

## Class 14: Gene Prediction (1)

**Held:** Tuesday, 13 October 2009

**Summary:** We consider ways in which we can computationally identify genes (or sequences likely to be genes) in longer DNA sequences.

### Related Pages:

- EBoard.
- Reading: Chapter 6.

### Notes:

- Don't forget that we have a reading and reflection for Thursday.

### Overview:

- What can we do with long sequences?
- Gene finding
- Ways of predicting useful ORFs
- Obstacles to finding genes
- Web exploration

## Gene Prediction: Going Beyond DNA Sequences

- We've now figured out how to get long sequences of DNA.
  - See previous classes for details?
- What can we do with those sequences?
  - If we sequence something new, we can *search* those sequences for a similar sequence.
  - We can *compare* related sequences to see how they change (e.g., to build trees)
  - ...
- One clearly useful thing would be to find the genes in the genome
  - But it's a very hard problem

## Strategies for Gene Prediction

- Match with existing genes in other organisms
  - Feels like a bit of a chicken-and-the-egg problem
- String matching
  - Computer science is simple; we're good at finding strings (even approximate strings)
  - Biology is messy: We don't know what strings to predict.
- Other characteristics, such as base-pair frequency

## Obstacles to Gene Prediction

- Incomplete knowledge: Promotor sequences not universal (and probably not always known)
- Introns and Exons
- ...

## Web Exploration

- If there is time, we'll work on the Web exploration from the book.

---

Copyright © 2009 Vida Praitis and Samuel A. Rebelsky. This work is licensed under a Creative Commons Attribution-NonCommercial 2.5 License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/2.5/> or send a letter to Creative Commons, 543 Howard Street, 5th Floor, San Francisco, California, 94105, USA.