

Martin John Bishop

Cambridge, UK

Milano, Italia

m.bishop@itb.cnr.it

The size of the DNA genome sequence data

◆ HIV virus	0.01 megabases
◆ Bacterium	4.5 megabases
◆ Yeast	12 megabases
◆ Nematode	100 megabases
◆ Fruit fly	165 megabases
◆ Human	3000 megabases

Mapping and sequencing

- ◆ Genome shotgun sequencing works for bacteria
- ◆ Higher organisms are usually sequenced from overlapping BAC clones
- ◆ A variety of methods and software are used to produce these physical maps

Mapping human genes

- ◆ Chromosome banding and visualization of probe location
- ◆ Rodent cell line hybrids containing one human chromosome
- ◆ Radiation hybrid mapping more powerful and needs sophisticated analysis software

Comparative method

- ◆ Life is more uniform at the molecular level than we might have imagined
- ◆ Genes from bacteria can be informative in human biology
- ◆ Model organisms are extremely informative
 - yeast, nematode, fruit fly
 - rice, thale cress (*Arabidopsis* - a brassica)
 - puffer fish, chicken, mouse

Gene numbers

◆ Yeast	12 Mb	6,183 genes
◆ Nematode	100 Mb	16,332 genes
◆ Fruit fly	165 Mb	13,601 genes
◆ Puffer fish	400 Mb	30,000 genes
◆ Chicken	1200 Mb	30,000 genes
◆ Human	3000 Mb	30,000 genes

Genes and proteins

- ◆ The number of human genes is about 30,000
- ◆ The number of human transcripts is about 60,000
- ◆ The number of human proteins is at least 300,000

Phenotype and genotype

- ◆ Genetic linkage mapping requires at least two variants of a characteristic
- ◆ The only way to map morphological characters
- ◆ Genetic Linkage User Environment (GLUE) is provided by LITBIO
<http://portal.litbio.org/menu-bin/GLUE/glue.pl>

Human single nucleotide polymorphism (SNP) data

- ◆ 1,420,000 human SNPs mapped by 2003
- ◆ Represents an SNP about every 1.9 Kbase
- ◆ Helps to define haplotype variation across genome
- ◆ Survey allelic variation for association with disease

Summary of Annotation on human Genome Build 36.1 dbSNP Build 126

- ◆ http://www.ncbi.nlm.nih.gov/SNP/snp_summary_byOrg.cgi?tax_id=9606&build_id=126

Genomic sequences - methods

- ◆ Comparative genomics
 - human/mouse/chicken/puffer fish
- ◆ Ensembl
 - EBI/Sanger project for automated annotation

Transcripts

- ◆ EST clustering
 - Microarrays
 - Alternative splicing
 - SNPs

- ◆ Analysis of expression data
 - Co-regulated genes
 - Hints about function

HAPMAP

- ◆ The HapMap is a catalog of common genetic variants that occur in human beings. It describes what these variants are, where they occur in our DNA, and how they are distributed among people within populations and among populations in different parts of the world. The International HapMap Project is not using the information in the HapMap to establish connections between particular genetic variants and diseases. Rather, the Project is designed to provide information that other researchers can use to link genetic variants to the risk for specific illnesses, which will lead to new methods of preventing, diagnosing, and treating disease.
<http://www.hapmap.org/whatishapmap.html.en>

HAPMAP 2

- ◆ The DNA samples for the HapMap have come from a total of 270 people. The Yoruba people of Ibadan, Nigeria, provided 30 sets of samples from two parents and an adult child (each such set is called a trio). In Japan, 45 unrelated individuals from the Tokyo area provided samples. In China, 45 unrelated individuals from Beijing provided samples. Thirty U.S. trios provided samples, which were collected in 1980 from U.S. residents with northern and western European ancestry by the Centre d'Etude du Polymorphisme Humain (CEPH).
<http://www.hapmap.org/hapmappopulations.html.en>

Application - Medicine

- ◆ Man and chimpanzee - language
- ◆ Man and mouse – genetics
- ◆ Man and rat – physiology
- ◆ Man and dog – QTLs, behaviour
- ◆ Man and pig – organ replacement? [More likely to be via human stem cell route.]

Human and chimpanzee

- ◆ <http://www.bushorchimp.com/>
- ◆ Genes for smell, digestion, long-bone growth, hairiness and hearing differ
- ◆ ASL 240 reliable signs
<http://www.mnsu.edu/emuseum/cultural/language/chimpanzee.html>

Man and mouse

- ◆ Human and mouse dysmorphologies can be related
- ◆ Mapping is possible in either species followed by gene identification for defect
- ◆ Large mutation programs exist for mouse e.g. <http://www.mut.har.mrc.ac.uk/introduction.htm>
- ◆ http://www.informatics.jax.org/searches/allele_report.cgi gave 232 matching Alleles (162 Genes/Markers represented) for deafness

Human and mouse 2

- ◆ <http://www.jax.org/>

- ◆ Oxford grid

- <http://www.informatics.jax.org/searches/oxfordgrid.cgi>

- ◆ Orthology map

- http://www.informatics.jax.org/reports/homologymap/mouse_human.shtml

Human and dog

- ◆ Boxers, Golden Retrievers, and St Bernards have high rates of lymphoma; whereas Great Danes, Irish Wolfhounds, and other large breeds are predisposed to bone cancer.
- ◆ Aggressive/Docile breeds
- ◆ <http://www.genome.gov/12511476>

Human and Fugu

- ◆ <http://fugu.biology.qmul.ac.uk/>
- ◆ **Highly Conserved Non-coding sequences are associated with vertebrate development**
<http://fugu.biology.qmul.ac.uk/frontPageMore.html>

Applications - Gambling

- ◆ Horse genome
- ◆ Greyhound/Whippet genome
- ◆ Fighting cock genome

Applications – Animal husbandry

- ◆ Cattle
- ◆ Sheep/ goat
- ◆ Deer
- ◆ Pig
- ◆ Chicken/ turkey/ ostrich
- ◆ Salmon/ trout/ tilapia

Applications – plant breeding

- ◆ Graminae -
rice/maize/sorghum/millet/oats/barley/
rye/wheat
- ◆ Brassicae – cabbage/turnip/rape
- ◆ Solanaceae – aubergines, peppers, potatoes,
tomatoes
- ◆ Chenopodiaceae – sugar beet, spinach
- ◆ Leguminosae – peas, beans