



# STRIKING DEVELOPMENT

## Changes for release 1.6.1 (04/2005):

- Added a MacOS X pb client, which can be invoked from the command line like the UNIX and Windows clients.
- Added documentation on how to integrate PB with a batch queue system such as LSF or SGE. This documentation can be found in a text file called BatchIntegration in the doc directory on the PB CDROM. This also includes a new script called pbsub which provides an easy interface to submitting pb commands to a batch queue system.
- Added the "pb fastacmd" command, which works like the NCBI blast fastacmd tool. This allows users to dump fasta-formatted data out of blast-formatted databases. Users may dump entire databases or ask for just particular sequences (if the database was formatted with -o T).
- Updated README.sparc to talk about adding a pbguest user so that non-UID users would be able to search (e.g., guest users, DOS clients).
- Fixed a minor issue where the pbd init script would fail to start the manager daemon properly if the hostname of the manager became fully qualified.
- Updated the README to describe proper installation of the PB rpms on differen architectures (x86 versus x86\_64).
- Updated the install script for PB to better support automatic installation on Rocks 2.3.2 and Rocks 3.2.0 systems. This causes worker nodes to automatically configure the PB daemons in the event that new workers are added or old ones are re-kickstarted.

- Fixed a bug that caused query-anchored output for gapped tblastn searches to show incorrect alignment information or garbage data.
- Fixed a bug in the NCBI 2.2.6 cache lookup feature that caused warning and error messages from some psi-blast and psi-tblastn searches. The NCBI 2.2.6 cache lookup was added to fix a problem with some large blastp searches.
- Disabled two-pass searching. This feature was not working, and in fact was abandoned in later NCBI releases. If one uses `-P 2` on the blastall command line, it is now treated the same as `-P 0`. This matches NCBI blast releases 2.2.4 and later.
- Fixed a series of bugs that prevented some searches with GI lists (the `-l` option) from working properly.
- Fixed a bug that could occasionally cause query splitting to fail with a message saying that some of the pieces were smaller than the overlap.
- Applied a fix from NCBI blast 2.2.6 that prevents trouble with some gapped nucleotide searches that would fail with messages about the start points of sequences being negative or larger than their sizes.
- Fixed a bug that would prevent long perfect-match searches (for instance, a long sequence searched against itself) from showing the entire alignment. Sometimes this bug would also appear as very poor alignment data after the first several thousand match positions.
- Fixed a bug that would cause some megablast searches on small word sizes to crash.
- Fixed a bug that caused some searches on older formats of database data to fail because of missing and/or incorrect OIDs when database splitting took place.
- Fixed a bug in the Windows pb client that prevented proper operation when dealing with PSSM checkpoint data from psi\_blast and psi\_tblastn searches. These searches now work properly.
- Fixed a bug that would cause command-line option parsing to fail

and crash the pb client under some circumstances.

- Updated the exit code status of many pb commands, such that they now return a non-zero (failure) status when they do not succeed. This aids in scripting, as parsing of the output from the commands is no longer necessary in many cases.
- Fixed a series of bugs in the way that wildcards were handled with the pb filesystem commands.
- Added -r and -f options to the "pb rm" command. These flags allow recursive and forceful deletion, respectively, just like the UNIX rm command.
- Updated the way that the -workerlist argument is handled, such that each name given indicates a single processor. Thus, if one wanted to use two cpus on a given node, one would specify that node's name twice in the workerlist. Previously, putting a worker node's name in the list would cause all cpus on that node to be available for use.
- Modified the way that tasks are selected for placement on worker nodes to prevent large jobs from starving each other out. Previously, if job A was submitted and allowed to become mostly complete, and then job B was submitted (both of roughly equivalent size), job A would stall in the queue until job B reached the same level of completion as job A, at which point they would both complete. Now, job A's last tasks will continue to run in parallel with job B's first tasks.
- Fixed a bug that caused "pb status" to crash if there were a large number of jobs in the queue. If the number of jobs in the queue exceeds thresholds, "pb status" will now report a number of jobs and then note that there are more remaining in the queue which aren't shown.
- Fixed a bug that caused some large databases to be unsearchable after copying or renaming.

---

Changes for release 1.6.0 (04/2004):

- Fixed a bug that caused searches to fail when using XML output [-m 7]

and database partitioning.

- The "-R" recursive option for the chown, chgrp, and chmod commands is no longer supported.
- Changed pb status so that non-root users can see the percentage of other people's blast job.
- Client and server can now take advantage of x86\_64 instructions on Opteron CPUs.
- Full integration with Rocks 3.1, 3.0, and 2.3.2 installation system.

---

Changes for release 1.5.6 [11/25/03]:

- pb formatdb can now be used to convert NCBI indexed BLAST indices into Paracel BLAST formatted indices. See the manual for more details.
- Numerous improvements to improve the accuracy of large query "chopping" searches.
- Fixed a partitioning bug that could result in the message, "Bad reckoning in querySplit".
- Enabled psi-blastn searches from within Paracel Blast. Performing psi-blastn searches involves two steps. In the first step, the user should run psi-blast to create and save a position-specific scoring matrix (PSSM) given a single protein query sequence and a protein database and using the "-j" option with an integer greater than 1 as follows:

```
pb blastpgp -i protein_sequence -j 2 -C protein_sequence.pssm -F F \  
-d protein_database
```

where the position-specific scoring matrix for the given protein query sequence is saved in "protein\_sequence.pssm".

In the second step, the user runs psi-blastn given the PSSM that was generated in the first step and the same protein query sequence that was used to generate the PSSM against a DNA database as follows:

```
pb blastall -p psitblastn -i protein_sequence -R protein_sequence.pssm -F F \  
-d dna_database
```

The "-R" option is required when using "pb blastall -p psitblastn".  
All other options that apply when using "pb blastall -p tblastn" also  
apply when using "pb blastall -p psitblastn", but the following  
restrictions should be observed:

- 1) The query sequence used in the psi-tblastn search should be the same as  
the one used to generate the PSSM.
  
- 2) By default, "pb blastpgp" has filtering off ("-F F") while "pb blastall"  
has filtering on ("-F T"). To ensure consistent usage of the  
blastpgp/psi-tblastn combination, the "-F" option should be set to the  
same value in both steps. So, it could either be set to "-F T" in the  
blastpgp step and not set in the psi-tblastn step (since it's set to  
"-F T" by default), or it should not be set in the blastpgp step (it is  
set to "-F F" by default) and set to "-F F" in the psi-tblastn step.  
If the user desires to use PFP to filter the query, then PFP should  
be used in both the "pb blastpgp" and the "pb blastall -p psitblastn"  
steps with the same PFP parameter file.

Note that the PSSM can be constructed using one database and then used to  
search a different database. Even if the two database names are the same,  
"pb blastpgp" uses the protein version while "pb blastall -p psitblastn"  
uses the DNA version.

- Updated "cache management" system, providing significant performance  
enhancements for certain types of searches, for example blastn  
searches of multi-GB databases.
  
- Fixed bugs that resulted in reporting wrong identities, positives, and  
their percentages, as well as missing score headers for some hits,  
especially for ungapped searches.
  
- Changed the size of the results hit list for ungapped searches to  
match that for NCBI 2.2.5 which is smaller than that for NCBI 2.2.3.  
Hence, there might be fewer hits reported for ungapped searches, such  
as tblastx, than before.

- Ranking of hits when databases were split or queries were chopped was inadvertently performed according to descending scores instead of ascending e-values in PB 1.5.2, 1.5.3 and 1.5.4. Now, hits are ranked according to ascending e-values, similar to NCBI, even when PB performs database splitting or query chopping.
- Improved the detection and elimination of contained alignments when query chopping is performed.
- The "Unable to calculate Karlin-Altschul params ..." error message now lists only the name of the query sequence for which it pertains, instead of listing the names of all query sequences in the sub-job.
- Fixed a bug that resulted in adding a very large number of insertions at the beginning of an alignment.
- Plugged several memory leaks.

---

#### Changes for release 1.5 (03/27/03):

- Fixed a bug in query chopping that would result in duplicate alignments being reported for hits in "overlap" regions.
- Fixed several bugs in query chopping that would cause jobs to fail.
- Fixed a bug that caused searches with query chopping to report incorrect e-values.
- Added the pbd-slave rpm to facilitate installation on systems with local disks on worker nodes.
- Updated the XML output format to match that of NCBI 2.2.5.
- Improved error handling in the pb client code so that pb will usually return with non-zero exit status when errors are encountered.
- Fixed the bug that caused formatting of the latest version of the

NR protein database to crash when the "-o T" option was used.

- Fixed the bug that caused some searches against the latest version of the NR database to crash.