

Outline

- Genetic variation
- Sequence alignments
- BLAST
- Exercise 2: Analyze Sanger sequencing data from week 1

Genetic Variation

Zygosity in diploid organisms

- Homozygous: two identical copies of an allele
- Heterozygous: two non-identical copies of an allele
- Hemizygous: only one copy of allele is present

Chromosome arm
Centromere
Allele

Sister chromatids

Homozygous Heterozygous Hemizygous

Autosomes
Sex chromosomes

Genetic Variation

Allele: one of multiple possible forms of a genetic locus.

- Different alleles of a gene often result in the same trait.
- Dominant allele: acts dominantly over a recessive allele.
- Recessive allele: acts recessively to a dominant allele.
- Codominance: phenotype is a mix of traits associated with two alleles.
- Incomplete dominance: phenotype is an intermediate between two alleles.

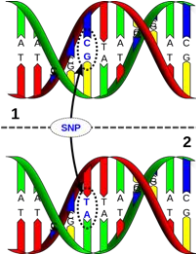
codominance

incomplete dominance

Genetic Variation

Genetic variation important in disease and research:

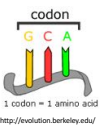
- Insertions and deletions (indels)
- Copy Number variants (CNVs)
- Single nucleotide polymorphisms (SNPs)
 - 10 million known SNPs in humans



~99% DNA similarity on average amongst individuals

Image: MIT, Christine Daniloff

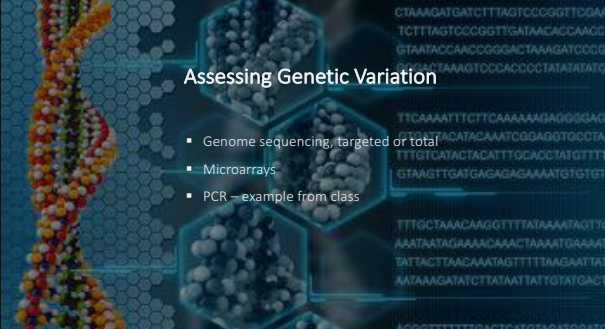
Genetic Variation



	Point mutations				
	No mutation	Silent	Nonsense	Missense	
				conservative	non-conservative
DNA level	TTC	TTT	ATC	TCC	TGC
mRNA level	AAG	AAA	UAG	AGG	ACG
protein level	Lys	Lys	STOP	Arg	Thr

- **Silent mutations:** don't change the amino acid sequence.
- **Nonsense mutations:** encode stop codons.
- **Missense mutations:** cause a change in the amino acid sequence.

Assessing Genetic Variation



- Genome sequencing, targeted or total
- Microarrays
- PCR – example from class

```

CTAAAGATGATCTTAGTCGCCGGTTCGAA
TCCTTAGTCGCCGGTTGATACCAACCAAC
GTAATACCAACCCGGACTAAGATCCCG
GGCACTAAAGTCCGCCCCCTATATATG
TTCAAAATTTCTTCAAAAAAGGGGAG
TTTCTTACATACAATCCGAGGTGGCTA
TTTGTATACAGATTTGCACCTATGTTT
GTAAGTTGATGAGAGAGAAAATGTGTGT
TTTGCTAAACAAGTTTTATAAATAGTTC
AAATAATAGAAAACAACATAAATGAAAT
TATACTTAACAATATTTTTAAGATAT
AAATAAGATATCTATAATATGTATGACT
ACGGTTTTTTTGACTCATGTAGATGGATC
    
```

Sequence Alignments

Arrangement of two or more DNA, RNA, or protein sequences to determine similarity.

Two categories

Global alignment: forced alignment across entire query sequence.

Local alignment: alignment of shorter regions of similarity within longer sequences.

```

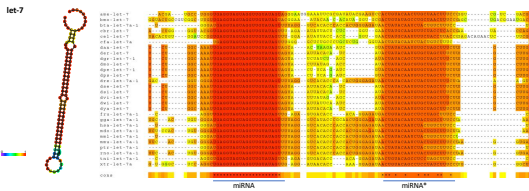
Global  FTFTALILLAVV
        F--TAL-LLA-AV

Local   FTFTALILL-AVAV
        --FTAL-LLAAV--
    
```

Sequence Alignments

Pairwise alignment: alignment of two sequences to identify the best match.

Multiple sequence alignment: alignment of three or more sequences.



Sequence Alignments

Some common alignment software:

Needle: software for pairwise global sequence alignment.

Water: software for pairwise local sequence alignment.

TCoffee: software for progressive alignment of two or more DNA, RNA, or protein sequences.



Clustal Omega: software for global alignment of multiple DNA, RNA, or protein sequences (up to several hundred thousand).



Basic Local Alignment Search Tool (BLAST): a popular search tool for comparing DNA or amino acid sequences.



BLAST

Basic Local Alignment Search Tool: compares sequences to large databases of sequences to identify homologous loci. Useful for identifying genes and homologs.

Query sequence is broken down into short fragments ("words") of 3 aa or 28 nt (by default)

↓

Words are matched to a database of words

↓

Two non-overlapping word matches within a short distance of each other are identified

↓

Maximal alignment identified by extending from better matching word in both directions

BLAST

Program	Query	Subject
blastn	DNA	DNA
blastp	Protein	Protein
blastx	DNA translated to protein	Protein
tblastn	Protein	DNA translated to protein
tblastx	DNA translated to protein	DNA translated to protein
blast2	DNA or protein	DNA or protein

Several databases to blast against:

- nr – all non-redundant sequences in GenBank
- Species-specific sequences (e.g. human, mouse, etc.)
- Custom databases
- Use on web, server, or personal computer

Exercise 2:

- Examine the sequence trace from your sequencing run.
- Align your sequence against the reference sequences.
- Blast your sequence and determine what gene it is.
- Determine the function of the gene.
- Answer the following questions and submit your answers on Canvas:
 1. What is the identity of the gene?
 2. What trait are certain alleles of the gene linked to?
