CHAPTER 3

DNA Fingerprinting: Single-locus Analysis

Introduction

In this experiment you will fingerprint your DNA using a single-locus probe. To this end you will use the membrane prepared in Chapter 2. While performing single-locus analysis you will learn two procedures: removing old probe from the membrane (membrane stripping) and a hybridization signal detection method called direct detection.

First, the multi-locus probe will be stripped off the membrane. Membrane stripping procedures are frequently used in research and forensic analysis. It permits hybridization of several different probes to the same Southern blot. Next, the membrane will be hybridized with a single-locus probe and the position of the hybridized band will be visualized by chemiluminescence. The procedure you will use is a standard procedure used in forensic laboratories. In this hybridization protocol, probe is conjugated directly to reporter enzyme – alkaline phosphatase. The presence of enzyme during hybridization makes it necessary to modify standard hybridization procedures in order to protect the enzyme. This technique is frequently called the direct detection procedure and is more sensitive and faster than the indirect detection used in Chapter 2 (Surzycki, 2000).

The D2S44 single-locus probe will be used in this experiment. This probe represents a tandem repeat region present on human chromosome 2. Chromosome 2 is the result of fusion between the two ancestral primate chromosomes. The human repeat unit, with one single base substitution, is present in both the gorilla and the chimpanzee, but it is not tandem repeated. Thus, this tandem repetition arose after the evolutionary split between chimpanzees and humans (Nakamura et al., 1987; Holmlund and Lindblom, 1995; Evertsson, 1999; Holmlund, 1999).

This experiment will be carried out over two laboratory periods. In the first laboratory period, the nylon membrane will be stripped, hybridized, and exposed to film. In the second laboratory period, you will analyze the results. Figure 3.1 presents a schematic outline of the experiment.

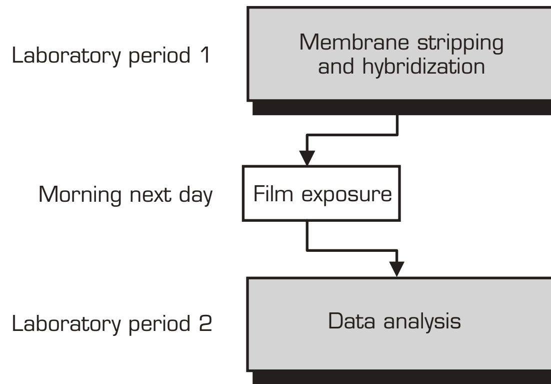

Figure 3.1 Schematic outline of the single-locus DNA fingerprinting experiment.

Background

DNA typing with mini-satellite (variable number tandem repeat or VNTR) loci is achieved by creating DNA fragment lengths containing variable numbers of repeated monomeric species. For this, genomic DNA is restricted with enzyme that cuts externally to the block of tandem repeats and analyzed using Southern blot hybridization. A Southern blot technique is essential in mini-satellite (VNTR) analysis due to a large fragment length for these loci (800 bp to several kilobases). This fragment size is outside the capability of PCR (polymerase chain reaction) techniques.

The multi-locus analysis described in Chapter 2 uses consensus core sequence as a hybridization probe. This probe, under low stringency hybridization conditions, hybridizes to a number of related VNTR sequences. Consequently, polymorphism at multiple loci is simultaneously identified and appears as a large number of bands on an autogram. The technique has its advantages and disadvantages. The advantage of this analysis is that it is highly informative because large numbers of loci are analyzed at the same time. The major disadvantages of the method are that alleles at different loci cannot be identified, thereby resulting in difficulties of statistical population analysis and it is difficult to reproduce the results due to difficulties in controlling the stringency of hybridization on which a number of the bands depend. These difficulties led to the development of a single-locus technique for analysis of VNTR allelic polymorphism.

The single-locus DNA typing (DNA fingerprinting) technique is very similar to the multi-locus method. First, DNA is digested with restriction enzyme that does not recognize the DNA sequence within the VNTR core. Next restricted DNA is hybridized by Southern blot using a specific probe that recognizes only one member of the VNTR family. The probe is usually

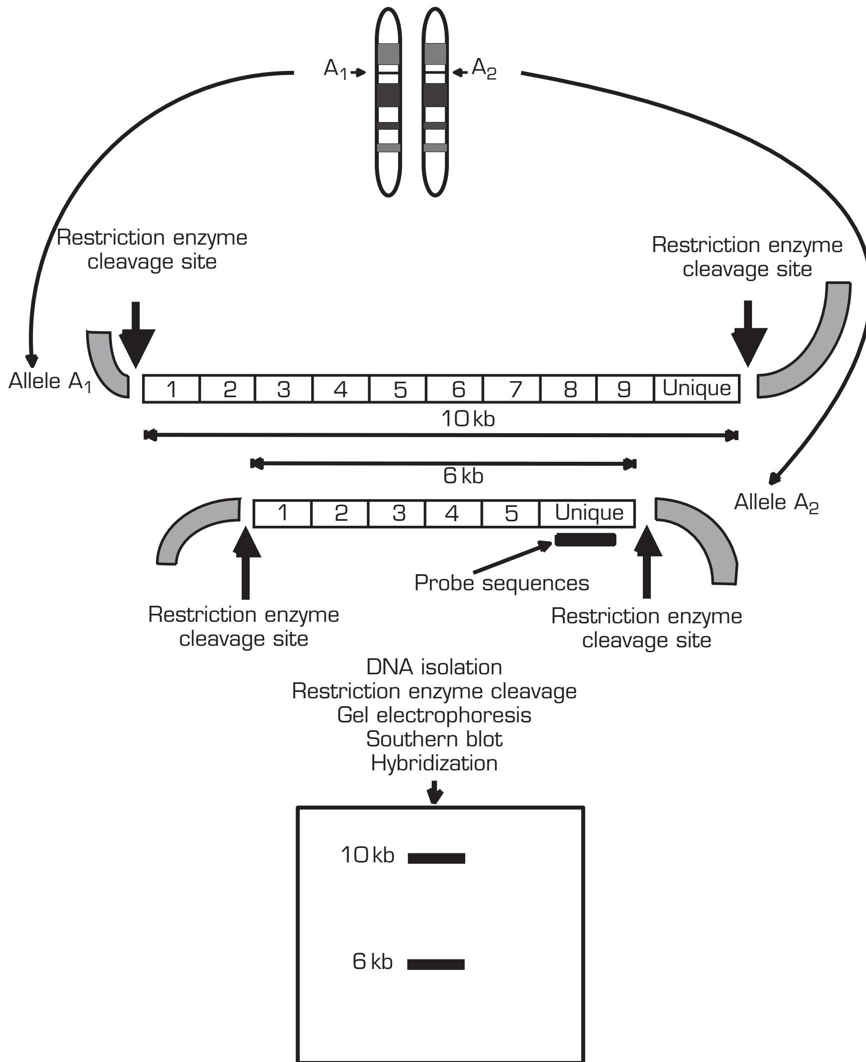

Figure 3.2 Principle of single-locus DNA fingerprinting. The single VNTR locus A is present on two homologous chromosomes at positions A_1 and A_2. The repeat of 1,000 bp is repeated nine times in allele A_1 and six times in allele A_2. Purified DNA is cut with restriction enzyme that does not have a recognition site in the repeated elements, but cuts DNA at some short distance from the beginning of the repeat (unique site). Restriction of the A_1 allele generates DNA fragments of 10 kb and restriction of the A_2 allele results in DNA fragments of 6 kb. Both A_1 and A_2 fragments contain the repeated element and DNA adjacent to it. This adjacent DNA is identical for both alleles and present only once in the entire genome. The fragments are separated by agarose gel electrophoresis and Southern blotted to a membrane. Hybridization is carried out with probe that is **not complementary** to the repeated element but to the **unique DNA** located adjacent to locus A. Hybridization generates two bands corresponding in size to two A alleles present in two homologous chromosomes.

a unique DNA sequence that is located between the end of the repeated array and the site of the restriction enzyme cut. The principle of this method is illustrated in Fig. 3.2. Consequently, as a result of hybridization, only two bands are obtained rather than the multiple bands present in multi-locus analysis. Because the single loci detected by this method are characterized, the bands represent the DNA genotype of the individual analyzed.

DNA profiling using single-locus probes is most frequently used in forensic analysis. This is because the use of these probes makes it possible to analyze data using statistical analysis of population genetics. In addition, interpretation of the results is easy since each individual has only two polymorphic variants, one inherited from the mother and one from the father. Multi-locus DNA fingerprinting analysis using single-locus probes is generated by the application of several probes (usually five to six), each one specific for a well-characterized polymorphic locus. The probes currently in use are those for the loci D1S7, D2S44, D4S139, D5S110, D10S28, and D17S79.

In the USA, the most widely used method for forensic profiling is based on a protocol developed by the FBI. It uses the enzyme *Hae*III for digesting genomic DNA samples and employs probes for single-locus VNTR markers (Budowle et al., 2000).

The locus that is used in this experiment is D2S44 VNTRs located on the long arm of chromosome 2. It has more than 30 alleles, ranging in size from 1.0 to 5.0 kb.

Protocol

Stripping the membrane

In order to hybridize the membrane with the new probe it is necessary to remove old probe from the membrane. Using this procedure one can strip and reprobe a membrane at least ten times.

1. Working in gloves, cut open the hybridization bag containing the membrane. Cut with scissors along three edges of the bag being careful not to cut into the membrane. Using forceps open the bag "like a book." Pick up the membrane with your gloved hands and roll it into a tube-like shape with the DNA side being inside.

2. Insert the rolled membrane into a roller bottle. Make sure that the side of the membrane with DNA is facing away from the glass. Do not worry that the rolled membrane is narrower than the bottle. The membrane will expand when you add liquid to the roller bottle.

3. Add 30 ml of sterile water to the roller bottle. Wash the membrane rotating at maximum speed for 5 minutes. Discard the water. This step removes CDP-Star (chemiluminescent substrate) from the membrane.

4. Add 20 ml of alkaline probe-stripping solution to the roller bottle and rotate the bottle at 37°C at maximum rotation speed for 10 minutes. Discard the alkaline probe-stripping solution. Invert the bottle over a paper towel and let the remaining liquid drain for 1 minute.

5. Add 20 ml of two times SSC solution to the roller bottle and wash the membrane at room temperature rotating the bottle at maximum speed for 20 minutes. Discard the solution and repeat the washing procedure one more time.

Pre-hybridization

1. Add 10 ml of hybridization solution to a 15 ml plastic centrifuge tube and preheat it for 10 minutes in a 55°C water bath. Add the solution to the roller bottle.

2. Close the roller bottle tightly and place it in the hybridization oven. Rotate slowly at 55°C for 20 minutes.

3. Start preparing the probe for hybridization. Pour 10 ml of hybridization solution into a 15 ml conical centrifuge tube. Add 1 µl of probe to the tube and incubate it at 55°C in a water bath until it is ready to add to the roller bottle. **Note**: your instructor will give you the D2S44-labeled probe.

4. After pre-hybridization has been completed, remove the bottle from the oven and pour off the pre-hybridization solution.

5. Add prewarmed hybridization solution prepared in step 3 to the roller bottle. **Note**: it is very important that prewarmed hybridization solution is added to the bottle as fast as possible and the bottle returned to the oven immediately.
6. Return the bottle to a hybridization oven and allow it to **rotate slowly** at 55°C for 20 minutes. Remember that the hybridization temperature is critical for specific hybridization and it must be 55°C for at least 20 minutes.
7. Take your bottle from the hybridization oven and pour off the hybridization solution into a 15 ml centrifuge tube. Return the tube with the probe to your instructor. This probe can be reused several times.
8. Drain liquid from the bottle by inverting it over a paper towel for 1–2 minutes.

Washing reaction

1. Add 20 ml of wash buffer 1 to the tube. Place the tube into the hybridization oven and allow it to rotate slowly for 10 minutes at 55°C. Repeat the washing procedure one more time.
2. Remove the roller bottle from the oven and pour off wash buffer 1. Drain the solution well by inverting the bottle over a paper towel for 1–2 minutes.

Signal detection

1. Add 20 ml of buffer C (detection buffer) to the roller bottle. Close the bottle and move the membrane towards the tube opening by gently shaking the tube.
2. Open the bottle and remove the membrane to a Pyrex dish filled with 200 ml of buffer C (detection buffer). Wash the membrane for 20 minutes on a rotary shaker.
3. Place a plastic bag on a sheet of Whatman 3MM paper and add 10 ml of buffer C. Wearing gloves transfer the membrane from the Pyrex dish into the plastic bag. Open the bag and insert the membrane into it. Place the membrane into a pool of buffer and move it with your fingers to the end of the bag. Leave as little space as possible between the membrane and the end of the bag. This will limit the amount of expensive chemiluminescent substrate needed.
4. Pour off buffer from the bag. Remove the remaining liquid by gently pressing it out with a Kimwipe tissue. **Note**: do not press strongly on the membrane because this will increase the background. Most of the liquid should be removed from the bag leaving the membrane slightly damp. At this point a very small amount of liquid will be visible at the edge of the membrane.
5. Open the end of the bag slightly, leaving the membrane side that does not contain DNA attached to the side of the bag. Add 0.9–1 ml of CDP-Star

solution directing the stream towards the side of the bag. Do not add solution directly onto the membrane.

6. Place the bag on a sheet of Whatman 3MM paper with the DNA side up and distribute the liquid over the surface of the membrane by gently moving the liquid around with a Kimwipe tissue. Make sure that the entire membrane is evenly covered. Do not press on the membrane because this will cause "press marks" on the film. Gently remove excess CDP-Star from the bag by guiding excess solution towards the open end of the bag and onto the Whatman paper with a Kimwipe tissue. Make sure that the membrane remains damp. **Note**: at this point, small liquid droplets will be visible on the edge of the membrane, but liquid should not be present on the membrane surface. Seal the bag with a heat sealer.

7. Place the bag in an X-ray film cassette DNA side up. In a darkroom, place X-ray film over the membrane. Expose the film for 5–10 minutes at room temperature. Open the cassette and develop the film using standard procedures for film development. **Note**: maximum light emission for CDP-Star is reached in 20–30 minutes, the light emission remains constant for approximately 24 hours, and the blot can be exposed to the film a number of times during this period. The best results are usually obtained when the membrane is exposed the next day.

Protocol

To analyze your data, follow the protocol and do the calculations as shown in the example shown in the section on expected results.

1. Align the X-ray film with the gel drawing made previously. The outline of the filter should be visible on the film. Measure the electrophoretic mobility of each band.

2. Prepare a semi-log graph of the standard DNA fragments. Use the plot of electrophoretic mobility of the known size fragments as a function of the log of fragment size in base pairs. The instructor will give you the sizes of the standard DNA fragments.

3. Measure the electrophoretic mobility of the DNA fragments that hybridize to the probe. Determine the sizes of these fragments using the graph prepared in step 2.

4. Using DNA band size, find to which bin(s) your allele belongs. Use Table 3.1 for this determination. Read the frequency of the bin to which each of your bands belongs from Table 3.1.

5. Prepare the uncertainty windows table for all bands. Next prepare a match windows table for your bands and calculate the probability that a randomly chosen person from the population will have the same profile as you. For all these calculations use the example described in the section on expected results.

Data analysis

This data analysis will simulate the single-locus analysis used in forensic cases. The DNA of control 1 will be taken as victim DNA. The DNA of the second control will be treated as DNA found at the crime scene. Both these DNAs are DNA of **evidentiary samples**. Your DNA and the DNA of your partner will be the DNA of two suspects.

Your task will be to calculate the statistical probability that the DNA profile of any one person will contain all the alleles represented by the bands of the evidentiary samples. Thus, the question addressed by this DNA analysis is what is the probability that a person chosen at random from the relevant population will likewise have a DNA profile matching that of the evidentiary sample? You will therefore determine what is the probability that your DNA single-locus profile is matching the "evidentiary sample," i.e. are you "guilty" or not of this imaginary crime.

The first step of this analysis is to compare the DNA patterns on your autogram in order to determine whether your DNA pattern matches the DNA pattern of the evidentiary samples. The patterns are visually

Table 3.1 Allelic frequencies of the D2S44 VNTR locus for DNA cut with *Hae*III enzyme

Bin number	Size range (bp)	*N*	Frequency
3	0–871	8	0.005
4	872–963	5	0.003
5	964–1,077	24	0.015
6	1,078–1,196	38	0.024
7	1,197–1,352	73	0.046
8	1,353–1,507	55	0.035
9	1,508–1,637	197	0.124
10	1,638–1,788	170	0.107
11	1,789–1,924	131	0.083
12	1,925–2,088	79	0.050
13	2,089–2,351	131	0.013
14	2,352–2,522	60	0.038
15	2,523–2,692	65	0.041
16	2,693–2,862	63	0.040
17	2,863–3,033	136	0.086
18	3,034–3,329	141	0.089
19	3,330–3,674	119	0.075
20	3,675–3,979	36	0.023
21	3,980–4,322	27	0.017
22	4,324–5,685	13	0.008
25	5,686–	13	0.008

evaluated in order to determine a likely match. At this step a suspect may be excluded from further consideration because the pattern will be noticeably different from the evidentiary samples. If a suspect pattern is close to the evidentiary samples (even for one band), this is considered a "likely match."

If a suspect's bands cannot be excluded, they are subjected to analysis in order to determine the length of the represented DNA fragments measured in base pairs. These measurements are taken by comparing the bands for the sample fragments with the bands for the size marker fragments of known base pair lengths.

The most straightforward means of making a "likely match" calculation is to apply the product rule to **allele frequencies** derived from major population groups. The essence of the product rule is multiplication of individual band probabilities in order to arrive at an overall probability, statistically expressed as a simple fraction, that one out of a stated number of persons in the population (e.g. one out of 100,000) will match the DNA profile of the evidentiary sample in question.

In court procedures, at least six to seven independent single-locus alleles are analyzed. Thus, the rule is applied in two stages: first, to determine the

allelic frequency at each locus and then to determine the alleles' combined frequency at all loci.

Measurement of band sizes using gel electrophoresis is not exact because the resolving power of the gel is limited. This results in an inability to separate alleles that are close in size. For comparison purposes therefore the database bands are sorted into ranges of size called "bins." There are two kinds of bins: "floating bins" and "fixed bins." Forensic analysis in the USA uses a fixed bins database of allele frequencies prepared by the FBI.

The fixed bins database contains a catalog of the entire spectrum of VNTR base pair sizes likely to appear as bands. A separate fixed bin table has been compiled for each locus. Each band is recorded within the bin that encompasses its base pair size. To protect a suspect against unduly small **frequencies**, any bin with four or fewer bands is combined with its neighbor until each bin contains a minimum of five bands.

Table 3.1 presents the bin size and allelic frequencies of the D2S44 locus for *Hae*III-restricted human genomic DNA. The fixed bin table shows not only each bin's range of sizes and number of bands (*N*), but also each bin's frequency, which is calculated from the ratio of the number of bands in the bin to the total number of bands in the table. The FBI laboratory prepared this table.

In fixed bin analysis the frequency of an evidentiary band is determined by assigning it the frequency of the fixed bin into which its base pair size falls. In order to calculate the frequency (probability) of occurrence of a particular allele arrangement for a heterozygote (two bands), the **frequencies** of those bands are multiplied by each other and the result multiplied by 2.0 ($2pq$). This is because the frequency of heterozygotes in a Hardy–Weinberg (random) population is equal to $2pq$, where p is the frequency of one allele in question and q is the frequency of the second allele. If there is only one band at the locus, either the donor is homozygous or there is a second **allele** that for some reason did not appear. In order to take both possibilities into account while avoiding prejudice to the suspect, the frequency of the first band is simply multiplied by 2.0.

The next step is to calculate the probability that a randomly chosen person has the same profile as the evidence sample. For this the "product rule" is applied. The frequencies of matched large and small bands for each of the probes are multiplied by each other and multiplied by 2.0 in order to determine the frequency of heterozygotes ($2pq$). This value is calculated for each probe and multiplied by each other. For example, let us assume that the matched frequency for probe 1 is 0.017, for probe 2 is 0.26 and 0.015, for probe 3 is 0.072 and 0.013, for probe 4 is 0.089, for probe 5 is 0.017, and for probe 6 is 0.089. Thus, the overall probability is equal to $2(0.017) \times 2(0.26)(0.015) \times 2(0.072)(0.013) \times 2(0.089) \times 2(0.017) \times 2(0.089) = 1.6 \times 10^{-11}$ or one in 60 billon. Thus, the probability of a random match with the evidentiary samples other than the suspect is very small because there are only 5 billion people living on earth.

Expected results

Figure 3.3 presents the autogram of single-locus analysis similar to your autogram. S1 and S2 are two suspects (your DNA and your partner's DNA) and E1 and E2 are evidentiary samples (control DNA 1 and 2).

The results presented in Fig. 3.3 indicate that both "suspects" cannot be excluded and represent a "match" because the bands in S1 and S2 are located very close to at least one of the bands in the evidentiary samples. The next step of analysis is to determine the sizes of all of the bands for the S1, S2, E1, and E2 data. The results are recorded in the size column of the uncertainty windows table (Table 3.2).

Next, the upper and lower limits of the uncertainty windows for each band are calculated. The size of these windows spans the range 2.5 percent below to 2.5 percent above the measured value. To calculate this value, the band size is multiplied by 0.025 and recorded in the column named 2.5 percent in Table 3.2. This value is added and subtracted to the corresponding band size and both values are recorded in the column named uncertainty window.

Subsequently, a comparison is made between the uncertainty windows for each suspect band and those of the evidentiary bands. For example, comparing the uncertainty window for a large band of S1 with windows for E1 shows that only one window of the E1 band (band 2) overlaps. Comparing the uncertainty window for the small S1 band with E1 windows shows that none of the windows overlap. Furthermore, a comparison of both S1 bands with the band of E2 also indicates no overlap.

This analysis is continued for suspect 2 (S2). In this example there are two overlaps for this subject. The uncertainty widow for the small band of S2 overlaps windows for band 4 of E2 and the small band of E2 (see Table 3.2).

The next step is to calculate the size of the match window that will be used for finding the frequency of these markers in the database for the allele frequency of D2S44 presented in Table 3.1. For this analysis, match windows are calculated with 5 percent error. These calculations are performed for each band of the suspect that overlaps bands of the evidentiary samples. The results are recorded in a table. In the above example, only two bands are analyzed. These are overlap of the large band of S1 with band 2 of E1 (4,549) and the overlap of the small band of S2 with two bands of S2 (small band and band 4). Table 3.3 shows the results of this calculation for the example given.

In order to calculate the "match window" the size of the large band of S1 (4,549) is multiplied by 0.05 and recorded in the 5 percent column of Table 3.3. The "match window" is calculated by the addition and subtraction of the 5 percent value to and from the size of the large band. The results are recorded in the match window column. Match window values are used for finding the bin number in Table 3.1 that overlaps the match window value.

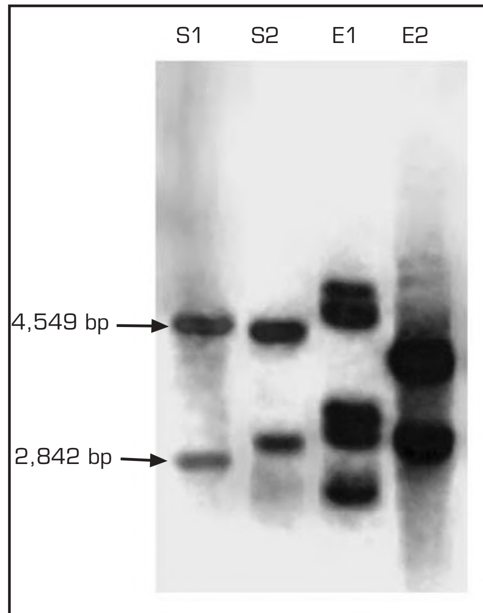

Figure 3.3 Autogram of single-locus analysis. Autogram of the DNA of two suspects and two evidentiary samples. S1, suspect 1; S2, suspect 2; E1, evidentiary DNA 1; E2, evidentiary DNA 2.

In the example, two bins overlap the value of the match window, i.e. bins 21 and 22. This is recorded in the bin(s) column. The frequencies for these bins are 0.017 and 0.008, respectively (see Table 3.1). Only the larger of these two frequencies is recorded in the table. The results of this analysis for the S2 suspect are also shown in Table 3.3.

In forensic analysis, the probe will be "stripped" from the membrane and then hybridized with the next single-locus probe. The results are calculated for the match with this probe. This process is repeated with at least six other probes.

In our example, we analyzed the result of hybridization with only one probe. Thus, the probability of a match (or non-match) is high. For S1 individuals, only one band matched the evidentiary sample. To calculate this probability, we multiply the frequency for a single band by 2.0. Thus, the probability is $2 \times 0.017 = 0.034$. This means that if we analyze DNA from 1,000 persons from any population, 34 of them will show a match with this single evidentiary band. This certainly does not exclude S1 as the suspect in this crime. However, analysis of six more single-locus probes, as required by courts, might prove S1 innocent.

For the S2 individual, this probability is $2 \times 0.089 = 0.178$ meaning that out of 1,000 persons from any population, 178 of them will show a match.

Table 3.2 The uncertainty windows data

Source	Band	Size (bp)	2.5% error	Uncertainty window
S1	Large	4,549	114	4,435–4,663
	Small	2,840	71	2,911–2,769
S2	Large	4,250	106	4,356–4,144
	Small	3,000	75	3,075–2,925
	Band 1	5,258	131	5,389–5,127
	Band 2	4,690	117	4,807–4,573
E1	Band 3	3,246	81	3,327–3,165
	Band 4	3,040	76	3,116–2,964
	Band 5	2,461	61	2,522–2,400
E2	Large	4,000	100	4,100–3,900
	Small	3,082	77	3,159–3,005

Table 3.3 Match windows and frequencies of D2S44 alleles

Source	Band	Size (bp)	5%	Match Window	Bin(s)	Frequency
S1	Large	4,549	227	4,322–4,776	21 and 22	0.017
S2	Small	3,000	150	3,150–2,850	16, 17, 18	0.089

Excerpt from the court procedures of the Simpson Trial D2S44 Probe

Mr Harmon: Okay. Mr Sims, yesterday you had described A16, which is the autorad for the probe D1S7 and today we have A17 which is the autorad for **D2S44** for the membrane AM626 and could you just describe again the samples that are up there and we will go through them and I will have you mark them as best you can.

Mr Sims: Okay. Just to reorient you again, in lane 1 this is one of the ladders, the size standards that we mentioned.

Mr Harmon: Could you get a different color for that. That will not show up really well.

Mr Sims: Is that better?

Mr Harmon: Sure.

Mr Sims: Again, this is lane 1. This is the size standard. Lane 2 is the K562. That is the national standard that we use. Lane 3 is the quality control, the blind sample. Lane 4 is another size standard. Lane 5 and the next two lanes now I will talk about, these are the reference bloodstains from Nicole Brown and then next we have the reference bloodstain from Ronald Goldman.

Mr Harmon: Okay. And is it easy to distinguish between Miss Brown and Mr Goldman at their – by their reference sample?

Mr Sims:	Yes, it is.
Mr Harmon:	You do not need a computer to do that?
Mr Sims:	No, you do not.

Day 3

Mr Harmon:	Could you – just the way we approached the other statistics, could you describe from among the three major population groups that you have used in your calculations the frequency estimate for those three groups and the match between Nicole Brown's reference blood and Greg Matheson's cut-out stain on the sock?
Mr Sims:	Yes. For those – the six loci that I looked at, these are RFLP loci D1S7, **D2S44**, D4S139, D5S110, D10S28, and D17S79, the profile detected in stain 42-A(1) occurs in approximately 1 in 21 billion Caucasians, 1 in 41 billion African Americans and 1 in 7.7 billion Hispanics, again indicating that this profile is a rare event and pointing out that these are for unrelated individuals.

References

Budowle, B., Smith, J., Moretti, T., and DiZinno, J. (2000) *DNA Typing Protocols: Molecular Biology and Forensic Analysis*. BioTechniques Book Publication, Eaton Publishing, Natick, MA.

Evertsson, U. (1999) The minisatellite D2S44 in chimpanzee has a different tandem repeated unit localized 200 bases upstream of the human repeat array. *Student J. Health Sci.*, **2**, 13.

Holmlund, G. (1999) The polymorphism of the minisatellite system D2S44. Medical dissertation, Linkoping University.

Holmlund, G. and Lindblom, B. (1995) Flanking region sequences and internal repeat structure of the pYNH24 (D2S44) 2 kbp insert analyzed by polymerase chain reaction and partial digestion with *Rsa*I. *Electrophoresis*, **16**, 1881–5.

Nakamura, Y., Gillian, S., O'Connell, P., Leppert, M., Lathrop, G., Lalouel, J.-M. et al. (1987) Isolation and mapping of a polymorphic DNA sequence pYNH24 on chromosome 2 (D2S44). *Nucleic Acids Res.*, **15**, 10073.

Surzycki, S. (2000) *Basic Techniques in Molecular Biology*. Springer-Verlag, Berlin, Heidelberg, and New York.

General reading

(1998) *The Evaluation of Forensic DNA Evidence. Committee of DNA Forensic Science: An Update*. National Academic Press, Washington, DC.