

A Theory of Limits in Artificial Selection with Many Linked Loci

A. ROBERTSON

Introduction

Though linkage has been clearly recognized to be an important factor in artificial selection, particularly by Mather and his colleagues, the absence of any predictive treatment of its effect has remained one of the most important gaps in selection theory. In the last decade, there have been several efforts to attack the problem when many loci are involved using computer simulation (Fraser, 1957; Cockerham and Martin, 1958; Gill, 1965; Qureshi and Kempthorne, 1968, and others). These have rather illustrated its importance than increased our fundamental understanding. The problem is certainly very complex. In this paper we deal initially with a rather simplified situation with a set of loci equally spaced along a chromosome, each with two alleles with the same initial allelic frequencies and effects on the character under selection. We further restrict the problem by assuming additive gene action and initial linkage equilibrium. Nevertheless seven parameters are needed to describe the initial state of the population and the selection process. The necessary parameters and the symbols for them are:

- (i) The effective population size in Wright's sense of the word (N).
- (ii) The number of loci (n).
- (iii) The difference between the two homozygotes at a locus in the character selected for (a).
- (iv) The initial frequency of the desirable allele at each locus (q).
- (v) A measure of the intensity of selection (i).
- (vi) The phenotypic variance of the character in the initial population (σ^2).
- (vii) The map distance between adjacent loci (c).

We are then interested in the change in the character under selection at different generations with the main emphasis in the present paper on the limit of selection when all loci are fixed, though not necessarily for the desirable allele.

Certainly one of the main reasons for the lack of progress has been a failure to recognize that the population size is one of the most important

factors determining the effect of linkage. A theory of selection in finite populations for single loci has only recently been developed. The situation with two linked loci with initial linkage equilibrium is fairly well understood from the work of Latter (1965) and Hill and Robertson (1966). Without the assumption of equivalence between the two loci, the specification of the two-locus problem requires seven parameters, but a satisfactory treatment is possible by a reduction to five parameter combinations. Here the initial frequency of the desirable alleles proves to be of major importance. In almost all other simulation work, with the exception of Qureshi and Kempthorne (1968), initial frequencies of 0.5 have been assumed. Another feature of the work with two loci referred to above is that it was built around a comparison of the extreme situations of no crossing-over ($r=0$) and free recombination ($r=1/2$) respectively between the loci. In a surprising number of other studies, no real zero point has been included. Very often simulation runs with a small amount of crossing-over (say $r=0.005$) have been used for this purpose. In some instances, this may be in effect a long way from the true zero, as will be discussed later in the paper.

This exploration of a seven-dimensional space is not easy. I do not find that the conventional statistical approach to the problem, as used by Qureshi, Kempthorne, and Hazel (1968) in which one takes a set of values of the relevant parameters, does runs with all possible combinations within the set, and analyzes the results in main effects, first order interactions, and so on, contributes much to our understanding of the problem. To say that we "understand" a problem implies that we have found that way of looking at it which gives the simplest possible description of the results – that which has the least number of interactions. We may by a conventional analysis discover that factor A and factor B have significant effects and that there is a significant interaction between them. It is however much more useful to find that an analysis in terms of, say, A^2/B accounts for almost all of the variation due to both factors. In statistical terms, we are seeking the best "re-parameterization". We are only using a computer to find a way of thinking about the problem without needing a computer.

This has been achieved by, first of all, dealing with those special cases which were amenable to algebra on the basis of existing theory to get some clue as to what combinations of parameters were useful and using these as a guide when looking for a pattern in the simulation results. In fact, the parameter combination which now appears most useful in dealing with intermediate amounts of linkage arose from seeing a particular regularity in the results and developing a theoretical explanation for it. The conventional statistical analysis in essence provides a description of a set of results *a posteriori*. Our approach was on the contrary