Computer Notes

Popmin: A Program for the Numerical Optimization of Population Sizes in Marker-Assisted Backcross Programs

F. Hospital and G. Decoux

In a backcross breeding program aimed at introgressing a "target" gene from a "donor" line into the genomic background of a "recipient" line, an important issue is to reduce the length of the intact chromosomal segment of donor type dragged along around the target gene (linkage drag), because this segment hosts most of the unwanted donor genes still segregating in the population after a few generations (Hospital 2001; Naveira and Barbadilla 1992; Stam and Zeven 1981; Young and Tanksley 1989). This reduction can be achieved by selecting for individuals that are heterozygous at the target locus and homozygous for recipient-type alleles at two markers flanking the target locus on each side (such individuals are termed "double homozygotes" herein).

The probability of obtaining such double homozygote individuals (probability of success) depends on the distances between the target gene and the flanking markers, on the number of successive backcross (BC) generations that are to be performed (total duration of the breeding program), and on the number of individuals that are genotyped at each generation (population sizes). For a better reduction of linkage drag, flanking markers should be chosen as closely linked to the target locus as possible (Hospital 2001). But the probability of obtaining double homozygote individuals for close markers in a single BC generation is very low. Hence it is generally preferable to perform selection on at least two successive BC generations, allowing for example selection for a single homozygote on one side of the target, then for a single homozygote on the other side (Young and Tanksley 1989). Moreover, genotyping effort is even reduced when considering more than two BC generations and when population sizes at each generation are optimized simultaneously (Hospital 2001).

In order to minimize genotyping efforts, it is then necessary to compute minimal population sizes—that is, the minimal number of individuals that should be genotyped at each generation so that at least one double homozygote is obtained by the end of the program. These computations over several successive BC generations are more complex than in the case of a single BC generation, because recombination between the target gene and each flanking marker can take place in any generation to provide a double-homozygote genotype. Such computations must be performed numerically.

Popmin is an interactive program designed for fast and easy numerical computation of such minimal population sizes in backcross programs. It is based on the theoretical framework provided in Hospital (2001). The user should refer to that article or to the documentation supplied with the program (see below) for more details. The theoretical framework assumes that, at each BC generation, a single individual is selected based on its genotype at the two flanking markers. The program provides the minimal number of individuals that should be genotyped at each generation, such that at least one individual with desired genotype is obtained by the end of the program. On an average, more than one individual may be obtained, but formally, minimal population sizes such that at least $k > 1$ individuals are obtained are not considered here.

First, popmin should be used before a marker-assisted backcross program is started in order to optimize the breeding scheme under various constraints: positions of the flanking markers, total number of BC generations, and number of individuals genotyped.

If rapid success is mandatory (e.g., for economic reasons) then one can compute the minimum number of individuals that should be genotyped in order to obtain a double homozygote in the requested number of BC generations. If rapid success is not mandatory, then one can use popmin to investigate breeding schemes of various durations (three, four, or more BC generations) in order to determine the appropriate balance between the duration and the cost of the program. Examples of such investigations are given in the user's manual, along with the corresponding outputs. In general, unless rapid success is really mandatory, allowing a not-too-low risk of failure in early BC generations (risk of not obtaining a double homozygote at that generation) can permit a drastic reduction of genotyping costs and should be recommended, which is converse to what is generally advocated (e.g., Frisch et al. 1999). Obviously the strategy and number of individuals to be genotyped should be reconsidered at each generation once the actual genotype of the individual selected is known. This is also possible using popmin with a relevant option. Finally, one can also use popmin simply to compute the probabilities of success associated with particular population sizes.

Popmin is written in ANSI C and runs under Unix, Linux, and Windows (DOS). Complete source code along with executables and documentation can be downloaded free of charge at http://moulon.inra.fr/~fred/programs/popmin.

From the Station de Génétique Végétale, INRA/UPS/INAPG, Ferme du Moulon, 91190 Gif sur Yvette,

France. Address correspondence to Frédéric Hospital at the address above, or e-mail: fred@moulon.inra.fr.

© 2002 The American Genetic Association

References

Frisch M, Bohn M, and Melchinger AE, 1999. Minimum sample size and optimal positioning of flanking markers in marker-assisted backcrossing for transfer of a target gene. Crop Sci 39:967–975.

Hospital F, 2001. Size of donor chromosome segments around introgressed loci and reduction of linkage drag in marker-assisted backcross programs. Genetics 158:1363–1379.

Naveira H and Barbadilla A, 1992. The theoretical distribution of lengths of intact chromosome segments around a locus held heterozygous with backcrossing in a diploid species. Genetics 130: 205–209.

Stam P and Zeven AC, 1981. The theoretical proportion of the donor genome in near-isogenic lines of self-fertilizers bred by backcrossing. Euphytica 30:227–238.

Young ND and Tanksley SD, 1989. RFLP analysis of the size of chromosomal segments retained around the *tm-2* locus of tomato during backcross breeding. Theor Appl Genet 77:353–359.

Received April 3, 2001
Accepted August 8, 2002

Corresponding Editor: Bruce S. Weir