Equilibrium Frequency of Endosymbionts in Multiple Infections Based on the Balance between Vertical Transmission and Cytoplasmic Incompatibility

Yuuki Kawasaki¹, Hiroshi Ito², Hisashi Kajimura¹

¹ Graduate School of Bioagricultural Sciences, Nagoya University, Nagoya, Japan, ² Faculty of Design, Kyushu University, Fukuoka, Japan

Abstract

Cytoplasmic incompatibility (CI)-inducing endosymbiotic bacteria, such as Wolbachia and Cardinium, have been well studied through field data and validations on the basis of numerical simulations. However, the analytically derived equilibrium frequency of multiple infections has not yet been determined, although the equilibrium for cases of single infection has been reported. In this study, we considered the difference equation for endosymbionts using three parameters: the probability of the failure of vertical transmission (µ), CI strength (z), and the level of host inbreeding (p). To analyze this model, we particularly focused on \( Q_N \), i.e., the frequency of host individuals completely infected with all \( N \)-bacterial strains in the population. \( Q_N \) at the equilibrium state, was analytically calculated in the cases where \( N = 1, 2 \) and \( N \) is an arbitrary value. We found that \( Q_N \) can be described using two parameters: \( N \) and \( z \), which is identical to \( u/z \). \( Q_N \) has a larger value in a system with a smaller \( z \). In addition, \( z \) determines the maximum number of strains that infect a single host. Our results revealed the following: i) three parameters can be reduced to a single parameter, i.e., \( z \) and ii) the threshold of the maximum number of infections is defined by \( z \), which prevents additional invasions by endosymbionts.

Introduction

Endosymbiotic bacteria such as Wolbachia and Cardinium are well known to be reproductive manipulators infecting insect cells [1,2]. Cytoplasmic incompatibility (CI) is considered to be the most common and efficient form of manipulation that can spread an infection throughout the host population. CI causes males infected with a bacterial strain reproductively incompatible with uninfected females [3]. Because of CI, infected females exhibit a relatively higher fitness than uninfected females; therefore, the number of infected individuals gradually increases and they become dominant in the host population. For example, approximately only 10% of Drosophila simulans individuals were infected with Wolbachia in the middle of the 1980s in California; however, the infection rate increased to 95% in 1995 [4,5]. In a rearing system of Encarsia pergandiella infected with Cardinium, the infection gradually spread to almost fixation within the population, irrespective of the initial rate of infection [6].

The dynamics of CI-inducing bacteria can be primarily determined by three parameters: vertical transmission efficiency, fecundity of infected females, and CI levels [7], although the dynamics may be affected by other parameters [8,9]. Vertical transmission, from a mother to her offspring, is the main route of transmission of the bacteria to other host individuals. The transmission rate is defined as the proportion of offspring infected from an infected mother, which is approximately 1 (≥0.95 in most cases) [10]. Successful transmission has crucial effects on the spread and maintenance of the bacteria [11]. CI-inducing bacteria sometimes affect the fitness of the infected host directly through fecundity. The direction of their effect through fecundity depends on the host-bacteria combination, and the effect may be negative [12] or positive [13], or often neutral [14]. The CI level is defined as the proportion of the number of offspring died because of CI relative to the number of total offspring reproduced. In addition, the effect of CI is dependent on the host-bacteria combination, which highly varies from zero to one, i.e., from nearly neutral to complete mortality [15–17].

Models have been developed to study the dynamics of CI-inducing bacterial symbionts, particularly in Wolbachia, using the three conventional parameters [4,10]. The long-term behavior depends on the initial frequency of bacterial infection [19]. If the initial frequency of bacterial infection is below the threshold, the frequency heads towards extinction; however, if the frequency stochastically exceeds the threshold (e.g., because of random genetic drift), then it is expected to spread to another equilibrium state where both infected and uninfected individuals exist. A higher vertical transmission rate and/or stronger CI will lead to higher infection frequency equilibrium. In most of the models on the basis of difference equations, the equilibria have rarely been analytically derived [20–22]. A few analytical results have been
reported for single [7] and double infections [23]. Farkas & Hinow [23] analytically examined the case of hosts infected with only a single bacterial strain when two strains were present in the field. However, the symbiotic bacterial dynamics of an arbitrary number of strains remains poorly understood [7,10], which prompts the questions: “how do the parameters determine the dynamics of infections and how many strains can infect a host population?” Experimental studies have reported multiple infections of Wolbachia in parasitic ants (nine strains) [24], byturid beetles (five strains) [25], and ambrosia beetles (five) [26]. However, whether a host population can be infected with a larger number of species is not known. Thus, the generalization of bacterial numbers is a big issue in endosymbiotic studies.

In addition, the modeling of endosymbiotic dynamics in the context of mating system has not been extensively studied. CI is a phenomenon that occurs via mating, thus it is considered that the mating system should affect the dynamics of the endosymbionts. Mating systems can be categorized roughly as random mating and inbreeding. Random mating or panmictic mating is mating between any individuals in the population. In contrast, inbreeding or sib mating, is mating between a brother and a sister. Therefore, CI should be less effective under inbreeding because the mating partner should have the same infection status. Strong inbreeding has been observed in some insects, such as fig wasps and ambrosia beetles, which are known to have Wolbachia infections [26,27]. Some theoretical studies have addressed the dynamics under inbreeding [28–30]. However, most of these models were examined numerically rather than analytically. In addition, the scope of these models was limited in the case of single or double infections.

This study aimed to develop an analytically solvable model of the infection dynamics of CI-inducing bacteria in sib/panmictic mating systems. We derived the analytical equilibrium of single and double infections, and generalized arbitrary numbers of strain infections using three parameters (vertical transmission rate, CI levels and the level of inbreeding). Our results show that bacterial infection thresholds can be simply expressed as a function of the ratio between the failure of vertical transmission and the CI level. In addition, we derived the maximum number of bacterial strains that are capable of infecting a host population.

Methods

Model

We consider a host population infected with N endosymbiotic strains. The integer $i (=0, 1, \ldots, 2^N - 1)$ represents the state of each host individual, infected or uninfected with bacteria, where $i$ is written in base 2 with N bits and each bit represents the state of infection with bacteria, i.e., 0 and 1 indicate absence and presence, respectively. For example, when $i=5$ (=101 in base 2), the host is infected with the first and third bacterial strains but uninfected with the second strain. For convenience, we introduce a binary function $f(i_1, i_2)$, which indicates whether the set of infected bacterial strains $i_1$ is included in that of $i_2$. If the set of $i_1$ is a subset of that of $i_2$, $f(i_1,i_2)=1$, otherwise $f(i_1,i_2)=0$. For example, $f(3,7)=1$ because the set of strains $i_2$ (=111 in base 2) contains any strains of $i_1$ (=011 in base 2). In addition, we define $n(i)$ as the number of strains carried by an individual of type $i$.

We develop an difference equation to express the dynamics of the frequencies of hosts infected with multiple bacterial strains. Let $q_i$ be the frequency of individuals infected with the bacterial set $i$. For simplicity, we assume that all of the bacteria have identical parameters in terms of the vertical transmission rate and CI strength among strains.

Next, we consider the formulation of CI among strains. CI results in a decreased hatching rate by a factor $1-z$; therefore, the fitness of individual mating incompatibility is reduced to $1-z$. The fitness of mating between a male with bacterial strains of $i$ and a female with bacterial strains of $j$ is generally described as $f(i,j)+(1-z)(1-f(i,j))$.

A bacterial strain can be transferred by vertical transmission from a mother to offspring at egg stage with a probability of $1-\mu$, whereas it fails to be transferred with a probability of $\mu$. In the field, measurements of $\mu$ are typically low ($<0.05$). An individual host infected with the bacterial set $j$ will produce eggs with $i$ with a probability of $(1-\mu)^{m(i)/n(i)}$ when $f(i,j)=1$. Note that our model assumes a diplodiplod sex determination system in the host insects, but the effects of CI can be quite different between diplodiploidy and haplodiploidy [31].

The occurrence frequency of CI depends on the mating pattern. Under random mating, CI occurs depending on the distribution of $q_i$. The average fitness of a female with the bacterial set $j$ is

$$q_j = \sum_{k=0}^{2^N-1} q_k \{ f(k,j)+(1-z)(1-f(k,j)) \}.$$  

In contrast, under inbreeding, CI depends on $\mu$. Mating partner tends to have the same infection status as themselves because the pair are possibly produced by the same mother. CI is less frequent under inbreeding than random mating because it occurs only under following conditions: (i) a mother lacks the bacterial strain(s) because of vertical transmission failure and (ii) a father carries the strain(s) because of their successful transmission from a common grandmother. It is difficult to formulate the exact effect of CI under inbreeding because the frequency in a previous generation is needed. For simplicity, we hereafter assume that $\mu$ is sufficiently small to regard the average fitness of a female with the bacterial set $j$ as 1. Let $q'$ be the frequency of infected individuals in the next generation and $p$ the probability of random mating. The difference equation of $q_i$ is described as follows:

$$wq'_i = \sum_{j=0}^{2^N-1} q_j f(i,j)(1-\mu)^{m(j)/n(j)} \left[ (1-p)+p \sum_{k=0}^{2^N-1} q_k \{ f(k,j)+(1-z)(1-f(k,j)) \} \right].$$  

where $w$ is a normalization factor that maintains the sum of the updated host frequencies as unity. Eq.(1) can be rewritten as follows:

$$wq'_i = \sum_{j=0}^{2^N-1} q_j f(i,j)(1-\mu)^{m(j)/n(j)} \left[ 1-pz(-1+\sum_{k=0}^{2^N-1} q_k f(k,j)) \right].$$  

Eq.(2) is symmetrical for $p$ and $z$, which means that the effect of $p$ on the infection dynamics is equivalent to that of $z$. In addition, Eq.(2) reveals that the dynamics can be described using only two parameters, $\mu$ and $pz$. If two systems have the same value of $pz$, then the dynamics of these systems should be identical. If we use $z=pz$ as the effective CI strength, then $z$ should correspond to $z$ in the systems with the equivalent dynamics under complete random mating ($p=1$). As a result, Eq.(1) can be rearranged to a more
simple form using \( z \) as follows:

\[
\begin{align*}
   wq_i' &= \sum_{j=0}^{2N-1} q_i j (1-\mu)^{p(i)} \mu^{p(j) - p(i)} \\
   &\quad \left[ 1 - z \sum_{k=0}^{2N-1} q_k [1-f(k,j)] \right].
\end{align*}
\]

Eq.(3) is identical to the equation developed by Frank [18], but \( z \) is substituted by \( \bar{z} \). Frank suggested that if all the bacterial strains share the common parameters \( \mu \) and \( \bar{z} \), the polymorphic equilibrium solutions should be symmetric for the frequencies, i.e., the frequency of the hosts with the same number of strains should be identical. For example, if \( q_{110} > 0 \), then \( q_{110} = q_{101} = q_{011} \) should be satisfied. Frank performed a reduction of Eq.(3) as follows:

\[
\begin{align*}
   wQ_i &= \sum_{j=0}^{N} Q_i \left( \begin{array}{c} N-i \\ j-i \end{array} \right) (1-\mu)^j \mu^{-i} \\
   &\quad \left[ 1 - \bar{z} \sum_{k=0}^{N} Q_k \left( \begin{array}{c} N \\ k \end{array} \right) j - \left( \begin{array}{c} j \\ k \end{array} \right) \right],
\end{align*}
\]

where \( Q_i \) is the frequency of a type that carries \( i \) different strains, in which there are \( \left( \begin{array}{c} N \\ i \end{array} \right) \) different combinations. We analyze Eq.(4) hereafter.

Frank also proved that a polymorphic equilibrium, where both infected and uninfected individuals exist, can be stably maintained only if there are host individuals infected with all of the bacterial strains in the population [14]. In other words, if the largest \( m(i) \) in the population is less than \( N \), polymorphism should be lost and some bacterial strains will eventually fail to achieve infection until stable maintenance is possible (Figure 1). Frank stated that the presence or absence of individuals infected with all bacterial strains in a population determines the fate of the population. Therefore, we focus on the dynamics of the frequency of hosts infected with all of the bacterial strains in an \( N \)-strain population, \( Q_N \).

Next, we analytically derive the equilibrium for the system [Eq.(4)]. We will first focus on simple cases with one or two bacterial strains in the population before generalizing the results to an arbitrary number of strains.

Results

Single infection

When \( N = 1 \), the system [Eq.(4)] can be represented as follows:

\[
\begin{align*}
   wQ_1 &= Q_1 (1-\mu),
\end{align*}
\]

where \( w = Q_1 + Q_0 (1-\bar{z}Q_1) \) and the total frequency \( Q_0 + Q_1 \) equals unity.

Because \( Q_1 - Q_1 = \frac{\bar{z}Q_1}{w} (-Q_1^2 + Q_1 + \frac{\mu}{2}) \), there are a maximum of three fixed points in the system: \( \left( Q_0^*, Q_1^* \right) = (1,0), \left( 1 + \sqrt{1-4\bar{z}}, 1 - \sqrt{1-4\bar{z}} \right) \) and \( \left( 1 - \sqrt{1-4\bar{z}}, 1 + \sqrt{1-4\bar{z}} \right) \). where \( \bar{z} \equiv \mu / \bar{z} \). The phase portrait of the system is shown in Figure 2a. The second and third fixed points can exist only when \( 0 < \bar{z} < \bar{z}^*(1) = \frac{1}{2} \) because a saddle-node bifurcation occurs at \( \bar{z} = \bar{z}^*(1) \). Thus, we can obtain the nontrivial fixed points except (1,0) only for \( 0 < \bar{z} < \frac{1}{2} \). (1,0) and \( \left( 1 - \sqrt{1-4\bar{z}}, 1 + \sqrt{1-4\bar{z}} \right) \) are stable and \( \left( 1 + \sqrt{1-4\bar{z}}, 1 - \sqrt{1-4\bar{z}} \right) \) is unstable (Figure 2a). The unstable fixed point shows the basin boundary. If the initial value of \( Q_1 \) is below the unstable fixed point, then \( Q_1 \) decreases asymptotically to 0, i.e., the population loses the bacterial infection. In contrast, if \( Q_1 \) exceeds the boundary due to random drift or other accidental events, \( Q_1 \) can converge towards \( \frac{1}{2} + \frac{\sqrt{1-4\bar{z}}}{2} \). Note that because \( \frac{1}{2} + \frac{\sqrt{1-4\bar{z}}}{2} > \frac{1}{2} \) whenever the bacterial infection reaches equilibrium, the frequency of infected individuals is certainly more than half.

Double infection

In the case of \( N = 2 \), the system [Eq.(4)] can be represented as follows:

\[
\begin{align*}
   wQ_2 &= Q_2 (1-\mu)^2 \\
   wQ_1 &= Q_2 (1-\mu) + Q_1 (1-\mu) [1-\bar{z}(Q_1 + Q_2)]
\end{align*}
\]

where \( w = Q_2 + 2Q_2 [1-\bar{z}(Q_1 + Q_2)] + Q_1 [1-\bar{z}(Q_2 + 2Q_1)] \) and the total frequency \( Q_0 + 2Q_1 + 2Q_2 \) equals unity. This system has five fixed points as follows: \( \left( Q_0^*, Q_1^*, Q_2^* \right) = (1,0,0), \left( 1 - \mu - \sqrt{1-8\bar{z}+2\mu^2}, 1+\mu + \sqrt{1-8\bar{z}+2\mu^2} \right), \left( 1 + \mu - \sqrt{1-8\bar{z}+2\mu^2}, 1-\mu + \sqrt{1-8\bar{z}+2\mu^2} \right) \).
was a stable fixed point and the others were unstable. This stable fixed point exists if $0 < z < \chi^2_z$, where $\chi^2_z = \frac{(4+\mu)+\sqrt{\mu^2+8\mu+32}}{8}$. When $z = \chi^2_z$, $(1-2z+\sqrt{1-(4+\mu)z-4z^2}) \frac{1}{2}$ and $1-2\chi - \sqrt{1-(4+\mu)\chi-4\chi^2} \frac{1}{2}$ coalesce due to saddle-node bifurcation as well as when $N=1$. Because $\chi^2_z < \chi^1_z$ for any value of $\mu>0$, the range in which $\chi^1_z$ and $\chi^2_z$ are located splits into two when $\mu$ is increased. When $\mu\approx 1.1$ (Figure 3), $\chi^1_z$ and $\chi^2_z$ show a similar range of stability. When $\mu<0.05$ and $z<0.38$, $0.24$, and $0.20$, respectively. The horizontal and vertical axes represent $Q_2$ and $Q_1$, respectively. Each arrow represents the difference in $Q_2$ and $Q_1$ between a generation, $\Delta Q = (Q_2 - Q_1, Q_1 - Q_0)$. The color of an arrow indicates the magnitude of the vector. Solid and open circles indicate stable and unstable equilibrium points. The shaded areas depict the basin of attraction for the extinction of endosymbionts. (E–G) Vector fields where $N=5$ with $\mu=0.05$ and $z=0.75, 0.50$ and $0.25$, respectively. Each arrow represents the difference of $Q_2$ and $Q_1$ between a generation, $\Delta Q = (Q_2 - Q_1, Q_4 - Q_3)$. doi:10.1371/journal.pone.0094900.g002

Because the absolute values of $Q_2$ and $Q_1$, and concluded that only $\left(1-2z+\sqrt{1-(4+\mu)z-4z^2}\right) \frac{1}{2}$ was a stable fixed point and the others were unstable. This stable fixed point exists if $0 < z < \chi^2_z$, where $\chi^2_z = \frac{(4+\mu)+\sqrt{\mu^2+8\mu+32}}{8}$. When $z = \chi^2_z$, $(1-2z+\sqrt{1-(4+\mu)z-4z^2}) \frac{1}{2}$ and $1-2\chi - \sqrt{1-(4+\mu)\chi-4\chi^2} \frac{1}{2}$ coalesce due to saddle-node bifurcation as well as when $N=1$. Because $\chi^2_z < \chi^1_z$ for any value of $\mu>0$, the range in which $\chi^1_z$ and $\chi^2_z$ are located splits into two when $\mu$ is increased. When $\mu\approx 1.1$ (Figure 3), $\chi^1_z$ and $\chi^2_z$ show a similar range of stability. When $\mu<0.05$ and $z<0.38$, $0.24$, and $0.20$, respectively. The horizontal and vertical axes represent $Q_2$ and $Q_1$, respectively. Each arrow represents the difference in $Q_2$ and $Q_1$ between a generation, $\Delta Q = (Q_2 - Q_1, Q_1 - Q_0)$. The color of an arrow indicates the magnitude of the vector. Solid and open circles indicate stable and unstable equilibrium points. The shaded areas depict the basin of attraction for the extinction of endosymbionts. (E–G) Vector fields where $N=5$ with $\mu=0.05$ and $z=0.75, 0.50$ and $0.25$, respectively. Each arrow represents the difference of $Q_2$ and $Q_1$ between a generation, $\Delta Q = (Q_2 - Q_1, Q_4 - Q_3)$. doi:10.1371/journal.pone.0094900.g002

Figure 3. Comparison of $Q_N$ in the numerical (colored points) and analytical (black curves) solutions where $N=1$, 2, 5 and 10. The parameters used in the simulations were selected from the ranges of $0 < \mu < 0.05$ and $0 < z < 1.0$. The convergence condition was $\sum_{n=0}^{N}(Q_n - Q_{n-1})^2 < 10^{-10}$. The analytically derived equilibria are shown by a solid line for $N=1$ ($Q^{*}_1 = 1 + \sqrt{1-4\chi^1_z}$), a dash-dotted line for $N=2$ [Eq.(7)], and dotted lines for $N=5$ and $10$ [Eq.(10)]. doi:10.1371/journal.pone.0094900.g003
completely infected host can exist at $N = 2$ is narrower than that when $N = 1$.

As stated in the previous section, we assume that $\mu$ is a relatively small parameter. By substituting zero for $\mu$ as a zero-order approximation, $Q_2^1$ at the stable fixed point and the critical value of $z^{(2)}$ are given as follows:

$$Q_2^1 = \frac{1 - 2z + \sqrt{1 - 4z - 4z^2}}{2} + O(\mu),$$  
(7)

$$z^{(2)} = -\frac{1 + \sqrt{2}}{2} + O(\mu).$$  
(8)

The numerical simulation confirmed that this approximation is adequate and that $\mu$ had less effect on $Q_2^1$ and $z^{(2)}$ than $z$ (Figure 3).

**N-strain infection**

Finally, we calculated the fixed points of Eq.(4) for arbitrary $N$. We assumed that $Q_1 \approx 0$, for $0 \leq n \leq N - 2$ in an equilibrium state. Numerical simulations supported this assumption (Figure S1). For example, when $N = 10$, $\bar{z} = 0.4$ and $\mu = 0.025$, 95.5% of the population possessed $N$ or $N - 1$ bacterial strains. The dominance of highly infected hosts was enhanced by the strain number $N$. Thus, this assumption is more plausible when the system has a greater $N$ and a lower $z$. The reduced system is described as follows:

$$wQ_N = Q_N(1 - \mu)^N$$  
(9)

where $Q_N + NQ_{N - 1} = 1$, $w \approx Q_N + NQ_{N - 1}[1 - \bar{z}(Q_N + (N - 1)Q_{N - 1})]$. Because this is a one-dimensional system, the fixed points can be found easily, $Q_N^* = -\frac{2 - N \pm \sqrt{N^2 - 4Nz + 4(1 - \mu)^N}}{2\bar{z}}$. If $\mu$ was regarded as a relatively small parameter, $(1 - \mu)^N \approx 1 - N\mu$. Then,

$$Q_N^* = 0, \frac{2 - N \pm N \sqrt{1 - 4z}}{2}$$  
(10)

which shows that the frequency of completely infected hosts depends only on $z$ when $\mu$ is sufficiently small. The difference in $Q_N$ between one generation, $\Delta Q_N$, is represented as follows:

$$\Delta Q_N \equiv Q_N^* - Q_N \approx -\frac{\bar{z}}{wN}(Q_N^* - Q^0)(Q_N^* - Q^+)(Q_N^* - Q^-).$$  
(11)

where $Q^0 = 0$, $Q^+ = \frac{2 - N \pm N \sqrt{1 - 4z}}{2}$. If $N \geq 2$ and $z < \frac{N - 1}{N - 2}$, the order of the three fixed points is $Q^- < Q^0 < Q^+$ and $Q^+$ is a stable equilibrium.

Figure 3 compares $Q^+$ at various $z$ with the results calculated by numerical simulations for $N = 5$ and 10. The approximated solution of $Q^+$ was generally consistent with the numerical results. In particular, the agreement was better when $z$ was sufficiently small.

Equation (11) indicates that an exchange of stability, so-called transcritical bifurcation, should occur between $Q^+$ and $Q^0$ at $z = \frac{N - 1}{N^2}$. When $z > \frac{N - 1}{N^2}$, the order of fixed points is $Q^- < Q^+ < Q^0$ and the stable fixed point is $Q^0$. However, we did not observe a transcritical bifurcation for any parameter set of $N$, $\bar{z}$ and $\mu$. Instead, a discrete jump into 0 was observed before $Q_N^*$ reached to 0, which implied that a saddle-node bifurcation also occurred at $N \geq 3$ (Figure 3). The reduced system did not replicate the disappearance of the stable fixed point; however, the analytical solution of $Q_N^*$ agreed with the numerical results until the discrete jump occurred. In addition, The numerical simulation confirmed that when the bacterial strain number $N$ was sufficiently high, a discrete jump occurred around $Q_N = 0$. Therefore, with a large $N$ limit, the value of $z$ at which the transcritical bifurcation occurs is equal to $z$ at the saddle-node bifurcation. Thus, we obtained the critical $z_N^{(N)}$ for a sufficiently large $N$ as follows:

$$z_N^{(N)} = \frac{N - 1}{N^2}. \quad (12)$$

We compared the analytically derived $z_N^{(N)}$ with numerical simulations. Figure 4 presents the numerically calculated $Q_N^*$ at various $\mu$ and $z$ when $N = 1, 2$ and 5. The lines in each figure indicate the analytically derived critical value of $z_N^{(N)} = \frac{1}{4}$.
n bacterial strains that can infect a population, approached the threshold derived by numerical simulation. The results, we plotted the critical value of bacteria can infect a population. On the basis of the analytical area derived by simulation indicates the values of positive value, irrespective of the value of \( m \) parameters where \( \{ \{ 1 \} \} \) consistence is reduced when \( N \) ranges from 0 to 1. In contrast, when \( \mu \geq x_1 \), \( Q_N \) cannot have a positive value, irrespective of the value of \( \bar{z} \). Figure 5 confirms the consistency between the numerical results and Eq(12). The shaded area derived by simulation indicates the values of \( x \) where \( N \) bacteria can infect a population. On the basis of the analytical results, we plotted the critical value of \( x \) where \( N = 1 \) and \( \bar{z} \),

\[
\frac{\zeta}{x} = \frac{1 + \sqrt{2} - 4}{2x} \text{ and } \frac{\zeta}{x} = \frac{4}{2x}.
\]

The lines successfully delineate the parameters where \( N \) bacterial infections can be maintained. The consistency is reduced when \( \mu \) is greater because \( \zeta(2) \) and \( \zeta(3) \) include \( O(\mu) \) errors. In particular, the ranges of \( \bar{z} \) and \( \mu \) that allow infections with a complete set of bacteria are distinct. The value of \( \bar{z} \) ranges from 0 to 1. In contrast, when \( \mu \geq x_1 \), \( Q_N \) cannot have a positive value, irrespective of the value of \( \bar{z} \). Figure 5 confirms the consistency between the numerical results and Eq(12). The shaded area derived by simulation indicates the values of \( x \) where \( N \) bacteria can infect a population. On the basis of the analytical results, we plotted the critical value of \( x \) where \( N = 1 \) and \( \bar{z} \),

\[
\frac{\zeta}{x} = \frac{1 + \sqrt{2} - 4}{2x} \text{ and } \frac{\zeta}{x} = \frac{4}{2x}.
\]

For \( N \geq 3 \), we plotted the critical \( x \) derived from Eq(12). For larger values of \( N \), the analytical results approached the threshold derived by numerical simulation.

The inverse function of Eq(12) defines the maximum number of bacterial strains that can infect a population, \( N_c \), which is given as follows:

\[
N_c = \frac{1 + \sqrt{1 - 4x}}{2x} \approx \frac{1}{x} - 1.
\]

\( N_c \) is approximately proportional to \( x^{-1} \). Therefore, it has a larger value with a lower value of \( x \).

**Discussion**

In this study, we demonstrated the analytic equilibrium solutions of the frequencies of individuals infected with all the bacterial strains in a population with single, double and arbitrary \( N \) infections. Our results provide a qualitative insight into the symbiont bacterial dynamics, in contrast to recently developed models that quantitatively simulate specific experimental results [4,32].

Our model used a parameter \( p \), the level of inbreeding, from the perspective of host behavior. We revealed that \( p \) has a completely negative effect on CI. Endosymbionts manipulates host reproduction to maintain their own infections in host populations. In contrast, host insects can reduce the prevalence of endosymbiont infections by inbreeding. This result is consistent with previous numerical simulations of inbreeding [28]. Engelstatter et al. claimed that the infection rate decreases with increasingly inbred hosts because uninfected females increasingly mate with uninfected males, which leads to fewer incompatible matings. Our analytic calculation determined a linear relationship \( \bar{z} = \mu \), which clearly indicates the negative effect of inbreeding on the CI strength. However, this simple relationship does not appear to hold in some cases of stochastic models. Genetic drift in stochastic island model might have effects on inbreeding [29]. Our result is also consistent with the previous analytical result reported by Dannowski et al. [30]. This model focused on double infections by male-killing bacterial strains and analytically showed that higher level of host inbreeding leads to lower frequency of bacterial infections. Although their result was limited in the case of double infections, our result is applicable to arbitrary number of CI-inducing bacterial strains.

The analytically derived equilibrium depends mainly on one parameter, \( x \). \( x \) is a novel index that we defined in this study, which is the ratio of the probability of vertical transmission failure (\( \mu \) relative to the effective CI level (\( \bar{z} \)). Previous studies have shown that these parameters are involved with the infection dynamics. Furthermore we demonstrated the simple relationship between these parameters, i.e., the bacterial infection dynamics are determined simply by a balance of both. The equilibrium frequencies of the fully infected host are at a higher rate when \( x \) is lower (Figure 3). In this case, CI-inducing bacteria can select two alternative strategies to maintain and spread themselves: increasing the vertical transmission efficiency [16] and/or increasing the CI strength. There have been no reports of the differences among these strategies. Figure 4 presents the asymmetric relationship between \( \mu \) and \( \bar{z} \). When \( \mu \) was higher than approximately 0.2, bacterial infections could not be sustained regardless of the value of \( \bar{z} \). Thus, the vertical transmission rate has a threshold value for infection. In contrast, bacterial infections occur with any value of \( \bar{z} \) provided an appropriate value of \( \mu \) is selected on the basis on \( \bar{z} \). Thus, the vertical transmission rate determines the infection state more strongly than CI.

Eqs(12) and (13) represent the relationship between the maximum number of bacterial strains and the bacterial parameters. Considering a host population containing \( N_0 \) bacterial strains, if \( N_0 \) increases to \( N_1 \) through the horizontal transfer of a bacterial strain, the critical value of \( x \) decreases from \( x^{(N_0)} \) to \( x^{(N_1)} \). Provided \( z < x^{(N_1)} \) is satisfied, the \( N_1 \)-infected population will be maintained. Otherwise, the infection cannot be maintained. For example, a population where \( z = 0.2 \) can maintain three strains at most because \( x^{(N)} \approx 0.18 < z < x^{(N)} \approx 0.22 \). If the population actually contains the maximum number of strains, the host population is “saturated” with bacteria, i.e. other strains will fail to invade the host population. The parameters are expected to have the same value in our model. Therefore, the frequency of the rarest strain necessarily declines to zero. A higher \( x \) value than that of other strains may be needed for a new strain to invade the saturated hosts.

Our theoretical results suggest that the equilibrium frequency of completely infected hosts decreases as \( N \) increases (Figure 3). In addition, the range of \( x \) where bacteria can be maintained becomes narrower as \( N \) increases (Figure 5). The severe conditions for multiple infections are caused by the increased possibility of vertical transmission failure the offspring by any strains. \( X \)-strains that infect a individual can successfully transmit all the strains to the offspring with a lower probability than \( X^{-1} \) strains. Consequently, these results support a noble idea of Wolbachia-induced speciation. Suppose that \( z \) is enough large to maintain single infection but not double infection (0.207–0.25), and...
endosymbiotic strains, A and B strains, infect some insects in two different geographic regions respectively. The infections in both regions will spread until a hybrid zone is produced. Then, horizontal transfer makes double-infected individuals of A and B strains. However, the double infection must be lost because $\alpha$ is too low to maintain the both infections. Therefore, the hybrid zone will be unstable and post-zygotic speciation will be occurred between A-infected and B-infected individuals.

To confirm that our results were consistent with experimental data, we compared the predicted equilibria with the values of CI-inducing Wolbachia and Cardinium obtained using natural populations or an artificial line (Figure 6, Table S1). No data were available for triple or higher multiple infection. Therefore, only five single infections and one double infection were used for reference. No data were available on the inbreeding frequency of the insects studied; hence, $p$ values of all were assumed to be zero, i.e. completely random mating. The experimental values were close to the predicted lines. Thus, our results can be used to estimate the frequency of infected hosts at equilibrium using the parameters ($\mu$ and $z$) as well as for estimating the parameters that are often difficult to be measured.

In this study, we aimed to provide a qualitative outline of the dynamics of multiple infections by introducing an analytically solvable model. By ignoring quantitative accuracy, our model sheds light on the mathematical structure related to the multiple infection dynamics. However, we introduced several assumptions to develop a solvable model. Finally, we roughly approximated the frequencies of hosts infected with $N$ to $n=2$ strains as being equals to 0 to calculate the equilibria for $N$ arbitrary strains. A more rigorous approximation method would reduce the error between the numerical and analytical results. The elimination of these assumptions needs to be addressed in future studies.

**Supporting Information**

**Figure S1** Dominancy of highly-infected individuals. (A) An example of distribution of frequencies $Q_i$ in equilibrium state. The frequencies were obtained by simulation with $\alpha = 0.001$ ($\mu = 0.005, z = 0.5$) and $N = 10$. (B) The ratio of $Q_N + Q_{N-1}$ in equilibrium state where $\alpha = 0.01$ (closed circles) and $\alpha = 0.05$ (open circles).

**Table S1** The values of parameters and the frequency of completely infected hosts in the field. These values were reported in the references listed above.

**Acknowledgments**

We thank A. Telschow for helpful comments on an earlier version of this manuscript and all members of the Forest Protection Laboratory at Nagoya University for advice related to this study.

**Author Contributions**

Analyzed the data: YK HI. Wrote the paper: YK HI HK. Conceived and designed the study: YK HI. Performed the analytic calculations and the numerical simulations: YK HI HK. Designed the study: YK HI. Performed the analytic calculations and the numerical simulations: YK HI HK. We thank A. Telschow for helpful comments on an earlier version of this manuscript and all members of the Forest Protection Laboratory at Nagoya University for advice related to this study.

**Author Contributions**

Analyzed the data: YK HI. Wrote the paper: YK HI HK. Conceived and designed the study: YK HI. Performed the analytic calculations and the numerical simulations: YK HI.

1. Werren J (1997) Biology of Wolbachia. Annu Rev Entomol 42: 587–609.
2. Zchori-Fein E, Perlman S (2004) Distribution of the bacterial symbiont Cardinium in arthropods. Mol Ecol 13: 2009–2016.
3. Raigon J, Scott T (2003) Wolbachia and cytoplasmic incompatibility in the California Culex pipiens mosquito species complex: parameter estimates and infection dynamics in natural populations. Genetics 165: 2029–2038.
4. Turelli M, Hoffmann A (1991) Rapid spread of an inherited incompatibility factor in California Drosophila. Nature 353: 440–442.

**References**

1. Werren J (1997) Biology of Wolbachia. Annu Rev Entomol 42: 587–609.
2. Zchori-Fein E, Perlman S (2004) Distribution of the bacterial symbiont Cardinium in arthropods. Mol Ecol 13: 2009–2016.
3. Raigon J, Scott T (2003) Wolbachia and cytoplasmic incompatibility in the California Culex pipiens mosquito species complex: parameter estimates and infection dynamics in natural populations. Genetics 165: 2029–2038.
4. Turelli M, Hoffmann A (1991) Rapid spread of an inherited incompatibility factor in California Drosophila. Nature 353: 440–442.
16. Charlat S, Ballard J, Mercot H (2004) What maintains noncytoplasmic
infection in Drosophila simulans? Heredity 93: 135–142.

17. Watanabe M, Miura K, Hunter M, Wajnberg E (2010) Superinfection of
Wolbachia in byturid beetles. J Evol Biol 23: 1807–1815.

18. Frank S (1998) Dynamics of cytoplasmic incompatibility with multiple Wolbachia infections. J Theor Biol 192: 213–218.

19. Jansen V, Turelli M, Godfray H (2008) Stochastic spread of Wolbachia and the evolution of reproductive isolation between Drosophila relict and Drosophila suzukii. Evolution 63: 1157–1164.

20. Mouton L, Dedene F, Heurteux H, Bouletracks M, Profizi N, et al. (2004) Virulence, multiple infections and regulation of symbiotic population in the Wolbachia-Asobara tabida symbiosis. Genetics 168: 181–189.