March 2008
Features

Natural selection, maths and milk

by Charlotte Mulcare

Let's recall the definitions:

F_X = observed frequency of the allele **X** (where **X** is one of **T**, **C**, **A** or **B**), regardless of which other allele it occurs with;

F_{XY} = observed frequency of **X** allele occurring with **Y** allele (where **X** is one of **T** or **C** and **Y** is one of **A** or **B**).

Each of the quantities above lie between zero and one, because they are *frequencies*: F_X is the number of times that the allele **X** was observed in our sample, divided by the total size of the sample.

Now, since in each individual case one of the **T** or **C** allele must always be observed, we have

$F_T + F_C = 1$.

Similarly,

$F_A + F_B = 1$.

We also have:

$F_{TA} + F_{TB} = \mathrm{F_T}$;

$F_{CA} + F_{CB} = \mathrm{F_C}$;

$F_{TA} + F_{CA} = F_A$;

$F_{TB} + F_{CB} = F_B$.

First we will show that

$D = F_{TA} - F_T F_A = F_{CB} - F_C F_B$.

This means that the value D can be calculated in two different ways and this will be useful later.

Clearly,

$F_{TA} = F_T - F_{TB} = 1 - F_C - F_{TB}$.

So

$D = F_{TA} - F_T F_A$

$= 1 - F_C - F_{TB} - (1 - F_C)(1 - F_B)$

$= F_B - F_{TB} - F_C F_B = F_{CB} - F_C F_B$,

which is what we wanted to show.

Now note that F_{TA} can take any value between 0 and F_T. If $F_{TA} = 0$, then this means that **T** was never observed with **A**, and if $F_{TA} = F_T$, then **T** was only ever observed with **A**.

Since $D = F_{TA} - F_T F_A$, we get

$-F_T F_A \;\; D \;\; F_T - F_T F_A = F_T F_B$.

Similarly, since D is also equal to $F_{CB} - F_C F_B$ and since $0 \;\; F_{CB} \;\; F_C$, we get

$-F_C F_B \;\; D \;\; F_C - F_T F_B = F_C F_A$.

For D^2 this means

$0 \;\; D^2 \;\; F_T F_C \; F_A \; F_B$.

Therefore, r^2 always lies between 0 and 1.

The last thing to notice is that D^2 can indeed take the value 1, but only when $F_{TA} = 0$ and $F_{CB} = 0$ (so that **T** only occurs with **B** and **C** only occurs with **A**), or when $F_{TA} = F_T$ and $F_{CB} = F_C$ (so that **T** only occurs with **A** and **C** only occurs with **B**).

In other words, $r^2 = 1$ precisely when there is complete association between alleles.

Return to main article

Plus is part of the family of activities in the Millennium Mathematics Project, which also includes the NRICH and MOTIVATE sites.