Reviewer 1  
  
Observations:  
  
The authors describe web based tool (SuperPhy) whose purpose is to simplify the workflow for the analysis of bacterial strains. SuperPhy integrate several sw to perform specific tasks (such as Panseq, FastTree) and databases (Chado). It includes about 100 E.coli strains in the current database. The web interface is availabe (in beta version) at [lfz.corefaciliy.ca](http://lfz.corefaciliy.ca/) and is operational. This article follows the article in BMC bioinforamtics 2010 descibing the pan-seq tool (Laing et al 2010). Essentially SuperPhy makes pan-seq easier to use and more useful for analyzing pan-genomes of bacteria. The paper is well written though could be shortened. Table 1 is probably redundant. More emphasis should be placed in the use cases expanding session 4, and the sw architecture could be reduced. Figure 1 is too big, while figure 3 is hard to read. Since emphasis is placed in doing analysis on-the-fly it could be useful to have some time measurements in which to distinguish user time and machine time to perform certain significant tasks.  
  
Suggestions for revision:  
  
1. High: More use cases, less architectural desciption  
  
2. High: Fifures need to be made more relevant to the use cases  
  
3. Med: Timing results could be also useful

Response:

Table 1 has been removed. Figure 1's size has been reduced and figure 3 re-arranged and enlarged. Unfortunately, we are not able to add additional figures due to space limitations.

The Design section has been reworked and condensed as recommended. We also elaborated on the functionality of SuperPhy by augmenting the Functionality section and further comparing the SuperPhy capabilites to other genome analysis platforms in a newly added Discussion section. Approximate times for the analysis results to be returned were added to the Uploading a Genome subsection.

Reviewer 2  
  
Observations:  
  
Excellent work and great and very useful website. Authors should correct minor typos in the text.

Response:

We have reviewed the text and fixed several typos.  
  
Reviewer 3  
  
Observations:  
  
This paper presents a new resource called SuperPhy for genome analysis of bacterial pathogens. The resource has a number of distinctive features compared to other similar services, of which the most notable in my view are the possibility of users to upload their own genome datasets and leave them private and tools to help determine virulence and antiimicrobial drug resistance factors. The paper would be stronger if it had some comparison with other similar systems. I can name a few: IMG/JGI, RAST, and PATRIC. The first two allow user genome upload, but they are generic prokaryote genome resources. The third does not allow user genome upload, but is very rich in terms of analysis tools, much richer than SuperPhy. Regarding metadata (section 2.3), the authors developed their own ontology, but it would have been much better if they had used already published proposals, such as Nature Biotechnology 29, 415–420 (2011).

Response:

A Discussion section was added in which we compared the functionality of SuperPhy to other genome analysis platforms including IMG, PATRIC and MicroScope. RAST was not included in the Discussion, as it appears to be purely a genome annotation tool. It is also a component of the PATRIC workflow.

We agree that using existing ontology specifications would improve the inter-operability of the SuperPhy database. We will revise and adapt the SuperPhy genome meta-data ontology to match the terms defined in minimum information about a genome sequence (MIGS) specification. Note, however, that SuperPhy is not intended as a primary sequence repository. Users are encouranged to submit their genome sequences to sources such as NCBI with the complete MIGS information. In SuperPhy, we then link back to these primary sources.  
  
Reviewer 4  
  
Observations:  
  
It is an interesting tool. It is a good contribution to the clinical genomics area. I recommend to include information about computer platforms, languages, etc used

Response:

A subsection "Implementation Details" was added to the Design section that lists the languages and packages used to build SuperPhy.  
  
Reviewer 5  
  
Observations:  
  
The back end details such as database schema and pipeline are of little interest to most readers, as would be details of programming techniques etc., thus this part of the paper could be reduced/compressed/summarised.  
  
Suggestions for revision:  
  
1. Med: The title .... analysis of Pathogens is a big overstatement of a resource that only contains E coli. Either the title should be changed to ... analysis of E Coli. Or the database contents should be expanded to include a wide and representative range of pathogens. If the database is in 'pilot' status then perhaps it's not ready for publication, or the title should say: SuperPhy: A pilot database for ...  
  
The authors should mention existing large-scale genome phylogenetic resources, e.g.  
Federhan, S. (2011) The NCBI Taxonomy database. Nucl acids res. 40(1) D136-D143  
Fang, H. et al. (2013) A daily-updated tree of (sequenced) life as a reference for genome research. Scientific Reports, doi:10.1038/srep02015.  
Letunic, I. and Bork, P. (2006) Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinforamtics, 23(1) 127-8

Response:

The Design section has been reworked and condensed as recommended. Also Table 1 has been removed.

A Discussion section was added in which we mention large-scale phylogenetic resources. These are compared to SuperPhy.

SuperPhy, in it's current form is a fully functional platform for the analysis of E.coli. The design and analyses generated by SuperPhy are applicable to many bacterial pathogens and it is our plan to incorporate other pathogens into the database. The reviewer indicated that this is not apparent from the title of the article. To reflect both the current state of SuperPhy and our future plans, we have modified the Title and Abstract to state that the database is currently a pilot for future expansion to include other bacterial pathogens.  
  
Reviewer 6  
  
Observations:  
  
The paper presents the new effort in the "comparison of multiple genomes" field. It is well written and easy to follow. The SuperPhy portal is available online and works fine. Only link to upload new genome gave me an error (at least on 17th - 22nd of October) Internal Server Error ([amanji@phac.gc.ca](mailto:amanji@phac.gc.ca)) Apache/2.2.22 (Ubuntu) Server at 199.116.235.48 Port 80 Comments on paper (sorted by paper structure): 1. The link to SuperPhy can be removed from Abstract 2. Introduction title font has a different from other sections format. 3. In the Introduction section: Maybe, Taverna and Galaxy are not the best candidates for comparison with SuperPhy. These systems are made to execute/describe more generic bio-workflows. Then, SuperSpy is a more research specific system. Do other candidates for comparison exist? 4. Fix format of this link "http:\\lfz corefacility ca/superphy" in Introduction. 5. Fix "A Contig Collection will" format in the section 2.1 Database Schema. 6. Critical discussion is absent. I cannot really judge how this system is useful for biologists in this research field. Maybe, the discussion section can give this information? Any numbers on users and system usage?  
  
Suggestions for revision:  
  
1. Med: The paper presents the new effort in the "comparison of multiple genomes" field.   
It is well written and easy to follow. The SuperPhy portal is available online and works fine.  
Only link to upload new genome gave me an error (at least on 17th - 22nd of October)  
  
Internal Server Error  
([amanji@phac.gc.ca](mailto:amanji@phac.gc.ca))  
Apache/2.2.22 (Ubuntu) Server at 199.116.235.48 Port 80  
  
Comments on paper (sorted by paper structure):  
  
1. The link to SuperPhy can be removed from Abstract  
2. Introduction title font has a different from other sections format.  
3. In the Introduction section: Maybe, Taverna and Galaxy are not the best candidates for comparison with  
SuperPhy. These systems are made to execute/describe more generic bio-workflows. Then, SuperSpy is a more research   
specific system. Do other candidates for   
comparison exist?  
4. Fix format of this link "http:\\lfz corefacility ca/superphy" in Introduction.  
5. Fix "A Contig Collection will" format in the section 2.1 Database Schema.  
6. Critical discussion is absent. I cannot really judge how this system is useful for biologists in this research field.  
Maybe, the discussion section can give this information? Any numbers on users and system usage?

Response:

We have fixed the Internal Server Error that the reviewer noted.

Items 1,2,4,5 were changed as suggested.

Regarding items 3 and 6, a Discussion section was added in which we compared the functionality of SuperPhy to other genome analysis platforms including IMG, PATRIC and MicroScope, as well as elaborate on the utility of the SuperPhy platform.