Evolutionary Ecology Research, 2013, **15**: 43–59

Evolution of transmission mode in obligate symbionts

Devin M. Drown, Peter C. Zee*, Yaniv Brandvain# and Michael J. Wade

Department of Biology, Indiana University, Bloomington, Indiana, USA

ABSTRACT

Background: A host obtains symbionts by horizontal transmission when infected from the environment or contagiously from other hosts in the same generation. In contrast, vertical transmission occurs when a host obtains its symbionts directly from its parents. Either vertical or horizontal transmission can sustain an association between a host and its symbiont.

Questions: What evolutionary forces are necessary to evolve from an ancestral state of horizontal transmission to a derived state of vertical transmission?

Mathematical methods: We explore a general model of fitness interaction, including both additive and epistatic effects, between host and symbiont genes. Recursion equations allow us to analyse the short-term behaviour of the model and to study long-term deterministic effects with numerical iterations.

Key assumptions: Obligate interaction between a symbiont and a single host species with genetically determined horizontal and vertical transmission. No free-living symbionts or uninfected hosts and each host is infected by only a single symbiont genetic lineage (no multiple infections). No population structure.

Conclusions: Epistasis for fitness between host and symbiont genes, like that in a matching alleles model, is a necessary condition for the evolution of vertical from horizontal transmission. Stochastic individual-based simulations show that (1) mutation facilitates the switch to vertical transmission and (2) vertical transmission is a stable evolutionary endpoint for a matching alleles model.

Keywords: co-evolution, epistasis, horizontal transmission, linkage disequilibrium, vertical transmission.

INTRODUCTION

Intimate host–symbiont associations are ubiquitous in natural communities (Bright and Bulgheresi, 2010; Toft and Andersson, 2010). Many exhibit life-history adaptations to ensure horizontal or vertical transmission, and some hosts employ both modes of symbiont acquisition (Bright and Bulgheresi, 2010; Sachs *et al.*, 2011). For instance, most plant species are known to form symbiotic associations with mycorrhizal fungi that colonize roots (Rodriguez *et al.*, 2009). Many legumes

Correspondence: D.M. Drown, 1001 E Third Street, Indiana University, Bloomington, IN 47405, USA. E-mail: dmdrown@indiana.edu

**Current address*: Department of Biology, Stanford University, Stanford, CA 94305, USA.

#*Current address*: Center for Population Biology, University of California, Davis, CA 95616, USA.

form beneficial genotype-specific symbioses when infected with nitrogen-fixing rhizobia (Sachs and Simms, 2008). Corals and certain sponges maintain mutualistic associations with photosynthetic symbionts acquired from the environment (Goulet, 2006; Taylor *et al.*, 2007; Usher, 2008). Obligate associations require mechanisms for trans-generation transmission (Bright and Bulgheresi, 2010). Many, if not most, intracellular symbioses are vertically transmitted from adult hosts to host progeny, a phenomenon called 'hereditary symbiosis' (Clay, 1994; Bright and Bulgheresi, 2010). Indeed, symbionts have been classified into two categories: (1) P-symbionts, obligate primary endosymbionts with historically long host associations; and (2) S-symbionts, facultative secondary symbionts with recent host associations and the capability of returning to an ancestral, free-living state (Moya *et al.*, 2008). In addition, the evolutionary merging of host and symbiont genomes from different species or kingdoms has facilitated major innovations in eukaryotic life (Maynard Smith and Szathmáry, 1997; Michod, 1999).

The two modes of symbiont transmission are horizontal or vertical. With horizontal transmission, hosts obtain symbionts either directly from the environment or contagiously by infection from other hosts in the same generation. With vertical transmission, a host obtains its symbionts directly from its parents, generating fidelity between host genes and symbionts.

The mode of transmission affects the conflicts caused when two different genomes exert dual control of a phenotype (Frank, 1996a, 1997; Wade, 2007). It also affects the covariance between genes in the symbiont genome and in the host nuclear and organelle genomes (Brandvain *et al.*, 2011). This covariance, called the 'inter-species disequilibrium' (Sanchez *et al.*, 2000), hereafter ID, is analogous to linkage disequilibrium (LD) between alleles at different loci within a single genome. When fitness is a function of the interaction between genomes, the selection responses of *both host and symbiont* are proportional to the ID between genomes (Goodnight, 1987, 1988, 1995; Barton and Turelli, 2004; Hill *et al.*, 2006). Vertical transmission increases the efficiency of response to selection on host–symbiont gene combinations and reduces conflict between them (Frank, 1996a, 1997; Wade and Goodnight, 2006; Wade, 2007).

We investigate the kind of selection necessary to maintain beneficial host–symbiont pairs. When some host–symbiont combinations have high fitness but others do not, we show that indirect selection favours mechanisms that maintain favoured host–symbiont combinations in much the same way that recombination modifiers are favoured by epistatic selection (Feldman *et al.*, 1980; Otto and Feldman, 1997; Otto and Barton, 2001; Otto and Lenormand, 2002). One such mechanism is a transition from horizontal to vertical symbiont transmission. We model an obligate mutualistic interaction between a symbiont and a single host species with genetically determined horizontal and vertical transmission. We explore a general model of fitness interaction, including both additive and epistatic effects, between host and symbiont genes to determine which kinds of selection generate genetic covariance (ID) between host and symbiont and thereby allow indirect selection to increase the frequency of a transmission modifier. Our analyses are deterministic as well as stochastic. We discuss the implications of our findings for other co-evolutionary interactions such as virulence.

MODEL

Our model population consists of two species, a host and its obligate mutualistic symbiont. We assume no free-living symbionts, uninfected hosts, or multiple infections. Each species has a biallelic locus that contributes to fitness (A, a: host; B, b: symbiont). Interactions

between these alleles in the host and symbiont affect fitness in both species. We consider both species to be haploids with clonal reproduction in discrete, non-overlapping fully synchronized generations. We model a lifecycle ordered as selection → reproduction → infection. The model is symmetric with respect to the transmission locus with alternative alleles, M and m, which determine the rate of vertical transmission (see below). That is, our results do not depend upon whether the transmission locus is in the symbiont genome or in the host genome. For the sake of exposition, we will assume that the host controls symbiont transmission, as supported by empirical evidence (Buchner, 1965; Frank, 1996b).

We let m hosts transmit their symbiont vertically at a rate V_m and M hosts at a rate V_M where $V_M = V_m + \delta$. Neither allele (m nor M) carries a direct fitness benefit or cost. By altering the sign and magnitude of δ, we can model increases ($\delta > 0$) or decreases ($\delta < 0$) in the rate of vertical transmission. Hence, for each host haplotype, a fraction of its offspring is infected with its symbionts (vertical transmission), while the remaining offspring are produced without symbionts and become infected with symbionts randomly chosen from other hosts (horizontal transmission). These two fractions are referred to as the *vertical pool* and *horizontal pool*, respectively.

Our model of fitness interaction between genomes includes multiplicative and epistatic effects (Table 1). Host–symbiont combinations with alleles *a*/*b* have their fitness decreased by α, the multiplicative effect. We also describe an epistatic effect, ε, as a deviation from the multiplicative effect of carrying both alternative alleles. We describe the system of recursions of each of the eight three-locus host–symbiont pairs (4 host haplotypes × 2 symbiont haplotypes; see Appendix A, equations A1–A5).

ANALYSES

Short-term evolution

For a general fitness model with multiplicative (α) and epistatic effects (ε) (Table 1), we calculate the complete set of recursions (equations A1–A5) and present our notation for allele frequencies (P_i) and two- and three-way disequilibria across species (ID_{ij} and ID_{ijk} respectively) (see Appendix B, equations B1–B7). From these, we describe the change in the frequency of the transmission modifier (P_M) across one generation given an arbitrary degree of existing linkage disequilibrium as

$$\Delta P_M = \frac{(\alpha^2 + \varepsilon)(ID_{ABM} + D_{AM}P_A + ID_{BM}P_B) + (\alpha - \alpha^2 - \varepsilon)(D_{AM} + ID_{BM})}{\bar{w}}, \tag{1}$$

where population mean fitness is

$$\bar{w} = (1 - \alpha)^2 + \varepsilon + (\alpha^2 + \varepsilon)(ID_{AB} + P_A P_B) + (\alpha - \alpha^2 - \varepsilon)(P_A + P_B). \tag{2}$$

Table 1. Fitness (w_{ij}) of host/symbiont

	Symbiont *B*	Symbiont *b*
Host *A*	1	$1 - \alpha$
Host *a*	$1 - \alpha$	$(1 - \alpha)^2 + \varepsilon$

From equations (1, 2), we see that selection on the transmission modifier is indirect and acts through both within- and between-genome disequilibria, D and ID respectively. Thus, to understand the evolution of the transmission modifier, we must understand the necessary and sufficient conditions for generating non-zero values of these disequilibria.

We assume that all loci are initially in complete linkage equilibrium (e.g. $P_{ABM} = P_A P_B P_M$), and find that the changes in disequilibria are

$$\Delta ID_{AB} = \frac{\varepsilon(1 - P_A)P_A(1 - P_B)P_B(V_m + \delta P_M)}{[(1 - \alpha)^2 + \varepsilon + (\alpha^2 + \varepsilon)(P_A P_B) + (\alpha - \alpha^2 - \varepsilon)(P_A + P_B)]^2}, \tag{3}$$

$$\Delta D_{AM} = 0, \tag{4}$$

$$\Delta ID_{BM} = 0, \tag{5}$$

$$\Delta ID_{ABM} = \frac{\varepsilon\delta(1 - P_A)P_A(1 - P_B)P_B(1 - P_M)P_M}{[(1 - \alpha)^2 + \varepsilon + (\alpha^2 + \varepsilon)(P_A P_B) + (\alpha - \alpha^2 - \varepsilon)(P_A + P_B)]^2}. \tag{6}$$

From equations (3–6), we see epistasis ($\varepsilon \neq 0$) is necessary for generating disequilibria across species (ID_{AB}) and between that interspecific ID and the modifier locus (ID_{ABM}). The signs of these disequilibria are also dependent on ε because, by assumption, $V_m \geq 0$, $\delta > 0$. (Note that the denominators are always positive.) Positive disequilibria are generated (equations 3–6) when $\varepsilon > 0$, and $ID_{ABM} > 0$ results in indirect selection for increased vertical transmission.

Numerical iterations

We explored the longer-term dynamics of our system numerically by iterating the full set of recursions (equations A1–A6) using MATLAB version R2010a (The Mathworks) for 10,000 generations. We tracked the frequency of the transmission modifier allele, *M*, assuming different parameter combinations of δ, α, and ε.

For each parameter combination, we investigated 19 initial allele frequencies for both host and symbiont across the range [0.05, 0.95] in 0.05 increments (19 frequencies for host alleles, *a* and *A*, factorially combined with 19 frequencies for symbiont alleles, *b* and *B*; a total of 361 frequency combinations). Initially, hosts and symbionts were associated with one another proportional to their frequencies (e.g. host and symbiont alleles were in linkage and interspecies equilibrium: $ID_{AB} = D_{AM} = ID_{BM} = ID_{ABM} = 0$; see Appendix B for detailed definitions), and the transmission modifier, which was also in linkage equilibrium with both loci, was initially present at a low frequency ($P_M = 0.05$).

For the fitnesses in Table 1, we varied the strength of the interaction across three orders of magnitude ($\alpha = \{0.01, 0.1, 1\}$). We varied the degree of epistasis from multiplicative ($\varepsilon = 0$), to additive ($\varepsilon = -\alpha^2$) to matching alleles ($\varepsilon = 2\alpha - \alpha^2$) (Frank, 1992). We also included the neutral case, $\alpha = \varepsilon = 0$.

The ancestral transmission was completely horizontal ($V_m = 0$). We introduced vertical transmission modifiers, ranging across three orders of magnitude ($\delta = \{0.01, 0.1, 1\}$). The case $\delta = 1$ represents a switch from completely horizontal to completely vertical transmission. We also examined hitchhiking of a neutral modifier by setting $\delta = 0$.

To measure selection, we compared the frequency of the transmission modifier allele to that of a neutral modifier at generation 10,000. Runs where the difference in frequency was greater than 1×10^{-5} were considered favourable selection. Table 2 summarizes the

Table 2. Proportion of combinations increasing modifier

α	ε	$V_M = 0.01$	$V_M = 0.1$	$V_M = 1$
(a) Neutral				
0	0	0	0	0
(b) Multiplicative $\varepsilon = 0$				
0.01	0	0	0	0
0.1	0	0	0	0
1	0	0	0	0
(c) Additive $\varepsilon = -\alpha^2$				
0.01	−0.0001	0	0	0
1	−0.01	0	0	0
(d) Matching alleles $\varepsilon = 2\alpha - \alpha^2$				
0.01	0.0199	0.9280	1	1
0.1	0.19	1	1	1
1	1	1	1	1

evolution of vertical transmission as a proportion of the 361 initial conditions of the A and B frequencies that showed an increase for each parameter combination.

Because selection on the transmission modifier was indirect, we tracked: (1) ID_{AB}, the interspecies disequilibrium at fitness loci; (2) D_{AM}, the disequilibrium between the selected locus and the modifier locus within the host; (3) ID_{BM}, the interspecies disequilibrium between the selected locus in the symbiont and the modifier locus in the host; and (4) ID_{ABM}, the three-way association between both host loci and symbiont locus. Table 3 summarizes maximum values and terminal values of ID_{AB} and ID_{ABM} averaged over the initial host–symbiont conditions for each parameter combination.

As expected from our single-generation analyses (above), with no selection on the interaction ($\alpha = 0$) or a lack of positive epistasis ($\varepsilon < 0$), the change in modifier frequency was equivalent to neutral (Table 2a–c) and no disequilibria were generated (Table 3a–c). Additive fitnesses (representative dynamic shown in Fig. 1) resulted in transient negative disequilibria, so that the transmission modifier decreased by a small amount (Fig. 1b). Table 3c shows that these negative associations were not sustained.

As expected from our single-generation analyses (above), with positive epistasis ($\varepsilon > 0$) and the matching alleles model (Frank, 1992), the change in frequency of the vertical transmission modifier exceeded that of the neutral case (Table 2d and Fig. 1b, d). Both of the necessary disequilibria were positive under the matching alleles regime (Fig. 1b and Table 3d).

Individual-based stochastic simulations

We created an individual-based stochastic simulation in MATLAB version R2010a (The Mathworks) to explore the effects of mutation and finite population size on the evolution of transmission under the matching alleles regime ($\varepsilon = 2\alpha - \alpha^2$). This is an important extension

Table 3. Generation of interspecies disequilibrium

		ID_{AB}			ID_{ABM}		
α	ε	$V_M = 0.01$	$V_M = 0.1$	$V_M = 1$	$V_M = 0.01$	$V_M = 0.1$	$V_M = 1$
(a) Neutral							
0	0	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)
(b) Multiplicative $\varepsilon = 0$							
0.01	0	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)
0.1	0	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)
1	0	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)
(c) Additive $\varepsilon = -\alpha^2$							
0.01	−0.0001	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)
1	−0.01	0(0)	0(0)	0(−0.0002)	0(0)	0(0)	0(−0.0002)
(d) Matching alleles $\varepsilon = 2\alpha - \alpha^2$							
0.01	0.0199	0(0)	0(0)	0.0387(0.0387)	0(0)	0(0)	0.019(0.0222)
0.1	0.19	0(0)	0(0)	0.0387(0.0387)	0(0)	0(0)	0.019(0.0223)
1	1	0(0.0001)	0.0003(0.0019)	0.0387(0.0387)	0(0.0001)	0(0.001)	0.019(0.0222)

Note: Mean final (maximum magnitude) ID value of 361 initial starting conditions of interspecies disequilibrium with 10,000 generations of evolution.

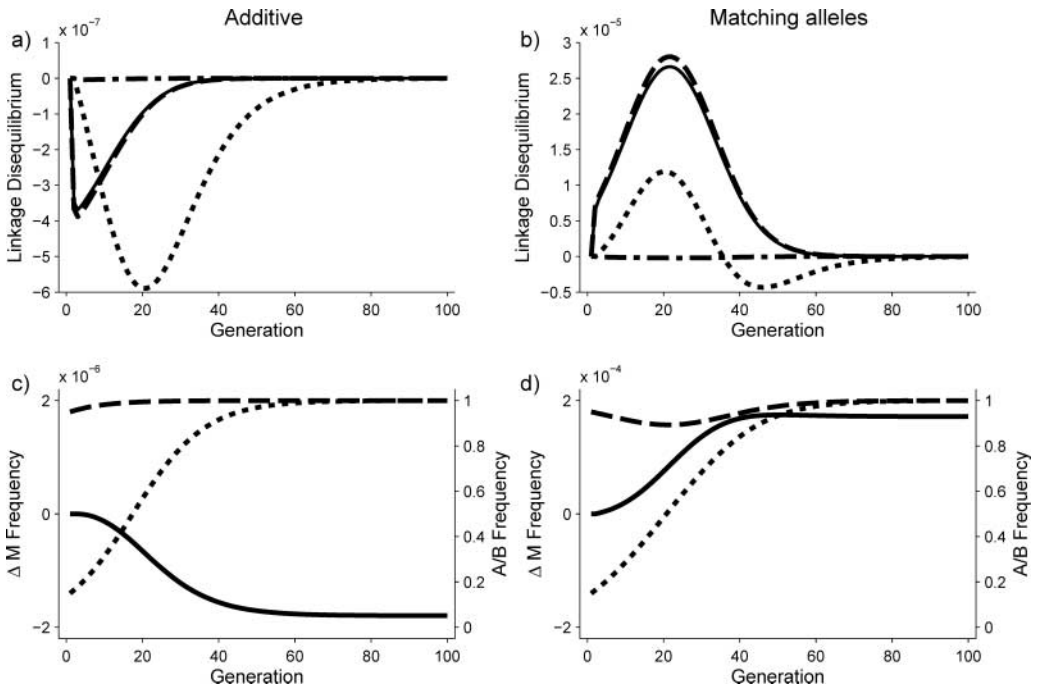

Fig. 1. Sample long-term dynamics of linkage disequilibrium and allele frequency changes given different regimes of fitness interaction. Left column (a, c): additive; right column (b, d): matching alleles. (a, b) The four different values of linkage disequilibrium are plotted (dashed line, ID_{AB}; dotted line, D_{AM}; dash-dot line, ID_{BM}; solid line, ID_{ABM}). (c, d) To show a long-term change in modifier frequency, we plot the change from the initial frequency ($M = 0.05$) (solid line) along with the frequency of the interaction alleles (dotted line: frequency of A) and symbiont (dashed line: frequency of B). Note the separate axes scales. We assumed the ancestral state was complete horizontal transmission ($V_m = 0$) and the modifier effect (δ) increased the rate of vertical transmission by 0.1. The strength of interaction (α) was 0.1 and $\varepsilon = -0.01$ (left column) and $\varepsilon = 0.19$ (right column). The initial allele frequencies for the interaction loci were $A = 0.15$ and $B = 0.95$.

of our model because it allows for the stochastic generation of ID, which can then be harnessed by selection.

Initial frequencies for the A and B loci were determined by Latin Hypercube sampling implemented in MATLAB to create 2000 initial conditions. In all cases, the M allele was started at a frequency of 0.05. We investigated the change in the frequency of the modifier allele and its fixation probability under a broad range of parameter values.

An initial population of N hosts each containing one symbiont (N symbionts in total) was created from the initial allele frequencies. The forwards and backwards mutation rates, μ, were symmetrical, equal, and independent at each interaction locus. There was no mutation at the modifier locus. To simulate selection, we drew N host–symbiont pairs with replacement weighted by the matching alleles fitness regime (Table 1c).

For all simulations, we assumed the unmodified rate of vertical transmission to be zero ($V_m = 0$), so that all m individuals were placed into the horizontal pool after selection, where host–symbiont pairs were dissociated and then randomly re-paired. Those with M alleles were placed at rate V_M in the vertical pool, wherein the host–symbiont pairs remained

associated. This process maintained population size and allele frequencies, but removed any pre-existing association of symbiont haplotypes with host haplotypes during horizontal transmission. The final step of the life cycle was to combine the horizontal and vertical pools to reconstitute the population for the next generation. We kept track of allele and haplotype frequency dynamics as well as the temporal dynamics of linkage disequilibrium. Each simulation was run for 100,000 generations.

Comparison of stochastic simulation with deterministic numerical iterations

As expected, results of the stochastic simulations were most similar to those of the numerical deterministic iterations when mutation was weak and population size was large. For instance in Fig. 2 (middle column), for $N = 10{,}000$ and no mutation ($\mu = 0$), the modifier allele does not evolve to high frequency and allelic diversity is quickly lost from the system as the interaction alleles approach fixation (Fig. 2b). In addition, the stochastic simulation produced the same transient dynamics of the linkage disequilibria (compare Fig. 2g and 2h). Unlike the numerical iterations where evolution at the transmission modifier stops, small changes in frequency persist as random genetic drift influences the system (compare Fig. 2a and 2b).

With a strong mutation rate ($\mu = 0.01$), the initial dynamics remain similar to the numerical iterations (Fig. 2, right column). The A and B alleles increase in frequency, as does the *ABm* association (Fig. 2f, green line). Mutations maintain polymorphism at the interaction loci, allowing continued evolution and fixation of the *M* allele (Fig. 2c, black line).

Probability of fixation

We calculated the probability of fixation of a vertical transmission modifier for a factorial combination of parameter values ($\alpha = \{0, 0.01, 0.1, 1\}$; $\delta = \{0, 0.01, 0.1, 1\}$; $\mu = \{0, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$; $N = \{1000, 10{,}000\}$; total of 160 combinations) under the matching allele regime ($\varepsilon = 2\alpha - \alpha^2$). For each combination, we ran 2000 independent simulations with Latin Hypercube sampled starting A and B allele frequencies across the range [0, 1] for a total of 320,000 simulations. The same 2000 initial allele frequencies were used across the 160 parameter combinations to minimize initial condition effects. We ran each simulation for a maximum of 100,000 generations, stopping if the *M* allele was either lost or fixed. All but two of the 320,000 simulations ended with either loss or fixation before the cut-off. The two cases were treated as a failure to fix. The probability of fixation was calculated as the number of modifier fixations divided by the number of runs (2000). These results are summarized in Fig. 3.

We included two controls. One with $V_M = 0$ ($\delta = 0$) represents a neutral modifier and controls for the modifier effect. The probability of fixation in every case was equal to the initial frequency as expected (Fig. 3, grey circles) (Hedrick, 2005, p. 306). We also included cases where the interaction had no effect ($\alpha = \varepsilon = 0$), eliminating indirect selection on the modifier. Just as in the neutral modifier case, we found no effect on a transmission modifier: it fixed with probability equal to its initial frequency (5%; Fig. 3, leftmost set of points within each panel).

Genetic drift limited the evolution of vertical transmission by allowing stochastic loss of the modifier even in the presence of mutation. When mutation rates were zero, our

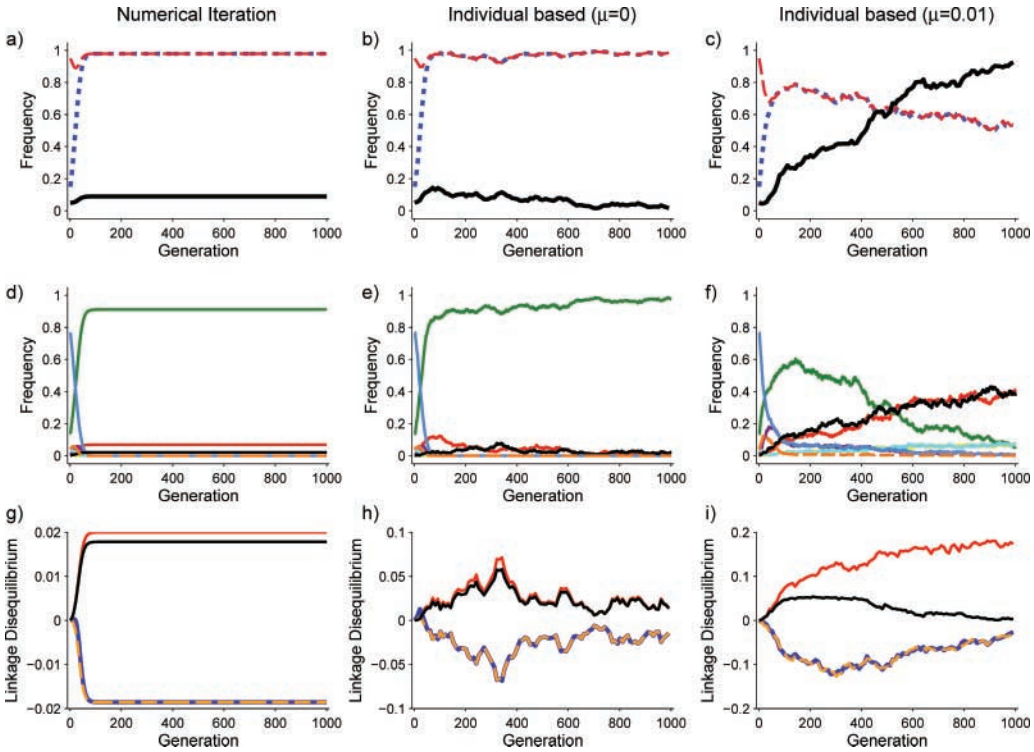

Fig. 2. Comparison of numerical iterations and individual-based simulations under the matching alleles regime and strong transmission modifier effect. The top row shows the long-term change in modifier frequency; we plot the change from the transmission modifier frequency (black) together with the frequency of the interaction alleles (blue: frequency of A) and symbiont (red: frequency of B). The second row plots the nine haplotype frequencies (red: *ABM*; green: *ABm*; yellow: *AbM*; purple: *Abm*; cyan: *aBM*; blue: *aBm*; black: *abM*; dashed orange: *abm*). The third row plots the four different values of linkage disequilibrium (red: ID_{AB}; blue: D_{AM}; orange: ID_{BM}; black: ID_{ABM}). The left column plots the numerical iterations. The middle and right columns plot the results from the individual-based simulations. The individual-based simulations differ in the presence of mutation at the interaction loci. We set $\mu = 0$ (middle column) or $\mu = 0.01$ (right column). In all cases, we assumed the ancestral state was complete horizontal transmission ($V_m = 0$) and the modifier effect (δ) increased the rate of vertical transmission to 1. The strength of interaction (α) was 0.1 and $\varepsilon = 0.19$. The initial allele frequencies for the interaction loci were $A = 0.15$ and $B = 0.95$. The transmission modifier (M) was introduced at a frequency of 0.05. For the individual-based simulations, $N = 10{,}000$.

simulations behaved similarly to the numeric iterations of the full recursions (Fig. 3a and 3f). Only in the cases where the modifier is especially strong ($\delta = 1$) and the alternative state is complete vertical transmission ($V_M = 1$), is there an increase in the frequency of fixation above the neutral prediction (Fig. 3a and 3f, red circles).

Modifiers to complete vertical transmission ($V_M = 1$) have the highest probability of fixation across all parameter space (Fig. 3, red circles). Weaker modifiers fix (Fig. 3e and 3j) only when the interaction is strong ($\alpha = 1$) or when mutation rates are high ($\mu = 0.1$). At low mutation rates, even strong selection and high V_M show only modest probabilities of fixation. The probability of fixation increases substantially with an increase in mutation rate.

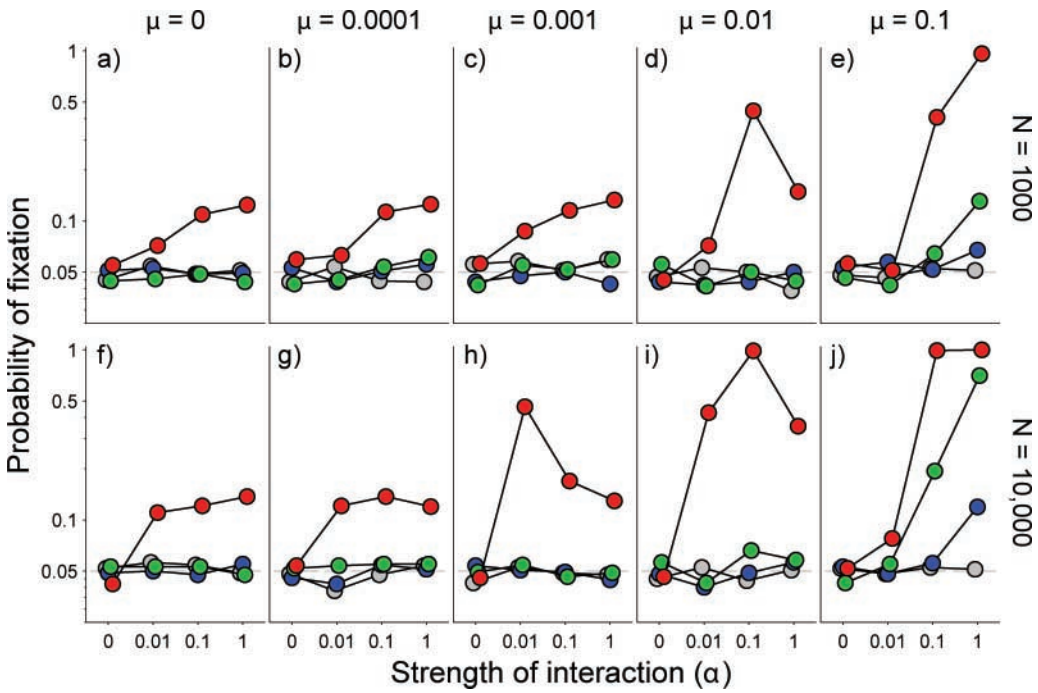

Fig. 3. Individual-based simulations of the probability of fixation assuming the matching allele regime of fitness interaction. Three different modified vertical transmission rates ($V_M = 0.01$, blue; 0.1, green; 1, red) were compared with a modifier of neutral effect ($V_m = V_M = 0$, grey). The probability of fixation is calculated as the fraction of the 2000 simulations that resulted in the fixation of the modifier allele (M) within 100,000 generations. Host population sizes varied ($N = 1000$, top row; 10,000, bottom row) as well as the rate of mutation ($\mu = 0$, first column; 10^{-4}, second column; 10^{-3}, third column; 10^{-2}, fourth column; 10^{-1}, fifth column). Note that the y-axis is on a log scale. The grey dashed line indicates the null expectation (0.05 probability of fixation).

Vertical transmission as an evolutionary attractor

To investigate the evolutionarily stable transmission rate, we modified our stochastic simulation to allow mutation at the transmission locus at a rate of 10^{-4} per individual per generation. Given a mutation, its small effect was drawn from the uniform distribution [–0.1, 0.1] with a 0.5 probability; otherwise, the mutation was of random effect and a new vertical transmission rate for that individual was drawn from the uniform distribution [0, 1]. Vertical transmission rates above 1 or below 0 were truncated to 1 or 0 respectively. The null expectation, in the absence of selection, is a vertical transmission rate of 0.5 owing to mutation pressure. All other aspects of the simulations were kept the same as the previous section.

We examined a factorial combination of parameter values ($\alpha = \{0, 0.01, 0.1, 1\}$; $\delta = \{0, 0.01, 0.1, 1\}$; $\mu = \{0, 10^{-3}, 10^{-2}, 10^{-1}\}$; $N = \{1000, 10{,}000\}$; total of 32 combinations). For each combination, 16 independent simulations with initial A and B frequencies were sampled from the range [0.2, 0.8] at intervals of 0.2. Simulations were run for 10 million generations or when population mean transmission rate reached a pseudo-equilibrium state.

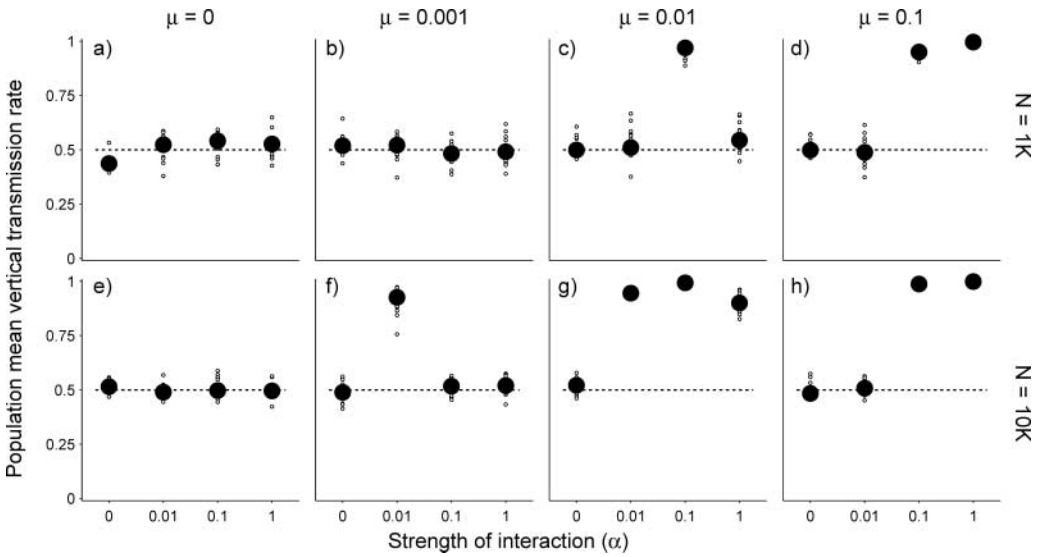

Fig. 4. Individual-based simulations of the evolutionary attracting vertical transmission rate assuming the matching allele regime of fitness interaction. Simulations were run for a maximum of 10 million generations. The population mean vertical transmission rate is calculated as the mean of the per generation population means across the final one million generations (small open circles). The mean value for the 16 independent runs for each parameter combination is shown (large black circle). Host population sizes varied ($N = 1000$, top row; 10,000, bottom row) as well as the rate of mutation at the interaction loci ($\mu = 0$, first column; 10^{-3}, second column; 10^{-2}, third column; 10^{-1}, fourth column). The rate of mutation at the transmission locus was fixed at 10^{-4} for all simulations. The dotted line indicates the null expectation (0.5 vertical transmission rate).

During a simulation run, every 100 generations, we calculated the population mean transmission rate. If the variance across 100,000 generations of the population mean was $< 10^{-3}$, we assumed we had reached a pseudo-equilibrium.

Results are summarized in Fig. 4 and mirror those of the fixation analyses. In cases where there was a high probability of a complete vertical modifier fixing, complete vertical transmission was also an evolutionary attractant (compare Fig. 3, red circles and Fig. 4). Selection never favoured complete horizontal transmission as a stable end-point; owing to mutation pressure, most simulations ended at 0.5 vertical transmission. We found that mutations at the interaction loci lead more often to the fixation of complete vertical transmission (compare Fig. 4e and 4g). As with our previous analysis, genetic drift reduced the opportunity for vertical transmission to evolve in the system. This is evident when contrasting results with small N (e.g. Fig. 4c) to those with large N (e.g. Fig. 4g).

DISCUSSION

Our model investigates how fitness interactions between host and symbiont affect the evolution of symbiont transmission mode. Previous theoretical work (Brandvain *et al.*, 2011) showed that even small amounts of horizontal transmission erode interspecific linkage disequilibrium in the absence of selection. In that work, transmission mode was not permitted to evolve. In our model, we allowed transmission itself to be under genetic control of

a modifier locus. We find that selection favours vertical transmission under limited circumstances in the absence of a direct fitness benefit. A transmission modifier is favoured only when there is positive interspecific, epistatic gene combinations, and not with only additive or multiplicative gene action.

With positive epistasis for fitness, alleles in one species are favoured only on specific genotypic backgrounds of the other, as is the case with the matching alleles model. Such positive epistasis for fitness generates sufficient interspecific disequilibria (equations 3 and 6) to permit hitchhiking and indirect selection of the vertical transmission modifier (equation 1) through a three-way disequilibrium. Individuals bearing a vertical transmission modifier of large effect transmit not only their genes but also the symbiont background on which they are most fit to their offspring with higher fidelity than do individuals lacking the modifier. Put differently, the vertical transmission modifier allows its bearers to effectively 'lock in' the high fitness environment, thereby increasing the heritability of fitness. In contrast, horizontal transmission immediately removes trans-specific disequilibria (Brandvain *et al.*, 2011), guaranteeing a random, non-heritable experience of the environment of the other species.

Strong epistasis in fitness generates positive interspecific disequilibria. However, genetic drift generates negative linkage disequilibrium (Barton, 2010) diminishing the probability of fixation of vertical transmission modifiers at small population sizes. Even with strong selection, the hitchhiking of the transmission modifier can be transient as selection fixation erodes genetic variation at the interacting loci. Mutation at the interaction loci allows continued selection for a transmission modifier. Complete vertical transmission appears to be a stable evolutionary end-point for matching allele selection regimes. While it might seem intuitive that vertical transmission would be selected for with an obligate host–symbiont interaction, we found that without a force maintaining polymorphism, like mutation, the conditions favouring the evolution of vertical transmission of an obligate symbiont were limited in our model. With other forms of selection, such as when symbionts are essential and contagious infection is uncertain, selection may favour vertical transmission as a mechanism to ensure that all offspring inherit a symbiont.

Although the conditions favouring vertical transmission appear limited in theory, vertical symbiont transmission, or 'hereditary symbiosis' (Clay, 1994), is ubiquitous in nature. For example, *Rickettsia peacockii*, which is only transmitted vertically after losing the ability to pass between host cells and enter the tick saliva (Simser *et al.*, 2005). This loss of molecular mechanism transforms horizontal into vertical transmission.

We can infer the mode of symbiont transmission and test the predictions of our model by estimating interspecific genetic covariance (i.e. ID). For example, Hurtado *et al.* (2003) found a strong association between mitochondrial loci of the vent clam, *Calyptogena magnica*, and loci in its sulphur-oxidizing proteo-bacterial symbiont and inferred high levels of maternal symbiont transmission. However, the absence of trans-specific disequilibria (ID) does not necessarily imply horizontal transmission (see Brandvain *et al.*, 2011) and, conversely, high values of ID may reflect strong selection more than strictly vertical transmission. For rigorous inference, the theoretical expectation of the distribution of ID between host and symbiont genomes is essential, especially as such studies based on genomic and sequence variation become more common (Wu *et al.*, 2006, 2009; Toft and Andersson, 2010).

Our results allow us to make predictions of patterns potentially observable in nature. We found that a high rate of mutation in the genes responsible for the host–symbiont fitness interaction is important to the fixation of vertical transmission. There is evidence of high

rates of substitution in intracellular bacteria that are vertically transmitted, consistent with this prediction (Douglas, 2010). Faster rates of evolutionary change at loci responsible for the interspecific interaction should also be observed as elevated levels of sequence divergence between symbiont populations or species.

Unlike our model with reciprocally positive host and symbiont fitness interactions, the interests of the host and symbiont are less likely to be perfectly aligned in nature (i.e. a genotype–genotype combination may be good for the host and bad for the symbiont). The transition from horizontal to vertical transmission mode may be associated with a switch from pathogenic to mutualistic interactions. Increased vertical transmission should more closely align the interests of the symbiont with those of the host and potentially reduce virulence (Ewald, 1987; Bull *et al.*, 1991). In theory, co-inheritance of genomes (host and symbiont) via vertical transmission increases the efficiency of selection and decreases the amount of genomic conflict (Wade and Goodnight, 2006; Wade, 2007). With obligate vertical transmission, 'the organism has no opportunity to escape from the consequences of any deleterious impacts it might have on its partner' (Douglas, 2010, p. 68). However, limited amounts of horizontal transmission immediately break down host–symbiont combinations and can eliminate selection favouring reduced virulence. Thus, partial vertical transmission may not be completely sufficient to alleviate the interspecies conflict (Douglas, 2010). For example, in hosts with sexual reproduction and strict maternal inheritance of symbionts, males can be seen as a dead end for symbiont infection (but see Unckless and Herren, 2009; Wade and Brandvain, 2009). Empirical evidence suggests that the host has won the conflict over sexual ratio manipulation (Douglas, 2010).

Epidemiological models have provided conflicting results about the relationship of transmission mode and virulence (Ewald, 1987; Yamamura, 1993; Lipsitch *et al.*, 1996). Cases of asymmetric fitness interactions may yield different selection pressures on host and symbiont, thereby changing selection on transmission. Investigation of a wider range of host–symbiont fitness interactions is warranted.

We propose that the statistical character of gene combinations determines whether host and symbiont fitnesses covary positively or negatively. This in turn determines whether or not selection favours 'inter-specific gene combinations' (positive fitness covariance) or an antagonistic, evolutionary arms race between host and symbiont (negative fitness covariance). Just as recombination between nuclear genes occurs *within local populations* and does not affect the among-population component of LD, host population genetic structure reduces mixing of host and symbiont haplotypes and preserves ID (Y. Brandvain and M.J. Wade, unpublished). When horizontal transmission occurs primarily between genetically related hosts (e.g. family members), its effect of reducing ID is greatly weakened. Furthermore, reductions in symbiont effective population size may play as important a role in the extreme reductive genome evolution seen in some mutualistic endo-symbionts as the selective elimination of metabolic redundancy and the integration of host and symbiont metabolism (Lynch and Conery, 2003; Lynch, 2006; Kuo *et al.*, 2009).

ACKNOWLEDGEMENTS

Discussion with the Theta working group at Indiana University improved this research. Amy Dapper and Dave Van Dyken provided specific help on our initial model. Jefferson Davis at the Stat/Math Center at Indiana University provided assistance optimizing Matlab code and running simulations on the Quarry computer cluster. The manuscript was improved by comments from Michael Doebeli,

Maren Friesen, Curt Lively, and anonymous reviewers. The work was supported by NIH grant R01GM084238 to M.J.W.

REFERENCES

Barton, N.H. 2010. Genetic linkage and natural selection. *Phil. Trans. R. Soc. Lond. B*, **365**: 2559–2569.

Barton, N.H. and Turelli, M. 2004. Effects of genetic drift on variance components under a general model of epistasis. *Evolution*, **58**: 2111–2132.

Brandvain, Y., Goodnight, C.J. and Wade, M.J. 2011. Horizontal transmission rapidly erodes disequilibria between organelle and symbiont genomes. *Genetics*, **189**: 397–404.

Bright, M. and Bulgheresi, S. 2010. A complex journey: transmission of microbial symbionts. *Nature Rev. Microbiol.*, **8**: 218–230.

Buchner, P. 1965. *Endosymbiosis of Animals with Plant Microorganisms.* New York: Wiley Interscience.

Bull, J.J., Molineux, I.J. and Rice, W.R. 1991. Selection of benevolence in a host–parasite system. *Evolution*, **45**: 875–882.

Clay, K. 1994. Hereditary symbiosis in the grass genus *Danthonia. New Phytol.*, **126**: 223–231.

Douglas, A.E. 2010. *The Symbiotic Habit.* Princeton, NJ: Princeton University Press.

Ewald, P.W. 1987. Transmission modes and evolution of the parasitism–mutualism continuum. *Ann. NY Acad. Sci.*, **503**: 295–306.

Feldman, M.W., Christiansen, F.B. and Brooks, L.D. 1980. Evolution of recombination in a constant environment. *Proc. Natl. Acad. Sci. USA: Biol. Sci.*, **77**: 4838–4841.

Frank, S.A. 1992. Models of plant pathogen coevolution. *Trends Genet.*, **8**: 213–219.

Frank, S.A. 1996a. Host–symbiont conflict over the mixing of symbiotic lineages. *Proc. R. Soc. Lond. B*, **263**: 339–344.

Frank, S.A. 1996b. Host control of symbiont transmission: the separation of symbionts into germ and soma. *Am. Nat.*, **148**: 1113–1124.

Frank, S.A. 1997. Models of symbiosis. *Am. Nat.*, **150**: S80–S99.

Goodnight, C.J. 1987. On the effect of founder events on epistatic genetic variance. *Evolution*, **41**: 80–91.

Goodnight, C.J. 1988. Epistasis and the effect of founder events on the additive genetic variance. *Evolution*, **42**: 441–454.

Goodnight, C.J. 1995. Epistasis and the increase in additive genetic variance: implications for phase-1 of Wright's shifting-balance process. *Evolution*, **49**: 502–511.

Goulet, T.L. 2006. Most corals may not change their symbionts. *Mar. Ecol. Progr. Ser.*, **321**: 1–7.

Hedrick, P.W. 2005. *Genetics of Populations.* Sudbury, MA: Jones & Bartlett.

Hill, W.G., Barton, N.H. and Turelli, M. 2006. Prediction of effects of genetic drift on variance components under a general model of epistasis. *Theor. Popul. Biol.*, **70**: 56–62.

Holsinger, K.E., Feldman, M.W. and Christiansen, F.B. 1984. The evolution of self-fertilization in plants: a population genetic model. *Am. Nat.*, **124**: 446–453.

Hurtado, L.A., Mateos, M., Lutz, R.A. and Vrijenhoek, R.C. 2003. Coupling of bacterial endosymbiont and host mitochondrial genomes in the hydrothermal vent clam *Calyptogena magnifica. Appl. Environ. Microbiol.*, **69**: 2058–2064.

Kuo, C.H., Moran, N.A. and Ochman, H. 2009. The consequences of genetic drift for bacterial genome complexity. *Genome Res.*, **19**: 1450–1454.

Lipsitch, M., Siller, S. and Nowak, M.A. 1996. The evolution of virulence in pathogens with vertical and horizontal transmission. *Evolution*, **50**: 1729–1741.

Lynch, M. 2006. Streamlining and simplification of microbial genome architecture. *Annu. Rev. Microbiol.*, **60**: 327–349.

Lynch, M. and Conery, J.S. 2003. The origins of genome complexity. *Science*, **302**: 1401–1404.

Maynard Smith, J. and Szathmáry, E. 1997. *The Major Transitions in Evolution.* Oxford: Oxford University Press.

Michod, R.E. 1999. *Darwinian Dynamics: Evolutionary Transitions in Fitness and Individuality.* Princeton, NJ: Princeton University Press.

Moya, A., Pereto, J., Gil, R. and Latorre, A. 2008. Learning how to live together: genomic insights into prokaryote–animal symbioses. *Nature Rev. Genet.*, **9**: 218–229.

Otto, S.P. and Barton, N.H. 2001. Selection for recombination in small populations. *Evolution,* **55**: 1921–1931.

Otto, S.P. and Feldman, M.W. 1997. Deleterious mutations, variable epistatic interactions, and the evolution of recombination. *Theor. Popul. Biol.*, **51**: 134–147.

Otto, S.P. and Lenormand, T. 2002. Resolving the paradox of sex and recombination. *Nature Rev. Genet.*, **3**: 252–261.

Robinson, W.P., Asmussen, M.A. and Thomson, G. 1991. 3-locus systems impose additional constraints on pairwise disequilibria. *Genetics*, **129**: 925–930.

Rodriguez, R.J., White, J.F., Jr., Arnold, A.E. and Redman, R.S. 2009. Fungal endophytes: diversity and functional roles. *New Phytol.*, **182**: 314–330.

Sachs, J.L. and Simms, E.L. 2008. The origins of uncooperative rhizobia. *Oikos*, **117**: 961–966.

Sachs, J.L., Essenberg, C.J. and Turcotte, M.M. 2011. New paradigms for the evolution of beneficial infections. *Trends Ecol. Evol.*, **26**: 202–209.

Sanchez, M.S., Arnold, J. and Asmussen, M.A. 2000. Symbiont survival and host–symbiont disequilibria under differential vertical transmission. *Genetics*, **154**: 1347–1365.

Simser, J.A., Rahman, M.S., Dreher-Lesnick, S.M. and Azad, A.F. 2005. A novel and naturally occurring transposon, isrpe1 in the *Rickettsia peacockii* genome disrupting the ricka gene involved in actin-based motility. *Mol. Microbiol.*, **58**: 71–79.

Taylor, M.W., Radax, R., Steger, D. and Wagner, M. 2007. Sponge-associated microorganisms: evolution, ecology, and biotechnological potential. *Microbiol. Mol. Biol. Rev.*, **71**: 295–347.

Toft, C. and Andersson, S.G. 2010. Evolutionary microbial genomics: insights into bacterial host adaptation. *Nature Rev. Genet.*, **11**: 465–475.

Unckless, R.L. and Herren, J.K. 2009. Population genetics of sexually antagonistic mitochondrial mutants under inbreeding. *J. Theor. Biol.*, **260**: 132–136.

Usher, K.M. 2008. The ecology and phylogeny of cyanobacterial symbionts in sponges. *Mar. Ecol.*, **29**: 178–192.

Wade, M.J. 2007. The co-evolutionary genetics of ecological communities. *Nature Rev. Genet.*, **8**: 185–195.

Wade, M.J. and Brandvain, Y. 2009. Reversing mother's curse: selection on male mitochondrial fitness effects. *Evolution*, **63**: 1084–1089.

Wade, M.J. and Goodnight, C.J. 2006. Cyto-nuclear epistasis: two-locus random genetic drift in hermaphroditic and dioecious species. *Evolution*, **60**: 643–659.

Wu, D., Daugherty, S.C., Van Aken, S.E., Pai, G.H., Watkins, K.L., Khouri, H. *et al.* 2006. Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. *PLoS Biol.*, **4**: 1079–1092.

Wu, D.Y., Hugenholtz, P., Mavromatis, K., Pukall, R., Dalin, E., Ivanova, N.N. *et al.* 2009. A phylogeny-driven genomic encyclopaedia of bacteria and archaea. *Nature*, **462**: 1056–1060.

Yamamura, N. 1993. Vertical transmission and evolution of mutualism from parasitism. *Theor. Popul. Biol.*, **44**: 95–109.

APPENDIX A: DERIVATION OF RECURSIONS

This appendix derives recursion equations that describe the evolution of vertical transmission in an obligate symbiont under the assumptions described in the main text. Because hosts always contain a single symbiont, we can consider the host–symbiont combination a single haplotype and write the frequency of a haplotype as P_{ijk}, where i is the host interaction allele (A or a), j is the symbiont interaction allele (B or b), and k is the transmission modifier allele (M or m).

Selection

After selection, the frequency of haplotypes in the population is

$$P'_{ijk} = \frac{P_{ijk} w_{ij}}{\bar{w}}, \tag{A1}$$

where haplotype fitness (w_{ij}) is found in Table 1 (Hedrick, 2005, p. 120). We assume that the transmission modifier locus has no effect on fitness and exclude it from the fitness notation in Table 1. The population mean fitness is

$$\bar{w} = \sum_i \sum_j w_{ij} \left(\sum_k P_{ijk} \right). \tag{A2}$$

Transmission

For clarity, we separate the transmission step into host–symbiont associations that are generated via vertical transmission and those generated via horizontal transmission. These processes are happening simultaneously. To determine the frequency of host–symbiont haplotypes after transmission, we calculate the frequencies of haplotypes generated by each transmission mode separately. We then combine these pools of host–symbiont pairs. Vertical transmission maintains the association among the host and symbiont haplotypes because offspring produced with vertical transmission of the symbiont have the same haplotype as the parental type. On the other hand, horizontal transmission randomly associates symbiont and hosts in a *horizontal pool.* We incorporate a transmission trade-off where an opportunity to transmit a symbiont vertically results in the loss of an opportunity to transmit a symbiont into the horizontal pool. This trade-off is analogous to pollen discounting opportunities in plant reproduction (Holsinger *et al.*, 1984). Assuming this trade-off, the frequency of the symbiont haplotypes in the horizontal pool is

$$X_j = \frac{\sum_k \left[(1 - V_k) \sum_i P'_{ijk} \right]}{\bar{X}}, \tag{A3}$$

where we sum symbionts across host types, i, and V_k is the vertical transmission rate of symbionts associated with hosts of type k, and the total fraction horizontally infecting symbionts is given by

$$\bar{X} = \sum_k \left[(1 - V_k) \sum_i \sum_j P'_{ijk} \right]. \tag{A4}$$

We assume that the fraction of offspring generated under vertical transmission retain their relationships between host and symbiont. We use equations (A3) and (A4) to calculate the

random association of symbionts infecting hosts via horizontal transmission. Combining these two modes of transmission, the frequency of haplotypes after transmission is

$$P''_{ijk} = V_k P'_{ijk} + (1 - V_k) X_j \sum_j P'_{ijk}. \tag{A5}$$

APPENDIX B: DEFINITIONS OF ALLELE FREQUENCIES AND LINKAGE DISEQUILIBRIUM

Our full system of recursions (equations A1–A5) is specified in terms of haplotype frequencies because selection is acting directly on the interaction of host and symbiont. Our transmission modifier does not have any direct effect on fitness, but selection acts indirectly on this locus through associations with particular genetic backgrounds. We can recast the system of haplotype frequencies in terms of allele frequencies and population linkage disequilibrium. This new notation allows us to track allele frequencies at each locus. The frequency of the M transmission modifier allele is

$$P_M = \sum_i \sum_j P_{ijM}. \tag{B1}$$

The frequency of the A allele at the interaction locus of the host is

$$P_A = \sum_j \sum_k P_{Ajk}, \tag{B2}$$

and in the symbiont the B allele frequency is given by

$$P_B = \sum_i \sum_k P_{iBk}. \tag{B3}$$

We define two-locus disequilibrium for all two-locus combinations (D within species, ID between species) in the system by

$$ID_{AB} = \left(\sum_k P_{ABk}\right)\left(\sum_k P_{abk}\right) - \left(\sum_k P_{Abk}\right)\left(\sum_k P_{aBk}\right), \tag{B4}$$

$$D_{AM} = \left(\sum_j P_{AjM}\right)\left(\sum_j P_{ajm}\right) - \left(\sum_j P_{Ajm}\right)\left(\sum_j P_{ajM}\right), \tag{B5}$$

$$ID_{BM} = \left(\sum_i P_{iBM}\right)\left(\sum_i P_{ibm}\right) - \left(\sum_i P_{iBm}\right)\left(\sum_i P_{ibM}\right), \tag{B6}$$

where D_{xy} is the statistical linkage between the x and y locus in the same species and ID_{xy} is used when the loci occur in different species. Using the definitions of the two-way linkage disequilibrium (equations B4–B6) and the definitions of the allele frequencies (equations B1–B3), following Robinson *et al.* (1991), we define a three-way interspecies disequilibrium by

$$ID_{ABM} = P_{ABM} - P_A ID_{BM} - P_B D_{AM} - P_M ID_{AB} - P_A P_B P_M. \tag{B7}$$