H A P L O G E N - User's Manual

Qualitative inheritance analysis of zymograms and DNA electropherograms in haploid gametophytes

Developed for free distribution by: Elizabeth M. Gillet
Abt. Forstgenetik, Universitaet Goettingen
Buesgenweg 2, D-37077 Goettingen, Germany

Described under original name HAPLOZYM in: Gillet, EM. "Qualitative inheritance analysis of isoenzymes in haploid gametophytes: Principles and a computerized method". Silvae Genetica 45, 1996, 8-16.

June 1997, rev. October 1998

Given the banding patterns of the zymograms or DNA electropherograms of a genetically closed sample of gametophytes, <HAPLOGEN> systematically generates all hypotheses for the mode of inheritance of these patterns that conform to certain qualitative rules for the genetic interpretation of single bands. These rules follow from formulation of the concept of << TRANSMISSION HOMOLOGY >> within single individuals and sets of individuals (Gillet 1996).

This manual is published and <HAPLOGEN> is available on the internet under URL:
http://www.uni-forst.gwdg.de/forst/fg/index.htm

CONTENTS:

1. INHERITANCE ANALYSIS

Gametophytes represent the haplophase generation in sexually reproducing organisms. The haplophase genotype (termed << HAPLOTYPE >> in diploid organisms) of each gametophyte is identical to that of the meiospore from which it developed by mitotic division. Thus the different banding patterns expressed by an individual's gametophytes reflect the meiotic segregation of << ALLELES >> (gene variants at a single locus) at loci at which the individual is heterozygous.

Even if an individual's own banding pattern is not known, inferences drawn from the relationships between bands among its gametophytes allow the << INHERITANCE ANALYSIS >> of the banding patterns. Inheritance analysis is performed in order to determine the << MODE OF INHERITANCE >> of the banding patterns, the two components of which are the

(1) << MODE OF TRANSMISSION >> : number of loci, identification of the alleles at each locus, and the

(2) << MODE OF GENE ACTION >> : intra- and interlocus interactions between alleles (dominance, codominance, epistasis).

In plants, the following tissues possess only the genetic information of a gametophyte and may be accessible for isoenzyme or DNA analysis: the primary endosperm of conifer seeds (macrogametophyte), single pollen or egg cells (requiring PCR methods), and haploid or double-haploid plants. In animals, single egg and sperm cells may be analyzable by PCR methods.

2. SAMPLING OF GAMETOPHYTES

GENETIC CLOSURE

<HAPLOGEN> ideally requires as input the banding patterns of a genetically closed sample of gametophytes, which is explained as follows. Assuming complete genetic control of the banding patterns, each pattern is the expression of the gametophyte's haplotype at the controlling loci. A sample of gametophytes is defined to be << GENETICALLY CLOSED >>, if it contains all possible haplotypes that can result as interlocus combinations of the alleles in the sample. The genetic closure of a sample can only be judged retrospectively, i.e., after inheritance analysis has been successful in identifying loci and alleles. Nevertheless, sampling strategies can be devised to increase the chances of obtaining a genetically closed sample.

SUFFICIENT SAMPLE SIZE

Given a desired probability for genetic closure of a sample, the sufficient sample size is a function of the number of haplotypes and the frequency of the rarest haplotype among the total gametophytes from which a random sample is drawn. If these can be estimated, the sufficient (or minimum) sample size of gametophytes required to ensure with the given probability that all haplotypes are detected can be calculated after Gregorius (1980, considering that the sampling of single haplotypes as the primary endosperm of conifer seeds, as single ovules, or as pollen grains is equivalent to sampling alleles in homozygous genotypes). The sufficient sample size increases for decreasing minimum haplotype frequency.

SAMPLING GAMETOPHYTES OF A SINGLE INDIVIDUAL

Genetic closure is easiest to achieve by sampling gametophytes of a single individual. The number of haplotypes and the expected frequency of the rarest haplotype depend on several unknown quantities: the number of loci at which the individual is heterozygous, the segregation proportions at each of these loci, and the recombination frequencies between these loci. For k heterozygous loci, the frequency of the rarest haplotype is maximal, if segregation at each locus is regular (1:1) and alleles between loci are randomly associated. In this case, the 2^k possible haplotypes are expected to be uniformly distributed, i.e., all have expected frequencies equal to 1/(2^k). In practice,

TABLE 1: For sampling of haplotypes produced by a single parent individual, minimum sample size to ensure a given probability of genetic closure of the sample is given under the following assumptions:
(1) the parent is heterozygous at m of the loci that control the banding pattern,
(2) segregation of the alleles at each of the m loci is regular (1:1), and
(3) the alleles at the different loci show stochastic independence (no linkage).

Parent heterozygous at m loci	Total number 2^m of haplotypes of equal frequencies (frequency=1/2^m)	Minimum sample size such that probability of detection of all haplotypes is greater than 95%	99%	99.9%
1	2 (0.500000)	6	8	11
2	4 (0.250000)	16	21	29
3	8 (0.125000)	38	51	68
4	16 (0.062500)	90	115	150
5	32 (0.031250)	203	255	327
6	64 (0.015625)	453	557	703

the expected frequency must be estimated by assuming limits on segregation distortion and recombination fractions based on information gained from other systems. In general, the wider these limits are allowed to be, the larger will be the sufficient sample size.

If regular segregation and random association can be assumed for the alleles at all loci controlling the banding patterns in the parent individual, then Table 1 gives sufficient sample sizes to ensure a given probability of genetic closure of a sample of the individual's gametophytes.

SAMPLING GAMETOPHYTES FROM SEVERAL INDIVIDUALS

A sample consisting of gametophytes of several individuals will rarely be genetically closed. In this case it is recommended to run <HAPLOGEN> on the gametophytes of each individual separately but using a common system of numbering of band positions. In such a system, band position n in one sample will also be labelled as position n in all other samples, even if for some individuals no band appears at this position in any of its gametophytes. Often the hypotheses for individual trees can be combined, allowing inference of homology of bands between individuals. An example is given in Bergmann & Gillet (1996).

SAMPLING GAMETOPHYTES IN BULK COLLECTIONS FROM POPULATIONS

Bulk collections of gametophytes from large natural populations may be genetically closed, but only if all possible multilocus genotypes are represented in the gametophyte-producing population. In addition to the above factors determining haplotype frequency among gametophytes of single individuals, minimum expected haplotype frequency for bulk sampling also depends on the frequency distribution of multilocus genotypes in the parental populations, the gametic phases in linkage groups in each parent, the individual gamete production (fecundity), gametic selection, and when analyzing conifer endosperms on the individual fertilities.

If the bulk collection can be assumed to be genetically closed, and if it is possible to estimate the minimum frequency of all haplotypes in the collection, then Table 2 gives sufficient sample sizes to ensure a given probability of genetic closure.

SEQUENTIAL SAMPLING OF GAMETOPHYTES

Since sufficient sample sizes are rarely exactly calculable, a sequential sampling scheme among gametophytes with the potential for genetic closure (i.e., not among bulked gametophytes of only a few individuals) may be most appropriate: sampling continues until <HAPLOGEN> succeeds in finding a hypothesis.

GENETICALLY UN-CLOSED SAMPLES

Some samples of haplotypes, though not genetically closed, still provide enough information for a hypothesis to be formulated. <HAPLOGEN> then states how many haplotypes are missing. Continued sampling should find these also, if the hypothesis is correct and the underlying population of haplotypes is genetically closed.

TABLE 2: For sampling haplotypes in a base collection of gametophytes (e.g. bulk population sample), minimum sample size to ensure a given probability of detecting all haplotypes that are present at relative frequencies not less than a given minimum frequency is given. The number of haplotypes actually present in the base collection is assumed to equal 1/(minimum haplotype frequency).
Word of warning: A sample of sufficient size to detect all haplotypes can, however, only be genetically closed if the base collection itself is genetically closed. This need not be the case for gametophytes sampled from a bulk population harvest, for example, unless the adult trees can collectively produce all haplotypes that can be constructed from all of the alleles present in the population at all of the loci.

All haplotypes should be detected that have frequency not less than	Minimum sample size such that probability of detection of all such haplotypes is greater than		
	95%	99%	99.9%
0.500	6	8	11
0.400	7	10	14
0.300	11	15	22
0.200	21	28	39
0.100	51	66	88
0.090	57	74	99
0.080	65	84	112
0.070	77	99	131
0.060	92	119	156
0.050	117	149	194
0.040	152	192	249
0.030	212	265	341
0.020	341	422	536
0.010	754	916	1146
0.009	850	1030	1285
0.008	972	1174	1462

Reproduced with author's permission from Gregorius (1980).

3. ELEMENTARY ZONES

Given a sample of banding patterns, the path of migration of bands is divided into << ELEMENTARY ZONES >>, abbreviated << EZONE >> in <HAPLOGEN>, such that

(1) each elementary zone contains a band of at least one banding pattern;
(2) any two bands of different banding patterns that appear in the same elementary zone are considered to represent the "same" band (in general, identical isoenzymes or DNA fragments).

Qualitative inheritance analysis of the banding patterns consists in interpretation of the patterns of band appearance in the elementary zones.

For this purpose, elementary zones are classified into the following types:

An elementary zone is << FIXED >>, if a band appears in this zone in all of the patterns. The lack of variation in a fixed zone prohibits its interpretation.

A non-fixed elementary zone is << DEPENDENT >> on a second non-fixed elementary zone, if whenever a band appears in the one zone of any pattern, a band is also present in the second zone. A zone can be dependent on more than one zone (besides itself).

A non-fixed elementary zone i is << INDEPENDENT >>, if it is not dependent on any other elementary zone, i.e., if for each other non-fixed elementary zone j, there exists a banding pattern that exhibits a band in i but no band in j.

Two non-fixed elementary zones are << EQUIVALENT >>, if each zone is dependent on the other, i.e., if in every banding pattern bands appear either in both zones or in neither zone. The relation "equivalence", denoted "~", partitions the set of elementary zones into << EQUIVALENCE CLASSES >> of elementary zones, since it is reflexive (Z~Z), symmetric (Z~Y==>Y~Z), and transitive (Z~Y and Y~X ==> Z~X).

4. ZYMOGRAMS

Isoenzymes are defined as "electrophoretically separable variants of one enzyme ... system" (Bergmann et al. 1989). For isoenzyme banding patterns (zymograms), the development of a computer program for the formulation of hypotheses on the mode of inheritance is a complex task, due to the different ways in which isoenzymes expressed in haploid tissue correspond to genes at loci. Whereas each enzyme molecule of a monomeric enzyme system is the product of the gene at a single locus, polymeric enzymes are formed from two or more enzyme subunits, each of which is the product of the gene at a locus.

TYPES OF ISOENZYMES

Three types of enzyme molecule can occur in haploid tissue:

<< HOMOMERIC >> isoenzymes consist of subunits that are all encoded by genes of the same type at the same locus. Monomeric isoenzymes, which consist of only a single subunit, are treated as homomerics.

<< INTERLOCUS HETEROMERIC >> isoenzymes consist of subunits encoded by genes at two (or more) different loci. (Intralocus heteromeric isoenzymes cannot be formed in haploid tissue.)

<< POST-TRANSLATIONAL MODIFICATION (PTM) >> is an enzyme molecule, the electrostatic charge or molecular conformation of which is modified, probably by the product of a gene considered to belong to the "genetic background" (i.e. not coding for subunits of the enzyme system being studied). PTM can affect migration velocity through the gel. If not all molecules of a particular subunit structure in an individual are modified, PTM results in the appearance of one or more additional bands in the zymogram. Two types of PTM of molecules of a given subunit structure can be distinguished within a collection of individuals in its environment: A PTM of a particular molecule will be termed << FIXED >>, if the PTM occurs in all members possessing the molecule, and otherwise << FACULTATIVE >>.

INTERPRETATION OF BANDING PATTERNS

Specification of the mode of inheritance involves identification of the origin of each band as a molecule consisting of how many subunits coded by which alleles at which loci. In some cases, the absence of a band must be interpreted as the presence of a "null allele" (that produces a defective subunit) at some locus.

The strategy formulated by Gillet (1996) is to identify all elementary zones that contain homomeric isoenzymes and then to

partition these zones into disjoint sets such that each set represents a complete set of transmission homologous gene types, i.e., the set of all (non-null) alleles of a locus. Thus, the alleles present at each locus in the sample of banding patterns are represented by a set of elementary zones, and the allele present in any given banding pattern is revealed by the appearance of a band in one of these zones; if no band appears in any of these zones, a null allele is assumed.

>> Independence of non-fixed homomeric elementary zones:
   If the sample of banding patterns is genetically closed, then all non-fixed elementary zones representing homomerics are independent. With one rare exception (see below), only elementary zones representing homomerics are independent.

>> Dependence of interlocus heteromeric isoenzymes:
   The appearance of a band representing an interlocus heteromeric depends on the appearance of the two bands representing the corresponding homomerics. An exception is the case in which one of the genes is a null allele that produces the heteromeric but not the homomeric.

>> Dependence of PTM:
   Appearance of a band in an elementary zone representing a PTM is dependent on the appearance of the unmodified isoenzyme (as long as not all molecules are modified and the zone of the unmodified isoenzyme is not fixed). This dependence distinguishes the elementary zones of heteromerics and PTM's from those of homomerics.

>> Exceptional case:
   The single, probably rare case of an independent zone that is not homomeric is given by a facultative PTM of a fixed zone.

>> Identification of loci:
   Considering only the independent elementary zones, the homomeric elementary zones corresponding to the same locus are recognizable by the appearance of a band in exactly one of them in each zymogram (transmission homology for diploidy). If a set of homomeric zones cannot be made complete, then the existence of a (recessive) null allele at the locus must be postulated. However, care must be taken if a locus is found to comprise only one active allele and a null allele; an alternative explanation is that the "active allele" is a facultative PTM of an isoenzyme in the fixed zone, which, as mentioned above is the only case of an non-homomeric independent zone. For interpretation of dependent elementary zones, see Table~3.

>> Number of banding patterns predicted by hypothesis:
   The number of banding patterns that should exist under each hypothesis is compared with the number found as a control of the genetic closure of the sample.

TABLE 3: Properties of the elementary zones representing the various types of isoenzyme under a codominant mode of gene action including the possibility of null alleles. Migration velocities of molecules encoded by different genes or sets of genes are assumed to be unequal. These properties hold in a genetically closed sample of gametophytes. (Reproduced from Gillet (1996)).

	Type of isoenzyme	Property of elementary zone
	- Unmodified enzymes and fixed PTM's (1) -	
I.	Homomeric of gene at variable locus together with its fixed PTM's	Independent
II.	Homomeric of gene at fixed locus together with its fixed PTM's	Fixed
III.	Heteromeric between genes at variable loci together with its fixed PTM's	Is dependent only on the two independent elementary zones of the respective homomerics
IV.	Heteromeric between gene at a variable locus and gene at a fixed locus together with its fixed PTM's	Is represented by the (extended) homomeric elementary zone of the gene at the variable locus
V.	Heteromeric between genes at two fixed loci together with its fixed PTM	Fixed
	- Facultative PTM appears as additional band -	
	Homomeric (I) and PTM	Elementary zone of PTM is dependent on independent elementary zone of unmodified homomeric
	Homomeric (II) and PTM	Unmodified homomeric is fixed band, PTM is independent elementary zone
	Heteromeric (III) and PTM	Elementary zone of PTM is dependent on dependent elementary zone of unmodified heteromeric
	Heteromeric (IV) and PTM	Elementary zone of PTM is dependent on elementary zone of homomeric of variable locus
	Heteromeric (V) and PTM	Unmodified heteromeric is fixed band, PTM is independent elementary zone

(1) PTM = post-translational modification

5. DNA ELECTROPHEROGRAMS

TYPES OF DNA FRAGMENT

In DNA electrophoresis, each elementary zone is represents a DNA fragment "encoded" by a single allele at a locus, since fragments analogous to heteromeric isoenzymes and post-translational modification are thought not to occur. In a genetically closed sample of haploid gametophytes, each allele appears in banding patterns independently of the alleles at all other loci. Moreover, since the presence of one allele in a banding pattern rules out the presence of other alleles at the same locus, the elementary zones of DNA electropherograms all fulfill the definition of independence of elementary zones given above.

INTERPRETATION OF BANDING PATTERNS

If an elementary zone in a given sample is found not to be independent, the sample cannot be genetically closed, and no hypothesis can be formulated.

If all elementary zones show independence for the given sample, the qualitative interpretation of the banding patterns is the same as for monomeric isoenzymes without post-translational modification. The "null alleles" that often occur in DNA analysis, especially in RAPD, are also accounted for in analogy to the "null alleles" of isoenzyme analysis.

Thus the genetic interpretation of DNA electropherograms is no different than for a monomeric enzyme system allowing for "null alleles" but without PTM. Thus only minor changes were necessary to extend the applicability of the original program HAPLOZYM developed for zymograms (Gillet 1996) to DNA electropherograms. The number of elementary zones can, however, be much larger (e.g. DNA fingerprints), requiring a much larger sample of gametophytes to ensure genetic closure.

6. GENERATING ADDITIONAL HYPOTHESES

PERMUTATION OF ELEMENTARY ZONES

The assignment of homomeric equivalence classes to alleles of loci is performed in consecutive order. This ordering may have an effect on the assignment, in that a different ordering would yield a different hypothesis.

If a hypothesis is formulated for which the sample is genetically closed, all permutations of the elementary zones can be generated and examined for further hypotheses.

SPLITTING ELEMENTARY ZONES

It frequently happens, especially but not only in DNA banding patterns, that isoenzymes or DNA fragments with different molecular structures, stemming from different loci, migrate to the same position in the gel. Since their respective bands are usually indistinguishable (except for the rare case that differences in band intensity are interpretable), they are assigned to the same elementary zone. Since this elementary zone has two different genetic interpretations, <HAPLOGEN> is unable to formulate a valid hypothesis for mode of inheritance.

To alleviate this problem, <HAPLOGEN> provides the option of splitting one elementary zone at a time into two zones, alternately assigning the band appearing in the original zone in a banding pattern to either one or to both of the new zones. All combinations of assignment to the first new zone, the second new zone, and to both new zones are produced among all of the banding patterns that exhibit a band in the original zone.

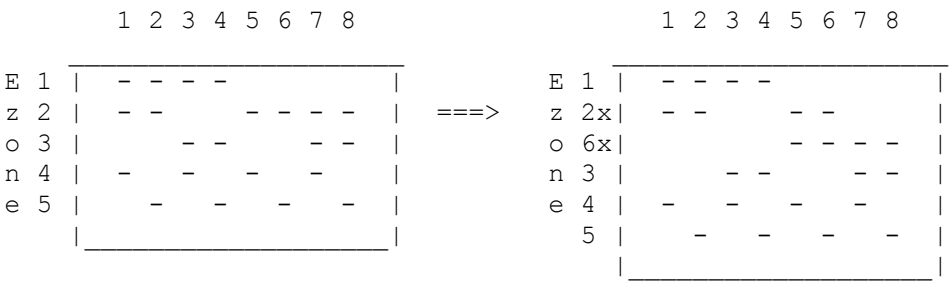
If a banding pattern exhibits a band in the elementary zone to be split, there are three ways of assigning this band to the two new zones:

- band appears in new zone 1 but not 2 (splitting code 1)
- band appears in new zone 2 but not 1 (splitting code 2)
- band appears in both new zones 1 and 2 (splitting code 0).

If N banding patterns exhibit a band in the elementary zone to be split, there are 3^N ways to assign the band to one or both of the new zones among the N patterns. Subtracting the three trivial cases where for each banding pattern the band is always assigned to the same new zone or to both zones, there remain 3^N-3 ways to split the bands appearing in the original elementary zone over the two new zones. Considering that each of these 3^N-3 cases has a symmetric counterpart that gives the same banding pattern (i.e., when all splitting codes 1 are replaced by 2 and all 2's by 1's), a total of (3^N-3)/2 different ways to distribute the N bands over two new elementary zones exist.

EXAMPLE 1: For the sample of banding patterns to the left, no hypothesis can be formulated for which the sample is genetically closed. By splitting elementary zone 2 into the two zones 2x and 6x, as done in the right diagram, hypothesis can be formulated for which the sample is genetically closed: Elementary zones 1 and 6x are alleles of locus 1, zones 2x and 3 are alleles of locus 2, and zones 4 and 5 are alleles of locus 3.

Schematic representation of banding patterns



7. INPUT FILE FOR <HAPLOGEN>

The input to the program is a schematic representation of the different banding patterns observed in a sample of gametophytes. After definition of the elementary zones represented by the patterns in the sample, each banding pattern can be described by a list of ones and zeros indicating presence or absence, respectively, of a band in the successive elementary zones.

To input m banding patterns, the user applies any text editor to prepare a data file (unformatted, ASCII characters only) consisting of m+3 lines (optionally m+2) as described in the following:

FORMAT OF INPUT FILE

Line 1: n = integer specifying number of different elementary zones

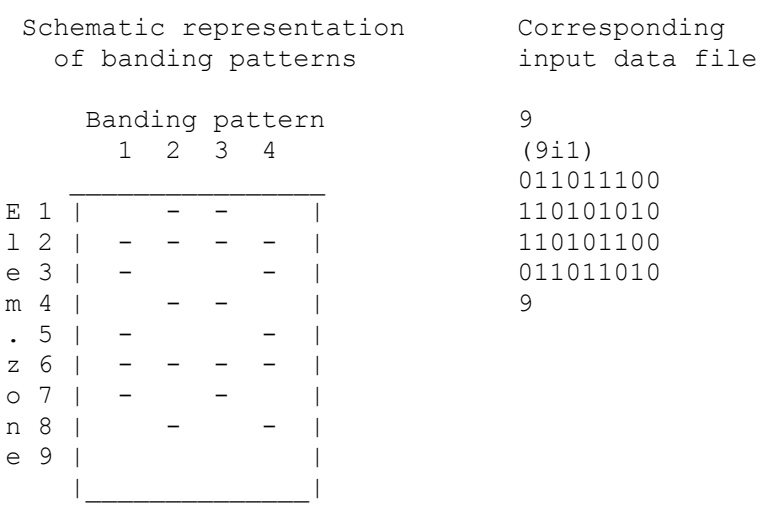
Line 2: (nI1) = usual FORTRAN format specification for reading banding patterns as a list of integers, each of width 1.
Other FORTRAN formats for reading list of integers can be specified, e.g. to include blanks of width w by wX (X-format) or define each of k integers to be of width w by kIw (see Example 4).

Lines 3 to m+2, one line for each banding pattern conforming to the format defined in Line 2:
A list of length n of 0's and 1's representing the banding pattern, where the entry in the j-th position of the list specifies the presence or absence of a band in the j-th elementary zone:
-> "1" signifies presence
-> "0" signifies absence

Line m+3: A "9" in the first position (according to format specification in Line 2) ends reading of the input file.
Optionally, the file can instead terminate at the end of the last line that defines a banding pattern, without a carriage return; otherwise, the following line and any further empty lines will be read as the banding pattern "000...0".

Banding patterns that are encountered more than once can be included in the input file as often as they appear, since <HAPLOGEN> recognizes redundant patterns and prints the number of times each pattern is encountered.

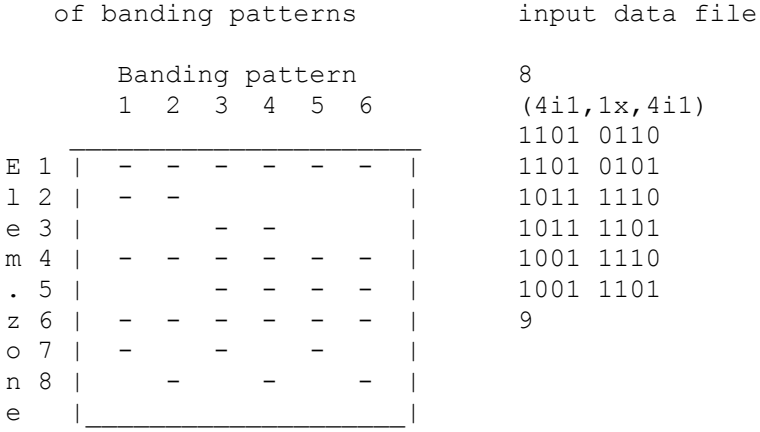
EXAMPLE 2



Corresponding input data file

```
9
(9i1)
011011100
110101010
110101100
011011010
9
```

EXAMPLE 3

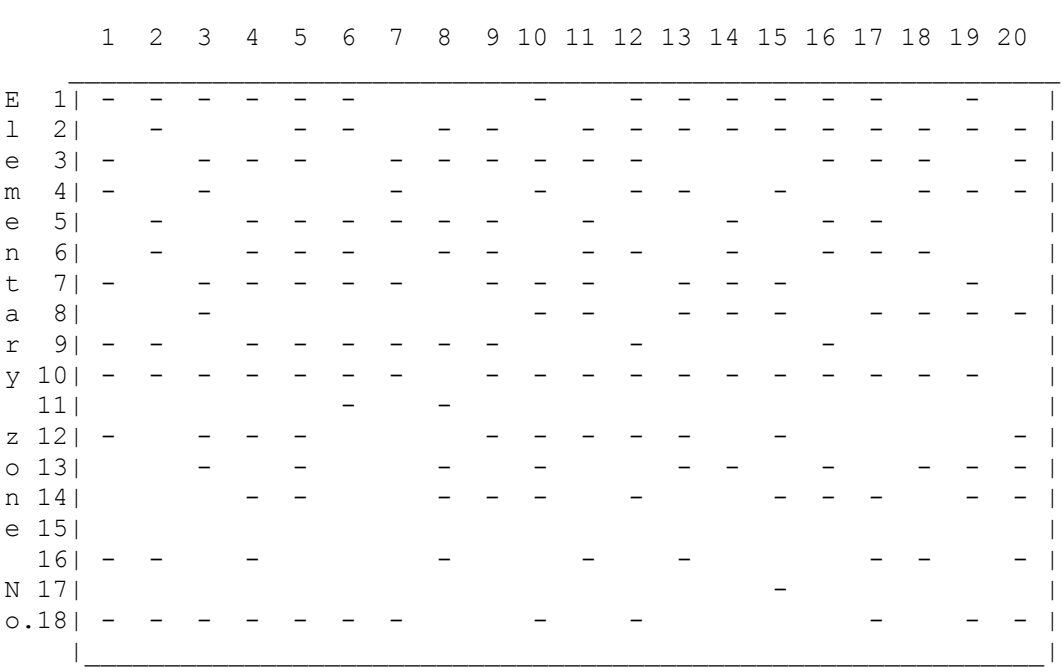


Corresponding input data file

```
8
(4i1,1x,4i1)
1101 0110
1101 0101
1011 1110
1011 1101
1001 1110
1001 1101
9
```

EXAMPLE 4

Schematic representation of banding patterns



Corresponding input file

```
18
(i1,17i2)
1  0  1  1  0  0  1  0  1  1  0  1  0  0  0  1  0  1
1  1  0  0  1  1  0  0  1  1  0  0  0  0  0  1  0  1
1  0  1  1  0  0  1  1  0  1  0  1  1  0  0  0  0  1
1  0  1  0  1  1  1  0  1  1  0  1  0  1  0  1  0  1
1  1  1  0  1  1  1  0  1  1  0  1  1  1  0  0  0  1
1  1  0  0  1  1  1  0  1  1  1  0  0  0  0  0  0  1
0  0  1  1  1  0  1  0  1  1  0  0  0  0  0  0  0  1
0  1  1  0  1  1  0  0  1  0  1  0  1  1  0  1  0  0
0  1  1  0  1  1  1  0  1  1  0  1  0  1  0  0  0  0
1  0  1  1  0  0  1  1  0  1  0  1  1  1  0  0  0  1
0  1  1  0  1  1  1  1  0  1  0  1  0  0  0  1  0  0
1  1  1  1  0  1  0  0  1  1  0  1  0  1  0  0  0  1
1  1  0  1  0  0  1  1  0  1  0  1  1  0  0  1  0  0
1  1  0  0  1  1  1  1  0  1  0  0  1  0  0  0  0  0
1  1  0  1  0  0  1  1  0  1  0  1  0  1  0  0  1  0
1  1  1  0  1  1  0  0  1  1  0  0  1  1  0  0  0  0
1  1  1  0  1  1  0  1  0  1  0  0  0  1  0  1  0  1
0  1  1  1  0  1  0  1  0  1  0  0  1  0  0  1  0  0
1  1  0  1  0  0  1  1  0  1  0  0  1  1  0  0  0  1
0  1  1  1  0  0  0  1  0  0  0  1  1  1  0  1  0  1
9
```

8. RUNNING <HAPLOGEN>

When started, <HAPLOGEN> asks for the name of an input file that was prepared previously using a standard text editor (see Sec. 7 above). It then poses the following questions:

CHOOSE INPUT FILE

>> Name of input data file [default extension = .dat]: >>

Include path if data file is not located in the same directory as the program. If the file's extension is ".dat", it can be omitted and is supplied by the program. For example, the file "d:\path\infile.dat" can be given as "d:\path\infile", but "e:\zymo.tst" must be fully given.

>> ** File does not exist: XXX

If named file, here XXX, is not found, you are prompted to retry. Check path designation.

CHOOSE OUTPUT DEVICE

>> Output device? Screen only="s" or File+Screen="f" [default="s"] : >>

- An answer of "s" causes all output to appear on the screen only - no output is saved for later reference.
- An answer of "f" causes all output to be saved in the output file and abbreviated output to simultaneously appear on the screen.

SPECIFY OUTPUT FILE

>> Name of output file [default = XXX.out] ? : >>

Press ENTER to give the output file the same path and filename (here represented by XXX) as the input file and the extension ".out". Otherwise, type complete path, filename and extension as desired.

>> ** File XXX.out already exists. Append="a", Overwrite="o"? : >>

- An answer of "a" causes new output to be appended to the end of the existing file XXX.out without changing previous contents of the file.
- An answer of "o" causes new output to be written at the beginning of XXX.out, and all previous contents of the file are lost.

SPECIFY TYPE OF BANDING PATTERNS

>> Type of pattern?:
Zymogram = "z", DNA electropherogram = "d" [default = "z"] >>

- If the answer is "d", <HAPLOGEN> treats all bands as alleles at some locus. In terms of programming technique, all bands are handled as if they were monomeric (thus homomeric) isoenzymes in a system allowing "null alleles" but without PTM.

GENERATING ADDITIONAL HYPOTHESES BY PERMUTING HOMOMERIC EQUIVALENCE CLASSES

>> Do you want to search for alternative hypotheses by checking the nnn permutations of the homomeric equivalence classes? : >>

- If the answer is "y", then the numbers of the homomeric equivalence classes are permuted in all possible ways to check for additional hypotheses.
- If the answer is "n", no permutations are performed.

GENERATING ADDITIONAL HYPOTHESES BY SPLITTING ONE EZONE INTO TWO OVERLAPPING EZONES

>> Do you want to search for overlapping Ezones (epistasis)? :
Yes="y", No="n" [default="n"] : >>

See Section 6 above.

>> Which Ezone should be split into two new Ezones?
Ezone N = "N", All Ezones = "0", End program = "-1"
[no default]: >>

- If the answer is a positive integer "N", then only elementary zone N is split.
- If the answer is "0", all elementary zones are split, one at a time.
- If the answer is "-1", the program is terminated.

>> If an Ezone exhibits a band in N patterns, there are (3^N-3)/2 ways to distribute the N bands over two new Ezones, such that for each of the N patterns, a band appears in at least one of the new zones.
Input maximal N not greater than nn for which Ezone splitting is to be performed [default="nn"] : >>

- An answer of "N" causes only those Ezones to be split in which min(N,nn) or fewer banding patterns exhibit a band, where nn is originally set in <HAPLOGEN> to equal 6.

>> When splitting of an Ezone yields a viable hypothesis, do you want to search for alternative hypotheses by checking the permutations of the homomeric equivalence classes?
Always="a", Sometimes="s", Never="n" [default="n"]: >>

If all elementary zones are to be split (answer "0" above), the answer to this question determines whether permutations of homomeric equivalence classes are carried out automatically or only upon request.

9. REFERENCES

Bergmann F, Gillet EM. 1996. Phylogenetic relationships among pine species inferred from different numbers of 6PGDH loci. Plant Systematics and Evolution 208, 25-34.

Bergmann F, Gregorius H-R, Scholz F. 1989. Isoenzymes, indicators of environmental impacts on plants or environmentally stable gene markers? In: Scholz F, Gregorius H-R, Rudin D (eds.): Genetic Effects of Air Pollutants in Forest Tree Populations. Springer-Verlag, Berlin, Heidelberg, New York, Tokyo, pp 3-6.

Gillet EM. 1996. Qualitative inheritance analysis of isoenzymes in haploid gametophytes: Principles and a computerized method. Silvae Genetica 45, 8-16.

Gregorius H-R. 1980. The probability of losing an allele when diploid genotypes are sampled. Biometrics 36, 632-652.

10. TECHNICAL CONSIDERATIONS

<HAPLOGEN> is written in Fortran~77. The version available on the internet under URL:

http://www.uni-forst.gwdg.de/forst/fg/index.htm

is compiled for DOS and runs with Windows.

E-Mail of author: egillet@gwdg.de