GenoMS Tutorial

GenoMS can now be launched using the executable 'main_specnets'. The 'main_specnets' executable sits in "**<sps_dir>/bin/main_specnets**". Usage is identical to when running CSPS, with one additional command line option '-q'

Example :

```
~/sps/bin/main_specnets sps.params -q -ll 9 -lf log.txt
```

See 'Analysis of MS.pdf' for explanation of the other command line options. Currently, GenoMS cannot be run in a distributed manner on a compute cluster. Fortunately, it tends to be pretty fast and should take only a couple hours for large experiments.

Parameter file

In addition to the parameters described in 'Analysis of MS.pdf', GenoMS has several other required and optional parameters that should be included in **sps.params**.

Required parameters

Parameter name	Valid values	Description
DBCOMBINED	Any valid file name	The file containing proteins to be used as templates
TEMPLATECONSTRAINTFILE	Any valid file name	The file describing the relationship of protein templates in the file DBCOMBINED

There are 3 built in databases that can be used.

Database File Name	Auxiliary File	Description
Mouse_Rat_HC_LC.fasta	Mouse_Rat_HC_LC.constraints	Contains all sequences from IMGT for mouse and rat.
Mouse_Rat_HC.fasta	Mouse_Rat_HC.constraints	Contains only the heavy chain sequences from IMGT for mouse and rat.
Mouse_Rat_LC.fasta	Mouse_Rat_LC.constraints	Contains only the light chain sequqnces from IMGT for mouse and rat

Optional parameters

Parameter names	Valid values	Default	Description
FDR_CUTOFF	0-1	0.01	The fdr cut-off for the database search results.
DIGEST	Trypsin,Chymotrypsin, Other	Other	The protease used for the experiment
FIXEDMOD	Amino Acid,Mass	None	Any modification fixed to all of a particular amino acid, such as C,+57 or C,+99
PEAK_PENALTY	Any	None	The peak penalty penalizes predicted peaks that have spectra which do not support the peak (recommended)

Example parameters file

```
# System parameters
INSTALLDIR=~/sps
REPORT_DIR=./report
EXE_DIR=$INSTALLDIR/bin

# SGE parameters
GRID_NUMNODES=100
GRID_NUMCPUS=1
GRID_SGE_EXE_DIR=/opt/sge62/bin/lx24-amd64
GRID_EXE_DIR=$INSTALLDIR/bin

# Input files
REPORT_TITLE=Test_project
FASTA_DATABASE=./data/homolog_prots_LC.fasta
AMINO_ACID_MASSES=../bin/AA_cys_iaa.txt
INPUT_SPECS_MS=./data/aBTLA_LC_AspN_042707.mgf;./data/aBTLA_LC_chymot
rypsin_042707.mgf;./data/aBTLA_LC_pepsin_30min_042707.mgf;./data/aBTL
A_LC_pepsin_3h_042707.mgf;./data/aBTLA_LC_trypsin_042707.mgf;./data/a
BTLA_hybrid_LC_DTT_IAA_AspN_ON_100407.mgf;./data/aBTLA_hybrid_LC_DTT_
IAA_chymotryp_30min_100407.mgf;./data/aBTLA_hybrid_LC_DTT_IAA_chymotr
yp_3h_100407.mgf;./data/aBTLA_hybrid_LC_DTT_IAA_tryp_30m_100407.mgf;.
/data/aBTLA_hybrid_LC_DTT_IAA_tryp_ON_100407.mgf

# Main parameters
TOLERANCE_PEAK=0.4
TOLERANCE_PM=1.0

# Preprocessing parameters
CLUSTER_MIN_SIZE=1
CLUSTER_MODEL=LTQ_TRYP
MIN_SPECTRUM_QUALITY=0.1
```

```
CORRECT_PM=no
GUESS_CHARGE=no
# Alignment parameters
MIN_OVERLAP_AREA=0.45
RESOLUTION=0.1
FILTER_TRIGS=yes
MIN_MOD_MASS=-100
MAX_MOD_MASS=100
MIN_RATIO=0.4
MAX_PVALUE=0.05
MIN_MATCHED_PEAKS=4
PARTIAL_OVERLAPS=1

# CSPS parameters
SPS_PROJECTS=sps_projects.txt

# Parameters for tag-based selection of homologous proteins
TAG_LEN=6
MAX_AA_JUMP=2
DOUBLE_AA_JUMPS=1
MATCH_TAG_FLANKING_MASSES=0
MAX_NUM_MODS=2
MIN_MATCHED_PEAKS_DB=7

#GenoMS Parameters
DBCOMBINED=Mouse_Rat_LC.fasta
TEMPLATECONSTRAINTFILE=Mouse_Rat_LC.constraints
FIXEDMOD=C,+57
PEAK_PENALTY=1
```