

Reflection for Final Project

Anindita Varshneya

Statement of Work

As coder, I was responsible for making all necessary customizations to GenMAPP Builder. I created the customized species profile, and then added additional modifications that allowed us to collect our gene IDs from the ORF tag instead of the OrderedLocusNames tag in the XML file. I also modified the TallyEngine to return results indicating how many gene IDs exist in the different data tables. I created a final build that fixed the data table labeling problem that appeared in build 4, but we were unable to run a new export with these new changes to confirm that they were done correctly due to a time crunch at the end of the semester. I also contributed to the presentation and report. At the beginning of the project, I also read and contributed to the Genome Presentation with the help of Brandon Litvak. I also checked that the Gene Database Schema was reflective of the information in our database and added a title to our schema accordingly with the help of Brandon.

As project manager, I helped create goals for the group project weekly. After creating the slides representing the coding work done in the presentation, I also created some of the introduction slides and did minor formatting changes throughout the presentation. In addition to writing about the coding results and inputting the changes that I made to the code, I wrote the abstract and a majority of the introduction, and worked with my group to input figures and tables and write appropriate captions for them, as well as properly include citations throughout our report. Our entire group also worked together to proofread the paper.

My wiki pages:

- https://xmllpipedb.cs.lmu.edu/biodb/fall2015/index.php/Anuvarsh_Week_11
- https://xmllpipedb.cs.lmu.edu/biodb/fall2015/index.php/Anuvarsh_Week_12
- https://xmllpipedb.cs.lmu.edu/biodb/fall2015/index.php/Anuvarsh_Week_14
- https://xmllpipedb.cs.lmu.edu/biodb/fall2015/index.php/Anuvarsh_Week_15

Assignments I worked on:

- https://xmllpipedb.cs.lmu.edu/biodb/fall2015/images/a/a1/Genome_Presentation_-_Anu_and_Brandon_-_Genialomics%28FIXED%29.pdf
- https://xmllpipedb.cs.lmu.edu/biodb/fall2015/images/5/5e/BioDB_Final_Presentation_GEN_20151214.pdf
- <https://xmllpipedb.cs.lmu.edu/biodb/fall2015/images/e/ee/Genialomics-DatabaseSchema-20151211.pdf>
- Final Report

Code I contributed to:

- <https://github.com/lmu-bioinformatics/xmllpipedb/blob/b-cenocepacia/gmbuilder/src/edu/lmu/xmllpipedb/gmbuilder/databasetoolkit/profiles/BurkholderiaCenocepaciaUniProtSpeciesProfile.java>

- <https://github.com/lmu-bioinformatics/xmlpipedb/blob/b-cenocepacia/gmbuilder/src/edu/lmu/xmlpipedb/gmbuilder/resource/properties/gmbuilder.properties>
- <https://github.com/lmu-bioinformatics/xmlpipedb/blob/b-cenocepacia/gmbuilder/src/edu/lmu/xmlpipedb/gmbuilder/databasetoolkit/profiles/UniProtDatabaseProfile.java>

Assessment of Project

Our group worked incredibly well throughout this project. Every class meeting, we all updated each other on the status of our personal contributions to the project, and helped set up deadlines for major milestones and assignments. The entire team worked together well and communicated well with each other throughout the project through our Facebook group message. The team was also super supportive of each other and always offered to help pick up pieces of one person was falling behind in their tasks, or reminded each other of due dates for personal assignments/reflections. If I could go back and do it again, I would probably have started the report earlier as it ended up being much more work and much longer than I had initially imagined. Despite having to complete it in the four days prior to its due date, I think it came out well, and that our entire group came together to contribute to it very well.

I think the database project is of good quality. The only improvement that should have been made earlier is the one regarding the TallyEngine results. I think the analysis of results from GenMAPP and MAPPFinder from Kevin and Veronica were insightful and very interesting. I think our Genialomics home page is organized fairly well as it is easy to access all of the required information within a few clicks. I had no problem finding important and relevant information from any of my teammates individual journals throughout the project. I did not check every teammates individual pages, but those that I found I needed access to were easy to find and easy to maneuver.

I think the group did complete all of our objectives. The only thing that was not completed (which is incredibly unfortunate and is something I would want to do as soon as I have the opportunity) is run an additional export ensuring that the changes I made to the TallyEngine data table names worked out the way I hoped.

Reflection on the Process

In terms of biology, I learned quite a bit about analyzing pathways from MAPPFinder data and GO terms. When we did our first run through GenMAPP and MAPPFinder, I had so much difficulty trying to put the pieces together, but after looking at the results this time, with help from Kevin, I learned quite a bit about how to analyze and put together the different GO terms to come up with a conclusion regarding how different pathways are affected. In terms of computer science, I learned quite a bit about how so many different pieces of data come together to create a comprehensive database.

I found that checking in with my team just to hear how they were doing in their part of the project really helped me put together the entire picture of what we were doing. I worry that if I did not take the initiative to do so from the beginning, I would have had a much harder time understanding the important results from our project. I also appreciated

how well our group worked together and contributed to the project. It helped me stay motivated and put out my best work because I knew that the rest of my group was working hard to make sure we had a quality end product as well. Communication was key to make sure this happened, and I think our group did a good job keeping everyone in the loop.

I learned a lot about maneuvering through different databases, and understanding what distinguishes a good database from a bad database. Looking back to the NAR journal articles we looked at, I think I would have had a better analysis of the different databases knowing what I know now about what information I might want to have at my fingertips right when I access a database. I also learned quite a bit about creating builds and using Eclipse as I had not used this IDE prior to this project.

Despite everything that I learned, one aspect of this project that will most likely stick with me the most was the process of analyzing changes in different biological pathways by understanding the different terms/aspects to that pathway that could have increased or decreased expression. I hope that in a year from now I will still be expanding my knowledge regarding this type of analysis, as it was probably the most exciting part of the project for me. A lesson more applicable to daily life that I will take away from this project is the importance of honest dedication and good communication in a team. If my entire team did not care about this project so much and communicate with each other so well, I don't think this project would have been quite as enjoyable as it was.