

Übungen zur Vorlesung Algorithmische Bioinformatik

Freie Universität Berlin, WS 2010/11

Roland Krause · Martin Vingron · Matthias Winkelmann

Blatt 9 · Ausgabe am 13.12.2010

Abgabe am 3.1.2011 vor Beginn der Vorlesung

Aufgabe 31 (Euler-Zyklus/Hamilton). Beweisen Sie für Graph 1 die Existenz oder Nichtexistenz eines Euler- und eines Hamiltonkreises.

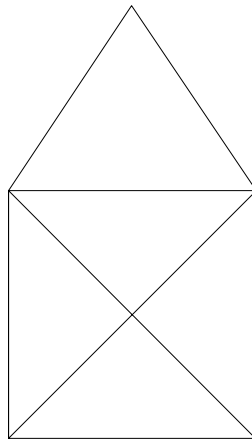


Abbildung 1: Graph 1

Aufgabe 32 (Sequencing by Hybridization (korrigiert)). Gegeben ist das Spektrum $S = \{GCA, CGC, CGT, CCG, CAC, ACG, ACC, TAC\}$. Rekonstruieren Sie die Ursprungssequenz mittels der beiden in der Vorlesung vor gestellten Verfahren.

Aufgabe 33 (Praktisches Assembly). Assemblieren Sie (simulierte) Reads, wie sie aus einem BAC erhalten¹.

1. Laden Sie die Reads für Ihre Arbeitsgruppe. Zu welchem Organismus gehören Ihre Reads?
2. Nutzen Sie den CAP3-Assembler² mit Standardparametern
3. Erstellen Sie einen Plot der Coverage entlang des Contigs.
4. Ermitteln Sie mit RepeatMasker³ repetitive Bereiche in Ihrem Fragment.
5. Bestimmen Sie mittels GenScan⁴ die Bereiche, für die Ihr Fragment für Proteine kodiert.
6. Vergleichen Sie die Vorhersage mit der Annotation, die Ihnen die Datenbanken liefern⁵. Dokumentieren Sie Ihr Vorgehen.

¹http://lectures.molgen.mpg.de/Algorithmische_Bioinformatik_WS1011/dokumente/u9data/

²<http://seq.cs.iastate.edu/cap3.html>, die Windows-Version läuft nur unter Cygwin

³<http://www.repeatmasker.org/>

⁴<http://genes.mit.edu/GENSCAN.html>

⁵Mögliche Quellen: <http://flybase.org>, <http://ensembl.org>

Aufgabe 34 (Implementierung eines Assemblers). Alternativ zum praktischen Assembly.

1. Überlegen Sie sich einen Assembler nach Ihren Vorstellungen und stellen Sie ihn dar. Implementieren Sie ihn.
2. Erstellen Sie ein Genom nach einer Hintergrundverteilung an $A, C, G, T = \frac{1}{4}$ und der Länge 20.000nt.
3. Simulieren Sie die Sequenzierung mit Readlängen von 500 nt und einer effektiven Coverage von 5 und einem θ von 0.15.
4. Assemblieren Sie Fragmente und erstellen Sie das Genom.
5. Wiederholen Sie das Experiment und ermitteln Sie, wie vollständig Sie das Genom jeweils erhalten. Die Anzahl der Wiederholungen sollte sich nach der Effizienz Ihres Assembly richten.
6. Wenden Sie Ihre Implementierung auf realistischeren Testdaten an⁶
7. Bewerten Sie die Leistungsfähigkeit Ihrer Implementierung. Wie lange luft ein Assembly?

⁶http://lectures.molgen.mpg.de/Algorithmische_Bioinformatik_WS1011/dokumente/u9data/