

Artificial Symbiogenesis and Differing Reproduction Rates

Larry Bull*

University of the West of England

Keywords

Coevolution, endosymbiosis, horizontal gene transfer, organelles, *NKCS* model

Abstract Symbiosis is the phenomenon in which organisms of different species live together in close association. Symbiogenesis is the name given to the process by which symbiotic partners combine and unify. This letter reconsiders previous work using the *NKCS* model of coevolution to explore symbiogenesis. In particular, the role of different replication rates between the coevolving partners is considered. This is shown to provide a broader scope for the emergence of endosymbioses and subsequent horizontal gene transfers.

1 Introduction

Symbioses are commonplace in the natural world, and it is therefore argued that the phenomenon is of great evolutionary significance (e.g., [11]). Symbiogenesis is the hypothesis that if the relationship between symbionts evolves in the direction of increasing dependency, potentially new morphologies and physiologies can emerge (e.g., [10]). This letter begins by considering the formation of an *endosymbiosis* between two coevolving species: the relationship under which one partner exists within another. In particular, it considers the fact that the partners may not reproduce at the same rate. This is done using a version of Kauffman and Johnsen's [9] abstract *NKCS* model, which allows for the systematic alteration of various aspects of a coevolving environment. Symbiogenesis also delineates the transfer of genes from one symbiont's genome to another—horizontal gene transfer—creating a more complex genome for the recipient. Again using the *NKCS* model, the evolutionary performance of endosymbionts that transfer increasing fractions of their genome to their partner is explored.

Computational models considering symbiogenesis have been presented since the beginnings of nature-inspired approaches (e.g., [1]). More recent examples include Ikegami and Kaneko's (e.g., [7]) demonstration of how the symbiogenetic merging of evolving artificial entities can be beneficial if the task faced is iteratively decomposable (see also [14]), and Tomlinson and Bull's (e.g., [13]) use of the process within an evolutionary reinforcement learner to solve perceptually ambiguous maze navigation problems. After introducing the *NKCS* model, this letter reconsiders the findings of Bull and Fogarty [3], who used it to explore symbiogenesis. In particular, it is shown that assumptions made by them restrict the potential for the process to emerge as beneficial.

2 The *NKCS* Model

Kauffman and Johnsen [9] introduced the abstract *NKCS* model to enable the study of various aspects of coevolution. In their model an individual is represented by a genome of N (binary) genes,

* Department of Computer Science, University of the West of England, Frenchay, Bristol BS16 1QY, UK. E-mail: Larry.Bull@uwe.ac.uk

each of which depends epistatically upon K other randomly chosen genes in its genome. Thus increasing K , with respect to N , increases the epistatic linkage, increasing the ruggedness of the fitness landscapes by increasing the number of fitness peaks, which increases the steepness of the sides of fitness peaks and decreases their typical heights. Each gene is also said to depend upon C randomly chosen traits in each of the other S species with which it interacts. The adaptive moves by one species may deform the fitness landscape(s) of its partner(s). Altering C , with respect to N , changes how dramatically adaptive moves by each species deform the landscape(s) of its partner(s). The model assumes all inter- and intra-genome interactions are so complex that it is only appropriate to assign random values to their effects on fitness. Therefore, for each of the possible $K + SC$ interactions, a table of $2^{K + SC + 1}$ fitnesses is created for each gene, with all entries in the range 0.0 to 1.0, such that there is one fitness for each combination of traits. The fitness contribution of each gene is found from its table; these fitnesses are then summed and normalized by N to give the selective fitness of the total genome for that species. Such tables are created for each species (Figure 1; the reader is referred to [8] for full details).

Kauffman and Johnsen’s [9] model uses populations of one individual (said to represent a converged species) and a genetic hill climber to evolve them in turn. That is, if a given single gene mutant is found to be fitter than its parent in the current context of the other species, that species as a whole moves to the configuration represented by the mutant. This is repeated for all species over a number of generations. They show how both inter-genome (C) and intra-genome (K) epistasis affects a coevolving system, particularly in the attainment of Nash equilibria (“a combination of actions by a set of agents such that, for each agent, granted that the other agents do not alter their own actions, its action is optimal” [8, p. 245]). Their model is used here with two species ($S = 1$), together with a probabilistic mutation rate of $1/N$ per gene for the hill climber and $N = 100$. All results presented are from 10 runs on each of 10 randomly created $NKCS$ models (altogether 100 runs) after 5000 generations.

3 Symbiogenesis I: Endosymbiosis

Previously, Bull and Fogarty [3] presented a population-based version of the $NKCS$ model with which to explore the conditions under which endosymbiosis may occur. In their model, two genetically separated species can form or dissolve an endosymbiotic association through a mutation-like event. They show how an endosymbiotic relationship between the two becomes more dominant as the degree of coupling between them (C) increases with respect to the degree of local epistatic coupling (K): When $K > C$, endosymbiosis appears to be beneficial. This is explained by the fact that

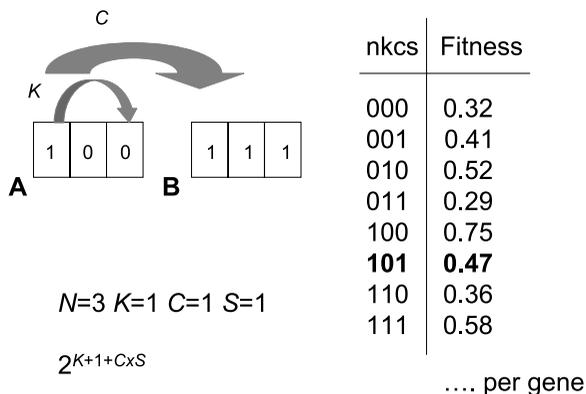


Figure 1. The $NKCS$ model: Each gene is connected to K randomly chosen local genes and to C randomly chosen genes in each of the S other species. A random fitness is assigned to each possible set of combinations of genes. These are normalized by N to give the fitness of the genome.

the coupling removes the effects of C between the two; all fitness landscape movement is removed, and C essentially becomes K .

However, they assumed that the endosymbiotic relationship resulted in a single, genetically joined symbiotic organism consisting of a part from each species. Here, as in Kaufman and Johnsen's original model, each species evolved one generation in turn and therefore at the same rate. Mutations to either joined part must therefore subsequently be beneficial to the whole. This can be seen as somewhat founded on the cooperative view of symbiogenesis as a natural process (e.g., [11]) wherein a network of mutually beneficial relationships between organisms is viewed as becoming more integrated with time. Others have suggested that the formation of endosymbioses may be seen as an example of slavery, with the endosymbiont being internalized purely for the benefit of the host (e.g., [12]). Perhaps more importantly, it is often the case that the endosymbionts exist as an evolving population *within* the host. This is typical, for example, for endosymbiotic relationships between animals and bacteria and is the case for the major organelles within eukaryotic cells. Bull and Fogarty's [3] assumption of a single, genetically linked genome resulting from the endosymbiosis is therefore a potentially significant simplification of the typical scenario; their assumption of equal rates of replication for the two species—and hence of no internalized population evolving to its own fitness function as before—may mask some of the underlying dynamics of the process.

Bull et al. [4] have used the *NKCS* model to show how, for increasing C , coevolving species can experience a phase-transition-like dynamic when one has a faster relative rate of reproduction. Here, the slower species' evolution can quickly reduce to a random walk as its relative reproduction rate R (number of rounds of mutation and selection) increases with respect to that of its ecologically coupled partner. Simply, the slower species is unable to track the continual motion of the optima within its own fitness landscape, caused by the changes in its partner.

Figure 2 shows examples of the effect for two coevolving species for various K and C , and for varying R . As can be seen, as one species performs an increasing number (R) of rounds of mutation and selection for every round by the other, its fitness increases (for $C > 1$ and all K). The same result was found for a variety of K , C , and R (not shown). Therefore, if the slower species were able to internalize the faster one, with the primary aim of reducing the relative difference in its reproduction rate from that of the faster species (upon which it depends or to which it is significantly coupled), its fitness would increase relative to those that do not form the endosymbiotic relationship. This would be the case despite the fact that the faster, internalized species would continue to coevolve with its host. Hence the scenario uses the slavery analogy rather than the purely cooperative relationship of Bull and Fogarty [3].

4 Symbiogenesis 2: Horizontal Gene Transfer

With the transfer of genes, a symbiosis becomes more closely integrated. Part of the genome of one symbiont is transferred to the genome of the other. The new genome may underlie metabolic pathways leading to an advantageous product that neither partner was capable of producing alone. [11, p. 189]

Within the *NKCS* model there is no scope for the emergence of novel functionalities; however, Bull and Fogarty [3] examined the selective performance of gene transfer as a way of configuring interdependent genes with a slightly altered version of the model previously described. They again assumed a genetically linked genome consisting of one set of genes from each of the two originally coevolving species and then assumed one of the species exploited recombination. They compared the scenario where recombination may occur only within the sexual species' part of the genome, the other species being inherited as a block, with a scenario where a 100% gene transfer is said to have occurred and therefore recombination can occur anywhere in the whole linked genome. They showed how, as the degree of coupling between the two species' increased (i.e., C , which subsequently behaves as K), the 100% gene transfer scenario becomes dominant. This is explained

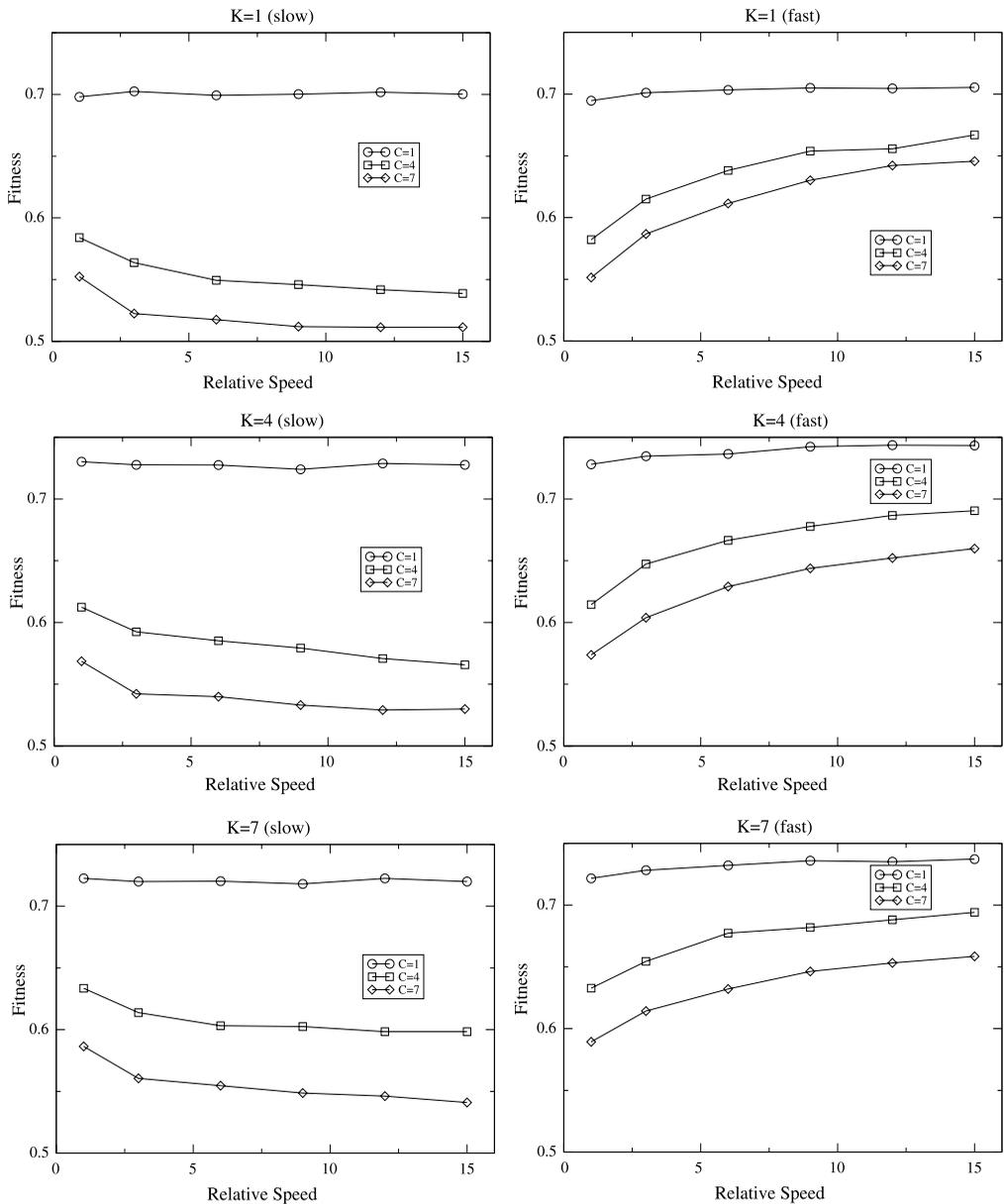


Figure 2. Showing the effect of one species evolving faster than another under different scenarios.

by the suggestion that only allowing recombination to occur in half of the linked genome effectively places a fixed recombination point in its middle, which is detrimental, because the degree of coupling (epistasis) increases across the whole genome.

This scenario is somewhat confused, however: By assuming a genome consisting of one part from the first originally coevolving species and a second part from the other originally coevolving species, using a combined total fitness, Bull and Fogarty were only dealing with the case under which 100% gene transfer had occurred.

As noted above, it is typically the case that a host maintains a still-coevolving population of endosymbionts internally. Thereafter, some genes from them may be transferred to the genome of the host. This process has been considered using the *NKCS* model containing two species coevolving at

different rates, as was shown in Figure 2. Here various percentages T of the genome of the faster species are assumed to have been transferred to the genome of the slower species. Thus the faster species is now viewed as the endosymbiont, and the slower species as the host. The varying percentage is taken from the left-hand end of the genome of the endosymbiont and placed on the right-hand end of the host's genome. The fitness contributions of the transferred genes to the host are calculated using the same table as before (i.e., from the endosymbiotic species' fitness function), and the total is now normalized by $N + T$, and the positions of all the original K and C connections of the original gene are carefully considered (Figure 3). Endosymbionts have their genes' fitness contributions normalized by $N - T$.

Figures 4 and 5 show typical examples with $T = 0\%$, 10% , 30% , 40% , and 50% . Both species are initialized with the same allele in each position over the $2N$ genes in each case. As can be seen, in the highest-percentage case (50%) the hosts benefit significantly from the transfer of genes from the endosymbionts (two-tailed t -test $P < .05$ for each K, C, R combination per T versus $T = 0$). This is simply because it markedly decreases the relative rate of change of those genes, and host-detrimental mutants in the transferred genes are discarded. Conversely, the endosymbionts experience a significant drop in fitness. Generally, however, up until T is approximately 30% there is typically no significant effect ($P > .05$) on the fitness of either, with respect to the equivalent $T = 0\%$ case. That is, in all cases, transferring a smaller number of genes appears to be selectively neutral—with a definite bound of less than 50% .

Previously, Harvey [6] has considered the use of a mutation operator that increases genome length. Using a version of the abstract NK fitness landscape model (e.g., [8]), Harvey showed, by including a bias, that gradual growth through small increases in genome length via mutation is sustainable, whereas large increases in genome length per growth event are not sustainable. This is explained as being due to the fact that a degree of correlation between the smaller fitness landscape and the larger one must be maintained; a fit solution in the former space must achieve a suitable level of fitness in the latter to survive into succeeding generations. Harvey's growth operator is a form of mutation that adds g random genes to an original genome of length G ; the larger g , the less correlated the two landscapes will be, regardless of the underlying degree of correlation of each. Bull [2] has also shown that increased growth can be obtained in the coevolutionary case with increasing C , that is, under the conditions for endosymbiosis as identified here.

The finding that a small percentage of an endosymbiont's genome can be moved without detriment to its fitness can be seen as essentially the inverse case of that considered by Harvey. That is, so long as the amount reduced leaves the endosymbiont on a smaller landscape that is sufficiently similar to the original larger one, the transfer of genes is selectively neutral from its perspective; optima of similar fitness must be found under the given selection and mutation pressures in the two cases. As noted above, such transfers can become beneficial to the host in the longer term, and thus the

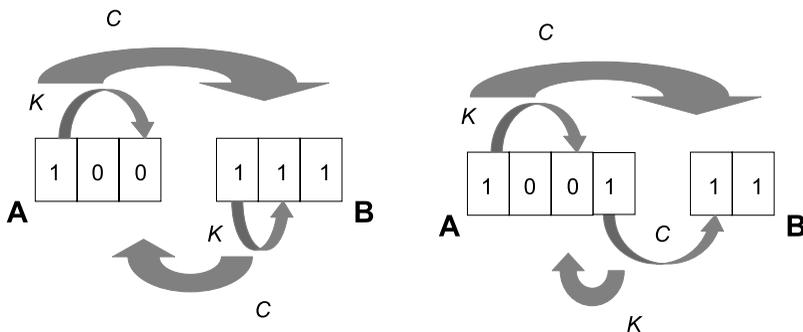


Figure 3. Horizontal gene transfer in the $NKCS$ model: The leftmost T genes ($T = 1$ here) in the faster species (B) are transferred to the rightmost end of the slower species (A). The internal (K) and external connections (C) are labeled appropriately, and the fitness table entries for the T genes are removed from B's and added onto A's. The total fitnesses are normalized by $N + T$ and $N - T$, respectively.

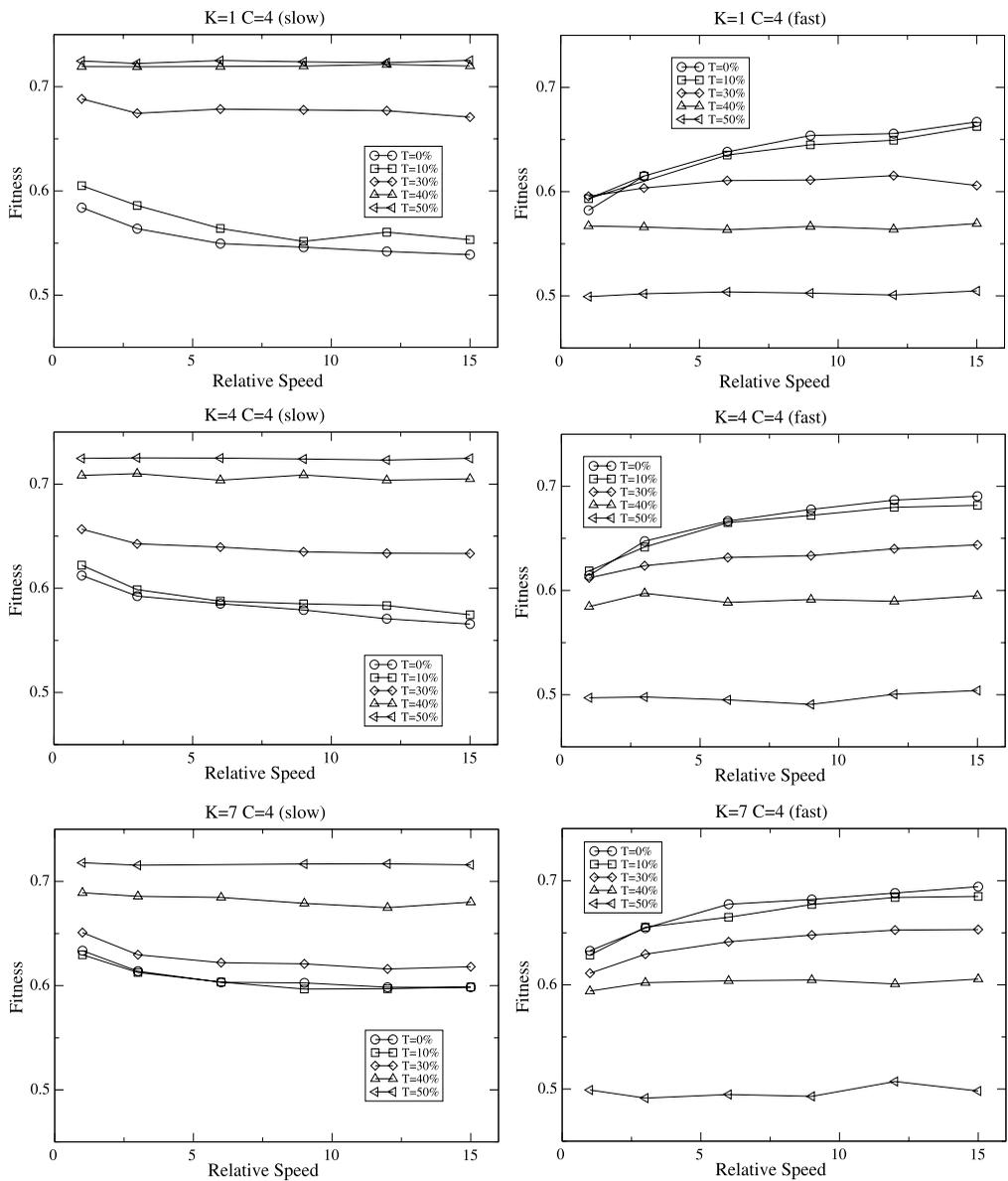


Figure 4. Showing the $C = 4$ case for degrees of gene transfer (T) at different speeds.

host and its shorter endosymbionts will take over from hosts where the transfer has not occurred. A slight metabolic advantage for the genetically shorter endosymbionts can be envisaged, which may give some direction to the roughly neutral scenario, thereby removing the transferred genes completely from the endosymbiont population over time.

Mitochondrial DNA can contain anywhere from 3 to 67 genes; and chloroplasts 60 to 100 genes, and it is assumed that the rest of their original DNA has transferred to the nucleus, although some may simply have been lost. However, it is not clear why such organelle DNA remains (e.g., see [5] for recent discussions). The results in Figures 4 and 5 show that up to around 30% transfer is generally sustainable. Importantly, the results suggest that gene transfers will eventually stop occurring. That is, the resultant endosymbiont (organelle) will eventually experience a significant relative drop in fitness from a transfer, because it is no longer possible to move such a small (coherent) amount,

namely, when 30% of an endosymbiont's genome is less than one gene's encoding. The variation in the number of genes remaining can be considered as due to factors such as variations in gene coding length (i.e., number of bases), the order of transfers occurring, gene loss, or error thresholds under mutation.

5 Conclusions

Symbiogenesis is a fundamental force within natural evolution. Previously, results from using the abstract *NKCS* model showed that increased coevolutionary coupling compared to internal epistasis (i.e., $C > K$) would result in the beneficial formation of endosymbioses [3]. The results presented

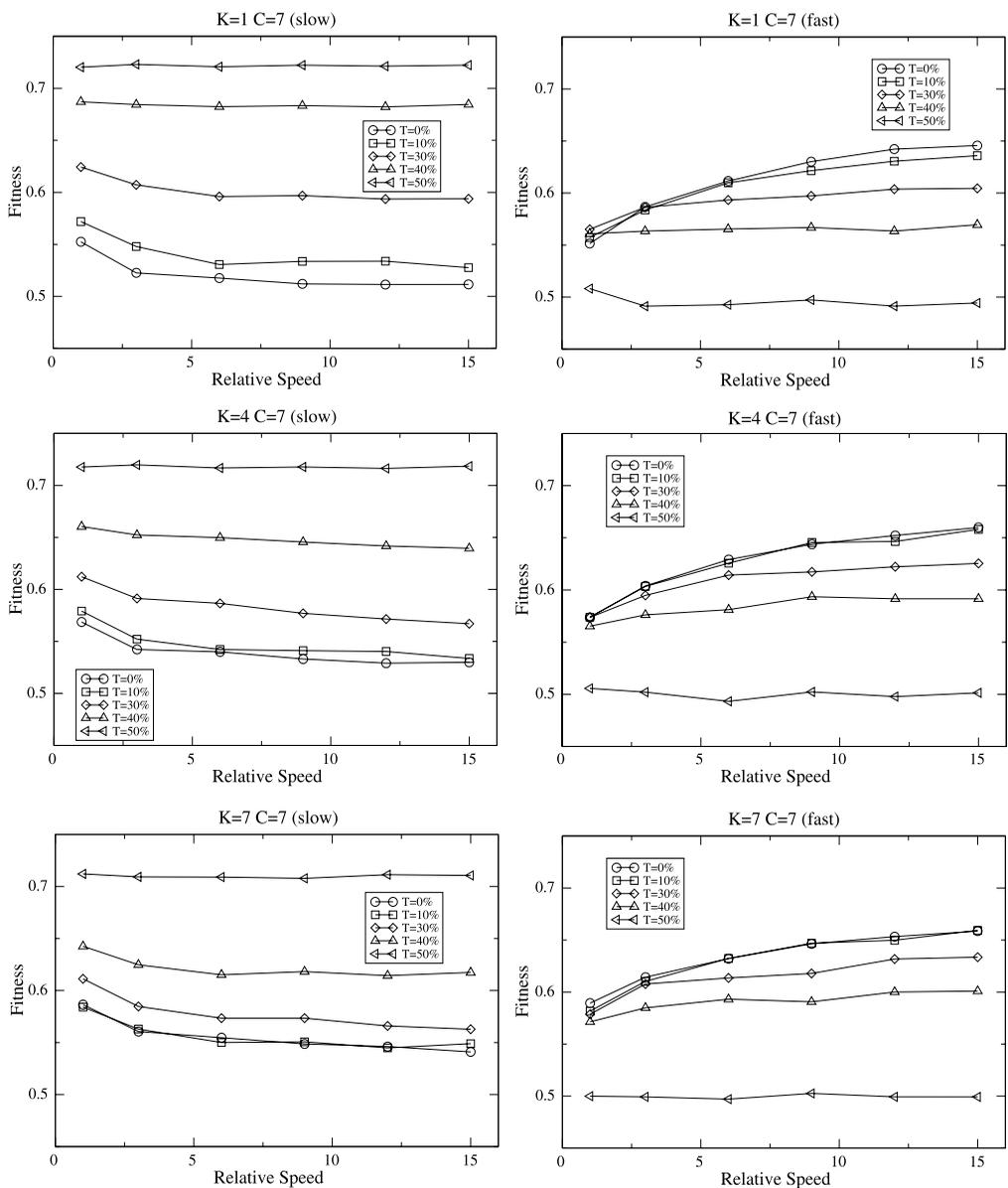


Figure 5. Showing the $C = 7$ case for gene transfer at different speeds.

here have shown that if relative replicator speed is considered, with the assumption that formation significantly slows the internalized species, the conditions for beneficial endosymbioses are wider—whenever $C > 1$. Further, under the same conditions, horizontal gene transfer will occur if a small fraction of the endosymbiont's genome is moved per event, as this is selectively neutral, with a limit predicted due to fitness landscape correlations.

References

1. Barricelli, N. A. (1957). Symbiogenetic evolution processes realized by artificial methods. *Methods*, IX(35/36), 143–182.
2. Bull, L. (2005). Coevolutionary species adaptation genetic algorithms: A continuing SAGA on coupled fitness landscapes. In *Proceedings of the Eighth European Conference on Artificial Life* (pp. 322–331). Berlin: Springer.
3. Bull, L., & Fogarty, T. C. (1996). Artificial symbiogenesis. *Artificial Life*, 2(3), 269–292.
4. Bull, L., Holland, O., & Blackmore, S. (2000). On meme-gene coevolution. *Artificial Life*, 6(3), 227–235.
5. De Grey, A. (2005). Forces maintaining organellar genomes: Is any as strong as genetic code disparity or hydrophobicity? *BioEssays*, 27, 436–446.
6. Harvey, I. (1992). Species adaptation genetic algorithms: A basis for a continuing SAGA. In F. J. Varela & P. Bourgine (Eds.), *Toward a Practice of Autonomous Systems: Proceedings of the First European Conference on Artificial Life* (pp. 346–354). Cambridge, MA: MIT Press.
7. Ikegami, T., & Kaneko, K. (1990). Genetic fusion. *Physical Review Letters*, 65(26), 3352–3355.
8. Kauffman, S. A. (1993). *The origins of order: Self-organisation and selection in evolution*. Oxford, UK: Oxford University Press.
9. Kauffman, S. A., & Johnsen, S. (1992). Co-evolution to the edge of chaos: Coupled fitness landscapes, poised states and co-evolutionary avalanches. In C. G. Langton, C. Taylor, J. D. Farmer, & S. Rasmussen (Eds.), *Artificial Life II* (pp. 325–370). Reading, MA: Addison-Wesley.
10. Khakhina, L. N. (Ed.) (1992). *Concepts of symbiogenesis: History of symbiogenesis as an evolutionary mechanism*. New Haven, CT: Yale University Press.
11. Margulis, L. (1992). *Symbiosis in cell evolution*. New York: W.H. Freeman.
12. Maynard Smith, J., & Szathmáry, E. (1995). *The major transitions in evolution*. New York: W.H. Freeman.
13. Tomlinson, A., & Bull, L. (2001). Symbiogenesis in learning classifier systems. *Artificial Life*, 7(1), 33–62.
14. Watson, R., & Pollack, J. (2001). Symbiotic composition and evolvability. In *Proceedings of the 6th European Conference on Artificial Life* (pp. 480–490). Berlin: Springer.