

# CLC Assembly Cell

## Features & Benefits

### All Major Sequencing Platforms Supported

- Illumina's Genome Analyzer
- SOLiD by Life Technologies
- 454 GS flx by Roche
- Helicoscope by Helicos
- Sanger sequencing data

### System Requirements

- Mac OS X 10.4 or later (incl. Intel-based Macs)
- Windows 2000, Windows XP, Windows Vista or Windows 7
- Linux: Redhat or SuSE
- 32 bit version and 64 bit version of operating system/computer on all platforms
- 2 GB RAM required
- 16-48 GB required for large assemblies

Version 3.2 for Windows, Mac OS X, and Linux  
CLC bio©Copyright 2011

### Helping overcome the challenges of Next Generation Sequencing

CLC Assembly Cell is a command-line program that makes your high-throughput sequencing data analysis pipeline fast, flexible and easy to maintain.

Next Generation Sequencing technologies present a number of challenges to bioinformatics in terms of data analysis. The first challenge is assembling the data. Existing software for assembly typically supports only one type of data (either short or long reads) which makes it hard to exploit the advantages of combining different platforms. In contrast, CLC Assembly Cell offers high-speed assembly of sequencing data from all platforms (see text to the left). The ability to combine these technologies in the same analysis is one of the major strengths of the CLC Assembly Cell, as well as the ability to analyze reads ranging from 35bp to many hundred bp.

The Assembly Cell supports reference assembly and *de novo* assembly of genomes of all sizes.

### De novo assembly

The *de novo* assembly of CLC Assembly Cell supports both short read and long read assembly, including 454/Titanium. CLC Assembly Cell also supports *de novo* assembly of paired end data (see figure 1 below) and hybrid assembly of multiple data types.

### Reference assembly

CLC Assembly Cell supports reference assembly of Illumina Genome Analyzer, SOLiD, 454, Helicos and Sanger sequencing data.

It also handles short and long read assembly (incl. 454/Titanium), as well as gapped and ungapped alignment.

### Human genome *de novo* assembly benchmarks

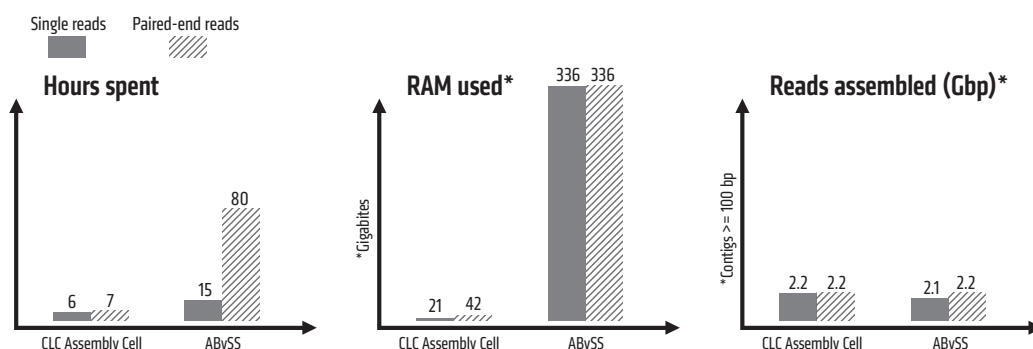


Figure 1: The data set used for the human genome assembly is produced with Illumina's Genome Analyzer and consists of approximately 3.6 billion reads resulting in 38 fold coverage of the human genome. Most reads have a length of 36 bp, summing up to 130 Gbp in total.

