Linkage Disequilibrium

Peter JP Croucher, *University of California at Berkeley, Berkeley, California, USA*

When two or more polymorphic loci are studied in a population, the interaction between the loci is often expressed in terms of linkage disequilibrium (LD). The loci are in LD if their respective alleles do not associate independently (randomly). The degree of over- or underrepresentation of the expected haplotypes measures the extent of LD. Recombination acts to reduce LD.

Advanced article

Article Contents

- Measures of Linkage Disequilibrium
- Sources of Linkage Disequilibrium
- Multilocus Models and Interaction
- Decay as a Function of Age

doi: 10.1002/9780470015902.a0005427.pub2

Measures of Linkage Disequilibrium

The genetics of two or more loci are considered in terms of haplotype frequencies. Consider a pair of segregating loci, one with alleles A and a at frequencies p_A and p_a and one with alleles B and b at frequencies p_B and p_b. If the alleles associate independently (there is no linkage disequilibrium (LD)) then the frequencies of the four possible haplotypes or gametes, AB, Ab, aB and ab, are given by the products of the allele frequencies p_Ap_B, p_Ap_b, p_ap_B and p_ap_b. In the presence of LD some of these haplotypes will be more frequent than expected and some will be more rare; this difference from expectation is typically measured with the LD coefficient D (Lewontin and Kojima, 1960) as tabulated below: **Table 1**. **See also**: Population Genetics: Multilocus

Consequently, $D = p_{AB}p_{ab} - p_{Ab}p_{aB}$. This is often expressed more simply as $D = p_{Ab} - p_Ap_b$. D can take values between -0.25 and $+0.25$. It is a common practise to work with D^2 to eliminate problems with sign. Normalized coefficients of LD are often employed, most commonly D' and r^2. D' (Lewontin, 1964) takes values between -1 and $+1$ and is less dependent on allele frequencies than is D. D' is given by dividing D by its maximum possible numerical value for the given allele frequencies:

$$\begin{aligned} D' &= D/\min(p_ap_B, p_Ap_b) \text{ if } D > 0 \text{ or} \\ D' &= D/\min(p_Ap_B, p_ap_b) \text{ if } D < 0 \end{aligned} \quad [1]$$

The other commonly used normalized coefficient, r^2 (Hill and Robertson, 1968), the square of the allelic correlation coefficient r, takes values from 0 to 1 and is calculated by dividing D^2 by the product of all four allele frequencies:

$$r^2 = (p_{AB}p_{ab} - p_{Ab}p_{aB})^2/(p_Ap_ap_Bp_b) \quad [2]$$

Numerous additional measures of LD have been proposed, especially within the context of mapping loci associated with disease (see, e.g. Devlin and Risch, 1995; Morton *et al.*, 2001).

Sources of Linkage Disequilibrium

Populations are not finite. Random drift generates LD because not all haplotypes are sampled proportionately from generation to generation. Consequently, population history (e.g. bottlenecks and founder effects) can generate marked differences in the extent of LD between populations. **See also**: Genetic Drift; Genetic Load; Population Genetics: Historical Aspects; Population History and Linkage Disequilibrium

Frequently, LD exists between loci simply because an insufficient number of generations have passed to allow recombination to randomize the haplotypes in the population. When a new mutation arises on a chromosome it will initially be in complete LD with one of the alleles of any neighbouring polymorphism. **See also**: Genetic Variation: Polymorphisms and Mutations

LD can be generated by selection. Epistatic effects, where the fitness of an allele at one locus depends on that at another locus, can lead to the preferential selection of certain haplotypes and therefore the maintenance of LD. The human leukocyte antigen (HLA) genes may be one example. **See also**: Disease Associations: Human Leukocyte Antigen (HLA) and Apolipoprotein E (APOE) Gene; Epistasis; Major Histocompatibility Complex (MHC); Population Genetics: Multilocus

LD also results when differentiated populations (i.e. the allele frequencies at loci differ) merge (admixture) or when individuals with different genotypes mate nonrandomly. **See also**: Population History and Linkage Disequilibrium

Multilocus Models and Interaction

The HLA system represents a multilocus system of highly variable, tightly linked genes involved in immune defence and self-recognition. Many haplotypes are overrepresented and this high level of LD may be the result of selection for heterozygote advantage. The HLA is a classical, albeit

Table 1 The LD coefficient D is the difference between the observed and the expected haplotype frequencies

Haplotype	Frequency
AB	$p_{AB} = p_Ap_B + D$
Ab	$p_{Ab} = p_Ap_b - D$
aB	$p_{aB} = p_ap_B - D$
ab	$p_{ab} = p_ap_b + D$

complex, example of a coadapted gene complex. Much current attention focuses on measures of LD for multiple-locus (and multiple-allele) systems. Many two-locus measures can be extended to multilocus measures, e.g. D and D' (Ayres and Balding, 2001; Weir, 1996). Most approaches do not account for the higher-order LD present in multiple-locus systems. Some novel measures such as the haplotype-based Normalized Entropy Difference (Nothnagel *et al*., 2002), together with 'top–down' mathematical decomposition (Gorelick and Laubichler, 2004), attempt to incorporate these difficulties. **See also**: Disease Associations: Human Leukocyte Antigen (HLA) and Apolipoprotein E (APOE) Gene; Multilocus Linkage Analysis; Major Histocompatibility Complex (MHC); Population Genetics: Multilocus

Loci separated by large distances or on different chromosomes may also exhibit 'LD' or allelic association, for instance, because of higher-level interactions or epistasis. Consequently, the term gametic phase disequilibrium is often used to specify that physically linked markers are being examined. **See also**: Epistasis; Genetic Maps: Integration; Population Genetics: Multilocus

Decay as a Function of Age

Given an infinite population and no selection, recombination will act over successive generations to reduce the amount of LD between two physically linked markers. LD decays exponentially at a rate that depends on the linkage distance or recombination fraction, r, such that from one generation to the next $D_{n+1} = (1-r)\, D_n$. If the loci are unlinked ($r = 0.5$) then D will halve each generation, but if r is small, as it may be for closely linked markers, then substantial levels of LD can remain for hundreds of generations. The persistence of LD between closely spaced markers is frequently exploited in attempts to map disease-causing mutations. Projects, such as the International HapMap Project (The International HapMap Consortium, 2005), are attempting to map LD patterns across the human genome to maximize this potential. **See also**: Blocks of Limited Haplotype Diversity; Human Variation Databases; Linkage Analysis; Linkage and Association Studies; Population History and Linkage Disequilibrium; Population Genetics: Multilocus; Susceptibility Genes: Detection

References

Ayres KL and Balding DJ (2001) Measuring gametic disequilibrium from multilocus data. *Genetics* **157**: 413–423.

Devlin B and Risch N (1995) A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics* **29**: 311–322.

Gorelick R and Laubichler MD (2004) Decomposing multilocus linkage disequilibrium. *Genetics* **166**: 1581–1583.

Hill WG and Robertson A (1968) Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics* **38**: 226–231.

Lewontin RC (1964) The interaction of selection and linkage I. General considerations; heterotic models. *Genetics* **49**: 49–67.

Lewontin RC and Kojima K (1960) The evolutionary dynamics of complex polymorphisms. *Evolution* **14**: 458–472.

Morton NE, Zhang W, Taillon-Miller P *et al*. (2001) The optimal measure of allelic association. *Proceedings of the National Academy of Sciences of the USA* **98**: 5217–5221.

Nothnagel M, Fürst R and Rhode K (2002) Entropy as a measure of linkage disequilibrium over multilocus haplotype blocks. *Human Heredity* **54**: 186–198.

The International HapMap Consortium (2005) A haplotype map of the human genome. *Nature* **437**: 1299–1320.

Weir BS (1996) *Genetic Data Analysis II*. Sunderland, MA: Sinauer Associates.

Further Reading

Daly M, Rioux JD, Schaffner SF, Hudson TJ and Lander ES (2001) High-resolution haplotype structure in the human genome. *Nature Genetics* **29**: 229–232.

Goldstein DB (2001) Islands of linkage disequilibrium. *Nature Genetics* **29**: 109–111.

Jorde JB (2000) Linkage disequilibrium and the search for complex disease genes. *Genome Research* **10**: 1435–1444.

Kruglyak L (1999) Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nature Genetics* **22**: 139–144.

Lewontin RC (1988) On measures of gametic disequilibrium. *Genetics* **120**: 849–852.

Maynard Smith J (1989) *Evolutionary Genetics*. Oxford, UK: Oxford University Press.

Nordborg M and Tavaré S (2002) Linkage disequilibrium: what history has to tell us. *Trends in Genetics* **18**: 83–90.

Reich DE, Cargill M, Bolk S *et al*. (2001) Linkage disequilibrium in the human genome. *Nature* **441**: 199–204.