Instructions for Curators:
Entering new data into the QTLdb

Version 0.7, Feb. 7, 2014

This instruction manual is for <u>general curators</u> using the on-line QTLdb Editor to add new and/or to update existing QTL data in the database.

1.   GETTING PREPARED – Data qualification and preparation

1) Identify a data source: Data to be curated must be from (1) original report of new QTL results from QTL experiment supported with statistical evidence; or (2) published paper in a publicly available scientific journal (i.e., no review papers or summary of other people's work).

2) Have electronic versions (PDF or MS Word) of the report or paper ready to be archived.

3) Find the PubMed ID of the paper (if published) from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed and save the paper with a file name as "pubmed_id.pdf" (e.g., 55526372.pdf). If no PubMed ID can be found for the paper, or the report is yet to be published, an "ISU" number will be assigned. (See Heading 3 below for details. When the PubMed ID becomes available, please notify the database administrator, and we will replace the temporary "ISU" number with the new ID.)

4) Read through the paper and make sure it contains the following information: 1) source of the paper (journal, issue, year, authors, etc.; ignore this if pre-publication report); 2) QTL experiment and analysis methods (animals, design, statistical model, etc.); and 3) QTL results (name and description of the traits, chromosomal locations, flanking markers, statistical evidence).

5) Get a printed copy of the paper and attach a blank page to it for taking necessary notes during the curation (e.g., write down the PubMed ID or ISU reference ID, mark things that you may need to check back later, etc.). *It is always a useful practice to take notes, or mark on the paper, what you have entered into the database, to make it possible for any error to be traced back.* **Keep this raw record on file for at least 12 months for possible data debugging needs.**

2.   USING THE EDITOR

1) Use a web browser to open the AnimalQTLdb editor with URL: http://www.animalgenome.org/cgi-bin/QTLdb/edit/login

2) Precautions using the editor. Please consider these tips:

   a)   When you are ready to submit a form, click the "submit" button only once; do not click multiple times – this may cause error.

b) Be sure to "logout" when you are done or will be away from your computer. The best way for security is to simply close the browser window *entirely*.

c) In the event of network hang, or situations when you have to quit while not sure if a filled form is saved into the database, you may consider saving your browser page so you can follow up when you return.

d) Read the instructions on each editor page/form carefully before you decide to click a button. It is always a good practice to be clear about where you are at and what you are doing.

3) Report any bugs. Although contingencies are built in and the editor was tested, bugs/exceptions may exist. Please include useful information such as the URL, title of the form, steps you took, and copy the errors, along with a brief description of the problem, to send as an email bug report to bioinfo-team@animalgenome.org. Details help a lot for proper diagnoses, correctly reproducing the error, and timely responses.

4) The editor interfaces with the database by organizing curation activities into **3 groups** of forms: <u>reference information</u>, <u>experiment information</u>, and <u>QTL information</u>. Within each group of forms, there are **2 sets** of forms for action: one is for <u>entering new data (*Add New*)</u>, and one for <u>updating existing data (*Edit*)</u>.

5) *Data privilege*: A general curator will only be able to change (update) the data originally entered by him/her.

6) The sequence to enter new data is:
Step 1. Enter reference data (once for a paper/report).
Step 2. Enter experiment data (once for a paper/report).
Step 3. Enter QTL data (maybe multiple times for a paper).

7) The Editor tool has made data entry a cascade process. You may take breaks after you have completely submitted a form (no error prompt), provided you know where it was when you return. It is not recommended that you leave a half-filled form on an unattended computer.

8) After you have submitted a form, DO NOT try to "Go Back" on the browser to "submit" again. This way, you would duplicate the same record in the database.

3.   ENTERING REFERENCE INFORMATION

The 2007 version of QTLdb included a new component in the curator/editor tools that does periodical (semi-) automated searches of PubMed for new QTL data publications, and provides curators with a new paper review panel to determine if a paper is one that needs to be curated.

If you have a paper in hand to enter, go to step "3" directly. Or, start at "1":

1) Upon login, click the "Check for new papers → **Get New** " button. This will list all recent QTL papers of your species that are not yet in the QTLdb.

2) Review each paper by its title/abstract to determine if it is a qualifying paper for entering:
   - If it is not a qualifying paper, **Dismiss** it.
   - If it is a qualifying paper, click the **Curate** button to bring the information into the "reference curation" form to start curation (*Go to step 4*).

   Hint: *It is discouraged to add information for multiple papers before you can actually finish curation on the previous one.*

3) Upon login, click the "Reference Data → **Add New** " button. This will bring up a form for you to enter Reference data (information about a paper).

4) If you already have a PUBMED ID at hand, we now (2013) have a tool to help you bringing data from PubMed database to the entry form. Simply enter the ID in "Check PUBMED ID" field and click the **Check** button to proceed.

5) **Tips**:

   a) Make sure you have a PubMed ID (always preferred), or an "ISU" ID in the "Ref_ID" field. Use this ID to refer to the work you have done when debugging may be needed.

   b) Use a consistent format for "Issue", i.e., **32(1):322-329**, or **32:322-329**; leave the year to "Year" field.

   c) For a report (un-published), fill in the "Journal" and "Issue" field with "**TBA**" (to be added) in order to proceed.

   d) You may copy an abstract from PubMed but not directly from the original paper due to copyright issues. You may have to come up with your own abstract if one is not available from PubMed.

4) **Continue**: After you submit reference information, there are buttons on the returning page that give you 2 options: edit the current reference info you just submitted; or continue to enter the experiment data (the editor remembers which paper it is for).

5) **Break**: If you take a break at this point (*make a note!*), when you return, click "Experiment Description → Add New ", then choose the Reference ID from a pull-down menu, and go on from there.

4.   ENTERING EXPERIMENT DATA

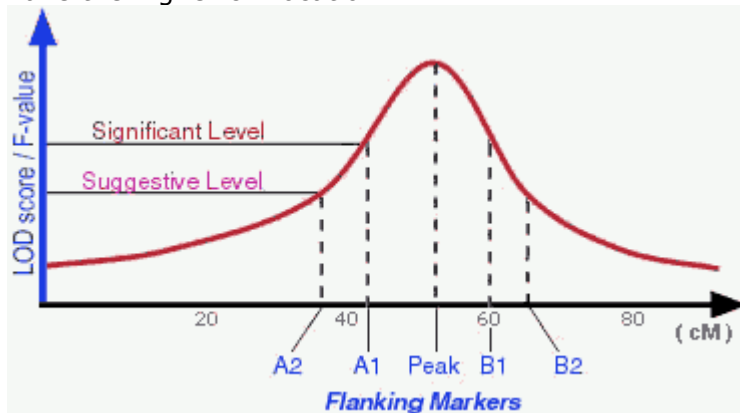1) You continued from 3.3 or 3.4. This will bring up a form for you to enter Experiment data.

2) **Tips**:

    a) The Animals field should include information on the breeds used in the study.

    b) The Design field should include the mating strategy (e.g., half-sib family, granddaughter design), as well as a list of the traits analyzed.

    c) The Analysis field should describe how the data were analyzed.

    d) It is required that you fill all blanks except "Notes", which is optional.

    e) Make sure there is a correct Reference ID (PubMed ID or ISU ID) in the field.

3) **Continue**: After you submit experiment information, there are buttons on the returning page that give you 2 options: | edit | the current experiment info you just submitted; or | continue | to enter the QTL data (the editor remembers which paper it is for).

4) **Break**: If you take a break at this point (*make a note!*), when you return, click "QTL Details → | Add New |", then choose the Reference ID from a pull-down menu, and go on from there.

5. ENTERING QTL DATA

1) You continued from 4.3 or 4.4. This will bring up a form for you to enter QTL data.

2) **Tips**:

    a) Please note all fields with a red star (*) by the field name are mandatory. No form will be accepted without all of them being correctly filled.

    b) QTL symbol: While you may create one with sensible abbreviations, it is recommended that you use one that is in existence.

    c) Breeds: Look up the numeric breed ID from the link and fill in the cell; In case of multiple IDs, separate them with a comma.

    d) Reference Map: MARC (cattle and pig), Wageningen (chicken), and International Mapping Flock (IMF) maps serve as the backbones of the underlying maps.   All other map locations are interpolated to these maps using anchoring markers (see below).

    e) Position (center of a QTL; cM): When MARC/Wageningen/IMF map location is given, use that; however, in most cases, one has to interpolate map locations in the QTL analysis to a MARC/Wageningen/IMF map location. Use the comparative map tool in the Arkdb for this purpose: http://www.thearkdb.org/. For map location (cM) between 2 **closest** markers, for example a QTL peak, the USDA-MARC tool is better for interpolation: http://www.marc.usda.gov/genome/swine/swine.html or http://www.marc.usda.gov/genome/cattle/cattle.html

f) QTL span interval (range): When the range information is available, enter it, as in "23.1-48.6" (with a dash in between). When there is not range information available, **leave it blank**.

g) Flanking markers A1 and B1 are landmarks for QTL that pass *significance threshold* (may or may not be available). By convention, numbering occurs from the "top" of the chromosome, so the "upper" marker will have the lower cM location, and the "lower" marker will have the higher cM location.

h) Flanking markers A2 and B2 are landmarks for QTL that pass *suggestive threshold* (may or may not be available). By convention, numbering occurs from the "top" of the chromosome, so the "upper" marker will have the lower cM location, and the "lower" marker will have the higher cM location.
Check to make sure the marker you enter is in the database. Use "**Find markers**" tool to check for existence of a marker in the database. See Heading 7 for more details.

i) Associated QTL: Enter QTL_ID for each QTL "associated". In case of multiple IDs, separate them with a comma. The "**associated QTL**" refer to QTL that share common markers, candidate genes, or other grounds, used in independent QTL/Association analysis or reports.

j) Haplotype: Fill in the haplotype designation in the order reported, and link each element with a hyphen (-), as in:
$GAGT_{10}-TGAT_{10}-GGGT_{10}-GGGT_9$
In cases where multiple haplotypes are involved at one locus (usually within a gene), enter all significant ones separated with a semicolon, as in:
$H1-H3 \; ; \; H2-H4$

k) Trait ID: Click "Find trait ID" to look up the ID to use for a Trait. If you don't find a particular one, see Heading 6 for more details.

l) Candidate gene ID/name/symbol: Fill when applicable.

m) Chromosome: Use drop-down menu to select numbers or "x"/"y".

n) Map span interval (cM): Map location interpolation; see 5.2(d) above.

o) Statistics: You must fill in at least one of the following: LOD score, LS mean, P-value, F Statistics, Variance, Bayes value, Likelihood ratio (if more available, fill all).

p) Test scale: Choose correct option from drop-down menu when available.

q) Model tested: Choose correct option from drop-down menu when available.

r) Significance level: Choose "significant" or "suggestive" from drop-down menu.

s) QTL effects: Fill when available.

t) Epistasis: Epistatic links between QTL are entered separately. Go from the main panel, Add New → Continue to enter epistasis data .

u) Comments: Write down your personal notes as curator, when deemed necessary. This is to note special information that is not covered by the existing form fields.

3) **Continue**: After you submit QTL information, there are buttons on the returning page that give you 2 options: edit the current QTL info you just submitted, or continue to enter another QTL from the same reference paper (the editor remembers which paper it is for).

4) **Break**: If you take a break at this point (*make a note!*), when you return, click "QTL Descriptions → Add New ", then choose the Reference ID from a pull-down menu, and go on from there.

5) **Update**: the "QTL Description → Edit " button can be used to update the QTL information by choosing a QTL entry from a pull-down menu.

6.   FINISHING STEPS

When you finish entering all data from a report/paper, you will need to
1) Upload the report/paper to the database using the "Reference Data -> Upload PDF" button, using the file name convention explained in 1.3, so that the report/paper will be correctly linked to the database queries.
2) After the paper has been uploaded, return to the New Candidate QTL Papers list and **click the Done button to mark it.**

7.   ABOUT TRAIT DATA

For general curators:
The Trait data curation has not been released to general QTL curators at this time. Trait curation is a complex and evolving process that is ongoing in a separate Animal Trait Ontology (ATO) project. The ATO traits are available to QTL curators once they have been approved by the ATO team. Most, if not all, traits are already available in the QTLdb for you to choose from. If you

don't find a particular trait within the current QTL database, send the trait description to the administrator to get it scrutinized against the ATO database and possibly entered before you may proceed. Once it is entered, you can use the "**Find trait ID**" function (5.2(g)) to bring it into your QTL data form.

For trait editors:
QTL Trait to VT/PT/CMO mapping tool:

QTL Trait ontology edit tool:

8. ABOUT LINKAGE MAP DATA

Curation for Linkage Map data has not been released for general curators at this time. All USDA-MARC, Wageningen, and Sheep Linkage Map v4.7 map markers (plus about 100 non-MARC markers for pigs) are currently in the database.

The chance for you to encounter a rare marker is small. However, if you do encounter one, send it along with the interpolated map location to the administrator to get it entered.

9. VIEWING YOUR DATA

Regardless if your data is released to the public or kept as "private" data, or is pending administrative actions, you can always view your data internally using the curator tools, either alone or in comparison to other public data.

Click the QTL details → **Preview** button to open up a new window, select the reference you wish to view, and further select the QTL to bring into viewing table or chromosome graph.
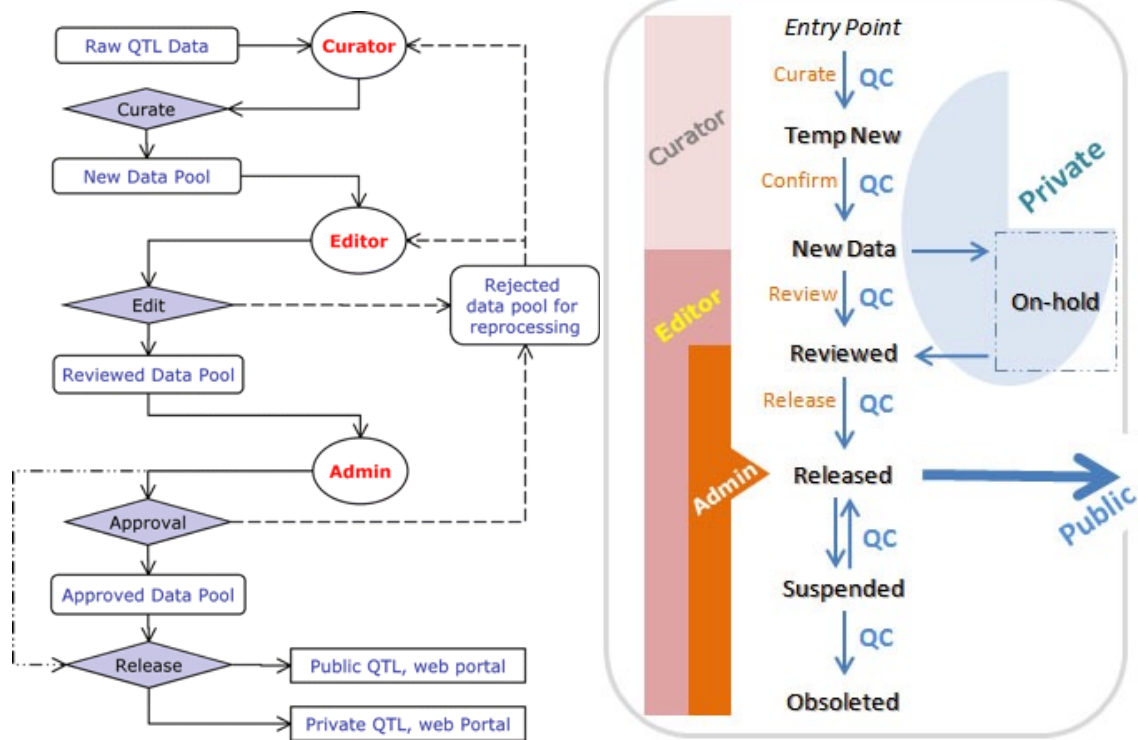
Tips: By browsing to QTL details from a public link, you can click the "Edit" link to update the information on that QTL using your curator's login/password.

10. ABOUT PRIVATE DATA AND PUBLIC DATA

Newly entered data are pending review (by "Editors") and approval (by the "Admin") before they can be officially accepted as part of QTLdb. Once the curated data pass **review** and **approval**, they can be seen from the public URL. If you wish to keep your data as "private", to be viewed only internally by yourself (using QTLdb as a research platform), you need to notify the administrator for the batch of data you wish to be "**released**" as private data.

11. UNDERSTANDING THE CURATION WORK FLOW

Data curation is an integral part of data processing flow and workflow in the QTLdb data management. It is useful for a curator to know the data flow (Figure 1) in order to properly communicate with the "editor" and the "administrator". Often, a curator can also serve as an editor to review or examine the QTL data in the pipeline.

Data status:

CONTACT: Human beings make mistakes. But don't panic – whenever you encounter problems, errors, or get confused, contact Zhiliang Hu (see below) with details. There may be pitfalls in the curation tools, or functions to be added.

Zhiliang Hu
Associate Scientist
Dept of Animal Science
Iowa State University

```
  Tel.: 901-878-2120
 Email: zhu@iastate.edu
 Skype: zhiliang.hu
```