dropdown for 'Select Trimmomatic operation to perform' (set to 'Sliding window trimming (SLIDINGWINDOW)'), a text input for 'Number of bases to average across' (set to '4'), and another text input for 'Average quality required' (set to '20'). A '+ Insert Trimmomatic Operation' button is at the bottom.
- History Panel (Right):** Shows a list of datasets under the heading 'GSE107337_3samples'. The list includes items like '18: FastQC on data 6: Raw Data', '17: FastQC on data 6: We bpage', '16: FastQC on data 5: Raw Data', '15: FastQC on data 5: We bpage', and '14: FastQC on data 4: Raw Data'. Each item has icons for viewing, editing, and deleting.

A red arrow with the number '1' points to the vertical scrollbar of the central panel, indicating that it should be moved down to adjust the view.

W5-4: reverse側

The screenshot shows the Galaxy web interface. The browser address bar displays <https://www.ebi.ac.uk/ena/data/> and the page title is 'Galaxy'. The main navigation bar includes 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', and 'ユーザー'. The 'Tools' sidebar on the left lists categories like 'GENOMIC FILE MANIPULATION', 'FASTA/FASTQ', and 'FASTQ Quality Control'. The main tool configuration area is for 'Trimmomatic Operation'. The input field is labeled 'Input: R1/first of pair (R2/second of pair)' and contains '6: SRR6322569_2'. A red arrow points to the 'Multiple datasets' button. Below the input field, there is a question 'Perform initial ILLUMINACLIP step?' with 'Yes' and 'No' buttons. The 'Trimmomatic Operation' section includes a dropdown for 'Select Trimmomatic operation to perform' (set to 'Sliding window trimming (SLIDINGWINDOW)'), a text input for 'Number of bases to average across' (set to '4'), and another text input for 'Average quality required' (set to '20'). A '+ Insert Trimmomatic Operation' button is at the bottom. The right sidebar shows a 'History' section with a search bar and a list of datasets, including '18: FastQC on data 6: Raw Data', '17: FastQC on data 6: We bpage', '16: FastQC on data 5: Raw Data', '15: FastQC on data 5: We bpage', and '14: FastQC on data 4: Raw Data'.

W5-4: reverse側

The screenshot shows the Galaxy web interface at <https://www.ebi.ac.uk/ena/data/>. The main content area is the 'Input FASTQ file (R2/second of pair)' tool configuration page. A red arrow with the number '2' points to the 'Browse Datasets' button. The input field contains a list of dataset IDs: 6: SRR6322569_2, 5: SRR6322569_1, 4: SRR6322567_2, 3: SRR6322567_1, 2: SRR6322564_2, and 1: SRR6322564_1. Below the input field, there is a note: 'This is a batch mode input field. Separate jobs will be triggered for each dataset selection.' The 'Perform initial ILLUMINACLIP step?' section has 'Yes' and 'No' buttons. The 'Trimmomatic Operation' section has a dropdown menu set to 'Sliding window trimming (SLIDINGWINDOW)'. The right sidebar shows the 'History' panel with a search bar and a list of jobs, including '18: FastQC on data 6: Raw Data', '17: FastQC on data 6: We bpage', '16: FastQC on data 5: Raw Data', '15: FastQC on data 5: We bpage', and '14: FastQC on data 4: Raw Data'. The top navigation bar includes 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', and 'ユーザー'.

入力として用いたいのはreverse側のFASTQファイルなので、①下部に移動。

W5-4: reverse側

The screenshot shows the Galaxy web interface with a search modal window open. The modal window contains a search bar and a table of search results. A red arrow with the number 1 points to the bottom of the table, indicating the reverse side FASTQ files.

Label	Details	Time
18: FastQC on data 6: RawData	txt	2019-08-16 04:01
17: FastQC on data 6: Webpage	html	2019-08-16 04:01
16: FastQC on data 5: RawData	txt	2019-08-16 04:01
15: FastQC on data 5: Webpage	html	2019-08-16 04:01
14: FastQC on data 4: RawData	txt	2019-08-16 04:02
13: FastQC on data 4: Webpage	html	2019-08-16 04:02
12: FastQC on data 3: RawData	txt	2019-08-16 04:00
11: FastQC on data 3: Webpage	html	2019-08-16 04:00

入力として用いたいのはreverse側のFASTQファイルなので、①下部に移動。移動後。

W5-4: reverse側

The screenshot shows the Galaxy web interface with a search results dialog box open. The dialog box contains a search bar and a list of search results. A red arrow with the number 1 points to the entry '2: SRR6322564_2' in the list.

File ID	File Name	Format	Date
9	FastQC on data 2: Webpage	html	2019-08-16 04:00
8	FastQC on data 1: RawData	txt	2019-08-16 04:00
7	FastQC on data 1: Webpage	html	2019-08-16 04:00
6	SRR6322569_2	fastqsanger.gz	2019-08-16 03:59
5	SRR6322569_1	fastqsanger.gz	2019-08-16 03:59
4	SRR6322567_2	fastqsanger.gz	2019-08-16 03:59
3	SRR6322567_1	fastqsanger.gz	2019-08-16 03:59
2	SRR6322564_2	fastqsanger.gz	2019-08-16 03:59
1	SRR6322564_1	fastqsanger.gz	2019-08-16 03:59

W5-4: reverse側

入力として用いたいのはreverse側のFASTQファイルなので、①下部に移動。移動後、reverse側に相当するのは、②履歴名の最後が_2となっているものたちなので、それをクリックして、③Ok。

The screenshot shows the Galaxy web interface with a search results dialog box open. The dialog box has a search bar at the top and a table of results below. The results are as follows:

File Name	Extension	Date
9: FastQC on data 2: Webpage	html	2019-08-16 04:00
8: FastQC on data 1: RawData	txt	2019-08-16 04:00
7: FastQC on data 1: Webpage	html	2019-08-16 04:00
6: SRR6322569_2	fastqsanger.gz	2019-08-16 03:59
5: SRR6322569_1	fastqsanger.gz	2019-08-16 03:59
4: SRR6322567_2	fastqsanger.gz	2019-08-16 03:59
3: SRR6322567_1	fastqsanger.gz	2019-08-16 03:59
2: SRR6322564_2	fastqsanger.gz	2019-08-16 03:59
1: SRR6322564_1	fastqsanger.gz	2019-08-16 03:59

Three items (6, 4, and 2) are highlighted in green and marked with a red circle containing the number 2. A red arrow with the number 3 points to the 'Ok' button at the bottom right of the dialog box.

W5-4: reverse側

The screenshot displays the Galaxy web interface. The top navigation bar includes 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', and 'ユーザー'. The main content area is divided into three panels:

- Tools Panel:** A search bar and a list of tools under the category 'FASTQ Quality Control'. Tools listed include 'FastQC Read Quality reports', 'Trimmomatic flexible read trimming tool for Illumina NGS data', 'MultiQC aggregate results from bioinformatics analyses into a single report', 'FASTQ Summary Statistics by column', 'Compute quality statistics', and 'Draw nucleotides distribution chart'.
- Input Panel:** Titled 'Input FASTQ file (R2/second of pair)'. It contains a list of six dataset IDs: 6: SRR6322569_2, 5: SRR6322569_1, 4: SRR6322567_2, 3: SRR6322567_1, 2: SRR6322564_2, and 1: SRR6322564_1. Below the list is a note: 'This is a batch mode input field. Separate jobs will be triggered for each dataset selection.' There is a checkbox for 'Perform initial ILLUMINACLIP step?' with 'Yes' and 'No' options. Below that is the text 'Cut adapter and other illumina-specific sequences from the read'.
- Trimmomatic Operation Panel:** A dropdown menu labeled 'Trimmomatic Operation' with '1: Trimmomatic Operation' selected. Below it is a section 'Select Trimmomatic operation to perform' with a dropdown menu currently showing 'Sliding window trimming (SLIDINGWINDOW)'.
- History Panel:** Titled 'History', it shows a list of previous jobs. The most recent jobs are highlighted in green: '18: FastQC on data 6: Raw Data', '17: FastQC on data 6: We bpage', '16: FastQC on data 5: Raw Data', '15: FastQC on data 5: We bpage', and '14: FastQC on data 4: Raw Data'.

W5-5: アダプター除去

入力データはIlluminaのRNA-seqリードなので、そのアダプター配列除去も行ってもらいたい。デフォルトはNoなので、①Yesに変更。

The screenshot shows the Galaxy web interface with the Trimmomatic tool selected. The 'Input FASTQ file (R2/second of pair)' field contains a list of six datasets: 6: SRR6322569_2, 5: SRR6322569_1, 4: SRR6322567_2, 3: SRR6322567_1, 2: SRR6322564_2, and 1: SRR6322564_1. Below the input field, a note states: "This is a batch mode input field. Separate jobs will be triggered for each dataset selection." The 'Perform initial ILLUMINACLIP step?' checkbox is checked, and a red arrow with a circled '1' points to it. The 'Trimmomatic Operation' dropdown is set to 'Sliding window trimming (SLIDINGWINDOW)'. The 'History' panel on the right shows a list of jobs, including 'FastQC on data 6: Raw Data' and 'FastQC on data 6: We bpage'.

W5-5: アダプター除去

入力データはIlluminaのRNA-seqリードなので、そのアダプター配列除去も行ってもらいたい。デフォルトはNoなので、①Yesに変更。変更後の状態。

The screenshot shows the Galaxy web interface. The main panel displays the 'Input FASTQ file (R2/second of pair)' section with a list of six SRR6322569_2 files. Below this, a dialog box asks 'Perform initial ILLUMINACLIP step?' with 'Yes' and 'No' buttons. A red arrow points to the 'Yes' button, which has a circled '1' next to it. The dialog also includes the text 'Cut adapter and other illumina-specific sequences from the read' and a dropdown menu for 'Select standard adapter sequences or provide custom?' set to 'Standard'. The right-hand side shows a 'History' panel with a search bar and a list of jobs, including '18: FastQC on data 6: Raw Data', '17: FastQC on data 6: We bpage', '16: FastQC on data 5: Raw Data', '15: FastQC on data 5: We bpage', and '14: FastQC on data 4: Raw Data'.

W5-5: アダプター除去

入力データはIlluminaのRNA-seqリードなので、そのアダプター配列除去も行ってもらいたい。デフォルトはNoなので、①Yesに変更。変更後の状態。①が上部にくるように、少しページ下部に移動

The screenshot shows the Galaxy web interface. The main content area displays the configuration for the 'Perform initial ILLUMINACLIP step?' tool. A red arrow with a circled '1' points to the 'Yes' button. The tool settings are as follows:

- Perform initial ILLUMINACLIP step? Yes No
- Cut adapter and other illumina-specific sequences from the read
- Select standard adapter sequences or provide custom?
- Adapter sequences to use
- Maximum mismatch count which will still allow a full match to be performed
- How accurate the match between the two 'adapter ligated' reads must be for PE palindrome read alignment

The right sidebar shows a history of jobs, with the top job highlighted in green:

- 18: FastQC on data 6: Raw Data
- 17: FastQC on data 6: We bpage
- 16: FastQC on data 5: Raw Data
- 15: FastQC on data 5: We bpage
- 14: FastQC on data 4: Raw Data

W5-6: 配列を指定

アダプター配列のデフォルトは①single-endのTruSeq2になっている。

The screenshot shows the Galaxy web interface. The main panel is titled "Perform initial ILLUMINACLIP step?". It has a "Yes" button selected. Below that, it asks to "Cut adapter and other illumina-specific sequences from the read". The "Select standard adapter sequences or provide custom?" dropdown is set to "Standard". The "Adapter sequences to use" dropdown is set to "TruSeq2 (single-ended, for Illumina GAII)", which is highlighted with a red arrow and a circled "1". Below that, there is a field for "Maximum mismatch count which will still allow a full match to be performed" set to "2". At the bottom, there is a field for "How accurate the match between the two 'adapter ligated' reads must be for PE palindrome read alignment" set to "30".

The "History" panel on the right shows a list of datasets under "GSE107337_3samples". The top three items are highlighted in green:

- 18: FastQC on data 6: Raw Data
- 17: FastQC on data 6: We bpage
- 16: FastQC on data 5: Raw Data

W5-6: 配列を指定

アダプター配列のデフォルトは①single-endのTruSeq2になっている。入力データはMiSeqのpaired-endだと分かっているので、②を選択。

The screenshot shows the Galaxy web interface. The main panel is titled "Perform initial ILLUMINACLIP step?". It has "Yes" and "No" buttons. Below that, it says "Cut adapter and other illumina-specific sequences from the read". There is a section "Select standard adapter sequences or provide custom?" with a dropdown menu set to "Standard". Below that is "Adapter sequences to use" with a dropdown menu. The dropdown menu is open, showing several options. The option "TruSeq3 (paired-ended, for MiSeq and HiSeq)" is highlighted in blue, and a red arrow with the number "2" points to it. Other options include "TruSeq2 (single-ended, for Illumina GAII)", "TruSeq2 (single-ended, for Illumina GAII)", "TruSeq3 (single-ended, for MiSeq and HiSeq)", "TruSeq2 (paired-ended, for Illumina GAII)", "TruSeq3 (additional seqs) (paired-ended, for MiSeq and HiSeq)", and "Nextera (paired-ended)". On the right side, there is a "History" panel with a search bar and a list of datasets. The dataset "18: FastQC on data 6: Raw Data" is highlighted in green. The "Tools" panel on the left shows a search bar and a list of tools under "GENOMIC FILE MANIPULATION", "FASTA/FASTQ", "FASTQ Quality Control", "SAM/BAM", and "BED".

W5-6: 補足説明

補足説明。①ENAに戻っています。②Illumina MiSeqの、③paired-endデータです。例えば④のリンク先を眺めると、バージョン番号までは分かりませんがTruSeqが用いられていることがわかります。

https://www.ebi.ac.uk/ena/data/ Galaxy
European Bioinformatics Institute [GB] | ebi.ac.uk/ena/data/view

PRJNA419802	SAMN08098215	SRS2714083	SRX3422362	SRR6322563	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098214	SRS2714082	SRX3422363	SRR6322564	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098213	SRS2714084	SRX3422364	SRR6322565	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098212	SRS2714085	SRX3422365	SRR6322566	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098211	SRS2714086	SRX3422366	SRR6322567	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098210	SRS2714087	SRX3422367	SRR6322568	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098218	SRS2714088	SRX3422368	SRR6322569	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2
PRJNA419802	SAMN08098217	SRS2714089	SRX3422369	SRR6322570	568703	Lactobacillus rhamnosus GG	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2

W5-6: 配列を指定

①無事変更されたことを確認。実際問題としては、どれを選択していいかわかりかねる場合があります。しかし、とりあえずTrimmomaticをやってみた結果を、もう一度FastQCにかけて、無事アダプター配列除去が行われていれば実用上は問題ないはずです。

The screenshot shows the Galaxy web interface. The main panel is titled "Perform initial ILLUMINACLIP step?". It has a "Yes" button selected. Below that, it asks to "Cut adapter and other illumina-specific sequences from the read". The "Select standard adapter sequences or provide custom?" dropdown is set to "Standard". The "Adapter sequences to use" dropdown is set to "TruSeq3 (paired-ended, for MiSeq and HiSeq)", which is highlighted with a red arrow and a circled "1". The "Maximum mismatch count which will still allow a full match to be performed" is set to 2. The "How accurate the match between the two 'adapter ligated' reads must be for PE palindrome read alignment" is set to 30. The right sidebar shows a history of jobs, with the most recent ones highlighted in green.

W5-7: 実行

残りのオプションは特に変更せず、①下部に移動して、②実行。

The screenshot shows the Galaxy web interface for the Trimmomatic tool. The main panel displays the tool configuration with the following options:

- Average quality required:** 20
- + Insert Trimmomatic Operation** (button)
- Output trimlog file?** Yes No
- Output trimmomatic log messages?** Yes No
- Execute** (button, highlighted with a red arrow and the number 2)

The right-hand panel shows the **History** section with a search bar and a list of datasets. The top dataset, **18: FastQC on data 6: Raw Data**, is highlighted with a green background and a red arrow and the number 1.

W5-8: 実行中

The screenshot shows the Galaxy web interface with the following components:

- Navigation Bar:** Includes the Galaxy logo, menu items like "データ解析", "ワークフロー", "可視化する", "共有データ", "ヘルプ", "ユーザー", and a "Using 0%" indicator.
- Tools Panel (Left):** Contains a search bar and categories such as "GENOMIC FILE MANIPULATION", "FASTA/FASTQ", "FASTQ Quality Control", "SAM/BAM", and "BED".
- Job Status Panel (Center):** A green box containing the text: "It produces 12 outputs:" followed by a list of jobs:
 - 30: Trimmomatic on SRR6322569_2 (R2 unpaired)
 - 29: Trimmomatic on SRR6322569_1 (R1 unpaired)
 - 28: Trimmomatic on SRR6322569_2 (R2 paired)
 - ...
 Below the list, it states: "You can check the status of queued jobs and view the resulting data by refreshing the History panel. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered."
- History Panel (Right):** Shows a search bar for datasets and a list of recent jobs:
 - GSE107337_3samples (2.16 GB)
 - 30: Trimmomatic on SRR6322569_2 (R2 unpaired)
 - 29: Trimmomatic on SRR6322569_1 (R1 unpaired)
 - 28: Trimmomatic on SRR6322569_2 (R2 paired)
 - 27: Trimmomatic on SRR6322569_1 (R1 paired)
 - 26: Trimmomatic on SRR6322569_1 (R1 paired)

W5-8: 実行中

https://www.ebi.ac.uk/ena/data/v Galaxy

usegalaxy.org

Galaxy データ解析 ワークフロー 可視化する 共有データ ヘルプ ユーザー Using 0%

Tools

search tools

GENOMIC FILE MANIPULATION

FASTA/FASTQ

FASTQ Quality Control

[FastQC](#) Read Quality reports

[Trimmomatic](#) flexible read trimming tool for Illumina NGS data

[MultiQC](#) aggregate results from bioinformatics analyses into a single report

[FASTQ Summary Statistics](#) by column

[Compute quality statistics](#)

[Draw nucleotides distribution chart](#)

[Draw quality score boxplot](#)

SAM/BAM

BED

History

search datasets

GSE107337_3samples

30 shown

2.16 GB

30: Trimmomatic on SRR6322569_2 (R2 unpaired)

29: Trimmomatic on SRR6322569_1 (R1 unpaired)

28: Trimmomatic on SRR6322569_2 (R2 paired)

27: Trimmomatic on SRR6322569_1 (R1 paired)

26: Trimmomatic on SRR6322569_1 (R1 unpaired)

Executed **Trimmomatic** and successfully added 3 jobs to the queue.

The tool uses 6 inputs:

6: SRR6322569_2

5: SRR6322569_1

4: SRR6322567_2

...

It produces 12 outputs:

30: Trimmomatic on SRR6322569_2 (R2 unpaired)

29: Trimmomatic on SRR6322569_1 (R1 unpaired)

28: Trimmomatic on SRR6322569_2 (R2 paired)

27: Trimmomatic on SRR6322569_1 (R1 paired)

26: Trimmomatic on SRR6322569_1 (R1 unpaired)

W5-9: 実行完了

無事実行完了したようです。このときは約10分かかりました。①サイズが3.47 GBに増えていますね。

The screenshot shows the Galaxy web interface. The main panel displays a green notification box with a checkmark, stating: "Executed **Trimmomatic** and successfully added 3 jobs to the queue." Below this, it lists the tool's inputs: 6: SRR6322569_2, 5: SRR6322569_1, and 4: SRR6322567_2. It also lists 12 outputs, including 30: Trimmomatic on SRR6322569_2 (R2 unpaired) and 29: Trimmomatic on SRR6322569_1 (R1 unpaired). The right-hand History panel shows a search for "GSE107337_3samples" with 30 datasets shown. A red arrow points to the "3.47 GB" size of the top dataset, which is labeled with a circled "1".

W5-10: 解説

① Trimmomaticの入力は、ヒストリー1から6のFASTQファイル。②出力は計12個。ヒストリー19から30に相当します。

The screenshot shows the Galaxy web interface. The main content area displays a green box with a checkmark indicating a successful Trimmomatic job. The text inside the box reads: "Executed Trimmomatic and successfully added 3 jobs to the queue." Below this, it lists the inputs: "The tool uses 6 inputs:" followed by "6: SRR6322569_2", "5: SRR6322569_1", and "4: SRR6322567_2". A red arrow labeled "1" points to these input lines. Below the inputs, it says "It produces 12 outputs:" followed by "30: Trimmomatic on SRR6322569_2 (R2 unpaired)", "29: Trimmomatic on SRR6322569_1 (R1 unpaired)", and "26: Trimmomatic on SRR6322567_2 (R2 unpaired)". A red arrow labeled "2" points to these output lines. On the right side, the "History" panel shows a list of datasets for "GSE107337_3samples", with 30 shown. The list includes entries for Trimmomatic on SRR6322569_2 (R2 unpaired), SRR6322569_1 (R1 unpaired), SRR6322569_2 (R2 paired), SRR6322569_1 (R1 paired), and SRR6322567_2 (R2 unpaired). The left sidebar shows the "Tools" menu with categories like "GENOMIC FILE MANIPULATION", "FASTA/FASTQ", "FASTQ Quality Control", "SAM/BAM", and "BED".

W5-10: 解説

例えば、①SRR6322569の入力はヒストリー5と6、②出力はヒストリー27から30。このうち、その後の解析に用いるのは通常、③ペアで生き残ったリードからなるヒストリー27と28のデータ。unpairedではなくpairedのほうだということ。

The screenshot displays the Galaxy web interface. The central panel shows a successful execution of the Trimmomatic tool. The input list includes:

- 6: SRR6322569_2 (highlighted with a red box and arrow ①)
- 5: SRR6322569_1 (highlighted with a red box and arrow ①)
- 4: SRR6322567_2

The output list includes:

- 30: Trimmomatic on SRR6322569_2 (R2 unpaired)
- 29: Trimmomatic on SRR6322569_1 (R1 unpaired)
- 28: Trimmomatic on SRR6322569_2 (R2 paired)
- 27: Trimmomatic on SRR6322569_1 (R1 paired)
- 26: Trimmomatic on SRR6322567_2 (R2 unpaired)

The History panel on the right shows the job history for 'GSE107337_3samples'. A red box and arrow ② highlight the outputs 27 and 28. A red box and arrow ③ highlight the outputs 27 and 28, indicating they are the paired reads used for subsequent analysis.

Contents

- W1: 公共データベースENA
- W2: クラウド解析環境Galaxy
- W3: ENAからGalaxyへのFASTQファイルの送付
- W4: クオリティチェック (FastQC)
- W5: クオリティコントロール (Trimmomatic)
- W6: クオリティチェック (FastQC)
- W7: ゲノム配列へのマッピング (Bowtie2)
- W8: カウント情報取得 (htseq-count)
- W9: カウント情報の連結 (Column join)
- W10: カウントデータの数値行列の行名について
- W11: 公共カウントデータ (GSE107337) を用いて、倍率変化と \log_2 (倍率変化)を算出
- W12: Galaxy上で得られたカウントデータについても同様に計算し、比較
- W13とW14: 原著論文中の記載事項を確認
- W15: 2つのカウントデータから共通遺伝子の情報を用いて類似度を計算

W6-1 : FastQC

Trimmomatic実行結果の妥当性を確認すべく、再度①FastQCを実行。第11回のW13とノリは同じ。

The screenshot shows the Galaxy web interface. The top navigation bar includes 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', and 'ユーザー'. The 'Tools' panel on the left is expanded to 'FASTQ Quality Control', with 'FastQC Read Quality reports' highlighted by a red lightning bolt icon with the number 1. The main workspace displays a green success message: 'Executed Trimmomatic and successfully added 3 jobs to the queue.' Below this, it lists 6 inputs and 12 outputs. The 'History' panel on the right shows a list of jobs, including '30: Trimmomatic on SRR6 322569_2 (R2 unpaired)', '29: Trimmomatic on SRR6 322569_1 (R1 unpaired)', '28: Trimmomatic on SRR6 322569_2 (R2 paired)', '27: Trimmomatic on SRR6 322569_1 (R1 paired)', and '26: Trimmomatic on SRR6 322567_2 (R2 unpaired)'. Each job entry has icons for viewing, editing, and deleting.

こんな感じになる。

W6-1 : FastQC

The screenshot shows the Galaxy web interface. The top navigation bar includes the Galaxy logo, a search bar, and various utility buttons. The main content area is divided into three panels:

- Tools Panel (Left):** Lists various tools under categories like "GENOMIC FILE MANIPULATION", "FASTA/FASTQ", and "FASTQ Quality Control". The "FastQC Read Quality reports" tool is highlighted.
- Tool Interface (Center):** Shows the configuration for the "FastQC Read Quality reports" tool. It includes options for "Short read data from your current history" (set to "30: Trimmomatic"), "Contaminant list" (set to "No tabular dataset"), "Adapter list" (set to "No tabular dataset"), and "Submodule and Limit specifying file" (set to "Nothing selected").
- History Panel (Right):** Displays a list of datasets from the current history. The datasets are listed with their IDs and descriptions, such as "30: Trimmomatic on SRR6 322569_2 (R2 unpaired)".

こんな感じになる。①Multiple datasets。

W6-2: 複数個指定

The screenshot shows the Galaxy web interface. The main tool panel is for 'FASTQ Quality Control'. The 'Multiple datasets' dropdown is highlighted with a red arrow and a circled '1'. The 'History' panel on the right shows a list of datasets, including '30: Trimmomatic on SRR6 322569_2 (R2 unpaired)', '29: Trimmomatic on SRR6 322569_1 (R1 unpaired)', '28: Trimmomatic on SRR6 322569_2 (R2 paired)', '27: Trimmomatic on SRR6 322569_1 (R1 paired)', and '26: Trimmomatic on SRR6 322567_2 (R2 unpaired)'.

W6-2: 複数個指定

こんな感じになる。①Multiple datasets。②Browse Datasets。

The screenshot shows the Galaxy web interface. The main content area displays the results of a 'Trimmomatic' job. A red arrow points to the 'Browse Datasets' button, which is highlighted with a red circle containing the number 1. The interface shows a list of datasets from a previous job, including '30: Trimmomatic on SRR6 322569_2 (R2 unpaired)' and others. The 'History' panel on the right shows the job details for 'GSE107337_3samples'.

W6-2: 複数個指定

こんな感じになる。①Multiple datasets。②Browse Datasets。こんな感じになる。実行したいのはTrimmomatic実行後のpairedのデータのみなので…

The screenshot shows the Galaxy web interface. A search modal window is open, displaying a table of Trimmomatic jobs. The table has three columns: Label, Details, and Time. The jobs listed are:

Label	Details	Time
30: Trimmomatic on SRR6322569_2 (R2 unpaired)	fastqsanger.gz	2019-08-16 07:34
29: Trimmomatic on SRR6322569_1 (R1 unpaired)	fastqsanger.gz	2019-08-16 07:34
28: Trimmomatic on SRR6322569_2 (R2 paired)	fastqsanger.gz	2019-08-16 07:34
27: Trimmomatic on SRR6322569_1 (R1 paired)	fastqsanger.gz	2019-08-16 07:34
26: Trimmomatic on SRR6322567_2 (R2 unpaired)	fastqsanger.gz	2019-08-16 07:30
25: Trimmomatic on SRR6322567_1 (R1 unpaired)	fastqsanger.gz	2019-08-16 07:30
24: Trimmomatic on SRR6322567_2 (R2 paired)	fastqsanger.gz	2019-08-16 07:30
23: Trimmomatic on SRR6322567_1 (R1 paired)	fastqsanger.gz	2019-08-16 07:30

The background shows the Galaxy dashboard with various tool categories and a dataset viewer. The dataset viewer shows a tabular dataset with 2 columns: name and sequence. The sequence shown is CAAGCAGAAGACGGCATACGA.

W6-2: 複数個指定

こんな感じになる。① Multiple datasets。② Browse Datasets。こんな感じになる。実行したいのは Trimmomatic 実行後の paired のデータのみなので、① これら4つと...

usegalaxy.org

Type to Search

Label	Details	Time
30: Trimmomatic on SRR6322569_2 (R2 unpaired)	fastqsanger.gz	2019-08-16 07:34
29: Trimmomatic on SRR6322569_1 (R1 unpaired)	fastqsanger.gz	2019-08-16 07:34
28: Trimmomatic on SRR6322569_2 (R2 paired)	fastqsanger.gz	2019-08-16 07:34
27: Trimmomatic on SRR6322569_1 (R1 paired)	fastqsanger.gz	2019-08-16 07:34
26: Trimmomatic on SRR6322567_2 (R2 unpaired)	fastqsanger.gz	2019-08-16 07:30
25: Trimmomatic on SRR6322567_1 (R1 unpaired)	fastqsanger.gz	2019-08-16 07:30
24: Trimmomatic on SRR6322567_2 (R2 paired)	fastqsanger.gz	2019-08-16 07:30
23: Trimmomatic on SRR6322567_1 (R1 paired)	fastqsanger.gz	2019-08-16 07:30

Cancel Ok

No tabular dataset

tab delimited file with 2 columns: name and sequence. For example: Illumina Small RNA RT Primer
CAAGCAGAAGACGGCATACGA

W6-2: 複数個指定

こんな感じになる。①Multiple datasets。②Browse Datasets。こんな感じになる。実行したいのはTrimmomatic実行後のpairedのデータのみなので、①これら4つと、②少し下に移動して、③残りの2つを選択して、④Ok。

The screenshot shows the Galaxy web interface with a search results dialog box. The dialog has a search bar and a list of datasets. The datasets are as follows:

ID	Description	File Name	Date
28	Trimmomatic on SRR6322569_2 (R2 paired)	fastqsanger.gz	2019-08-16 07:34
27	Trimmomatic on SRR6322569_1 (R1 paired)	fastqsanger.gz	2019-08-16 07:34
26	Trimmomatic on SRR6322567_2 (R2 unpaired)	fastqsanger.gz	2019-08-16 07:30
25	Trimmomatic on SRR6322567_1 (R1 unpaired)	fastqsanger.gz	2019-08-16 07:30
24	Trimmomatic on SRR6322567_2 (R2 paired)	fastqsanger.gz	2019-08-16 07:30
23	Trimmomatic on SRR6322567_1 (R1 paired)	fastqsanger.gz	2019-08-16 07:30
22	Trimmomatic on SRR6322564_2 (R2 unpaired)	fastqsanger.gz	2019-08-16 07:29
21	Trimmomatic on SRR6322564_1 (R1 unpaired)	fastqsanger.gz	2019-08-16 07:29
20	Trimmomatic on SRR6322564_2 (R2 paired)	fastqsanger.gz	2019-08-16 07:29
19	Trimmomatic on SRR6322564_1 (R1 paired)	fastqsanger.gz	2019-08-16 07:29

The dialog also has 'Cancel' and 'Ok' buttons at the bottom right.

こんな感じになる。①下部に移動して、実行ボタンを探す。

W6-3: FastQC実行

The screenshot shows the Galaxy web interface. On the left, the 'Tools' panel lists 'FASTQ Quality Control' with 'FastQC Read Quality reports' selected. The main workspace shows the 'FastQC Read Quality reports (Galaxy Version 0.72+galaxy1)' tool interface. Below the tool name, there is a section for 'Short read data from your current history' with a list of datasets (25: Trimmomatic on SR to 19: Trimmomatic on SR). A note below this list states: 'This is a batch mode input field. Separate jobs will be triggered for each dataset selection.' At the bottom, there is a 'Contaminant list' section with a dropdown menu set to 'No tabular dataset' and a text area containing 'tab delimited file with 2 columns: name and sequence. For example: Illumina Small RNA RT Primer CAAGCAGAAGACGGCATACGA'. On the right, the 'History' panel shows a search for 'GSE107337_3samples' with 30 results shown, totaling 3.47 GB. A red arrow labeled '1' points to the 'search datasets' input field in the History panel.

こんな感じになる。①下部に移動して、実行ボタンを探す。②Execute。

W6-3: FastQC実行

The screenshot shows the Galaxy web interface. On the left, the 'Tools' sidebar is expanded to 'FASTQ Quality Control', with 'FastQC Read Quality reports' selected. The main panel shows the configuration for 'Lower limit on the length of the sequence to be shown in the report' (empty input) and 'length of Kmer to look for' (set to 7). A red arrow labeled '1' points to the 'Execute' button. A second red arrow labeled '2' points to the 'Execute' button. The right sidebar shows the 'History' panel with a list of datasets, including '30: Trimmomatic on SRR6 322569_2 (R2 unpaired)', '29: Trimmomatic on SRR6 322569_1 (R1 unpaired)', '28: Trimmomatic on SRR6 322569_2 (R2 paired)', '27: Trimmomatic on SRR6 322569_1 (R1 paired)', and '26: Trimmomatic on SRR6 322567_2 (R2 unpaired)'.

W6-4: 実行中

The screenshot shows the Galaxy web interface with the following components:

- Browser:** Address bar shows `https://www.ebi.ac.uk/ena/data/v` and `usegalaxy.org`.
- Navigation:** Galaxy logo and menu items: データ解析, ワークフロー, 可視化する, 共有データ, ヘルプ, ユーザー. A "Using 1%" indicator is visible.
- Tools Panel (Left):** Search tools input. Categories include GENOMIC FILE MANIPULATION, FASTA/FASTQ, FASTQ Quality Control, and SAM/BAM. Specific tools listed include [FastQC](#), [Trimmomatic](#), [MultiQC](#), [FASTQ Summary Statistics](#), [Compute quality statistics](#), [Draw nucleotides distribution chart](#), and [Draw quality score boxplot](#).
- Job Status Panel (Center):**
 - Green checkmark icon: Executed **FastQC** and successfully added 6 jobs to the queue.
 - The tool uses 6 inputs:
 - 28: Trimmomatic on SRR6322569_2 (R2 paired)
 - 27: Trimmomatic on SRR6322569_1 (R1 paired)
 - 24: Trimmomatic on SRR6322567_2 (R2 paired)
 - ...
 - It produces 12 outputs:
 - 42: FastQC on data 28: RawData
- History Panel (Right):**
 - Search datasets input.
 - Dataset: **GSE107337_3samples** (42 shown, 3.47 GB).
 - Job list:
 - 42: FastQC on data 2 (8: RawData)
 - 41: FastQC on data 2 (8: Webpage)
 - 40: FastQC on data 2 (7: RawData)
 - 39: FastQC on data 2 (7: Webpage)
 - 38: FastQC on data 2 (4: RawData)

W6-5: 実行完了

https://www.ebi.ac.uk/ena/data/v Galaxy

usegalaxy.org

Galaxy データ解析 ワークフロー 可視化する 共有データ ヘルプ ユーザー Using 1%

Tools

search tools

GENOMIC FILE MANIPULATION

FASTA/FASTQ

FASTQ Quality Control

[FastQC](#) Read Quality reports

[Trimmomatic](#) flexible read trimming tool for Illumina NGS data

[MultiQC](#) aggregate results from bioinformatics analyses into a single report

[FASTQ Summary Statistics](#) by column

[Compute quality statistics](#)

[Draw nucleotides distribution chart](#)

[Draw quality score boxplot](#)

SAM/BAM

BED

History

search datasets

GSE107337_3samples

42 shown

3.48 GB

42: FastQC on data 28: RawData

41: FastQC on data 28: Webpage

40: FastQC on data 27: RawData

39: FastQC on data 27: Webpage

38: FastQC on data 24: RawData

Executed **FastQC** and successfully added 6 jobs to the queue.

The tool uses 6 inputs:

28: Trimmomatic on SRR6322569_2 (R2 paired)

27: Trimmomatic on SRR6322569_1 (R1 paired)

24: Trimmomatic on SRR6322567_2 (R2 paired)

...

It produces 12 outputs:

42: FastQC on data 28: RawData

W4-7と同様の手順で、①htmlを保存。②中央パネルをリフレッシュ。

W6-6: htmlを保存

The screenshot shows the Galaxy web interface. The browser address bar is <https://www.ebi.ac.uk/ena/data/>. The Galaxy logo is in the top left. The navigation bar includes 'データ解析' (Data Analysis), 'ワークフロー' (Workflow), '可視化する' (Visualize), '共有データ' (Shared Data), 'ヘルプ' (Help), and 'ユーザー' (User). The 'Tools' panel on the left lists various tools under categories like 'GENOMIC FILE MANIPULATION', 'FASTA/FASTQ', and 'FASTQ Quality Control'. The central panel shows a green success message: 'Executed **FastQC** and successfully added 6 jobs to the queue. The tool uses 6 inputs: 28: Trimmomatic on SRR6322569_2 (R2 paired), 27: Trimmomatic on SRR6322569_1 (R1 paired), 24: Trimmomatic on SRR6322567_2 (R2 paired), ... It produces 12 outputs: 42: FastQC on data 28: RawData'. The right panel shows the 'History' section with a search bar and a list of jobs. Two red arrows labeled '1' point to the 'FastQC on data 28: W ebpage' and 'FastQC on data 27: W ebpage' entries. A red arrow labeled '2' points to the 'データ解析' menu item.

Contents

- W1: 公共データベースENA
- W2: クラウド解析環境Galaxy
- W3: ENAからGalaxyへのFASTQファイルの送付
- W4: クオリティチェック (FastQC)
- W5: クオリティコントロール (Trimmomatic)
- W6: クオリティチェック (FastQC)
- W7: ゲノム配列へのマッピング (Bowtie2)
- W8: カウント情報取得 (htseq-count)
- W9: カウント情報の連結 (Column join)
- W10: カウントデータの数値行列の行名について
- W11: 公共カウントデータ (GSE107337) を用いて、倍率変化と \log_2 (倍率変化)を算出
- W12: Galaxy上で得られたカウントデータについても同様に計算し、比較
- W13とW14: 原著論文中の記載事項を確認
- W15: 2つのカウントデータから共通遺伝子の情報を用いて類似度を計算

次は、Bowtie2プログラムを用いてマッピング。①ツール選択パネル下部に移動して、②Mapping。

W7-1 : Bowtie2

The screenshot shows the Galaxy web interface. The top navigation bar includes 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', and 'ユーザー'. The 'Tools' panel on the left lists various genomic analysis tools under 'GENOMICS ANALYSIS' and 'METAGENOMICS'. A red lightning bolt with the number '2' points to the 'Mapping' tool. A red arrow with the number '1' points to the 'Galaxy Help' overlay, which contains the text: 'Galaxy Help Got Questions? Get Answers. help.galaxyproject.org'. The 'History' panel on the right shows a list of datasets, including 'GSE107337_3samples' and several 'FastQC on data' entries.

W7-1 : Bowtie2

次は、Bowtie2プログラムを用いてマッピング。①ツール選択パネル下部に移動して、②Mapping。③Bowtie2。

The screenshot shows the Galaxy web interface. The browser address bar displays 'https://www.ebi.ac.uk/ena/data/v' and 'Galaxy'. The main navigation bar includes 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', and 'ユーザー'. The 'Tools' panel on the left is expanded to 'GENOMICS ANALYSIS', with 'Mapping' selected. The 'Bowtie2 - map reads against reference genome' tool is highlighted, indicated by a red arrow with the number '3'. The central content area displays a 'Galaxy Help' banner with the text 'Got Questions? Get Answers.' and the URL 'help.galaxyproject.org'. The 'History' panel on the right shows a list of datasets, including 'GSE107337_3samples' and several 'FastQC on data' entries.

W7-1 : Bowtie2

次は、Bowtie2プログラムを用いてマッピング。①ツール選択パネル下部に移動して、②Mapping。③Bowtie2。こんな感じになる。

The screenshot shows the Galaxy web interface. The top navigation bar includes the Galaxy logo and tabs for 'データ解析' (Data Analysis), 'ワークフロー' (Workflow), '可視化する' (Visualize), '共有データ' (Shared Data), 'ヘルプ' (Help), and 'ユーザー' (User). The left sidebar contains a 'Tools' section with a search bar and a list of tools under 'GENOMICS ANALYSIS' > 'Mapping'. The main content area displays the configuration for the 'wtie2 - map reads against reference genome' tool. The configuration includes a 'Favorite' button, 'Versions' dropdown, and 'Options' dropdown. The 'Is this single or paired library' dropdown is set to 'Single-end'. The 'FASTA/Q file' section shows a file upload icon, a '30: Trimmomatic' dropdown, and a folder icon. Below this, there are instructions: 'Must be of datatype "fastqsanger" or "fasta"'. There are two sections for writing unaligned and aligned reads to separate files, each with 'Yes' and 'No' buttons. The right sidebar shows the 'History' panel with a search bar and a list of datasets, including 'GSE107337_3samples' and several 'FastQC on data' entries.

①デフォルトはSingle-endなので…。

W7-2: オプション

The screenshot shows the Galaxy web interface for configuring the 'wtie2' tool. The browser address bar shows 'https://www.ebi.ac.uk/ena/data/v' and 'usegalaxy.org'. The Galaxy logo and navigation menu are at the top. The left sidebar lists tool categories: GENOMICS ANALYSIS, Assembly, Annotation, and Mapping. The central area shows the 'wtie2 - map reads against reference genome' tool configuration. A dropdown menu for 'Is this single or paired library' is set to 'Single-end', with a red arrow and a circled '1' pointing to it. Below this, there are options for 'FASTA/Q file' (30: Trimmomatic) and checkboxes for 'Write unaligned reads (in fastq format) to separate file(s)' and 'Write aligned reads (in fastq format) to separate file(s)'. The right sidebar shows a 'History' panel with a search bar and a list of datasets, including '42: FastQC on data 28: RawData', '41: FastQC on data 28: Webpage', '40: FastQC on data 27: RawData', '39: FastQC on data 27: Webpage', and '38: FastQC on data 24: RawData'.

①デフォルトはSingle-endなので…②Paired-endに変更。

W7-2: オプション

The screenshot shows the Galaxy web interface with the 'wtie2' tool configuration page. The 'Is this single or paired library' dropdown menu is open, showing options: Single-end, Paired-end (selected), Paired-end Dataset Collection, and Paired-end data from single interleaved dataset. Red arrows labeled 1 and 2 point to the dropdown and the selected 'Paired-end' option respectively.

Tools

search tools

GENOMICS ANALYSIS

Assembly

Annotation

Mapping

[Bowtie2](#) - map reads against reference genome

[LASTZ](#) : align long sequences

[LASTZ_D](#) : estimate substitution scores matrix

[Map with BWA-MEM](#) - map medium and long reads (> 100 bp) against reference genome

[Map with BWA](#) - map short reads (< 100 bp) against reference genome

[STAR-Fusion](#) detect fusion genes in RNA-Seq data

wtie2 - map reads against reference genome (Galaxy Version 2.3.4.2)

☆ Favorite Versions Options

Is this single or paired library

Single-end

Single-end

Paired-end

Paired-end Dataset Collection

Paired-end data from single interleaved dataset

Yes No

--un/--un-conc (possibly with -gz or -bz2); This triggers --un parameter for single reads and --un-conc for paired reads

Write aligned reads (in fastq format) to separate file(s)

Yes No

History

search datasets

GSE107337_3samples

42 shown

3.48 GB

42: FastQC on data 28: RawData

41: FastQC on data 28: Webpage

40: FastQC on data 27: RawData

39: FastQC on data 27: Webpage

38: FastQC on data 24: RawData

①デフォルトはSingle-endなので…②Paired-endに変更。こんな感じになる。

W7-2: オプション

Galaxy

データ解析 ワークフロー 可視化する 共有データ ヘルプ ユーザー Using 1%

Tools

search tools

GENOMICS ANALYSIS

Assembly

Annotation

Mapping

[Bowtie2](#) - map reads against reference genome

[LASTZ](#) : align long sequences

[LASTZ_D](#) : estimate substitution scores matrix

[Map with BWA-MEM](#) - map medium and long reads (> 100 bp) against reference genome

[Map with BWA](#) - map short reads (< 100 bp) against reference genome

[STAR-Fusion](#) detect fusion genes in RNA-Seq data

wtie2 - map reads against reference genome (Galaxy Version 2.3.4.2)

Favorite Versions Options

Is this single or paired library

Paired-end

FASTA/Q file #1

30: Trimmomatic

Must be of datatype "fastqsanger" or "fasta"

FASTA/Q file #2

30: Trimmomatic

Must be of datatype "fastqsanger" or "fasta"

Write unaligned reads (in fastq format) to separate file(s)

Yes No

--un/--un-conc (possibly with -gz or -bz2); This triggers

History

search datasets

GSE107337_3samples

42 shown

3.48 GB

42: FastQC on data 28: RawData

41: FastQC on data 28: Webpage

40: FastQC on data 27: RawData

39: FastQC on data 27: Webpage

38: FastQC on data 24: RawData

W7-2: オプション

①がpaired-endのforward側(1つめ)、②がreverse側(2つめ)のリードに相当するデータの履歴番号を指定するところなので。1つ1つやっていく。

The screenshot shows the Galaxy web interface for the 'wtie2' tool. The tool configuration is as follows:

- Tool: **wtie2** - map reads against reference genome (Galaxy Version 2.3.4.2)
- Is this single or paired library: Paired-end
- FASTA/Q file #1: [File selection icon] [Copy icon] [Folder icon] 30: Trimmomatic [Folder icon]
- FASTA/Q file #2: [File selection icon] [Copy icon] [Folder icon] 30: Trimmomatic [Folder icon]
- Write unaligned reads (in fastq format) to separate file(s): Yes

The 'History' panel on the right shows a list of datasets for 'GSE107337_3samples' (42 shown, 3.48 GB). The top four datasets are highlighted in green:

- 42: FastQC on data 28: RawData
- 41: FastQC on data 28: Webpage
- 40: FastQC on data 27: RawData
- 39: FastQC on data 27: Webpage

Red boxes and arrows labeled 1 and 2 point to the 'FASTA/Q file #1' and 'FASTA/Q file #2' input fields, respectively, indicating where to specify the dataset IDs from the history panel.

W7-3: forward側

The screenshot shows the Galaxy web interface for the 'wtie2' tool. The browser address bar shows 'https://www.ebi.ac.uk/ena/data/v' and 'usegalaxy.org'. The Galaxy navigation bar includes 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', and 'ユーザー'. The 'Tools' sidebar on the left lists 'GENOMICS ANALYSIS' tools like 'Bowtie2', 'LASTZ', and 'Map with BWA-MEM'. The main tool configuration area for 'wtie2' shows 'Is this single or paired library' set to 'Paired-end'. Under 'FASTA/Q file #1', the 'Multiple datasets' checkbox is checked, indicated by a red arrow with the number '1'. Below it, 'FASTA/Q file #2' is also configured. The 'History' sidebar on the right shows a list of datasets, including 'GSE107337_3samples' and several 'FastQC on data' entries.

W7-3: forward側

The screenshot shows the Galaxy web interface. The main content area displays the configuration for the 'wtie2' tool. A red arrow with the number '2' points to the 'Browse Datasets' button in the FASTA/Q file #1 input field. The 'History' panel on the right shows a list of datasets, including 'GSE107337_3samples' and several 'FastQC on data' entries.

Galaxy

データ解析 ワークフロー 可視化する 共有データ ヘルプ ユーザー Using 1%

Tools

search tools

GENOMICS ANALYSIS

Assembly

Annotation

Mapping

[Bowtie2](#) - map reads against reference genome

[LASTZ](#) : align long sequences

[LASTZ_D](#) : estimate substitution scores matrix

[Map with BWA-MEM](#) - map medium and long reads (> 100 bp) against reference genome

[Map with BWA](#) - map short reads (< 100 bp) against reference genome

[STAR-Fusion](#) detect fusion genes in RNA-Seq data

Favorite Versions Options

wtie2 - map reads against reference genome (Galaxy Version 2.3.4.2)

Is this single or paired library

Paired-end

FASTA/Q file #1

Browse Datasets

30: Trimmomatic on S
29: Trimmomatic on S
28: Trimmomatic on S
27: Trimmomatic on S
26: Trimmomatic on S
25: Trimmomatic on S
24: Trimmomatic on S

This is a batch mode input field. Separate jobs will be triggered for each dataset selection.

Must be of datatype "fastqsanger" or "fasta"

FASTA/Q file #2

History

search datasets

GSE107337_3samples

42 shown

3.48 GB

42: FastQC on data 28: RawData

41: FastQC on data 28: Webpage

40: FastQC on data 27: RawData

39: FastQC on data 27: Webpage

38: FastQC on data 24: RawData

W7-3: forward側

①Multiple datasets。②Browse Datasets。こんな感じになる。③下部に移動して、入力として与えたい Trimmomatic実行後のデータを探す。

Galaxy

usegalaxy.org

Type to Search

Label	Details	Time
42: FastQC on data 28: RawData	txt	2019-08-16 09:08
41: FastQC on data 28: Webpage	html	2019-08-16 09:08
40: FastQC on data 27: RawData	txt	2019-08-16 09:09
39: FastQC on data 27: Webpage	html	2019-08-16 09:09
38: FastQC on data 24: RawData	txt	2019-08-16 09:08
37: FastQC on data 24: Webpage	html	2019-08-16 09:08
36: FastQC on data 23: RawData	txt	2019-08-16 09:08
35: FastQC on data 23: Webpage	html	2019-08-16 09:08

Cancel Ok

FASTA/Q file #2

Must be of datatype "fastqsanger" or "fasta"

ebpage

38: FastQC on data 24: RawData

W7-3: forward側

①Multiple datasets。②Browse Datasets。こんな感じになる。③下部に移動して、入力として与えたい Trimmomatic実行後のデータを探す。④forward側の R1 pairedを選択して、⑤Ok。

ID	Description	File Type	Date
28	Trimmomatic on SRR6322569_2 (R2 paired)	fastqsanger.gz	2019-08-16 09:07
27	Trimmomatic on SRR6322569_1 (R1 paired)	fastqsanger.gz	2019-08-16 09:07
26	Trimmomatic on SRR6322567_2 (R2 unpaired)	fastqsanger.gz	2019-08-16 07:30
25	Trimmomatic on SRR6322567_1 (R1 unpaired)	fastqsanger.gz	2019-08-16 07:30
24	Trimmomatic on SRR6322567_2 (R2 paired)	fastqsanger.gz	2019-08-16 09:06
23	Trimmomatic on SRR6322567_1 (R1 paired)	fastqsanger.gz	2019-08-16 09:06
22	Trimmomatic on SRR6322564_2 (R2 unpaired)	fastqsanger.gz	2019-08-16 07:29
21	Trimmomatic on SRR6322564_1 (R1 unpaired)	fastqsanger.gz	2019-08-16 07:29
20	Trimmomatic on SRR6322564_2 (R2 paired)	fastqsanger.gz	2019-08-16 09:06
19	Trimmomatic on SRR6322564_1 (R1 paired)	fastqsanger.gz	2019-08-16 09:06

W7-3: forward側

①Multiple datasets。②Browse Datasets。こんな感じになる。③下部に移動して、入力として与えたい Trimmomatic実行後のデータを探す。④forward側の R1 pairedを選択して、⑤Ok。こんな感じになる。これでforward側データの指定完了。次は、⑥reverse (R2 paired)側。

The screenshot shows the Galaxy web interface. The main panel displays the configuration for the 'wtie2' tool. The 'Is this single or paired library' dropdown is set to 'Paired-end'. The 'FASTA/Q file #1' field is highlighted with a red arrow and the number 6, indicating the selection of a dataset from the history panel on the right. The history panel shows a list of datasets, including '42: FastQC on data 28: RawData', '41: FastQC on data 28: Webpage', '40: FastQC on data 27: RawData', '39: FastQC on data 27: Webpage', and '38: FastQC on data 24: RawData'. The 'Tools' panel on the left lists various genomics analysis tools, including 'Bowtie2', 'LASTZ', and 'Map with BWA-MEM'.

W7-4: reverse側

①Multiple datasets。②Browse Datasets。こんな感じになる。③下部に移動して、入力として与えたい Trimmomatic実行後のデータを探す。④forward側の R1 pairedを選択して、⑤Ok。こんな感じになる。これでforward側データの指定完了。次は、⑥reverse (R2 paired)側。

The screenshot shows the Galaxy web interface. The browser address bar displays <https://www.ebi.ac.uk/ena/data/> and the Galaxy logo. The main navigation bar includes 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', and 'ユーザー'. The 'Tools' sidebar on the left lists various genomics analysis tools under 'GENOMICS ANALYSIS', including 'Assembly', 'Annotation', and 'Mapping'. The central tool configuration area is for 'Trimmomatic', showing options for 'FASTA/Q file #2' (set to '30: Trimmomatic') and checkboxes for 'Write unaligned reads (in fastq format) to separate file(s)' and 'Write aligned reads (in fastq format) to separate file(s)'. A red arrow points to the 'FASTA/Q file #2' dropdown menu, which is labeled with a circled '6'. The right sidebar shows the 'History' section with a search bar and a list of datasets, including 'GSE107337_3samples' and several 'FastQC on data' entries.

W7-4: reverse側

The screenshot shows the Galaxy web interface for the FASTX-TO-FASTQ tool. The browser address bar shows <https://www.ebi.ac.uk/ena/data/> and the Galaxy logo. The navigation bar includes 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', and 'ユーザー'. The 'Tools' sidebar on the left lists categories like 'GENOMICS ANALYSIS', 'Assembly', 'Annotation', and 'Mapping'. The main tool configuration area shows the 'FASTX-TO-FASTQ' tool with a dropdown menu set to '30: Trimmomatic'. A red arrow points to the 'Multiple datasets' checkbox, which is checked. Below this, there are two sections for writing unaligned and aligned reads to separate files, each with 'Yes' and 'No' buttons. The 'History' sidebar on the right shows a list of datasets, including 'GSE107337_3samples' and several 'FastQC on data' entries.

W7-4: reverse側

The screenshot shows the Galaxy web interface at <https://www.ebi.ac.uk/ena/data/>. The browser address bar shows <https://www.ebi.ac.uk/ena/data/> and the page title is "usegalaxy.org". The Galaxy logo is visible in the top left, and the user is logged in as "Using 1%".

The main content area displays a workflow step titled "FASTA/Q file #2". The input field is empty, and a red arrow with a circled "2" points to the "Browse Datasets" button. Below the input field, there is a list of jobs:

- 30: Trimmomatic on S
- 29: Trimmomatic on S
- 28: Trimmomatic on S
- 27: Trimmomatic on S
- 26: Trimmomatic on S
- 25: Trimmomatic on S
- 24: Trimmomatic on S

Below the list, there is a note: "This is a batch mode input field. Separate jobs will be triggered for each dataset selection." Below this, there is another input field with the same note and a "Write unaligned reads (in fastq format) to separate file(s)" section with "Yes" and "No" buttons. Below that, there is a note: "--un/--un-conc (possibly with -gz or -bz2); This triggers --un parameter for single reads and --un-conc for paired reads".

The right sidebar shows the "History" section with a search bar and a list of jobs:

- GSE107337_3samples (42 shown, 3.48 GB)
- 42: FastQC on data 28: RawData
- 41: FastQC on data 28: Webpage
- 40: FastQC on data 27: RawData
- 39: FastQC on data 27: Webpage
- 38: FastQC on data 24: RawData

The left sidebar shows the "Tools" section with a search bar and a list of tools under "GENOMICS ANALYSIS":

- Assembly
- Annotation
- Mapping
- Bowtie2 - map reads against reference genome
- LASTZ : align long sequences
- LASTZ_D : estimate substitution scores matrix
- Map with BWA-MEM - map medium and long reads (> 100 bp) against reference genome
- Map with BWA - map short reads (< 100 bp) against reference genome
- STAR-Fusion detect fusion genes in RNA-Seq data

W7-4: reverse側

①Multiple datasets. ②Browse Datasets. ③下部に移動して、④reverse側のR2 pairedを選択して、⑤Ok。

The screenshot shows the Galaxy web interface with a search results dialog box. The dialog box contains a search bar and a list of search results. The results are as follows:

ID	Description	Source	Time
28	Trimmomatic on SRR6322569_2 (R2 paired)	fastqsanger.gz	2019-08-16 09:07
27	Trimmomatic on SRR6322569_1 (R1 paired)	fastqsanger.gz	2019-08-16 09:07
26	Trimmomatic on SRR6322567_2 (R2 unpaired)	fastqsanger.gz	2019-08-16 07:30
25	Trimmomatic on SRR6322567_1 (R1 unpaired)	fastqsanger.gz	2019-08-16 07:30
24	Trimmomatic on SRR6322567_2 (R2 paired)	fastqsanger.gz	2019-08-16 09:06
23	Trimmomatic on SRR6322567_1 (R1 paired)	fastqsanger.gz	2019-08-16 09:06
22	Trimmomatic on SRR6322564_2 (R2 unpaired)	fastqsanger.gz	2019-08-16 07:29
21	Trimmomatic on SRR6322564_1 (R1 unpaired)	fastqsanger.gz	2019-08-16 07:29
20	Trimmomatic on SRR6322564_2 (R2 paired)	fastqsanger.gz	2019-08-16 09:06
19	Trimmomatic on SRR6322564_1 (R1 paired)	fastqsanger.gz	2019-08-16 09:06

Red arrows indicate the following actions:

- Arrow 4 points to row 28, row 24, and row 20.
- Arrow 3 points to row 24.
- Arrow 5 points to the 'Ok' button.

W7-4: reverse側

①Multiple datasets. ②Browse Datasets. ③下部に移動して、④reverse側のR2 pairedを選択して、⑤Ok. こんな感じになる。

The screenshot shows the Galaxy web interface. The browser address bar displays <https://www.ebi.ac.uk/ena/data/> and the Galaxy logo. The main navigation bar includes 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', and 'ユーザー'. The 'Tools' sidebar on the left lists 'GENOMICS ANALYSIS' with sub-sections for 'Assembly', 'Annotation', and 'Mapping'. The central workspace shows a workflow step configuration for 'Write unaligned reads (in fastq format) to separate file(s)'. It includes a dropdown menu for 'FASTA/Q file #2' with options 20 through 26, all labeled 'Trimmomatic on S'. Below the dropdown is a note: 'This is a batch mode input field. Separate jobs will be triggered for each dataset selection.' There are 'Yes' and 'No' buttons. The 'History' sidebar on the right shows a search bar and a list of datasets under 'GSE107337_3samples', including '42: FastQC on data 28: RawData', '41: FastQC on data 28: Webpage', '40: FastQC on data 27: RawData', '39: FastQC on data 27: Webpage', and '38: FastQC on data 24: RawData'.

他のオプションとして、①paired-end optionsがNoになっているので…

W7-5: オプション

The screenshot shows the Galaxy web interface for the tool 'Write aligned reads (in fastq format) to separate file(s)'. The configuration options are as follows:

- Do you want to set paired-end options?**: No (highlighted with a red arrow and the number 1)
- Will you select a reference genome from your history or use a built-in index?**: Use a built-in genome index
- Select reference genome**: Baboon (Papio anubis): papHam1

The 'History' panel on the right shows a list of datasets:

- 42: FastQC on data 28: RawData
- 41: FastQC on data 28: Webpage
- 40: FastQC on data 27: RawData
- 39: FastQC on data 27: Webpage
- 38: FastQC on data 24: RawData

他のオプションとして、①paired-end optionsがNoになっているので…②Yesに変更しておく。

W7-5: オプション

https://www.ebi.ac.uk/ena/data/v Galaxy

usegalaxy.org

Galaxy データ解析 ワークフロー 可視化する 共有データ ヘルプ ユーザー Using 1%

Tools

search tools

GENOMICS ANALYSIS

Assembly

Annotation

Mapping

[Bowtie2](#) - map reads against reference genome

[LASTZ](#) : align long sequences

[LASTZ_D](#) : estimate substitution scores matrix

[Map with BWA-MEM](#) - map medium and long reads (> 100 bp) against reference genome

[Map with BWA](#) - map short reads (< 100 bp) against reference genome

[STAR-Fusion](#) detect fusion genes in RNA-Seq data

Write aligned reads (in fastq format) to separate file(s)

Yes No

--al/--al-conc (possibly with -gz or -bz2); This triggers --al parameter for single reads and --al-conc for paired reads

Do you want to set paired-end options?

No ①

No

Yes ②

Use a built-in genome index

Built-ins were indexed using default options. See `indexes` section of help below

Select reference genome

Baboon (Papio anubis): papHam1

If your genome of interest is not listed, contact the Galaxy team

History

search datasets

GSE107337_3samples

42 shown

3.48 GB

42: FastQC on data 28: RawData

41: FastQC on data 28: Webpage

40: FastQC on data 27: RawData

39: FastQC on data 27: Webpage

38: FastQC on data 24: RawData

他のオプションとして、①paired-end optionsがNoになっているので…②Yesに変更しておく。変更後。

W7-5: オプション

The screenshot shows the Galaxy web interface. The main content area displays the configuration for the tool "Write aligned reads (in fastq format) to separate file(s)". The "Do you want to set paired-end options?" dropdown menu is set to "Yes", which is highlighted by a red arrow and a circled "2". Below this, the "Set the minimum fragment length for valid paired-end alignments" field is set to "0". The right sidebar shows the "History" panel with a list of datasets, including "42: FastQC on data 28: RawData", "41: FastQC on data 28: Webpage", "40: FastQC on data 27: RawData", "39: FastQC on data 27: Webpage", and "38: FastQC on data 24: RawData".

W7-6: リファレンス

中央パネル下部に移動して、リファレンスゲノム配列を指定するところ。①デフォルトはGalaxy上で指定可能なもののみ。

The screenshot shows the Galaxy web interface. The central panel displays the configuration for a tool, with a dropdown menu set to "Use a built-in genome index". A red arrow with the number "1" points to this dropdown. Below it, the "Select reference genome" dropdown is set to "Baboon (Papio anubis): papHam1". The "Set read groups information?" dropdown is set to "Do not set". The "Select analysis mode" dropdown is set to "1: Default setting only". The right panel shows the "History" section with a search bar and a list of datasets, including "GSE107337_3samples" and several "FastQC on data" entries.

W7-6: リファレンス

中央パネル下部に移動して、リファレンスゲノム配列を指定するところ。①デフォルトはGalaxy上で指定可能なもののみ。ここでは、Ensembl Bacteria から提供されているASM2650v1 のゲノム(ASM2650v1.fa)を利用する(第13回のW13)。ついでに後に利用するアノテーションファイル(ASM2650v1.gff3)もアップロードしておく。そしてそれをヒストリーから選択できるように…

The screenshot shows the Galaxy web interface. The left sidebar lists tool categories: GENOMICS ANALYSIS, Assembly, Annotation, and Mapping. Under Mapping, several tools are listed, including Bowtie2, LASTZ, LASTZ_D, Map with BWA-MEM, Map with BWA, and STAR-Fusion. The central panel displays the configuration for a tool, with a dropdown menu for 'Select reference genome' set to 'Baboon (Papio anubis): papHam1'. A red arrow with the number '1' points to this dropdown. The right sidebar shows a list of datasets under 'GSE107337_3samples', including '42: FastQC on data 28: RawData', '41: FastQC on data 28: Webpage', '40: FastQC on data 27: RawData', '39: FastQC on data 27: Webpage', and '38: FastQC on data 24: RawData'.

W7-6: リファレンス

中央パネル下部に移動して、リファレンスゲノム配列を指定するところ。①デフォルトはGalaxy上で指定可能なもののみ。ここでは、Ensembl Bacteria から提供されているASM2650v1 のゲノム(ASM2650v1.fa)を利用する(第13回のW13)。ついでに後に利用するアノテーションファイル(ASM2650v1.gff3)もアップロードしておく。そしてそれをヒストリーから選択できるように…②に変更しておく。

The screenshot shows the Galaxy web interface. The top navigation bar includes the URL <https://www.ebi.ac.uk/ena/data/v> and the Galaxy logo. The main content area is divided into three panels. The left panel, titled 'Tools', lists various genomics analysis tools under categories like 'GENOMICS ANALYSIS', 'Assembly', 'Annotation', and 'Mapping'. The central panel displays a configuration form for a tool. The first section, 'Will you select a reference genome from your history or use a built-in index?', has a dropdown menu open. The selected option is 'Use a genome from the history and build index', which is highlighted in blue. A red arrow with the number '2' points to this option. Another red arrow with the number '1' points to the dropdown menu. The right panel shows a list of datasets, including '42: FastQC on data 28: RawData', '41: FastQC on data 28: Webpage', '40: FastQC on data 27: RawData', '39: FastQC on data 27: Webpage', and '38: FastQC on data 24: RawData'. Each dataset entry has a search icon, a pencil icon, and a close icon.

W7-6: リファレンス

中央パネル下部に移動して、リファレンスゲノム配列を指定するところ。①デフォルトはGalaxy上で指定可能なもののみ。ここでは、Ensembl Bacteria から提供されているASM2650v1 のゲノム(ASM2650v1.fa)を利用する(第13回のW13)。ついでに後に利用するアノテーションファイル(ASM2650v1.gff3)もアップロードしておく。そしてそれをヒストリーから選択できるように…②に変更しておく。変更後。

The screenshot shows the Galaxy web interface. The top navigation bar includes the Galaxy logo and menu items like 'データ解析', 'ワークフロー', '可視化する', and '共有'. The left sidebar contains a 'Tools' section with a search bar and a list of tools under 'GENOMICS ANALYSIS', including 'Assembly', 'Annotation', and 'Mapping'. The main panel displays a form for selecting a reference genome. A dropdown menu is set to 'Use a genome from the history and build index', with a red arrow pointing to it and a circled '2'. Below this, there are sections for 'Select reference genome', 'Set read groups information?', 'Select analysis mode', and 'Do you want to use presets?'. The right sidebar shows a search bar for datasets and a list of datasets, including 'GSE107337_3samples' and several 'FastQC on data' entries.

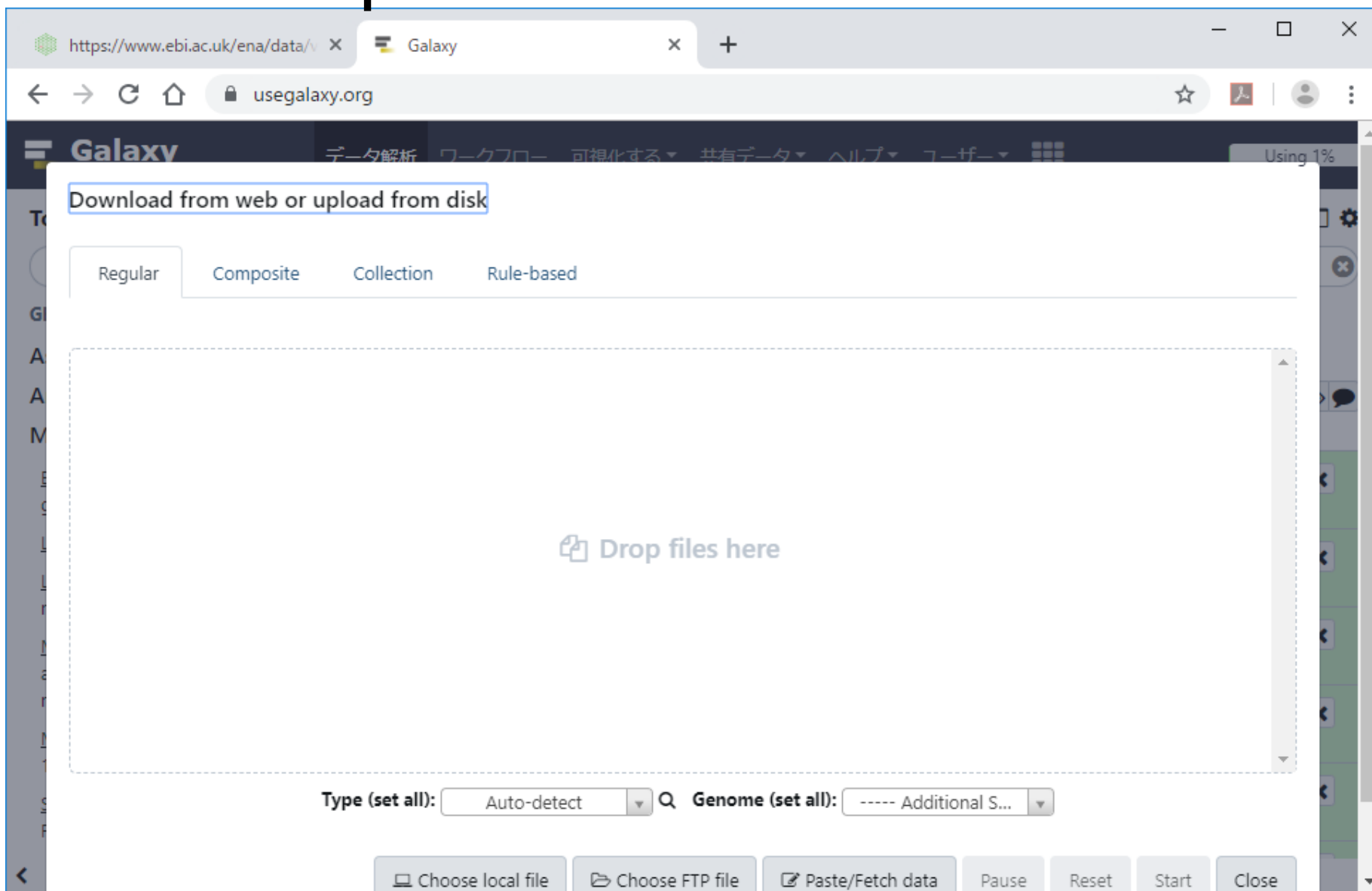
W7-6: リファレンス

まだヒストリーから選択可能な状態ではないので、
①No fasta datasetとなっている。ASM2650v1 のゲノムファイル(ASM2650v1.fa)とアノテーションファイル(ASM2650v1.gff3)をアップロードすべく、②を押す(第11回のW5-3)。

The screenshot shows the Galaxy web interface. The top navigation bar includes 'データ解析' (Data Analysis) with a red arrow labeled '2' pointing to it. The left sidebar contains 'Tools' with a search bar and categories like 'GENOMICS ANALYSIS', 'Assembly', 'Annotation', and 'Mapping'. The central panel shows a tool configuration interface with sections: 'Will you select a reference genome from your history or use a built-in index?' (dropdown: 'Use a genome from the history and build index'), 'Select reference genome' (dropdown: 'No fasta dataset' with a red arrow labeled '1' pointing to it), 'Set read groups information?' (dropdown: 'Do not set'), 'Select analysis mode' (dropdown: '1: Default setting only'), and 'Do you want to use presets?'. The right sidebar shows the 'History' panel with a search bar and a list of datasets, including 'GSE107337_3samples' and several 'FastQC on data' entries.

こんな感じになる。

W7-7: Upload



W7-7: Upload

こんな感じになる。2つのファイルをドラッグ&ドロップで置いたところ。①Start。

Download from web or upload from disk

Regular Composite Collection Rule-based

You added 2 file(s) to the queue. Add more files or click 'Start' to proceed.

Name	Size	Type	Genome	Settings	Status
ASM2650v1.fa	2.9 MB	Auto-det...	----- Additional S...	⚙️	0%
ASM2650v1.gff3	1.7 MB	Auto-det...	----- Additional S...	⚙️	0%

Type (set all): Auto-detect Genome (set all): ----- Additional S...

Choose local file Choose FTP file Paste/Fetch data Pause Reset **Start** Close

W7-7: Upload

こんな感じになる。2つのファイルをドラッグ&ドロップで置いたところ。①Start。アップロード中…。

Download from web or upload from disk

Regular Composite Collection Rule-based

Please wait...2 out of 2 remaining.

Name	Size	Type	Genome	Settings	Status
ASM2650v1.fa	2.9 MB	Auto-det...	----- Additional S...	⚙️	Adding to history...
ASM2650v1.gff3	1.7 MB	Auto-det...	----- Additional S...	⚙️	0%

Type (set all): Auto-detect Genome (set all): ----- Additional S...

Choose local file Choose FTP file Paste/Fetch data Pause Reset Start Close

W7-7: Upload

こんな感じになる。2つのファイルをドラッグ&ドロップで置いたところ。①Start。アップロード完了。②Close

Download from web or upload from disk

Regular Composite Collection Rule-based

Name	Size	Type	Genome	Settings	Status
ASM2650v1.fa	2.9 MB	Auto-det...	----- Additional S...	⚙️	100% ✓
ASM2650v1.gff3	1.7 MB	Auto-det...	----- Additional S...	⚙️	100% ✓

Type (set all): Auto-detect Genome (set all): ----- Additional S...

Choose local file Choose FTP file Paste/Fetch data Pause Reset Start Close

W7-8: 確認

The screenshot shows the Galaxy web interface at <https://www.ebi.ac.uk/ena/data/v>. The main navigation bar includes 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', and 'ユーザー'. The 'Tools' sidebar on the left lists 'GENOMICS ANALYSIS' with sub-sections for 'Assembly', 'Annotation', and 'Mapping'. Under 'Mapping', several tools are listed, including 'Bowtie2 - map reads against reference genome'. The central configuration panel for Bowtie2 includes the following options:

- Will you select a reference genome from your history or use a built-in index?** (Dropdown: Use a genome from the history and build index)
- Select reference genome** (Dropdown: 43: ASM2650v1.fa)
- Set read groups information?** (Dropdown: Do not set)
- Select analysis mode** (Dropdown: 1: Default setting only)
- Do you want to use presets?** (Dropdown: [unselected])

The right-hand 'History' panel shows a list of datasets under the heading 'GSE107337_3samples' (44 shown, 3.49 GB). The visible datasets are:

- 44: ASM2650v1.gff3
- 43: ASM2650v1.fa
- 42: FastQC on data 28: RawData
- 41: FastQC on data 28: Webpage
- 40: FastQC on data 27: RawData
- 39: FastQC on data 27: Webpage

W7-8: 確認

こんな感じになる。①アップロードした2つのファイルが、履歴に追加されているのがわかる。②リファレンスゲノムの候補として、W7-6の「No fasta dataset」だった状態から、履歴43のものに切り替わっていることもわかる。

The screenshot shows the Galaxy web interface. The main content area displays the configuration for a tool, with a dropdown menu for 'Select reference genome' set to '43: ASM2650v1.fa'. A red arrow labeled '2' points to this dropdown. The right-hand panel shows the 'History' section, which lists several datasets. The top two datasets, '44: ASM2650v1.gff3' and '43: ASM2650v1.fa', are highlighted in green, and a red arrow labeled '1' points to them. The 'Tools' sidebar on the left lists various genomics analysis tools under categories like Assembly, Annotation, and Mapping.

W7-9: 実行

The screenshot shows the Galaxy web interface. The central panel displays the configuration for a Bowtie2 job. A red arrow with the number 1 points to the 'Execute' button. The right panel shows a list of datasets, including 'GSE107337_3samples' and several FastQC reports.

Galaxy

データ解析 ワークフロー 可視化する 共有データ ヘルプ ユーザー Using 1%

Tools

search tools

GENOMICS ANALYSIS

Assembly

Annotation

Mapping

[Bowtie2](#) - map reads against reference genome

[LASTZ](#) : align long sequences

[LASTZ_D](#) : estimate substitution scores matrix

[Map with BWA-MEM](#) - map medium and long reads (> 100 bp) against reference genome

[Map with BWA](#) - map short reads (< 100 bp) against reference genome

[STAR-Fusion](#) detect fusion genes in RNA-Seq data

Do you want to tweak SAM/BAM Options?

No

See "Output Options" section of Help below for information

Save the bowtie2 mapping statistics to the history

Yes No

Job Resource Parameters

Use default job resource parameters

Execute ①

Bowtie2 Overview

Bowtie2 is an ultrafast and memory-efficient tool for aligning sequencing reads to long reference sequences. It is particularly good at aligning reads of about 50 up to 100s or 1,000s of characters to relatively long (e.g. mammalian) genomes. Bowtie 2 supports gapped, local, and paired-end alignment modes. Galaxy wrapper for Bowtie 2 outputs alignments in BAM format, enabling

History

search datasets

GSE107337_3samples

44 shown

3.49 GB

44: ASM2650v1.gff3

43: ASM2650v1.fa

42: FastQC on data 28: RawData

41: FastQC on data 28: W ebpage

40: FastQC on data 27: RawData

39: FastQC on data 27: W ebpage

W7-10: 実行中

Galaxy

データ解析 ワークフロー 可視化する 共有データ ヘルプ ユーザー Using 1%

Tools

search tools

GENOMICS ANALYSIS

Assembly

Annotation

Mapping

[Bowtie2](#) - map reads against reference genome

[LASTZ](#) : align long sequences

[LASTZ_D](#) : estimate substitution scores matrix

[Map with BWA-MEM](#) - map medium and long reads (> 100 bp) against reference genome

[Map with BWA](#) - map short reads (< 100 bp) against reference genome

[STAR-Fusion](#) detect fusion genes in RNA-Seq data

Executed **Bowtie2** and successfully added 3 jobs to the queue.

The tool uses 7 inputs:

43: ASM2650v1.fa

28: Trimmomatic on SRR6322569_2 (R2 paired)

27: Trimmomatic on SRR6322569_1 (R1 paired)

...

It produces 3 outputs:

47: Bowtie2 on data 43, data 28, and data 27: aligned reads (BAM)

History

search datasets

GSE107337_3samples

47 shown

3.49 GB

47: Bowtie2 on data 43, data 28, and data 27: aligned reads (BAM)

46: Bowtie2 on data 43, data 24, and data 23: aligned reads (BAM)

45: Bowtie2 on data 43, data 20, and data 19: aligned reads (BAM)

44: ASM2650v1.gff3

43: ASM2650v1.fa

W7-10: 実行中

https://www.ebi.ac.uk/ena/data/v x Galaxy x +

usegalaxy.org

Galaxy データ解析 ワークフロー 可視化する 共有データ ヘルプ ユーザー Using 1%

Tools

search tools

GENOMICS ANALYSIS

Assembly

Annotation

Mapping

[Bowtie2](#) - map reads against reference genome

[LASTZ](#) : align long sequences

[LASTZ_D](#) : estimate substitution scores matrix

[Map with BWA-MEM](#) - map medium and long reads (> 100 bp) against reference genome

[Map with BWA](#) - map short reads (< 100 bp) against reference genome

[STAR-Fusion](#) detect fusion genes in RNA-Seq data

Executed **Bowtie2** and successfully added 3 jobs to the queue.

The tool uses 7 inputs:

43: ASM2650v1.fa

28: Trimmomatic on SRR6322569_2 (R2 paired)

27: Trimmomatic on SRR6322569_1 (R1 paired)

...

It produces 3 outputs:

47: Bowtie2 on data 43, data 28, and data 27: aligned reads (BAM)

History

search datasets

GSE107337_3samples

47 shown

3.49 GB

47: Bowtie2 on data 43, data 28, and data 27: aligned reads (BAM)

46: Bowtie2 on data 43, data 24, and data 23: aligned reads (BAM)

45: Bowtie2 on data 43, data 20, and data 19: aligned reads (BAM)

44: ASM2650v1.gff3

43: ASM2650v1.fa

W7-11: 実行完了

https://www.ebi.ac.uk/ena/data/v x Galaxy x +

usegalaxy.org

Galaxy データ解析 ワークフロー 可視化する 共有データ ヘルプ ユーザー Using 1%

Tools ☆ ↑

search tools x

GENOMICS ANALYSIS

Assembly

Annotation

Mapping

[Bowtie2](#) - map reads against reference genome

[LASTZ](#) : align long sequences

[LASTZ_D](#) : estimate substitution scores matrix

[Map with BWA-MEM](#) - map medium and long reads (> 100 bp) against reference genome

[Map with BWA](#) - map short reads (< 100 bp) against reference genome

[STAR-Fusion](#) detect fusion genes in RNA-Seq data

✓ Executed **Bowtie2** and successfully added 3 jobs to the queue.

The tool uses 7 inputs:

43: ASM2650v1.fa

28: Trimmomatic on SRR6322569_2 (R2 paired)

27: Trimmomatic on SRR6322569_1 (R1 paired)

...

It produces 3 outputs:

47: Bowtie2 on data 43, data 28, and data 27: aligned reads (BAM)

44: ASM2650v1.gff3

43: ASM2650v1.fa

History ↻ + □ ⚙

search datasets x

GSE107337_3samples

47 shown

4.52 GB

☑ 🗑️ 💬

47: Bowtie2 on data 43, data 28, and data 27: aligned reads (BAM) 👁️ ✎ ✕

46: Bowtie2 on data 43, data 24, and data 23: aligned reads (BAM) 👁️ ✎ ✕

45: Bowtie2 on data 43, data 20, and data 19: aligned reads (BAM) 👁️ ✎ ✕

44: ASM2650v1.gff3 👁️ ✎ ✕

43: ASM2650v1.fa 👁️ ✎ ✕

Contents

- W1: 公共データベースENA
- W2: クラウド解析環境Galaxy
- W3: ENAからGalaxyへのFASTQファイルの送付
- W4: クオリティチェック (FastQC)
- W5: クオリティコントロール (Trimmomatic)
- W6: クオリティチェック (FastQC)
- W7: ゲノム配列へのマッピング (Bowtie2)
- W8: カウント情報取得 (htseq-count)
- W9: カウント情報の連結 (Column join)
- W10: カウントデータの数値行列の行名について
- W11: 公共カウントデータ (GSE107337) を用いて、倍率変化と \log_2 (倍率変化)を算出
- W12: Galaxy上で得られたカウントデータについても同様に計算し、比較
- W13とW14: 原著論文中の記載事項を確認
- W15: 2つのカウントデータから共通遺伝子の情報を用いて類似度を計算

W8-1 : htseq-count

次はカウント情報取得。htseq-countというプログラムを用いる。①を押してヒストリーパネル以外をリフレッシュした後の状態。

The screenshot shows the Galaxy web interface. The browser address bar displays `https://www.ebi.ac.uk/ena/data/` and the page title is "Galaxy". The main navigation bar includes "データ解析" (Data Analysis), "ワークフロー" (Workflow), "可視化する" (Visualize), "共有データ" (Shared Data), "ヘルプ" (Help), and "ユーザー" (User). The "Tools" panel on the left lists various tool categories such as "Get Data", "Send Data", "Collection Operations", "GENERAL TEXT TOOLS", "Text Manipulation", "Filter and Sort", "Join, Subtract and Group", "Datamash", "GENOMIC FILE MANIPULATION", "FASTA/FASTQ", "FASTQ Quality Control", "SAM/BAM", and "BED". The central content area features a tutorial card titled "Running Your Own Understanding how Galaxy works" with the subtitle "An in-depth tutorial". The "History" panel on the right shows a list of datasets, including "GSE107337_3samples" (47 shown, 4.52 GB) and several Bowtie2 alignment jobs (e.g., "47: Bowtie2 on data 43, data 28, and data 27: aligned reads (BAM)"). A red arrow with the number "1" points to the "refresh" icon in the History panel's search bar.

W8-1 : htseq-count

W7で行っていたのは、①Mapping。カウント情報取得を行うhtseq-countプログラムは、②RNA-seqのカテゴリ内に存在する。そのことを知っているからそうしています。もしプログラム名しかわからない場合は、③の検索窓でhtseq-countで探してもよい。

The screenshot shows the Galaxy web interface. On the left, the 'Tools' sidebar is visible with a search bar (labeled ③) and a list of categories. Under 'GENOMICS ANALYSIS', 'Mapping' is highlighted with a red lightning bolt and a circled 1 (labeled ①), and 'RNA-seq' is also highlighted with a red lightning bolt and a circled 2 (labeled ②). The central panel displays a 'Galaxy 101' tutorial card with the text 'Galaxy is an open source, web-based platform for data intensive biomedical research...' and a 'Galaxy Training Network' logo. On the right, the 'History' panel shows a list of datasets, including 'GSE107337_3samples' and several 'Bowtie2 on data...' entries.

W8-1 : htseq-count

②をクリックした結果。実は③featureCountsというプログラムでもカウント情報取得という目的を達成することはできる。がここではhtseq-countを利用すべく、④下部を少しずつ眺めながらhtseq-countを探す。

The screenshot shows the Galaxy web interface. The 'Tools' panel on the left lists various genomic analysis tools. A red arrow labeled '2' points to 'RNA-seq'. Another red arrow labeled '3' points to 'featureCounts'. A red arrow labeled '4' points to the 'Tweets' section at the bottom of the main content area. The 'History' panel on the right shows a list of datasets, including 'GSE107337_3samples' and several Bowtie2 alignment jobs.

Galaxy is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy start here or consult our help resources. You can install your own Galaxy by following the tutorial and choose from thousands of tools from the Tool Shed.

Tools

search tools

GENOMICS ANALYSIS

Assembly

Annotation

Mapping

Variant Calling

ChIP-seq

RNA-seq

[limma](#) Perform differential expression with limma-voom or limma-trend

[edgeR](#) Perform differential expression of count data

[HISAT2](#) A fast and sensitive alignment program **updated**

[featureCounts](#) Measure gene expression in RNA-Seq experiments from SAM or BAM files.

History

search datasets

GSE107337_3samples

47 shown

4.52 GB

47: Bowtie2 on data 43, data 28, and data 27: aligned reads (BAM)

46: Bowtie2 on data 43, data 24, and data 23: aligned reads (BAM)

45: Bowtie2 on data 43, data 20, and data 19: aligned reads (BAM)

44: ASM2650v1.gff3

43: ASM2650v1.fa

Try Galaxy on the Cloud

Now you can have a personal Galaxy within the infinite Universe

Tweets by @galaxyproject

①htseq-countを発見したのでクリック。

W8-1 : htseq-count

The screenshot shows the Galaxy web interface. The 'Tools' sidebar on the left contains a search bar and a list of tools. The tool 'htseq-count' is highlighted with a red arrow and a circled '1'. The main content area displays a welcome message and a tutorial titled 'Running Your Own Understanding how Galaxy works'. The 'History' panel on the right shows a list of datasets and jobs, including 'Bowtie2 on data 43, data 28, and data 27: aligned reads (BAM)'.

W8-1 : htseq-count

The screenshot shows the Galaxy web interface. The browser address bar displays `https://www.ebi.ac.uk/ena/data/v` and `usegalaxy.org`. The Galaxy navigation bar includes options like "データ解析", "ワークフロー", "可視化する", "共有データ", "ヘルプ", and "ユーザー".

Tools Panel: A search bar for tools is visible. The `htseq-count` tool is selected, showing its description: "Count aligned reads in a BAM file that overlap features in a GFF file (Galaxy Version 0.9.1)".

Configuration Panel:

- Aligned SAM/BAM File:** Set to "47: Bowtie2 on".
- GFF File:** Set to "44:".
- Mode:** Set to "Union".
- Stranded:** Set to "Yes".

History Panel: Shows a list of datasets. The current dataset is "47: Bowtie2 on data 43, data 28, and data 27: aligned reads (BAM)". Other visible datasets include "46: Bowtie2 on data 43, data 24, and data 23: aligned reads (BAM)", "45: Bowtie2 on data 43, data 20, and data 19: aligned reads (BAM)", "44: ASM2650v1.gff3", and "43: ASM2650v1.fa".

計3つのBAMファイルを一度に取り扱いたいのので、
①Multiple datasets。

W8-2: 複数個指定

The screenshot shows the Galaxy web interface. The main tool configuration area is for 'htseq-count'. A red arrow points to the 'Align to BAM File' section, where the 'Multiple datasets' radio button is selected. Below this, there are two input fields for BAM files: '47: Bowtie2 on' and '44:'. The 'Mode' is set to 'Union' and 'Stranded' is set to 'Yes'. The right panel shows a history of datasets, including '47: Bowtie2 on data 43, data 28, and data 27: aligned reads (BAM)', '46: Bowtie2 on data 43, data 24, and data 23: aligned reads (BAM)', '45: Bowtie2 on data 43, data 20, and data 19: aligned reads (BAM)', '44: ASM2650v1.gff3', and '43: ASM2650v1.fa'.

計3つのBAMファイルを一度に取り扱いたいのので、
①Multiple datasets。②Browse Datasets。

W8-2: 複数個指定

The screenshot shows the Galaxy web interface. The main tool panel is for **htseq-count**. The **Aligned SAM/BAM File** section contains a list of datasets: 47: Bowtie2 on data 43, 46: Bowtie2 on data 43, and 45: Bowtie2 on data 43. A red arrow with the number 2 points to the **Browse Datasets** button next to the first dataset. The **History** panel on the right shows a list of datasets: 47: Bowtie2 on data 43, data 28, and data 27: aligned reads (BAM), 46: Bowtie2 on data 43, data 24, and data 23: aligned reads (BAM), 45: Bowtie2 on data 43, data 20, and data 19: aligned reads (BAM), 44: ASM2650v1.gff3, and 43: ASM2650v1.fa.

W8-2: 複数個指定

The screenshot shows the Galaxy web interface with a search modal open. The modal contains a search bar and a table of results. The table has three columns: Label, Details, and Time. The results are as follows:

Label	Details	Time
47: Bowtie2 on data 43, data 28, and data 27: aligned reads (BAM)	bam	2019-08-19 05:15
46: Bowtie2 on data 43, data 24, and data 23: aligned reads (BAM)	bam	2019-08-19 05:09
45: Bowtie2 on data 43, data 20, and data 19: aligned reads (BAM)	bam	2019-08-19 05:09
44: ASM2650v1.gff3	gff3	2019-08-19 04:52
43: ASM2650v1.fa	fasta	2019-08-19 05:05
42: FastQC on data 28: RawData	txt	2019-08-16 09:08
41: FastQC on data 28: Webpage	html	2019-08-16 09:08
40: FastQC on data 27: RawData	txt	2019-08-16 09:09

The background interface shows the Galaxy search bar with the text "Type to Search" and a list of tools including Sailfish, goseq, DESeq2, RNA STAR, cummeRbund, htseq-count, GffCompare, and Cuffmerge. The search results modal is currently open, displaying the table above.

こんな感じになる。入力はBAMファイルなので…それらを選択して、①Ok。

W8-2: 複数個指定

The screenshot shows the Galaxy web interface with a search results dialog box. The dialog box has a search input field and a table of search results. The table has three columns: Label, Details, and Time. The first three rows are highlighted in green, indicating they are selected. A red arrow with the number '1' points to the 'Ok' button at the bottom right of the dialog box.

Label	Details	Time
47: Bowtie2 on data 43, data 28, and data 27: aligned reads (BAM)	bam	2019-08-19 05:15
46: Bowtie2 on data 43, data 24, and data 23: aligned reads (BAM)	bam	2019-08-19 05:09
45: Bowtie2 on data 43, data 20, and data 19: aligned reads (BAM)	bam	2019-08-19 05:09
44: ASM2650v1.gff3	gff3	2019-08-19 04:52
43: ASM2650v1.fa	fasta	2019-08-19 05:05
42: FastQC on data 28: RawData	txt	2019-08-16 09:08
41: FastQC on data 28: Webpage	html	2019-08-16 09:08
40: FastQC on data 27: RawData	txt	2019-08-16 09:09

こんな感じになる。ここまででマッピング結果ファイルの指定は完了。

W8-2: 複数個指定

The screenshot displays the Galaxy web interface for the **htseq-count** tool. The tool description states: "Count aligned reads in a BAM file that overlap features in a GFF file (Galaxy Version 0.9.1)".

Aligned SAM/BAM File: A list of dataset IDs is entered: 47: Bowtie2 on data 43, 46: Bowtie2 on data 43, 45: Bowtie2 on data 43.

GFF File: Dataset ID 44 is selected.

Mode: Union

History: A list of datasets is shown, including the ones used in the current job:

- 47: Bowtie2 on data 43, data 28, and data 27: aligned reads (BAM)
- 46: Bowtie2 on data 43, data 24, and data 23: aligned reads (BAM)
- 45: Bowtie2 on data 43, data 20, and data 19: aligned reads (BAM)
- 44: ASM2650v1.gff3
- 43: ASM2650v1.fa

W8-3: GFFファイル

次は、どの領域のカウント情報を取得するかを指定する。①で見えているgff3ファイルを、②の部分で認識させるべく、②をクリック。

The screenshot shows the Galaxy web interface. The main tool panel displays the **htseq-count** tool configuration. The **Aligned SAM/BAM File** section shows a list of datasets: 47: Bowtie2 on data 43, 46: Bowtie2 on data 43, and 45: Bowtie2 on data 43. The **GFF File** section has a dropdown menu set to '44:'. A red arrow labeled '2' points to this dropdown. The **Mode** section is set to 'Union'. The **History** panel on the right shows a list of datasets, with '44: ASM2650v1.gff3' highlighted in green. A red arrow labeled '1' points to this entry.

W8-3: GFFファイル

次は、どの領域のカウント情報を取得するかを指定する。①で見えているgff3ファイルを、②の部分で認識させるべく、②をクリック。②の部分では履歴番号しか表示されてなかったのが不安だったが、③無事認識されているようだ。

The screenshot shows the Galaxy web interface with the following components:

- Tools Panel:** Lists various tools including `htseq-count`, `goseq`, `DESeq2`, `RNA STAR`, `cummeRbund`, `htseq-count`, `GffCompare`, and `Cuffmerge`.
- Tool Configuration:** Shows the `htseq-count` tool with options for Favorite, Versions, and Options. The description states: "Count aligned reads in a BAM file that overlap features in a GFF file (Galaxy Version 0.9.1)".
- Input Fields:** Includes an "Aligned SAM/BAM File" field with a list of files (47, 46, 45) and a "GFF File" dropdown menu currently showing "44:". A "Mode" dropdown is set to "Union".
- History Panel:** Displays a list of datasets under "GSE107337_3samples". The list includes:
 - 47: Bowtie2 on data 43, data 28, and data 27: aligned reads (BAM)
 - 46: Bowtie2 on data 43, data 24, and data 23: aligned reads (BAM)
 - 45: Bowtie2 on data 43, data 20, and data 19: aligned reads (BAM)
 - 44: ASM2650v1.gff3
 - 43: ASM2650v1.fa

W8-3: GFFファイル

次は、どの領域のカウント情報を取得するかを指定する。①で見えているgff3ファイルを、②の部分で認識させるべく、②をクリック。②の部分では履歴番号しか表示されてなかったのが不安だったが、③無事認識されているようだ。ここまででアノテーションファイル(履歴-44)を認識させるところが完了

The screenshot shows the Galaxy web interface. The browser address bar displays <https://www.ebi.ac.uk/ena/data/> and the Galaxy logo. The main navigation bar includes "データ解析", "ワークフロー", "可視化する", "共有データ", "ヘルプ", and "ユーザー". The "Tools" panel on the left lists various tools, with "htseq-count" selected. The tool configuration area shows the "Aligned SAM/BAM File" field containing a list of datasets: "47: Bowtie2 on data 43, 46: Bowtie2 on data 43, 45: Bowtie2 on data 43". Below this, the "GFF File" field is set to "44:" and the "Mode" is set to "Union". The "History" panel on the right shows a list of datasets, including "47: Bowtie2 on data 43, data 28, and data 27: aligned reads (BAM)", "46: Bowtie2 on data 43, data 24, and data 23: aligned reads (BAM)", "45: Bowtie2 on data 43, data 20, and data 19: aligned reads (BAM)", "44: ASM2650v1.gff3", and "43: ASM2650v1.fa".

W8-4: Feature type

①少し下部に移動。次は何の情報をカウントしたいのかを指定する。②デフォルトはexonだが、ここでは原著論文と同じくgeneのカウント情報を取得したいので、geneに変更。

The screenshot shows the Galaxy web interface. The main panel displays the configuration for the 'Feature type' tool. The 'Feature type' dropdown is set to 'exon', with a red arrow and the number '2' pointing to it. Below it, the 'ID Attribute' is set to 'gene_id'. The 'Set advanced options' dropdown is set to 'Default settings'. The 'Execute' button is visible at the bottom. On the right, the 'History' panel shows a search bar and a list of datasets, including 'GSE107337_3samples' and several 'Bowtie2 on data' entries. A red arrow and the number '1' point to the 'History' search bar.

W8-4: Feature type

①少し下部に移動。次は何の情報をカウントしたいのかを指定する。②デフォルトはexonだが、ここでは原著論文と同じくgeneのカウント情報を取得したいので、geneに変更。③変更後。

The screenshot shows the Galaxy web interface. The main panel is titled "Feature type" and contains the following fields and options:

- Feature type:** A text input field containing "gene". A red arrow with the number "3" points to this field.
- ID Attribute:** A text input field containing "gene_id".
- Set advanced options:** A dropdown menu set to "Default settings".
- Execute:** A blue button with a checkmark icon.

The right-hand panel is titled "History" and shows a list of datasets:

- 47: Bowtie2 on data 43, data 28, and data 27: aligned reads (BAM)
- 46: Bowtie2 on data 43, data 24, and data 23: aligned reads (BAM)
- 45: Bowtie2 on data 43, data 20, and data 19: aligned reads (BAM)
- 44: ASM2650v1.gff3
- 43: ASM2650v1.fa

W8-5: 実行

The screenshot shows the Galaxy web interface. The main content area is for a tool configuration. The 'Feature type' field is set to 'gene'. The 'ID Attribute' field is set to 'gene_id'. The 'Set advanced options' dropdown is set to 'Default settings'. A red arrow with the number '1' points to the 'Execute' button. The right sidebar shows the 'History' panel with a list of datasets, including '47: Bowtie2 on data 43, data 28, and data 27: aligned reads (BAM)', '46: Bowtie2 on data 43, data 24, and data 23: aligned reads (BAM)', '45: Bowtie2 on data 43, data 20, and data 19: aligned reads (BAM)', '44: ASM2650v1.gff3', and '43: ASM2650v1.fa'. The top navigation bar includes 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', and 'ユーザー'. The top right corner shows 'Using 1%'.

W8-6: 実行中

https://www.ebi.ac.uk/ena/data/v Galaxy

usegalaxy.org

Galaxy データ解析 ワークフロー 可視化する 共有データ ヘルプ ユーザー Using 1%

Tools

search tools

on RNA-Seq transcripts

[Sailfish](#) transcript quantification from RNA-seq data

[goseq](#) tests for overrepresented gene categories

[DESeq2](#) Determines differentially expressed features from count tables

[RNA STAR](#) Gapped-read mapper for RNA-seq data

[cummeRbund](#) visualize Cuffdiff output

[htseq-count](#) - Count aligned reads in a BAM file that overlap features in a GFF file

[GffCompare](#) compare assembled transcripts to a reference annotation

[Cuffmerge](#) merge together several Cufflinks assemblies

History

search datasets

GSE107337_3samples

53 shown

4.52 GB

53: htseq-count on data 44 and data 47 (no feature)

52: htseq-count on data 44 and data 47

51: htseq-count on data 44 and data 46 (no feature)

50: htseq-count on data 44 and data 46

49: htseq-count on da

Executed **htseq-count** and successfully added 3 jobs to the queue.

The tool uses 4 inputs:

47: **Bowtie2** on data 43, data 28, and data 27: aligned reads (BAM)

46: **Bowtie2** on data 43, data 24, and data 23: aligned reads (BAM)

45: **Bowtie2** on data 43, data 20, and data 19: aligned reads (BAM)

...

It produces 6 outputs:

53: **htseq-count** on data 44 and data 47 (no feature)

W8-6: 実行中

https://www.ebi.ac.uk/ena/data/v Galaxy

usegalaxy.org

Galaxy データ解析 ワークフロー 可視化する 共有データ ヘルプ ユーザー Using 1%

Tools

search tools

on RNA-Seq transcripts

[Sailfish](#) transcript quantification from RNA-seq data

[goseq](#) tests for overrepresented gene categories

[DESeq2](#) Determines differentially expressed features from count tables

[RNA STAR](#) Gapped-read mapper for RNA-seq data

[cummeRbund](#) visualize Cuffdiff output

[htseq-count](#) - Count aligned reads in a BAM file that overlap features in a GFF file

[GffCompare](#) compare assembled transcripts to a reference annotation

[Cuffmerge](#) merge together several Cufflinks assemblies

History

search datasets

GSE107337_3samples

53 shown

4.52 GB

53: htseq-count on data 44 and data 47 (no feature)

52: htseq-count on data 44 and data 47

51: htseq-count on data 44 and data 46 (no feature)

50: htseq-count on data 44 and data 46

49: htseq-count on data 4

Executed **htseq-count** and successfully added 3 jobs to the queue.

The tool uses 4 inputs:

47: **Bowtie2** on data 43, data 28, and data 27: aligned reads (BAM)

46: **Bowtie2** on data 43, data 24, and data 23: aligned reads (BAM)

45: **Bowtie2** on data 43, data 20, and data 19: aligned reads (BAM)

...

It produces 6 outputs:

53: **htseq-count** on data 44 and data 47 (no feature)

W8-7: 実行完了

The screenshot shows the Galaxy web interface at usegalaxy.org. The main content area displays a green notification box with a checkmark icon, indicating a successful execution of the **htseq-count** tool. The message states: "Executed **htseq-count** and successfully added 3 jobs to the queue." Below this, it lists the tool's inputs and outputs.

Inputs:

- 47: Bowtie2 on data 43, data 28, and data 27: aligned reads (BAM)
- 46: Bowtie2 on data 43, data 24, and data 23: aligned reads (BAM)
- 45: Bowtie2 on data 43, data 20, and data 19: aligned reads (BAM)
- ...

Outputs:

- 53: htseq-count on data 44 and data 47 (no feature)
- 50: htseq-count on data 44 and data 46
- 49: htseq-count on data 44 and data 47 (no feature)
- 51: htseq-count on data 44 and data 46 (no feature)
- 52: htseq-count on data 44 and data 47
- 53: htseq-count on data 44 and data 47 (no feature)

The right-hand panel shows the 'History' section, which includes a search bar for datasets and a list of recent jobs. The top job in the history is 'GSE107337_3samples' (4.52 GB). Below it, several 'htseq-count' jobs are listed, each with icons for viewing, editing, and deleting the job details.

W8-8: 解説

計6ファイルの結果のうち、①no featureが付加されたもの(履歴49, 51, 53)は、ほしいカウントデータ以外の統計情報に相当するものなので基本無視でよい。残りの履歴48, 50, 52が、指定したfeatureに対するカウント情報を含むものです。

The screenshot shows the Galaxy web interface. The top navigation bar includes 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', 'ユーザー', and 'Using 1%'. The left sidebar lists various tools, with 'htseq-count' selected. The main content area shows the execution log for 'htseq-count', which successfully added 3 jobs to the queue. The log lists four inputs: 47 (Bowtie2 on data 43, 28, 27), 46 (Bowtie2 on data 43, 24, 23), 45 (Bowtie2 on data 43, 20, 19), and 53 (htseq-count on data 44 and 47). It also lists six outputs, with three marked with a red circle containing the number 1: 53 (no feature), 51 (no feature), and 49 (no feature). The right sidebar shows the history of datasets, with three entries (53, 51, 49) marked with a red circle containing the number 1, indicating they are the focus of the explanation.

Contents

- W1: 公共データベースENA
- W2: クラウド解析環境Galaxy
- W3: ENAからGalaxyへのFASTQファイルの送付
- W4: クオリティチェック (FastQC)
- W5: クオリティコントロール (Trimmomatic)
- W6: クオリティチェック (FastQC)
- W7: ゲノム配列へのマッピング (Bowtie2)
- W8: カウント情報取得 (htseq-count)
- W9: カウント情報の連結 (Column join)
- W10: カウントデータの数値行列の行名について
- W11: 公共カウントデータ (GSE107337) を用いて、倍率変化と \log_2 (倍率変化)を算出
- W12: Galaxy上で得られたカウントデータについても同様に計算し、比較
- W13とW14: 原著論文中の記載事項を確認
- W15: 2つのカウントデータから共通遺伝子の情報を用いて類似度を計算

W9-1 : Column join

次はカウントデータ取得結果の連結。①を押してヒストリーパネル以外をリフレッシュした後の状態。

The screenshot shows the Galaxy web interface. The browser address bar is <https://www.ebi.ac.uk/ena/data/>. The Galaxy logo is in the top left, and the navigation menu includes 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', and 'ユーザー'. The 'Tools' panel on the left lists various categories like 'Get Data', 'Send Data', 'Collection Operations', 'GENERAL TEXT TOOLS', 'Text Manipulation', 'Filter and Sort', 'Join, Subtract and Group', 'Datamash', 'GENOMIC FILE MANIPULATION', 'FASTA/FASTQ', 'FASTQ Quality Control', 'SAM/BAM', and 'BED'. The main content area features a tweet from @galaxyproject with the text '125+ ways to use Galaxy' and a grid of tool icons. The 'History' panel on the right shows a search bar for datasets and a list of recent operations, including '53: htseq-count on data 4 and data 47 (no feature)', '52: htseq-count on data 4 and data 47', '51: htseq-count on data 4 and data 46 (no feature)', '50: htseq-count on data 4 and data 46', and '49: htseq-count on data 4'. A red arrow labeled '1' points to the 'データ解析' menu item.

W9-1 : Column join

次はカウントデータ取得結果の連結。①を押してヒストリーパネル以外をリフレッシュした後の状態。② Collection Operations。

The screenshot shows the Galaxy web interface. The browser address bar displays 'https://www.ebi.ac.uk/ena/data/'. The Galaxy logo is in the top left, and the navigation menu includes 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', and 'ユーザー'. The 'Tools' sidebar on the left has a search bar and categories like 'Get Data', 'Send Data', and 'Collection Operations'. A red arrow labeled '2' points to 'Collection Operations'. The main content area features a '125+ ways to use Galaxy' banner with a grid of tool icons. The 'History' panel on the right shows a list of datasets, including 'GSE107337_3samples' with 53 items. A red arrow labeled '1' points to the 'refresh' icon in the top right of the history panel.

W9-1 : Column join

次はカウントデータ取得結果の連結。①を押してヒストリーパネル以外をリフレッシュした後の状態。② Collection Operations。③Column join。

The screenshot shows the Galaxy web interface. The top navigation bar includes 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', and 'ユーザー'. The 'Tools' panel on the left lists various operations, with 'Column Join on Collections' highlighted by a red arrow labeled '3'. The main content area displays a welcome message and a carousel titled '125+ ways to use Galaxy'. The 'History' panel on the right shows a list of datasets, including 'GSE107337_3samples' and several 'htseq-count' results.

W9-1 : Column join

次はカウントデータ取得結果の連結。①を押してヒストリーパネル以外をリフレッシュした後の状態。② Collection Operations。③ Column join。こんな感じになる。

The screenshot shows the Galaxy web interface. The browser address bar displays `https://www.ebi.ac.uk/ena/data/v` and `usegalaxy.org`. The main navigation bar includes 'Galaxy', 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', 'ユーザー', and 'Using 1%'. The left sidebar contains 'Tools' with a search bar and 'Collection Operations' with links for 'Column Join on Collections', 'Unzip Collection', 'Zip Collection', 'Filter empty datasets from a collection', 'Filter failed datasets from a collection', 'Flatten Collection into a flat list of datasets', 'Merge Collections into single list of datasets', 'Relabel List Identifiers from contents of a file', and 'Filter List from contents of a file'. The main workspace shows the 'Column Join on Collections (Galaxy Version 0.0.3)' tool configuration. Under 'Tabular files', a list of files is shown, with the first item selected: '53: htseq-count on data 44 ar'. Below this, the 'Identifier column' is set to '1', the 'Number of Header lines in each item' is '0', and 'Keep original column header' is set to 'Yes'. The right sidebar shows the 'History' panel with a search bar and a list of datasets under 'GSE107337_3samples'. The first dataset is highlighted in green: '53: htseq-count on data 4 4 and data 47 (no feature)'. Other datasets in the list include '52: htseq-count on data 4 4 and data 47', '51: htseq-count on data 4 4 and data 46 (no feature)', '50: htseq-count on data 4 4 and data 46', and '49: htseq-count on data 4'.

W9-2: 連結対象指定

The screenshot shows the Galaxy web interface with the 'Column Join on Collections' tool selected. The interface is divided into several sections:

- Tools:** A search bar and a list of tool categories including 'Get Data', 'Send Data', and 'Collection Operations'.
- Tool Configuration:**
 - Column Join on Collections (Galaxy Version 0.0.3):** Includes buttons for 'Favorite', 'Versions', and 'Options'.
 - Tabular files:** A list of files with a 'Browse Datasets' button highlighted by a red arrow with the number '1'.
 - Identifier column:** A text input field containing '1'.
 - Number of Header lines in each item:** A text input field containing '0'.
 - Keep original column header:** Radio buttons for 'Yes' and 'No'.
- History:** A list of datasets under the heading 'GSE107337_3samples', showing 53 items with a total size of 4.52 GB. Each item has a preview and control icons.

W9-2: 連結対象指定

①Browse Datasets。連結したいものはヒストリー48, 50, 52なので…。

The screenshot shows the Galaxy web interface with a search dialog box open. The dialog box has a search input field and a table of search results. The table has three columns: Label, Details, and Time. The results are as follows:

Label	Details	Time
53: htseq-count on data 44 and data 47 (no feature)	tabular	2019-08-19 09:03
52: htseq-count on data 44 and data 47	tabular	2019-08-19 09:03
51: htseq-count on data 44 and data 46 (no feature)	tabular	2019-08-19 08:48
50: htseq-count on data 44 and data 46	tabular	2019-08-19 08:48
49: htseq-count on data 44 and data 45 (no feature)	tabular	2019-08-19 08:39
48: htseq-count on data 44 and data 45	tabular	2019-08-19 08:39
47: Bowtie2 on data 43, data 28, and data 27: aligned reads (BAM)	bam	2019-08-19 08:30
46: Bowtie2 on data 43, data 24, and data 23: aligned reads (BAM)	bam	2019-08-19 08:30

Below the table are 'Cancel' and 'Ok' buttons. The background shows the Galaxy interface with a 'Keep original column header' dialog box and a list of datasets.

W9-2: 連結対象指定

①Browse Datasets。連結したいものは履歴48, 50, 52なので…それらを指定して、②Ok。

The screenshot shows the Galaxy web interface with a search results dialog box open. The dialog box has a search input field at the top and a table of results below. The table has three columns: 'Label', 'Details', and 'Time'. The following table represents the data shown in the dialog box:

Label	Details	Time
53: htseq-count on data 44 and data 47 (no feature)	tabular	2019-08-19 09:03
52: htseq-count on data 44 and data 47	tabular	2019-08-19 09:03
51: htseq-count on data 44 and data 46 (no feature)	tabular	2019-08-19 08:48
50: htseq-count on data 44 and data 46	tabular	2019-08-19 08:48
49: htseq-count on data 44 and data 45 (no feature)	tabular	2019-08-19 08:39
48: htseq-count on data 44 and data 45	tabular	2019-08-19 08:39
47: Bowtie2 on data 43, data 28, and data 27: aligned reads (BAM)	bam	2019-08-19 08:30
46: Bowtie2 on data 43, data 24, and data 23: aligned reads (BAM)	bam	2019-08-19 08:30

At the bottom of the dialog box, there are 'Cancel' and 'Ok' buttons. A red arrow with the number '2' points to the 'Ok' button.

W9-2: 連結対象指定

①Browse Datasets。連結したいものは履歴48, 50, 52なので…それらを指定して、②Ok。こんな感じになる。

The screenshot shows the Galaxy web interface. The main panel displays the configuration for the 'Column Join on Collections' tool. The 'Tabular files' section shows a list of files, with items 49, 50, and 52 highlighted. The 'Identifier column' is set to 1, and the 'Number of Header lines in each item' is set to 0. The 'Keep original column header' option is set to 'Yes'. The 'History' panel on the right shows a list of datasets, with items 49, 50, and 52 highlighted in green, indicating they are selected for the join operation.

W9-3: 実行

The screenshot shows the Galaxy web interface. The main panel displays the configuration for the 'Join' tool. The 'Fill character' field contains a period. The 'Additional datasets to create' section has a 'Select/Unselect all' checkbox. The 'Execute' button is highlighted with a red arrow labeled '2'. The 'History' panel on the right shows a list of datasets, with the top entry '53: htseq-count on data 4 and data 47 (no feature)' highlighted in green and marked with a red arrow labeled '1'.

composed of the input dataset names

Fill character

.

Additional datasets to create

Select/Unselect all

Execute

Joins lists of tabular datasets together on a field.

Example

To join three files, with headers, based on the first column:

First file (in_1.tabular):

#KEY	c2	c3	c4
one	1-1	1-2	1-3
two	1-4	1-5	1-6
three	1-7	1-8	1-9

History

search datasets

GSE107337_3samples

53 shown

4.52 GB

53: htseq-count on data 4 and data 47 (no feature)

52: htseq-count on data 4 and data 47

51: htseq-count on data 4 and data 46 (no feature)

50: htseq-count on data 4 and data 46

49: htseq-count on data 4

W9-4: 実行中

The screenshot shows the Galaxy web interface. The top navigation bar includes 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', and 'ユーザー'. The main content area is divided into three panels:

- Tools:** A sidebar on the left with a search bar and a list of tool categories: 'Get Data', 'Send Data', and 'Collection Operations'. Under 'Collection Operations', several tools are listed, including 'Column Join on Collections', 'Unzip Collection', 'Zip Collection', 'Filter empty datasets from a collection', 'Filter failed datasets from a collection', 'Flatten Collection into a flat list of datasets', 'Merge Collections into single list of datasets', 'Relabel List Identifiers from contents of a file', and 'Filter List from contents of a file'.
- Job Execution Status:** A central green panel with a checkmark icon. It states: 'Executed **Column Join** and successfully added 1 job to the queue.' Below this, it lists the tool's inputs: '52: htseq-count on data 44 and data 47', '50: htseq-count on data 44 and data 46', and '48: htseq-count on data 44 and data 45'. It then states: 'It produces this output:' followed by '54: Column Join on data 52, data 50, and data 48'. At the bottom, it provides instructions: 'You can check the status of queued jobs and view the resulting data by refreshing the History panel. When the job has been run'.
- History:** A panel on the right with a search bar and a list of datasets. The top dataset is 'GSE107337_3samples' (54 shown, 4.52 GB). Below it, a list of jobs is shown, including '54: Column Join on data 52, data 50, and data 48', '53: htseq-count on data 44 and data 47 (no feature)', '52: htseq-count on data 44 and data 47', and '51: htseq-count on data 44 and data 46 (no feature)'. Each job entry has icons for viewing, editing, and deleting.

W9-5: 実行完了

The screenshot shows the Galaxy web interface at usegalaxy.org. The main content area displays a green success message: "Executed **Column Join** and successfully added 1 job to the queue." It details the tool's inputs: "52: htseq-count on data 44 and data 47", "50: htseq-count on data 44 and data 46", and "48: htseq-count on data 44 and data 45". The output is listed as "54: Column Join on data 52, data 50, and data 48". A note at the bottom suggests checking the job status in the History panel. The History panel on the right shows a list of jobs, including "GSE107337_3samples" (4.52 GB) and several "htseq-count" jobs. The left sidebar contains tool categories like "Get Data", "Send Data", and "Collection Operations".

W9-6: 解説

The screenshot shows the Galaxy web interface. The main content area displays a green box with a checkmark indicating a successful job execution. The text reads: "Executed **Column Join** and successfully added 1 job to the queue." It lists three inputs: "52: htseq-count on data 44 and data 47", "50: htseq-count on data 44 and data 46", and "48: htseq-count on data 44 and data 45". It also states "It produces this output:" followed by "54: Column Join on data 52, data 50, and data 48".

The right-hand panel, titled "History", shows a list of datasets. A red arrow with the number "1" points to the entry "54: Column Join on data 52, data 50, and data 48". Other entries include "53: htseq-count on data 44 and data 47 (no feature)", "52: htseq-count on data 44 and data 47", "51: htseq-count on data 44 and data 46 (no feature)", and "50: htseq-count on data 44".

W9-6: 解説

出力結果は、①1つ(履歴-54)。②データを中央パネル上で表示させて、大まかに眺める。

The screenshot shows the Galaxy web interface. The main panel displays a green notification box with a checkmark, indicating a successful job execution. The notification text reads: "Executed **Column Join** and successfully added 1 job to the queue." Below this, it lists the tool's inputs: "52: htseq-count on data 44 and data 47", "50: htseq-count on data 44 and data 46", and "48: htseq-count on data 44 and data 45". It then states "It produces this output:" followed by "54: Column Join on data 52, data 50, and data 48". At the bottom, it says "You can check the status of queued jobs and view the resulting data by refreshing the History panel. When the job has been run".

The History panel on the right shows a list of datasets. A red arrow with the number "2" points to the top entry: "54: Column Join on data 52, data 50, and data 48". A tooltip with the text "データを表示" (Show data) is visible over the eye icon for this entry. Other entries in the history include "53: htseq-count on data 44 and data 47 (no feature)", "52: htseq-count on data 44 and data 47", "51: htseq-count on data 44 and data 46 (no feature)", and "50: htseq-count on data 44".

W9-7: 行列

こんな感じになる。①1列目はgene ID情報。②2列目は、③ヒストリー48の情報。ヒストリー48の名前が④で見えているものになります。

The screenshot shows the Galaxy web interface. The main panel displays a table with two columns: '#KEY' and 'htseq-count on data 44 and data 45_2'. The table lists various gene IDs and their corresponding counts. The right panel shows the 'History' section with a list of datasets. A green box highlights the selected history entry, '54: Column Join on data 52, data 50, and data 48'. Other history entries are visible below it.

#KEY	htseq-count on data 44 and data 45_2
EBG00001128470	130
EBG00001128476	15
EBG00001128500	99
EBG00001128509	166
EBG00001128529	0
LGG_00001	195
LGG_00002	513
LGG_00003	21
LGG_00004	138
LGG_00005	239
LGG_00006	680
LGG_00007	6
LGG_00008	54
LGG_00009	11
LGG_00010	13
LGG_00011	1412
LGG_00012	1373
LGG_00013	703

W9-7: 行列

こんな感じになる。①1列目はgene ID情報。②2列目は、③ヒストリー48の情報。ヒストリー48の名前が④で見えているものになります。ちなみに、⑤ヒストリー45は、SRR6322564 (pH4.5_1h_rep3) のマッピング結果のBAMファイルに相当します。

The screenshot shows the Galaxy web interface. The main content area displays a table with two columns: '#KEY' and 'htseq-count on data 44 and data 45_2'. The table lists various gene IDs and their corresponding counts. A red arrow points to the 5th row of the table, which is highlighted in green. The right sidebar shows the 'History' panel with a search bar and a list of datasets. The 5th item in the history list is highlighted in green, matching the row in the table. The left sidebar contains navigation options like 'Tools', 'Get Data', 'Send Data', and 'Collection Operations'.

1	2
#KEY	htseq-count on data 44 and data 45_2
EBG00001128470	130
EBG00001128476	15
EBG00001128500	99
EBG00001128509	166
EBG00001128529	0
LGG_00001	195
LGG_00002	513
LGG_00003	21
LGG_00004	138
LGG_00005	239
LGG_00006	680
LGG_00007	6
LGG_00008	54
LGG_00009	11
LGG_00010	13
LGG_00011	1412
LGG_00012	1373
LGG_00013	703

W9-7: 行列

①中央パネルを右のほうに移動させたところ。②3列目が、③ヒストリー50の情報。

The screenshot shows the Galaxy web interface. The central panel displays a table with columns labeled '3' and '4'. The table contains numerical data for various rows. The right panel shows the 'History' section with a list of datasets. The 50th entry in the history is highlighted in green and labeled with a red arrow '3'. A red arrow '1' points to the scrollbar of the central panel, and a red arrow '2' points to the column header '3'.

	3	4
and data 45_2	130	75
htseq-count on data 44 and data 46_2	15	17
htseq-c	99	52
	166	97
	0	0
	195	248
	513	589
	21	32
	138	171
	239	389
	680	910
	6	5
	54	22
	11	19
	13	19
	1412	817
	1373	1140
	703	500

History entries:

- 54: Column Join on data 52, data 50, and data 48
- 53: htseq-count on data 44 and data 46_2
- 52: htseq-count on data 44 and data 47
- 51: htseq-count on data 44 and data 46 (no feature)
- 50: htseq-count on data 44

W9-7: 行列

①中央パネルを一番右まで移動させたところ。②4列目が、③ヒストリー52の情報。

The screenshot shows the Galaxy web interface. The central panel displays a table with the following data:

data 44 and data 46_2	htseq-count on data 44 and data 47_2
75	161
17	9
52	52
97	211
0	0
248	252
589	842
32	31
171	174
389	814
910	1640
5	1
22	279
19	24
19	22
817	806
1140	884
500	627

The History panel on the right shows a list of datasets. The 52nd entry is highlighted in green:

- 54: Column Join on data 52, data 50, and data 48
- 53: htseq-count on data 4 4 and data 47 (no feature)
- 52: htseq-count on data 4 4 and data 47**
- 51: htseq-count on data 4 4 and data 46 (no feature)
- 50: htseq-count on data 4

①をクリックして保存する。

W9-8: 保存

The screenshot shows the Galaxy web interface. The top navigation bar includes 'データ解析', 'ワークフロー', '可視化する', '共有データ', 'ヘルプ', and 'ユーザー'. The main content area is divided into three sections: 'Tools', a central table, and 'History'.

Tools

- Get Data
- Send Data
- Collection Operations
 - Column Join on Collections
 - Unzip Collection
 - Zip Collection
 - Filter empty datasets from a collection
 - Filter failed datasets from a collection
 - Flatten Collection into a flat list of datasets
 - Merge Collections into single list of datasets
 - Relabel List Identifiers from contents of a file
 - Filter List from contents of a file

Table

Tool	Input	Output
Column Join on Collections	data 44 and data 46_2	htseq-count on data 44 and data 47_2
	75	161
	17	9
	52	52
	97	211
	0	0
	248	252
	589	842
	32	31
	171	174
	389	814
	910	1640
	5	1
	22	279
	19	24
	19	22
	817	806
	1140	884
	500	627

History

- 54: Column Join on data 52, data 50, and data 48 (highlighted with a red arrow and '1')
- 53: htseq-count on data 4 4 and data 47 (no feature)
- 52: htseq-count on data 4 4 and data 47
- 51: htseq-count on data 4 4 and data 46 (no feature)

W9-8: 保存

①をクリックして保存する。②(ヘッダー行を除く)行数は2,949。つまり、2,949遺伝子からなるカウントデータだということがわかる。

The screenshot shows the Galaxy web interface. The main table displays the results of an htseq-count operation on data 44 and data 47_2. The table has two columns: the first column contains the number of reads for each gene, and the second column contains the gene identifiers. A red arrow labeled '2' points to the '2,949 lines, 1 comments' information in the preview window.

Gene	Count
data 44 and data 46_2	161
75	17
17	9
52	52
97	211
0	0
248	252
589	842
32	31
171	174
389	814
910	1640
5	1
22	279
19	24
19	22
817	806
1140	884
500	627

Preview window details:

- Operation: 54: Column Join on data 52, data 50, and data 48
- 2,949 lines, 1 comments
- フォーマット: tabular, データベース: ?
- Table structure:

1	2
#KEY	htseq-count on data 44 and
EBG00001128470	130
EBG00001128476	15
EBG00001128500	99
EBG00001128509	166

W9-8: 保存

①をクリックして保存する。②(ヘッダー行を除く)行数は2,949。つまり、2,949遺伝子からなるカウントデータだということがわかる。③ダウンロード。

The screenshot shows the Galaxy web interface. The main table displays the results of an htseq-count operation on data 44 and data 47_2. The table has two columns: the first column contains the number of reads for each gene, and the second column contains the gene identifiers. A preview window is open for the dataset '54: Column Join on data 52, data 50, and data 48', showing a tabular format with 2,949 lines and 1 comment. The preview window includes a 'ダウンロード' (Download) button and a '2' indicating the number of items. The URL at the bottom of the browser is https://usegalaxy.org/datasets/bbd44e69cb8906b5d6a17b680c498537/display?to_ext=tabular.

Gene ID	Count
data 44 and data 46_2	4
75	161
17	9
52	52
97	211
0	0
248	252
589	842
32	31
171	174
389	814
910	1640
5	1
22	279
19	24
19	22
817	806
1140	884
500	627

54: Column Join on data 52, data 50, and data 48
2,949 lines, 1 comments
フォーマット: tabular, データベース: ?
ダウンロード 2

```
#KEY htseq-count on data 44 and  
EBG00001128470 130  
EBG00001128476 15  
EBG00001128500 99  
EBG00001128509 166
```

W9-9: Excelで概観

①Excelで眺めるとこんな感じになります。これは最初のほう。

#KEY	htseq-count on data 44 and data 45_2	htseq-count on data 44 and data 46_2	htseq-count on data 44 and data 47_2
EBG00001128470	130	75	161
EBG00001128476	15	17	9
EBG00001128500	99	52	52
EBG00001128509	166	97	211
EBG00001128529	0	0	0
LGG_00001	195	248	252
LGG_00002	513	589	842
LGG_00003	21	32	31
LGG_00004	138	171	174
LGG_00005	239	389	814
LGG_00006	680	910	1640
LGG_00007	6	5	1
LGG_00008	54	22	279
LGG_00009	11	19	24
LGG_00010	13	19	22
LGG_00011	1412	817	806

W9-9: Excelで概観

①Excelで眺めるとこんな感じになります。これは最後のほう。

#KEY	htseq-count on data 44 and data 45_2	htseq-count on data 44 and data 46_2	htseq-count on data 44 and data 47_2
LGG_02928	51	75	131
LGG_02929	6	5	3
LGG_02930	6	6	5
LGG_02931	40	50	159
LGG_02932	16	8	16
LGG_02933	34	39	53
LGG_02934	111	27	84
LGG_02935	9	9	3
LGG_02936	36	52	49
LGG_02937	156	252	289
LGG_02938	69	131	82
LGG_02939	120	80	141
LGG_02940	386	197	342
LGG_02941	201	112	224
LGG_02942	120	69	72
LGG_02943	2111	499	879
LGG_02944	4255	2668	14963

W9-10: サンプル名

①私はこの段階で、後の解析結果を解釈しやすくするため、サンプル名などを手動で変更しておく。

gene_ID	pH4.5_1h_rep3	pH4.5_24h_rep3	pH7_CCG_rep2
EBG00001128470	130	75	161
EBG00001128476	15	17	9
EBG00001128500	99	52	52
EBG00001128509	166	97	211
EBG00001128529	0	0	0
LGG_00001	195	248	252
LGG_00002	513	589	842
LGG_00003	21	32	31
LGG_00004	138	171	174
LGG_00005	239	389	814
LGG_00006	680	910	1640
LGG_00007	6	5	1
LGG_00008	54	22	279
LGG_00009	11	19	24
LGG_00010	13	19	22
LGG_00011	1412	817	806

①

W9-11: 確認

得られたカウントデータ行列の行数 (= 2,949) が、Ensembl Bacteria中の*L. rhamnosus* GGの統計情報と一致するか調べるべく、①をクリック。結果的に同じ情報なので、②をクリックしてもよい。

The screenshot shows the Ensembl Bacteria website for *Lactobacillus rhamnosus* GG. The page includes a search bar, navigation links (HMMER, BLAST, Tools, Downloads, More), and a main content area with several sections:

- About *Lactobacillus rhamnosus* GG**: Includes an information icon and a link to "Information and statistics" (marked with a red arrow and circled number 2).
- Genome assembly**: Shows the assembly ID [ASM2650v1](#) and a link to "More information and statistics" (marked with a red arrow and circled number 1). It also includes links for "Download DNA sequence (FASTA)" and "Display your data in Ensembl Bacteria".
- Gene annotation**: Provides information on what can be found (protein-coding and non-coding genes, splice variants, cDNA, protein sequences, non-coding RNAs) and includes links for "More about this genebuild", "Download genes, cDNAs, ncRNA, proteins - FASTA - GFF3", and "Update your old Ensembl IDs".
- Comparative genomics**: Offers information on gene families based on HAMAP and PANTHER classification, with links for "More about comparative analyses" and "Phylogenetic overview of gene families".

http://bacteria.ensembl.org/Lactobacillus_rhamnosus_gg/Info/Index

W9-11: 確認

得られたカウントデータ行列の行数 (= 2,949) が、Ensembl Bacteria中の*L. rhamnosus* GGの統計情報と一致するか調べるべく、①をクリック。結果的に同じ情報なので、②をクリックしてもよい。③2,944個であることがわかります。

Lactobacillus rhamnosus GG

Lactobacillus rhamnosus GG Assembly and Gene Annotation

Statistics

Summary

Assembly	ASM2650v1, INSDC Assembly GCA_000026505.1 , Feb 2015
Database version	97.1
Base Pairs	3,010,111
Golden Path Length	3,010,111
Genebuild by	ENA
Genebuild method	Generated from ENA annotation
Data source	European Nucleotide Archive

Gene counts

Coding genes	2,944
Non coding genes	156
Small non coding genes	156
Gene transcripts	3,100



Contents

- W1: 公共データベースENA
- W2: クラウド解析環境Galaxy
- W3: ENAからGalaxyへのFASTQファイルの送付
- W4: クオリティチェック (FastQC)
- W5: クオリティコントロール (Trimmomatic)
- W6: クオリティチェック (FastQC)
- W7: ゲノム配列へのマッピング (Bowtie2)
- W8: カウント情報取得 (htseq-count)
- W9: カウント情報の連結 (Column join)
- W10: カウントデータの数値行列の行名について
- W11: 公共カウントデータ (GSE107337) を用いて、倍率変化と \log_2 (倍率変化)を算出
- W12: Galaxy上で得られたカウントデータについても同様に計算し、比較
- W13とW14: 原著論文中の記載事項を確認
- W15: 2つのカウントデータから共通遺伝子の情報を用いて類似度を計算

W10-1: 行名情報

GFFファイル(ASM2650v1.gff3)をExcelで開いて、EBG00001128470で検索した結果。

The screenshot shows the Microsoft Excel interface with the file 'ASM2650v1.gff3' open. A search dialog box titled '検索と置換' (Search and Replace) is displayed over the spreadsheet. The search criteria are set to 'EBG00001128470'. The search results are visible in the spreadsheet, showing rows with coordinates and feature names.

Line No.	Start	End	Feature Type	Feature Name
2519	FM179322	ena	mRNA	505972
2520	FM179322	ena	exon	505972
2521	FM179322	ena	CDS	505972
2522	###			
2523	FM179322	ena	gene	506286
2524	FM179322	ena	mRNA	506286
2525	FM179322	ena	exon	506286
2526	FM179322	ena	CDS	506286
2527	###			
2528	FM179322	Rfam	gene	506642
2529	FM179322	Rfam	transcript	506642
2530	FM179322	Rfam	exon	506642
2531	###			
2532	FM179322	Rfam	gene	507097
2533	FM179322	Rfam	transcript	507097
2534	FM179322	Rfam	exon	507097
2535	###			
2536	FM179322	ena	gene	507589
2537	FM179322	ena	mRNA	507589
2538	FM179322	ena	exon	507589

W10-1: 行名情報

W8-4で見られるFeature typeをgeneに変更した
ということは、GFFファイル中の①3列目が②
geneとなっている行の座標情報を用いてカウント
データを取得したということ。

自動保存 ASM2650v1.gff3 - Excel

ファイル ホーム 挿入 ページレイアウト 数式 データ 校閲 表示 ヘルプ 検索 共有

I2528 : X ✓ fx ID=gene:EBG00001128470;Name=rli28;biotype=sRNA;gene_id=EBG00001128470;logic_name=rfam_g ^
enes

検索と置換 ? X

検索(D) 置換(P)

検索する文字列(N): EBG00001128470

オプション(I) >>

すべて検索(I) 次を検索(E) 閉じる

	A	B	C	D	E	F	G	H	I	J	K	L	M
2519	FM179322	ena	mRNA	505972	50								GG_00497;bioty
2520	FM179322	ena	exon	505972	50								R86392-1;const
2521	FM179322	ena	CDS	505972	50								AR86392;prote
2522	###												
2523	FM179322	ena	gene	506286	50								ding;description
2524	FM179322	ena	mRNA	506286	50								GG_00498;bioty
2525	FM179322	ena	exon	506286	50								R86393-1;const
2526	FM179322	ena	CDS	506286	50								AR86393;prote
2527	###												
2528	FM179322	Rfam	gene	506642	506842	.	.	.	ID=gene:EBG00001128470;Name=rli28;biotype=sRNA;				
2529	FM179322	Rfam	transcript	506642	506842	.	.	.	ID=transcript:EBT00001719480;Parent=gene:EBG00001				
2530	FM179322	Rfam	exon	506642	506842	.	.	.	Parent=transcript:EBT00001719480;Name=EBG000011				
2531	###												
2532	FM179322	Rfam	gene	507097	507297	.	.	.	ID=gene:EBG00001128509;Name=rli28;biotype=sRNA;				
2533	FM179322	Rfam	transcript	507097	507297	.	.	.	ID=transcript:EBT00001719564;Parent=gene:EBG00001				
2534	FM179322	Rfam	exon	507097	507297	.	.	.	Parent=transcript:EBT00001719564;Name=EBG000011				
2535	###												
2536	FM179322	ena	gene	507589	509007	.	.	.	ID=gene:LGG_00499;Name=is27;biotype=protein_codin				
2537	FM179322	ena	mRNA	507589	509007	.	.	.	ID=transcript:CAR86394;Parent=gene:LGG_00499;Nan				
2538	FM179322	ena	exon	507589	509007	.	.	.	Parent=transcript:CAR86394;Name=CAR86394-1;const				

ASM2650v1

準備完了 ScrollLock - 100%

W10-2:4,5,7列目

W8-4で見られるFeature typeをgeneに変更した
ということは、GFFファイル中の①3列目が②
geneとなっている行の座標情報を用いてカウント
データを取得したということ。座標は、[③start, ④
end]の範囲が用いられる。⑤ストランド情報も。

	A	B	C	D	E	F	G	H	I	J	K	L	M
2519	FM179322	ena	mRNA	505972	506328	.	-	.	ID=transcript:CAR86392;Parent=gene:LGG_00497;bioty				
2520	FM179322	ena	exon	505972	506328	.	-	.	Parent=transcript:CAR86392;Name=CAR86392-1;const				
2521	FM179322	ena	CDS	505972	506328	.	-	0	ID=CDS:CAR86392;Parent=transcript:CAR86392;prote				
2522	###												
2523	FM179322	ena	gene	506286	506450	.	-	.	ID=gene:LGG_00498;biotype=protein_coding;description				
2524	FM179322	ena	mRNA	506286	506450	.	-	.	ID=transcript:CAR86393;Parent=gene:LGG_00498;bioty				
2525	FM179322	ena	exon	506286	506450	.	-	.	Parent=transcript:CAR86393;Name=CAR86393-1;const				
2526	FM179322	ena	CDS	506286	506450	.	-	0	ID=CDS:CAR86393;Parent=transcript:CAR86393;prote				
2527	###												
2528	FM179322	Rfam	gene	506642	506842	.	-	.	ID=gene:EBG00001128470;Name=rli28;biotype=sRNA;				
2529	FM179322	Rfam	transcript	506642	506842	.	-	.	ID=transcript:EBT00001719480;Parent=gene:EBG00001				
2530	FM179322	Rfam	exon	506642	506842	.	-	.	Parent=transcript:EBT00001719480;Name=EBG000011				
2531	###												
2532	FM179322	Rfam	gene	507097	507297	.	-	.	ID=gene:EBG00001128509;Name=rli28;biotype=sRNA;				
2533	FM179322	Rfam	transcript	507097	507297	.	-	.	ID=transcript:EBT00001719564;Parent=gene:EBG00001				
2534	FM179322	Rfam	exon	507097	507297	.	-	.	Parent=transcript:EBT00001719564;Name=EBG000011				
2535	###												
2536	FM179322	ena	gene	507589	509007	.	-	.	ID=gene:LGG_00499;Name=is27;biotype=protein_codin				
2537	FM179322	ena	mRNA	507589	509007	.	-	.	ID=transcript:CAR86394;Parent=gene:LGG_00499;Nan				
2538	FM179322	ena	exon	507589	509007	.	-	.	Parent=transcript:CAR86394;Name=CAR86394-1;const				

W10-3:9列目

W8-4の画面上でID Attributeのところ指定したgene_idが、カウントデータを取得した際に、用いる行名情報に相当。

	A	B	C	D	E	F	G	H	I	J	K	L	M
2519	FM179322	ena	mRNA	505972	506328	.	-	.	ID=transcript:CAR86392;Parent=gene:LGG_00497;biotype=transcript;Name=CAR86392-1;logic_name=car86392-1				
2520	FM179322	ena	exon	505972	506328	.	-	.	Parent=transcript:CAR86392;Name=CAR86392-1;logic_name=car86392-1				
2521	FM179322	ena	CDS	505972	506328	.	-	0	ID=CDS:CAR86392;Parent=transcript:CAR86392;logic_name=car86392				
2522	###												
2523	FM179322	ena	gene	506286	506450	.	-	.	ID=gene:LGG_00498;biotype=protein_coding;description=LGG_00498				
2524	FM179322	ena	mRNA	506286	506450	.	-	.	ID=transcript:CAR86393;Parent=gene:LGG_00498;biotype=transcript;Name=CAR86393-1;logic_name=car86393-1				
2525	FM179322	ena	exon	506286	506450	.	-	.	Parent=transcript:CAR86393;Name=CAR86393-1;logic_name=car86393-1				
2526	FM179322	ena	CDS	506286	506450	.	-	0	ID=CDS:CAR86393;Parent=transcript:CAR86393;logic_name=car86393				
2527	###												
2528	FM179322	Rfam	gene	506642	506842	.	-	.	ID=gene:EBG00001128470;Name=rli28;biotype=sRNA;logic_name=rfam_genes				
2529	FM179322	Rfam	transcript	506642	506842	.	-	.	ID=transcript:EBT00001719480;Parent=gene:EBG00001128470;logic_name=ebt00001719480				
2530	FM179322	Rfam	exon	506642	506842	.	-	.	Parent=transcript:EBT00001719480;Name=EBG00001128470;logic_name=rfam_genes				
2531	###												
2532	FM179322	Rfam	gene	507097	507297	.	-	.	ID=gene:EBG00001128509;Name=rli28;biotype=sRNA;logic_name=rfam_genes				
2533	FM179322	Rfam	transcript	507097	507297	.	-	.	ID=transcript:EBT00001719564;Parent=gene:EBG00001128509;logic_name=ebt00001719564				
2534	FM179322	Rfam	exon	507097	507297	.	-	.	Parent=transcript:EBT00001719564;Name=EBG00001128509;logic_name=rfam_genes				
2535	###												
2536	FM179322	ena	gene	507589	509007	.	-	.	ID=gene:LGG_00499;Name=is27;biotype=protein_coding;description=LGG_00499				
2537	FM179322	ena	mRNA	507589	509007	.	-	.	ID=transcript:CAR86394;Parent=gene:LGG_00499;Name=CAR86394-1;logic_name=car86394-1				
2538	FM179322	ena	exon	507589	509007	.	-	.	Parent=transcript:CAR86394;Name=CAR86394-1;logic_name=car86394-1				

W10-3:9列目

指定したgene_idは、GFFファイル中の①9列目 (attribute列と呼ばれる) に存在します。例えば、EBG00001128470の検索結果としてハイライトされている②のセルの場合は、③のところにある。ゆえに④がカウントデータの行名として使われる

	A	B	C	D	E	F	G	H	I	J	K	L	M
2519	FM179322	ena	mRNA	505972	506328	.	-	.	ID=transcript:CAR86392;Parent=gene:LGG_00497;biotype=transcript;Name=CAR86392-1;constitutive_exon				
2520	FM179322	ena	exon	505972	506328	.	-	.	Parent=transcript:CAR86392;Name=CAR86392-1;constitutive_exon				
2521	FM179322	ena	CDS	505972	506328	.	-	0	ID=CDS:CAR86392;Parent=transcript:CAR86392;protein_coding				
2522	###												
2523	FM179322	ena	gene	506286	506450	.	-	.	ID=gene:LGG_00498;biotype=protein_coding;description=ribosomal_L16S				
2524	FM179322	ena	mRNA	506286	506450	.	-	.	ID=transcript:CAR86393;Parent=gene:LGG_00498;biotype=transcript;Name=CAR86393-1;constitutive_exon				
2525	FM179322	ena	exon	506286	506450	.	-	.	Parent=transcript:CAR86393;Name=CAR86393-1;constitutive_exon				
2526	FM179322	ena	CDS	506286	506450	.	-	0	ID=CDS:CAR86393;Parent=transcript:CAR86393;protein_coding				
2527	###												
2528	FM179322	Rfam	gene	506642	506842	.	-	.	ID=gene:EBG00001128470;Name=rli28;biotype=sRNA;description=23S_rRNA				
2529	FM179322	Rfam	transcript	506642	506842	.	-	.	ID=transcript:EBT00001719480;Parent=gene:EBG00001128470;Name=EBT00001719480-1;constitutive_exon				
2530	FM179322	Rfam	exon	506642	506842	.	-	.	Parent=transcript:EBT00001719480;Name=EBG00001128470				
2531	###												
2532	FM179322	Rfam	gene	507097	507297	.	-	.	ID=gene:EBG00001128509;Name=rli28;biotype=sRNA;description=23S_rRNA				
2533	FM179322	Rfam	transcript	507097	507297	.	-	.	ID=transcript:EBT00001719564;Parent=gene:EBG00001128509;Name=EBT00001719564-1;constitutive_exon				
2534	FM179322	Rfam	exon	507097	507297	.	-	.	Parent=transcript:EBT00001719564;Name=EBG00001128509				
2535	###												
2536	FM179322	ena	gene	507589	509007	.	-	.	ID=gene:LGG_00499;Name=is27;biotype=protein_coding;description=ribosomal_L16S				
2537	FM179322	ena	mRNA	507589	509007	.	-	.	ID=transcript:CAR86394;Parent=gene:LGG_00499;Name=CAR86394-1;constitutive_exon				
2538	FM179322	ena	exon	507589	509007	.	-	.	Parent=transcript:CAR86394;Name=CAR86394-1;constitutive_exon				

W10-4: 情報提供元

この画面上では、①EBG00001128470と②EBG00001128509が見られる。

	A	B	C	D	E	F	G	H	I	J	K	L	M
2519	FM179322	ena	mRNA	505972	506328	.	-	.	ID=transcript:CAR86392;Parent=gene:LGG_00497;bioty				
2520	FM179322	ena	exon	505972	506328	.	-	.	Parent=transcript:CAR86392;Name=CAR86392-1;const				
2521	FM179322	ena	CDS	505972	506328	.	-	0	ID=CDS:CAR86392;Parent=transcript:CAR86392;prote				
2522	###												
2523	FM179322	ena	gene	506286	506450	.	-	.	ID=gene:LGG_00498;biotype=protein_coding;description				
2524	FM179322	ena	mRNA	506286	506450	.	-	.	ID=transcript:CAR86393;Parent=gene:LGG_00498;bioty				
2525	FM179322	ena	exon	506286	506450	.	-	.	Parent=transcript:CAR86393;Name=CAR86393-1;const				
2526	FM179322	ena	CDS	506286	506450	.	-	0	ID=CDS:CAR86393;Parent=transcript:CAR86393;prote				
2527	###												
2528	FM179322	Rfam	gene	506642	506842	.	-	.	ID=gene:EBG00001128470;Name=rli28;biotype=sRNA;				
2529	FM179322	Rfam	transcript	506642	506842	.	-	.	ID=transcript:EBT00001719480;Parent=gene:EBG00001				
2530	FM179322	Rfam	exon	506642	506842	.	-	.	Parent=transcript:EBT00001719480;Name=EBG000011				
2531	###												
2532	FM179322	Rfam	gene	507097	507297	.	-	.	ID=gene:EBG00001128509;Name=rli28;biotype=sRNA;				
2533	FM179322	Rfam	transcript	507097	507297	.	-	.	ID=transcript:EBT00001719564;Parent=gene:EBG00001				
2534	FM179322	Rfam	exon	507097	507297	.	-	.	Parent=transcript:EBT00001719564;Name=EBG000011				
2535	###												
2536	FM179322	ena	gene	507589	509007	.	-	.	ID=gene:LGG_00499;Name=is27;biotype=protein_codin				
2537	FM179322	ena	mRNA	507589	509007	.	-	.	ID=transcript:CAR86394;Parent=gene:LGG_00499;Nan				
2538	FM179322	ena	exon	507589	509007	.	-	.	Parent=transcript:CAR86394;Name=CAR86394-1;const				



W10-4: 情報提供元

この画面上では、①EBG00001128470と②EBG00001128509が見られる。2列目に記載されている、③これらの情報提供元(source)は、いずれもRfamであることがわかる。その一方で、LGGから始まるIDのもののsourceはena。

自動保存 ASM2650v1.gff3 - Exo

ファイル ホーム 挿入 ページレイアウト 数式 データ 校閲 表示

I2532 ID=gene:EBG00001128509;Name=rli28;biotype=sRNA;gene_id=EBG00001128509;logic_name=rfam_genes

	A	B	C	D	E	F	G	H	I	J	K	L	M
2519	FM179322	ena	mRNA	505972	506328	.	-	.	ID=transcript:CAR86392;Parent=gene:LGG_00497;biotype=protein_coding;description=				
2520	FM179322	ena	exon	505972	506328	.	-	.	Parent=transcript:CAR86392;Name=CAR86392-1;constitutive_element=				
2521	FM179322	ena	CDS	505972	506328	.	-	0	ID=CDS:CAR86392;Parent=transcript:CAR86392;protein_coding=				
2522	###												
2523	FM179322	ena	gene	506286	506450	.	-	.	ID=gene:LGG_00498;biotype=protein_coding;description=				
2524	FM179322	ena	mRNA	506286	506450	.	-	.	ID=transcript:CAR86393;Parent=gene:LGG_00498;biotype=protein_coding;description=				
2525	FM179322	ena	exon	506286	506450	.	-	.	Parent=transcript:CAR86393;Name=CAR86393-1;constitutive_element=				
2526	FM179322	ena	CDS	506286	506450	.	-	0	ID=CDS:CAR86393;Parent=transcript:CAR86393;protein_coding=				
2527	###												
2528	FM179322	Rfam	gene	506642	506842	.	-	.	ID=gene:EBG00001128470;Name=rli28;biotype=sRNA;description=				
2529	FM179322	Rfam	transcript	506642	506842	.	-	.	ID=transcript:EBT00001719480;Parent=gene:EBG00001128470;Name=EBT00001719480;biotype=				
2530	FM179322	Rfam	exon	506642	506842	.	-	.	Parent=transcript:EBT00001719480;Name=EBG00001128470;biotype=sRNA;description=				
2531	###												
2532	FM179322	Rfam	gene	507097	507297	.	-	.	ID=gene:EBG00001128509;Name=rli28;biotype=sRNA;description=				
2533	FM179322	Rfam	transcript	507097	507297	.	-	.	ID=transcript:EBT00001719564;Parent=gene:EBG00001128509;Name=EBT00001719564;biotype=				
2534	FM179322	Rfam	exon	507097	507297	.	-	.	Parent=transcript:EBT00001719564;Name=EBG00001128509;biotype=sRNA;description=				
2535	###												
2536	FM179322	ena	gene	507589	509007	.	-	.	ID=gene:LGG_00499;Name=is27;biotype=protein_coding;description=				
2537	FM179322	ena	mRNA	507589	509007	.	-	.	ID=transcript:CAR86394;Parent=gene:LGG_00499;Name=CAR86394;biotype=protein_coding;description=				
2538	FM179322	ena	exon	507589	509007	.	-	.	Parent=transcript:CAR86394;Name=CAR86394-1;constitutive_element=				

ASM2650v1 - + 100%

準備完了 ScrollLock

Contents

- W1: 公共データベースENA
- W2: クラウド解析環境Galaxy
- W3: ENAからGalaxyへのFASTQファイルの送付
- W4: クオリティチェック (FastQC)
- W5: クオリティコントロール (Trimmomatic)
- W6: クオリティチェック (FastQC)
- W7: ゲノム配列へのマッピング (Bowtie2)
- W8: カウント情報取得 (htseq-count)
- W9: カウント情報の連結 (Column join)
- W10: カウントデータの数値行列の行名について
- W11: 公共カウントデータ (GSE107337) を用いて、倍率変化と \log_2 (倍率変化)を算出
- W12: Galaxy上で得られたカウントデータについても同様に計算し、比較
- W13とW14: 原著論文中の記載事項を確認
- W15: 2つのカウントデータから共通遺伝子の情報を用いて類似度を計算

W11-1: GSE107337

原著論文著者らがGEO上で公開している、①
GSE107337のページ。②下部に移動。

NCBI > GEO > **Accession Display** Not logged in | [Login](#)

GEO help: Mouse over screen elements for information.

Scope: Format: Amount: GEO accession:

Series GSE107337 [Query DataSets for GSE107337](#)

Status	Public on Nov 29, 2017
Title	RNA-seq analysis of Lactobacillus at acidic stress
Organism	Lactobacillus rhamnosus GG
Experiment type	Expression profiling by high throughput sequencing
Summary	To understand transcriptional regulation of probiotic bacteria under acidic condition, RNAseq analysis was carried out over different growth conditions
Overall design	Comparison of acidic (pH4) and neutral (pH7) conditions by differentially expressed genes
Contributor(s)	Choi I, Oh S
Citation missing	<i>Has this study been published? Please login to update or notify GEO.</i>
Submission date	Nov 25, 2017
Last update date	May 15, 2019
Contact name	kucsbl submitter
E-mail(s)	kucsbl.group@gmail.com
Organization name	Korea University

<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE107337>

W11-1: GSE107337

原著論文著者らがGEO上で公開している、① GSE107337のページ。②下部に移動。③をクリックして得られたのが、次のスライドのファイル。

GEO Accession viewer

ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE107337

Platforms (1) [GPL24302](#) Illumina MiSeq (Lactobacillus rhamnosus GG)

Samples (9) [GSM2864941](#) pH4_1h rep1
[GSM2864942](#) pH4_1h rep2
[GSM2864943](#) pH4_1h rep3

Relations

BioProject [PRJNA419802](#)
SRA [SRP125628](#)

Download family

Download family	Format
SOFT formatted family file(s)	SOFT ?
MINiML formatted family file(s)	MINiML ?
Series Matrix File(s)	TXT ?

Supplementary file	Size	Download	File type/resource
GSE107337_RPKM.csv.gz	21.1 Kb	(ftp) (http)	CSV
GSE107337_RawCounts.csv.gz	20.8 Kb	(ftp) (http)	CSV

[SRA Run Selector](#) [?](#)

Raw data are available in SRA
Processed data are available on Series record

[NLM](#) | [NIH](#) | [GEO Help](#) | [Disclaimer](#) | [Accessibility](#)

<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE107337>

原著論文著者らがGEO上で公開している、①
GSE107337のraw countデータ。

W11-1: GSE107337

	A	B	C	D
1	Gene	pH4_1h_average_rawcount	pH4_24h_average_rawcount	pH7_CCG_average_rawcount
2	LGG_00001	286	676	545
3	LGG_00002	776	1601	1450
4	LGG_00003	71	114	91
5	LGG_00004	201	470	368
6	LGG_00005	368	1160	1447
7	LGG_00006	1041	2742	2880
8	LGG_00007	7	10	2
9	LGG_00008	73	76	499
10	LGG_00009	16	45	42
11	LGG_00010	20	65	57
12	LGG_00011	2281	2596	2150
13	LGG_00012	2510	3631	3097
14	LGG_00013	971	1301	1424

ftp://ftp.ncbi.nlm.nih.gov/geo/series/GSE107nnn/GSE107337/suppl/GSE107337_RawCounts.csv.gz



W11-2: 平均値

原著論文著者らがGEO上で公開している、① GSE107337のraw countデータ。②各群について3反復の平均値だということがわかる。

自動保存 GSE107337_RawCounts.csv - Excel 門田 幸二

ファイル ホーム 挿入 ページレイアウト 数式 データ 校閲 表示 ヘルプ 検索 共有

F13

	A	②	②	②
1	Gene	pH4_1h_average_rawcount	pH4_24h_average_rawcount	pH7_CCG_average_rawcount
2	LGG_00001	286	676	545
3	LGG_00002	776	1601	1450
4	LGG_00003	71	114	91
5	LGG_00004	201	470	368
6	LGG_00005	368	1160	1447
7	LGG_00006	1041	2742	2880
8	LGG_00007	7	10	2
9	LGG_00008	73	76	499
10	LGG_00009	16	45	42
11	LGG_00010	20	65	57
12	LGG_00011	2281	2596	2150
13	LGG_00012	2510	3631	3097
14	LGG_00013	971	1301	1424

GSE107337_RawCounts

ScrollLock 100%

W11-2: 列名変更

原著論文著者らがGEO上で公開している、① GSE107337のraw countデータ。②各群について3反復の平均値だということがわかる。平均値だということはわかったので、列名を短くしておく。

	A	B	C	D
1	Gene	pH4_1h	pH4_24h	pH7_CCG
2	LGG_00001	286	676	545
3	LGG_00002	776	1601	1450
4	LGG_00003	71	114	91
5	LGG_00004	201	470	368
6	LGG_00005	368	1160	1447
7	LGG_00006	1041	2742	2880
8	LGG_00007	7	10	2
9	LGG_00008	73	76	499
10	LGG_00009	16	45	42
11	LGG_00010	20	65	57
12	LGG_00011	2281	2596	2150
13	LGG_00012	2510	3631	3097
14	LGG_00013	971	1301	1424

W11-3: 倍率変化

①コントロール(中性状態)に対する、②短期(1hなのでshort-term)の酸性ストレスの倍率変化を計算したのが③。④LGG_00001の場合は、 $286/545 = 0.5248$ 。

	A	B	C	D	E	F
1	Gene	pH4_1h	pH4_24h	pH7_CCG	pH4_1h/CCG	pH4_24h/CCG
2	LGG_00001	286	676	545	0.5248	1.2404
3	LGG_00002	776	1601	1450	0.5352	1.1041
4	LGG_00003	71	114	91	0.7802	1.2527
5	LGG_00004	201	470	368	0.5462	1.2772
6	LGG_00005	368	1160	1447	0.2543	0.8017
7	LGG_00006	1041	2742	2880	0.3615	0.9521
8	LGG_00007	7	10	2	3.5000	5.0000
9	LGG_00008	73	76	499	0.1463	0.1523
10	LGG_00009	16	45	42	0.3810	1.0714
11	LGG_00010	20	65	57	0.3509	1.1404
12	LGG_00011	2281	2596	2150	1.0609	1.2074
13	LGG_00012	2510	3631	3097	0.8105	1.1724
14	LGG_00013	971	1301	1444	0.6819	0.9136

W11-3: 倍率変化

①コントロール(中性状態)に対する、②長期(24hなのでlong-term)の酸性ストレスの倍率変化を計算したのが③。④LGG_00001の場合は $676/545 = 1.2404$ 。

	A	B	C	D	E	F
1	Gene	pH4_1h	pH4_24h	pH7_CCG	pH4_1h/CCG	pH4_24h/CCG
2	LGG_00001	286	676	545	0.5248	1.2404
3	LGG_00002	776	1601	1450	0.5352	1.1041
4	LGG_00003	71	114	91	0.7802	1.2527
5	LGG_00004	201	470	368	0.5462	1.2772
6	LGG_00005	368	1160	1447	0.2543	0.8017
7	LGG_00006	1041	2742	2880	0.3615	0.9521
8	LGG_00007	7	10	2	3.5000	5.0000
9	LGG_00008	73	76	499	0.1463	0.1523
10	LGG_00009	16	45	42	0.3810	1.0714
11	LGG_00010	20	65	57	0.3509	1.1404
12	LGG_00011	2281	2596	2150	1.0609	1.2074
13	LGG_00012	2510	3631	3097	0.8105	1.1724
14	LGG_00013	971	1301	1444	0.6819	0.9136

W11-4: log比

「①コントロール(中性状態)に対する短期(1hなのでshort-term)の酸性ストレスの倍率変化」の、
②対数(底は2)。LGG_00001の場合は、③
0.5248の \log_2 をとるということ。つまり、④
 $\log_2(0.5248) = -0.9302$ 。

自動保存 オ GSE107337_RawCounts

ファイル ホーム 挿入 ページレイアウト 数式 データ 校閲 表示

G2 =LOG(E2,2)

	A	B	C	D	E	F	G	H
1	Gene	pH4_1h	pH4_24h	pH7_CCG	pH4_1h/CCG	pH4_24h/CCG	$\log_2(\text{pH4_1h/CCG})$	$\log_2(\text{pH4_24h/CCG})$
2	LGG_00001	286	676	545	0.5248	1.2404	-0.9302	0.3108
3	LGG_00002	776	1601	1450	0.5352	1.1041	-0.9019	0.1429
4	LGG_00003	71	114	91	0.7802	1.2527	-0.3580	0.3251
5	LGG_00004	201	470	368	0.5462	1.2772	-0.8725	0.3530
6	LGG_00005	368	1160	1447	0.2543	0.8017	-1.9753	-0.3189
7	LGG_00006	1041	2742	2880	0.3615	0.9521	-1.4681	-0.0708
8	LGG_00007	7	10	2	3.5000	5.0000	1.8074	2.3219
9	LGG_00008	73	76	499	0.1463	0.1523	-2.7731	-2.7150
10	LGG_00009	16	45	42	0.3810	1.0714	-1.3923	0.0995
11	LGG_00010	20	65	57	0.3509	1.1404	-1.5110	0.1895
12	LGG_00011	2281	2596	2150	1.0609	1.2074	0.0853	0.2720
13	LGG_00012	2510	3631	3097	0.8105	1.1724	-0.3032	0.2295
14	LGG_00013	971	1301	1424	0.6819	0.9136	-0.5524	-0.1303

GSE107337_RawCounts 100%

W11-4: log比

「①コントロール(中性状態)に対する長期(24hなのでlong-term)の酸性ストレスの倍率変化」の、②対数(底は2)。LGG_00001の場合は、③1.2404の \log_2 をとるということ。つまり、④ $\log_2(1.2404) = 0.3108$ 。

自動保存 オ GSE107337_RawCounts

ファイル ホーム 挿入 ページレイアウト 数式 データ 校閲 表示

H2 =LOG(F2,2)

	A	B	C	D	E	F	G	H
1	Gene	pH4_1h	pH4_24h	pH7_CCG	pH4_1h/CCG	pH4_24h/CCG	$\log_2(\text{pH4_1h/CCG})$	$\log_2(\text{pH4_24h/CCG})$
2	LGG_00001	286	676	545	0.5248	1.2404	-0.9302	0.3108
3	LGG_00002	776	1601	1450	0.5352	1.1041	-0.9019	0.1429
4	LGG_00003	71	114	91	0.7802	1.2527	-0.3580	0.3251
5	LGG_00004	201	470	368	0.5462	1.2772	-0.8725	0.3530
6	LGG_00005	368	1160	1447	0.2543	0.8017	-1.9753	-0.3189
7	LGG_00006	1041	2742	2880	0.3615	0.9521	-1.4681	-0.0708
8	LGG_00007	7	10	2	3.5000	5.0000	1.8074	2.3219
9	LGG_00008	73	76	499	0.1463	0.1523	-2.7731	-2.7150
10	LGG_00009	16	45	42	0.3810	1.0714	-1.3923	0.0995
11	LGG_00010	20	65	57	0.3509	1.1404	-1.5110	0.1895
12	LGG_00011	2281	2596	2150	1.0609	1.2074	0.0853	0.2720
13	LGG_00012	2510	3631	3097	0.8105	1.1724	-0.3032	0.2295
14	LGG_00013	971	1301	1424	0.6819	0.9136	-0.5524	-0.1303

GSE107337_RawCounts 100%

W11-5:LGG_02240

おさらい。今眺めているものは、著者らがGEO上で公開しているGSE107337のraw countデータから、我々が倍率変化や \log_2 (倍率変化)を再計算したもの。①LGG_02240の結果。

自動保存 印刷 戻る 進む GSE107337_RawCounts

ファイル ホーム 挿入 ページレイアウト 数式 データ 校閲 表示 ヘルプ 検索 共有

A2158 : LGG_02240

	A	B	C	D	E	F	G	H
1	Gene	pH4_1h	pH4_24h	pH7_CCG	pH4_1h/CCG	pH4_24h/CCG	$\log_2(\text{pH4_1h/CCG})$	$\log_2(\text{pH4_24h/CCG})$
2152	LGG_02233	227	331	539	0.4212	0.6141	-1.2476	-0.7035
2153	LGG_02234	406	791	1093	0.3715	0.7237	-1.4287	-0.4665
2154	LGG_02235	582	999	1289	0.4515	0.7750	-1.1472	-0.3677
2155	LGG_02236	277	467	647	0.4281	0.7218	-1.2239	-0.4703
2156	LGG_02238	768	674	577	1.3310	1.1681	0.4125	0.2242
2157	LGG_02239	5146	18308	5690	0.9044	3.2176	-0.1450	1.6860
①	LGG_02240	1106	2444	422	2.6209	5.7915	1.3900	2.5339
2159	LGG_02241	87	107	130	0.6692	0.8231	-0.5794	-0.2809
2160	LGG_02242	377	340	540	0.6981	0.6296	-0.5184	-0.6674
2161	LGG_02244	36	116	44	0.8182	2.6364	-0.2895	1.3985
2162	LGG_02245	19	101	29	0.6552	3.4828	-0.6101	1.8002
2163	LGG_02246	23	63	8	2.8750	7.8750	1.5236	2.9773
2164	LGG_02247	384	531	556	0.6906	0.9550	-0.5340	-0.0664
2165	LGG_02248	245	405	318	0.7704	1.2736	-0.3762	0.3489

GSE107337_RawCounts 表示設定 100%

W11-5:LGG_02240

おさらい。今眺めているものは、著者らがGEO上で公開しているGSE107337のraw countデータから、我々が倍率変化や \log_2 (倍率変化)を再計算したもの。①LGG_02240の結果。原著論文中で、②2.53-fold、③1.39-foldと書かれていた数値は、我々が再計算した結果中の、 \log_2 (倍率変化)の数値と完全に一致。

	A	B	C	D	E	F	G	H
1	Gene	pH4_1h	pH4_24h	pH7_CCG	pH4_1h/CCG	pH4_24h/CCG	\log_2 (pH4_1h/CCG)	\log_2 (pH4_24h/CCG)
2152	LGG_02233	227	331	539	0.4212	0.6141	-1.2476	-0.7035
2153	LGG_02234	406	791	1093	0.3715	0.7237	-1.4287	-0.4665
2154	LGG_02235	582	999	1289	0.4515	0.7750	-1.1472	-0.3677
2155	LGG_02236	277	467	647	0.4281	0.7218	-1.2239	-0.4703
2156	LGG_02238	768	674	577	1.3310	1.1681	0.4125	0.2242
2157	LGG_02239	5146	18308	5690	0.9044	3.2176	-0.1450	1.6860
①	LGG_02240	1106	2444	422	2.6209	5.7915	1.3900	2.5339
2159	LGG_02241	87	107	130	0.6692	0.8231	-0.1821	-0.1559
2160	LGG_02242	377	340	540	0.6981	0.6296	-0.5184	-0.6674
2161	LGG_02244	36	116	44	0.8182	2.6364	-0.2895	1.3985
2162	LGG_02245	19	101	29	0.6552	3.4828	-0.6101	1.8002
2163	LGG_02246	23	63	8	2.8750	7.8750	1.5236	2.9773
2164	LGG_02247	384	531	556	0.6906	0.9550	-0.5340	-0.0664
2165	LGG_02248	245	405	318	0.7704	1.2736	-0.3762	0.3489

W11-5:LGG_02372

おさらい。今眺めているものは、著者らがGEO上で公開しているGSE107337のraw countデータから、我々が倍率変化や \log_2 (倍率変化)を再計算したもの。①LGG_02473の結果。原著論文中で、②3.34-fold、③1.34-foldと書かれていた数値は、我々が再計算した結果中の、 \log_2 (倍率変化)の数値と完全に一致。

	A	B	C	D	E	F	G	H
1	Gene	pH4_1h	pH4_24h	pH7_CCG	pH4_1h/CCG	pH4_24h/CCG	\log_2 (pH4_1h/CCG)	\log_2 (pH4_24h/CCG)
2280	LGG_02366	91	138	79	1.1519	1.7468	0.2040	0.8047
2281	LGG_02367	15	31	11	1.3636	2.8182	0.4475	1.4948
2282	LGG_02368	10	31	7	1.4286	4.4286	0.5146	2.1468
2283	LGG_02369	12	46	5	2.4000	9.2000	1.2630	3.2016
2284	LGG_02370	15	64	7	2.1429	9.1429	1.0995	3.1926
2285	LGG_02371	14	68	6	2.3333	11.3333	1.2224	3.5025
①	LGG_02372	33	132	13	2.5385	10.1538	1.3440	3.3440
2287	LGG_02373	103	330	97	1.0619	3.4021	0.3556	1.5444
2288	LGG_02374	37	109	41	0.9024	2.6585	-0.1481	1.4106
2289	LGG_02375	314	339	773	0.4062	0.4386	-1.2997	-1.1892
2290	LGG_02376	52	69	48	1.0833	1.4375	0.1155	0.5236
2291	LGG_02377	33	58	18	1.8333	3.2222	0.8745	1.6881
2292	LGG_02378	42	85	49	0.8571	1.7347	-0.2224	0.7947
2293	LGG_02379	42	85	64	0.6563	1.3281	-0.6077	0.4094

Contents

- W1: 公共データベースENA
- W2: クラウド解析環境Galaxy
- W3: ENAからGalaxyへのFASTQファイルの送付
- W4: クオリティチェック (FastQC)
- W5: クオリティコントロール (Trimmomatic)
- W6: クオリティチェック (FastQC)
- W7: ゲノム配列へのマッピング (Bowtie2)
- W8: カウント情報取得 (htseq-count)
- W9: カウント情報の連結 (Column join)
- W10: カウントデータの数値行列の行名について
- W11: 公共カウントデータ (GSE107337) を用いて、倍率変化と \log_2 (倍率変化)を算出
- W12: Galaxy上で得られたカウントデータについても同様に計算し、比較
- W13とW14: 原著論文中の記載事項を確認
- W15: 2つのカウントデータから共通遺伝子の情報を用いて類似度を計算

W12-1:9 samples

W9-10までと同様の手順で、9サンプル分のカウントデータ取得まで行った結果。

The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D	E	F	G	H	I	J
1		pH4.5_1h			pH4.5_24h			pH7_CCG		
2	gene_ID	rep1	rep2	rep3	rep1	rep2	rep3	rep1	rep2	rep3
3	EBG00001128470	7	133	130	69	158	75	124	161	68
4	EBG00001128476	3	10	15	15	44	17	9	9	1
5	EBG00001128500	25	73	99	58	176	52	107	52	21
6	EBG00001128509	11	166	166	72	206	97	154	211	69
7	EBG00001128529	0	0	0	0	0	0	0	0	0
8	LGG_00001	58	141	195	158	491	248	392	252	76
9	LGG_00002	113	413	513	412	1233	589	911	842	263
10	LGG_00003	1	29	21	20	69	32	29	31	8
11	LGG_00004	17	93	138	128	335	171	255	174	60
12	LGG_00005	53	159	239	272	882	389	878	814	258
13	LGG_00006	204	585	680	742	2230	910	1849	1640	549
14	LGG_00007	0	3	6	0	11	5	3	1	0
15	LGG_00008	5	43	54	15	63	22	253	279	154
16	LGG_00009	3	11	11	6	50	19	42	24	6

W12-1:9 samples

W9-10までと同様の手順で、9サンプル分のカウントデータ取得まで行った結果。赤枠の3サンプル分のみのカウントデータ(W9-10)も表示。

自動保存 data_9samples.xlsx... 門田 幸二

ファイル ホーム 挿入 ページレイアウト 数式 データ 校閲 表示 ヘルプ 検索

A1 : X ✓ fx

	A	B	C	D	E	F	G	H	I	J
1		pH4.5_1h			pH4.5_24h			pH7_CCG		
2	gene_ID	rep1	rep2	rep3	rep1	rep2	rep3	rep1	rep2	rep3
3	EBG00001128470	7	133	130	69	158	75	124	161	68
4	EBG00001128476	3	10	15	15	44	17	9	9	1
5	EBG00001128500	25	73	99	58	176	52	107	52	21
6	EBG00001128509	11	166	166	72	206	97	154	211	69
7	EBG00001128529	0	0	0	0	0	0	0	0	0
8	LGG_00001	58	141	195	158	491	248	392	252	76
9	LGG_00002	113	413	513	412	1233				
10	LGG_00003	1	29	21	20	69				
11	LGG_00004	17	93	138	128	335				
12	LGG_00005	53	159	239	272	882				
13	LGG_00006	204	585	680	742	2230				
14	LGG_00007	0	3	6	0	11				
15	LGG_00008	5	43	54	15	63				
16	LGG_00009	3	11	11	6	50				

gene_ID	pH4.5_1h_rep3	pH4.5_24h_rep3	pH7_CCG_rep2
EBG00001128470	130	75	161
EBG00001128476	15	17	9
EBG00001128500	99	52	52
EBG00001128509	166	97	211
EBG00001128529	0	0	0
LGG_00001	195	248	252
LGG_00002	513	589	842
LGG_00003	21	32	31
LGG_00004	138	171	174
LGG_00005	239	389	814

W12-1:9 samples

W9-10までと同様の手順で、9サンプル分のカウントデータ取得まで行った結果。赤枠の3サンプル分のみのカウントデータ(W9-10)も表示。対応する列の数値は、確かに一致している。

自動保存 data_9samples.xlsx... 門田 幸二

ファイル ホーム 挿入 ページレイアウト 数式 データ 校閲 表示 ヘルプ 検索

	A	B	C	D	E	F	G	H	I	J
1		pH4.5_1h			pH4.5_24h			pH7_CCG		
2	gene_ID	rep1	rep2	rep3	rep1	rep2	rep3	rep1	rep2	rep3
3	EBG00001128470	7	133	130	69	158	75	124	161	68
4	EBG00001128476	3	10	15	15	44	17	9	9	1
5	EBG00001128500	25	73	99	58	176	52	107	52	21
6	EBG00001128509	11	166	166	72	206	97	154	211	69
7	EBG00001128529	0	0	0	0	0	0	0	0	0
8	LGG_00001	58	141	195	158	491	248	392	252	
9	LGG_00002	113	413	513	412	1233				
10	LGG_00003	1	29	21	20	69				
11	LGG_00004	17	93	138	128	335				
12	LGG_00005	53	159	239	272	882				
13	LGG_00006	204	585	680	742	2230				
14	LGG_00007	0	3	6	0	11				
15	LGG_00008	5	43	54	15	63				
16	LGG_00009	3	11	11	6	50				

gene_ID	pH4.5_1h_rep3	pH4.5_24h_rep3	pH7_CCG_rep2
EBG00001128470	130	75	161
EBG00001128476	15	17	9
EBG00001128500	99	52	52
EBG00001128509	166	97	211
EBG00001128529	0	0	0
LGG_00001	195	248	252
LGG_00002	513	589	842
LGG_00003	21	32	31
LGG_00004	138	171	174
LGG_00005	239	389	814

W12-2: sum

	A	B	C	D	E	F	G	H
1	gene_ID	pH4.5_1h	pH4.5_24h	pH7_CCG				
2	EBG00001128470	270	302	353				
3	EBG00001128476	28	76	19				
4	EBG00001128500	197	286	180				
5	EBG00001128509	343	375	434				
6	EBG00001128529	0	0	0				
7	LGG_00001	394	897	720				
8	LGG_00002	1039	2234	2016				
9	LGG_00003	51	121	68				
10	LGG_00004	248	634	489				
11	LGG_00005	451	1543	1950				
12	LGG_00006	1469	3882	4038				
13	LGG_00007	9	16	4				
14	LGG_00008	102	100	686				
15	LGG_00009	25	75	72				
16	LGG_00010	21	85	73				

W12-2: sum

同一群内の3反復の和(sum)をとった結果。赤枠の原著論文のほうはaverage countであったが(W11-2)、averageとsumは数学的に等価と考えてよいので、似た数値分布になるsumのほうを採用した。

	A	B	C	D	E	F	G
1	gene_ID	pH4.5_1h	pH4.5_24h	pH7_CCG			
2	EBG00001128470	270	302	353			
3	EBG00001128476	28	76	19			
4	EBG00001128500	197	286	180			
5	EBG00001128509	343	375	434			
6	EBG00001128529	0	0	0			
7	LGG_00001	394	897	720			
8	LGG_00002	1039	2234	2016			
9	LGG_00003	51	121	68			
10	LGG_00004	248	634	489			
11	LGG_00005	451	1543	1950			
12	LGG_00006	1469	3882	4038			
13	LGG_00007	9	16	4			
14	LGG_00008	102	100	686			
15	LGG_00009	25	75	72			
16	LGG_00010	21	85	73			

	A	B	C	D
1	Gene	pH4_1h	pH4_24h	pH7_CCG
2	LGG_00001	286	676	545
3	LGG_00002	776	1601	1450
4	LGG_00003	71	114	91
5	LGG_00004	201	470	368
6	LGG_00005	368	1160	1447
7	LGG_00006	1041	2742	2880
8	LGG_00007	7	10	2
9	LGG_00008	73	76	499
10	LGG_00009	16	45	42
11	LGG_00010	20	65	57
12	LGG_00011	2281	2596	2150
13	LGG_00012	2510	3631	3097
14	LGG_00013	971	1301	1424

W12-3: 倍率変化

自動保存 検索

data_9samples.xlsx - 保存しました 門田 幸二

ファイル ホーム 挿入 ページレイアウト 数式 データ 校閲 表示 ヘルプ

E7 =B7/D7

	A	B	C	D	E	F	G	H
	gene_ID	pH4.5_1h	pH4.5_24h	pH7_CCG	pH4_1h/CCG	pH4_24h/CCG	log ₂ (pH4_1h/CCG)	log ₂ (pH4_24h/CCG)
1	EBG00001128470	270	302	353	0.7649	0.8555	-0.3867	-0.2251
2	EBG00001128476	28	76	19	1.4737	4.0000	0.5594	2.0000
3	EBG00001128500	197	286	180	1.0944	1.5889	0.1302	0.6680
4	EBG00001128509	343	375	434	0.7903	0.8641	-0.3395	-0.2108
5	EBG00001128529	0	0	0	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
6	LGG_00001	394	897	720	0.5472	1.2458	-0.8698	0.3171
7	LGG_00002	1039	2234	2016	0.5154	1.1081	-0.9563	0.1481
8	LGG_00003	51	121	68	0.7500	1.7794	-0.4150	0.8314
9	LGG_00004	248	634	489	0.5072	1.2965	-0.9795	0.3746
10	LGG_00005	451	1543	1950	0.2313	0.7913	-2.1123	-0.3377
11	LGG_00006	1469	3882	4038	0.3638	0.9614	-1.4588	-0.0568
12	LGG_00007	9	16	4	2.2500	4.0000	1.1699	2.0000
13	LGG_00008	102	100	686	0.1487	0.1458	-2.7496	-2.7782
14	LGG_00009	25	75	72	0.3472	1.0417	-1.5261	0.0589
15	LGG_00010	21	85	73	0.2877	1.1644	-1.7975	0.2196

data_9samples summed averaged

表示設定 100%

赤枠が \log_2 (倍率変化)算出部分。

W12-4: \log_2 (倍率変化)

The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D	E	F	G	H
1	gene_ID	pH4.5_1h	pH4.5_24h	pH7_CCG	pH4_1h/CCG	pH4_24h/CCG	$\log_2(\text{pH4_1h/CCG})$	$\log_2(\text{pH4_24h/CCG})$
2	EBG00001128470	270	302	353	0.7649	0.8555	-0.3867	-0.2251
3	EBG00001128476	28	76	19	1.4737	4.0000	0.5594	2.0000
4	EBG00001128500	197	286	180	1.0944	1.5889	0.1302	0.6680
5	EBG00001128509	343	375	434	0.7903	0.8641	-0.3395	-0.2108
6	EBG00001128529	0	0	0	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
7	LGG_00001	394	897	720	0.5472	1.2458	-0.8698	0.3171
8	LGG_00002	1039	2234	2016	0.5154	1.1081	-0.9563	0.1481
9	LGG_00003	51	121	68	0.7500	1.7794	-0.4150	0.8314
10	LGG_00004	248	634	489	0.5072	1.2965	-0.9795	0.3746
11	LGG_00005	451	1543	1950	0.2313	0.7913	-2.1123	-0.3377
12	LGG_00006	1469	3882	4038	0.3638	0.9614	-1.4588	-0.0568
13	LGG_00007	9	16	4	2.2500	4.0000	1.1699	2.0000
14	LGG_00008	102	100	686	0.1487	0.1458	-2.7496	-2.7782
15	LGG_00009	25	75	72	0.3472	1.0417	-1.5261	0.0589
16	LGG_00010	21	85	73	0.2877	1.1644	-1.7975	0.2196

W12-5:LGG_02240

Excel spreadsheet showing the calculation results for LGG_02240. The spreadsheet is titled "data_9samples.xlsx" and is open in Excel. The active cell is A2246, containing the formula "LGG_02240".

	A	B	C	D	E	F	G	H
1	gene_ID	pH4.5_1h	pH4.5_24h	pH7_CCG	pH4_1h/CCG	pH4_24h/CCG	log ₂ (pH4_1h/CCG)	log ₂ (pH4_24h/CCG)
2239	LGG_02233	313	483	752	0.4162	0.6423	-1.2646	-0.6387
2240	LGG_02234	563	1064	1452	0.3877	0.7328	-1.3668	-0.4485
2241	LGG_02235	820	1379	1751	0.4683	0.7875	-1.0945	-0.3446
2242	LGG_02236	401	691	939	0.4271	0.7359	-1.2275	-0.4424
2243	LGG_02237	55	87	79	0.6962	1.1013	-0.5224	0.1392
2244	LGG_02238	1130	1014	810	1.3951	1.2519	0.4803	0.3241
2245	LGG_02239	6952	25000	7758	0.8961	3.2225	-0.1583	1.6882
2246	LGG_02240	1231	2914	490	2.5122	5.9469	1.3290	2.5721
2247	LGG_02241	121	147	174	0.6954	0.8448	-0.5241	-0.2433
2248	LGG_02242	668	567	1063	0.6284	0.5334	-0.6702	-0.9067
2249	LGG_02243	163	131	419	0.3890	0.3126	-1.3621	-1.6774
2250	LGG_02244	47	143	47	1.0000	3.0426	0.0000	1.6053
2251	LGG_02245	18	115	35	0.5143	3.2857	-0.9594	1.7162
2252	LGG_02246	22	61	8	2.7500	7.6250	1.4594	2.9307
2253	LGG_02247	534	752	809	0.6601	0.9295	-0.5993	-0.1054

The spreadsheet also shows tabs for "data_9samples", "summed", and "averaged". The zoom level is set to 100%.

W12-5:LGG_02372

Excel spreadsheet showing the calculation results for LGG_02372. The spreadsheet is titled "data_9samples.xlsx" and is displayed in the "summed" worksheet. The data is organized in columns A through H, representing various parameters for different gene IDs. The row for LGG_02372 is highlighted in red, and a red arrow points to the gene ID cell (A2377).

	A	B	C	D	E	F	G	H
1	gene_ID	pH4.5_1h	pH4.5_24h	pH7_CCG	pH4_1h/CCG	pH4_24h/CCG	log ₂ (pH4_1h/CCG)	log ₂ (pH4_24h/CCG)
2371	LGG_02365	219	345	297	0.7374	1.1616	-0.4395	0.2161
2372	LGG_02366	123	179	104	1.1827	1.7212	0.2421	0.7834
2373	LGG_02367	22	46	14	1.5714	3.2857	0.6521	1.7162
2374	LGG_02368	19	49	8	2.3750	6.1250	1.2479	2.6147
2375	LGG_02369	15	69	6	2.5000	11.5000	1.3219	3.5236
2376	LGG_02370	20	91	11	1.8182	8.2727	0.8625	3.0484
2377	LGG_02371	22	90	10	2.2000	9.0000	1.1375	3.1699
2378	LGG_02372	56	188	15	3.7333	12.5333	1.9005	3.6477
2379	LGG_02373	131	490	131	1.0000	3.7405	0.0000	1.9032
2380	LGG_02374	49	139	52	0.9423	2.6731	-0.0857	1.4185
2381	LGG_02375	416	445	1096	0.3796	0.4060	-1.3976	-1.3004
2382	LGG_02376	74	102	73	1.0137	1.3973	0.0196	0.4826
2383	LGG_02377	35	68	26	1.3462	2.6154	0.4288	1.3870
2384	LGG_02378	53	108	79	0.6709	1.3671	-0.5759	0.4511
2385	LGG_02379	67	153	116	0.5776	1.3190	-0.7919	0.3994

Contents

- W1: 公共データベースENA
- W2: クラウド解析環境Galaxy
- W3: ENAからGalaxyへのFASTQファイルの送付
- W4: クオリティチェック (FastQC)
- W5: クオリティコントロール (Trimmomatic)
- W6: クオリティチェック (FastQC)
- W7: ゲノム配列へのマッピング (Bowtie2)
- W8: カウント情報取得 (htseq-count)
- W9: カウント情報の連結 (Column join)
- W10: カウントデータの数値行列の行名について
- W11: 公共カウントデータ (GSE107337) を用いて、倍率変化と \log_2 (倍率変化)を算出
- W12: Galaxy上で得られたカウントデータについても同様に計算し、比較
- W13とW14: 原著論文中の記載事項を確認
- W15: 2つのカウントデータから共通遺伝子の情報を用いて類似度を計算

W13-1:LGG_00113

原著論文のp1608-1609にかけての文で述べられている、①LGG_00113の酸ストレス条件下における発現量低下を②公共データで確認。

	A	B	C	D	E	F	G	H
1	Gene	pH4_1h	pH4_24h	pH7_CCG	pH4_1h/CCG	pH4_24h/CCG	log ₂ (pH4_1h/CCG)	log ₂ (pH4_24h/CCG)
107	LGG_00107	132	151	134	0.9851	1.1269	-0.0217	0.1723
108	LGG_00108	92	351	912	0.1009	0.3849	-3.3093	-1.3776
109	LGG_00109	119	402	1615	0.0737	0.2489	-3.7625	-2.0063
110	LGG_00110	165	392	2403	0.0687	0.1631	-3.8643	-2.6159
111	LGG_00111	72	127	1101	0.0654	0.1153	-3.9347	-3.1159
112	LGG_00112	178	341	1967	0.0905	0.1734	-3.4660	-2.5282
113	LGG_00113	176	277	1159	0.1519	0.2390	-2.7192	-2.0649
114	LGG_00114	99	220	940	0.1053	0.2340	-3.2472	-2.0952
115	LGG_00115	136	296	860	0.1581	0.3442	-2.6607	-1.5387
116	LGG_00116	202	198	298	0.6779	0.6644	-0.5610	-0.5898
117	LGG_00117	99	93	37	2.6757	2.5135	1.4199	1.3297
118	LGG_00118	13	31	8	1.6250	3.8750	0.7004	1.9542
119	LGG_00119	32	79	17	1.8824	4.6471	0.9125	2.2163
120	LGG_00120	52	78	30	1.7333	2.6000	0.7935	1.3785

W13-1:LGG_00113

原著論文のp1608-1609にかけての文で述べられている、①LGG_00113の酸ストレス条件下における発現量低下を②自分らのデータで確認。

自動保存 門田 幸二 data_9samples.xlsx

ファイル ホーム 挿入 ページレイアウト 数式 データ 校閲 表示 ヘルプ 検索 共有

A119 : X ✓ fx LGG_00113

	A	B	C	D	E	F	G	H
1	gene_ID	pH4.5_1h	pH4.5_24h	pH7_CCG	pH4_1h/CCG	pH4_24h/CCG	log ₂ (pH4_1h/CCG)	log ₂ (pH4_24h/CCG)
112	LGG_00106	78	225	44	1.7727	5.1136	0.8260	2.3543
113	LGG_00107	187	232	196	0.9541	1.1837	-0.0678	0.2433
114	LGG_00108	120	490	1226	0.0979	0.3997	-3.3529	-1.3231
115	LGG_00109	147	554	2275	0.0646	0.2435	-3.9520	-2.0379
116	LGG_00110	221	536	3400	0.0650	0.1576	-3.9434	-2.6652
117	LGG_00111	80	167	1331	0.0601	0.1255	-4.0564	-2.9946
118	LGG_00112	284	459	2715	0.1046	0.1691	-3.2570	-2.5644
119	LGG_00113	224	368	1561	0.1435	0.2357	-2.8009	-2.0847
120	LGG_00114	107	296	1200	0.0892	0.2467	-3.4874	-2.0194
121	LGG_00115	186	409	1192	0.1560	0.3431	-2.6800	-1.5432
122	LGG_00116	261	275	432	0.6042	0.6366	-0.7270	-0.6516
123	LGG_00117	161	125	52	3.0962	2.4038	1.6305	1.2653
124	LGG_00118	16	46	9	1.7778	5.1111	0.8301	2.3536
125	LGG_00119	36	95	22	1.6364	4.3182	0.7105	2.1104
126	LGG_00120	66	97	46	1.4348	2.1087	0.5208	1.0764

data_9samples summed averaged

表示設定 100%

W13-2:LGG_00115

原著論文のp1608-1609にかけての文で述べられている、①LGG_00115の酸ストレス条件下における発現量低下を②公共データで確認。

自動保存 GSE107337_RawCounts.xlsx 門田 幸二

ファイル ホーム 挿入 ページレイアウト 数式 データ 校閲 表示 ヘルプ 検索 共有

A115 LGG_00115

	A	B	C	D	E	F	G	H
1	Gene	pH4_1h	pH4_24h	pH7_CCG	pH4_1h/CCG	pH4_24h/CCG	log ₂ (pH4_1h/CCG)	log ₂ (pH4_24h/CCG)
107	LGG_00107	132	151	134	0.9851	1.1269	-0.0217	0.1723
108	LGG_00108	92	351	912	0.1009	0.3849	-3.3093	-1.3776
109	LGG_00109	119	402	1615	0.0737	0.2489	-3.7625	-2.0063
110	LGG_00110	165	392	2403	0.0687	0.1631	-3.8643	-2.6159
111	LGG_00111	72	127	1101	0.0654	0.1153	-3.9347	-3.1159
112	LGG_00112	178	341	1967	0.0905	0.1734	-3.4660	-2.5282
113	LGG_00113	176	277	1159	0.1519	0.2390	-2.7192	-2.0649
114	LGG_00114	99	220	940	0.1053	0.2340	-3.2472	-2.0952
115	LGG_00115	136	296	860	0.1581	0.3442	-2.6607	-1.5387
116	LGG_00116	202	198	298	0.6779	0.6644	-0.5610	-0.5898
117	LGG_00117	99	93	37	2.6757	2.5135	1.4199	1.3297
118	LGG_00118	13	31	8	1.6250	3.8750	0.7004	1.9542
119	LGG_00119	32	79	17	1.8824	4.6471	0.9125	2.2163
120	LGG_00120	52	78	30	1.7333	2.6000	0.7935	1.3785

GSE107337_RawCounts 表示設定 100%

W13-2:LGG_00115

原著論文のp1608-1609にかけての文で述べられている、①LGG_00115の酸ストレス条件下における発現量低下を②自分らのデータで確認。

Excel spreadsheet showing data for LGG_00115. The table is titled 'LGG_00115' and contains the following data:

	A	B	C	D	E	F	G	H
1	gene_ID	pH4.5_1h	pH4.5_24h	pH7_CCG	pH4_1h/CCG	pH4_24h/CCG	log ₂ (pH4_1h/CCG)	log ₂ (pH4_24h/CCG)
112	LGG_00106	78	225	44	1.7727	5.1136	0.8260	2.3543
113	LGG_00107	187	232	196	0.9541	1.1837	-0.0678	0.2433
114	LGG_00108	120	490	1226	0.0979	0.3997	-3.3529	-1.3231
115	LGG_00109	147	554	2275	0.0646	0.2435	-3.9520	-2.0379
116	LGG_00110	221	536	3400	0.0650	0.1576	-3.9434	-2.6652
117	LGG_00111	80	167	1331	0.0601	0.1255	-4.0564	-2.9946
118	LGG_00112	284	459	2715	0.1046	0.1691	-3.2570	-2.5644
119	LGG_00113	224	368	1561	0.1435	0.2357	-2.8009	-2.0847
120	LGG_00114	107	296	1200	0.0892	0.2467	-3.4874	-2.0194
121	LGG_00115	186	409	1192	0.1560	0.3431	-2.6800	-1.5432
122	LGG_00116	261	275	432	0.6042	0.6366	-0.7270	-0.6516
123	LGG_00117	161	125	52	3.0962	2.4038	1.6305	1.2653
124	LGG_00118	16	46	9	1.7778	5.1111	0.8301	2.3536
125	LGG_00119	36	95	22	1.6364	4.3182	0.7105	2.1104
126	LGG_00120	66	97	46	1.4348	2.1087	0.5208	1.0764

W13-3:LGG_00108

原著論文のp1608-1609にかけての文で述べられている、①LGG_00108の酸ストレス条件下における発現量低下を②公共データで確認。

	A	B	C	D	E	F	G	H
1	Gene	pH4_1h	pH4_24h	pH7_CCG	pH4_1h/CCG	pH4_24h/CCG	log ₂ (pH4_1h/CCG)	log ₂ (pH4_24h/CCG)
107	LGG_00107	132	151	134	0.9851	1.1269	-0.0217	0.1723
108	LGG_00108	92	351	912	0.1009	0.3849	-3.3093	-1.3776
109	LGG_00109	119	402	1615	0.0737	0.2489	-3.7625	-2.0063
110	LGG_00110	165	392	2403	0.0687	0.1631	-3.8643	-2.6159
111	LGG_00111	72	127	1101	0.0654	0.1153	-3.9347	-3.1159
112	LGG_00112	178	341	1967	0.0905	0.1734	-3.4660	-2.5282
113	LGG_00113	176	277	1159	0.1519	0.2390	-2.7192	-2.0649
114	LGG_00114	99	220	940	0.1053	0.2340	-3.2472	-2.0952
115	LGG_00115	136	296	860	0.1581	0.3442	-2.6607	-1.5387
116	LGG_00116	202	198	298	0.6779	0.6644	-0.5610	-0.5898
117	LGG_00117	99	93	37	2.6757	2.5135	1.4199	1.3297
118	LGG_00118	13	31	8	1.6250	3.8750	0.7004	1.9542
119	LGG_00119	32	79	17	1.8824	4.6471	0.9125	2.2163
120	LGG_00120	52	78	30	1.7333	2.6000	0.7935	1.3785

W13-3:LGG_00108

原著論文のp1608-1609にかけての文で述べられている、①LGG_00108の酸ストレス条件下における発現量低下を②自分らのデータで確認。

Excel spreadsheet showing gene expression data for LGG_00108. The row for LGG_00108 is highlighted in red. A red circle with the number 1 is next to the gene ID, and a red circle with the number 2 is next to the Excel title bar.

	A	B	C	D	E	F	G	H
1	gene_ID	pH4.5_1h	pH4.5_24h	pH7_CCG	pH4_1h/CCG	pH4_24h/CCG	log ₂ (pH4_1h/CCG)	log ₂ (pH4_24h/CCG)
112	LGG_00106	78	225	44	1.7727	5.1136	0.8260	2.3543
113	LGG_00107	187	232	196	0.9541	1.1837	-0.0678	0.2433
114	LGG_00108	120	490	1226	0.0979	0.3997	-3.3529	-1.3231
115	LGG_00109	147	554	2275	0.0646	0.2435	-3.9520	-2.0379
116	LGG_00110	221	536	3400	0.0650	0.1576	-3.9434	-2.6652
117	LGG_00111	80	167	1331	0.0601	0.1255	-4.0564	-2.9946
118	LGG_00112	284	459	2715	0.1046	0.1691	-3.2570	-2.5644
119	LGG_00113	224	368	1561	0.1435	0.2357	-2.8009	-2.0847
120	LGG_00114	107	296	1200	0.0892	0.2467	-3.4874	-2.0194
121	LGG_00115	186	409	1192	0.1560	0.3431	-2.6800	-1.5432
122	LGG_00116	261	275	432	0.6042	0.6366	-0.7270	-0.6516
123	LGG_00117	161	125	52	3.0962	2.4038	1.6305	1.2653
124	LGG_00118	16	46	9	1.7778	5.1111	0.8301	2.3536
125	LGG_00119	36	95	22	1.6364	4.3182	0.7105	2.1104
126	LGG_00120	66	97	46	1.4348	2.1087	0.5208	1.0764

W13-4:LGG_00109

原著論文のp1608-1609にかけての文で述べられている、①LGG_00109の酸ストレス条件下における発現量低下を②公共データで確認。

	A	B	C	D	E	F	G	H
1	Gene	pH4_1h	pH4_24h	pH7_CCG	pH4_1h/CCG	pH4_24h/CCG	log ₂ (pH4_1h/CCG)	log ₂ (pH4_24h/CCG)
107	LGG_00107	132	151	134	0.9851	1.1269	-0.0217	0.1723
108	LGG_00108	92	351	912	0.1009	0.3849	-3.3093	-1.3776
109	LGG_00109	119	402	1615	0.0737	0.2489	-3.7625	-2.0063
110	LGG_00110	165	392	2403	0.0687	0.1631	-3.8643	-2.6159
111	LGG_00111	72	127	1101	0.0654	0.1153	-3.9347	-3.1159
112	LGG_00112	178	341	1967	0.0905	0.1734	-3.4660	-2.5282
113	LGG_00113	176	277	1159	0.1519	0.2390	-2.7192	-2.0649
114	LGG_00114	99	220	940	0.1053	0.2340	-3.2472	-2.0952
115	LGG_00115	136	296	860	0.1581	0.3442	-2.6607	-1.5387
116	LGG_00116	202	198	298	0.6779	0.6644	-0.5610	-0.5898
117	LGG_00117	99	93	37	2.6757	2.5135	1.4199	1.3297
118	LGG_00118	13	31	8	1.6250	3.8750	0.7004	1.9542
119	LGG_00119	32	79	17	1.8824	4.6471	0.9125	2.2163
120	LGG_00120	52	78	30	1.7333	2.6000	0.7935	1.3785

W13-4:LGG_00109

原著論文のp1608-1609にかけての文で述べられている、①LGG_00109の酸ストレス条件下における発現量低下を②自分らのデータで確認。

Excel spreadsheet showing gene expression data for LGG_00109. The spreadsheet has columns for gene ID, pH4.5_1h, pH4.5_24h, pH7_CCG, pH4_1h/CCG, pH4_24h/CCG, log₂(pH4_1h/CCG), and log₂(pH4_24h/CCG). The row for LGG_00109 is highlighted in red. A red circle with the number 1 is next to the gene ID, and a red circle with the number 2 is next to the CCG column header.

	A	B	C	D	E	F	G	H
1	gene_ID	pH4.5_1h	pH4.5_24h	pH7_CCG	pH4_1h/CCG	pH4_24h/CCG	log ₂ (pH4_1h/CCG)	log ₂ (pH4_24h/CCG)
112	LGG_00106	78	225	44	1.7727	5.1136	0.8260	2.3543
113	LGG_00107	187	232	196	0.9541	1.1837	-0.0678	0.2433
114	LGG_00108	120	490	1226	0.0979	0.3997	-3.3529	-1.3231
115	LGG_00109	147	554	2275	0.0646	0.2435	-3.9520	-2.0379
116	LGG_00110	221	536	3400	0.0650	0.1576	-3.9434	-2.6652
117	LGG_00111	80	167	1331	0.0601	0.1255	-4.0564	-2.9946
118	LGG_00112	284	459	2715	0.1046	0.1691	-3.2570	-2.5644
119	LGG_00113	224	368	1561	0.1435	0.2357	-2.8009	-2.0847
120	LGG_00114	107	296	1200	0.0892	0.2467	-3.4874	-2.0194
121	LGG_00115	186	409	1192	0.1560	0.3431	-2.6800	-1.5432
122	LGG_00116	261	275	432	0.6042	0.6366	-0.7270	-0.6516
123	LGG_00117	161	125	52	3.0962	2.4038	1.6305	1.2653
124	LGG_00118	16	46	9	1.7778	5.1111	0.8301	2.3536
125	LGG_00119	36	95	22	1.6364	4.3182	0.7105	2.1104
126	LGG_00120	66	97	46	1.4348	2.1087	0.5208	1.0764

W14-1:LGG_01064

原著論文のp1611の左中央付近で述べられている、①LGG_01064の酸ストレス条件下における発現量上昇を②公共データで確認。

	A	B	C	D	E	F	G	H
1	Gene	pH4_1h	pH4_24h	pH7_CCG	pH4_1h/CCG	pH4_24h/CCG	log ₂ (pH4_1h/CCG)	log ₂ (pH4_24h/CCG)
1010	LGG_01058	19	35	46	0.4130	0.7609	-1.2756	-0.3943
1011	LGG_01059	6	24	10	0.6000	2.4000	-0.7370	1.2630
1012	LGG_01060	20	68	47	0.4255	1.4468	-1.2327	0.5329
1013	LGG_01061	1644	1981	2169	0.7580	0.9133	-0.3998	-0.1308
1014	LGG_01062	1204	3690	2140	0.5626	1.7243	-0.8298	0.7860
1015	LGG_01063	19	131	25	0.7600	5.2400	-0.3959	2.3896
1016	LGG_01064	27	268	11	2.4545	24.3636	1.2955	4.6067
1017	LGG_01065	33	389	8	4.1250	48.6250	2.0444	5.6036
1018	LGG_01066	144	707	152	0.9474	4.6513	-0.0780	2.2176
1019	LGG_01067	2	12	1	2.0000	12.0000	1.0000	3.5850
1020	LGG_01068	150	602	183	0.8197	3.2896	-0.2869	1.7179
1021	LGG_01069	228	940	271	0.8413	3.4686	-0.2493	1.7944
1022	LGG_01070	165	557	196	0.8418	2.8418	-0.2484	1.5068
1023	LGG_01071	92	343	105	0.8762	3.2667	-0.1907	1.7078

W14-1:LGG_01064

原著論文のp1611の左中央付近で述べられている、①LGG_01064の酸ストレス条件下における発現量上昇を②自分らのデータで確認。

Excel spreadsheet showing gene expression data for LGG_01064. The spreadsheet is titled "data_9samples.xlsx" and displays data for various genes across different pH conditions and CCG treatments. The row for LGG_01064 is highlighted in red.

	A	B	C	D	E	F	G	H
1	gene_ID	pH4.5_1h	pH4.5_24h	pH7_CCG	pH4_1h/CCG	pH4_24h/CCG	log ₂ (pH4_1h/CCG)	log ₂ (pH4_24h/CCG)
1063	LGG_01057	380	825	978	0.3885	0.8436	-1.3638	-0.2454
1064	LGG_01058	46	130	168	0.2738	0.7738	-1.8688	-0.3699
1065	LGG_01059	39	111	74	0.5270	1.5000	-0.9241	0.5850
1066	LGG_01060	35	88	52	0.6731	1.6923	-0.5712	0.7590
1067	LGG_01061	2258	2677	2986	0.7562	0.8965	-0.4032	-0.1576
1068	LGG_01062	1625	5164	3043	0.5340	1.6970	-0.9051	0.7630
1069	LGG_01063	26	229	44	0.5909	5.2045	-0.7590	2.3798
1070	LGG_01064	49	389	18	2.7222	21.6111	1.4448	4.4337
1071	LGG_01065	37	536	12	3.0833	44.6667	1.6245	5.4811
1072	LGG_01066	195	983	209	0.9330	4.7033	-0.1000	2.2337
1073	LGG_01067	2	52	2	1.0000	26.0000	0.0000	4.7004
1074	LGG_01068	233	871	247	0.9433	3.5263	-0.0842	1.8182
1075	LGG_01069	328	1359	355	0.9239	3.8282	-0.1141	1.9367
1076	LGG_01070	202	701	248	0.8145	2.8266	-0.2960	1.4991
1077	LGG_01071	130	445	152	0.8553	2.9276	-0.2256	1.5497

W14-2:LGG_02032

原著論文のp1611の左中央付近で述べられている、①LGG_02032の酸ストレス条件下における発現量低下を②公共データで確認。

自動保存 検索

GSE107337_RawCounts.xlsx 門田 幸二

ファイル ホーム 挿入 ページレイアウト 数式 データ 校閲 表示 ヘルプ 検索 共有

A1959 : X ✓ fx LGG_02032

	A	B	C	D	E	F	G	H
1	Gene	pH4_1h	pH4_24h	pH7_CCG	pH4_1h/CCG	pH4_24h/CCG	log ₂ (pH4_1h/CCG)	log ₂ (pH4_24h/CCG)
1953	LGG_02025	20	148	78	0.2564	1.8974	-1.9635	0.9241
1954	LGG_02026	26	144	95	0.2737	1.5158	-1.8694	0.6001
1955	LGG_02027	65	238	468	0.1389	0.5085	-2.8480	-0.9755
1956	LGG_02028	143	175	197	0.7259	0.8883	-0.4622	-0.1708
1957	LGG_02030	13	44	16	0.8125	2.7500	-0.2996	1.4594
1958	LGG_02031	88	152	180	0.4889	0.8444	-1.0324	-0.2439
1959	LGG_02032	86	207	617	0.1394	0.3355	-2.8429	-1.5756
1960	LGG_02033	59	179	95	0.6211	1.8842	-0.6872	0.9140
1961	LGG_02034	109	504	729	0.1495	0.6914	-2.7416	-0.5325
1962	LGG_02035	191	41092	381	0.5013	107.8530	-0.9962	6.7529
1963	LGG_02036	168	272	472	0.3559	0.5763	-1.4903	-0.7952
1964	LGG_02037	1734	1257	3475	0.4990	0.3617	-1.0029	-1.4670
1965	LGG_02038	338	371	482	0.7012	0.7697	-0.5120	-0.3776
1966	LGG_02039	62	93	84	0.7381	1.1071	-0.4381	0.1468

GSE107337_RawCounts

表示設定 100%

W14-2:LGG_02032

原著論文のp1611の左中央付近で述べられている、①LGG_02032の酸ストレス条件下における発現量低下を②自分らのデータで確認。

	A	B	C	D	E	F	G	H
1	gene_ID	pH4.5_1h	pH4.5_24h	pH7_CCG	pH4_1h/CCG	pH4_24h/CCG	log ₂ (pH4_1h/CCG)	log ₂ (pH4_24h/CCG)
2031	LGG_02025	20	183	99	0.2020	1.8485	-2.3074	0.8863
2032	LGG_02026	36	214	130	0.2769	1.6462	-1.8524	0.7191
2033	LGG_02027	83	336	652	0.1273	0.5153	-2.9737	-0.9564
2034	LGG_02028	230	289	305	0.7541	0.9475	-0.4072	-0.0777
2035	LGG_02029	40	51	65	0.6154	0.7846	-0.7004	-0.3499
2036	LGG_02030	18	65	19	0.9474	3.4211	-0.0780	1.7744
2037	LGG_02031	139	215	259	0.5367	0.8301	-0.8979	-0.2686
2038	LGG_02032	116	290	831	0.1396	0.3490	-2.8407	-1.5188
2039	LGG_02033	1	6	1	1.0000	6.0000	0.0000	2.5850
2040	LGG_02034	178	727	1057	0.1684	0.6878	-2.5700	-0.5399
2041	LGG_02035	242	58154	525	0.4610	110.7695	-1.1173	6.7914
2042	LGG_02036	260	409	666	0.3904	0.6141	-1.3570	-0.7034
2043	LGG_02037	2416	1779	4907	0.4924	0.3625	-1.0222	-1.4638
2044	LGG_02038	468	510	648	0.7222	0.7870	-0.4695	-0.3455
2045	LGG_02039	75	110	112	0.6696	0.9821	-0.5785	-0.0260

W14-3:LGG_00418

原著論文のp1611の左中央付近で述べられている、①LGG_00418の酸ストレス条件下における発現量低下を②公共データで確認。

自動保存 検索

GSE107337_RawCounts.xlsx 門田 幸二

ファイル ホーム 挿入 ページレイアウト 数式 データ 校閲 表示 ヘルプ 検索 共有

A402 : X ✓ fx LGG_00418

	A	B	C	D	E	F	G	H
1	Gene	pH4_1h	pH4_24h	pH7_CCG	pH4_1h/CCG	pH4_24h/CCG	log ₂ (pH4_1h/CCG)	log ₂ (pH4_24h/CCG)
396	LGG_00412	102	256	285	0.3579	0.8982	-1.4824	-0.1548
397	LGG_00413	14	47	9	1.5556	5.2222	0.6374	2.3847
398	LGG_00414	76	131	48	1.5833	2.7292	0.6630	1.4485
399	LGG_00415	80	289	580	0.1379	0.4983	-2.8580	-1.0050
400	LGG_00416	42	144	306	0.1373	0.4706	-2.8651	-1.0875
401	LGG_00417	209	730	2188	0.0955	0.3336	-3.3880	-1.5836
402	LGG_00418	156	443	1710	0.0912	0.2591	-3.4544	-1.9486
403	LGG_00419	65	178	569	0.1142	0.3128	-3.1299	-1.6766
404	LGG_00420	80	207	289	0.2768	0.7163	-1.8530	-0.4814
405	LGG_00421	11	44	70	0.1571	0.6286	-2.6699	-0.6699
406	LGG_00422	9	33	71	0.1268	0.4648	-2.9798	-1.1054
407	LGG_00423	25	102	42	0.5952	2.4286	-0.7485	1.2801
408	LGG_00424	25	59	21	1.1905	2.8095	0.2515	1.4903
409	LGG_00425	13	73	23	0.5652	3.1739	-0.8231	1.6663

GSE107337_RawCounts

表示設定 100%

W14-3:LGG_00418

原著論文のp1611の左中央付近で述べられている、①LGG_00418の酸ストレス条件下における発現量低下を②公共データで確認。

Excel spreadsheet showing gene expression data for LGG_00418. The row for LGG_00418 is highlighted in red. A red arrow points to the cell containing '1' in the first column, and another red arrow points to the cell containing '2' in the second column of the header row.

	A	B	C	D	E	F	G	H
1	gene_ID	pH4.5_1h	pH4.5_24h	pH7_CCG	pH4_1h/CCG	pH4_24h/CCG	log ₂ (pH4_1h/CCG)	log ₂ (pH4_24h/CCG)
417	LGG_00411	32	82	20	1.6000	4.1000	0.6781	2.0356
418	LGG_00412	2	9	7	0.2857	1.2857	-1.8074	0.3626
419	LGG_00413	16	64	14	1.1429	4.5714	0.1926	2.1926
420	LGG_00414	96	184	67	1.4328	2.7463	0.5189	1.4575
421	LGG_00415	111	389	818	0.1357	0.4756	-2.8815	-1.0723
422	LGG_00416	46	170	353	0.1303	0.4816	-2.9400	-1.0541
423	LGG_00417	282	980	3031	0.0930	0.3233	-3.4260	-1.6289
424	LGG_00418	222	609	2233	0.0994	0.2727	-3.3304	-1.8745
425	LGG_00419	98	230	789	0.1242	0.2915	-3.0092	-1.7784
426	LGG_00420	103	271	380	0.2711	0.7132	-1.8834	-0.4877
427	LGG_00421	9	54	87	0.1034	0.6207	-3.2730	-0.6881
428	LGG_00422	17	46	93	0.1828	0.4946	-2.4517	-1.0156
429	LGG_00423	28	149	58	0.4828	2.5690	-1.0506	1.3612
430	LGG_00424	32	68	26	1.2308	2.6154	0.2996	1.3870
431	LGG_00425	12	111	27	0.4444	4.1111	-1.1699	2.0395

Contents

- W1: 公共データベースENA
- W2: クラウド解析環境Galaxy
- W3: ENAからGalaxyへのFASTQファイルの送付
- W4: クオリティチェック (FastQC)
- W5: クオリティコントロール (Trimmomatic)
- W6: クオリティチェック (FastQC)
- W7: ゲノム配列へのマッピング (Bowtie2)
- W8: カウント情報取得 (htseq-count)
- W9: カウント情報の連結 (Column join)
- W10: カウントデータの数値行列の行名について
- W11: 公共カウントデータ (GSE107337) を用いて、倍率変化と \log_2 (倍率変化)を算出
- W12: Galaxy上で得られたカウントデータについても同様に計算し、比較
- W13とW14: 原著論文中の記載事項を確認
- W15: 2つのカウントデータから共通遺伝子の情報を用いて類似度を計算

W15-1 : data_original.txt

Gene	pH4_1h	pH4_24h	pH7_CCG
LGG_00001	286	676	545
LGG_00002	776	1601	1450
LGG_00003	71	114	91
LGG_00004	201	470	368
LGG_00005	368	1160	1447
LGG_00006	1041	2742	2880
LGG_00007	7	10	2
LGG_00008	73	76	499
LGG_00009	16	45	42
LGG_00010	20	65	57
LGG_00011	2281	2596	2150
LGG_00012	2510	3631	3097
LGG_00013	971	1301	1424
LGG_00014	17	78	16
LGG_00015	82	255	33
LGG_00016	30	149	21
LGG_00017	30	127	20

W15-2: data_galaxy.txt

gene_ID	pH4.5_1h	pH4.5_24h	pH7_CCG
EBG00001128470	270	302	353
EBG00001128476	28	76	19
EBG00001128500	197	286	180
EBG00001128509	343	375	434
EBG00001128529	0	0	0
LGG_00001	394	897	720
LGG_00002	1039	2234	2016
LGG_00003	51	121	68
LGG_00004	248	634	489
LGG_00005	451	1543	1950
LGG_00006	1469	3882	4038
LGG_00007	9	16	4
LGG_00008	102	100	686
LGG_00009	25	75	72
LGG_00010	21	85	73
LGG_00011	2536	3127	2407
LGG_00012	2667	4325	3536

W15-3: 類似度計算

```

RGui (64-bit)
ファイル 編集 閲覧 その他 パッケージ ウィンドウ ヘルプ Vignettes

R Console
> data_ori <- read.table("data_original.txt", header=TRUE, sep="\t", row.names=1)
> data_gal <- read.table("data_galaxy.txt", header=TRUE, sep="\t", row.names=1)
> dim(data_ori)
#行数と列数を表示
[1] 2838 3
> dim(data_gal)
#行数と列数を表示
[1] 2949 3
> common <- intersect(rownames(data_ori), rownames(data_gal)) #共通した行名情報を$
> length(common)
#要素数(遺伝子数)を表示
[1] 2838
> data <- cbind(data_ori[common, ], data_gal[common, ]) #列方向で結合した結果をda$
> cor(data, method="spearman")
#総当たりのSpearman相関係数を計算
      pH4_1h  pH4_24h  pH7_CCG  pH4.5_1h  pH4.5_24h  pH7_CCG
pH4_1h  1.0000000  0.9023052  0.8919116  0.9544823  0.8670647  0.8617587
pH4_24h  0.9023052  1.0000000  0.8183148  0.8583811  0.9544563  0.7889634
pH7_CCG  0.8919116  0.8183148  1.0000000  0.8533259  0.7849992  0.9662319
pH4.5_1h  0.9544823  0.8583811  0.8533259  1.0000000  0.9059912  0.8948864
pH4.5_24h 0.8670647  0.9544563  0.7849992  0.9059912  1.0000000  0.8279148
pH7_CCG  0.8617587  0.7889634  0.9662319  0.8948864  0.8279148  1.0000000
> |

```

公開カウントデータのほうで、総カウント数の最大と最小の間には、①2.18倍もの差がある！

W15-4: 総カウント数

```
RGui (64-bit)
ファイル 編集 閲覧 その他 パッケージ ウィンドウ ヘルプ Vignettes

R Console
> dim(data_ori) #行数と列数を表示
[1] 2838 3
> dim(data_gal) #行数と列数を表示
[1] 2949 3
> common <- intersect(rownames(data_ori), rownames(data_gal)) #共通した行名情報を$
> length(common) #要素数(遺伝子数)を表示
[1] 2838
> data <- cbind(data_ori[common, ], data_gal[common, ]) #列方向で結合した結果をda$
> cor(data, method="spearman") #総当たりのSpearman相関係数を計算
      pH4_1h  pH4_24h  pH7_CCG  pH4.5_1h  pH4.5_24h  pH7_CCG
pH4_1h  1.000000  0.9023052  0.8919116  0.9544823  0.8670647  0.8617587
pH4_24h  0.9023052  1.0000000  0.8183148  0.8583811  0.9544563  0.7889634
pH7_CCG  0.8919116  0.8183148  1.0000000  0.8533259  0.7849992  0.9662319
pH4.5_1h  0.9544823  0.8583811  0.8533259  1.0000000  0.9059912  0.8948864
pH4.5_24h  0.8670647  0.9544563  0.7849992  0.9059912  1.0000000  0.8279148
pH7_CCG  0.8617587  0.7889634  0.9662319  0.8948864  0.8279148  1.0000000
> colSums(data_ori)
      pH4_1h  pH4_24h  pH7_CCG
      774376  1670016  1690218
> max(colSums(data_ori)) / min(colSums(data_ori))
[1] 2.182684
> |
```

