GEP Project Management System: TSS Project Submission

Author

Wilson Leung   wleung@wustl.edu

Document History

Initial Draft   08/21/2015

Version

GEP Project Management System (Version alpha)

Introduction

In addition to the coding region annotations, the GEP has also created a collection of transcription start site (TSS) projects where students are asked to estimate the TSS position and the TSS search region for each gene. These TSS projects will generally rely on coding region annotations that have previously been submitted by GEP students and reconciled by the GEP staff.

In order to facilitate the verification and reconciliation of the TSS annotations submitted to the GEP, all GEP annotation project submissions should contain a completed TSS report form. This report form describes the evidence that were used to support the TSS annotations and the distribution of core promoter motifs surrounding the TSS. The TSS report is part of the TSS package that you have downloaded when you claimed the project. (The TSS report form is also available under the "Specific Issues in GEP Annotation Projects" section of the GEP web site.)

In some cases, the reconciled gene models might need to be revised (*e.g.*, because of a misannotation or because of revisions to the reference *D. melanogaster* gene models by FlyBase). In these cases, you will need to prepare three additional supplemental files that contain data for all the **revised gene models** in your project in three different formats: General Feature Format (GFF), transcript sequences, and peptide sequences. You can use the Gene Model Checker to validate each gene model and generate these supplemental files for each revised isoform. For each type of supplemental file, you can use the Annotation Files Merger to combine the individual files for each isoform into a single project file. Consensus errors that interfere with the TSS annotations should also be documented using the Sequence Updater.

Submitting TSS Projects

Procedure:

1. Login to the GEP Project Management System (Figure 1)
   - Navigate to the GEP Project Management System through the GEP home page (http://gep.wustl.edu ➔ "Projects" ➔ "Project Management System")
   - Enter the credentials issued to you by the GEP:
     - Username: [<first letter of first name><last name>]
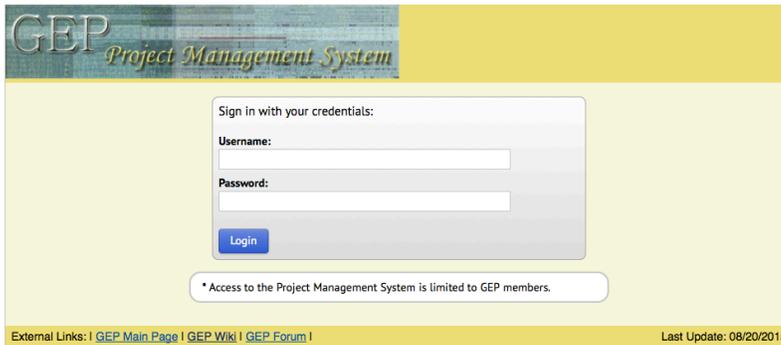   - Password: [<issued by the GEP>]
   - Click "Login"

Figure 1 Login screen for the GEP Project Management System.

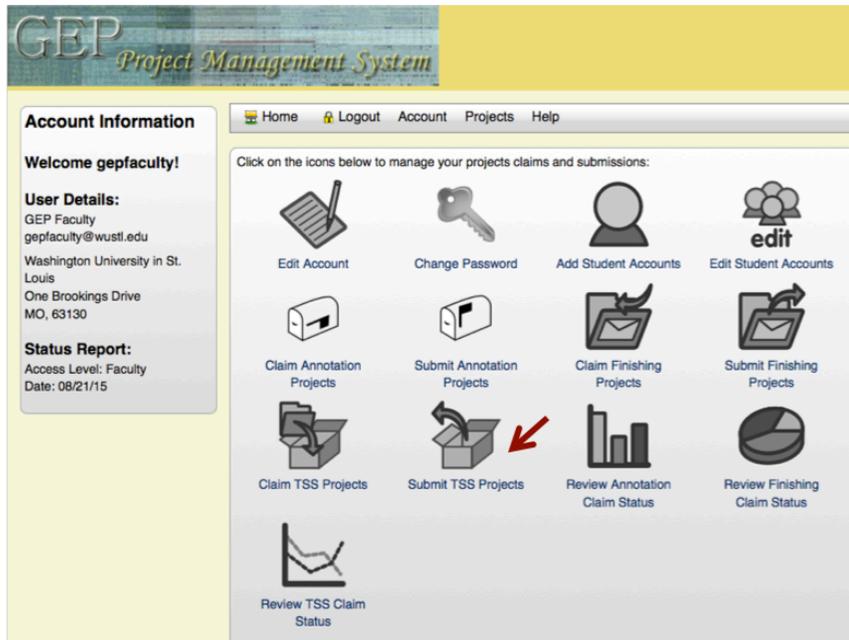2. Click on the link "Submit TSS Projects" link on the main page (Figure 2).

Figure 2 Click on the "Submit TSS Project" link to submit TSS projects.

3. Select the appropriate project and the completed TSS report form. Then click on the "Submit Project" button (Figure 3).

Figure 3 Submitting a TSS project by selecting the project (*e.g.*, dbiarmipes_dot_Aug2013_contig1_tss) under the "Project Name" field and the project report under the "TSS Project Report" field.

4. If all goes well, you should see a message that confirms the project has been successfully submitted to the GEP (Figure 4).
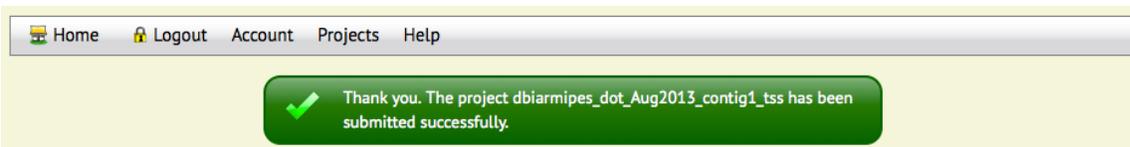

Figure 4 Confirmation message from the Project Management System indicating that the TSS project has been submitted successfully to the GEP.

> **Note:** Please send an email to Wilson (wleung@wustl.edu) to confirm that the GEP has received your project submission and that it contains all the information required for project reconciliation.

Submitting Annotation Projects with Revised Gene Models

During the TSS annotation process, you might encounter cases where the reconciled gene models are incorrect because of a misannotation or because of a subsequent update of the *D. melanogaster* gene models (*e.g.*, isoforms that have been added or removed by FlyBase). In these cases, your students should construct new gene models using the GEP annotation protocol and document the evidence used to support the revised gene model in the "Revised gene models report form" section of the TSS report form.

For each revised gene model, your students should use the Gene Model Checker to verify the gene model and to create the three corresponding set of files that are required for all GEP gene annotations (i.e. GFF, transcript, and peptide sequence files). For each file type, the supplemental files for all the revised gene models should be combined into a single project file prior to project submission. See the "Gene Model Checker User Guide" and the "Annotation Files Merger User Guide" for details (available under "Help" ➔ "Documentations" ➔ "Web Framework" on the GEP web site).

To submit a project with revised gene models, select the project under the "Project Name" field, select the completed TSS report form under the "TSS Project Report" field, and then select the checkbox next to the "This submission contains revised gene models" field. Three additional fields will appear in the submission interface where you can upload the GFF, transcript, and peptide sequence files for all of the revised gene models (Figure 5).

Figure 5   Select the "This submission contains revised gene models" checkbox to upload the merged GFF, peptide, and transcript sequence files for the revised gene models.

Submitting TSS Projects with Consensus Sequence Errors

For projects with errors in the consensus sequence that affect the TSS annotations, students should use the Sequence Updater (available through the GEP home page under "Projects" ➜ "Annotation Resources") to document the changes to the original project sequence. If there are multiple consensus errors in the project, students can use the Annotation Files Merger to combine all the VCF files generated by the Sequence Updater into a single VCF file prior to project submission.

> **Note:** Consensus errors that have been identified during the coding region annotations are available through the "Consensus Errors" track on the GEP UCSC Genome Browser (under the "Mapping and Sequencing Tracks" section). The submission should only include the VCF file if the project contains **additional** consensus errors that have not been documented previously.

To submit a TSS project with errors in the project sequence, select the project under the "Project Name" field, select the completed TSS report under the "TSS Project Report" field, and then select the checkbox next to the "This submission documents additional consensus errors" field. A new "VCF File" field will appear on the submission form that will allow you to upload the combined VCF file. For projects with annotated genes, you will also need to upload the combined GFF, transcript and peptide sequence files as described above (Figure 6).

Similar to the protocol for the annotation of the coding regions, all the TSS and coding regions coordinates reported in the submission should be **relative to the coordinates of the original project sequence** irrespective of whether there are errors in the consensus project sequence.

Figure 6 For projects with additional errors in the consensus sequence, you should select the "This submission documents additional consensus errors" checkbox and upload the combined VCF file as part of the submission.

Additional Resources

Please contact Wilson (wleung@wustl.edu) if you encounter any problems or have any questions regarding the protocol for submitting TSS projects.