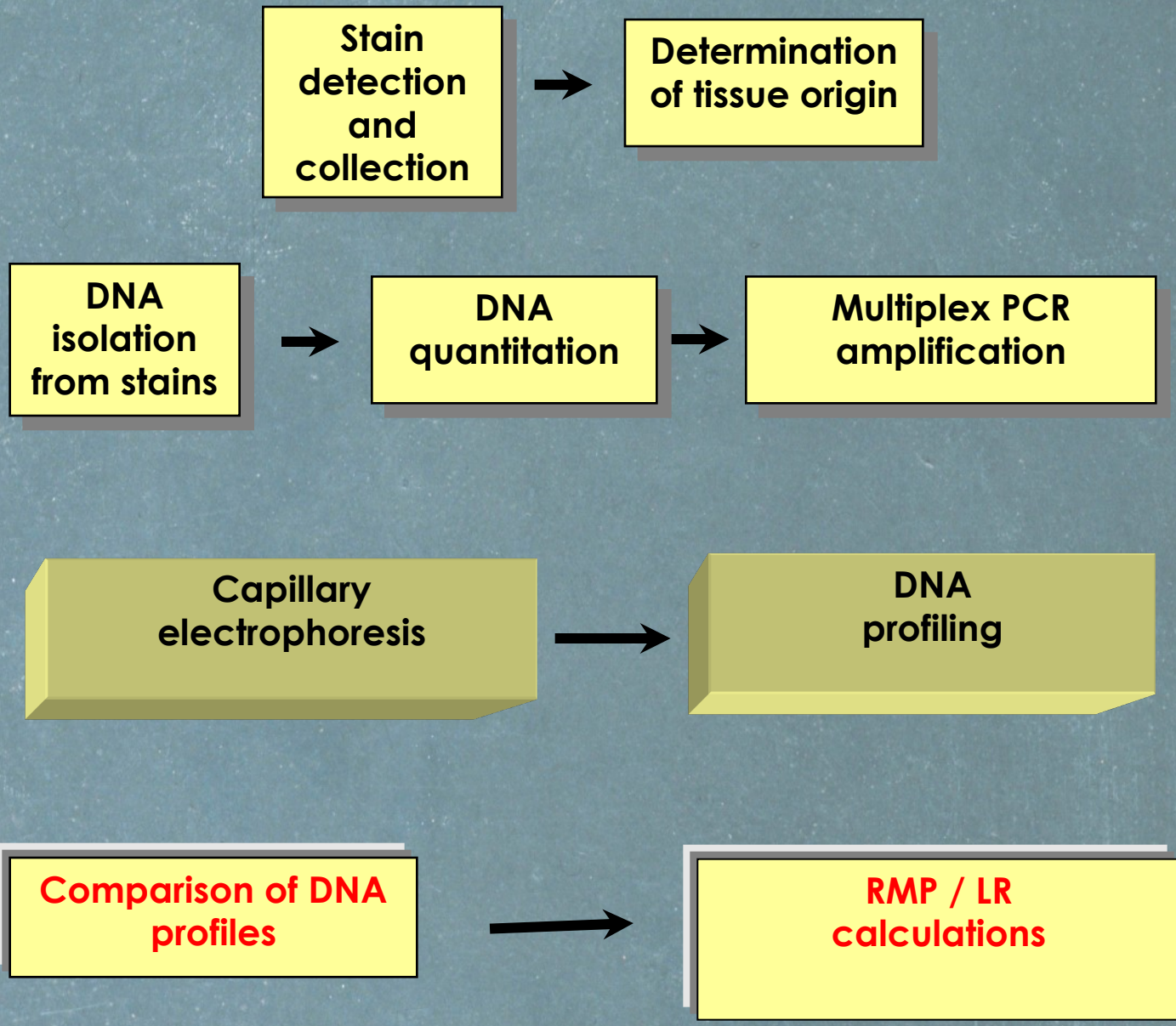
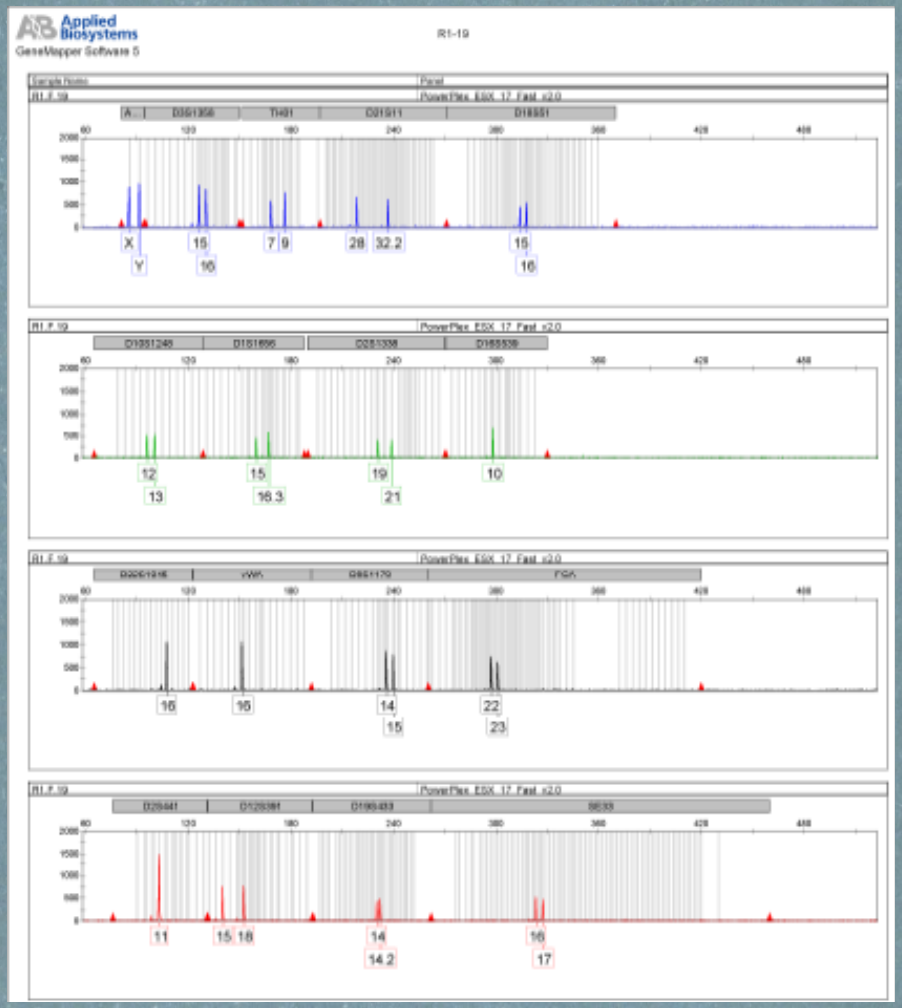
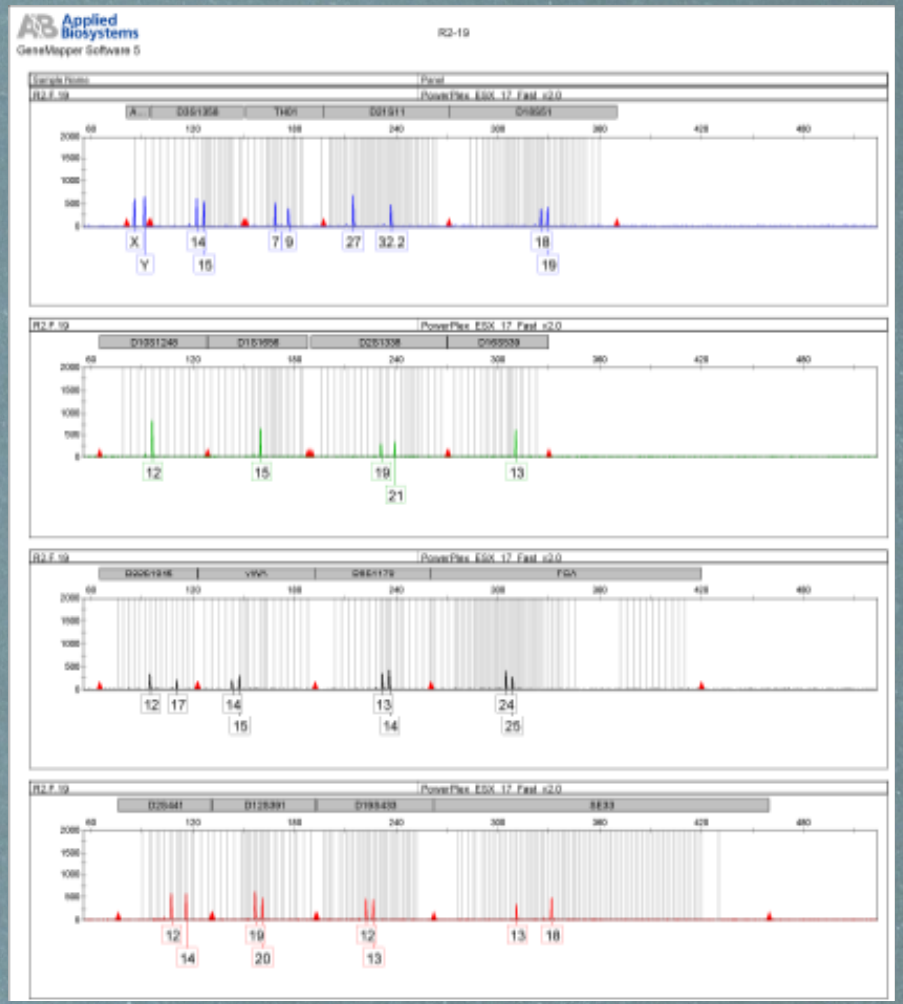

Stain detection and collection
Determination of tissue origin
DNA isolation from stains
DNA quantitation
Multiplex PCR amplification
Capillary electrophoresis
DNA profiling
Comparison of DNA profiles
RMP / LR calculations

✓ the combination of genotypes of a stain (or of an individual) for a particular set of STRs is called a **DNA profile**

✓ after CE, epgs of relevant samples and their DNA profiles are compared in order to draw conclusions

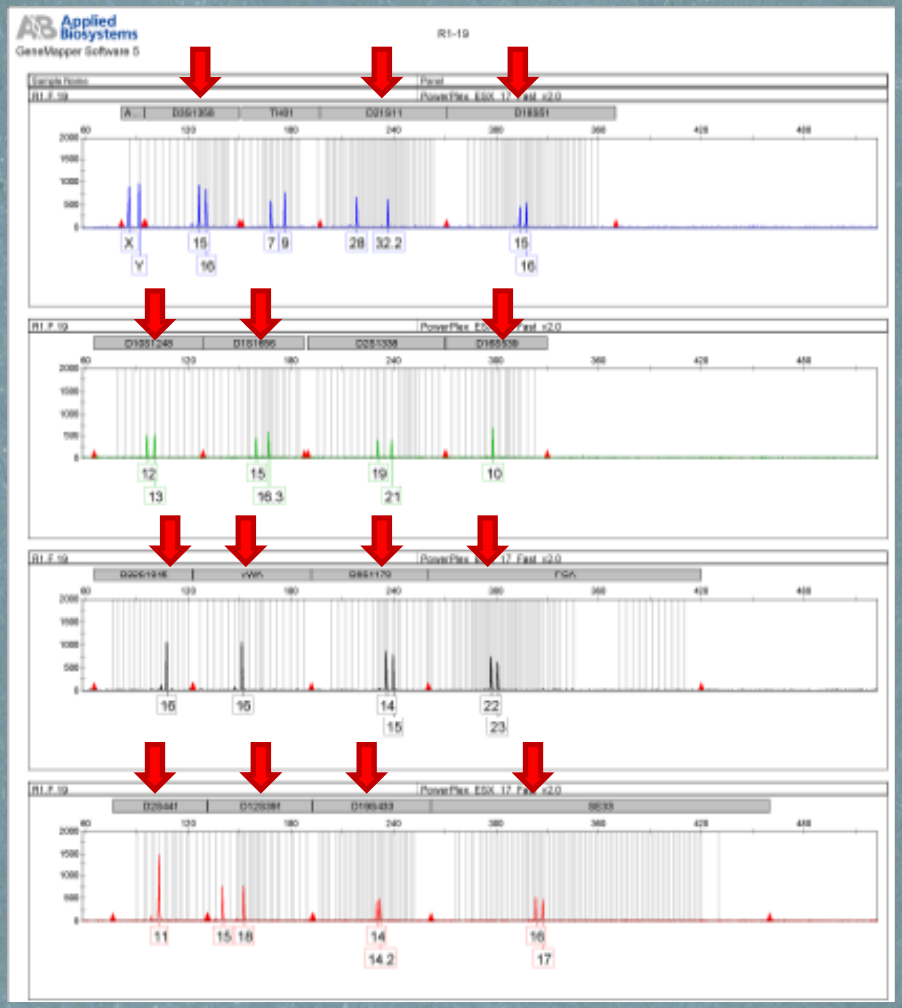


Suspect

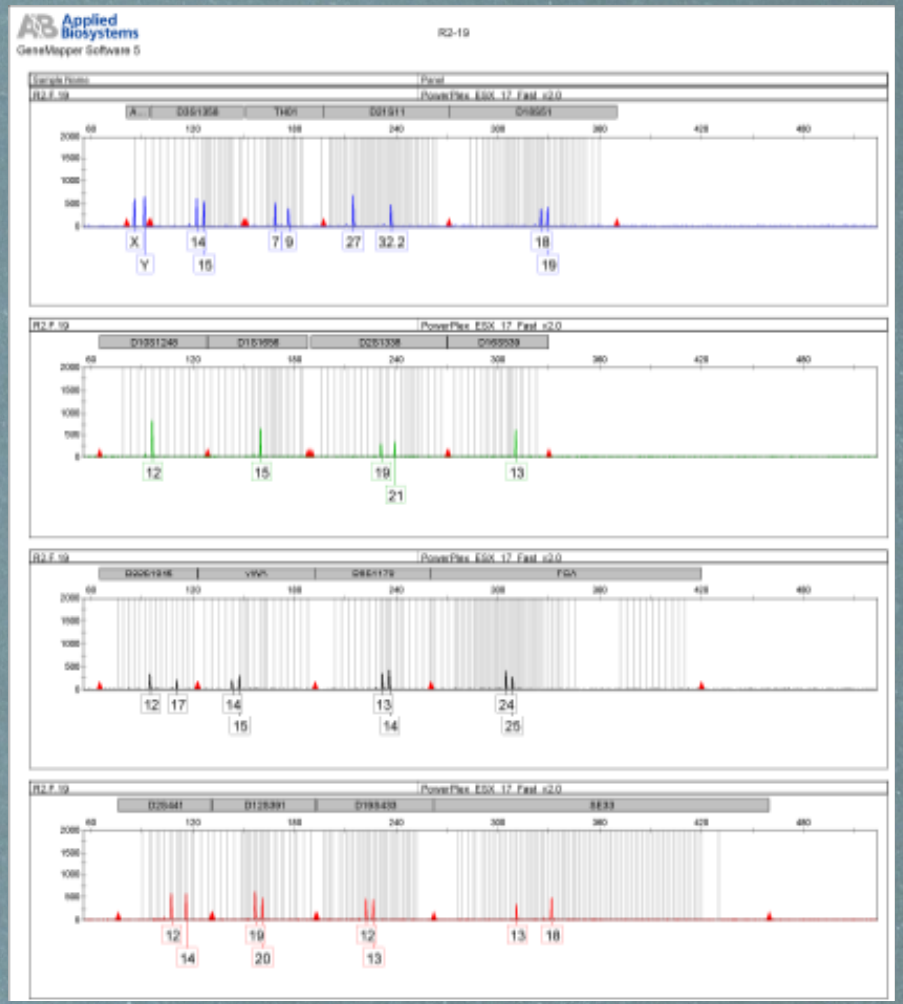


Stain

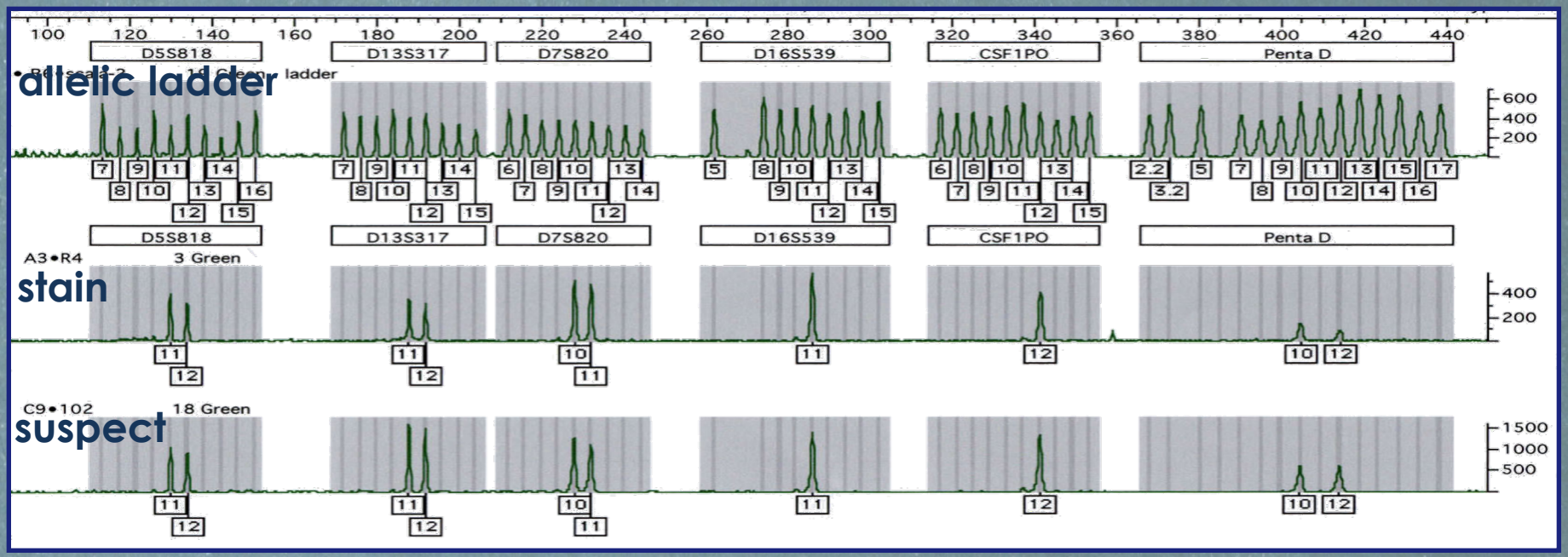
«It can be excluded that the stain originated from the suspect»

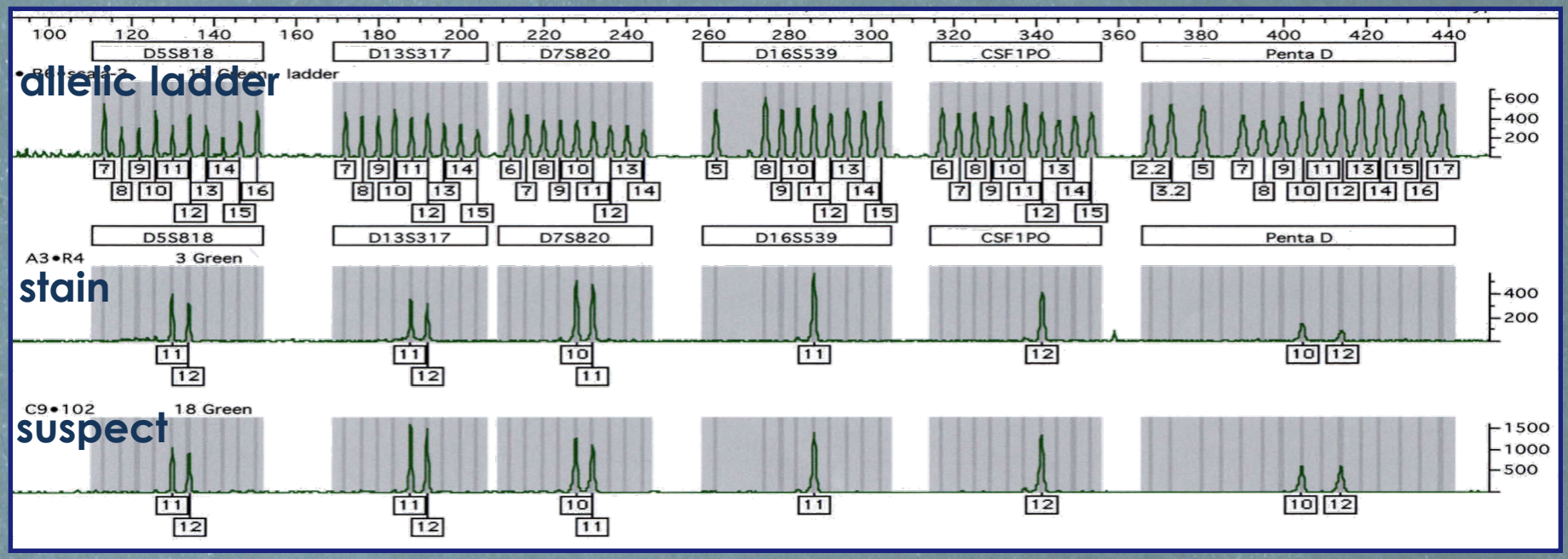


Suspect



Stain

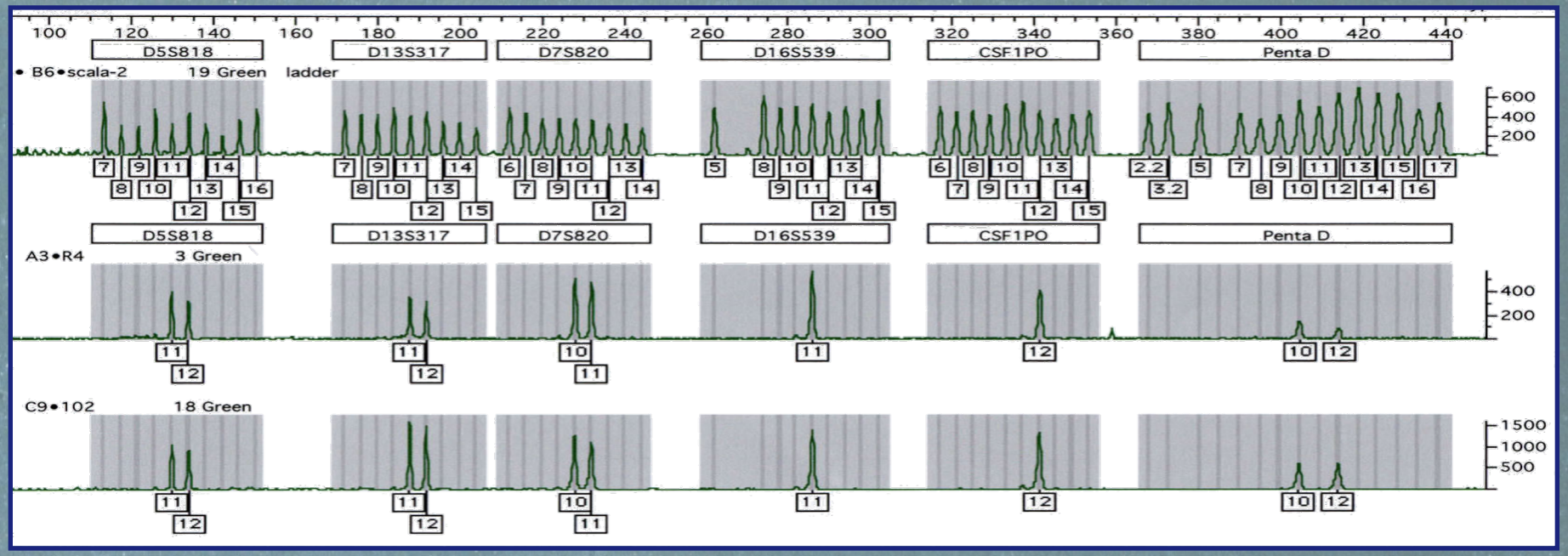

allelic ladder
stain
suspect
D5S818
D13S317
D7S820
D16S539
CSF1PO
Penta D
A3•R4
3 Green
C9•102
18 Green



"it cannot be excluded that the stain originated from the suspect"

↓

Need for mathematical interpratation of the DNA "match"



Random match probability (RMP)

Alleles are not infinite in the population.

What is the probability that the observed match is adventitious, i.e. the stain does not belong to the suspet but to another individual who, by chance, carries the same DNA profile of the suspect?

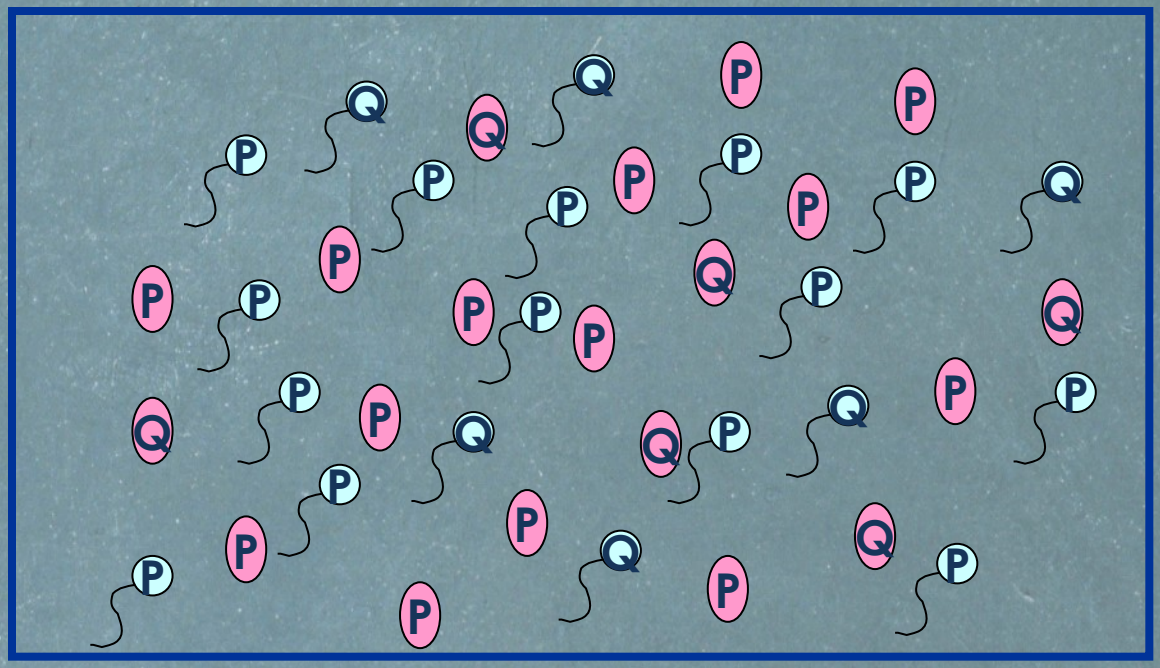
In order to calculate RMP we need to perform population genetics studies and establish reference databases of allele frequencies.

Let's imagine a very simple population study, done on 10 subjects who are genotyped for a single STR marker with only two possible alleles P and Q

	Genotype
Subject #1	PP
Subject #2	PQ
Subject #3	PQ
Subject #4	PP
Subject #5	PP
Subject #6	PQ
Subject #7	PP
Subject #8	QQ
Subject #9	PP
Subject #10	PQ

- ✓ Frequency of allele P = 70% (14/20)
- ✓ Frequency of allele Q = 30% (6/20)

Now let's imagine that population, not as a combination of subjects, but of gamete



If gametes combines randomly (as commonly in human populations…) then:

Frequency of homozygotes PP = P x P = 70% x 70% = 49% (P^2)

Frequency of homozygotes QQ = Q x Q = 30% x 30% = 9% (Q^2)

Frequency of heterozygotes PQ = P x Q + Q x P

= 70% x 30% + 30% x 70% = 42% (2PQ)

The law relating allele and genotype frequencies is called Hardy-Weinberg (HW) law.

HW law makes several assumptions:

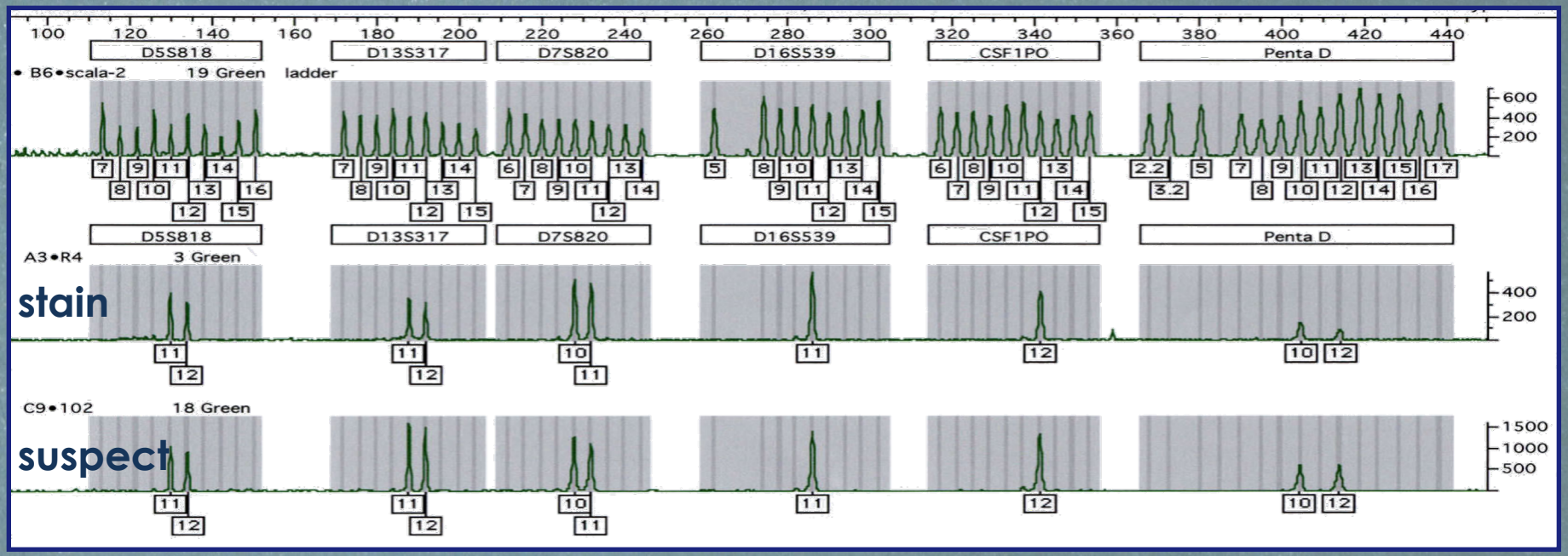
- random mating in the population
- no selection
- population sufficiently large that there is no random fluctuation of frequencies between generations
- no change in population structure across generations because of migration

However, STR allele and genotype frequencies observed in most natural human populations respect HW law, i.e. they are in «HW equilibrium» (HWE).

Deviations from HWE can be observed because of:

- inbreeding (excess of homozygous genotypes)
- alleles that are deleterious in the homozygous form but give a selective advantage in heterozygous form (excess of heterozygous genotypes)
- primer binding site mutation recurring with high frequency within a population (apparent excess of homozygous genotypes)

Provided we know the allele frequencies of the tested STRs in the population relevant to the case, we can now calculate the RMP for the DNA profile below



.258 x .187 x .111 x .102 x 109 x .055

RMP = 3 X 10^{-6} (about one subject out of 300,000, Italian population)

Product rule (joint probability of two independent events): expected genotype frequencies of each STR can be multiplied since CODIS/ESS STRs are independent from each other (on different chromosomes or distant enough on the same chromosome so that they are inherited in random association)

On average RMPs for variable numbers of tested STR loci are:

- ✓ 6 STR: 1×10^{-5} (one out of 100,000)
- ✓ 9 STR: 1×10^{-10} (one out of 10,000,000,000)
- ✓ ESS (12 STRs): 10^{-16} (one out of 1,000,000,000,000,000)
- ✓ CODIS (13 STRs): 10^{-16} (one out of 1,000,000,000,000,000)
- ✓ 15 STR: 1×10^{-17} (one out of 10,000,000,000,000,000)
- ✓ Expanded CODIS (20 STRs): 1×10^{-24} (one out of 1,000, 000,000,000,000,000,000,000,000)

✓ an alternative way to express the weight of the evidence in case of a DNA profile match is the **likelihood ratio (LR)** approach

✓ the LR is a fraction which compares the probability of the observing the genotypes given two opposite hypothesis

$$LR = \frac{\text{Probability to observe a DNA profile (matching that of the suspect), given that the stain comes from the suspect}}{\text{Probability to observe a DNA profile (matching that of the suspect), given that the stain does not come from the suspect}}$$

$$LR = \frac{1}{\text{Frequency of the DNA profile In the relevant population (RMP)}} \quad \text{e.g.} \quad \frac{1}{1 \times 10^{-18}} = 10^{18}$$

The LR approach becomes extremely useful, and is therefore preferred, when facing challanging samples, as degraded or mixed DNA profiles

TO BE CONTINUED...

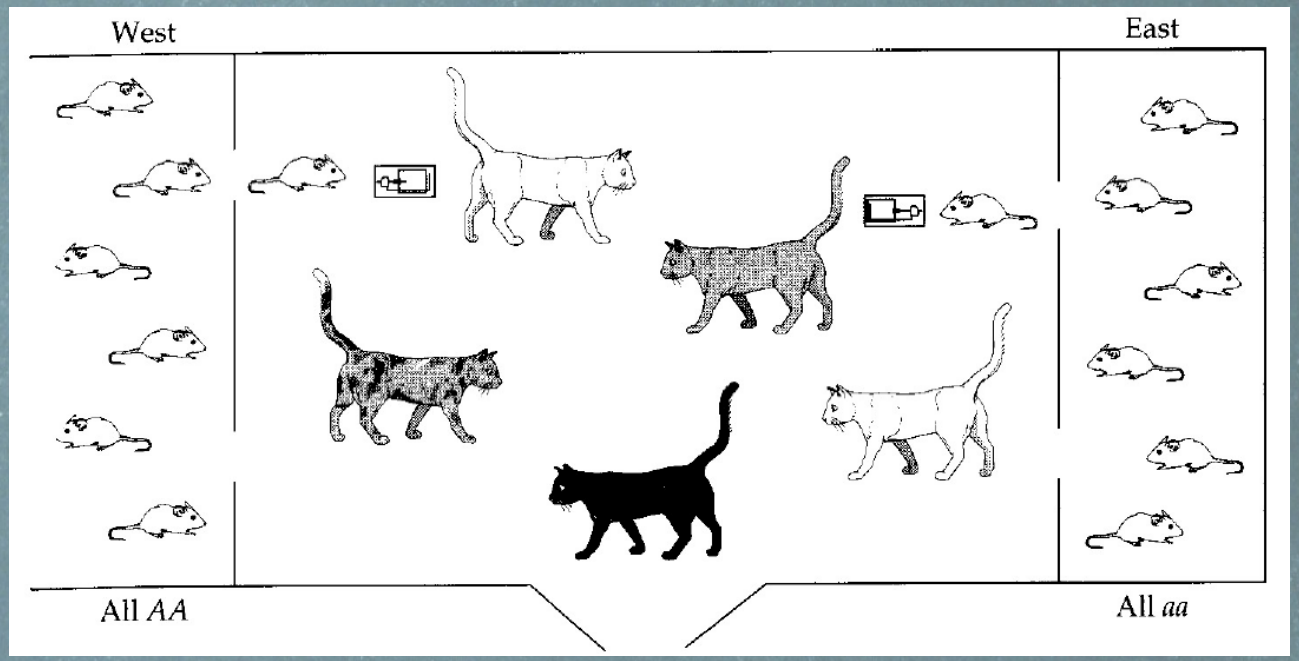
What is the relevant population for genotype frequency calculations?

- The relevant population is that of the offender, not that of the suspect
- However when a match is found between stain and suspect calculations are routinely made using allefe frequencies from the suspect's population (assuming that DNA profiles matching that of the suspect will have maximum possible frequency in the population he comes from)?
- An alternative approach is to replicate RMP/LR calculations using allele frequencies from different broad bio-geographic groups

But what if we do not know allele frequencies in the relevant population? or we suspect the presence of stratification/substructuring (subgroups carryinng significant genetic differences) in the broad bio-geographic reference sample?

Population genetic studies support the conclusion that, for forensic STRs, differences in allele frequencies between subgroups within broad bio-geographic areas (continents or subcontinents) are negligible, at least in cosmopolitan (not isolated) populations.

A correction (called fixation index «F_{ST}», or «θ») that takes account of population structure (allele sharing by common descent) can be applied to RMP/LR calculations.



$F_{ST} = (H_T - H_S) / H_T$

Can range from 0 (no differentiation) to 1 (total differentiation) as in the present case

(0.5 – 0) / 0.5

H_T: expected heterozygosity in random mating total population

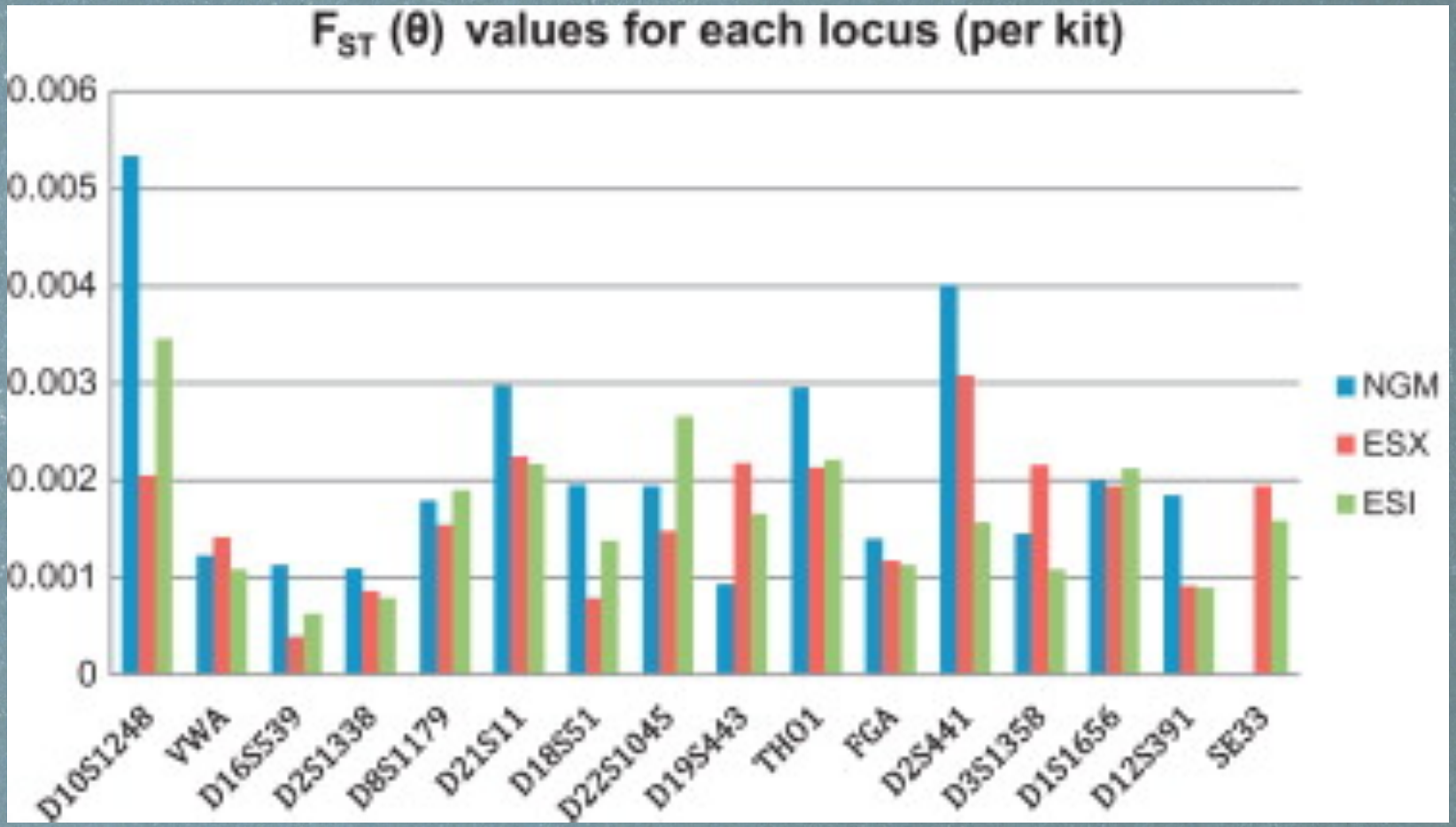
H_S: mean expected heterozygosity within random mating subpopulations

In humans, excluding highly isolated and inbred populations F_{ST} is normally < 1%

		Under HWE	Unconditional (NRC II recommendation 4.1)	Conditional with substructure adjustment (NRC II recommendation 4.10a) (NRC II recommendation 4.10b)
Homozygote	Formula	p^2	$p^2 + p(1-p)\theta$	$\Pr(PP\|PP) = \frac{[p(1-\theta)+2\theta]\,[p(1-\theta)+3\theta]}{(1+\theta)(1+2\theta)}$
D18S51 17,17 $p = 0.139$ $\theta = 0.01$	Calculation	$(0.139)^2$	$(0.139)^2 + (0.139)\times$ $(1\text{-}0.139) \times (0.01)$ $= 0.0193 + 0.00120$	$= \frac{[0.139(1-0.01)+2(0.01)][0.139(1-0.01)+3(0.01)]}{(1+0.01)(1+2(0.01))}$
	Result	$= \mathbf{0.0193}$	$= \mathbf{0.0205}$	$= \mathbf{0.0256}$
Heterozygote	Formula	$2pq$	$2pq$	$\Pr(PQ\|PQ) = \frac{2[p(1-\theta)+\theta]\,[q(1-\theta)+\theta]}{(1+\theta)(1+2\theta)}$
CSF1PO 11,12 $p = 0.309$ $q = 0.360$ $\theta = 0.01$	Calculation	$2(0.309)(0.360)$	$2(0.309)(0.360)$	$= \frac{2[0.309(1-0.01)+0.01][0.360(1-0.01)+0.01]}{(1+0.01)(1+2(0.01))}$
	Result	$= \mathbf{0.222}$	$= \mathbf{0.222}$	$= \mathbf{0.225}$

Corrects for excess of homozygous genotypes in populations due to common ancestry

Corrects for uncertainty of allele frequencies in subpopulation compared to reference general population



Welch et al. Forensic Sci Int Genet 2012

- 5700 European individuals from 24 European nations
- d = log10 (RMP calculated using national database / RMP calculated using combined european database)
- d > 1 is not conservative (RMP > in national rather than european database)
- each unit of d represents 1 order of magniture

Fst	d	S.D.	P(d > 0)	P(d > 1)	P(d > 2)	P(d > 3)	P(d > 4)	P(d > 5)
0.000	0.250	0.475	0.690	0.078	0.0026	0.0004	0.0002	0.0000
0.003	−0.001	0.441	0.443	0.027	0.0002	0.0000	0.0000	0.0000
0.005	−0.152	0.449	0.310	0.013	0.0000	0.0000	0.0000	0.0000
0.010	−0.473	0.483	0.145	0.002	0.0000	0.0000	0.0000	0.0000
0.020	−1.0.28	0.556	0.031	0.000	0.0000	0.0000	0.0000	0.0000
0.030	−1.514	0.622	0.004	0.000	0.0000	0.0000	0.0000	0.0000

Gill et al. Forensic Sci Int 2003

Mean d value calculated across all 5700 individuals assuming different F_{ST} values

Even without F_{ST} correction, less than 1% of samples display a 100-fold non conservative change in RMP

Same calculations with 5700 European individuals using national database and an Afro-Caribbean database

Fst	d	S.D.	$P(d > 0)$	$P(d > 1)$	$P(d > 2)$	$P(d > 3)$	$P(d > 4)$	$P(d > 5)$
0.000	1.240	1.06	0.89	0.594	0.2347	0.0456	0.0047	0.0000
0.010	0.443	1.01	0.68	0.291	0.0547	0.0035	0.0000	0.0000
0.017	0.0005	1.00	0.52	0.158	0.0172	0.0002	0.0000	0.0000
0.020	−0.192	1.00	0.44	0.112	0.0084	0.0000	0.0000	0.0000
0.030	−0.748	1.01	0.23	0.032	0.0004	0.0000	0.0000	0.0000
0.040	−1.249	1.01	0.10	0.007	0.0000	0.0000	0.0000	0.0000
0.050	−1.709	1.03	0.04	0.001	0.0000	0.0000	0.0000	0.0000
0.100	−3.604	1.09	0.00	0.000	0.0000	0.0000	0.0000	0.0000

Gill et al. Forensic Sci Int 2003

RMP calculations assume unrelated individuals, but the defense may argue that the stain was left not by the suspect but by a close relative: since relatives' DNA is partly identical by descent, a correction needs to be applied to RMP.

Table 21.6

Example calculations with corrections for relatives using the NRC II recommended formula.

From U.S. Caucasian (N = 302); Appendix II – sample in database

						Under HWE		*NRCII Recommendation 4.4*				
	A1	A2	Allele 1 freq (p)	Allele 2 freq (q)		Calc freq		F= 1/4 (parent)	F= 1/8 (half sib)	F= 1/16 (1st cousin)		Full sib
D13S317	11	14	0.33940	0.04801	2pq	0.0326	eq. 4.8b	0.1937	0.1131	0.0729	eq. 4.9b	0.3550
TH01	6	6	0.23179	—	p^2	0.0537	eq. 4.8a	0.2318	0.1428	0.0982	eq. 4.9a	0.3793
D16S539	9	11	0.11258	0.32119	2pq	0.0723	eq. 4.8b	0.2169	0.1446	0.1085	eq. 4.9b	0.3765
D18S51	14	16	0.13742	0.13907	2pq	0.0382	eq. 4.8b	0.1382	0.0882	0.0632	eq. 4.9b	0.3287
D21S11	28	30	0.15894	0.27815	2pq	0.0884	eq. 4.8b	0.2185	0.1535	0.1209	eq. 4.9b	0.3814
D3S1358	16	17	0.25331	0.21523	2pq	0.1090	eq. 4.8b	0.2343	0.1717	0.1403	eq. 4.9b	0.3944
D5S818	12	13	0.38411	0.14073	2pq	0.1081	eq. 4.8b	0.2624	0.1853	0.1467	eq. 4.9b	0.4082
D7S820	9	9	0.17715	—	p^2	0.0314	eq. 4.8a	0.1772	0.1043	0.0678	eq. 4.9a	0.3464
D8S1179	12	14	0.18543	0.16556	2pq	0.0614	eq. 4.8b	0.1755	0.1184	0.0899	eq. 4.9b	0.3531
CSF1PO	10	10	0.21689	—	p^2	0.0470	eq. 4.8a	0.2169	0.1320	0.0895	eq. 4.9a	0.3702
FGA	21	22	0.18543	0.21854	2pq	0.0810	eq. 4.8b	0.2020	0.1415	0.1113	eq. 4.9b	0.3713
TPOX	8	8	0.53477	—	p^2	0.2860	eq. 4.8a	0.5348	0.4104	0.3482	eq. 4.9a	0.5889
VWA	17	18	0.28146	0.20033	2pq	0.1128	eq. 4.8b	0.2409	0.1768	0.1448	eq. 4.9b	0.3986
AMEL	X	Y										
						1.20E–15		3.17E–09	1.68E–11	3.74E–13		4.04E–06 1 in 247 616