Hi everyone, I’m Yuanrong Liu. I have just graduated from a bioinformatics program this year, and I am working for professor Manoj after my graduation. I will present the process of my work today. It’s am online implementation of a RNA sequencing pipeline.

The pipeline I used is called iRAP, the integrated RNA sequencing analysis pipeline. It’s a very flexible pipeline, and contains more than 30 methods for the RNA sequencing data analysis, including the mapper, quantification methods, differential expression methods. And the basic run for the pipeline is just use the function irap in a unix or gnu system. And the parameter is a configuration file that contains all the options and parameters for the analysis. I will talk about more later.

My goal is to create a user friendly interface, and let user to run the iRAP pipeline remotely on the web server. The methods I used to build the website includes HTML, CSS, and the I used PHP, JavaScript, MySQL for the backend, and python to manipulate some data. The website uses an Apache server, and is deployed by the Amazon Web Service, Elastic Compute Cloud.

The website includes these pages. And the structure of the sever looks like the right side. The html folder includes all php pages I wrote for the web server, includes staticfiles, scripts. And I set a separate file users in the html to store all user analysis result. In the users folder, there will be many folder for each registered user, and inside the use folder, there are folders with a job ids, that corresponds to each job the individual user run. Thus, user could run their analysis separately. Outside the html folder, it will have a data folder, this folder has the similar structure with the users folder to store user’s own data. There are also other folders or ap[lications locally. Include the genome data, MySql database data, and the pipeline application.

Next, I want to show a quick demo for the website.

For the work and features we already had. User could perform analysis online, and download and show the result like I shown before, we also have a login system to store user data separately, and the website has a user friendly interface.

There are also lots of problems and features need to be solve and improve. The biggest problem is the environment of the pipeline. Because it’s a big pipeline that contains 30 or 40 methods, the dependencies of the package is also huge. Currently, I have some problems with the dependency, so there are very limited combinations could be used now, and the result doesn’t contain any figure. Maybe it just because lack some modules in the environment. Also, although the website is user friendly, it loses some flexibility for the pipeline. Lastly, the analysis online now is just running directly on the ec2 instance, it will caused some problems like the efficiency, and freezing page on the website.

So for future works, I will fix the dependencies problem, create more options for users to customize their analysis, and I will also upload more species genome data to make the web application more flexible. AWS could computing services will also help us to solve the efficiency problem.