RNA-Seq

de novo

ITMO 2016

TP

Xi LIU
Erwan CORRE
Strategy for transcriptome analyses and application on Galaxy

1. RNA-seq reads
   - History importation
2. Quality control
   - FastQC
3. Read cleanup
   - Trimmomatic, SortMeRNA
4. Quality control
   - FastQC ok
5. De novo assembly
   - Trinity
6. Abundance estimation
   - Trinity Statistics
7. Transcriptome filtering
   - Build expression matrix
   - Filter low expression transcripts
8. Abundance estimation
   - Align reads and estimated abundance
9. Protein coding regions identification
   - TransDecoder
10. Functional annotation
    - NCBI BLAST+blastx / Diamond
    - NCBI BLAST+blastp / Diamond
    - hmmscan
    - SignalP
    - TMHMM
    - Generate gene to transcript map
    - Trinotate
11. Differential expression analysis
    - Align reads and estimated abundance

Pn° : Practice ppt page
Tn° : Theories ppt page
Galaxy tools

2. Shared Data

3. Histories

4. RNA-seq de novo TP 2016

5. Import history

2 conditions (A & B) x 2 replicates
Quality Control for raw reads

1. Fastq Manipulation
2. FastQC Read Quality reports
3. Outputs for A1_left.fq
4. Execute: FastQC (0.67)
5. Outputs for A1_left.fq

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

The main functions of FastQC are:
- Import of data from BAM, SAM or FastQ/FastQ.gz files (any variant),
- Providing a quick overview to tell you in which areas there may be problems
- Summary graphs and tables to quickly assess your data
- Export of results to an HTML based permanent report
- Offline operation to allow automated generation of reports without running the interactive application

FastQC
This is a Galaxy wrapper. It merely exposes the external package FastQC which is documented at FastQC. Kindly acknowledge it as well as this tool if you use it. FastQC incorporates the Picard-tools libraries for sam/bam processing.
Reads cleanup

1. Trimmomatic flexible read trimming tool for Illumina NGS data
2. Yes
3. All « left »
4. Yes
5. All illumina

Remove adapter(s)
Reads cleanup

Depend on the results, check fastqc

7. 5’ quality trimming

8. 3’ quality trimming

9. average quality trimming by segment/window
Reads cleanup

9. Average quality filtering

10. Trimmed reads length filtering

11.

What it does

Trimmomatic performs a variety of useful trimming tasks for Illumina paired-end and single ended data.

This tool allows the following trimming steps to be performed:

- **ILLUMINACLIP**: Cut adapter and other Illumina-specific sequences from the read
- **SLIDINGWINDOW**: Perform a sliding window trimming, cutting once the average quality within the window falls below a threshold
- **MINLEN**: Drop the read if it is below a specified length
- **LEADING**: Cut bases off the start of a read, if below a threshold quality
- **TRAILING**: Cut bases off the end of a read, if below a threshold quality
- **CROP**: Cut the read to a specified length
- **HEADCROP**: Cut the specified number of bases from the start of the read
- **AVGQUAL**: Drop the read if the average quality is below a specified value
- **MAXINFO**: Trim reads adaptively, balancing read length and error rate to maximise the value of each read

Outputs

For reads A1

If ILLUMINACLIP is requested then it is always performed first; subsequent options can be mixed and matched and will be
Reads cleanup

Analysis Data

<table>
<thead>
<tr>
<th>Tool: Trimmomatic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number: 25</td>
</tr>
<tr>
<td>Name: Trimmomatic on A1_left.fq (R1 paired)</td>
</tr>
<tr>
<td>Created: Mon 21 Nov 2016 09:23:21 AM (UTC)</td>
</tr>
<tr>
<td>Filesize: 1.5 MB</td>
</tr>
<tr>
<td>Dbkey:</td>
</tr>
<tr>
<td>Format: fastqsanger</td>
</tr>
<tr>
<td>Galaxy Tool ID: tools.g2.bx.psu.edu/repos/pjbriggs/trimmomatic/trimmomatic/0.36.0</td>
</tr>
<tr>
<td>Galaxy Tool Version: 0.36.0</td>
</tr>
<tr>
<td>Tool Version:</td>
</tr>
<tr>
<td>Tool Standard Output: stdout</td>
</tr>
<tr>
<td>Tool Standard Error: stderr</td>
</tr>
<tr>
<td>Tool Exit Code: 0</td>
</tr>
<tr>
<td>History Content API ID: a46d5f94d2d85a8</td>
</tr>
<tr>
<td>Job API ID: ae3968a4e9e28dc0</td>
</tr>
<tr>
<td>History API ID: b2df5ddbbe728974</td>
</tr>
<tr>
<td>UUID: 8cd988cf-70b6-4b99-9cc9-515273d06bd</td>
</tr>
<tr>
<td>Full Path: /w/galaxy/galaxy3/database/files/008/dataset_8646.dat</td>
</tr>
</tbody>
</table>

Input Parameter |
| Value |
| Note for rerun |
| Paired end data? yes |
| Input Type pair_of_files |
| Input FASTQ file (R1/first of pair) 1: A1_left.fq |
| Input FASTQ file (R2/second of pair) 2: A1_right.fq |
| Perform initial ILLUMINA CLIP step? yes |
| Adapter sequences to use All illumina |
| Maximum mismatch count which will still allow a full match to be performed 2 |
| How accurate the match between the two ‘adapter ligated’ reads must be for PE palindrome read alignment 30 |
| How accurate the match between any adapter etc. sequence must be against a read 10 |
| Select Trimmomatic operation to perform SLIDINGWINDOW |
| Number of bases to average across 4 |
| Averate quality required 20 |
| Select Trimmomatic operation to perform MINLEN |
| Minimum length of reads to be kept 20 |
| Select Trimmomatic operation to perform LEADING |
| Minimum quality required to keep a base 3 |
| Select Trimmomatic operation to perform TRIMMING |

History

| Tool Version: |
| Value |
| Note for rerun |
| 27: Trimmomatic on A1_left.fq (R1 unpaired) |
| 26: Trimmomatic on A1_right.fq (R2 paired) |
| 25: Trimmomatic on A1_left.fq (R1 paired) |
| 24: FastQC on data |

Arguments:

* -mx8G
* -jar
* /w/galaxy/galaxy3/shed_tools_dependencies/trimmomatic/0.36/pjbriggs/trimmomatic/14d05f2d511d/trimmomatic-0.36.jar
* PE
* -threads
* 6
* -phred33
* /w/galaxy/galaxy3/database/files/008/dataset_8593.dat
* /w/galaxy/galaxy3/database/files/8601/
4. Reads cleanup

Input Read Pairs: 10000
Both Surviving: 9305 (93.05%)
Forward Only Surviving: 118 (1.18%)
Reverse Only Surviving: 273 (2.73%)
Dropped: 304 (3.04%)
Quality Control for cleaned reads

1. FastQC
   Read Quality reports (Galaxy Version 0.67)
   Short read data from your current history
   - 40: Trimmomatic on B2_right.fq (R2 unpaired)
   - 39: Trimmomatic on B2_left.fq (R1 unpaired)
   - 38: Trimmomatic on B2_right.fq (R2 paired)
   - 37: Trimmomatic on B2_left.fq (R1 paired)
   - 36: Trimmomatic on B1_right.fq (R2 unpaired)
   - 35: Trimmomatic on B1_left.fq (R1 unpaired)
   - 34: Trimmomatic on A2_right.fq (R2 paired)
   - 33: Trimmomatic on A2_left.fq (R1 paired)
   - 32: Trimmomatic on A1_right.fq (R2 paired)
   - 31: Trimmomatic on A1_left.fq (R1 paired)
   - 30: Trimmomatic on A1_right.fq (R2 unpaired)
   - 29: Trimmomatic on A1_left.fq (R1 unpaired)
   - 28: Trimmomatic on A1_right.fq (R2 unpaired)
   - 27: Trimmomatic on A1_left.fq (R1 unpaired)
   - 26: Trimmomatic on A1_right.fq (R2 paired)
   - 25: Trimmomatic on A1_left.fq (R1 paired)
   - 8: B2_right.fq
   - 7: B2_left.fq
   - 6: B1_right.fq
   - 5: B1_left.fq
   - 4: A2_right.fq
   - 3: A2_left.fq
   - 2: A1_right.fq
   - 1: A1_left.fq
   This is a batch mode input field. Separate jobs will be triggered for each dataset selection.

2: all « paired »

If the cleaned reads of good enough quality, we can go to assembly them

if not return to the cleanup step and change some parameters
De novo assembly

1. **Trinity de novo assembly of RNA-Seq data**

2. **Paired or Single-end data?**
   - Paired

3. **Left/Forward strand reads**
   - All « left paired of cleaned reads »
   - (left)

4. **Right/Reverse strand reads**
   - All « right paired of cleaned reads »
   - (right)
4. Depend on your data!

5. Depend on your data!

6. Depend on your data!

<table>
<thead>
<tr>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>38: Trimmmomatic on B2_right.fq (R2 paired)</td>
</tr>
<tr>
<td>37: Trimmmomatic on B2_left.fq (R1 paired)</td>
</tr>
<tr>
<td>36: Trimmmomatic on B1_right.fq (R2 unpaired)</td>
</tr>
<tr>
<td>35: Trimmmomatic on B1_left.fq (R1 unpaired)</td>
</tr>
</tbody>
</table>

Run in silico normalization of reads
- Yes
- No

Jaccard Clip options
- Yes
- No

Set if you expect high gene density with UTR overlap (--jaccardClip)

Additional Options

Execute

---

**Output**

| >TRINITY_DN0_c0_g1_i1 len=2298 path=[4551:0-2297] [-1, 4551, -2] |
| GGACCAAGGAGTCTAGCATCTATGCGAGTGTTTGGGTGATGAAAACC CATCCGCGAAATGAAAGTGAATGCAGGTGGGAACGCCCTTGTGGCG TGCACCATCGACCGACCCGGAAGT |

---

**Trinity version:** v2.2.0
**NOT:** Latest version of Trinity is Trinity-v2.3.2, and can be obtained at: [https://github.com/trinityrnaseq/trinityrnaseq/releases](https://github.com/trinityrnaseq/trinityrnaseq/releases)

Monday, November 21, 2016:
12:59:50 CMD: java -Xmx4m -XX:ParallelGCThreads=2 -jar /opt/trinity/trinityrnaseq.jar
Assembly assessment: simple metrics

1. Trinity suite:
   - Trinity de novo assembly of RNA-Seq data

   Trinity Statistics: Obtain basic stats for the number of genes and isoforms and contiguity of the assembly (Galaxy Version 2.2.0.0)

2. Trinity assembly:
   - 58: Trinity on data 38, data 34, and others: Assembled Transcri...  
   - 1. Execute

3. Output:
   - Stats based on ALL transcript contigs:
     - Contig N10: 1393
     - Contig N20: 924
     - Contig N30: 719
     - Contig N40: 563
     - Contig N50: 471
     - Median contig length: 315
     - Average contig: 426.88
     - Total assembled bases: 222406

   - Stats based on ONLY LONGEST ISOFORM per 'GENE':
     - Contig N10: 1381
     - Contig N20: 899
     - Contig N30: 704
     - Contig N40: 561
     - Contig N50: 468
     - Median contig length: 314.5
     - Average contig: 424.03
     - Total assembled bases: 220494
Abundance estimation for a better transcriptome

1. Align reads and estimate abundance on a de novo assembly of RNA-Seq data.

2. Check step 5 on page 13.

3. One mapping by sample, but we can run with multiple datasets.

4. Depend on your data!

Abundance estimation for a better transcriptome

<table>
<thead>
<tr>
<th>Left/Forward strand reads</th>
</tr>
</thead>
<tbody>
<tr>
<td>40: Trimmomatic on B2_right.fq (R2 unpaired)</td>
</tr>
<tr>
<td>39: Trimmomatic on B2_left.fq (R1 unpaired)</td>
</tr>
<tr>
<td>38: Trimmomatic on B2_right.fq (R2 paired)</td>
</tr>
<tr>
<td>37: Trimmomatic on B2_left.fq (R1 paired)</td>
</tr>
<tr>
<td>36: Trimmomatic on B1_right.fq (R2 unpaired)</td>
</tr>
<tr>
<td>35: Trimmomatic on B1_left.fq (R1 unpaired)</td>
</tr>
<tr>
<td>34: Trimmomatic on B1_right.fq (R2 paired)</td>
</tr>
<tr>
<td>33: Trimmomatic on B1_left.fq (R1 paired)</td>
</tr>
<tr>
<td>32: Trimmomatic on B1_right.fq (R2 unpaired)</td>
</tr>
<tr>
<td>31: Trimmomatic on B1_left.fq (R1 unpaired)</td>
</tr>
<tr>
<td>30: Trimmomatic on B1_right.fq (R2 paired)</td>
</tr>
<tr>
<td>29: Trimmomatic on B1_left.fq (R1 paired)</td>
</tr>
<tr>
<td>28: Trimmomatic on A2_right.fq (R2 unpaired)</td>
</tr>
<tr>
<td>27: Trimmomatic on A2_left.fq (R1 unpaired)</td>
</tr>
<tr>
<td>26: Trimmomatic on A2_right.fq (R2 paired)</td>
</tr>
<tr>
<td>25: Trimmomatic on A2_left.fq (R1 paired)</td>
</tr>
</tbody>
</table>

All « left paired of cleaned reads »

<table>
<thead>
<tr>
<th>Right/Reverse strand reads</th>
</tr>
</thead>
<tbody>
<tr>
<td>40: Trimmomatic on B2_right.fq (R2 unpaired)</td>
</tr>
<tr>
<td>39: Trimmomatic on B2_left.fq (R1 unpaired)</td>
</tr>
<tr>
<td>38: Trimmomatic on B2_right.fq (R2 paired)</td>
</tr>
<tr>
<td>37: Trimmomatic on B2_left.fq (R1 paired)</td>
</tr>
<tr>
<td>36: Trimmomatic on B1_right.fq (R2 unpaired)</td>
</tr>
<tr>
<td>35: Trimmomatic on B1_left.fq (R1 unpaired)</td>
</tr>
<tr>
<td>34: Trimmomatic on B1_right.fq (R2 paired)</td>
</tr>
<tr>
<td>33: Trimmomatic on B1_left.fq (R1 paired)</td>
</tr>
<tr>
<td>32: Trimmomatic on B1_right.fq (R2 unpaired)</td>
</tr>
<tr>
<td>31: Trimmomatic on B1_left.fq (R1 unpaired)</td>
</tr>
<tr>
<td>30: Trimmomatic on B1_right.fq (R2 paired)</td>
</tr>
<tr>
<td>29: Trimmomatic on B1_left.fq (R1 paired)</td>
</tr>
<tr>
<td>28: Trimmomatic on A2_right.fq (R2 unpaired)</td>
</tr>
<tr>
<td>27: Trimmomatic on A2_left.fq (R1 unpaired)</td>
</tr>
<tr>
<td>26: Trimmomatic on A2_right.fq (R2 paired)</td>
</tr>
<tr>
<td>25: Trimmomatic on A2_left.fq (R1 paired)</td>
</tr>
</tbody>
</table>

All « right paired of cleaned reads »
Abundance estimation for a better transcriptome

9305 reads; of these:
9305 (100.00%) were paired; of these:
6634 (71.30%) aligned concordantly 0 times
2608 (28.03%) aligned concordantly exactly 1 time
63 (0.68%) aligned concordantly >1 times

28.70% overall alignment rate

28.7% <<<<~80%:
→ parameters?
→ assembly?
→ In this case: quantity of reads is not enough
1. Transcribe the isoforms counts for each sample and name the sample.

3. Outputs

4. Additional Options

5. Execute
Transcriptome filtering

Output of 1.
Raw counts

Output of 2.
TPM normalization
(Transcripts Per Million)
Transcriptome filtering

6. Filter low expression transcripts from a Trinity assembly
7. Trinity assembly
8. Expression matrix
9. Minimum expression level required across any sample
10. Execute
11. Output
12. Change data type
13. fasta
14. Save
Transcriptome filtering

15. **Trinity Statistics**
   - Obtain basic stats for the number of genes and isoforms and contiguity of the assembly (Galaxy version 2.2.0.0)

16. **Output**
   - Before filtering
   - After filtering

### Before filtering

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Trinity 'genes':</td>
<td>520</td>
</tr>
<tr>
<td>Total Trinity transcripts:</td>
<td>521</td>
</tr>
<tr>
<td>Percent GC: 44.23</td>
<td></td>
</tr>
</tbody>
</table>

### After filtering

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total trinity 'genes':</td>
<td>437</td>
</tr>
<tr>
<td>Total trinity transcripts:</td>
<td>438</td>
</tr>
<tr>
<td>Percent GC: 44.31</td>
<td></td>
</tr>
</tbody>
</table>

**Do you have a good or bad threshold?**
Abundance estimation for a differential expression analysis

1. Align reads and estimate abundance on de novo assembly of RNA-Seq data

2. Filter low expression transcripts on data 69 and data 38: filtered low expression transcripts de novo assembly of RNA-Seq data

3. Paired or Single-end data? Paired

4. All « left paired of cleaned reads »

5. All « right paired of cleaned reads »

6. Depends on your data !

7. Outputs for sample A1
Protein coding regions identification

1. Depend on your data!

Outputs
Functional annotation: BLASTX

1. NCBI BLAST+ blastx Search protein database with translated nucleotide query sequence(s)
2. Nucleotide query sequence(s)
   - 58: Trinity on data 38, data 34, and others: Assembled Transcripts
3. Subject database/sequences
   - Locally installed BLAST database
   - Protein BLAST database
     - Select/Unselect all
     - UniProt Swiss-Prot for Trinotate
4. Query genetic code
5. Type of BLAST
   - Standard
   - blastx - Traditional BLASTX to compare translated nucleotide query to protein database
   - blastx-fast - Use longer words for seeding, faster but less accurate
6. Set expectation value cutoff
   - 0.001
7. Output format
   - Tabular (standard 12 columns)
8. Advanced Options
   - Hide Advanced Options
9. Execute

Note: Database searches may take a substantial amount of time. For large input datasets it is advisable to...
Functional annotation: BLASTX

Because Galaxy focuses on processing tabular data, the default output of this tool is tabular. The standard BLAST+ tabular output contains 12 columns:

<table>
<thead>
<tr>
<th>Column</th>
<th>NCBI name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>qseqid</td>
<td>Query Seq-id (ID of your sequence)</td>
</tr>
<tr>
<td>2</td>
<td>sseqid</td>
<td>Subject Seq-id (ID of the database hit)</td>
</tr>
<tr>
<td>3</td>
<td>pident</td>
<td>Percentage of identical matches</td>
</tr>
<tr>
<td>4</td>
<td>length</td>
<td>Alignment length</td>
</tr>
<tr>
<td>5</td>
<td>mismatch</td>
<td>Number of mismatches</td>
</tr>
<tr>
<td>6</td>
<td>gapopen</td>
<td>Number of gap openings</td>
</tr>
<tr>
<td>7</td>
<td>qstart</td>
<td>Start of alignment in query</td>
</tr>
<tr>
<td>8</td>
<td>qend</td>
<td>End of alignment in query</td>
</tr>
<tr>
<td>9</td>
<td>sstart</td>
<td>Start of alignment in subject (database hit)</td>
</tr>
<tr>
<td>10</td>
<td>send</td>
<td>End of alignment in subject (database hit)</td>
</tr>
<tr>
<td>11</td>
<td>evalue</td>
<td>Expectation value (E-value)</td>
</tr>
<tr>
<td>12</td>
<td>bitscore</td>
<td>Bit score</td>
</tr>
</tbody>
</table>
Functional annotation: BLASTP

1. NCBI BLAST+ blastp
2. Protein query sequence(s)
3. Subject database/sequences
   - Locally installed BLAST database
   - Protein BLAST database
     - UniProt Swiss-Prot for Trinotate
4. Type of BLAST
   - blastp - Traditional BLASTP to compare a protein query to a protein database
   - blastp-fast - Use longer words for searching, faster but less accurate
   - blastp-short - BLASTP optimized for queries shorter than 30 residues
5. Set expectation value cutoff
   - 0.001
6. Output format
   - Tabular (standard 12 columns)
7. Advanced Options
   - Hide Advanced Options
   - Execute

Note: Database searches may take a substantial amount of time. For large input datasets it is advisable to allow overnight processing.
Functional annotation: Diamondx

1. Diamond alignment tool for short sequences against a protein database

2. Will you select a reference genome from your history or use a built-in index?
   - Use a built-in index
   - Built-ins were indexed using default options

3. Select a reference genome
   - Uniprot Swissprot
   - If your genome of interest is not listed, contact your Galaxy admin

4. What do you want to align
   - Align DNA query sequences (blastx)

---

23/11/2016

Title

2. Input query file in FASTA format.

3. Will you select a reference genome from your history or use a built-in index?
   - Use a built-in index
   - Built-ins were indexed using default options
   - Select a reference genome:
     - Uniprot Swissprot

4. What do you want to align?
   - Align amino acid query sequences (blastp)
     - (--blastp|--blastx)

   Gap open penalty
   - 11
     - (--gapopen)

   Gap extend penalty
   - 1
     - (--gapextend)

   Select scoring matrix
   - BLOSUM62
     - (--matrix)

   Filter by score
   - Maximum e-value to report alignments
     - (--eval/e--/min-score)
Functional annotation: protein domains identification

1. **hmmscan** search sequence(s) against a profile database

2. **Select a HMM model database**
   - Pfam-a for Trinotate

3. **Outputs**
   - Table of per-sequence/ per-domain hits from HMM matches of TransDecoder on data 58: pep against the profile database
   - Table of per-domain hits from HMM matches of TransDecoder on data 58: pep against the profile database
   - Table of per-sequence hits from HMM matches of TransDecoder on data 58: pep against the profile database
Target name: domain name  
Accession: Pfam id  
Query name: transcript id  
E-value  
Description of target: domain description
Functional annotation:
Signal peptides prediction

1. **Signal peptides prediction**

2. **FASTA file of protein sequences**

3. **Organism**
   - Eukaryote

4. **Truncate sequences to this many amino acids**
   - 70

5. **Use zero for no truncation, maximum value 6000**

6. **Execute**

7. **Output**
   - **Depend on your data!**

---

23/11/2016

**Titre**
Functional annotation: Transmembrane regions prediction

1. Tools
   - Galaxy / ABiMS
   - Protein sequence analysis
     - TMHMM 2.0: Find transmembrane domains in protein sequences
     - SignalP 3.0: Find signal peptides in protein sequences
   - Fasta file of protein sequences

2. Output
   - 93: TMHMM results
     - 92: SignalP euk results
     - 91: Table of per-sequence per-domain hits from HMM matches

3. Table
   - ID: TRINITY_DN0_c0.g1_i1.1
   - Length: 105
   - ExpAA: 0.09
   - First60: 0.09
   - PredHel: 0
   - Topology: 0
   - Localisation: o: no transmembranal, i: transmembranal and movement

Time(s)
Functional annotation: correspondance between gene and transcript ids

1. Generate gene to transcript map for Trinity assembly

Trinity assembly

58: Trinity on data 38, data 34, and others: Assembled Transcripts

Execute

3. Trinity assembles transcript sequences from Illumina RNA-Seq data. This tool produces a file containing correspondence between gene ids and transcript ids based on the name of transcripts assembled by Trinity. The output file is intended to be used by the “Align reads and estimate abundance” tool.

Output

1. Generate gene to transcript map on data 58: Genes to transcripts map

93: TMHMM results

92: SignalP euk results

91: Table of euk results
Diamond results can be used for steps 4 and 5
Functional annotation and analysis

<table>
<thead>
<tr>
<th>gene_id</th>
<th>transcript_id</th>
<th>sprot_Top_BLASTX_hit</th>
</tr>
</thead>
<tbody>
<tr>
<td>#gene_id</td>
<td>transcript_id</td>
<td>sprot_Top_BLASTX_hit</td>
</tr>
<tr>
<td>TRINITY_DN0_c0_g1</td>
<td>TRINITY_DN0_c0_g1</td>
<td>ART2.YEAST^ART2.YEAST^Q:2164-1997,H1:564/85.71%</td>
</tr>
<tr>
<td>TRINITY_DNB_c0_g1</td>
<td>TRINITY_DNB_c0_g1</td>
<td>HSP70.SCHPO^HSP70.SCHPO^Q:847-2,H2:231-512^100%</td>
</tr>
<tr>
<td>TRINITY_DN9_c0_g1</td>
<td>TRINITY_DN9_c0_g1</td>
<td>RS14B.SCHPO^RS14B.SCHPO^Q:518-135,H1:128^100%</td>
</tr>
<tr>
<td>TRINITY_DN1_c0_g1</td>
<td>TRINITY_DN1_c0_g1</td>
<td>FF3.SCHPO^FF3.SCHPO^Q:1273-266,H1:432^100%</td>
</tr>
</tbody>
</table>

Output

The output has the following column headers:

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>#gene_id</td>
</tr>
<tr>
<td>1</td>
<td>transcript_id</td>
</tr>
<tr>
<td>2</td>
<td>sprot_Top_BLASTX_hit</td>
</tr>
<tr>
<td>3</td>
<td>RNAMMER</td>
</tr>
<tr>
<td>4</td>
<td>prot_id</td>
</tr>
<tr>
<td>5</td>
<td>prot_coords</td>
</tr>
<tr>
<td>6</td>
<td>sprot_Top_BLASTP_hit</td>
</tr>
<tr>
<td>7</td>
<td>custom_pombe_pep_BLASTX</td>
</tr>
<tr>
<td>8</td>
<td>custom_pombe_pep_BLASTP</td>
</tr>
<tr>
<td>9</td>
<td>Pfam</td>
</tr>
<tr>
<td>10</td>
<td>SignalP</td>
</tr>
<tr>
<td>11</td>
<td>TmHMM</td>
</tr>
<tr>
<td>12</td>
<td>eggnog</td>
</tr>
<tr>
<td>13</td>
<td>Kegg</td>
</tr>
<tr>
<td>14</td>
<td>gene_ontology_blast</td>
</tr>
<tr>
<td>15</td>
<td>gene_ontology_pfam</td>
</tr>
<tr>
<td>16</td>
<td>transcript</td>
</tr>
<tr>
<td>17</td>
<td>peptide</td>
</tr>
</tbody>
</table>
Dear class,

THANKS FOR YOUR ATTENTION
AND
PLEASE DON’T ASK TOO MUCH