Requirements Specification

for

<Gene to Protein: Human Mitochondrial Genes and Their Proteins>

Version 1.0

Prepared by <Erin Ledford>

<University of Nebraska at Omaha>

<February 12th, 2021>

Table of Contents

Table of Contents ii

Revision History ii

1. Introduction 1

1.1 Purpose 1

1.2 Document Conventions 1

1.3 Intended Audience 1

1.4 Project Scope 1

1.5 References 1

2. Overall Description 1

2.1 Project Perspective: Background 1

2.2 Project Functions 2

2.3 User Classes and Characteristics 2

2.4 Operating Environment 2

2.5 Design and Implementation Constraints 2

2.6 User Documentation 2

2.7 Assumptions and Dependencies 2

3. Interface Requirements 3

3.1 Web User Interfaces 3

3.2 Database Interfaces 3

4. Other Nonfunctional Requirements 3

4.1 Privacy Requirements 3

4.2 License Requirements 3

4.3 Accessibility of the Web Interface 3

5. Other Requirements 4

Appendix A: Glossary 4

Appendix B: To Be Determined List 5

Revision History

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Date** | **Reason For Changes** | **Version** |
|  |  |  |  |
|  |  |  |  |

# Introduction

## Purpose

<Describe the purpose of your project.>

The purpose of this project is to create a small database of human mitochondrial genes and the proteins that they are related to.

## Document Conventions

<Describe any standards or typographical conventions that were followed when writing this SRS, such as fonts or highlighting that have special significance, acronymcs, etc.>

There are no standards or typographical conventions that were followed when writing this SRS.

## Intended Audience

The intended audience of this project is student(s) and instructor(s) in BIOI 4870/CSCI 8876.

## Project Scope

<Provide a short description of the project being specified and its purpose, including:

* relevant benefits,
* objectives, and
* goals

>

The purpose of the project is to make a database that shows human mitochondrial genes and the proteins that are involved in human mitochondria. The objective of this project is to learn how NCBI and other databases create user friendly databases that show genes and proteins [1]. The benefit is a smaller database that focuses specifically on human mitochondria genes and proteins. The goal of this project is to make a small, user friendly database with only human mitochondria genes and proteins.

## References

< *At least 5 peer-reviewed resources should be used with in-text citations. Whatever citation style used should be identifiable and should be followed consistently. If you need help, please contact our STEM librarian, Heidi Blackburn by clicking "Library Resources" on the left Canvas panel under Grades.*

Also, list any other documents or Web addresses to which this SRS refers, including relevant biomedical literature. These may include user interface style guides, contracts, standards, system requirements specifications, use case documents, or a vision and scope document. Provide enough information so that the reader could access a copy of each reference, including title, author, version number, date, and source or location.>

[1] E. A. Schon, S. DiMauro, and M. Hirano, “Human mitochondrial DNA: roles of inherited and somatic mutations,” Nature Reviews Genetics, vol. 13, no. 12, pp. 878–890, 2012.

[2] National Center for Biotechnology Information, 1988. [Online]. Available: https://www.ncbi.nlm.nih.gov/. [Accessed: 14-Feb-2021].

[3] O. G. Berg and C. G. Kurland, “Why Mitochondrial Genes are Most Often Found in Nuclei,”

Molecular Biology and Evolution, vol. 17, no. 6, pp. 951–961, 2000.

[4] R. W. Taylor and D. M. Turnball, “Mitochondrial DNA mutations in human disease,” Nature reviews. Genetics, 06-May-2005. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/15861210/. [Accessed: 14-Feb-2021].

[5] T. UniProt&nbsp;Consortium, “UniProt: the universal protein knowledgebase,” Nucleic Acids Research, vol. 46, no. 5, pp. 2699–2699, 2018.

# Overall Description

## Project Perspective: Background

<Describe the context and origin of the project being specified in this SRS, including:

* What background information is needed for a fellow BIOI 4870/CSCI 8876 student to understand the project?
* Why is this project significant or needed in this field?
* Novelty - Does a version of this project exist elsewhere? How is your proposed project different, if at all?
  + I.e. If you are proposing a nucleotide database you could note the existence of NCBI Nucleotide and why yours is different (obvisouly yours would be smaller but may be for a specific set of bacterial isolates, for example)
  + OPTIONAL: If desired, include a simple diagram that shows the major components of the overall system, subsystem interconnections, and external interfaces.>

The human mitochondrial genome is important to understand for those in the health field. There are many diseases associated with mitochondrial genes and proteins [1,3,4]. Mitochondrial DNA is inherited differently from compared to nuclear DNA [1,3,4]. This makes it harder for those that need to find information about the diseases caused by those genes as they are less researched compared to nuclear DNA [1,3,4]. This project is needed in the field to provide a smaller, more specific database to search for the genes or proteins that may cause these diseases or to allow users to study specific genes and proteins. A version of this project exists elsewhere, in NCBI, as a giant database that can be hard for users to search [2]. The difference between NCBI and my project is the both the size (mine is far smaller) and that my project will be an aggregation of all reputable sources of human mitochondrial genes and proteins sequences that I can find.

## Project Functions

<Summarize the major functions the project must perform or must let the user perform. Details will be provided in Section 3, so only a high level summary (such as a bullet list) is needed here.>

* Search the database by accession number.
* A user FAQ.
* Search specifically for genes.
* Genes will have the proteins they may make in their description.
* Proteins will have what genes they originate from in their description.
* Genes will have how many variants lead to disease and the disease associated to the highest amount of variants.

## User Classes and Characteristics

<Identify the various users that you anticipate will use this project. Users may be differentiated based on frequency of use, subset of project functions used, technical expertise, security or privilege levels, educational level, or experience.>

I expect naïve and sophisticated users will use this project. The license I have chosen will allow others to use my work as they see fit. Naïve users will be those that do not have any bioinformatics experience or education, they may have to use the FAQ in the web interface to understand basic concepts. Sophisticated users will be those in this class, or those that have any bioinformatics experience.

## Operating Environment

<Describe the environment in which the software will operate, including the hardware platform, operating system and versions, and any other software components or applications with which it must peacefully coexist

If you will run the project on Odin (highly recommended), you can find OS information here: <https://www.unomaha.edu/college-of-information-science-and-technology/about/odin.php>>

The software will operate on Odin, which is a Linux server. Odin users are limited to 1 GB of storage, 80 processes, and 100 minutes per job. The project will use SQL version 15.1 for Debian Linux. There are no other software components or applications with which it must peacefully coexist.

## Design and Implementation Constraints

<Describe any items or issues that will limit the options available to the developers. This might include – limited ability to download and install new tools, 1GB of data storage (on Odin), etc .>

There will be a limited ability to download and install tools, and there is a limit on data storage of 1 GB as I will be using Odin.

## User Documentation

<List the user documentation components (such as user manuals, on-line help, and tutorials) that will be delivered along with the software. Identify any known user documentation delivery formats or standards. Will it be stored on Github? Bitbucket? Shared as a .zip?>

User manuals and tutorials will be stored on Github or shared as a .zip. A small FAQ will be added to the web interface to answer basic questions.

## Assumptions and Dependencies

<Identify any dependencies the project has on external factors, such as software components that you intend to reuse from another project, unless they are already documented elsewhere (for example, in the vision and scope document or the project plan).>

Any dependencies of the project are described elsewhere in the vision and scope and the project plan.

# Interface Requirements

## Web User Interfaces

A simple PHP interface will be prepared for database users according to tutorials and code demonstrations presented in class.

## Database Interfaces

*<Describe Where you will obtain the data and How you plan to get the data to your database (data dump, API, etc). (This can always change – just show me that you have investigated how the databases you will use make their datasets available. Also describe:*

* *Approximate specs of the data (size, format, etc)*
* *Language and server you will use to make your database*
* *Frequency of data upload – one time, daily, weekly, monthly, etc?*

*>*

I will obtain the data for this project from NCBI and SWISS-PROT [2, 5]. I plan to data dump my data into the database by grabbing data from NCBI and making a secondary database [2, 5]. I will use either FASTA or Genbank format for genes and proteins [2]. I will use Odin for my server and SQL as the language for my database. I will only upload the data one time and most likely not update it unless necessary.

# Other Nonfunctional Requirements

## Privacy Requirements

The project does not involve the use of any personal, private, or secure information.

## License Requirements

<Select a license to use for your project depending on how you would permit others to use it outside the scope of the class and include that license below. You can use this website to help you choose a license: <https://choosealicense.com/>>

This project will fall under the MIT License that only requires preservation of copyright and license notices.

## Accessibility of the Web Interface

*<Universal design is defined as “the design and composition of an environment so that it may be accessed, understood and used to the greatest possible extent, in the most independent and natural manner possible, in the widest possible range of situations, without the need for adaptation, modification, assistive devices or specialised solutions, by any persons of any age or size or having any particular physical, sensory, mental health or intellectual ability or disability, and means, in relation to electronic systems, any electronics-based process of creating products, services or systems so that they may be used by any person.” Universal design encompasses accessibility and provides a framework for evaluating the usability of our project from a broad user scope. There are a number of ways to begin to make your project accessible – select at least 1-2 of the following examples that you would like to try to make your project more accessible and include that in this section:*

Improving accessibility of the project will be achieved by:

* The web interface for users will be checked to see that it works on mobile phone or tablet browser
* Any videos provided on the web interface will be captioned
* Images on the web interface will be accompanied by alternative text
* The web interface for users will be checked to see that it works with a screen reader
* The web interface for users will be checked to see that it works with keyboard accessibility only (i.e. no use of a mouse)
* Time requirements will not be included in the project unless absolutely necessary
* All hyperlinks on the website will be provided in full text
* The website will not use audio-only files
* The web interface will include a link at the bottom to report accessibility problems (i.e. this can go to a Google form)
* The web interface will be tested to ensure it works for both right- and left-handed users
* Complex graphics (graphs, charts, etc.) are accompanied by equivalent text, either through a description in the body of the page
* Color will not be used solely to convey important information
* Data tables have column and/or row headers appropriately identified
* Others/propose your own – there are many more examples of accessibility checkpoints online: <https://www.w3.org/TR/WAI-WEBCONTENT/full-checklist>

>

Accessibility of the project will be improved by including a link at the bottom of the web interface to report accessibility problems via a Google form. It will also be improved by including alternative text for any web images.

# Other Requirements

<Define any other requirements not covered elsewhere in the SRS. This might include database requirements, internationalization requirements, legal requirements, reuse objectives for the project, and so on. Add any new sections that are pertinent to the project.>

Appendix A: Glossary

<Define all the terms necessary to properly interpret the SRS, including acronyms and abbreviations. You may wish to build a separate glossary that spans multiple projects or the entire organization, and just include terms specific to a single project in each SRS.>

Appendix B: To Be Determined List

<Collect a numbered list of the TBD (to be determined) references that remain in the SRS so they can be tracked to closure.>

<https://www.uniprot.org/uniprot/?query=human+mitochondria&sort=score>

<https://uswest.ensembl.org/Human/Search/Results?q=mitochondria;site=ensembl;facet_species=Human>

<https://www.ncbi.nlm.nih.gov/gene/?term=human%20mitochondria>