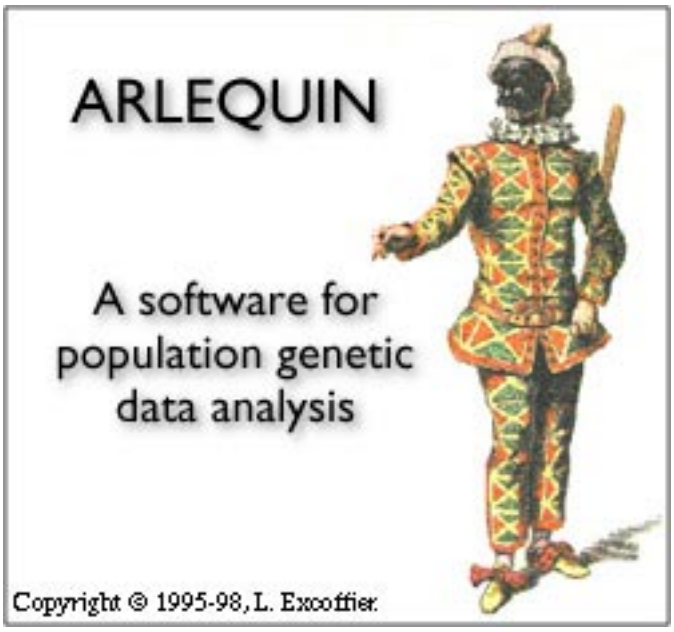
ARLEQUIN
A software for
population genetic
data analysis
Copyright © 1995-98, L. Excoffier.

ver 1.1

ARLEQUIN ver 1.1

A software for population genetic data analysis

Authors:
Stefan Schneider, Jean-Marc Kueffer, David Roessli, and Laurent Excoffier

Genetics and Biometry Laboratory
Dept. of Anthropology and Ecology
University of Geneva
CP 511
1211 Geneva 24
Switzerland

E-mail : arlequin@sc2a.unige.ch
URL: http://anthropologie.unige.ch/arlequin

December 1997

Table of contents:

1 Introduction 6
1.1 Why Arlequin? 6
1.2 Arlequin philosophy 6
1.3 About this manual 6
1.4 Data types handled by Arlequin 7
1.4.1 DNA sequences 8
1.4.2 RFLP Data 8
1.4.3 Microsatellite data 8
1.4.4 Standard data 8
1.4.5 Allele frequency data 9
1.5 Methods implemented in Arlequin 9
1.6 System requirements 10
1.7 Installing and uninstalling Arlequin 10
1.8 List of files included in the Arlequin package 11
1.9 Arlequin limitations 12
1.10 How to cite Arlequin 12
1.11 Acknowledgements 12
1.12 Bug report and comments 12
1.13 How to get the last version of the Arlequin software? 12
1.14 What is new in version 1.1 compared to version 1.0 13
1.15 Forthcoming developments 13
1.16 Remaining problem 14
2 Getting started 15
2.1 Preparing input files 15
2.2 Loading project files into Arlequin 15
2.3 Selecting analyses to be performed on your data 15
2.4 Creating and using Setting Files 15
2.5 Performing the analyses 16
2.6 Stopping the computations 16
2.7 Consulting the results 16
3 Input files 17
3.1 Format of Arlequin input files 17
3.2 Project file structure 17
3.2.1 Profile section 17
3.2.2 Data section 19
3.2.2.1 Haplotype list (optional) 19
3.2.2.2 Distance matrix (optional) 20
3.2.2.3 Samples 21
3.2.2.4 Genetic structure 23
3.3 Eexample of an input file 24
3.4 Automatically creating the outline of a project file 25
3.5 Conversion of data files 26
3.6 Arlequin batch files 27
4 Output files 28
4.1 Result file 28

4.2 View your results in HTML browser *28*
4.3 Arlequin Log file *28*
4.4 Back-up file *29*
4.5 Linkage Disequilibrium Result File *29*
4.6 Variance components null distribution histograms *29*
5 Examples of input files **30**
5.1 Example of allele frequency data *30*
5.2 Example of standard data (Genotypic data, unknown gametic phase, recessive alleles) *30*
5.3 Example of DNA sequence data (Haplotypic) *31*
5.4 Example of microsatellite data (Genotypic) *32*
5.5 Example of RFLP data(Haplotypic) *33*
5.6 Example of standard data (Genotypic data, known gametic phase) *34*
6 Arlequin interface **35**
6.1 Menus *35*
6.1.1 File Menu 35
6.1.2 Edit Menu 35
6.1.3 Project Menu 35
6.1.4 Setup Menu 37
6.1.5 Special Menu 37
6.1.6 Window Menu 37
6.1.7 Help Menu 38
6.2 Toolbar *38*
6.3 Status Bar *39*
6.4 Dialog boxes *39*
6.4.1 General Settings 39
6.4.2 Diversity indices 41
6.4.3 Neutrality tests 43
6.4.4 Gametic disequilibrium 44
6.4.5 Genetic structure 46
6.4.6 Launch Pad 48
7 Methodological outlines **50**
7.1 Intra-population level methods *51*
7.1.1 Standard diversity indices 51
7.1.1.1 Gene diversity 51
7.1.1.2 Number of usable loci 51
7.1.1.3 Number of polymorphic sites (S) 51
7.1.2 Molecular indices 51
7.1.2.1 Mean number of pairwise differences (π) 51
7.1.2.2 Nucleotide diversity or average gene diversity over L loci (RFLP and DNA data) 52
7.1.2.3 Theta estimators 52
7.1.2.3.1 Theta(Hom) 52
7.1.2.3.2 Theta(*S*) 53
7.1.2.3.3 Theta(*k*) 53
7.1.2.3.4 Theta(π) 54
7.1.2.4 Mismatch distribution 54
7.1.2.5 Estimation of genetic distances between DNA sequences 55
7.1.2.5.1 Pairwise difference 56

7.1.2.5.2 Percentage difference 56
7.1.2.5.3 Jukes and Cantor 56
7.1.2.5.4 Kimura 2-parameters 57
7.1.2.5.5 Tamura 57
7.1.2.5.6 Tajima and Nei 58
7.1.2.5.7 Tamura and Nei 58
7.1.2.6 Estimation of genetic distances between RFLP haplotypes 59
7.1.2.6.1 Number of pairwise difference 59
7.1.2.6.2 Proportion of difference 60
7.1.2.7 Estimation of distances between Microsatellite haplotypes 60
7.1.2.7.1 No. of different alleles 60
7.1.2.7.2 Sum of squared size difference 60
7.1.2.8 Estimation of distances between Standard haplotypes 61
7.1.2.8.1 Number of pairwise differences 61
7.1.3 Haplotype frequency estimation 61
7.1.3.1 Haplotypic data or Genotypic data with known Gametic phase 61
7.1.3.2 Genotypic data with unknown Gametic phase 61
7.1.4 Linkage disequilibrium between pairs of loci 62
7.1.4.1 Exact test of linkage disequilibrium (haplotypic data) 62
7.1.4.2 Likelihood ratio test of linkage disequilibrium (genotypic data, gametic phase unknown) 64
7.1.4.3 Measures of gametic disequilibrium (haplotypic data) 64
7.1.5 Hardy-Weinberg equilibrium. 65
7.1.6 Neutrality tests. 66
7.1.6.1 Ewens-Watterson homozygosity test 66
7.1.6.2 Ewens-Watterson-Slatkin exact test 66
7.1.6.3 Chakraborty's test of population amalgamation 67
7.1.6.4 Tajima's test of selective neutrality 67
7.2 Inter-population level methods *67*
7.2.1 Population genetic structure inferred by analysis of variance (AMOVA) 67
7.2.1.1 Haplotypic data, one group of populations 69
7.2.1.2 Haplotypic data, several groups of populations 70
7.2.1.3 Genotypic data, one group of populations, no within- individual level 70
7.2.1.4 Genotypic data, several groups of populations, no within- individual level 71
7.2.1.5 Genotypic data, one population, within- individual level 72
7.2.1.6 Genotypic data, one group of populations, within- individual level 72
7.2.1.7 Genotypic data, several groups of populations, within- individual level 73
7.2.2 Population pairwise genetic distances 73
7.2.3 Exact tests of population differentiation 75
8 Appendix **76**
8.1 Overview of input file keywords *76*
9 References **79**

1 INTRODUCTION

1.1 Why Arlequin?

Arlequin is the French translation of "Arlecchino", a famous character of the Italian "Commedia dell'Arte". As a character he has many aspects, but he has the ability to switch among them very easily according to its needs and to necessities. This polymorphic ability is symbolized by his colorful costume, from which the Arlequin icon was designed.

1.2 Arlequin philosophy

The goal of Arlequin is to provide the average user in population genetics with quite a large set of methods and statistical tests, in order to extract information on genetic and demographic features of a collection of population samples.

The graphical interface has been designed such as to allow the user to rapidly select the different analyses he wants to perform on his data. We felt important to be able to explore the data, to analyze several times the same data set from different perspectives, with different selected options.

The statistical tests implemented in Arlequin have been chosen such as to minimize hidden assumptions and to be as powerful as possible. Thus, they often take the form of either permutation tests or exact tests, with some exceptions.

Finally, we wanted Arlequin to be able to handle genetic data under many different forms, and to try to carry out the same types of analyses irrespective of the format of the data.

Because Arlequin has a rich set of features and many options, it means that the user has to spend some time in learning them. However, we hope that the learning curve will not be that steep.

Arlequin is made available free of charge, as long as we have enough local resources to support the development of the program.

1.3 About this manual

The main purpose of this manual is to allow you to use Arlequin on your own, in order to limit as far as possible e-mail exchange with us.

In this manual, we have tried to provide a description of

1. the data types handled by Arlequin
2. the way these data should be formatted before the analyses
3. the graphical interface
4. the impact of different options on the computations
5. methodological outlines describing which computations are actually performed by Arlequin.

Even though this manual contains the description of some theoretical aspects, it should not be considered as a textbook in basic population genetics. We strongly recommend you to consult the original references provided with the description of a given method if you are in doubt with any aspect of the analysis.

1.4 Data types handled by Arlequin

Arlequin can handle several types of data either in *haplotypic* or *genotypic* form. The basic data types are:

- DNA sequences
- RFLP data
- Microsatellite data
- Standard data
- Allele frequency data

By *haplotypic form* we mean that genetic data can be presented under the form of haplotypes (i.e. a combination of alleles at one or more loci). This haplotypic form can result from the analyses of haploid genomes (mtDNA, Y chromosome, prokaryotes), or from diploid genomes where the gametic phase could be inferred by one way or another. Note that allelic data are treated here as a single locus haplotype.

Ex 1: haplotypic RFLP data : `1001101001010010`10`
Ex 2: haplotypic standard HLA data : `DRB1*0101 DQB1*0102 DPB1*0201`

By *genotypic form*, we mean that genetic data is presented under the form of diploid genotypes (i.e. a combination of pairs of alleles at one or more loci).

Ex1: genotypic DNA sequence data:

```
ACGGCATTTAAGCATGACATACGGATTGACA
ACGGGATTTTAGCATGACATTCGGATAGACA
```

Ex 2: genotypic Microsatellite data :

```
63    24    32
62    24    30
```

The gametic phase of a multi-locus genotype may be either known or unknown. If the gametic phase is known, the genotype can be considered as made up of two well-defined haplotypes. For genotypic data with unknown gametic phase, you can consider the two alleles present at each locus as codominant, or you can allow for the presence of a recessive allele. This gives finally four possible forms of genetic data:

- Haplotypic data,
- Genotypic data with known gametic phase,
- Genotypic data with unknown gametic phase (no recessive alleles)
- Genotypic data with unknown gametic phase (recessive alleles).

1.4.1 DNA sequences

DNA sequences of arbitrary length can be accommodated by Arlequin. Each nucleotide is considered as a distinct locus. The four nucleotides "C", "T", "A", "G" are considered as unambiguous alleles for each locus, and the "-" is used to indicate a deleted nucleotide. Usually the question mark "?" codes for an unknown nucleotide. The following notation for ambiguous nucleotides are also recognized:

R: A/G (purine)

Y: C/T (pyrimidine)

M: A/C

W: A/T

S: C/G

K: G/T

B: C/G/T

D: A/G/T

H: A/C/T

V: A/C/G

N: A/C/G/T

1.4.2 RFLP Data

RFLP haplotypes of arbitrary length can be handled by Arlequin. Each restriction site is considered as a distinct locus. The presence of a restriction site should be coded as a "1", and its absence as a "0". The "-" character should be used to denote the deletion of a site, not its absence due to a point mutation.

1.4.3 Microsatellite data

The raw data consist here of the allelic state of one or an arbitrary number of microsatellite loci. For each locus, one should in principle provide the number of repeats of the microsatellite motif as the allelic definition, if one wants his data to be analyzed according to the step-wise mutation model (for the analysis of genetic structure). It may occur that the absolute number of repeats is unknown. If the difference in length between amplified products is the direct consequence of changes in repeat numbers, then the minimum length of the amplified product could serve as a reference, allowing to code the other alleles in terms of additional repeats as compared to this reference. If this strategy is impossible, then any other number could be used as an allelic code, but the step-wise mutation model cold not be assumed for theses data.

1.4.4 Standard data

Data for which the molecular basis of the polymorphism is not particularly defined, or when different alleles are considered as mutationally equidistant from each other. Standard data haplotypes are thus compared for their content at each locus, without taking special care about the nature of the alleles, which can be either similar or different. For instance, HLA data (human MHC) enters the category of standard data.

8 APPENDIX

8.1 Overview of input file keywords

Keywords	Description	Possible values
[Profile]		
Title	A title describing the present analysis	A string of alphanumeric characters within double quotes
NbSamples	The number of different samples listed in the data file	A positive integer larger than zero
DataType	The type of data to be analyzed (only one type of data per project file is allowed)	STANDARD, DNA, RFLP, MICROSAT, FREQUENCY
GenotypicData	Specifies if genotypic or gametic data is available	0 (haplotypic data), 1 (genotypic data)
LocusSeparator	The character used to separate adjacent loci	WHITESPACE, TAB, NONE, or any character other than "#", or the character specifying missing data Default: WHITESPACE
GameticPhase	Specifies if the gametic phase is known (for genotypic data only)	0 (gametic phase not known), 1 (known gametic phase) Default: 1
RecessiveData	Specifies whether recessive alleles are present at all loci (for genotypic data)	0 (co-dominant data), 1 (recessive data) Default: 0
RecessiveAllele	Specifies the code for the recessive allele	Any string within quotation marks This string can be explicitly used in the input file to indicate the occurrence of a recessive homozygote at one or several loci. Default: "null"
MissingData	A character used to specify the code for missing data	"?" or any character within quotes, other than those previously used Default: "?"
Frequency	Specifies the format of haplotype frequencies	ABS (absolute values), REL (relative values: absolute values will be found by multiplying the relative frequencies by the sample sizes) Default: ABS

CompDistMatrix	Specifies if the distance matrix has to be computed from the data	0 (use any specified distance matrix), 1 (compute distance matrix from haplotypic information) Default: 0
FrequencyThreshold	The minimum frequency a haplotype has to reach for being listed in any output file	A real number between 1e-2 and 1e-7. Default: 1e-5
EpsilonValue	The EM algorithm convergence criterion. (For advanced users only)	A real number between 1e-7 and 1e-12. Default: 1e-7

Keywords	**Description**	**Possible values**
[Data]		
[[HaplotypeDefinition]]	(facultative section)	
HaplListName	The name of a haplotype definition list	A string within quotation marks
HaplList	The list of haplotypes listed within braces ({...})	A series of haplotype definitions given on separate lines for each haplotype. Each haplotype is defined by a haplotype label and a combination of alleles at different loci. The Keyword EXTERN followed by a string within quotation marks may be used to specify that a given haplotype list is in a different file

Keywords	**Description**	**Possible values**
[Data]		
[[DistanceMatrix]]	(facultative section)	
MatrixName	The name of the distance matrix	A string within quotation marks
MatrixSize	The size of the matrix	A positive integer larger than zero (corresponding to the number of haplotypes listed in the haplotype list)
LabelPosition	Specifies whether haplotypes labels are entered by row or by column	ROW (the haplotype labels will be entered consecutively on one or several lines, within the MatrixData segment, before the distance matrix elements), COLUMN (the haplotype labels will be entered as the first column of each row of the distance matrix itself)
MatrixData	The matrix data itself listed within braces ({...})	The matrix data will be entered as a format-free lower-diagonal matrix. The haplotype labels can be either entered consecutively on one or several lines (if LabelPosition=ROW), or entered at the first column of each row (if labelPosition=COLUMN). The special keyword EXTERN may be used followed by a file name within quotation marks, stating that the data must be read in an another file

Keywords	Description	Possible values
[Data]		
[[Samples]]		
SampleName	The name of the sample. This keyword is used to mark the beginning of a sample definition	A string within quotation marks
SampleSize	Specifies the sample size	An integer larger than zero. For haplotypic data, it must specify the number of gene copies in the sample. For genotypic data, it must specify the number of individuals in the sample.
SampleData	The sample data listed within braces ({...})	The keyword EXTERN may be used followed by a file name within quotation marks, stating that the data must be read in a separate file. The SampleData keyword ends a sample definition

Keywords	Description	Possible values
[Data]		
[[Structure]]	(facultative section)	
StructureName	The name of a given genetic structure to test	A string of characters within quotation marks
NbGroups	The number of groups of populations	An integer larger than zero
IndividualLevel	Specifies whether the level of genetic variability within individuals has to be taken into account (for genotypic data only)	0 (the component of variance due to differences between haplotypes within individuals will be ignored) 1 (the component of variance due to differences between haplotypes within individuals, and its associated statistics will be computed)
Group	The definition of a group of samples, identified by their SampleName listed within braces ({...})	A series of strings within quotation marks all enclosed within braces, and, if desired, on separate lines

9 REFERENCES

Abramovitz, M., and I. A. Stegun, 1970 Handbook of Mathematical Functions. Dover, New York.

Aris-Brosou, S., and L. Excoffier, 1996 The impact of population expansion and mutation rate heterogeneity on DNA sequence polymorphism. Mol. Biol. Evol. 13: 494-504.

Cavalli-Sforza, L. L., and W. F. Bodmer, 1971 The Genetics of Human Populations. W.H. Freeman and Co., San Francisco, CA.

Chakraborty, R. 1990 Mitochondrial DNA polymorphism reveals hidden heterogeneity within some Asian populations. Am. J. Hum. Genet. 47:87-94.

Chakraborty, R., and K. M. Weiss, 1991 Genetic variation of the mitochondrial DNA genome in American Indians is at mutation-drift equilibrium. Am. J. Hum. Genet. 86: 497-506.

Cockerham, C. C., 1969 Variance of gene frequencies. Evolution 23: 72-83.

Cockerham, C. C., 1973 Analysis of gene frequencies. Genetics 74: 679-700.

Dempster, A., N. Laird and D. Rubin, 1977 Maximum likelihood estimation from incomplete data via the EM algorithm. J Roy Statist Soc 39: 1-38.

Efron, B. 1982 The Jacknife, the Bootstrap and other Resampling Plans. Regional Conference Series in Applied Mathematics, Philadelphia:.

Ewens, W.J. 1972 The sampling theory of selectively neutral alleles. Theor. Popul. Biol. 3:87-112.

Ewens, W.J. 1977 Population genetics theory in relation to the neutralist-selectionist controversy. In: Advances in human genetics, edited by Harris, H. and Hirschhorn, K.New York:Plenum Press,p. 67-134.

Excoffier, L., Smouse, P., and Quattro, J. 1992 Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. Genetics 131:479-491.

Excoffier, L. and M. Slatkin. 1995 Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. Mol. Biol. Evol. 12:921-927

Excoffier, L., and M. Slatkin, 1998 Incorporating genotypes of relatives into a test of linkage disequilibrium. Am. J. Hum. Genet. (January issue)

Goudet, J., M. Raymond, T. de Meeüs and F. Rousset, 1996 Testing differentiation in diploid populations. Genetics 144: 1933-1940.

Guo, S. and Thompson, E. 1992 Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 48:361-372.

Harpending, R. C., 1994 Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. Hum. Biol. 66: 591-600.

Hudson, R. R., 1990 Gene genealogies and the coalescent proces, pp. 1-44 in Oxford Surveys in Evolutionary Biology, edited by Futuyama, and J. D. Antonovics. Oxford University Press, New York.

Jukes, T. and Cantor, C. 1969 Evolution of protein molecules. In: Mammalian Protein Metabolism, edited by Munro HN, New York:Academic press, p. 21-132.

Kimura, M. 1980 A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. J. Mol. Evol. 16:111-120.

Kumar, S., Tamura, K., and M. Nei. 1993 MEGA, Molecular Evolutionary Genetic Analysis ver 1.0.The Pennsylvania State University, University Park, PA 16802.

Lange, K., 1997 Mathematical and Statistical Methods for Genetic Analysis. Springer, New York.

Levene H. (1949). On a matching problem arising in genetics. Annals of Mathematical Statistics 20, 91-94.

Lewontin, R. C. (1964) The interaction of selection and linkage. I. General considerations; heterotic models. Genetics 49: 49-67.

Lewontin, R. C., and K. Kojima. (1960) The evolutionary dynamics of complex polymorphisms. Evolution 14: 450-472.

Long, J. C., 1986 The allelic correlation structure of Gainj and Kalam speaking people. I. The estimation and interpretation of Wright's F-statistics. Genetics 112: 629-647.

Michalakis, Y. and Excoffier, L. , 1996 A generic estimation of population subdivision using distances between alleles with special reference to microsatellite loci. Genetics 142:1061-1064.

Nei, M., 1987 Molecular Evolutionary Genetics. Columbia University Press, New York, NY, USA.

Raymond M. and F. Rousset. 1994 GenePop. ver 3.0. Institut des Sciences de l'Evolution. Université de Montpellier, France.

Raymond M. and F. Rousset. 1995 An exact tes for population differentiation. Evolution 49:1280-1283.

Reynolds, J., Weir, B.S., and Cockerham, C.C. 1983 Estimation for the coancestry coefficient: basis for a short-term genetic distance. Genetics 105:767-779.

Rice, J.A. 1995 Mathematical Statistics and Data Analysis. 2nd ed. Duxburry Press: Belmont, CA

Rogers, A., 1995 Genetic evidence for a Pleistocene population explosion. Evolution 49: 608-615.

Rogers, A. R., and H. Harpending, 1992 Population growth makes waves in the distribution of pairwise genetic differences. Mol. Biol. Evol. 9: 552-569.

Rousset, F., 1996 Equilibrium values of measures of population subdivision for stepwise mutation processes. Genetics 142: 1357-1362.

Slatkin, M., 1991 Inbreeding coefficients and coalescence times. Genet. Res. Camb. 58: 167-175.

Slatkin, M. 1994a Linkage disequilibrium in growing and stable populations. Genetics 137:331-336.

Slatkin, M. 1994b An exact test for neutrality based on the Ewens sampling distribution. Genet. Res. 64(1):71-74.

Slatkin, M., 1995 A measure of population subdivision based on microsatellite allele frequencies. Genetics 139: 457-462.

Slatkin , M., 1996 A correction to the exact test based on the Ewens sampling distribution. Genet. Res. 68: 259-260.

Slatkin, M. and Excoffier, L. 1996 Testing for linkage disequilibrium in genotypic data using the EM algorithm. Heredity 76:377-383.

Stewart, F. M. 1977 Computer algorithm for obtaining a random set of allele frequencies for a locus in an equilibrium population. Genetics 86:482-483.

Strobeck, K., 1987 Average number of nucleotide differences in a sample from a single subpopulation: A test for population subdivision. Genetics 117: 149-153.

Tajima, F., 1983 Evolutionary relationship of DNA sequences in finite populations. Genetics 105: 437-460.

Tajima, F. 1989a. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585-595,.

Tajima, F. 1989b. The effect of change in population size on DNA polymorphism. Genetics 123:597-601,.

Tajima, F. 1993. Measurement of DNA polymorphism. In: Mechanisms of Molecular Evolution. Introduction to Molecular Paleopopulation Biology, edited by Takahata, N. and Clark, A.G., Tokyo, Sunderland, MA:Japan Scientific Societies Press, Sinauer Associates, Inc., p. 37-59.

Tajima, F. and Nei, M. 1984. Estimation of evolutionary distance between nucleotide sequences. Mol. Biol. Evol. 1:269-285.

Tajima, F., 1996 The amount of DNA polymorphism maintained in a finite population when the neutral mutation rate varies among sites. Genetics 143: 1457-1465.

Tamura, K., 1992 Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C content biases. Mol. Biol. Evol. 9: 678-687.

Tamura, K., and M. Nei, 1993 Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. 10: 512-526.

Uzell, T., and K. W. Corbin, 1971 Fitting discrete probability distribution to evolutionary events. Science 172: 1089-1096.

Watterson, G., 1975 On the number of segregating sites in genetical models without recombination. Theor.Popul.Biol. 7: 256-276.

Watterson, G. 1978. The homozygosity test of neutrality. Genetics 88:405-417

Watterson, G. A., 1986 The homozygosity test after a change in population size. genetics 112: 899-907.

Weir, B. S., 1996 Genetic Data Analysis II: Methods for Discrete Population Genetic Data. Sinauer Assoc., Inc., Sunderland, MA, USA.

Weir, B.S. and Cockerham, C.C. 1984 Estimating F-statistics for the analysis of population structure. Evolution 38:1358-1370.

Wright, S., 1951 The genetical structure of populations. Ann.Eugen. 15: 323-354.

Wright, S., 1965 The interpretation of population structure by F-statistics with special regard to systems of mating. Evol 19: 395-420.

Zouros, E., 1979 Mutation rates, population sizes and amounts of electrophoretic variation of enzyme loci in natural populations. Genetics 92: 623-646.