Bioinformatics, computer lab I
University of Potsdam WS 2019/2020, S. Hartmann

1 Getting a Linux account for the Computer pools

For the exercises in this course you will need a LINUX Account for the Computer Pools. Information about the computer pools, including PDF versions of the application, can be reached via `http://www.uni-potsdam.de/ibb` $\longrightarrow$ Studium und Lehre $\longrightarrow$ Computerpool
Please fill out and submit the application as soon as possible.

Please log in under the Linux operating system for all exercises of this course: Some of the software we'll use is installed only under Linux. But even if you're just using the browser (like today), it allows for consistency and help you to get used to the interface more quickly.

2 Globins

Several of the computer exercises in this course will focus on the analysis of a family of globular proteins: the *globins*. Among the members of this family are *hemoglobin* and *myoglobin*.

Use today's exercise to get somewhat familiar with globins – and with different online resources! Use the search terms in *italics* (above) to find out more about these proteins. Use additional and more specific search terms based on the information you find.

Take notes as you browse and collect the information. It may be a good idea to create a new folder for this course (e.g., "bioinfWS19") in your home directory. Within this directory, you can create separate directories for the different exercises (e.g., "lab01" for today's exercise). This allows to better organize data files and notes for each exercise you'll complete during the semester.

3 Search engines

3.1 General search engine (e.g., Google)

Anyone can publish a web page, and it sometimes is not immediately clear whether to trust the information presented in a given online article: The leading scientist on the subject matter may have published it – or a 5th-grade student. For this reason, it is important to pay attention to the following when browsing through Google results:

- who is the author or the organization that prepared the web page?
- does the article contain references for the information given?
- when was the article created or last modified?
- does the site contain the author's contact information?

3.2 Scientific publications

Google offers a specialized search engine for scientific literature, Google Scholar, at `http://scholar.google.com/`.

✎ Again, search for one or more of the three search terms given above.

✎ How is the result different from your previous Google search? What kind of information is available for each publication?

4 Databases

The results of Google searches can probably be used effectively as a source for background information by anyone. Accessing and using specific biological databases, however, is a different matter.

Entries within a databases have a consistent structure and format, and curators have pre-selected what gets included in the database. Although most scientific databases can be accessed free of charge, navigating them and accessing their content is often not very intuitive. We'll cover a few biological databases in more detail starting next week, here are a few to get you started:

- annotated proteins: `http://www.uniprot.org`
- scientific literature: `http://www.ncbi.nlm.nih.gov/pubmed` or `http://webofknowledge.com/`
- metabolic pathways and associated genes: `http://www.genome.jp/kegg/`

✎ Use today's keywords (globin, myoglobin, hemoglobin) to explore these resources.

✎ What types of information do these databases contain? Which sections make sense to you, which don't?

5 Summary

✎ Briefly summarize what you have learned about globins today.

✎ Briefly summarize what you have learned about different (types of) online resources today.

✎ What is the difference between a search engine and a database?