Finding hidden messages from DNA

Bioinformatics course (MTAT.03.239)
22.09.16
Priit Adler
Introduction round!
Fundamentals of Biology

Course Features
- Video lectures
- Assignments and solutions
- Recitation videos
- Subtitles/transcript
- Exams and solutions

Course Description
Fundamentals of Biology focuses on the basic principles of biochemistry, molecular biology, genetics, and recombinant DNA. These principles are necessary to understanding the basic mechanisms of life and anchor the biological knowledge that is required to understand many of the challenges in everyday life, from human health and disease to loss of biodiversity and environmental quality.

Course Format
This course has been designed for independent study. It consists of four units, one for each topic. The units can be used individually or in combination. The materials for each unit include:
Bioinformatics Algorithms

An Active Learning Approach

1st Edition:

www.ester.ee/record=b4422219

2nd Edition:

www.ester.ee/record=b4517979

https://youtu.be/yfXeKPt0nw4
Chapter 1 videos

https://youtu.be/M6qVRaNjcY0?
list=PLQ-85lQIPqFNFDlaNMZ1JVN73rO1a0dpn
DNA replication
What is DNA?

Shape of DNA? human vs bacteria

Why replicate?
Finding Origin of Replication

Finding oriC Problem: Finding oriC in a genome.

• **Input.** A genome.
• **Output.** The location of oriC in the genome.

OK – let’s cut out this DNA fragment. Can the genome replicate without it?

This is not a computational problem!
Value of OriC

Biotechnology

ARS / plasmids

Biosynthesis
Value of OriC

OriDB provides a catalogue of confirmed and predicted DNA replication origin sites. At present this is limited to budding yeast (\textit{S. cerevisiae}) and fission yeast (\textit{S. pombe}). If you use OriDB, please cite our paper.

This website provides access to and searching of the database. The data have been collated by Conrad Nieduszynski as the culmination of a number of genome-wide studies to identify the location of replication origins throughout the budding yeast genome. In addition to the genome-wide studies data from a large number of other origin mapping and characterization studies have been included.
Insulin

insulin biosynthesis in *e. coli* - 1978.

on the market 1982

Now mostly produced using *e. coli* or *s. cerevisiae*

Potential to use plants for even cheaper alternative
How frequent is frequent?

Maybe there is a pattern I could look for, but I have no idea where to start.

http://www.csbio.unc.edu/mcmillan/Comp555S16/Lecture02.html
Does DNA have direction?

What is complementary match?

Do you know OriC region?

Probability of 3, 6, 9, 12-mer?
Better way to find OriC region

https://www2.estrellamountain.edu/faculty/farabee/biobk/BioBookDNAMOLGEN.html
Okazaki fragments

ssDNA mutation rate:

Deamination of C and ultimately mutation of C to T

#G - #C skew

k-mers with mismatches
#G - #C skew

http://www.g-language.org/wiki/tutorialgcskewenglish
#G - #C skew

Bacteria - nice

Archea - not so nice

Eukaryotes - a mess
#G - #C skew

Bacteria - nice

- OriC (mostly 1)
- Genome rearrangement
#G - #C skew

Archea - not so nice

- multiple OriC
- genome rearrangement
- transcription
#G - #C skew

Eukaryotes - a mess

- multiple OriC
- genome rearrangements
- transcription: 3’ and 5’
GC skew at the 5' and 3' ends of human genes links R-loop formation to epigenetic regulation and transcription termination

Paul A. Ginno1,2,4, Yoong Wearn Lim1,3, Paul L. Lott2, Ian Korf1,2 and Frédéric Chédin1,2,5

http://genome.cshlp.org/content/23/10/1590.full
Summary

Finding OriC as computational problem!

How much biology you need to know to solve problems?

Talk to people! Especially biologists!
Locations

Rosalind is a platform for learning bioinformatics and programming through problem solving. Take a tour to get the hang of how Rosalind works.

If you don’t know anything about programming, you can start at the Python Village. For a collection of exercises to accompany Bioinformatics Algorithms book, go to the Textbook Track. Otherwise you can try to storm the Bioinformatics Stronghold right now.

Python Village

If you are completely new to programming, try these initial problems to learn a few basics about the Python programming language. You’ll get familiar with the operations needed to start solving bioinformatics challenges in the Stronghold.

Bioinformatics Stronghold

Discover the algorithms underlying a variety of bioinformatics topics: computational mass spectrometry, alignment, dynamic programming, genome assembly, genome rearrangements, phylogeny, probability, string algorithms and others.

Bioinformatics Armory

Ready-to-use software tools abound for bioinformatics analysis. Whereas in the Stronghold you implement algorithms on your own, in the Armory you solve similar problems by using existing tools.

Bioinformatics Textbook Track

A collection of exercises to accompany Bioinformatics Algorithms: An Active-Learning Approach by Phillip Compeau & Pavel Pevzner. A full version of this text is hosted on stepic.org

Algorithmic Heights

A collection of exercises in introductory algorithms to accompany "Algorithms", the popular textbook by Dasgupta, Papadimitriou, and Vazirani.

http://rosalind.info/problems/locations/
Log in with OpenID

Do you already have an account on one of these sites? Click the logo to log in with it here:

- Google
- twitter
- facebook
- Stack Exchange
- OpenID

Why OpenID?
OpenID is service that allows you to log in to many different websites using a single identity. Find out more about OpenID and how to get an OpenID enabled account.

Want to use with an existing account?
If the existing account's email address matches your OpenID login, we will attempt to link the two. Otherwise, sign in with the existing account, click your username at the top of the page, then click my accounts to load the page for associating new OpenID accounts.

Log in with a Rosalind account

Username:

Password:

Log in

Forgot your Rosalind account information?
Click here to have an email sent with your username, along with instructions on how to reset your password.

Don't receive an activation e-mail?
Click here to re-send an activation e-mail.

Don't have a Rosalind account?
Click here to create a Rosalind account.

http://rosalind.info/problems/locations/
Rosalind

http://rosalind.info/classes/337/

http://rosalind.info/classes/enroll/e173f20008/
Chapter 1 homework

http://rosalind.info/classes/337/

Due 05.10.16 midnight

tasks 1-14: choose min 8, max 12 tasks