Installation of RECOG local server

Building data on RECOG local server

Supported OS †

- Linux (currently tested on CentOS5 [http://www.centos.org/])
- Mac OS X (10.4 or later, Intel)

↑

Requirements †

The following external programs are needed to be installed on your computer before you install the RECOG system.

Software	Mandatory	Explanation	Remarks
Perl	Yes	Scripting language	Version 5.8 or more
GCC	Yes	Compiling C programs	
MySQL	Yes	Database management system	Version 5.0 or more
Apache	Yes	HTTP server	Version 2.0 or more
Curl	Yes	File transfer	
BLAST	Yes	Similarity search	Version 2.0 or more
ClustalW	No	Constructing multiple alignments	
MAFFT	No	Constructing multiple alignments	
PHYLIP	No	Constructing phylogenetic trees	
GD Library	No	Drawing comparative genome maps	

↑

Download †

The latest version of RECOG is available at:

http://bias.nibb.ac.jp/RECOG/

Downloaded file is recog-server-x.y.z.tar.gz, where x.y.z is a version number.

↑

User setting †

We recommend you to create a specific UNIX user account for administrating the RECOG system because some

setting required for RECOG administration (e.g. MySQL setting) might conflict with your own setting. In the following, we assume the following administrator user account.

Item	Value
User name	recogadm
Group name	recog

All of the administrative operations below should be done by this user.

<div align="right">↑</div>

File extraction ↑

In the following, we assume you will install the recog server on the user's home directory ($HOME) and downloaded file is saved in /tmp directory.

```
% cd $HOME
% tar xfz /tmp/recog-server-x.y.z.tar.gz
```

A new directory named RECOG-server is created. Hereafter, we call this directory RECOG_HOME.

<div align="right">↑</div>

Setup the RECOG server ↑

We assume RECOG_HOME is $HOME/RECOG-server and operations are done by the user "recogadm".

1. Setting environmental variables
   - sh/bash

     ```
     % export RECOG_HOME=$HOME/RECOG-server
     % export PERLLIB=$RECOG_HOME/perllib
     ```

   - csh/tcsh

     ```
     % setenv RECOG_HOME $HOME/RECOG-server
     % setenv PERLLIB $RECOG_HOME/perllib
     ```

2. Execute the configure script
   - For a single node server (without any queueing system)
     Execute the configure script as follows

     ```
     % cd $RECOG_HOME/perllib
     % autoconf
     % ./configure ¥
     > --with-recogadm-passwd=recogadm_password ¥
     > --with-blast-dir=/usr/local/blast ¥
     > --with-mysql-dir=/usr/local/mysql
     ```

     - If the configure does not detect some commands, use the following options

Option	Explanation
--with-recogadm-passwd	The password of the administrator（recogadm）in the web interface.
--with-blast-dir	The path for the BLAST commands
--with-clustalw-dir	The path for the clustalw command
--with-mafft-dir	The path for the mafft command
--with-imagemagick-dir	The path for the ImageMagick commands
--with-mysql-dir	The path for the MySQL commands
--with-curl-dir	The path for the curl command

   - For a multiple node cluster with a Network Queuing System (currently RECOG supports Sun Grid Engine

only):

```
% cd $RECOG_HOME/perllib
% autoconf
% ./configure ¥
> --with-recogadm-passwd=recogadm_password ¥
> --with-blast-dir=/usr/local/blast ¥
> --with-mysql-dir=/usr/local/mysql ¥
> --with-nqs-dir=/usr/local/sge/OS ¥
> --with-nqs-pub-dir=/home/USER/local/pub_nqs ¥
> --with-nqs-max-nodes=3 ¥
> --with-nqs-queue=queue_name
```

- If the configure does not detect some NQS specific commands, use the following options

Option	Explanation
--with-nqs-dir	Directory for qsub/qstat commands
--with-nqs-queue	Queue name for submitting jobs to the NQS system Default: default queue set in the NQS system
--with-nqs-pub-dir	Directory shared among NQS nodes Results are stored in a directory named *UserName* under this directory.
--with-nqs-max-nodes	The maximum number of nodes used for submitting NQS sjobs
--with-sge-root	Set the environmental variable SGE_ROOT used in Sun Grid Engine
--with-sge-cell	Set the environmental variable SGE_CELL used in Sun Grid Engine
--with-sge-cluster-name	Set the environmental variable SGE_CLUSTER_NAME used in Sun Grid Engine
--with-sge-execd-port	Set the environmental variable SGE_EXECD_PORT used in Sun Grid Engine
--with-sge-qmaster-port	Set the environmental variable SGE_QMASTER_PORT used in Sun Grid Engine

3. Setup file for environmental variables
   Environment setup files are generated in the $RECOG_HOME/etc directory, which can be used to set the environmental variables required for the administrator's tasks. Read the files as follows before any administrative operation described below.

   - For sh/bash

     ```
     % . $RECOG_HOME/etc/recog.sh
     ```

   - For csh/tcsh

     ```
     % source $RECOG_HOME/etc/recog.csh
     ```

MySQL setup [↑]

Each RECOG server uses an independent MySQL server, so you need to set up a MySQL server even if your system already have one. The following procedures set up MySQL database for the RECOG server.

1. Configuring MySQL
   Move to the $RECOG_HOME/build/mysql directory and execute configure command. This command setup a new MySQL server including the setup of MySQL users and generating database. You must determine and remember the root password, name and password of mysql user for storing RECOG database.

Option	Explanation
--with-base-dir	Path for the MySQL server commands
--with-data-dir	Path for storing the MySQL database
--with-root-passwd	Password for the MySQL administrator (=root)
--with-user-name	MySQL user account name used by RECOG server
--with-user-passwd	Password for the RECOG user

- Example

```
% cd $RECOG_HOME/build/mysql
% autoconf
% ./configure ¥
> --with-base-dir=/usr/local/mysql ¥
> --with-data-dir=$RECOG_HOME/mysql ¥
> --with-root-passwd=root_password ¥
> --with-user-name=recog ¥
> --with-user-passwd=password
```

2. Testing MySQL server
   Execute the following command to check whether mysql server can be started correctly

```
% /usr/local/mysql/bin/mysql --defaults-file=$RECOG_HOME/etc/my.cnf -e 'show status'
```

- If it has successfully started up, then the following information is displayed.

```
+---------------------------------+----------+
| Variable_name                   | Value    |
+---------------------------------+----------+
| Aborted_clients                 | 0        |
```

(snip)

```
| Uptime_since_flush_status       | 23       |
+---------------------------------+----------+
```

- The following message is displayed if the MySQL server is failed to start.

```
ERROR 2002 (HY000): Can't connect to local MySQL server through socket '/mysql/mysql.sock' (2)
```

Tips on MySQL server [↑]

We assume the server machine on which MySQL server runs has at least 2GB of memory.
If your machine has larger memory size, increasing the following parameter by modifying
$RECOG_HOME/etc/my.cnf can enhance the performance.

- key_buffer_size
  The value is in default 512M. The following example increases it to 2GB.

```
key_buffer_size=2G
```

Utility commands for MySQL server [↑]

The following commands can be used to manage the MySQL server.

- Starting and stopping MySQL server
  - Start the MySQL server
    The following command start the mysql server. This should be done whenever you reboot the server
    machine.

    ```
    % cd $RECOG_HOME/bin
    % ./startup_mysql.sh
    ```

  - Stop the MySQL server
    The following command stop the mysql server. This should be done when you shutdown the server
    machine.
    You need to enter the MySQL administrator's password

    ```
    % cd $RECOG_HOME/bin
    % ./shutdown_mysql.sh
    ```

- Initializing the MySQL database [#a0def30a] The following command can be used to re-initialize the MySQL
  database.

[NOTE] This command delete all data and you cannot recover them.

```
% cd $RECOG_HOME/build/mysql
% ./create_recog_databases.sh
```

Setting up Apache httpd server

RECOG server also uses an independent httpd server. Execute the following to configure the server.

1. Copying httpd.conf file
Copy the system standard httpd.conf file into the RECOG system. The location of the standard httpd.conf file may be different between systems. For example, /etc/httpd/conf/httpd.conf in Red Hat Enterprise Linux and /etc/apache2/httpd.conf in MacOS 10.6. After copying it, you should add a line indicating to include the RECOG specific configuration as follows:
    - For Red Hat Linux (CentOS) (change the location of httpd.conf file for your system)

```
% cd $RECOG_HOME/WWW/conf
% cp /etc/httpd/conf/httpd.conf .
% echo "Include $RECOG_HOME/WWW/conf/httpd-recog.conf" >> httpd.conf
```

2. Edit the configuration file $RECOG_HOME/WWW/conf/httpd.conf
Modify the copied httpd.conf file to make the description in httpd-recog.conf effective.
   i. Modify the address and the port number of the httpd server.
      - Before

```
#Listen 12.34.56.78:80
Listen 80
```

      - After

```
#Listen 12.34.56.78:80
Listen 192.168.123.45:8080
```

      (192.168.123.45 is an example of IP address. Replace it with the address of your own machine. Also, 808 is the default port number in the RECOG system, but you can modify it if this number has already been used by other service)
   ii. Nullify the CustomLog directive (location of the httpd log files), which activate the location described in httpd-recog.conf file.
      - Before

```
CustomLog logs/access_log common
```

      The location might be different. For example, it might be

```
CustomLog /private/var/log/apache2/access_log common
```

      - After（Comment out by adding '#'）

```
#CustomLog logs/access_log common
```

   iii. Comment out the directive to load mod_ssl module.
      - Before

```
LoadModule ssl_module libexec/apache2/mod_ssl.so
```

      - After（Comment out by adding '#'）

```
#LoadModule ssl_module libexec/apache2/mod_ssl.so
```

3. Access control
[NOTE] For security reason, all connections are disabled in the initial setting. So you MUST modify it to enable connection at least from your machine on which you will execute the RECOG client
Initially, the access control is described in $RECOG_HOME/WWW/conf/httpd-recog.conf as follows:

```
    Order deny,allow
    Deny from all
#    Allow from host.example.com
```

- If you intend to allow access from any site, edit $RECOG_HOME/WWW/conf/httpd-recog.conf as follows (comment out the Deny directive):

```
#    Deny from all
#    Allow from host.example.com
```

- If you intend to allow access from limited sites, edit $RECOG_HOME/WWW/conf/httpd-recog.conf to add the IP addresses or the host/domain names of the allowed sites

```
    Order deny,allow
    Deny from all
    Allow from host.example.com
    Allow from .example.com
    Allow from 192.168.1.23
    Allow from 192.168.2.0/24
```

↑

Starting the httpd server ↑

1. Starting the httpd server
   Execute the following command to invoke the httpd server. Note that the command path of httpd might be different in your system.

```
% /usr/sbin/httpd -f $RECOG_HOME/WWW/conf/httpd.conf
```

2. Checking the invocation of the httpd server
   Access the following URL on your web browser
   http://(server IP address):8080/index.html
   (The port number 8080 should also be modified if you set the other port number)
   The server works well if the browser displays 'RECOG-server'.

↑

Stopping the httpd server ↑

1. Stopping the httpd server
   Execute the following command to stop the httpd server.

```
% kill `cat $RECOG_HOME/WWW/logs/httpd.pid`
```

Last-modified: 2010-12-08 (水) 14:36:35 (0m)

Site admin: anonymous

PukiWiki 1.4.6 Copyright © 2001-2005 PukiWiki Developers Team. License is GPL.
Based on "PukiWiki" 1.3 by yu-ji. Powered by PHP 5.2.12. HTML convert time: 0.217 sec.

Building data on RECOG local server

Installation of RECOG local server

Building data on RECOG local server [†]

- Building data on RECOG local server
- Data update (command line interface)

[↑]

Data update (command line interface) [†]

1. Setup environmental variables
   Set environmental variables by reading the setup file. Note that $RECOG_HOME is the top directory of RECOG server.
   - For sh/bash

     ```
     % . $RECOG_HOME/etc/recog.sh
     ```

   - For csh/tcsh

     ```
     % source $RECOG_HOME/etc/recog.csh
     ```

2. Update taxonomy database
   Download and update the taxonomy database from the following NCBI site
   - ftp://ftp.ncbi.nih.gov/pub/taxonomy/taxdump.tar.gz

     ```
     % cd $RECOG_HOME/build
     % ./m0110convTaxonomyDb.pl
     ```

3. Update CDD database
   Download and update the CDD database from the following NCBI site
   - ftp://ftp.ncbi.nih.gov/pub/mmdb/cdd/cdd.tar.gz

     ```
     % cd $RECOG_HOME/build
     % ./m0020cddDb.pl
     ```

4. Update additional databases
   Download and update additional databases from the following sites
   - KEGG: http://www.genome.jp/kegg/catalog/org_list.html
   - NCBI Genomes: http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi?dump=selected
   - GOLD: http://www.genomesonline.org/DBs/gold_published_table.txt

     ```
     % cd $RECOG_HOME/build
     % ./m0050kegg.pl
     % ./m0051ncbi.pl
     % ./m0052gold.pl
     ```

5. Update motif identifier table
   Download and update the motif identifier table from NCBI
   - ftp://ftp.ncbi.nih.gov/pub/mmdb/cdd/cddid.tbl.gz

```
% cd $RECOG_HOME/build
% ./m0120motifTab.pl -DOWNLOAD
```

6. Download published genome data from the MBGD server
   Download and update specified published genome data. To find the genome identifiers (gm numbers) see the following URL
   http://mbgd.genome.ad.jp/data/species/spid.tab

   ```
   % cd $RECOG_HOME/build
   % ./mirrorSpidData.pl gm00001 gm00002 gm00003 gm00004 gm00005
   ```

7. Input user genome data
   Skip here if you need not add or update your original genome data.
   See [Preparation of user genome data] to prepare and input the user genome data.
8. Check $RECOG_HOME/etc/spid.tab file Check the list of target species
   The following information is listed for each genome with the tab-delimited format.
   i. spid : species ID
      The public RECOG data is numbered in 'gm99999' format
      User genome data is numbered in 'gu99999' format
   ii. spname : species code
      The species code in 3-5 letters
   iii. dir-name : FTP directory
      This field is used only for published data. For user genome data, species name is filled here
   【例】
   http://mbgd.genome.ad.jp/data/species/spid.tab
9. Insert species data into the MySQL databae

   ○ Insert species data

   ```
   % ./m0400InsertSpecInfo.pl
   ```

   ○ Remove checkpoint files
     After execution, checkpoint file is created. Remove the corresponding checkpoint files if you want to re-execute some failed procedure.

   ```
   % /bin/rm -f $RECOG_HOME/work/.m04*
   ```

   ○ Check inserted data
     ■ check all genome data

   ```
   % ./viewUserGenomeData.pl -ALL
   ```

     ■ Check only specified species (with species ID)

   ```
   % ./viewUserGenomeData.pl gu00001
   ```

   ○ Remove data from the MySQL database
     Remove all data of specified genome from the MySQL database when there are some mistake. [NOTE] You cannot recover deleted data.
     ■ Remove data with species ID

   ```
   % ./removeUserGenomeData.pl gu01234
   ```

10. Definition of species set
    By default, all-against-all comparison will be done among all the defined genome sequences, but you can limit calculation by defining "species sets": if a species set is defined, RECOG calculates all-against-all similarity only among the defined species sets.

- Editing the species set file

```
% vi $RECOG_HOME/etc/spid_set/spid_set.tab
```

  - Format
    The following items are listed in the tab-delimited format.

Name	Explanation
name	Name of species set
species list	List of spid or species_code multiple species should be separated by a comma

11. Execute homology search
    - Execute homology search

```
% ./m0500ExecSearch.pl
```

    - Remove check point file
      Remove the corresponding checkpoint files if you want to re-execute some failed procedure.

```
% /bin/rm –f $RECOG_HOME/work/.m05*
```

12. Insert homology search result into the MySQL database
    - Insert data into the database

```
% ./m1000InsertSearchRes.pl
```

    - Remove checkpoint file
      Remove the corresponding checkpoint files if you want to re-execute some failed procedure.

```
% /bin/rm –f $RECOG_HOME/work/.m10*
```

13. Expand homology search result
    - Execute expansion and additional post-processing

```
% ./m1100BuildData.pl –all
```

    If you need to re-execute 1120CopyHomologyInfo.pl in the middle of the execution, you must remove the following files.

```
% /bin/rm –f $RECOG_HOME/work/homologyMd5.$PID
% /bin/rm –f $RECOG_HOME/work/homologySp.$PID
```

    - Remove checkpoint file
      Remove the corresponding checkpoint files if you want to re-execute some failed procedure.

```
% /bin/rm –f $RECOG_HOME/work/.m11*
% /bin/rm –f $RECOG_HOME/work/homologyMd5.INPUT
% /bin/rm –f $RECOG_HOME/work/homologySp.INPUT
```

14. Publishing the database

```
% ./m1200PublishInfo.pl
```

Last-modified: 2011-02-14 (月) 22:52:27 (0m)

Site admin: anonymous

PukiWiki 1.4.6 Copyright © 2001-2005 PukiWiki Developers Team. License is GPL.
Based on "PukiWiki" 1.3 by yu-ji. Powered by PHP 5.2.12. HTML convert time: 0.239 sec.

Preparation of user genome data

[Building data on RECOG local server](#)

Preparation of user genome data [†]

Your data should be prepared in GenBank format or a combination of FASTA sequence files and tab-delimited annotation files.

[↑]

GenBank format files [†]

If your data is already in GenBank format, what you should do is minimum.

1. Determine the species code of your genome
   Determine the species code of your genome based on the following rules:
   i. The length of species code is between 3 and 5.
   ii. The first letter should be a lower alphabetical character.
   iii. The other letters should be lower alphabetical or numeric characters.
   iv. The code should not be used yet for other species in the database.
      Species codes of the published genomes are basically same as that of the KEGG database, which are found in http://mbgd.genome.ad.jp/data/species/spid.tab
      (e.g., the species code of Haemophilus_influenzae is hin)
2. Generate a data directory and template files
   Execute the following command to create a data directory and template data files
   In addition to the above species code, you need to specify the number of chromosomes/plasmid (using -nchr option).

```
% cd $RECOG_HOME/build
% ./createUgSkeleton.pl -sp=species_code -nchr=num_chromosomes
Create skeleton for guXXXXX
```

   where guXXXXX is an automatically generated species identifier (XXXXX is a serial number). In this case, a data directory is $RECOG_HOME/species/guXXXXX and template files are created there.
3. Put your data in the created directory
   Put your data in the "GB" directory in the created directory ($RECOG_HOME/species/*guXXXXX*/GB). Create directory manually if there exists no such direcgtory.
   Name each file as 1.gbk, 2.gbk, ... according to the chromosome numbers (or your desired order of chromosomes).

```
% cd $RECOG_HOME/species/guXXXXX
% mkdir  GB
```

4. Conversion the GenBank file
   Execute the following command to extract data from the GenBank files.

```
% cd $RECOG_HOME/build
% ./m0310ConvGB2Mbgd.pl  guXXXXX
```

   Add -contig option when processing incomplete genome data.

```
% ./m0310ConvGB2Mbgd.pl  -contig  guXXXXX
```

Add also -NO_LOCATION option if you need not extract location information (beginning and ending positions on the chromosome sequence) of each gene.This is especially useful for metagenomic data.

```
% ./m0310ConvGB2Mbgd.pl  -contig  -NO_LOCATION  guXXXXX
```

FASTA sequence files and Tab-delimited text files †

You can also put the information with a combination of FASTA sequence files and tab-delimited text files. To do this, you should prepare the following:

1. Generate a data directory and template files
   Determine the species code of your genomes referring to the above "GenBank format files" section. Also, count the number of chromosomes/plasmids.
   Execute the following command to create a data directory and template data files

   ```
   % cd $RECOG_HOME/build
   % ./createUgSkeleton.pl  -sp=species_code  -nchr=num_chromosomes
   Create skeleton for guXXXXX
   ```

   where guXXXXX is an automatically generated species identifier and a data directory named $RECOG_HOME/species/guXXXXX is created.
2. Edit genome data
   Edit the file $RECOG_HOME/species/guXXXXX/gm/genome.txt to put genomic information. The file should include the following information:

Name	Mandatory	Explanation	Remarks
sp	Yes	species code	e.g.) 'hin'
abbrev	Yes	Abbreviated species name	e.g.) 'H.influenzae'
orgname	Yes	Full species name	e.g.) 'Haemophilus influenzae'
strain	Yes	Strain name	e.g.) 'Rd KW20'
taxid	Yes	taxonomy ID	e.g.) '71421'
type	No	genome type	complete / incomplete / metagenome default: complete
date_release	Yes	Release date (year-month-date)	e.g.) 1995-07-28'
specweight	No	Weight of species	default: (20300000−published_date)*100

   An example of genome data:

   ```
   sp <TAB> hin
   abbrev <TAB> H.influenzae
   orgname <TAB> Haemophilus influenzae
   strain <TAB> Rd KW20
   taxid <TAB> 71421
   date_release<TAB> 1995-07-28
   specweight <TAB>
   ```

3. Putting data for each chromosome/gene information
   Put your chromosome/gene data according to the following directions.
   You should specify at least gene information and protein sequences. Other information can be

omitted if no chromosome/contig map information is available.
In the following, $(ChrNo) should be replaced with a serial number of chromosome
(1,2,3,...).

i. Chromosome (or contig) information (optional)
   Edit the file $RECOG_HOME/species/guXXXXX/gm/data/$(ChrNo).chromosome.txt.

Name	Mandatory	Explanation	Remarks
name	Yes	Chromosome name	e.g.) 'chromosome 1'
shape	Yes	Shape of chromosome	'circular' or 'linear'
type	Yes	Type	'chromosome' , 'plasmid', or 'contig'
length	No*	Length	e.g.) '1830138'

(*) Omissible if you specify a chromosomal sequence file (DNAseq).
An example of chromosome data:

```
name <TAB> chromosome 1
shape <TAB> circular
type <TAB> chromosome
length <TAB> 1830138
```

ii. Gene information (mandatory)
   For each chromosome, put gene data in that chromosome as a file named
   $RECOG_HOME/species/guXXXXX/gm/data/'$(ChrNo).gene, where each line contains a
   gene information consisting the following items, which are described as a tab-delimited
   format.

Name	Mandatory	Explanation	Remarks
Gene_ID	Yes	Gene Identifier	Must be unique in the genome. e.g.) locus tag
Gene name	No	Gene name	
Start position	No	Start position on the chromosome sequence	
End position	No	End position on the chromosome sequence	Start < End
Direction	Yes	Direction	1/-1 Specify 1 if you omit Start/End positions
type	Yes	Type of gene	CDS/misc_RNA/rRNA/tRNA/scRNA/snRNA/snoRNA/...
Product name	No	Name of product	

iii. Protein sequence (mandatory)
    Put protein sequence data as a FASTA-formatted file named
    $RECOG_HOME/species/guXXXXX/gm/data/$(ChrNo).protseq.
    Entry name of the sequence should correspond to the gene identifier (Gene_ID) in the
    gene information file described above.

```
>Gene_ID
MAIKIGINGFGRIGRIVFRAAQHRDDIEVVGINDLIDVEYMAYMLKYDSTHGRFDGTVEV
KDGNLVVNGKTIRVTAERDPANLNWGAIGVDIAVEATGLFLTDETARKHITAGAKKVVLT
```

```
GPSKDATPMFVRGVNFNAYAGQDIVSNASCTTNCLAPLARVVHETFGIKDGLMTTVHATT
ATQKTVDGPSAKDWRGGRGASQNIIPSSTGAAKAVGKVLPALNGKLTGMAFRVPTPNVSV
VDLTVNLEKPASYDAIKQAIKDAAEGKTFNGELKGVLGYTEDAVVSTDFNGCALTSVFDA
DAGIALTDSFVKLVSWYDNETGYSNKVLDLVAHIYNYKG
```

    iv. Chromosomal DNA sequence (DNA seq) (optional)
Put chromosomal sequence data as a file named
$RECOG_HOME/species/guXXXXX/gm/data/$(ChrNo).chrseq.

    v. Gene DNA sequence (optional)

- If you already have gene sequence information,
put the sequence data as a FASTA-formatted file named
$RECOG_HOME/species/guXXXXX/gm/data/$(ChrNo).geneseq.
- If you do not have gene sequence information, but have chromosome sequences,
execute the following command to generate gene sequences using Start/End
column in the gene information file.

```
% cd $RECOG_HOME/build
% ./convertDna2GeneSeq.pl guXXXXX
```

- If you have neither gene sequences nor chromosomal DNA sequences,
do nothing and skip registering gene sequences.

4. Checking the data
The following command checks whether the data you input is defined correctly.
Correct data if you encounter some errors.

```
% cd $RECOG_HOME_build
% ./checkUserGenomeData.pl –spid=guXXXXX
```

- The following problems are reported as 'Error'.
    a. In genome.txt, one of sp/orgname/taxid/release field is null.
    b. In genome.txt, the specified species_code has already been used in
       $RECOG_HOME/etc/spid.tab.
    c. In genome.txt, Illegal date_release.
    d. In chromosome.txt, shape value is nether 'circular' nor 'linear'.
    e. In chromosome.txt, type is neither 'chromosome', 'plasmid', nor 'contig'.
    f. In chromosome.txt , both length and chrseq is empty.
    g. In gene information, some mandatory items (Gene_ID, dir) are null.
    h. In gene information, only one of the boundary positions (Start/End) is specified.
    i. In gene information, Gene_ID is not uniquely defined.
    j. No protein seq file exists for some genes.
- The following possible problems are also reported as 'Warning'; some of them will be
  automatically fixed by the program when you specify -UPDATE option (see below).
    a. In genome.txt, abbrev field is empty. Generate based on orgname defined in
       genome.txt .
    b. In genome.txt, specified taxid cannot be found in the Taxonomy database. Set taxid
       as 32644 (unknown).
    c. No chromosome/chrseq file corresponding to some genes.
    d. In chromosome.txt, length field is null. Automatically generate from the DNA seq
       data.
    e. In gene information, boundary position does not satisfy the required condition Start
       < End. In such a case, two positions are automatically exchanged.
       However, if it satisfies (chromosome length / 2) < abs(End - Start), RECOG
       considers that the gene spans the origin of a circular genome and does not
       exchange the positions.
    f. Protein or DNA sequences contain non-alphabetical characters. Such characters are
       automatically removed.

5. Automatic problem fix

After having solved the detected problems, execute the following command to fix the problems reported in the Warning messages. If you encounter errors again, fix them manually and re-execute the command.

```
% cd $RECOG_HOME_build
% ./checkUserGenomeData.pl -spid=guXXXXX -UPDATE
```

Last-modified: 2010-12-24 (金) 10:47:01 (52d)

Site admin: anonymous

PukiWiki 1.4.6 Copyright © 2001-2005 PukiWiki Developers Team. License is GPL.
Based on "PukiWiki" 1.3 by yu-ji. Powered by PHP 5.2.12. HTML convert time: 0.237 sec.