

HLA and Anthropology: The Hardy-Weinberg Story

10th INTERNATIONAL SUMMER
SCHOOL ON IMMUNOGENETICS

Stintino, Sardinia, 18 September, 2013

Outline

- Requirements
- HWE and a very simplified and arbitrary example
- Guidelines
- Selected readings

Step-wise

- Estimate allele and haplotype frequencies
 - E.g. Expectation-maximisation (EM) algorithm
- Test Hardy-Weinberg equilibrium
 - E.g. goodness-of-fit (χ^2 or likelihood-ratio G test) algorithm
- Test Selective neutrality
 - (Ewens–Watterson–Slatkin algorithm)
- *Test linkage disequilibrium*
 - *goodness-of-fit (χ^2 or likelihood-ratio G test)*
- *Compare different populations*
 - (*Fst genetic distance*)

HWE - Principle and Usage

- allele and genotype frequencies in an ideal population will remain constant over time and can be predicted from each other
- crucial in the validation of many current approaches in population genetic and disease association studies

Two Allele Model

Alleles	Frequency
A	a
B	b

Phenotypes	Frequency
A	a^2
AB	$2ab$
B	b^2

	A	B
A	a^2	ab
B	ab	b^2

Population I (n=50)

Phenotype , Observed	n
A	12
AB	25
B	13

Alleles	Frequency
A	0.49
B	0.51

Phenotype , Expected	Calculation	n
A	$0,49*0,49*50$	12,005
AB	$2*0,49*0,51*50$	24,99
B	$0,51*0,51*50$	13,005

Chi squared equals 0.000 with 2 degrees of freedom.

The two-tailed P value equals 1.0000

By conventional criteria, this difference is considered to be not statistically significant.

Population II (n=50)

Phenotype , Observed	n
A	25
AB	0
B	25

Alleles	Frequency
A	0.50
B	0.50

Phenotype , Expected	Calculation	n
A	$0,49*0,49*50$	12,5
AB	$2*0,49*0,51*50$	25
B	$0,51*0,51*50$	12,5

Chi squared equals 50.160 with 2 degrees of freedom.

The two-tailed P value is less than 0.0001

Multiallele Model

General

$$(p_1 + \dots + p_n)^2 = \sum_{i=1}^n \sum_{j=1}^n p_i p_j = 1$$

Frequencies

Homozygote

$$p_i^2$$

Heterozygotes

$$2p_i p_j$$

Software packages

- Arlequin
- Gene[rate]
- Pypop
- ...
- www.allelefrequencies.net

If the null hypothesis of the HWE is rejected ...

- Sample size?
- Data correct?
- Mutation?
- Migration?
- Selective Pressure?
- Non random mating?
 - ...

Population definitions and sampling strategies

- Avoid outdated racial classifications and population names (e.g. 'Caucasian') and using instead geographic and / or cultural (e.g. linguistic) criteria to describe human populations (e.g. 'pan-European')
- Questionnaire (http://hla-net.eu/population_questionnaire)

HLA typing standards for population genetics analyses

- Retain maximal information when reporting HLA typing results
- Report all ambiguities
- Take into account a preliminary list of alleles defined by polymorphisms outside the peptide-binding sites

HLA population data storage and analysis

- Software should be able to handle ambiguous data, e.g. Gene[rate]

Ethical Issues

- Heterogeneity in national requirements (Ethics committees, consent etc)
- <http://hla-net.eu>

Selected readings

- Estimation of evolutionary forces (*Tissue Antigens* 54:533, 1999, *Science* 272:67, 1996, *Hum Immunol.* 69:443, 2008)
- Peopling history (*Am J Phys Anthropol* 145:81, 2011, *Nature* 488:370, 2012 (not HLA))
- Registries (*PLoS One.* 7(7):e41400) 2012