

DNAMAN

Sequence Analysis Software

*Sequence
Assembly*

BIOINFORMATICS PLATFORM

Sequence Assembly

- ▶ De novo Assembly
- ▶ Re-Sequencing (Reference-Based)
- ▶ Assembly Parameters
- ▶ Assembly with Trace Files

De Novo Assembly

- No reference sequence available
- Assemble overlapping short reads into long contigs
- Each short reads is aligned with others to produce a score matrix
- Contigs are assembled based on scores
 - High score pairs aligned first
 - Low score pairs last Contigs are assembled based on scores
- Computation intensive

Re-Sequencing Assembly

- Reference sequence available
- Each short reads is aligned to Reference (No score matrix needed)
- Computation less intensive than De Novo Assembly

Assembly Parameters

1. Select sequences for assembly from File/Folder/Channel
2. Quality of short sequences can be improved by filtering ambiguous regions
3. Choose type of assembly
4. Set overlap end qualification requirements
5. Set fast alignment parameters
6. Choose Final Assembly method
7. Press Assemble button to start

The screenshot shows the 'Sequence Assembly' window with the following settings and annotations:

- 1**: File list containing:
 - Sequence#: 2001
 - \\dnaman\AssemblySamples.seq
 - \\dnaman\lambda-assembled.seq
- 2**: 'Remove flanking regions when > 1 ambiguous bases in 10'.
- 3**: 'Assembly Method' set to 'Re-sequencing (longest seq as template)'.
- 4**: 'Minimum overlap' set to 80.
- 5**: 'Quick alignment' parameters: K-tuple 4, Gap open 4, Gap extend 0, Window 4, Diagonals per 500bp 3.
- 6**: 'Final Assembly' method set to 'Optimal'.
- 7**: 'Assemble' button.

Other visible settings include: 'Remove vector sequence' (unchecked), 'Identity >= 90 %', 'End Comparison' (unchecked), 'Maximum overlap 300', 'De novo' (unchecked), 'Minimize window while running' (unchecked), and 'Re-order sequences' (checked).

Assembly with Trace Files

Sequence Assembly

Sequence#: 2

C:\dnaman\test1.seq
C:\dnaman\abi_sample1.abi ← Trace File

Add file
Folder
Channel
Database
Remove
Clear

Remove flanking regions when > ambiguous bases in
 Remove vector sequence:

Minimum overlap Identity >= % Load Parameters Save Parameters

Quick alignment
K-tuple Gap open Gap extend
Window Diagonals per 500bp

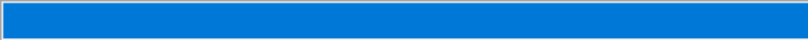
End Comparison
Maximum overlap

Final Assembly
 Quick
 Optimal

Alignment Method Quick Alignment End Comparison

Assembly Method De novo Re-sequencing (longest seq as template)

Minimize window while running Re-order sequences

Assemble Complete 1 of 1 

1 overlaps for 1 contig(s)

Show result Cancel

Assembly with Trace Files

Consensus
Test1
sample1.abi

```
:560 :570 :580 :590 :600 :610 :620 :630 :640 :650 :660  
GCTTGTGATGGCAGAATGGGCTAAGCTTAGTGTCTTAAGGnCATCTCCGaaGGGnTTCCCATGGGGtAGCAgATTGGGTACCACCCAAGTATTGACTCACCCATCAACAAC  
GCTTGTGATGGCAGAATGGGCTAAGCTTAGTGTCTTAAGG.CATCTCCGAAGGG.TTCCCATGGGGTAGCAGATTGGGTACCACCCAAGTATTGACTCACCCATCAACAAC  
TCTTAAGGNNATCTCCGANGGGNTTCCCATGGGGNAGCANATTGGGTACCACCCAAGTATTGACTCACCCATCAACAAC
```

Ambiguous 'N' 1

Options
Output

Seq Num=2

Coverage=1.15

1 1227

Test1
abi_sample1.abi

1. Double-click on ambiguous 'N' position
2. Popup Trace window showing 'G' in the 'N' position

abi_sample1.abi

abi_sample1.abi

T CTCC GAN GGG N TTCCCA TGG G N AGC A ATT TGG T ACC ACCCA AGT ATT T G ACT CA

20 30 40 50 60

'N' may be called to 'G' 2