

Bioinformatics CSM17

Week1:What is Bioinformatics?

A Multidisciplinary Subject incorporating:

- ▶ **Biology**

- the study of living systems

- ▶ **Informatics**

- the representation, organisation, manipulation, distribution, maintenance and use of information

Bioinformatics includes...

- ▶ Computer applications in biology
- ▶ IT and IS for management & analysis of biological data
 - e.g. classification/identification systems, databases, development of analysis tools
- ▶ Molecular analysis (e.g. DNA sequences)
- ▶ Artificial intelligence
- ▶ Simulations of biological systems

Why is it Important to Mankind?

- ▶ Analysis of our 'own' data (human genome project –1990s)
- ▶ biotechnology / genetic engineering
- ▶ gene therapy
- ▶ resulting in...
 - cures for diseases
 - prevention of infection
 - increase in food production

Why is it Important to *Us*?

- ▶ Storage, access and presentation of large amounts of data (*databases, HCI*)
- ▶ Design of data analysis tools (*software eng.*)
- ▶ Pattern recognition (*neural networks/expert systems*)
 - e.g. identification of unknown specimens or gene sequences
- ▶ Prediction (*neural networks*)

Biological Data

- ▶ Often extremely variable
- ▶ Can be difficult to acquire
- ▶ Definitions often fuzzy or difficult to define
- ▶ Often dynamic – can change over time

Kinds of Biological Data

- ▶ Phenotype (the ‘hardware’)
 - *the actual organism as seen (WYS)*
- ▶ Genotype (the ‘software’, or ‘firmware’)
 - *the actual genetic code (‘design’, WYG)*

Kinds of Biological Data

- ▶ Phenetic (the ‘hardware’)
 - *analogue*
 - e.g. length of leaves, number of legs
- ▶ Genotypic (the ‘software’, or ‘firmware’)
 - *digital*
 - e.g. DNA, RNA, gene sequences
- ▶ Protein structures (the ‘components’)

Phenetic (Phenotypic) Data

- ▶ Stored as...*living* specimens
- ▶ In...
 - a botanic garden (plants & sometimes fungi)
 - a zoo (animals)
 - amateur collections
 - culture collections (bacteria, viruses, etc.)
- ▶ And...
 - (usually) local databases

Phenetic (Phenotypic) Data

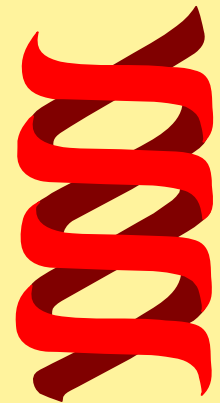
- ▶ Stored as... *preserved* specimens
 - as actual dried or 'pickled' plants & animals
- ▶ In...
 - a herbarium (plants & fungi)
 - a museum (animals)
- ▶ And...
 - books, published papers, illustrations, sounds
- ▶ ...essentially *multimedia* in nature

Databases for Phenetic Data

- ▶ Living Collection Accession Databases
 - e.g. at Royal Botanic Gardens, Kew (UK)
- ▶ Online Floras : www.eFloras.org
 - e.g. Flora of North America

Genotypic (Genomic) Data

- ▶ Molecular sequence data
 - stored as sequences
- ▶ In.... Large Databases
- ▶ Nucleic Acids
 - DNA (Deoxyribonucleic Acid)
 - RNA (Ribonucleic Acid)



Databases for Genomic Data

▶ GenBank

- NCBI National Center for Biotechnology Information.
- <http://www.ncbi.nlm.nih.gov/GenBank/index.html>

▶ EMBL–EBI

- European Bioinformatics Institute
- <http://www.ebi.ac.uk/embl/index.html>

Databases for Protein Data

- ▶ (RCSB) PDB
 - Research Collaboratory for Structural Bioinformatics
 - Protein Data Bank
 - <http://www.rcsb.org/pdb/>

A Note on Biological Names

- ▶ *Latin* name (binomial) is the primary key
- ▶ names *do* change (although they try not to)
- ▶ there is only *one* specimen (+ a few copies) that 'definitely' has that name (the type)

Summary

▶ Bioinformatics

- is a new multidisciplinary subject involving IT and IS applied to the management and analysis of biological data
- it requires the design & use of databases, software engineering, pattern recognition, expert systems & artificial neural networks (i.e. computing expertise)
- has the potential for resulting in...
 - cures for, and prevention of diseases
 - an increase in food production

Useful Websites & Journals

- ▶ *Bioinformatics* Journal (formerly *Computer Applications in the Biosciences*)
- ▶ Bioinformatics Organization:
<http://www.bioinformatics.org>
- ▶ Kew Gardens (UK):
<http://www.kew.org>
- ▶ Natural History Museum (UK):
<http://www.nhm.ac.uk>

References

- ▶ Gibas & Jambeck (2001). *Bioinformatics Computer Skills*. Chapter 1: Biology in the computer age.
- ▶ Attwood & Parry-Smith (1999). *Introduction to Bioinformatics*. Chapter 1: Introduction.
- ▶ Krane & Raymer (2003). Fundamental concepts of Bioinformatics (Preface)