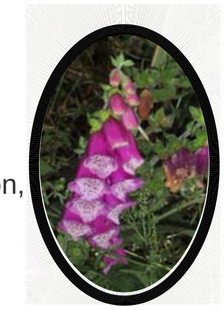
computing science and biology (1)

introduction and overview

connections between cs and biology



- biology is the science of life
- progress through observation, experimentation, theory
- technology in part drives advances in biology

example: bacteria

- von Leewenhoek (1683) discovered that in the white matter between his teeth there were millions of microscopic "animals – more, in fact, than there were human beings in the united Netherlands ... very prettily a-moving"
- Lister (1867) linked bacteria with disease

… today, we have treatments, prevention for many bacterial diseases; appreciation for roles of bacteria in our environment

example: evolution and genes

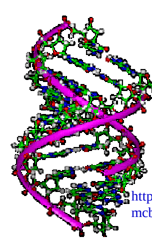
- Mendel (1865) experimented with pea plants to show inheritance of organism's traits
- Avery et al. (1944) established that genes, coded in DNA, carry our hereditary information

… today, these insights are leading to diagnoses and treatments of genetic diseases

the more we know, the more we know we don't know

- 99% of bacteria are unidentified, since they can't be cultured (grown) in a lab environment
- we don't know how many genes we have or what functions are associated with most of these genes

clues to further understanding lie at the molecular level



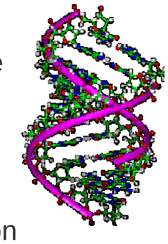
http://www.umaryland.edu/graduate/mcb/images/DNA2-smallest.gif

new technologies, including computers, are essential to the study of molecular biology

goal for today

- see some roles that computing science plays in advancing research in molecular biology
- but first, let's look at some of the molecules in the cell

our genome



- stores our genetic information in DNA molecules
- a double-stranded bead necklace with four different kinds of beads (bases, nucleotides): A,C,G, T
- beads are paired: A-T, G-C
- 3 billion base pairs in each cell
- on the order of one hundred trillion cells in an adult body (including bacterial cells, which have their own genomes)
- to store the raw information in our cells would require 30 trillion CDs!

proteins

- proteins are the body's activists: carry blood, digest food, form hair, fingernails, and much more
- beads on a necklace, with 20 different bead types (amino acids)
- the beads fold into interesting shapes
- the shape is key to the function of the molecule

http://www.nigms.nih.gov/news/science_ed/structlife/

genes

- to keep our body functioning, proteins are constantly manufactured in our cells
- genes – segments of our DNA – contain codes for proteins
- a codon – three bases of DNA – codes for one amino acid
- the *genetic code* specifies the correspondence between codons and amino acids

the genetic code

TTT phenylalanine TTC phenylalanine TTA leucine TTG leucine	TCT serine TCC serine TCA serine TCG serine	TAT tyrosine TAC tyrosine TAA stop TAG stop	TGT cysteine TGC cysteine TGA stop TGG tryptophan
CTT leucine CTC leucine CTA leucine CTG leucine	CCT proline CCC proline CCA proline CCG proline	CAT histidine CAC histidine CAA glutamine CAG glutamine	CGT arginine CGC arginine CGA arginine CGG arginine
ATT isoleucine ATC isoleucine ATA isoleucine ATG methionine (start)	ACT threonine ACC threonine ACA threonine ACG threonine	AAT asparagine AAC asparagine AAA lysine AAG lysine	AGT serine AGC serine AGA arginine AGG arginine
GTT valine GTC valine GTA valine GTG valine	GCT alanine GCC alanine GCA alanine GCG alanine	GAT aspartic acid GAC aspartic acid GAA glutamic acid GAG glutamic acid	GGT glycine GGC glycine GGA glycine GGG glycine

example: Methionine – Isoleucine – Phenelalanine – Aspartic Acid – Glycine … is coded by ATGATCTTTGACGGG … (as well as by other codes)

more on genes

- by the recent estimates, humans have perhaps as few as 20,000 genes
- we share
  - 99.9% of our genome with each other
  - 98% of our genome with chimpanzees
  - 50% of our genome with the roundworm
- *mutations* – changes in the bases of a gene – can cause genetic diseases

challenges in molecular biology

- what are our genes?
- what are our proteins?
- what to these proteins do?
- what genes, proteins do other organisms have?

success in answering these questions will lead to understanding and ultimately to better prevention and cure of diseases

what do computers provide?

- *tools to determine genomic sequences*
- *access to data*: annotated databases of genomic and protein data
- *tools for analyzing data*: learning what the data means: what are the structure of molecules, where are the genes
- *tools for visualizing data*: enabling visual interpretation of data

let's see some concrete examples

example: determining genomic sequences

- on 12 April 2003, a group at the BC Cancer Agency's Genome Sciences Centre in Vancouver, lead by Caroline Astell, became the first group worldwide to sequence the genomic material of the SARS virus
- computer assembly of sequence data was a major part of the effort

examples: providing access to data

- NCBI National Center for Biotechnology Information: repository of sequence data, including whole genomes of over 800 organisms
- Protein Information Resource: protein databases and analysis tools
  - founded in 1984, building on work of Margaret Dayhoff, who published the first comprehensive "Atlas of Protein Sequence and Structure" and who pioneered development of computer methods for comparing protein sequences
- specialized sites for organisms

example: how bacteria cause disease

- "Our laboratory is using computer-based analysis, combined with laboratory experimentation, to gain a better understanding of how some bacteria cause disease." – Fiona Brinkman, SFU, winner of the 2003 B.C. Science Council Young Innovator Award
- the genome sequence of a bacterium can be analyzed by computer to gain knowledge about virulent proteins produced by the bacterium
- many advantages over traditional approaches to understanding bacteria

example: how organisms are related

- Charles Darwin and his successors relied on comparison of visible traits of organisms to guess at evolutionary tree
- nowadays, DNA of organisms is compared, yielding more reliable trees

example: how organisms are related

"My research arose from a fascination with the diversity of forms and behaviours of jumping spiders, which lead to systematics, which led to phylogenetic theory and computer programming. "
– Wayne Maddison, Professor and Canada Research Chair, UBC

- Wayne and his brother maintain the 'Tree of Life' and MacClade websites; MacClade is a tool for analyzing phylogenetic trees.

what about influence of biology on computing?

- viruses, worms have taken on new meanings!
- evolution through genetic mutation is successful at "finding good solutions" to nature's "optimisation problems"; similar methods can be used in computations
- nature's ways of communication (e.g. ants) are also emulated in computational settings
- if DNA is such a remarkable means for information storage, could DNA be used for computing?

summary

- molecular approach to biology, with its associated vast quantities of sequence data, relies on sophisticated computational tools
  - databases
  - visualization and graphics
  - software engineering
  - algorithms
  - human-computer interaction
- at the same time, nature has made its mark on computational methods for solving problems

resources

- the structure of life
  - http://www.nigms.nih.gov/news/science_ed/structlife/
- sequencing SARS
  - http://www.vanmag.com/0306/sars.html
- Fiona Brinkman's lab:
  - http://www.pathogenomics.sfu.ca/brinkman/index.html
- Wayne Maddison's 'Tree of Life' site:
  - http://tolweb.org/tree/phylogeny.html
- UBC's Bioinformatics Centre BioTeach site:
  - http://www.bioteach.ubc.ca/Bioinformatics/