

3-3-2-3-8 最もよく使われる連鎖不平衡係数 D' と r^2 の特徴と関係について

r^2 は R^2 とも表記されることがある

連鎖平衡・連鎖不平衡の最も特徴的な状態は以下のように3つである。

- 1 連鎖平衡にて $D' = 0$
- 2 つの biallelic markers が作る「完全連鎖」には大きく2種類ある。
存在し得る4つの haplotypes のうち、
- 2 2つの haplotypes しか存在しない場合と、
- 3 3つの haplotypes が存在する場合

前者を absolute disequilibrium、後者を complete disequilibrium と分けて表現することもある

	Haplotype AB	Haplotype Ab	Haplotype aB	Haplotype ab
連鎖平衡	$P(A) \times P(B)$	$P(A) \times (1 - P(B))$	$(1 - P(A)) \times P(B)$	$(1 - P(A)) \times (1 - P(B))$
Absolute disequilibrium	$P(A)$	0	0	$1 - P(A)$
Complete disequilibrium	$P(A)$	0	$P(B) - P(A)$	$1 - P(B)$

但し $P(A)$ は SNP A のアレル頻度、 $P(B)$ は SNP B のアレル頻度、AB、Ab、aB、ab はそれぞれ haplotype を表し、 $P(A)$ 、 $P(B)$ を用いた数式で表されたセル内の式は、連鎖の状態別のハプロタイプ頻度を表す。

この特徴的な3状態における D' と r^2 の値は

	D'	r^2
連鎖平衡	0	0
Absolute disequilibrium	1	1
Complete disequilibrium	1	0より大、1未満

このことからわかるように、Absolute disequilibrium と Complete disequilibrium との連鎖不平衡係数を区別したいときには r^2 が適切であり、どちらも「完全連鎖」しているとして一まとめにして扱いたい場合には D' が適当である。

一般的に r^2 の減衰は D' のそれよりも急峻であることが知られている。参考となる文献は

[Linkage Disequilibrium in Humans: Models and Data. JK Prichard and M Przeworski, Am J Hum Genet. 69:1-14,2001](#)

その他の事情として次のような関係が知られている。

多型の発生と消長を決定しているのは、変異の発生率(μ)と組替え率(c)、及び random drift であるとするのが、一般的であるが、そのような立場にたつとき、 r^2 は μ 及び c と以下のような関係にある。

$$E(r^2) = (4N(\mu + c))^{-1}$$

ここで $E(r^2)$ は r^2 の期待値である。