CubeX: Cubic Exact Solution

Results

For an explanation of the analysis and results please see notes below.

Number of biologically possible solutions: 1 . See X^2 table below 3x3

	Haplotype frequencies				LD statistics		
Solution	f_{11}	f_{12}	f_{21}	f_{22}	D'	r^2	X^2
α	0.7723	0.0277	0.1027	0.0973	0.723	0.2984	5.97

3x3 table of observed and expected diplotype numbers

Black numbers on white are original data entered
Coloured numbers on coloured background represent the solutions from the table above.

		SNP 2		
		11	12	22
SNP 1	11	12 11.9	1 0.9	0 0.0
	12	3 3.2	3 3.1	0 0.1
	22	0 0.2	1 0.4	0 0.2

Solution	X^2 of 3x3
α	1.4635

This is a X^2 of the 3x3 table. The higher the value, the less good the fit of the observed haplotypes to Hardy-Weinberg equilibrium. Please see footnote regarding degrees of freedom. If there are two or more solutions, the lower values are more likely (although note the different degrees of freedom if there are empty cells). However, a significant value indicates genotype data out of Hardy-Weinberg equilibrium, a problem that should be addressed before interpreting these results.

Other statistics

Minimum *biologically* possible f_{11}: 0.7
Maximum *biologically* possible f_{11}: 0.775

Number of impossible solutions: 2

β: f_{11} = 0.48188

γ: f_{11} = 0.65836

SNP 1 allele 1 frequency = 0.8
SNP 2 allele 1 frequency = 0.875

Notes

- Here f_{11} refers to an estimated haplotype frequency for allele 1 at locus 1 and allele 1 at locus 2 (likewise for other haplotypes 12,21 and 22). The character f should have a "hat" to indicate that it is estimated but html limitations prevent this.
- f_{11} is based on direct solution of the cubic equation expressing the phase uncertain double heterozygotes (middle square of the 3x3) and overall model in terms of estimated f_{11} assuming:
  1. random mating
  2. Hardy-Weinberg equilibrium at both loci
- Expectation-maximisation algorithms rely on an iteration rather than direct solution. The direct solution will display both the most likely (which EM should ideally reach) and other possible solutions where iterations may converge in error.
- X^2 values under the 3x3 represent difference between observed and expected diplotype frequencies. The number of degrees of freedom is equal to the number of observations (diplotype counts) minus four estimated parameters which are either three haplotypes (the fourth can be inferred) and D, or one haplotype, two allele frequencies and D. If nine different diplotypes are observed the number of degrees of freedom is therefore five. For each empty cell in the 3x3 the number of degrees of freedom is reduced by one. If the user knows there are only three haplotypes present (and therefore six diplotypes) then there are only three estimated parameters (D is inferred by the three haplotype frequencies) and 3df. It is important to note that in the latter case neither cubic solution nor iteration is necessary as the haplotype frequencies can be directly counted from the diplotype data. If the user believes that there are only three alleles and hence six diplotypes, but there are non-zero values for any of the other three possible diplotypes, then reconsideration of the technical veracity of the data and of the homogeneity of the population sample would be wise.
- In perfect Hardy-Weinberg equilibrium "expected" numbers will be whole numbers matching the "observed" diplotype numbers. However, imperfect Hardy-Weinberg proportions will result in impossible "fractions" of individuals, which are shown for comparison to observations.

Return to input