

ILASセミナー 2022年7月

RNAシーケンシング (RNA-seq) 解析

# 内容

---

1. RNA-seqについて
2. リードから発現解析まで
3. 発現解析結果を見してみる
  - 3a. 遺伝子配列データ
  - 3b. 発現解析データ
  - 3c. BLAST検索データ
  - 3d. GO termデータ

# 内容

---

## 1. RNA-seqについて

## 2. リードから発現解析まで

## 3. 発現解析結果を見てみる

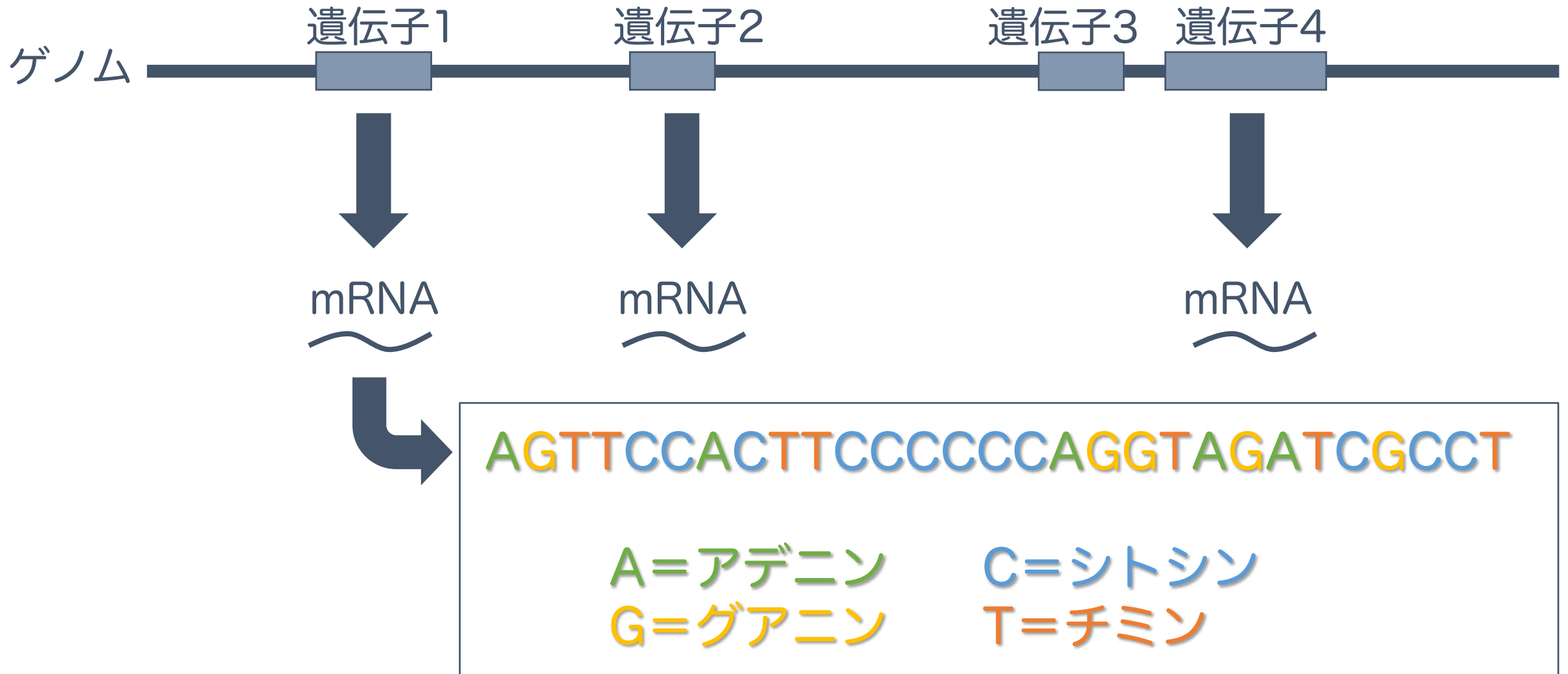
3a. 遺伝子配列データ

3b. 発現解析データ

3c. BLAST検索データ

3d. GO termデータ

# RNA-seqは発現遺伝子の配列を得る方法



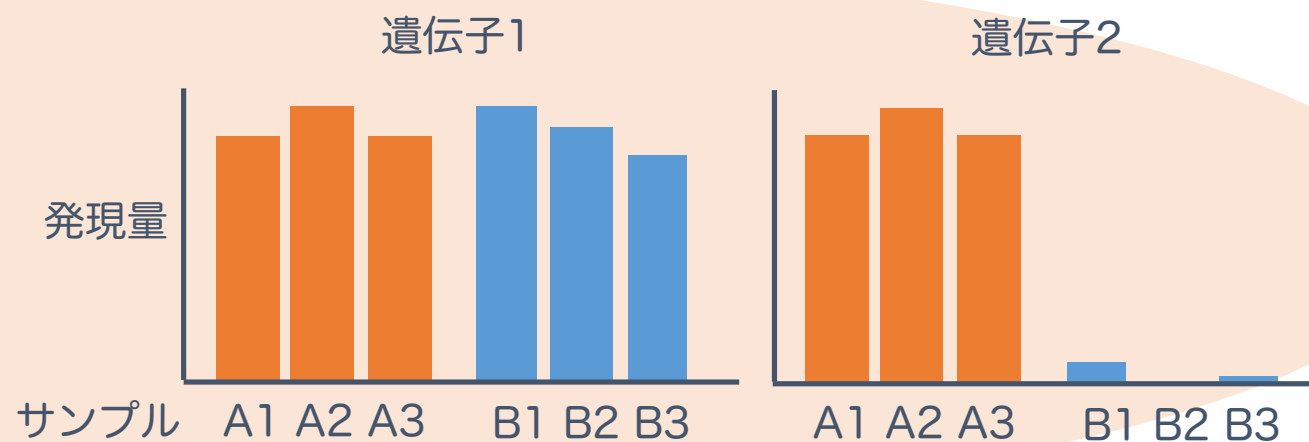
# RNA-seqの解析で出来ること

---

塩基配列の違いを見つける

サンプルA AGTTCCACTTCCC  
サンプルB AGTTCCACTT**G**CC

発現量の違いを調べる



# RNA-seqサンプル

---

ヤブガラシ

ねじれ部位 - ねじれなし部位

カイヅカイブキ

ネズ葉 - ヒノキ葉

水草

水上葉 - 水中葉

コムギ近縁野生種

芒あり穂 - 芒なし穂

どんな遺伝子に違いがあるのか？

形質の違いを決めている遺伝子は何か？

# RNA-seqまで手順

## サンプリング



## RNA抽出 & RNA-seq前処理



## シーケンシング



AGTTCCACTT  
CTAGCTGGC  
GGTACAAAT

Illumina NextSeq 550

<https://www.illumina.com/>

# リードデータ (FastQファイル)

@リードのID

リードの塩基配列 [A,T,G,C,N(未定義塩基)の文字列]

+

各塩基のクオリティ情報

```
@E00461:120:GW170612271:3:1101:18609:1520 1:N:0:NCTTCCAT
NGCTCGCGGGGCAGAAGACAAAAGCAAGGAGAAGATGTTGTACATCCTCTCAAGGTCTCCCTCGAGGAATTGTACA
+
#AAAFJJJJ<JJFJ<FAJJJJJJFAJJJJJJFAFFJFFJJJJJJFJJFJ<AFFJJJJFJJJ<<<JJ<JF7JJJJJJF
```

1個目のリードの情報

```
@E00461:120:GW170612271:3:1101:6603:1538 1:N:0:NCTTCCAT
NGATCATCTTCGGTGCAAGAGCGCTCGAGCAGTAAGCTATTACGGCACTCTTTCAAGGATCGCTGCTTCTAGGCGAA
+
#AAFAJJ<JJJJJJJJFJJFJJFJ<JJJ7F<J<JFJAJJFJJ-F<FJJJJ<JJJJJFFJJJ<AJJJFJJFFFFF<
```

2個目のリードの情報

```
@E00461:120:GW170612271:3:1101:8958:1538 1:N:0:NCTTCCAT
NTCTCGCTAATAATAAGCTGCAAGGAGAATTACATGGAAAACCTTATTGAAGGACAAAGGCATAATACCAAAAACAA
+
#-A<FJAFJJFJJJJFJJJJJJFJFAFFJJJJJJJJFJJJJ7JFAJJJJJJJJJJJJJJJJJA<AFJJJJJJJJJJJJJA
```

3個目のリードの情報

```
@E00461:120:GW170612271:3:1101:10067:1538 1:N:0:NCTTCCAT
```



# RNA-seqリード量

サンプル名	植物	形質	サンプリング地点	# Reads
A17	ヤブガラシ	曲がった茎	物集女	3,372,645
A18	ヤブガラシ	曲がった茎	物集女	7,848,589
A19	ヤブガラシ	曲がった茎	物集女	5,793,892
A20	ヤブガラシ	曲がっていない茎	物集女	8,449,159
A21	ヤブガラシ	曲がっていない茎	物集女	7,131,044
A22	ヤブガラシ	曲がっていない茎	物集女	15,569,039
B01	カイツカイブキ	ヒノキ葉	本部構内	4,468,317
B02	カイツカイブキ	ヒノキ葉	本部構内	4,508,321
B05	カイツカイブキ	ネズ葉	本部構内	7,643,174
B06	カイツカイブキ	ネズ葉	本部構内	6,747,773
B10	カイツカイブキ	ヒノキ葉	北部構内	9,481,746
B14	カイツカイブキ	ネズ葉	北部構内	4,614,752
C01	Ae. caudata	芒あり親系統	-	17,256,208
C02	Ae. caudata	芒あり親系統	-	22,269,475
C03	Ae. caudata	芒あり親系統	-	28,804,149
C04	Ae. caudata	芒なし親系統	-	18,462,632
C05	Ae. caudata	芒なし親系統	-	2,770,968
C06	Ae. caudata	芒なし親系統	-	26,130,602
C07	Ae. caudata	芒なしF2個体バルク	-	25,808,638
C08	Ae. caudata	芒なしF2個体バルク	-	10,782,017
C09	Ae. caudata	芒なしF2個体バルク	-	26,204,667
C10	Ae. caudata	芒ありF2個体バルク	-	24,381,158
C11	Ae. caudata	芒ありF2個体バルク	-	19,460,297
C12	Ae. caudata	芒ありF2個体バルク	-	34,376,536
5A1	水草No.5	水上葉	-	6,805,482
5A2	水草No.5	水上葉	-	7,809,672
5A3	水草No.5	水上葉	-	6,971,707
5W1	水草No.5	水中葉	-	3,660,951
5W2	水草No.5	水中葉	-	7,105,096
5W3	水草No.5	水中葉	-	9,984,782
6A1	水草No.6	水上葉	-	21,159,017
6A2	水草No.6	水上葉	-	19,701,326
6A3	水草No.6	水上葉	-	25,221,295
6W1	水草No.6	水中葉	-	5,404,468
6W2	水草No.6	水中葉	-	3,134,729
6W3	水草No.6	水中葉	-	7,175,190

配布プリント参照

# 内容

---

1. RNA-seqについて

2. リードから発現解析まで

3. 発現解析結果を見してみる

3a. 遺伝子配列データ

3b. 発現解析データ

3c. BLAST検索データ

3d. GO termデータ

# 解析の流れ

---

参照配列を準備



アライメント



リードをカウント  
発現量を比較

# 解析の流れ

---

参照配列を準備



アライメント



リードをカウント  
発現量を比較

リードをつなぎ合わせる

CTACCTAGCTCA  
CTACCTAGCTCA  
AGCTAGCTACCT  
CTCCCAGCTAGC  
CCGCCTCCCAGC  
CCTCCCGCCTCC  
TCCTCCTCCCGC



TCCTCCTCCCGCCTCCCAGCTAGCTACCTAGCTCA

遺伝子配列を再構築

# 解析の流れ

---

参照配列を準備



アライメント



リードをカウント  
発現量を比較

リードを並べる

CTACCTAGCTCA  
CTACCTAGCTCA  
AGCTAGCTACCT  
CTCCCAGCTAGC  
CCGCCTCCCAGC  
CCTCCCGCCTCC  
TCCTCCTCCCGC

TCCTCCTCCCGCCTCCCAGCTAGCTACCTAGCTCA

参照配列

# 解析の流れ

参照配列を準備



アライメント

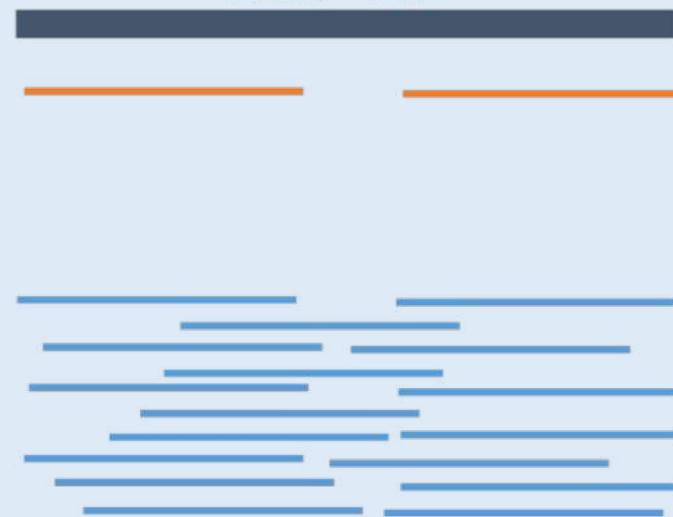


リードをカウント  
発現量を比較

Sample A

Sample B

遺伝子1



# 内容

---

1. RNA-seqについて

2. リードから発現解析まで

3. 発現解析結果を見してみる

3a. 発現解析結果

3b. 遺伝子配列

3c. BLAST検索結果

## 發現解析結果

Genes	WS	WC	WE	WB	PS	PI	PA	sample	length	logFC	logCPM	Pvalue	DR
TENNYT_SAME_c1_g1	809	865	5048	48	1367	130	158	362 white	Purple	-0.086852512542441	7.717330		
TENNYT_SAME_c1_g2	1065	1065	2527	1009	1456	2623	1006	100% white	Purple	0.004276751031743	9.993891		
TENNYT_SAME_c1_g3	129	129	129	129	129	129	129	129 white	Purple	-0.10489725425425425	7.717330		
TENNYT_SAME_c1_g4	100	233	263	22	139	184	26	96 white	Purple	0.00000000000000000	8.989000		
TENNYT_SAME_c1_g5	2042	2097	2361	1268	1851	2485	2391	96 white	Purple	-0.09161446103103103	1.507467		
TENNYT_SAME_c2_g1	32	62	21	8	63	78	8	46 white	Purple	-0.04876121343434343	4.276739		
TENNYT_SAME_c2_g2	3089	3089	7793	3021	5527	8661	3068	100% white	Purple	0.01212121212121212	9.993891		
TENNYT_SAME_c2_g3	2060	2502	2200	1466	1749	2757	2351	1188 white	Purple	-0.07386371212121212	7.717330		
TENNYT_SAME_c2_g4	0	38	0	0	0	0	0	17 white	Purple	-0.0268610343751	1		
TENNYT_SAME_c2_g5	176	105	185	30	103	230	68	82 white	Purple	-0.08191609160916091	8.978139		
TENNYT_SAME_c2_g6	176	105	185	30	103	230	68	82 white	Purple	-0.10162131016213101	9.008489		
TENNYT_SAME_c2_g7	184	169	184	70	221	190	130	126 white	Purple	-0.03989311343434343	8.989000		
TENNYT_SAME_c2_g8	19	16	0	17	59	0	17	0 white	Purple	0.07746021609160916	1.507467		
TENNYT_SAME_c2_g9	0	0	10	10	33	39	0	0 white	Purple	-0.05234714031031031	3.994923		
TENNYT_SAME_c2_g10	218	218	817	170	300	445	325	123 white	Purple	-0.03002016091609160	10.07795		
TENNYT_SAME_c2_g11	209	409	206	143	247	429	207	407 white	Purple	0.00000000000000000	8.989000		
TENNYT_SAME_c2_g12	0	0	0	0	0	0	27	0 white	Purple	1.07911214031031031	1.507467		
TENNYT_SAME_c2_g13	21	0	0	0	19	0	0	0 white	Purple	6.7691011123002103042	1.507467		
TENNYT_SAME_c2_g14	137	133	200	70	164	338	62	50 white	Purple	-0.02496216091609160	9.64131		
TENNYT_SAME_c2_g15	0	0	0	0	0	0	0	0 white	Purple	0.00000000000000000	1.47187842		
TENNYT_SAME_c3_g1	9020	10812	15462	8798	14564	13652	9051	7688 white	Purple	-0.01100000000000000	1.507467		
TENNYT_SAME_c3_g2	57	64	76	34	63	59	46	56 white	Purple	0.02686121343434343	1.507467		
TENNYT_SAME_c3_g3	0	5	22	15	8	12	26	0 white	Purple	-0.07254216091609160	8.989000		

# 発現量に差がある 遺伝子の塩基配列情報

[illegible]

Google Scholar

論文検索  
関連情報を集める

# BLAST検索

NIH National Library of Medicine  
National Center for Biotechnology Information

BLAST<sup>®</sup> » blastx

Translated BLAST: blastx

blastn blastp **blastx** tblastn tblastx

BLAST<sup>®</sup> search protein databases using a translated nucleotide query. [more...](#)

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#)

Query subrange [?](#)

From

To

Or, upload file [ファイルを選択](#) 選択されていません [?](#)

Genetic code

Job Title

[Enter a descriptive title for your BLAST search](#) [?](#)

☐ Align two or more sequences [?](#)

Choose Search Set

Database  [?](#)

Organism

☐ exclude [Add organism](#)

Blast is a trademark of the National Center for Biotechnology Information. All other trademarks are the property of their respective owners. All rights reserved. All other trademarks are the property of their respective owners.