



CLC Bioinformatics Cell

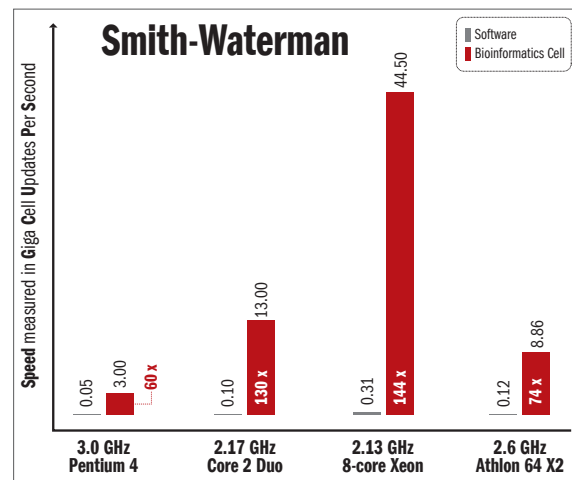
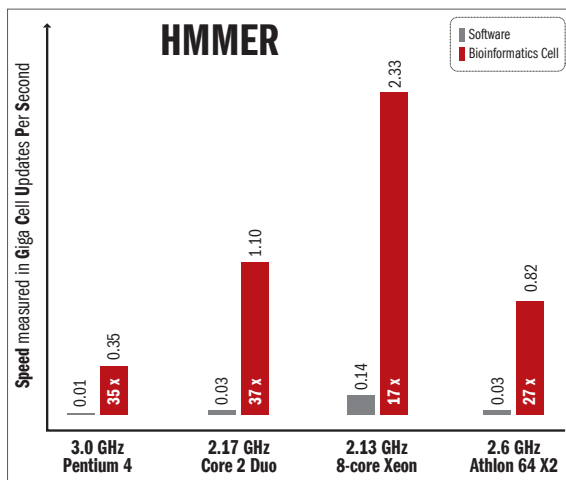
© Copyright CLC bio, August 2007

High-performance computing

With a CLC Bioinformatics Cell, you will realize significant accelerations of a number of important and commonly used bioinformatics algorithms.

By utilizing the hidden computational powers in your desktop computer, server, or computer cluster, the Cell will give you a fast ClustalW, the fastest Smith-Waterman implementation (130x), and the fastest HMMER implementation (37x) ever made on standard hardware.

Using the Cell will speed up your research significantly:



Integrated

The Cell is fully compatible and integrated with the most comprehensive bioinformatics desktop software applications in the market: CLC DNA Workbench, CLC RNA Workbench, CLC Protein Workbench, and CLC Combined Workbench.

The Cell thus benefits from all the features of CLC bio's workbenches such as the graphical user interface, the option of running multiple analyses in batch runs, the data storage opportunities, the graphical and tabular overview of the results, etc. Using high performance computing has never been easier or more effective.

Command-line execution is also an option. This means you have the option of integrating high-performance calculations in your existing bioinformatics workflow of scripts and in-house programs.

You can of course also use the Cell through our integrated Web portal tool, giving you and your colleagues access to using the Cell from all over the world.

Flexible

This innovative approach to hardware-accelerated bioinformatics provides flexibility in a variety of ways:

- **Physical location:** No need for a cooled server room or new expensive hardware to perform high performance computing. All you need is your computer.
- **Operating system:** As long as your computer has an Intel or AMD processor, the Cell works on both Windows, Linux, and Mac OS X operating systems.
- **Software use:** The Cell can be used as an integrated part of all CLC bio's workbenches, or through a command line interface.
- **Customization:** With CLC Developer Kit, you can create your own workflows including the functionalities of the Cell.



CLC Bioinformatics Cell



CLC Bioinformatics Cell uses the Single Instruction, Multiple Data (SIMD) technique to parallelize and thereby accelerate bioinformatics algorithms on Intel and AMD computer processors

Utilize the full power of your computer cluster

If you have a computer cluster, or if you have a computer with multiple CPUs or cores, the Cell provides you with a solution that accelerates your calculations proportionally with the number of Cell's implemented.

HMMER searches

HMMER is an implementation of profile HMM models for protein sequence analysis.

Many researchers use HMMER for very computational intensive and time-consuming tasks, and we have therefore developed a SIMD-based version of HMMER that accelerates Hmmpfam and Hmmssearch up to 30 times on a modern desktop computer, resulting in significant reductions in search times.

Smith Waterman searches – when you need ALL the answers

Finding homologue DNA, RNA, or protein sequences in a database can be done in many ways. Smith-Waterman based searches is the only method that identifies all true hits, but the algorithm is very slow when working on large datasets and most scientists therefore use the much faster BLAST search algorithm.

However, the speed of BLAST comes on the expense of the quality. The sensitivity of BLAST is in fact so low that there is significant risk of missing important sequences of interest. Compared to Smith Waterman based searches, up to 50% of the search hits are thus not found using BLAST.

With the Cell, you can speed up a Smith-Waterman search previously taking half a day to around 6 minutes, and the Cell thus removes the argument for using BLAST – at least in situations where you search through data where you not only need some of the answers but all of the answers.

**BLAST gives you
some of the answers.**

**The Cell gives you all
of them.**

ClustalW alignments

ClustalW is one of the most used methods for doing multiple sequence alignments. When aligning many sequences and when aligning long sequences, the speed of the algorithm is however not impressing.

ClustalW being so widely used as it is, we have implemented a SIMD-version of it in the Cell, resulting in an acceleration of 10 times on most computers.