

PARACEL® BLAST

Accelerated BLAST software optimized for Sun clusters

Paracel BLAST algorithms join with Sun clusters for high speed and scientifically valid results

Paracel BLAST is the most advanced BLAST software written specifically for large-scale cluster systems such as those offered by Sun Microsystems. Paracel started with the NCBI BLAST source code and rewrote significant portions of it to produce a native-parallel BLAST application and eliminate significant performance bottlenecks. Consistent with Paracel's commitment to scientific validity, these enhancements are obtained without loss of accuracy, sensitivity, or selectivity. This allows users of Sun computers to obtain their desired results in record time with unprecedented cost effectiveness.

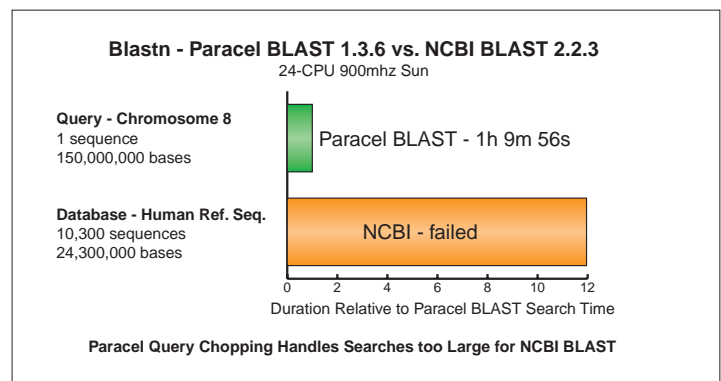
Paracel BLAST Innovations

Paracel BLAST is an integrated native-parallel application with optimizations, enhancements, and features not available in any other BLAST system.

- **Automatic parallelization, queuing, and scheduling** – High performance is attained without complex user interactions. When you submit a search from your Sun workstation, either through the command-line or the web-based graphical user interface, Paracel BLAST takes advantage of the high throughput of your Sun cluster and automatically executes your searches in parallel on multiple processors. You are not required to master complicated scheduling system software.
- **Search large databases without loss of performance** – Paracel BLAST operates at the same high speed on databases of all sizes, even databases much larger than the physical RAM of the cluster. Problems that previously failed because of their sheer size can now be rapidly solved on your existing Sun systems. Example: with Paracel BLAST, your Sun cluster can quickly search all of GenBank, instead of being restricted to just small parts of it.
- **Work with queries 100 times larger than NCBI** – Paracel's proprietary query-chopping system allows you to search chromosome-sized queries in parallel without loss of accuracy. Instead of your IT department spending valuable time writing

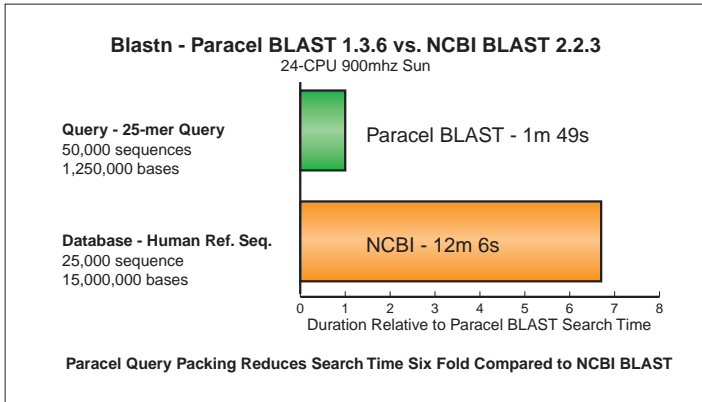
scripts to overcome NCBI BLAST's limitations, you can directly perform the biological analyses that are so important to the success of your projects.

- **Rapidly perform large-scale analyses** – Paracel BLAST's superior performance lets you complete searches routinely that were previously prohibitively long. A Paracel tblastx search of human transcripts against the human reference sequences can be completed in 2½ hours on a 24-CPU 900mhz Sun system, compared with 30½ hours using NCBI BLAST. A 24-CPU Sun system running Paracel BLAST can search Human Chromosome 8 (a 150 Mbase sequence) against the human reference sequences in less than one hour and ten minutes.

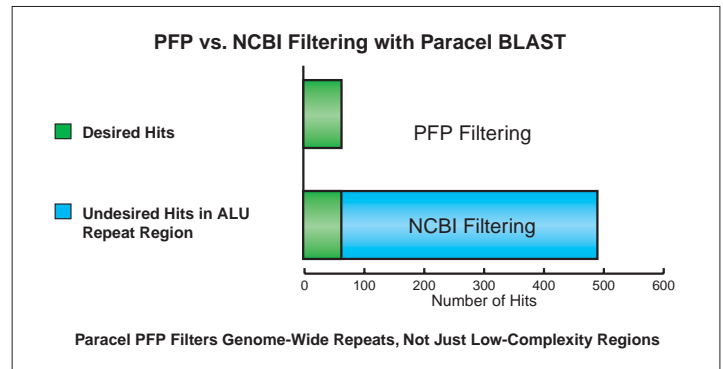


- **Rapid searches of probes and primer sequences** – Paracel's proprietary "query packing" technology provides unprecedented acceleration of searches made up of short query sequences. Paracel's software yields a blastn search of 25-mers that is over *six times* faster than with NCBI BLAST.
- **NCBI BLAST file compatibility** – Paracel BLAST uses the same input and output file formats as NCBI BLAST. If you currently use NCBI BLAST, you can easily integrate Paracel BLAST into your existing computing environment and pipeline.
- **NCBI BLAST look and feel** – There is no learning curve when moving from NCBI to Paracel BLAST. The Paracel user

interface has been designed so that users of NCBI BLAST immediately find themselves in a familiar work environment.



Paracel BLAST is preconfigured with the most commonly used PFP parameters, so researchers can exploit the flexibility and sensitivity of PFP to filter out noise and minimize the number of false hits directly from the Paracel BLAST interface.



- **Superior scalability** – Paracel BLAST’s queuing and scheduling system provides high parallel efficiency even on large cluster systems, providing high throughput for complex analyses and multi-user configurations.
- **Enhanced sequence filtering** – Paracel BLAST integrates seamlessly with SEG, DUST and PFP, the Paracel Filtering Package. PFP filters contaminants such as mitochondrial DNA and *E. coli* and masks unwanted elements such as vector sequences, SINEs, LINEs, repeats and low-complexity regions.
- **Immediate portability to your Sun system** – Paracel BLAST searches can be implemented on your existing Sun cluster. A typical system would consist of the following:
 - A Sun 34 node Netra X1 Grid
 - Solaris 5.6 or later
 - At least 1 GB RAM/CPU
 - Ethernet network (100 baseT or 1000 baseT)
 - At least 100 GB shared storage (for searchable databases)

Paracel BLAST Algorithms

Paracel BLAST provides accelerated versions of the following standard algorithms:

- **Blastn** – compares a DNA query against a DNA database.
- **Blastp** – compares a protein query against a protein database.
- **Blastx** – compares a DNA query translated into all six reading frames against a protein database.
- **TBlastn** – compares a protein query against a translated DNA database.
- **TBlastx** – compares the six reading frames of a DNA query against the six reading frames of a DNA database.
- **Phi-Blast** – the Pattern Hit Initiated BLAST algorithm combines the matching of regular expressions with local alignments surrounding the match. This allows the researcher to find database sequences that match a user-defined regular expression and are homologous in the area around the match.
- **Psi-Blast** – the Position Specific Iterative BLAST algorithm is useful both for identifying distant members of a protein family, whose relationship is not recognizable by straight sequence comparison, and also for deducing the function of proteins that are unannotated in the database. The added sensitivity of this algorithm over ordinary BLAST searches comes from the use of a profile that is automatically generated from a multiple alignment of the highest scoring hits in an initial Blastp or Phi-Blast search. The profile is used when performing subsequent Psi-Blast searches, and the results of each iteration are used to refine the profile.
- **MegaBlast** – concatenates many queries together to reduce the time spent scanning the database. MegaBlast is optimized for aligning sequences that differ slightly. This search algorithm is up to ten times faster than some more common sequence searching algorithms, but because of its long word size is less sensitive than other BLAST searches.

PARACEL[®]
A CELERA BUSINESS

1055 East Colorado Boulevard • Fifth Floor
Pasadena, CA 91106-2341
Phone: 626.744.2000 • Fax: 626.744.2001
www.paracel.com

Regional Sales
North America
West Coast: 510.583.6175
East Coast: 732.245.9439

Europe
United Kingdom: +44 (0) 1908.357.700
Asia Pacific
U.S. Office: 408.395-3114