Linkage disequilibrium.

We consider first genetic markers of two alleles in diploid case. Then the allelic marker has two levels ”+” and ”−”, and the genotype marker has three levels ”++”, ”+−” and ”−−”.

1: Allelic linkage disequilibrium (LDE) is based on the alele frequencies p_+ and p_-. For any two fixed locuses denote p_{xy} is the frequency of the variant x at the first locus and variant y at the second one. In these notations we have the following frequency table:

Locus 1\Locus 2	”+”	”−”	Total
”+”	p_{++}	p_{+-}	$p_{+\bullet}$
”−”	p_{-+}	p_{--}	$p_{-\bullet}$
Total	$p_{\bullet+}$	$p_{\bullet-}$	1

Introduce the coefficient $D = p_{++} - p_{\bullet+}p_{+\bullet}$, which is the covariance between the two random variables taking values 0 under + variant and 1 under − variant (or vice versa).

There are two ways to determine the allelic LDE coefficients:

$$r = \frac{D}{\sqrt{p_{+\bullet}(1-p_{+\bullet})p_{\bullet+}(1-p_{\bullet+})}}$$

and r^2 is commonly used, and

$$D' = |D|/D_{\max},$$

where

$$D_{\max} = \begin{cases} \min(p_{+\bullet}p_{\bullet+}, p_{-\bullet}p_{\bullet-}), \text{ under } D < 0 \\ \min(p_{+\bullet}p_{\bullet-}, p_{-\bullet}p_{\bullet+}), \text{ under } D > 0. \end{cases}$$

We are using the signed version of the D' coefficient

$$D'_s = D/D_{\max}$$

(or the r coefficient) to measure allelic LDE in two locuses.

2: Composite LDE. In the diploid organisms all the genome variants in two locuses (haplotypes) are collected in the following table:

Locus 1\Locus 2	”++”	”+−”	”−+”	”−−”	Total
”++”	p_{++++}	p°_{+++-}	p°_{++-+}	p_{++--}	$p_{++\bullet}$
”+−”	p°_{+-++}	p°_{+-+-}	p°_{+--+}	p°_{+---}	$p_{+-\bullet}$
”−+”	p°_{-+++}	p°_{-++-}	p°_{-+-+}	p°_{-+--}	$p_{-+\bullet}$
”−−”	p_{--++}	p°_{--+-}	p°_{---+}	p_{----}	$p_{--\bullet}$
Total	$p_{\bullet++}$	$p_{\bullet+-}$	$p_{\bullet-+}$	$p_{\bullet--}$	1

After unification cells with identical combination we obtain 10-cells table:

<table>
<tr><th>Locus 1\Locus 2</th><th>”++”</th><th>”+−”</th><th>”−+”</th><th>”−−”</th><th>Total</th></tr>
<tr><td>”++”</td><td>p_{++++}</td><td colspan="2">p_{+++-}</td><td>p_{++--}</td><td>$p_{++\bullet}$</td></tr>
<tr><td>”+−”</td><td>p_{+-++}</td><td>p_{+-+-}</td><td>p_{+--+}</td><td>p_{+---}</td><td>$p_{+-\bullet}$</td></tr>
<tr><td>”−−”</td><td>p_{--++}</td><td colspan="2">p_{--+-}</td><td>p_{----}</td><td>$p_{--\bullet}$</td></tr>
<tr><td>Total</td><td>$p_{\bullet++}$</td><td colspan="2">$p_{\bullet+-}$</td><td>$p_{\bullet--}$</td><td>1</td></tr>
</table>

were $p_{ijks} = 2p^\circ_{ijks} = 2p^\circ_{ijsk} = 2p^\circ_{jiks}$.

The haploid (gametic) part of the LDE: $D_g = p_g - p_{+\bullet}p_{\bullet+}$, where

$$p_g = p_{++++} + p_{+++-}/2 + p_{+-++}/2 + (p_{+-+-} + p_{+--+})/4.$$

The diploid (non gametic) part LDE: $D_d = p_d - p_{+\bullet}p_{\bullet+}$ linkage disequilibrium, where

$$p_d = p_{++++} + p_{+++-}/2 + p_{+-++}/2 + p_{+-+-}/2.$$

The composite LDE:

$$\Delta = D_g + D_d = p_g + p_d - 2p_{+\bullet}p_{\bullet+}.$$

If genotypes are known only the composite LDE cannot be obtained. Then one use the within chromosome equilibrium $p_{+-+-} = p_{+--+}$. Under the within chromosome equilibrium $\Delta = 2D_g$.

The composite linkage disequilibrium correlation is defined as follows:

$$r = \frac{\Delta}{\sqrt{(p_{+\bullet}p_{-\bullet} + D_1)(p_{\bullet+}p_{\bullet-} + D_2)}},$$

where $D_1 = HWD_1 = p_{++\bullet} - p_{+\bullet}^2$, $D_2 = HWD_2 = p_{\bullet++} - p_{\bullet+}^2$.

Under within chromosome equilibrium assumption if to set

$$\xi_1 = \begin{cases} -1, \text{ if genotype is "++"}, \\ 0, \text{ if genotype is "+-"}, \\ 1, \text{ if genotype is "--"}, \end{cases} \quad \text{and} \quad \xi_2 = \begin{cases} -1, \text{ if genotype is "++"}, \\ 0, \text{ if genotype is "+-"}, \\ 1, \text{ if genotype is "--"}, \end{cases}$$

then $\Delta = \mathbf{cov}(\xi_A, \xi_B)$ and $r = \mathbf{corr}(\xi_A, \xi_B)$.