

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 'EMBL-EBI' and menu items: 'Services', 'Research', 'Training', and 'About us'. Below this is a teal banner with the ENA logo (a green wave) and the text 'ENA European Nucleotide Archive'. A search bar is located on the right side of the banner, containing the text 'GSE107337'. Below the search bar, there are examples: 'Examples: BN000065, histone'. To the right of the search bar is a 'Search' button. Below the search bar, there are links for 'Advanced' and 'Sequence'. A red arrow labeled '①' points to the ENA logo. A red arrow labeled '②' points to the search bar. A red arrow labeled '③' points to the 'Search' button. Below the banner is a navigation bar with 'Home', 'Search & Browse', 'Submit & Update', 'Software', 'About ENA', and 'Support'. The main content area is divided into two columns. The left column has a large heading 'European Nucleotide Archive' and a paragraph: 'The European Nucleotide Archive (ENA) provides a comprehensive record of the world's nucleotide sequencing information, covering raw sequencing data, sequence assembly information and functional annotation. [More about ENA](#)'. Below this is another paragraph: 'Access to ENA data is provided though the browser, through search tools, large scale file download and through the API.' The right column has a 'Popular' section with a list of links: 'Submit and update', 'Sequence submissions', 'Genome assembly submissions', 'Submitting environmental sequences', 'Citing ENA data', 'Rest URLs for data retrieval', and 'Rest URLs to search ENA'. Below this is a 'Latest ENA news' section with a link: '08 Jul 2019: Release 140 of ENA's assembled/annotated sequences is now'. At the bottom left, there is a 'Text Search' section with a search bar and examples: 'Examples: BN000065, histone'.

# 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 (European Nucleotide Archive) website. The browser address bar shows the URL: <https://www.ebi.ac.uk/ena/data/view/PRJNA419802>. The page header includes the EMBL-EBI logo and navigation links for Services, Research, Training, and About us. The main content area features the ENA logo and a search bar with a 'Search' button. Below the search bar is a navigation menu with links for Home, Search & Browse, Submit & Update, Software, About ENA, and Support. The main content displays the study details for PRJNA419802, including the title 'RNA-seq analysis of Lactobacillus at acidic stress', view options (Project XML, Study XML), and download options (Project XML, Study XML). A table lists the study's metadata, including the Name, Submitting Centre, and Organism. A secondary accession number (SRP125628) is also listed. The description of the study is provided at the bottom.

EMBL-EBI

Services Research Training About us

ENA  
European Nucleotide Archive

Search

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

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	<a href="#">Lactobacillus rhamnosus GG</a>

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)
<a href="#">PRJNA419802</a>	<a href="#">SAMN08098216</a>	<a href="#">SRS2714081</a>	<a href="#">SRX3422361</a>	<a href="#">SRR6322562</a>	568703	<a href="#">Lactobacillus rhamnosus GG</a>	Illumina MiSeq	PAIRED	<a href="#">File 1</a> <a href="#">File 2</a>	<a href="#">File 1</a> <a href="#">File 2</a>
<a href="#">PRJNA419802</a>	<a href="#">SAMN08098215</a>	<a href="#">SRS2714083</a>	<a href="#">SRX3422362</a>	<a href="#">SRR6322563</a>	568703	<a href="#">Lactobacillus rhamnosus GG</a>	Illumina MiSeq	PAIRED	<a href="#">File 1</a> <a href="#">File 2</a>	<a href="#">File 1</a> <a href="#">File 2</a>
<a href="#">PRJNA419802</a>	<a href="#">SAMN08098214</a>	<a href="#">SRS2714082</a>	<a href="#">SRX3422363</a>	<a href="#">SRR6322564</a>	568703	<a href="#">Lactobacillus rhamnosus GG</a>	Illumina MiSeq	PAIRED	<a href="#">File 1</a> <a href="#">File 2</a>	<a href="#">File 1</a> <a href="#">File 2</a>
<a href="#">PRJNA419802</a>	<a href="#">SAMN08098213</a>	<a href="#">SRS2714084</a>	<a href="#">SRX3422364</a>	<a href="#">SRR6322565</a>	568703	<a href="#">Lactobacillus rhamnosus GG</a>	Illumina MiSeq	PAIRED	<a href="#">File 1</a> <a href="#">File 2</a>	<a href="#">File 1</a> <a href="#">File 2</a>
<a href="#">PRJNA419802</a>	<a href="#">SAMN08098212</a>	<a href="#">SRS2714085</a>	<a href="#">SRX3422365</a>	<a href="#">SRR6322566</a>	568703	<a href="#">Lactobacillus rhamnosus GG</a>	Illumina MiSeq	PAIRED	<a href="#">File 1</a> <a href="#">File 2</a>	<a href="#">File 1</a> <a href="#">File 2</a>
<a href="#">PRJNA419802</a>	<a href="#">SAMN08098211</a>	<a href="#">SRS2714086</a>	<a href="#">SRX3422366</a>	<a href="#">SRR6322567</a>	568703	<a href="#">Lactobacillus rhamnosus GG</a>	Illumina MiSeq	PAIRED	<a href="#">File 1</a> <a href="#">File 2</a>	<a href="#">File 1</a> <a href="#">File 2</a>

# 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)
<a href="#">PRJNA419802</a>	<a href="#">SAMN08098216</a>	<a href="#">SRS2714081</a>	<a href="#">SRX3422361</a>	<a href="#">SRR6322562</a>	568703	<a href="#">Lactobacillus rhamnosus GG</a>	Illumina MiSeq	PAIRED	<a href="#">File 1</a> <a href="#">File 2</a>	<a href="#">File 1</a> <a href="#">File 2</a>
<a href="#">PRJNA419802</a>	<a href="#">SAMN08098215</a>	<a href="#">SRS2714083</a>	<a href="#">SRX3422362</a>	<a href="#">SRR6322563</a>	568703	<a href="#">Lactobacillus rhamnosus GG</a>	Illumina MiSeq	PAIRED	<a href="#">File 1</a> <a href="#">File 2</a>	<a href="#">File 1</a> <a href="#">File 2</a>
<a href="#">PRJNA419802</a>	<a href="#">SAMN08098214</a>	<a href="#">SRS2714082</a>	<a href="#">SRX3422363</a>	<a href="#">SRR6322564</a>	568703	<a href="#">Lactobacillus rhamnosus GG</a>	Illumina MiSeq	PAIRED	<a href="#">File 1</a> <a href="#">File 2</a>	<a href="#">File 1</a> <a href="#">File 2</a>
<a href="#">PRJNA419802</a>	<a href="#">SAMN08098213</a>	<a href="#">SRS2714084</a>	<a href="#">SRX3422364</a>	<a href="#">SRR6322565</a>	568703	<a href="#">Lactobacillus rhamnosus GG</a>	Illumina MiSeq	PAIRED	<a href="#">File 1</a> <a href="#">File 2</a>	<a href="#">File 1</a> <a href="#">File 2</a>
<a href="#">PRJNA419802</a>	<a href="#">SAMN08098212</a>	<a href="#">SRS2714085</a>	<a href="#">SRX3422365</a>	<a href="#">SRR6322566</a>	568703	<a href="#">Lactobacillus rhamnosus GG</a>	Illumina MiSeq	PAIRED	<a href="#">File 1</a> <a href="#">File 2</a>	<a href="#">File 1</a> <a href="#">File 2</a>
<a href="#">PRJNA419802</a>	<a href="#">SAMN08098211</a>	<a href="#">SRS2714086</a>	<a href="#">SRX3422366</a>	<a href="#">SRR6322567</a>	568703	<a href="#">Lactobacillus rhamnosus GG</a>	Illumina MiSeq	PAIRED	<a href="#">File 1</a> <a href="#">File 2</a>	<a href="#">File 1</a> <a href="#">File 2</a>



# 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 main content area displays a welcome message and a 'Galaxy 101' tutorial banner. The left sidebar contains a 'Tools' section with a search bar and various tool categories. The right sidebar shows a 'History' section with a search bar and a message indicating that the history is empty.

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 a "Tools" section with a search bar and various tool categories like "Get Data", "Send Data", "Collection Operations", "GENERAL TEXT TOOLS", "GENOMIC FILE MANIPULATION", and "FASTA/FASTQ". 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" and the page title is "Galaxy".

# 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 '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 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 central content area features 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: '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 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 is visible, showing a search bar and a list of datasets. The top item is "Unnamed history". A red circle with the number "1" is placed over the "Unnamed history" entry. A tooltip box is overlaid on this entry, containing the text: "ヒストリーの名前を変更するにはクリック". Below the history panel, there is a blue information box that reads: "i ヒストリーは空です。 You can load your own data or get data from an external source". The left sidebar contains a "Tools" section with a search bar and various tool categories like "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". The browser address bar shows "usegalaxy.org".

# 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 interface is divided into three main sections:

- Tools:** A sidebar on the left 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".
- History:** A panel on the right showing a search bar for datasets. A dataset named "GSE107337\_3samples" is listed with "(empty)" below it. A red arrow with the number "2" points to this dataset name. 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".
- Content Area:** The main workspace contains a text block: "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". At the bottom, there is a "Tweets by @galaxyproject" section.

# 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

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)
<a href="#">PRJNA419802</a>	<a href="#">SAMN08098216</a>	<a href="#">SRS2714081</a>	<a href="#">SRX3422361</a>	<a href="#">SRR6322562</a>	568703	<a href="#">Lactobacillus rhamnosus GG</a>	Illumina MiSeq	PAIRED	<a href="#">File 1</a> <a href="#">File 2</a>	<a href="#">File 1</a> <a href="#">File 2</a>
<a href="#">PRJNA419802</a>	<a href="#">SAMN08098215</a>	<a href="#">SRS2714083</a>	<a href="#">SRX3422362</a>	<a href="#">SRR6322563</a>	568703	<a href="#">Lactobacillus rhamnosus GG</a>	Illumina MiSeq	PAIRED	<a href="#">File 1</a> <a href="#">File 2</a>	<a href="#">File 1</a> <a href="#">File 2</a>
<a href="#">PRJNA419802</a>	<a href="#">SAMN08098214</a>	<a href="#">SRS2714082</a>	<a href="#">SRX3422363</a>	<a href="#">SRR6322564</a>	568703	<a href="#">Lactobacillus rhamnosus GG</a>	Illumina MiSeq	PAIRED	<a href="#">File 1</a> <a href="#">File 2</a>	<a href="#">File 1</a> <a href="#">File 2</a>
<a href="#">PRJNA419802</a>	<a href="#">SAMN08098213</a>	<a href="#">SRS2714084</a>	<a href="#">SRX3422364</a>	<a href="#">SRR6322565</a>	568703	<a href="#">Lactobacillus rhamnosus GG</a>	Illumina MiSeq	PAIRED	<a href="#">File 1</a> <a href="#">File 2</a>	<a href="#">File 1</a> <a href="#">File 2</a>
<a href="#">PRJNA419802</a>	<a href="#">SAMN08098212</a>	<a href="#">SRS2714085</a>	<a href="#">SRX3422365</a>	<a href="#">SRR6322566</a>	568703	<a href="#">Lactobacillus rhamnosus GG</a>	Illumina MiSeq	PAIRED	<a href="#">File 1</a> <a href="#">File 2</a>	<a href="#">File 1</a> <a href="#">File 2</a>
<a href="#">PRJNA419802</a>	<a href="#">SAMN08098211</a>	<a href="#">SRS2714086</a>	<a href="#">SRX3422366</a>	<a href="#">SRR6322567</a>	568703	<a href="#">Lactobacillus rhamnosus GG</a>	Illumina MiSeq	PAIRED	<a href="#">File 1</a> <a href="#">File 2</a>	<a href="#">File 1</a> <a href="#">File 2</a>

# 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、②File 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。

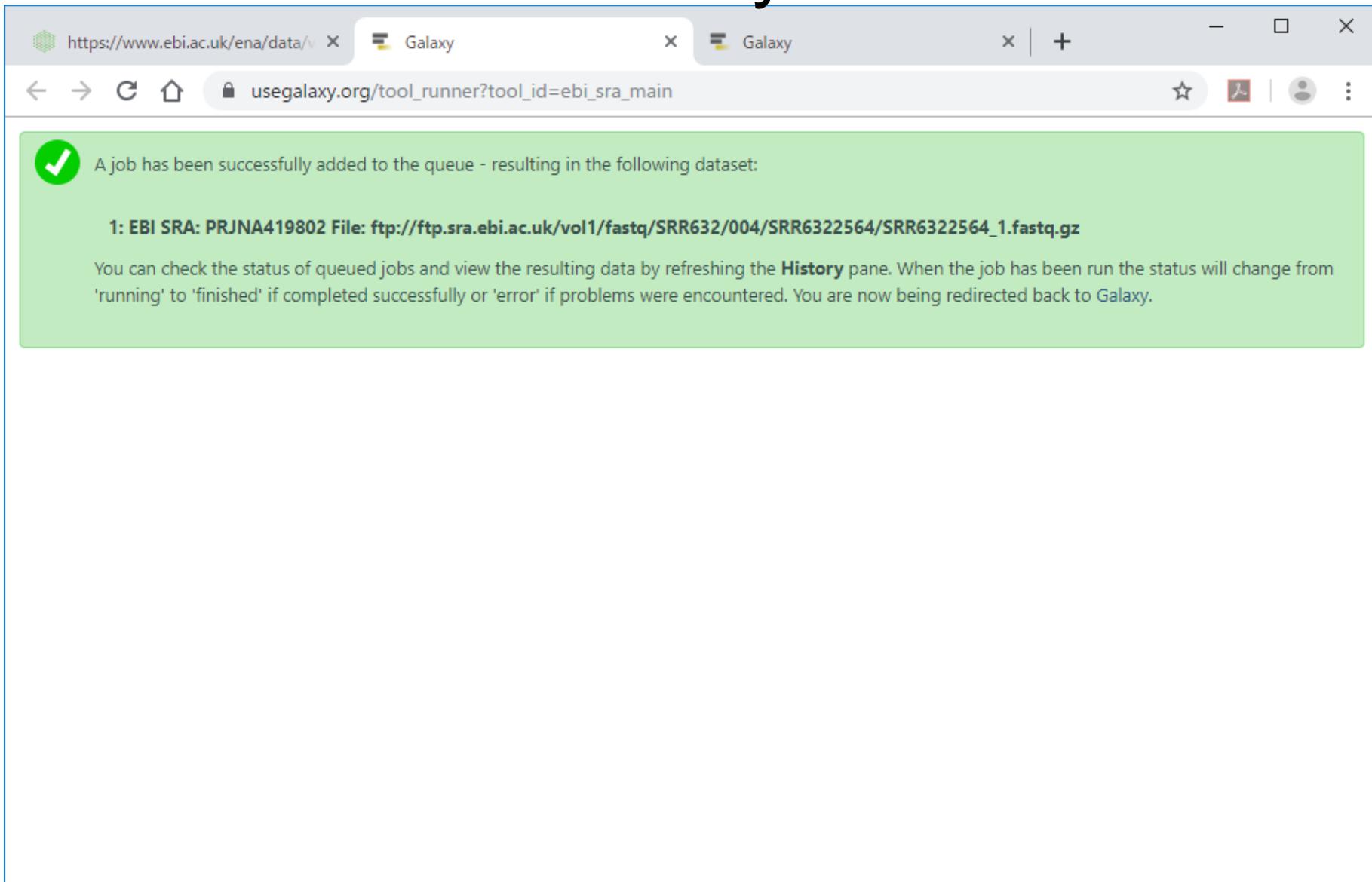
Showing results 1 - 9 of 9 results

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)
<a href="#">PRJNA419802</a>	<a href="#">SAMN08098216</a>	<a href="#">SRS2714081</a>	<a href="#">SRX3422361</a>	<a href="#">SRR6322562</a>	568703	<a href="#">Lactobacillus rhamnosus GG</a>	Illumina MiSeq	PAIRED	<a href="#">File 1</a> <a href="#">File 2</a>	<a href="#">File 1</a> <a href="#">File 2</a>
<a href="#">PRJNA419802</a>	<a href="#">SAMN08098215</a>	<a href="#">SRS2714083</a>	<a href="#">SRX3422362</a>	<a href="#">SRR6322563</a>	568703	<a href="#">Lactobacillus rhamnosus GG</a>	Illumina MiSeq	PAIRED	<a href="#">File 1</a> <a href="#">File 2</a>	<a href="#">File 1</a> <a href="#">File 2</a>
<a href="#">PRJNA419802</a>	<a href="#">SAMN08098214</a>	<a href="#">SRS2714082</a>	<a href="#">SRX3422363</a>	<a href="#">SRR6322564</a>	568703	<a href="#">Lactobacillus rhamnosus GG</a>	Illumina MiSeq	PAIRED	<a href="#">File 1</a> <a href="#">File 2</a>	<a href="#">File 1</a> <a href="#">File 2</a>
<a href="#">PRJNA419802</a>	<a href="#">SAMN08098213</a>	<a href="#">SRS2714084</a>	<a href="#">SRX3422364</a>	<a href="#">SRR6322565</a>	568703	<a href="#">Lactobacillus rhamnosus GG</a>	Illumina MiSeq	PAIRED	<a href="#">File 1</a> <a href="#">File 2</a>	<a href="#">File 1</a> <a href="#">File 2</a>
<a href="#">PRJNA419802</a>	<a href="#">SAMN08098212</a>	<a href="#">SRS2714085</a>	<a href="#">SRX3422365</a>	<a href="#">SRR6322566</a>	568703	<a href="#">Lactobacillus rhamnosus GG</a>	Illumina MiSeq	PAIRED	<a href="#">File 1</a> <a href="#">File 2</a>	<a href="#">File 1</a> <a href="#">File 2</a>
<a href="#">PRJNA419802</a>	<a href="#">SAMN08098211</a>	<a href="#">SRS2714086</a>	<a href="#">SRX3422366</a>	<a href="#">SRR6322567</a>	568703	<a href="#">Lactobacillus rhamnosus GG</a>	Illumina MiSeq	PAIRED	<a href="#">File 1</a> <a href="#">File 2</a>	<a href="#">File 1</a> <a href="#">File 2</a>



すぐにこんな感じになる。

# 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 "Collection Operations". A top navigation bar contains "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 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 item is "GSE107337\_3samples" with a red arrow labeled "2" pointing to it. Below it, the first history item is expanded, showing "1: EBI SRA: PRJNA41980" and "2 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR632/004/SRR6322564/SRR6322564\_1.fastq.gz" with a red arrow labeled "1" pointing to it. The left sidebar contains a "Tools" section with a search bar and various tool categories like "Get Data", "Send Data", "Collection Operations", "GENERAL TEXT TOOLS", "GENOMIC FILE MANIPULATION", and "FASTA/FASTQ".

# 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 central text block: "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 text is a "Galaxy 101 Introduction to Galaxy tutorial" banner with a "help." button. 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 a "History" section with a search bar and a list of datasets. The first dataset is highlighted in orange and contains the following information: "1: EBI SRA: PRJNA41980", "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 dataset entry. The bottom of the page shows a "Tweets by @galaxyproject" section.

# 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。

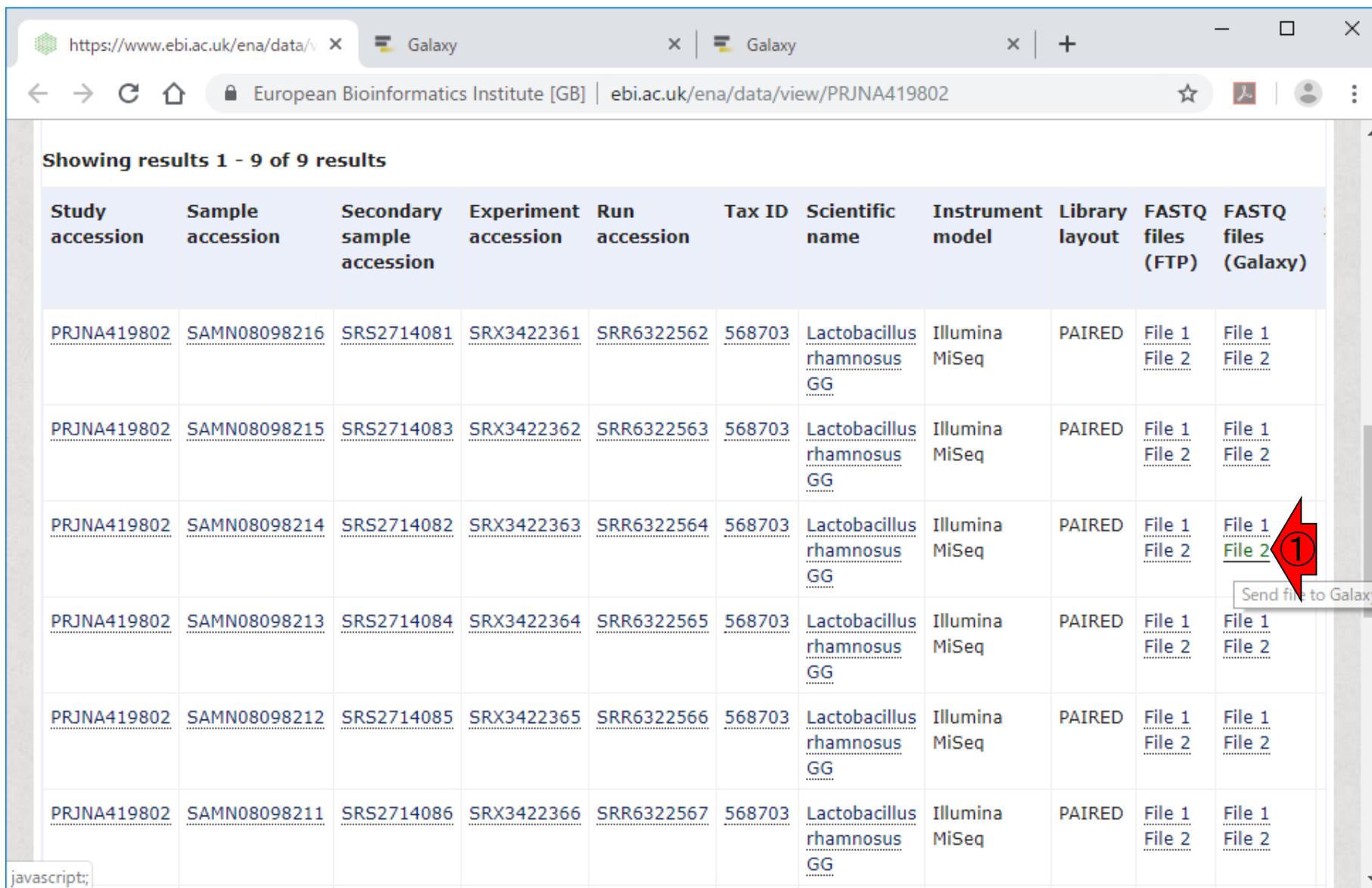
The screenshot shows the Galaxy web interface. The main content area displays a welcome message: "Galaxy is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy start here or consult our help resources. You can install your own Galaxy by following the tutorial 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" with the ISCB logo. At the bottom, there is a "Tweets by @galaxyproject" section.

The left sidebar contains a "Tools" section with a search bar and 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. It lists a dataset: "GSE107337\_3samples" with a size of "304.99 MB". A red arrow labeled "2" points to this size. Below it, a green box highlights the dataset details: "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

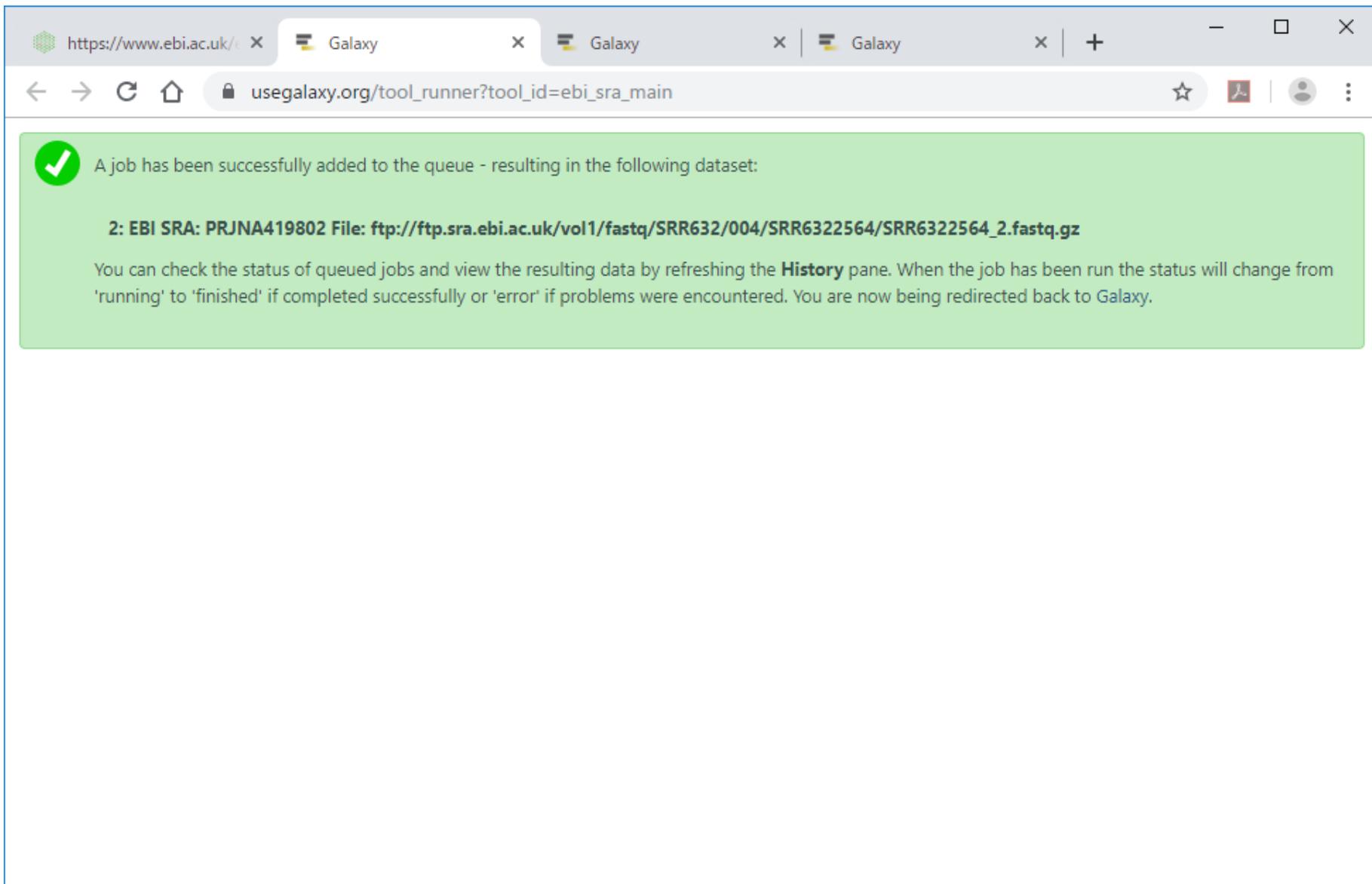


Showing results 1 - 9 of 9 results

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)
<a href="#">PRJNA419802</a>	<a href="#">SAMN08098216</a>	<a href="#">SRS2714081</a>	<a href="#">SRX3422361</a>	<a href="#">SRR6322562</a>	568703	<a href="#">Lactobacillus rhamnosus GG</a>	Illumina MiSeq	PAIRED	<a href="#">File 1</a> <a href="#">File 2</a>	<a href="#">File 1</a> <a href="#">File 2</a>
<a href="#">PRJNA419802</a>	<a href="#">SAMN08098215</a>	<a href="#">SRS2714083</a>	<a href="#">SRX3422362</a>	<a href="#">SRR6322563</a>	568703	<a href="#">Lactobacillus rhamnosus GG</a>	Illumina MiSeq	PAIRED	<a href="#">File 1</a> <a href="#">File 2</a>	<a href="#">File 1</a> <a href="#">File 2</a>
<a href="#">PRJNA419802</a>	<a href="#">SAMN08098214</a>	<a href="#">SRS2714082</a>	<a href="#">SRX3422363</a>	<a href="#">SRR6322564</a>	568703	<a href="#">Lactobacillus rhamnosus GG</a>	Illumina MiSeq	PAIRED	<a href="#">File 1</a> <a href="#">File 2</a>	<a href="#">File 1</a> <a href="#">File 2</a> ①
<a href="#">PRJNA419802</a>	<a href="#">SAMN08098213</a>	<a href="#">SRS2714084</a>	<a href="#">SRX3422364</a>	<a href="#">SRR6322565</a>	568703	<a href="#">Lactobacillus rhamnosus GG</a>	Illumina MiSeq	PAIRED	<a href="#">File 1</a> <a href="#">File 2</a>	<a href="#">File 1</a> <a href="#">File 2</a>
<a href="#">PRJNA419802</a>	<a href="#">SAMN08098212</a>	<a href="#">SRS2714085</a>	<a href="#">SRX3422365</a>	<a href="#">SRR6322566</a>	568703	<a href="#">Lactobacillus rhamnosus GG</a>	Illumina MiSeq	PAIRED	<a href="#">File 1</a> <a href="#">File 2</a>	<a href="#">File 1</a> <a href="#">File 2</a>
<a href="#">PRJNA419802</a>	<a href="#">SAMN08098211</a>	<a href="#">SRS2714086</a>	<a href="#">SRX3422366</a>	<a href="#">SRR6322567</a>	568703	<a href="#">Lactobacillus rhamnosus GG</a>	Illumina MiSeq	PAIRED	<a href="#">File 1</a> <a href="#">File 2</a>	<a href="#">File 1</a> <a href="#">File 2</a>

# W3-2: SRR6322564

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。



The screenshot shows a web browser window with the URL `usegalaxy.org/tool_runner?tool_id=ebi_sra_main`. The browser has multiple 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 browser address bar shows 'usegalaxy.org'. The main content area displays a workflow execution status: 'Using 0%'. Below this, there is a 'History' panel on the right with a search bar and a list of datasets. The first dataset is highlighted in green: '1: EBI SRA: PRJNA419802 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR632/004/SRR6322564/SRR6322564\_1.fastq.gz'. The second dataset is '2: EBI SRA: PRJNA41980 2 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR632/004/SRR6322564/SRR6322564\_2.fastq.gz'. The main workspace shows a 'Galaxy 101' tutorial graphic and a 'Tweets by @galaxyproject' section at the bottom.

# 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 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 categories like 'Get Data', 'Send Data', and 'GENERAL TEXT TOOLS'. The central content area displays a tweet from @galaxyproject. The 'History' panel on the right shows two datasets, both highlighted in green, indicating they are ready for analysis. The datasets 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

# 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. On the right, the 'History' panel lists datasets. Two entries are highlighted with a red box:

- 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

A red arrow with the number 1 points to the edit icon (pencil) of the second dataset. A tooltip '変数を編集する' (Edit variable) is displayed over the edit icon. The left sidebar shows tool categories like 'GENERAL TEXT TOOLS' and 'GENOMIC FILE MANIPULATION'. The main content area displays a 'Galaxy 101' tutorial slide.

# 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 browser address bar shows <https://www.ebi.ac.uk/ena/data/> and the page URL is [usegalaxy.org/datasets/edit](https://usegalaxy.org/datasets/edit). The main content area is titled "Edit dataset attributes" and includes sections for "Attributes", "Permissions", and "Name". The "Name" field contains the text "EBI SRA: PRJNA419802 File: ftp://ftp.sra.ebi.ac.uk/vol1/1". A red arrow labeled "3" points to the end of this text. The "History" panel on the right shows a list of datasets, with the second entry highlighted in green and a red box around its name: "1: EBI SRA: PRJNA419802 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR632/004/SRR6322564/SRR6322564\_1.fastq.gz". A red arrow labeled "2" points to the edit icon of this 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 main content area is titled "Edit dataset attributes" and contains several sections:

- Attributes:** Includes a search bar, "Convert" button, and "Datatypes" dropdown.
- Permissions:** A section for managing dataset permissions.
- Buttons:** "変数を編集する" (Edit variables), "Auto-detect", and "Save" (highlighted with a red arrow and number 6).
- Name:** A text input field containing "SRR6322564\_1" (highlighted with a red arrow and number 5).
- Info:** A text area containing the dataset's original information: "EBI SRA: PRJNA419802 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR632/004/SRR6322564/SRR6322564\_1.fastq.gz" (highlighted with a red arrow and number 4).
- Annotation:** A text area for adding annotations.

The right sidebar shows the "History" section, listing datasets. The top entry is "GSE107337\_3samples" (643.12 MB). Below it, two dataset entries are visible, each with a red arrow and number 2 pointing to the "Info" field:

- 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 tab is labeled 'Galaxy' and the URL is [usegalaxy.org/datasets/edit](https://usegalaxy.org/datasets/edit). The interface includes a navigation bar with '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 'Using 0%' indicator is visible in the top right.

The main content area is divided into three panels:

- Tools:** A search bar for tools 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'.
- Edit dataset attributes:** A panel with a green 'Attributes updated.' message. It contains tabs for 'Attributes', 'Convert', and 'Datatypes', and a 'Permissions' section. Below these are buttons for '変数を編集する', 'Auto-detect', and 'Save'. The 'Name' field contains 'SRR6322564\_1'. The 'Info' section displays 'EBI SRA: PRJNA419802 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR632/004/SRR6322564/SRR6322564\_1.fastq.gz'. The 'Annotation' section is currently empty.
- History:** A panel with a search bar for datasets. It shows a list of datasets under the heading 'GSE107337\_3samples'. The first entry is '2: EBI SRA: PRJNA419802 File' with a size of 643.12 MB. The second entry is '1: SRR6322564\_1'. A red arrow labeled '2' points to the first entry, and a red arrow labeled '1' points to the second entry.

# W3-3: rename

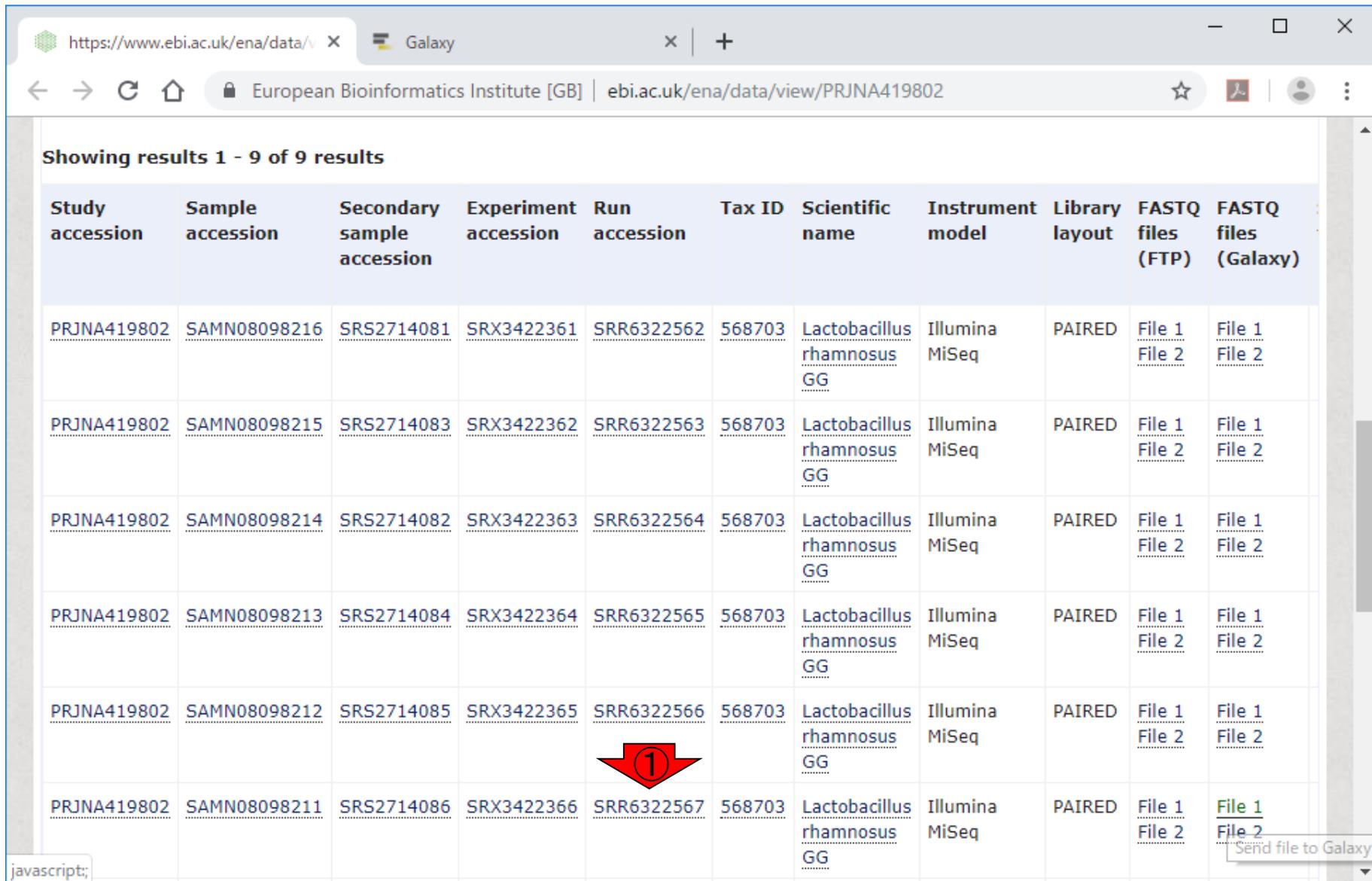
①forward側ファイルのrename完了。②reverse側についても同様に行う。作業完了。

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. The dataset "GSE107337\_3samples" is listed with a size of 643.12 MB. Below it, two datasets are shown in a list: "2: SRR6322564\_2" and "1: SRR6322564\_1". Each entry has icons for view, edit, and delete.

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)
<a href="#">PRJNA419802</a>	<a href="#">SAMN08098216</a>	<a href="#">SRS2714081</a>	<a href="#">SRX3422361</a>	<a href="#">SRR6322562</a>	568703	<a href="#">Lactobacillus rhamnosus GG</a>	Illumina MiSeq	PAIRED	<a href="#">File 1</a> <a href="#">File 2</a>	<a href="#">File 1</a> <a href="#">File 2</a>
<a href="#">PRJNA419802</a>	<a href="#">SAMN08098215</a>	<a href="#">SRS2714083</a>	<a href="#">SRX3422362</a>	<a href="#">SRR6322563</a>	568703	<a href="#">Lactobacillus rhamnosus GG</a>	Illumina MiSeq	PAIRED	<a href="#">File 1</a> <a href="#">File 2</a>	<a href="#">File 1</a> <a href="#">File 2</a>
<a href="#">PRJNA419802</a>	<a href="#">SAMN08098214</a>	<a href="#">SRS2714082</a>	<a href="#">SRX3422363</a>	<a href="#">SRR6322564</a>	568703	<a href="#">Lactobacillus rhamnosus GG</a>	Illumina MiSeq	PAIRED	<a href="#">File 1</a> <a href="#">File 2</a>	<a href="#">File 1</a> <a href="#">File 2</a>
<a href="#">PRJNA419802</a>	<a href="#">SAMN08098213</a>	<a href="#">SRS2714084</a>	<a href="#">SRX3422364</a>	<a href="#">SRR6322565</a>	568703	<a href="#">Lactobacillus rhamnosus GG</a>	Illumina MiSeq	PAIRED	<a href="#">File 1</a> <a href="#">File 2</a>	<a href="#">File 1</a> <a href="#">File 2</a>
<a href="#">PRJNA419802</a>	<a href="#">SAMN08098212</a>	<a href="#">SRS2714085</a>	<a href="#">SRX3422365</a>	<a href="#">SRR6322566</a>	568703	<a href="#">Lactobacillus rhamnosus GG</a>	Illumina MiSeq	PAIRED	<a href="#">File 1</a> <a href="#">File 2</a>	<a href="#">File 1</a> <a href="#">File 2</a>
<a href="#">PRJNA419802</a>	<a href="#">SAMN08098211</a>	<a href="#">SRS2714086</a>	<a href="#">SRX3422366</a>	<a href="#">SRR6322567</a>	568703	<a href="#">Lactobacillus rhamnosus GG</a>	Illumina MiSeq	PAIRED	<a href="#">File 1</a> <a href="#">File 2</a>	<a href="#">File 1</a> <a href="#">File 2</a>

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)
<a href="#">PRJNA419802</a>	<a href="#">SAMN08098216</a>	<a href="#">SRS2714081</a>	<a href="#">SRX3422361</a>	<a href="#">SRR6322562</a>	568703	<a href="#">Lactobacillus rhamnosus GG</a>	Illumina MiSeq	PAIRED	<a href="#">File 1</a> <a href="#">File 2</a>	<a href="#">File 1</a> <a href="#">File 2</a>
<a href="#">PRJNA419802</a>	<a href="#">SAMN08098215</a>	<a href="#">SRS2714083</a>	<a href="#">SRX3422362</a>	<a href="#">SRR6322563</a>	568703	<a href="#">Lactobacillus rhamnosus GG</a>	Illumina MiSeq	PAIRED	<a href="#">File 1</a> <a href="#">File 2</a>	<a href="#">File 1</a> <a href="#">File 2</a>
<a href="#">PRJNA419802</a>	<a href="#">SAMN08098214</a>	<a href="#">SRS2714082</a>	<a href="#">SRX3422363</a>	<a href="#">SRR6322564</a>	568703	<a href="#">Lactobacillus rhamnosus GG</a>	Illumina MiSeq	PAIRED	<a href="#">File 1</a> <a href="#">File 2</a>	<a href="#">File 1</a> <a href="#">File 2</a>
<a href="#">PRJNA419802</a>	<a href="#">SAMN08098213</a>	<a href="#">SRS2714084</a>	<a href="#">SRX3422364</a>	<a href="#">SRR6322565</a>	568703	<a href="#">Lactobacillus rhamnosus GG</a>	Illumina MiSeq	PAIRED	<a href="#">File 1</a> <a href="#">File 2</a>	<a href="#">File 1</a> <a href="#">File 2</a>
<a href="#">PRJNA419802</a>	<a href="#">SAMN08098212</a>	<a href="#">SRS2714085</a>	<a href="#">SRX3422365</a>	<a href="#">SRR6322566</a>	568703	<a href="#">Lactobacillus rhamnosus GG</a>	Illumina MiSeq	PAIRED	<a href="#">File 1</a> <a href="#">File 2</a>	<a href="#">File 1</a> <a href="#">File 2</a>
<a href="#">PRJNA419802</a>	<a href="#">SAMN08098211</a>	<a href="#">SRS2714086</a>	<a href="#">SRX3422366</a>	<a href="#">SRR6322567</a>	568703	<a href="#">Lactobacillus rhamnosus GG</a>	Illumina MiSeq	PAIRED	<a href="#">File 1</a> <a href="#">File 2</a>	<a href="#">File 1</a> <a href="#">File 2</a>

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 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 left sidebar lists various tool categories such as 'Get Data', 'Send Data', 'Collection Operations', 'GENERAL TEXT TOOLS', 'GENOMIC FILE MANIPULATION', and 'FASTA/FASTQ'. The main content area features a banner for '125+ ways to use Galaxy' and a tweet from @galaxyproject. The right-hand 'History' panel displays 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: PRJNA41980 (2 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR632/007/SRR6322567/SRR6322567\_1.fastq.gz)
- 2: SRR6322564\_2
- 1: SRR6322564\_1

## 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 "FASTA/FASTQ".

The right sidebar shows the "History" section with a search bar and a list of datasets. The dataset "GSE107337\_3samples" is expanded, showing 4 datasets. The first two are EBI SRA files:

- 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

The bottom of the page shows a "Tweets by @galaxyproject" section.

## W3-5: rename

The screenshot displays the Galaxy web interface for editing dataset attributes. The browser address bar shows `usegalaxy.org/datasets/edit`. 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', 'GENOMIC FILE MANIPULATION', and 'FASTA/FASTQ'.
- Edit dataset attributes:** The central panel shows a green notification 'Attributes updated.' Below it are tabs for 'Attributes', 'Convert', and 'Datatypes'. A 'Permissions' section is also visible. A button '変数を編集する' (Edit variables) is present, along with 'Auto-detect' and 'Save' buttons. The 'Name' field contains 'SRR6322567\_2'. The 'Info' field contains the file path: 'EBI SRA: PRJNA419802 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR632/007/SRR6322567/SRR6322567\_2.fastq.gz'.
- History:** The right panel shows a search bar and a list of datasets under the heading 'GSE107337\_3samples'. The list includes:
 

Dataset ID	View	Edit	Delete
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 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", "GENOMIC FILE MANIPULATION", and "FASTA/FASTQ".
- Edit dataset attributes:** The central panel shows a green notification "Attributes updated." and tabs for "Attributes", "Convert", "Datatypes", and "Permissions". Below the tabs, there is a button "変数を編集する" (Edit variables) 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".
- History:** A sidebar on the right showing a list of datasets. The current dataset is highlighted in green. The list includes:
  - 6: SRR6322569\_2
  - 5: SRR6322569\_1
  - 4: SRR6322567\_2
  - 3: SRR6322567\_1
  - 2: SRR6322564\_2
  - 1: SRR6322564\_1

# 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), 'ヘルプ' (Help), and 'ユーザー' (User), and a 'Using 0%' indicator. The 'Tools' panel on the left lists various tool categories like 'Get Data', 'Send Data', 'Collection Operations', 'GENERAL TEXT TOOLS', 'GENOMIC FILE MANIPULATION', and 'FASTA/FASTQ'. The central panel is titled 'Edit dataset attributes' and shows a green message 'Attributes updated.' with options for 'Attributes', 'Convert', 'Datatypes', and 'Permissions'. Below this are buttons for '変数を編集する' (Edit variables), 'Auto-detect', and 'Save'. The 'Name' field contains 'SRR6322569\_2' and the 'Info' field contains 'EBI SRA: PRJNA419802 File: ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR632/009/SRR6322569/SRR6322569\_2.fastq.gz'. The right panel shows the 'History' of datasets, listing 'GSE107337\_3samples' with 6 items shown, totaling 2.15 GB. The items are listed in descending order of size: 6: SRR6322569\_2, 5: SRR6322569\_1, 4: SRR6322567\_2, 3: SRR6322567\_1, 2: SRR6322564\_2, and 1: SRR6322564\_1. A red arrow points to the 'データ解析' menu item in the top navigation bar.

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' (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 like 'Get Data', 'Send Data', 'Collection Operations', 'GENERAL TEXT TOOLS', 'GENOMIC FILE MANIPULATION', and 'FASTA/FASTQ'. The central panel displays a tutorial 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 six individual samples (SRR6322569\_2, SRR6322569\_1, SRR6322567\_2, SRR6322567\_1, SRR6322564\_2, SRR6322564\_1).

# 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 with the number 1. The main content area displays a 'Galaxy 101' tutorial slide. The 'History' panel on the right shows a list of datasets, including 'GSE107337\_3samples' and several individual samples like 'SRR6322569\_2'.

# 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 `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 "Using 0%" indicator is visible in the top right.

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", and "Trimmomatic flexible read trimming". A red arrow with the number "2" points to the "FASTQ Quality Control" category.

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 right sidebar shows the "History" panel with a search bar and a list of datasets under the heading "GSE107337\_3samples". The list includes 6 datasets, with the top one highlighted in green:

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"/>

# W4-1: 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 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", "ユーザー", and a "Using 0%" indicator.

**Tools Panel:** A search bar for tools is visible. Under "GENERAL TEXT TOOLS", "FASTQ Quality Control" is selected, showing "FastQC Read Quality reports" and "Trimmomatic flexible read trimming".

**FastQC Tool Interface:** The tool is titled "FastQC Read Quality reports (Galaxy Version 0.72+galaxy1)". It includes options for "Favorite", "Versions", and "Options". Below, there 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). A description for the contaminant list reads: "tab delimited file with 2 columns: name and sequence. For example: Illumina Small RNA RT Primer CAAGCAGAAGACGGCATACGA".

**History Panel:** Titled "History", it shows a search bar for datasets. The current dataset is "GSE107337\_3samples" (2.15 GB). A list of 6 datasets is shown, each with a view, edit, and delete icon:

- 6: SRR6322569\_2
- 5: SRR6322569\_1
- 4: SRR6322567\_2
- 3: SRR6322567\_1
- 2: SRR6322564\_2
- 1: SRR6322564\_1

# 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 a circled '3'. The 'History' panel on the right shows 6 datasets selected, indicated by a red arrow and a circled '2'. A red arrow and a circled '1' points to the 'Load data from your current history' button.

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

FastQC Read Quality reports (Galaxy Version 0.72+galaxy1)

Favorite Versions Options

Load data from your current history

6: SRR6322569\_2

Content type: Multiple datasets

No tabular dataset

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

Adapter list

No tabular dataset

list of adapters adapter sequences which will be explicitly searched against the library. tab delimited file with 2 columns: name and sequence. (--adapters)

Submodule and Limit specifying file

Nothing selected

History

search datasets

GSE107337\_3samples

6 shown

2.15 GB

6: SRR6322569\_2

5: SRR6322569\_1

4: SRR6322567\_2

3: SRR6322567\_1

2: SRR6322564\_2

1: SRR6322564\_1

# 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. The 'Short read data from your current history' section 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. A red arrow with the number 1 points to the 'Multiple datasets' button. The 'History' panel on the right shows the same six datasets listed in a table with icons for viewing, editing, and deleting each entry.

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

# 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. A red arrow with the number 1 points to the first file in the list.

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

# 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. The files are listed in descending order of time. A red arrow labeled '1' points to the first file in the list, and another red arrow labeled '2' points to the 'Ok' button at the bottom right of the dialog.

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 for tools. Under 'GENERAL TEXT TOOLS', 'FASTQ Quality Control' is expanded to show 'FastQC Read Quality reports'.
- FastQC Configuration:** Shows 'QC Read Quality reports (Galaxy Version 0.72+galaxy1)'. It includes buttons for 'Favorite', 'Versions', and 'Options'. Below this is a section for 'Short read data from your current history' 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 note states: 'This is a batch mode input field. Separate jobs will be triggered for each dataset selection.' Below that is a 'Contaminant list' section with a dropdown menu currently set to 'No tabular dataset'.
- History:** Shows a search bar for datasets. The current history is 'GSE107337\_3samples' (2.15 GB). A list of six datasets is shown, each with an eye icon, a pencil icon, and a close icon. A red arrow with the number '1' points to the 'Options' button in the FastQC configuration 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。

# W4-3: FastQC実行

The screenshot shows the Galaxy web interface. The main content area displays the configuration for the FastQC tool. The 'Lower limit on the length of the sequence to be shown in the report' field is empty. The 'length of Kmer to look for' is set to 7. A red arrow labeled '1' points to the 'Execute' button. Another red arrow labeled '2' points to the 'Execute' button. The right sidebar shows the 'History' panel with a list of datasets, including 'SRR6322569\_2', 'SRR6322569\_1', 'SRR6322567\_2', 'SRR6322567\_1', 'SRR6322564\_2', and 'SRR6322564\_1'. The 'Execute' button is highlighted in blue.

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 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 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', and 'FASTQ Quality Control'. The main content area shows the configuration for the tool 'Lower limit on the length of the sequence to be shown in the report'. It includes a search bar for tools, a text input field for the sequence length, a slider for 'length of Kmer to look for' set to 7, and a 'Sending...' button. Below the configuration, there is an 'i Purpose' section explaining that FastQC aims to provide a simple way to do some quality control checks on raw sequence data. On the right side, the 'History' panel shows a list of datasets under the heading 'GSE107337\_3samples', with 6 datasets shown, including SRR6322569\_2, SRR6322569\_1, SRR6322567\_2, SRR6322567\_1, SRR6322564\_2, and SRR6322564\_1. Each dataset entry has icons for viewing, editing, and deleting.

# 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 browser address bar displays `https://www.ebi.ac.uk/ena/data/v` 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". The left sidebar lists various tool categories such as "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", and "Trimmomatic flexible read trimming". The central panel displays the output of a tool, stating "It produces 12 outputs:" followed by a list of jobs: "18: FastQC on data 6: RawData", "17: FastQC on data 6: Webpage", and "16: FastQC on data 5: RawData". Below this list, a message reads: "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 datasets, including "GSE107337\_3samples" (2.15 GB) and several "FastQC on data" jobs. The "FastQC on data 5: RawData" job is highlighted in orange, indicating it is currently running.

# 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。

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

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

...

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.

History

search datasets

GSE107337\_3samples

18 shown

2.15 GB

18: FastQC on data 6: RawData

17: FastQC on data 6: Webpage

16: FastQC on data 5: Raw Data

15: FastQC on data 5: Webpage

14: FastQC on data 4: RawData

# 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 content area displays a green message: "It produces 12 outputs: 18: FastQC on data 6: RawData 17: FastQC on data 6: Webpage 16: FastQC on data 5: RawData". Below this, 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 status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered." The right-hand "History" panel shows a list of jobs, 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". 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 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", and "Trimmomatic flexible read trimming".

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 green message: "Executed **FastQC** and successfully added 6 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 "18: FastQC on data 6: RawData" and "17: FastQC on data 6: Webpage". On the right, the History panel shows a list of datasets, with the top entry being "18: FastQC on data 6: Raw Data". 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 and the text: "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 12 outputs: 18: FastQC on data 6: RawData, 17: FastQC on data 6: Webpage, and 14: FastQC on data 4: RawData.
- History:** A panel on the right showing a list of datasets. Item 17, "FastQC on data 6: Webpage", is highlighted with a red arrow and a circled number 1. Other items include "FastQC on data 6: Raw Data", "FastQC on data 5: Raw Data", "FastQC on data 5: Webpage", and "FastQC on data 4: Raw Data".

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。

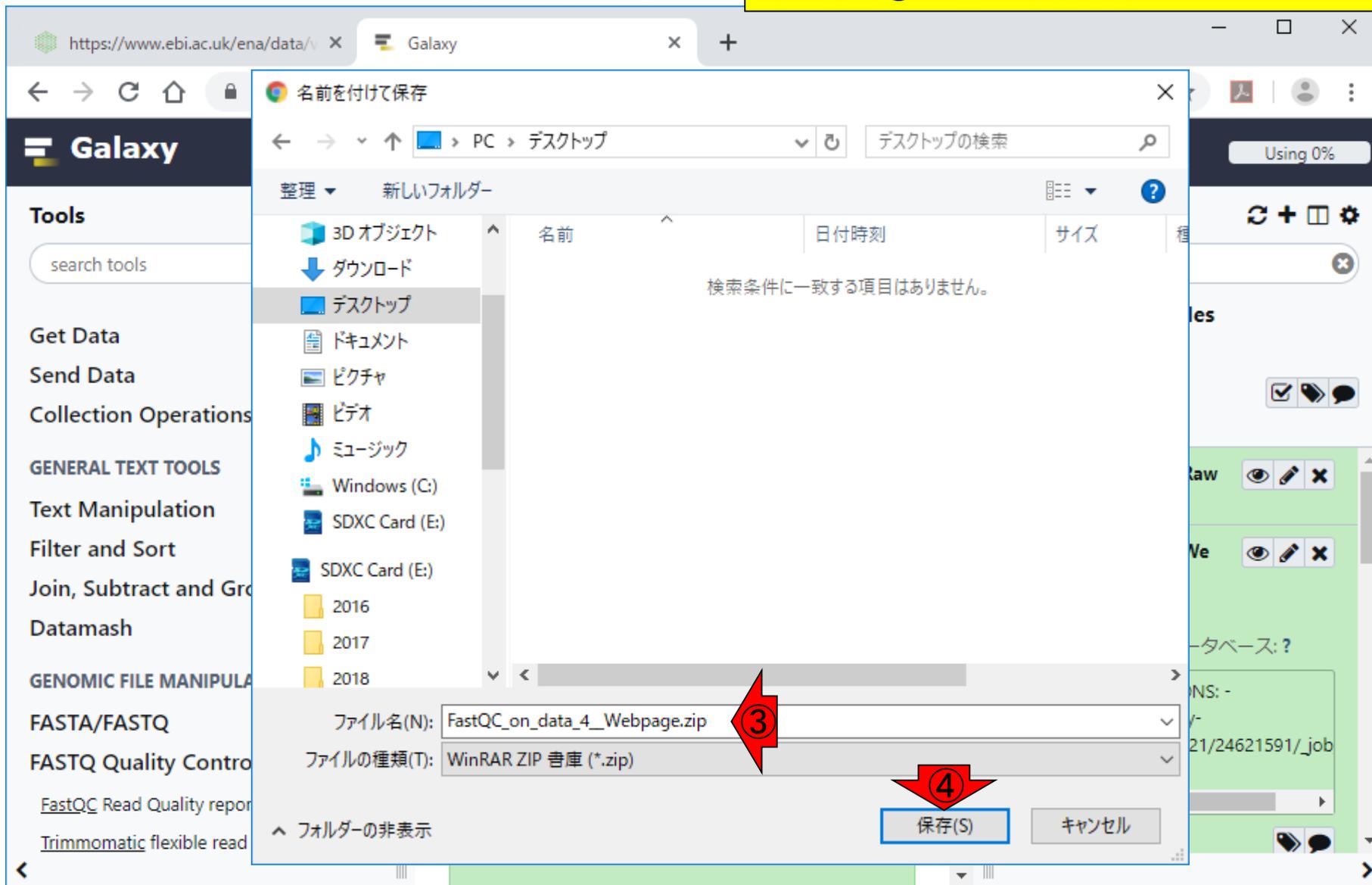
# W4-7: htmlを保存

The screenshot shows the Galaxy web interface. The 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', 'FastQC Read Quality reports', 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 '18: FastQC on data 6: RawData'. It also lists 12 outputs, including '17: FastQC on data 6: Webpage'.
- History:** A right-hand panel showing a search bar for datasets. It lists 'GSE107337\_3samples' (18 shown, 2.16 GB) and '18: FastQC on data 6: Raw Data'. Below it, '17: FastQC on data 6: Webpage' is shown with a size of 621.3 KB and format 'html'. A red arrow with the number '2' points to the save icon (a floppy disk) at the bottom of this 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'.
- Execution Log:** A central green panel showing a successful execution of 'FastQC' with 6 jobs added to the queue. It lists inputs: 6: SRR6322569\_2, 5: SRR6322569\_1, 4: SRR6322567\_2, and outputs: 18: FastQC on data 6: RawData, 17: FastQC on data 6: Webpage.
- History:** A right-hand panel showing a search bar and a list of datasets. The selected dataset is 'GSE107337\_3samples' (18 shown, 2.16 GB). Below it, a job '17: FastQC on data 6: Webpage' is highlighted with a red arrow and a circled '1' pointing to its 'Data' icon.

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



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 Help**  
Got Questions?  
Get Answers.  
[help.galaxyproject.org](http://help.galaxyproject.org)

**History**

- 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

Tweets 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 with a search bar and a list of tool categories. The 'FASTQ Quality Control' tool is highlighted with a red arrow labeled '2'. A 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' and a tweet from @galaxyproject. The right panel shows the 'History' section with a list of recent jobs, including 'FastQC on data 6: Raw Data'.

# 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 <https://www.ebi.ac.uk/ena/data/> and the page URL is [usegalaxy.org](https://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 active, showing a search bar and a list of tool categories. Under 'FASTQ Quality Control', the tool 'Trimmomatic flexible read trimming tool for Illumina NGS data' is highlighted with a red arrow and a circled number '3'. The central workspace contains a 'Galaxy 101' tutorial banner. The 'History' panel on the right shows a list of recent jobs, including 'FastQC on data 6: Raw Data' and 'FastQC on data 6: We bpage'.

# 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" panel on the left lists categories like "GENOMIC FILE MANIPULATION", "FASTA/FASTQ", and "FASTQ Quality Control". The Trimmomatic tool configuration includes a search bar, "Favorite", "Versions", and "Options" buttons. The "Single-end or paired-end reads?" dropdown is set to "Single-end". The "Input FASTQ file" field shows "6: SRR6322569\_2". The "Perform initial ILLUMINACLIP step?" section has "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 "History" panel on the right shows a list of datasets for "GSE107337\_3samples", 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".

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 selected. The tool name 'Trimmomatic' is highlighted with a red arrow labeled '1'. The dropdown menu for 'Single-end or paired-end reads?' is also highlighted with a red arrow labeled '2'. The interface includes a search bar for tools, a list of datasets in the history, and various tool options like 'Favorite', 'Versions', and 'Options'.

# 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 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 flexible read trimming tool for Illumina NGS data (Galaxy Version 0.36.6)'. A red arrow labeled '1' points to the tool title. Below the title, the 'Single-end or paired-end reads?' dropdown menu is open, showing options: 'Single-end', 'Paired-end (two separate input files)' (highlighted in blue), and 'Paired-end (as collection)'. A red arrow labeled '2' points to the dropdown menu, and another red arrow labeled '3' points to the 'Paired-end (two separate input files)' option. Below the dropdown, the 'Trimmomatic Operation' section shows '1: Trimmomatic Operation' and 'Select Trimmomatic operation to perform' with 'Sliding window trimming (SLIDINGWINDOW)' selected. 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.

# 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?**: 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?**: No
- Trimmomatic Operation**: Cut adapter and other illumina-specific sequences from the read

The History panel on the right shows a list of datasets for "GSE107337\_3samples" (2.16 GB). The most recent dataset is "18: FastQC on data 6: Raw Data".

# 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)' and 'Input FASTQ file (R2/second of pair)' fields are both set to '6: SRR6322569\_2'. Red arrows labeled 1 and 2 point to these fields. The History panel on the right shows a list of datasets, with '18: FastQC on data 6: Raw Data' and '17: FastQC on data 6: We bpage' highlighted in green.

# 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 the input dataset 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. The main content area displays the configuration for the **Trimmomatic** tool. The tool description is "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)".

The "Input FASTQ file (R1/first of pair)" section 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. A red arrow with the number "2" points to the "Browse Datasets" button next to this list.

The right sidebar shows the "History" panel with a search bar and a list of jobs. The top job is "18: FastQC on data 6: Raw Data". Below it are "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".

# W5-3: 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。

usegalaxy.org

Galaxy

Tools

search tools

GENOMIC FILE MAN

FASTA/FASTQ

FASTQ Quality Co

FastQC Read Quality

Trimmomatic flexible

tool for Illumina NGS

MultiQC aggregate r

bioinformatics analys

report

FASTQ Summary Stat

Compute quality stat

Draw nucleotides dis

Draw quality score boxplot

SAM/BAM

BED

input field. Separate jobs

will be triggered for each

dataset selection.

Input FASTQ file (R2/second of pair)

6: SRR6322569 2

bpage

14: FastQC on data 4: Raw

Data

Using 0%

Type to Search

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

Cancel Ok

# 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 user should 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 results below. The results are sorted by date, with the most recent at the top. The results are as follows:

ID	Filename	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

Red arrows with the number 2 point to the filenames of results 5, 3, and 1. A red arrow with the number 3 points to the 'Ok' button at the bottom right of the dialog box.

# 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** flexible read trimming tool. The tool description is "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)". The "Input FASTQ file (R1/first of pair)" field contains a list of six SRR IDs: 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 "Options" button in the top right of the tool configuration panel. 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".

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 main panel displays the configuration for the Trimmomatic tool. The 'Input FASTQ file (R2/second of pair)' is set to '6: SRR6322569\_2'. The 'Perform initial ILLUMINACLIP step?' section has 'No' selected. The 'Trimmomatic Operation' section shows '1: Trimmomatic Operation' selected, with 'Sliding window trimming (SLIDINGWINDOW)' chosen for the operation, 4 bases to average across, and an average quality required of 20. A red arrow with the number 1 points to the 'Perform initial ILLUMINACLIP step?' section. The right-hand 'History' panel shows a list of datasets, with '18: FastQC on data 6: Raw Data' highlighted in green. 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 left shows the URL 'https://www.ebi.ac.uk/ena/data/v' and the Galaxy logo.

# W5-4: reverse側

The screenshot shows the Galaxy web interface. The browser address bar is <https://www.ebi.ac.uk/ena/data/> and the page URL is [usegalaxy.org](https://usegalaxy.org). The Galaxy logo and navigation menu are at the top. The left sidebar lists tool categories: GENOMIC FILE MANIPULATION, FASTA/FASTQ, FASTQ Quality Control, and SAM/BAM. The central panel shows the Trimmomatic tool configuration. 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 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 a text input for 'Average quality required' (set to '20'). A '+ Insert Trimmomatic Operation' button is at the bottom. The right sidebar shows the 'History' panel 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. The browser address bar displays `https://www.ebi.ac.uk/ena/data/v` and `usegalaxy.org`. 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 "GENOMIC FILE MANIPULATION", "FASTA/FASTQ", and "FASTQ Quality Control".
- Input FASTQ file (R2/second of pair):** The central panel shows 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 "2" points to the "Browse Datasets" button above the list. Below the list, a note states: "This is a batch mode input field. Separate jobs will be triggered for each dataset selection."
- History:** A sidebar on the right showing 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".

Below the input field, there is a section titled "Perform initial ILLUMINACLIP step?" with "Yes" and "No" buttons. Further down, the "Trimmomatic Operation" section is visible, showing "1: Trimmomatic Operation" and a dropdown menu set to "Sliding window trimming (SLIDINGWINDOW)".

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 open. The modal displays a table of search results. A red arrow with the number 1 points to the entry '14: FastQC on data 4: Raw Data'.

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'.

File ID	File Name	File Type	Created 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 contains a search bar and a list of search results. 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 points to the 'Ok' button, marked with a red circle containing the number 3.

# W5-4: reverse側

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

- Tools Panel:** Search tools, GENOMIC FILE MANIPULATION, FASTA/FASTQ, FASTQ Quality Control (FastQC, Trimmomatic, MultiQC), FASTQ Summary Statistics, Compute quality statistics, Draw nucleotides distribution chart, Draw quality score boxplot, SAM/BAM, BED.
- Input FASTQ file (R2/second of pair):** A list of six datasets: 6: SRR6322569\_2, 5: SRR6322569\_1, 4: SRR6322567\_2, 3: SRR6322567\_1, 2: SRR6322564\_2, 1: SRR6322564\_1.
- History Panel:** Search datasets, GSE107337\_3samples (18 shown, 2.16 GB), and a list of jobs including FastQC on data 6: Raw Data, FastQC on data 6: We bpage, FastQC on data 5: Raw Data, FastQC on data 5: We bpage, and FastQC on data 4: Raw Data.
- Confirmation Dialog:** "Perform initial ILLUMINACLIP step?" with "Yes" and "No" buttons. Below it, text explains: "Cut adapter and other illumina-specific sequences from the read".
- Trimmomatic Operation:** "1: Trimmomatic Operation" with a dropdown menu set to "Sliding window trimming (SLIDINGWINDOW)".

# 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', 'FastQC on data 6: We bpage', 'FastQC on data 5: Raw Data', 'FastQC on data 5: We bpage', and '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。

The screenshot shows the Galaxy web interface. The main panel displays the 'Perform initial ILLUMINACLIP step?' dialog box. A red arrow with the number 1 points to the 'Yes' button. The dialog box contains the following text:

**Perform initial ILLUMINACLIP step?**

Yes  No

Cut adapter and other illumina-specific sequences from the read

**Select standard adapter sequences or provide custom?**

Standard

**Adapter sequences to use**

The background shows the 'Input FASTQ file (R2/second of pair)' section with a list of files:

- 6: SRR6322569\_2
- 5: SRR6322569\_1
- 4: SRR6322567\_2
- 3: SRR6322567\_1
- 2: SRR6322564\_2
- 1: SRR6322564\_1

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

- 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

# W5-5: 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 (selected), No
- Cut adapter and other illumina-specific sequences from the read**
- Select standard adapter sequences or provide custom?**: Standard
- Adapter sequences to use**: TruSeq2 (single-ended, for Illumina GAII)
- Maximum mismatch count which will still allow a full match to be performed**: 2
- How accurate the match between the two 'adapter ligated' reads must be for PE palindrome read alignment**: 30

The right sidebar shows the 'History' section with a search bar and a list of jobs. The top job, '18: FastQC on data 6: Raw Data', is highlighted in green. Below it are jobs 17, 16, 15, and 14, each with eye, edit, and delete icons.

# 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, 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 "History" panel on the right shows 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".

# W5-6: 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. The "History" panel on the right shows a list of datasets, with "18: FastQC on data 6: Raw Data" highlighted in green.

# 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

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

# 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 for configuring the 'Perform initial ILLUMINACLIP step?'. The 'Yes' button is selected. The configuration includes:

- Perform initial ILLUMINACLIP step?**: Yes
- Cut adapter and other illumina-specific sequences from the read**
- Select standard adapter sequences or provide custom?**: Standard
- Adapter sequences to use**: TruSeq3 (paired-ended, for MiSeq and HiSeq) (highlighted with a red arrow and '1')
- Maximum mismatch count which will still allow a full match to be performed**: 2
- How accurate the match between the two 'adapter ligated' reads must be for PE palindrome read alignment**: 30

The right sidebar shows a history of jobs for 'GSE107337\_3samples' (2.16 GB). The jobs listed 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

# 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 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" panel on the left lists various genomic file manipulation tools, including "FASTQ Quality Control" and "FASTQ Summary Statistics by column". The main content area shows the Trimmomatic tool configuration with a text input for "Average quality required" set to "20" and a "+ Insert Trimmomatic Operation" button. Below this are two "Output trimlog file?" and "Output trimmomatic log messages?" sections, each with "Yes" and "No" radio buttons. The "Execute" button is highlighted with a red arrow and the number 2. The "History" panel on the right shows a list of jobs, with the top job "18: FastQC on data 6: Raw Data" highlighted in green and a red arrow and the number 1 pointing to it.

# W5-8: 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 Panel (Left):** Contains a search bar for tools and a list of tool categories: GENOMIC FILE MANIPULATION, FASTA/FASTQ, and FASTQ Quality Control. Under FASTQ Quality Control, there are links for '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'. Below these are categories for SAM/BAM and BED.
- Central Panel:** A green box contains the text 'It produces 12 outputs:'. Below this, a list of jobs is shown:
  - 30: Trimmomatic on SRR6322569\_2 (R2 unpaired)
  - 29: Trimmomatic on SRR6322569\_1 (R1 unpaired)
  - 28: Trimmomatic on SRR6322569\_2 (R2 paired)
  - ...
 Below the list, a paragraph explains: '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):** Contains a search bar for datasets and a list of datasets. The top dataset is 'GSE107337\_3samples' (2.16 GB). Below it, a list of jobs is shown:
  - 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)

# 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。

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**

30 shown

3.47 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 SRR6322567\_2 (R2 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)

# 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 main content area shows a successful execution of the Trimmomatic tool. The summary indicates that 6 inputs were used and 12 outputs were produced. The inputs are listed as follows:

- 6: SRR6322569\_2
- 5: SRR6322569\_1
- 4: SRR6322567\_2
- ...

The outputs are listed as follows:

- 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 a search for datasets, with 30 datasets found. A red box highlights the outputs 27-30. Red arrows and numbers 1, 2, and 3 point to the input and output entries.

# 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 main content area is divided into three panels:

- Tools:** A sidebar on the left with a search bar and categories like 'GENOMIC FILE MANIPULATION', 'FASTA/FASTQ', and 'FASTQ Quality Control'. A red lightning bolt icon with the number '1' points to the 'FastQC Read Quality reports' link.
- Execution Log:** A central green box with a checkmark icon. It states: 'Executed Trimmomatic and successfully added 3 jobs to the queue.' It lists 6 inputs: '6: SRR6322569\_2', '5: SRR6322569\_1', and '4: SRR6322567\_2'. It also lists 12 outputs, including '30: Trimmomatic on SRR6322569\_2 (R2 unpaired)', '29: Trimmomatic on SRR6322569\_1 (R1 unpaired)', and '26: Trimmomatic on SRR6322567\_2 (R2 unpaired)'.
- History:** A sidebar on the right showing a list of datasets. The top dataset is 'GSE107337\_3samples' (3.47 GB). Below it, several Trimmomatic jobs are listed, such as '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 entry has icons for viewing, editing, and deleting.

# W6-1 : 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 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 'FastQC Read Quality reports (Galaxy Version 0.72+galaxy1)' tool configuration. It includes sections 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 under the heading 'GSE107337\_3samples'. The list shows 30 datasets, with 3.47 GB of data. The following table shows the first five entries:

Dataset ID	Description	View	Edit	Delete
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)			
26: Trimmomatic on SRR6 322567_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 panel displays the configuration for the 'FASTQ Quality Control' tool. A red arrow points to the 'Multiple datasets' dropdown menu, which is highlighted. The 'History' panel on the right shows a list of datasets, including '30: Trimmomatic on SRR6 322569\_2 (R2 unpaired)' and others.

**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**

**FASTQ Quality Control**

Favorite Versions Options

QC Read Quality reports (Galaxy Version 0.72+galaxy1)

Show data from your current history

30: Trimmomatic

Multiple datasets

No tabular dataset

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

**Adapter list**

No tabular dataset

list of adapters adapter sequences which will be explicitly searched against the library. tab delimited file with 2 columns: name and sequence. (--adapters)

**Submodule and Limit specifying file**

Nothing selected

**History**

search datasets

**GSE107337\_3samples**

30 shown

3.47 GB

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)

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 browser address bar is <https://www.ebi.ac.uk/ena/data/>. The page title 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", "ユーザー", and "Using 1%".

**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

**Short read data from your current history**

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

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

**Contaminant list**

No tabular dataset

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

**History**

GSE107337\_3samples

30 shown

3.47 GB

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)	👁️ ✎️ ✕
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。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 interface with a search bar, tool panels for GENOMIC FILE MANIPULATION, FASTA/FASTQ, and FASTQ Quality Control, and a dataset viewer showing a tabular dataset with 2 columns: name and sequence. The dataset content 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

Red arrows and numbers indicate the selection process:

- ①: Points to the first four rows (IDs 28, 27, 26, 25).
- ②: Points to the second two rows (IDs 24, 23).
- ③: Points to the last two rows (IDs 20, 19).
- ④: Points to the 'Ok' button.

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. The browser address bar displays `https://www.ebi.ac.uk/ena/data/v` 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'. The 'Tools' sidebar on the left lists categories like 'GENOMIC FILE MANIPULATION', 'FASTA/FASTQ', and 'FASTQ Quality Control', with 'FastQC Read Quality reports' selected. The central workspace shows the 'FastQC Read Quality reports (Galaxy Version 0.72+galaxy1)' tool configuration. A red arrow with the number '1' points to the 'Options' dropdown menu. Below the tool configuration, there is a section for 'Short read data from your current history' with a list of jobs (25 to 19) and a 'Contaminant list' section. The right sidebar shows the 'History' panel with a search bar and a list of jobs, including '30: Trimmomatic on SRR6 322569\_2 (R2 unpaired)' and others.

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 with the following elements:

- 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 selected.
- Configuration Panel (Center):** Shows settings for the selected tool. The 'Lower limit on the length of the sequence to be shown in the report' is set to an empty field. The 'length of Kmer to look for' is set to 7. A blue 'Execute' button with a checkmark is visible.
- History Panel (Right):** Shows a list of datasets. The top entry is '30: Trimmomatic on SRR6 322569\_2 (R2 unpaired)'. A red arrow labeled '1' points to this entry.

Red arrows labeled '1' and '2' indicate the steps described in the text: '1' points to the dataset entry in the history, and '2' points to the 'Execute' button.

# W6-4: 実行中

The screenshot shows the Galaxy web interface with a central job execution status panel. The interface includes a top navigation bar with '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' menus. A 'Tools' sidebar on the left lists various genomic file manipulation tools. The central panel displays a green success message: 'Executed **FastQC** and successfully added 6 jobs to the queue.' It details the tool's inputs: 28: Trimmomatic on SRR6322569\_2 (R2 paired), 27: Trimmomatic on SRR6322569\_1 (R1 paired), and 24: Trimmomatic on SRR6322567\_2 (R2 paired). It also lists the output: 42: FastQC on data 28: RawData. On the right, a 'History' panel shows a list of recent jobs, including '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', and '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 displays <https://www.ebi.ac.uk/ena/data/>. The Galaxy logo and navigation menu are visible at the top. A red arrow with the number '2' points to the 'データ解析' (Data Analysis) menu item. The main content area is divided into three panels:

- Tools:** A sidebar on the left with 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'.
- Execution Panel:** A central green panel showing a successful execution of 'FastQC'. It states: 'Executed **FastQC** and successfully added 6 jobs to the queue.' Below this, it lists the tool's inputs: '28: Trimmomatic on SRR6322569\_2 (R2 paired)', '27: Trimmomatic on SRR6322569\_1 (R1 paired)', and '24: Trimmomatic on SRR6322567\_2 (R2 paired)'. It also lists one output: '42: FastQC on data 28: RawData'.
- History:** A panel on the right showing a list of datasets. The top dataset is 'GSE107337\_3samples' (42 shown, 3.48 GB). Below it, a list of jobs is shown, each with a green background and a red arrow with the number '1' pointing to the 'Save HTML' icon (a document with a checkmark). The jobs listed are: '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'.

# 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/> 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" panel on the left is active, showing a search bar and a list of tools under "GENOMICS ANALYSIS". The "Mapping" category is expanded, and "Bowtie2 - map reads against reference genome" is highlighted with a red arrow and a circled number 3. Other tools listed include "LASTZ", "LASTZ\_D", "Map with BWA-MEM", "Map with BWA", and "STAR-Fusion". The main content area displays a welcome message for Galaxy. The "History" panel on the right shows a list of datasets, including "GSE107337\_3samples" and several "FastQC on data" entries. A "Galaxy Help" banner is overlaid in the center, with the text "Got Questions? Get Answers." and the URL "help.galaxyproject.org".

# 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 displays the Galaxy web interface for configuring the Bowtie2 tool. 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', 'ユーザー', and 'Using 1%'. The left sidebar lists 'Tools' and 'GENOMICS ANALYSIS' categories, with 'Mapping' selected. The central tool configuration area for 'Bbowtie2' shows the following options:

- Favorite, Versions, Options
- wtie2 - map reads against reference genome (Galaxy Version 2.3.4.2)
- Is this single or paired library: Single-end
- FASTA/Q file: 30: Trimmomatic
- Must be of datatype "fastqsanger" or "fasta"
- Write unaligned reads (in fastq format) to separate file(s): Yes
- Write aligned reads (in fastq format) to separate file(s): Yes

The right sidebar shows the 'History' panel with a search bar and a list of 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…。

# W7-2: オプション

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 'Galaxy'. The page title is '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' sidebar on the left lists 'GENOMICS ANALYSIS' with sub-categories: 'Assembly', 'Annotation', and 'Mapping'. Under 'Mapping', several tools are listed, including 'Bowtie2 - map reads against reference genome'. The central tool configuration area for 'wtie2' shows the following options: 'Favorite', 'Versions', 'Options', 'Is this single or paired library' (set to 'Single-end'), 'FASTA/Q file' (set to '30: Trimmomatic'), and two 'Write unaligned reads' and 'Write aligned reads' sections, each with 'Yes' and 'No' buttons. A red arrow points to the 'Single-end' dropdown with a circled '1'. 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.

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

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)
- Library Type:** Paired-end
- FASTA/Q file #1:** 30: Trimmomatic
- FASTA/Q file #2:** 30: Trimmomatic
- Write unaligned reads (in fastq format) to separate file(s):** No

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

# 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] 30: Trimmomatic [File selection]
- FASTA/Q file #2: [File selection] 30: Trimmomatic [File selection]
- 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 arrows labeled 1 and 2 point to the 'FASTA/Q file #1' and 'FASTA/Q file #2' input fields, respectively, indicating the need to specify dataset IDs for paired-end reads.

## W7-3: forward側

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 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 central tool configuration area for 'wtie2' shows the following settings:

- Favorite, Versions, Options buttons
- Tool name: **wtie2** - map reads against reference genome (Galaxy Version 2.3.4.2)
- Library type: Paired-end (dropdown)
- FASTA/Q file #1: Multiple datasets (checked), 30: Trimmomatic
- FASTA/Q file #2: 30: Trimmomatic
- Must be of datatype "fastqsanger" or "fasta"
- Write unaligned reads (in fastq format) to separate file(s): Yes
- Command line: --un/--un-conc (possibly with -gz or -bz2); This triggers

The 'History' sidebar on the right shows a list of datasets, including 'GSE107337\_3samples' (3.48 GB) and several 'FastQC on data' entries (e.g., 42: FastQC on data 28: RawData, 41: FastQC on data 28: Webpage, etc.).

# 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' section. The 'History' panel on the right shows a list of datasets, including 'GSE107337\_3samples' and several 'FastQC on data' entries.

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

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

Browse Datasets

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。

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

# 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。

The screenshot shows the Galaxy web interface with a search for Trimmomatic datasets. The search results are as follows:

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

Red arrows with circled numbers indicate the following steps:

- ③: Points to the date column of the dataset with ID 24.
- ④: Points to the description column of datasets with IDs 27, 23, and 19.
- ⑤: Points to the 'Ok' button at the bottom of the search results.

# 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' dropdown menu is open, showing a list of datasets. A red arrow points to the number '6' in the dropdown list. The 'History' panel on the right 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'.

# 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 'GENOMICS ANALYSIS' with sub-sections for 'Assembly', 'Annotation', and 'Mapping'. The central tool configuration area is for 'Trimmomatic'. It includes a search bar for tools, a dropdown menu for 'FASTA/Q file #2' set to '30: Trimmomatic', and several configuration options with 'Yes/No' buttons. A red arrow points to the 'FASTA/Q file #2' dropdown, which is labeled with a circled '6'. 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 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 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. The browser address bar is <https://www.ebi.ac.uk/ena/data/>. The page title is "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%".

**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

**FASTA/Q file #2**

Must be of datatype "fastqsanger" or "fasta"

**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"

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

Yes No

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

**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

# W7-4: 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:

Item ID	Description	Source	Timestamp
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 with circled numbers indicate the following steps:

- ③: Points to the 'Ok' button at the bottom right of the dialog box.
- ④: Points to the 'R2 paired' status of items 24 and 20.
- ⑤: Points to the 'Ok' button at the bottom right of the dialog box.

# 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' tools such as 'Bowtie2', 'LASTZ', and 'Map with BWA-MEM'. The central workspace shows a workflow step 'FASTA/Q file #2' with a dropdown menu open, listing files from 20 to 26. The 'History' sidebar on the right shows a list of datasets, with '42: FastQC on data 28: RawData' highlighted in green.

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:

- Write aligned reads (in fastq format) to separate file(s)**: Yes (selected), No
- Do you want to set paired-end options?**: No (highlighted with a red arrow and a circled '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, 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…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

# W7-5: オプション

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。

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 points to this dropdown, which is labeled with a circled "1". Below this, the "Select reference genome" dropdown is set to "Baboon (Papio anubis): papHam1". The right panel shows 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 top navigation bar includes 'Galaxy' and 'データ解析' (Data Analysis). The left sidebar lists 'Tools' under 'GENOMICS ANALYSIS', with categories like 'Assembly', 'Annotation', and 'Mapping'. The main panel displays a form for selecting a reference genome. A red arrow points to the dropdown menu labeled 'Use a built-in genome index', which is marked with a circled '1'. Below this, 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'. On the right, a 'search datasets' 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'.

# 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 central panel displays the question: "Will you select a reference genome from your history or use a built-in index?". Below this, there are three options in a dropdown menu: "Use a built-in genome index", "Use a built-in genome index", and "Use a genome from the history and build index". The third option is selected and highlighted in blue. Red arrows labeled 1 and 2 point to the dropdown menu and the selected option, respectively. 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".

# 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 and a circled '2' pointing to it. 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.

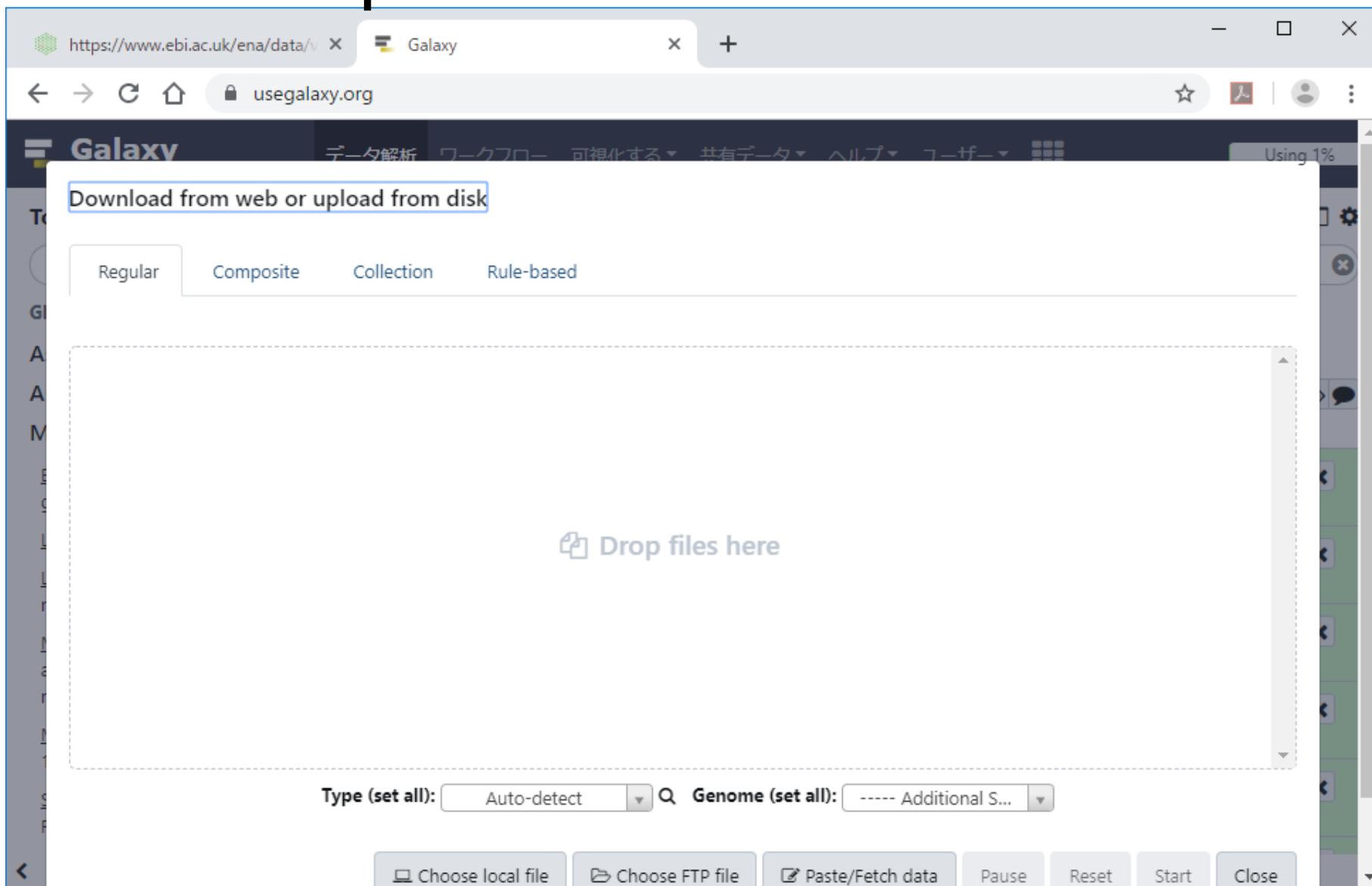
# W7-6: リファレンス

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 'Galaxy' and 'Using 1%'. The left sidebar contains 'Tools' and 'GENOMICS ANALYSIS' categories. The central panel displays configuration options for a tool, including 'Will you select a reference genome from your history or use a built-in index?' (set to 'Use a genome from the history and build index'), 'Select reference genome' (set to 'No fasta dataset'), 'Set read groups information?' (set to 'Do not set'), and 'Select analysis mode' (set to '1: Default setting only'). The right sidebar shows a '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…。

The screenshot shows the Galaxy web interface at usegalaxy.org. A modal window titled "Download from web or upload from disk" is open, displaying the upload progress for two files. The interface includes tabs for "Regular", "Composite", "Collection", and "Rule-based". A progress indicator shows "Please wait...2 out of 2 remaining." The file list table is as follows:

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%

At the bottom of the modal, there are filters for "Type (set all): Auto-detect" and "Genome (set all): ----- Additional S...". Below the modal, the main interface shows buttons for "Choose local file", "Choose FTP file", "Paste/Fetch data", "Pause", "Reset", "Start", and "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 displays 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' panel on the left lists 'GENOMICS ANALYSIS' with sub-sections for 'Assembly', 'Annotation', and 'Mapping'. The 'Mapping' section includes tools like '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', and 'STAR-Fusion detect fusion genes in RNA-Seq data'. The central workspace shows the configuration for 'Bowtie2', with options for 'Will you select a reference genome from your history or use a built-in index?' (set to 'Use a genome from the history and build index'), 'Select reference genome' (set to '43: ASM2650v1.fa'), 'Set read groups information?' (set to 'Do not set'), 'Select analysis mode' (set to '1: Default setting only'), and 'Do you want to use presets?'. The right-hand 'History' panel shows a list of datasets under 'GSE107337\_3samples', including '44: ASM2650v1.gff3', '43: ASM2650v1.fa', '42: FastQC on data 28: RawData', '41: FastQC on data 28: Webpage', '40: FastQC on data 27: RawData', and '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 with the number '2' points to this dropdown. On the right side, the 'History' panel shows a list of datasets. The top two datasets, '44: ASM2650v1.gff3' and '43: ASM2650v1.fa', are highlighted in green and circled with a red arrow and the number '1'. The 'Tools' panel on the left shows 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, indicating the step to be taken.

**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

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)

**History** ↻ + □ ⚙

search datasets x

**GSE107337\_3samples**

47 shown

3.49 GB

☑ 🗑️ 💬

🕒 47: Bowtie2 on data 4 3, data 28, and data 27: al igned reads (BAM) 👁️ ✎ ✕

🕒 46: Bowtie2 on data 4 3, data 24, and data 23: al igned reads (BAM) 👁️ ✎ ✕

🕒 45: Bowtie2 on data 4 3, data 20, and data 19: al igned 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 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)

**History** ↻ + □ ⚙

search datasets x

**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

**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

**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

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)

# 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 'Tools' panel on the left lists various categories like 'Get Data', 'Send Data', and 'GENERAL TEXT TOOLS'. The central area displays a message about Galaxy and a tutorial titled 'Running Your Own Understanding how Galaxy works'. The 'History' panel on the right shows a list of datasets, including 'GSE107337\_3samples' and several Bowtie2 alignment jobs. A red arrow with the number 1 points to the refresh icon in the 'History' panel.

# 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 (marked with a red lightning bolt and '3') and a list of categories. Under 'GENOMICS ANALYSIS', 'Mapping' is marked with a red lightning bolt and '1', and 'RNA-seq' is marked with a red lightning bolt and '2'. The central panel displays a 'Galaxy 101' tutorial card with a workflow diagram. 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 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 'GENOMICS ANALYSIS', with sub-categories like 'Assembly', 'Annotation', 'Mapping', 'Variant Calling', 'ChIP-seq', and 'RNA-seq'. The 'RNA-seq' category is highlighted with a red arrow and the number 2. Below it, the 'featureCounts' tool is listed with a red arrow and the number 3. The main content area displays a welcome message and a 'Try Galaxy on the Cloud' banner. The 'History' panel on the right shows a list of datasets, with the top entry being '47: Bowtie2 on data 43, data 28, and data 27: aligned reads (BAM)'. A red arrow and the number 4 point to the 'RNA-seq' category in the Tools panel.

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. A red arrow points to the 'htseq-count' tool, which is circled with a red '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, 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.

**Running Your Own Understanding how Galaxy works**  
An in-depth tutorial

**Tools**

search tools

on RNA-Seq transcripts

[Salmon](#) 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

**①** [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**

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

# 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:**

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

**History Panel:** Shows a list of datasets. The top entry is "GSE107337\_3samples" (47 shown, 4.52 GB). Below it, several BAM files are listed, such as "47: Bowtie2 on data 43, data 28, and data 27: aligned reads (BAM)".

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'. The 'Align to BAM File' section has a red arrow pointing to the 'Multiple datasets' button. Below it, two dataset selection boxes are visible: '47: Bowtie2 on' and '44:'. The 'Mode' is set to 'Union' and 'Stranded' is set to 'Yes'. 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
- 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 entry. Below this list, there is a note: 'This is a batch mode input field. Separate jobs will be triggered for each dataset selection.' The 'GFF File' section has a dropdown menu with '44:' selected. The 'Mode' section has 'Union' selected. The right-hand 'History' panel shows a search bar and 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; and 43: ASM2650v1.fa.

# W8-2: 複数個指定

The screenshot shows the Galaxy web interface with a search modal window open. The modal window has a search bar at the top and a table of results below. 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

At the bottom of the modal window, there are 'Cancel' and 'Ok' buttons. The background shows the Galaxy interface with a search bar and a list of tools.

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 results. The table has three columns: Label, Details, and Time. 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 shows the Galaxy web interface for the **htseq-count** tool. The tool description is: "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 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 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:** 47: Bowtie2 on data 43, 46: Bowtie2 on data 43, 45: Bowtie2 on data 43.

**GFF File:** 44: (indicated by a red arrow labeled ②)

**Mode:** Union

**History:** 47 shown, 4.52 GB. The history 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 (indicated by a red arrow labeled ①)
- 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。

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 files: 47: Bowtie2 on data 43, 46: Bowtie2 on data 43, and 45: Bowtie2 on data 43. The **GFF File** section shows a dropdown menu with the selected file **44: ASM2650v1.gff3**. The **Mode** is set to **Union**. The **History** panel on the right shows a list of datasets, including **44: ASM2650v1.gff3** (highlighted with a red arrow ①) and **43: ASM2650v1.fa**. The **Tools** panel on the left lists various tools for RNA-Seq analysis.

# 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 'Aligned SAM/BAM File' with a list of datasets: 47: Bowtie2 on data 43, 46: Bowtie2 on data 43, and 45: Bowtie2 on data 43. Below this, the 'GFF File' is set to '44:' and the 'Mode' is 'Union'. The 'History' panel on the right shows a search for 'GSE107337\_3samples' with 47 datasets shown, including the selected ones: 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 list of datasets, including 'GSE107337\_3samples' and several 'Bowtie2 on data' entries. A red arrow and the number '1' point to the 'History' panel.

# 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 with the 'Feature type' tool configuration. The 'Feature type' field is set to 'gene' and is highlighted with a red arrow and the number 3. The 'ID Attribute' field is set to 'gene\_id'. The 'History' panel on the right shows a list of datasets including Bowtie2 alignments and GFF files.

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

minimum value. (--minaaqual)

**Feature type**

gene

Feature type (3rd column in GFF file) to be used. All features of other types are ignored. The default, suitable for RNA-Seq and Ensembl GTF files, is exon. (--type)

**ID Attribute**

gene\_id

GFF attribute to be used as feature ID. Several GFF lines with the same feature ID will be considered as parts of the same feature. The feature ID is used to identify the counts in the output table. All features of the specified type MUST have a value for this attribute. The default, suitable for RNA-Seq and Ensembl GTF files, is gene\_id. (--idattr)

**Set advanced options**

Default settings

Execute

**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

# 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 at the bottom of the configuration panel.

**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

minimum value. (--minaaqual)

**Feature type**

gene

Feature type (3rd column in GFF file) to be used. All features of other types are ignored. The default, suitable for RNA-Seq and Ensembl GTF files, is exon. (--type)

**ID Attribute**

gene\_id

GFF attribute to be used as feature ID. Several GFF lines with the same feature ID will be considered as parts of the same feature. The feature ID is used to identify the counts in the output table. All features of the specified type MUST have a value for this attribute. The default, suitable for RNA-Seq and Ensembl GTF files, is gene\_id. (--idattr)

**Set advanced options**

Default settings

Execute ①

**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

# 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

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)

**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

# W8-6: 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, 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 1%" indicator.
- Tools Panel (Left):** Search tools input, list of tools including `htseq-count`.
- Job Execution Panel (Center):** Green box with a checkmark: "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)
- History Panel (Right):** Search datasets input, list of datasets for "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 (highlighted in green)

# W8-7: 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

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 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

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)

# W8-8: 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 central panel displays the execution details of the **htseq-count** tool. It indicates that 3 jobs were added 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)

The right panel shows the history of datasets for **GSE107337\_3samples**. Three items are highlighted with red arrows and a circled '1':

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

# 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。

# W9-1 : Column join

The screenshot shows the Galaxy web interface. The top navigation bar includes a 'Refresh' icon (a circular arrow) which is highlighted with a red arrow and the number 1. Below the navigation bar, the 'Tools' panel on the left lists various categories like 'Get Data', 'Send Data', and 'Collection Operations'. The main content area features a tweet from @galaxyproject with the text '125+ ways to use Galaxy' and a grid of tool icons. The right-hand 'History' panel displays a list of datasets, with the top entry being '53: htseq-count on data 4 and data 47 (no feature)'. The browser address bar shows 'usegalaxy.org'.

# 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 navigation links like 'データ解析', 'ワークフロー', and '可視化する' are in the top center. A red arrow labeled '①' points to the 'usegalaxy' link in the address bar. On the left sidebar, the 'Collection Operations' menu item is highlighted with a red arrow labeled '②'. The main content area features a welcome message and a carousel titled '125+ ways to use Galaxy'. On the right, the 'History' panel shows a list of datasets, including 'GSE107337\_3samples' with 53 items shown and a total size of 4.52 GB. The history items are listed as 'htseq-count on data 4 and data 47 (no feature)'. At the bottom, there is a 'Tweets by @galaxyproject' section.

# 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 left sidebar contains a 'Tools' section with a search bar and a list of 'Collection Operations' 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'. A red arrow labeled '3' points to 'Column Join on Collections'. The main content area displays a welcome message and a carousel titled '125+ ways to use Galaxy' with the text 'Right now.' The right sidebar shows a 'History' panel with a search bar and a list of datasets, including 'GSE107337\_3samples' and several 'htseq-count' datasets. The bottom of the page shows a 'Tweets by @galaxyproject' section.

# 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 a 'Using 1%' indicator.

The 'Tools' panel on the left lists various operations under 'Collection Operations', 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'.

The main tool panel shows the configuration for 'Column Join on Collections (Galaxy Version 0.0.3)'. It includes a 'Tabular files' list with the following entries:

- 53: htseq-count on data 44 ar
- 52: htseq-count on data 44 ar
- 51: htseq-count on data 44 ar
- 50: htseq-count on data 44 ar
- 49: htseq-count on data 44 ar
- 48: htseq-count on data 44 ar
- 44: ASM2650v1.aff3

The configuration fields are:

- Identifier column:** 1
- Number of Header lines in each item:** 0
- Keep original column header:** Yes

The 'History' panel on the right shows a list of datasets under 'GSE107337\_3samples' (53 shown, 4.52 GB). The visible datasets are:

- 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 4 and data 46
- 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. The first few items are highlighted in green.

# 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 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 following table represents the data in the search dialog:

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 two buttons: "Cancel" and "Ok". 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 tool is set to join 'Tabular files' with 'Identifier column' 1 and 'Number of Header lines in each item' 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.

**Tools**

search 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](#)

**Column Join on Collections (Galaxy Version 0.0.3)**

Favorite Versions Options

**Tabular files**

- 53: htseq-count on data 44 ar
- 52: htseq-count on data 44 ar
- 51: htseq-count on data 44 ar
- 50: htseq-count on data 44 ar
- 49: htseq-count on data 44 ar
- 48: htseq-count on data 44 ar
- 44: ASM2650v1.aff3

**Identifier column**

1

**Number of Header lines in each item**

0

**Keep original column header**

Yes No

Disable if you want columns headers to be only composed of the input dataset names

**History**

search datasets

**GSE107337\_3samples**

53 shown

4.52 GB

- 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 4 and data 46
- 49: htseq-count on data 4

# 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 checkbox for 'Select/Unselect all' which is unchecked. 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'. Below the 'Execute' button, there is an example of the tool's output:

```

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

```

# W9-4: 実行中

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 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 contains the following text:

Executed **Column Join** and successfully added 1 job to the queue.

The tool uses 3 inputs:

  - 52: htseq-count on data 44 and data 47
  - 50: htseq-count on data 44 and data 46
  - 48: htseq-count on data 44 and data 45

It produces this output:

  - 54: Column Join on data 52, data 50, and data 48

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 showing a list of datasets. The top entry is **GSE107337\_3samples** (54 shown, 4.52 GB). Below it, a list of jobs is shown, each with an eye icon, a pencil icon, and a close icon (X). The jobs listed are:
  - 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
  - 51: htseq-count on data 44 and data 46 (no feature)

# W9-5: 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, indicating a successful execution of the **Column Join** tool. 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 also lists the output: "54: Column Join on data 52, data 50, and data 48". A note at the bottom of the notification 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. The top dataset is **GSE107337\_3samples** (4.52 GB). Below it, several jobs are listed, including "54: Column Join on data 52, data 50, and data 48", "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)", and "50: htseq-count on data 4". Each job entry has icons for viewing, editing, and deleting.

The **Tools** panel on the left shows a search bar and a list of tools under "Collection Operations", 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](#).

# W9-6: 解説

The screenshot shows the Galaxy web interface. The main content area 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 right-hand side of the interface features a "History" panel with a search bar and a list of datasets. The top dataset is "GSE107337\_3samples" with 54 items shown and a size of 4.52 GB. Below it, a list of jobs is displayed. The job "54: Column Join on data 52, data 50, and data 48" is highlighted in green and has a red arrow with the number "1" pointing to its icon. Other jobs in the list include "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)", and "50: htseq-count on data 4".

# W9-6: 解説

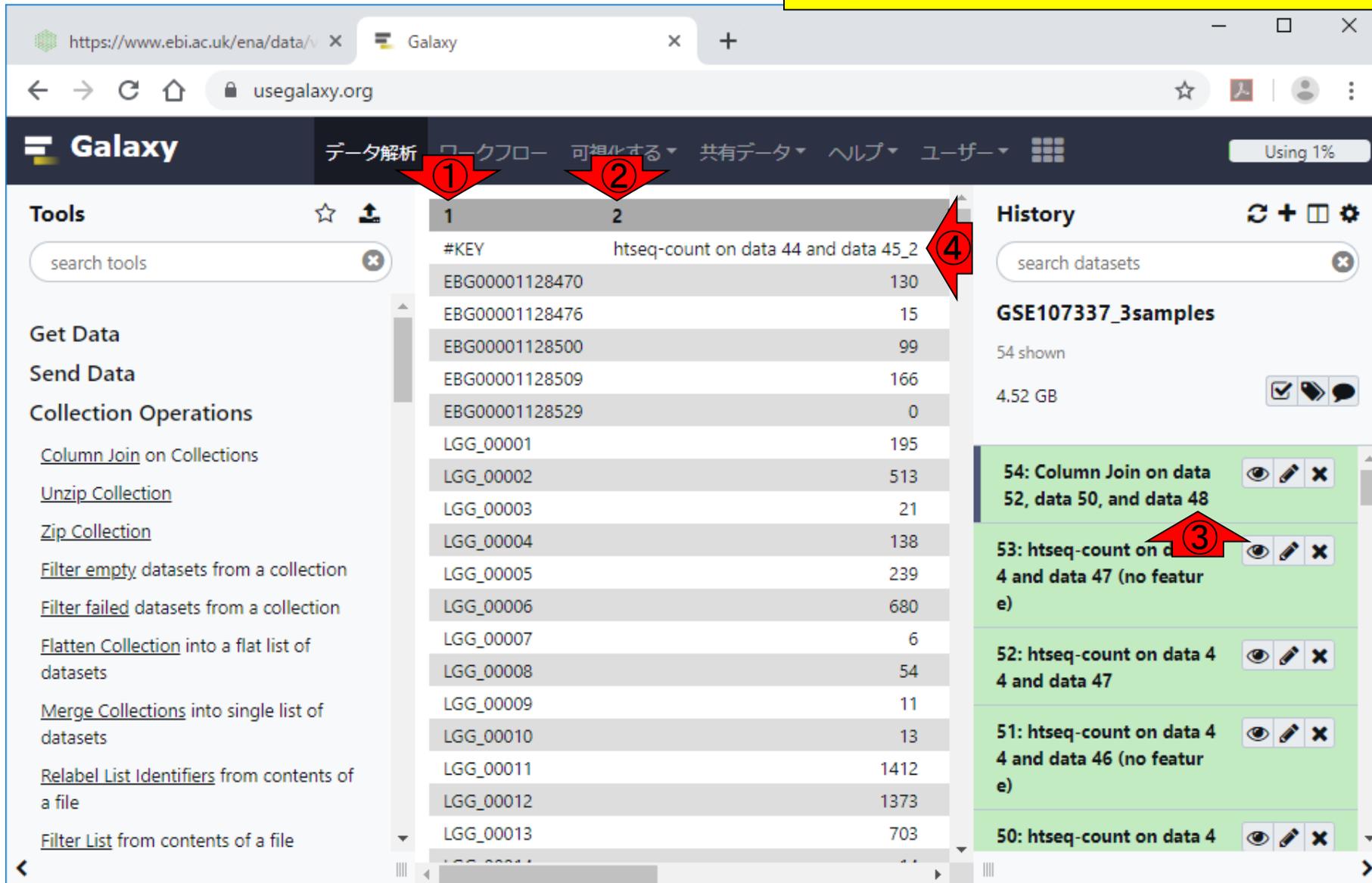
出力結果は、①1つ(ヒストリー54)。②データを中央パネル上で表示させて、大まかに眺める。

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 table displays gene IDs in column 1 and htseq-counts in column 2. The history panel on the right shows a list of operations, with the entry for '54: Column Join on data 52, data 50, and data 48' highlighted in green. Red arrows point to the following elements:

- ①: Column header '1' in the table.
- ②: Column header '2' in the table.
- ③: History entry '53: htseq-count on data 4 and data 47 (no feature)'.
- ④: Cell containing 'htseq-count on data 44 and data 45\_2' in the table.

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

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'. A red arrow labeled '5' points to the top of the table. The table lists various gene IDs and their corresponding counts. The right sidebar shows the 'History' panel with a search box and a list of datasets. The dataset '54: Column Join on data 52, data 50, and data 48' is highlighted in green. Other datasets in the history include '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)', and '50: htseq-count on data 4'.

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 operations. The 50th entry in the history is highlighted with a red arrow labeled '3'. The interface includes a search bar, navigation buttons, and a top navigation bar with 'Galaxy' and 'データ解析'.

	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の情報の。

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 operations. The 52nd entry is highlighted with a red arrow labeled ③:

- 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 circled '1')
- 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)

# 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 ID	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

History: GSE107337\_3samples, 54 shown, 4.52 GB

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: ?

```
#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 panel displays a table with 4 columns. The first column is labeled 'data 44 and data 46\_2' and the second is 'htseq-count on data 44 and data 47\_2'. The table contains 15 rows of data. A 'Tools' sidebar is on the left, and a 'History' sidebar is on the right. A preview window for a dataset is open, showing a table with 2 columns: '#KEY' and 'htseq-count on data 44 and'. The preview window has a 'ダウンロード' (Download) button highlighted with a red circle and the number 3.

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: GSE107337\_3samples, 54 shown, 4.52 GB

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 interface for *Lactobacillus rhamnosus* GG. The main navigation bar includes links for HMMER, BLAST, Tools, Downloads, and More. A search bar is present with the text "Search Ensembl Bacteria...". The page title is "Lactobacillus rhamnosus GG". Below the title, there is a search box for "Lactobacillus rhamnosus GG..." and a "Go" button. The page is divided into several sections:

- About *Lactobacillus rhamnosus* GG**: Contains an "Information and statistics" link, which is highlighted with a red arrow labeled "2".
- Genome assembly: ASM2650v1**: Contains "More information and statistics" (highlighted with a red arrow labeled "1"), "Download DNA sequence (FASTA)", and "Display your data in Ensembl Bacteria".
- Gene annotation**: Includes "What can I find?" (Protein-coding and non-coding genes, splice variants, cDNA and protein sequences, non-coding RNAs), "More about this genebuild", "Download genes, cDNAs, ncRNA, proteins - FASTA - GFF3", and "Update your old Ensembl IDs".
- Comparative genomics**: Includes "What can I find?" (Gene families based on HAMAP and PANTHER classification), "More about comparative analyses", and "Phylogenetic overview of gene families".
- Variation**: States "This species currently has no variation database. However you can process your own variants using the Variant Effect Predictor." and includes a "Variant Effect Predictor" link.

[http://bacteria.ensembl.org/Lactobacillus\\_rhamnosus\\_gg/Info/Index](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 <a href="#">GCA_000026505.1</a> , 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	<a href="#">European Nucleotide Archive</a>

#### 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 biological features.

	A	B	C	D	E	F	G	H	I	J	K	L	M
2519	FM179322	ena	mRNA	505972	50								
2520	FM179322	ena	exon	505972	50								
2521	FM179322	ena	CDS	505972	50								
2522	###												
2523	FM179322	ena	gene	506286	50								
2524	FM179322	ena	mRNA	506286	50								
2525	FM179322	ena	exon	506286	50								
2526	FM179322	ena	CDS	506286	50								
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				

# W10-1: 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。

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

I2528 : X ✓ fx ID=gene:EBG00001128470;Name=rli28;biotype=sRNA;gene\_id=EBG00001128470;logic\_name=rfam\_g ^

	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				

ASM2650v1 (+)

準備完了 ScrollLock 100%

# 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。

自動保存 ASM2650v1.gff3 - Excel 門田 幸二 検索 共有

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

I2528 : ID=gene:EBG00001128470;Name=rli28;biotype=sRNA;gene\_id=EBG00001128470;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_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=				
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_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;logic_name=rfam_genes				
2529	FM179322	Rfam	transcript	506642	506842	.	-	.	ID=transcript:EBT00001719480;Parent=gene:EBG00001128470;Name=EBT00001719480;logic_name=rfam_genes				
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;Name=EBT00001719564;logic_name=rfam_genes				
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=				
2537	FM179322	ena	mRNA	507589	509007	.	-	.	ID=transcript:CAR86394;Parent=gene:LGG_00499;Name=CAR86394;logic_name=rfam_genes				
2538	FM179322	ena	exon	507589	509007	.	-	.	Parent=transcript:CAR86394;Name=CAR86394-1;constitutive_exon=				

ASM2650v1 準備完了 ScrollLock 100%



# 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;const				
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;biotype				
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

自動保存 ASM2650v1.gff3 - 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

I2528 ID=gene:EBG00001128470;Name=rli28;biotype=sRNA;gene\_id=EBG00001128470;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=transcript;Name=CAR86392-1;const				
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;biotype				
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				

ASM2650v1 準備完了 ScrollLock 100%



①

# 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\_g enes

	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				

ASM2650v1

準備完了 ScrollLock - + 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

# 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	<a href="#">Lactobacillus rhamnosus GG</a>
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)	<a href="#">Choi I, Oh S</a>
Citation missing	<i>Has this study been published? Please <a href="#">login</a> to update or <a href="#">notify GEO</a>.</i>
Submission date	Nov 25, 2017
Last update date	May 15, 2019
Contact name	kucsbl submitter
E-mail(s)	<a href="mailto:kucsbl.group@gmail.com">kucsbl.group@gmail.com</a>
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
<a href="#">SOFT formatted family file(s)</a>	SOFT <a href="#">?</a>
<a href="#">MINiML formatted family file(s)</a>	MINiML <a href="#">?</a>
<a href="#">Series Matrix File(s)</a>	TXT <a href="#">?</a>

Supplementary file	Size	Download	File type/resource
<a href="#">GSE107337_RPKM.csv.gz</a>	21.1 Kb	<a href="#">(ftp)</a> <a href="#">(http)</a>	CSV
<a href="#">GSE107337_RawCounts.csv.gz</a>	20.8 Kb	<a href="#">(ftp)</a> <a href="#">(http)</a>	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](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 $\log_2$ をとるとのこと。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  編集 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

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

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

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	-0.2895	-0.5184
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	0.2556	1.5254
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



# W12-2: sum

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				
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 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 検索 共有

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 <sub>2</sub> (pH4_1h/CCG)	log <sub>2</sub> (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
	gene_ID	pH4.5_1h	pH4.5_24h	pH7_CCG	pH4_1h/CCG	pH4_24h/CCG	$\log_2(\text{pH4\_1h/CCG})$	$\log_2(\text{pH4\_24h/CCG})$
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

## 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". The spreadsheet displays a table with columns A through H, representing various gene expression metrics. The row for LGG\_02240 is highlighted in red, and a red arrow points to the cell A2246.

	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 <sub>2</sub> (pH4_1h/CCG)	log <sub>2</sub> (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

## 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 cell containing the gene ID "LGG\_02372".

	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 <sub>2</sub> (pH4_1h/CCG)	log <sub>2</sub> (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	$\log_2(\text{pH4\_1h/CCG})$	$\log_2(\text{pH4\_24h/CCG})$
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。

Excel spreadsheet showing gene expression data for LGG\_00113. The spreadsheet has columns for gene ID, pH4.5\_1h, pH4.5\_24h, pH7\_CCG, pH4\_1h/CCG, pH4\_24h/CCG, log<sub>2</sub>(pH4\_1h/CCG), and log<sub>2</sub>(pH4\_24h/CCG). The row for LGG\_00113 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 log<sub>2</sub>(pH4\_24h/CCG) value.

	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 <sub>2</sub> (pH4_1h/CCG)	log <sub>2</sub> (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

# W13-2: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 関数 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

GSE107337\_RawCounts.xlsx 門田 幸二

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 <sub>2</sub> (pH4_1h/CCG)	log <sub>2</sub> (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。

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

A121 : X ✓ fx LGG\_00115

	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 <sub>2</sub> (pH4_1h/CCG)	log <sub>2</sub> (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%

# W13-3: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	$\log_2(\text{pH4\_1h/CCG})$	$\log_2(\text{pH4\_24h/CCG})$
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 spreadsheet has columns for gene ID, pH4.5\_1h, pH4.5\_24h, pH7\_CCG, pH4\_1h/CCG, pH4\_24h/CCG, log<sub>2</sub>(pH4\_1h/CCG), and log<sub>2</sub>(pH4\_24h/CCG). The row for LGG\_00108 is highlighted with a red border. 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
	gene_ID	pH4.5_1h	pH4.5_24h	pH7_CCG	pH4_1h/CCG	pH4_24h/CCG	log <sub>2</sub> (pH4_1h/CCG)	log <sub>2</sub> (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
①	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 <sub>2</sub> (pH4_1h/CCG)	log <sub>2</sub> (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。

	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 <sub>2</sub> (pH4_1h/CCG)	log <sub>2</sub> (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。

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

A1016 : X ✓ fx LGG\_01064

	A	B	C	D	E	F	G	H
1	Gene	pH4_1h	pH4_24h	pH7_CCG	pH4_1h/CCG	pH4_24h/CCG	log <sub>2</sub> (pH4_1h/CCG)	log <sub>2</sub> (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

GSE107337\_RawCounts 表示設定 100%

# 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 contains a table with columns A through H. The row for LGG\_01064 is highlighted in red. A red arrow points to the gene ID "LGG\_01064" in cell A1070, and another red arrow points to the "2" in 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 <sub>2</sub> (pH4_1h/CCG)	log <sub>2</sub> (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 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 <sub>2</sub> (pH4_1h/CCG)	log <sub>2</sub> (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。

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

A2038 LGG\_02032

	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 <sub>2</sub> (pH4_1h/CCG)	log <sub>2</sub> (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

data\_9samples summed averaged 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。

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

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 <sub>2</sub> (pH4_1h/CCG)	log <sub>2</sub> (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 spreadsheet's 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 <sub>2</sub> (pH4_1h/CCG)	log <sub>2</sub> (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	<u>0.9544823</u>	0.8670647	0.8617587
pH4_24h	0.9023052	1.0000000	0.8183148	0.8583811	<u>0.9544563</u>	0.7889634
pH7_CCG	0.8919116	0.8183148	1.0000000	0.8533259	0.7849992	<u>0.9662319</u>
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
> |
```

