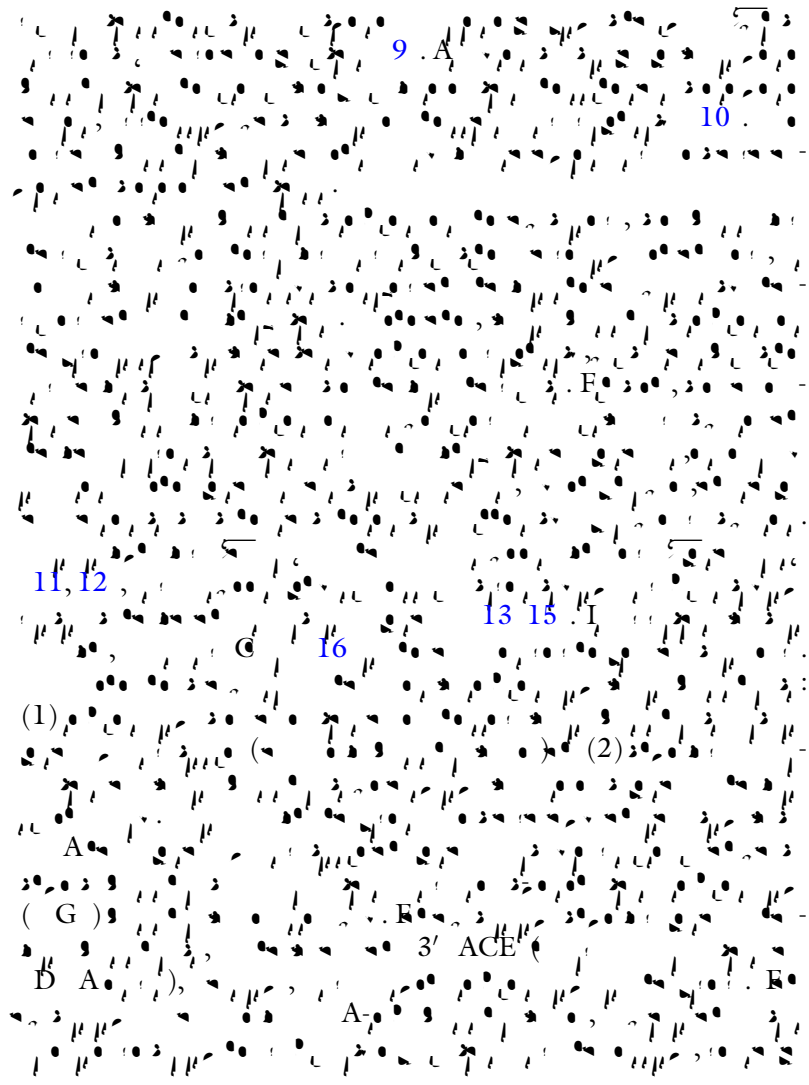


2.2 3′ ACE(ı A DNA E ı)

1. 3′ ACE ... D A ... (... -
   F ... ®).
2. ... : ... Note 1
3. ... H ... F ... (... F ... -
   ... ®

2.3.2 NGS Data Quality Checks

1. F C (:// /) 18 .
2. (:// ? =) 19 .
3. EA (:// / /) 20 .
4. A H (:// / A H/) 21 .

2.3.3 De Novo Transcriptome Assembly

1. (:// / /) 17 .
2. E 22 , .
3. B 23 , .
4. E G (:// E G /) 24 .
5. E (:// /) 25 .
6. CD-HI (:// /) 26 .

2.3.4 Toxin Gene Identification and Expression Quantification

1. A + (:// B 279690/) 27 .
2. DIA D (:// /)

3. EA (:// /) 40 ,
4. G I(:// /) 41 .

3 Methods

3.1 NA I

A . E A A 20 42 . I A A (Note 3 A).

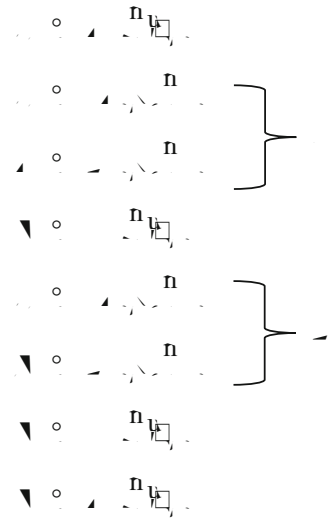
1. A 100 500 2 (1 50) 1 I . I 10 100 I ( ).
2. I 5 .
3. A 200 .
4. C 15 .
5. I 2 3 .
6. G 12,000 × 4 °C 15 .
7. (50%) . D .
8. A 500 100% .
9. I 10 .
10. G 12,000 × 4 °C 10 .
11. A ( ).
12. 1 75% .
13. G 7500 × 4 °C 5 .
14. A 300 100% 40 3 .

15. F [illegible] −20 °C [illegible].
16. G [illegible] 10,000 × [illegible] 15 [illegible] 4 °C.
17. [illegible] 10 [illegible].
18. A [illegible] 10 16 [illegible] Note 4 [illegible] A [illegible]

3.2 3′ ACE (A DNA E): 3′ ACE [illegible] 3′ ACE [illegible] F [illegible] ®

6 0.5

2. 10 70



H 4 10 °C (10 °C).

13. −20 °C 1%

14. E ( ) 1% D A C C .

15. GE - E . A 0.5 :

5 2×

1 GE - E

3 C D A

1 D A

16.

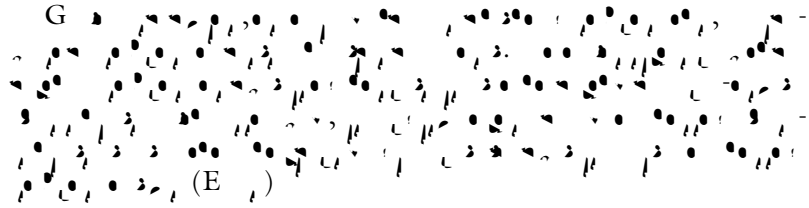
17. I 4 °C .

18. B . A I G, E . DH5α , 5 50 . Note 7

19. F
D
20. I 30
21. H 20 42 °C,
22. I 2 1 B
23. I 60 37 °C
24. 200
25. 37 °C
(16 18 ;).
26. 4 °C
27. / ( ) GE - E - 2 B + (1). E . A 42 .
28. 37 °C
29. I C 5
30. 200 . E GE - E , 7 6

3.3 N -G (NG) :

C D N



2. (E) : ( ). I (>150)

EA (-E A) 20

A H (F A H) 21

3. . I

A

://

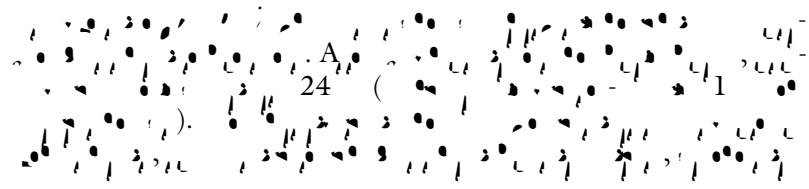
Box 1 Abridged Pipeline Example Commands. A few command examples are given; documentation for each program should be referenced for all command arguments and parameters, and only examples are provided. All CPU / thread arguments should be modified based on computing resources:

```
##############################
### FASTQC example command ###
##############################
SYNOPSIS

Usage:
fastqc seqfile1 seqfile2 .. seqfileN
fastqc [-o output dir] [--(no)extract] [-f fastq|bam|sam]
[-c contaminant file] seqfile1 .. seqfileN

fastqc RAWDATA_PAIR_1.fastq.gz RAWDATA_PAIR_2.fastq.gz -o
OUTPUT_DIRECTORY

###################################
### TRIMMOMATIC example command ###
###################################
SYNOPSIS

Usage:
PE [-threads <threads>] [-phred33|-phred64] [-trimlog
<trimLogFile>] [-quiet] [-validatePairs] [-basein <inputBase> |
<inputFile1> <inputFile2>] [-baseout <outputBase> | <outputFile1P>
<outputFile1U> <outputFile2P> <outputFile2U>] <trimmer1>...
or:
SE [-threads <threads>] [-phred33|-phred64] [-trimlog
<trimLogFile>] [-quiet] <inputFile> <outputFile> <trimmer1>...

java -jar trimmomatic-0.35.jar PE -threads 4 -phred33 RAWDATA_-
PAIR_1.fastq.gz RAWDATA_PAIR_2.fastq.gz OUTPUT_R1-paired.fastq
OUTPUT_R1-unpaired.fastq OUTPUT_R2-paired.fastq OUTPUT_R2-
unpaired.fastq ILLUMINACLIP:TruSeq3-PE-2.fa:2:40:15 SLIDINGWIN-
DOW:4:15 LEADING:20 TRAILING:20 MINLEN:50 HEADCROP:9

############################
### PEAR example command ###
############################
SYNOPSIS

Usage:
pear <options>
```

```
Standard (mandatory):
-f, --forward-fastq <str> Forward paired-end FASTQ file.
-r, --reverse-fastq <str> Reverse paired-end FASTQ file.
-o, --output <str> Output filename.

pear -f INPUT_R1-paired.fastq -r INPUT_R2-paired.fastq -o
OUTPUT_NAME

##############################
### FLASH example command ###
##############################
SYNOPSIS

Usage:
flash [OPTIONS] MATES_1.FASTQ MATES_2.FASTQ
flash [OPTIONS] --interleaved-input (MATES.FASTQ | -)
flash [OPTIONS] --tab-delimited-input (MATES.TAB | -)

flash -o OUTPUT_PREFIX -t 5 INPUT_R1-paired.fastq INPUT_R2-paired.
fastq -r 140 -f 350 -s 50 -d OUTPUT_DIRECTORY

################################
### TRINITY example command ###
################################
SYNOPSIS

#Usage:
# --seqType <string> :type of reads: ('fa' or 'fq')
#
# --max_memory <string> :suggested max memory to use by #Trinity where
limiting can be enabled. (jellyfish, sorting, etc)
#provided in Gb of RAM, ie. '--max_memory 10G'
#
# If paired reads:
# --left <string> :left reads, one or more file names #(separated by
commas, no spaces)
# --right <string> :right reads, one or more file names #(separated by
commas, no spaces)
#
# Or, if unpaired reads:
# --single <string> :single reads, one or more file names, #comma-
```

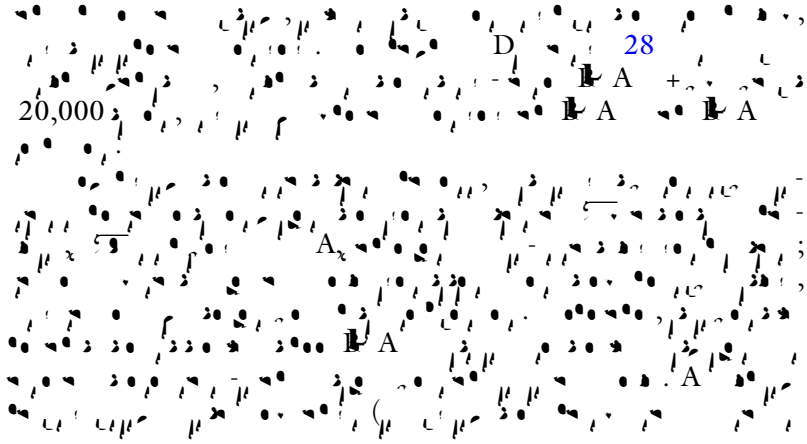
```
# --samples_file <string> tab-delimited text file #indicating
biological replicate relationships.
#ex.
#cond_A cond_A_rep1 A_rep1_left.fq A_rep1_right.fq
#cond_A cond_A_rep2 A_rep2_left.fq A_rep2_right.fq
#cond_B cond_B_rep1 B_rep1_left.fq B_rep1_right.fq
#cond_B cond_B_rep2 B_rep2_left.fq B_rep2_right.fq
#

Trinity --seqType fq --max_memory 50G --left INPUT_R1-paired.fastq.
gz --right INPUT_R2-paired.fastq.gz --CPU 6 --full_cleanup --min_-
contig_length 100 --verbose

################################
### CD-HIT example command ###
################################
SYNOPSIS

Usage:
cd-hit-est [Options]

cd-hit-est -i INPUT_SEQUENCE -o OUTPUT_SEQUENCE -c 1 -n 8

################################
### BLAST+ example command ###
################################
SYNOPSIS

Usage:
blastx [-h] [-help] [-import_search_strategy filename]
[-export_search_strategy filename] [-task task_name] [-db
database_name]
[-dbsize num_letters] [-gilist filename] [-seqidlist filename]
[-negative_gilist filename] [-entrez_query entrez_query]
[-db_soft_mask filtering_algorithm] [-db_hard_mask
filtering_algorithm]
[-subject subject_input_file] [-subject_loc range] [-query
input_file]
[-out output_file] [-evalue evalue] [-word_size int_value]
[-gapopen open_penalty] [-gapextend extend_penalty]
[-qcov_hsp_perc float_value] [-max_hsps int_value]
[-xdrop_ungap float_value] [-xdrop_gap float_value]
[-xdrop_gap_final float_value] [-searchsp int_value]
[-sum_stats bool_value] [-max_intron_length length] [-seg
SEG_options]
```

```
[-soft_masking soft_masking] [-matrix matrix_name]
[-threshold float_value] [-culling_limit int_value]
[-best_hit_overhang float_value] [-best_hit_score_edge
float_value]
[-window_size int_value] [-ungapped] [-lcase_masking] [-query_loc
range]
[-strand strand] [-parse_deflines] [-query_gencode int_value]
[-outfmt format] [-show_gis] [-num_descriptions int_value]
[-num_alignments int_value] [-line_length line_length] [-html]
[-max_target_seqs num_sequences] [-num_threads int_value]
[-remote]
[-comp_based_stats compo] [-use_sw_tback] [-version]

blastx -query INPUT_SEQUENCE -db nr -max_target_seqs 3 -num_threads
8 -outfmt '6 std stitle' -out Blastx_nr_outfmt6

################################
### RSEM example command ###
################################
SYNOPSIS

Usage:
rsem-prepare-reference [options] reference_fasta_file
(s) reference_name
rsem-calculate-expression [options] upstream_read_file
(s) reference_name sample_name
rsem-calculate-expression [options] --paired-end upstream_read_-
file(s) downstream_read_file(s) reference_name sample_name
rsem-calculate-expression [options] --alignments [--paired-end]
```



[illegible]

3.3.4 Toxin Gene Identification and Expression Quantification



E (A-) 31 .

B /B 2 32

B .B

B 2 (

E),

B 2

(E

48

)

3.4 -

68, 69 . I

70

13 . A

71 .

E

( )

72 .

ω,

/ ω < 1

, ω = 1

). BEB 8

7

χ^2

A

24

A

H 77 . H

A

H D

38 . D 78 .

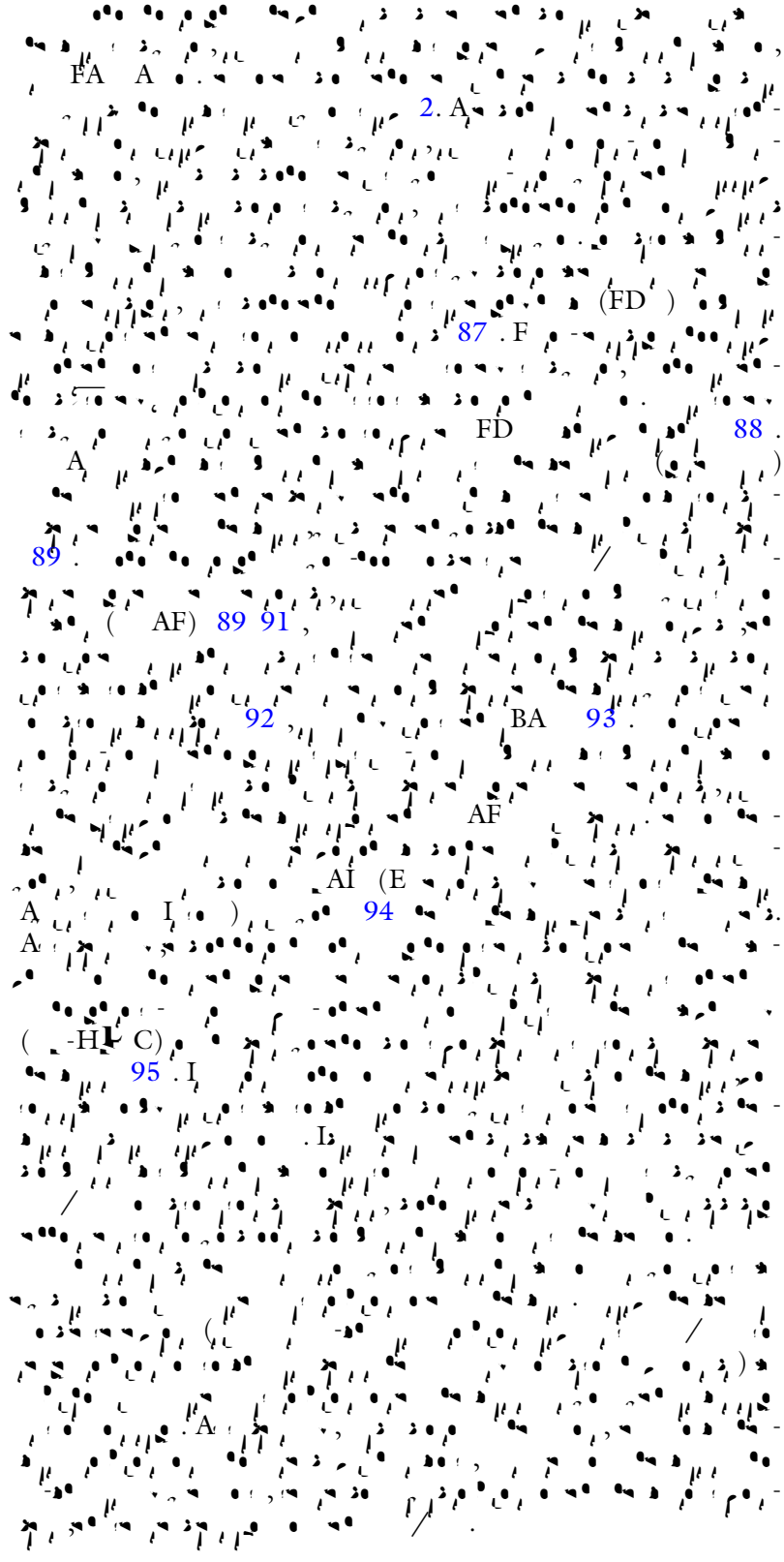
3.5 H -

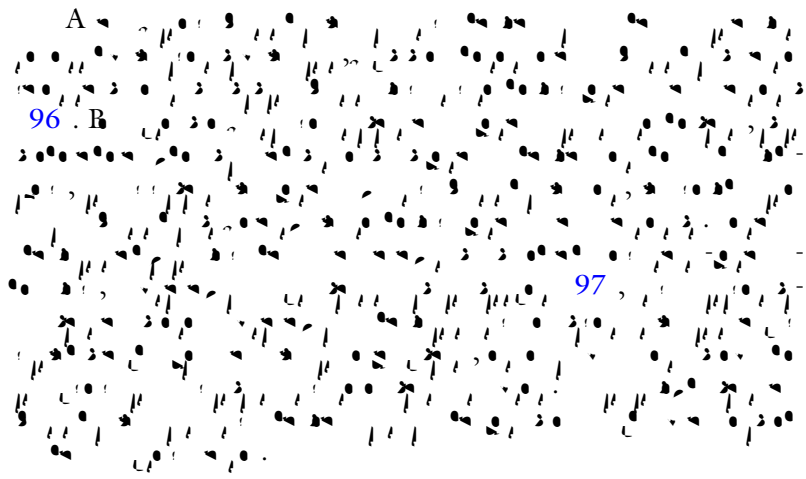
I

A

(/)

( ), G -C, /





4 Notes

1. I (G) (A), , l 4 . I C (2016) C 3′ ACE.
2. I G 50 °C.
3. A ( ),

A
A (F). A ,
A (,
DE C). I
A.
A
A A
A
A
A
A
A −80 °C
A
−80 °C, A
. I 1 -
2
I
A . I A
.

4. I A -
, D I (; I)
D A,
3′ ACE .
D A

5. C
C
3′ ACE
D A (F).
C
F
42 .

6. C : (1) I
25 , 50
. (2) I
H ,
C . (3) I
. H , (>40)
. (4) I

68 °C
(5) D A C . 1:2
1:10 D A C .

7.
A
1 B . A

8. A B . I
(A I) B
A (A)
28 18 98 ; .A I 7 8

11(3):315 329
13. A D, A , -B A,

38. D , AF , E D , (2010) D 2010:

62. , E (2013) A . B C G 14:465
63. F D, C , E (2011) H E : A 39: 29 37
64. F D, G , E , E , A , C, A, GA, , B A (2016) : A 44(D1):D279 D285
65. H , A, A , A , B , A, B , B D, B , B , C E, G , G , D , D , D , F D, F , G , H D, H , D, E, I, D, , A C, D , C, H, D, C, , AF, C, A, D, C , , D, CH, C, (2012) I 2011: A 40:D306 D312
66. C (2017) G B 27 (3):230 240
67. A, BA, C , B (2008) A 5(7):621 628
68. , (2015) G 11(10) 1005596
69. EA, H, G EC, G C, CE, I, H , GF, A A, F BG (2014) I ( ): 99:68 83
70. D , A E , C B (2011) A ( ) 57(5):657 671
71. A D, A , B A, , A (2017)
72. , E, A A, A, F BG (2013) A E : 5(11):2172 2208
73. (1998) HI -1 G 148(3):929 936
74. (2000) E 51:423 432
75. , B , (2005) B B E 22 (4):1107 1118
76. , , (2005) E B E 22(12):2472 2479
77. , F D, (2005) H : B 21(5):676 679
78. , F D (2005) D B 21 (10):2531 2533
79. C A, A, C , C , C , (2018) 10(9):E373
80. (1988) G 1:92 101
81. , , F, D A, G (2003) G 42 (4):405 411
82. A , D, , C A, C , (2015) G ( ) G : G F ( ). 121:28 43
83. D , , , EE (2017) ( ). 5 3249
84. D , C , EE, , C , B (2012)

[illegible]