Genomes: Now that we have them, what can we do with them?

Shifra Ben-Dor
Bioinformatics and Biological Computing Unit,
Weizmann Institute of Science

INN COBI
What do genome browsers look like?

- They have some representation of the sequence
  - Sometimes vertical (NCBI)
  - Sometimes horizontal (UCSC, Ensembl, TAIR)
  - Sometimes circular (Microbial Browsers)

- They have some form of annotation
What do we use genome sequence for?

- Genes
- Splice Variants
- Variation analysis (SNP, mutations)
- Promoters
- Comparative Genomics
- Evolution
What do we use genome sequence for?

- Genes
- Splice Variants
- Variation analysis (SNP, mutations)
- Promoters
- Comparative Genomics
- Evolution
Looking for genes:

- In the lab:
  - Two hybrid system
  - SSH
  - Exon trapping
  - Linkage analysis
  - Database Searches
Looking for genes

• In the browser:
  - Existing mRNAs or ESTs from large scale projects
  - Gene Prediction programs
    • Notoriously unreliable
  - Genetic Maps and Markers
  - Comparative Genomics
Sources of mRNA’s

• Experimental
  - Clone new gene
  - Clone gene from database
  - 2 hybrid system

• Database
  - “Typical” cDNA
  - Full length cDNA
  - EST
Looking for genes

• In the browser:
  - Existing mRNAs or ESTs from large scale projects
  - Gene Prediction programs
    • Notoriously unreliable
  - Genetic Maps and Markers
  - Comparative Genomics
Gene Prediction Programs

• Many available, based on different algorithms, mainly checking for coding regions

• In one paper checking for accuracy, the measure used was “at least one exon correct”!
Gene Prediction Problems

• UTR’s (non-coding exons) usually wrong, or not predicted at all
• Interleaved genes
• Nested Genes
• Cassette genes
Looking for genes

• In the browser:
  - Existing mRNAs or ESTs from large scale projects
  - Gene Prediction programs
    • Notoriously unreliable
  - Genetic Maps and Markers
  - Comparative Genomics
Genetic Maps and Markers

• Go to an area with linkage disequilibrium

• Pull out a list of candidate transcripts
What do we use genome sequence for?

- Genes
- Splice Variants
- Variation analysis (SNP, mutations)
- Promoters
- Comparative Genomics
- Evolution
Splice Variants

- Can generally be seen by comparison of expressed sequences (mRNA and EST) to genomic sequence
What do we use genome sequence for?

- Genes
- Splice Variants
- Variation analysis (SNP, mutations, editing)
- Promoters
- Comparative Genomics
- Evolution
Variation (SNP, mutations, editing)

- Comparison of Genome sequence to transcribed sequences: mRNA, EST
- Look for differences
What do we use genome sequence for?

- Genes
- Splice Variants
- Variation analysis (SNP, mutations)
- Promoters
- Comparative Genomics
- Evolution
Genetic Variation

Since the DNA used in the various genome sequencing projects came from different donors, we can start looking at genetic variation.

There are many changes in our DNA. Some cause disease, some cause no change, some may be the reason people differ in reaction to various drug treatments.
DNA Genetic Code Dictates Amino Acid Identity and Order

DNA Sequence Is the Genetic Code.

Growing Protein Chain

GCA AGA GAT AAT TGT...

Ala  Arg  Asp  Asn  Cys

1  2  3  4  5

Slide taken from DOE Human Genome Program website http://www.ornl.gov/hgmis
DNA Sequence Variation in a Gene Can Change the Protein Produced by the Genetic Code

**Gene A from Person 1**
- GCA
- AGA
- GAT
- AAT
- TGT...
- Protein Products
  - Ala
  - Arg
  - Asp
  - Asn
  - Cys...

**Gene A from Person 2**
- GCG
- AGA
- GAT
- AAT
- TGT...
- Codon change made no difference in amino acid sequence
  - Ala
  - Arg
  - Asp
  - Asn
  - Cys...

**Gene A from Person 3**
- GCA
- AAA
- GAT
- AAT
- TGT...
- Codon change resulted in a different amino acid at position 2
  - Ala
  - Lys
  - Asp
  - Asn
  - Cys...
Health or Disease?

**DNA Sequence**

**Person 1**
A A A T T T T

↑

Normal protein

**Person 2**
A A T T T T T

↑

Some DNA variations have no negative effects

Low or nonfunctioning protein

**Person 3**
A A C T T T T

↑

Other variations lead to disease (e.g., sickle cell) or increased susceptibility to disease (e.g., lung cancer)

Slide taken from DOE Human Genome Program website
http://www.ornl.gov/hgmis
Promoters

• Make sure you have the 5’ end of the sequence (look at expressed sequences)
• Look for alternate starts
• Extract upstream regions
What do we use genome sequence for?

- Genes
- Splice Variants
- Variation analysis (SNP, mutations)
- Promoters
- Comparative Genomics
- Evolution
Comparative Genomics

- Look for genes in one species not found in another
- Look for genes in one species not present in another
- Gene Clusters / Synteny
- Look for conserved regulatory elements
How do we look at a browser

Three main ways to enter:

• Text Search
• Sequence Search
• Browse
Text Search

Type in name of what you’re looking for: gene, marker, locus, location...

• Problems: You get where they think the best location is. If there is an alternate location, you won’t see it.

• If there are related sequences (psuedogenes, family members, duplications...) you won’t see them.

• If there are problems with the assembly, you won’t see them
Sequence Search

• If you have ANY sequence, use it!
• Run searches against NR and Genome
  - Some genome sequences aren’t in the genome assembly
  - Some mRNA sequences are deposited as DNA in genbank
• If you are running a cross-species comparison, start with the protein
• Problem: you need sequence
Browse

• Usually, you can click on a part of the genome, and scroll in
• Usually, doesn’t give you much information unless you know where you’re going, or unless you’re preparing examples for a course
How do we look at what comes out of the search?

- Various annotations are kept in “tracks” or “maps”
- There is generally a way to control which ones are visible
- There are generally controls to zoom in/out, or jump to particular locations
Keep in mind what the input is

• If you are looking at markers, expect a point to be highlighted

• If you are looking at an mRNA/genomic comparison there will be regions of match and regions of mismatch