Linkage Maps

Maps that reflect the degree of linkage between loci can be constructed from observed recombination frequencies (see Centimorgan (cM)). The distances on these linkage maps will accurately reflect physical distances on the chromosome only if exchange frequencies are constant along the chromosome.

Linkage in Prokaryotes

In bacteria with a single chromosome, loci that are close enough together on the chromosome to be transmitted together in phage-mediated transduction may be referred to as 'linked.' Similarly, when transformation is conducted with chromosomal DNA, markers are 'linked' if they are cotransformed as a result of sometimes being on the same fragment created by artifactual breakage of the chromosome. Unlinked markers are transduced or transformed into the same recipient cell at a frequency about equal to the product of the transduction or transformation frequencies of the individual markers.

In crosses with bacteriophages, as standardly conducted, recombination frequencies less than 50% cannot be taken as evidence of linkage because a fraction of the progeny phage particles has lacked the opportunity to assort its genes. Linkage is implied by a pair of loci that gives a significantly lower recombination frequency than do the loci with the largest observed values.

See also: **Centimorgan (cM); Genetic Recombination; Mapping Function; Tetrad Analysis**

Linkage Disequilibrium

N E Morton

doi: 10.1006/rwgn.2001.0767

Dependence of gene frequencies at two or more loci is called allelic association, gametic disequilibrium, or linkage disequilibrium (LD). Whereas unlinked loci reach independence (Hardy–Weinberg equilibrium) in a single generation, linked loci with recombination rate $\theta < 0.5$ reduce initial LD in an infinite population to a proportion $e^{-t\theta}$ after t generations. The time required to go halfway to equilibrium is therefore $T = (1n2)/\theta$, or more than a million years if $\theta = 10^{-5}$ and there are 20 years per generation. A convenient but inaccurate rule of thumb is that $\theta = 0.01$ corresponds to about 1 megabase (Mb). By this approximation, $\theta = 10^{-5}$ corresponds to 1 kb. If θ is as small as 10^{-6}, the time since apes and hominids diverged is not long enough to go halfway to equilibrium. Therefore selection is not required to explain persistence of disequilibrium, which depends to a considerable extent on episodes of population contraction. There have been two major bottlenecks in human evolution. The first was when two chromosomes that are nonhomologous in apes fused to form the chromosome 2 inherited by our species. The second bottleneck was when we migrated out of Africa in the last 100 000 years. As a consequence, LD is least for sub-Saharan Africa. Lesser bottlenecks have occured in the history of particular populations.

LD may be measured in many ways. Some are confounded with significance tests, and therefore with sample size. All are to some degree confounded with allele frequencies. The most reliable and best validated is the association probability ρ_t, which is made up of two parts. Association that has diminished from an initial value ρ_0 in founders is $\rho_{rt} = \rho_0 e^{-(1/2N+\theta)t}$, where N is the effective size over t generations. Association that has built up by genetic drift since the founders is $\rho_{ct} = L\,(1 - e^{-(1/2N+\theta)t})$, and $\rho_t = \rho_{rt} + \rho_{ct}$. If N is constant, the equilibrium value as $t \to \infty$ is $L = 1/(1 + 2N\theta)$ if θ is small and $1/(1 + 2N)$ if $\theta = 0.5$. The latter is negligible in real populations. If $1/2\,N$ is small compared to θ, ρ_d follows the Malecot model for isolation by distance, equating $t\theta$ to εd, where d is distance between loci. On the genetic scale d measures recombination directly, with relatively larger sampling error over small distances. On the physical scale d is only indirectly related to recombination, but is more accurate if sequence-based. Choice should be based on goodness of fit to the best available maps. Analyzed as isolation by distance, LD provides a way to compare allelic association for chromosome regions in different populations, and therefore to detect variations in recombination, selective sweeps that reduced haplotype diversity, and effects of population history and structure. This information determines the optimal populations and density of markers for positional cloning of genes affecting normal physiology and disease. Localization is more precise by LD than by linkage. An alternative for multilocus haplotypes is cladistic analysis when its assumptions to reduce the number of independent variables are valid and the causal region has been made small by LD or other evidence.

See also: **Bottleneck Effect; Genetic Drift**