

Hinweise zum Modul Grundlagen der Sequenzanalyse Wintersemester 2022/2023

Stefan Kurtz

16. Oktober 2022

1 Vorlesung

1. 14 Vorlesungstermine: ab 18.10.2019 jeweils dienstags von 14:15–15:45 Uhr
2. Die Vorlesung vermittelt grundlegende Methoden zur Analyse von biologischen Sequenzen, insbesondere zu den folgenden Themen:
 - Edit-Distanz Modell
 - Berechnung optimaler Alignments mit Verfahren der dynamischen Programmierung
 - Heuristische Verfahren zum schnellen Sequenzvergleich
 - Multiple Sequenzalignments
 - Phylogenetische Analyse
3. Die Vorlesungsinhalte werden in Form von Folien und Skripten präsentiert.
4. Das Material steht jeweils in `gitlab@rrz` zur Verfügung (siehe unten). Sie werden jeweils per E-mail darüber informiert, wenn neues Material zur Verfügung steht.
5. Fragen und Anmerkungen zur Vorlesung bitte per E-mail an `stefan.kurtz@uni-hamburg.de`

2 Übungen

1. 13 Übungstermine: jeweils dienstags um 16:15–17:45 Uhr (Beginn 25.10.2022, also eine Woche später als in Stine angekündigt)

2. Die Übungen werden von Herrn Kurtz geleitet.
3. Beginnend mit der zweiten Woche des Semesters wird jeweils dienstags ein Übungszettel mit Aufgaben zusammen mit Materialien in `gitlab@rrz` bereitgestellt.
4. Bei den Aufgaben handelt es sich ab Blatt 4 teilweise um Programmieraufgaben. Die Lösungen der Programmieraufgaben müssen in Python3 angefertigt werden. Da die Tests Ihrer Lösungen auf Basis von Standard-Unix Werkzeugen wie z.B. `make`, `diff`, `head` durchgeführt werden, benötigen Sie einen Rechner mit Linux oder macOS Betriebssystem oder WSL unter MS-Windows. Hier sind die meisten Entwicklerwerkzeuge bereits vorinstalliert, bzw. lassen sich leicht installieren.
5. In den Aufgabenstellung werde ich darauf achten, dass die entsprechenden Programmierkenntnisse bereits in PfN1 vermittelt wurden, wenn die GSA-Übungen gelöst werden müssen. In Einzelfällen gebe ich auf den Übungsblättern Hinweise, welche Python3-Sprachelemente für die Lösung verwendet werden können. Bei allen Teilnehmern, die nicht an PfN1 teilnehmen, gehe ich davon aus, dass die Programmierkenntnisse für die Lösung der GSA-Aufgaben ausreichen. Bitte weisen Sie mich darauf hin, wenn das aus Ihrer Sicht nicht der Fall sein sollte.
6. Sie müssen Lösungen in schriftlicher Form abgeben. Diese werden mit Punkten bewertet.
7. Die genaue Form der Abgaben Ihrer Lösungen wird in der ersten Übung erläutert.
8. Die Lösungen sollen in 2-er oder 3-er Gruppen (Kleingruppen) angefertigt und gruppenweise abgegeben werden.
9. Sie haben bis spätestens 22:00 Uhr des nachfolgenden Sonntags Zeit, Ihre Lösungen per E-mail an `gsa@zbh.uni-hamburg.de` zu schicken. Die jeweiligen Abgabetermine werden auf den entsprechenden Übungsblättern genannt.
10. Im Betreff/Subject der E-mail muss

`GSA-Uebung BlattXY.Name1.Name2 ...`

stehen, wobei XY die zweistellige Nummer des Übungsblattes mit führenden Nullen ist und Name1 bzw. Name2 die Nachnamen der Mitglieder der Kleingruppe.

11. Besondere Zeichen in den Namen: Sie dürfen keine Umlaute verwenden, d.h. ü wird durch ue geschrieben, ß durch ss etc. Falls ein Nachname aus mehreren Worten besteht, werden diese Worte durch einen Unterstrich (`_`) verbunden. Die Betreffzeile enthält also genau ein Leerzeichen.
12. Sie müssen dieses Format unbedingt einhalten, denn Ihre Lösungen werden von einem Mail-Server empfangen und Ihre Lösungen automatisch extrahiert.

13. Der Mail-Server ist nicht fehlertolerant, d.h. wenn Sie sich nicht an die Vorgaben halten, könnte es sein, dass Ihre Lösung bei der Extraktion vom Mail-Server unbeachtet bleibt und dann auch nicht korrigiert wird.
14. Beachten Sie bei der Abgabe entsprechende Hinweise für die Benennung der Dateien Ihrer Lösung.
15. Die Übungstermine dienen hauptsächlich dazu, erste Lösungsansätze für das aktuelle Aufgabenblatt zu erarbeiten. Das Ziel ist es, dass in jeder Übung der Übungsleiter mindestens einmal mit jeder Kleingruppe über die Lösungsansätze spricht. Das Gespräch wird kein Prüfungsgespräch sein. Es soll lediglich sichergestellt werden, dass alle Studierenden gleichermassen von der Übung und der damit verbundenen fachlichen Hilfestellung profitieren. Bitte informieren Sie den Übungsleiter vor einer Übung per E-mail, wenn absehbar ist, dass Sie nicht teilnehmen können.
16. Die Übungen sind so konzipiert, dass die meisten Studierenden die Lösungen nicht während der Übungszeiten vollständig erstellen können. Nach unserer Erfahrung bewegt sich der Median der Arbeitszeit für die Bearbeitung der Übungsblätter zwischen 4 und 7.5 Stunden pro Woche.
17. Als Teil der abgegebenen Lösungen erfragen wir jeweils von allen Kleingruppen den Zeitaufwand und die Einschätzung des Schwierigkeitsgrads der Übungen. Hierzu erhalten Sie eine Rückmeldung, in der diese Information zusammengefasst für alle Studierenden statistisch aufbereitet wird.
18. Geforderte Studienleistung für den Übungsabschluss ist die aktive und regelmäßige Teilnahme an den Übungen. Dazu müssen insgesamt im Semester 50% der möglichen Punkte der Übungsaufgaben erreicht werden.

3 Modulabschlussprüfungen (Klausur)

- 15. Februar 2023, 10:30–12:00 Uhr, ESA M
- 15. März 2023, 09:30–11:00 Uhr, ESA H

4 Material

1. Das Material zur Vorlesung und zur Übung wird via `git` verteilt. Dazu verwenden wir die Infrastruktur von `gitLab` am RRZ der Uni Hamburg.
2. Nach der Freischaltung führen Sie in einem Verzeichnis auf Ihrem Rechner einmalig den folgenden Befehl aus:

```
git clone https://gitlab.rrz.uni-hamburg.de/Bae4410/gsa_2022.git
```

Sie müssen dann Ihre B-Kennung angeben und anschließend das Passwort. Dann wird ein Verzeichnis `gsa_2022` mit dem Material angelegt.

3. Vor jeder Übung und Vorlesung wird das Material aktualisiert. Um das aktualisierte Material zu erhalten, wechseln Sie jeweils in das genannte Verzeichnis und führen den Befehl

```
git pull
```

aus.

4. Da der gesamte Verzeichnisbaum für Sie nur zum Lesen gedacht ist, dürfen Sie dort selbst keine Dateien verändern oder neue Dateien anlegen. Ihre Lösungen zu den Übungen müssen Sie daher in einem anderen Verzeichnis, z.B. `my_gsa_2022` ablegen.
5. Nachdem Sie sich in Linux eingearbeitet haben, wechseln Sie in das Verzeichnis

```
gsa_2022/bin
```

Dort finden Sie eine Datei mit einigen Shell-Befehlen. Um diese Befehle benutzen zu können, müssen Sie folgende Schritte durchführen, wenn Sie als Kommandozeileninterpreter das Programm `bash` verwenden.¹

- a) Fügen Sie in der Datei `.bashrc` in Ihrem HOME-Verzeichnis die Zeile (den Punkt am Anfang nicht vergessen)

```
. $HOME/gsa_2022/bin/add_to_bash.sh
```

ein.

- b) Falls die Datei `.bashrc` in Ihrem HOME-Verzeichnis nicht existiert, dann legen Sie sie neu an.
- c) Falls Sie `gsa_2022` in einem anderen Verzeichnis als `$HOME` klonen, müssen Sie die obige Zeile und die Definition von `ROOT2GSA` in `add_to_bash.sh` entsprechend anpassen.

Wenn Sie einen anderen Kommandozeileninterpreter als `bash` verwenden, müssen Sie ggf. andere Dateien verändern bzw. eine andere Syntax für die Definition der Umgebungsvariablen verwenden. Bei der Verwendung der `zsh` heißt die Datei `.zshrc` anstatt `.bashrc`.

¹Um das herauszufinden, führen Sie im Terminal den Befehl `echo $SHELL` aus. Das Ergebnis ist der Pfad der Shell, z.B. `/bin/bash`. Zum Ändern der Shell können Sie das Programm `chsh` aufrufen.