**Colorado Technical University**

 **Course:** MATH366 – Probability and Statistics

#### Unit 6 Part 11 Readings: DNA Probability

**DNA Probability**

DNA profiling is a technique employed by forensic scientists to assist in the identification of individuals on the basis of their respective DNA profiles

Molecular biologists have identified regions of human DNA called "markers" which are highly variable between individuals

Because they vary so much, these markers are good for identifying and comparing individuals

They can be used, for example, for parental testing and rape investigation

The same lengths at all markers mean it is the same person or an identical twin

Similar lengths mean people are related, different lengths mean they are unrelated

**The CODIS markers**

In the United States the FBI has organized the CODIS (COmbined Dna Index System) database using 13 markers for forensic identification in criminal cases

The tables below show the number of possible values for all 13 CODIS markers and the AMEL (gender):

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Marker | D3S1358 | vWA | FGA | D8S1179 | D21S11 | D18S51 | D5S818  |
| Alleles | 8  | 11  | 14  | 12  | 22  | 21  | 10  |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Marker | D13S317 | D7S820 | D16S539 | THO1 | TPOX | CSF1PO | AMEL |
| Alleles | 8  | 10  | 9  | 7  | 8  | 10  | X Y |

If there are two possible values, A and B, at a genetic marker in a population, there are three possible pairs (genotypes): AA, BB, and AB

If there are three values, A or B or C, there are six possible genotypes: AA, BB, CC, AB, AC, and BC

For *n* different alleles, the total number of possible genotypes is:

### DNA Profile Frequency Calculations

### Genotype Probability at any STR Marker

Part of the work of forensic DNA analysis is the creation of population databases for the STR markers studied. Probability calculations are based on knowing allele frequencies for each STR marker for a representative human population. Allele frequency is defined as the number of copies of the allele in a population divided by the sum of all alleles in a population. For an individual, if the two alleles have frequencies of *p* and *q* in a population, the probability (P) of an individual of having both alleles at a single marker is:

P = 2 *p* *q*

### Probability for a DNA profile of Multiple Markers

If allele frequency for different markers can be shown to be independently inherited, the probability for the combined genotype can be determined by the multiplication (product) rule.

The probability (P) for a DNA profile is the product of the probability (P1, P2, ... Pn) for each individual marker:

Profile Probability = (P1) (P2) ... (Pn)

The probability can be an extremely low number when all 13 CODIS STR markers are included in the DNA profile.