

Supporting Information

Appendix S1: Statistical models

The statistical models used to test the Hypotheses 2-4 are described below. We fit all models using a Bayesian framework and the probabilistic programming language Stan (Carpenter *et al.* 2017). For each model, we fit 3 chains with 3000 iterations and 1500 warm-up iterations. We assessed model convergence visually with traceplots and by ensuring that $\hat{R} < 1.1$ for all parameters in the fitted model (Gelman *et al.* 2014). The models and the code used to run the models can be found at github.com/mqwilber/neutral_microbiome.

Hypothesis 2: Dispersal and drift vary across local communities

A sampled frog j is equivalent to a local community and equation 2 in the main text can be updated to allow I to vary across frogs depending on frog-specific covariates such as Bd load. Specifically,

$$\begin{aligned}\boldsymbol{\pi}_j &\sim \text{Dirichlet}(I_j p_1, \dots, I_j p_S) \\ \log(I_j) &= \mathbf{x}'_j \boldsymbol{\beta} \\ \boldsymbol{\beta} &\sim \text{MVN}(0, \mathbf{I} \times 9)\end{aligned}\tag{S1}$$

\mathbf{x}_j is a row vector of covariates, $\boldsymbol{\beta}$ is a column vector of effect sizes for each covariate, and \mathbf{I} is the identity matrix. The multiplication by nine gives the standard deviation for each parameter in $\boldsymbol{\beta}$, indicating a weak regularizing prior about zero for these parameters.

Hypothesis 3: Dispersal and drift vary by bacterial species

We allowed I to vary with both frog-level covariates and between all $i = 1, \dots, S - 1$ bacterial species. We only considered $S - 1$ species as the information contained in the S th species was redundant based on the sum-to-zero constraint of relative abundances.

The model we fit was

$$\begin{aligned}\pi_{i,j} &\sim \text{Beta}(I_{i,j} p_i, I_{i,j} (1 - p_i)) \\ \log(I_{i,j}) &= \mathbf{x}'_j \boldsymbol{\beta} + \mathbf{z}'_j \boldsymbol{\omega}_i \\ \boldsymbol{\beta} &\sim \text{MVN}(0, \mathbf{I} \times 9) \\ \boldsymbol{\omega}_{i,g} &\sim \text{Normal}(0, \sigma_g^2) \\ \sigma_g &\sim \text{Half-Normal}(0, 2)\end{aligned}\tag{S2}$$

\mathbf{x}_j is a column vector of covariates for fixed effects on recruitment, \mathbf{z}_j is a column vector of covariates from the j th sample that have species-level random effects, and $\boldsymbol{\omega}_i$ is a column vector of G random effects for species i . We modeled each random effect as an independent normal distribution with variance σ_g^2 . The above model tested whether species experienced different fundamental recruitment rates and whether Bd affected these fundamental recruitment rates differently.

Hypothesis 4: Selection processes drive community assembly

Testing for selection effects corresponded to allowing metacommunity relative abundance p_i to vary with local community characteristics. The full model we fit was

$$\begin{aligned}
\pi_{i,j} &\sim \text{Beta}(I_{i,j}p_i, I_{i,j}(1-p_i)) \\
\log(I_{i,j}) &= \mathbf{x}_j'\boldsymbol{\beta} + \mathbf{z}_j'\boldsymbol{\omega}_i \\
\log\left(\frac{p_{i,j}}{1-p_{i,j}}\right) &= \mathbf{c}_j'\boldsymbol{\Phi}_i \\
\boldsymbol{\beta} &\sim \text{MVN}(0, \mathbf{I} \times 9) \\
\boldsymbol{\omega}_{i,g} &\sim \text{Normal}(0, \sigma_g^2) \\
\sigma_g &\sim \text{Half-Normal}(0, 2) \\
\boldsymbol{\Phi}_i &\sim \text{MVN}(0, \mathbf{I})
\end{aligned} \tag{S3}$$

The model is similar to equation S2, but OTU-specific selection effects are captured in how “apparent” metacommunity abundance changes with frog-level covariates \mathbf{c}_j , such as Bd load. $\boldsymbol{\Phi}_i$ is a column vector of the effects of covariates \mathbf{c}_j on “apparent” metacommunity relative abundance of species i . We put a stronger regularizing prior on $\boldsymbol{\Phi}_i$ as decreasing OTU abundance decreased our ability to infer selection processes from the data itself. Therefore, we erred toward inferring no effect of selection processes unless the data strongly suggested otherwise.

Sampling Model

All we can observe from a local community is a sample of relative abundance and not true relative abundance $\boldsymbol{\pi}$. This would require sampling all N_T individuals that comprise the local community. Assuming that the number of sampled bacteria N_t is $N_t \ll N_T$, we can approximate the sampling model as (Harris *et al.* 2017)

$$\begin{aligned}
y_{i,j} &\sim \text{Binomial}(\pi_{i,j}, N_{t,j}) \\
\pi_{i,j} &\sim \text{Beta}(I_{i,j}p_i, I_{i,j}(1-p_i))
\end{aligned} \tag{S4}$$

36 where $y_{i,j}$ are the observed abundance for OTU i on frog j .

37 **Appendix S2: Marginal distribution of species relative abundance** 38 **when I varies among species**

39 Sloan *et al.* (2006) used a diffusion approximation to compute the marginal relative abundance distribution
40 of a species at equilibrium under the neutral model given in equation 1 in the main text. When I is assumed
41 constant across species, the marginal distribution of relative abundance of species in i in local community j
42 is given by

$$\pi_{i,j} \sim \text{Beta}(I_j p_i, I_j (1 - p_i)) \quad (\text{S5})$$

43 When I varies among species, equation S5 no longer holds. In the main text, we tested whether equation
44 S5 is sufficient to describe the observed bacterial abundances or if there is evidence that dispersal and drift
45 vary among bacterial species. We tested for differences in dispersal and drift among species by fitting the
46 model

$$\pi_{i,j} \sim \text{Beta}(I_{i,j} p_i, I_{i,j} (1 - p_i)) \quad (\text{S6})$$

47 While equation S6 is not the exact marginal distribution for the relative abundance of species i predicted
48 by the neutral model with varying I_i , it allows us to statistically test whether equation S5 fails to capture
49 observed patterns of bacteria abundance. The reason this distinction is important is because $I_{i,j}$ cannot
50 necessarily be interpreted as the fundamental recruitment number for species i in local community j as
51 equation S6 does not directly link back to a neutral model with varying I_i .

52 To illustrate this point, consider the following neutral model with species-specific dispersal

$$\begin{aligned} P(N_i + 1 | N_i) &= (1 - x_i) \left(\frac{m_i p_i + (1 - m_i) x_i}{1 + m_c - \sum_{j=1}^S x_j m_j} \right) \\ P(N_i - 1 | N_i) &= x_i \left(1 - \frac{m_i p_i + (1 - m_i) x_i}{1 + m_c - \sum_{j=1}^S x_j m_j} \right) \\ P(N_i | N_i) &= 1 - P(N_i + 1 | N_i) - P(N_i - 1 | N_i) \end{aligned} \quad (\text{S7})$$

53 where x_i is the relative abundance of species i , N_i/N_T , m_i is the species-specific dispersal, and m_c is the
54 weighted dispersal from the metacommunity $\sum_{j=1}^S m_j p_j$. When $m_i = m$ for all species i , we recover equation
55 1 in the main text with the approximation that $N_T - 1 = N_T$. The numerator $1 + m_c - \sum_{j=1}^S x_j m_j$ normalizes

species-specific probabilities of increasing by one in a time-step across species.

For a given species i , the expected change in relative abundance x_i $E[\Delta x_i|x_i]$ is given by

$$E[\Delta x_i|x_i] = (1/N_T)P(N_i + 1|N_i) - (1/N_T)P(N_i - 1|N_i) = \frac{m_i(p_i - x_i) + x_i(\sum_{j=1}^S x_j m_j - m_c)}{(1 + m_c - \sum_{j=1}^S x_j m_j)N_T} \quad (\text{S8})$$

The expected change in the mean relative abundance is zero when $x_i = p_i \forall i$, such that $x_i = p_i \forall i$ is an equilibrium. We tested the stability of this equilibrium point more generally (though not exhaustively), to understand whether, deterministically, a microbial community following equation S7 would remain near the equilibrium p_i if perturbed. The goal of this stability analysis was to generally confirm that the expected value would stay near p_i once it was reached. For each value of S between 3 and 50, we performed a local stability analysis on 5000 communities with random configurations of p_i and m_i . The equilibrium $x_i = p_i \forall i$ was stable for every community tested.

We simulated equation S7 with $S = 6$ species in the community, total community size fixed at $N_T = 1000$, and species-specific dispersal rates. We then numerically computed the equilibrium distributions and compared these distributions to equation S6 (Figure S5). While equation S6 was a good approximation for five species, equation S6 underestimated the variance of the most abundant species. This was because the low variances (high I_s) of species 2 and 3 restricted the variance of species 1 as there was never enough empty space for a species to obtain the high level of abundances as predicted by equation S6. The accuracy of equation S6 depends on the interaction between species-specific I_i . Therefore, while a fitted species-specific I_i may directly relate to the fundamental recruitment number for some species in a community, for other species I_i can only be interpreted as a measure of dispersion in abundance.

Appendix S3: Sensitivity of analyses to the number of OTUs used

To test how sensitive our analyses were to our choice of 19 OTUs, we fitted two additional independent species neutral models using 9 and 29 OTUs. We did this for both the field data and the laboratory data. For the field data, the three additional models that we fit are given in Table S2 and Table S3. Of the three models, the best predictive model in terms of LOO-IC was the model with a quadratic Bd selection effect and OTU-specific Bd effects on dispersal and drift. Removing Bd as a dispersal and drift effect and removing Bd as a selection effect lead to worse models in terms of LOO-IC for both 9 and 29 OTUs (Table S2, S3). This was consistent with our analysis for 19 OTUs presented in the main text. Moreover, the main effects of body size on dispersal and drift were significantly negative for 9 and 29 OTUs (95% credible interval for

body size effect for 9 and 29 OTUs, respectively: $[-0.24, -0.01]$, $[-0.22, -0.06]$) and the main effects of Bd on dispersal and drift were null (95% credible interval for Bd effect for 9 and 29 OTUs, respectively: $[-0.51, 0.2]$, $[-0.21, 0.21]$), consistent with the analysis using 19 OTUs presented in the main text.

For the laboratory data, we also found that the model with a quadratic effect of Bd on selection and OTU-specific effects of Bd on dispersal and drift was the best predictive model when we used 9 and 29 OTUs (Table S4, S5), compared to models without Bd effects on dispersal, drift, and selection. The main effects of frog body size and Bd on dispersal and drift were generally consistent when using 9 or 29 OTUs, compared to the 19 OTUs used in the main text. The body size effects on dispersal and drift were not significant for 9, 19 or 29 OTUs, though they all were trending positively (95% credible intervals for body size effects for 9, 19, and 29 OTUs, respectively: $[-0.02, 0.1]$, $[-0.01, 0.08]$, and $[-0.01, 0.08]$). The Bd effects on dispersal and drift were significantly positive for 9 OTUs and non-significant but trending positive for 19 and 29 OTUs (95% credible intervals for Bd effects on dispersal and drift for 9, 19, and 29 OTUs, respectively: $[0.03, 0.57]$, $[-0.08, 0.42]$, and $[-0.07, 0.38]$).

Table S1: Sampling dates, sample sizes, population-level status of frog populations sampled for Bd and microbial communities.

Lake ID	Dates sampled	Number of frogs	Status
YOSE 4	2010-08-18	18	Enzootic
SEKI 1	2010-08-22	8	Pre-epizootic
	2010-09-09	10	
SEKI 2	2010-08-23	8	Epizootic
YOSE 3	2010-08-27	8	Enzootic
SEKI 3	2010-08-30	20	Epizootic
	2010-09-14	10	
YOSE 1	2010-09-01	8	Enzootic
	2010-09-16	18	
YOSE 2	2010-08-29	7	Enzootic
	2010-09-13	18	

Table S2: Community assembly models for the amphibian microbiome in the field, fit with the 9 most abundant OTUs. The “Full Model” is the same model structure as the best predictive model Model F8 in Table 1 in the main text. “No Bd dispersal, drift” removes the effect of Bd on dispersal and drift and “No Bd dispersal, drift, selection” removes the effect of Bd on dispersal, drift, and selection.

Model name	$\Delta\text{LOO-IC}$	Drift and dispersal	Selection
Full Model	0	Body size, Bd, Lake, OTU-specific Bd	Lake \times Bd ²
No Bd dispersal, drift	8.63	Body size, Lake, OTU-specific I	Lake \times Bd ²
No Bd dispersal, drift, selection	303.96	Body size, Lake, OTU-specific I	Lake

Table S3: Community assembly models for the amphibian microbiome in the field data, fit with the 29 most abundant OTUs. The “Full Model” is the same model structure as the best predictive model Model F8 in Table 1 in the main text. “No Bd dispersal, drift” removes the effect of Bd on dispersal and drift and “No Bd dispersal, drift, selection” removes the effect of Bd on dispersal, drift, and selection.

Model name	$\Delta\text{LOO-IC}$	Drift and dispersal	Selection
Full Model	0	Body size, Bd, Lake, OTU-specific Bd	Lake \times Bd ²
No Bd dispersal, drift	30.34	Body size, Lake, OTU-specific I	Lake \times Bd ²
No Bd dispersal, drift, selection	558.07	Body size, Lake, OTU-specific I	Lake

Table S4: Community assembly models for the amphibian microbiome in the laboratory, fit with the 9 most abundant OTUs. The “Full Model” is the same model structure as the best predictive model Model L7 in Table 2 in the main text. “No Bd dispersal, drift” removes the effect of Bd on dispersal and drift and “No Bd dispersal, drift, selection” removes the effect of Bd on dispersal, drift, and selection.

Model name	$\Delta\text{LOO-IC}$	Drift and dispersal	Selection
Full Model	0	Body size, Bd, OTU-specific Bd	Treatment \times Bd ²
No Bd dispersal, drift	57.93	Body size, OTU-specific I	Treatment \times Bd ²
No Bd dispersal, drift, selection	247.21	Body size, OTU-specific I	Treatment

Table S5: Community assembly models for the amphibian microbiome in the laboratory, fit with the 29 most abundant OTUs. The “Full Model” is the same model structure as the best predictive model Model L7 in Table 2 in the main text. “No Bd dispersal, drift” removes the effect of Bd on dispersal and drift and “No Bd dispersal, drift, selection” removes the effect of Bd on dispersal, drift, and selection.

Model name	$\Delta\text{LOO-IC}$	Drift and dispersal	Selection
Full Model	0	Body size, Bd, OTU-specific Bd	Treatment \times Bd ²
No Bd dispersal, drift	185.44	Body size, OTU-specific I	Treatment \times Bd ²
No Bd dispersal, drift, selection	769.72	Body size, OTU-specific I	Treatment

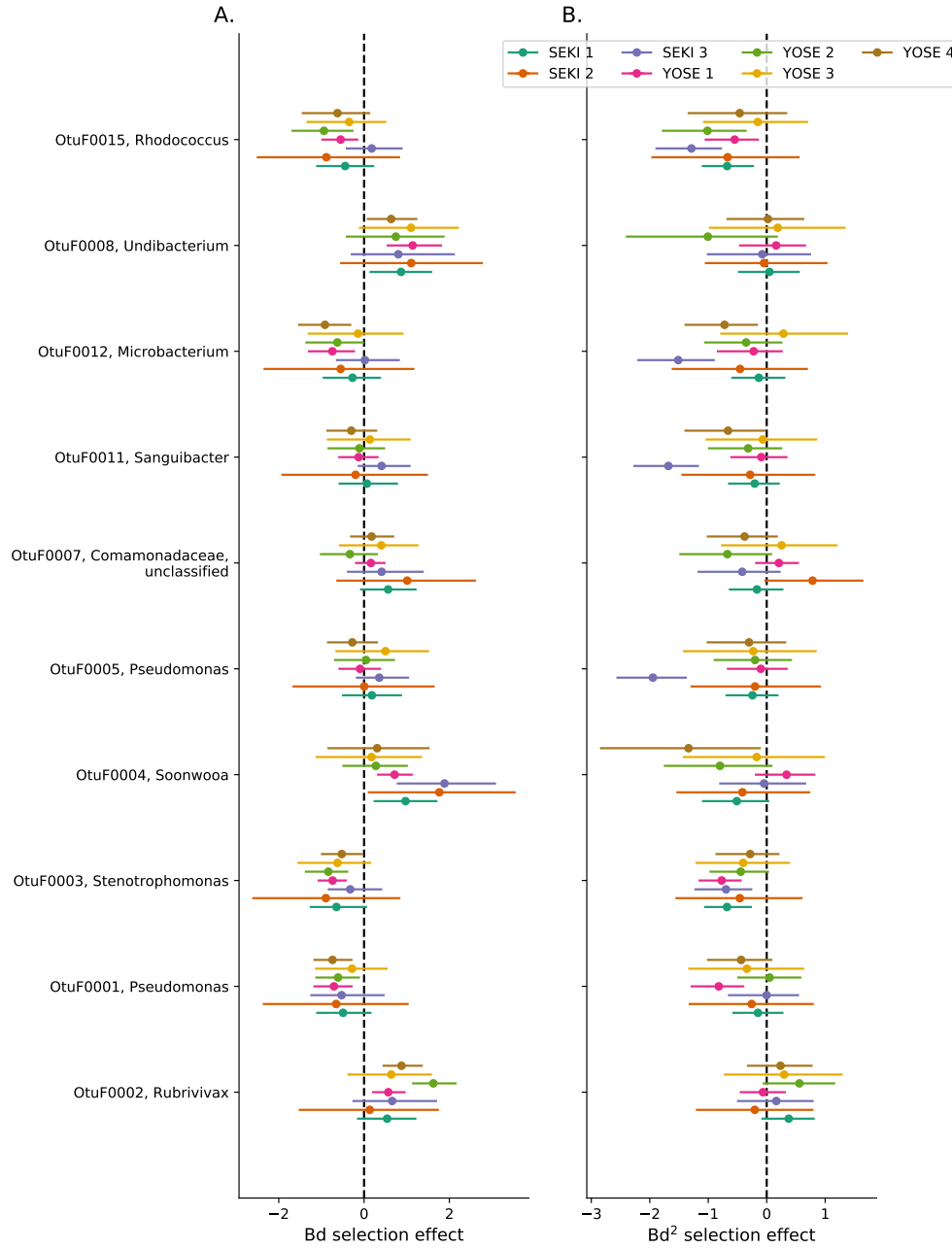


Figure S1: Bd effected the selection processes shaping the amphibian skin microbiome in the field. **A.** Estimated Bd selection effects on the relative abundance for ten OTUs across seven lakes in the field. Each point gives the median predicted Bd effect for a given lake and a given OTU from Model F8 with selection, dispersal, and drift effects. The error bars are 95% credible intervals about the estimate. Plots of OTU relative abundance are given in Fig. S2. **B.** Same as A., but the quadratic selection effect of Bd load on OTU relative abundance. The large negative quadratic effects of Bd on Psuedomonas, Sanguibacter, and Microbacterium for lake SEKI 3 are heavily influence by a single frog with low Bd load (Fig. S2).

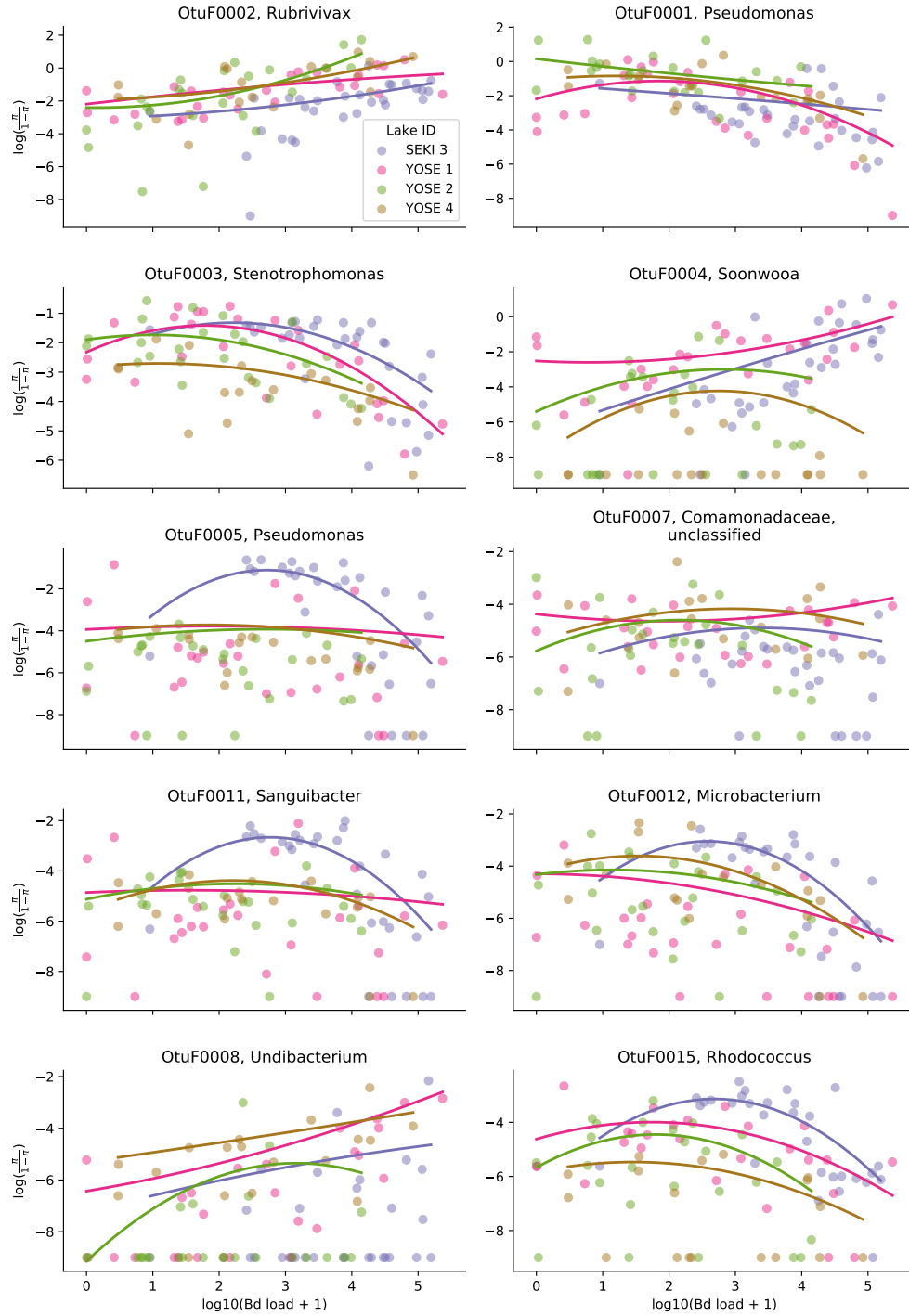


Figure S2: The effect of Bd on OTU relative abundance (π) for ten OTUs in the field data across the four lakes in the field with the most frogs sampled. The colored points represent the observed OTU relative abundance in a sample transformed as $\log(\frac{\pi}{1-\pi})$. Zeros are plotted at value -9. The colored lines give the predicted effect of Bd on the relative abundance of an OTU from Model F8 with selection, dispersal and drift.

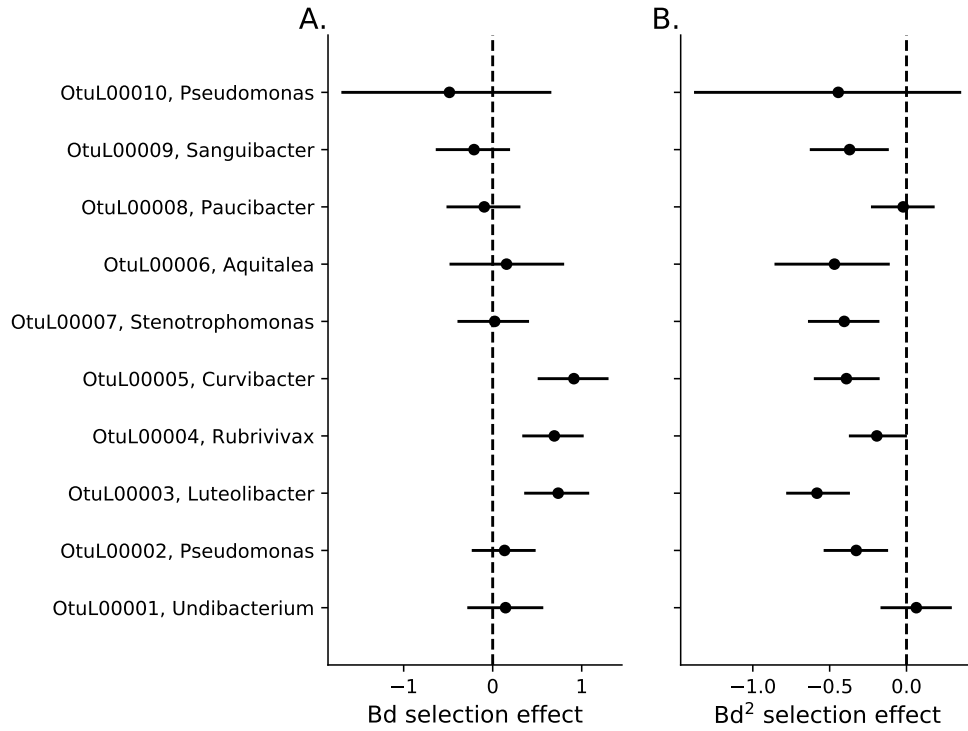


Figure S3: Bd effected the selection processes shaping the amphibian skin microbiome in the laboratory. **A.** Estimated Bd selection effects on the relative abundance for ten OTUs in the laboratory experiment. Each point gives the median predicted Bd effect from Model L7 with selection, dispersal, and drift effects. The errorbars are 95% credible intervals about the estimate. Plots of OTU relative abundance are given in Fig. S3. **B.** Same as A., but the quadratic selection effect of Bd load on OTU relative abundance.

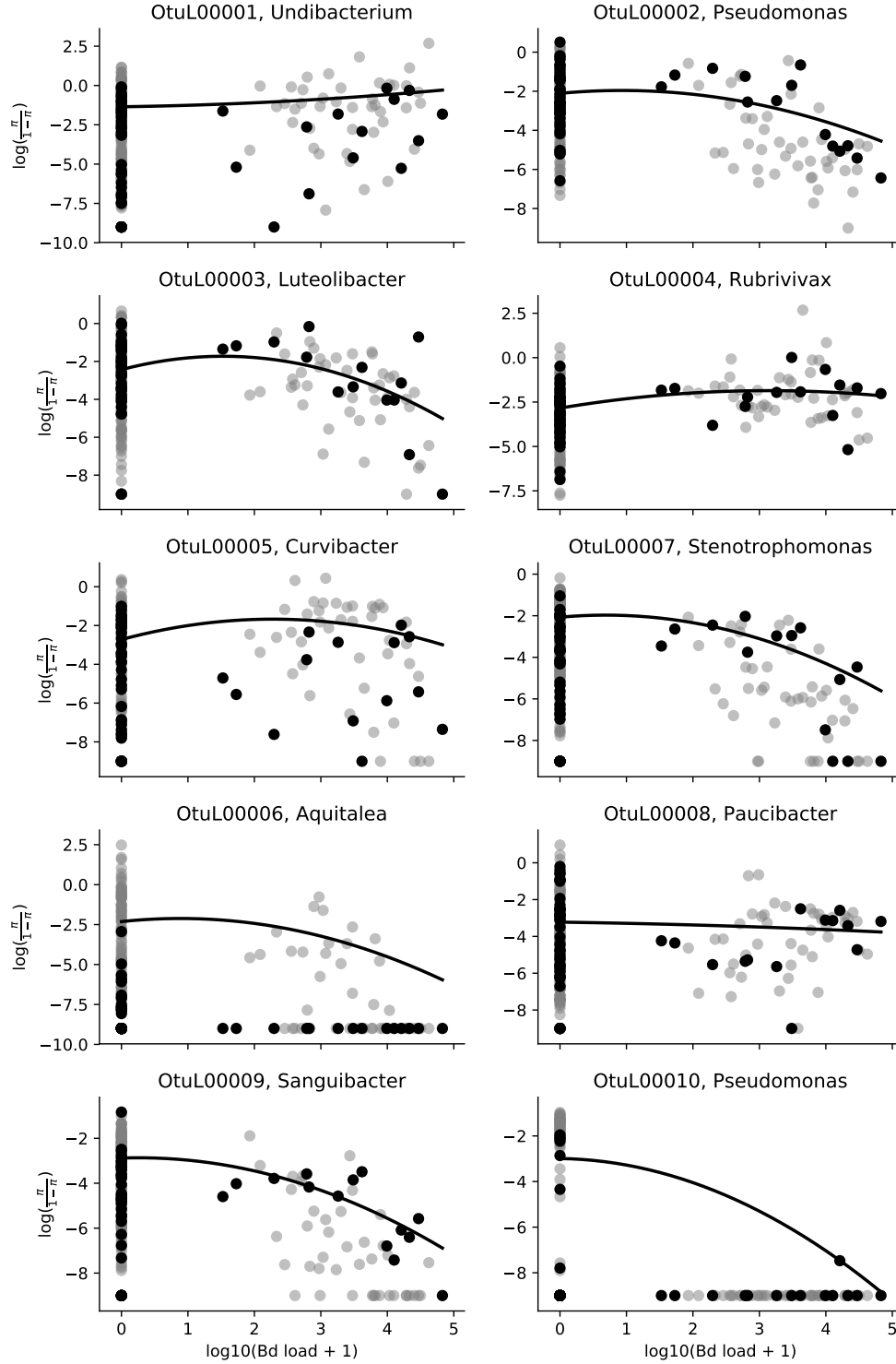


Figure S4: The effect of Bd on OTU relative abundance (π) for 10 OTUs in the laboratory experiment across all 14 treatments (two Frogs sources crossed with seven lake waters). The points represent the observed OTU relative abundance in a sample transformed as $\log(\frac{\pi}{1-\pi})$. Black points show the observed abundance for the treatment with frogs from SEKI 3 and water from YOSE 1. Gray points comprise all other treatments. Zeros are plotted at value -9. The black line gives the predicted effect of Bd on the relative abundance of each OTU for the SEKI 3-YOSE 1 treatment from Model L7 with selection, dispersal and drift effects. Model L7 only included a main effect of treatment on selection, such that all of the lab data contributed to estimating Bd selection effects.

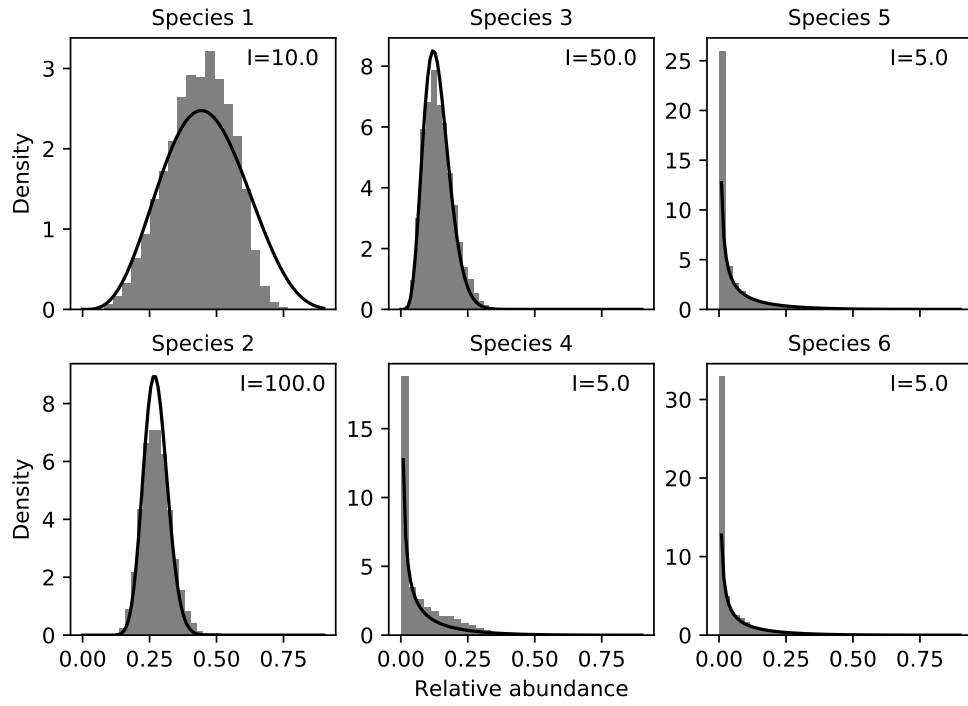


Figure S5: Comparison of the neutral approximation given by equation S6 (black line) and the simulated relative abundance distribution from a neutral model with six species where species have varying fundamental recruitment numbers I (gray histogram).

References

- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P. & Riddell, A. (2017). Stan: A probabilistic programming language. *J. Stat. Soft.*, 76, 1–32.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. & Rubin, D. B. (2014). *Bayesian Data Analysis*. 3rd edn. Taylor & Francis Group, LLC, Boca Raton.
- Harris, K., Parsons, T. L., Ijaz, U. Z., Lahti, L., Holmes, I. & Quince, C. (2017). Linking statistical and ecological theory: Hubbell’s unified neutral theory of biodiversity as a hierarchical Dirichlet process. *P. IEEE*, 105, 516 – 529.
- Sloan, W. T., Lunn, M., Woodcock, S., Head, I. M., Nee, S. & Curtis, T. P. (2006). Quantifying the roles of immigration and chance in shaping prokaryote community structure. *Environ. Microbiol.*, 8, 732–40.