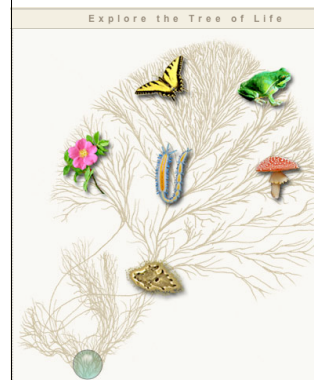


Lab section moved to Hitchcock 220

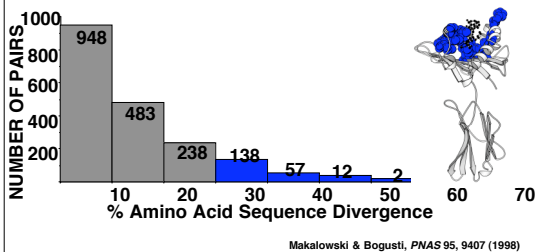
<http://tolweb.org/tree/phylogeny.html>



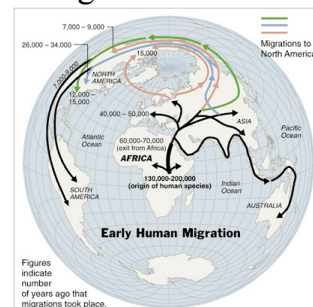
What is the
relationship of
life on earth?

What can rates of evolution tell us about function?

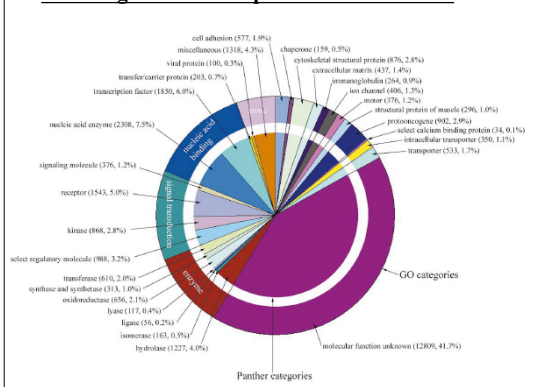
Rodent x Human 1880 Orthologous Sequence Pairs (~4% of genes)



Identifying human populations using molecular data



Human genome is complete: How is it used?



Prerequisites to Understanding Molecular Evolution

- Before you can analyze sequence, you have to understand their structure.
- Basic Biology so you can understand the language of the molecular evolution: Central Dogma
 - Transcription
 - Translation

Central Dogma of Molecular Biology

1) The concept of genes is historically defined on the basis of genetic inheritance of a phenotype. (Mendelian Inheritance)

2) The DNA an organism encodes the genetic information. It is made up of a double stranded helix composed of ribose sugars.

Adenine (A), Cytosine (C)

Guanine (G) and Thymine (T).

Central Dogmas of Molecular Biology

3) Each side of the double helix faces it's complementary base.

A $\downarrow$ T, and G $\downarrow$ C.

4) Biochemical process that read off the DNA always read it from the 5' side towards the 3' side. (replication and transcription).

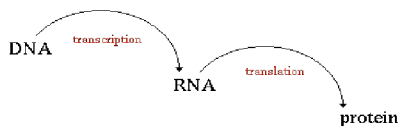
5) A gene can be located on either the 'plus strand' or the minus strand.

If the sequence on the + strand is ACGTGATCGATGCTA, the - strand must be read off by reading the reverse complement of this sequence going 'backwards'

e.g. TAGCATCGATCACGT

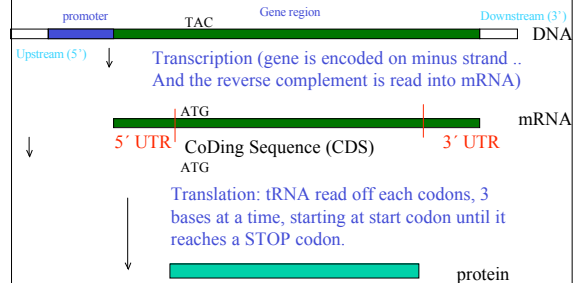
Central Dogmas of Molecular Biology

6) DNA information is copied over to mRNA that acts as a template to produce proteins.



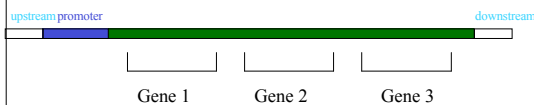
Prokaryotic genes

Prokaryotes (intronless protein coding genes)



Prokaryotic genes (operons)

Prokaryotes (operon structure)

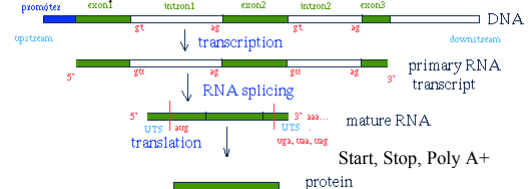


In prokaryotes, sometimes genes that are part of the same operational pathway are grouped together under a single promoter. They then produce a pre-mRNA which eventually produces 3 separate mRNA's.

Eukaryotic Central Dogma

In Eukaryotes, the DNA does not contain a continuous stretch of the coding gene, rather exons must be spliced. (However, some eukaryotes genes contain no introns)

mRNA – (messenger RNA) Contains the assembled copy of the gene. The mRNA acts as a messenger to carry the information stored in the DNA in the nucleus to the cytoplasm where the ribosomes can make it into protein.



introns

• Transcript region, interrupted by introns. Each intron

• starts with a donor site consensus
(G₁₀₀T₁₀₀A₆₂A₆₈G₈₄T₆₃..)

• Has a branch site near 3' end of intron (one not very conserved consensus UACUAAC)

• ends with an acceptor site consensus.
(12Py..NC₆₅A₁₀₀G₁₀₀)



Exons

• The exons of the transcript region are composed of:

• 5'UTR (mean length of 769 bp) with a specific base composition, that depends on local G+C content of genome)

• AUG (or other start codon)

• Remainder of coding region

• Stop Codon

• 3' UTR (mean length of 457, with a specific base composition that depends on local G+C content of genome)

Genetic Code

How does an mRNA specify amino acid sequence? Each amino acid is specified by a particular combination of three nucleotides, called a **codon**:

Each 3 nucleotide code for one amino acid.

• The first codon is the start codon, and usually coincides with the Amino Acid Methionine. (M which has codon code 'ATG')

• The last codon is the stop codon and does NOT code for an amino acid. It is sometimes represented by '*' to indicate the 'STOP' codon.

• A coding region (CDS,) starts at the START codon and ends at the STOP codon. Also called open reading frame (ORF).

Open Reading Frames (ORF)

On a given piece of DNA, there can be 6 possible frames. The ORF can be either on the + or minus strand and on any of 3 possible frames

Codon table

First Nucleotide		Second Nucleotide				Third Nucleotide	
		U	C	A	G	U	C
U	UUU	UCU	UAU	UGU	U	U	
	UUC	UCC	UAC	UGC	U	C	
	UUA	UCA	UAA	UGA	U	A	
	UUG	UCG	UAG	UGG	U	G	
C	CUU	CCU	CAU	CGU	C	U	
	CUC	CCC	CAC	CGC	C	C	
	CUA	CCA	CAA	CGA	C	A	
	CUG	CCG	CAG	CGG	C	G	
A	AUU	AUC	AUA	AUG	A	U	
	AUC	AUU	AUA	AUG	A	C	
	AUA	AUU	AUA	AUG	A	A	
	AUG	AUU	AUA	AUG	A	G	
G	GUU	GUU	GUU	GUU	G	U	
	GUC	GUC	GUC	GUC	G	C	
	GUA	GUA	GUA	GUA	G	A	
	GUG	GUG	GUG	GUG	G	G	

Start
AUG

The Genetic Code

- Code is triplet, 64 codons (4³ = 64 bases)
- Start and Stop codons
- Codon synonyms for 18 codons

Abbreviation	Amino acid
Ala	Alanine
Arg	Arginine
Asn	Asparagine
Asp	Aspartic acid
Cys	Cysteine
Gln	Glutamine
Glu	Glutamic acid
Gly	Glycine
His	Histidine
Ile	Isoleucine
Leu	Leucine
Lys	Lysine
Met	Methionine
Phe	Phenylalanine
Pro	Proline
Ser	Serine
Thr	Threonine
Tyr	Tyrosine
Val	Valine

Note the **degeneracy** of the genetic code. Each amino acid might have up to six codons that specify it.

- Different organisms have different frequencies of codon usage.
- A handful of species vary from the codon association described above, and use different codons for different amino acids.

~2% of the human genome encodes protein

Non-Coding Eukaryotic DNA

Untranslated regions (UTR's)

• introns (can be genes within introns of another gene!)

• intergenic regions.

- repetitive elements

- pseudogenes ("dead" genes)

Pseudogenes

Pseudogenes:

DNA sequence that might code for a gene, but that is unable to result in a protein. This deficiency might be in transcription (lack of promoter, for example) or in translation or both.

Processed pseudogenes:

Gene retroposed back in the genome after being processed by the splicing apparatus. Thus it is fully spliced and has polyA tail.

Insertion process flanks mRNA sequence with short direct repeats.

Thus no promoters. Unless is accidentally retroposed downstream of the promoter sequence.

Do not confuse with single-exon genes.

More Non-Protein genes

rRNA - ribosomal RNA

is one of the structural components of the ribosome. It has sequence complementarity to regions of the mRNA so that the ribosome knows where to bind to an mRNA it needs to make protein from.

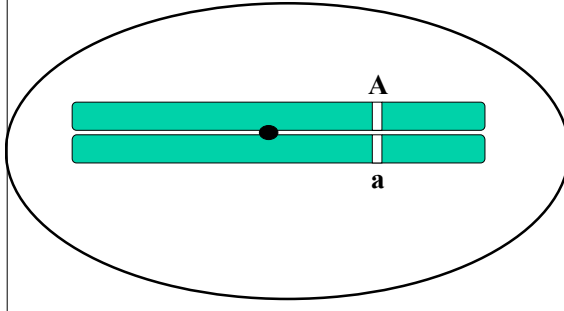
snRNA - small nuclear RNA

is involved in the machinery that processes RNA's as they travel between the nucleus and the cytoplasm.

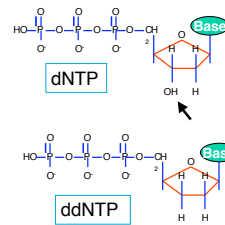
hnRNA - hetero-nuclear RNA.

small RNA involved in transcription.

What about variation? define allele



DNA Sequencing

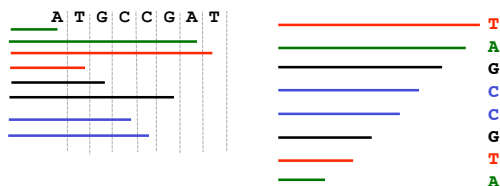


•Sanger et al. (1977)– 1980 Nobel Prize.

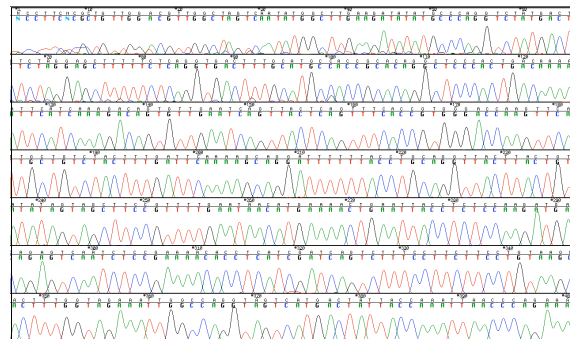
Basic Ingredients:

- Deoxy nucleoside triphosphate (dNTP)
- Dideoxy nucleoside triphosphate (ddNTP)
- ddNTPs act as chain terminators as they lack 3'-OH group
- Primers, polymerase

ddNTPS are labeled with fluorescent dyes.
Random incorporation of ddNTPS results in different sized fragments.
Separate by electrophoresis, visualize by exciting dyes.

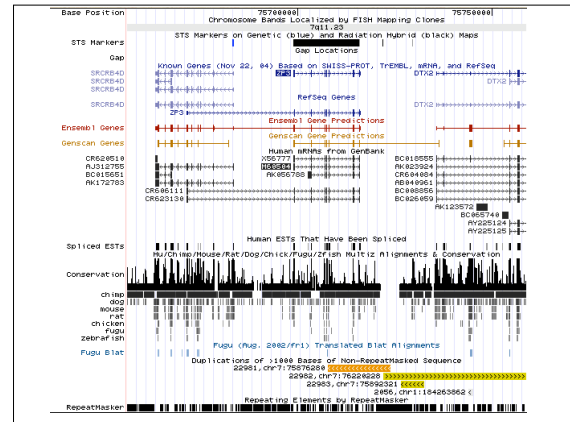


NOTE: Spacing should be roughly even since adding one nucleotide each elongation.



Types of sequencing projects

- Genome sequencing (draft, finished, annotated)
- Expressed Sequence Tags (ESTs)
- Library screening for specific genes.
- PCR amplification
- Phylogenetic sampling
- Single Nucleotide Polymorphisms (SNPs)



Lab section
moved to
Hitchcock 220

Wednesday Jan 5th
**Genome Sciences &
Combi Seminar:**
Dr. Martin Kreitman
"Deciphering rules
governing enhancer
functional evolution"
3:30, Hitchcock 132