Symbiosis Enables the Evolution of Rare Complexes in Structured Environments

Rob Mills and Richard A. Watson

Natural Systems Research Group, University of Southampton, UK
`rmm05r@ecs.soton.ac.uk`

Abstract. We present a model that considers evolvable symbiotic associations between species, such that one species can have an influence over the likelihood of other species being present in its environment. We show that this process of 'symbiotic evolution' leads to rare and adaptively significant complexes that are unavailable via non-associative evolution.

1 Introduction

The organisation of biological systems depends strongly on the interactions between the organisms that share environments [1]. These interactions not only shape the selective forces on organisms, but in addition can be subject to change themselves. Symbiosis, the collaboration between organisms of different types [2], is very common in nature [3, 4]. Symbiotic associations have the capability of altering an organism's biotic environment: the selective context in which it will appear. This modification in environment is clear where symbiosis is taken to its logical conclusion, symbiogenesis: where the symbionts involved become reproductively inseparable [5]. This has the potential for evolutionary significance: symbiogenesis is thought to be implicated in some of the major transitions in evolution [6, 7]. There are also less extreme symbioses between free-living species that still a have significant impact on their likely biotic environments [1].

Given the ubiquity of symbiotic associations, and the assumption that such relationships can modify the likelihood of co-occurrence between species, we ask the following questions: What kind of association formation mechanism can lead to the evolution of complexes that are unevolvable in the absence of associations? Under what conditions is such a distinction available?

In this paper we describe a model framework of symbiotic evolution, where the ecosystem composition adapts rapidly according to local dynamics, symbiotic associations adapt gradually between co-occurring species in the ecosystem, and the associations in turn modify local dynamics. We investigate an implementation of this framework that idealises the separation in timescales of these two adaptive processes. We find that evolving symbiotic associations on a slower timescale than the dynamical changes in ecosystem composition is sufficient to lead to adaptively significant complexes with many dependencies resolved, and are unavailable without association formation. We use a structured adaptive landscape, where only very specific changes are sufficient to traverse the

ruggedness. The challenge then for associative evolution is to evolve compatible groupings that resolve conflicts, and transfer competition to a higher level, which is sufficient to traverse this ruggedness. This provides insight into why there is a distinction in the evolutionary outcomes with and without association formation. In other work in this volume [8], we show conditions under which individual selection leads to the reinforcement of associations between individuals of different species that co-occur in the ecosystem dynamics. In the present paper, we use a higher-level model where associations evolve between species according to their co-occurrence. This allows us to examine the adaptive significance of the coupled processes, when compared to non-associative evolution.

Prior models have been suggested that investigate the evolution of symbiogenic encapsulation [10–12], or abstractions thereof [13]. These models use a variety of mechanisms to determine the suitability of symbiogenic joins, including Pareto dominance [10], context-optimality [13], and maximising reciprocal synergy [12]. In [14] the evolution of 'observers' provides groupings to coarsegrain an adaptive landscape. A related approach is applied to physical models in [15]. As noted, we use different temporal scales for ecosystem dynamics and association formation. In evolutionary computation, [16] uses results of multiple hill climbing runs to build a model of dependencies, which provides a similar timescale separation to successfully solve hierarchical problems. Memetic algorithms [17] also use search at two levels, but importantly, neither search process modifies the variational units for the other.

2   Modelling Symbiotic Variation

In our model ecosystem there are many species. Since we want to investigate the effect of evolving the interactions *between* species, the species themselves have a trivial representation. For each of N niches in the environment, two particular species compete to inhabit it. There are $2N$ species in total. Each species also has a set of association strengths, one for each of the other $2N-1$ species. These associations can evolve, whereas their niche is immutable.

Initially the ecosystem has a random constitution of N species, such that all niches are occupied. All association strengths are initialised at zero. Changes in the ecosystem composition occur by following the system dynamics – by introducing random migrants that compete with the current ecosystem occupants. The symbiotic associations strengthen between species that are present in the ecosystem at any given timestep. Within a reasonably short timescale, no further changes in ecosystem composition occur: the ecosystem is at a locally stable state. Periodically, the ecosystem is perturbed such that its constitution is randomised, and local dynamics will again cause ecosystem changes. Note that these perturbations do not modify the symbiotic associations that have evolved.

The changes in the ecosystem composition are effected by allowing a random species to immigrate, and if the overall utility of the ecosystem is higher with this migrant than without, it is retained. When the migrant species has non-zero associations with other species, these are interpreted as probabilities that those

Table 1: Symbiosis Model Main Procedure

1. allow d demes to run to their local attractor
2. measure the co-occurrence of each pair of species within all deme attractors, and reinforce symbiotic associations according to Eqn. 1.
3. randomise each ecosystem composition and go to step 1.

other species will migrate at once. Thus, if a pair of species has symbiotic associations of maximal strength, they will always migrate together. These associations have the effect of correlating the possible variation in ecosystem composition.

We use an external fitness function to define the overall utility of each ecosystem composition, and this is used in a 'black box' manner: a utility value is only defined if all niches are occupied. However, by comparing two compositions that only differ by the occupant of a single niche, we identify which of the alternatives is more suited to the current context, thus effecting individual-level selection. The case is similar when a migrant group modifies multiple niches. Therefore, this reveals the context-dependent utility contribution of a migrant group.

In order to examine the behaviour we instantiate a model within this framework, with two additional assumptions: i) the length of time the system spends at local attractors dominates the time spent in transients; ii) instead of modifying the associations gradually over several trajectories, we use several independent demes in parallel, and modify the associations in proportion to the co-occurrence of species across the ensemble of demes. Using only attractor states to inform association evolution makes explicit the separation of timescales between changes to association and ecological changes.

2.1 Model M-S: Continuous Associations, Parallelised

If the ecosystem spends most of its time at attractors (or close to), a suitable approximation is to only modify the associations according to the species that are present at attractors. Procedurally, this model is described in Table 1. Following the dynamics of the ecosystem is a simple process that takes into account the symbiotic associations. The procedure is described in Table 2.

Observed co-occurrence, $O_{i,j}$, is calculated from the proportion of demes where both species i and j are present. The expected co-occurrence frequency is calculated from the product of univariate frequencies: $E_{i,j} = X_i X_j$. Using these values we construct a deviance from expected metric (essentially, a measure of surprise with respect to $E_{i,j}$). Associations are formed according to the rule in Eqn 1, where h is the threshold below which all $S_{i,j}$ values are set to zero.

$$S_{i,j} = \begin{cases} \frac{O_{i,j}-E_{i,j}}{\min(X_i,X_j)-E_{i,j}}, & \text{if } h \cdot \frac{O_{i,j}-E_{i,j}}{\min(X_i,X_j)-E_{i,j}} < O_{i,j} \leq \min(X_i,X_j) \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

The exit condition in step 6) when following dynamics can either be to make a pre-specified number of migrations, or alternatively to wait for P trial migrations

Table 2: Symbiosis-Informed Ecosystem Dynamics

---

1. Evaluate the initial ecosystem composition $(\rightarrow f_p)$
2. Form a migrant group g:
    (a) Randomly select a migrant species m, and add to g
    (b) Select without replacement a random species $x \neq m$
    (c) With probability $S_{m,x}$, add x to g, unless that niche is already filled in g
    (d) If any species has not been sampled, goto (b)
3. Temporarily introduce g to the ecosystem, allowing g to take precedence over the current occupants
4. Evaluate the modified ecosystem composition $(\rightarrow f_m)$
5. If $f_m \geq f_p$, allow the migrant group g to remain permanently, and set $f_p \leftarrow f_m$
6. If exit conditions not met, go to step 2).

---

without a change in composition. Note that in step 5), if the overall utility of the modified ecosystem is not higher than the original ecosystem, *all* species changes are reverted. Implications of this are discussed further in Sec. 4.

3 Simulation Experiments

We investigate the behaviour of the described models, and in particular, the complexes that evolve. We measure the evolutionary timescales required for each model to find the globally optimal configuration in the entire landscape (examined over several system sizes). The control model (M-C) is equivalent to M-S except that no associations are evolved – the S matrix is held at 0 throughout.

Watson and Jansen [18] introduce a synthetic problem class where instances comprise several large modules of binary variables, or 'building blocks'. Each module has two optima, one of higher utility than the other, both with equal sized basins. These are concatenated with no inter-module dependencies to construct the full problem. Eqn. 2 defines the utility contribution of a single module.

$$f(x) = \begin{cases} k & \text{if } U(x) = k \\ \frac{U(x)}{2} & \text{if } k > U(x) > \frac{k}{2} \\ \frac{(k-U(x))}{2} & \text{else} \end{cases} \qquad (2)$$

Given that each of Z modules has k variables, x is a configuration of variables within that module, and $U(x)$ is the unitation (number of variables set to '1'). We set $k = Z = \sqrt{N}$, such that the size of the modules scales with the size of the system, and thus refer to this problem as the scalable building blocks (SBB) problem. Each variable in the problem corresponds to a niche, and we refer to the two species that can occupy this niche as the '0-species' and '1-species' – giving a total of $2N$ different species in the ecosystem. The resultant landscape is very rugged, with 2^Z locally optimal configurations. Of these configurations, only one is globally optimal: when all niches are occupied by the 1-species. From any local optimum, the nearest configuration of higher utility differs in k niches – all within one module that is currently entirely occupied with 0-species.
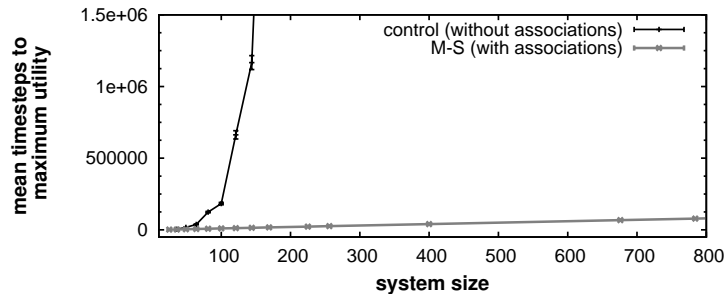
Fig. 1: Timesteps that each model takes before visiting the highest utility configuration in the SBB landscape. The symbiotic model visits this configuration with ease, whereas the control model very rarely does (thus requiring exponentially many epochs to sample the appropriate basin). The parameter settings are as follows: for M-S, we set $d = 50$, $h = 0.6$; In both M-S and the control M-C, following the dynamics uses $2N$ migrations, sampling on the initially selected species m without replacement; each model continues to restart until the globally optimal configuration is sampled. 30 repeats are performed for M-S, 100 repeats are performed for M-C (on account of higher stochasticity).

4   Discussion

As Fig. 1 shows, the symbiosis model provides an efficient processes that finds the very particular configuration with highest overall utility. This is in contrast to the control model, where the number of timesteps required to find the same configuration increases exponentially with the system size. The comparison indicates the significance of the result that M-S can find such a configuration.

The symbiotic model is very efficient at discovering the globally optimal configuration in this landscape. How is this so? First, consider the attractors visited by the initial set of demes. In each one, some modules will be occupied by all-0, and some by all-1 species – but no single attractor will have all-1 in all modules. Whenever a 1-species occurs, it always co-occurs with other 1-species in the other niches *in that module*. Note that it is not the case that 1-species always occur in any particular module: the all 0-species attractor also has a 50% basin. Furthermore, there is no correlation between the attractor that each module finds, since the landscape is separable. Associations are formed between species that co-occur frequently across the set of demes. Thus, within any module strong symbioses will form amongst all k of the 1-species, and likewise amongst all k of the 0-species. No associations will form between 0-species and 1-species that occupy niches in the same module, since at attractors there are no co-occurrences. The between-module co-occurrences are predicted by the univariate frequencies (*i.e.,* there is little or no surprise), so no associations will form here either. Fig. 2 (a) shows the resultant associations from the described process.

Now let us consider the local dynamics with these associations. Each migration is likely to introduce all of the compatible species in a particular module. This transfers competition to the module-level. To start with, introducing either
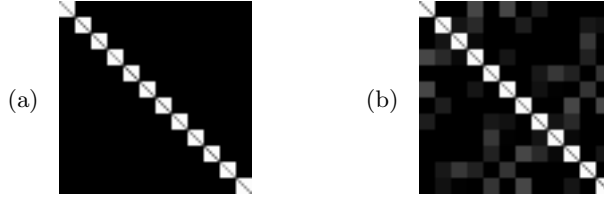
Fig. 2: Symbiosis matrices for a 72 species (N=36, k=6) ecosystem. (a) Ideal target to represent the landscape structure. (b) Calculated from the surprise metric (Eqn. 1) on an example set of local optima, without a threshold (*i.e., h*=0). This leads to a largely accurate representation, with some spurious interactions. An appropriate threshold can recover the target – see text. Lighter shades indicate stronger associations.

an all-0 group or an all-1 group will remain, since both are significantly higher utility than a random composition in the corresponding niches. However, as soon as an all-1 group is introduced, it has exactly the right variation to move to the highest utility possible in that module, and will thus replace any other composition. When appropriate all-1 groups have migrated into each of the modules, the overall utility will be maximised. After the associations develop as described above, this reliably occurs.

While each of the configurations that is discovered in one of the demes has N species present, the subsequent competition is not directly between these local optima. The associations that evolve form groups of k members, corresponding to each module, as described. This is important for a significant result. Because the module sub-functions are independent, these small groups are both sufficient and selectively efficient in the sense that they create Z independent competitions between the two sub-solutions in each module, rather than a single competition between all 2^Z possible local optima [18].

The success of the symbiosis model depends on the formation of associations appropriate to construct these per-module competitions. Fig. 2 (a) shows the target associations, which comprises strong associations between all compatible species within each module, and between-module associations. Frame (b) shows an example calculation of the surprise metric, without a threshold applied (Eqn. 1). In addition to the correct within-module interactions, this measure indicates some spurious interactions. Using Eqn. 1 with $h = 0.6$ is sufficient to recover just the appropriate associations, as in frame (a).

There are several control models that we could have used, but would any have a better chance than M-C? Selecting on groups at the level of entire ecosystems does not allow one ecosystem configuration to have any correlation with the next. Without the ability to follow fitness gradients, all possible ecosystems must be enumerated. In principle, migrations of uncorrelated groups of species can move between the local optima that defeat single-migrant dynamics. However, to move between local optima in the SBB landscape, the k species that are in one module must all change at once. The number of different possible k-species groups is

$2^k \binom{N}{k}$, so any uncorrelated group formation process will require a prohibitively large number of attempts before finding the exact group necessary to move from one local optimum to another of higher utility. Even if the decomposition were somehow known, randomly forming a group of the particular membership required is still exponential in the module size, which scales with the system size. Considering all 2^Z local attractors as M-C does is therefore the fastest control when $k = Z$, despite it not requiring any structural information. Therefore, any reasonable control that does not evolve symbiotic associations will not reliably find ecosystems of the same level of utility.

As described above, we use a mechanism that reinforces associations at the inter-species level – based on the surprise in the co-occurrence of species across several demes, with respect to the expected frequencies predicted from individual occurrence levels. The proposed mechanism is simpler than those suggested in previous models [10–12], and gives rise to qualitatively distinct results. In other work, we use a model that has an explicit population within each species, and each individual can evolve species-specific symbiotic associations (*i.e.,* the associations are individual traits) [8]. We use this model to explore the types of population structure that lead to the evolution of adaptively significant associations. In particular, we show that the high-level mechanism used in M-S need not be imposed at the species level, but can in fact can be manifested via the evolution of individual traits. As in the present model, the evolved associations create higher-level groupings and selective units.

Note that due to the transparent simplicity of the SBB landscape, it is actually possible to simplify M-S to only use 'all-or-nothing' associations in this case. However, such a simplification narrows the applicability. Elsewhere we have investigated landscapes in which associations with intermediate strengths find high-utility configurations that cannot be found with 'all-or-nothing' associations [19, 20]. It is worth noting that the algorithm proposed in [16] is able to efficiently solve hierarchical problems with some similar abstractions, in particular by model building from information at local optima.

In our model, we apply selection on migration groups such that an entire group is rejected if overall utility is not improved. This effectively causes the units of variation to be synonymous with the units of selection. An alternative scheme might allow groups to migrate together, but select on individual species. We suggest that because the individual selection is performed in the new context, with the entire migration group, the ultimate changes in ecosystem composition will not be significantly different than if selecting on groups as an entire unit. Recall that migration groups typically comprise species that were frequently found to co-occur in locally stable contexts. Thus, the individual species would be selected in the context of the particular group. We leave the verification that both schemes have qualitatively equivalent results for future work.

We have presented a model of the evolution of symbiotic associations where a separation of the temporal scales of changes in ecosystem composition and changes in species associations leads to the evolution adaptively significant complexes. This is in contrast to previous models of symbiosis with similar timescales

for both levels of adaptation, and results in a simpler and more biologically plausible model. Provided that the ecosystem is perturbed at a low frequency such that the transients are shorter than the average time between perturbations, the associations that form give rise to specific species groupings that can traverse rugged landscapes.

References

1. Thompson, J.N.: The Coevolutionary Process. Chicago (1994)
2. Mayr, E.: What Evolution is. Phoenix, London (2001)
3. Dunbar, H.E., Wilson, A.C.C., Ferguson, N.R., Moran, N.A.: Aphid thermal tolerance is governed by a point mutation in bacterial symbionts. PLoS Biology **5**(5) (2007) e96
4. Margulis, L.: The Symbiotic Planet. Phoenix, London (1998)
5. Khakhina, L.N.: Concepts of Symbiogenesis. Yale University Press (1992)
6. Maynard Smith, J., Szathmáry, E.: The major transitions in evolution. Oxford University Press (1995)
7. Margulis, L., Dolan, M.F., Guerrero, R.: Origin of the nucleus from the karyomastigontin amitochondriate protists. PNAS **97**(13) (2000) 6954–6959
8. Watson, R.A., Palmius, N., Mills, R., Powers, S., Penn, A.: Can selfish symbioses effect higher-level selection? In: ECAL. (2009)
9. Mills, R., Watson, R.A.: Variable discrimination of crossover versus mutation using parameterized modular structure. In: GECCO. (2007) 1312–1319
10. Watson, R.A., Pollack, J.B.: A computational model of symbiotic composition in evolutionary transitions. Biosystems **69**(2-3) (2003) 187–209
11. Defaweux, A., Lenaerts, T., van Hemert, J.I.: Evolutionary transitions as a metaphor for evolutionary optimisation. In: ECAL. (2005) 342–352
12. Mills, R., Watson, R.A.: Symbiosis, synergy and modularity: introducing the reciprocal synergy symbiosis algorithm. In: ECAL. (2007) 1192–1201
13. de Jong, E.D., Watson, R.A., Thierens, D.: On the complexity of hierarchical problem solving. In: GECCO. (2005) 1201–1208
14. Philemotte, C., Bersini, H.: A gestalt genetic algorithm: less details for better search. In: Procs GECCO. (2007) 1328–1334
15. Houdayer, J., Martin, O.C.: Renormalization for discrete optimization. Physical Review Letters **83** (1999) 1030–1033
16. Iclanzan, D., Dumitrescu, D.: Overcoming hierarchical difficulty by hill-climbing the building block structure. In: GECCO. (2007) 1256–1263
17. Krasnogor, N., Smith, J.: A tutorial for competent memetic algorithms. IEEE Transactions on Evolutionary Computation **9**(5) (2005) 474–488
18. Watson, R.A., Jansen, T.: A building-block royal road where crossover is provably essential. In: GECCO. (2007) 1452–1459
19. Mills, R., Watson, R.A.: Adaptive units of selection can evolve complexes that are provably unevolvable under fixed units of selection (abstract). In: ALIFE XI. (2008) 785
20. Mills, R., Watson, R.A.: Dynamic problem decomposition via evolved symbiotic associations. (in preparation)