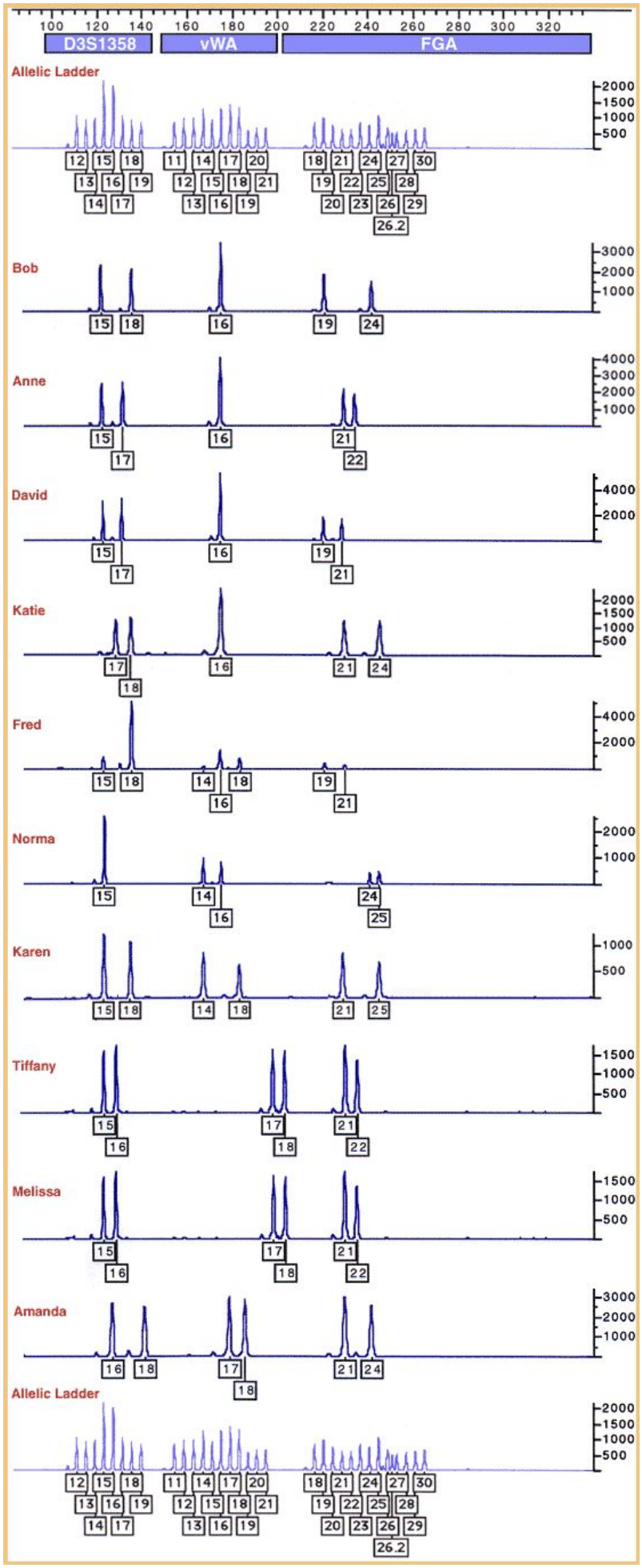
100 120 140 160 180 200 220 240 260 280 300 320
D3S1358
vWA
FGA
Allelic Ladder
2000
1500
1000
500
12 15 18
13 16 19
14 17
11 14 17 20
12 15 18 21
13 16 19
18 21 24 27 30
19 22 25 28
20 23 26 29
26.2
Bob
3000
2000
1000
15 18
16
19
24
Anne
4000
3000
2000
1000
15
17
16
21
22
David
4000
2000
15
17
16
19
21
Katie
2000
1500
1000
500
17
18
16
21
24
Fred
4000
2000
15 18
14
16
18
19
21
Norma
2000
1000
15
14
16
24
25
Karen
1000
500
15 18
14
18
21
25
Tiffany
1500
1000
500
15
16
17
18
21
22
Melissa
1500
1000
500
15
16
17
18
21
22
Amanda
3000
2000
1000
16 18
17
18
21 24
Allelic Ladder
2000
1500
1000
500
12 15 18
13 16 19
14 17
11 14 17 20
12 15 18 21
13 16 19
18 21 24 27 30
19 22 25 28
20 23 26 29
26.2

What are the 13 core CODIS loci?

A National DNA Databank

The Federal Bureau of Investigation (FBI) of the US has been a leader in developing DNA typing technology for use in the identification of perpetrators of violent crime. In 1997, the FBI announced the selection of 13 STR loci to constitute the core of the United States national database, CODIS. All CODIS STRs are tetrameric repeat sequences. All forensic laboratories that use the CODIS system can contribute to a national database. DNA analysts like Bob Blackett can also attempt to match the DNA profile of crime scene evidence to DNA profiles already in the database.

There are many advantages to the CODIS STR system:

- The CODIS system has been widely adopted by forensic DNA analysts
- STR alleles can be rapidly determined using commercially available kits.
- STR alleles are discrete, and behave according to known principles of population genetics
- The data are digital, and therefore ideally suited for computer databases
- Laboratories worldwide are contributing to the analysis of STR allele frequency in different human populations
- STR profiles can be determined with very small amounts of DNA

A DNA Profile: The 13 CODIS STR loci

As part of his training and proficiency testing for DNA Profile analysis of STR (Short Tandem Repeat) Polymorphisms, Forensic Scientist and DNA Analyst Bob Blackett created a DNA profile on his own DNA. Here is Bob's DNA Profile for the 13 core Genetic Loci of the United States national database, CODIS (Combined DNA Index System):

Locus	D3S1358	vWA	FGA	D8S1179	D21S11	D18S51	D5S818
Genotype	15, 18	16, 16	19, 24	12, 13	29, 31	12, 13	11, 13
Frequency	8.2%	4.4%	1.7%	9.9%	2.3%	4.3%	13%

Locus	D13S317	D7S820	D16S539	THO1	TPOX	CSF1PO	AMEL
Genotype	11, 11	10, 10	11, 11	9, 9.3	8, 8	11, 11	X Y
Frequency	1.2%	6.3%	9.5%	9.6%	3.52%	7.2%	(Male)

For each genetic locus, Bob has determined his "genotype", and the expected frequency of his genotype at each locus in a representative population sample. For example, at the genetic locus known as "D3S1358", Bob has the genotype of "15, 18". This genotype is shared by about 8.2% of the population. By combining the frequency information for all 13 CODIS loci, Bob can calculate that the frequency of his profile would be 1 in 7.7 quadrillion Caucasians (1 in 7.7 times 10 to the 15th power!

In Bob's forensic DNA work, he often compares the DNA profile of biological evidence from a crime scene with a known reference sample from a victim or suspect. If any two samples have matching genotypes at all 13 CODIS loci, it is a virtual certainty that the two DNA samples came from the same individual (or an identical twin).

Calculating Genotype frequency:

For a heterozygous 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 locus is

$P = 2pq$

If an individual is homozygous for an allele with a frequency of p, the probability (P) of the genotype is

$P = p^2$

We saw earlier that Bob Blackett has the genotype 15, 18 at the locus D3S1358. In a reference database of 200 U.S. Caucasians, the frequency of the alleles 15 and 18 was 0.2825 and 0.1450, respectively. The frequency of the 15, 18 genotype is therefore

$P = 2\ (0.2825)\ (0.1450) = .0819$, or 8.2%.

Calculating Frequency on Multiple Loci

If databases of allele frequency for different loci can be shown to be independently inherited by appropriate statistical tests, 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 (P_1, P_2, ... P_n) for each individual locus, i.e.

Profile Probability = $(P_1)\ (P_2)\ \ldots\ (P_n)$

The probability can be an extremely low numbers when all 13 CODIS STR markers are included in the DNA profile. As mentioned earlier, Bob Blackett calculated his own profile probability at 1.3 times 10^{-16}, or no more frequent than 1 in 7.7 quadrillion individuals (7.7 million billion), which is more than a million times the population of the planet.