Chapter 1
Nature of Discrete Genetic Data

1. **Forensic science & indididual identification.**
2. **Evolution.**
3. **Search for genes affecting human disease or traits of economically importance in domestic plants and animals.**
4. **Characterizing features of DNA, amino acid and protein Sequence.**
5. **Prediction of protein structure from the amino acid sequence features.**

Mendelizing unit

Example of genetic data

Phenotypic data

Codomonant phenotypic data

Allozyme data

Protein sequence data

Restriction fragment data

Polymerase chain reaction: RAPD loci.

VNTR and STR loci

DNA sequence data

Microarray data

Continuous data?

Notation and terminology

1. **Population size** N **and sample size** n
2. **A genetic loci, A, either with two alleles,** A, a**, or with a series of alleles** A_{uu}
3. **Allele frequencies:** p_A, p_b, p_{AB}

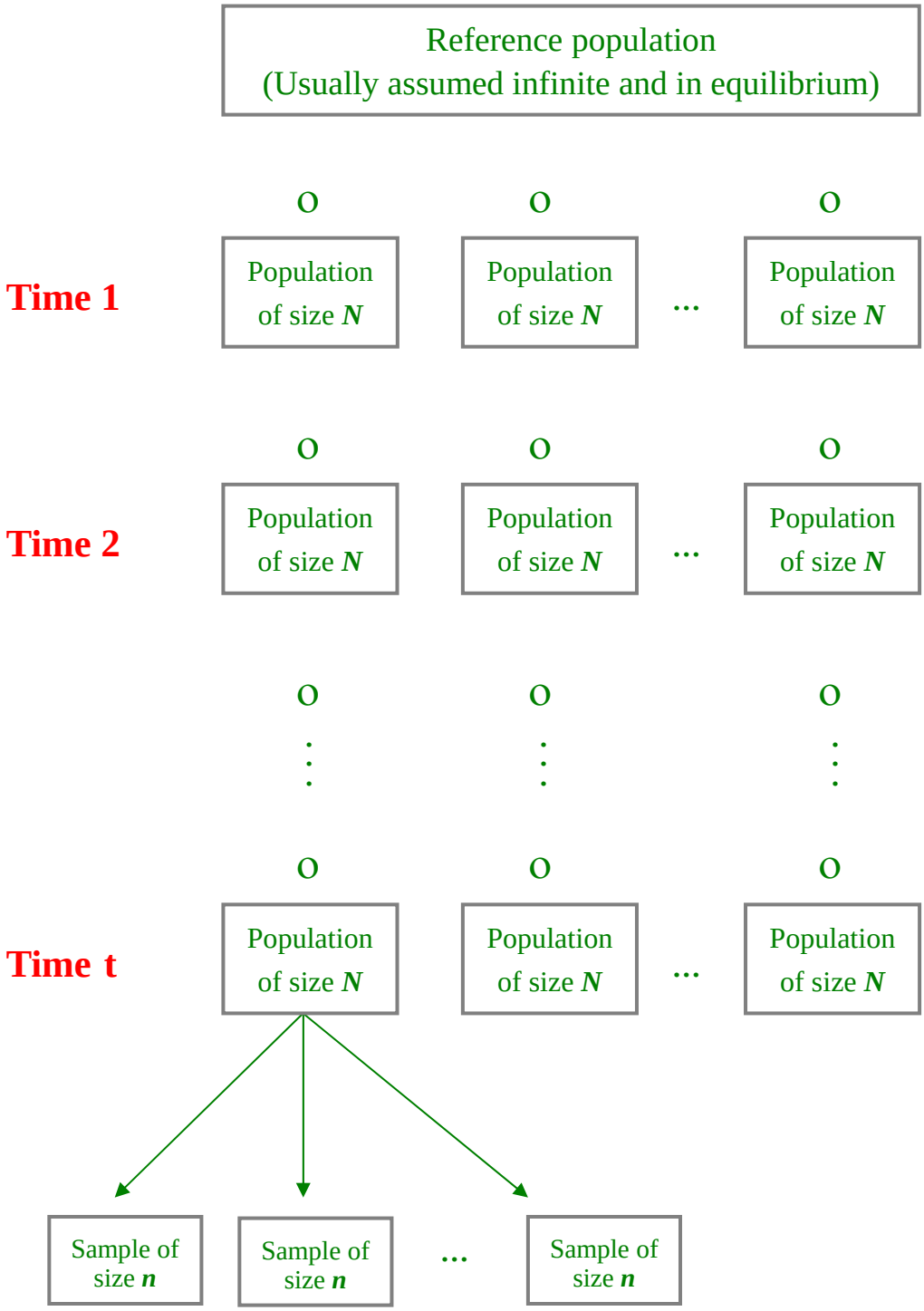
   Geneotypic frequencies: P_{AA}, P_{aa} **…**

   Sample counts: n_{AA}, n_A, **…**

   Sample frequencies: $\tilde{p}_A$, $\tilde{P}_{AA}$

   Estimated frequencies: $\hat{p}_A$, $\hat{P}_{AA}$

Genetic and statistical sampling



Experiment by Buri (1956)

Mendel and Fisher

Summary of the results of Mendel's experiments with seven traits of the garden pea.

PARENTAL TRAITS	F_1	F_2(NUMBER)		F_2(RATIO)
		DOMINANT	RECESSIVE	
(1) seed coat texture (round vs. wrinkled)	all round	5474 round	1850 wrinkled	2.96:1
(2) cotyledon color (yellow vs.green)	all yellow	6022 yellow	2001 green	3.01:1
(3) flower color (purple vs. white)	all purple	705 purple	224 white	3.15:1
(4) pod shape (inflated vs. constricted)	all inflated	882 inflated	299 constricted	2.95:1
(5) pod color (green vs. yellow)	all green	428 green	152 yellow	2.82:1
(6) flower location (axial vs. terminal)	all axial	651 axial	207 terminal	3.14:1
(7) sterm length (long vs. short)	all long	787 long	277 short	2.84:1

Mendel's Experiments for Garden Pea

Parental (P) Cross Phenotypes:	Dominant (e.g., long stems)	×	Recessive (e.g., short stems)
Genotypes:	AA		aa
Gametes:	A		a
F_1 generation Phenotypes:	Dominant (e.g., long stems)		
Genotypes:	Aa		
F_2 generation Phenotypes:	3/4 dominant	+	1/4 recessive
Genotypes:	(1 AA：2 Aa)		(aa)

F_3 generation Phenotypes:	all dominant	3/4 dominant	+	1/4 recessive	all recessive
Genotypes:	(AA)	(1 AA：2 Aa)		(aa)	(aa)

Homework:

Derive the genotypic frequencies of AA, Aa and aa in the F_3 and F_4 generation.

(1) consider the case of selfing

(2) consider the case of random mating (intercrossing)

(3) Explain how the gene frequencies and genotypic frequencies change

Table 1.1 Mendel's results for seven dominant characters in *Pisum sativum*.

	Character	Dominant Form		Recessive Form	
		Seed characters			
A	Seed shape	5474	Round	1850	Wrinkled
B	Cotyledon color	6022	Yellow	2001	Green
		Plant characters			
C	Seed coat color	705	Grey-brown	224	White
D	Pod shape	882	Simply inflated	299	Constricted
E	Unripe pod color	428	Green	152	Yellow
F	Flower position	651	Axial	207	Terminal
G	Stem length	787	Long	277	Short

Source: Mendel (1866).

Table 1.2 MNS blood group types for families reported by Race et al. (1949).

	Father	Mother	Child 1	Child 2	Child 3	Child 4	Child 5	Child 6
1	MSMs	MsMs	MsMs	MSMs	MSMs			
2	MM.S	MsMs	MSMs	MSMs				
3	MM.S	MM.S	MM.S	MM.S	MM.S			
4	MM.S	MM.S	MM.S					
5	MM.S	MM.S	MM.S					
6	MM.S	MM.S	MM.S	MM.S	MM.S			
7	MsMs	MsNs	MsNs	MsNs				
8	MsNs	MsMs	MsMs					
9	MsNs	MsMs	MsNs					
10	MsMs	MsNs	MsMs=	=MsMs				
11	MsMs	MsNs	MsNs	MsMs	MsNs	MsMs		
12	MsNs	MSMs	MSNs–	–MsMs				
13	MSMs	MsNs	MsNs	MSNs	MsMs			
14	MsNs	MSMs	MsMs	MSNs	MSMs			
15	MM.S	MsNs	MSMs	MSMs	MSNs	MSNs	MSNs	MSNs
16	MsNs	MM.S	MSMs	MSMs=	=MSMs			
17	MM.S	MsNs	MSMs	MSMs				
18	MsNs	MM.S	MSMs=	=MSMs				
19	MsNs	MM.S	MSNs					
20	MsNs	MM.S	MSMs	MSMs				
21	MsMs	MSNs	MSMs	MsNs	MSMs	MsNs		
22	MsMs	MSNs	MSMs	MSMs	MSMs–	–MSMs		
23	MSNs	MSMs	MM.S	MsNs				
24	MSNs	MSMs	MM.S	MsNs				
25	MsNS	MSMs	MsMs					
26	MM.S	MN.S	MM.S	MN.S	MN.S			
27	MM.S	MN.S	MM.S	MN.S				
28	MM.S	MN.S	MM.S	MM.S				
29	MN.S	MM.S	MN.S	MM.S				
30	MM.S	MN.S	MN.S					
31	MM.S	MN.S	MM.S	MN.S				
32	MM.S	MN.S	MN.S	MM.S				
33	MN.S	MM.S	MN.S–	–MM.S				
34	MM.S	MN.S	MM.S	MM.S				
35	MM.S	MN.S	MM.S	MM.S	MN.S			
36	NsNs	MsMs	MsNs	MsNs				
37	NsNs	MsMs	MsNs	MsNs	MsNs	MsNs	MsNs	MsNs
38	NsNs	MsMs	MsNs	MsNs	MsNs			
39	MsMs	NsNs	MsNs					
40	MSMs	NsNs	MsNs	MsNs				
41	MSMs	NsNs	MSNs	MsNs				
42	NsNs	MSMs	MSNs	MsNs	MSNs			
43	NsNs	MM.S	MSNs					
44	NN.S	MsMs	MsNS					
45	NsNs	MSMs	MsNs	MN.S				
46	MM.S	NN.S	MN.S					

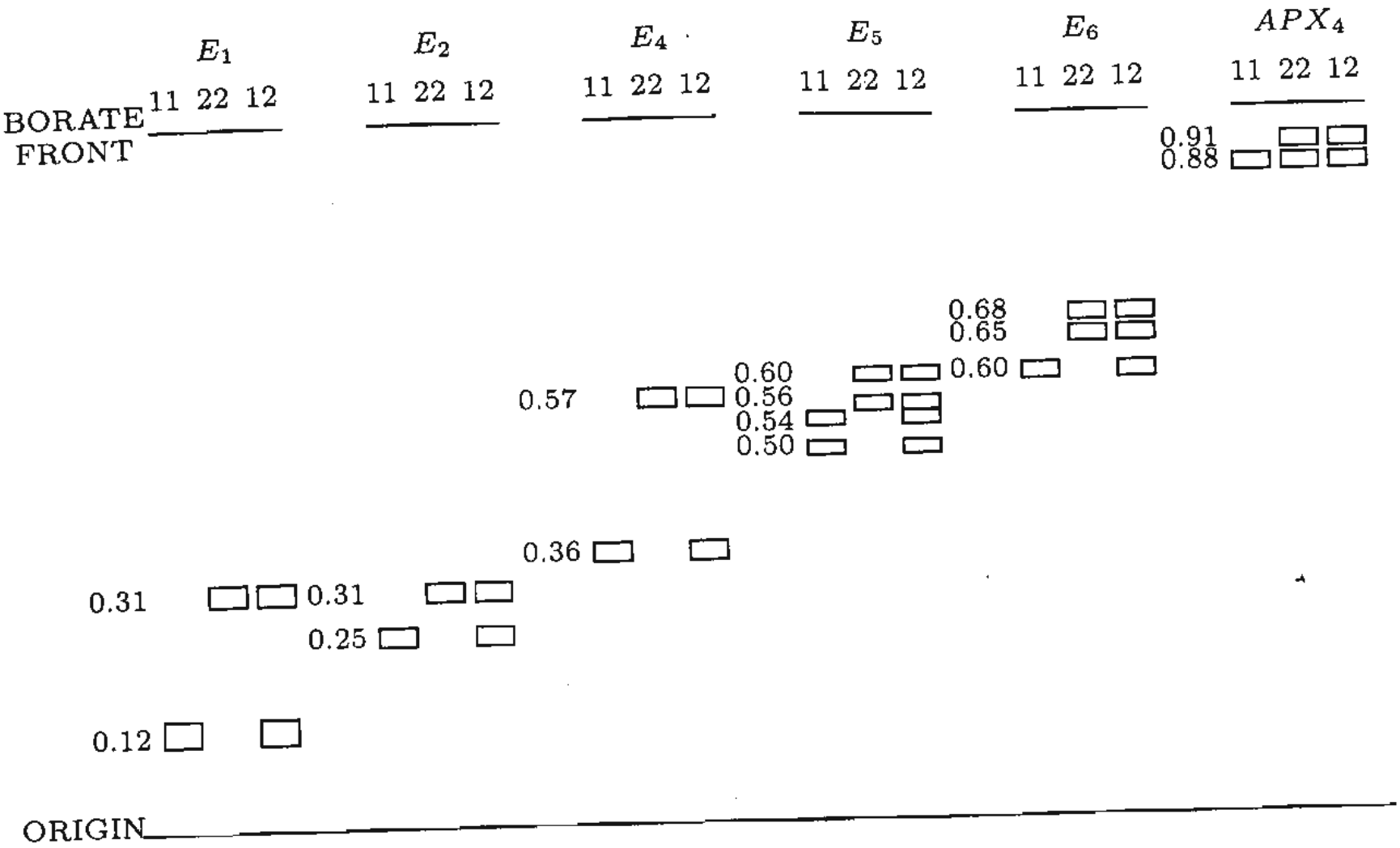


Figure 1.1 Electrophoretic banding pattern for six loci in wild oats. Source: Clegg and Allard (1973).

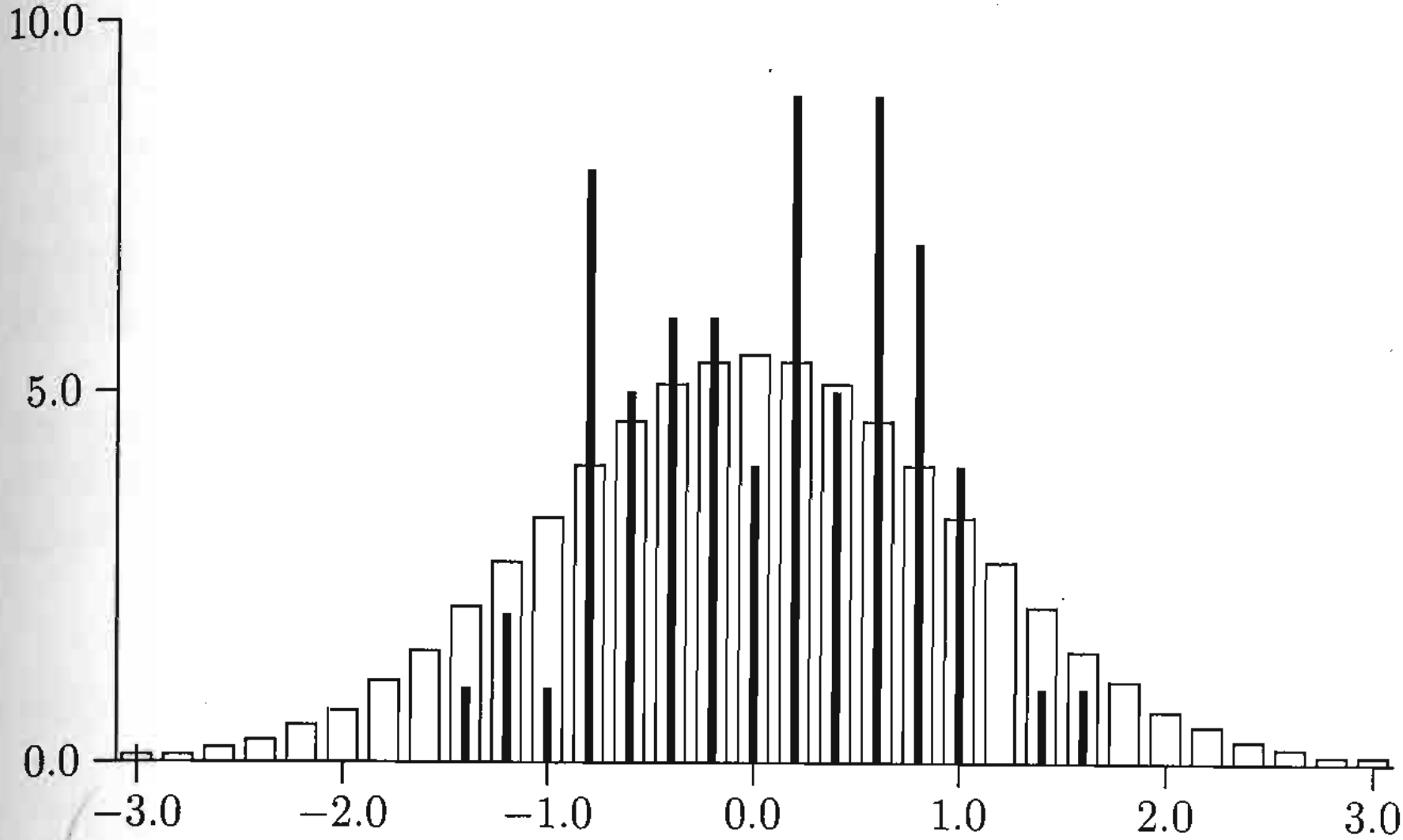


Figure 1.6 Histogram of X values for Mendel's data. Open bars are for a normal distribution, and solid bars are for Mendel's data. Source: Edwards (1986).

```
1   Tuna Y2      PRRRR--QASRPVRRRRRYRRSTAARRRRRVVRRRR
2   Tuna Z2      PRRRR--RSSRPVRRRRRYRRSTAARRRRRVVRRRR
3   Tuna Z1      PRRRR--RSSRPVRRRRRYRRSTAARRRRRVVRRRR
4   Salmon AII   PRRRRRRSSSRPIRRRR-YRRAS--RRRRRGGRRRR
5   Trout IB     PRRRRRRSSSRPIRRRR-PRRVS--RRRRRGGRRRR
6   Salmon AI    PRRRR--SSSRPVRRRRRPR-VSR-RRRRRGGRRRR
7   Trout IA     PRRRR--SSSRPVRRRRRPRRVSR-RRRRRGGRRRR
8   Trout II     PRRRR--SSSRPVRRRR-ARRVSR-RRRRRGGRRRR
9   Herring YII  PRRR-TRRASRPVRRRR-PRRVS--RRRR--ARRRR
10  Herring Z    ARRRRSRRASRPVRRRR-PRRVS--RRRR--ARRRR
11  Herring YI   ARRRRS--SSRPIRRRR-PRRRTT-RRRR-AGRRRR
12  Sturgeon B   ARRRRR--SSRPQRRRRR-RRHG--RRRR--GRR--
13  Sturgeon A   ARRRRRHASTKLKRRRRR-RRHQ--KK----SHK--
```

Figure 1.2 Alignment of fish protamine sequences. Source: Hunt and Dayhoff (1982).

Table 1.4 One-letter and three-letter codes for the 20 amino acids.

	Amino Acid	Codes			Amino Acid	Codes	
1	Alanine	Ala	A	11	Leucine	Leu	L
2	Arginine	Arg	R	12	Lysine	Lys	K
3	Asparagine	Asn	N	13	Methionine	Met	M
4	Aspartic acid	Asp	D	14	Phenylalanine	Phe	F
5	Cysteine	Cys	C	15	Proline	Pro	P
6	Glutamine	Gln	Q	16	Serine	Ser	S
7	Glutamic acid	Glu	E	17	Threonine	Thr	T
8	Glycine	Gly	G	18	Tryptophan	Trp	W
9	Histidine	His	H	19	Tyrosine	Tyr	Y
10	Isoleucine	Ile	I	20	Valine	Val	V

Table 1.5 Some common restriction enzymes and their recognition sequences.

Enzyme	Recognition Sequence
*Bam*HI	GGATCC CCTAGG
*Eco*RI	GAATTC CTTAAG
*Hind*III	AAGCTT TTCGAA
*Taq*I	TCGA AGCT
*Xho*I	CTCGAG GAGCTC

Table 1.6 Restriction map variants among *Drosophila melanogaster* lines.

		Restriction Sites				Insertions/Deletions*				
		*Bam*HI	*Hind*III	*Hind*III	*Xho*I	Δa	Δb	Δc	Δd	Δf
Line	*Adh*	−7.1†	−3.0†	+2.7†	+1.2†	20	550	900	180	30
R1	*S*	+	−	−	+	−	−	−	+	−
R2	*S*	+	−	−	+	−	−	−	+	−
M1	*S*	+	−	−	+	−	−	−	−	−
R3	*S*	+	−	−	+	−	−	−	−	−
N1	*S*	+	−	−	−	−	−	−	−	+
N2	*S*	+	−	−	−	−	−	−	−	+
R4	*S*	+	−	−	−	−	−	−	−	−
R5	*S*	+	−	−	−	−	−	−	−	−
R6	*S*	+	−	+	−	−	−	−	−	−
K1	*S*	+	−	+	−	−	−	+	−	−
K2	*S*	+	−	+	+	−	−	−	−	−
K3	*F*	−	−	−	+	−	−	−	−	−
K4	*F*	−	−	−	+	−	−	−	−	−
R8	*F*	−	+	−	+	−	−	−	−	−
M2	*F*	−	+	−	+	−	−	−	−	−
R9	*F*	−	−	−	+	+	−	−	−	−
R10	*F*	−	−	−	+	+	−	−	−	−

Source: Langley et al. (1982).

* a, d, f are deletions; b, c are insertions. Estimated sizes are given in numbers of nucleotides.

† These figures (thousands of nucleotides) indicate the positions of the restriction sites.

```
1   GTAAATATAG TTTAACCAAA ACATCAGATT GTGAATCTGA CAACAGAGGC TTACGACCCC TTATTTACC
2   GTAAATATAG TTTAACCAAA ACATCAGATT GTGAATCTGA CAACAGAGGC TCACGACCCC TTATTTACC
3   GTAAATATAG TTTAACCAAA ACATCAGATT GTGAATCTGA TAACAGAGGC TCACAACCCC TTATTTACC
4   GTAAATATAG TTTAACCAAA ACATTAGATT GTGAATCTAA TAATAGGGCC CCACAACCCC TTATTTACC
5   GTAAACATAG TTTAATCAAA ACATTAGATT GTGAATCTAA CAATAGAGGC TCGAAACCTC TTGCTTACC
```

Figure 1.3 Five mitochondrial DNA sequences. 1: human; 2: chimpanzee; 3: gorilla; 4: orangutan; 5: gibbon. Source: Brown et al. (1982).