Written Project Proposal

Jesse Valliere
Dr. Stuart
Advanced Bioinformatics
Spring 2022

Background & History

Intracellular bacteria have been a notable area of study in recent years, providing examples of symbiosis with hosts that are still being studied today. Notable examples of intracellular bacteria and their symbiotic relationships include mitochondria and chloroplasts, which have evolved into every animal and plant species seen today. More recently, species of the *Wolbachia* genus have been of interest for their unique invasion of many species, and their selfish manipulation techniques they employ for their own benefits. Originally discovered in the early 20th century, *Wolbachia* are known to be maternally inherited intracellular bacteria that can manipulate reproduction of their host species in order to benefit themselves. These species infect insects, arthropods, spiders, mites, and even crustaceans, and thus demonstrate a wide array of possible hosts they can proliferate with. This proliferation is done through male killing, feminization, parthenogenesis, and cytoplasmic incompatibility (CI), and varies on the type of *Wolbachia* species being examined.

Cytoplasmic incompatibility has been found to be the most common form of reproductive manipulation, and involves embryonic lethality in crosses of infected males and uninfected females. This mechanism is fairly efficient in ensuring success of the *Wolbachia* species, as reproduction is rescued when both males and females both express two key genes, cytoplasmic incompatibility factor A (*cifA*), and cytoplasmic incompatibility factor B (*cifB*).[1] These genes are still being studied extensively, and are even being applied in large-scale attempts to control the spread of malaria, dengue, and the Zika virus from arthropods to humans.[2] An important thing to mention, however, is that these genes are rapidly evolving and have notable genetic variation, and so a need for a source of all this varying information has been seen.

Main Concept

The main concept I am proposing for in regards to my Advanced Bioinformatics project is to create a database, tentatively named the *Wolbachia* CIF Gene Database, that can accurately store and retrieve entries of CIF genes and their sequences. This would include vital information for each entry, such as a Gene ID, the arthropod host it was found in, the strain of *Wolbachia*, information on where it was collected (city, state, country), the CIF type, and the respective sequence. An example of an entry is shown under the Example Interface outline. This information could then be searched, primarily based on the *Wolbachia* strain or arthropod host of interest, and their resulting entries would be shown. This concept was provided to me in collaboration with Christine Girtain and Sarah Bordenstein, and all areas of the database will be consulted with Mrs. Bordenstein to ensure it meets the requirements and expectations she may have, as well as her collaborators in the field.

Along with general entries for each unique gene ID, I also plan to incorporate a BLAST tool for users to use if needed. This tool would allow users to input a sequence of interest to them that could then be screened against the database entries in order to find similarities, or even a matching, existing entry.

Two other concepts I hope to implement with the database include an upload tool, in which users could upload sequences and information that could be screened and uploaded to the database if approved, and a user profile creation and log-in tool. The mechanism for how this will work is still undecided, but I hope to have it done autonomously or with minimal routine checking so that the database won't require continuous manual maintenance. The profile creation and log-in would work with this upload tool so that entries could be traced back to individual users if needed, and to let them know if there was an issue with the sequence and information they tried to upload. If time permits, I would also like to set this up so it could be collaborative for users to find others also working on *Wolbachia* research, but as of right now I am unsure with how to implement this effectively.

Resources & Skills Needed to Implement

The relevant resources and skills needed to implement this project include a multitude of different techniques and sources. These will be subsequent to change as more parts of the project are completed, as I do not know the entirety of where this project and its tools/implementations will be finalized.

In regards to the programming aspect of this database, I plan to work primarily with Python, PSQL, HTML, and CSS. Python and PSQL will help me generate, maintain, and alter the database as needed so that it contains the most relevant and new information, and the HTML and CSS languages will help with front-end development of the database, such as the general look and supporting tabs. Under Python and PSQL I will primarily use the Flask web framework, which will help maintain the overall look of the database, and WTForms, which is a rendering library to also help maintain the databases look and interactions.

The data sources for this project will be provided by Sarah Bordenstein and the Bordenstein lab as an initial beginning to the project, and with the help of the upload tool, could eventually be provided by anyone from virtually anywhere. The data management aspect, as mentioned earlier, is still unclear, but it is hoped that there could be some recognizable similarities between all entries that could then be screened for in future entry uploads to ensure accuracy.

Finally, most of the supporting papers and research articles provided will be discovered as the project progresses, but I primarily plan to work with the papers described in the references section and their supporting references, as well as any more I may find in the near future. I also plan to continuously check and work with the existing *Wolbachia* database framework, found at https://wolbachiaprojectdb.org/, so that it can properly mirror their existing tool.

Expected Timeline

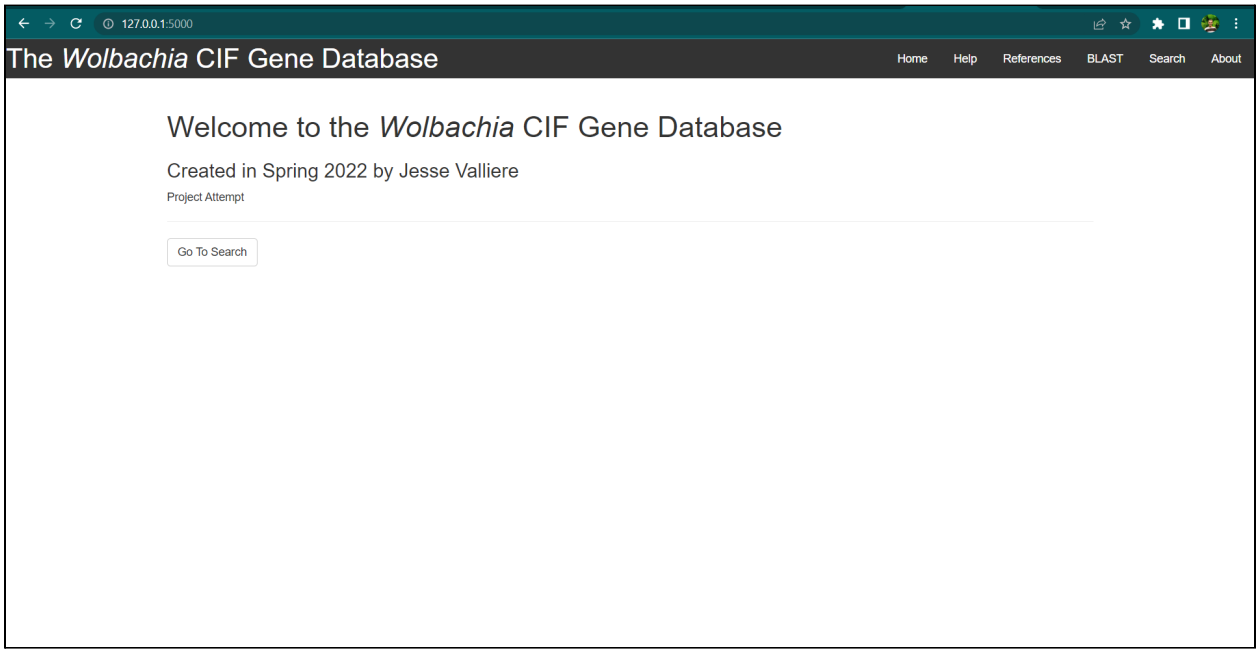
The expected timeline for this project has been described below to the best of my ability. Following Dr. Stuart's request, I plan to do routine biweekly check-ins to ensure the scope of the project is being met in a timely manner, and so that it will be completed on time before the end of the semester. Any changes made to this timeline would first be consulted with Dr. Stuart and Sarah Boredenstein, and only applied if both agree.

Dates	Objectives To Be Completed
2/15/22 - 3/1/22	● Continue setting up framework of website and mirroring The Wolbachia Project Database. (pictures, labels, etc.) ● Have most supporting tabs (Home, Help, References, About) set up and accurate.
3/1/22 - 3/15/22	● Have all data loaded in. ● Clean up Search tab and supporting results. ● Accurately be able to utilize Search function to have data appear correctly. ● Begin trying to incorporate BLAST tool
3/15/22 - 3/29/22	● Learn how to incorporate BLAST tool into the site. ● Have BLAST tab cleaned up and accurately reflecting what Bordenstein Lab is hoping for. ● Consult Bordenstein Lab, Dr. Hague for advice on areas to improve for users.
3/29/22 - 4/12/22	● Continue setting up any remaining framework of website and mirroring The Wolbachia Project Database. (pictures, labels, etc.) ● Include an upload sequence tab and begin researching ways this could be checked autonomously. ● Incorporate log-in / register action for users
4/12/22 - 4/26/22	● Incorporate database onto server so it can be accessed on any desktop without downloading software. ● Continue working on upload sequence tab and finding a way to have it checked autonomously.
4/26/22 - 5/6/22	● Consult Bordenstein Lab, Dr. Hague for advice on areas to improve for users. ● Clean up loose ends and prepare to display.

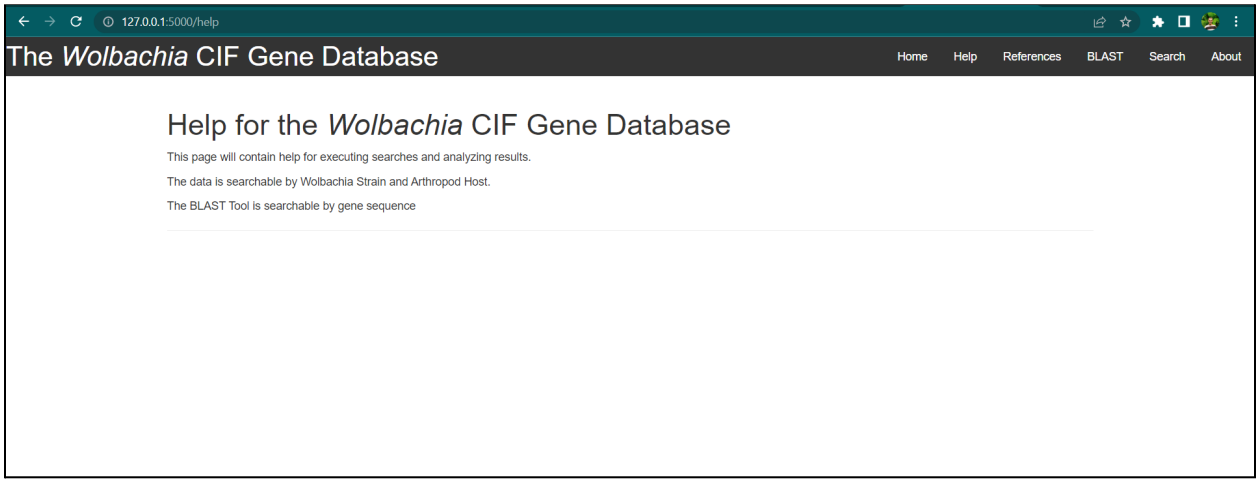
Example Interface

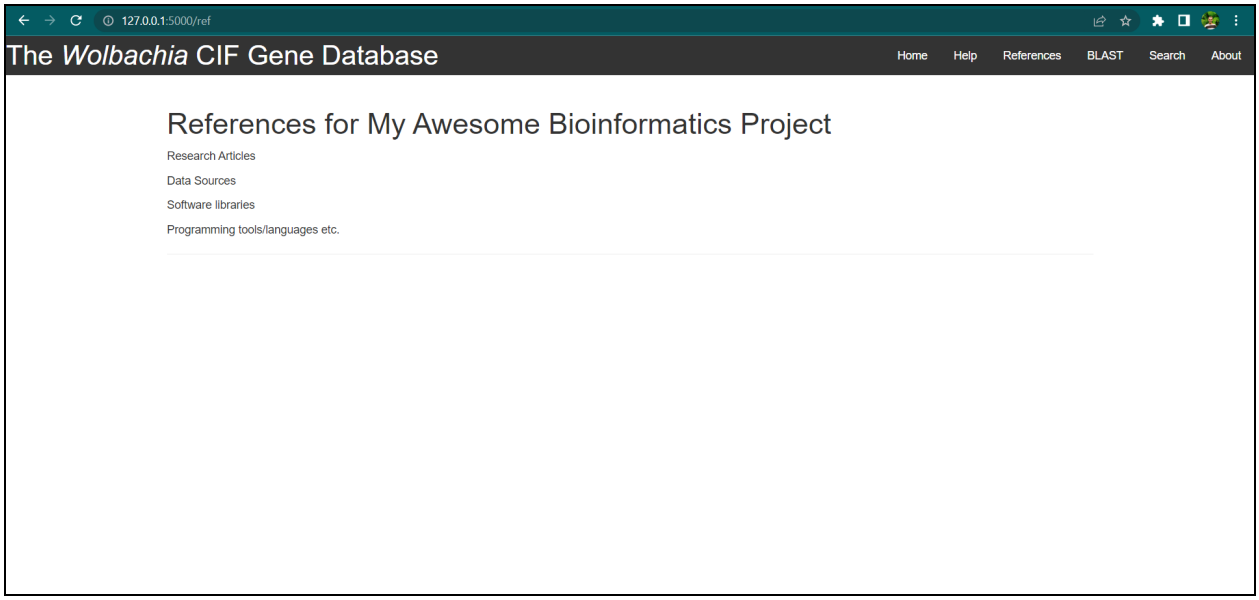
It is important to mention before showing examples of the interface that these are preliminary images of the final product, and will include much more updating and modernization before they reach their final appearances. I plan to cooperate with the Bordenstein Lab and their colleagues to make sure it accurately reflects what they hope for, and that it matches their proposed standards.

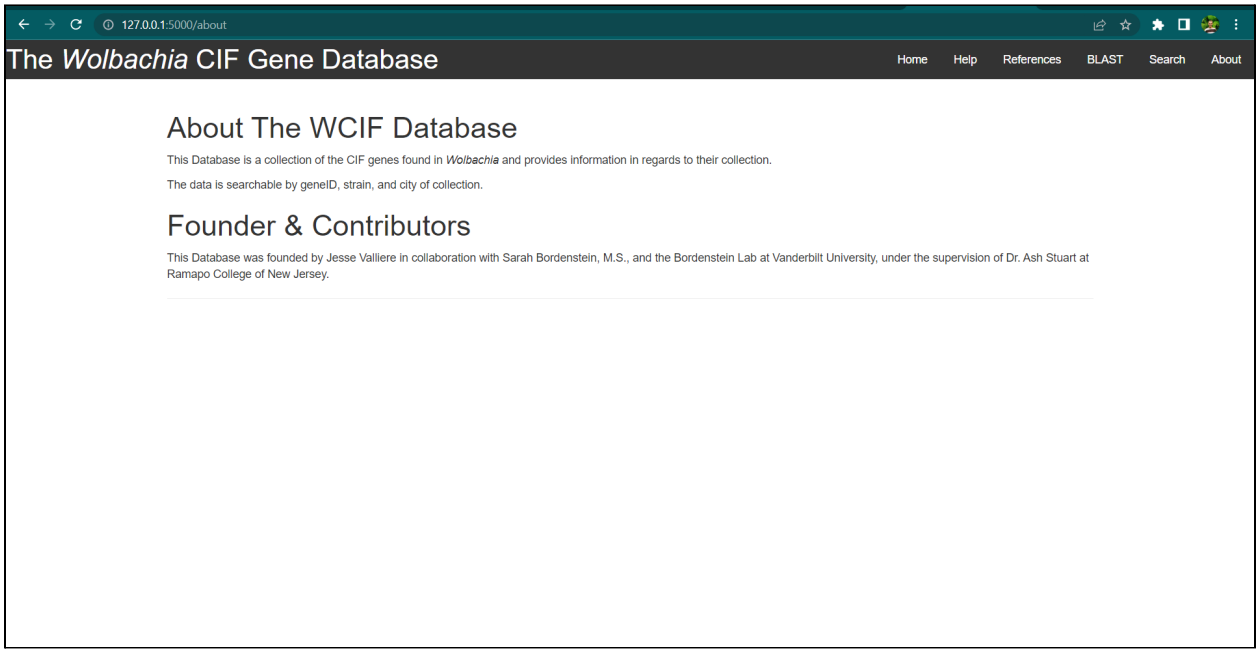
Example Image of the Home page users will see when navigating to the database.



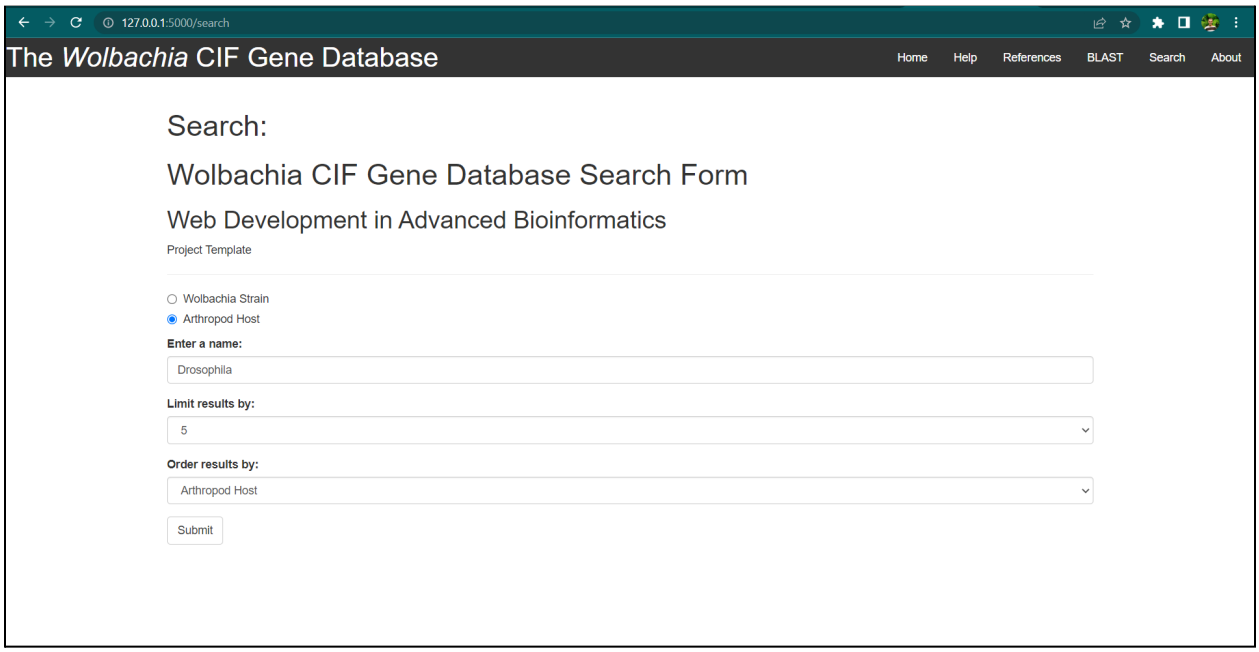
Example Image of the Help, References, and About tabs for where users can find more information. These will continuously be updated until the database is finished so that it reflects everything I used to create it.

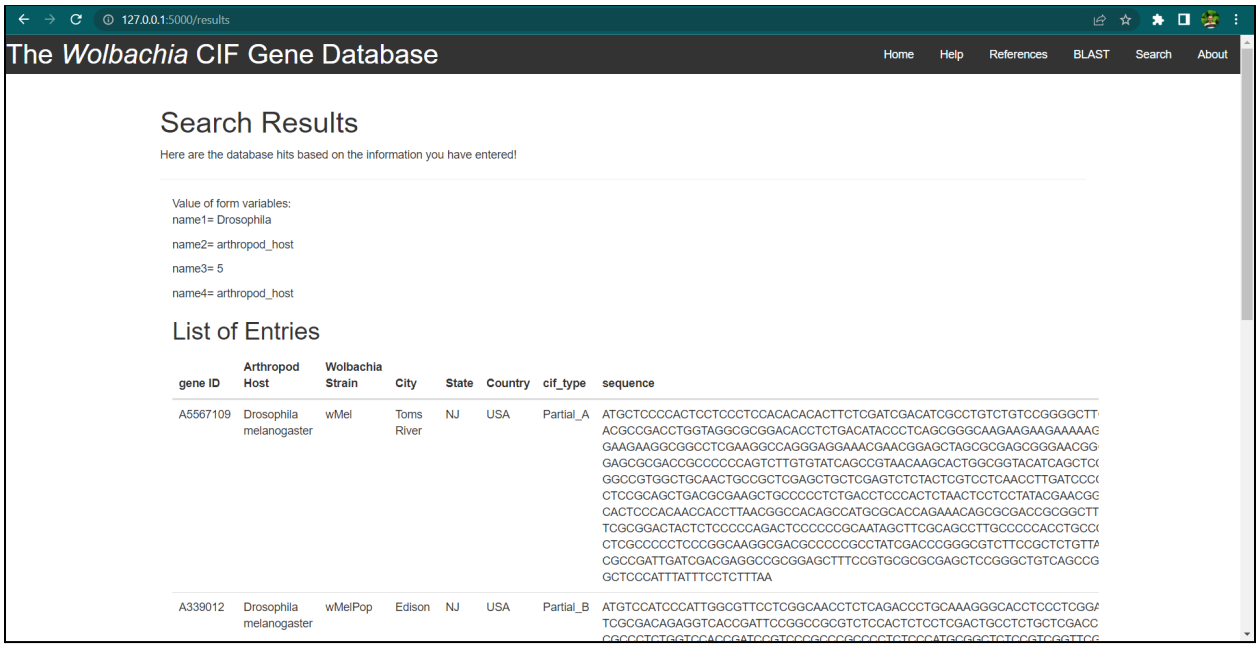



127.0.0.1:5000/ref
The Wolbachia CIF Gene Database
Home
Help
References
BLAST
Search
About
References for My Awesome Bioinformatics Project
Research Articles
Data Sources
Software libraries
Programming tools/languages etc.

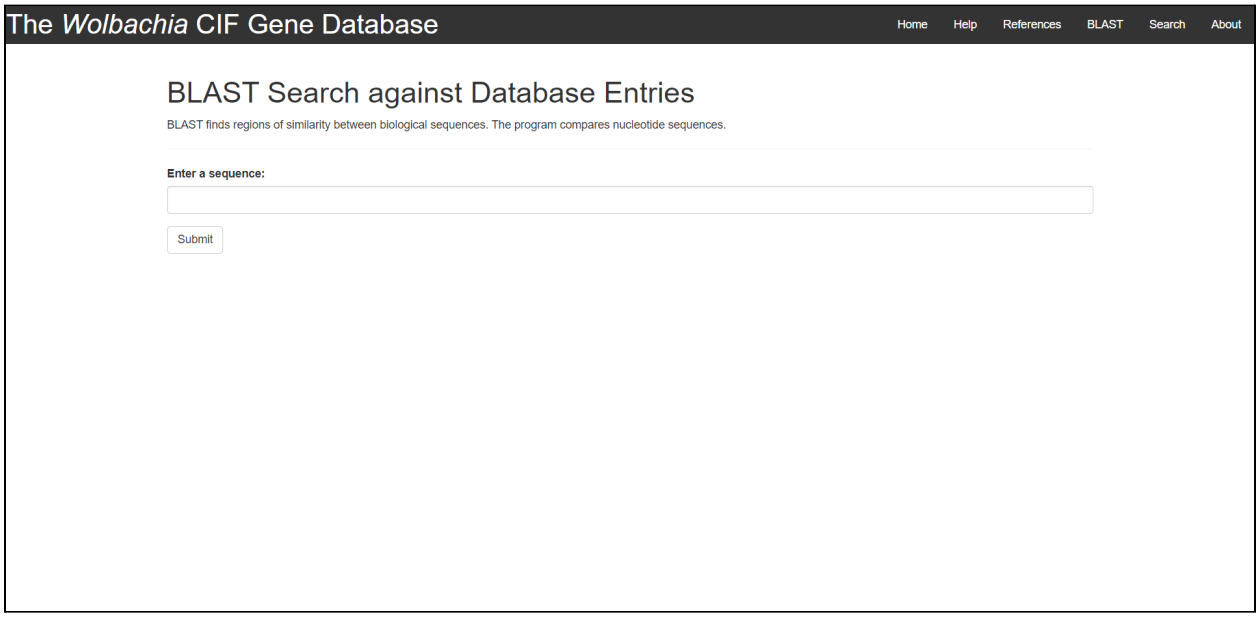

127.0.0.1:5000/about
The Wolbachia CIF Gene Database
Home
Help
References
BLAST
Search
About
About The WCIF Database
This Database is a collection of the CIF genes found in Wolbachia and provides information in regards to their collection.
The data is searchable by geneID, strain, and city of collection.
Founder & Contributors
This Database was founded by Jesse Valliere in collaboration with Sarah Bordenstein, M.S., and the Bordenstein Lab at Vanderbilt University, under the supervision of Dr. Ash Stuart at Ramapo College of New Jersey.

Example Image of a potential Search and how its results would be displayed. The database does not have the accurate data as of yet, so these are example hits I created myself.

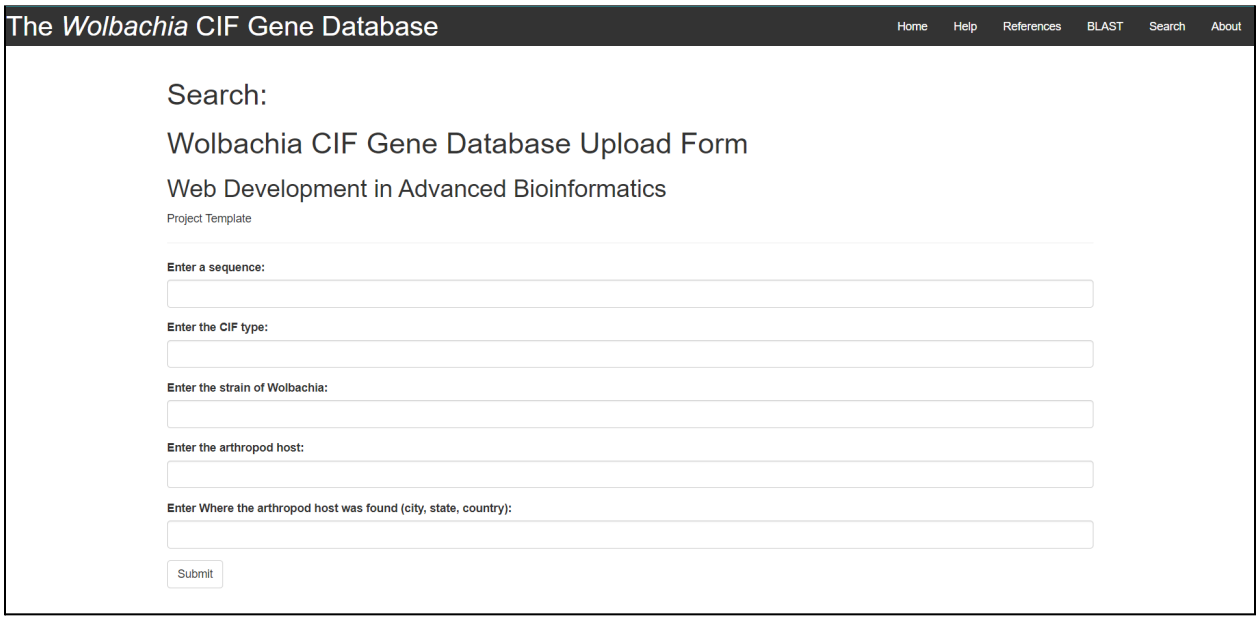




Example Image of BLAST tool input screen for users. Note that the results would be displayed in a tabular format, with entries being shown based on percent similarity (in decreasing order).



Example Image of the upload sequence tool. After uploading the information will be checked before being incorporated into the existing data to ensure it is accurate and realistic. This is most likely going to be found on the Home Page of the database.



Conclusion

Overall, the need to have a supporting database like this for *Wolbachia* research is clear, and I believe that this database will be a useful tool for the research community to consult whenever they may need. With the collaboration with Sarah Bordenstein and the Bordenstein Lab at Vanderbilt University, I believe that the database will be able to match their standards and their existing database. Cytoplasmic incompatibility in *Wolbachia* is a very prominent area of research as the need to control the spread of diseases from arthropods continues, and I believe this database will be a useful contribution for researchers to consult.

References

1. LePage, Daniel P., et al. "Prophage WO genes recapitulate and enhance Wolbachia-induced cytoplasmic incompatibility." *Nature* 543.7644 (2017): 243-247.
2. Zhiyong Xi, Deepak Joshi, Chapter 14 - Genetic Control of Malaria and Dengue Using Wolbachia, Editor(s): Zach N. Adelman, Genetic Control of Malaria and Dengue, Academic Press, 2016, Pages 305-333, ISBN 9780128002469, https://doi.org/10.1016/B978-0-12-800246-9.00014-4.