Updated version (June 2006, 20th EFI meeting in Oslo) of the teaching session "Biostatistics" presented at the 16th European Histocompatibility Conference, Strasbourg, March 19-22, 2002

HLA data analysis in anthropology: some practical applications

Alicia Sanchez-Mazas

Laboratory of Anthropology, Genetics and Peopling history (AGP), Department of Anthropology and Ecology, University of Geneva (http://agp.unige.ch)

The following set of applications show how to use several programs online (available at the Gene[VA] site http://geneva.unige.ch) and to download (Arlequin, http://anthro.unige.ch/software/arlequin/) to estimate HLA allele and haplotype frequencies, to test Hardy-Weinberg equilibrium, linkage disequilibrium, allelic and disease associations for HLA, and to compare different populations tested for this system.

To learn how to manage HLA data with ambiguities (not possible here) and other useful procedures, we recommend another set of programs – Gene[rate] – available at http://geneva.unige.ch/generate, and the companion document explaining their use.

1. Estimating allele and haplotype frequencies

Recommended algorithm: EM (expectation-maximisation)

We show below an application with Arlequin:

Creating and loading an input file

```
 [Profile]

        Title="HLA-B-Cw Mandenka"
        NbSamples=1
        DataType=STANDARD
        GenotypicData=1
        LocusSeparator=WHITESPACE
        GameticPhase=0
        RecessiveData=1
        RecessiveAllele="blank"
        MissingData='?'
        FrequencyThreshold=0.00001
        EpsilonValue=0.0000001

[Data]
        [[Samples]]

          SampleName="Mandenka"
          SampleSize=163
          SampleData={
         101  1      B*7801      Cw*03
                        B70    Cw*1601
         102  1      B*7801      Cw*03
                        B70    Cw*1601
         103  1      B*7801      Cw*03
                        B70    Cw*1601
         105  1         B47    Cw*0202
                        B70    Cw*0701
         106  1         B35      Cw*03
                        B70    Cw*0401
         108  1      B*5702      Cw*18
                     B*7801    Cw*0701
         109  1         B35    Cw*0401
                        B70    Cw*0401
         110  1          B7    Cw*1601
         ….
         ….
         ….
         }
```

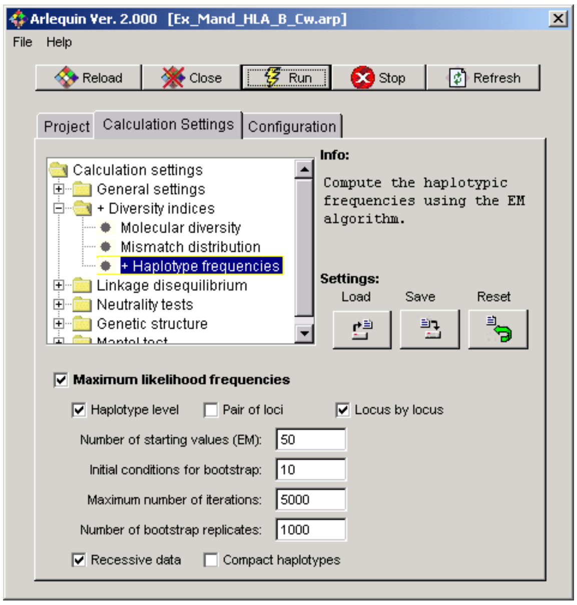
(data updated from Sanchez-Mazas et al. 2000, TA 56:303-312)

This file is called a project by Arlequin. It should have the extension .arp (for arlequin project). In the Arlequin program, *open your project* by first *adding it to your list*, *selecting it* and *answering OK*. The project is loaded and a summary of your options is shown in the *Project* tab.

A new directory is automatically created with the same name as your file, where Arlequin will store the results. In case of any problem (for example, an invalid keyword in your input file), a file Arlequin_log explaining you the error is generated in the new directory.

Choosing calculation settings

- Open the *Diversity indices* directory in the tree structure, and select *Haplotype frequencies*. Click *maximum likelihood frequencies* and the option(s) you want (haplotype level, all pairs of loci and/or locus by locus) for estimating frequencies.
- *Bootstrap initial values* (e.g. 10) and *number of replicates* (set 1000 as a minimum) serve to compute the standard deviations of gene frequencies.
- **Click *<u>recessive data</u>*** (otherwise the calculation will ignore the presence of a blank allele despite your declaration of recessive data in the [PROFILE] section of your input file).
- For information about the other parameters, see Arlequin user's guide.



Info:

Compute the haplotypic frequencies using the EM algorithm.

☑ Maximum likelihood frequencies

☑ Haplotype level ☐ Pair of loci ☑ Locus by locus

Number of starting values (EM): 50

Initial conditions for bootstrap: 10

Maximum number of iterations: 5000

Number of bootstrap replicates: 1000

☑ Recessive data ☐ Compact haplotypes

Running the program

- Once your parameters are chosen, click the *Run* button.

Results file (html format, loaded with your usual browser):

```
//////////////////////////////////////////////////////////////////////
RUN NUMBER 1 (06/02/02 at 23:29:42)
//////////////////////////////////////////////////////////////////////

Project information:
--------------------
           NbSamples      =  1
           DataType       =  STANDARD
           GenotypicData  =  1
           GameticPhase   =  0
           RecessiveData  =  1

==============================
Settings used for Calculations
==============================

 General settings:
 -----------------
        Deletion Weight                = 1
        Transition Weight Weight       = 1
        Tranversion Weight Weight      = 1
        Epsilon Value                  = 1e-07
        Significant digits for output  = 5
        Use original haplotype definition
        Alllowed level of missing data = 0.05


  Active Tasks:
  -------------

    Haplotypic Frequency estimation:
    --------------------------------
        Make estimations at the:
            Haplotypic level
            Locus level
        Initial Conditions              = 50
        Maximum No. Of Iterations       = 5000
        Bootstrap Replicates            = 1000
        Initial Conditions for Bootstrap = 10
        Recessive Data
        Do not Compact Haplotypes


==============================================================================
== ANALYSES AT THE INTRA-POPULATION LEVEL
==============================================================================

===============================================================================
== Sample :           Mandenka
===============================================================================

=====================================
== Haplotype frequencies estimation : (Mandenka)
=====================================

Reference: Dempster, A., N. Laird and D. Rubin, 1977.
           Excoffier, L. and M. Slatkin. 1995.
           Lange, K., 1997.
           Weir, B. S., 1996.
No. of gene copies in sample                : 326
No. of random initial conditions for EM     : 50
No. of different maximum likelihoods found  : 4
Epsilon value for stopping iterations       : 1.000000e-07
Logarithm of the sample maximum-likelihood  : -1011.76
No. of bootstraps for generating s.d.'s      : 1000

-------------------------------------------
Maximum-likelihood haplotype frequencies :
-------------------------------------------

Total number of possible haplotypes      :            192
Minimum frequency to reach for output    :       1.00e-05

    #   Haplotype          Freq.      s.d.
    1     UNKNOWN       0.003067    0.002937           B*4101 Cw*17
    2     UNKNOWN       0.018405    0.007244           B*4102 Cw*17
    3     UNKNOWN       0.003067    0.003071           B*5001 Cw*0602
    4     UNKNOWN       0.024540    0.008569           B*5601 Cw*0102
   ...
   ...
   51     UNKNOWN       0.055215    0.012580           B8 Cw*03
   52     UNKNOWN       0.003933    0.003803           blank Cw*0202
   53     UNKNOWN       0.004718    0.004787           blank Cw*1601

Sum of all 192 haplotype frequencies : 1.000000

Sum of 53 listed frequencies : 0.999997

 ************************************************************

Allele frequencies :

(1000 bootstrap replicates)
Allele frequencies for the locus 1

No. of gene copies in sample                : 326
No. of random initial conditions for EM     : 50
No. of bootstrap replicates                 : 1000
No. of different maximum likelihoods found  : 8
Epsilon value for stopping iterations       : 1.00000e-07
Logarithm of the sample maximum-likelihood  : -842.238

-------------------------------------------
Maximum-likelihood haplotype frequencies :
-------------------------------------------

Total number of possible haplotypes      :             27
Minimum frequency to reach for output    :       1.00e-05
```

```
    #   Haplotype         Freq.      s.d.
    1     UNKNOWN       0.003067    0.003181           B*4101
    2     UNKNOWN       0.018405    0.007297           B*4102
    3     UNKNOWN       0.003067    0.002958           B*5001
    4     UNKNOWN       0.027607    0.009007           B*5601
   ...
   ...
   24     UNKNOWN       0.064417    0.013922           B7
   25     UNKNOWN       0.128834    0.018576           B70
   26     UNKNOWN       0.061350    0.013317           B8

Sum of all 27 haplotype frequencies : 1.000000

Sum of 26 listed frequencies : 1.000000


Allele frequencies for the locus 2

No. of gene copies in sample                : 326
No. of random initial conditions for EM     : 50
No. of bootstrap replicates                 : 1000
No. of different maximum likelihoods found  : 4
Epsilon value for stopping iterations       : 1.00000e-07
Logarithm of the sample maximum-likelihood  : -673.263

------------------------------------------
Maximum-likelihood haplotype frequencies :
------------------------------------------

Total number of possible haplotypes      :             17
Minimum frequency to reach for output    :       1.00e-05

    #   Haplotype         Freq.      s.d.
    1     UNKNOWN       0.028436    0.008832           Cw*0102
    2     UNKNOWN       0.068764    0.013631           Cw*0202
    3     UNKNOWN       0.155452    0.018724           Cw*03
    4     UNKNOWN       0.196915    0.020674           Cw*0401
   ...
   ...
   15     UNKNOWN       0.049080    0.011319           Cw*17
   16     UNKNOWN       0.025300    0.008836           Cw*18
   17     UNKNOWN       0.038399    0.015427           blank

Sum of all 17 haplotype frequencies : 1.000000

Sum of 17 listed frequencies : 1.000000

////////////////////////////////////////////////////////////////////
END OF RUN NUMBER 1 (06/02/02 at 23:32:44))
Total computing time for this run : 0h 3m 1s 691 ms
////////////////////////////////////////////////////////////////////
```

Ignore the term *UNKNOWN* given with the list of alleles or haplotypes (haplotypes are defined in other applications).

2. Testing Hardy-Weinberg equilibrium

a) Recommended algorithm: goodness-of-fit (χ^2 or likelihood-ratio G test)

We show below an application with HWHLA (at http://geneva.unige.ch/hla_data_analysis/)

Warning: using Arlequin's exact test is **forbidden** in the presence of a blank allele.

Input files:

- Observed phenotypic distribution

  Note that a phenotype [A1] should be written A1,A1 as if it was a genotype, even though a blank allele is taken into account.

```
1    dqb1*0601,dqb1*0301
2    dqb1*0301,dqb1*0302
3    dqb1*0501,dqb1*0301
4    dqb1*0603,dqb1*0301
5    …
6    …
```

- Estimated allelic frequencies (e.g. Arlequin result)
  Warning: the blank allele must be written in the first line, even if its frequency is zero

```
blank         0.000000
dqb1*02       0.182292
dqb1*0301     0.526041
dqb1*0302     0.041667
dqb1*0401     0.007812
…
…
```

Results file:

```
Genotype                  #Obs     Exp        Qui2cont
dqb1*02-dqb1*02           #6       6.3802     0.022660
dqb1*02-dqb1*0301         #42      36.8229    0.727861
dqb1*02-dqb1*0302         #1       2.9167     1.259549
dqb1*02-dqb1*0401         #0       0.5468     0.546841
…
…
dqb1*0603-dqb1*0604       #0       0.0938     0.093755
dqb1*0604-dqb1*0604       #0       0.1055     0.105473

Total Q2                  74.85241
DF(collapsed)             65
Pvalue                    0.1889536
```

The result is not significant here (P > 0.05), meaning that the null hypothesis of Hardy-Weinberg equilibrium is not rejected.

Second example of results file:

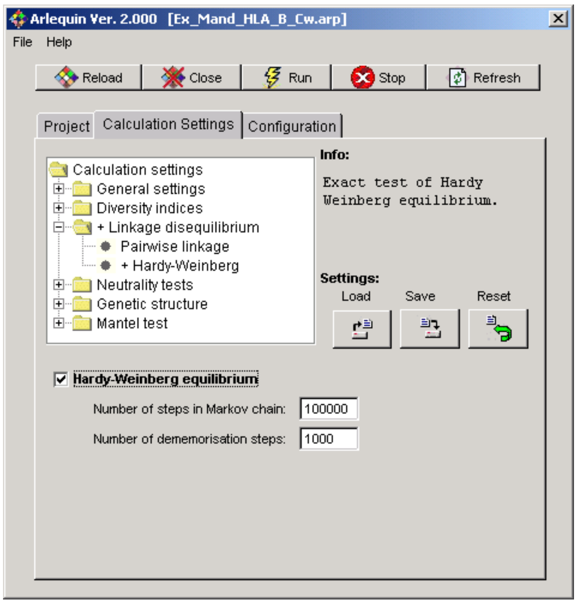
```
Genotype                  #Obs     Exp        Qui2cont
…
…
drb1*07-drb1*1302         #0       1.2864     1.286448
drb1*07-drb1*1303         #2       0.8125     1.735597
drb1*07-drb1*1304         #10      6.4841     1.90643
drb1*07-drb1*1312         #0       0.0677     0.067704
drb1*07-drb1*1401         #2       0.2031     15.896776
drb1*07-drb1*1601         #0       0.8125     0.812496
drb1*0801-drb1*0801       #1       0.3639     1.111706
drb1*0801-drb1*0802       #0       0.3729     0.372876
drb1*0801-drb1*0803       #0       0.0374     0.037418
drb1*0801-drb1*0804       #0       0.131      0.130969
…
…
Total Q2                  370.0592
DF(collapsed)             300
Pvalue                    0.003568546
```

In this case, the result is significant (P < 0.01), but this is only due to phenotypes with expected values lower than 5 individuals (like the phenotype in bold). No decision can then be taken about Hardy-Weinberg equilibrium.

b) Possible algorithm in the **strict absence** of a blank allele: exact test

We show below an application with Arlequin:

Input file as before (see section 1.) but set RecessiveData to zero in the [PROFILE] section.
In *Calculation settings*, open *Linkage disequilibrium* directory, and then set *Hardy-Weinberg* options.



Info:

Exact test of Hardy Weinberg equilibrium.

☑ Hardy-Weinberg equilibrium

Number of steps in Markov chain: 100000

Number of dememorisation steps: 1000

- Results file (html format) is loaded with your usual browser:

```
////////////////////////////////////////////////////////////////
RUN NUMBER 1 (07/02/02 at 15:04:48)
////////////////////////////////////////////////////////////////

   Project information:
--------------------
        NbSamples      =  1
        DataType       =  STANDARD
        GenotypicData  =  1
        GameticPhase   =  0
        RecessiveData  =  0

==============================
Settings used for Calculations
==============================

 General settings:
 -----------------
        Deletion Weight                = 1
        Transition Weight Weight       = 1
        Tranversion Weight Weight      = 1
        Epsilon Value                  = 1e-07
        Significant digits for output  = 5
        Use original haplotype definition
        Alllowed level of missing data = 0.05

 Active Tasks:
 -------------

    Hardy-Weinberg equilibrium test:
    --------------------------------
        No. of steps in Markov chain      = 100000
        No. of Dememorisation Steps       = 1000
        Required precision on Probability = 0
        Significance level                = 2
        Test association at the Locus level

==============================================================
== ANALYSES AT THE INTRA-POPULATION LEVEL
==============================================================
   ==============================================================
== Sample :        Suisses
==============================================================

===============================
== Hardy-Weinberg equilibrium : ( Suisses)
===============================

Reference: Guo, S. and Thompson, E. 1992. Levene H. (1949).  Exact test using a
Markov chain (for all Loci):
Forecasted chain length     :100000
```

```
Dememorization steps        :1000

Locus  #Genot     Obs.Heter.   Exp.Heter.  P. value     s.d.  Steps done
    1      80       0.83750      0.91006    0.03869  0.00037      100172


//////////////////////////////////////////////////////////////
END OF RUN NUMBER 1 (07/02/02 at 15:04:49))
Total computing time for this run : 0h 0m 1s 222 ms
//////////////////////////////////////////////////////////////
```

The result is here significant at the 5% level (P value < 0.05).

If your data include a recessive allele (blank), Arlequin will not perform the exact test. You may ignore the presence of a blank, but the result may be erroneous.

To avoid errors with Arlequin, do not forget to *reset* all *calculation settings* between different applications (for example if you change the recessive data option and run the program again).

3. Testing linkage disequilibrium*

*In fact, what one tests is always the null hypothesis of linkage equilibrium. A significant result (e.g. P value < 0.01) means that equilibrium is rejected, and hence that haplotypes or loci are in linkage disequilibrium.

a) Individual level (linkage disequilibrium coefficients for haplotypes)

Recommended algorithm: goodness-of-fit (χ^2 or likelihood-ratio G test)

We show below an application with LDHLA (at http://geneva.unige.ch/hla_data_analysis/)

Input files:

- Estimated haplotype frequencies (e.g. Arlequin results):

```
372
drb1*0102    dqb1*0501             0.005376
drb1*0301    dqb1*02               0.095227
drb1*0301    dqb1*0402             0.002688
drb1*0301    blank                 0.006924
…
…
drb1*1401    dqb1*0503             0.008065
drb1*1601    dqb1*0501             0.002688
drb1*1601    dqb1*0502             0.02957
blank        dqb1*0301             0.013073
blank        dqb1*0302             0.006107
blank        dqb1*0601             0.005376
```

- Estimated allelic frequencies (e.g. Arlequin result), one file per locus:

```
blank          0.034255
drb1*0102      0.005376
drb1*0301      0.100479
drb1*0302      0.024194
drb1*0403      0.002688
drb1*0405      0.008357
…
…
```

```
blank          0.000000
dqb1*02        0.182292
dqb1*0301      0.526041
dqb1*0302      0.041667
dqb1*0401      0.007812
dqb1*0402      0.015625
…
…
```

Results file

(here only χ2 results are shown but the program also performs a likelihood-ratio G test):

```
Nb Tests:    k=300

Sign.level (α)     |    Sign. Level (α’) after Bonferroni’s correction
0.05               |    1.67E-04
0.01               |    3.33E-05

                            Allele Freqs.         Haplotype freqs.     Linkage dis.         Significance
Locus 1      Locus 2          1        2           Obs.      Exp.        D         D'         X2       P value
…
drb1*1301    dqb1*0602      0.0430   0.0618      0.0027   0.0027      0.0000   0.0007       0.0001   0.9914
drb1*1301    dqb1*0603      0.0430   0.0108      0.0108   0.0005      0.0103   1.0000       85.2000  0.0000
drb1*1301    dqb1*0604      0.0430   0.0242      0.0056   0.0010      0.0045   0.1950       7.3100   0.0069
drb1*1301    blank          0.0430   0.0000      0.0035   0.0000      0.0035   nan          nan      nan
drb1*1302    dqb1*02        0.0484   0.1880      0.0000   0.0091     -0.0091  -1.0000       3.3900   0.0657
drb1*1302    dqb1*0301      0.0484   0.5190      0.0029   0.0251     -0.0222  -0.8840       7.3000   0.0069
…
…
Nb haplotypes Freq > 3%           8
Nb haplotypes Natt > 5            17
```

b) Global level (linkage disequilibrium between loci pairs)

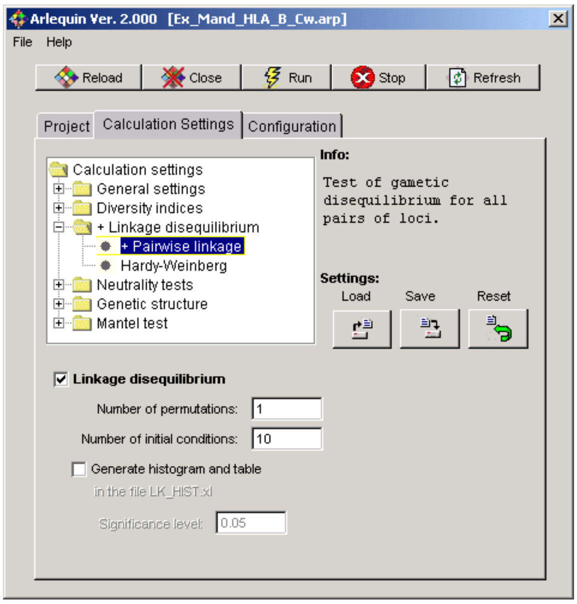
- Use goodness-of-fit tests as before (e.g. program LDHLA), but correct the level of significance by Bonferroni’s correction*. If at least one result is found significant after Bonferroni’s correction, the global test may be considered as significant.

  Warning: do not reject the null hypothesis if the P value is significant due to genotypes exhibiting expected frequencies lower than 5.

  *Bonferroni’s correction must be applied when many independent tests are performed simultaneously. The new level of significance (α’ instead of α) is given by $\alpha' = 1-(1-\alpha)^{\frac{1}{k}}$ which is approximately similar to $\alpha' = \frac{\alpha}{k}$ for a total of k tests.

- The global likelihood-ratio G test can also be performed with Arlequin

  Warning: as using Arlequin’s permutation procedure on the likelihood-ratio is **forbidden** in the presence of a blank allele, it is necessary to **set the number of permutations to 1.**



Info:

Test of gametic disequilibrium for all pairs of loci.

☑ Linkage disequilibrium

Number of permutations: 1

Number of initial conditions: 10

□ Generate histogram and table in the file LK_HIST.xl

Significance level: 0.05

4. Testing independence of allelic products (or antigens) of two different loci

Typical *contingency (or 2x2) table*:

	dqb1*0602	dqb1*other	
drb1*1304	9	75	Σ=84
drb1*other	14	88	Σ=102
	Σ=23	Σ=163	N=186

Recommended algorithm: Fisher exact test

Warning: this test is not easily done manually.

We show below an application with FIHLA (at http://geneva.unige.ch/hla_data_analysis/)

This programs tests all possible associations at two loci.

Input file:

- Observed phenotypic distribution at two loci:

```
drb1*1304,drb1*1304                dqb1*0601,dqb1*0301
drb1*1304,drb1*1304                dqb1*0301,dqb1*0302
drb1*1301,drb1*1302                dqb1*0501,dqb1*0301
drb1*1102,drb1*1301                dqb1*0603,dqb1*0301
drb1*1102,drb1*1104                dqb1*0602,dqb1*0301
drb1*0804,drb1*0806                dqb1*0602,dqb1*0301
drb1*1302,drb1*1601                dqb1*0501,dqb1*0502
drb1*07,drb1*1303                  dqb1*02,dqb1*0301
drb1*0301,drb1*1601                dqb1*0502,dqb1*02
drb1*1301,drb1*1302                dqb1*0501,dqb1*0301
drb1*1101,drb1*1303                dqb1*0602,dqb1*0301
…
…
```

Results file

```
Locus 1       Locus 2        YY     YN     NY     NN                 Pvalue
drb1*1304     dqb1*0601      1      83     1      101                1.0000
drb1*1304     dqb1*0301      81     3      70     32                 0.0000
drb1*1304     dqb1*0302      9      75     5      97                 0.1669
drb1*1304     dqb1*0501      12     72     20     82                 0.4355
drb1*1304     dqb1*0603      2      82     2      100                1.0000
drb1*1304     dqb1*0602      9      75     14     88                 0.6559
drb1*1304     dqb1*0502      1      83     10     92                 0.0133
drb1*1304     dqb1*02        24     60     40     62                 0.1628
drb1*1304     dqb1*0401      0      84     3      99                 0.2530
drb1*1304     dqb1*0503      0      84     3      99                 0.2530
drb1*1304     dqb1*0402      3      81     3      99                 1.0000
drb1*1304     dqb1*0604      1      83     8      94                 0.0423
drb1*1301     dqb1*0601      0      16     2      168                1.0000
drb1*1301     dqb1*0301      14     2      137    33                 0.7404
drb1*1301     dqb1*0302      1      15     13     157                1.0000
drb1*1301     dqb1*0501      2      14     30     140                1.0000
drb1*1301     dqb1*0603      4      12     0      170                0.0000
drb1*1301     dqb1*0602      1      15     22     148                0.6976
drb1*1301     dqb1*0502      0      16     11     159                0.6027
drb1*1301     dqb1*02        3      13     61     109                0.2701
drb1*1301     dqb1*0401      0      16     3      167                1.0000
…
…
```

The association in bold (not significant) corresponds to the example of contingency table given above. For another pair, like drb1*1304-dqb1*0301, the alleles are found together more often than by chance in individuals (P=0.0000). This may be due both to their linkage disequilibrium on the same chromosome (cis position), and to an association in trans position, due, for example, to a selective effect favouring the co-occurrence of their allelic products.

5. Testing disease associations

Typical *contingency (or 2x2) table*:

	dqb1*0602	dqb1*other	
patients	75	9	Σ=84
controls	14	88	Σ=102
	Σ=23	Σ=163	N=186

The Fisher exact test is also recommended to test the association of a given allele to a given disease. A global test may also be performed to assess whether a given locus is globally associated to a given disease. After multiple Fisher exact tests, apply Bonferroni's correction and conclude as in 3.b.

6. Comparing different populations

Possible procedure: Fst genetic distance and test of significance

We show below an example with Arlequin:

The example below concerns the case where only relative allele frequency data are available for one or more populations. The [PROFILE] section must then include the options Datatype=FREQUENCY and frequency=REL (for relative). If phenotypic data are available for all populations, use data file format as in chapter 1.

Input file including estimated allele frequencies (e.g. Arlequin output)

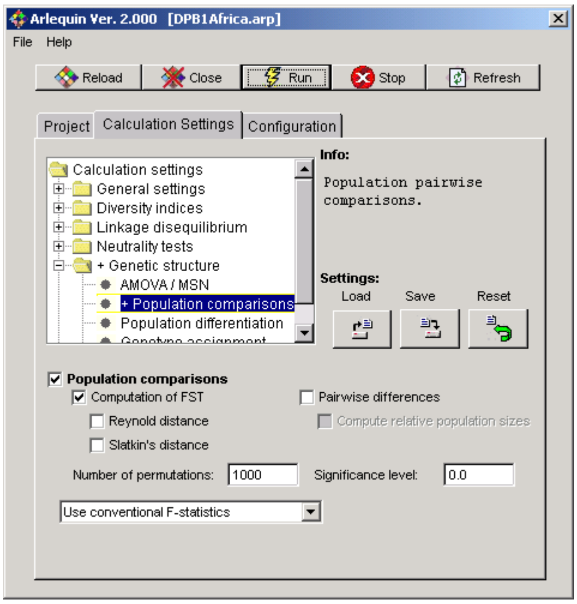
```
[Profile]

        Title="HLA-DPB1"
        NbSamples=2
        DataType=FREQUENCY
        GenotypicData=0
       Frequency=REL

[Data]

        [[Samples]]

          SampleName="Mandenka"
          SampleSize=197
          SampleData={
               DP0101  0.1350
               DP0201  0.1790
               DP0202  0.0000
               DP0301  0.0260
               ...
               ...
               DP5001  0.0080
               DP5101  0.0000
               DP5501  0.0000
               DPX     0.0000
               }
          SampleName="Aka"
          SampleSize=93
          SampleData={
               DP0101  0.0490
               DP0201  0.0300
               DP0202  0.0000
               ...
               ...
               DP5101  0.0000
```

```
                 DP5501   0.0000
                 DPX      0.0040
                 }
```

(data from Renquin et al. 2001, TA 58:211-222)

Choosing calculation settings

- Open the *Genetic structure* directory in the tree structure, and select *Population comparisons*. Click *Computation of Fst.*
- Choose the number of permutations for the statistical test (here 1000).



Info:
Population pairwise comparisons.

☑ Population comparisons
☑ Computation of FST
☐ Reynold distance
☐ Slatkin's distance
☐ Pairwise differences
☐ Compute relative population sizes
Number of permutations: 1000
Significance level: 0.0
Use conventional F-statistics

Warning: using Arlequin’s *exact test* for population differentiation is **forbidden** in the presence of a blank allele.

Results file (html format, loaded with your usual browser):

```
  ////////////////////////////////////////////////////////////////////
RUN NUMBER 1 (11/02/02 at 12:49:53)
////////////////////////////////////////////////////////////////////

   Project information:
--------------------
           NbSamples      =  2
           DataType       =  FREQUENCY
           GenotypicData  =  0

==============================
Settings used for Calculations
==============================

 General settings:
 -----------------
        Deletion Weight                = 1
        Transition Weight Weight       = 1
        Tranversion Weight Weight      = 1
        Epsilon Value                  = 1e-07
        Significant digits for output  = 5
        Use original haplotype definition
        Alllowed level of missing data = 0.05

 Active Tasks:
 -------------

    Population pairwise Fst values:
    ------------------------------
        No. of permutations for significance = 1000
        No. of permutations for Mantel test  = 1000
```

```
       Distance matrix:
            Compute F-statistics on haplotype frequencies only

  ===============================================================================
== GENETIC STRUCTURE ANALYSIS
===============================================================================

   ======================================================================================
== Comparisons of pairs of population samples
======================================================================================

List of labels for population samples used below:
-------------------------------------------------

Label      Population name
-----      ---------------
  1:       Mandenka
  2:       Aka

------------------------
Population pairwise FSTs
------------------------

Computing conventional F-Statistics from haplotype frequencies
                     1         2
           1   0.00000
           2   0.25001   0.00000

------------
FST P values
------------

Number of permutations : 1023

                       1                 2
           1           *
           2   0.00000+-0.0000           *

------------
Matrix of significant Fst P values
Significance Level=0.0500
------------

Number of permutations : 1023

                     1         2
           1                   +
           2         +

////////////////////////////////////////////////////////////////////
END OF RUN NUMBER 1 (11/02/02 at 12:49:53))
Total computing time for this run : 0h 0m 0s 421 ms
////////////////////////////////////////////////////////////////////
```

The two populations are significantly different. This is indicated by a “+” (5% level here).

Acknowledgments

I thank the Swiss National Science Foundation (SNF grant #3100A0-112651).

.