# Computer Note

## AUTOTET: A Program for Analysis of Autotetraploid Genotypic Data

**P. H. Thrall and A. Young**

AUTOTET is a program written in ANSI-C that calculates allele frequencies, and estimates genetic diversity and fixation statistics for autotetraploids based on codominant genotypic data. In general, genetic analysis of autotetraploid data is complicated by two main factors. The first is the existence of multiple heterozygous states: three partial heterozygotes (AAAB, AABB, AABC) and the full heterozygote (ABCD). In particular, estimation of expected heterozygosity ($H_e$) and the fixation coefficient ($F$) is burdensome under a tetraploid model, where even relatively low numbers of alleles at a locus can result in very large numbers of possible genotypic classes for which relative frequencies must be calculated based on population gene frequencies.

A second problem for analysis of heterozygotes is that at meiosis, individual loci can undergo either random chromosomal segregation (RceS) or random chromatid segregation (RcdS) (Bever and Felber 1992; Geiringer 1949; Haldane 1930). Under RceS, and assuming random mating, expected genotype frequencies can be estimated directly from the products of population gene frequencies as for diploids. However, under RcdS, sister chromatids can segregate into the same gamete, resulting in double reduction. In this case, equilibrium heterozygosity will be lower than expected under chromosomal segregation due to gametic disequilibrium, and genotype frequencies must be estimated on the basis of gametic frequencies (Geiringer 1949). The probability of double reduction occurring at a locus ($\alpha$) depends on the amount of quadrivalent formation and the proximity of the locus to the centromere, and may range from $\alpha = 0$ to a theoretical maximum of $\alpha = 1/7$ (Wricke and Weber 1986).

This intractability has led to the common use of approximations of these parameters where all heterozygotes are treated equally, chromosomal segregation is assumed, and $H_e$ is calculated as for a diploid system:

$$1 - \sum_{i=1}^{s} p_i^4, \qquad (1)$$

where $s$ is the number of alleles at a locus, and $p_i$ is the frequency of the $i$th allele. Such an approach is unsatisfactory because heterozygosity estimates and fixation indices (when calculated as above) will respond only to changes in the number of full homozygotes. This genotypic class is considerably less responsive to inbreeding than the same class in true diploids. This makes assessment of interpopulation differences in heterozygosity and gene fixation excessively conservative. It also prevents meaningful comparative analysis of diversity between tetraploid and diploid populations within species.

AUTOTET allows rigorous analysis of heterzygosity statistics, with partial heterozygotes weighted by 1 minus the likelihood of any two alleles being identical by descent, and heterozygosity and fixation indices calculated under both full chromosomal and full chromatid segregation following Geiringer (1949) and Li (1955).

### Input

Each input file should contain data for a single population. It can be either a space or tab delimited ASCII file with rows as individuals and columns as loci. The first line of the input file must consist of the names of the loci (up to five characters in length), separated by either spaces or tabs. Subsequent rows should consist of individual four-allele genotypes. Alleles are coded as letters and, because the program is case sensitive, up to 52 different alleles can be allocated. Missing or incomplete genotypes for a particular locus (i.e., those with less than four alleles) must be coded as four dots ('. . . .'); these will be excluded from analyses for that locus. The program uses dynamic memory allocation to construct arrays, therefore there are no fixed requirements with respect to the number of individuals in the dataset to be analyzed (currently AUTOTET assumes a maximum of 30 loci per population, but this can be easily modified).

### Output

Results of the analysis are written to a separate file specified by the user. Output consists of a list of alleles and their frequencies, followed by a table of diversity statistics for each locus, their averages across loci, standard deviations, and the sample sizes upon which these are based (Table 1). Chi-square goodness-of-fit tests for observed-to-expected genotype frequencies are then given. The statistics calculated are

1. Allelic richness ($A$)—the number of alleles at a locus.
2. Genotypic richness ($G$)—the number of four allele genotypes at a locus.
3. Allelic richness within individuals ($A_i$)—the average number of alleles per individual at a locus.
4. Observed heterozygosity ($H_o$)—the observed heterozygosity at a locus, with the five possible classes of genotypes being weighted inversely to the probability of any two of their alleles being identical by descent, such that the heterozygosity weights used are AAAA = 0, AAAB = 0.5, AABB = 0.667, AABC = 0.833, and ABCD = 1 (see Bever and Felber 1992).
5. Expected heterozygosity [$H_e(Ce)$]—the heterozygosity expected under ran-

**Table 1. Sample output file from AUTOTET**

Input file: mcc.txt Date: 16-6-1998 Time: 16:25

I. Allele frequencies for individual loci

| Locus | Allele | Frequency |
|---|---|---|
| *gpi1* | b | 0.034314 |
| *gpi1* | f | 0.088235 |
| *gpi1* | g | 0.303922 |
| *gpi1* | h | 0.044118 |
| *gpi1* | l | 0.004902 |
| *gpi1* | m | 0.495098 |
| *gpi1* | s | 0.029412 |
| *mdr1* | m | 0.750000 |
| *mdr1* | s | 0.250000 |
| *got1* | f | 0.151042 |
| *got1* | m | 0.828125 |
| *got1* | s | 0.020833 |
| *got2* | m | 0.848958 |
| *got2* | p | 0.005208 |
| *got2* | s | 0.145833 |
| *got3* | b | 0.024510 |
| *got3* | f | 0.191176 |
| *got3* | h | 0.053922 |
| *got3* | m | 0.715686 |
| *got3* | s | 0.014706 |

II. Population genetic statistics

| | | | | | | Chromosome | | Chromatid | |
|---|---|---|---|---|---|---|---|---|---|
| Locus | *N* | *A* | *AI* | *G* | $H_o$ | $H_e$ | *F* | $H_e$ | *F* |
| *gpi1* | 51 | 7 | 2.431 | 23 | 0.663 | 0.651 | −0.019 | 0.607 | −0.092 |
| *mdr1* | 51 | 2 | 1.706 | 4 | 0.389 | 0.375 | −0.037 | 0.350 | −0.111 |
| *got1* | 48 | 3 | 1.521 | 6 | 0.271 | 0.291 | 0.069 | 0.272 | 0.003 |
| *got2* | 48 | 3 | 1.458 | 4 | 0.253 | 0.258 | 0.017 | 0.241 | −0.053 |
| *got3* | 51 | 5 | 1.804 | 10 | 0.441 | 0.448 | 0.014 | 0.418 | −0.056 |
| Means: | 49.800 | 4.000 | 1.784 | 9.400 | 0.404 | 0.404 | 0.002 | 0.377 | −0.069 |
| SD: | 1.643 | 2.000 | 0.388 | 7.987 | 0.165 | 0.156 | 0.014 | 0.146 | 0.022 |

III. Goodness-of-fit (observed-expected genotype frequencies)

| | Chromosome | | Chromatid | |
|---|---|---|---|---|
| Locus | $\chi^2$ | *P* | $\chi^2$ | *P* |
| *gpi1* | 2.95 | >.05* | 4.45 | >.05* |
| *mdr1* | 0.16 | >.05 | 1.40 | >.05 |
| *got1* | 0.34 | >.05 | 0.86 | >.05 |
| *got2* | 2.50 | >.05 | 0.91 | >.05 |
| *got3* | 14.60 | <.01 | 9.69 | <.05 |

Data are from the autotetraploid *Rutidosis leptorrhynchoides* which undergoes significant quadrivalent formation. For chromatid segregation, $\alpha = 0.143$.

* Warning: one or more genotypic classes have expected values < 5.0.

dom mating, and assuming random chromosomal segregation (i.e., no double reduction). Expected population genotypic frequencies are AAAA = $p^4$, AAAB = $4p^3q$, AABB = $6p^2q^2$, AABC = $12p^2qr$, ABCD = $24pqrs$, and partial heterozygotes are weighted as for observed heterozygosity.

6. Expected heterozygosity [$H_e(Cd)$]—the heterozygosity expected under random mating, and assuming some level of chromatid segregation. The default calculation assumes maximum double reduction where $\alpha = 1/7$, but other values of $\alpha$ can also be input directly by the user. Expected population genotypic frequencies are AAAA = $g_{11}^2$, AAAB = $4g_{11}g_{12}$, AABB = $2(g_{11}g_{22} + 2g_{12}^2)$, AABC = $4(g_{11}g_{23} + 2g_{12}g_{13})$, ABCD = $8(g_{12}g_{34} + g_{13}g_{24} + g_{14}g_{23})$, and partial heterozygotes are weighted as for observed heterozygosity. The $g_{ij}$ are gamete frequencies, such that $g_{11} = p^2 + [3\alpha/(2 + \alpha)]p(1 - p)$, $g_{12} = pq(4 - 4\alpha)/(2 + \alpha)$, and so on for the other possible gamete types (Geiringer 1949; Wricke and Weber 1986; Bever and Felber 1992).

6. Fixation coefficients [$F(Ce)$ and $F(Cd)$]—under both chromosomal and chromatid segregation calculated as $1 - (H_o/H_e)$. Note that the mean and standard deviation for *F* in both cases are weighted by the expected heterozygosities.

## Computer Requirements

AUTOTET runs on any IBM compatible computer under MS-DOS. There will be an exponential increase in the number of possible four-allele genotypes as the number of alleles increases; the general formula for a polyploid (Elandt-Johnson 1971) is given by

$$\binom{s + 2m - 1}{2m}, \qquad (2)$$

where *s* is as defined in equation 1 above and *m* is the number of homologous chromosomes used to form a gamete. Thus the time the program takes to analyze a dataset depends primarily on the number of alleles at a locus, as it has to calculate genotypic frequencies for all possible combinations of alleles. This will generally be quite rapid when the number of alleles at a locus is 10 or less. A dataset with an unusually high level of allelic diversity [100 individuals, 7 loci with an average of 7 alleles per locus and a maximum of 10 alleles at two of the loci (a total of 715 possible genotypes at these loci)] took 5.5 minutes to run on a 66 Mhz 486 laptop, while on a 200 Mhz Pentium the analysis took just over a minute.

A copy of the program and source code may be obtained from the authors by sending a DOS-formatted 3.5 in. high-density diskette to the address below, or it can be downloaded directly from the Centre for Plant Biodiversity Research and Australian National Herbarium homepage on the Internet at http://www.anbg.gov.au/cpbr/cpbr.html.

From the Centre for Plant Biodiversity Research, CSIRO Division of Plant Industry, GPO Box 1600 Canberra, ACT 2601, Australia. Address correspondence to P. H. Thrall at the address above. P.H.T. acknowledges the support of a Queen Elizabeth II Fellowship. A.H.D. Brown provided useful comments on an earlier draft of this manuscript.


**References**

Bever JD and Felber F, 1992. The theoretical population genetics of autopolyploidy. In: Oxford surveys in evolutionary biology (Antonovics J and Futuyma D, eds). New York: Oxford University Press; 185–217.

Elandt-Johnson RC, 1971. *Probability models and statistical methods in genetics.* New York: John Wiley & Sons.

Geiringer H, 1949. Chromatid segregation of tetraploids and hexaploids. *Genetics*, 34:665–684.

Haldane JBS, 1930. Theoretical genetics of autopolyploids. *J Genet* 22:359–372.

Li CC, 1955. *Population genetics.* Chicago: University of Chicago Press.

Wricke G and Weber WE, 1986. *Quantitative genetics and selection in plant breeding.* Berlin: deGruyter.

Received August 9, 1999
Accepted March 10, 2000

Corresponding Editor: Robert Angus