

ARGA pipeline user guide

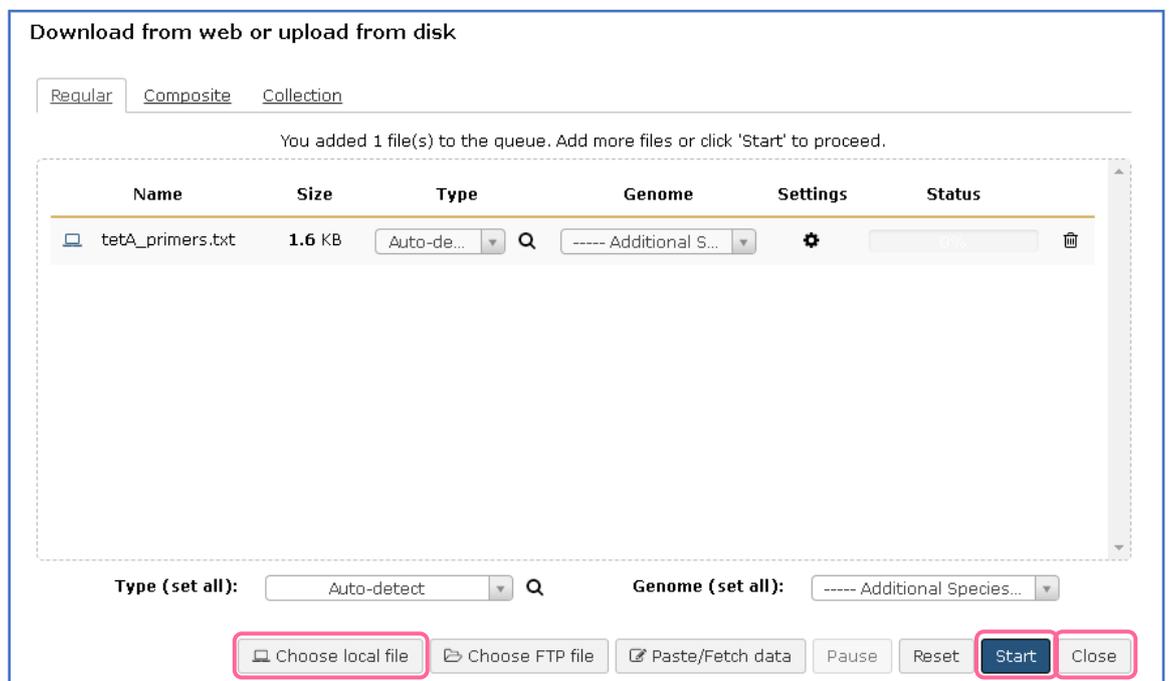
<http://mem.rcees.ac.cn:8083/>

1. Get Data

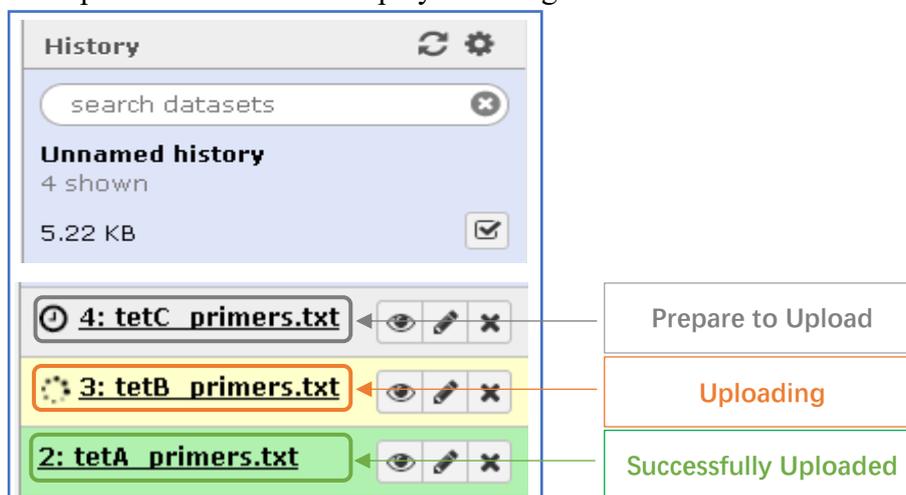
Get Data

[Upload File](#) from your computer

1.1 Click the [Upload File](#), then click the [Choose local file](#) at the pop-up window, select the object files to upload, Click the [Start](#), then click the [Close](#) to close the pop-up window.



1.2 The uploaded data would display in the right of the website.



1.3 Once the file is successfully uploaded, the dataset can be viewed, saved and deleted.

6: tetE primers.txt [View] [Edit] [Close]

18 lines
格式: tabular, 数据库: ?

uploaded tabular file

[Save] [Info]

1	2
tetE_1F	TTGGCGCTGTATGCAATGAT
tetE_1R	CGACGACCTATGCGATCTGA
tetE_2F	GGTCAGATCGCATAGGTCGT
tetE_2R	GCGTATGGAGGATGTGTTT
tetE_3F	ATGAACCGCACTGTGATGATG

View dataset content | Edit dataset attributes | Delete the dataset

Save the dataset | View dataset information

2. Primer Assessment

Primer Assessment

Primer Coverage BLASTn program to calculate the number of matched sequences in targeted database

Primer Specificity BLASTn program to calculate the number of matched sequences in targeted database

2.1 Prepare the tab-delimited txt file as requested, and upload it to the pipeline.

Input File Format (tab-delimited txt file):

The content of each column (3 columns, no headline in the txt file):

PrimerName PrimerSequence PrimerPairNumber(Arabic numerals, from 1 to n)

- tetA_1F CCGCGCTTTGGGTCATT 1
- tetA_1R CCGCGCTTTGGGTCATT 1
- tetA_2F CCGCGCTTTGGGTCATT 2
- tetA_2R CCGCGCTTTGGGTCATT 2
- tetA_6F CCGCGCTTTGGGTCATT 6
- tetA_6R CCGCGCTTTGGGTCATT 6
-

Notice:

Please do not use any symbols like ";", "(", ")", "#", "-" in your sample names. They will cause some potential errors. You can replace them with "_". "_" is not permitted here.

Example File:

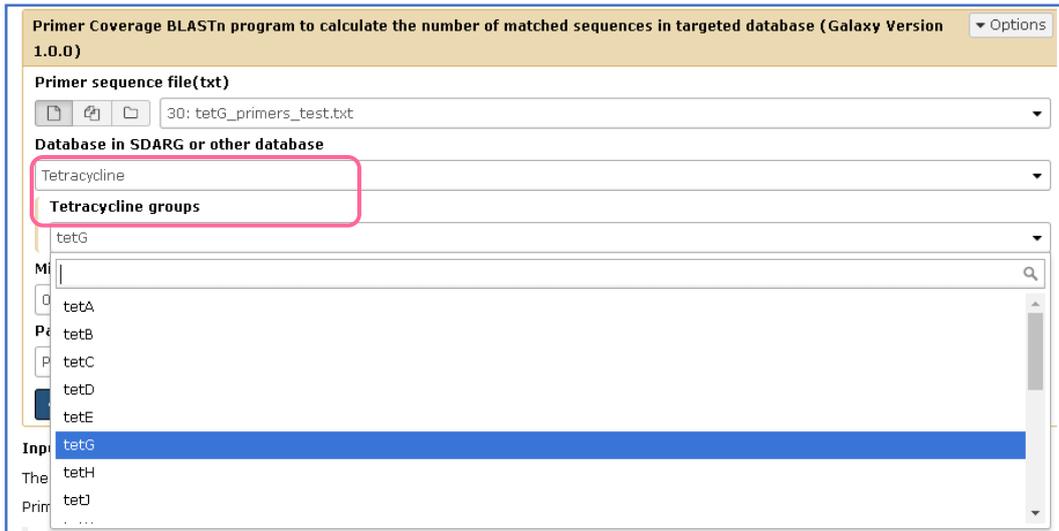
tetG_1F	CATCAGCGCCGGTCTTATG	1
tetG_1R	CCCCATGTAGCCGAACCA	1
tetG_2F	GCAGAGCAGGTCGCTGG	2
tetG_2R	CCYGCAAGAGAAGCCAGAAG	2

Degenerate Primer

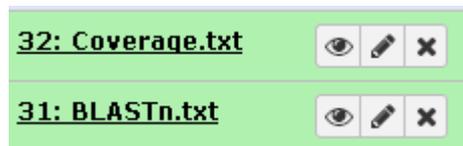
2.2 Select the Primer sequence file to do primer coverage calculation, and then choose the corresponding database (both database already integrated in SDARG or own database is acceptable) to do primer coverage calculation. When calculating the primer coverage, 0 or 1 Mismatch is allowed, the Paired Calculating is optional, paired is proper for primer pairs, this option would output both the coverage of single primer and coverage of primer pairs, while the unpaired only output the coverage of each primer. Click Execute.

Example of Database in SDARG or other database:

If the primers are targeted tetG gene (Tetracycline resistance gene), we can firstly select the AR group - Tetracycline from the first option, then select the ARG type – tetG from the second option, chose 0 mismatch and paired calculation, execute.



There are two output file, one is the BLASTn result, and the other is the coverage calculation result in txt file.



```

# BLASTN 2.2.31+
# Query: tetG_2R_1_Length_20_Degenerate_1_Position_1_FR_2
# Database: mergedatabase
# Fields: query id, subject id, % identity, alignment length, mismatches, gap opens, q. start, q. end, s. start, s. end, evaluate, bit score
# 18 hits found
tetG_2R_1_Length_20_Degenerate_1_Position_1_FR_2   ARID91942 |tetg|tetracycline 100.00 20 0 0 1 20 239 220 2e-06
37.4
tetG_2R_1_Length_20_Degenerate_1_Position_1_FR_2   ARID91942 |tetg|tetracycline 100.00 20 0 0 1 20 215 196 2e-06
37.4
tetG_2R_1_Length_20_Degenerate_1_Position_1_FR_2   ARID91942 |tetg|tetracycline 100.00 20 0 0 1 20 191 172 2e-06
37.4
tetG_2R_1_Length_20_Degenerate_1_Position_1_FR_2   ARID91943 |tetg|tetracycline 95.00 20 1 0 1 20 239 220 1e-04
31.9
tetG_2R_1_Length_20_Degenerate_1_Position_1_FR_2   ARID91943 |tetg|Tetracycline 95.00 20 1 0 1 20 239 220 1e-04
31.9
tetG_2R_1_Length_20_Degenerate_1_Position_1_FR_2   ARID91941 |tetg|Tetracycline 95.00 20 1 0 1 20 239 220 1e-04
31.9
tetG_2R_1_Length_20_Degenerate_1_Position_1_FR_2   ARID91940 |tetg|tetracycline 95.00 20 1 0 1 20 239 220 1e-04
31.9
tetG_2R_1_Length_20_Degenerate_1_Position_1_FR_2   ARID663988 |tetg|tetracycline 95.00 20 1 0 1 20 239 220 1e-04
31.9
tetG_2R_1_Length_20_Degenerate_1_Position_1_FR_2   ARID91940 |tetg|tetG 95.00 20 1 0 1 20 239 220 1e-04 31.9
tetG_2R_1_Length_20_Degenerate_1_Position_1_FR_2   ARID91940 |tetg|tetracycline 95.00 20 1 0 1 20 215 196 1e-04
31.9
tetG_2R_1_Length_20_Degenerate_1_Position_1_FR_2   ARID91944 |tetg|tetracycline 95.00 20 1 0 1 20 191 172 1e-04
31.9
tetG_2R_1_Length_20_Degenerate_1_Position_1_FR_2   ARID91944 |tetg|tetracycline 95.00 20 1 0 1 20 191 172 1e-04
31.9
tetG_2R_1_Length_20_Degenerate_1_Position_1_FR_2   ARID91944 |tetg|tetracycline 95.00 20 1 0 1 20 191 172 1e-04
31.9
tetG_2R_1_Length_20_Degenerate_1_Position_1_FR_2   ARID91943 |tetg|tetracycline 95.00 20 1 0 1 20 191 172 1e-04
31.9
tetG_2R_1_Length_20_Degenerate_1_Position_1_FR_2   ARID91941 |tetg|tetracycline 95.00 20 1 0 1 20 191 172 1e-04
31.9
tetG_2R_1_Length_20_Degenerate_1_Position_1_FR_2   ARID91941 |tetg|tetracycline 95.00 20 1 0 1 20 191 172 1e-04
31.9
tetG_2R_1_Length_20_Degenerate_1_Position_1_FR_2   ARID91941 |tetg|tetracycline 95.00 20 1 0 1 20 191 172 1e-04
31.9
tetG_2R_1_Length_20_Degenerate_1_Position_1_FR_2   ARID91940 |tetg|tetracycline 95.00 20 1 0 1 20 191 172 1e-04
31.9
# BLASTN 2.2.31+
# Query: tetG_2R_2_Length_20_Degenerate_1_Position_1_FR_2
# Database: mergedatabase
# Fields: query id, subject id, % identity, alignment length, mismatches, gap opens, q. start, q. end, s. start, s. end, evaluate, bit score
# 18 hits found
tetG_2R_2_Length_20_Degenerate_1_Position_1_FR_2   ARID91943 |tetg|tetracycline 100.00 20 0 0 1 20 239 220 2e-06
37.4
tetG_2R_2_Length_20_Degenerate_1_Position_1_FR_2   ARID91943 |tetg|tetracycline 100.00 20 0 0 1 20 239 220 2e-06

```

The **31: BLASTn.txt** file contains the detailed information of the sequences matched to each primer sequence. In the example file, tetG_2R (CCYGAAGAGAAGCCAGAAG) is a degenerate primer, correspond to two traditional

primers (CCCGCAAGAGAAGCCAGAAG and CCTGCAAGAGAAGCCAGAAG), hence in the BLASTn result, there are two tetG_2R primers displayed.

Primer	Coverage
tetG_1R	0.23
tetG_1R	0.19
tetG_2F	0.16
tetG_2R	0.17

Submitted Primer Sequence Result – single primer

PRIMER Pair	Coverage
1	0.19
2	0.16

Submitted Primer Sequence Result – primer pairs

Recommended Primer Pairs

Primer Database	Primer Names	Forward Primer	Coverage of Forward Primer (%)	Specificity of Forward Primer (%)	Reverse Primer	Coverage of Reverse Primer (%)	Specificity of Reverse Primer (%)	Coverage of Primer Pairs (%)	Specificity of Primer Pairs (%)
tetG Primer database	tetG_1	GCTCGGTGGTATCTCTGCTC	70.00	100.00	AGCAACAGAATCGGGAACAC	67.00	100.00	67.00	100.00
	tetG_2	TCARCCATTCGGATTCGA	23.00	100.00	TGGCCCGCARTCAATG	22.00	91.67	21.00	100.00
	tetG_3	CATCAGCCCGGCTTATG	23.00	100.00	CCCCATGTAGCCGAACCA	19.00	100.00	19.00	100.00
	tetG_4	GCAGAGCAGGTCGCTGG	16.00	100.00	CCYCAAGAGAAGCCAGAAG	17.00	100.00	16.00	100.00
	tetG_5	TTTCGGATTCTTACGGTC	15.00	100.00	TCCTGCCATAGAGCTTAGA	15.00	100.00	15.00	100.00
	tetG_6	CAGCTTTCGGATTCACGG	15.00	100.00	GATGGTGAGGCTCGTTAGC	15.00	100.00	15.00	100.00
	tetG_7	GCTCGGTGGTATCTCTGCTC	70.00	100.00	CAAAGCCCTTGCTTGTAC	15.00	100.00	15.00	100.00
	tetG_8	GGTGCTTCGGCTTCTCTTG	3.00	100.00	CAATGGTTGAGGCAGTACA	7.00	100.00	1.00	100.00
	tetG_9	CATTCGCTCTGATCG	7.00	77.78	TTGGTGAGGCTTGAAGC	2.00	100.00	1.00	100.00

Integrated Primer Quality Report

The 32: Coverage.txt file displays both the coverage of the input primer pairs of *tetG* gene, and the coverage of the previous collected *tetG* gene primers, in this case, the users can choose the primers with higher coverage.

Example of Database in SDARG or other database:

If the primers are targeted *tetG* gene (Tetracycline resistance gene), we can either select database as above, or we can firstly upload own database

33: tetG_nucl database.fasta , and select Others (own database) option, then select the uploaded database as the targeted database in the primer coverage calculation, execute.

Primer Coverage BLASTn program to calculate the number of matched sequences in targeted database (Galaxy Version 1.0.0) Options

Primer sequence file(txt)

Database in SDARG or other database

- Tetracenomycin
- Tetracycline
- Thimethoprim
- Thiopeptide
- Tunicamycin
- Vancomycin
- Others (own database)

Primer Coverage BLASTn program to calculate the number of matched sequences in targeted database (Galaxy Version 1.0.0) Options

Primer sequence file(txt)

Database in SDARG or other database

Database(FASTA) If the primers target ARGs, there is no need to select your own database

Mismatch

Paired Calculating

There are also two output files as above, one is the BLASTn result, and the other is the coverage calculation result in txt file.

35: Coverage.txt

34: BLASTn.txt

```
Primer Coverage
tetG_1F 0.23
tetG_1R 0.19
tetG_2F 0.16
tetG_2R 0.17

PRIMER Pair      Coverage
1      0.19
2      0.16
```

The difference is that in the **35: Coverage.txt** file, only the coverage of the submitted primers are displayed in the result.

2.3 The primer specificity calculation is similar to the primer coverage calculation.

3. Sequence Annotation

Sequence Annotation

Nucleotide Sequence Annotation -- Blastn program to query nucleotide sequences in SDARG

Protein Sequence Annotation -- HMMsearch program to search protein sequences against HMM models of SDARG

3.1 Prepare the FASTA format file as requested, and upload it to the pipeline.

Example File:

```
>ARID822840|dihydropteroate synthase [Escherichia coli]
ATCCCATCGTGGAGTCGCTGCGGAGGCCGACGCCATGGTGACGGTGTTCGGCATTCTGAAATCTCACGGAGGACTCCTTCTTGGATGAGAGCCGGCGGCTAGACCCCGCCGGCGCTGTCACCGGGGG
ATCCAAATGCTGCGAGTCGGATCAGACGTCGTGGATGTGGACCGGCCGCCATCCGGACGGGAGGCGCTGTATGCGCGCGATGAGATCAGACGTAATGCGCGCGCTTAGACGCCCTGTCCGAT
CAGATGCACCGTGTTCGAATGACAGCTTCCAACCGGAAACCCAGCGCTATGGCTCAAGCGCGGGGTGGGCTACCTGAACGATATCCAAGGATTTCCTGACCCGTGCGCTATCCCGATATTGCTGAG
GGGACTGCAGCTGTGGTTATGCACCTCAGCGCAGCGGGATGGCATGCCACCGCACCGGTCACCTTCGACCCGAAGACGGGCTCGACGAGATTGTGCGGTTCTTCGAGGGCGGGTTTCGGCTTG
CGACGGAGCGGGTGCCTGCGACCGGCTCATCTCGATCCGGGATGGGATTTTCTTGAGCCCCGACCGGAAACATCGCTGCACGTGCTGTCGAACTTCAAAGCTGAAGTCGGCGTTGGGGCTT
CGCTATTGGTCTCGGTGTCGCGGAAATCCTTCTTGGGGGCCACCGTGGCTTCTGTAAGGATCTGGGTCCAGCGAGCCTTGGCGGGAACCTCACGGCATCGCAATGGCGCTGACTACGTCGCG
ACCCACGCGCTGGAGATCTGCGAAGCGCAATCACTTCTCGGAAACCTCGGAAATTTTCGAGTCGCGACGCGAGAGACCGAGGGTTAGATCATGCCTAG

>ARID822880|dihydropteroate synthase [Enterobacter cloacae]
ATCCCATCGTGGAGTCGCTGCGGAGGCCGACGCCATGGTGACGGTGTTCGGCATTCTGAAATCTCACGGAGGACTCCTTCTTGGATGAGAGCCGGCGGCTAGACCCCGCCGGCGCTGTCACCGGGGG
ATCGAAATGCTGCGAGTCGGATCAGACGTCGTGGATGTGGACCGGCCGCCATCCGGACGGGAGGCGCTGTATGCGCGCGATGAGATCAGACGTAATGCGCGCGCTTAGACGCCCTGTCCGAT
CAGATGCACCGTGTTCGAATGACAGCTTCCAACCGGAAACCCAGCGCTATGGCTCAAGCGCGGGGTGGGCTACCTGAACGATATCCAAGGATTTCCTGACCCGTGCGCTATCCCGATATTGCTGAG
GGGACTGCAGCTGTGGTTATGCACCTCAGCGCAGCGGGATGGCATGCCACCGCACCGGTCACCTTCGACCCGAAGACGGGCTCGACGAGATTGTGCGGTTCTTCGAGGGCGGGTTTCGGCTTG
CGACGGAGCGGGTGCCTGCGACCGGCTCATCTCGATCCGGGATGGGATTTTCTTGAGCCCCGACCGGAAACATCGCTGCACGTGCTGTCGAACTTCAAAGCTGAAGTCGGCGTTGGGGCTT
CGCTATTGGTCTCGGTGTCGCGGAAATCCTTCTTGGGGGCCACCGTGGCTTCTGTAAGGATCTGGGTCCAGCGAGCCTTGGCGGGAACCTCACGGCATCGCAATGGCGCTGACTACGTCGCG
ACCCACGCGCTGGAGATCTGCGAAGCGCAATCACTTCTCGGAAACCTCGGAAATTTTCGAGTCGCGACGCGAGAGACCGAGGGTTAGATCATGCCTAG
```

3.2 Select the nucleotide sequence file to do BLASTn with the integrated SDARG nucleotide sequence database, chose the proper E-value (better $\leq 10^{-2}$), and Execute.

Nucleotide Sequence Annotation -- Blastn program to query nucleotide sequences in SDARG (Galaxy Version 1.0.0) Options

nucleotide sequence file(FASTA)

245: nud_test.fasta

E-value

0.0001

Execute

The result is as following, the test example sequence should be annotated as *sull* gene:

Query= ARID822840|dihydropteroate synthase [Escherichia coli]

Length=876

							Score	E												
							(Bits)	Value												
Sequences producing significant alignments:																				
gi		99867132		gb		ABF67777.1		DQ517526		sul1		High		3.00E-164		SulI (...		1581		0.0
gi		90265407		emb		CAJ77089.1		CT025832		sul1		High		3.00E-164		sul1d...		1581		0.0
gi		90265368		emb		CAJ77050.1		CT025832		sul1		High		3.00E-164		sul1d...		1581		0.0
gi		85721150		gb		ABC76093.1		CP000252		sul1		High		3.00E-164		dihydr...		1581		0.0
gi		820773496		gb		AKG90356.1		KP742988		sul1		High		3.00E-164		dihyd...		1581		0.0
gi		766547723		gb		AJS16031.1		KP116299		sul1		High		3.00E-164		dihyd...		1581		0.0
gi		765526620		gb		AJS10216.1		KP453775		sul1		High		3.00E-164		dihyd...		1581		0.0
gi		762229514		dbj		BAQ55583.1		LC030435		sul1		High_gene		3.00E-164...				1581		0.0
gi		700587998		gb		AIU94568.1		KM649682		sul1		High		3.00E-164		dihyd...		1581		0.0
gi		693581793		dbj		BAP68573.1		AP014650		sul1		High		3.00E-164		dihy...		1581		0.0
gi		693581787		dbj		BAP68567.1		AP014650		sul1		High		3.00E-164		dihy...		1581		0.0
gi		690385531		gb		AIR72316.1		KM023773		sul1		High		3.00E-164		dihyd...		1581		0.0
gi		641398833		gb		AIA58914.1		KJ510410		sul1		Middle		3.00E-164		sul...		1581		0.0
gi		633259752		dbj		BAO79444.1		AB926431		sul1		High		3.00E-164		sulf...		1581		0.0
gi		593804565		gb		AHL68236.1		KF017283		sul1		High		3.00E-164		dihyd...		1581		0.0
gi		575870568		emb		CDF66134.1		HG003695		sul1		High_gene		3.00E-164...				1581		0.0
gi		572099405		dbj		BAO27817.1		AB898038		sul1		High		3.00E-164		sulf...		1581		0.0
gi		546201898		dbj		BAN87141.1		AP013357		sul1		High_gene		3.00E-164...				1581		0.0
gi		537705814		dbj		BAN81907.1		AB846838		sul1		High		3.00E-164		dihy...		1581		0.0
gi		528172304		gb		AGS27666.1		CP006631		sul1		High		3.00E-164		Dihyd...		1581		0.0
gi		514883720		gb		AG062374.1		JX442974		sul1		High_gene		3.00E-164 ...				1581		0.0
gi		452916843		gb		AGG19217.1		KC189475		sul1		High_gene		3.00E-164 ...				1581		0.0
gi		444506920		gb		AGE11239.1		JQ996150		sul1		High		3.00E-164		dihyd...		1581		0.0
gi		443417939		emb		CCO61950.1		HF546976		sul1		High_gene		3.00E-164...				1581		0.0
gi		429345522		gb		AFZ84491.1		JX988621		sul1		High_gene		3.00E-164 ...				1581		0.0
gi		410066879		gb		AFV58080.1		JX517203		sul1		High		3.00E-164		dihyd...		1581		0.0
gi		407804968		gb		EKF76221.1		JH930402		sul1		High_gene		3.00E-164 ...				1581		0.0
gi		394349075		gb		AFN35097.1		CP003684		sul1		High		3.00E-164		Dihyd...		1581		0.0
gi		394348984		gb		AFN35006.1		CP003684		sul1		High		3.00E-164		Dihyd...		1581		0.0
gi		385721348		gb		AFI72853.1		JQ364967		sul1		High_gene		3.00E-164 ...				1581		0.0
gi		380714704		gb		AFE02726.1		JQ639792		sul1		High_gene		3.00E-164 ...				1581		0.0
gi		377806718		gb		AFB76403.1		JN676148		sul1		High		3.00E-164		Dihyd...		1581		0.0
gi		374413623		gb		AEZ49746.1		JN687470		sul1		High_gene		3.00E-164 ...				1581		0.0
gi		374413610		gb		AEZ49733.1		JN687470		sul1		High_gene		3.00E-164 ...				1581		0.0
gi		374413594		gb		AEZ49717.1		JN687470		sul1		High_gene		3.00E-164 ...				1581		0.0
gi		365819152		gb		AEX00942.1		JF274992		sul1		High_gene		3.00E-164 ...				1581		0.0

3.3 Select the protein sequence file to do HMMsearch with the integrated SDARG

Hidden Markov Models database, and Execute.

Protein Sequence Annotation -- HMMsearch program to search protein sequences against HMM models of SDARG (Galaxy) Options

Version 1.0.0)

Protein Sequence file(FASTA)

Example File:

```
>ARID823383|sul1|dihydropteroate synthase, partial (plasmid) [Proteus mirabilis]
ARVSLRKRSGVAADRLILDPCMGFFLSPAPETSLHVLNQLKLSALGLPLLVSYSRKSFLGATVGLPVKDLGPASLAAELHAIGNGADYVRTHAPCDLRSATITSETLAKFRSRDARDRGLDHA
>ARID823346|sul1|dihydropteroate synthase type 1, partial [Salmonella enterica subsp. enterica serovar Pullorum]
CRLVVIHSAQRDGLATRTGHLRPNALDEIVRFFKARVSALQRSGVAADRLILNFCMGFFLSPAPETSLHVLNQLKLSAWGLPLLVSYSRKSFLGATVGLPVKDLGPASLAAELHAIGNGADYFRK
>ARID823108|sul1|SulI, partial [Pseudomonas aeruginosa]
CRLVVMHSAQRDGLATRTGHLRPNALDEIVRFFPEARVSALRSGVAADRLILDPCMGFFLSPAPETSLHVLNQLKLSALGLPLLVSYSRKSFLGATVGLPVKDLGPASLAAELHAIGNGADYVRTHAPG
DLRSATITSETLAKFRSRDARDRGLDHA
```

Output result:

```
Query:      aac2i  [M=207]
Scores for complete sequences (score includes all domains):
--- full sequence ---  --- best 1 domain ---  -#dom-
  E-value  score  bias    E-value  score  bias    exp  N  Sequence Description
-----
[No hits detected that satisfy reporting thresholds]

Domain annotation for each sequence (and alignments):
  [No targets detected that satisfy reporting thresholds]

Internal pipeline statistics summary:
-----
Query model(s):                1  (207 nodes)
Target sequences:              3  (413 residues searched)
Passed MSV filter:             0  (0); expected 0.1 (0.02)
Passed bias filter:            0  (0); expected 0.1 (0.02)
Passed Vit filter:             0  (0); expected 0.0 (0.001)
Passed Fwd filter:             0  (0); expected 0.0 (1e-05)
Initial search space (Z):      3  [actual number of targets]
Domain search space (domZ):    0  [number of targets reported over threshold]
# CPU time: 0.01u 0.01s 00:00:00.02 Elapsed: 00:00:00.09
# Mc/sec: 0.95
//
Query:      aac2ia [M=178]
Scores for complete sequences (score includes all domains):
--- full sequence ---  --- best 1 domain ---  -#dom-
  E-value  score  bias    E-value  score  bias    exp  N  Sequence Description
-----
[No hits detected that satisfy reporting thresholds]

Domain annotation for each sequence (and alignments):
  [No targets detected that satisfy reporting thresholds]
```

```

Query:      sul1 [M=283]
Scores for complete sequences (score includes all domains).
----- full sequence ----- best 1 domain ----- #dom-
E-value score bias E-value score bias exp N Sequence Description
1.1e-82 265.4 0.0 1.2e-82 265.2 0.0 1.0 1 ARID823108|sul1|Sul1, partial [Pseudomonas aeruginosa]
3e-64 204.8 0.0 3.4e-64 204.6 0.0 1.0 1 ARID823346|sul1|dihydropteroate synthase type 1, partial [Salmon
1.8e-57 182.6 0.0 1.9e-57 182.5 0.0 1.0 1 ARID823383|sul1|dihydropteroate synthase, partial (plasmid) [Pro

Domain annotation for each sequence (and alignments):
>> ARID823108|sul1|Sul1, partial [Pseudomonas aeruginosa]
# score bias c-Evalue i-Evalue hmmfrom hmm to alifrom ali to envfrom env to acc
1 ! 265.2 0.0 1.2e-82 1.2e-82 124 283 .] 1 160 [] 1 160 [] 1.00

Alignments for each domain:
= domain 1 score: 265.2 bits; conditional E-value: 1.2e-82
sul1 124 crlvvmrsaqrdGi atrtGhLRpedaldEivrffearvsalrrsgvaadrililpGmgfflsdapetslhvlsnlqklsalglrllv 211
crlvvn+saqrdGi atrtGhLRpedaldEivrffearvsalrrsgvaadrililpGmgffls+apetslhvlsnlqklsalgl+llv
ARID823108|sul1|Sul1, 1 CRLVVMHSAQRDGIATRIGHLRPEALDEIVRFFEARVSAALRRSGVAADRILILDPGMGFLLSPAPETSLHVLNQLKLSALGLPLLV 88
9*****PP

sul1 212 svsrksflgatvlpvkdGGeaslaaelhai gngadyvrvhakgdlrsaitfseilakfrsrdardrvldea 283
svsrksflgatvlpvkdG+aslaaelhai gngadyvrtha+gdlrsaitfse+lakfrsrdardrvldea
ARID823108|sul1|Sul1, 89 SVSRKSLFGATVGLPVKDLGPASLAELHAI GNGADYVRVTHAPGDLESAITFSETLAKFRSARDRGLDHA 160
*****997 PP

>> ARID823346|sul1|dihydropteroate synthase type 1, partial [Salmonella enterica subsp. enterica serovar Fullorum]
# score bias c-Evalue i-Evalue hmmfrom hmm to alifrom ali to envfrom env to acc
1 ! 204.6 0.0 3.4e-64 3.4e-64 124 250 .. 1 127 [. 1 128 [] 0.99

Alignments for each domain:
= domain 1 score: 204.6 bits; conditional E-value: 3.4e-64
sul1 124 crlvvmrsaqrdGi atrtGhLRpedaldEivrffearvsalrrsgvaadrililpGmgfflsdapetslhvlsnlqkl 201
crlvvn+saqrdGi atrtGhLRpe+aldEivrff+arvsal+rsgvaadrilil+tpGmgffls+apetslhvlsnlqkl
ARID823346|sul1|dihydropteroate 1 CRLVVMHSAQRDGIATRIGHLRPEALDEIVRFFKARVSALRRSGVAADRILILNPGMGFFLSPAPETSLHVLNQLK 78
9*****PP

sul1 202 ksalglrllvsvrksflgatvlpvkdGGeaslaaelhai gngadyvr 250
ksa gl+llvsvrksflgatvlpvkdG+aslaaelhai gngadyvr
ARID823346|sul1|dihydropteroate 79 KSAWGLPFLVSVSRKSLFGATVGLPVKDLGPASLAELHAI GNGADYVR 127
*****98 PP

>> ARID823383|sul1|dihydropteroate synthase, partial (plasmid) [Proteus mirabilis]

```

The output file of protein sequence annotation displays the result of the submitted sequence to each ARG type in SDARG, and part of the results are as above, we can annotate this test example sequence as *sul1* gene related protein sequences.

Example Dataset:

If there is any problem with the format of uploading file, please see example dataset in the [2019 Example Dataset](#) folder from the Shared Data in pipeline:

<http://mem.rcees.ac.cn:8083/library/list#folders/F40876639881ca029>.