CMSC 838T – Lecture 14

- Genetics
  - Analyzing genes & inheritance
- Comparative genomics
  - Extracting information from cross-genome comparisons

Outline

- Genetics
  - Inheritance
  - Meiosis & recombination
  - Linkage maps
  - Genetic diseases
  - Polymorphisms
- Comparative genomics
Genetics

- **Inheritance**
  - Mendel studied inheritance in garden peas in 1865
  - Tested 34 varieties of peas, growing 28,000 pea plants
  - Found attributes determined by gene from each parent

- **Biology**
  - (Diploid) organisms maintain 2 copies of each chromosome
  - Each parent contributes 1 copy during reproduction (meiosis)
  - Chromosomes may be mixed during meiosis (recombination)
  - Different versions exist for each gene (alleles)
    - Example – blue eyes vs. brown eyes

Genetics – Inheritance

- **Phenotype determined by inherited genes**

![Diagram of inheritance process](image)
Genetics – Meiosis & Recombination

**Diploid parent**

- Ss

**Interphase**

- Ss

**Meiosis I**

- The alleles have segregated

**Meiosis II**

- S

**Gametes**

- S

- S

- S

- S

- Cross Over

- Homologous Chromosomes Pair

- Recombination

- Recombinant Chromatids

---

**Genetics – Linkage Maps**

**Biology of reproduction**

- Genes located on same chromosome (syntenic) usually transmitted together
- Recombination may separate genes
- Probability → distance on chromosome

**Linkage map**

- Examine inheritance data for population
  - Usually for genetic diseases
  - Estimate relative distance between genes
- May use SNP genetic markers
- Create linkage map
- Try to identify gene(s)

---

**Linkage map for corn**
Genetics – Genetic Diseases

- **Genetic diseases**
  - May be caused by single gene
    - Huntington’s, cystic fibrosis, sickle-cell anemia, etc...
  - May be caused by interaction between multiple genes
    - Asthma, heart disease, cancer
    - Gene may be a risk factor for disease
  - May be caused by multiple groups of genes
    - All showing same symptom

- **Clinical manifestation of genetics**
  - Susceptibility vs. resistance
  - Variations in disease severity or symptoms
  - Reaction to drugs (pharmacogenetics)

---

**Chromosome 4**

- **Huntington disease**
  - Wolf-Hirschhorn syndrome
  - PKU due to dihydropteridine reductase deficiency

- **Dentinogenesis imperfecta-1**
  - Acute lymphocytic leukemia

- **C3b inactivator deficiency**
  - Aspartylglucosaminuria
  - Williams-Beuren syndrome, type II

- **Sideroblastosis**
  - Anterior segment
  - Mesenchymal dysgenesis
  - Pseudohypopaldosteronism

- **Hepatocellular carcinoma**
  - Glutaric acidemia type II
  - Factor XI deficiency
  - Fletcher factor deficiency

- **MPS 1 (Hurler and Scheie syndromes)**
  - Mucopolysaccharidosis I
  - Periodontitis, juvenile

- **Dysalbuminemic hyperzincemia**
  - Dysalbuminemic hyperthyroxinemia
  - Aleukocytosis

- **Hereditary persistence of alpha-fetoprotein**
  - Aplastic anemia

- **Polycystic kidney disease, adult, type II**
  - Mucolipidosis II
  - Mucolipidosis III

- **Severe combined immunodeficiency due to 112 deficiency**
  - Bierer syndrome

- **Dysfibrinogemienemia, gamma types**
  - Hypofibrinogenemia, gamma types

- **Dysfibrinogenemia, alpha types**
  - Amyloidosis, hereditary renal, 105.200

- **Dysfibrinogenemia, beta types**
  - Facioscapulohumeral muscular dystrophy
Genetics – Polymorphisms & SNPs

♦ Polymorphism
  - Genetic variant appearing in > 1% of population
    ● Filter out spontaneous mutations
  - May form alleles (versions of genes)

♦ SNPs (Single Nucleotide Polymorphism)
  - Very common, 3.7 million (human) in dbSNP as of April 2003
  - Occurring every ~1250 bases (on average) between individuals
  - Most (estimated 99+%) have no effect on phenotype
    ● Occur in non-coding DNA, degenerate codons
  - Serve as markers for genes in laboratory
    ● If SNP is known to be physically close to gene (linked)
    ● Find SNP → identify allele, genetic disease
  - A few thousand SNPs can characterize human genome

Outline

♦ Genetics
♦ Comparative genomics
  - Genomes & model organisms
  - Genomic rearrangement / synteny
  - Genomic alignment
  - Clusters of orthologous genes (COGs)
  - Comparative gene analysis & prediction
Comparative Genomics

- **Description**
  - Large scale comparison of genomes
  - Understand biology of individual genome
  - Discover principles governing genomes

- **Assumption**
  - Biology is shared by different species
  - Analyzing multiple species together increases information
  - Can understand human genes by studying their relatives (orthologs) in simpler organisms
    - Mouse & rat used extensively as model for human

Genomics – Genomes & Model Organisms

- **Genome sequences in GenBank (April 2003)**
  - 519 eukaryotes
  - 426 bacteria
  - 1310 viruses
    (incl. SARS corona virus)

- **Model organisms**
  - Focus of multiple studies due to scientific interest
  - E. Coli, yeast, nematode worm, fruit fly, mustard plant, zebrafish, mouse, human

<table>
<thead>
<tr>
<th>Species</th>
<th>Genome (Mbp)</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mycoplasma</td>
<td>0.58</td>
<td>470</td>
</tr>
<tr>
<td>Rickettsia</td>
<td>1.11</td>
<td>834</td>
</tr>
<tr>
<td>Bacillus</td>
<td>4.2</td>
<td>4,100</td>
</tr>
<tr>
<td>E. Coli</td>
<td>4.6</td>
<td>4,288</td>
</tr>
<tr>
<td>Saccharomyces</td>
<td>13.5</td>
<td>6,034</td>
</tr>
<tr>
<td>Arabidopsis</td>
<td>119</td>
<td>25,498</td>
</tr>
<tr>
<td>Drosophila</td>
<td>165</td>
<td>13,601</td>
</tr>
<tr>
<td>Caenorabditus</td>
<td>97</td>
<td>18,424</td>
</tr>
<tr>
<td>Homo sapien</td>
<td>3,300</td>
<td>30,000+</td>
</tr>
</tbody>
</table>
Genomics – Gene Order Comparison

- **Chromosomal rearrangement**
  - Chromosomes can break at random location
  - Fragments rejoined at random by DNA repair mechanisms

- **Comparing genomes between species**
  - Analyze rearrangements using locations of orthologs
  - Gene order changed by rearrangements over time
  - Genes w/ similar biological function tend to remain localized

- **Synteny**
  - Same species – genes on same chromosome
  - Multiple species – matching sections of chromosomes (with same genes in same order)

Chromosomal Rearrangement – Genome Plot

- **Genome plot**
  - Dot matrix plot of genes
  - **Ortholog** (common ancestor & function)
  - **Paralog** (duplication)

- **Plot shows order of many orthologs preserved**
Chromosomal Rearrangement – Human Mouse

- **Human and mouse genomes**
  - 99% similar
  - Can cut human genome into >100 pieces and map onto mouse genome fairly accurately

![Genome Comparison Diagram]

Genomics – Genome Alignment

- **Aligning genomes**
  - Must be able to align very long sequences
  - Better alignments possible from similar genomes
  - GLASS – recursively align genomes starting with long matches
  - WABA – break genome into small overlapping pieces, align

- **Genome visualization tools**
  - VISTA
    - Sequence alignment visualization tool
    - Emphasize regions of high similarity
  - VCMAP
    - Visualize locations of orthologs between genomes
VISTA – Comparing Human & Rat Genomes

Virtual Chromosome Map – Human Rat Mouse

- **VCMap**
  - Rat map framework markers in red
  - Chromosome numbers in blue
  - Conserved regions in common color
Genomics – Clusters of Orthologous Genes

- **Comparing genomes to find orthologs**
  - Cluster orthologous genes (COG) between genomes
  - COGs usually represent classes of metabolic function
  - NCBI COG database stores relationships, annotations

- **Example**
  - Relationship between several yeast and bacterial orthologs in NCBI COG database

---

Genomics – Clusters of Orthologous Genes

- **Using COG database to predict gene function**
Genomics – Comparative Analysis & Prediction

- **Comparing genomes to extract information**
  - Conserved regions identify genes & regulation factors
    - ROSETTA – gene model + GLASS alignment

- **Predict gene function based on**
  - Cross annotation between genomes
  - Genes for proteins in same pathway should be correlated
    - Find correlated genes using clustering / statistics
  - Gene fusion
    - Genes producing interacting proteins sometimes fuse to produce single protein
    - Look for fused genes

Genetics & Comparative Genomics

- **Genetics**
  - Evaluation of inheritance based on phenotypes
  - Use variation in inheritance rate to derive linkage maps

- **Comparative genomics**
  - Extract information from multiple genomes
  - Techniques for mapping and comparing genomes

- **Analysis of evolutionary history**
  - Reveals selective pressure on mutations
  - Can provide hints to gene location & function