

READ ME FIRST!

Installing Mascot for the first time?

If you are installing Mascot for the first time, please read the 'last minute additions' on page 2 and the 'platform specific notes' on page 5 before following the installation instructions in Chapter 2 of the Installation and Setup Manual. A complete, up-to-date copy of the Installation and Setup Manual can be found on the Mascot installation CD, (*manual.pdf*).

Upgrading from an earlier version of Mascot?

Please read the 'last minute additions' on page 2 and the 'platform specific notes' on page 5 before following the upgrade instructions on page 3. Note that the latest version of the Installation and Setup Manual can be found on the CD-ROM (*manual.pdf*).

New: Protein Family Report

The Protein Family Summary displays search results a page at a time, allowing the results of even the largest searches to be browsed without client-side memory or browser limitations. Proteins are grouped into families using hierarchical clustering, making protein inference more intuitive and more accurate.

New: Percolator Support

Percolator is a utility from the University of Washington that uses machine learning to re-rank the peptide matches. The goal is to maximise sensitivity for a given false discovery rate.

New: Search Multiple Databases

Multi-select more than one database for a search. This is useful when you want to search a single organism database and include the sequences of common contaminants, such as BSA and trypsin.

And lots more

- Export search results as mzIdentML.
- Batch automate quantitation with Mascot Daemon.
- Support for mzML format peak lists.
- 64-bit executables for Windows.

Changes since last release

Details of individual changes since the previous major release of Mascot are listed on page 7 of these notes.

Last minute additions to documentation

See page 2

Last minute additions to documentation

- Some third party applications require helper scripts to be installed on the Mascot web server. If Mascot security is enabled, you should be aware that such scripts may create security holes.
- Microsoft Internet Explorer 8 no longer allows text to be pasted into a file upload control. This may break the integration between some older, third party software and the Mascot search form. If the software vendor does not have a fix, you may need to downgrade to Internet Explorer 7 or 6.
- Growth in the size of NCBIInr and UniRef100 has meant that there was no room on the Databases DVD for the IPI files referred to in Chapter 5 of the manual
- The executable called `nph-mascot1.exe` is for Mascot TD (“BIG” Mascot, where the precursor mass limit of 16 kDa has been removed). It will only be used for searches if enabled in the licence.
- The filesystem (NFS or a local file system) needs to support file locking and memory mapping. The following files will be locked/unlocked using the `fcntl(F_SETLKW,)` system call: `mascot.job`, `getseq.job`, `mascot.control`, `mascotnode.control`. If Mascot Daemon, Mascot Distiller on any application using the task management functions in `client.pl` are used, then there will be a `task_id` file in each `data/yyyymmdd` directory that will be locked/unlocked. The following files will be memory mapped for r/w: `mascot.control`, `mascotnode.control`. The location of these files can all be specified in the options section of `mascot.dat` so that if necessary they can be put on a local filesystem.
- Microsoft Internet Explorer 6 is noticeably slower than other browsers when displaying the new, Protein Family Summary. If possible, use one of the other supported browsers (`help/browser_help.html`).
- Keyword Aliasing (page 16 of the Setup & Installation manual): A step has been missed from the instructions. It is necessary to add an alias to the Apache configuration. Add the following `ScriptAlias` entry, immediately **before** the `ScriptAlias` for `/mascot/cgi`:

```
ScriptAlias /mascot/cgi/htsearch /usr/lib/cgi-bin/htsearch
```

On CentOS/RHEL, `/usr/lib/cgi-bin` should be replaced with `/var/www/cgi-bin`

You may also need to add the following if you get 403 errors, especially if you have Mascot defined in a separate virtual host:

```
<Directory /usr/lib/cgi-bin>  
    Order allow,deny  
    Allow from all  
</Directory>
```

Upgrading Mascot

To upgrade from a previous version of Mascot

1. Ensure that no-one will try to use Mascot during the upgrade procedure. Ideally, halt the web server.
2. Kill the *ms-monitor.exe* process
3. Rename your existing Mascot parent directory to (say) *mascot_old*. Create a new parent directory with the original name, e.g. *mascot*.
4. Follow the instructions in chapter 2 of the Setup and Installation manual to perform a new installation of Mascot into the new parent directory. Follow the procedure through to completion, and verify that the new system is functioning correctly with the supplied SwissProt database.

NOTE: If you were using non-default values for *MaxSequenceLen*, this should be changed in the new copy of *mascot.dat* **before** starting *ms-monitor.exe*

5. Kill the *ms-monitor.exe* process
6. If you have customised any of the Perl scripts, you may wish to merge your changes into the new Perl scripts. Mascot Perl scripts all have the extension *.pl* and are found in the following directories:
 - mascot/bin*
 - mascot/cgi*
 - mascot/x-cgi*
7. If you have customised any of the HTML pages, you may wish to merge your changes into the new HTML pages found in the *mascot/html* directory and subdirectories
8. Move any sequence databases that were under the old sequence directory to the new one.
9. Move all of the data subdirectories and their contents from beneath the old data directory to the new one. **Do not** move the *test* subdirectory or any of the files found in the *data* directory itself apart from *mascot.job*.
10. Move *searches.log* from the old *logs* directory to the new one. This will preserve the log of your existing results. You probably won't want to preserve the old *errorlog.txt* and *monitor.log*.
11. There have been changes to the main configuration file, *mascot.dat*. We recommend that you do a file compare (diff) between your old version of *mascot.dat* and the new one created by the installation. Make a note of any existing settings that you want to preserve.

Changes in the Databases, Parse, WWW or Options sections can be made using a text editor or using the browser based database

maintenance utility. This can be launched from a link on the home page. (If you halted the web server, you'll need to re-start it).

Any changes in the Taxonomy, Cluster or UniGene sections of *mascot.dat* must be made using a text editor.

12. In Mascot 2.1 and earlier, modification definitions were stored in a configuration file called *mod_file*. Mascot now takes its modification definitions direct from an XML representation of the Unimod database. **Note:** Whenever *unimod.xml* is updated, *mod_file*, *masses*, and *substitutions* files are created automatically to support old client applications that require these files. Do not be tempted to edit these files, because any changes will be lost the next time *unimod.xml* is updated. If you have custom modifications, you will need to add them to *unimod.xml* by using the new Configuration Editor
13. Other files that you may have modified are *enzymes*, *taxonomy*, and *fragmentation_rules*.
14. Restart Monitor:

```
cd /usr/local/mascot/bin
./ms-monitor.exe
```
15. The upgrade is now complete. If any compressed sequence files have to be rebuilt, the system may be busy for some time.

To upgrade Mascot Daemon:

Follow the link from the Mascot home page

Platform Specific Notes

Perl

The following compilations of Perl are supported by Mascot Parser 2.3 when a component of Mascot Server 2.3:

.so filename	Linux	Solaris
mspartner58.so	X	X
mspartner58-thread-multi.so	X	X
mspartner58-64int.so		X
mspartner58_64.so	X	X
mspartner58-thread-multi_64.so	X	
mspartner510.so	X	
mspartner510-thread-multi.so	X	
mspartner510_64.so	X	
mspartner510-thread-multi_64.so	X	

Linux

Kernels / Installations Supported

Mascot is tested during development with CentOS 4.6 (equivalent to RedHat Enterprise Linux 4.6) and Debian GNU/Linux 5.0 "Lenny".

Any distribution that includes a 2.6 kernel, glibc 2.3.4 or later, and libstdc++ 3.4.3 or later should be sufficient.

32 bit Linux

To use the 32 bit binaries, *after* untarring mascot.tar, overlay this with the 32 bit binaries:

```
tar xvf mascot-32.tar
```

Starting and stopping the Monitor service

In the mascot/bin directory, there is a file called mascot. Move this to the /etc/init.d directory with permissions rwxr-xr-x and owner root:root. As root, type:

```
chkconfig --add mascot
```

Files Greater than 2 GB

Fasta files greater than 2 GB are fully supported on ext2 and ext3 partitions.

Apache Server

If Apache is installed using a RedHat RPM, then the default user is “apache” rather than “nobody”. The “mascot.tar” file has been created using “nobody:nobody” (UID and GID 99) rather than “apache:apache”. Therefore you should chown -R all the files to “apache:apache”. Similar considerations apply to Ubuntu and Debian, where the default user is www-data:www-data.

Solaris

Solaris versions supported

Mascot is supported on Solaris 8, 9 and 10. Solaris on Intel (or AMD) is not supported.

UltraSPARC II Processors

UltraSPARC II processors are no longer supported. Mascot 2.1 was the last release that support the old 32 bit processors.

32 bit Solaris

To use the 32 bit binaries, *after* untarring mascot.tar, overlay this with the 32 bit binaries:

```
tar xvf mascot-32.tar
```

Recommended Patches

The patches listed below (or later versions) are recommended on Solaris systems running Mascot.

Solaris 8	108434-14, 108435-14
Solaris 9	111711-12, 111712-12,
Solaris 10	No patches required!

Use the “showrev -p” command to see what patches are installed. These patches and more information is available at:

<http://sunsolve.sun.com/>

Changes since Version 2.2.00

ID	Title
203	Support multiple Fasta databases
599	Installation - service name missing under Windows 2000
781	Delete a database from mascot.dat, still shows up in Database Status
1108	ignore_ions not working properly
1754	Non-linear progress reports in error tolerant search of large data set
1765	New installation program required
1792	Spurious M00251 messages. Warning : Taxonomy 'Viruses' ignored.
1834	Some URL's in Entrez reports in Protein View are relative
1883	Daemon check for Analyst 1.x is not correct
2062	Master results needs to display flag for unique vs shared
2136	Select summary (unassigned) has check boxes but no repeat search button.
2191	Add /LARGEADDRESSAWARE flag for Windows
2196	Crash when .dat or taxonomy file being updated on server when search starts
2209	New functions for client.pl
2233	Installer fails to set permissions adequately
2236	mzData ignores spectra with precursorList count > 1
2246	ms-getseq.exe and ms-gettaxonomy.exe not working for guest user
2338	Accession numbers should not be displayed in the status bar.
2341	Modifications on search form should use single master list
2378	Reference file entry not completely retrieved for first seq in SwissProt using AC rather than ID
2382	Changes to not.mascot.dat for Windows installer
2393	Support for QUERYLIST to enable repeat search from Select summary
2433	ET search or search with accessions on NA db is wrong
2434	peptide match parameter discrepancies for multiple NLs
2438	Reproducible crash in CMassComponent::makeQueryArray
2441	RDBI Daemon parameter fails to clear decoy/error tolerant
2442	DNA modifications are not output in Peptide Summary
2443	progress report is slightly wrong for NA searches with FRAMES
2445	nph-daemon_helper.pl missing from shipper (required by GPS Explorer)
2447	Confusing error message from Daemon with security and Integra user
2448	get_params.pl fails with Integra system user when security is enabled on Linux
2449	Modifications from hidden list not escaped in search form

ID	Title
2454	mod_file should be re-written if a user edits it
2455	Changes requested to modifications configuration editor
2457	Installer fails if IIS web site is bound to a specific interface or URL contains an IP address
2458	Installer needs to set permissions on unimod.xml and quantitation.xml
2464	Add the fixed mod delta to the XML export
2465	Mascot Daemon 2.2 no longer works with IIS integrated authentication
2466	Need to leak bytes in additional place for quant reports
2467	ms-monitor fails when shutting down under Windows with ms-lockmem enabled
2470	Support higher charge states (current limit 8+)
2471	PMF performance improvements
2472	Multiple instances of too many ions error in mzData causes nph-mascot.exe to crash
2473	Cannot have space in lcq_dta.exe path name in Daemon
2475	Quant. Want option to require modified residue to be present.
2476	Mascot is ignoring mzData retention time information
2477	Local modification being used in the wrong spectra?
2478	Perl scripts ouputting http headers with charset=ISO-8859-1
2479	Invalid etags can get matches
2481	Error tolerant search: discrepancies between Peptide Summary and Peptide View
2482	performance improvements for etag searches
2483	In reports, modifications could be hyperlinked to the Unimod entry
2484	Export help not clear
2485	getseq.pl takes a sessionID parameter, but doesn't pass it on to ms-getseq.exe
2486	Kludge modification names for GPS
2487	Add proper support for Vista
2488	Installer needs to configure IIS 6 to allow cgi and x-cgi
2489	The Mascot paths in the registry now have a trailing slash.
2490	Installer needs to install new config file called mod_aliases
2491	Configuration Editor bugs roll-up
2492	Mascot Security issue with ms-getseq.exe. Change to getseq.pl
2494	Daemon fails to report all Distiller errors
2496	SwissProt taxonomy is partly broken...
2497	NA searches sometimes crash while writing results file
2498	Missing graphs - improve error reporting in scripts
2502	nph-mascot.exe opens all compressed files on node. Should only open required ones?

ID	Title
2503	emPAI documentation updates
2505	Add button to display all matches in Peptide View
2506	URL arguments for quant formatting not being passed to export script
2507	Daemon drag and drop to repeat a search problems
2510	Search crash after failing to lock with ms-lockmem
2511	MS/MS search speed improvements
2512	primary_nl string confused if two matches have identical score
2515	Cannot remove multiple privileges with single click in security admin
2516	Add raw result file to export list
2518	Daemon should not halt a task if Distiller produces multiple peak lists
2519	State of hide error tolerant checkbox not passed to protein view
2520	Protein view and enzymes created by config editor are incompatible
2522	Increase range of charge states available in lcq_dta_shell
2524	Intermittent crash with integrated ET mode
2529	Table of matches at bottom of peptide view doesn't always agree with master results
2530	Scripts and parser had different upper limit on significance threshold
2533	ms-makesearchlog.exe crashes when corrupt result files encountered
2534	Create windows installer patch package
2535	Daemon external process parameter "<resultfilepath>" doesn't return a full path like the help says.
2536	automatic BIG mascot mode
2538	Intermediate file name is missing or empty. [M00398] - misleading error message.
2539	Command string for Daemon external processes is truncated unnecessarily at 255 characters
2540	If ITOL not set Mascot will search anyway
2541	If heavy or light isotope not found in Unimod, error messages are unclear
2542	Master results slow after fix for 2529
2543	Legacy upgrade doesn't remove old ARP entry
2544	High PMF score for single large protein with no enzyme (crazy search)
2550	Isotope correction for multiplex quantitation not implemented correctly
2559	Decoy stats shouldn't be displayed in protein summary for mixed ms and ms-ms
2561	Configuration editor limits number of residues in cleave and restrict fields to 10. Limit should be 20
2562	Protein View fails for hit in peptide summary with a protein score below threshold
2563	Spurious error message: Error [M00267 - Job -1 - X00875:ms_fileutil] - Failed to lock file
2564	Zero length fasta file causes ms-monitor.exe to crash

ID	Title
2566	Specific case where sequence tag fails
2567	ms-searchcontrol.exe should not output warnings if they will confuse client
2568	quantitation normalisation not consistent with help
2569	Quantitation not reporting protein ratio when there is a single peptide ratio
2571	Daemon should automatically remove MGF header lines when merging files
2572	Database maintenance corrupts extended characters which can lead to SetEntriesInAcl errors
2573	Swapping database with SeparateLockMem=1 and /3GB /PAE
2574	Update Daemon help re connection strings for SQL Server
2575	Add security to client.pl
2578	ms-gettaxonomy.exe fails if you request a tax id of -5 or lower
2580	Small memory leaks
2581	minor ms/ms scoring bugs (no visible symptoms)
2582	Need paged report for very large result files
2584	Decoy search on cluster only performs decoy part on first node
2589	ms-review - outputs invalid html if there is no searches.log file
2595	User parameters not HTML escaped during report generation
2597	No reason why tolerance needs to be set wide for multiplex quantitation
2598	Changing the case of a database name causes problems in Windows
2601	Supply a (sample) contaminants database
2602	Increase limit for number of databases to 256
2605	Improved reporting of protein hit
2609	PMF Mixture false positive
2612	Crash with large mod and very large peptide
2613	negative modification deltas with BIG mascot are output to result file as huge numbers
2616	Over-merging peptide matches when delta is the same
2618	Charge state in result different from input file
2619	Incorrect delta mass written to result file when fixed and variable apply to terminus
2620	Should probably include merged.dmp in mascot.dat (and supply it) to reduce M00384 warnings.
2622	Order of etag-results and of pmf queries is variable in result file
2623	Results slow to load in a browser because javascript var lines too long / large
2626	Protein view gives: "Fatal error no source defined" when database name contains parentheses
2629	Rare crash on Windows cluster. No message in error log files
2631	Progress reports for ET search inaccurate

ID	Title
2632	Upgrade projects to VS2005
2633	No point having sort unassigned control in a (protein) select summary
2635	Uniref100 help page needs updating
2637	Daemon failing to post correctly to https (ssl enabled) Mascot server
2639	Add support and doco for new MGF field "RAWSCANS".
2644	ProteinsInResultsFile set to 2 doesn't include rank 2 peptides with score same as rank 1
2646	Protein view: Fatal Error protein sequence has changed since the search was performed when there are base substitution in an error tolerant search
2647	Add support for UltraSPARC T1 and T2, and IV+ processors
2648	Minor CSV format changes for export script
2650	Mascot Daemon crash if number of formats supported by Distiller is reduced (e.g. uninstall Agilent MassHunter support)
2659	Don't use peptide length limit in error tolerant search
2662	Configuration Editor should use the system locale
2663	Crash with multiple subclusters
2664	PKL files with charge as float breaks Daemon merge
2666	Quantitation method can not be found [M00473] when there are strings containing multiple consecutive spaces in name.
2668	protein_view.pl should take enzyme definition from the results file
2669	BIG mascot mode has a few problems
2670	Error M00248 - Invalid taxonomy name class (authority) and (unpublished name)
2671	new version Uniprot SwissProt 56 has changed format for fasta
2673	Spurious 'Too many modifications' message for quantitation method
2674	Need to change certain text labels in ms_config Instruments section
2675	small databases with mutiple threads are not iterated properly
2676	Daemon follow up loses track of Distiller project after first follow up
2678	Illegal division by zero at ./quant_subs.pl line 3653
2679	Cannot have same modification appearing in more than one component in quantitation method
2680	ms-config crashes on empty environment
2681	Crash when writing to ipc.log, or rarely other log files. The ctime function is not thread safe
2683	Inconsistent report of quant method mods at the bottom of peptide summary.
2685	Mascot Integra Daemon not starting tasks with a long parameterset name
2686	If two local definitions in a quantitation method, second is ignored
2688	Change minimum ms-ms fragment tolerance in search form so that users don't have to switch units to mmu

ID	Title
2689	Parser functions requiring a string must not be passed a perl number
2692	implement mzML input data format
2693	More problems with large variable mods
2694	Can't lock both SwissProt and NCBIInr on Server 2003 (x64) cluster
2696	Selecting Merge MS/MS into single search and Distiller peak list format mzData gives error
2697	Crash with average quantitation protocol and no variable mods.
2699	pass spectrumID as SCANS in mzData
2701	Masses in help page are out of date
2702	Daemon fatal error with myODBC 5.1.5
2705	Need documentation for different neutral loss types somewhere
2706	index.html on databases DVD needs updating for 64 bit windows
2707	lcq_dta_shell.pl still using 1.008 for proton mass
2709	Change Daemon to have default Min. Scans / Group of 1 for lcq_dta.exe
2711	Crash on public web site with multiple very large modifications (glycosylations)
2717	Export with show_queries=1 or query_master=1 will run out of memory on large result file
2719	Some very short peptides with very low scores missed in multi-threaded search
2720	taxonomy ignored warning is reported twice
2721	Support IPv6 in the search log
2722	Search of Trembl crashes before it can even get started when "constrain search" is true.
2723	Crash with ETD-CID Instrument
2725	Configuration Editor allows duplicate values for report_ratio name
2726	Select summary should not collapse matches for different charge states
2733	Database maintenance doesn't set number of threads correctly in some cases
2735	ms-ms mzData file fails with out of memory if first spectrum is ms with 8000 peaks
2738	Add support for Distiller quantitation into Daemon so as to enable batch automation
2740	Add charge states "2+, 3+, and 4+" to search form for ETD
2741	In Daemon, support MDRO IPeakListFormatOptions4 functions
2742	mascot.dll memory leak has become unacceptable. Add work-around to Daemon
2743	Add a security task that set a separate query limit for no enzyme searches
2744	Test search fails when there are multiple sub-clusters
2745	Daemon should provide proper UI for web authentication username and password
2746	semi-specific enzyme or ET search on NA databases doesn't properly count number of Xs

ID	Title
2748	Failure to read .stats file doesn't report an error
2751	Error M00251 - taxonomy indexes ignored when searching SwissProt with large number of threads.
2752	homology thresholds can differ in cluster and standalone mode or with different chunks
2753	Mass of electron not considered when reading from dta files.
2754	discrepancies in mod deltas as output to dat-file
2757	Update Configuration Editor to support quantitation_2.xsd
2760	quant_subs.pl crashes when normalisation either average or median and some peptide matches have no ratios
2761	Integrate Mascot Percolator
2763	Add support for analysisXML (now called mzIdentML)
2764	Export script fails to include error tolerant matches in pepXML
2766	Update IPI help pages
2767	Allowed p value for significance threshold should be $1E-18 < p < 1$
2768	iTRAQ reporter ion masses don't allow for loss of electron
2769	Mascot Daemon Parameter editor needs to allow multiple fasta
2770	ms-monitor checks far too frequently for fasta file changes
2771	If a set of values is identical, should report SD(geo) as 1.0, not NN
2778	Add MALDI-ISD option to Instruments file
2780	Daemon needs heartbeat function for Integra
2784	Add copy and paste for command line to export script
2785	If file path and name are too long Daemon can not open the data file
2786	Mascot Daemon hyperlink to Protein Family Summary for large MS/MS
2790	Peptide view not listing fixed mods specified in quant method
2791	Config editor hangs when deleting quant ratios
2793	Save an index into the mgf/pkl/dta file for mzIdentML export
2797	Update test search and change template name to do_not_delete.asc
2799	Duplicate fasta filenames in different paths cause confusion
2804	Update multiplex to handle isobaric peptides (IPTL)
2805	Taxonomy entry with apostrophe breaks search form cookies
2814	Speed up for mzIdentML exporter when creating fragment ions
2815	Fragment masses wrong when ET mod applies to fixed mod site
2823	If IgnoreIonsScoreBelow is not zero in mascot.dat, cannot set to zero in form
2824	ms-monitor.exe crash when multiple threads write to monitor.log at same time