Supplemental materials to the dissertation titled “Evolution of virulence in facultative pathogens”

by

Aakash Pandey

B.Tech in Biotechnology, Kathmandu University, 2016

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Division of Biology
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2022

Supplementary Materials for "Virulence evolution of pathogens that can grow in reservoir environments"

Aakash Pandey[1, †], Nicole Mideo[2], and Thomas G. Platt[1]

[1] Division of Biology, Kansas State University, Manhattan, KS, USA

[2] Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, ON, Canada

[†] Corresponding Author: apandey@ksu.edu

The content of this material is already published along with Pandey, A., N. Mideo, and T. G. Platt. 2022. Virulence evolution of pathogens that can grow in reservoir environments. *The American Naturalist* 199:141–158.

A1. Threshold parameter calculation

The Jacobian of the infected host – free-living pathogen subsystem for system *1* of the main text at the disease-free equilibrium is given as,

$$\boldsymbol{J} = \begin{bmatrix} \beta S_0 - (d + \rho + \alpha) & \gamma S_0 \\ \theta & -\sigma + r \end{bmatrix} \quad A1.1.$$

Here, we are assuming $\tau = 1$ and $\phi = 1$. $\boldsymbol{J}$ can be decomposed into the $\boldsymbol{F}$ and $\boldsymbol{V}$ matrices such that, $\boldsymbol{J} = \boldsymbol{F} - \boldsymbol{V}$, where $\boldsymbol{F}$ represents the transmission matrix and $\boldsymbol{V}$ represents the transition matrix:

$$\boldsymbol{F} = \begin{bmatrix} \beta S_0 & \gamma S_0 \\ \theta & r \end{bmatrix} \quad A1.2$$

$$\boldsymbol{V} = \begin{bmatrix} d + \rho + \alpha & 0 \\ 0 & \sigma \end{bmatrix} \quad A1.3.$$

Here, we have included shedding, θ, in the transmission matrix as it represents the 'birth' of pathogens in the environmental compartment.

The next generation matrix (*NGM*) is given as,

$$\boldsymbol{NGM} = \boldsymbol{F}.\boldsymbol{V}^{-1}$$

$$\boldsymbol{NGM} = \begin{bmatrix} \frac{\beta S_0}{d+\rho+\alpha} & \frac{\gamma S_0}{\sigma} \\ \frac{\theta}{d+\rho+\alpha} & \frac{r}{\sigma} \end{bmatrix} \quad A1.4.$$

The maximum eigenvalue of the *NGM* determines the stability of the disease-free equilibrium (DFE) and is given as,

$$\lambda = \frac{1}{2}\left(\frac{\beta S_0}{(d+\rho+\alpha)} + \frac{r}{\sigma} + \sqrt{\left(\frac{\beta S_0}{(d+\rho+\alpha)} - \frac{r}{\sigma}\right)^2 + 4\frac{\gamma S_0}{\sigma}\frac{\theta}{(d+\rho+\alpha)}}\right) \qquad A1.5,$$

where $\lambda \leq 1$ indicates stability of the DFE, whereas, $\lambda > 1$ indicates that the DFE is unstable. When the free-living pathogen death rate (σ) is less than its growth rate in the environmental reservoir (r), the second term in *A1.5* is greater than 1 and the sum of the first term and the radical term is greater than 1, such that the full sum exceeds 2 giving λ greater than 1, thus indicating an unstable system.

A2. **General recipe for R_0 calculation for pathogens with complex life-histories**

The reproduction number for facultative pathogens can be defined as a form of the "type" reproduction number introduced by Roberts and Heesterbeek (2003) (Roberts and Heesterbeek 2003; Bani-Yaghoub et al. 2012). Hence, to calculate the R_0, we start with a single infected case (i.e., the primary case) and no free-living pathogens in the reservoir. From this initial condition, we use the next generation matrix (*A1.4*) to project the vector composed of the number of infected hosts and pathogen propagules in the reservoir, $[\boldsymbol{I} \quad \boldsymbol{P}]^{\boldsymbol{T}} = [\mathbf{1} \quad \mathbf{0}]^{\boldsymbol{T}}$, one generation forward giving $\left[\frac{\boldsymbol{\beta S_0}}{\boldsymbol{d+\rho+\alpha}} \quad \frac{\boldsymbol{\theta}}{\boldsymbol{d+\rho+\alpha}}\right]^{\boldsymbol{T}}$. Here, the superscript "$\boldsymbol{T}$" denotes the transpose of the indicated matrices. The first element in the vector is the expected number of new infections caused by direct transmission from the primary case. The second element is the expected number of pathogens shed by the primary infected case during the course of infection. At the next step, we set the first element of this new vector as 0 and project this new vector another generation using the next generation matrix (Roberts and Heesterbeek 2003). We iterate this until there are no more free-living pathogens that are derived from the primary infected case. $\boldsymbol{R_0}$ is

then obtained by adding all first elements of the vectors obtained after each transformation. Following this procedure, a general expression for R_0 can be derived as

$$R_0 = \frac{\beta S_0}{d+\rho+\alpha} + \frac{\gamma S_0}{\sigma}\frac{\theta}{d+\rho+\alpha}\left(1 + \frac{r}{\sigma} + \left(\frac{r}{\sigma}\right)^2 + \cdots\right) \qquad A2.1.$$

As described in the main text, we use this approach to derive an expression for the R_0 of facultative pathogens (Equation *3*). The limit of the series in this expression exists if and only if $r < \sigma$. In this case,

$$1 + \sum_{n=1}^{\infty}\left(\frac{r}{\sigma}\right)^n = \frac{1}{1-\frac{r}{\sigma}} \qquad A2.2$$

and the basic reproduction number reduces to:

$$R_0 = \frac{\beta S_0}{d+\rho+\alpha} + \frac{\gamma S_0}{\sigma}\frac{\theta}{d+\rho+\alpha}\left(\frac{1}{1-\frac{r}{\sigma}}\right) \qquad A2.3.$$

However, if $r > \sigma$,

$$\sum_{n=1}^{\infty}\left(\left(\frac{r}{\sigma}\right)^n\right) = \infty \qquad A2.4.$$

In this scenario, R_0 is infinitely large because whenever the growth rate of the pathogen is greater than the decay rate, one infected host is enough to drive the system towards an endemic equilibrium because the reservoir population can supply an infinite number of environmental infections. In this way, R_0 has limited utility in describing the epidemic potential of facultative pathogens. To highlight the importance of growth in the reservoir to disease dynamics, we can restrict analysis to new infections resulting directly from a single infected individual and the propagules it sheds. We term this quantity the direct reproduction number (R_{0d}). With this constraint, $\frac{\beta S_0}{d+\rho+\alpha}$ is the expected number of new

hosts directly infected by a single infected host and $\frac{\theta}{d+\rho+\alpha}$ is the expected number of free-living pathogens shed into the reservoir by that host. Each of these free-living pathogens result in $\frac{\gamma S_0}{\sigma}$ new infections through environmental transmission. Hence, an average of $\frac{\beta S_0}{d+\rho+\alpha} + \frac{\gamma S_0}{\sigma}\frac{\theta}{d+\rho+\alpha}$ total new infections are directly descendent from the primary infection in a susceptible population. This is the expression for R_{0d}, the direct reproduction number (Equation 5, main text). We note that $R_0 = R_{0d} + \frac{\gamma S_0}{\sigma}\frac{\theta}{d+\rho+\alpha}\left(\frac{r}{\sigma} + \left(\frac{r}{\sigma}\right)^2 + \cdots\right)$.

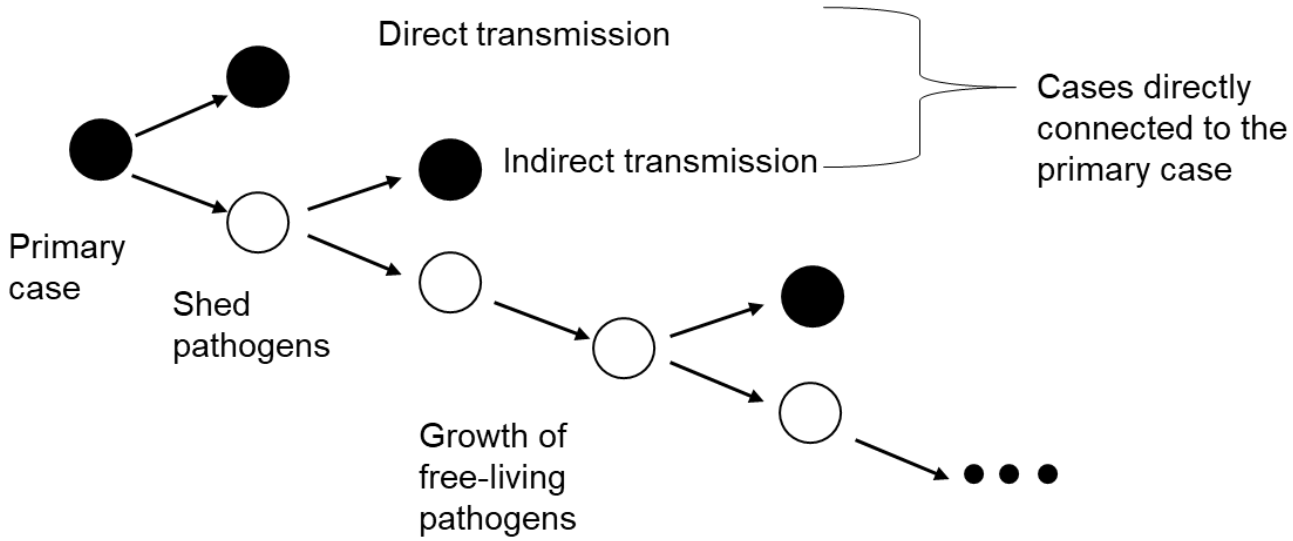

Figure S2.1: Schematic representation of R_0 calculation. To calculate the R_0, all new cases (shown in filled circles) are summed (Equation *3*, main text), whereas calculation of R_{0d} only includes those indicated by the bracket in the illustration. Open circles represent free-living pathogens in the environmental reservoir.

Comparison of the threshold parameters, λ and R_0

The basic reproduction number, R_0, is an important threshold parameter used to evaluate potential epidemiological outcomes in a population. Its derivation for pathogens with complex life-histories can pose a challenge. The next generation matrix method provides an intuitive approach for deriving R_0 of different systems. But how we structure the next generation matrix changes the biological interpretation of the threshold parameter obtained and not all threshold parameters obtained from the next generation matrices represent R_0. For example, including contributions from shedding in the transmission matrix for the Boldin and Kisdi (2012) model of an environmentally transmitted pathogen that cannot reproduce in the environmental reservoir results in the following expression for the maximum eigenvalue (λ) of the next generation matrix,

$$\lambda = \frac{1}{2}\left(\frac{\beta S_0}{(d+\rho+\alpha)} + \sqrt{\left(\frac{\beta S_0}{(d+\rho+\alpha)}\right)^2 + 4\frac{\gamma S_0}{\sigma}\frac{\theta}{(d+\rho+\alpha)}}\right) \qquad A2.4.$$

This expression and Boldin & Kisdi (2012)'s expression for R_0 are quantitatively different under the same parameter values (Figure S2). While our expression for λ and Boldin & Kisdi (2012)'s expression for R_0 are both threshold parameters making qualitatively similar predictions for system behavior, they have different biological meanings and thus are quantitatively different. The latter includes only the epidemiological births in the transmission matrix ($\boldsymbol{F}$). Placing shedding in addition to host-independent growth of free-living pathogens in the $\boldsymbol{F}$ matrix is justified for facultative pathogens since it reflects the 'birth' of new pathogens. Hence, our expression for λ includes the combined average birth of different types, whereas the R_0 expression represents average 'birth' of infected individuals.

We can use our intuitive approach presented above to derive the R_0 for the system presented in Boldin & Kisdi (2012). We have, $r = 0$, $\sigma > 0$, and hence, $\sigma > r$. Then from Equation *3* (main text), we obtain,

$$R_0 = \frac{\beta S_0}{d + \rho + \alpha} + \frac{\gamma S_0}{\sigma} \frac{\theta}{d + \rho + \alpha} \qquad A2.5.$$

This is the same expression obtained by Boldin & Kisdi (2012).

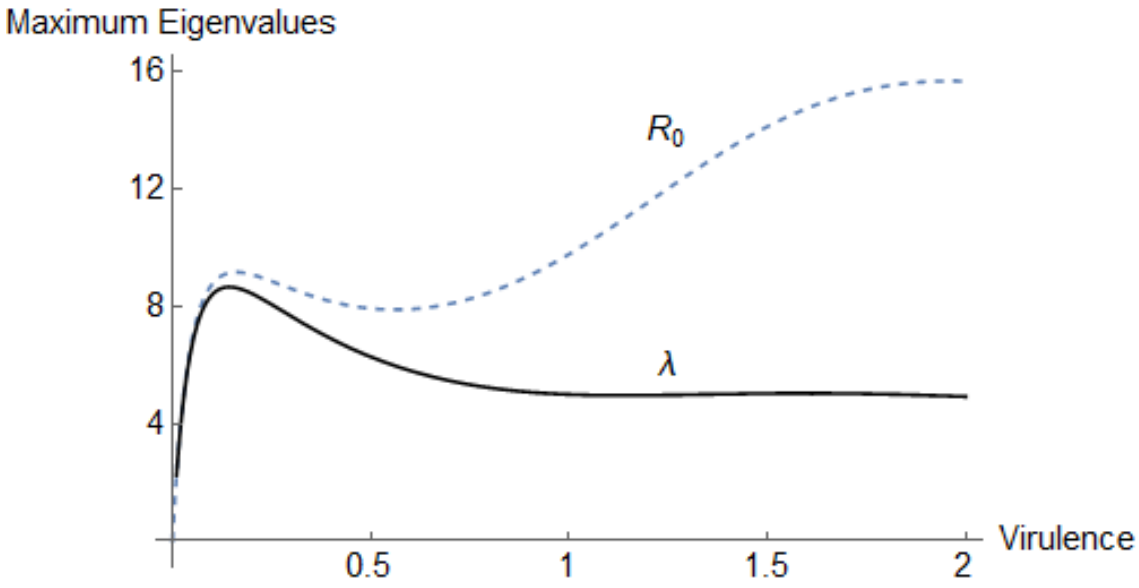

Figure S2.2: Behavior of two threshold parameters, R_0 and λ, obtained from different next generation matrices consideration for the same system presented in Boldin & Kisdi (2012). The dashed line represents the R_0 expression from Boldin & Kisdi (2012) (*A2.5*) and the bold line represents the λ expression (*A2.4*). R_0 and λ give different values for the same parameters, however they both equal 1 at the same trait value. The value of R_0 and λ can be quantitatively different at other trait values. Parameters: $b = 1$, $d = 0.2$, $v = 0.02$, $u = 3$, $g = 1$, $f = 1$, $c = 1$, $h = 0.1$, $\gamma = 0.35$.

A3. Finding Equilibria

To perform the evolutionary invasion analysis, we need to find endemic equilibria as a function of trait values. All potential equilibrium points can be found by solving the following system of equations:

$$b - \beta\hat{S}\hat{I} - \gamma\hat{S}\frac{\hat{P}}{\hat{P}+1} - d\hat{S} = 0 \qquad A3.1$$

$$\beta\hat{S}\hat{I} + \gamma\hat{S}\frac{\hat{P}}{\hat{P}+1} - (d + \rho + \alpha)\hat{I} = 0 \qquad A3.2$$

$$\theta\hat{I} - \sigma\hat{P} + r\hat{P}\left(1 - \frac{\hat{P}}{K}\right) = 0 \qquad A3.3.$$

Adding equations *A3.1* and *A3.2*, we obtain the following expression for equilibria numbers of infected hosts,

$$\hat{I} = \frac{b - d\,\hat{S}}{d + \rho + \alpha} \qquad A3.4.$$

From equations *A3.1* and *A3.4*, we obtain the following relationship,

$$\gamma\hat{S}\frac{\hat{P}}{\hat{P}+1} = b - \beta\hat{S}\left(\frac{b - d\,\hat{S}}{d + \rho + \alpha}\right) - d\hat{S} \qquad A3.5.$$

Similarly, we obtain the following relationship from equation *A3.4*,

$$\theta\left(\frac{b - d\,\hat{S}}{d + \rho + \alpha}\right) = \sigma\hat{P} + r\hat{P}\left(1 - \frac{\hat{P}}{K}\right) \qquad A3.6.$$

Equilibrium points are then the solution of *A3.5* and *A3.6*. Individual expressions of the solution of $\hat{S}$, $\hat{I}$, and $\hat{P}$ are complicated. However, these points can be computed for specific sets of parameter values. These solutions can also be found and visualized graphically. In the case of equations *A3.5* and *A3.6,* equilibrium points occur at the intersection of the two equations (Figure S2.3). The intersection where $\hat{P}$ is zero is the

disease-free equilibrium point, while the intersection where $\hat{S}$, and $\hat{P}$ are positive is the endemic equilibrium point.

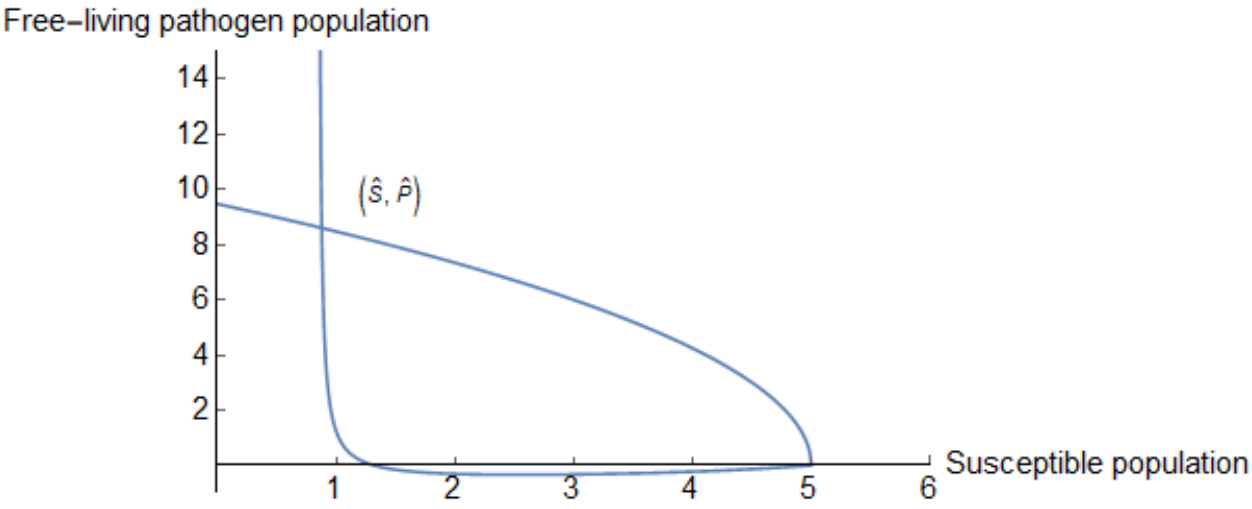

Figure S2.3: Graphical solution of equilibrium points. Parameter values are the same as given in the text where, $b = 1$, $\gamma = 0.35$, $d = 0.2$, $\rho = 0$, $v = 0.02$, $g = 1$, $f = 1$, $c = 1$, $h = 0.1$, $K = 15$, $r = 0.07$ and $\alpha = 1$. The graph intersects at two points in non-negative space. One is $\hat{P} = 0$ and $\hat{S} = b/d$. This is the disease-free equilibrium. The endemic equilibrium occurs at the intersection point where $\hat{S} > 0$ and $\hat{P} > 0$).

A4. Routh-Hurwitz condition for stability

Once the equilibrium points are calculated, we can evaluate the stability of these equilibria. For the disease-free equilibrium, its stability can be found using the techniques present in section A1. The stability of the endemic equilibrium can similarly be found by determining the eigenvalue of the Jacobian of the system (1) evaluated at that equilibrium. With the three non-linear trade-off functions described in the main text, explicit calculation of eigenvalues becomes tedious. Alternatively, Routh-Hurwitz criteria

can be used to evaluate stability. The Jacobian of system 1 at the endemic equilibrium is given as,

$$\boldsymbol{J_{end}} = \begin{bmatrix} \beta\hat{S} - (d+\rho+\alpha) & \dfrac{\gamma\hat{S}}{\left(1+\hat{P}\right)^2} \\ \theta & -\sigma + r - \dfrac{2r}{K}\hat{P} \end{bmatrix} \qquad A4.1.$$

For the endemic point $(\hat{S}, \hat{I}, \hat{P})$ to be stable, the trace of *A4.1* must be negative.

Condition 1:

$$2r\frac{\hat{P}}{K} + \sigma + (d+\rho+\alpha) > \beta\hat{S} + r \qquad A4.2$$

In addition, the determinant of *A4.1* must be positive.

Condition 2:

$$\left\{\beta\hat{S} - (d+\rho+\alpha)\right\}\left\{-\sigma + r - \frac{2r}{K}\hat{P}\right\} > \frac{\theta\gamma\hat{S}}{\left(1+\hat{P}\right)^2} \qquad A4.3$$

Parameter values for all analyses in the manuscript meet both Condition 1 and Condition 2, resulting in systems with a stable endemic equilibrium.

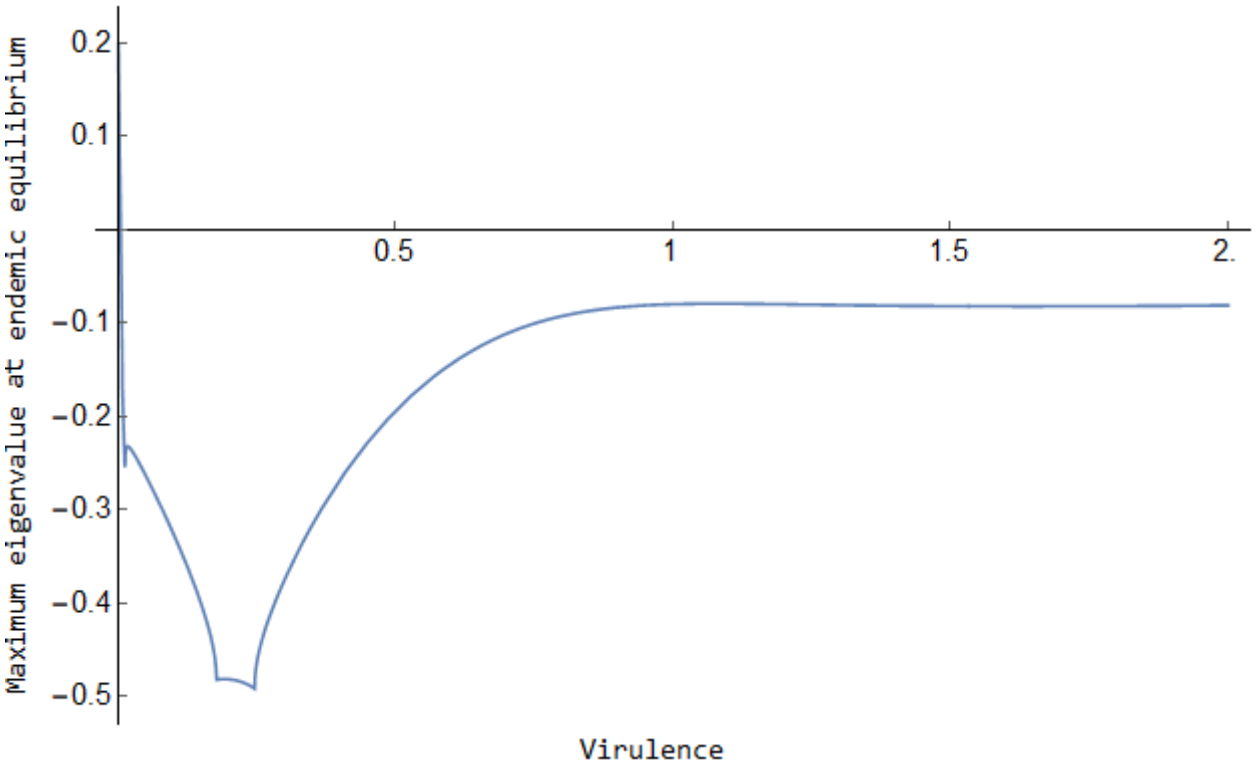

Figure S2.4: Evaluation of the stability of endemic equilibrium for the range of virulence used in the analysis in the main text. The eigenvalues are numerically computed using the parameter values given in Figure S2.3.

A5. Invasion Fitness

We used the Adaptive Dynamics (Dieckmann and Law 1996; Metz et al. 1996; Geritz et al. 1998) approach to assess how several potential life history trade-offs influence virulence evolution. The dynamics of mutant subsystem is given as,

$$\frac{dI_m}{dt} = \beta_m S I_m + \gamma S \frac{P_m}{P + P_m + 1} - (d + \rho + \alpha_m) I_m \qquad A5.1$$

$$\frac{dP_m}{dt} = \theta_m I_m - \sigma_m P_m + r P_m \left(1 - \frac{(P + P_m)}{K}\right) \qquad A5.2,$$

where the subscript '*m*' denotes the mutant trait values, and mutations affecting virulence also impact transmission, shedding, and persistence parameters.

The Jacobian of mutant sub-block when the mutant is rare is given as,

$$J_m|_{P_m=0,I_m=0} = \begin{bmatrix} \beta_m \hat{S} - (d+\rho+\alpha_m) & \gamma \hat{S} \dfrac{1}{1+\hat{P}} \\ \theta_m & -\sigma_m + r\left(1 - \dfrac{\hat{P}}{K}\right) \end{bmatrix} \quad A5.3.$$

Decomposing the matrix, we get,

$$F = \begin{bmatrix} \beta_m \hat{S} & \gamma \hat{S} \dfrac{1}{1+\hat{P}} \\ \theta_m & r\left(1 - \dfrac{\hat{P}}{K}\right) \end{bmatrix} \quad A5.4$$

$$V = \begin{bmatrix} (d+\rho+\alpha_m) & 0 \\ 0 & \sigma_m \end{bmatrix} \quad A5.5.$$

F is the transmission matrix and *V* is transition matrix of the mutant sub-system. The next generation matrix is given as,

$$\boldsymbol{NGM} = F.(V)^{-1}$$

$$\boldsymbol{NGM} = \begin{bmatrix} \dfrac{\beta_m \hat{S}}{(d+\rho+\alpha_m)} & \dfrac{\gamma \hat{S}}{(1+\hat{P})\sigma_m} \\ \dfrac{\theta_m}{(d+\rho+\alpha_m)} & \dfrac{r\left(1 - \dfrac{\hat{P}}{K}\right)}{\sigma_m} \end{bmatrix} \quad A5.6.$$

Here, $\hat{S}$ and $\hat{P}$ represent endemic equilibria. To compute the invasion fitness of a mutant, we follow Hurford et al. (2010). This allows us to decompose the invasion fitness into different components aiding in the analysis and interpretation. The fate of a mutant is given by its long-term growth rate in the environment established by the resident trait. This long-term growth rate, also known as the invasion fitness, is given as

the largest eigenvalue of ***NGM*** (Equation *A5.6*). Using the Perron-Frobenius theorem, the maximum eigenvalue of $\boldsymbol{NGM}$ is given as,

$$\rho(K) = \frac{\vec{v}^{\boldsymbol{T}}\boldsymbol{NGM}\vec{u}}{\vec{v}^{T}\vec{u}} \qquad A5.7.$$

Where, $\vec{v}$ and $\vec{u}$ are left and right eigenvectors, respectively. Elements in $\vec{v}$ are proportional to the reproductive values of each class (i.e. v_1 for the infected class and v_2 for the free-living pathogen class). Similarly, components of $\vec{u}$ are proportional to the equilibrium population of respective classes. We choose the eigenvectors such that $\vec{v}^{T}\vec{u} = 1$, equation *A5.7* can be approximated using Taylor's series as,

$$\rho(K) \approx 1 + \vec{v}^{T}\frac{\partial \boldsymbol{NGM}}{\partial \alpha_m}|_{\alpha=\alpha_m}\vec{u}\,(\alpha_m - \alpha) \qquad A5.8.$$

By definition, $\rho(K)|_{\alpha=\alpha_m} = 1$. Consequently, if the resident and mutant trait values are the same then the invasion fitness is 1. The left and right eigenvectors in $A5.8$ can be calculated for the special case of $\alpha = \alpha_m$.

Expanding *A*5.8 using *A*5.6, we get,

$$W(\alpha_m, \alpha) = 1 + \left(s_d\frac{\partial}{\partial \alpha_m}\left(\frac{\beta_m}{d+\rho+\alpha_m}\right) + s_s\frac{\partial}{\partial \alpha_m}\left(\frac{\theta_m}{d+\rho+\alpha_m}\right) + s_e\frac{\partial}{\partial \alpha_m}\left(\frac{1}{\sigma_m}\right)\right)\Big|_{\alpha_m=\alpha}(\alpha_m - \alpha)\ A5.9$$

where,

$$s_d = \frac{\theta}{d+\rho+\alpha}\,\frac{\gamma\hat{S}}{\sigma\left(1+\hat{P}\right)}\hat{S}$$

$$s_s = \left(1 - \frac{\beta\hat{S}}{d+\rho+\alpha}\right)\frac{\gamma\hat{S}}{\sigma\left(1+\hat{P}\right)}$$

$$s_e = \left(1 - \frac{\beta\hat{S}}{d+\rho+\alpha}\right)\frac{\theta}{d+\rho+\alpha}\frac{\gamma\hat{S}}{(1+P)} + \left(1 - \frac{\beta\hat{S}}{d+\rho+\alpha}\right)^2 r\left(1 - \frac{\hat{P}}{K}\right).$$

Depending on the trade-offs assumed, we can modify *A5.9*. For example, if we assume that virulence is not correlated with pathogen persistence in the environment, then *A5.9* reduces to,

$$W(\alpha_m, \alpha) = 1 + \left(s_d \frac{\partial}{\partial \alpha_m} \left(\frac{\beta_m}{d + \rho + \alpha_m} \right) + s_s \frac{\partial}{\partial \alpha_m} \left(\frac{\theta_m}{d + \rho + \alpha_m} \right) \right) (\alpha_m - \alpha) \qquad A5.10.$$

A6. Threshold value for non-zero free living pathogen population

The Jacobian of system 1 at the disease-free equilibrium, but with pathogen present in environmental reservoir is given as:

$$J = \begin{bmatrix} \beta S_0 - (d + \rho + \alpha) & \frac{\gamma S_0}{1 + P_0} \\ \theta & -\sigma + r\left(1 - \frac{2P_0}{K}\right) \end{bmatrix} \qquad A6.1.$$

where P_0 is the free pathogen population when $I = 0$. In this condition, the third equation of system 1 is independent of the rest of the system because there is no host shedding, and the free-pathogen population reaches an equilibrium given by,

$$P_0 = K\left(1 - \frac{\sigma}{r}\right).$$

If $r > \sigma(\alpha)$, then P_0 is positive but if $r < \sigma(\alpha)$ then P_0 goes to zero. Here we consider situations where the pathogen is present in the reservoir at equilibrium.

The transmission matrix is then given as,

$$\boldsymbol{F} = \begin{bmatrix} \beta S_0 & \frac{\gamma S_0}{1 + P_0} \\ \theta & r\left(1 - \frac{2P_0}{K}\right) \end{bmatrix} \qquad A6.2$$

and the transition matrix is given as,

$$\boldsymbol{V} = \begin{bmatrix} d + \rho + \alpha & 0 \\ 0 & \sigma \end{bmatrix} \qquad A6.3.$$

Again, we have kept shedding $\theta(\alpha)$ in the transmission matrix as it represents the birth of a new individuals in the environmental compartment.

The next generation matrix is then given as,

$$\boldsymbol{NGM} = \boldsymbol{F}.\boldsymbol{V}^{-1}$$

$$\boldsymbol{NGM} = \begin{bmatrix} \dfrac{\beta S_0}{d+\rho+\alpha} & \dfrac{\gamma S_0}{\sigma(1+P_0)} \\ \dfrac{\theta}{d+\rho+\alpha} & \dfrac{r}{\sigma}\left(1-\dfrac{2P_0}{K}\right) \end{bmatrix} \qquad A6.4.$$

λ in this condition is then given as,

$$\lambda = \frac{1}{2}\left(\frac{\beta S_0}{(d+\rho+\alpha)} + \frac{r}{\sigma}\left(1-\frac{2P_0}{K}\right) + \sqrt{\left(\frac{\beta S_0}{(d+\rho+\alpha)} - \frac{r}{\sigma}\left(1-\frac{2P_0}{K}\right)\right)^2 + 4\frac{\gamma S_0}{\sigma(1+P_0)}\frac{\theta}{(d+\rho+\alpha)}}\right). \qquad A6.5.$$

If we substitute the equilibrium expression of $P_0 = K\left(1-\frac{\sigma}{r}\right)$, then the λ expression reduces to:

$$\lambda = \frac{1}{2}\left(\frac{\beta S_0}{(d+\rho+\alpha)} + \frac{2\sigma - r}{\sigma} + \sqrt{\left(\frac{\beta S_0}{(d+\rho+\alpha)} - \frac{2\sigma-r}{\sigma}\right)^2 + 4\frac{r\gamma S_0}{\sigma(r+K(r-\sigma)}\frac{\theta}{(d+\rho+\alpha)}}\right). \qquad A6.6.$$

As in *A*1.5, $\lambda \leq 1$ indicates a stable disease-free equilibrium whereas $\lambda > 1$ indicates an unstable disease-free equilibrium. This threshold parameter (*A6.6*) can be used to evaluate the potential for disease outbreaks when a rare environmental infection occurs or the environmental transmission coefficient (γ) increases suddenly. The latter may

occur when contact between susceptible hosts and free-living pathogens in the reservoir increases or the probability of infection upon contact increases thereby facilitating a sudden increase in environmental transmission.

A7. Critical Function Analysis

Trade-off between persistence and virulence. We follow Kisdi (2006) to show the results of critical function analysis for a single trade-off considered in Section III of the main-text where we consider a single trade-off of free-living pathogen persistence-virulence. From the equation 8 of main text, the singular points are obtained whenever

$$\left(s_d\beta \frac{\partial}{\partial \alpha_m}\left(\frac{1}{d+\rho+\alpha_m}\right) + s_s\theta \frac{\partial}{\partial \alpha_m}\left(\frac{1}{d+\rho+\alpha_m}\right) + s_e \frac{\partial}{\partial \alpha_m}\left(\frac{1}{\sigma(\alpha_m)}\right)\right)\Bigg|_{\alpha_m=\alpha} = 0 \quad A7.1.$$

The critical function $\sigma_{crit}(\alpha)$, in terms of pathogen decay rate, is a solution of the differential equation

$$\sigma'_{crit}(\alpha) + \sigma^2_{crit}(\alpha)\frac{s_d\beta + s_s\theta}{s_e(d+\rho+\alpha)^2} = 0 \quad A7.2,$$

where the subscript $crit$ is written to differentiate it from the trade-off. Solutions of equation $A7.2$ with different initial conditions are a family of critical functions. The point where a trade-off is tangent to the critical function is a singular point. If the trade-off function is more concave than the critical function at the point of tangent, then the singular strategy is convergence stable. Mathematically, the condition for convergence stable at singular point is given as,

$$\frac{d}{d\alpha}\left(s_d\beta \frac{\partial}{\partial \alpha_m}\left(\frac{1}{d+\rho+\alpha_m}\right) + s_s\theta \frac{\partial}{\partial \alpha_m}\left(\frac{1}{d+\rho+\alpha_m}\right) + s_e \frac{\partial}{\partial \alpha_m}\left(\frac{1}{\sigma(\alpha_m)}\right)\right)\Bigg|_{\alpha_m=\alpha} < 0 \quad A7.3$$

Evolutionary stability depends on the following second derivative,

$$E = \left(s_d \beta \frac{\partial^2}{\partial \alpha_m^2} \left(\frac{1}{d + \rho + \alpha_m} \right) + s_s \theta \frac{\partial^2}{\partial \alpha_m^2} \left(\frac{1}{d + \rho + \alpha_m} \right) + s_e \frac{\partial^2}{\partial \alpha_m^2} \left(\frac{1}{\sigma(\alpha_m)} \right) \right) \Bigg|_{\alpha_m = \alpha} < 0 \quad A7.4.$$

The evolutionary stability ($A7.4$) along with the convergence stability given by the convexity of trade-off determine the nature of singular strategy. For example, if the singular strategy is convergence stable but evolutionarily unstable (i.e., $E > 0$), then the singular strategy is an evolutionary branching point. By varying the shape of trade-offs, one can determine the location and type of strategies attained by evolution in trait space. As an example, the solutions of critical functions for the equation $A7.2$ with different initial conditions is shown in Figure S2.5 along with different trait correlations assumed in the Section III of the main text. The critical function will only be tangent to the trait correlation of the shape shown by brown lines thus resulting in singular strategies in the given strategy space (see Figure 2.5a). Similarly, the black lines will not be tangent to the critical functions in the given space and hence, no singular strategy will exist in this range (see Figure 2.5b). Thus, critical function analysis allows us to rule out trait correlations where singular strategies will not exist.

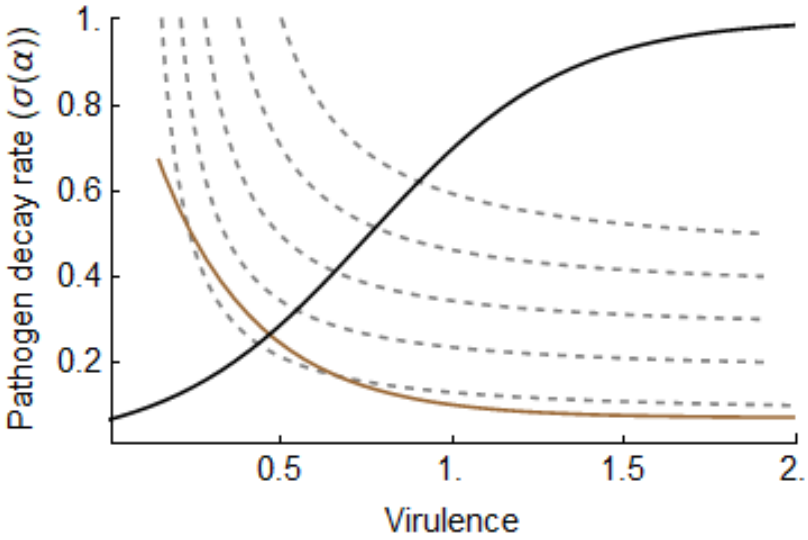

Figure S2.5. Critical function analysis for a single trade-off. Dashed curves are solutions to $A7.2$ for different initial conditions. Brown and black lines represent different trait correlations assumed in section III of the main text. The equation $\sigma(\alpha) = e^{-u\alpha} + v$ represent the brown line whereas the equation $\sigma(\alpha) = \frac{a}{e^{-u\alpha}+v}$ represents the black line. Parameter values: $b = 2$, $\beta = 0.12$, $\theta = 0.3$, $r = 0.3$, $u = 3.5$, $v = 0.07$, $\tau = 1$, $a = 0.07$, $K = 15$, $d = 0.5$, $\gamma = 0.15$.

Multiple trade-offs. Now we consider the case of multiple trait correlations where we assume all three trait correlations (transmission-virulence, shedding-virulence, persistence-virulence) are acting at the same time. Singular trait value is then obtained when,

$$\left(s_d \frac{\partial}{\partial \alpha_m}\left(\frac{\beta(\alpha_m)}{d+\rho+\alpha_m}\right) + s_s \frac{\partial}{\partial \alpha_m}\left(\frac{\theta(\alpha_m)}{d+\rho+\alpha_m}\right) + s_e \frac{\partial}{\partial \alpha_m}\left(\frac{1}{\sigma(\alpha_m)}\right)\right)\Bigg|_{\alpha_m=\alpha} = 0 \qquad A7.5.$$

As mentioned in the main text, if all three trait correlations are unknown, then this becomes an underdetermined system in the sense that there are three unknows for a single equation. For systems where there are at least two environmental feedback variables and just as many trade-offs, it is possible to get any arbitrary evolutionary outcomes (Kisdi 2015). In our system, there are at least two environmental feedback variables given by the density of susceptible and density of free-living pathogens in the reservoir. Given these feedback variables along with the three trait correlations, it is possible to construct the trade-off shapes to obtain arbitrary singularities. This technique can be used to identify potential trade-off shapes for a given evolutionary outcome.

In the main text, we fixed the transmission-virulence and shedding-virulence trade-offs and explore potential outcomes for persistence-virulence relationship. Here, we present an alternate scenario where we fix persistence-virulence and shedding-virulence relationship to explore outcomes for transmission-virulence relationship. The critical function in terms of transmission-virulence relationship is then the solution of

$$\beta'_{crit}(\alpha) - \frac{\beta_{crit}(\alpha)}{d+\rho+\alpha} + \frac{s_s}{s_d}\left(\theta'(\alpha) - \frac{\theta(\alpha)}{d+\rho+\alpha}\right) = \frac{s_e}{s_d}\frac{\sigma'(\alpha)}{\sigma^2(\alpha)}(d+\rho+\alpha) \qquad A7.6.$$

Figure S6 shows an example of the solution of *A7.6* for different initial values. As before, singular trait is where the trade-off is tangent to the critical function. Convergence stability is determined by the condition in $A7.3$ whereas evolutionary stability is determined by the condition in $A7.4$.

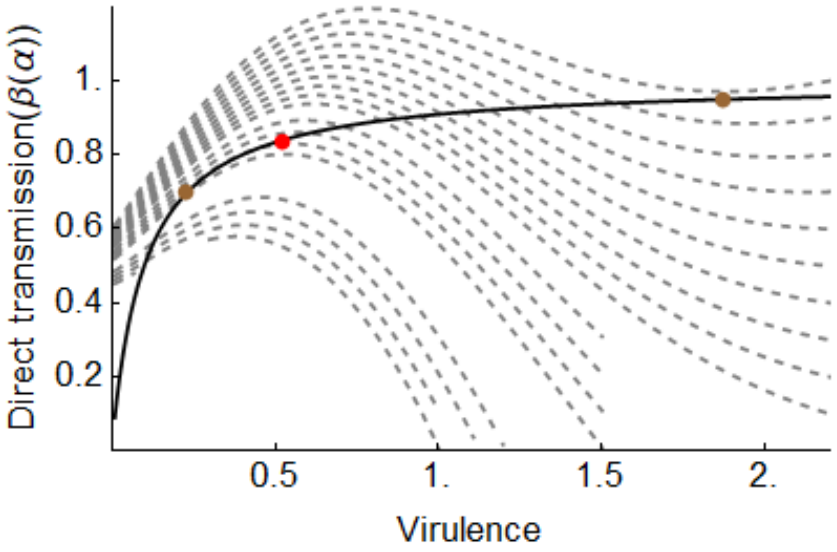

Figure S2.6. Critical function analysis for transmission-virulence relationship. Shedding-virulence relationship is given as $\theta(\alpha) = g\alpha/(g\alpha + f)$ and pathogen decay rate-virulence is given as $\sigma(\alpha) = Exp(-u\alpha) + v$. Dashed lines represent solutions to equation $A7.6$. Black line represents the trade-off function given as $\beta(\alpha) = \frac{c\alpha}{h+c\alpha}$. Brown points represent convergence stable singular points whereas red dot represents

repeller. The trade-off function is more concave than critical function near brown points whereas reverse is true in case of repeller point. Parameter values: $b = 1$, $d = 0.2$, $\theta = 0.3$, $r = 0.3$, $u = 3$, $v = 0.02$, $\tau = 1$, $K = 10$, $f = 1$, $g = 1$, $c = 1$, $h = .1$.

Similar critical function analysis can be performed for shedding-virulence relationship by fixing the transmission-virulence and persistence-virulence trade-offs. In this case, the critical function in terms of shedding-virulence is solution of the following differential equation,

$$\theta'_{crit}(\alpha) - \frac{\theta_{crit}(\alpha)}{d+\rho+\alpha} + \frac{s_d}{s_s}\left(\beta'(\alpha) - \frac{\beta(\alpha)}{d+\rho+\alpha}\right) = \frac{s_e}{s_s}\frac{\sigma'(\alpha)}{\sigma^2(\alpha)}(d+\rho+\alpha) \qquad A7.7$$

Supplemental Materials References

Bani-Yaghoub, M., R. Gautam, Z. Shuai, P. van den Driessche, and R. Ivanek. 2012. Reproduction numbers for infections with free-living pathogens growing in the environment. Journal of Biological Dynamics 6:923–940.

Boldin, B., and É. Kisdi. 2012. On the evolutionary dynamics of pathogens with direct and environmental transmission. Evolution 66:2514–2527.

Dieckmann, U., and R. Law. 1996. The dynamical theory of coevolution: a derivation from stochastic ecological processes. Journal of Mathematical Biology 34:579–612.

Geritz, S. A. H., É. Kisdi, G. Meszéna, and J. A. J. Metz. 1998. Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. Evolutionary Ecology 12:35–57.

Hurford, A., D. Cownden, and T. Day. 2010. Next-generation tools for evolutionary invasion analyses. Journal of The Royal Society Interface 7:561–571.

Kisdi, É. 2006. Trade-off geometries and the adaptive dynamics of two co-evolving species. Evolutionary Ecology Research 8:959–973.

———. 2015. Construction of multiple trade-offs to obtain arbitrary singularities of adaptive dynamics. Journal of Mathematical Biology 70:1093–1117.

Metz, J. A. J., S. A. H. Geritz, G. Meszéna, F. J. . Jacobs, and J. S. van Heerwaarden. 1996. Adaptive Dynamics: A geometrical study of the consequences of nearly faithful reproduction. Stochastic and Spatial Structures of Dynamical Systems 45:183–231.

Roberts, M. G., and J. A. P. Heesterbeek. 2003. A new method for estimating the effort required to control an infectious disease. Proceedings of the Royal Society B: 270:1359–1364.

Supplemental materials for: Coincidental selection stemming from interference competition promotes higher virulence in *Stenotrophomonas maltophilia*

Aakash Pandey[1*], Brandi Lynn Cobe[2], Tyler Hanson[1], Nicholas P. Cianciotto[2], Thomas G. Platt[1]

[1]Division of Biology, Kansas State University, Manhattan, KS, USA

[2]Department of Microbiology-Immunology, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA

*Corresponding author: apandey@ksu.edu

Table S3.1. Strains and strain's sources used in the experiments.

Name	Source
Stenotrophomonas maltophilia strains K279a, R551-3, JCMS, JV3	Michael Herman (University of Nebraska-Licoln)
E. coli OP50	Michael Herman (University of Nebraska-Licoln)
C. elegans wild-type N2	Anna Zinovyeva (Kansas State University)
E. coli MG1655	ATCC®
S. maltophilia K279a mutants Δ*vir*B10, Δ*vir*B10/*vir*B10+, 14245, 14245/14245+, 14255, 14255/14255+, 14245/14255	Nicholas P. Cianciotto (Northwestern university)

Table S3.2. Log-rank test for mean survival of *C. elegans* on 4 different *S. maltophilia* strains.

Condition	Statistics		
	χ^2	P-value	Bonferroni P-value
JV3 v.s. JCMS	44.7	0	0
JV3 v.s. R551-3	49.97	0	0
JV3 v.s. K279a	74.55	0	0
JCMS v.s. JV3	44.7	0	0
JCMS v.s. R551-3	5.11	0.0237	0.0712
JCMS v.s. K279a	13.88	0.0002	0.0006
R551-3 v.s. JV3	49.97	0	0
R551-3 v.s. JCMS	5.11	0.0237	0.0712
R551-3 v.s. K279a	2.23	0.1355	0.4064
K279a v.s. JV3	74.55	0	0
K279a v.s. JCMS	13.88	0.0002	0.0006
K279a v.s. R551-3	2.23	0.1355	0.4064

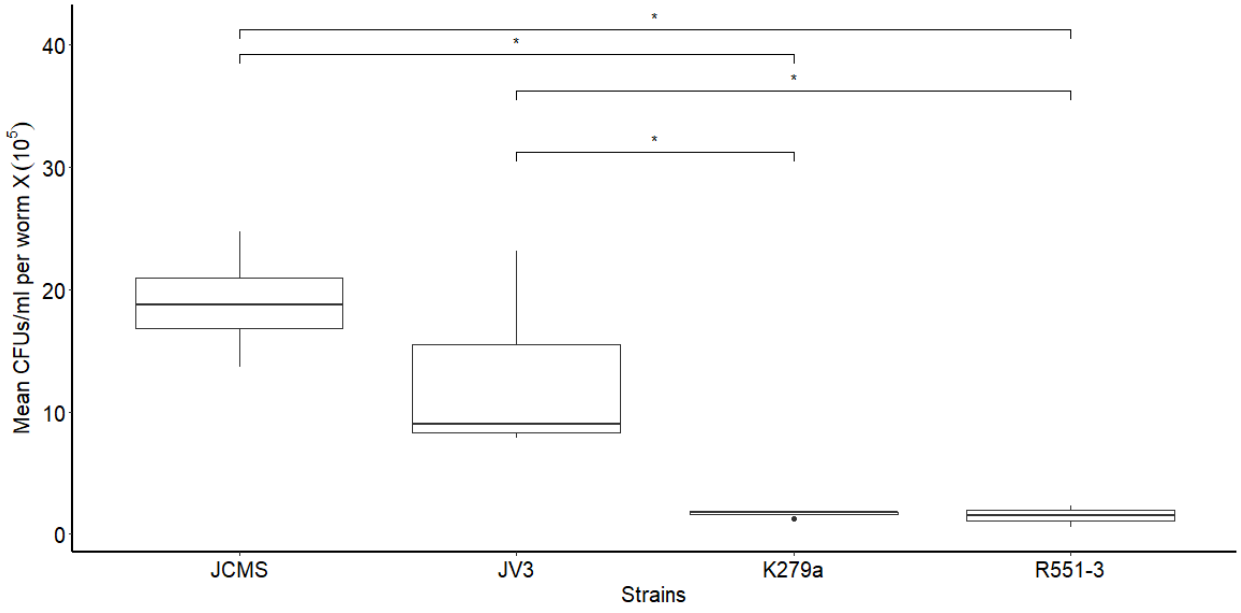

Figure S3.1. Mean CFUs obtained from *C. elegans* interacting with different *S. maltophilia* strains. Significance is evaluated using the Wilcoxon test.

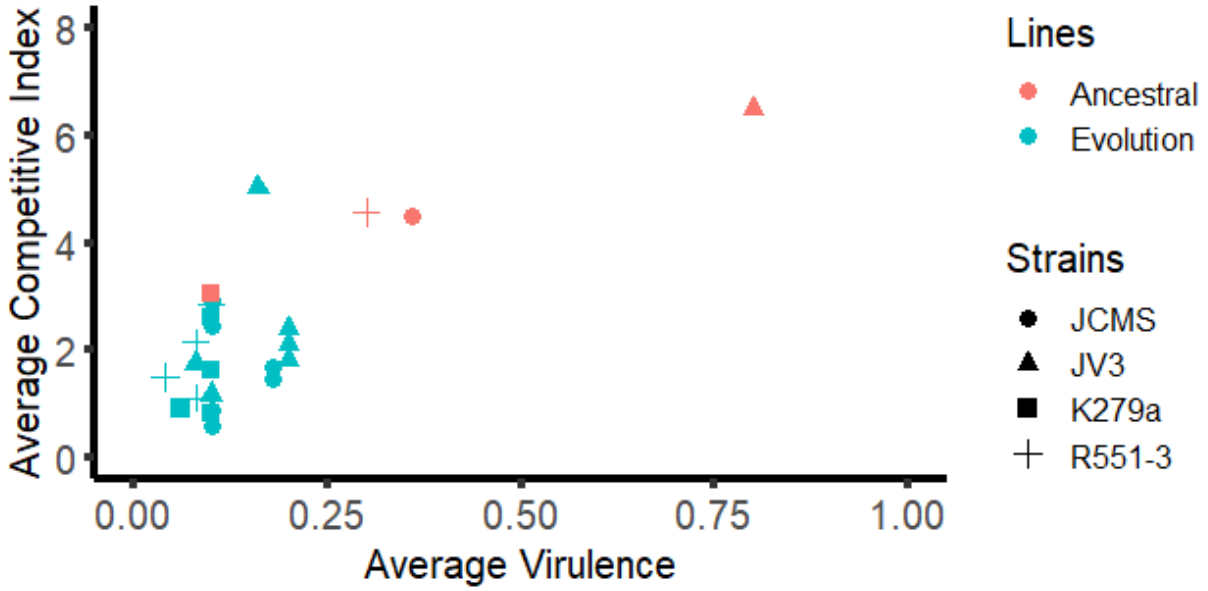

Figure S3.2. Correlation between average virulence and competitive ability of wildtype and evolved *S. maltophilia* strains. Virulence is measured here as the proportion of dead worms after 72 hours of infection. Competitive index is the ratio of *S. maltophilia*:*E.coli* CFUs at time point 2 hrs to the ratio of *S. maltophilia*:*E.coli* CFUs time point 0 hrs.

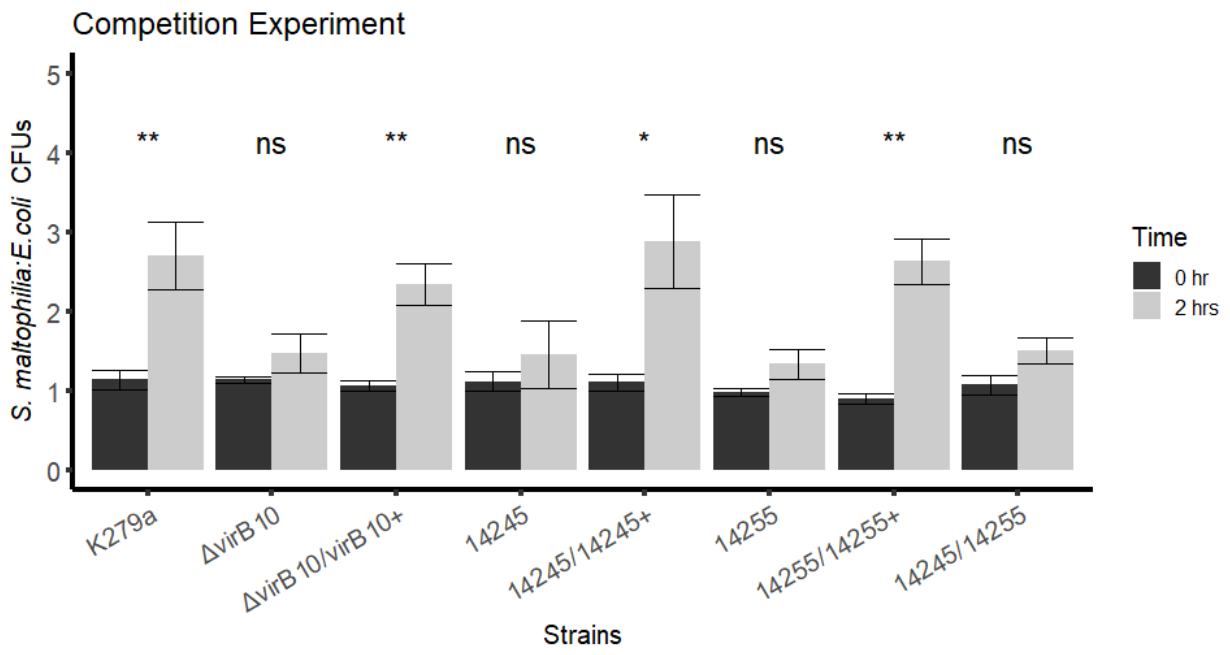

Figure S3.3. Competition experiment showing the effect of type IV secretion system and its effector proteins on interference antagonism of *E. coli* MG1655. The significance is evaluated using the Wilcoxon test (n=6).

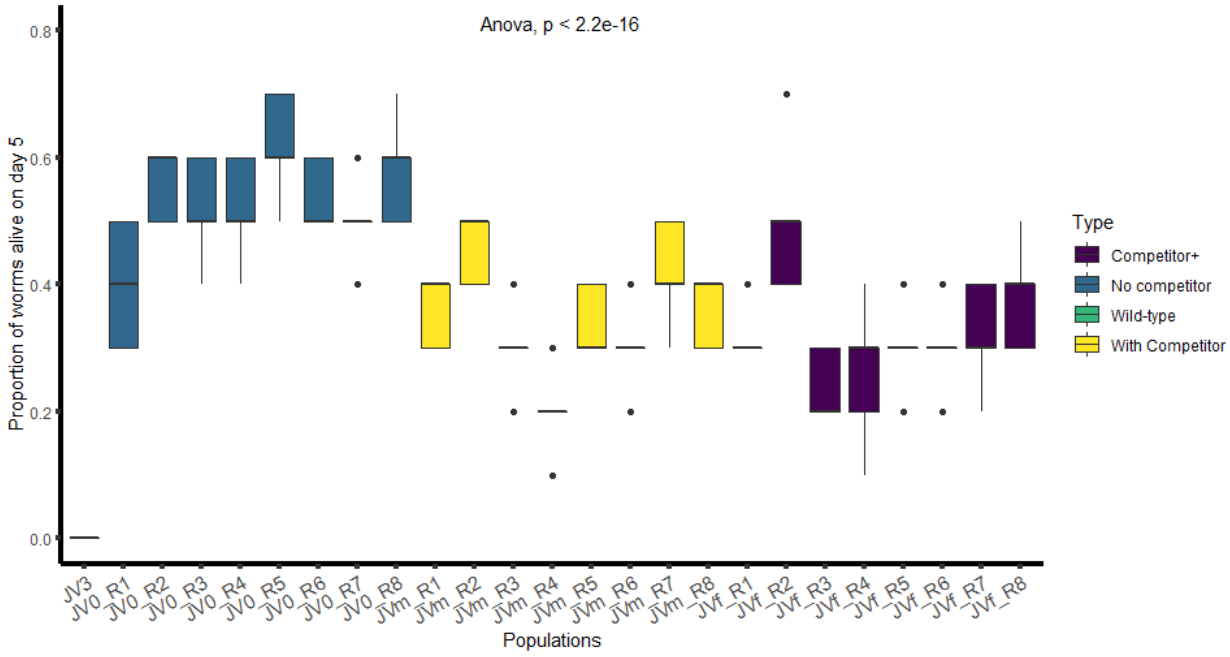

Figure S3.4. Survival of *C. elegans* on day 5 post infection from population lines evolved under different conditions. Blue boxes represent populations where competitor *E. coli* were not present, yellow boxes represent populations where *E. coli* was present at the start of the experiment, and purple boxes represent the population where *E. coli* cells were added to the media at each transfer.

TableS3.3. Log-rank test for mean survival of *C. elegans* on populations derived from evolution experiments.

Condition	Statistics		
	χ^2	P-value	Bonferroni P-value
Ancestral v.s. No competitor	61.13	0	0
Ancestral v.s. With competitor	37.16	0	0
Ancestral v.s. Competitor+	28.39	9.9e-8	3e-7
No competitor v.s. Ancestral	61.13	0	0
No competitor v.s. With competitor	20.2	0.000007	0.000021
No competitor v.s. Competitor+	28.73	8.3e-8	2.5e-7
With competitor v.s. Ancestral	37.16	0	0
With competitor v.s. No competitor	20.2	0.000007	0.000021
With competitor v.s. Competitor+	0.93	0.3346	1
Competitor+ v.s. Ancestral	28.39	9.9e-8	3e-7
Competitor+ v.s. No competitor	28.73	8.3e-8	2.5e-7
Competitor+ v.s. With competitor	0.93	0.3346	1

Supplementary materials for. Neutral and niche-based assembly of microbiome communities in agricultural and prairie soils

Aakash Pandey[1*], Paige Hansen[2, 3], Willow Kessler[2], Marcos Sarto[4], Ligia Souza[2], Austin Yoder[5], James D. Bever[2], Sharon Billings[2], Matthew Kirk[6], Terry Loecke[5,7], Charles W. Rice[4], Benjamin A. Sikes[2,7], Thomas G. Platt[1]

[1]Division of Biology, Kansas State University, Manhattan, KS, USA

[2]Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, KS, USA

[3]Department of Soil and Crop Sciences, Colorado State University, Fort Collins, CO, USA

[4]Department of Agronomy, Kansas State University, Manhattan, KS, USA

[5]Kansas Biological Survey, University of Kansas, Lawrence, KS, USA

[6]Department of Geology, Kansas State University, Manhattan, KS, USA

[7]Department of Environmental Studies, University of Kansas, Lawrence, KS, USA

*** Correspondence.**

Aakash Pandey

evomathbio@gmail.com

DNA extraction optimization

Given the many different sample types and soils we collected, we tested multiple different protocols and kits to optimize DNA extraction for all samples. In this, we trialed ~14 different protocols on 0-5 cm and 75+ cm samples that had random combinations of annual precipitation and land use treatments. We selected the extraction methodology that allowed us to get as high a DNA concentration as possible, out of as many sample types as possible. This ultimately led us to extract total genomic DNA from 0.25 g of soil and from 1 blank (molecular-grade water) control per season using the DNeasy PowerSoil Pro Kit (Qiagen, Germantown, MD, USA), following the manufacturer's protocol.

Despite our DNA extraction optimization, soil clay content was still too high in many samples to extract a workable quantity of DNA. This was especially the case in soils at the 30-75 cm depth and in soils deeper than 75 cm. Of the 356 total samples extracted, 17 (6 spring and 11 fall) had DNA concentrations less than 0.5 ng/mL. However, because prior efforts have indicated that some of these failed extractions could still produce useful sequencing libraries, we prepared libraries for all DNA extracts following the amplicon sequencing protocols described in the main text and below, regardless of whether the extraction "failed." To sequence as many samples as possible without diluting our final library pool, we excluded samples that had a lower final DNA concentration than the blank control extraction at the amplicon purification step immediately prior to pooling. Further information on samples excluded from the final 16S and ITS pools is located in Table S4.1.

Detailed 16S and ITS amplicon sequencing protocol

To generate bacterial community data, we amplified the V4 region of the 16S small subunit ribosomal gene using the Earth Microbiome Project (Thompson et al., 2017) primers combined with Illumina adaptor sequences. Our forward primer included the Illumina® (Illumina, Inc., San Diego, CA, USA) i5 overhang adapter sequence (5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3') attached to primer 515F-Y (5'-GTGYCAGCMGCCGCGGTAA-3'; Parada et al., 2016), and our reverse primer included the Illumina i7 overhang adapter sequence (5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-3') attached to primer 806R (5'-GGACTACNVGGGTWTCTAAT-3'; Apprill et al., 2015). Each 16S PCR reaction was prepared using 1X Q5® (New England Biolabs, Ipswich, MA, USA) reaction buffer, 1X Q5 GC enhancer, 0.02 U/μl High Fidelity (HF) Q5 polymerase, 2 mM dNTPs, 2 μM each forward and reverse primer, 1 μl DNA, and molecular-grade water to 25 μl. PCR amplifications consisted of a 30 s denaturation at 98°C; 25 cycles of 10 s at 98°C, 30 s at 57°C, and 30 s at 72°C; and a final elongation for 2 min at 72°C. 16S amplicons were then purified using the NucleoMag® PCR prep kit (Macherey-Nagel Gmbh & Co., Düren, Germany). Illumina Nextera™ indices were ligated to the purified amplicons in a second PCR reaction consisting of 1X Q5 reaction buffer, 1X Q5 GC enhancer, 0.02 U/μl Q5 HF polymerase, 0.3 mM dNTPs, 3 μl of each Nextera XT v2 index, 1 μl PCR product, and molecular-grade water to 50 μl. These amplifications consisted of a 30 s denaturation at 98°C; 8 cycles of 10 s at 98°C, 30 s at 55°C, and 30 s at 72°C; and a 5 min final elongation at 72°C. Amplicons were purified a second time using the NucleoMag PCR prep kit, and then pooled in equimolar concentrations in two groups that corresponded to samples

taken in spring and fall seasons. We chose to pool and sequence samples when they were first available because we did not have enough Nextera index combinations to sequence all samples in a single pool. At the library pooling step, four and three samples were excluded from the spring and fall pools, respectively, because their final DNA concentrations were lower than the blank negative control (Table S4.1). Both pooled libraries were sequenced at the University of Kansas Genomic Sequencing Core (Lawrence, KS, USA) using Illumina 2x300 bp MiSeq v2 chemistry, resulting in 71,148 ± 20,223 and 86,905 ± 55,127 (mean ± sd) high-quality reads per sample in the spring and fall pools.

To generate fungal community data, we amplified the ITS2 region between the 5.8S and the 28S ribosomal genes. Our forward primer consisted of the Illumina i5 overhang adapter sequence attached to primer ITS7 (5'- GTGARTCATCGAATCTTTG-3', Ihrmark et al., 2012), and our reverse primer consisted of the Illumina i7 overhang adapter sequence attached to primer ITS4 (5'- TCCTCCGCTTATTGATATGC -3', White et al., 1990). Each ITS2 PCR reaction was prepared using 1X Q5 reaction buffer, 1X Q5 GC enhancer, 0.02 U/μl Q5 HF polymerase, 2 mM dNTPs, 2 μM each forward and reverse primer, 1 μl DNA, and molecular-grade water to 40 μl. From this point on, our ITS protocol is identical to our 16S protocol. We amplified ITS amplicons using the same thermocycling conditions, purified ITS amplicons using the NucleoMag PCR prep kit, and ligated Nextera indices using the reaction mix and thermocycling conditions described above. These PCR products were purified using the NucleoMag PCR prep kit once more, and pooled in equimolar concentrations in two groups that correspond to samples taken in the spring and fall. The spring pool excluded 6 and the fall pool excluded 7 samples with a final DNA

concentration lower than the negative control (Table S4.1). Both pools were sequenced at the Kansas State University Genomic Sequencing Core (Manhattan, KS, USA) with Illumina 2x300 bp MiSeq v2 sequencing, resulting in 77,779 ± 19,124 and 86,365 ± 30,201 (mean ± sd) high-quality reads per sample in the spring and fall pools.

Table S4.1. Number of samples from each treatment excluded at various steps in the library generation and sequence processing pipeline, as well as total number of samples with high-quality data included in statistical analyses. Treatments that have less than 4 replicates are highlighted in darker colors in the rightmost columns.

Region	**Land use**	**Soil depth (cm)**	**Number of replicates sampled**		**Number of replicates excluded at library pooling**				**Number of replicates excluded at quality filtering**				**Total number of replicates included in final data set**			
			Spring	**Fall**	**Spring**		**Fall**		**Spring**		**Fall**		**Spring**		**Fall**	
					16S	**ITS**	**16S**	**ITS**	**16S**	**ITS**	**16S**	**ITS**	**16S**	**ITS**	**16S**	**ITS**
West	Ag	0-5	4	4									4	4	4	4
West	Ag	5-15	4	4									4	4	4	4
West	Ag	15-30	4	4									4	4	4	4

West	Ag	30-75	4	4				1					4	4	4	3
West	Ag	75+	4	4	1	2	1	3	1				2	2	2	1
West	Restored	0-5	4	4									4	4	4	4
West	Restored	5-15	4	4									4	4	4	4
West	Restored	15-30	4	4									4	4	4	4
West	Restored	30-75	4	4									4	4	4	4
West	Restored	75+	4	4	1			1					3	4	4	3
West	Native	0-5	4	4									4	4	4	4
West	Native	5-15	4	4									4	4	4	4
West	Native	15-30	4	4									4	4	4	4
West	Native	30-75	4	4									4	4	4	4
West	Native	75+	4	4		1					1		4	3	3	4

Central	Ag	0-5	4	4					1	1			3	3	4	4
Central	Ag	5-15	4	4									4	4	4	4
Central	Ag	15-30	4	4									4	4	4	4
Central	Ag	30-75	4	4					1				3	4	4	4
Central	Ag	75+	4	4		1	2		1				3	3	2	4
Central	Restored	0-5	4	4									4	4	4	4
Central	Restored	5-15	4	4						1			4	3	4	4
Central	Restored	15-30	4	4						1			4	3	4	4
Central	Restored	30-75	4	4									4	4	4	4
Central	Restored	75+	4	4								2	4	4	4	2
Central	Native	0-5	4	4									4	4	4	4

Central	Native	5-15	3	4									3	3	4	4
Central	Native	15-30	4	4									4	4	4	4
Central	Native	30-75	4	4									4	4	4	4
Central	Native	75+	4	4		1							4	3	4	4
East	Ag	0-5	4	4									4	4	4	4
East	Ag	5-15	4	4									4	4	4	4
East	Ag	15-30	4	4								1	4	4	4	3
East	Ag	30-75	4	4						1			4	3	4	4
East	Ag	75+	4	4	1			1					3	4	4	3
East	Restored	0-5	4	4									4	4	4	4
East	Restored	5-15	4	4									4	4	4	4
East	Restored	15-30	4	4									4	4	4	4

East	Restored	30-75	4	4					1				4	4	4	4
East	Restored	75+	4	4	1				1		1		3	4	3	4
East	Native	0-5	4	4									4	4	4	4
East	Native	5-15	4	4									4	4	4	4
East	Native	15-30	4	4							1		4	4	3	4
East	Native	30-75	4	4				1					4	4	4	3
East	Native	75+	3	0		1			2				1	2	0	0

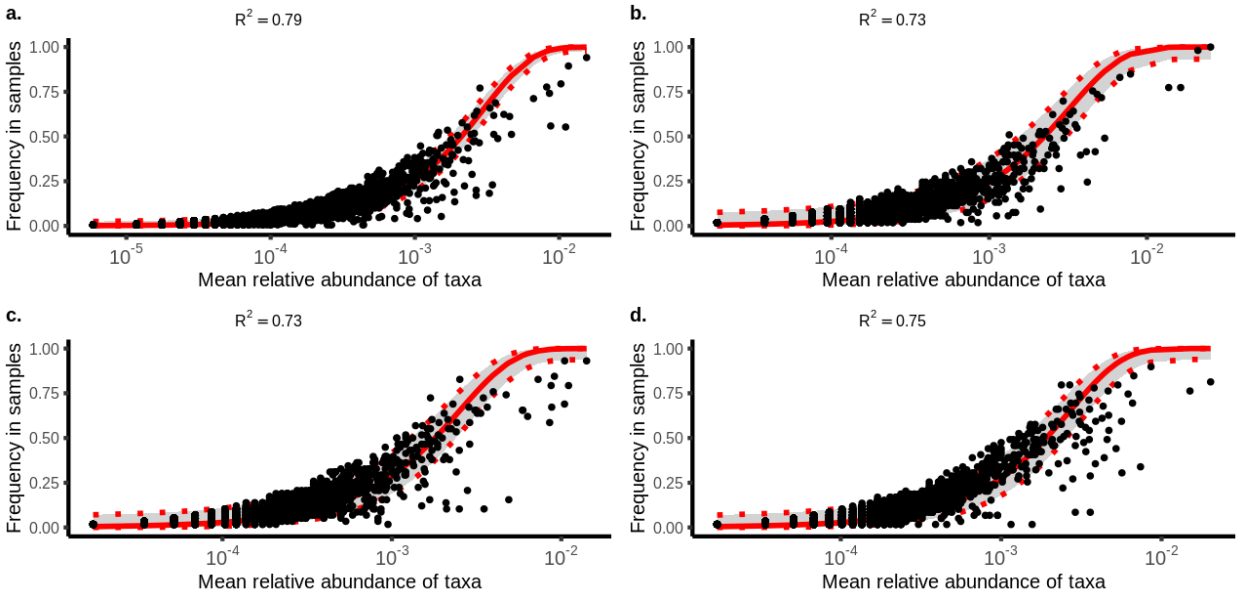

Figure S4.1. Neutral model fit for bacterial samples (ASVs) at different scales. a. shows fit for all samples collected in spring 2018. b. shows fit for the samples from site 1, c. shows fit for the samples from site 2, and d. shows fit for the samples from site 3. Red line represents prediction from the neutral model, gray shading represents 95% confidence interval, and black dot represents the observed taxa. Site 1 is in eastern Kansas, site 2 is within Konza Prairie, and site 3 is near Fort Hays.

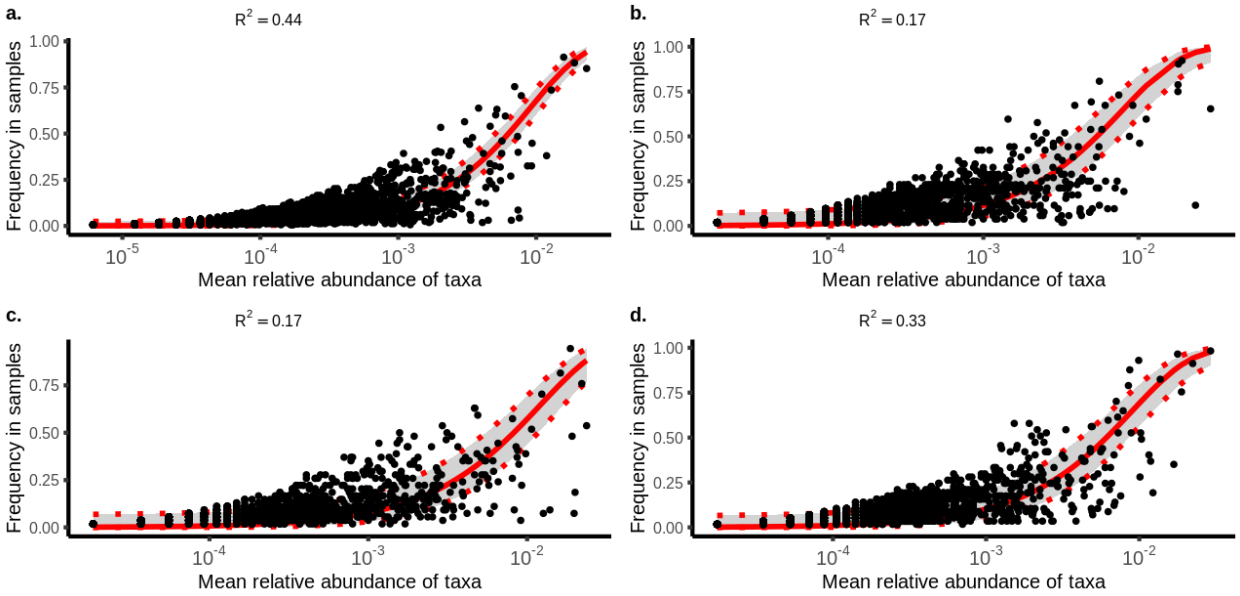

Figure S4.2. Neutral model fit for fungal samples (ASVs) at different scales. a. shows fit for all samples collected in fall 2018. b. shows fit for the samples from site 1, c. shows fit for the samples from site 2, and d. shows fit for the samples from site 3. Red line represents prediction from the neutral model, gray shading represents 95% confidence interval, and black dot represents the observed taxa.

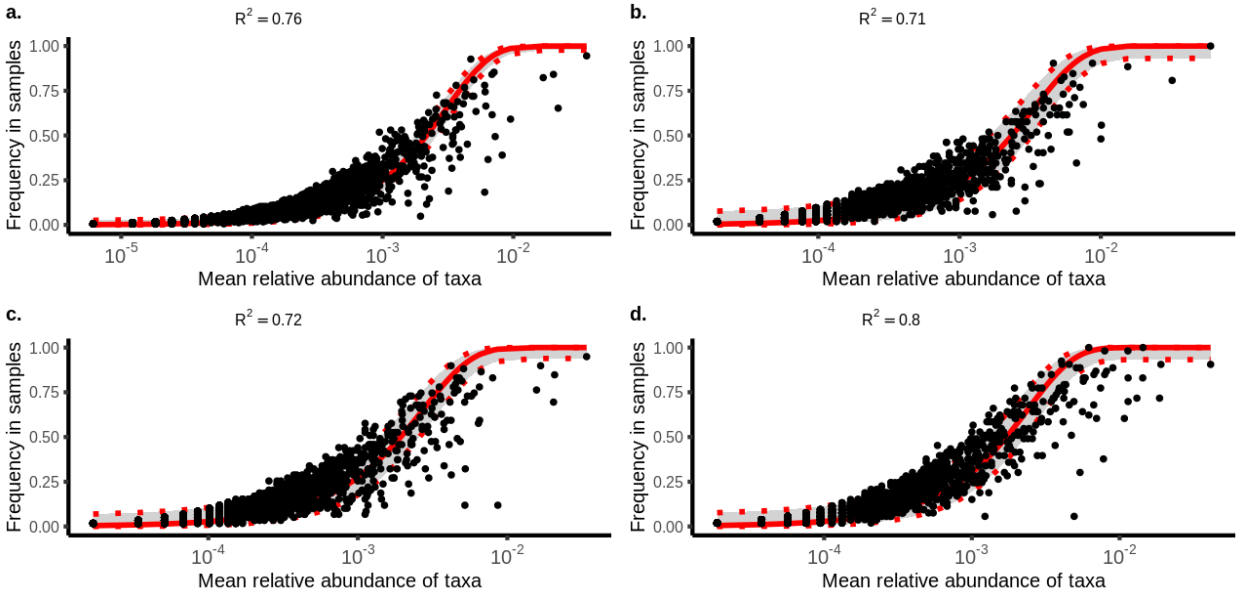

Figure S4.3. Neutral model fit for bacterial samples (OTUs) at different scales. a. shows fit for all samples collected in spring 2018. b. shows fit for the samples from site 1, c. shows fit for the samples from site 2, and d. shows fit for the samples from site 3. Red line represents prediction from the neutral model, gray shading represents 95% confidence interval, and black dot represents the observed taxa. Site 1 is in eastern Kansas, site 2 is within Konza Prairie, and site 3 is near Fort Hays.

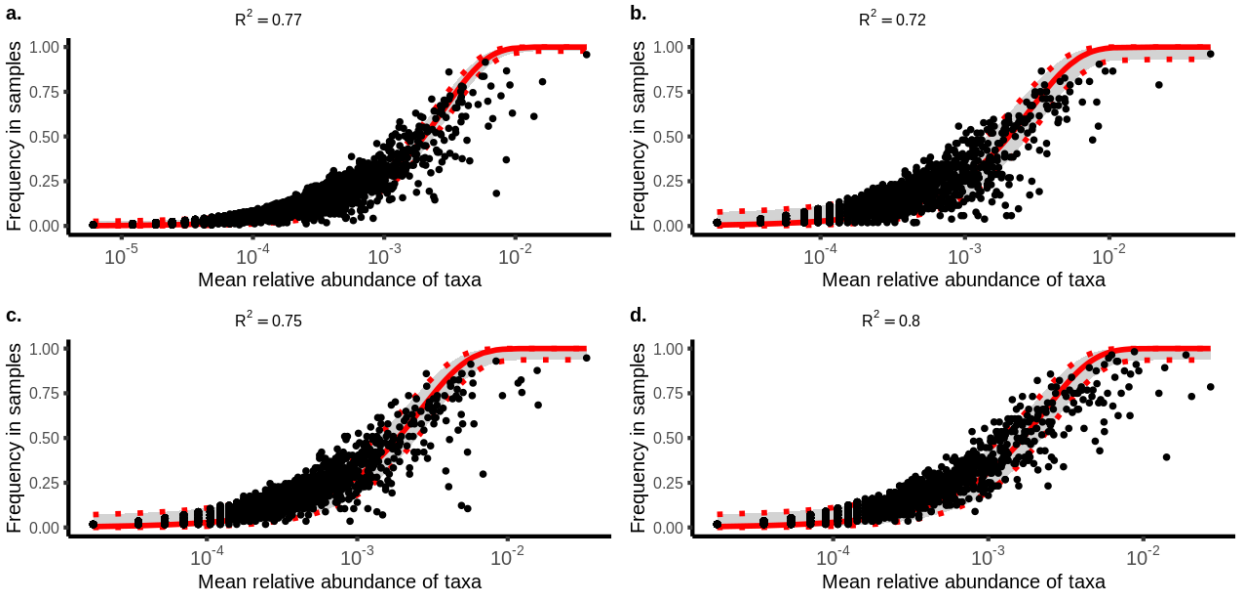

Figure S4.4. Neutral model fit for bacterial samples (OTUs) at different scales. a. shows fit for all samples collected in fall 2018. b. shows fit for the samples from site 1, c. shows fit for the samples from site 2, and d. shows fit for the samples from site 3. Red line represents prediction from the neutral model, gray shading represents 95% confidence interval, and black dot represents the observed taxa. Site 1 is in eastern Kansas, site 2 is within Konza Prairie, and site 3 is near Fort Hays.

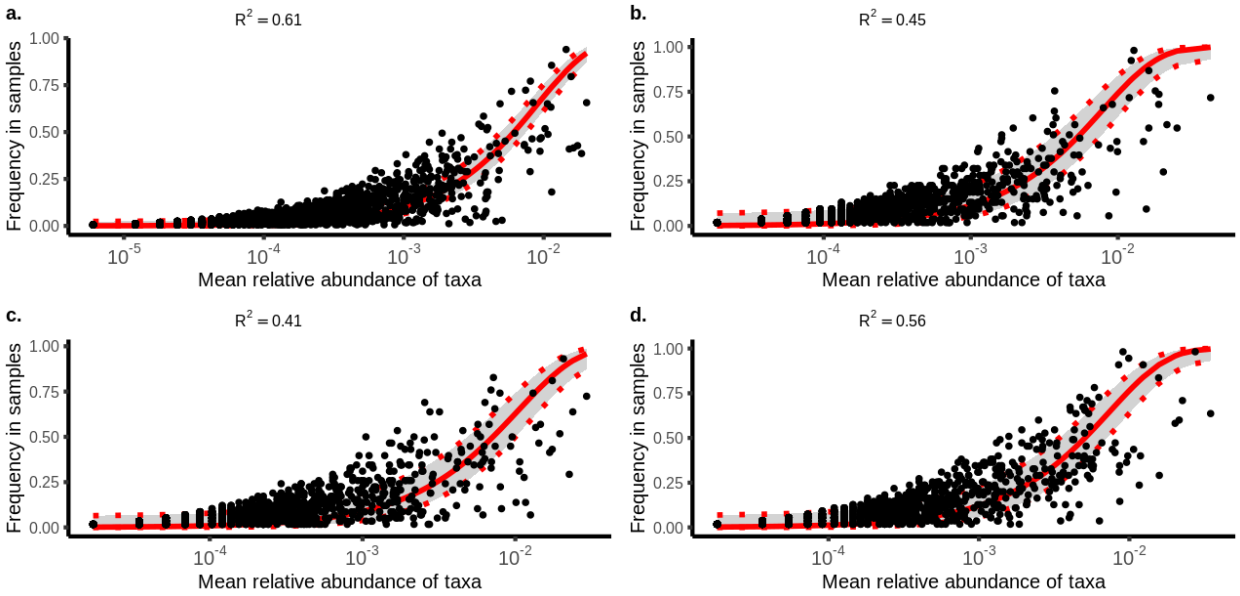

Figure S4.5. Neutral model fit for fungal samples (OTUs) at different scales. a. shows fit for all samples collected in spring 2018. b. shows fit for the samples from site 1, c. shows fit for the samples from site 2, and d. shows fit for the samples from site 3. Red line represents prediction from the neutral model, gray shading represents 95% confidence interval, and black dot represents the observed taxa. Site 1 is in eastern Kansas, site 2 is within Konza Prairie, and site 3 is near Fort Hays.

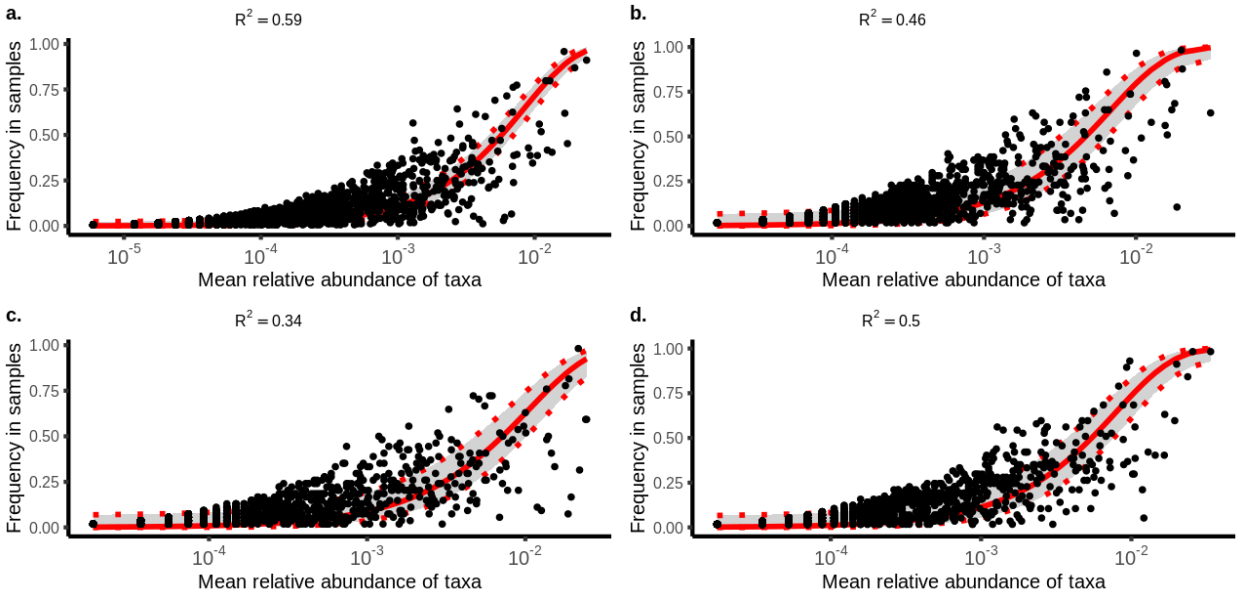

Figure S4.6. Neutral model fit for fungal samples (OTUs) at different scales. a. shows fit for all samples collected in fall 2018. b. shows fit for the samples from site 1, c. shows fit for the samples from site 2, and d. shows fit for the samples from site 3. Red line represents prediction from the neutral model, gray shading represents 95% confidence interval, and black dot represents the observed taxa. Site 1 is in eastern Kansas, site 2 is within Konza Prairie, and site 3 is near Fort Hays.

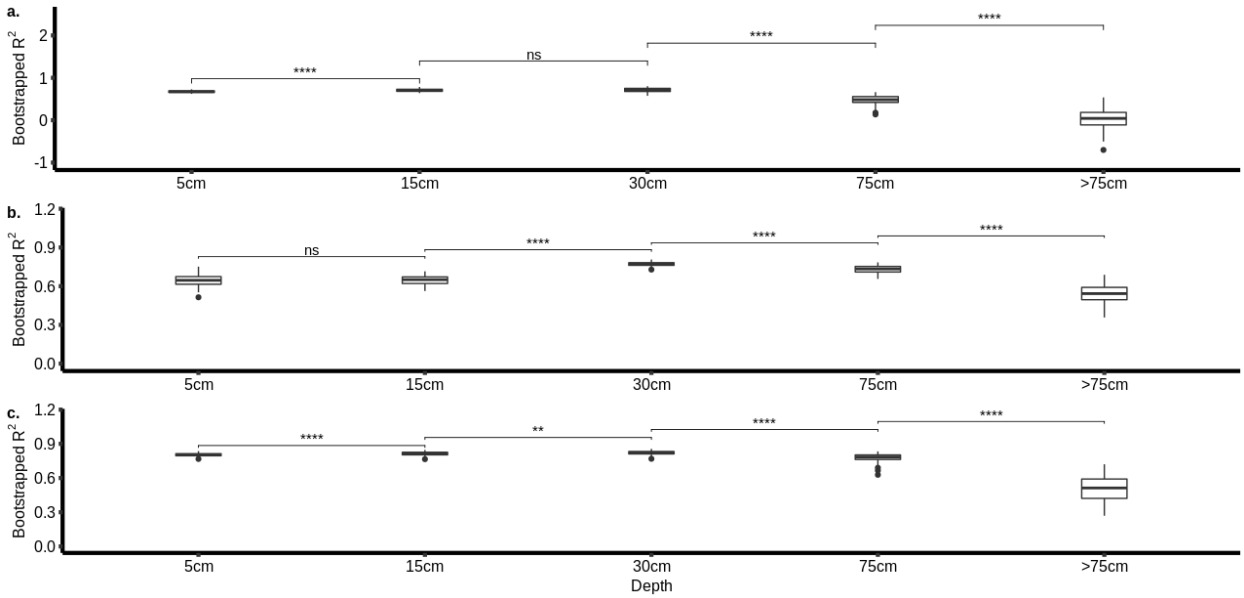

Figure S4.7. Bootstrapped confidence interval of goodness of fit to the neutral model of bacterial samples (ASVs) obtained from different depths. a. Goodness of fit to samples from site 1. b. Goodness of fit to samples from site 2. c. Goodness of fit to samples from site 3.

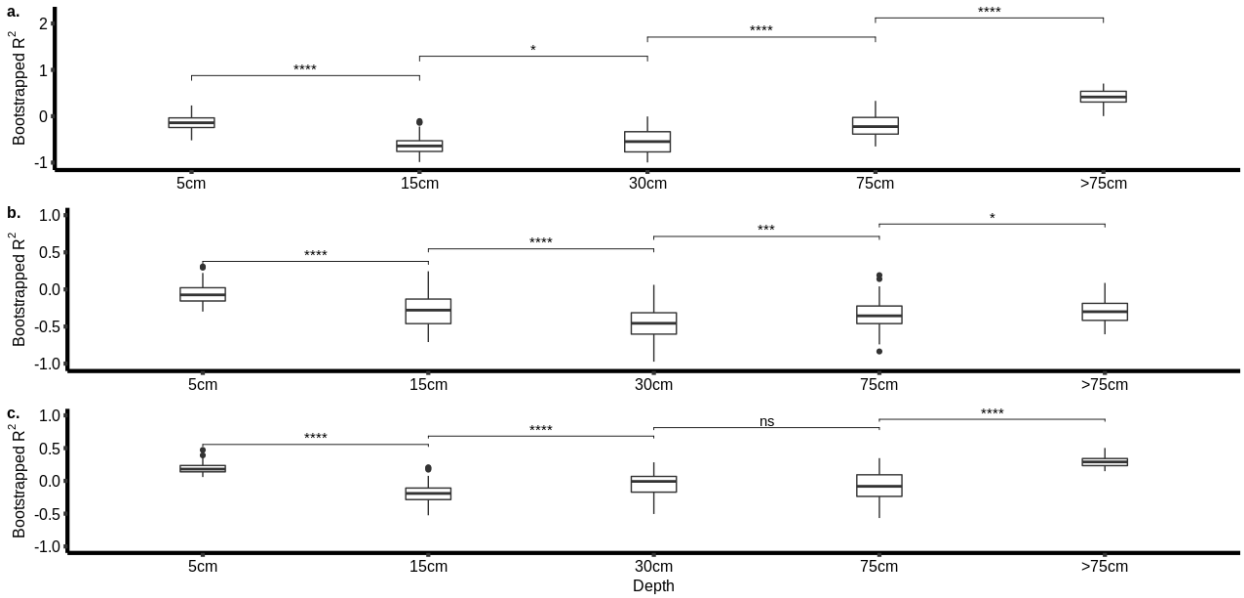

Figure S4.8. Bootstrapped confidence interval of goodness of fit to the neutral model of fungal samples (ASVs) obtained from different depths. a. Goodness of fit to samples from

site 1. b. Goodness of fit to samples from site 2. c. Goodness of fit to samples from site 3.

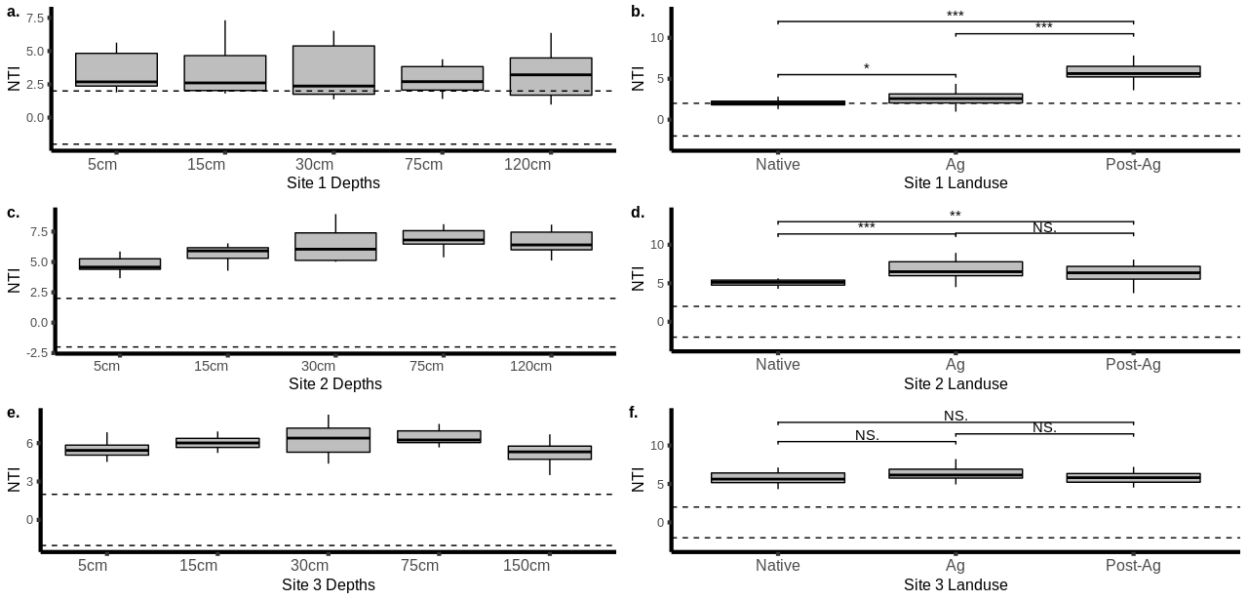

Figure S4.9. Distribution of Near Taxon Index (NTIs) values for bacterial samples (ASVs) obtained from different depth and land use profiles in fall 2018. The dashed line represents significant levels (-2 and +2). NTIs greater than 2 represents significant clustering whereas less than -2 represents significant overdispersion compared to random expectations. The top panel (a. and b.) represents sample from site 1, the middle panel (c. and d.) represents samples from site 2, and the bottom panel (e. and f.) represents samples from site 3.

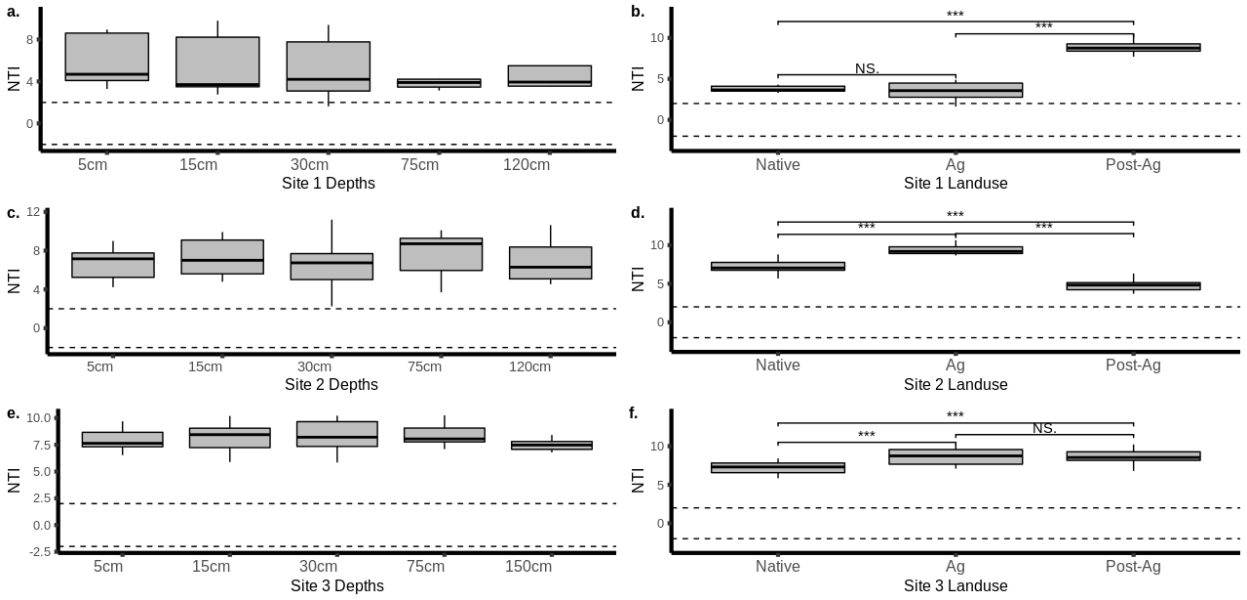

Figure S4.10. Distribution of Near Taxon Index (NTIs) values for bacterial samples (OTUs) obtained from different depth and land use profiles in fall 2018. The dashed line represents significant levels (-2 and +2). The top panel (a. and b.) represents sample from site 1, the middle panel (c. and d.) represents samples from site 2, and the bottom panel (e. and f.) represents samples from site 3.

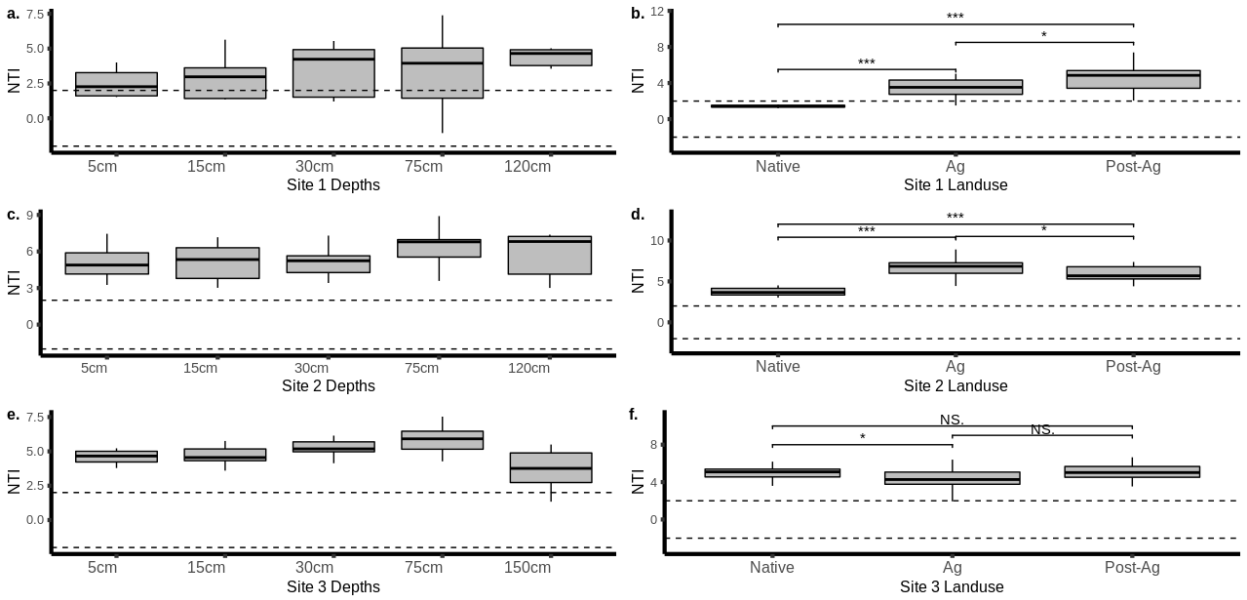

Figure S4.11. Distribution of Near Taxon Index (NTIs) values for bacterial samples (ASVs) obtained from different depth and land use profiles in spring 2018. The dashed line represents significant levels (-2 and +2). The top panel (a. and b.) represents sample from site 1, the middle panel (c. and d.) represents samples from site 2, and the bottom panel (e. and f.) represents samples from site 3.

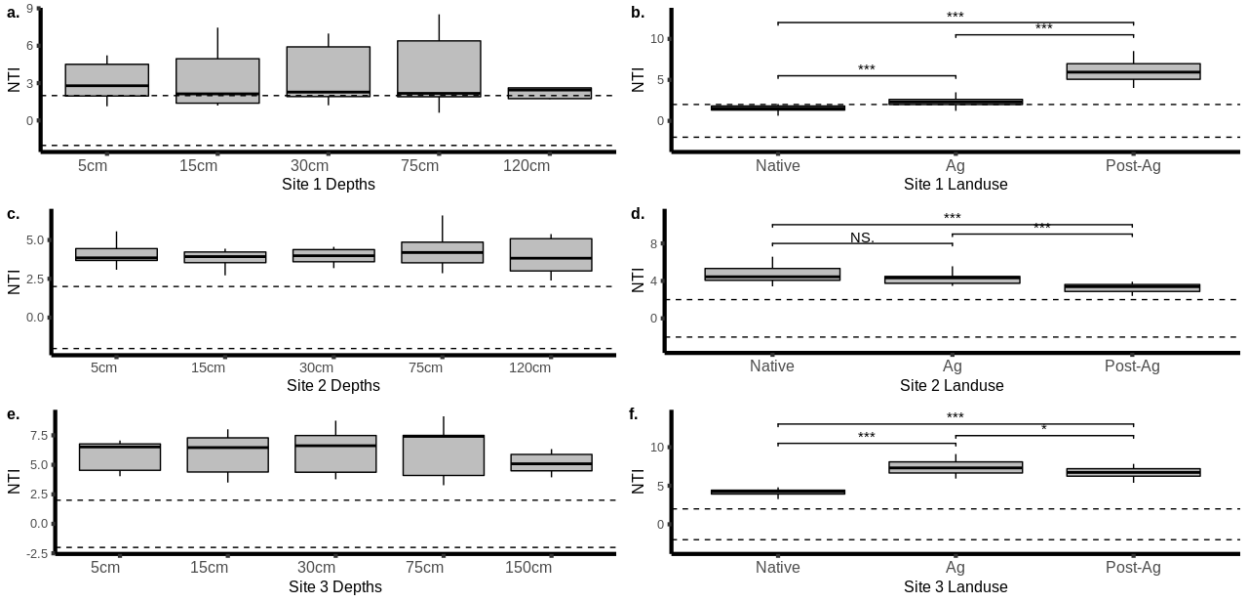

Figure S4.12. Distribution of Near Taxon Index (NTIs) values for bacterial samples (OTUs) obtained from different depth and land use profiles in spring 2018. The dashed line represents significant levels (-2 and +2). The top panel (a. and b.) represents sample from site 1, the middle panel (c. and d.) represents samples from site 2, and the bottom panel (e. and f.) represents samples from site 3.

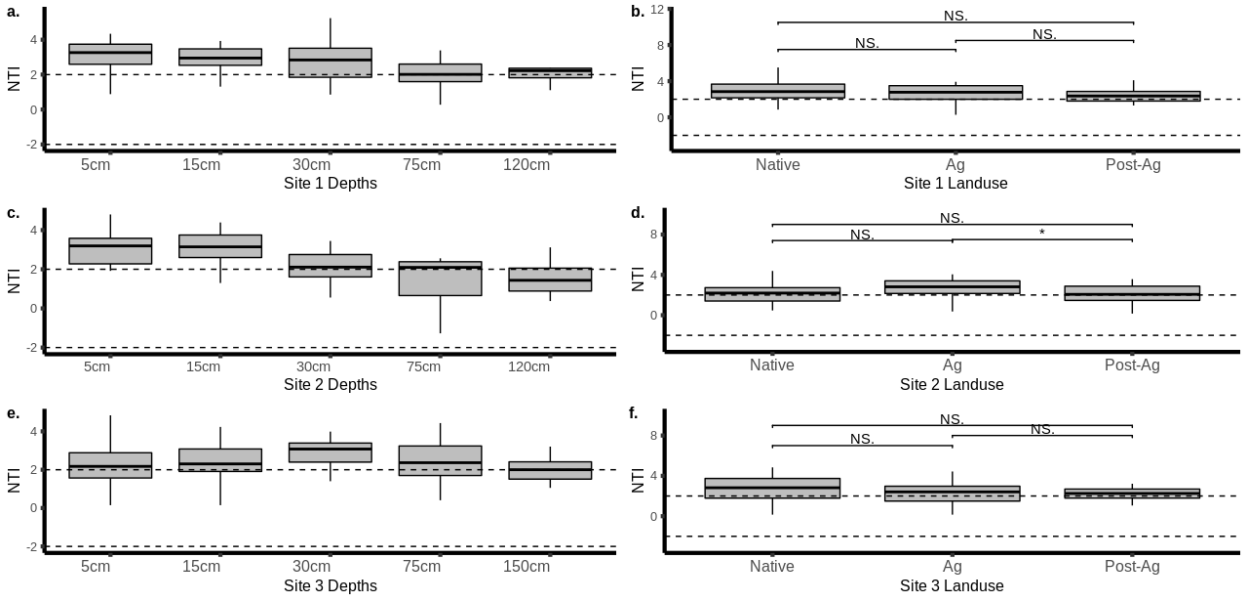

Figure S4.13. Distribution of Near Taxon Index (NTIs) values for fungi samples (ASVs) obtained from different depth and land use profiles in fall 2018. The dashed line represents significant levels (-2 and +2). The top panel (a. and b.) represents sample from site 1, the middle panel (c. and d.) represents samples from site 2, and the bottom panel (e. and f.) represents samples from site 3.

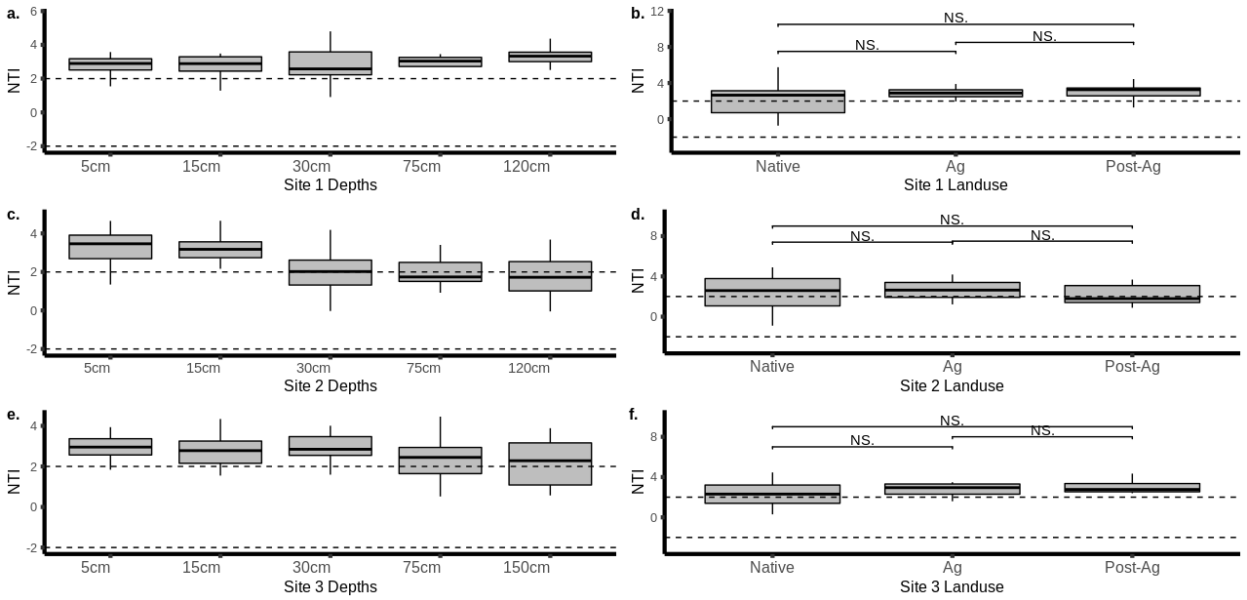

Figure S4.14. Distribution of Near Taxon Index (NTIs) values for bacterial samples (ASVs) obtained from different depth and land use profiles in spring 2018. The dashed line represents significant levels (-2 and +2). The top panel (a. and b.) represents sample from site 1, the middle panel (c. and d.) represents samples from site 2, and the bottom panel (e. and f.) represents samples from site 3.

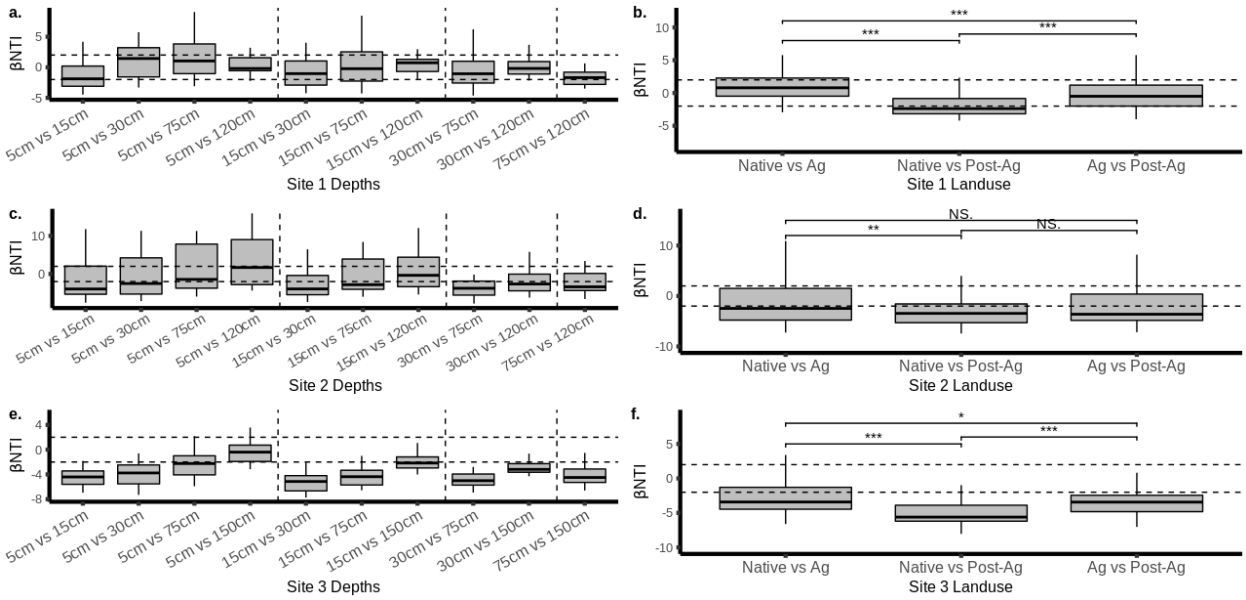

Figure S4.15. βNTI between bacterial samples (ASVs) obtained from different depths and landuse history. All samples are from 2018 spring. In figures b., d., and f., Ag represents samples from agricultural land and Post-Ag represents sample from post-agricultural/restored land. Horizontal dashed lines show upper and lower significance threshold values at $\beta NIT = +2$ and $\beta NTI = -2$, respectively.

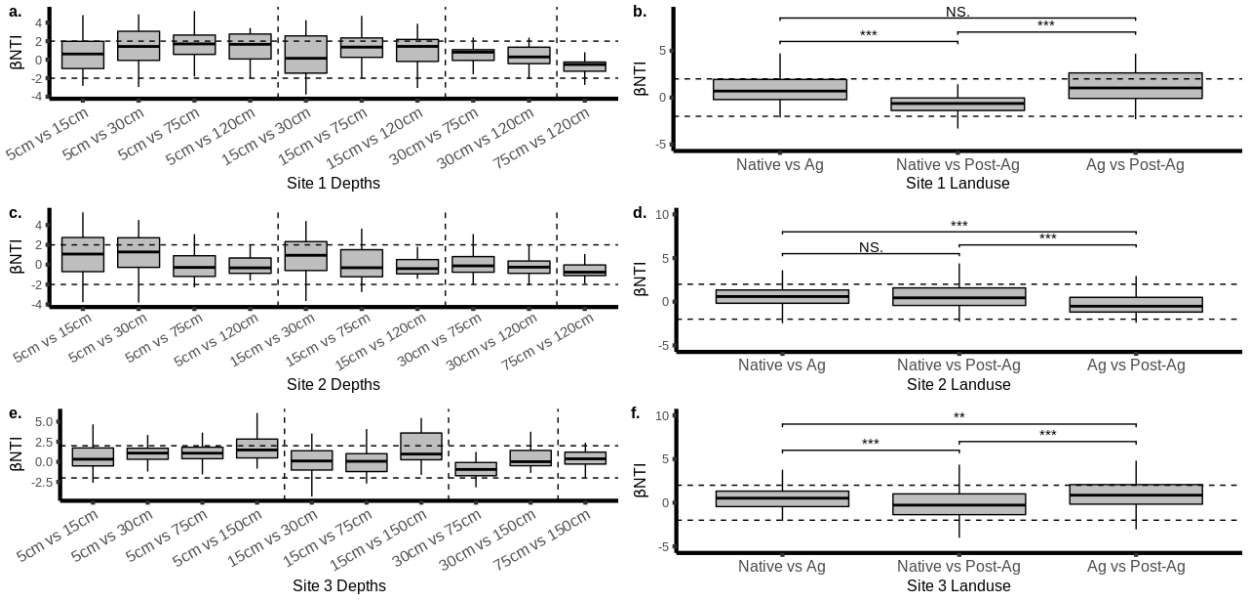

Figure S4.16. βNTI between fungal samples (ASVs) obtained from different depths and landuse history. All samples are from 2018 spring. In figures b., d., and f., Ag represents samples from agricultural land and Post-Ag represents sample from post-agricultural/restored land. Horizontal dashed lines show upper and lower significance threshold values at $\beta NIT = +2$ and $\beta NTI = -2$, respectively.

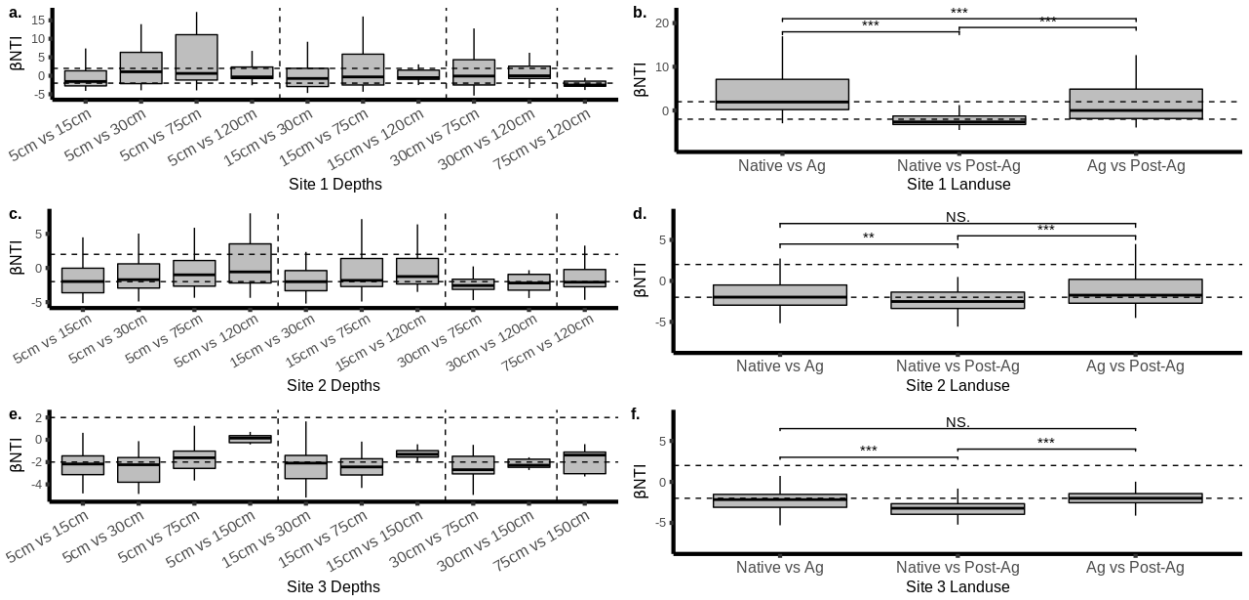

Figure S4.17. βNTI between bacterial samples (OTUs) obtained from different depths and land use history. All samples are from fall 2018. In figures b., d., and f., Ag represents samples from agricultural land and Post-Ag represents sample from post-agricultural/restored land. Horizontal dashed lines show upper and lower significance threshold values at $\beta NTI = +2$ and $\beta NTI = -2$, respectively.

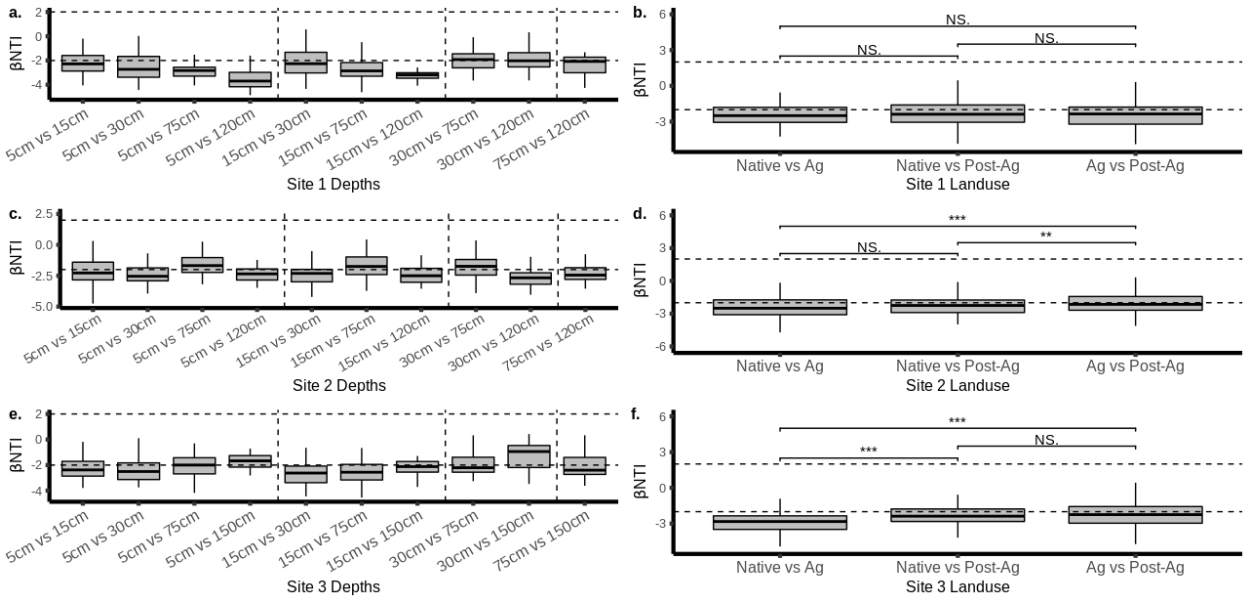

Figure S4.18. βNTI between fungal samples (OTUs) obtained from different depths and land use history. All samples are from fall 2018. In figures b., d., and f., Ag represents samples from agricultural land and Post-Ag represents sample from post-agricultural/restored land. Horizontal dashed lines show upper and lower significance threshold values at $\beta\mathrm{NTI} = +2$ and $\beta\mathrm{NTI} = -2$, respectively

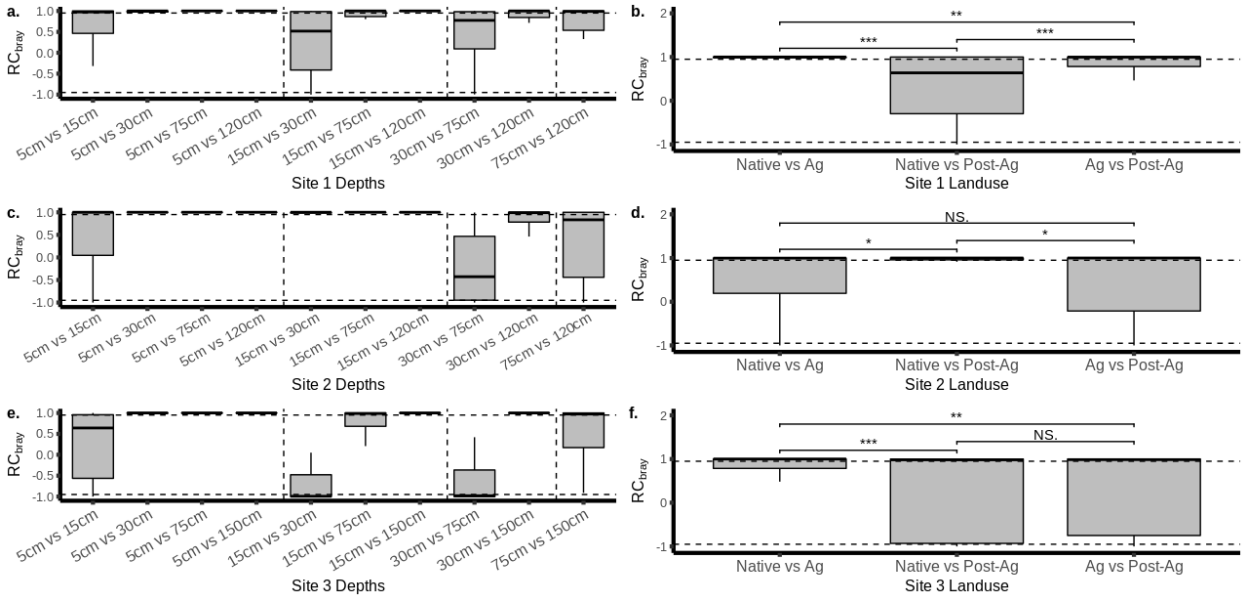

Figure S4.19. $\mathrm{RC}_{\mathrm{bray}}$ between bacterial samples (ASVs) obtained from different depths and land use history. All samples are from fall 2018. In figures b., d., and f., Ag represents samples from agricultural land and Post-Ag represents sample from post-agricultural/restored land. Horizontal dashed lines show upper and lower significance threshold values at $\mathrm{RC}_{\mathrm{bray}} = 0.95$ and $\mathrm{RC}_{\mathrm{bray}} = -0.95$, respectively.

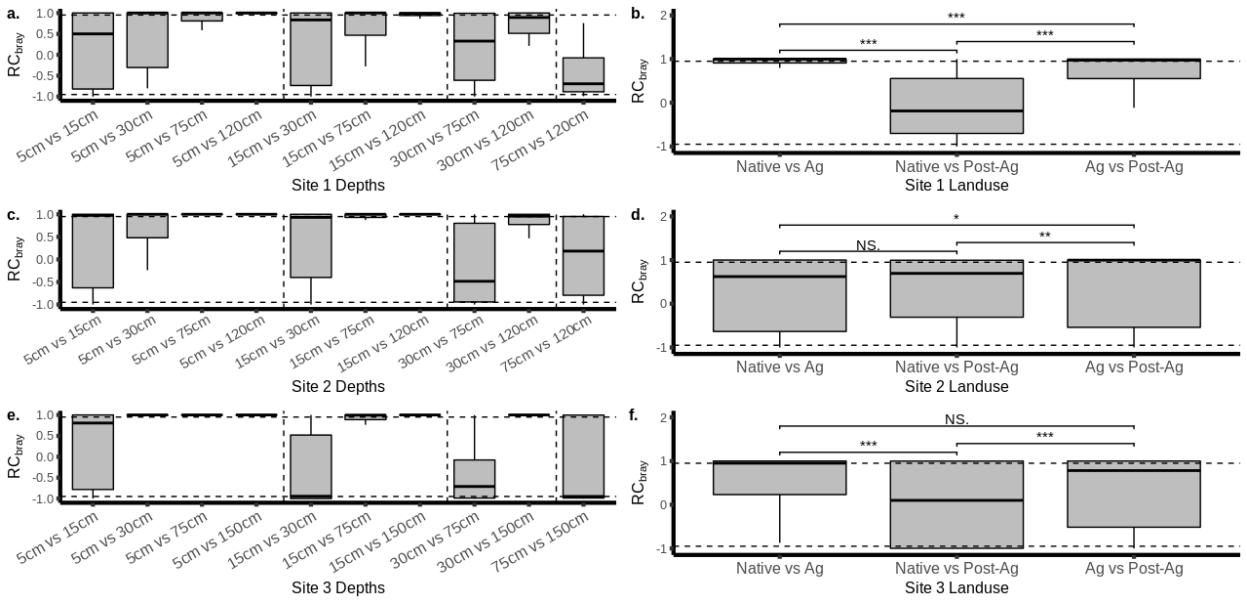

Figure S4.20. $\mathrm{RC}_{\mathrm{bray}}$ between bacterial samples (OTUs) obtained from different depths and land use history. All samples are from fall 2018. In figures b., d., and f., Ag represents samples from agricultural land and Post-Ag represents sample from post-agricultural/restored land. Horizontal dashed lines show upper and lower significance threshold values at $\mathrm{RC}_{\mathrm{bray}} = 0.95$ and $\mathrm{RC}_{\mathrm{bray}} = -0.95$, respectively.

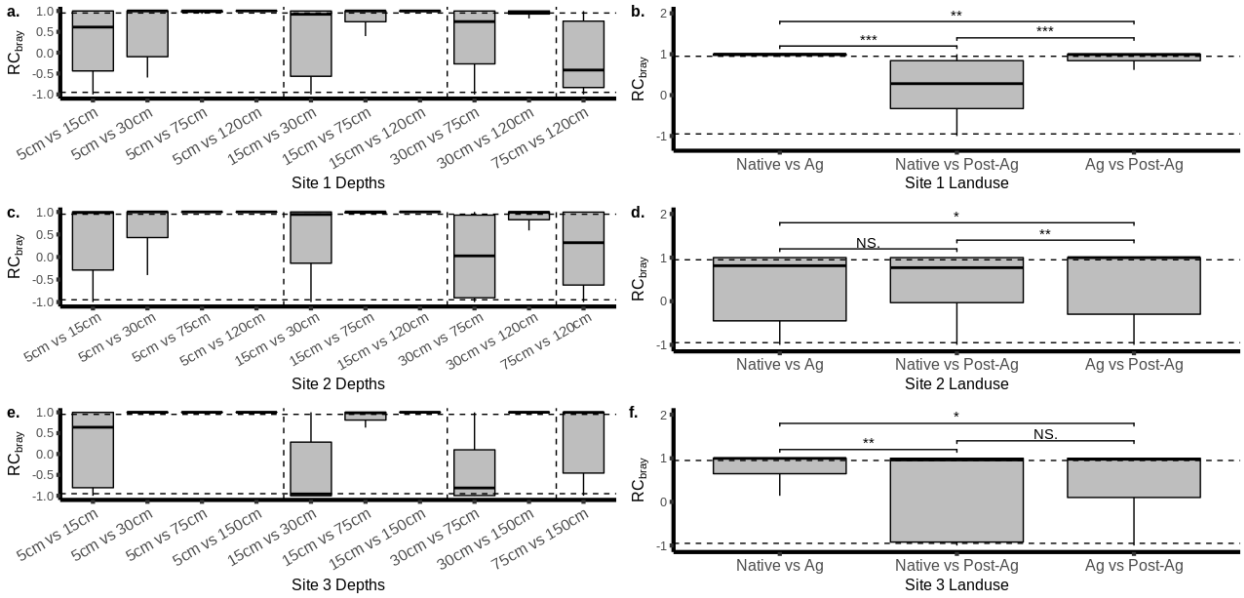

Figure S4.21. $\mathrm{RC}_{\mathrm{bray}}$ between bacterial samples (ASVs) obtained from different depths and land use history. All samples are from spring 2018. In figures b., d., and f., Ag represents samples from agricultural land and Post-Ag represents sample from post-agricultural/restored land. Horizontal dashed lines show upper and lower significance threshold values at $\mathrm{RC}_{\mathrm{bray}} = 0.95$ and $\mathrm{RC}_{\mathrm{bray}} = -0.95$, respectively.

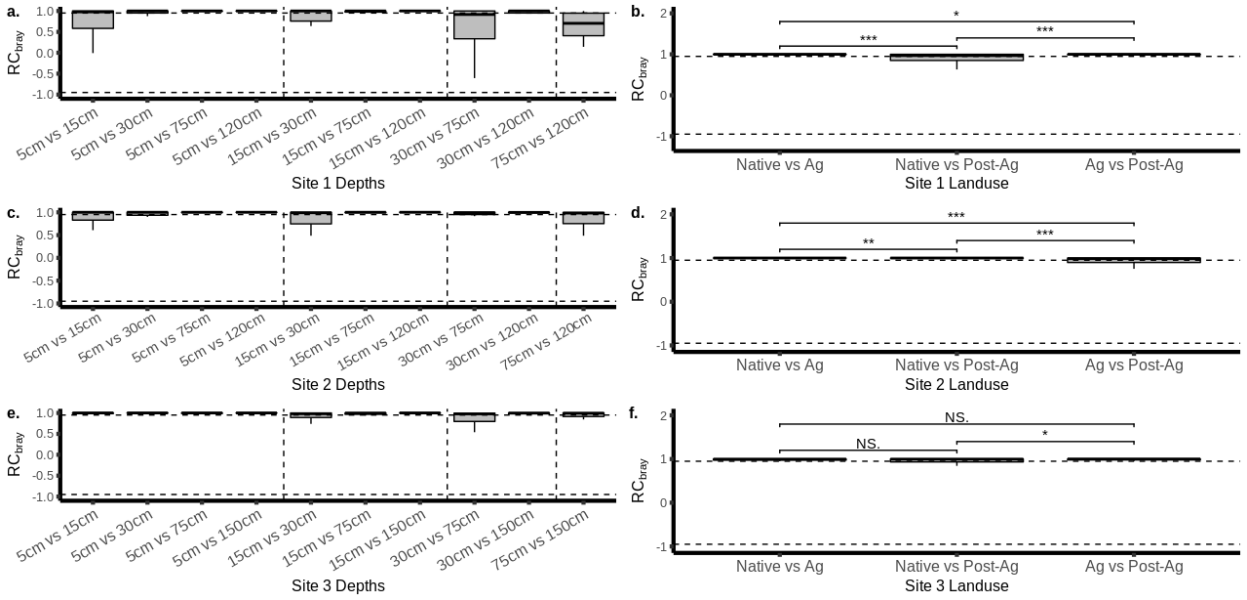

Figure S4.22. $\mathrm{RC}_{\mathrm{bray}}$ between fungal samples (ASVs) obtained from different depths and land use history. All samples are from fall 2018. In figures b., d., and f., Ag represents samples from agricultural land and Post-Ag represents sample from post-agricultural/restored land. Horizontal dashed lines show upper and lower significance threshold values at $\mathrm{RC}_{\mathrm{bray}} = 0.95$ and $\mathrm{RC}_{\mathrm{bray}} = -0.95$, respectively.

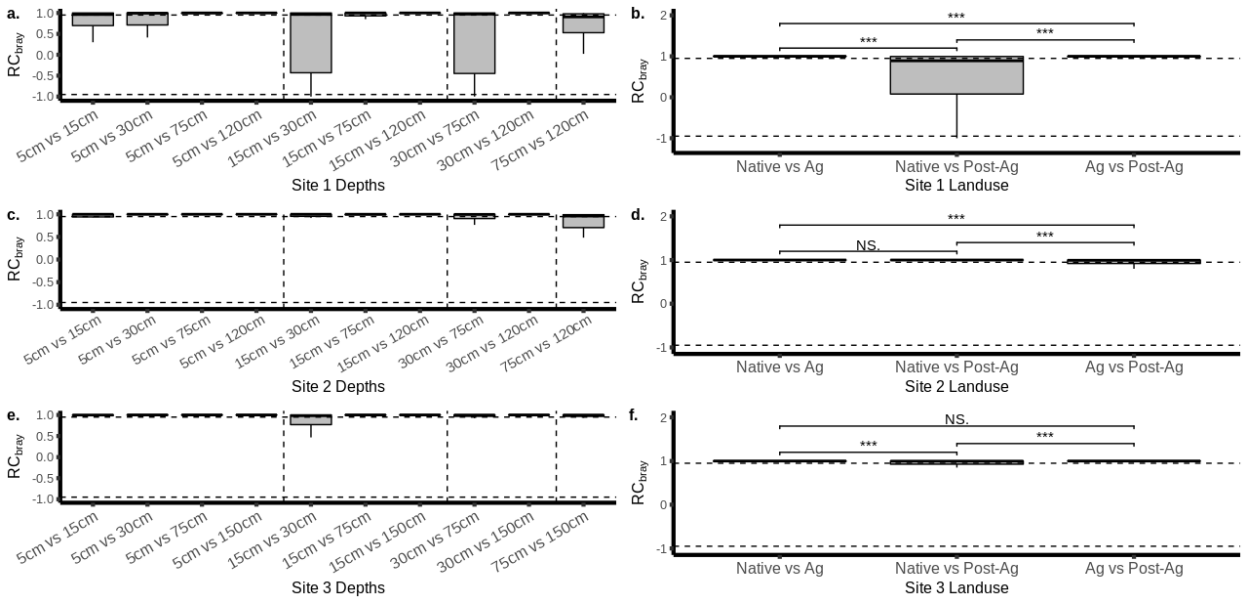

Figure S4.23. $\mathrm{RC}_{\mathrm{bray}}$ between fungal samples (OTUs) obtained from different depths and land use history. All samples are from fall 2018. In figures b., d., and f., Ag represents samples from agricultural land and Post-Ag represents sample from post-agricultural/restored land. Horizontal dashed lines show upper and lower significance threshold values at $\mathrm{RC}_{\mathrm{bray}} = 0.95$ and $\mathrm{RC}_{\mathrm{bray}} = -0.95$, respectively.

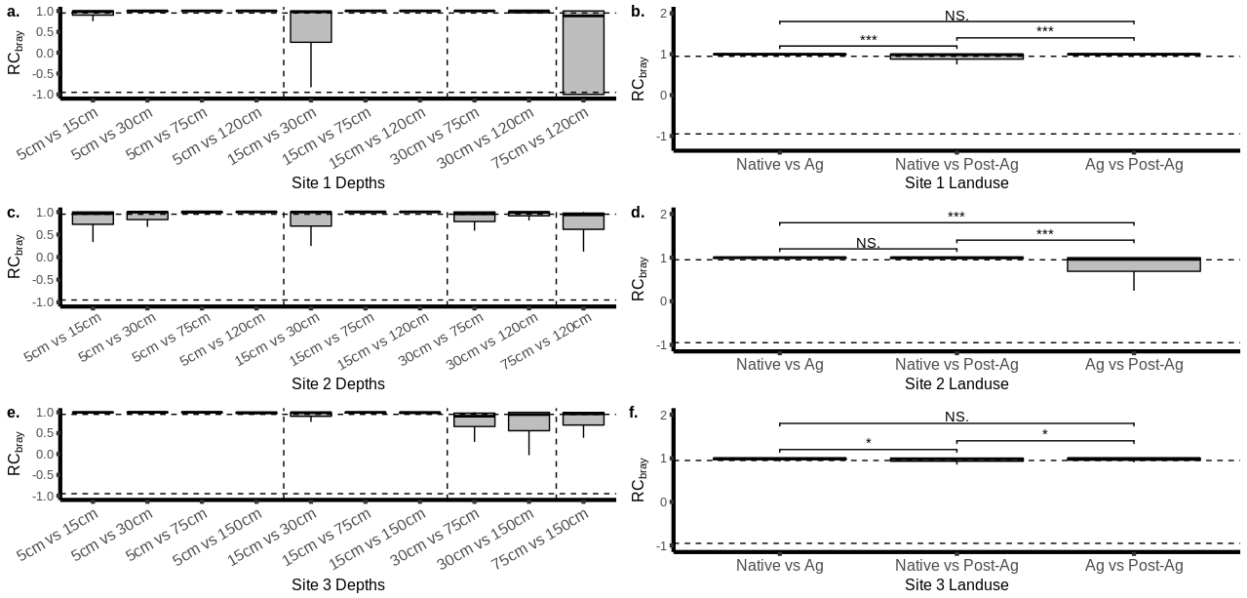

Figure S4.24. $\mathrm{RC}_{\mathrm{bray}}$ between fungal samples (ASVs) obtained from different depths and land use history. All samples are from spring 2018. In figures b., d., and f., Ag represents samples from agricultural land and Post-Ag represents sample from post-agricultural/restored land. Horizontal dashed lines show upper and lower significance threshold values at $\mathrm{RC}_{\mathrm{bray}} = 0.95$ and $\mathrm{RC}_{\mathrm{bray}} = -0.95$, respectively.

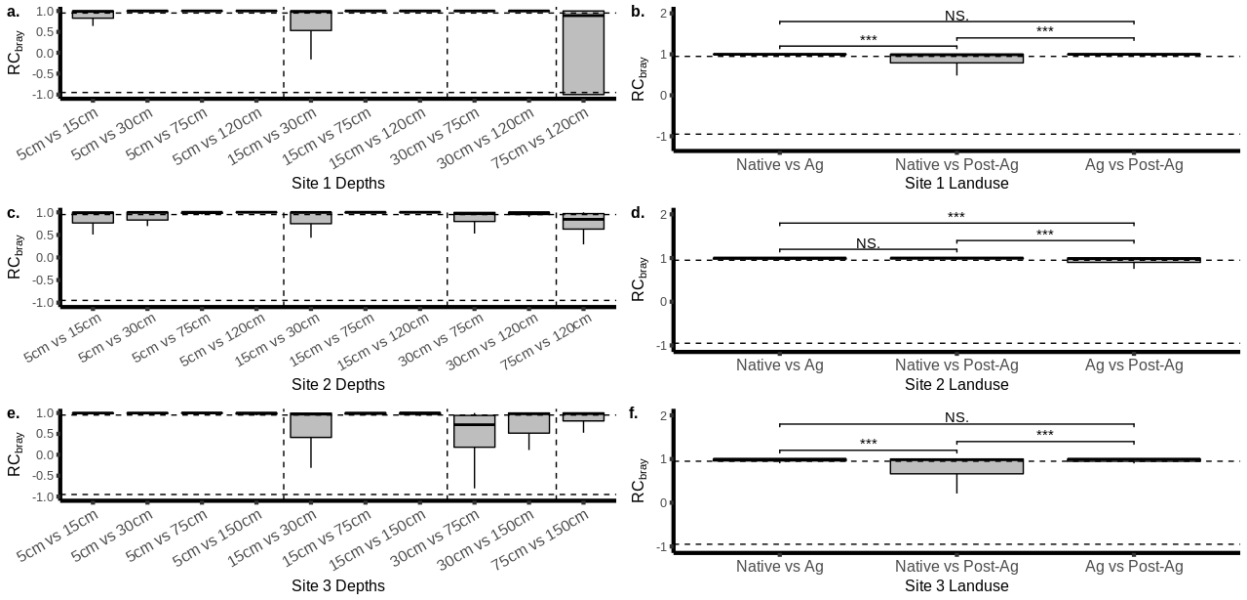

Figure S4.25. $\mathrm{RC}_{\mathrm{bray}}$ between fungal samples (OTUs) obtained from different depths and land use history. All samples are from spring 2018. In figures b., d., and f., Ag represents samples from agricultural land and Post-Ag represents sample from post-agricultural/restored land. Horizontal dashed lines show upper and lower significance threshold values at $\mathrm{RC}_{\mathrm{bray}} = 0.95$ and $\mathrm{RC}_{\mathrm{bray}} = -0.95$, respectively.

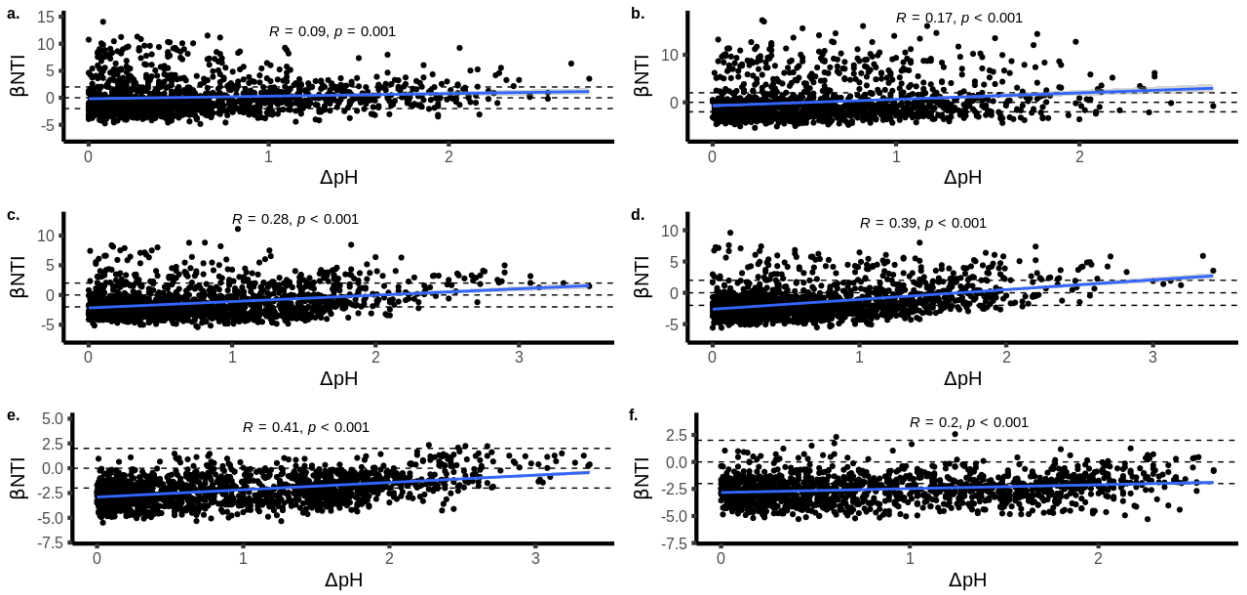

Figure S4.26. Correlation between pH differences and βNTI of paired samples based on bacterial OTUs. Left panel represents samples obtained from spring 2018 and the right panel represents the samples obtained from fall 2018.

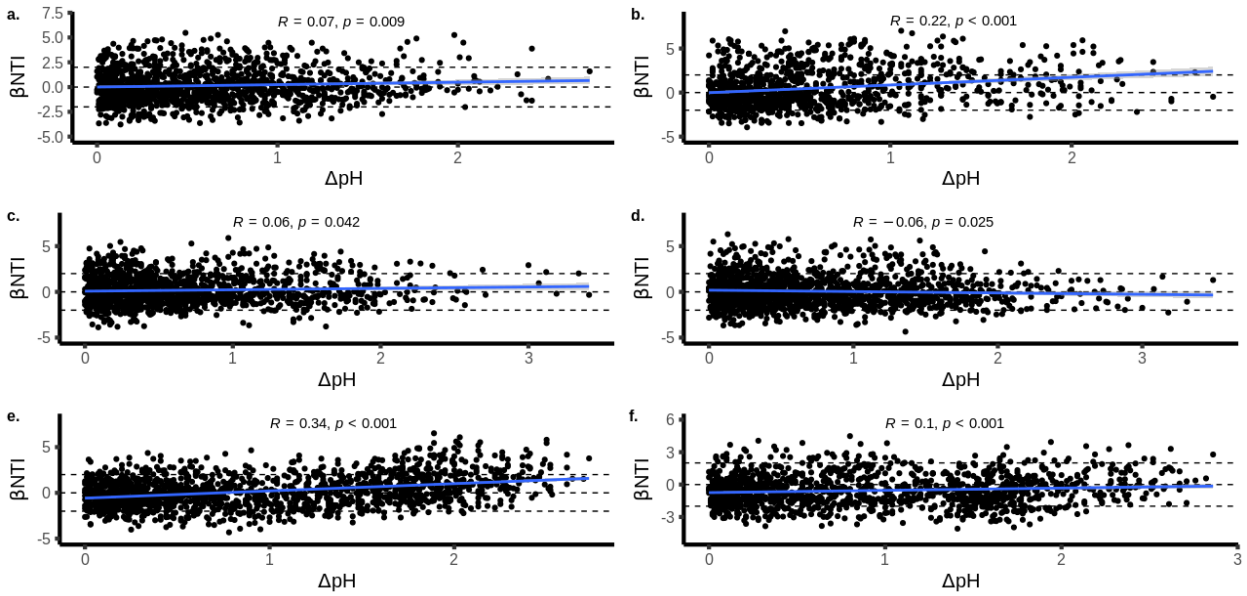

Figure S4.27. Correlation between pH differences and βNTI of paired samples based on fungal ASVs. Left panel represents samples obtained from spring 2018 and the right panel represents the samples obtained from fall 2018.

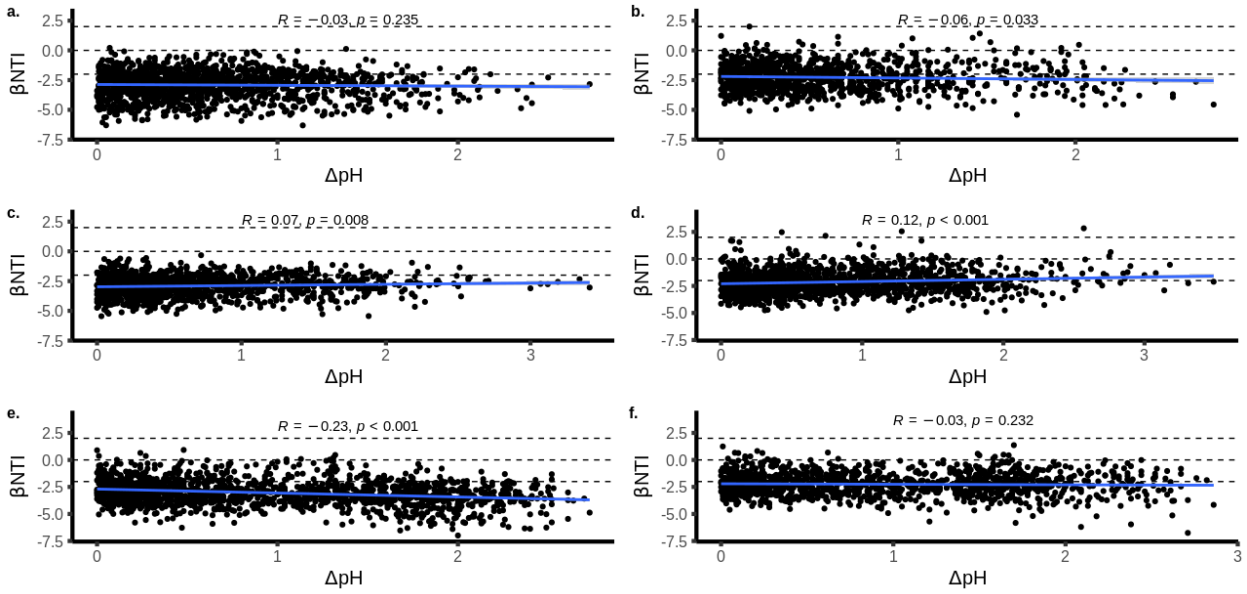

Figure S4.28. Correlation between pH differences and βNTI of paired samples based on fungal OTUs. Left panel represents samples obtained from spring 2018 and the right panel represents the samples obtained from fall 2018.

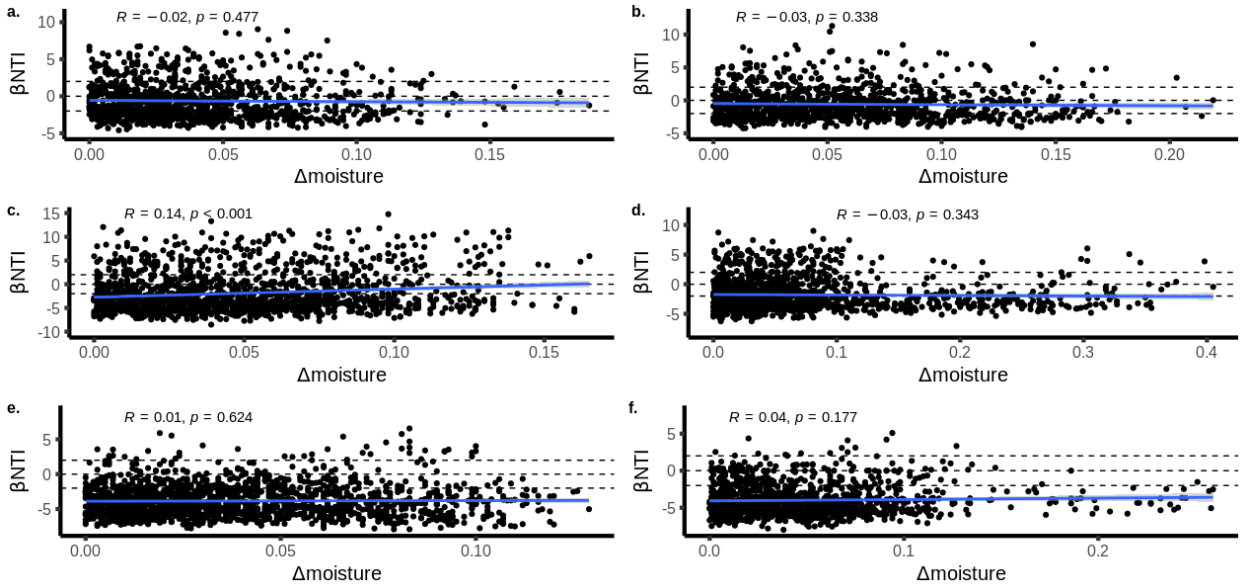

Figure S4.29. Correlation between moisture content differences and βNTI of paired samples based on bacterial ASVs. Left panel represents samples obtained from spring 2018 and the right panel represents the samples obtained from fall 2018.

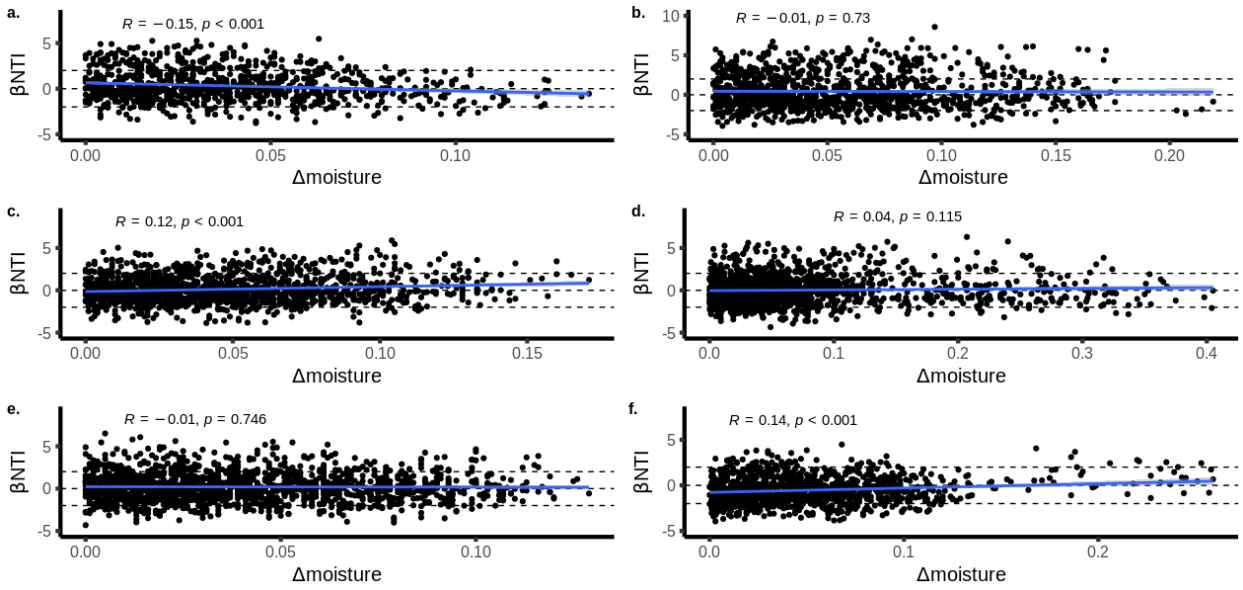

Figure S4.30. Correlation between moisture content differences and βNTI of paired samples based on fungal ASVs. Left panel represents samples obtained from spring 2018 and the right panel represents the samples obtained from fall 2018.

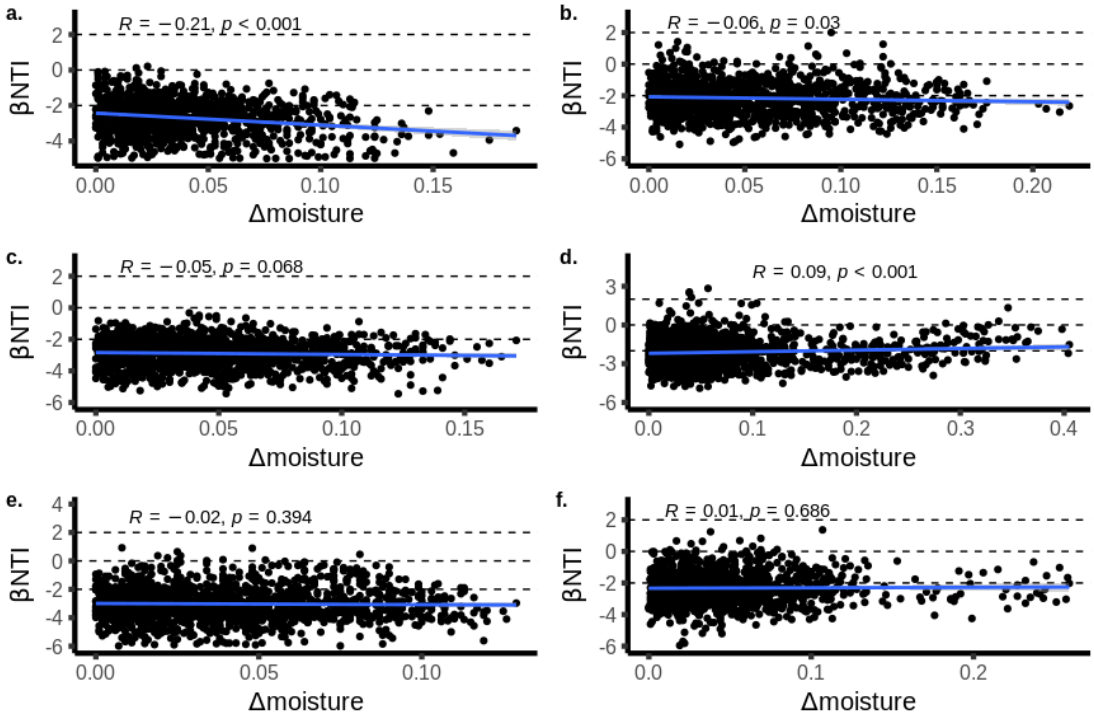

Figure S4.31. Correlation between moisture content differences and βNTI of paired samples based on fungal OTUs. Left panel represents samples obtained from spring 2018 and the right panel represents the samples obtained from fall 2018.

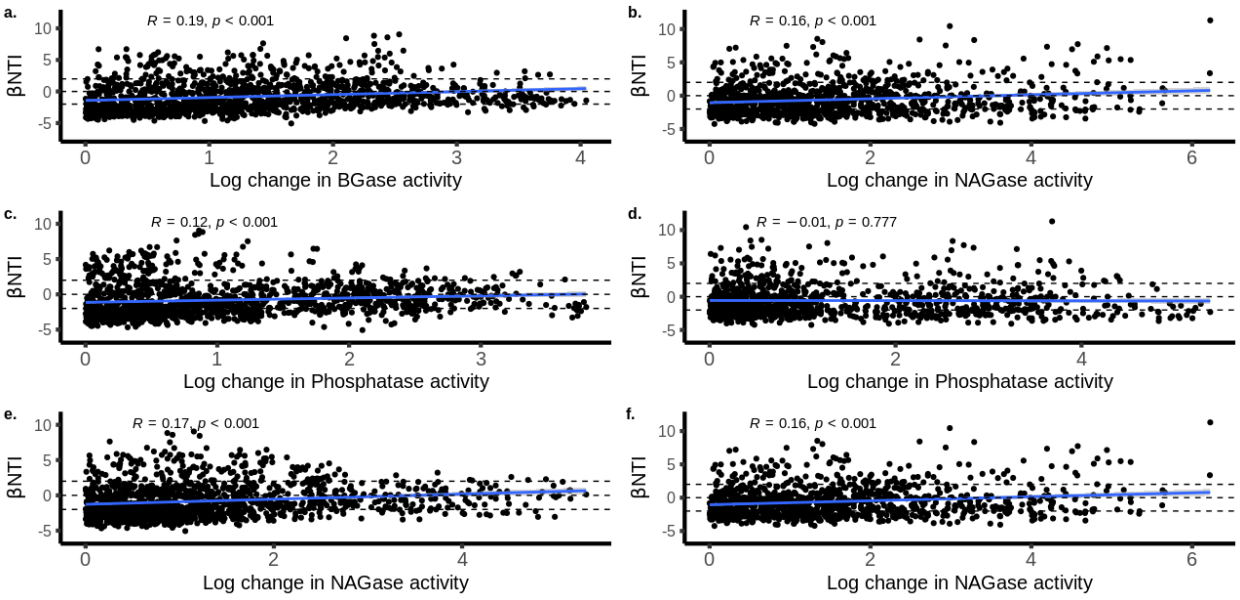

Figure S4.32. Correlation between fold change in enzymatic activity and βNTI of paired samples based on bacterial ASVs at site 1. Left panel represents samples obtained from spring 2018 and the right panel represents the samples obtained from fall 2018.

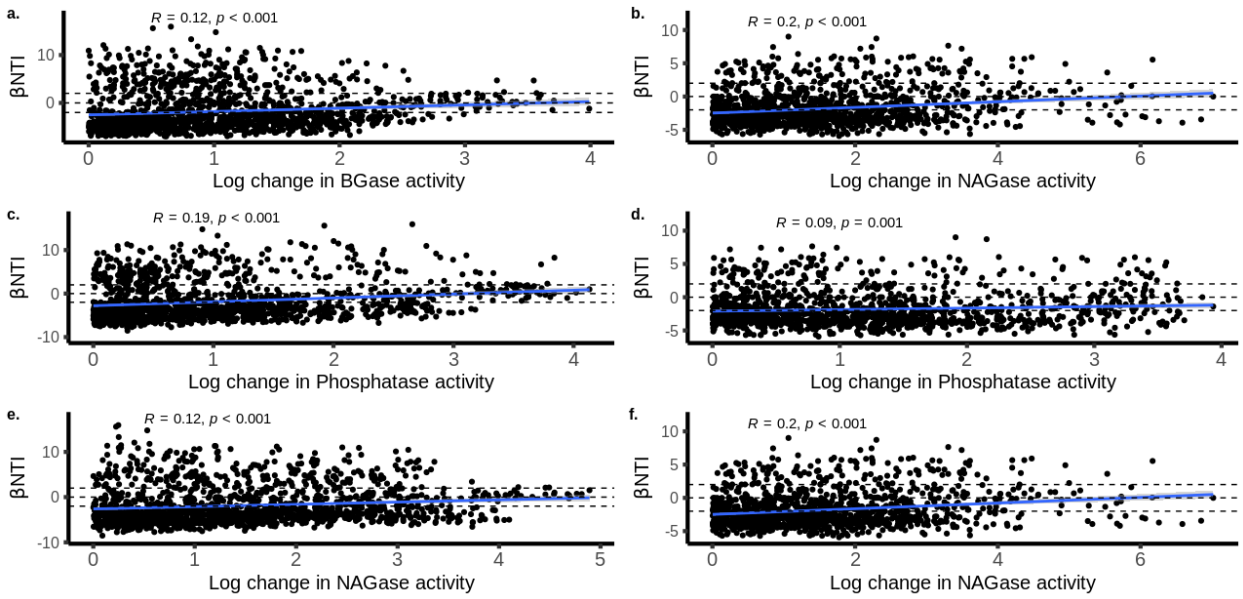

Figure S4.33. Correlation between fold change in enzymatic activity and βNTI of paired samples based on bacterial ASVs at site 2. Left panel represents samples obtained from spring 2018 and the right panel represents the samples obtained from fall 2018.

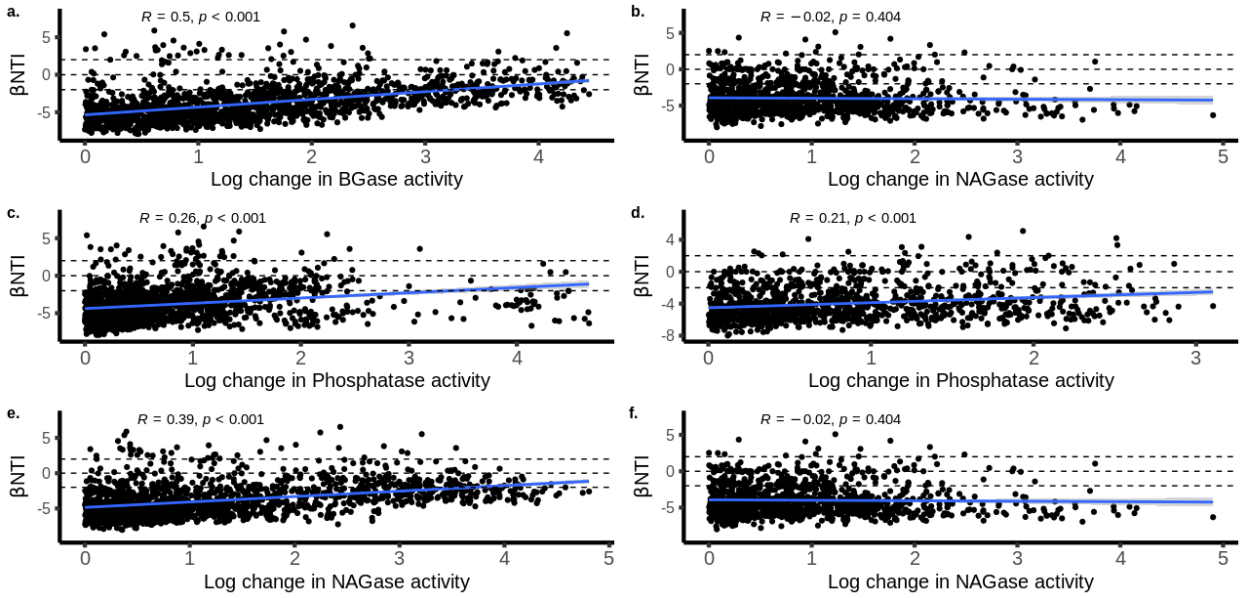

Figure S4.34. Correlation between fold change in enzymatic activity and βNTI of paired samples based on bacterial ASVs at site 3. Left panel represents samples obtained from spring 2018 and the right panel represents the samples obtained from fall 2018.

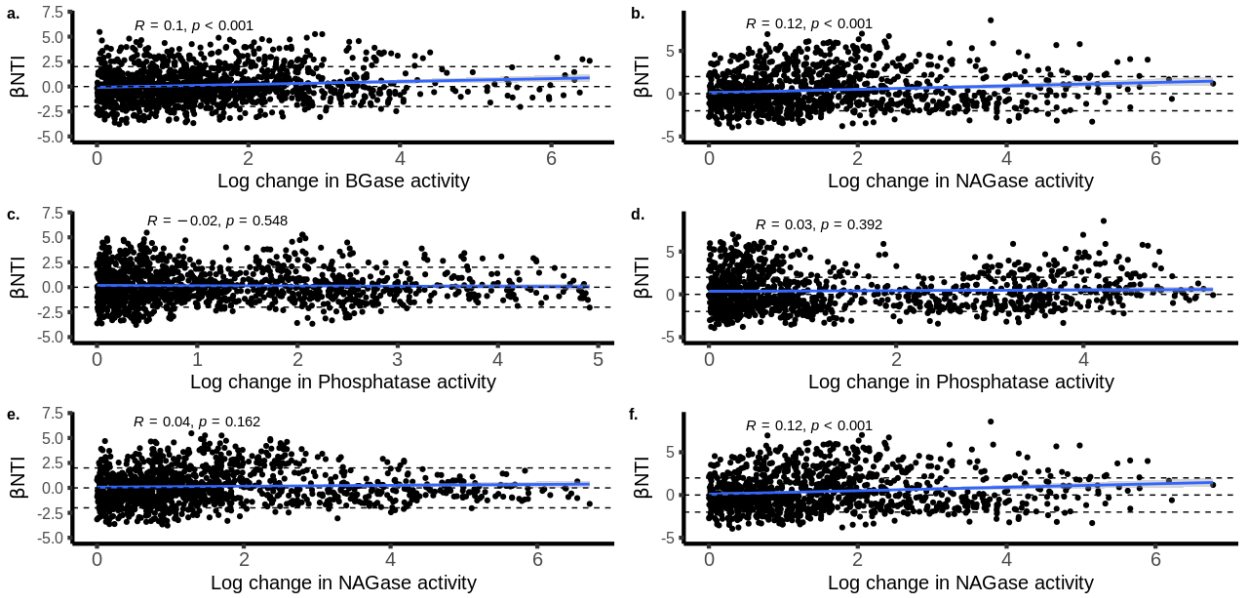

Figure S4.35. Correlation between fold change in enzymatic activity and βNTI of paired samples based on fungal ASVs at site 1. Left panel represents samples obtained from spring 2018 and the right panel represents the samples obtained from fall 2018.

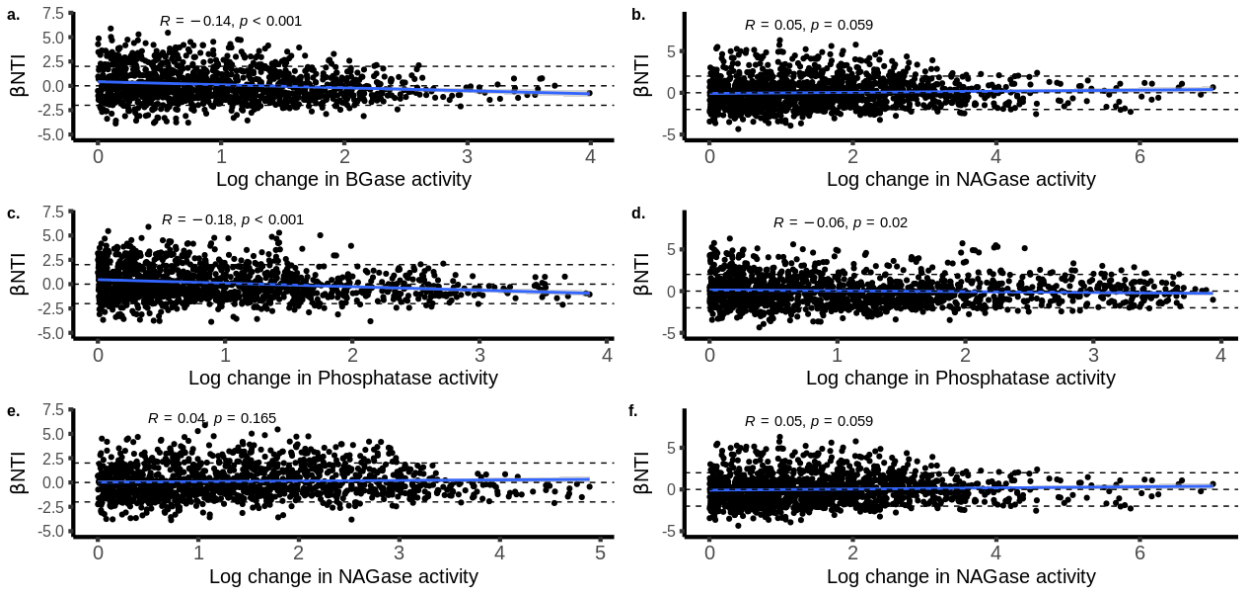

Figure S4.36. Correlation between fold change in enzymatic activity and βNTI of paired samples based on fungal ASVs at site 2. Left panel represents samples obtained from spring 2018 and the right panel represents the samples obtained from fall 2018.

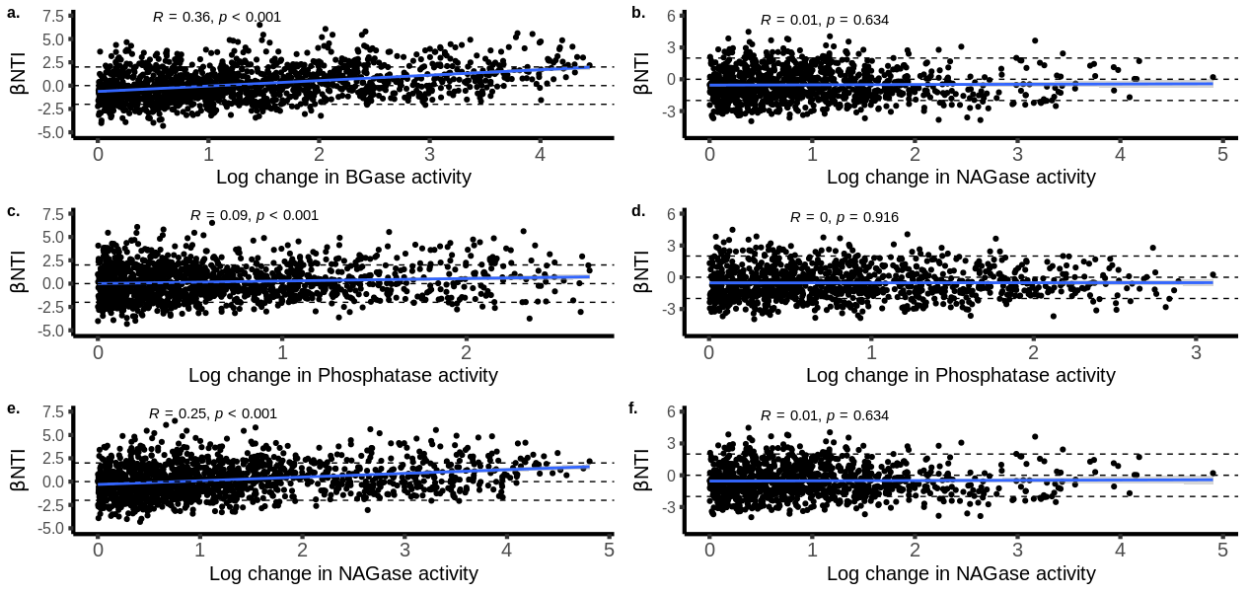

Figure S4.37. Correlation between fold change in enzymatic activity and βNTI of paired samples based on fungal ASVs at site 3. Left panel represents samples obtained from spring 2018 and the right panel represents the samples obtained from fall 2018.

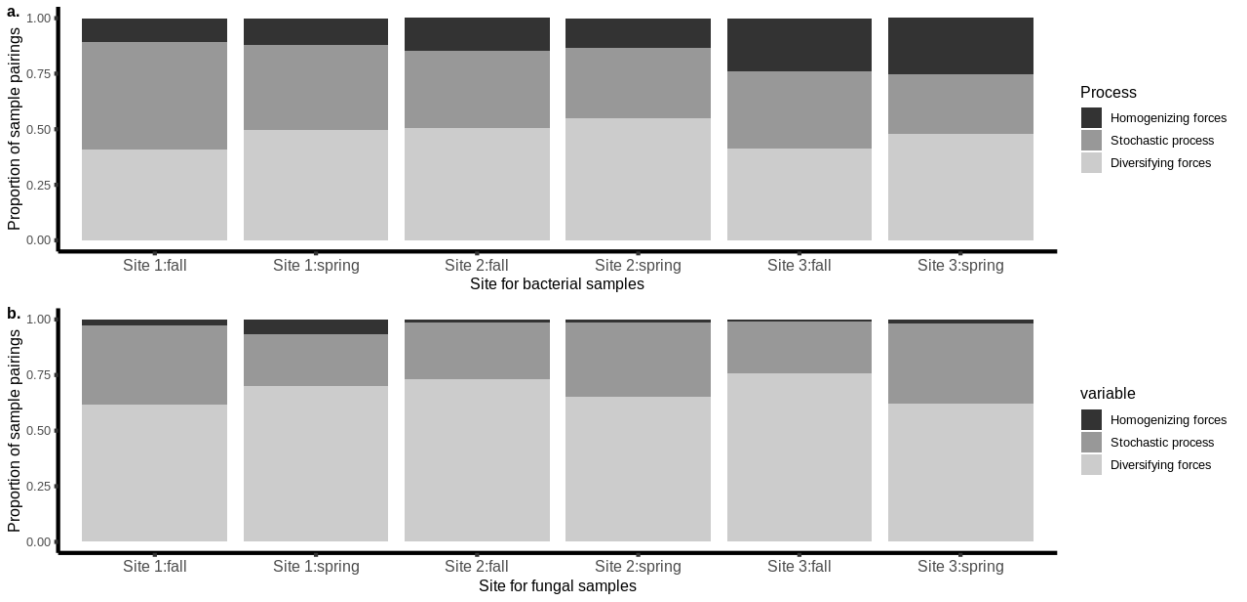

Figure S4.38. Proportions of differences between sample pairings explained by different mechanisms based on RC_{bray} values for both bacterial and fungal samples. Computation was done based on OTU table.