RNA-seq analysis
- de novo assembly

2011/03/25

Outline

- Contig assembly
- Functional annotation

- Unix command quick reference
  - http://163.25.92.61/course/unix_command.pdf

Contig Assembly

- Load and Import data
  - load data to server
    - Open Terminal

```
wget   http://163.25.92.61/course/testread.fq
```

Always check data first

- Check line number

```
wc   -l   testread.fq
```

- Check tail of file

```
tail  -8  testread.fq
```

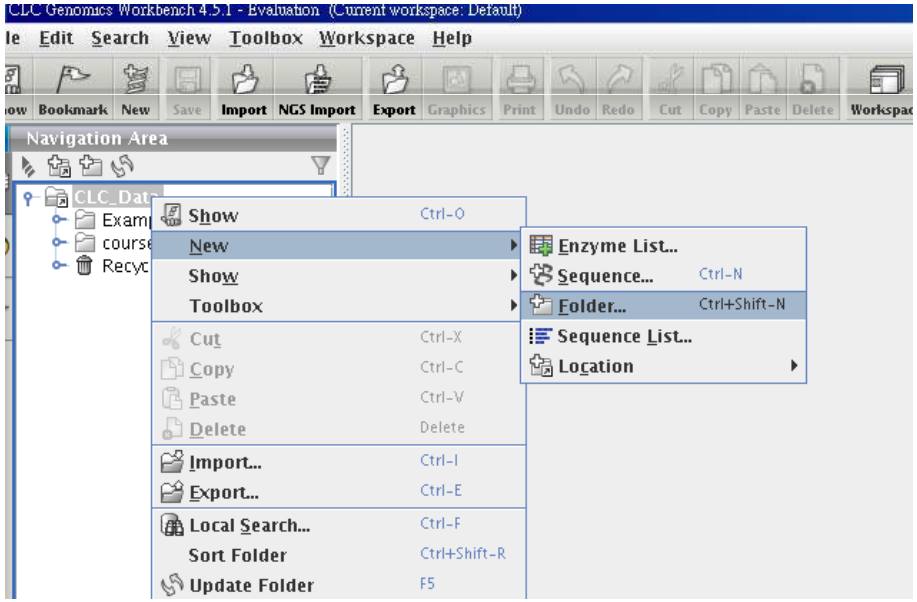
Fastq format

.....
.....
@A80E9RABXX:2:1:16008:20136#TAGCTTAT/1
CGCGGGTTGGTGGGCAGTTCTTGCCGTGGATCCAGATGCGGGAGTTAAGCTCTGGGTCGATCTTAACAACATCTGTGTGCATCATCTTCT
+
fffffffffdeeeeffffffeffeeffffffeffffeefafbeedfeffcdfd_ee^dbada^bb^baaecd\dadd`eeeccefed
@A80E9RABXX:2:1:16165:20136#TAGCTTAT/1
CATTATCATTAGCTACTTCAGAGGCGGAGAATAATGGGTTGTCGCGTGTTGCCTCTTCCATTGGAATTGAGGTTGCAATATGAGTCTCAC
+
ffffdefeffffffffffefdcefeddddYccca`cdddd cddb\c`b``d^b\a]]S_]`b^cT`L`\^]`G^]QUTVF\W[ZT\^[Y_
@A80E9RABXX:2:1:16354:20135#TAGCTTAT/1
CCCTTGATGACGTACTCCTTGAATGGGCCGAAAATGGCGGGGTAATAGTAGAGGGGAGCATCTTGGTAGCCTTCCCCTTCCCCTGGTGGG
+
fffffffffffffffffffffffffffffffeeeeeBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBBB
.....
.....

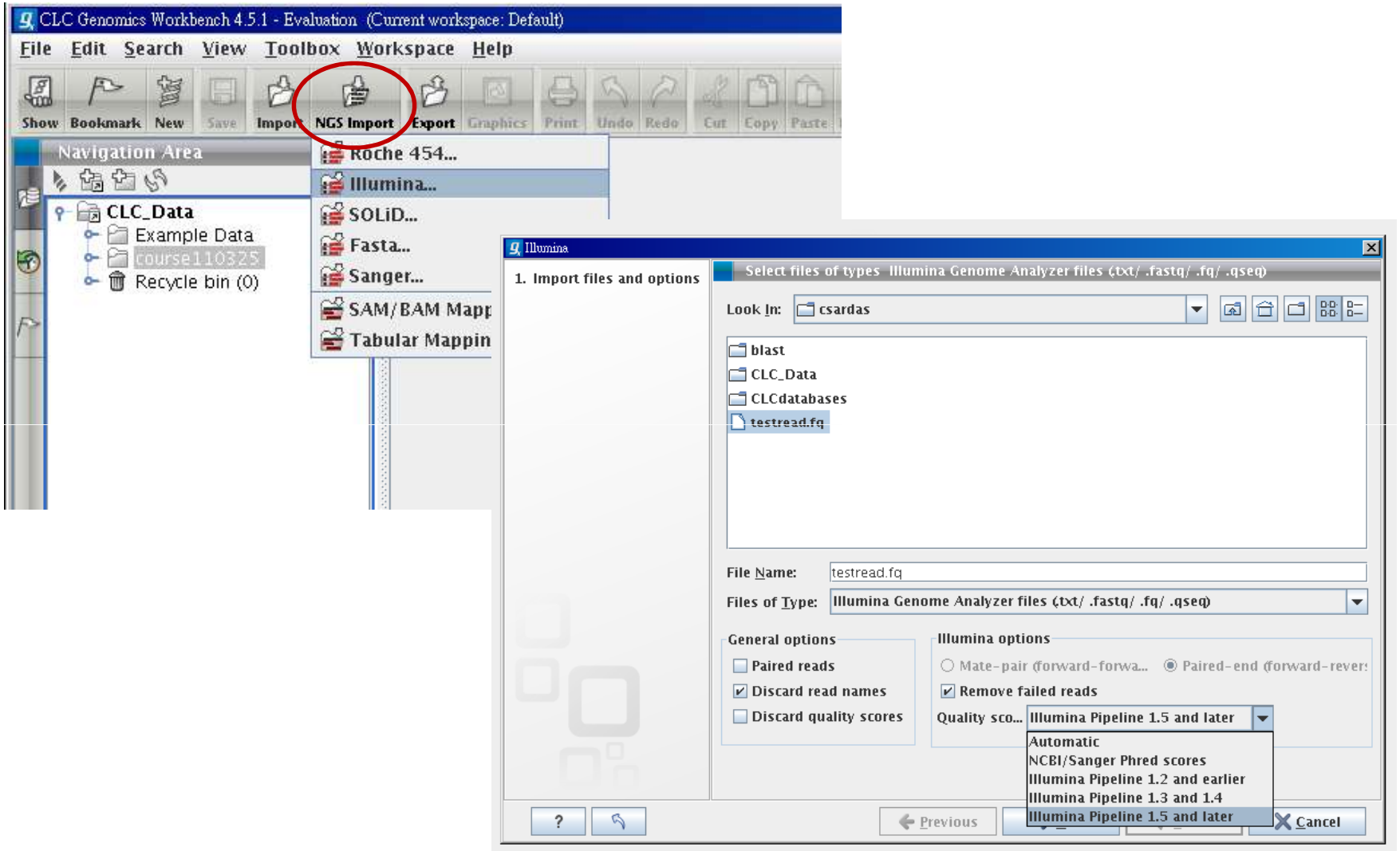
1. Read id
2. Read sequence
3. +
4. Read quality score

– Open CLC Genomics workbench

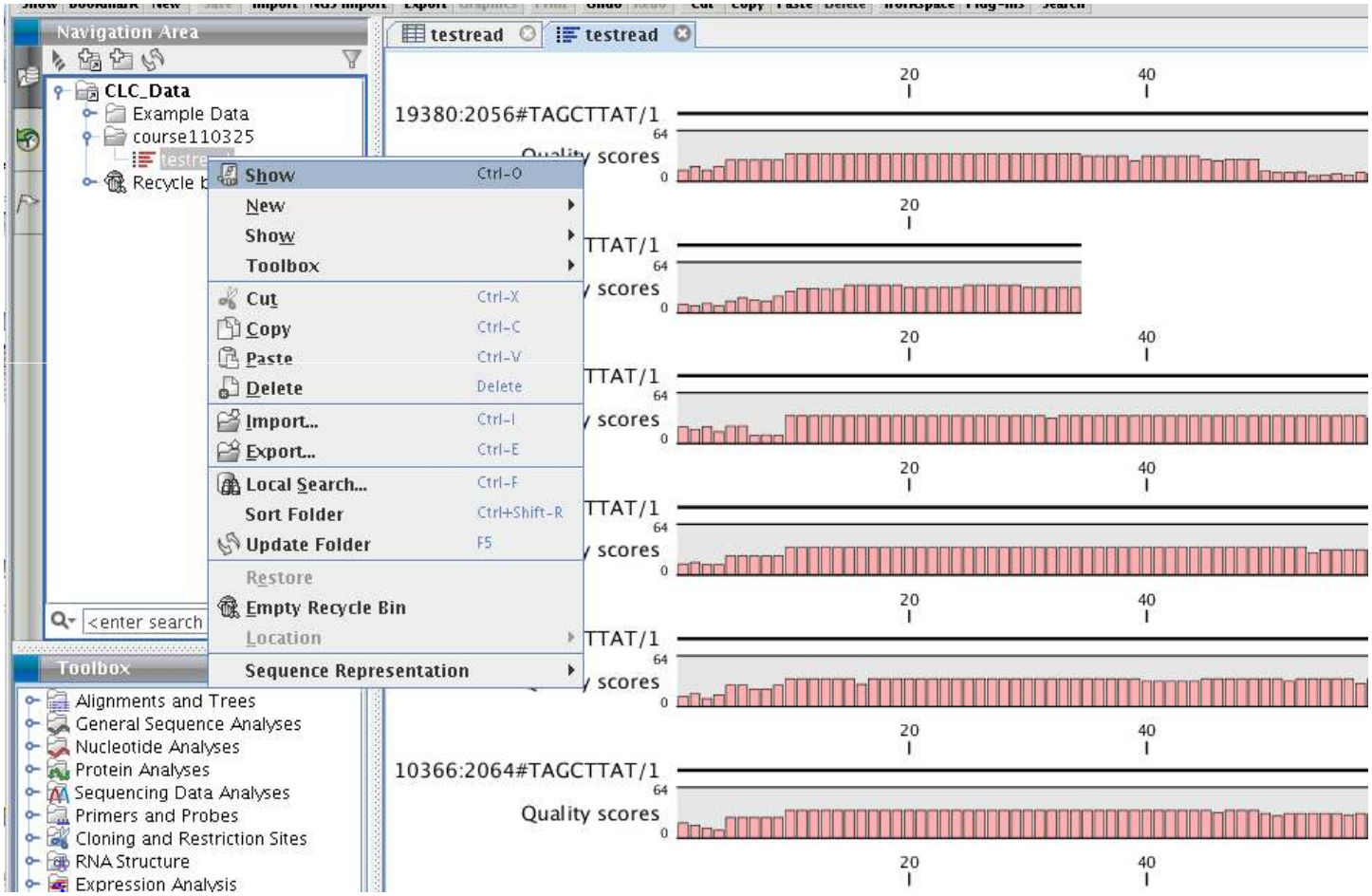
– Create folder “course110325”



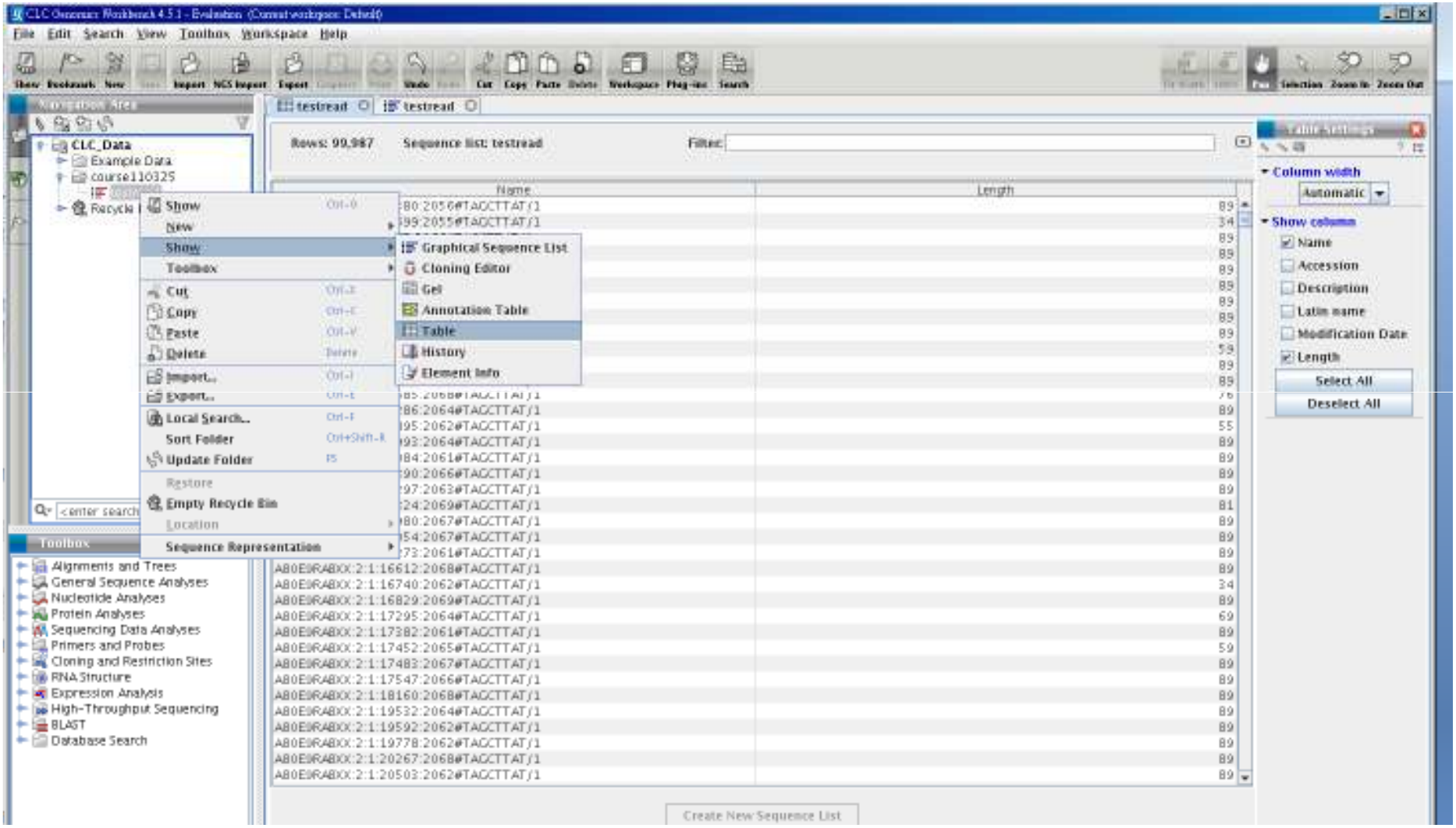
Use “NGS Import”



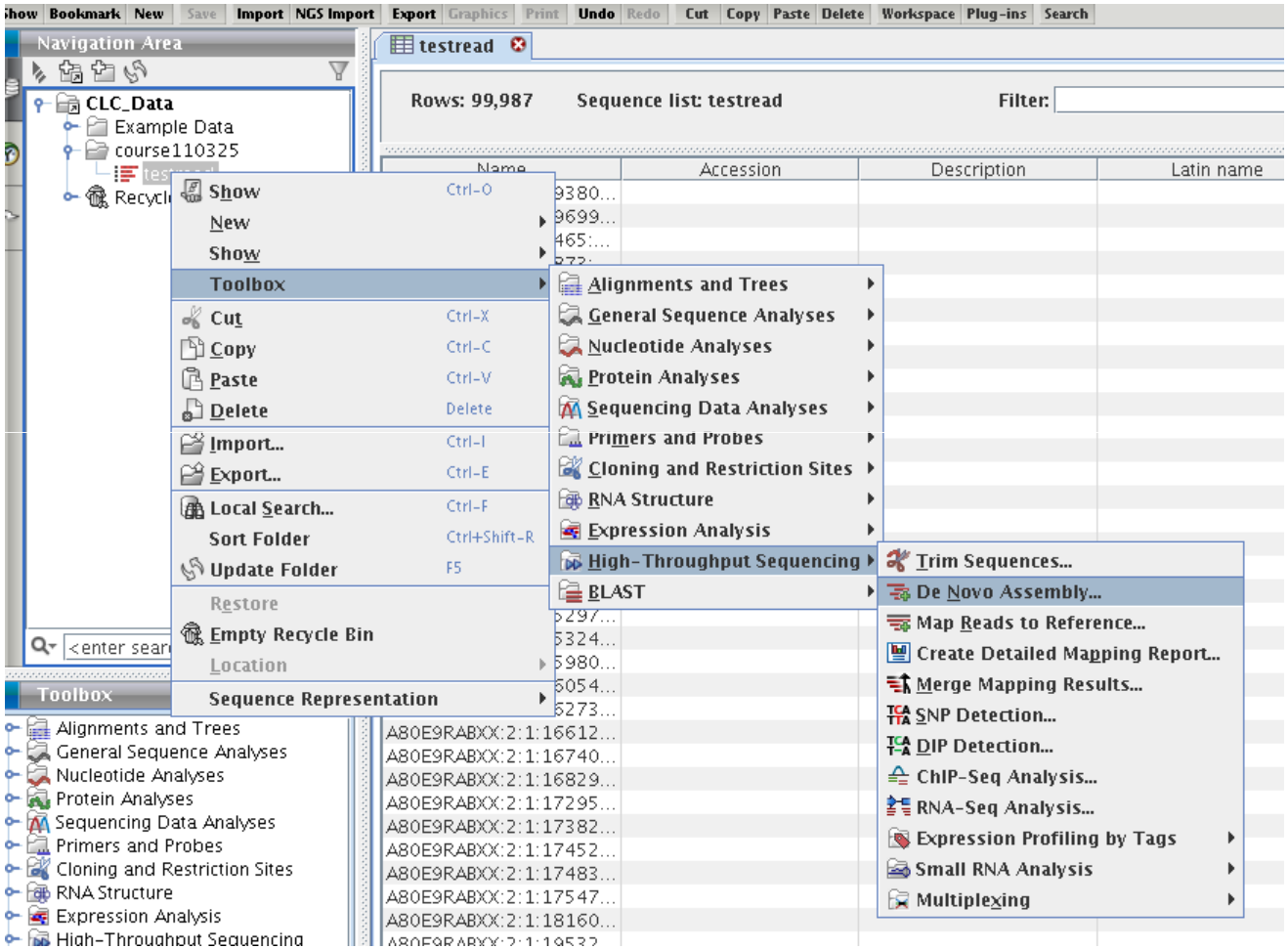
Show data

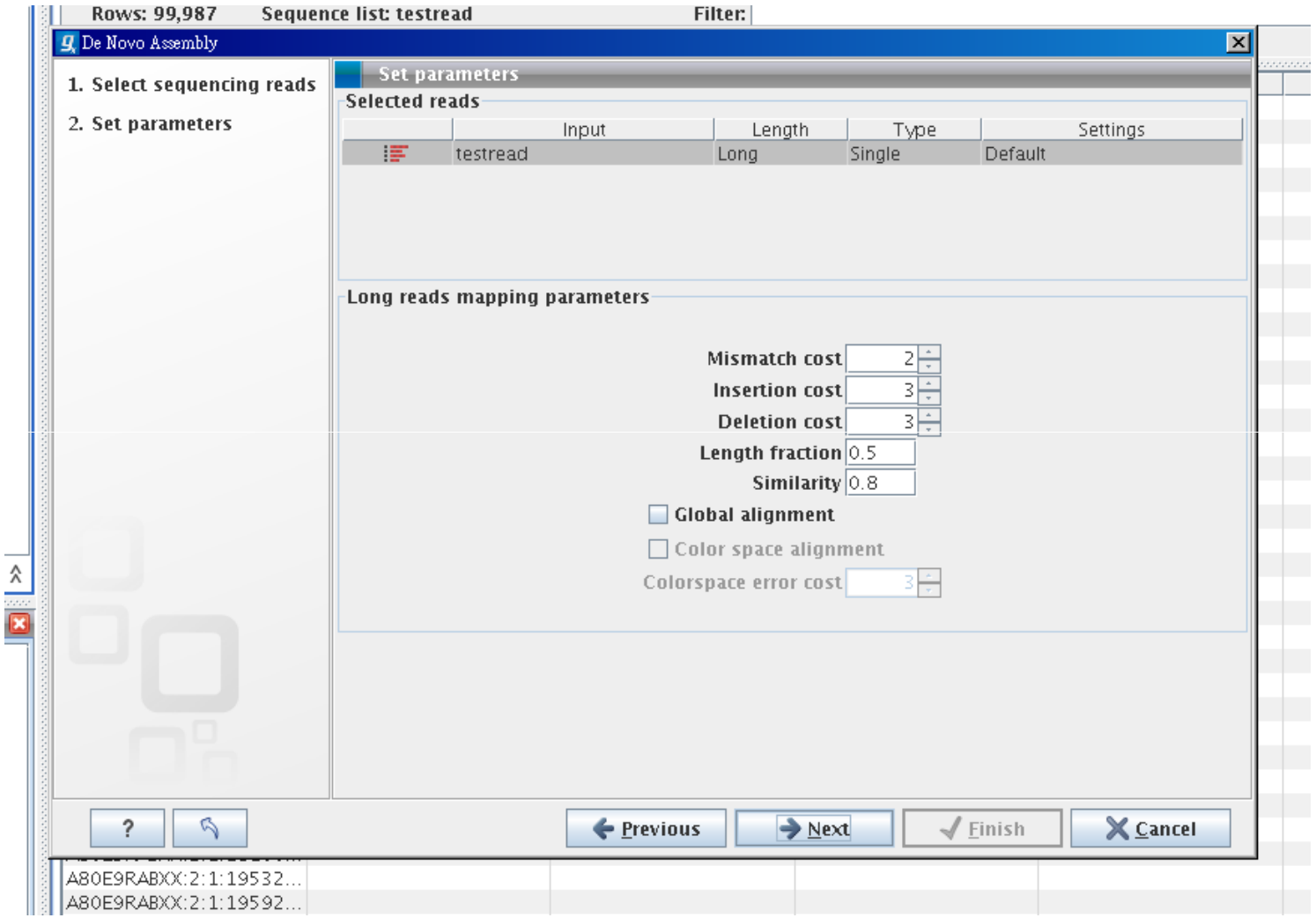

Navigation Area
CLC_Data
Example Data
course110325
testre
Recycle b
Show Ctrl-O
New
Show
Toolbox
Cut Ctrl-X
Copy Ctrl-C
Paste Ctrl-V
Delete Delete
Import... Ctrl-I
Export... Ctrl-E
Local Search... Ctrl-F
Sort Folder Ctrl+Shift-R
Update Folder F5
Restore
Empty Recycle Bin
Location
Sequence Representation
<enter search
Toolbox
Alignments and Trees
General Sequence Analyses
Nucleotide Analyses
Protein Analyses
Sequencing Data Analyses
Primers and Probes
Cloning and Restriction Sites
RNA Structure
Expression Analysis
testread
testread
19380:2056#TAGCTTAT/1
Quality scores
10366:2064#TAGCTTAT/1
Quality scores
20
40
64
0

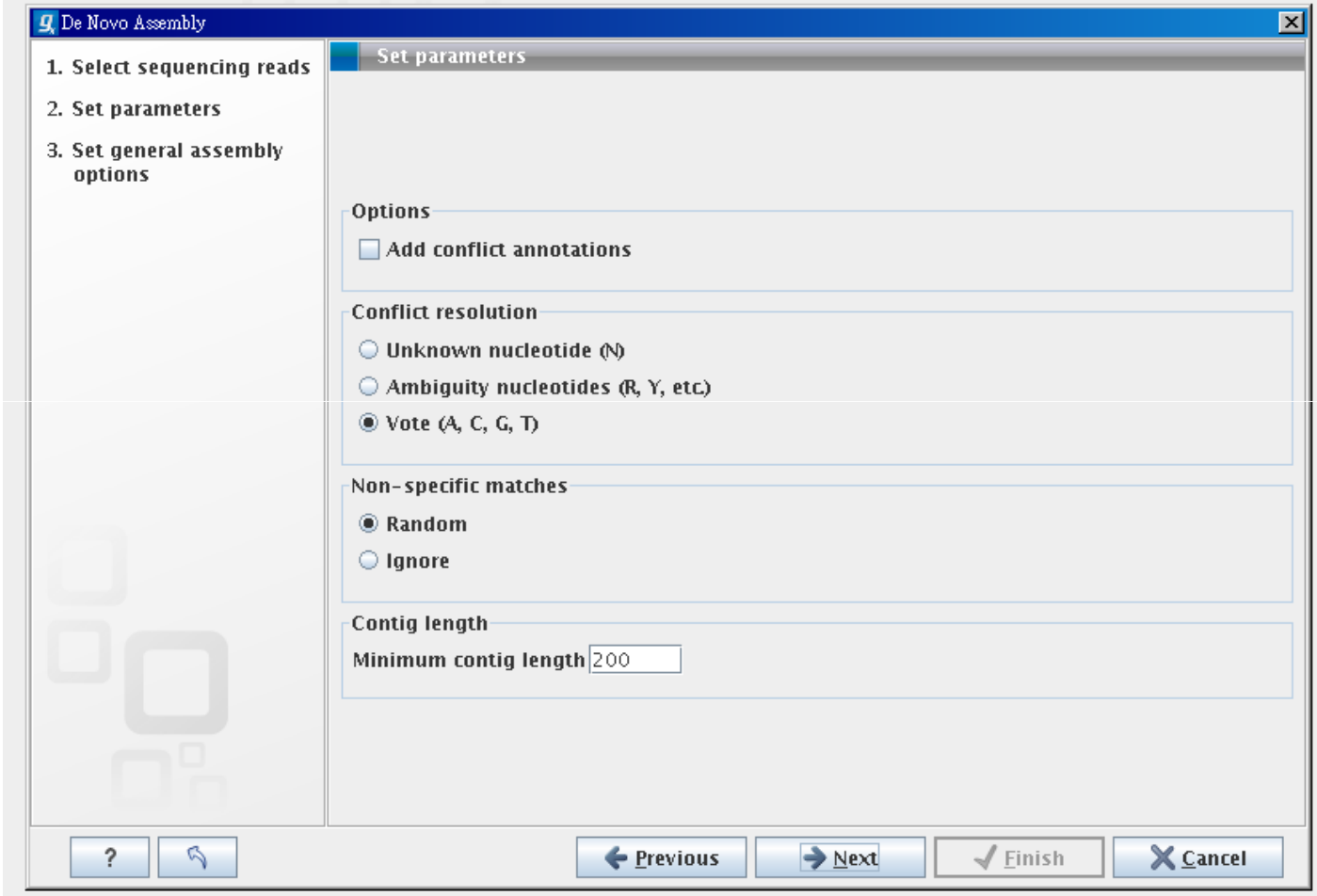
Show table

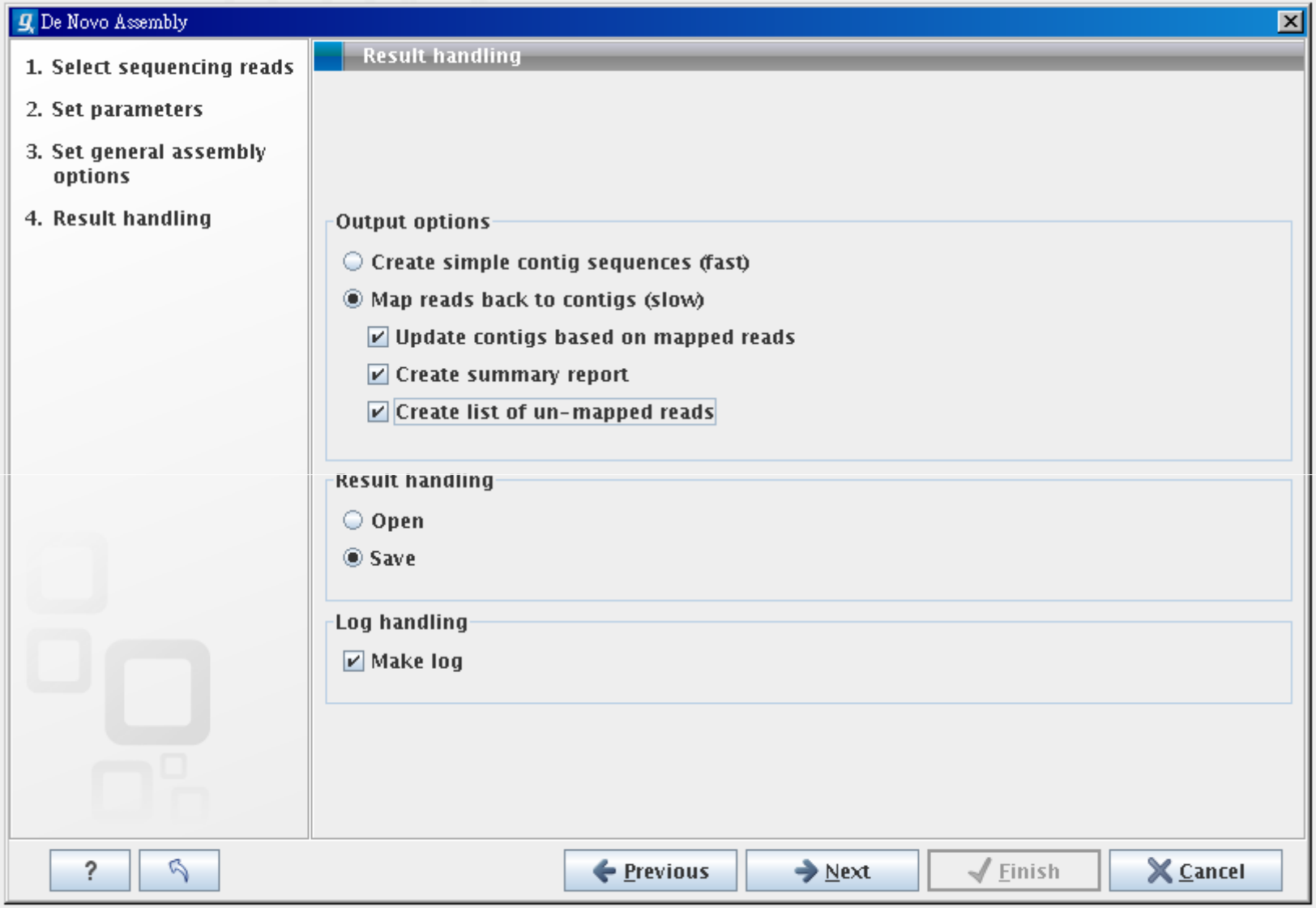

File Edit Search View Toolbox Workspace Help
testread
testread
Rows: 99,987
Sequence list: testread
Filter:
Name
Length
CLC_Data
Example Data
course110325
Recycle
Show
New
Show
Toolbox
Cut
Copy
Paste
Delete
Import...
Export...
Local Search...
Sort Folder
Update Folder
Restore
Empty Recycle Bin
Location
Sequence Representation
Graphical Sequence List
Cloning Editor
Gel
Annotation Table
Table
History
Element Info
Column width
Automatic
Show column
Name
Accession
Description
Latin name
Modification Date
Length
Select All
Deselect All
Toolbox
Alignments and Trees
General Sequence Analyses
Nucleotide Analyses
Protein Analyses
Sequencing Data Analyses
Primers and Probes
Cloning and Restriction Sites
RNA Structure
Expression Analysis
High-Throughput Sequencing
BLAST
Database Search
Create New Sequence List

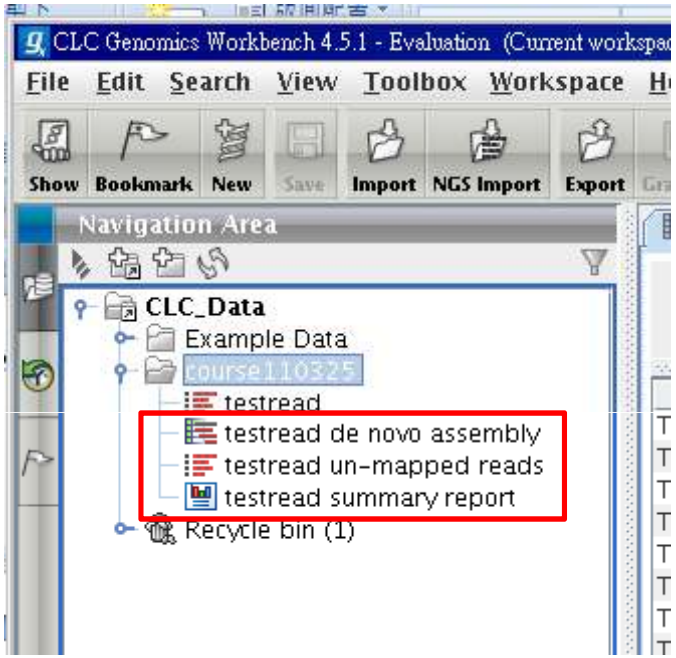
De Novo Assembly

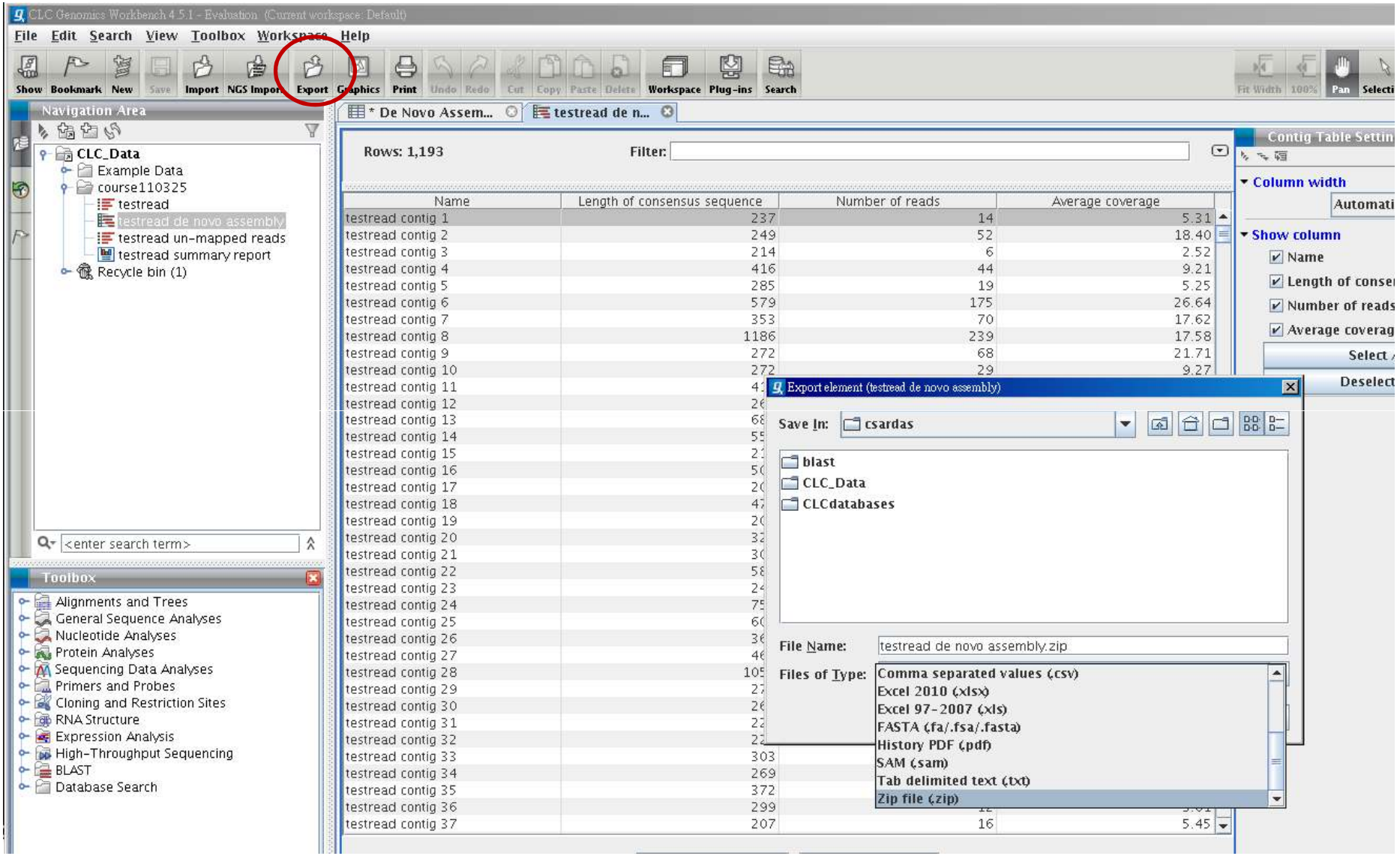



Rows: 99,987
Sequence list: testread
Filter:
De Novo Assembly
1. Select sequencing reads
2. Set parameters
Set parameters
Selected reads
Input
Length
Type
Settings
testread
Long
Single
Default
Long reads mapping parameters
Mismatch cost
2
Insertion cost
3
Deletion cost
3
Length fraction
0.5
Similarity
0.8
Global alignment
Color space alignment
Colorspace error cost
3
?
Previous
Next
Finish
Cancel
A80E9RABXX:2:1:19532...
A80E9RABXX:2:1:19592...

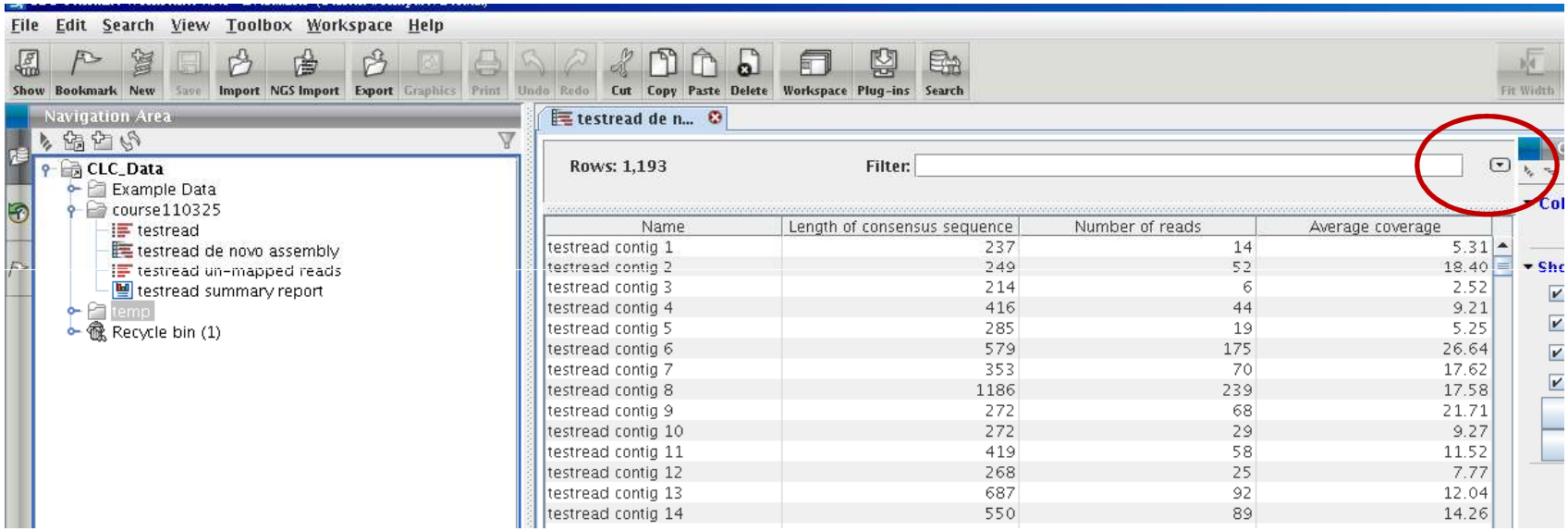

De Novo Assembly
1. Select sequencing reads
2. Set parameters
3. Set general assembly options
Set parameters
Options
Add conflict annotations
Conflict resolution
Unknown nucleotide (N)
Ambiguity nucleotides (R, Y, etc.)
Vote (A, C, G, T)
Non-specific matches
Random
Ignore
Contig length
Minimum contig length 200
?
Previous
Next
Finish
Cancel


De Novo Assembly
1. Select sequencing reads
2. Set parameters
3. Set general assembly options
4. Result handling
Result handling
Output options
Create simple contig sequences (fast)
Map reads back to contigs (slow)
Update contigs based on mapped reads
Create summary report
Create list of un-mapped reads
Result handling
Open
Save
Log handling
Make log
?
Previous
Next
Finish
Cancel

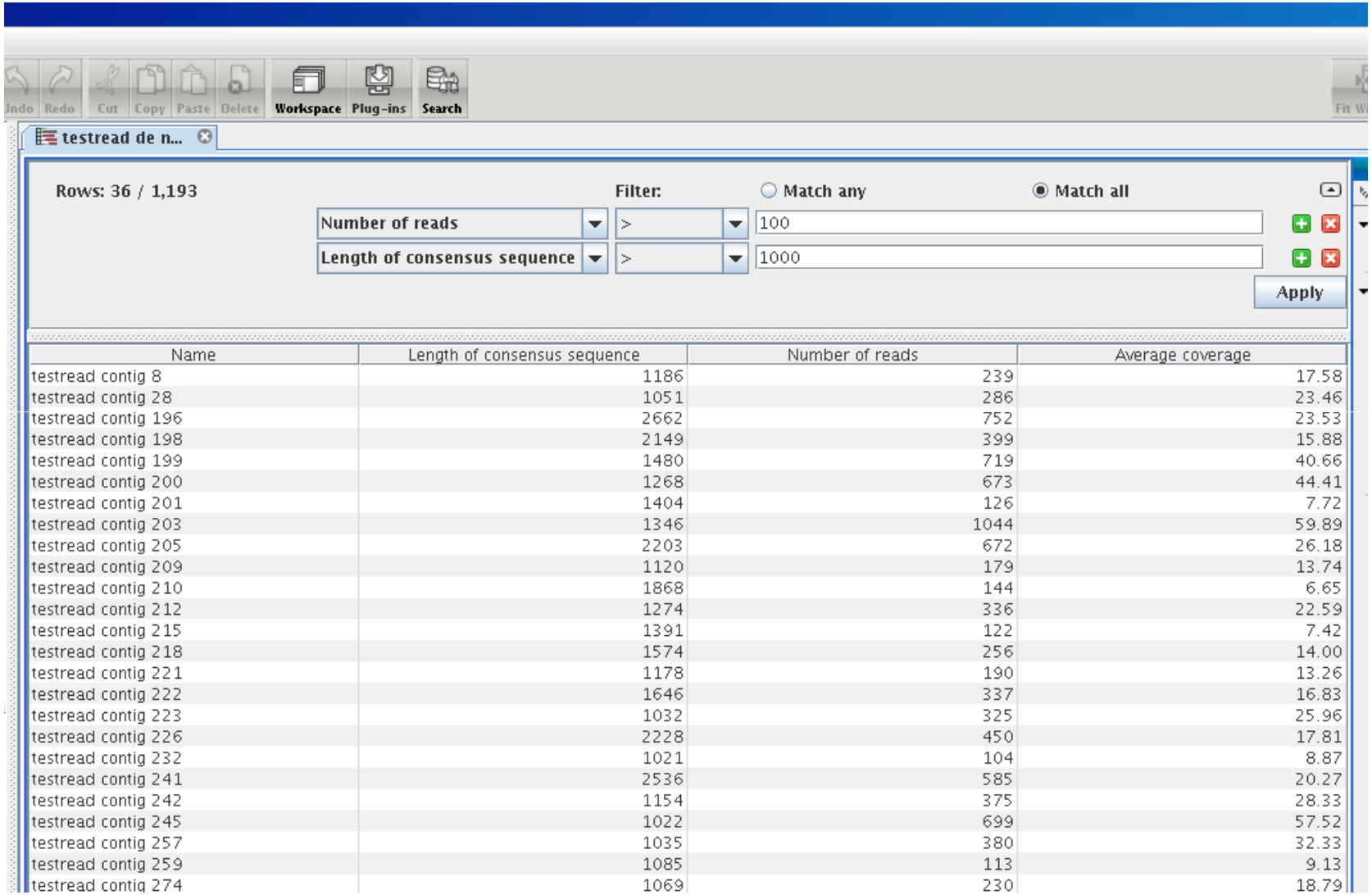

CLC Genomics Workbench 4.5.1 - Evaluation (Current workspa
File Edit Search View Toolbox Workspace H
Show Bookmark New Save Import NGS Import Export
Navigation Area
CLC_Data
Example Data
course110325
testread
testread de novo assembly
testread un-mapped reads
testread summary report
Recycle bin (1)


CLC Genomics Workbench 4.5.1 - Evaluation (Current workspace: Default)
File Edit Search View Toolbox Workspace Help
Show Bookmark New Save Import NGS Import Export Graphics Print Undo Redo Cut Copy Paste Delete Workspace Plug-ins Search
Navigation Area
CLC_Data
Example Data
course110325
testread
testread de novo assembly
testread un-mapped reads
testread summary report
Recycle bin (1)
<enter search term>
Toolbox
Alignments and Trees
General Sequence Analyses
Nucleotide Analyses
Protein Analyses
Sequencing Data Analyses
Primers and Probes
Cloning and Restriction Sites
RNA Structure
Expression Analysis
High-Throughput Sequencing
BLAST
Database Search
* De Novo Assem...
testread de n...
Rows: 1,193
Filter:
Name
Length of consensus sequence
Number of reads
Average coverage
testread contig 1 237 14 5.31
testread contig 2 249 52 18.40
testread contig 3 214 6 2.52
testread contig 4 416 44 9.21
testread contig 5 285 19 5.25
testread contig 6 579 175 26.64
testread contig 7 353 70 17.62
testread contig 8 1186 239 17.58
testread contig 9 272 68 21.71
testread contig 10 272 29 9.27
testread contig 37 207 16 5.45
Contig Table Settin
Column width
Automati
Show column
Name
Length of conse
Number of reads
Average coverag
Select
Deselect
Export element (testread de novo assembly)
Save In: csardas
blast
CLC_Data
CLCdatabases
File Name: testread de novo assembly.zip
Files of Type:
Comma separated values (csv)
Excel 2010 (xlsx)
Excel 97-2007 (xls)
FASTA (fa/.fsa/.fasta)
History PDF (pdf)
SAM (sam)
Tab delimited text (txt)
Zip file (zip)

filter

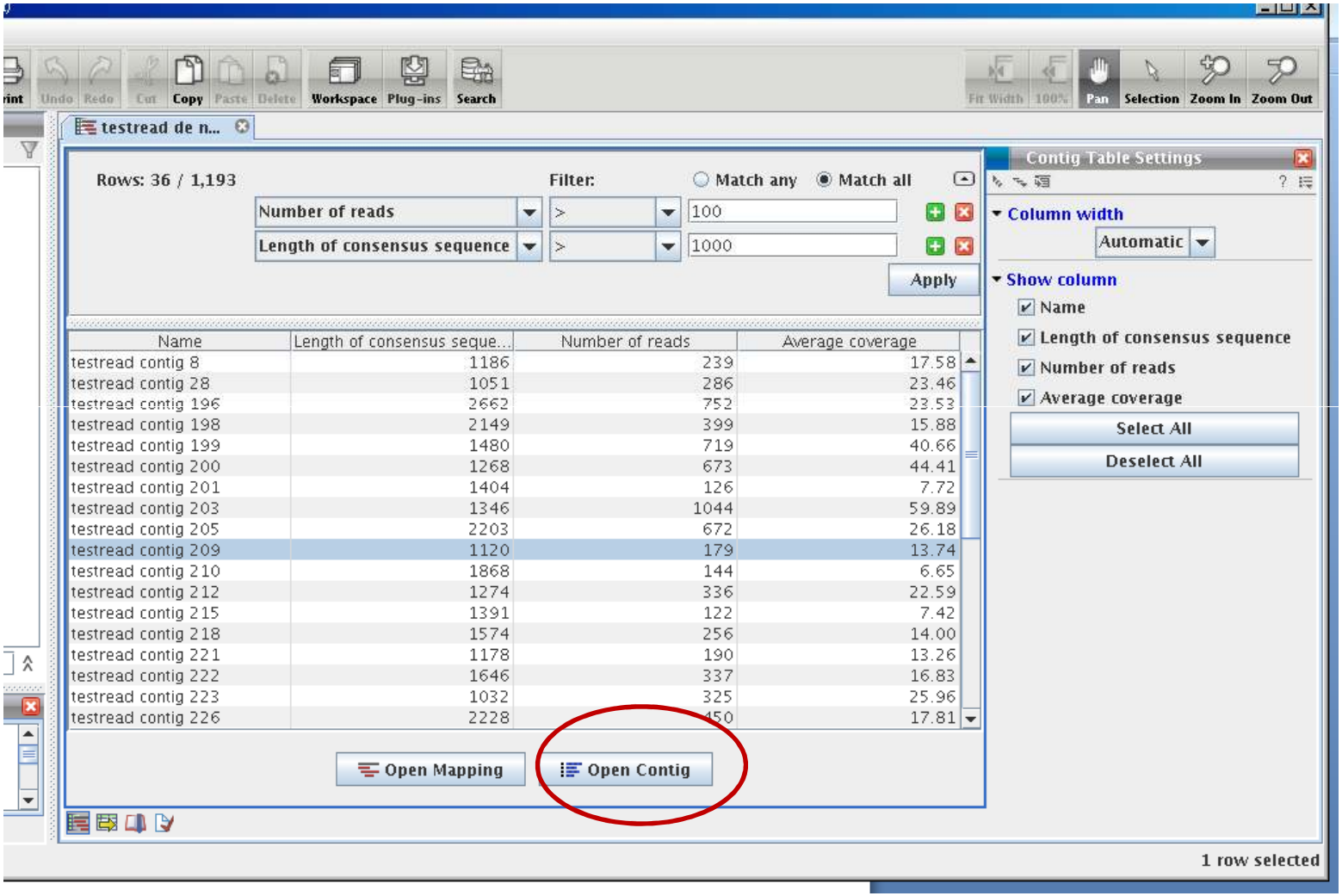


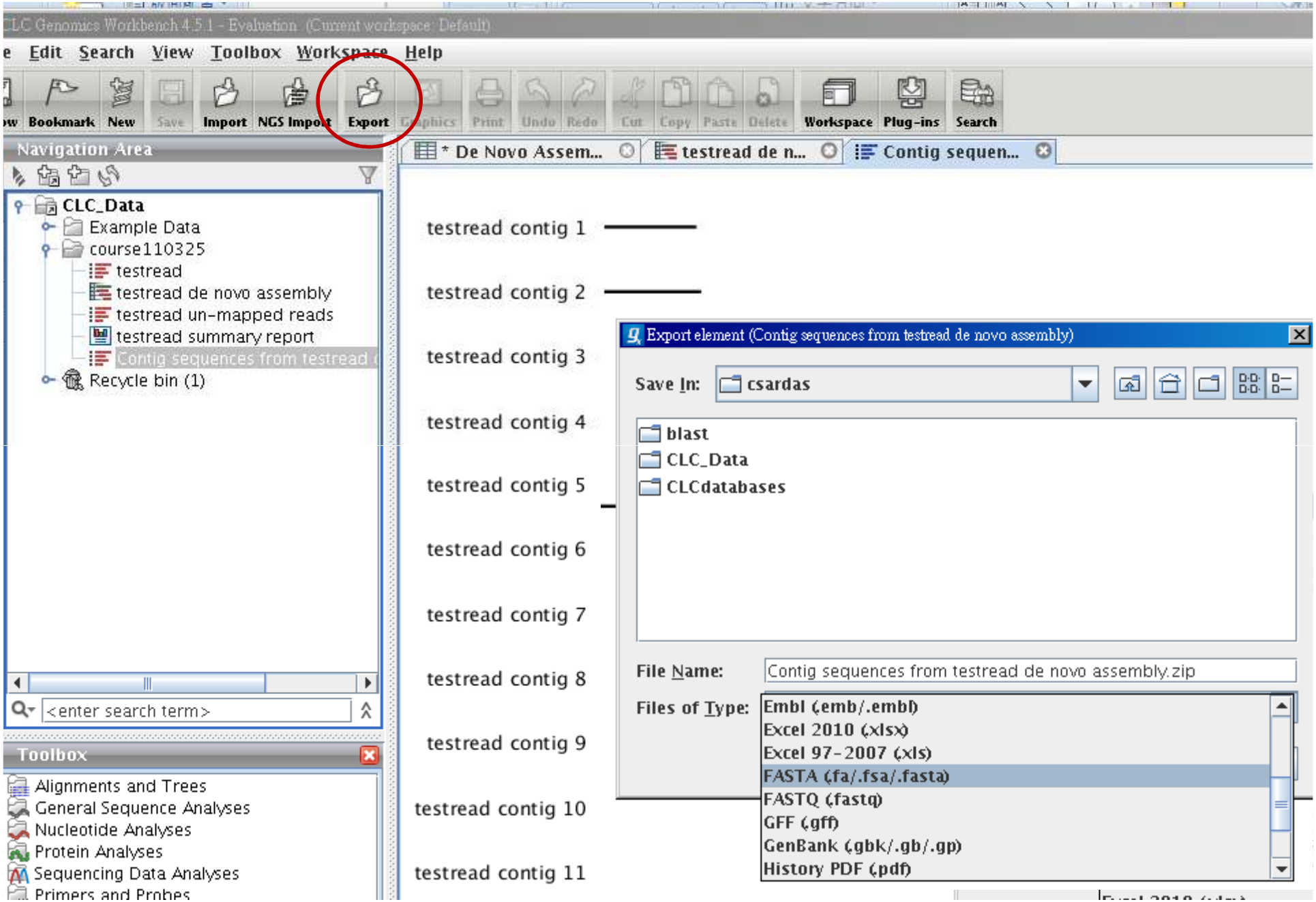
Name	Length of consensus sequence	Number of reads	Average coverage
testread contig 1	237	14	5.31
testread contig 2	249	52	18.40
testread contig 3	214	6	2.52
testread contig 4	416	44	9.21
testread contig 5	285	19	5.25
testread contig 6	579	175	26.64
testread contig 7	353	70	17.62
testread contig 8	1186	239	17.58
testread contig 9	272	68	21.71
testread contig 10	272	29	9.27
testread contig 11	419	58	11.52
testread contig 12	268	25	7.77
testread contig 13	687	92	12.04
testread contig 14	550	89	14.26



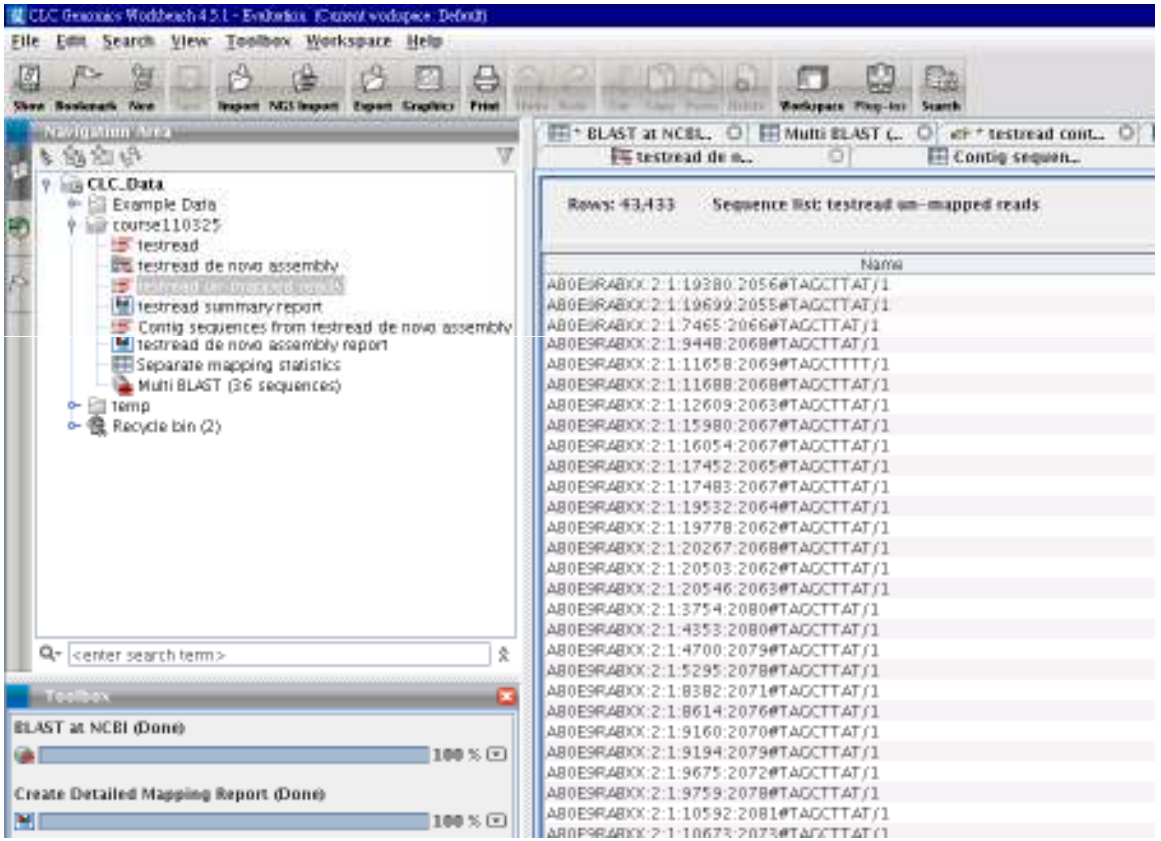
Name	Length of consensus sequence	Number of reads	Average coverage
testread contig 8	1186	239	17.58
testread contig 28	1051	286	23.46
testread contig 196	2662	752	23.53
testread contig 198	2149	399	15.88
testread contig 199	1480	719	40.66
testread contig 200	1268	673	44.41
testread contig 201	1404	126	7.72
testread contig 203	1346	1044	59.89
testread contig 205	2203	672	26.18
testread contig 209	1120	179	13.74
testread contig 210	1868	144	6.65
testread contig 212	1274	336	22.59
testread contig 215	1391	122	7.42
testread contig 218	1574	256	14.00
testread contig 221	1178	190	13.26
testread contig 222	1646	337	16.83
testread contig 223	1032	325	25.96
testread contig 226	2228	450	17.81
testread contig 232	1021	104	8.87
testread contig 241	2536	585	20.27
testread contig 242	1154	375	28.33
testread contig 245	1022	699	57.52
testread contig 257	1035	380	32.33
testread contig 259	1085	113	9.13
testread contig 274	1069	230	18.79

Use “open contig” to create sequence list

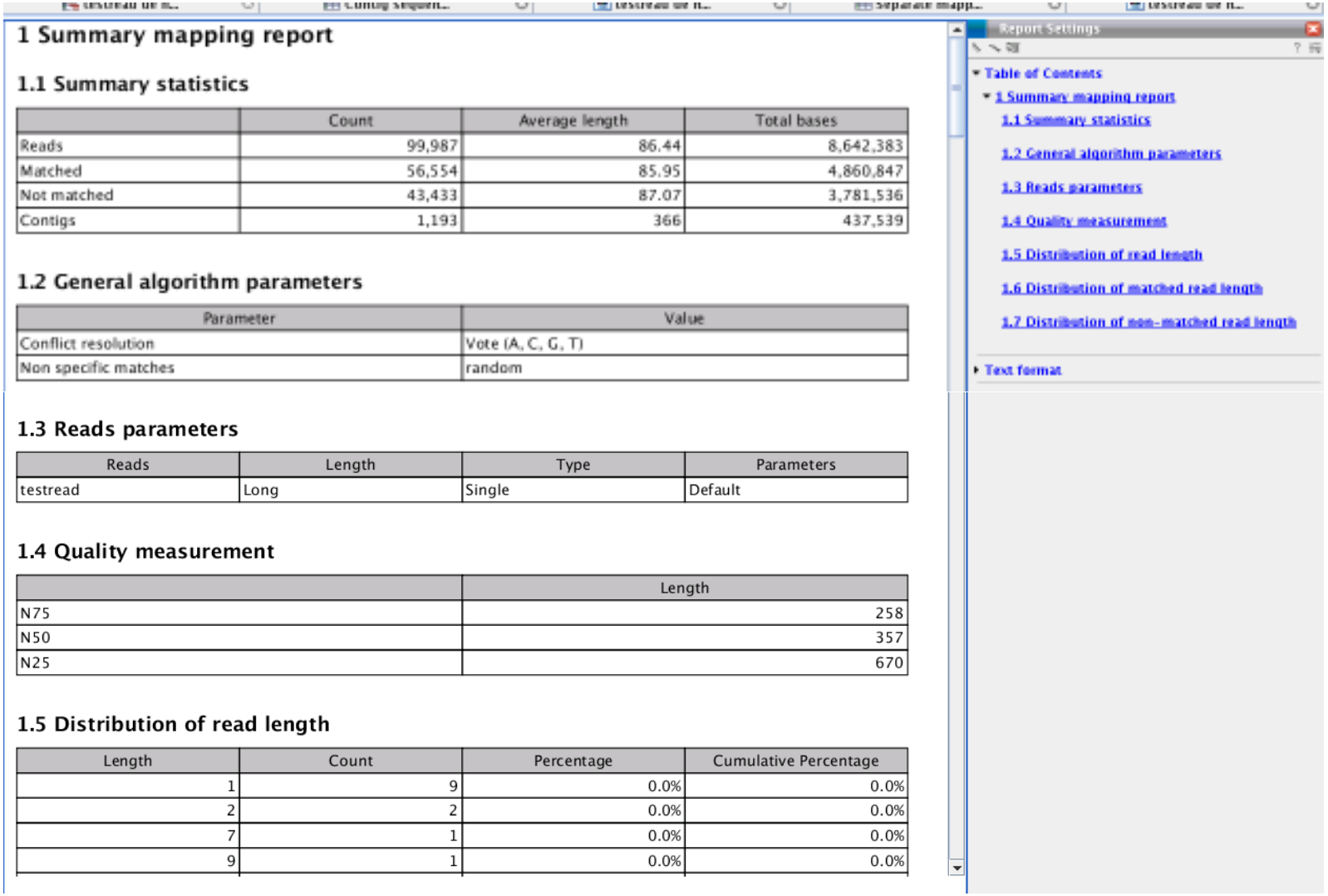



CLC Genomics Workbench 4.5.1 - Evaluation (Current workspace: Default)
Edit Search View Toolbox Workspace Help
Bookmark New Save Import NGS Import Export Graphics Print Undo Redo Cut Copy Paste Delete Workspace Plug-ins Search
Navigation Area
CLC_Data
Example Data
course110325
testread
testread de novo assembly
testread un-mapped reads
testread summary report
Contig sequences from testread
Recycle bin (1)
<enter search term>
Toolbox
Alignments and Trees
General Sequence Analyses
Nucleotide Analyses
Protein Analyses
Sequencing Data Analyses
Primers and Probes
* De Novo Assem...
testread de n...
Contig sequen...
testread contig 1
testread contig 2
testread contig 3
testread contig 4
testread contig 5
testread contig 6
testread contig 7
testread contig 8
testread contig 9
testread contig 10
testread contig 11
Export element (Contig sequences from testread de novo assembly)
Save In: csardas
blast
CLC_Data
CLCdatabases
File Name: Contig sequences from testread de novo assembly.zip
Files of Type:
Embl (.emb/.embl)
Excel 2010 (.xlsx)
Excel 97-2007 (.xls)
FASTA (.fa/.fsa/.fasta)
FASTQ (.fastq)
GFF (.gff)
GenBank (.gbk/.gb/.gp)
History PDF (.pdf)

Un mapped reads



Summary report



1 Summary mapping report

1.1 Summary statistics

	Count	Average length	Total bases
Reads	99,987	86.44	8,642,383
Matched	56,554	85.95	4,860,847
Not matched	43,433	87.07	3,781,536
Contigs	1,193	366	437,539

1.2 General algorithm parameters

Parameter	Value
Conflict resolution	Vote (A, C, G, T)
Non specific matches	random

1.3 Reads parameters

Reads	Length	Type	Parameters
testread	Long	Single	Default

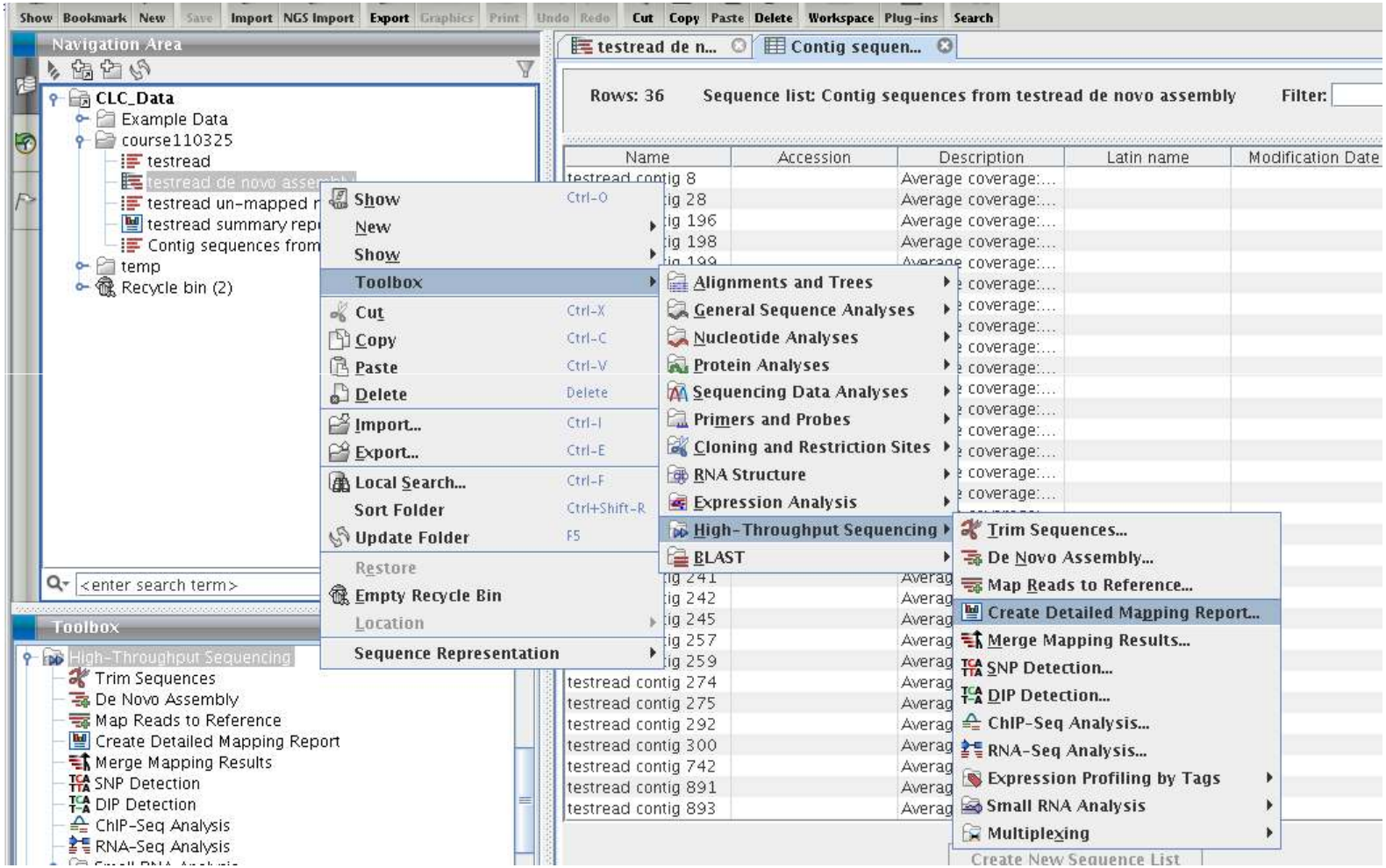
1.4 Quality measurement

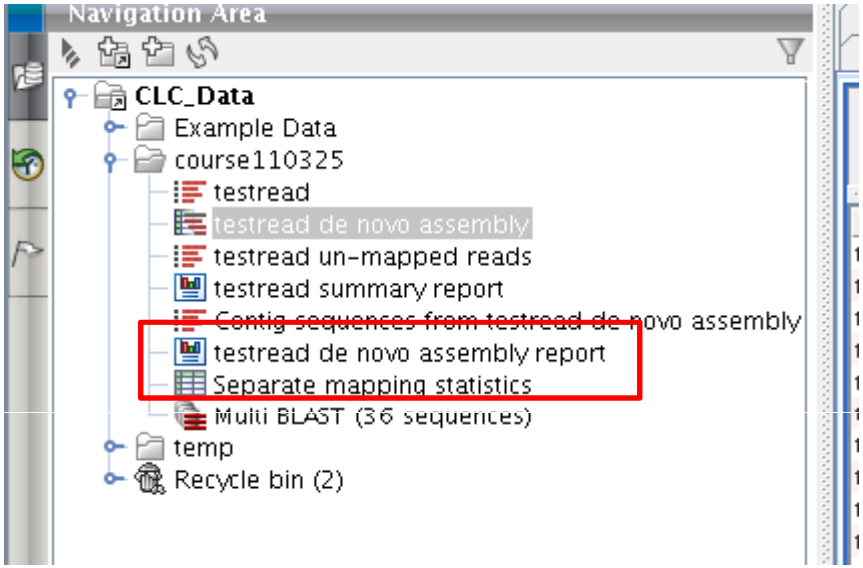
	Length
N75	258
N50	357
N25	670

1.5 Distribution of read length

Length	Count	Percentage	Cumulative Percentage
1	9	0.0%	0.0%
2	2	0.0%	0.0%
7	1	0.0%	0.0%
9	1	0.0%	0.0%

Detail mapping report




Navigation Area
CLC_Data
Example Data
course110325
testread
testread de novo assembly
testread un-mapped reads
testread summary report
Contig sequences from testread de novo assembly
testread de novo assembly report
Separate mapping statistics
Multi BLAST (36 sequences)
temp
Recycle bin (2)

1 Summary

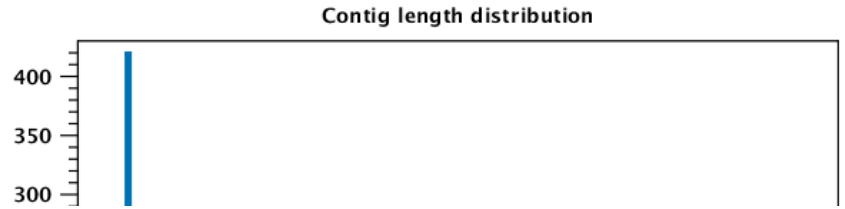
Contig count	1,193
Type	De novo assembly
Total read count	56,554
Mean read length	85.95
Total read length	4,860,847
Mean contig length	366
Total contig length	437,539
GC contents in %	44.49

2 Contig sequences

2.1 Contigs

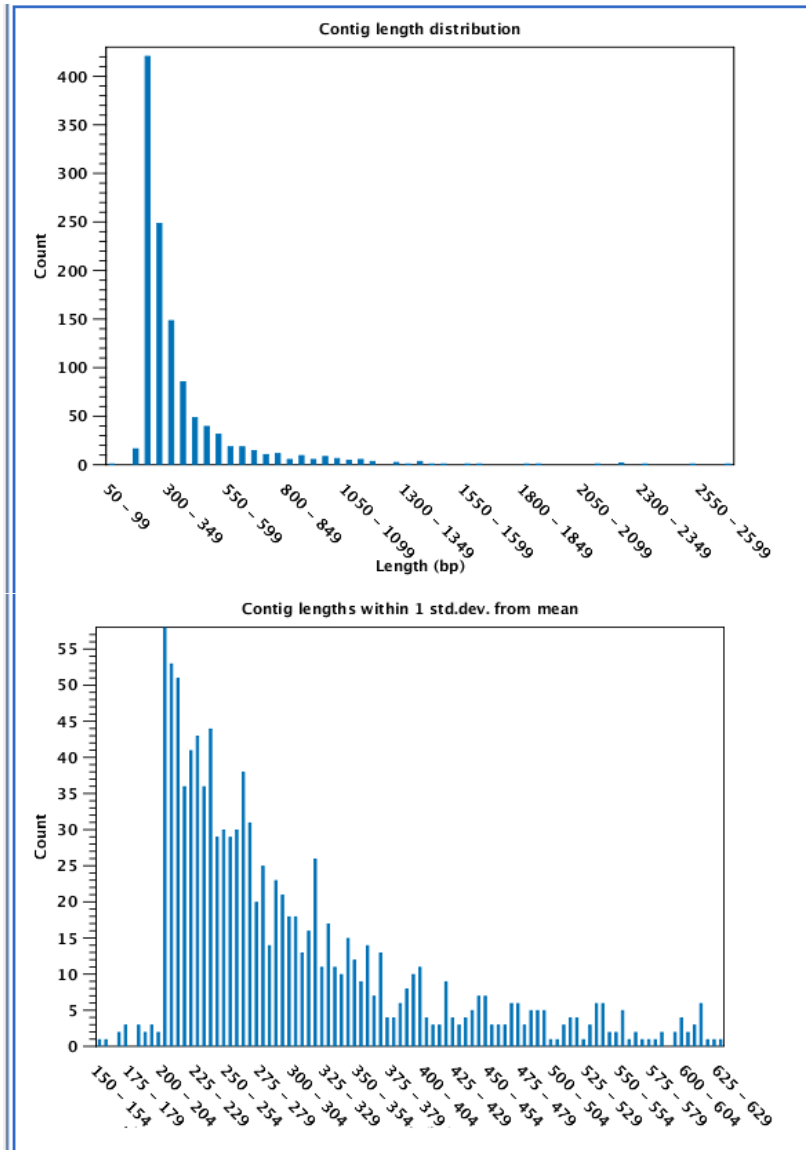
Contig set	Contig count	Min contig length	Max contig length	Mean contig length	Standard deviation	Total contig length	% GC
N25 contigs	105	670	2,662	1,045.89	407.73	109,818	46.60
N50 contigs	338	357	2,662	647.72	357.68	218,931	45.64
N75 contigs	704	258	2,662	466.36	303.69	328,315	45.08
All contigs	1,193	91	2,662	366.76	262.42	437,539	44.49
Long contigs (>10,000bp)	0						
Short contigs (<200bp)	29	91	200	184.34	22.62	5,346	41.51

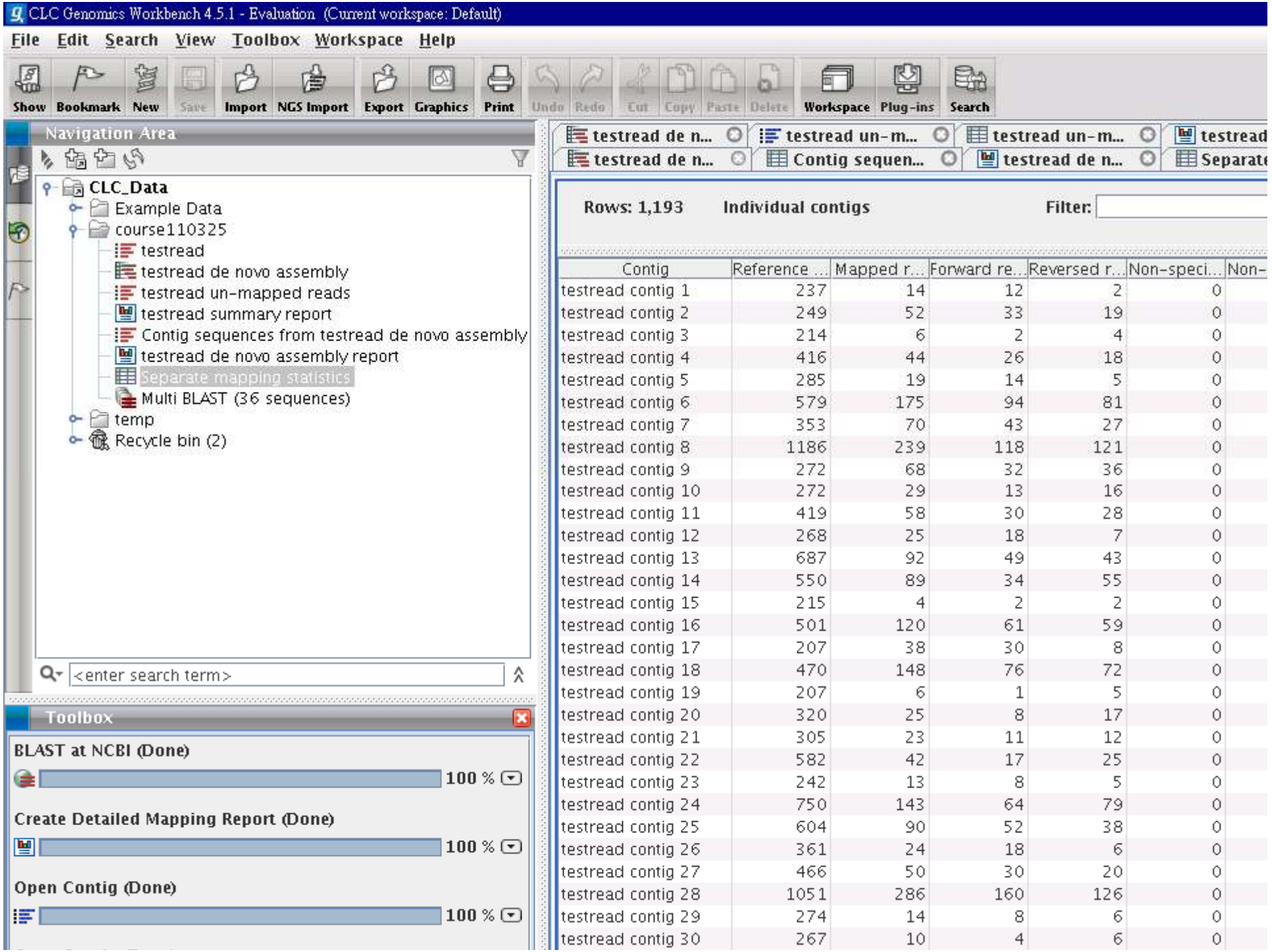
2.2 Contig length distribution



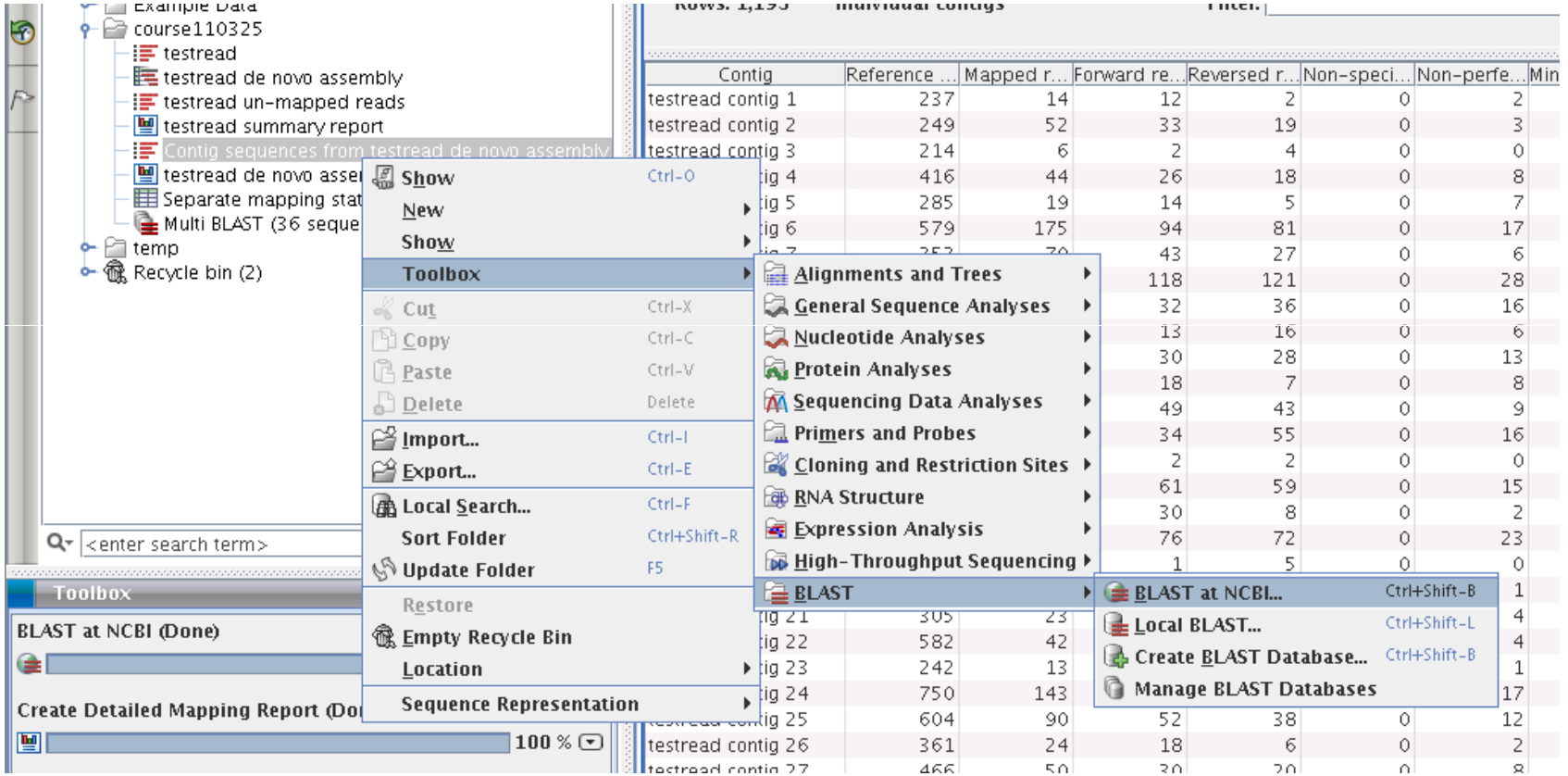
Report Settings

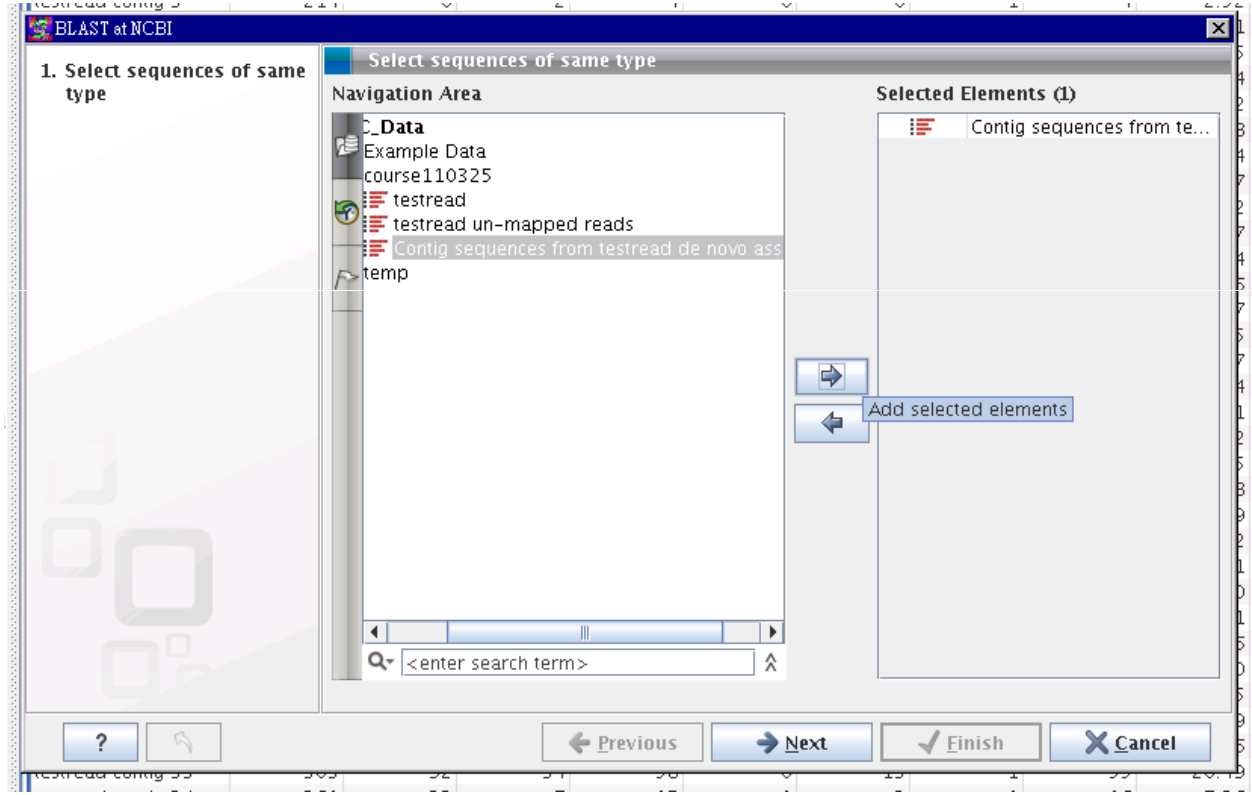
Table of Contents
1 Summary
2 Contig sequences
2.1 Contigs
2.2 Contig length distribution
2.3 Coverage level distribution
3 Mapped reads
3.1 All mapped reads
3.2 Non-specific matches
3.2.1 Non-specific matches
3.2.2 Non-specific distribution
3.2.3 Non-specific distribution (1-10 matches)
3.3 Non-perfect matches
3.3.1 Non-perfect matches
3.3.2 Non-perfect distribution
3.4 Read length distribution
Text format

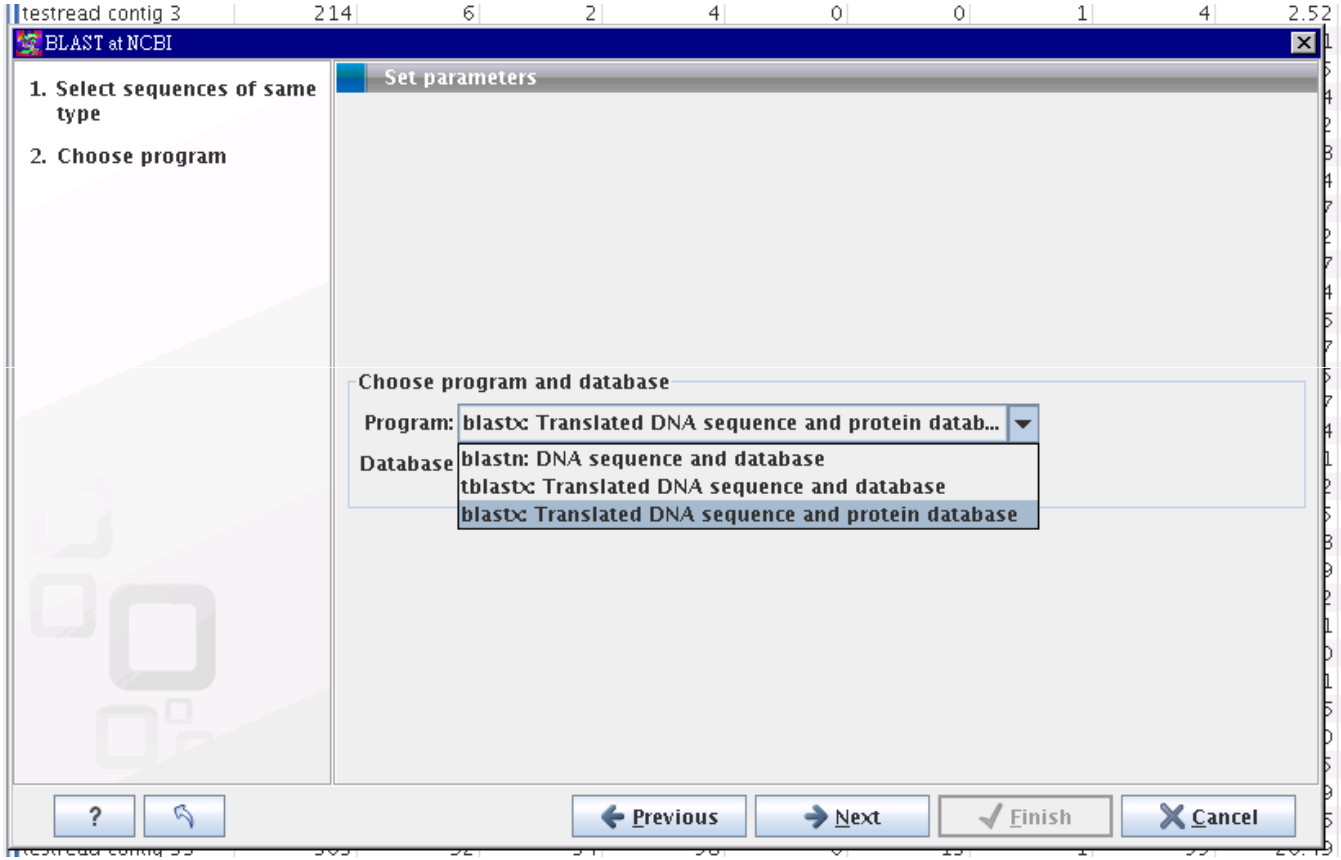

Contig length distribution
Count
Length (bp)
50 - 99
300 - 349
550 - 599
800 - 849
1050 - 1099
1300 - 1349
1550 - 1599
1800 - 1849
2050 - 2099
2300 - 2349
2550 - 2599
Contig lengths within 1 std.dev. from mean
Count
150 - 154
175 - 179
200 - 204
225 - 229
250 - 254
275 - 279
300 - 304
325 - 329
350 - 354
375 - 379
400 - 404
425 - 429
450 - 454
475 - 479
500 - 504
525 - 529
550 - 554
575 - 579
600 - 604
625 - 629

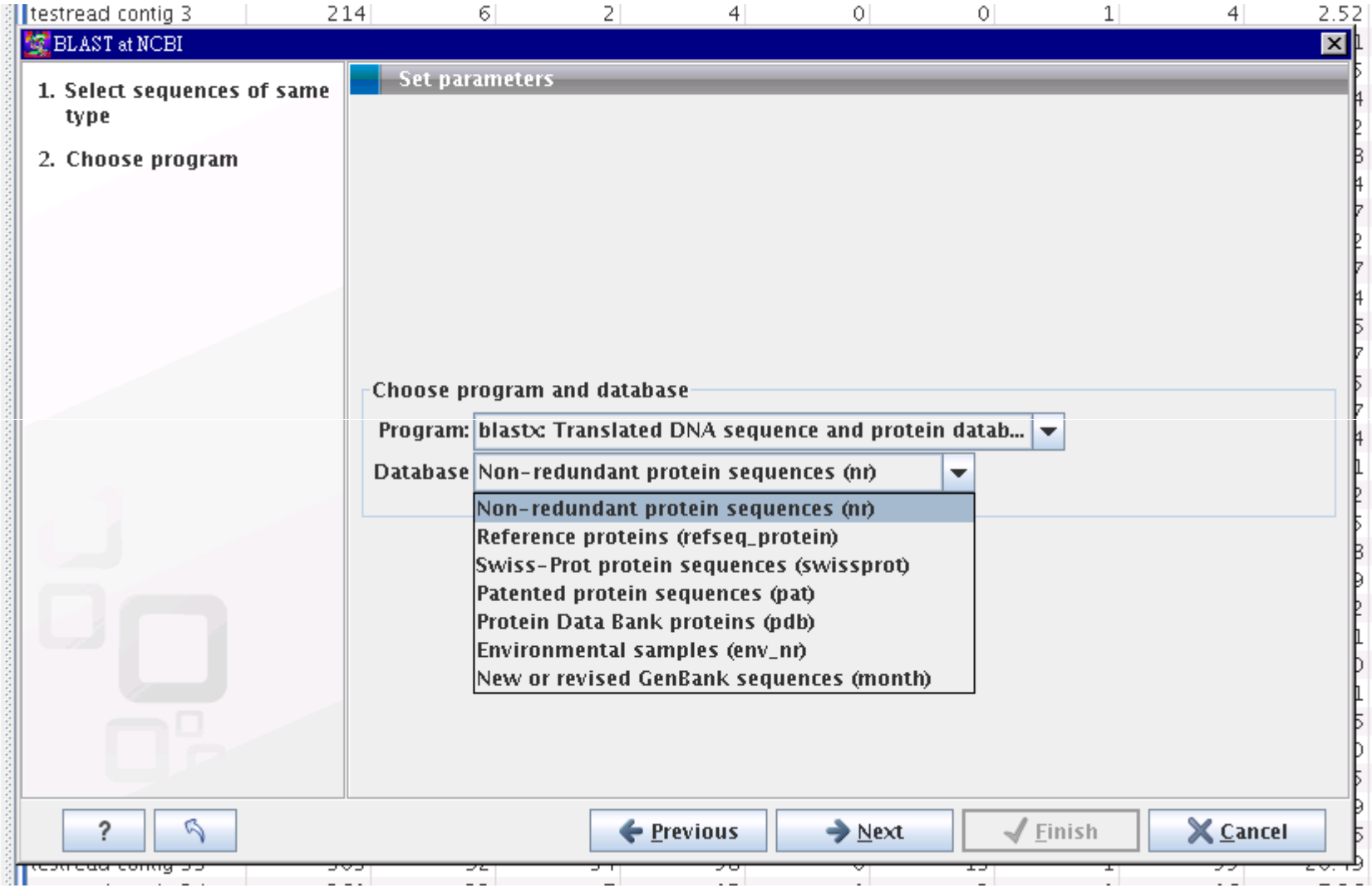

CLC Genomics Workbench 4.5.1 - Evaluation (Current workspace: Default)
File Edit Search View Toolbox Workspace Help
Show Bookmark New Save Import NGS Import Export Graphics Print Undo Redo Cut Copy Paste Delete Workspace Plug-ins Search
Navigation Area
CLC_Data
Example Data
course110325
testread
testread de novo assembly
testread un-mapped reads
testread summary report
Contig sequences from testread de novo assembly
testread de novo assembly report
Separate mapping statistics
Multi BLAST (36 sequences)
temp
Recycle bin (2)
<enter search term>
Toolbox
BLAST at NCBI (Done)
100 %
Create Detailed Mapping Report (Done)
100 %
Open Contig (Done)
100 %
testread de n...
testread un-m...
testread un-m...
testread de n...
Contig sequen...
testread de n...
Rows: 1,193
Individual contigs
Filter:
Contig
Reference ...
Mapped r...
Forward re...
Reversed r...
Non-speci...
testread contig 1 237 14 12 2 0
testread contig 2 249 52 33 19 0
testread contig 3 214 6 2 4 0
testread contig 4 416 44 26 18 0
testread contig 5 285 19 14 5 0
testread contig 6 579 175 94 81 0
testread contig 7 353 70 43 27 0
testread contig 8 1186 239 118 121 0
testread contig 9 272 68 32 36 0
testread contig 10 272 29 13 16 0
testread contig 11 419 58 30 28 0
testread contig 12 268 25 18 7 0
testread contig 13 687 92 49 43 0
testread contig 14 550 89 34 55 0
testread contig 15 215 4 2 2 0
testread contig 16 501 120 61 59 0
testread contig 17 207 38 30 8 0
testread contig 18 470 148 76 72 0
testread contig 19 207 6 1 5 0
testread contig 20 320 25 8 17 0
testread contig 21 305 23 11 12 0
testread contig 22 582 42 17 25 0
testread contig 23 242 13 8 5 0
testread contig 24 750 143 64 79 0
testread contig 25 604 90 52 38 0
testread contig 26 361 24 18 6 0
testread contig 27 466 50 30 20 0
testread contig 28 1051 286 160 126 0
testread contig 29 274 14 8 6 0
testread contig 30 267 10 4 6 0

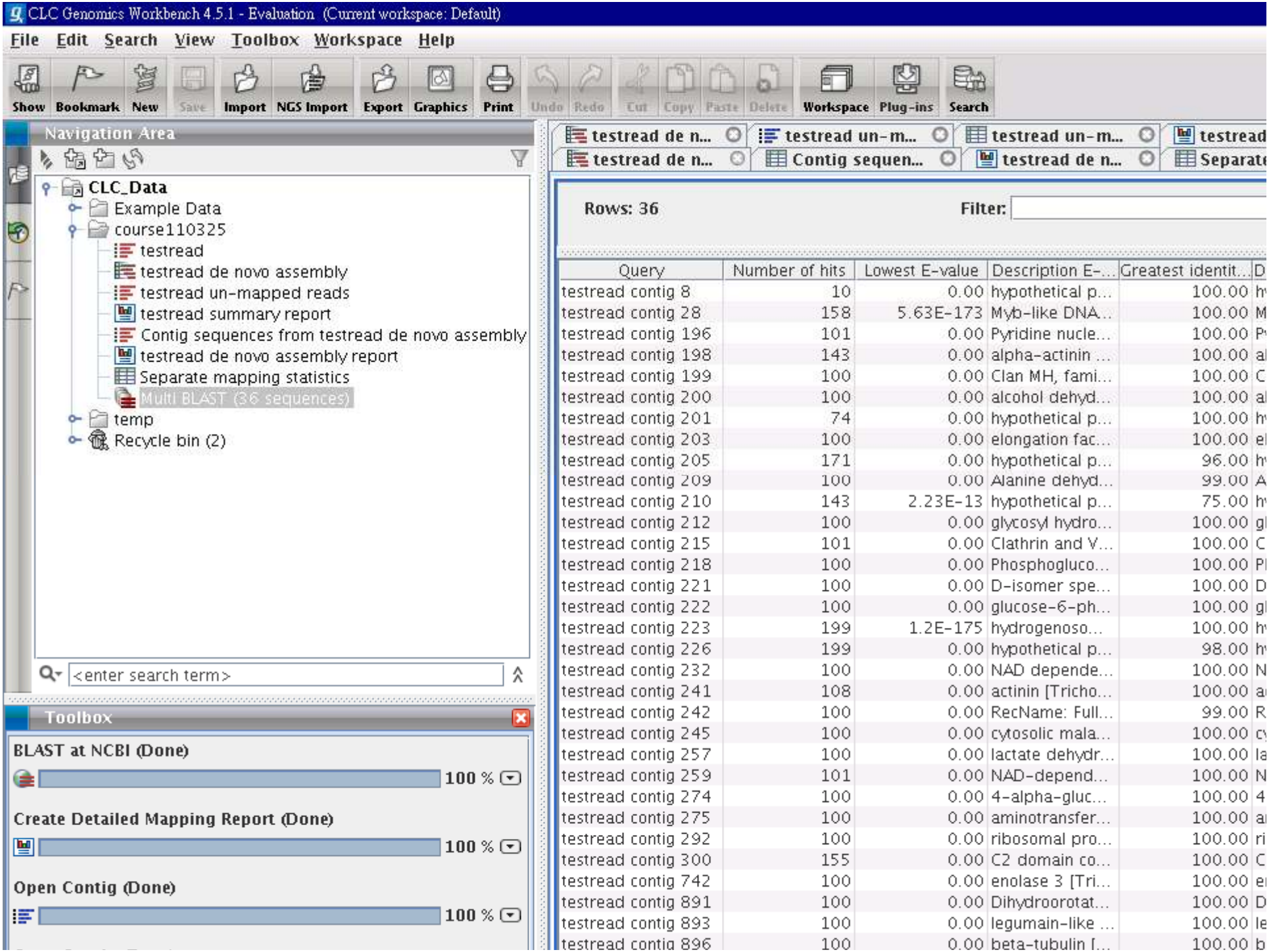
Blast to NCBI




BLAST at NCBI
1. Select sequences of same type
Select sequences of same type
Navigation Area
C_Data
Example Data
course110325
testread
testread un-mapped reads
Contig sequences from testread de novo ass
temp
<enter search term>
Selected Elements (1)
Contig sequences from te...
Add selected elements
?
Previous
Next
Finish
Cancel

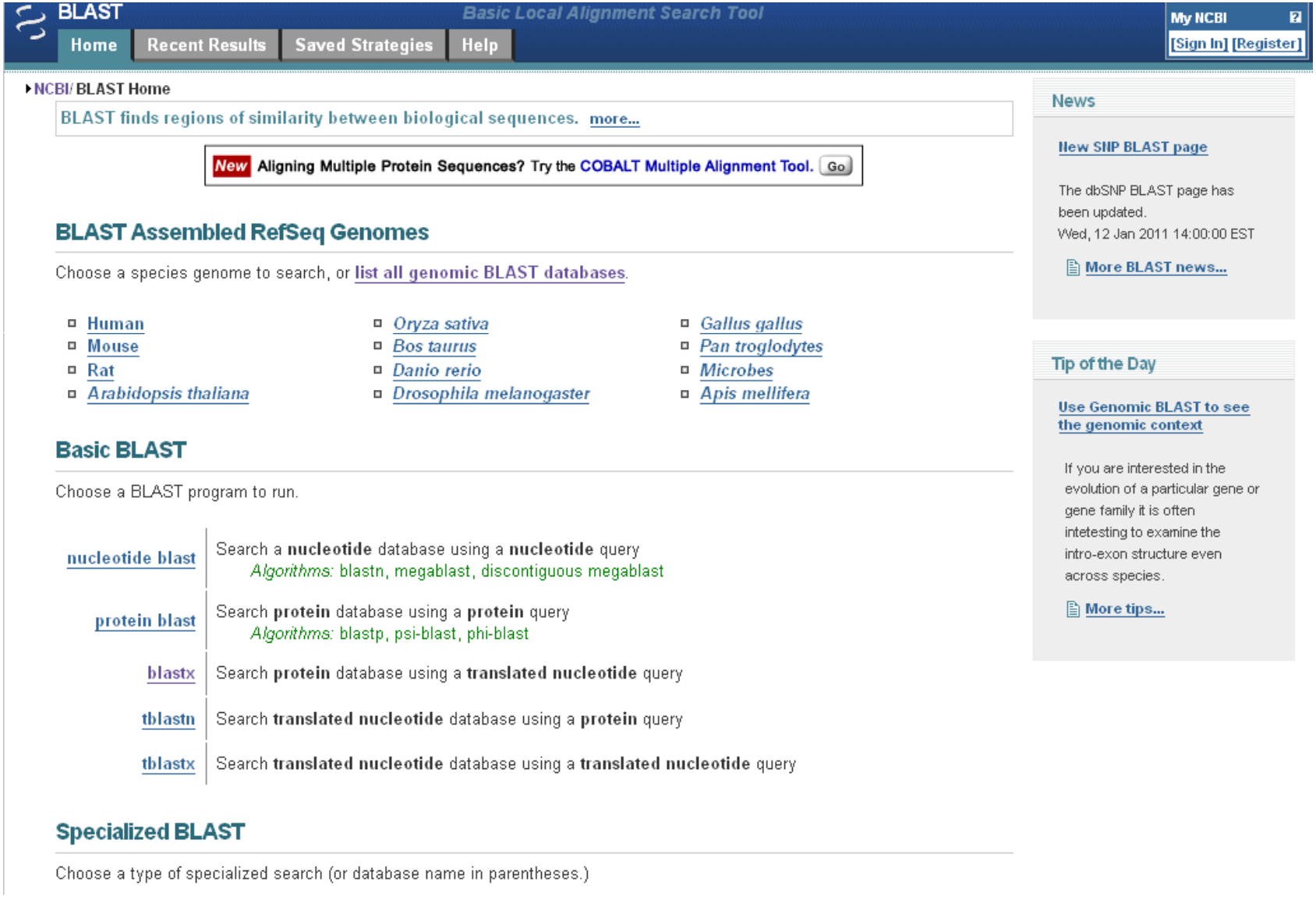

testread contig 3
214
6
2
4
0
0
1
4
2.52
BLAST at NCBI
1. Select sequences of same type
2. Choose program
Set parameters
Choose program and database
Program:
blastx Translated DNA sequence and protein datab...
Database
blastn: DNA sequence and database
tblastx Translated DNA sequence and database
blastx Translated DNA sequence and protein database
?
Previous
Next
Finish
Cancel

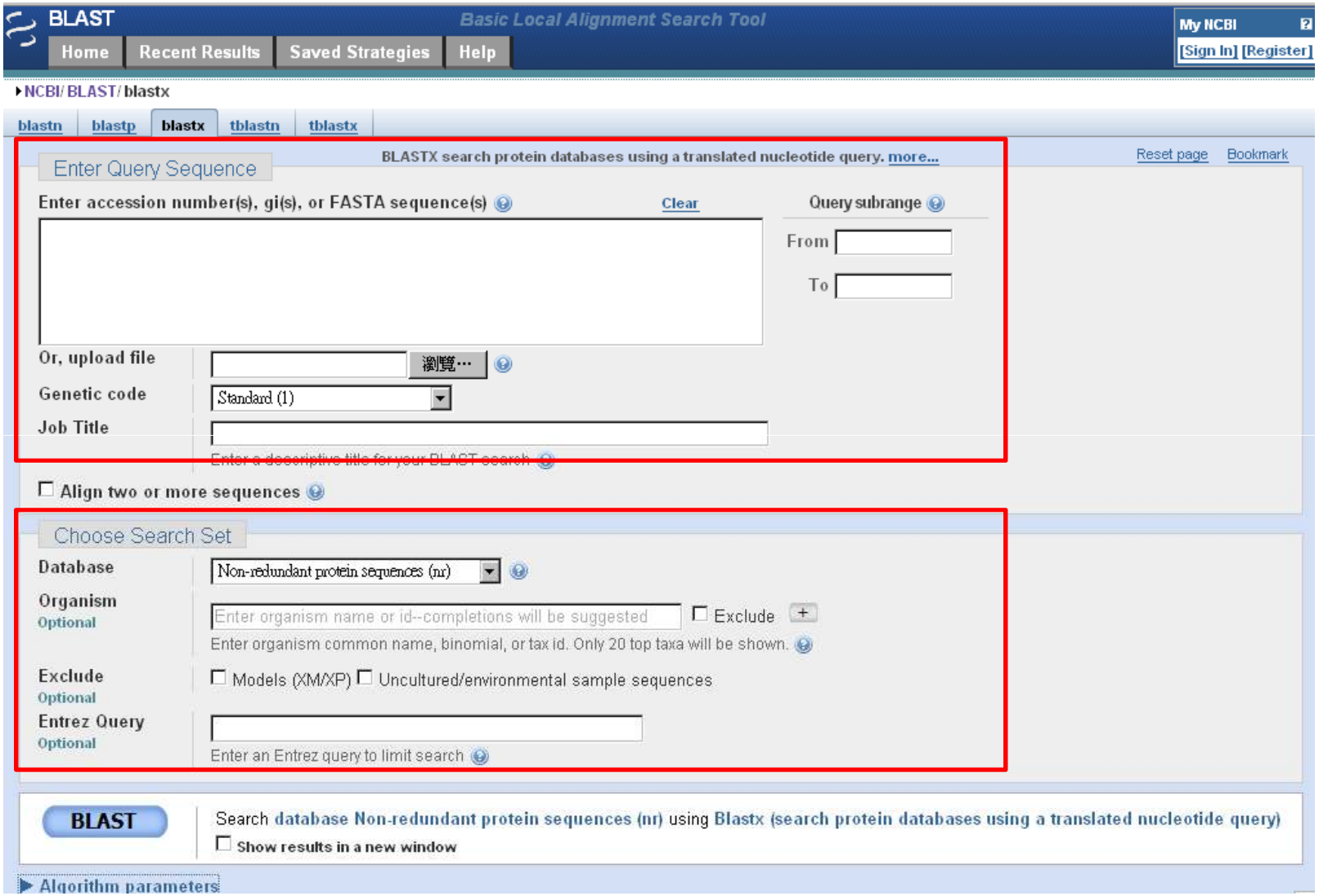

testread contig 3 214 6 2 4 0 0 1 4 2.52
BLAST at NCBI
1. Select sequences of same type
2. Choose program
Set parameters
Choose program and database
Program: blastx: Translated DNA sequence and protein datab...
Database Non-redundant protein sequences (nr)
Non-redundant protein sequences (nr)
Reference proteins (refseq_protein)
Swiss-Prot protein sequences (swissprot)
Patented protein sequences (pat)
Protein Data Bank proteins (pdb)
Environmental samples (env_nr)
New or revised GenBank sequences (month)
?
Previous
Next
Finish
Cancel


CLC Genomics Workbench 4.5.1 - Evaluation (Current workspace: Default)
File Edit Search View Toolbox Workspace Help
Show Bookmark New Save Import NGS Import Export Graphics Print Undo Redo Cut Copy Paste Delete Workspace Plug-ins Search
Navigation Area
CLC_Data
Example Data
course110325
testread
testread de novo assembly
testread un-mapped reads
testread summary report
Contig sequences from testread de novo assembly
testread de novo assembly report
Separate mapping statistics
Multi BLAST (36 sequences)
temp
Recycle bin (2)
<enter search term>
Toolbox
BLAST at NCBI (Done)
100 %
Create Detailed Mapping Report (Done)
100 %
Open Contig (Done)
100 %
testread de n...
testread un-m...
testread un-m...
testread de n...
Contig sequen...
testread de n...
Rows: 36
Filter:
Query | Number of hits | Lowest E-value | Description E-... | Greatest identit...
testread contig 8 | 10 | 0.00 | hypothetical p... | 100.00
testread contig 28 | 158 | 5.63E-173 | Myb-like DNA... | 100.00
testread contig 196 | 101 | 0.00 | Pyridine nucle... | 100.00
testread contig 198 | 143 | 0.00 | alpha-actinin ... | 100.00
testread contig 199 | 100 | 0.00 | Clan MH, fami... | 100.00
testread contig 200 | 100 | 0.00 | alcohol dehyd... | 100.00
testread contig 201 | 74 | 0.00 | hypothetical p... | 100.00
testread contig 203 | 100 | 0.00 | elongation fac... | 100.00
testread contig 205 | 171 | 0.00 | hypothetical p... | 96.00
testread contig 209 | 100 | 0.00 | Alanine dehyd... | 99.00
testread contig 210 | 143 | 2.23E-13 | hypothetical p... | 75.00
testread contig 212 | 100 | 0.00 | glycosyl hydro... | 100.00
testread contig 215 | 101 | 0.00 | Clathrin and V... | 100.00
testread contig 218 | 100 | 0.00 | Phosphogluco... | 100.00
testread contig 221 | 100 | 0.00 | D-isomer spe... | 100.00
testread contig 222 | 100 | 0.00 | glucose-6-ph... | 100.00
testread contig 223 | 199 | 1.2E-175 | hydrogenoso... | 100.00
testread contig 226 | 199 | 0.00 | hypothetical p... | 98.00
testread contig 232 | 100 | 0.00 | NAD depende... | 100.00
testread contig 241 | 108 | 0.00 | actinin [Tricho... | 100.00
testread contig 242 | 100 | 0.00 | RecName: Full... | 99.00
testread contig 245 | 100 | 0.00 | cytosolic mala... | 100.00
testread contig 257 | 100 | 0.00 | lactate dehydr... | 100.00
testread contig 259 | 101 | 0.00 | NAD-depend... | 100.00
testread contig 274 | 100 | 0.00 | 4-alpha-gluc... | 100.00
testread contig 275 | 100 | 0.00 | aminotransfer... | 100.00
testread contig 292 | 100 | 0.00 | ribosomal pro... | 100.00
testread contig 300 | 155 | 0.00 | C2 domain co... | 100.00
testread contig 742 | 100 | 0.00 | enolase 3 [Tri... | 100.00
testread contig 891 | 100 | 0.00 | Dihydroorotat... | 100.00
testread contig 893 | 100 | 0.00 | legumain-like ... | 100.00
testread contig 896 | 100 | 0.00 | beta-tubulin [... | 100.00

Use BLAST to get functional annotation

http://blast.ncbi.nlm.nih.gov/Blast.cgi




BLAST
Basic Local Alignment Search Tool
Home
Recent Results
Saved Strategies
Help
My NCBI
[Sign In] [Register]
NCBI/ BLAST/ blastx
blastn
blastp
blastx
tblastn
tblastx
BLASTX search protein databases using a translated nucleotide query. more...
Reset page
Bookmark
Enter Query Sequence
Enter accession number(s), gi(s), or FASTA sequence(s)
Clear
Query subrange
From
To
Or, upload file
瀏覽...
Genetic code
Standard (1)
Job Title
Enter a descriptive title for your BLAST search
Align two or more sequences
Choose Search Set
Database
Non-redundant protein sequences (nr)
Organism
Optional
Enter organism name or id--completions will be suggested
Exclude
Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown.
Exclude
Optional
Models (XM/XP)
Uncultured/environmental sample sequences
Entrez Query
Optional
Enter an Entrez query to limit search
BLAST
Search database Non-redundant protein sequences (nr) using Blastx (search protein databases using a translated nucleotide query)
Show results in a new window
Algorithm parameters

Download

- Latest Blast can be downloaded here
  - ftp://ftp.ncbi.nih.gov/blast/executables/release/LATEST/

```
wget  http://163.25.92.61/course/blast-2.2.24-x64-linux.tar.gz
```

- extract the file

```
tar   zxvf   blast-2.2.24-x64-linux.tar.gz
```

- executable binary file is at blast-2.2.24/bin/

```
cd   blast-2.2.24/bin/
ls   -l
```

blastall

- blastall -p program -i input_sequence -d dbname -o output_file

  -p (blastn, blastx, blastp, tbastn tblastx)

  -i (query file)

  -d (database name)

  nr, NCBI non-redundant database

  -o (output file)

Blast programs

-p program	-i Query sequence	-d database sequence
blastn	nucleotide	nucleotide
blastp	amino acid	amino acid
blastx	translated nucleotide	amino acid
tblastn	amino acid	translated nucleotide
tblastx	translated nucleotide	translated nucleotide

formatdb

- download proteins fasta

```
mkdir   blastdb
cd   blastdb/
wget   http://163.25.92.61/course/testproteins.fasta
```

- ncbi nr
  - ftp://ftp.ncbi.nih.gov/blast/db/

```
../blast-2.2.24/bin/formatdb   -i   testproteins.fasta  -p T
```

```
ls
```

- formatdb.log
- testproteins.fasta
- testproteins.fasta.phr
- testproteins.fasta.pin
- testproteins.fasta.psq

- formatdb 2.2.24 arguments:

- -t Title for database file [String] Optional
- -i Input file(s) for formatting [File In] Optional
- -l Logfile name: [File Out] Optional
- default = formatdb.log
- -p Type of file
- T - protein
- F - nucleotide [T/F] Optional
- default = T
- -o Parse options
- T - True: Parse SeqId and create indexes.
- F - False: Do not parse SeqId. Do not create indexes.
- [T/F] Optional
- default = F
- -a Input file is database in ASN.1 format (otherwise FASTA is expected)
- T - True,
- F - False.
- [T/F] Optional
- default = F
-

- -b ASN.1 database in binary mode
- T - binary,
- F - text mode.
- [T/F] Optional
- default = F
- -e Input is a Seq-entry [T/F] Optional
- default = F
- -n Base name for BLAST files [String] Optional
- -v Database volume size in millions of letters [Integer] Optional
- default = 4000
- -s Create indexes limited only to accessions - sparse [T/F] Optional
- default = F
- -V Verbose: check for non-unique string ids in the database [T/F] Optional
- default = F
- -L Create an alias file with this name
- use the gifile arg (below) if set to calculate db size
- use the BLAST db specified with -i (above) [File Out] Optional
- -F Gifile (file containing list of gi's) [File In] Optional
- -B Binary Gifile produced from the Gifile specified above [File Out] Optional
- -T Taxid file to set the taxonomy ids in ASN.1 deflines [File In] Optional

blastall

- blastall -p program -i input_sequence -d dbname -o output_file

  -p (blastn, blastx, blastp, tbastn tblastx)

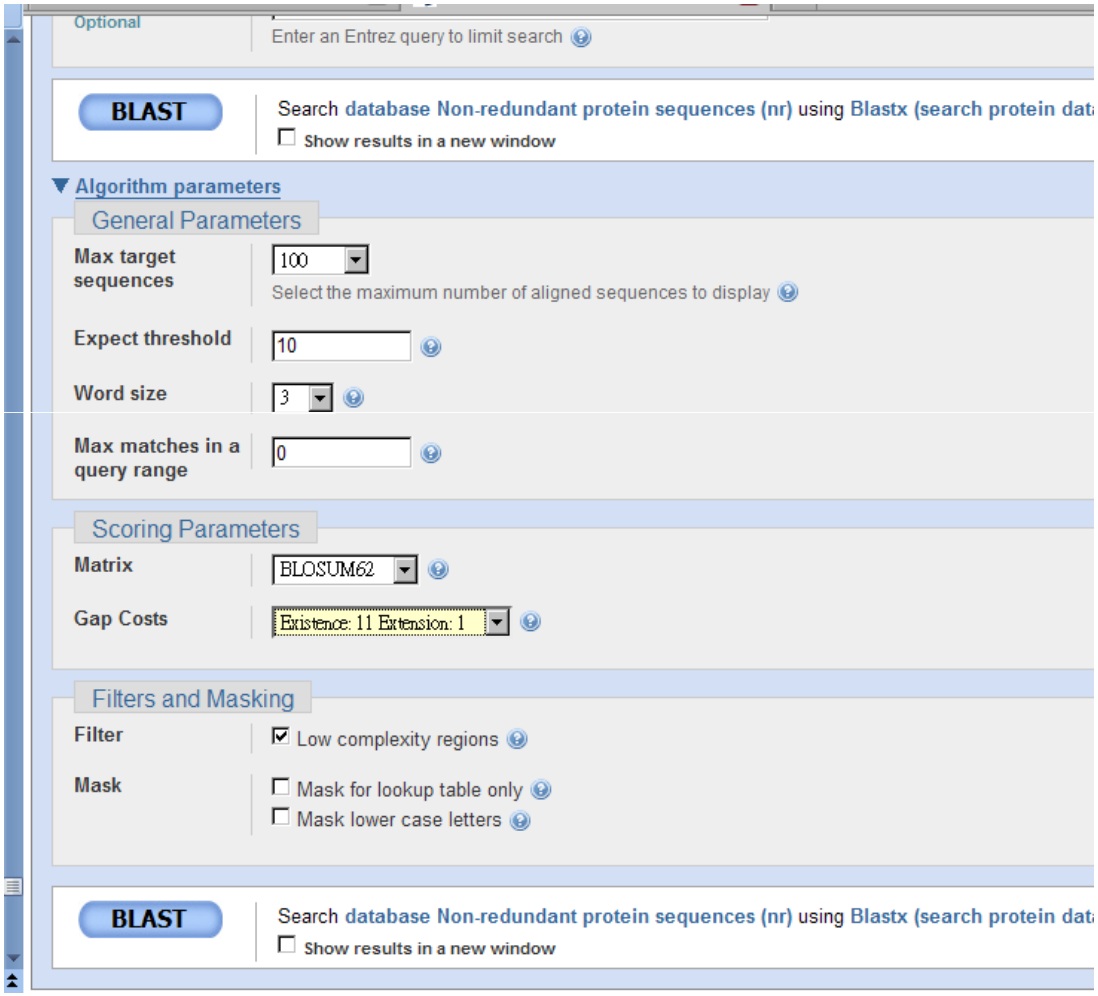
  -i (query file)

  -d (database name)

  nr, NCBI non-redundant database

  -o (output file)

blastall parameters



- blastall 2.2.24 arguments:

- -p Program Name [String]
- -d Database [String]
- default = nr
- -i Query File [File In]
- default = stdin
- -e Expectation value (E) [Real]
- default = 10.0
- -m alignment view options:
  - 0 = pairwise,
  - 1 = query-anchored showing identities,
  - 2 = query-anchored no identities,
  - 3 = flat query-anchored, show identities,
  - 4 = flat query-anchored, no identities,
  - 5 = query-anchored no identities and blunt ends,
  - 6 = flat query-anchored, no identities and blunt ends,
  - 7 = XML Blast output,
  - 8 = tabular,
  - 9 tabular with comment lines
  - 10 ASN, text
  - 11 ASN, binary [Integer]
- default = 0
- range from 0 to 11

- -Q Query Genetic code to use [Integer]
- default = 1
- -D DB Genetic code (for tblast[nx] only) [Integer]
- default = 1
- -a Number of processors to use [Integer]
- default = 1
- -O SeqAlign file [File Out] Optional
- -J Believe the query defline [T/F]
- default = F
- -M Matrix [String]
- default = BLOSUM62
- -W Word size, default if zero (blastn 11, megablast 28, all others 3) [Integer]
- default = 0
- -z Effective length of the database (use zero for the real size) [Real]
- default = 0
- -K Number of best hits from a region to keep. Off by default.
- If used a value of 100 is recommended. Very high values of -v or -b is also suggested [Integer]
- default = 0
- -P 0 for multiple hit, 1 for single hit (does not apply to blastn) [Integer]
- default = 0
- -Y Effective length of the search space (use zero for the real size) [Real]
- default = 0
- -S Query strands to search against database (for blast[nx], and tblastx)
- 3 is both, 1 is top, 2 is bottom [Integer]
- default = 3
- -T Produce HTML output [T/F]
- default = F

- -o BLAST report Output File [File Out] Optional
- default = stdout
- -F Filter query sequence (DUST with blastn, SEG with others) [String]
- default = T
- -G Cost to open a gap (-1 invokes default behavior) [Integer]
- default = -1
- -E Cost to extend a gap (-1 invokes default behavior) [Integer]
- default = -1
- -X X dropoff value for gapped alignment (in bits) (zero invokes default behavior)
- blastn 30, megablast 20, tblastx 0, all others 15 [Integer]
- default = 0
- -I Show GI's in deflines [T/F]
- default = F
- -q Penalty for a nucleotide mismatch (blastn only) [Integer]
- default = -3
- -r Reward for a nucleotide match (blastn only) [Integer]
- default = 1