Genes and DNA

- The human genome is estimated to contain between 20,000 and 30,000 genes.
- Genes only comprise about 5% of chromosomal DNA
- 95% is non-coding found between and within genes.
Genes and DNA

- **Introns** – non-coding DNA sequences
- **Exons** – coding DNA sequences
- Genomic DNA (introns and exons) $\rightarrow$ pre-mRNA (introns and exons) $\rightarrow$ mRNA (exons) $\rightarrow$ protein

Genes and DNA

- Genes are relatively the same form individual to individual
- **Introns** vary in size among individuals
- This variation is the basis of DNA identification.
- The variation in the number bases in introns is the evolutionary result of transposition.
Transposons

- **Transposition** = genetic recombination that moves DNA from one site to another.
- **Transposable elements (transposons)** = mobile genetic elements
- Also known as “Jumping genes”
- Makes up 30% of our genome!

Transposons

- Discovered by Barbara McClintock in the 1940s
  - Won the Nobel prize in 1983
  - Studied transposon activity in Maize.
  - 50% of the Maize genome consists of transposons.
  - Insertions, deletions and translocations cause the color of the corn kernels to change.
Transposons

- Can insert within genes disrupting gene function
- Can insert within the regulatory sequence of a gene.
  - Can alter how a gene is expressed.
- Most common source of new mutations in many organisms

Transposons

- Can lead to genetic disease if gene function is altered.
  - Hemophilia
  - Predisposition to colon cancer
  - Duchene muscular dystrophy
- Transposons can be used to study evolutionary relationships.
  - Related species have similar transposon sequences
- Different organisms have a different distribution of transposons or sequences as a result of transposons.
- Introns are partly the result of transposon activity
Transposons

- Contain the following
  - Flanking host DNA
  - Target site duplication
    - Generated during the process of transposition
  - Terminal inverted repeats
    - Recognition sequences for Transposase
  - Transposase gene
    - Codes for transposase which cuts at the inverted repeat sequences.

Transposons

- **Autonomous transposon** – have inverted repeats and transposase gene.
- **Nonautonomous** – only terminal repeats but no transposase gene
  - Needs an autonomous transposon in order to transpose
Cut and Paste Transposition

- Cut and paste transposition
  - Remove DNA from its original location and insert DNA into new site.
- Transposase binds to terminal inverted repeats and cleaves the DNA.
- The excised DNA is inserted into the target DNA.
- A nick at the 3’ end of the target DNA is repaired to fill the gaps.
  - This creates a target site duplication
- The DNA strand that the transposon was removed from must now be repaired

Viral Like Retrotransposons

- Retroviruses such as HIV use this method of transposition.
- RNA is reverse transcribed into double stranded cDNA.
- The cDNA is then inserted into a target DNA sequence
SINEs and Alu elements

- SINEs – short interspersed nuclear elements
  - A type of retrotransposon.
  - 100-400 base pairs in length.
- The Alu sequence is a type of SINE
  - The Alu name comes from the Alu I restriction enzyme recognition site found within the sequence.
- The Alu sequence is repeated more than 500,000 times in our genome
- ~10.5% of the human genome

SINEs and Alu elements

- Can be used to study evolutionary relationships.
  - Ancestors share Alu sequence insertions only if they have a common ancestor.
- Sometimes found within the introns of genes associated with certain diseases.
  - Act as markers for certain genetic disorders such as diabetes, colon cancer and breast cancer.
- Can be used to determine relatedness of individuals.
DNA Fingerprinting

• **DNA profiling** - technique used by forensic scientists in order to identify individuals.
• .1% of the human genome varies from individual to individual.

DNA Fingerprinting

• DNA profiling uses **STR regions**
• **STR**- Short tandem repeat
• STRs are short nucleotide sequences that repeat.
• STRs are normally found in introns.
• STRs are 2-10 base pairs in length.
• There are over 10,000 published STR sequences.
DNA Fingerprinting

- STRs are similar between related individuals but very different between unrelated individuals.
- These variations are the basis of DNA profiling

DNA Fingerprinting - STR

- These regions are targeted with sequence specific primers and amplified with PCR.
DNA Fingerprinting - STR

- The fragments are then separated by gel electrophoresis.
- A special ladder is used called an Allelic ladder
  - Contains all the possible repeat sizes so the EXACT size of the repeat can be determined.

DNA Fingerprinting - STR

- Benefits
  - At least 13 loci are used which assort independently.
    - High degree of accuracy based on statistics
    - The probably of a particular combination of these 13 loci is one in a quintillion (1 with 18 zeros after it).
    - This means that it is statistically impossible for two people to have the same DNA fingerprint.
  - Degraded DNA can be used.
Fingerprint Considerations

• Could the match be an accidental due to contamination.
  – If not, could the DNA sample have been planted?
  – If not, did the accused leave the DNA sample at the exact time of the crime?
  – If yes, does that mean the accused is guilty of the crime?
• Lab errors.

CODIS

• CODIS - Combined DNA Index System
• DNA profiles of offenders convicted of certain misdemeanors to sexual assault and murder.
  – Each state has its own qualifying offenses.
• Allows law enforcement to compare DNA profiles electronically.
Lab Overview

- We will measure genetic variation by looking for the frequency of Alu repeats within a population (This class).
- We will look at a single Alu sequence within the PV92 locus on chromosome 16.
  - This is a noncoding region of the PV92 locus.
- Using your own DNA, you will use PCR to amplify this particular region on Chromosome 16
- The PCR primers will flank the region were the Alu element would be found.
- A gel electrophoresis will be used to determine your genotype.

Lab Overview

- This particular Alu element is dimorphic
  - **Dimorphic** = present in some individuals but not in others.
- The following combinations exist
  - Homozygous for the Alu sequence
  - Heterozygous for the Alu sequence
  - Homozygous without the Alu sequence
Lab Overview

- Gel electrophoresis
  - The Alu insertion is about 300bp in length
  - The PV92 locus is about 641bp in length
  - Homozygous for Alu (+/+) • One band at 941bp
  - Homozygous for no Alu (-/-) • One band at 641bp
  - Heterozygous for Alu (+/-) • One band at 941bp
  • One band at 641bp

Fig. 3. The presence or absence of the Alu insert within the PV92 locus on chromosome 16.
Lab #1 – DNA extraction

• You will use your cheek cells as a DNA template.
• Rinse your mouth with a salt solution and collect cells with a centrifuge.
  – A visible pellet of cells about the size of a match head should be visible.
• The cells will be lysed by heating.
• This will release enzymes found in the lysosomes such as DNAase.
  – DNAase breaks down DNA!
• The InstaGene matrix contains negatively charged beads that chelate metal ions out of the solution.
  – Chelate = grab
• Metal ions act are necessary for DNAase activity

Labs #2 and #3 – PCR Amplification and Gel Electrophoresis

• Your DNA will be amplified at the PV92 locus.
  – Do not transfer any Instagene beads to the PCR reaction.
  – The beads will inhibit the PCR reaction by removing the Mg$^{2+}$ beads needed by Taq polymerase.
• The amplified DNA will be separated by gel electrophoresis and analyzed for the Alu sequence.
Gel Electrophoresis Considerations

- Competition during amplification of heterozygous samples
  - Amplification of heterozygous samples is difficult due to competition between both reactions.
  - The smaller is amplified more efficiently and will appear darker.
  - The larger band may be very faint.
  - Carefully interpret your gels
  - Soaking your gel in Ethidium bromide overnight will make visualization easier.

Gel Electrophoresis Considerations

- Larger band in heterozygous sample
  - Heteroduplexes may form during amplification causing the larger band to run slower.
  - Band(s) may appear at about 1,000 to 1,700 bp.
Lab #4 – Analysis of Results

• Determine the size of your fragments and determine your genotype.
• Collect class data.
• Determine the allelic frequency for the class and for the US population.

\[ p = \text{frequency of (+) allele} = \frac{\text{number of (+) alleles}}{\text{total number of alleles (both + and -)}} \]
\[ = \frac{2(\# \text{ of } +/+ \text{ students}) + 1(\# \text{ of } +/- \text{ students})}{\text{total number of alleles (both + and -)}} \]
\[ = \text{frequency of } (+/+ \text{ students) + } \frac{1}{2} \text{(frequency of } (+/- \text{ students))} \]

\[ q = \text{frequency of (-) allele} = \frac{\text{number of (-) alleles}}{\text{total number of alleles (both + and -)}} \]
\[ = \frac{2(\# \text{ of } -/- \text{ students}) + 1(\# \text{ of } +/- \text{ students})}{\text{total number of alleles (both + and -)}} \]
\[ = \text{frequency of } (-/- \text{ students) + } \frac{1}{2} \text{(frequency of } (+/- \text{ students))} \]
Lab #4 – Analysis of Results

• Hardy-Weinberg Equation
  \[ p^2 + 2pq + q^2 = 1 \]
  • Describes the frequencies of genotypes in a population that is at “genetic equilibrium”
    – Population is large
    – Members mate randomly and produce offspring with equal success.
    – No migration of individuals into or out of the population
    – No excessive mutation

\[ p^2 = \text{the expected frequency of the (+/+)} \]
\[ 2pq = \text{the expected frequency of the (+/-)} \]
\[ q^2 = \text{the expected frequency of the (-/-)} \]
Lab #4 – Analysis of Results

• These results are theoretical where the Hardy Weinberg conditions are met.
• Real life population sample may vary.
• Complete the Hardy Weinberg calculations for both your class data and the US population data
• If your observed genotypic frequencies for the data match your Hardy Weinberg calculations then your data is in genetic equilibrium.

DNA Fingerprinting – RFLP

• RFLP – Restriction Fragment Length Polymorphism
  – Variation in lengths of restriction fragments generates fingerprints
  – Extract DNA from cells
  – Cut DNA into fragments using restriction enzymes
  – Separate band using gel electrophoresis
  – Transfer DNA to a nitrocellulose membrane using a technique called Southern Blotting
  – Treat with a radioactively labeled DNA probes which binds to specific sequences on the membrane
  – Wash off excess DNA probes
  – Take a picture

• Problems
  – The EXACT sizes of the bands is unknown since can only estimate next to a ladder
  – Time consuming
  – High quantity of high quality DNA
  – Laboratory error rate is high