

probes). A disadvantage of PCR is the required design and synthesis of specific primer pairs for each region to be amplified.

A **single-nucleotide polymorphism** (SNP) is variation in the identity of the base appearing at a particular position in the genome. There are an estimated $1.5 \times 10^6 - 10^7$ **common** polymorphic SNP sites in the human genome (Cargill et al., 1999; International SNP Map Working Group, 2001; **International HapMap Consortium**, 2005), corresponding to resolution at a level of hundreds of base pairs. Detection can employ either DNA sequencing or oligonucleotide array technologies, which means that SNP analysis can be readily automated and multiplexed. If detection is by hybridization, the underlying principle is that mismatched bases reduce the melting temperature of DNA-DNA hybrids. In the simplest example, four different oligonucleotide features, each having one of the four different bases at the polymorphic site, can be arrayed on a solid substrate. The state of the base at the polymorphic site is indicated by which of the four features hybridizes with the DNA being tested. An alternative detection method is primer extension. The underlying principle in this case is the ability of DNA polymerase to add the correct base to a growing chain during synthesis along a DNA template. The oligonucleotide probe sequences arrayed on a solid substrate act as primers, so designed that the next base to be added corresponds to the polymorphic site. The polymorphism can be detected by supplying dXTP molecules bearing distinctive fluorescent tags and detecting the wavelength of fluorescence emission after DNA synthesis. When the primers are not immobilized, other detection methods such as standard DNA sequencing gels or MALDI-TOF mass spectrometry may be employed.

13.5 Linkage Disequilibrium (LD)

Linkage disequilibrium (hereafter abbreviated **LD**) refers to the nonrandom association of alleles in haplotypes. It is measured by comparing the proportion of an observed haplotype with the proportion that would be predicted based upon the population frequencies of the alleles at each locus. In this section, we discuss a number of properties and consequences of LD.

13.5.1 Quantitative Description of LD

Linkage disequilibrium may be quantified as the difference between the observed and predicted frequencies for allele combinations at two or more loci. The predicted frequencies are computed using the population frequencies of the alleles. For a system in which each locus has two alleles, let $p_{A_1B_1}$ and $p_{A_2B_2}$ be the probabilities in the population of haplotypes A_1B_1 and A_2B_2 , respectively. The allele frequencies of the genes contributing to these haplotypes are p_{A_1} , p_{B_1} , p_{A_2} , and p_{B_2} . Note that the allele frequencies are population quantities, which are unaffected by recombination. We introduce a quantity