

統合データベース講習会：AJACS筑波2
2012年8月6日、7日

DDBJパイプラインと DBCLS p-Galaxyの紹介

情報・システム研究機構 (ROIS)
ライフサイエンス統合データベースセンター (DBCLS)
河野 信

1

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

新世代シーケンサ



Roche GS FLX



Illumina HiSeq2500



Life Technology SOLiD4

上記以外にも

Helicos
IonTorrent
PacBio
Oxford Nanopore
...

大量並列化による大規模データの産出

データの大容量化
解析には高機能コンピュータが必要

3

©2012 統合データベース講習会 Licensed Under CC 表示 2.1. © 2012 DBCLS licensed under CC表示2.1日本

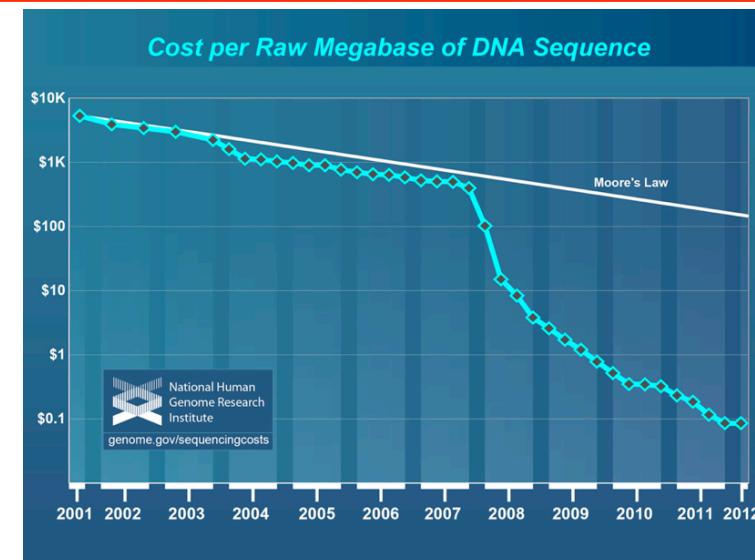
本日のアウトライン

- ◆ SRA/DRA
- ◆ D-way (MetaDefine)
- ◆ DDBJ Read Annotation Pipeline
- ◆ DBCLS p-galaxy
- ◆ MiGAP

2

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

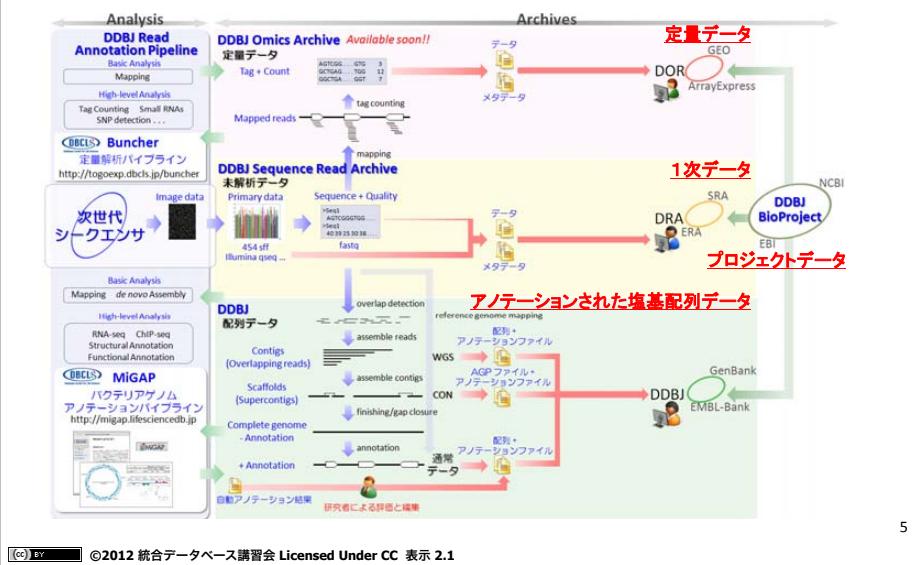
1Mbase当たりのシーケンスコスト



4

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

データの流れ



5

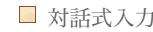
データの登録

◇一次配列データ

- D-wayを使ってSRA/DRAに登録

◇解析済み配列データ

- SAKURAを使ってDDBJに登録



- Mass Submission Systemを使ってDDBJに登録



◇定量データ

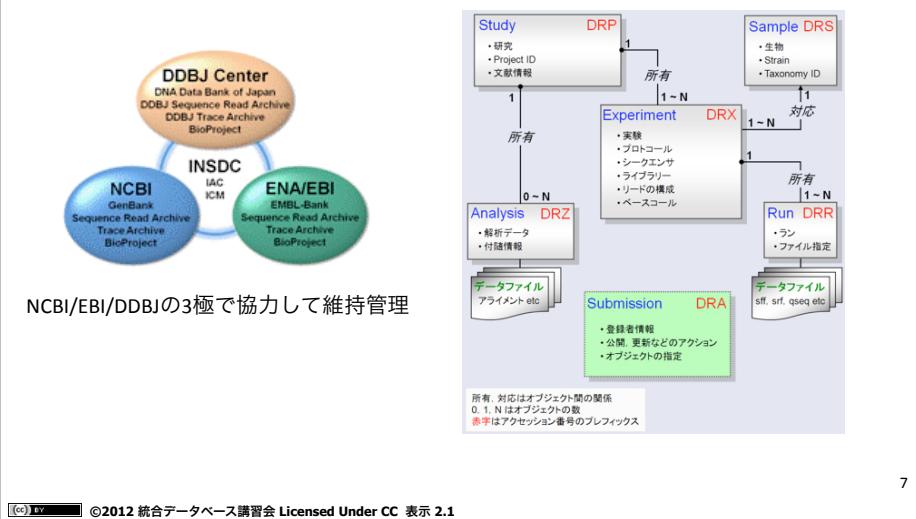
- DOR (GEO)

6

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

SRA/DRA

◆Sequence Read Archive/DDBJ Sequence Read Archive



7

<http://ddbj.nig.ac.jp/>

“DDBJ”で検索

The screenshot shows the DDBJ website search results for "DDBJ".

Search Results:

- DDBJ : DNA Data Bank of Japan**
- Hot Topics:**
 - 2012.07.12 DDBJ Read Annotation Pipeline サービス再開
 - 2012.07.06 ナミズウムシ (*Dugesia japonica*) EST データの公開
 - 2012.07.04 アフリカツメガエル (*Xenopus laevis*) GSS データの公開
- Maintenance:**
 - スクリーンの利用方法
 - スクリーンの利用方法
 - スクリーンマニュアル
- Information:**
 - DDBJ Web Magazine No.72

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

8

DDBJ Sequence Read Archive

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

登録に必要なメタデータとデータファイル
登録方法
データの検索・ダウンロード
統計

「DRAへの登録」動画マニュアル » DDBJ Youtube チャンネル

1. 必要なデータ (4:09)
Sequence Read Archive

DRAへの登録1
必要なデータ

2012.04.24 ©2012 統合データベース講習会 Licensed Under CC 表示 2.1

NGSデータの登録

◆ メタデータ

○ 登録者、生物種、シーケンサ、解析方法、配列長、etc...

○ XML形式で記述する

○ DDBJが提供しているMetaDefineで登録に最低限必要な部分は作成可能

XML形式

```
<?xml version="1.0" encoding="UTF-8"?>
<STUDY_SET xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <STUDY xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" accession="SRP001055" center="Cold Spring Harbor Laboratory" alias="Rapid Genomic Characterization of the Genus Vitis">
    <DESCRIPTION>
      <STUDY_TITLE>Rapid Genomic Characterization of the Genus Vitis</STUDY_TITLE>
      <STUDY_TYPE existing_study_type="Population Genomics"/>
      <STUDY_ABSTRACT>Next-generation sequencing technologies promise to...</STUDY_ABSTRACT>
      <CENTER_NAME>Cold Spring Harbor Laboratory</CENTER_NAME>
      <PROJECT_NAME/>
      <STUDY_DESCRIPTION>Genomic DNA libraries constructed for 17 Vitis...</STUDY_DESCRIPTION>
    </DESCRIPTION>
    <STUDY_LINKS>
      <ENTREZ_LINK>
        <ID>pubmed</ID>
        <ID>20084295</ID>
      </STUDY_LINK>
    </STUDY_LINKS>
  </STUDY>
</STUDY_SET>
```

11

D-way

To submit your data to the DDBJ Sequence Read Archive (DRA), create an account by e-mail request. Please read the following D-way Submission Account Manual.

D-way Submission Account Manual (Japanese)
D-way Submission Account Manual (English)

If you have questions, please [contact us](#).

Site Policy | Privacy | © DNA Data Bank of Japan

Webからアカウント作成は可能だが機能が制限されている

- BioProjectの登録のみ可能、配列の登録はできない
- 配列を登録するには公開鍵を添えてメールでアカウントを申請する

10

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

MetaDefine公開版

Submission Comment :

Center Name* : Individual

Laboratory Name* : name, lab, Your, Enter

Hold / Release* : Hold Until

Contact

#	Name*	E-mail*
New		
Copy		
Delete		

12

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

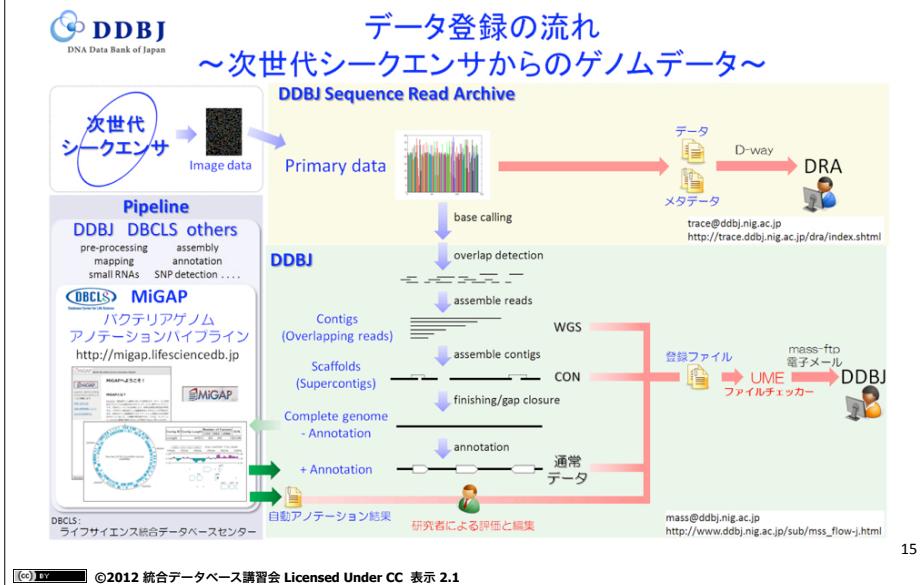
<http://trace.ddbj.nig.ac.jp/tools/>

MetaDefine

- ◆ 必要項目を入力後、各メタデータを関連付ける
 - ◆ Validate All → SubMit で登録（できるはず）

Submission	Study	Sample	Experiment	Run	Analysis (optional)	Relation
<p>Relate the created objects by drag & dropping them from left to right. In the Relation box, the objects will be listed in the following order from the top: Submission > Study > Sample > Experiment > Run > Analysis</p> <div style="border: 1px solid #ccc; padding: 5px; margin-bottom: 10px;"> ? </div> <div style="display: flex; justify-content: space-between;"> <div style="width: 45%;"> <p>Objects :</p> <ul style="list-style-type: none"> Undefined-0000_Run_0003 Undefined-0000_Experiment_0003 </div> <div style="width: 45%;"> <p>Relation :</p> <ul style="list-style-type: none"> ▼ Undefined-0000_Submission <ul style="list-style-type: none"> ▼ Undefined-0000_Study_0001 <ul style="list-style-type: none"> ▼ Undefined-0000_Sample_0001 <ul style="list-style-type: none"> ▼ Undefined-0000_Experiment_0001 <ul style="list-style-type: none"> <input type="checkbox"/> Undefined-0000_Run_0001 ▼ Undefined-0000_Experiment_0002 <ul style="list-style-type: none"> <input type="checkbox"/> Undefined-0000_Run_0002 </div> </div>						

データの流れ

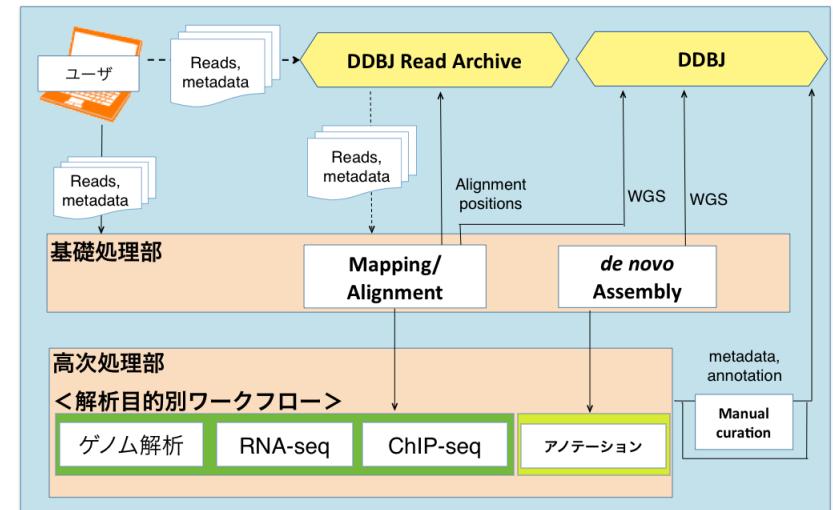


本日のアウトライン

- ◆ SRA/DRA
 - ◆ D-way (MetaDefine)
 - ◆ DDBJ Read Annotation Pipeline
 - ◆ DBCLS p-galaxy
 - ◆ MiGAP

©2012 総合データベース講習会 | licensed Under CC 表示 2.1

DDBJ Read Annotation Pipeline



©2012 統合データベース講習会 Licensed Under CC 表示 2.1

DDBJ Read Annotation pipeline 使用例

- ◆ SRA/DRAに登録されているピノノワール種の次世代シーケンスデータ
 - パイプラインにログイン
 - DRAの検索システムを使ってデータを検索
 - データをパイプラインにインポート
 - de novo assembly
 - 既知ブドウゲノムにmapping

17

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

DDBJ Read Annotation Pipeline

ENGLISH サイト内検索

HOME 塩基配列の登録 利用の手引き 検索・解析 FTP・WebAPI レポート・統計 お問い合わせ

DDBJ : DNA Data Bank of Japan
DDBJ (日本DNAデータバンク)は欧洲と米国との対応機関
(EBIおよびNCBI)と密接に協力しながら DDBJ/EMBL/GenBank
国際塩基配列データベースを構築している三大学間DNAデータバンクのひとつです
Photo by: Hidetaka Nagai

SAKURA
大鼠遺伝子システム(MSS)
マークの修正・更新
DDBJ Sequence Read Archive
DDBJ Trace Archive

Hot Topics
2012/07/12 DDBJ Read Annotation Pipeline サービス再開
2012/07/06 ナミカズミ (Dugesia japonica) EST データの公開
2012/07/04 ゼンマイカツオガエル (Xenopus laevis) GSS データの公開

Maintenance
2012/07/02 DAD リリース59.0における不具合のお詫び

Information
DDBJ Web Magazine No.72

塩基配列の登録・更新
塩基配列の登録手順を御案内します。
登録データの修正・更新
登録データが修正される方は、最初にお読みください。

FTP・Web API
FTP (ftp.ddbj.nig.ac.jp)
DDBJリリースなどのデータファイルをダウンロードできます。
Web API
DDBJが提供するサービスをユーザーが作成したプログラムから御利用頂けます。

日本DNAデータバンク INDBA

18

DDBJ Read Annotation Pipeline



English Japanese

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

LOGIN

新規アカウント作成 ゲストとしてログイン

User ID: _____
Password: _____
Login

動作中JOBの確認
PipelineのIDをお持ちでない場合、ゲストとしてログインすることができます。

マニュアルおよびチュートリアル
• 日本語マニュアル
• 英語マニュアル
• DBCLS 統合TV チュートリアル1 - 今日からはじめるDDBJ Read Annotation Pipeline
• DBCLS 統合TV チュートリアル2 - DDBJ Read Annotation Pipelineによるde novo Assembly解析

DRAアカウント登録
DRAアカウントの登録に関しては [こちらをご覧ください。](#)

pipeline_info Please check DDBJ Pipeline's help pages for further explanation. goo.gl/qsjnc
13 days ago · reply · retweet · favorite
pipeline_info New features:1)Velvet revival 2)A

19

Data setup 画面

FTPを利用したファイルアップロード (準備中)

DDBJ DNA Data Bank of Japan
ACCOUNT login ID [prenoddbj]
Logout Change password
ANALYSIS Data setup DRA Start FTP upload HTTP upload DRA Import Preprocessing Start step-1 Preprocessing Mapping / de novo Assembly step-2 Workflow Genome (SNP/Short Indel) RNA-seq (Tag count) ChIP-seq
JOB STATUS step 1 Preprocessing step 1 Mapping step 1 de novo Assembly step 2 All status
HELP

DRAからデータをインポート

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

FTP upload Private DRA entry Import public DRA Preprocessing HTTP upload
Import public FASTQ files from DRA database.
Here is the section of automatic download of public DRA/ERA/SRA entries.
Please input DRA/ERA/SRA accession number. Then the pipeline system import metadata and FASTQ files from DRA database.

Input DRA/ERA/SRA Accession Number Add my DRA entry
Accession Number can find here. DRA Search
Your request. (Here is display only, can not select.)
To select your downloaded entries. See Private DRA entry tab.
When the status makes "done", your requested entry is added in "Private DRA entry" tabs.
When the status makes "failed" or "preparing", please retry it.
queued : waiting or during download, done : file is ready, failed : please retry it, preparing : file is not yet in Status Submission Request date

20

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

DRA登録データの検索

The screenshot shows the DRASearch interface with a search bar at the top. A red arrow points from the 'Organism' input field to the search results table below. The table lists various organisms and their study counts, with 'Vitis' highlighted in green.

Organism	#	Organism Name	Study	#	Study Type	Study	#	Center Name	Study
Homo sapiens	1		1154	1	Whole Genome Sequencing	6264	1	JCVI	1888
unidentified	2		884	2	Transcriptome Analysis	1727	2	GEO	1831
Mus musculus	3		725	3	Metagenomics	1515	3	JGI	1768
Drosophila melanogaster	4		263	4	Epigenetics	914	4	UMIGS	1350
Caenorhabditis elegans	5		186	5	Resequencing	628	5	B1	936
metagenome sequence	6		181	6	RNASeq	617	6	SC	427
marine metagenome	7		162	7	Other	557	7	WUGSC	425
Arabidopsis thaliana	8		142	8	Population Genomics	164	8	BCM	311
Escherichia coli str. K-12 substr. MG1655	9		135	9	Gene Regulation Study	72	9	BioProject	292
soil metagenome	10		121	10	Pooled Clone Sequencing	72	10	UT-MGS	77

Copyright©DNA Data Bank of Japan. All Rights Reserved.

21

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

検索結果

The screenshot shows the DRASearch interface with a search result for 'Vitis'. A red arrow points to the study ID 'SRP001055' in the first row of the results table. Another red arrow points to the 'Sample' section in the right-hand panel, which lists multiple samples with their corresponding SRA IDs.

STUDY	SUBMISSION	STUDY_TITLE	STUDY_TYPE	ORGANISM	BASES	SUBMITTED	CENTER_NAME
1 SRP001055	SRA009211	Rapid Genomic Characterization of the Genus Vitis	Population Genomics	Vitis	1000000000	2009-07-23	Cold Spring Harbor Laboratory

Copyright©DNA Data Bank of Japan. All Rights Reserved.

22

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

SRA/DRAデータのインポート

The screenshot shows the 'Selecting Query Files' interface. A red arrow points from the 'Add my DRA entry' button to a confirmation dialog box. The dialog box contains instructions to click 'OK' to start import and includes checkboxes for 'Option', 'Send a mail when completed importing.', and 'Show a accessions list.'.

Confirmation

Click a OK button to start import.
This operation may take several minutes to several hours.

Option
Send a mail when completed importing.
Show a accessions list.

Input DRA/ERA/SRA Accession Number
SRA009211 Add my DRA entry

Accession Number can find here.
[DRA Search](#)

Your request. (Here is display only, can not select.)

To select your downloaded entries. See Private DRA entry tab.
When the status makes "done", your requested entry is added in "Private DRA entry".
When the status makes "failed" or "preparing", please retry it.

queued : waiting or during download, done : file is ready, failed : download is ok, but md5 was not check.

Status	Submission	Request
done	SRA009211	2012-0
preparing	SRA026538	2012-0

23

インポートされたデータ

The screenshot shows the DDBJ interface with a 'Selecting Query Files' step. A red arrow points to the 'Preprocessing' step in the pipeline. The main panel displays a table of imported SRA entries, with a red circle highlighting the 'SRP001055' row. Another red arrow points to the 'Study' section in the right-hand panel, which shows the study details for 'SRP001055'.

TYPE	ACCESSION	ALIAS	FILENAME	DL	VIEW
Submission	SRP009211	Vitisvirens	SRA009211 submission.xml	Download	View
Sample	SRP009211		SRA009211 sample.xml	Download	View
Study	SRP001055	Rapid Genomic Characterization of the Genus Vitis	SRA009211 study.xml	Download	View

画面右の「Preprocessing」から、配列のクオリティーなどをチェックできる

Preprocessing -> 計算対象ランデータを選択 -> NEXT

Copyright©DNA Data Bank of Japan. All Rights Reserved.

24

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

3

Preprocessing

Set Parameters for Preprocessing

Your selected queries
Run ACCESSION Read length Quality Score Read Layout
SRR031112 bp single

Steps of preprocessing workflow

- Step1: Calculating Quality Score(QS) of reads.
 - [Graph1] Mean and Standard deviation of Phred quality score by read position.
 - [Graph2] Frequency of bases by Phred quality score.
- Step2: BASE TRIMMING with 5'end and 3'end of each read.
 - Trimming bases with less than QS : 20
- Step3: READ REMOVING for remained low quality bases. (Optional) NA.
 - Set a THRESHOLD of QS : 0
 - Set a percentage of bases in a read with (QS=>THRESHOLD) to be held(NOT REMOVED).
 - [Graph3] Percentage of remained bases by read position after trimming and removing.

For paired-end reads, all read pairs without one of them are removed.

確認画面

Run Confirmation

Email notification
Send email notification when the job is completed or aborted with error:
kawano@dbcls.jp * Required

Confirmation of entries
Query sets
• SRR031112 -inbred_pinothor

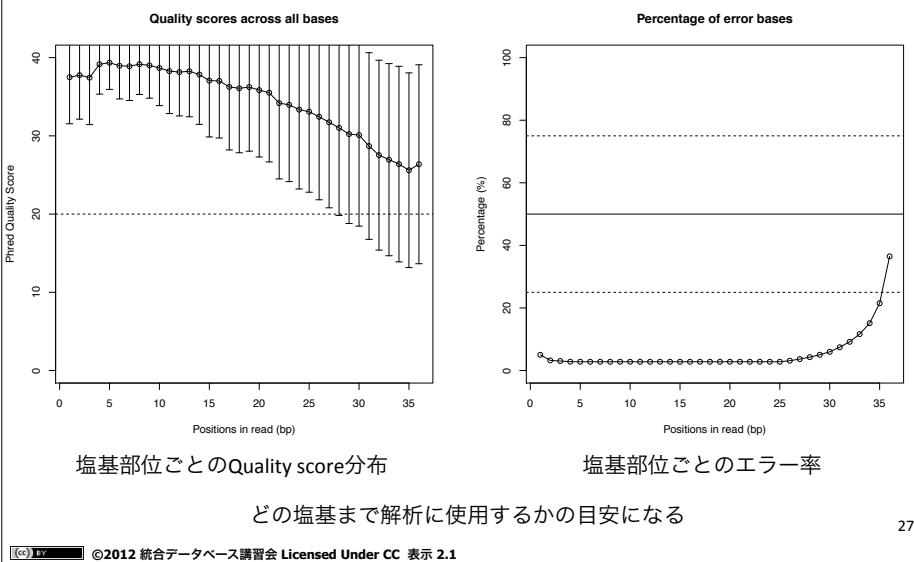
パラメータ設定

NEXT

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

25

Preprocessing 結果



27

進捗確認画面-Preprocessing

Status - Preprocessing

Detail view

ID	Tool (Version)	RunAccession or Filename	Download	Read length	Alias
4095	(1.0)	SRR031112.fastq.bz2	download (541.1 MB)	N.A. bp	inbred_pinothor

Order

Sort by: ID Descending Show Only Your Own Job Reload

Time

ID	User ID	Files	P/S	Status	Read #	Read length	Detail	Start time	End time
4166	--	Preprocess_Sai	P	complete	--	--	--	2012-07-31 14:28:19	2012-07-31 14:28:19
4165	--	e_coli_MAPPIN	P	complete	--	--	--	2012-07-31 14:28:19	2012-07-31 14:28:19
4098	--	SRA029956	S	complete	--	--	--	2012-07-25 18:24:51	2012-07-25 18:24:51
4095	orenoddbj	SRA009211	S	complete	--	--	--	2012-07-25 18:23:43	2012-07-25 18:23:43
4094	orenoddbj	SRA009211	S	complete	--	--	--	2012-07-25 17:04:58	2012-07-25 17:04:58
4093	orenoddbj	SRA009211	S	complete	--	--	--	2012-07-25 15:37:42	2012-07-25 15:37:42

Command

ID	Start time	End time	Log1	Log2	Result	MDS
4095	2012-07-25 16:56:32	2012-07-25 16:56:32	View	View	Success	
4094	2012-07-25 17:04:45	2012-07-25 17:04:45	View	View	Success	
4093	2012-07-25 17:04:46	2012-07-25 17:04:46	View	View	Success	

背景無色（白）が他のユーザ
背景黄色が自分のジョブ

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

26

パイプラインの実行～de novoアセンブル

Select Query Files

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

ANALYSIS

FTP upload Private DRA entry Import public DRA Preprocessing HTTP upload

Metadata of the DRA entry:

Sample

TYPE	ACCESSION	ALIAS	FILENAME	DL	VIEW
Submission	SRA009211	VitisVitis	SRA009211 submission.xml	Download	View
SR030099	muscat_solo1				
SR030102	thompson_solo1				
SR030104	malvasia_solo1				
SR030106	gewurztraminer_solo1				
SR030108	malvasia_solo1				
SR030110	plavennais_solo1				
SR030112	vermentino_solo1				
SR030113	vpramata_solo1				
SR030115	nesting_solo1				
SR030117	inbred_pinothor				

Study

No.	Experiment	Sample	Run	ACCESSION	STRAIN	Run_date	Read #	Read length	Instrument model	Layout
1	SR031389	SR500899	SR031097						ILLUMINA	single
2	SR031390	SR5008102	SR031098						ILLUMINA	single
3	SR031391	SR5008103	SR031099						ILLUMINA	single
4	SR031392	SR5008104	SR031100						ILLUMINA	single
5	SR031393	SR5008105	SR031101						ILLUMINA	single
6	SR031395	SR5008107	SR031102						ILLUMINA	single
7	SR031396	SR5008108	SR031103						ILLUMINA	single
8	SR031397	SR5008109	SR031104						ILLUMINA	single
9	SR031398	SR5008110	SR031105						ILLUMINA	single
10	SR031399	SR5008111	SR031106						ILLUMINA	single
11	SR0314000	SR5008112	SR031107						ILLUMINA	single
12	SR0314001	SR5008113	SR031108						ILLUMINA	single
13	SR0314002	SR5008115	SR031109						ILLUMINA	single
14	SR0314003	SR5008116	SR031110						ILLUMINA	single
15	SR0314004	SR5008114	SR031111						ILLUMINA	single
16	SR0314005	SR5008117	SR031112						ILLUMINA	single
17	SR0313994	SR5008106	SR031124						ILLUMINA	single

From metadata Counted from FASTQ (Sequence length is calculated from the first entry)

DRA Startで初期画面に戻る

計算対象ランデータを選択 -> NEXT

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

28

3

解析ツールの選択

Selecting Tools for Basic Analysis of DDBJ ANNOTATION PIPELINE

マッピング系ツール

Tool	Help	Version	Base space	Color space	Paired-end	Depth	Coverage	Error rate	SNP	Indel	.gff	.bed	.SAM	Comment	
BLAT		34													Single-end analysis only
MAQ		0.7.1													
bwa		0.5.9													
SOAP		2.21													
Bowtie		0.12.7													
TopHat		1.0.11													

de novo Assembly
Total limit is 22 Gbp

Tool	Help	Version	Base space	Color space	Paired-end	MSS(WGS)	Comment								
SOAPdenovo		1.05													Maximum K-mer value is 64.
ABySS		1.3.2													We severely recommend when performing Velvet, total length of those reads is up to 22G bp. Maximum K-mer value is 64.
Velvet		1.2.03													
Trinity		r2012-06-08													RNA-Seq De novo Assembly

Mapping Contigs by de novo Assembly to Reference Sequences.
The contigs will be aligned to reference genome.

Tool	Comment
BLAT	Single-end analysis only

BACK NEXT 29

解析するランの指定

Generating Query Sets from Query Read Files

Single analysis
Layout of single sequence.
5' Linker(1) Target Linker(2) 3'

Run ACCESSION	Read length	Quality Score
SRR031112	bp.	

1 (highlighted)
2 (highlighted)
3 (highlighted)

confirm

QUERY SET

RESET BACK NEXT

Generating Query Sets from Query Read Files

Single analysis
Layout of single sequence.
5' Linker(1) Target Linker(2) 3'

Run ACCESSION	Read length	Quality Score
SRR031112	bp.	

confirm

QUERY SET
Query set!
PairedOrientation RunAccession RunAlias RowLength QualityScore1 QualityScore2

single	SRR031112	inbred_pintnoir			
--------	-----------	-----------------	--	--	--

RESET BACK NEXT 31

解析ツールの詳細

マッピング系

BLAT	高速シーケンサ登場以前からあるアライメントツール。発現データはイントロンを想定したギャップを考慮
MAQ	高速シーケンサ登場初期にショートリードに対応。リード長が長くなるにしたがい開発はBWAに引き継がれる
BWA	MAQより速く、Titaniumのリードもオプションで対応
SOAP	メモリ消費量が少なく、より高速。精度はBWAより若干落ちる
Bowtie	ギャップは考慮しないが処理は速い。BWA、SOAP2、BowtieはBurrows-Wheeler変換というアルゴリズムでゲノムDNAに対してインデックスを作成、高速でマッピングする
TopHat	RNA-Seqのリードを内部でBowtieを利用してマッピング。スプライスジャンクションを特定する

アセンブリ系

SOAPdenovo	ヒト、パンダ等大型ゲノムのアセンブリで使用された。比較的高速
Abyss	初期に並列処理に対応したアセンブリ
Velvet	高速シーケンサ登場初期に開発された。メモリ消費多め
Trinity	RNA-Seq配列のアセンブリ。上記3つとともにde bruijn graphというアルゴリズムを使用

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

30

解析ツールのパラメータ設定

Setting for De Novo Assembly

soapdenovo
Set optional parameters of the single-end analysis

Step1) Make SOAP denovo configuration file
Please select a file format:
FASTQ file : -fq
*Please select FASTQ or FASTA correctly.

Please input average of insert size from all query files.
200

Please input maximum of read length from all query files.
50
perl make_config_file.pl query.fastq

Step2) Assembly
[Optional] You can input a soapdenovo custom options. (Some option is limited.)
soapdenovo all -s config_file

Step3) Create assembled sequences in FASTA file from pileupped reads to submit WGS division of DDBJ.
Set filtered length for contigs
perl lengthfilter.pl pileupFile 100 out_WGS.txt

BACK NEXT

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

32

確認画面

Run Confirmation

Destination of mail
When the request is completed, the system sends an email to this address.
kawano@dbcs.jp *Required

Assembly (soapdenovo)
Query sets
Query set1
PairedOrientation RunAccession RunAlias RowLength QualityScore1 QualityScore2
single SRR031112 inbred_pinothor

Assembly commands
soapdenovo
Set optional parameters of the single-end analysis
Step1) Make SOAP denovo configuration file
Please select a file format:
FASTQ file : 4q
* Please select FASTQ or FASTA correctly.
Please input average of insert size from all query files.
200
Please input maximum of read length from all query files.
50
perl make_config_file.pl query.fasta
Step2) Assembly
[Optional] You can input a soapdenovo custom options. (Some option is limited.)
soapdenovo all -s config_file
Step3) Create assembled sequences in FASTA file from pileupped reads to submit WGS division of DDBJ.
Set filtered length for contigs
perl lengthfilter.pl pileupFile 100 out_WGS.txt

JOB STATUS
step1 Preprocessing
step1 Mapping / de novo Assembly
step1 de novo Assembly
step2 All status

HELP
HELP (?)
TUTORIAL (?)
Contact Us
DDBJ Real Annotation
Development Team

進捗状況・計算結果の確認

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

33

背景無色（白）が他のユーザ
背景黄色が自分のジョブ

進捗確認画面-de novo Assembly

Status - de novo Assembly

Order: Sort by: ID Descending Show Only Your Own Job Reload

ID	User ID	Submission accession	P/S	Status	Tool	Read #	Read Assembly length	Mapping detail	Start time	End time	Elapsed time
4169	orenoddyb	SRA039211	S	generating	SOAPdenovo	—	—	View	—	2012-07-27 00:16:53	—
4160	—	DDA_fishmt_15	S	complete	ABYSS	602,461	—	—	2012-07-27 00:16:53	—	
4159	—	DDA_fishmt_15	S	complete	SOAPdenovo	602,461	—	—	2012-07-27 00:16:53	—	

Detail view

Job info

ID: 4082
Tool (Version): SOAPdenovo (1.05)
RunAccession or filename: SRR031112
Download: SRR031112.fastq.bz2
Read length: N.A. bp
Alias: inbred_pinothor

Download modified queries: SRR031112.fastq (574.2 MB)

Download wgs file: out_WGS.fasta (110.0 KB) コンティグファイルのダウンロード

Assembly statistics: download

Time: Wait time 0:03:1 Start time 2012-07-24 11:42:16 End time 2012-07-24 11:43:58

Command: SOAPdenovo27mer all -s soapdenovo.conf -o output
Start time: 2012-07-24 11:42:16
End time: 2012-07-24 11:43:27
Log1: View Log2: View Result: Download (5.6 MB) MDS: 34

©2012 統合データベース講習会 Licensed Under CC

パイプラインの実行～マッピング

Selecting Tools for Basic Analysis of DDBJ ANNOTATION PIPELINE

Reference Genome Mapping

Tool	Help	Version	Input data space	Base space	Color space	Paired-end	Evaluation	Error rate	Analysis	Output format	Comment
BLAT	34	✓	✓	✓	✓	✓	✓	✓	✓	gff	Single-end analysis only
Map2P	0.7.1	✓	✓	✓	✓	✓	✓	✓	✓	FASTA	
soap	2.21	✓	✓	✓	✓	✓	✓	✓	✓	FASTA	
Blastn	0.12.7	✓	✓	✓	✓	✓	✓	✓	✓	FASTA	
Total: 10.11											

de novo Assembly
Total limit = 20 Gbs

Mapping Contigs by de novo Assembly to Reference Sequences.
The contigs will be aligned to reference genome.

Tool	Comment
BLAT	Single-end analysis only

Generating Query Sets from Query Read Files

Single analysis
Layout of single sequence.
5' [Linker1] Target [Linker2] 3'
Run ACCESSION: Read length: Quality Score: SRR031112 bp

QUERY SET

Generating Query Sets from Query Read Files

Single analysis
Layout of single sequence.
5' [Linker1] Target [Linker2] 3'
Run ACCESSION: Read length: Quality Score:

QUERY SET

解析ツールの選択

解析するランの選択

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

35

- その他のリファレンスゲノムは
- 自分でゲノム配列をアップロードする
- DDBJからインポートする

リファレンスゲノムの指定

Specifying Database of Reference Genome

Major genome sets

Organisms: Arabidopsis thaliana
Genome sets: TAIR8
all check all clear
chr01.fa chr02.fa chr03.fa chr04.fa chr05.fa ChrC.fa chrM.fa

User original sets

Download or upload reference

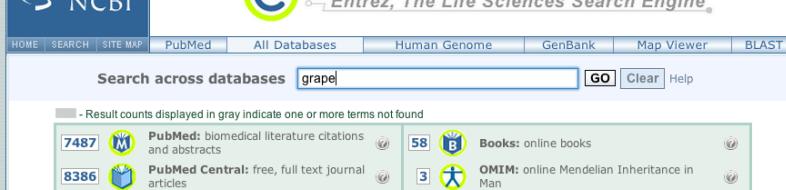
©2012 統合データベース講習会 Licensed Under CC 表示 2.1

メジャー モデル生物のリファレンスゲノムは
あらかじめプリセットされている
ヒト
マウス
線虫
イネ
シロイヌナズナ
酵母
...

36

ブドウゲノムIDの検索～NCBI Entrez

<http://www.ncbi.nlm.nih.gov/Entrez/>



The screenshot shows the NCBI Entrez search interface. The search term "grape" was entered into the search bar. Below the search bar, a message indicates that result counts displayed in gray indicate one or more terms not found. The search results are organized into two main sections: a grid of links and a list of detailed entries.

Search across databases grape **GO** **Clear** **Help**

- Result counts displayed in gray indicate one or more terms not found

7487  PubMed: biomedical literature citations and abstracts	58  Books: online books
8386  PubMed Central: free, full text journal articles	OMIM: online Mendelian Inheritance in Man
8  Site Search: NCBI web and FTP sites	

6086  Nucleotide: Core subset of nucleotide sequence records	18  dbGaP: genotype and phenotype
220547  EST: Expressed Sequence Tag records	none  UniGene: gene-oriented clusters of transcript sequences
31907  GSS: Genome Survey Sequence records	1  CDD: conserved protein domain database
60852  Protein: sequence database	Clone: integrated data for clone resources
9  Genome: whole genome sequences	none  UnISTS: markers and mapping data
381  Structure: three-dimensional macromolecular structures	171  PopSet: population study data sets
	22283  GEO Profiles: expression and molecular abundance profiles

GenBank (DDBJ) ID をコピー

Assembly Advanced Browse by organism Search

12X

Organism name: *Vitis vinifera*
Submitter: International Grape Genome Program
Assembly level: Chromosome
Genbank representation: full
GenBank Assembly ID: GCA_000003745.2 (latest)
RefSeq Assembly ID: GCA_000003745.2 (latest)
RefSeq Assembly and GenBank Assembly Identical: no
WGS Project: CAAFP3

[History](#)(Show revision history)

[Global statistics](#)

Total sequence length	466,260,629
Total assembly gap length	15,058,008
Gaps between scaffolds	154
Number of scaffolds	2,067
Scaffold N50	3,426,264
Number of contigs	14,857
Contig N50	102,754
Total number of chromosomes and plasmids	21

Assembly Definition [Assembly Statistics](#)

[Global assembly definition](#) Download the full sequence report

Click on the table row to see sequence details in the table to the right

Assembly Unit Name	Assembly Unit: Primary Assembly
Primary Assembly	Molecule name
non-nuclear	FN097015.. FN097017.. FN097018.. FN097020.. FN097022.. FN097024.. FN097025..
	GenBank ID
	RefSeq ID
	Unlocalized sequences count

39

ブドウゲノムのページ

リファレンスゲノムデータのインポート

Specifying Database of Reference Genome

RESET | **BACK** | **NEXT**

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

Retrieving a chromosome from DDBJ-DB by using HTTP REST

Input Accession Number (NSD) or (RefSeqID)
FN57015
LOAD

Pipeline

* Representative item Transfer(REST)

Uploading reference from local drive.

FASTA only (ファイルを選択) 選択されていません
2GB Flesize Limit
UPLOAD

RESET | **BACK** | **NEXT**

Vitis vinifera 各クロマソームのINSD IDを

→ 入力、LOAD を繰り返す

→ CREATE DATASET

→ CREATE GENOMESET

でリファレンスゲノムを登録する

INSD ID List

<input checked="" type="checkbox"/> FN5701705/FN5701705_1 Vitis vinifera, whole genome shotgun sequence of line PN4024, unoriented chromosome 2, chr2	DELETE
<input checked="" type="checkbox"/> FN5701708/FN5701708_1 Vitis vinifera, whole genome shotgun sequence of line PN4024, unoriented chromosome 3, chr3	DELETE
<input checked="" type="checkbox"/> FN5701709/FN5701709_1 Vitis vinifera, whole genome shotgun sequence of line PN4024, unoriented chromosome 4, chr4	DELETE
<input checked="" type="checkbox"/> FN57022/FN57022_1 Vitis vinifera, whole genome shotgun sequence of line PN4024, unoriented chromosome 5, chr5	DELETE
<input checked="" type="checkbox"/> FN57024/FN57024_1 Vitis vinifera, whole genome shotgun sequence of line PN4024, unoriented chromosome 6, chr6	DELETE
<input checked="" type="checkbox"/> FN57025/FN57025_1 Vitis vinifera, whole genome shotgun sequence of line PN4024, unoriented chromosome 7, chr7	DELETE
<input checked="" type="checkbox"/> FN57027/FN57027_1 Vitis vinifera, whole genome shotgun sequence of line PN4024, unoriented chromosome 8, chr8	DELETE
<input checked="" type="checkbox"/> FN57028/FN57028_1 Vitis vinifera, whole genome shotgun sequence of line PN4024, unoriented chromosome 9, chr9	DELETE
<input checked="" type="checkbox"/> FN57030/FN57030_1 Vitis vinifera, whole genome shotgun sequence of line PN4024, unoriented chromosome 10, chr10	DELETE
<input checked="" type="checkbox"/> FN57032/FN57032_1 Vitis vinifera, whole genome shotgun sequence of line PN4024, unoriented chromosome 11, chr11	DELETE
<input checked="" type="checkbox"/> FN57034/FN57034_1 Vitis vinifera, whole genome shotgun sequence of line PN4024, unoriented chromosome 12, chr12	DELETE
<input checked="" type="checkbox"/> FN57036/FN57036_1 Vitis vinifera, whole genome shotgun sequence of line PN4024, unoriented chromosome 13, chr13	DELETE
<input checked="" type="checkbox"/> FN57038/FN57038_1 Vitis vinifera, whole genome shotgun sequence of line PN4024, unoriented chromosome 14, chr14	DELETE
<input checked="" type="checkbox"/> FN57039/FN57039_1 Vitis vinifera, whole genome shotgun sequence of line PN4024, unoriented chromosome 15, chr15	DELETE
<input checked="" type="checkbox"/> FN57040/FN57040_1 Vitis vinifera, whole genome shotgun sequence of line PN4024, unoriented chromosome 16, chr16	DELETE
<input checked="" type="checkbox"/> FN57042/FN57042_1 Vitis vinifera, whole genome shotgun sequence of line PN4024, unoriented chromosome 17, chr17	DELETE
<input checked="" type="checkbox"/> FN57044/FN57044_1 Vitis vinifera, whole genome shotgun sequence of line PN4024, unoriented chromosome 18, chr18	DELETE
<input checked="" type="checkbox"/> FN5857047/FN5857047_1 Vitis vinifera, whole genome shotgun sequence of line PN4024, unoriented chromosome 19, chr19	DELETE

CREATE DATABASE

RESET | **BACK** | **NEXT**

リファレンスゲノムの選択

Specifying Database of Reference Genome

RESET BACK NEXT

Major genome sets

User original sets

Genome sets
Vitis vinifera genome

>FN597017|FN597017.1 Vitis vinifera, whole genome shotgun sequence of line PN40024
>FN597018|FN597018.1 Vitis vinifera, whole genome shotgun sequence of line PN40024
>FN597020|FN597020.1 Vitis vinifera, whole genome shotgun sequence of line PN40024
>FN597022|FN597022.1 Vitis vinifera, whole genome shotgun sequence of line PN40024
>FN597024|FN597024.1 Vitis vinifera, whole genome shotgun sequence of line PN40024
>FN597025|FN597025.1 Vitis vinifera, whole genome shotgun sequence of line PN40024
>FN597027|FN597027.1 Vitis vinifera, whole genome shotgun sequence of line PN40024
>FN597028|FN597028.1 Vitis vinifera, whole genome shotgun sequence of line PN40024
>FN597030|FN597030.1 Vitis vinifera, whole genome shotgun sequence of line PN40024
>FN597032|FN597032.1 Vitis vinifera, whole genome shotgun sequence of line PN40024
>FN597034|FN597034.1 Vitis vinifera, whole genome shotgun sequence of line PN40024
>FN597036|FN597036.1 Vitis vinifera, whole genome shotgun sequence of line PN40024
>FN597038|FN597038.1 Vitis vinifera, whole genome shotgun sequence of line PN40024
>FN597039|FN597039.1 Vitis vinifera, whole genome shotgun sequence of line PN40024
>FN597040|FN597040.1 Vitis vinifera, whole genome shotgun sequence of line PN40024
>FN597042|FN597042.1 Vitis vinifera, whole genome shotgun sequence of line PN40024
>FN597044|FN597044.1 Vitis vinifera, whole genome shotgun sequence of line PN40024
>FN597046|FN597046.1 Vitis vinifera, whole genome shotgun sequence of line PN40024
>FN597048|FN597048.1 Vitis vinifera, whole genome shotgun sequence of line PN40024

Download or upload reference

RESET BACK NEXT

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

41

進捗確認画面-Mapping

DDJB

ACCOUNT
User ID (prescribed)
Logout
Change password

ANALYSIS
Data setup
FTP upload
HTTP upload
Preprocessing
Mapping /
de novo Assembly
Workflow
Genome (N)PShort
SNPs (Tag count)
Job Status

JOB STATUS
Mapping
de novo Assembly

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

Status - Mapping

Mapping Job de novo Assembly Job Preprocessing Job

Job info
ID: 4170
Tool (Version): bwa (0.5.9)

RunAccession or filename: SKR031112

Detail view

ID User ID Submission accession PGS Status Tool Read # Read length Start time Elapsed time

4170 inbred_pindnor inbred_pindnor bwa 123,456,345 100 2012-08-01 11:54:31 18:47:13

4158 --- Kishinokawa g running bwa 123,456,345 100 2012-08-01 11:54:31 18:47:13

4157 --- r110_2 P complete Bowtie 4,296,782 3M 2012-07-31 11:11:54 2012-07-31 12:00:00

4154 --- r110_2 P running bwa 4,296,782 3M 2012-07-30 15:24:01 2012-07-30 15:24:01

4156 --- r110_2 S error bwa 0 3M ---

4155 --- r110_2 P error bwa 0 3M ---

Download modified queries
• SKR031112.fasta (583.0 MB)

Download merged pileup file
• merged.var.fl.vcf (2.5 GB)
• merged.sam (2.1 GB)

Download wgs file
• out_WGS.fasta (95.7 MB) コンティグファイルのダウンロード

Position errors Map ratio Depth, Coverage

Time

Wait time Start time End time
0:138 2012-08-01 11:04:45 2012-08-01 11:54:30

背景無色（白）が他のユーザ
背景黄色が自分のジョブ

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

43

パラメータ設定、確認、実行

Setting for Reference Genome Mapping

bwa
Set optional parameters of the single-end analysis
Please select an algorithm.
for short query [defaults]

If you are using LS-454 reads, select "for long query(BWA-SW)".

Step1) Convert reference sequence
bwa index refgenome.fasta

Step2) Map
bwa aln refgenome.fasta query1.fasta[?fasta] > out.sam

Step3) Remove reads, which are mapped multiple on the genome from out.sam.
 Uniq (optional)

Step4) Convert the read alignment to BAM format
samtools view -bS -o out.bam out.sam

Step5) Detect DNA polymorphism
samtools view -b var.raw.bcf > out.var.vcf

Please choose one of the following.

Step6) Analysis by Depth, Coverage
samtools mpileup -u -C0 -BQ0 -D10000000 -r refgenome.fasta.out.bam | bcftools view -bvcg > ->

Step7) Create assembled sequences in FASTA file from pileup reads to submit WGS division of DDBJ.
 getConsGeno_4pipeline.pl pileup [Not to include insertion of pileup reads...] out_WGS.txt

Destination of mail
When the request is completed, the system sends an email to this address.
kawano@dbcli.jp * Required

Reference Genome Map [bwa]

Query sets
Query set
PairedOrientation RunAccession RunAlias RowLength QualityScore1 QualityScore2
single SKR031112 inbred_pindnor

Run Confirmation
BACK RUN

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

42

マッピング結果のビューア

Tablet

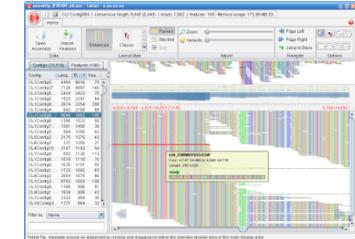
- http://bioinf.scri.ac.uk/tablet/
- windows, Mac OSX, Linux

IGVtools

- http://www.broadinstitute.org/igv/
- windows, Mac OSX, Linux

samtools tview

- http://samtools.sourceforge.net
- Mac OSX, Linux



©2012 統合データベース講習会 Licensed Under CC 表示 2.1

44

本日のアウトライン

- ◆ SRA/DRA
- ◆ D-way (MetaDefine)
- ◆ DDBJ Read Annotation Pipeline
- ◆ DBCLS p-galaxy
- ◆ MiGAP

45

 ©2012 統合データベース講習会 Licensed Under CC 表示 2.1

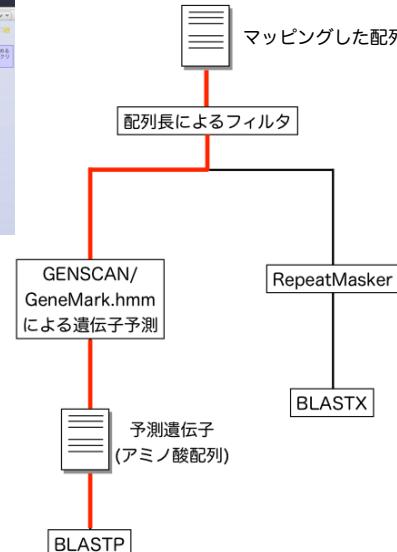
p-Galaxyを使った配列アノテーション



The screenshot shows the p-Galaxy web interface with several tool panels visible. On the left, there's a 'Workflow' panel with a green checkmark icon. In the center, there's a 'History' panel showing a recent job named 'usegalaxy.org'. The main workspace is mostly empty at this point.

<https://p-galaxy.genes.nig.ac.jp>

(開発中のサービスのため、一部サービスが
使えなかったり、一時的にアクセスできなくな
ったりする可能性があります)



47

 ©2012 統合データベース講習会 Licensed Under CC 表示 2.1

Galaxyとは？

- ◆ ゲノムなど生物学データを対象とした、データ解析、共有、公開のためのプラットフォーム
- ◆ オリジナルはペンシルバニア州立大学、エモリー大学を中心としたGalaxy teamが開発
- ◆ 独自のツールを追加して公開可能
 - DBCLS Galaxy
 - テキスト系ツール
 - p-Galaxy
 - 次世代解析ツール

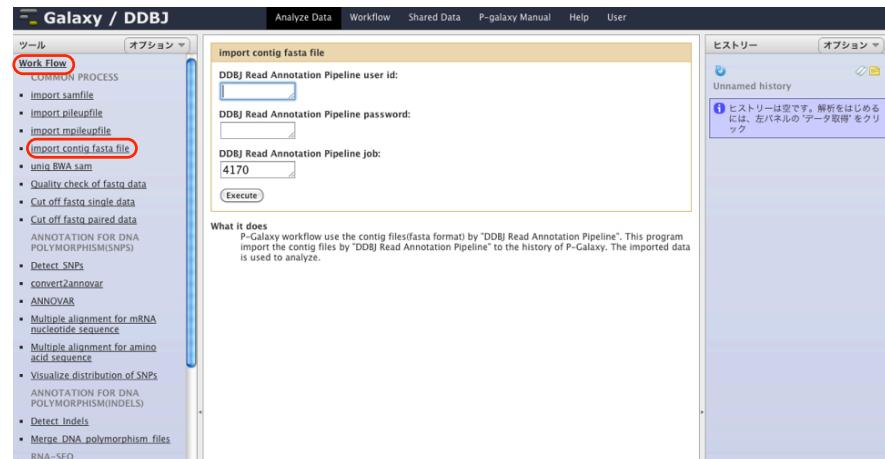


<http://usegalaxy.org/>

46

 ©2012 統合データベース講習会 Licensed Under CC 表示 2.1

DDBJパイプラインからのデータインポート



The screenshot shows the Galaxy web interface with the 'Work Flow' panel open. A specific step in the workflow is highlighted: 'import contig fasta file'. To the right, there's a detailed configuration form for the 'DDBJ Read Annotation Pipeline'. The form includes fields for 'user id', 'password', and 'job'. Below the form, there's a note explaining that the pipeline uses contig files to import data into the history for analysis.

DDBJ Read Annotation Pipelineで計算した結果をインポートできます

が、現在、うまくいきません

48

 ©2012 統合データベース講習会 Licensed Under CC 表示 2.1

ダウンロードしたファイルをアップロード

The screenshot shows the Galaxy DDBJ web interface. The left sidebar contains a 'Work Flow' section with a red box highlighting the 'Get Data' item, which has a sub-item 'Upload File from your computer' also highlighted with a red box. The main content area is titled 'Upload File' and contains fields for 'File Format' (set to 'Auto-detect'), 'File' (with a browse button), 'URL/Text' (a large text input field), 'Convert spaces to tabs' (checkbox checked), 'Genome' (a dropdown menu with 'Click to Search or Select' and an 'Execute' button), and a detailed 'Auto-detect' note at the bottom. The right sidebar shows a history panel with an entry for 'contig.fasta file'.

49

データの確認

50

操作が失敗すると赤色

コンテイグの配列長でフィルタリング

Galaxy / DDBJ

Analyze Data Workflow Shared Data P-galaxy Manual Help User

ツール オプション ▾

ANNOTATION FOR DE NOVO ASSEMBLED SEQ.

- FASTA File Length Filter **選択**
- Gene Prediction Choose GENSCAN or GeneMark.hmm
- Gene Prediction to FASTA Converts GENSCAN or GeneMark.hmm output file to FASTA format
- BLASTP
- RepeatMasker
- BLASTX

PHYLOGENETIC ANALYSIS

- sam_to_fasta for get mapping fasta data

FASTA File Length Filter

Input FASTA File: 2: out_WGS_pino_bwa.fasta

Base length of data removing from input file: 500

The sequence data that length is the same as or less than this value is removed.

Execute

ヒストリー オプション ▾

Unnamed history

2: out_WGS_pino_bwa.fasta

1: contig fasta file

Galaxy / DDBJ

Analyze Data Workflow Shared Data P-galaxy Manual Help User

ツール オプション ▾

ANNOTATION FOR DE NOVO ASSEMBLED SEQ.

- FASTA File Length Filter
- Gene Prediction Choose GENSCAN or GeneMark.hmm
- Gene Prediction to FASTA Converts GENSCAN or GeneMark.hmm output file to FASTA format
- BLASTP
- RepeatMasker
- BLASTX

PHYLOGENETIC ANALYSIS

- sam_to_fasta for get mapping fasta data

FASTA File Length Filter

Show all I Save This dataset is large and only the first megabyte is shown below.

>ENTRY_1_986
GAAGATTCGACGCGCTATGAAAATACATGCCTCTACACAAAGCAGTAAGAAGACTTCGTCGAATGAAAGACTCCCTCAAATTAGA
>ENTRY_1_2288
TGTTCATGTTGGAGCTGAGTTAGGGCTGGTTAGCTGAGTCATCTTGGAGACGAGCTTGAGTTCAGATGAGCAGATCTATATCCG
>ENTRY_1_2292
GTTGGTGTGAAAGATCCTTGGAGACTATACAGCAGCTAGCATGAGACATGGGATGATGATGCRAGCTGGGCCCTTCCCTATGCACT
>ENTRY_1_2300
AAGATATTCGAGAAATGGTATCATATGAACTTATGAGTGAAGATACTTAAAGATTTATATTANTCCATTATTCCTGCGCAAGACTTAACT
>ENTRY_1_2311
AAGATATTCGAGAAATGGTATCATATGAACTTATGAGTGAAGATACTTAAAGATTTATATTANTCCATTATTCCTGCGCAAGACTTAACT
>ENTRY_1_3486
AGGACAGGGGAGACTACATGAGCTTGGAGCTGAGACTGACATGAGCTTGGAGCTGAGACTGACATGAGCTTGGAGCTGAGACTGAG
>ENTRY_1_3487
AGGACAGGGGAGACTACATGAGCTTGGAGCTGAGACTGACATGAGCTTGGAGCTGAGACTGAGCTTGGAGCTGAGACTGAG

ヒストリー オプション ▾

Unnamed history

3: FASTA File Length Filter

2: out_WGS_pino_bwa.fasta

1: contig fasta file

57

ORF予測

52

予測結果からペプチドの配列を抽出

The screenshot shows two panels. The left panel displays a history step titled "Gene Prediction to FASTA" with a dropdown menu set to "Gene Prediction on data 3". The right panel shows a history step titled "BLASTP" with a dropdown menu set to "out_WGS_pino_bwa.fasta". Both panels have a "History" tab at the top.

Left Panel (Gene Prediction to FASTA):

- Input Gene Prediction file: 4: Gene Prediction on data 3
- Select extracting sequences: Peptide sequences
- Execute

Right Panel (BLASTP):

- Query file: 5: Gene Prediction t_A on data 4
- Select database: nr
- Expectation value: 0.00001
- Execute

予測ORF配列の機能アノテーション

The screenshot shows three panels. The left panel displays a history step titled "BLASTP" with a dropdown menu set to "nr". The middle panel shows a history step titled "Gene Prediction to FASTA" with a dropdown menu set to "FASTA on data 4". The right panel shows a history step titled "BLASTP" with a dropdown menu set to "out_WGS_pino_bwa.fasta". All panels have a "History" tab at the top.

Left Panel (BLASTP):

- Query file: 5: Gene Prediction t_A on data 4
- Select database: nr
- Expectation value: 0.00001
- Execute

Middle Panel (Gene Prediction to FASTA):

- Annotation for de novo assembled seq.
- FASTA File Length Filter
- Gene Prediction Choose GENSCAN or GeneMark.hmm
- Gene Prediction to FASTA Converts GENSCAN or GeneMark.hmm output file to FASTA format
- BLASTP
- RepeatMasker

Right Panel (BLASTP):

- Annotation for de novo assembled seq.
- FASTA File Length Filter
- Gene Prediction Choose GENSCAN or GeneMark.hmm
- Gene Prediction to FASTA Converts GENSCAN or GeneMark.hmm output file to FASTA format
- BLASTP
- RepeatMasker
- BLASTX
- PHYLOGENETIC ANALYSIS
- sam_to_fasta for get mapping

塩基配列から直接アノテーション: BLASTX

Diagram: A flowchart showing the direct annotation workflow. It starts with "マッピングした配列" (mapped sequence) which goes through "配列長によるフィルタ" (length filter) and "BLASTX" (highlighted with a red box). This leads to "BLASTP" (highlighted with a red box), "RepeatMasker", and finally "GENSCAN/GeneMark.hmmによる遺伝子予測" (Genes/GeneMark.hmm gene prediction). The "BLASTX" step also has a "sam_to_fasta for get mapping fasta data" option.

Screenshot:

The screenshot shows two panels. The left panel displays a history step titled "BLASTX" with a dropdown menu set to "Swiss-Pro-Plus". The right panel shows a history step titled "BLASTP" with a dropdown menu set to "out_WGS_pino_bwa.fasta". Both panels have a "History" tab at the top.

Left Panel (BLASTX):

- Annotation for de novo assembled seq.
- FASTA File Length Filter
- Gene Prediction Choose GENSCAN or GeneMark.hmm
- Gene Prediction to FASTA Converts GENSCAN or GeneMark.hmm output file to FASTA format
- BLASTP
- RepeatMasker
- BLASTX
- PHYLOGENETIC ANALYSIS
- sam_to_fasta for get mapping fasta data

Right Panel (BLASTP):

- Annotation for de novo assembled seq.
- FASTA File Length Filter
- Gene Prediction Choose GENSCAN or GeneMark.hmm
- Gene Prediction to FASTA Converts GENSCAN or GeneMark.hmm output file to FASTA format
- BLASTP
- RepeatMasker
- BLASTX
- PHYLOGENETIC ANALYSIS
- sam_to_fasta for get mapping fasta data
- Phylogenetic tree of mark gene for Plant NGS data
- Phylogenetic tree of rbcL gene for Plant NGS data

ワークフローの保存・公開

The screenshot shows three panels. The left panel displays a history step titled "BLASTX on data 3" (highlighted with a red box). The middle panel shows a "History Lists" dialog with "Saved Histories" selected. The right panel shows a "Share or Publish History 'kaju2'" dialog.

Left Panel (BLASTX on data 3):

- Annotation for de novo assembled seq.
- FASTA File Length Filter
- Gene Prediction Choose GENSCAN or GeneMark.hmm
- Gene Prediction to FASTA Converts GENSCAN or GeneMark.hmm output file to FASTA format
- BLASTP
- RepeatMasker
- BLASTX
- PHYLOGENETIC ANALYSIS
- sam_to_fasta for get mapping fasta data
- Phylogenetic tree of mark gene for Plant NGS data
- Phylogenetic tree of rbcL gene for Plant NGS data

Middle Panel (History Lists):

- History Lists
- Saved Histories
- Histories Shared with Me
- Current History
- Create New
- Clone
- Share or Publish

Right Panel (Share or Publish History 'kaju2'):

- Switch Share or Publish
- Rename
- Delete
- For Selected histories: Rename | Delete | Undo

Share or Publish History 'kaju2':

- Making History Accessible via Link and Publishing it
- This history accessible via link.
- Anyone can view and import this history by visiting the following URL: <https://p-galaxy.genomes.nig.ac.jp/u/orenndbbi/kaju2>
- You can:
 - Disable Access to History Link
 - Disables history's link so that it is not accessible.
 - Publish History
 - Publishes the history to Galaxy's Published Histories section, where it is publicly listed and searchable.
- Sharing History with Specific Users
- You have not shared this history with any users.
- Share with a user

Galaxyを使ってみる

- ヒトの22番染色体のコーディングエキソンからSNPsの多いエキソンを取り出す

https://main.g2.bx.psu.edu/galaxy101 の例

- Galaxyサーバ

http://galaxy.dbcls.jp/
http://usegalaxy.org/

57

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

UCSC Mainで22番染色体のエキソンを取り出す

The screenshot shows the UCSC Table Browser interface. The search parameters are set as follows:

- clade: Mammal
- genome: Human
- assembly: Feb. 2009 (GRCh37/hg19)
- group: Genes and Gene Prediction Tracks
- track: UCSC Genome Browser
- table: knownGene
- region: genome □ ENCODE Pilot region □ position chr22
- identifiers (names/accessions): (paste list) (upload list)
- filter: (create)
- intersection: (create)
- correlation: (create)
- output format: BED - browser extensible data (selected) □ Send output to Galaxy □ GREAT
- output file: (leave blank to keep output in browser)
- file type returned: plain text □ gzip compressed

Buttons at the bottom are highlighted with red circles: "get output" and "summary/statistics". A green circle highlights the "Galaxy" button under "Send output to". A red circle highlights the "knownGene" track entry. A green circle highlights the "knownGene" table entry.

58

UCSC Mainで22番染色体のエキソンを取り出す

The dialog box has the following settings:

- Include custom track header: checked
- name: tb_knownGene
- description: table browser query on knownGene
- visibility: pack
- url: (empty)
- Create one BED record per:
 - Whole Gene
 - Upstream by 200 bases
 - Exons plus 0 bases at each end
 - Introns plus 0 bases at each end
 - 5' UTR Exons
 - Coding Exons** (radio button selected)
 - 3' UTR Exons
 - Downstream by 200 bases
- Note: If a feature is close to the beginning or end of a chromosome and upstream/downstream bases are added, they may be truncated in order to avoid extending past the edge of the chromosome.
- Buttons: "Send query to Galaxy" (highlighted with a red circle), "Cancel".

59

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

UCSC Mainで22番染色体のSNPsを取り出す

The screenshot shows the UCSC Table Browser interface. The search parameters are set as follows:

- clade: Mammal
- genome: Human
- assembly: Feb. 2009 (GRCh37/hg19)
- group: Variation and Repeats
- track: Common SNPs (hg19)
- table: snp135Common
- region: genome □ ENCODE Pilot region □ position chr22-1-5104566
- identifiers (names/accessions): (paste list) (upload list)
- filter: (create)
- intersection: (create)
- correlation: (create)
- output format: BED - browser extensible data (selected) □ Send output to Galaxy □ GREAT
- output file: (leave blank to keep output in browser)
- file type returned: plain text □ gzip compressed

Buttons at the bottom are highlighted with red circles: "get output" and "summary/statistics". A green circle highlights the "Galaxy" button under "Send output to". A red circle highlights the "Common SNPs (hg19)" track entry. A green circle highlights the "snp135Common" table entry.

60

UCSC Mainで22番染色体のSNPsを取り出す

Output SNP135Common as BED

Include custom track header:
name= tb_snp135Common
description= table browser query on SNP135Common
visibility= pack
url=

Create one BED record per:
 Whole Gene
 Upstream by 200 bases
 Downstream by 200 bases

Note: if a feature is close to the beginning or end of a chromosome and upstream/downstream bases are added, they may be truncated in order to avoid extending past the edge of the chromosome.

[Send query to Galaxy](#) (Red box)
[Cancel](#)

ヒストリーにデータが格納される

ヒストリー オプション ▾

Unnamed history

2: UCSC Main on Human: SNP135Common (chr22:1-51304566)
1: UCSC Main on Human: knownGene (chr22:1-51304566)

61

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

エキソンにあるSNPsの数をかぞえる

Galaxy データ解析 ワークフロー 共有データ ヘルプ ユーザ

Tools オプション ▾

Group

Select data:
3: Join on data 2 and data 1 (Red box)

Query missing? See TIP below.

Group by column:
c4 (Red box)

Ignore case while grouping?:

Operations

Operation 1

Type:
 Count (Red box)

On column:
c4 (Red box)

Round result to nearest integer?:
 NO (Red box)
 Yes (Red box)

Add new Operation
実行 (Red box)

ヒストリー オプション ▾

Unnamed history

3: Join on data 2 and data 1 (Red box)
1
6,377 regions, フォーマット: interval, データベース: hg19
情報:
I view in GeneTrack | display at Ensembl Current

chrom	2.Start	3.End	4.Name
22	16287253	16287881	uc010gap.2
22	16448823	16449804	uc011agd.2
22	17071766	17073446	uc002ip.1_
22	17071766	17073446	uc002ip.1_
22	17071766	17073446	uc002ip.1_
22	17264508	17265295	uc002zv.3

2: UCSC Main on Human: SNP135Common (chr22:1-51304566)
1: UCSC Main on Human: knownGene (chr22:1-51304566)

Add new OperationをクリックするとOperations画面が現れる

63

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

エキソンとSNPsのデータをjoinする

Galaxy データ解析 ワークフロー 共有データ ヘルプ ユーザ

Tools オプション ▾

Join

Join:
1: UCSC Main on Human: 1-51304566 (Red box)

with:
2: UCSC Main on Human: 1-51304566 (Red box)

Second query

with min overlap:
1 (Red box)
(bp)

Return:
Only records that are joined (INNER JOIN) (Red box)
実行 (Red box)

TIP: If your query does not appear in the pull-down menu, it means that it is not in interval format. Use "edit attributes" to set chromosome, start, end, and strand columns.

Screencasts!
See Galaxy Interval Operation Screencasts (right click to open this link in another window).

Syntax

- Where overlap specifies the minimum overlap between intervals that allows them to be joined.
- Return only records that are joined returns only the records of the first query that join to a record in the second query. This is analogous to an INNER JOIN.
- Return all records of first query (fill null with ".") returns all intervals of the first query, and any intervals that do not join an interval from the second query are filled in with a period(.)
- Return all records of second query (fill null with ".") returns all intervals of the second query, and any intervals that do not join an interval from the first query are filled in with a period(.) Note that this may produce an invalid interval file, since a period(.) is not a valid chrom, start, end or strand.
- Return all records of both queries (fill null with ".") returns all records from both queries, and fills on either the right or left with periods. Note that this may produce an invalid interval file, since a period(.) is not a valid chrom, start, end or strand.

Example

If First query is:

62

SNPsの出現頻度順にソートする

Galaxy データ解析 ワークフロー 共有データ ヘルプ ユーザ

Tools オプション ▾

Sort

Sort Query:
4: Group on data 3 (Red box)

on column:
c2 (Red box)

with flavor:
 Numerical sort (Red box)

everything in:
 Descending order (Red box)

Column selections
Add new Column selection (Red box)

Join, Subtract and Group
Convert Formats

ヒストリー オプション ▾

Unnamed history

4: Group on data 3 (Red box)
3,729 lines, フォーマット: tabular, データベース: hg19
情報: --Group by c4: count[c4]
002zp_1_cds_0_0_chr22_17071767_3
002zw_3_cds_0_0_chr22_17264509_4
002zw_3_cds_1_0_chr22_17280661_1
002zw_3_cds_1_0_chr22_17444615_1
002zw_3_cds_2_0_chr22_1745656_1
002zw_3_cds_3_0_chr22_17446068_1

64

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

上位5個のエキソンを取り出す

The screenshot shows the Galaxy interface with a workflow history. A specific step is highlighted:

- Tools**: Tools menu.
- Get Data**, **Send Data**, **ENCODE Tools**, **Lift-Over**, **Text Manipulation** (selected).
- Select first** tool configuration:
 - 選択行数: 5 (highlighted with a red box)
 - 対象データセット: 5: Sort on data 4 (highlighted with a green box)
 - 実行 (Run) button
- ヒストリー** (History):
 - Unnamed history
 - 5: Sort on data 4 (highlighted with a green box)
 - 4: Group on data 3
 - 3: Join on data 2 and data 1
 - 2: UCSC Main on Human: SNP135Common (chr22:1-51304566)
 - 1: UCSC Main on Human: knownGene (chr22:1-51304566)

Below the history, a note says: "このツールを利用してデータセットの行頭行を選択することができます。" (You can use this tool to select the first few lines of a dataset.)

65

ワークフローの作成・保存 (ログインユーザーのみ)

The screenshot shows the Galaxy interface with a workflow history. A specific step is highlighted:

- ヒストリーリスト**
- 保存されたヒストリー**
- 共有されているヒストリー**
- 現在のヒストリー**
- 新たに作成する**
- 複製する**
- 共有する/公開する**
- ワークフロー展開** (highlighted with a red box)
- データセットのセキュリティ**
- 削除したデータセットを表示する**
- 隠れているデータセットを表示する**
- 構造を表示する**
- ファイルへエクスポート**
- 削除する**
- その他のアクション**
- ファイルからのインポート**

ツール (Workflow tools):

- UCSC Main (disabled)
- Join (highlighted with a yellow box)
- Group (highlighted with a yellow box)
- Sort (highlighted with a yellow box)
- Select first (highlighted with a yellow box)
- Convert Genome Intervals To Strict BED (disabled)
- Compare two Queries (disabled)

ヒストリーアイテムが生成されました (Workflow items generated):

- 1: UCSC Main on Human: knownGene (chr22:1-51304566)
- 2: UCSC Main on Human: SNP135Common (chr22:1-51304566)
- 3: Join on data 2 and data 1
- 4: Group on data 3
- 5: Sort on data 4
- 6: Select first on data 5
- 7: Compare two Queries on data 6 and data 1
- 8: Compare two Queries on data 6 and data 1

67

エキソンの座標情報を取り出す

The screenshot shows the Galaxy interface with a complex workflow history. A specific step is highlighted:

- Tools**: Tools menu.
- Get Data**, **Send Data**, **ENCODE Tools**, **Lift-Over**, **Text Manipulation**, **Filter and Sort**, **Join, Subtract and Group** (selected).
- Compare two Queries** tool configuration:
 - Compare: 1: UCSC Main on Human: 1-51304566 (highlighted with a red box)
 - Using column: c4 (highlighted with a red box)
 - against: 6: Select first on data 5 (highlighted with a green box)
 - and column: c1 (highlighted with a green box)
 - To find: Matching rows of 1st query (highlighted with a green box)
 - 実行 (Run) button
- ヒストリー** (History):
 - 6: Select first on data 5 (highlighted with a green box)
 - 5: Sort on data 4 (highlighted with a green box)
 - 4: Group on data 3 (highlighted with a green box)
 - 3: Join on data 2 and data 1
 - 2: UCSC Main on Human: SNP135Common (chr22:1-51304566)
 - 1: UCSC Main on Human: knownGene (chr22:1-51304566)

最終的な結果 (Final result):

- chr22 21044318 21045692 uc002zsw_2_cds_0_0_chr22_21044319_f 0 +
- chr22 32108068 32113221 uc003alo_2_cds_5_0_0_chr22_32108069_r 0 -
- chr22 32108068 32113277 uc003alp_4_cds_5_0_0_chr22_32108069_r 0 -
- chr22 46652457 46659219 uc003bh3_3_cds_0_0_chr22_46652458_r 0 -
- chr22 32108068 32113221 uc010gjw_1_cds_4_0_0_chr22_32108069_r 0 -

66

ワークフローの編集

The screenshot shows the Galaxy interface with a workflow editor. A specific step is highlighted:

- あなたのワークフロー** (Your Workflow):
 - 名前: Workflow constructed from history 'kaju'
 - ステップ数: 10
- 他の人から共有してもらっているワークフロー** (Workflows shared by others):
 - あなたと共有されているワークフローはありません
- その他オプション** (Other Options):
 - ワークフローメニューの設定 (Workflow menu settings)

ワークフローの編集 (Edit Workflow):

- Input dataset → Join (highlighted with a yellow box) → Group (highlighted with a yellow box) → Sort (highlighted with a yellow box) → Select first (highlighted with a yellow box) → Compare two Queries (highlighted with a yellow box)
- Input dataset → Join (highlighted with a yellow box) → Select data (highlighted with a yellow box) → Sort (highlighted with a yellow box) → Compare two Queries (highlighted with a yellow box)

68

他の染色体でのワークフロー実行

21番染色体のエキソンデータとSNPsデータ

The screenshot shows the Galaxy web interface. At the top, there's a navigation bar with 'ヒストリー' (History) and 'オプション' (Options). Below it is a 'Unnamed history' section containing two items:

- 2: UCSC Main on Human: 0 / 0 (sup135Common (chr21:1-48129895))
- 1: UCSC Main on Human: 0 / 0 (knownGene (chr21:1-48129895))

On the left, there are two tabs: 'Step 1: Input dataset' and 'Step 2: Input dataset'. The 'Step 1' tab has a 'SNPs' section with one item: 2: UCSC Main on Human: 1-48129895. The 'Step 2' tab has a 'Exons' section with one item: 1: UCSC Main on Human: 1-48129895.

On the right, there's a 'ワークフロー' (Workflow) section with a 'Workflow constructed from history "kaju"' button. A red arrow points from this button to a context menu that appears as a light gray box. The menu includes the following options:

- 編集 (Edit) - highlighted with a red circle
- 実行 (Run) - highlighted with a red circle
- 共有する/公開する (Share/Publicize)
- あなたと共有する (Share with you)
- ダウンロード/エクスポート (Download/Export)
- 複製する (Copy)
- 名称変更する (Change name)
- ワークフローメニュー (Workflow menu)
- 削除する (Delete)

At the bottom left, there's a note: 'Return Only records that are joined (INNER JOIN)' and '操作: Hide this dataset.' At the bottom right, there's a page number '69'.

今日のアウトライン

- ◆ SRA/DRA
 - ◆ D-way (MetaDefine)
 - ◆ DDBJ Read Annotation Pipeline
 - ◆ DBCLS p-galaxy
 - ◆ MiGAP

7

本家Galaxyに豊富なTutorialがあります

The screenshot shows the Galaxy web interface. At the top, there's a navigation bar with links for Analyze Data, ワークフロー, Shared Data, Visualization, Cloud, Help, and User. The main area displays a search bar and a list of available tools:

- Get Data
- Send Data
- ENCODE Tools
- Liftover
- Tab Manipulation
- Convert Formats
- FASTA manipulation
- Filter and Sort
- Join, Subtract and Group
- Extract Features
- Fetch Sequences
- Fetch Alignments
- Get Genomic Scores
- Operate on Genomic Intervals
- Statistical Tests
- Graph/Display Data
- Regional Variation
- Multivariate regression
- Multivariate Analysis
- Evolution
- Motif Tools
- Multiple Alignments
- Metagenomic analysis
- Human Genome Variation
- Genome Diversity
- EMBOSS

Below the tools, there's a section titled "Live Quickies" featuring six tool cards:

- Mapping against custom genome
- Illumina mapping: Single End
- Illumina mapping: Paired Ends
- Basic fastQ manipulation
- Advanced fastQ manipulation
- 454 Mapping Single End

A central box contains a message about Galaxy's purpose and its support by Penn State and Emory University. Below this is a note about the Galaxy server's price.

On the right side, there's a sidebar titled "ヒストリー" (History) which is currently empty, showing "0 bytes". A blue button at the bottom right of the sidebar says "ヒストリーをはじめに表示" (Show history first).

At the bottom left, there's a footer with the CC-BY license information and a copyright notice: "©2012 統合データベース講習会 Licensed Under CC 表示 2.1".

Microbial Genome Annotation Pipeline

<http://www.migap.org>

MiGAP Microbial Genome Annotation Pipeline

■ ホーム ■ フォーラム ■ FAQ ■ ヘルプ ■ 管理情報

■ トップメニュー

- MiGAPについて
- ハイブリッド
- お問い合わせ
- バイオリンクについて
- MiGAPバージョン連携主

■ パイプラインにログイン

■ フォーラムにログイン

■ ログアウト

■ パスワード

■ オートログイン

■ ログイン

パスワードをお忘れましたか?
ユーザ名をお忘れましたか?

■ アカウント

アカウントを取得する

Home

MiGAPにおけるtRNA予測機能の不具合について

2012年5月6日(木曜日)12:52 | 最終更新: 2012年5月6日(木曜日)12:06 | 作成者: Administrator |

再開後のMiGAPにおいてtRNA抽出プログラムであるtRNAscan-SEが正常に動作していなかったため、tRNAが検出されない可能性があったことをご了承ください。このため、本日（2012年5月2日）11：30にtRNAscanを再設定いたしました。この結果、現在はtRNAscan-SEはその後正常に動作しております。

2012年5月1日移行2日11:30以前に解析が終了したジョブにつきましては、ジョブの再投入・再解析をお願いいたします。

遺伝研究新スパンコでのMiGAPサービスの再開

2012年4月29日(日曜日)20:00 | 最終更新: 2012年5月6日(木曜日)11:47 | 作成者: Administrator |

遺伝研究新スパンコでのMiGAPサービスを以下のメッセージ通り再開いたします。

新規ジョブ投入 : 2012年5月1日

なお、旧スパンコでの実行結果はまだ閲覧できません。

MiGAPとは？

2009年12月31日(木曜日)00:00 | 最終更新: 2012年5月6日(木曜日)17:32 | 作成者: Administrator |

MiGAPは、微生物ゲノム解読において評価するデータベース・定評あるアルゴリズムを組合合わせたアノテーション実行バイブルシステムです。次世代シーケンサの出力により、現在は複数の研究室や研究グループが日々と微生物ゲノムを基配列に入手することが予想されます。従来はゲノム配列別のアノテーション作業は大きな負担がかかっていました。今後機や研究室やグループでは、アノテーションのための要員を確保できないのが現状であると考えられます。

組み込む...

■ 最新ニュース

- Introduction & Practice (as May 2012)
- MiGAPにおけるRNA予測機能の不具合について
- 遺伝研究新スパンコでのMiGAPサービスの再開
- 遺伝研究新クラスタークレーンシステム移行に伴うMiGAP投入・受付停止のお知らせ
- ヨブ投入実験をしました
- 遺伝研究新スパンコにてMiGAP投入受付停止のお知らせ
- 順次延長およびバッファイブインの実験を行なっています
- MiGAPの運営の目的についています
- 遺伝研究新スパンコでのMiGAP投入受付停止のお知らせ
- ヨブ投入実験をしました

■ 閲覧ランキング

- MiGAPとは？
- フォーラムヘルプ
- ユーザ登録登出ヘルプ
- MiGAPバージョン連携主
- 分析結果をグラフ化にする
- データベース登録登出ヘルプ
- MiGAPの運営の目的についています
- 在籍システム緊急保護のためヨブ投入を停止しています
- ヨブ投入実験をしました
- ヨブ登録登出ヘルプ

検索...

7

ユーザレベルの設定

MiGAP Microbial Genome Annotation Pipeline ver2.03

Logout Help Contact Us

Pipe Line Change User Level

Pipe Line History Change User Level Current Process

b-MiGAP
s-MiGAP
g-MiGAP

Change User Level

Set

73

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

◆ b-MiGAP

- ツール・パラメータ固定（ブロンズ）

◆ s-MiGAP

- パラメータ設定が可能（シルバー）

◆ g-MiGAP

- DBやツール、パラメータ設定が可能（ゴールド）

MiGAPのアノテーションワークフロー

◆ ORF (CDS)予測

- MetaGeneAnnotator, Glimmer, Augustus

◆ rRNA予測

- 類似性検索 (16S)、RNAmmer (5S, 23S)

◆ tRNA予測

- tRNAScan-SE

◆ ORF (CDS)のアノテーション (3段階)

- COGデータベースに対する検索・マッピング

- RefSeq Microbial Proteinに対する検索

- TrEMBL Proteinに対する検索

第一段階、第二段階でアノテーションがつかなかった場合のみ

74

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

配列を入力

MiGAP Microbial Genome Annotation Pipeline ver2.03

Logout Help Contact Us

Pipe Line [Running: 1] Waiting: 0

Pipe Line Name:

Upload Filename: (ファイルを選択) 選択されていません

or paste data in box below. ([Sample data](#))

```
>SS2 ATCGATATATCGAGAATAGTCGAATCTGATTGAGCTCATATAATGGACAACTACT
CCACAAATTCTGATGAAATATTCAGTCACTCGAGAAATCTGCTACGTTGCTTGATTA
TGCATAGCGCTATGCTTCCCGTGTCTCCGGATGTTGCTGAGTGTAAAAC
CGGTCTATAACGGATTGTATGCAATGAATGATTAGTCGGCATGACAGTCACAA
GCCCTTAAATACCGCGGTATCTGGAGAACATATCGGAAATACCAACCG
CACCGTTATGCTGCTGATGCACTGGTACGGTACGGAACCTTGGGTTCTGAGGAGACT
ACCGTTATGCTGCTGATGCACTGGTACGGTACGGAACCTTGGGTTCTGAGGAGACT
AGCGGGGCATGCTTACAGAACACCAATCTAAATCTCAATGAGACATT
```

Linear Circular Bacteria Archaea Eukarya tomato

Run Clear

75

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

進捗状況確認

MiGAP Microbial Genome Annotation Pipeline ver2.03

Logout Help Contact Us

Pipe Line List

JOB CANCEL

Start	Status
2012/07/30 12:51:31	Search ORF Search rRNA Search rRNA
	Annotation Phase-1 Annotation Phase-2 Annotation Phase-3
2012/08/01 14:38:37	Search ORF Search rRNA Search rRNA
	Annotation Phase-1 Annotation Phase-2 Annotation Phase-3

76

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

計算結果の表示

MiGAP Microbial Genome Annotation Pipeline ver2.03

Pipe Line History LDAP_kawano (b-MiGAP) 2012/08/01 17:11:34 [View Menu] [Hide Menu]

Pipe Line History List contig Hide

2012/08/01 14:38:37 contig S. lycopersicum DNA contig: SISBM_S00131_04

Number of Feature

Number of CDS	Number of tRNA	Number of RBS	Number of rRNA	Software Version	COG[20030417]	RefSeq release54[2012071]
3	2	3	2	NCBI BLAST 2.2.18	3	

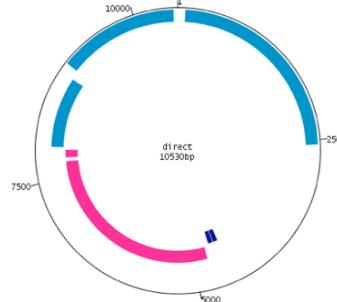
ログファイル、配列ファイル、GenBank/EMBL形式ファイル

Log File	N.A.	A.A.	Genbank	EMBL
pipeline.log	result-na.fasta.tar.gz	result-aa.fasta.tar.gz	result_gbk.tar.gz	result-a.gbk.tar.gz
			result.embl.tar.gz	result-a.embl.tar.gz

Log File 計算実行時のパラメータ、実行ステータス等
N.A. CDSにアノテーションされた部分のDNA塩基配列
A.A. CDSにアノテーションされた部分のアミノ酸配列
GenBank GenBank形式のアノテーションファイル
EMBL EMBL形式のアノテーションファイル

※ -a がついていないファイルは構造アノテーション (CDSがどこにあるか) のみの結果
-a がついているものは構造+機能アノテーションの結果が書かれている

©2012 統合データベース講習会 Licensed Under CC 表示 2.1



計算結果

アノテーション数、ツールのバージョン

Number of Feature				Annotation		
Number of CDS	Number of tRNA	Number of RBS	Number of rRNA	Software Version	COG[20030417]	RefSeq release54[2012071]
3	2	3	2	NCBI BLAST 2.2.18	3	

ログファイル、配列ファイル、GenBank/EMBL形式ファイル

Log File	N.A.	A.A.	Genbank	EMBL
pipeline.log	result-na.fasta.tar.gz	result-aa.fasta.tar.gz	result_gbk.tar.gz	result-a.gbk.tar.gz
			result.embl.tar.gz	result-a.embl.tar.gz

Log File 計算実行時のパラメータ、実行ステータス等

N.A. CDSにアノテーションされた部分のDNA塩基配列

A.A. CDSにアノテーションされた部分のアミノ酸配列

GenBank GenBank形式のアノテーションファイル

EMBL EMBL形式のアノテーションファイル

※ -a がついていないファイルは構造アノテーション (CDSがどこにあるか) のみの結果

-a がついているものは構造+機能アノテーションの結果が書かれている

78

©2012 統合データベース講習会 Licensed Under CC 表示 2.1

計算結果

DDBJ形式ファイル、GFFファイル

Download

DDBJ

result.fasta.tar.gz, result.annt.tar.gz, result.ddbj.tar.gz, result.ddbj.tar.gz(multi), result-a.fasta.tar.gz, result-a.annt.tar.gz, result-a.ddbj.tar.gz

GFF

result-a.ddbj.tar.gz(multi), result_gff.tar.gz

result.fasta 入力塩基配列

result.annt アノテーション結果

result.ddbj DDBJ形式のアノテーションファイル (コンティグ別)

result.ddbj(multi) DDBJ形式のアノテーションファイル (全コンティグ)

GFF General Feature Format ゲノムのアノテーションや場所を指定するフォーマット

※ -a がついていないファイルは構造アノテーション (CDSがどこにあるか) のみの結果

-a がついているものは構造+機能アノテーションの結果が書かれている

79

LSQA

<http://qa.lifesciencebedb.jp>

ログイン 検索 よくある質問

Q&A

質問 タグ ユーザー バッジ 未回答 検索する

welcome to #LSQA ただいまベータ版テスト中です。そのため頻繁に停止されがちあります。質問するならライフサイエンスQA

リード長の長いRNA-Seqのタグカウント 479 回答 479 開質 20 分前 hono ♦♦ 81

上流配列とプロモーター配列で同じ意味? 129 回答 129 開質 Jul 19 at 21:52 nob_fj ♦♦ 400

プロモーター間連鎖転写因子の抽出 404 回答 404 開質 Jul 06 at 12:02 かどた 151

samtoolsの使い方 13k 回答 13k 開質 Jun 05 at 23:02 mn3 ♦♦ 493

DBCLS提供のEMBOSS Explorerにアクセスできないのはたまたま? 327 回答 327 開質 May 25 at 15:51 hono ♦♦ 81

Bowtieのpair-end mappingのオプションについて 401 回答 401 開質 Apr 24 at 16:41 nob_fj ♦♦ 400

ペアエンドのFASTQの命名の選択肢 300 回答 300 開質 Apr 18 at 00:16 nob_fj ♦♦ 400

Excel fileからBEDフォーマットのfileを作るには? 391 回答 391 開質 Apr 14 at 15:56 mn3 ♦♦ 493

最近更新された質問

119 質問 233 回答

よくある質問

ただいまベータ版テスト運用中です。さうぞお使いください。ただし、ソースの変更やサービスの停止変更されることがあります。質問するならライフサイエンスQA

このページのRSSフィード @lsqa_all · @lsqa_nrgs

80

©2012 統合データベース講習会 Licensed Under CC 表示 2.1