Human Genome, ENCODE Project:

- CpG islands: C followed G on particular strand, should occur on frequency of 4.41% but instead shows 1% in the human genome, therefore areas of higher CG frequency named CpG islands.
    - Associated with promoters/transcription start sites about ½ of the CpG islands are part of the ubiquitous Alu repeat, leaving 28,000 assoc with no repeats
- Transposon: less than ½ of human genome from transposon-like repeats
- Gene Density: GC% content at around 41%, differs by region
    - Theory that mammalian genomes have 5 large genomic rearrangements called isochores: 3 GC rich, 2 GC poor
- Paralog Clusters: most not near one another, more recent dups are closer to each other
- ENCODE Project: Encyclopedia of DNA Elements, catalogs all functional elements in genome, uses bioinfo./lab methods to catalog 1% of human genome (from 44 manually and randomly selected regions of the genome).
    - Looked for: protein coding genes, non-protein coding genes, transcriptional regulatory elements, sequences that mediate chromosome structure, and yet to be defined functional sequences
    - Mouse ENCODE: apply comparative genomics to ENCODE

Fungal(yeast) Genomes:

- *Saccharomyces cerrivisiae* budding yeast:
    - Well studied, grows like proks, small genome(~13MB), first eukaryote sequenced, very different from human genome, 4% of genes have introns (and not all ORFs = genes)
- Yeast Genome Browsers: NCBI, UCSC, and
    - Saccharomyces Genome Database(SGD): one of best curated species specific webe resources
- Gene Duplication: *S. cerrivisiae* dup'd whole genome 100+ mil years ago. When gene dup'd:
    - Both copies remain, little mutation, extra transcript
    - One copy is deleted or permanently damaged
    - One copy becomes a pseudogene through mutation
    - One/both copies evolve different
- Fission Yeast: *Schizosaccharomyces pombe* resembles higher level euks, only has three chromosomes, and has fewer genes than *S. cerrivisiae*.
    - More introns makes it better for studying splicing
    - Mechanism for nucleosome positioning different in fission yeasts, 1 extra gene found not found in budding yeasts.
- Separate DB for Aspergillus related species of yeast from the Broad institute

Non-Metazoan Genomes:

- Lower euk model systems: some well studied non fungal euks = *Giardia lamblia, Trichomonas vaginalis, leishmania*
- More single celled organisms studied in depth: Malaria, ciliphora, paramecium, tetrahymena thermophilia, stramenophilia
- Plants such as green algae, osteococcus tauri, and Arabidopsis thaliana
- Slime molds are the closest unicellular organism to metazoans

Metazoan Genomes:

- *Caenorhabiditis elegans* (roundworm) helped discover RNAi, source for first major genome database, good for studying developmental events
- *Drosophila melanogaster* (fruit fly) study developmental patterns though don't resemble mammals, far apart enough that candidate genes found here might be necessary for eukaryotic life

Euk Chromosomes and Structures:

- Features differences between euks and proks:
    o Genome size and non-coding DNA greater in euks
    o Chromosomes usually linear and more than one in euks
    o Some euks can be haploid
    o Splicing a significant process in euks
    o Chromosomes contain centromeres and telomeres in euks
- Repeats: 10% human genome = Alu repeats
    o Interspersed
    o Pseudogenes
    o Simple repeats
    o Segmental duplications
    o Tandem repeats
- Aneuploidy: abnormal chromosome number
- Inversions: region of 1 chr inverted, and can occur within populations
- Chr # and Synteny: lower chromosome # means they will be bigger, and vice versa. Synteny maps are designed to map related regions of genomes between different species.
- Sequence Tag Sites(STS): short unique DNA seqs suggested that these are loci for DNA primers, useful for cross genome comparison
- dbSTS hosted by NCBI, now moved to the Probe Database.
- ePCR: electric PCR useful to start a sequence and find STS sites to design PCR primers.