

DeCypherSW™

Algorithm Module

The solution for sensitive, frameshift-tolerant sequence analysis

DeCypherSW™ rapidly processes Smith-Waterman searches using SeqCruncher™, an accelerator card designed with TimeLogic's field programmable gate array (FPGA) technology. With algorithms that are more rigorous—and therefore more compute-intensive—than BLAST, DeCypherSW allows you to utilize data from hundreds of genome projects for sensitive cross-species protein comparisons, and can also help you identify highly specific oligonucleotides for microarray, small RNA experiments.

DeCypherSW advantages:

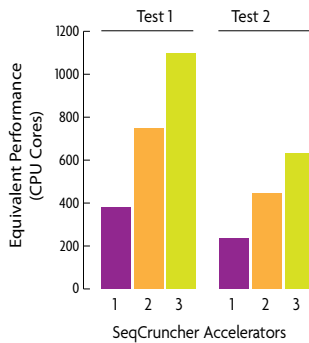
- Optimal local alignments can span long introns for genomic data analysis
- Accommodates insertion/deletion errors by interleaving alignments across all potential reading frames
- Solution scales easily with near-linear performance improvements
- Short-read and oligo (16-50 mer) mapping for remapping workflows now 10-15X faster

Smith-Waterman has been optimized for the SeqCruncher™ accelerator to deliver hundreds of CPU cores of performance

SeqCruncher™
FPGA Accelerator



DeCypherSW delivers 236-1099 CPU cores of performance within a single server



Due to the highly parallel SeqCruncher circuitry, DeCypherSW is able to substantially outperform traditional software such as SSEARCH. For typical Smith-Waterman searches (Test 1, Test 2), DeCypherSW—using a single SeqCruncher—ran 379X and 236X faster than SSEARCH on 1 CPU core. With 3 SeqCruncher cards installed, DeCypherSW demonstrated the equivalent of 1099 and 631 CPU cores. Additional details are presented in our Smith-Waterman Whitepaper.

Tests were completed on a Dell PowerEdge 2950 III server with 2 quad-core Xeon (3.0 GHz) processors and 8 GB of RAM (running 64-bit Fedora Core 6).



A 2U DeCypher host server can drive up to 3 SeqCruncher accelerators

Applications of DeCypherSW accelerated algorithms

Algorithm	Description	Applications
Smith-Waterman	Highly sensitive similarity searching with optimal local alignments (NT-NT, AA-AA)	<ul style="list-style-type: none"> • Minimize cross-hybridization and maximize specificity for microarray design, antibody production and siRNA experiments • “No compromise” homology modeling: find global alignments of your protein sequence to proteins for which protein structure is known • Find regulatory regions and mask repetitive elements
Double Affine Smith-Waterman	Powerful method that spans introns and delivers optimal alignments (NT-NT, AA-AA)	<ul style="list-style-type: none"> • Intron-tolerant mapping of coding sequences to genomes • Identify splice variants, truncated proteins, and recombinant deletions
Framesearch	Highly sensitive search using either the translated query or translated target database (AA-NT, NT-AA)	<ul style="list-style-type: none"> • Alignments span gaps and frameshifts, for effective cross-species analysis • Assign tentative functions to EST sequences <p><i>More tolerant of gaps and frameshifts than BLASTX</i></p>
Symmetric Frames Independent	Highly sensitive protein domain search between your nucleic query and nucleic database(s), employing frame-shifting as needed (NT-NT)	<ul style="list-style-type: none"> • Exploration of low-coverage genome data to find genes and pseudogenes • Annotate potential protein functions within low coverage nucleic sequences <p><i>More tolerant of gaps and frameshifts than TBLASTX</i></p>

Framesearch extends DeCypherSW utility

Framesearch builds upon Smith-Waterman by adding the capability to shift alignments across reading frames for comprehensive EST, gene and pseudogene annotation. For each comparison of your sequences to a group of protein sequences, the program finds the optimal alignment between your sequence and all possible codons from each nucleic strand.

A symmetric version of framesearch, Symmetric Frame Independent (SFI), completes sensitive protein domain searches between translated query and translated target databases for better cross-species comparisons. Symmetric Frame Independent excels in comparing data with low homology, as well as ESTs with frequent indels, that can result in missed or erroneous hits when using translated BLAST methods. SFI represents a 36-way Smith-Waterman search, yet is very usable on the TimeLogic solutions.

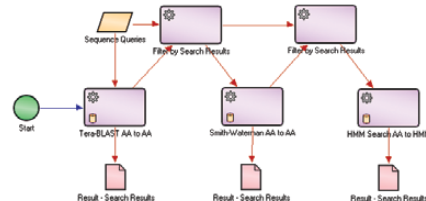
Interfaces for all levels of bioinformatics expertise

DeCypherSW includes an intuitive graphical user interface that lets you submit thousands of sequences per search. Color-coded search results, returned as HTML or text, reflect biologically relevant alignment details, including % identity, gapping & frameshifts (on each strand), mismatches and multiple domains. Target databases may also be built via the web (up to 200MB), simplifying administration.



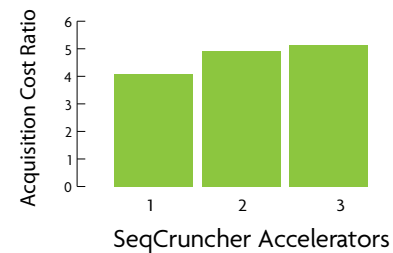
Larger, more complex searches may be processed through the Command Line client, which can be installed on Windows, Linux, and Solaris operating systems. This efficient connection to the DeCypher system relies on a set of concise command arguments for total control over your search jobs. These commands are easily scripted for automated pipelines.

With our optional PipeWorks™ client/server software for Windows, DeCypherSW searches may be interconnected with other search methods and filters to create comprehensive analysis workflows.



DeCypher vs. the commodity cluster

Computational performance is only meaningful when solution cost is considered as well. In the graph below, the cost of a DeCypher system (including the DeCypherSW module, 1, 2, or 3 SeqCruncher cards and a Dell 2950 III server) was compared to the cost of a cluster with equivalent performance (Dell 1950 IU servers with 2 quad-core CPUs). The cluster required 4-5.2X more budget to acquire.



Once you factor additional cluster operational costs (power, HVAC cooling, IT staff, server room rackspace), the DeCypher system becomes even more cost-effective.

Integration with additional DeCypher solutions

The DeCypher® Enterprise system is available with DeCypherSW™, DeCypherHMM™, TeraBLAST®, and the GeneDetective™ algorithm modules. With built-in clustering software that enables searches to scale across multiple SeqCruncher accelerator cards and multiple accelerated nodes, DeCypher enables you to configure an optimal biocomputing solution for your research environment.

DeCypherSW is also included with CodeQuest™, a plug-and-play workstation for personal or lab-wide genomics projects. CodeQuest includes all four algorithm modules, a powerful desktop workstation and 1 or 2 SeqCruncher accelerators.

A plug-and-play biocomputing solution

DeCypherSW includes all you need to execute fast Smith-Waterman searches for microarray, microRNA and siRNA projects. Its optimal alignments and excellent handling of short sequence comparisons make it an ideal tool for accurate mapping of oligos and short-read data. The Double Affine and Frameshift variants of Smith-Waterman provide rigorous capabilities for genome exploration, particularly for genome comparisons between evolutionarily divergent species.

Contact TimeLogic today to learn more about DeCypherSW or to schedule a free performance benchmark on your data!