SHORT COMMUNICATION

‘Social heterosis’ as a process that maintains genetic variation – a comment

M. SANTOS

Departament de Genètica i de Microbiologia, Grup de Biologia Evolutiva (GBE), Universitat Autònoma de Barcelona, Bellaterra (Barcelona), Spain

Introduction

In a recent paper Nonacs & Kapheim (2007; NK07 for short) claim that in the ‘mating pool’ mode of reproduction (Okasha, 2006, p. 51), where selection acts on groups of individuals that contribute offspring to a common pool which again forms new groups before selection by random assortment, allelic diversity can be easily maintained as long as genetically heterogeneous groups produce more offspring – no density regulation occurs within each group and, hence, allele frequencies are weighted by the average fitness of the group (i.e. hard selection; Christiansen, 1975); a process which they call ‘social heterosis’ and consider a potentially powerful mechanism that accounts for a significant amount of genetic variation.

Contrary to their suggestions, I show here that there is a reduced efficiency of selection and an increased probability of fixation of segregating alleles when many loci are simultaneously undergoing social heterosis even if independence of action of the different loci and linkage equilibrium are assumed. The critical issue is that linkage disequilibria are generated by the sampling process of creating small groups and interfere with selection.

Keywords:

- genetic diversity;
- genetic hitchhiking;
- group selection;
- linkage disequilibrium;
- social heterosis.

Abstract

In the ‘mating pool’ mode of reproduction, offspring genotypes at each generation are taken randomly from a common population and subdivided into groups where individuals representing a finite sample from the pooled distribution reproduce proportional to their fitness. Assuming that genetically diverse groups contribute more offspring, a recent article by Nonacs & Kapheim [J. Evol. Biol. 20 (2007) 2253] shows that allelic diversity can be easily maintained and proposes the process of ‘social heterosis’ as a potentially powerful mechanism that accounts for a significant amount of genetic variation. Contrary to their suggestions, I show here that there is a reduced efficiency of selection and an increased probability of fixation of segregating alleles when many loci are simultaneously undergoing social heterosis even if independence of action of the different loci and linkage equilibrium are assumed. The critical issue is that linkage disequilibria are generated by the sampling process of creating small groups and interfere with selection.
point out that the efficacy of the type of selection now envisaged by NK07 will also depend on the extent to which fitness is affected by multiple loci (an evolutionary possible scenario according to the authors).

In this note I first briefly summarize NK07’s model. Then I present the results from numerical simulations where it is shown that the efficacy of social heterosis does indeed depend on the number of loci, but the reason for this is more complex than in the situation modelled by Hoffmann & Nielsen (1985) where the productivity of a patch is simply proportional to the variance of genotypic values.

Model and simulations

NK07 consider a haploid population initially polymorphic for a varying number of alleles (from \( i = 2 \) to \( i = 10 \)) at equal frequencies where the total population size \( N = m \times n \) is subdivided into \( m \) groups with \( n \) individuals per group. With random grouping the distributions of the different compositions of groups for each of the \( i \) alleles will be binomial with parameters \((p_i,n)\). Assuming that allele 1 is the fittest one, the fitness of individuals within a group is modelled as follows:

\[
\begin{align*}
\text{Genetically homogeneous groups: } & W_1 = 1 \\
& W_{ij} = f_i; f_i \leq 1; \\
\text{Genetically heterogeneous groups: } & W_{1j} = 1 + y(a^2 - 1) \\
& W_{jk} = f_d[1 + y(a^2 - 1)]; \forall k \neq j, f_d \leq 1,
\end{align*}
\]

where \( y \) is a constant for the proportional increase of group’s fitness as a function of allele diversity \( a \) (i.e. number of alleles in the group); and \( z \) models the benefits of allele diversity upon fitness. Therefore, without genetic diversity increasing group’s fitness directional selection drives allele 1 to fixation whenever \( f_i \) or \( f_d < 1 \).

Here I simply extend NK07 model by considering the case where reproductive success of groups is simultaneously determined by \( g \geq 1 \) loci (i.e. conditions in eqn 1 apply to each locus). Independence of action of the different loci is assumed, which is taken to mean that the fitness values at different loci combine multiplicatively. Also linkage equilibrium is assumed (an assumption which is discussed below), which means here that the groups were created by random assortment with allele frequencies independently filled for each locus by drawing alleles according to their frequency in the common pool from the previous generation. The simulation program was implemented in MATLAB (V7; The Math-Works Inc., 2005).

A short digression is in order here. In his book on levels of selection Okasha (2006, pp. 56–59) considers that a key issue is whether the ‘particles’ (i.e. the different alleles in NK07 model) or the collectives (groups), or both, constitute the focal level. In the present model, it is quite obvious that allele diversity is the focal level, a situation referred to as ‘multi-locus selection’ (MLS1) by Okasha. As he points out, a number of authors claim that MLS1 is just individual-level selection with frequency-dependent fitness, so involves only one level of selection. However, I will freely use here the labels of ‘group selection’ when considering the among-group variance in fitness from eqn 1, and ‘allelic selection’ when considering genetically heterogeneous groups. Although this distinction is controversial, I think it may be useful in the present context because social heterosis can easily accommodate both the standard heterosis case of one locus with two alleles from a gene’s-eye perspective (i.e. \( n = 2 \) and \( f_d = 1 \); namely, no meiotic drive and, hence, all fitness differences are between, rather than within, organisms) as well as selection at the gene level by assuming meiotic drive (\( n = 2 \) and \( f_d < 1 \); see Okasha, 2006, pp. 162–169, for a comprehensive discussion).

Results

I will focus in the more interesting case where allele 1 has the highest fitness, a scenario extensively analysed by NK07. To facilitate comparisons with some of their results, it is further assumed here \( i = 3 \) alleles per locus, \( z = 1 \) and \( f_d = 0.95 \). Similarly, \( m = 30 \) groups were assumed unless otherwise stated. It is clear that average allelic diversity maintained by social heterosis decreases with an increasing number of loci (Fig. 1), and this behaviour is qualitatively independent of the initial parameter values (results not shown). However, some caution is needed to correctly interpret the results when the process is acting at multiple (independent) loci simultaneously. It is also evident from Fig. 1 that average allelic diversity initially decreases faster than in the pure drift case with a high enough number of loci, which clearly suggests that they encounter some kind of interference that promotes allelic fixation. Thus, it is interesting to note that the probability of fixation (\( P_{\text{fix}} \)
Fig. 1  Allele diversity over time maintained by social heterosis. In all panels $m = 30$ groups and group size ($n$) as indicated on the top; $z = 1$ and $f_s = f_d = 0.95$ were assumed except for the drift case (light-coloured heavy grey lines with $y = 0$ and $f_s = f_d = 1$). The dark-coloured heavy grey lines plot the results from simulations with only one locus ($g = 1$; averaged across 250 independent replicates) and can be compared with results from NK07. The thin lines show the results when social heterosis is acting on multiple (independent) loci simultaneously (averaged across loci and replicates).
for the fittest allele increased with the number of loci; e.g. for the case with \( n = 4 \) and \( y = 0.2 \), it rose from \( P_{\text{fix}} = 0.52 \) with one locus, to \( P_{\text{fix}} = 0.73 \) with 40 loci, and to \( P_{\text{fix}} = 0.82 \) with 80 and 120 loci. Even more remarkable, loss of the fittest allele was observed in some replicates with multiple loci (5% with 120 loci and the former parameter values), a situation never observed with one locus. (Note that with \( y = 0.2 \) the equilibrium frequencies estimated from 500 replicates with one locus, and \( n = 4 \) and \( m = 300 \) to minimize drift, were 0.52 for the fittest allele and 0.24 for each of the two remaining alleles.) The interference is more important as the constant for the proportional increase of group selection \( y \) raises; with \( n = 4 \) and \( y = 0.3 \) the probability of fixation of the fittest allele was approximately 18 times higher with 120 loci when compared to the one locus situation, but its loss was also observed in 13% of the cases. Overall, these figures immediately suggest that under some circumstances, invasion of the fittest allele is not guaranteed when social heterosis is simultaneously acting on many loci.

The interference effect deserves some further comments. As stated above, linkage equilibrium was assumed and the groups were created by random assortment of both individuals and genes, but linkage disequilibrium or nonrandom association of alleles between loci are generated within groups by the random sampling of few individuals, a process that interferes with selection—the so-called ‘Hill-Robertson’ effect that describes that linkage between loci under selection will reduce the overall effectiveness of selection in finite populations (Hill & Robertson, 1966; Felsenstein, 1974; Barton, 2000). Thus, although on average within-group linkage disequilibrium (LD) in the simulations is zero, when fittest alleles occur in coupling (LD > 0) they increase to a higher frequency and, conversely, when they occur in repulsion (LD < 0) they compete against each other, which slows their rate of increase. (Notice that with many loci the distinction between genetically homogeneous or heterogeneous groups is blurred as all individuals are genetically diverse.) The net effect is that in the common pool the average frequencies will generally not be the same than they would have been when considering only one locus at a time. Figure 2 illustrates the average frequency changes over the first 100 generations for two extreme situations with one and 120 loci to better appreciate the interference effect.

What is the maximum number of loci that can remain polymorphic under a given set of conditions? This question was approached by calculating the number of loci where allelic diversity \( a \geq 2 \) in the simulations with \( g = 40 \), \( g = 80 \) and \( g = 120 \) (Fig. 3). Take again the favourable case with \( n = 4 \) and \( y = 0.2 \), which translates into a huge gain in relative fitness of 33% for a group heterogeneous for the two unfit alleles (approximately 5% of the groups at the equilibrium frequencies above) relative to a group homogeneous for the fittest allele (approximately 7% of the groups). The numerical results suggest that up to 20 loci remain polymorphic when the initial number of loci experiencing social heterosis is 120. Obviously this is a very crude estimate because the total population size \( N = 120 \) is quite low, but an upper bound of approximately 42 loci can be obtained by dividing the previous figure by one minus the probability of fixation of the fittest allele in the simulations with one locus (i.e. 1 - 0.52). The same upper bound of 42 loci was obtained with \( n = 4 \) and \( y = 0.3 \), which apparently suggest that there is a trade-off between the strength of social heterosis and the interference effect.

**Conclusions**

My aim here was not to dismiss the potential evolutionary relevance of social heterosis, but to raise some caveats about its theoretical importance as a mechanism that can account for a significant amount of genetic variation when many loci are simultaneously being selected. With small groups the average among-group variance in relative fitness per locus increases with the number of loci under the initial conditions. However, linkage disequilibria are generated by the sampling process and this interferes with selection and increases the probability of fixation of segregating alleles. These unavoidable effects in the model clearly contradict NK07 concluding remark that social heterosis ‘can maintain allelic diversity in the face of both genetic drift and directional selection’. If haploid genotypes in the simulations were more realistically stored as lists of alleles on a single chromosome, or on several chromosomes, with a genome of map length \( R \) and recombination frequency between adjacent loci \( r < 1/2 \) (i.e. by assuming the following life cycle...
mimicking a patchy population structure: within-group selection, mating pool formation, recombination, subdivision), the results in Fig. 1 would have likely been more exaggerated (but see below). Both the sampling effect and the inherent directional selection ($f_s < 1, f_d < 1$ in eqn 1) would increase linkage disequilibria and situations in which LD < 0 tend to last longer (generating, on average, negative LD), something that did not

Fig. 3  Number of loci that remain polymorphic ($a \geq 2$) in all panels from Fig. 1 and given for each 250 generations.
happen in the present simulations (Fig. 1) as disequilibria were generated afresh each generation by sampling (see Model and simulations).

It is worth briefly noting that eqn 1 can be used to model the standard heterosis situation in diploid organisms from a gene’s-eye perspective by assuming groups with \( n = 2 \) from a gamete pool with bi-allelic loci and \( f_a = 1 \) (i.e. no meiotic drive). In a classical paper Franklin & Lewontin (1970) studied the effect of linkage in multilocus systems affecting fitness with multiplicative heterosis and tightly linked loci. With 36 loci and strong selective coefficients they found fixation a various loci, particularly in the asymmetrical case with different fitness values for the two homozygous genotypes. From an initial survey of a range of recombinational values (the recombination routine was simulated using the stochastic multi-locus method of Fraser & Burnell, 1970) with \( n = 4 \) and parameter values as in Fig. 1, I also found higher probabilities of fixation with loose linkage (0.1 \( \leq r \leq 0.4 \)), although the effect was not dramatic (results not shown). Larger groups should suffer less genetic interference arising from the sampling process but, as already pointed out by NK07, the efficiency of social heterosis would also decrease and eventually become negligible as populations with larger groups will have a lower among-group variance (i.e. group selection will not counter-balance allelic selection): a purely statistical principle from the Central Limit Theorem that states that variance among sample means is inversely related to sample size. To sum up, when multiple loci are considered social heterosis dwells somewhere between Scylla and Charybdis.

The numerical results in this note raise an intriguing possibility which, admittedly beforehand, I do not take very seriously. From the conditions in eqn 1, it is easy to envisage the evolution of ‘altruism’ [here I follow Wilson’s (1980) concept of ‘weakly altruistic’ behaviours as groups in social heterosis are formed at random] – a behaviour which is costly to the actor and beneficial to the recipient – by simply assuming a one-locus, two-allele situation (altruistic and selfish types; for a recent empirical example see Nedelcu & Michod, 2006), where selfish is the fittest allele (\( f_d < 1 \) and \( f_s > 1 \) (i.e. homogeneous groups for the altruistic allele are more productive than homogeneous groups for selfish types. With \( n > 2 \) the model as it is does not pay off for the frequency of the altruistic allele in heterogeneous groups, but this is not a critical point). If social heterosis is acting simultaneously on many loci, there is a nonzero probability that the altruistic allele goes to fixation, and there is no guarantee that a selfish mutant would invade the population.

**Acknowledgments**

This work was supported by grant CGL2006-13423-01/BOS from the Ministerio de Ciencia y Tecnología (Spain). This manuscript is supported by funding of the INCORE project under the Sixth Research Framework Programme of the European Union. The views expressed in this publication are the sole responsibility of the author and do not necessarily reflect the views of the European Commission.

I thank E. Szathmáry for helpful discussions and encouraging comments on the manuscript, and P. Nonacs, K. Kapheim and an anonymous reviewer for critical remarks on a previous draft.

**References**


Received 31 August 2007; revised 13 November 2007; accepted 23 November 2007.