RICE UNIVERSITY

Population Dynamics of Heritable Symbionts when Accounting for the Life History Complexity of their Host

by

Andrew Bibian

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

Master of Arts

APPROVED, THESIS COMMITTEE

Tom E.X. Miller, Chair
Assistant Professor of BioSciences

Amy Dunham
Assistant Professor of BioSciences

Volker Rudolf
Associate Professor of BioSciences

HOUSTON, TEXAS
May 2015
ABSTRACT

Population Dynamics of Heritable Symbionts when Accounting for the Life History Complexity of their Host

by

Andrew Bibian

I developed theory and experiments to understand how complex life cycles of a host affect the dynamics of their heritable symbionts, important ecological and evolutionary agents. With symbiont persistence and prevalence being a function of their effects on host fitness and transmission efficiencies, accounting for host demographic “storage,” in the form of non-reproductive or dormant host life stages, leads to unexpected results. Symbiont loss from demographic storage affects persistence similarly to loss from a host reproductive stage. Loss from host dormancy, however, affect dynamics if symbiont passage through the dormant stage occurs at a high rate, which we observed with experiments. Demographic rescue and symbiont persistence was possible, and observed, whereby stage specific symbiont benefits compensate for symbiont loss. Empirically, accounting for host dormancy in the form of a plant seed bank facilitated symbiont persistence and prevalence. Our results emphasize the importance of accounting for realistic complexity in host-symbiont dynamics.
Acknowledgments

There are many people to thank who have helped me along the way and I would first like to thank my advisor, Tom Miller. Tom, thank you for giving me the chance to join your lab, I don’t think I could have ever asked for a better advisor. Working with you has been my greatest intellectual endeavor and I’ve learned more than I ever imagined I could. Thanks for all your guidance and for being a role model that I will always strive to be like, you’ve made me feel more like a colleague than a student. Also, thanks for all the field days and lab trips, they’ve made my time here at Rice that much more memorable.

With that said, I’d like to thank my dear friend, colleague, and collaborator Brad Ochocki. Many of my ideas would not have come to fruition without our numerous strolls around campus and your infinite programming wisdom, thanks buddy. Thank you to the former and present Miller Lab folk who have all influenced and aided me in one way, shape, or form: Teresa Bohner, Aldo Compagnoni, Marion Donald, Rande Patterson, Emily Schultz, Michelle Sneck, and Michelle Downey.

To my original mentors turned colleagues, Louise Conrad and Andy Sih, I thank you. Thanks for giving me the opportunities to explore the various facets of our research project and supporting my development as a quantitative ecologist. Matt Young, Denise de Carion, and Kelly Weinersmith, thanks for all the good times on the Delta and in the lab.

Thank you to all my family and friends who supported me along the way. Thanks Mom, Dad, Anthony, Lori, Michy, Amy, Elva, Chloe, Kayla, Bob, Grandma Dora,
Grandma Nena, Devin, Casey, Travis, Kelby, Scott, Rose, the Powelson family, and the Schenck family.

To my other collaborators not already mentioned, Jen Rudgers and Alan Shadow, thanks for your logistical and intellectual support. Also, thanks to Volker Rudolf and Amy Dunham for their added wisdom and feedback on my work. Lastly, thank you to my funding sources, the National Science foundation (NSF-GRFP-Bibian and DEB-1145588) and the Goodwin Assistant Professorship (TEXM).
# Contents

**Acknowledgments** .......................................................................................................................... iii

**Contents** ........................................................................................................................................ v

**List of Figures** ..................................................................................................................................... vii

**List of Tables** .................................................................................................................................... ix

**List of Equations** .............................................................................................................................. x

**The role of host demographic storage in the dynamics of vertically transmitted symbionts** ......................... 11

1.1. Model for heritable symbioses in a stage-structured host: methods and results ... 17

1.1.1. Obligate seed dormancy ............................................................................................................. 18

1.1.1.1. Obligate seed dormancy: Persistence ...................................................................................... 20

1.1.1.2. Obligate seed dormancy: Prevalence .................................................................................... 22

1.1.1.3. Obligate seed dormancy summary ......................................................................................... 26

1.1.2. Facultative seed dormancy ........................................................................................................ 27

1.1.2.1. Facultative seed dormancy: Persistence ................................................................................ 28

1.1.2.2. Facultative seed dormancy: Prevalence ............................................................................... 30

1.1.2.3. Facultative seed dormancy summary .................................................................................... 33

1.1.3. Long-term facultative seed dormancy with variable germination ........................................... 34

1.1.3.1. Long-term facultative seed dormancy with variable germination: Persistence ....................... 35

1.1.3.2. Long-term facultative seed dormancy with variable germination: Prevalence ...................... 38

1.1.3.3. Variable germination summary .............................................................................................. 38

1.2. Discussion ...................................................................................................................................... 38

**Population-level effects of heritable symbionts during host dormancy: Bayesian estimation of latent host-symbiont interactions** ..................................................................................... 46

2.1. Methods ......................................................................................................................................... 52

2.1.1. Study system and plant material ............................................................................................... 52

2.1.2. Overview of seed bank experiments .......................................................................................... 53

2.1.3. Initial endophyte prevalence and seed viability ......................................................................... 55

2.1.4. Initial germination and endophyte status of seedlings ............................................................... 56

2.1.5. Seed bank experiments .............................................................................................................. 57

2.1.6. Overview of statistical analysis ............................................................................................... 59
2.1.7. Population model for *Agrostis hyemalis* ................................................................. 60

2.2. Results ................................................................................................................................. 62

2.2.1. Parameter estimates for seed demographic rates. ....................................................... 62

2.2.2. Parameter estimates for plant demographic rates. ...................................................... 65

2.2.3. Population-level predictions for *A. hyemalis* ............................................................ 66

2.3. Discussion ............................................................................................................................ 67

2.4. Conclusion ........................................................................................................................... 72

References .................................................................................................................................. 74

Appendix A: Exploring the converted seed assumption for an obligate seed dormancy model .......................................................................................................................... 83

Appendix B: Exploring the converted seed assumption for a facultative seed dormancy model .......................................................................................................................... 88

Appendix C: Overview of the statistical approach for Bayesian estimation of latent processes ........................................................................................................................................ 91

Appendix D: Assessing the accuracy and precision of Bayesian estimation for seed bank demographics ................................................................................................. 104

Appendix E: Methods and data collection for above-ground demographics .......... 113
List of Figures

Figure 1-1 Phase plot depicting regions of endophyte persistence (above the isoclines) or extinction (below the isoclines) for the obligate seed dormancy model (Eq. 4.1). Axes represent the effects of endophytes on seed survival in the seed bank ($S = \frac{s^+}{s^-}$) and host-plant seed production ($F = \frac{f^+}{f^-}$). Thin gray lines indicate neutral demographic effects. Isoclines show different levels of endophyte loss via imperfect vertical transmission ($\tau$) or retention ($\phi$), which have interchangeable effects (solid: $\tau$ or $\phi = 1$, dash: $\tau$ or $\phi = 0.75$, dotted: $\tau$ or $\phi = 0.25$).

Figure 1-2 Equilibrium endophyte frequencies ($E^+$) predicted by the obligate seed dormancy model (Eq. 5.1) in relation to the effects of endophytes on host-plant seed production ($F = \frac{f^+}{f^-}$). Panels show effects of endophytes on seed survival that are negative (A, $S = \frac{s^+}{s^-} = 0.5$), neutral (B, $S = 1$), and positive (C, $S = 2$). Within each panel, line shading and type represent different levels of endophyte loss from imperfect vertical transmission ($\tau$, gray) or retention ($\phi$, black): dashed, $\tau$ or $\phi = 0.75$; dotted, $\tau$ or $\phi = 0.25$. Solid black lines at $E^+ = 1$ show perfect transmission and retention ($\tau = \phi = 1$).

Figure 1-3 Grid of simulated phase plots depicting regions of endophyte persistence (above the isoclines) or extinction (below the isoclines) for the facultative seed bank model. Thin gray lines indicate neutral demographic effects, as in Fig. 1. Within each plot, line shading represents different types of endophyte loss: imperfect vertical transmission ($\tau$, gray) or retention ($\phi$, black). Line type represents different levels of endophyte loss: Solid lines show perfect transmission and retention, $\tau$ or $\phi = 1$; dashed, $\tau$ or $\phi = 0.75$; dotted, $\tau$ or $\phi = 0.25$. Each row of panel corresponds to different levels of seed bank probability: row 1 (A & B), $b = 0.3$; row 2 (C & D) $b = 0.8$, row 3 (E & F) $b = 1$.

Figure 1-4 Equilibrium endophyte frequencies ($E^+$) are shown in relation to the probability of banking ($b$) for the facultative seed bank model. Panels show endophyte effects on fertility that are neutral, center column ($F = \frac{f^+}{f^-} = 1$, B & E & H), negative, left column ($F = \frac{f^+}{f^-} = 0.5$, A & D & G) and positive, right column ($F = \frac{f^+}{f^-} = 2$, C & I & F). Rows show endophyte effects on seed survival that are neutral (row 1, $S = \frac{s^+}{s^-} = 1$), negative (row 2, $S = \frac{s^+}{s^-} = 0.5$), and positive (row 3, $S = \frac{s^+}{s^-} = 5$). Line shading represents endophyte loss pathways for imperfect retention ($\phi$, black) or imperfect vertical transmission ($\tau$, grey). Line types represent different levels of loss through transmission pathways: solid lines, $\tau$ or $\phi = 1$; dashed, $\tau$ or $\phi = 0.75$; dotted, $\tau$ or $\phi = 0.25$. 
Figure 1-5 Grid of simulated phase plots depicting regions of endophyte persistence (above the isoclines) or extinction (below the isoclines) for the full seed bank model. Within each plot, line shading represents the relationship between $F$ & $S$ (black) or $F$ & $G$ (grey). Rows represent different levels of our third parameter, depicting neutral effects, benefits, or cost of symbionts ($G$ or $S$). Row 1, $G$ or $S$ =1; row 2, $G$ or $S$ >1; row 3, $G$ or $S$ < 1. Line type represents different levels of endophyte loss: dashed, $\tau$ or $\phi$ = 0.75; dotted, $\tau$ or $\phi$ = 0.25 .......................................................... 37

Figure 2-1 Flow chart representation of the methods used to estimate seed bank demography. Ovals represent seeds stocks, squares represent the data generation process in the lab or the field, and arrows connect the origin of information to a give process. For example, from the initial seed stocks, seeds were assessed for viability and endophyte status, seeds were also used for germination trials, and finally seeds were also used for creating an artificial seed bank. .......................................................... 54

Figure 2-2 Figure panel showing the means and 95% credible intervals of seed demographics from the initial and post seed bank data, as estimated from our model. Columns are specific to each species (Fig. 1A,D, E: Agrostis hyemalis; Fig. 1B,F, H: Lolium arundinaceum; Fig. 1C,F, I: Lolium multiflorum). Rows corresponds groups of demographic parameters, where row 1 (Fig.2A,B, C), show estimates of seed viability ($s0 +/-, s1 +/-$). Row 2 shows estimates of seed germination ($g0 +/-, g1 +/-$). Row 3 shows estimates of imperfect retention ($\phi0, \phi1$). Colors distinguish endophyte status (black= E+, grey= E-). .......................................................... 64

Figure 2-3 Joint posterior probability distributions for the equilibrium endophyte frequency in above ground plants under two scenarios. Scenario one, where seed bank dynamics are unaltered (solid black line) and scenario two, the theoretical condition where no seeds survive in the seed bank (solid grey lines). Dashed lines correspond to means of the joint distributions................................................. 67
List of Tables

Table 1-1 Documented results for endophyte effects on seed viability and germination. Benefit (+) or cost (-) refers to the endophyte effect on the specified seed bank vital rate. Symbiont removal refers to whether or not endophytes were experimentally removed from host plants. Treatment refers to experimental treatments imposed by each study. An “*” indicates the resulting endophyte effects were dependent on other factors (treatment or time). ................................................................................................................. 45

Table 2-1 Parameter definitions for the population model, estimated means and 95% credible intervals for all species and host types. Not all vital rates were estimated for every species. Bolded estimates indicate non-overlapping credible intervals and a significant difference between E+ and E- vital rates. An asterisk indicates the quantity was not used in the population model................................................................................................................. 65
List of Equations

Equation 1-1 Obligate seed dormancy model, ................................................................. 18
Equation 1-2 Ratio of reproductive rates ................................................................. 19
Equation 1-3 Ratio of seed survival............................................................................ 19
Equation 1-4 Conditions for symbiont persistence (Obligate seed dormancy) ....... 20
Equation 1-5 Equilibrium symbiont frequency of host plants (Obligate seed dormancy) .................................................................................................................................................. 22
Equation 1-6 Limit of the equilibrium symbiont frequency as $F \to \infty$ ................. 25
Equation 1-7 Limit of the equilibrium symbiont frequency as $S \to \infty$ ................. 25
Equation 1-8 Facultative seed dormancy model .................................................... 27
Equation 1-9 Long-term facultative seed dormancy with variable germination model .................................................................................................................................................. 34
Equation 2-1 Population model for Agrostis hyemalis ............................................. 62
Chapter 1

The role of host demographic storage in the dynamics of vertically transmitted symbionts

Symbiotic microbes can have important effects on the fitness and population dynamics of their hosts (Oliver et al. 2008, Yule et al. 2013) and can even modify community and ecosystem processes (Rudgers & Clay 2007, Ferrari and Vavre 2011). For example, the symbiotic relationship between chemoautrophic bacteria and deep sea vent bivalves (*Baathymodiolus* spp.) and giant tubeworms (*Riftia pachyptila*) form the base of an extensive food web, enabling biomass at vent communities to be 500-1000 times greater than the surrounding environment (Hay et al. 2004). Understanding the processes that govern symbiont persistence and dynamics is thus a valuable area of inquiry.

Among hosts, transmission of symbionts may be horizontal (contagious spread) and / or vertical (from parent to offspring) and is an important dimension of the ecology and evolution of symbiotic microbes. Transmission that is predominantly or exclusively
vertical tightly couples the fitness of the symbiont to that of the host. The dependence of symbiont fitness on host fitness is expected to favor the evolution of host-symbiont mutualism (Ewald 1987, Sachs et al. 2004). Vertical transmission is a key feature of many ecologically important symbioses, including interactions between *Wolbachia* and arthropod hosts (Werren 1997), chemosynthetic microbes and marine invertebrate hosts (reviewed in Dubilier et al. 2008), fungal endophytes and grass hosts (Clay 1993) and gut microbiota and vertebrates (Ley et al. 2008).

Due to the fitness feedbacks imposed by vertical transmission, heritable symbionts that have positive effects on their hosts are expected to become fixed in host populations, whereas parasitic symbionts are expected to be eliminated if opportunities for horizontal transfer are rare or absent (e.g., Clay 1993). Despite this expectation, intermediate frequencies of heritable symbionts (between 0 and 100%) are common in host populations (Schulthess & Faeth 1998, Granath et al. 2007, Hilgenboeker et al. 2008, Rudgers et al. 2009, Miller and Rudgers 2014). Theoretical models have shown that imperfect vertical transmission, where some offspring of symbiotic parents fail to inherit the symbiont, can explain intermediate or low symbiont prevalence, even when effects on hosts are strongly beneficial (Turelli 1994, Ravel 1997, Gundel 2008). These models demonstrate that symbiont dynamics can be highly sensitive to variation in the rate of vertical transmission, and that this rate may be as or more important than symbiont effects on hosts demographic rates. Additionally, empirical evidence from a diversity of microbial symbioses indicates that imperfect vertical transmission is widespread (Jiggins et al. 2000, Darby and Douglas 2003, Narita et al. 2007, Afkhami & Rudgers 2008) and contributes to intermediate symbiont frequencies in host populations (Yule et al. 2013).
While theoretical and empirical work has highlighted the importance of vertical transmission from reproductive hosts to their offspring, other transitions in the host life cycle may provide additional opportunities for symbiont loss. Host demographic “storage” in the form of dormant or pre-reproductive life stages could lead to imperfect transmission of the symbiont prior to host reproduction. We refer to the loss of symbionts during somatic development across life history stages of an individual as “imperfect retention”. For example, heritable bacteria may fail to be retained across pre-reproductive metamorphic stages of arthropod hosts (Darby and Douglas 2003, Geib et al. 2009). Similarly, in plant symbioses, seed dormancy or persistent vegetative stages provide opportunities for symbiont death prior to host reproduction (e.g., Gundel et al 2011). Despite broad potential for imperfect retention in nature, most theory for heritable symbiosis employs simple models where host reproduction is the only opportunity for symbiont loss (Ravel 1997, Gundel et al. 2008, Miller and Rudgers 2014). Furthermore, host life history complexity provides not only additional pathways of symbiont loss but also new pathways for symbiont effects on hosts (Rudgers et al. 2012). For example, effects of symbionts can vary in magnitude and even direction across ontogenetic stages of their hosts (Yule et al. 2013). We know little about how realistic host life history complexity may influence host-symbiont dynamics via stage-specific symbiont loss and effects on hosts.

Here, we develop general theory to examine how host demographic “storage” influences the ecological dynamics of vertically transmitted symbionts. This theory applies broadly to heritable symbioses with some form of host demographic structure including a non-reproductive stage of the life cycle. To facilitate interpretation, we
ground our model in a particular ecological context: the widespread symbiosis between grasses and vertically transmitted fungal endophytes, an ecologically and economically important model system of heritable microbial symbiosis (Cheplick and Faeth 2009). Epichloid fungi live intercellularly within above-ground host tissues and are systemic symbionts of at least 80 genera of host grasses (Saikkonen 2006). It is estimated that up to 30% of all grass species may harbor these microbial symbionts (Leuchtmann 1993). Fungal endophytes can produce alkaloids (Schardl et al. 2013), which are thought to have beneficial effects on various aspects of plant performance, including increased drought tolerance (Morse et al. 2002, Kannadan and Rudgers 2008, Davitt et al. 2011), herbivore resistance (Muller and Krauss 2005, Rudgers and Clay 2007), and competitive ability (Clay et al. 1993, Rudger & Clay 2007). However, the strength and direction of endophyte effects may be dependent on ecological context (Faeth and Sullivan 2003, Saikkonen et al. 2006), including water availability (Rudgers & Swafford 2009) and herbivore pressure (Saikkonen et al. 2006, Afkhami & Rudgers 2009). Additionally, combinations of costs and benefits on different components of host demography have been reported (Rudgers et al. 2012; Yule et al. 2013). Most epichloid fungal endophytes are predominantly or exclusively vertically transmitted and imperfect vertical transmission is well documented (Afkhami and Rudgers 2008, Yule et al. 2013). Population-level endophyte frequencies in natural grass populations are variable and often intermediate (Rudgers et al. 2009, Miller and Rudgers 2014). Given the applied significance of grasses-endophytes in agriculture (Hoveland 1993), there is a need to better understand the processes that determine the prevalence and persistence of endophytes in a grass population.
Like many plants, grass life cycles often include potentially long-lived seed banks (Bakker et al. 1996), or storage stages, raising the possibility of imperfect retention during seed dormancy. Indeed, it is well recognized that fungal endophytes in the seed stage lose viability at a faster rate than do the seeds themselves, providing a convenient means to eliminate unwanted endophytes from agronomic grasses ("storage death": Rolston et al. 1986, Gundel et al. 2009). This observation suggests that passage through a seed bank should promote endophyte extinction from the host population as it increases the chances of endophyte loss. However, it may also be important to consider effects of endophytes on demographic processes at the seed stage, including seed survival in the seed bank and germination probability. Few empirical studies have considered the role of endophytes at the seed stage and available evidence indicates that endophyte effects may vary in direction and magnitude and may depend on environmental context. For example, positive, neutral, and negative effects of endophytes on seed bank survival and germination have been reported (Table 1-1). Mechanisms of endophyte effects on seed demography are poorly understood but may include changes in seed moisture content (Gundel et al. 2007) or maternal effects via seed provisioning (e.g., symbiotic maternal plants may experience a performance advantage that benefits their seeds, without any direct effects of the fungus inside the seed). We know very little about how such effects could influence the prevalence of endophytes in host populations, or about the relative importance of processes in the seed bank vs the plant stage. A stronger theoretical framework for the role of “storage stages” in host-symbiont interactions is needed to guide and interpret empirical work.
Our specific goals were to (1) test the hypothesis that, all else equal, accounting for symbiont loss from storage will promote symbiont extinction because it introduces additional opportunity for symbiont loss, (2) compare the consequences of imperfect vertical transmission vs. imperfect retention for long-run symbiont persistence and prevalence (frequency in the host population, given persistence), (3) compare the consequences of symbiont effects on host demography during storage vs. reproductive stages, and (4) explore the interactions between symbiont loss and effects on host demography. These goals will be addressed under a series of scenarios describing the dynamics for i) obligate storage of a fixed duration, ii) facultative storage of a fixed duration, and iii) facultative storage with a variable duration. Our results allowed us to generate insight into when and how storage stages matter for host-symbiont interactions and when and how they do not, providing a roadmap for future empirical studies. While our theory is motivated by the implications of seed banks in grass-endophyte interactions, this system provides a useful lens for understanding heritable symbiosis more generally because it includes features common to other symbioses: the host life cycle includes a persistent non-reproductive storage stage and the symbiont is predominantly or exclusively vertically transmitted, with potential for imperfect transmission and / or retention. The generality of our theory was maximized with the use of multiple scenarios that differ in whether entry into and exit from the storage stage are obligate or facultative.
1.1. Model for heritable symbioses in a stage-structured host:

methods and results

Our model builds upon an established framework for heritable symbiosis in unstructured annual hosts (Gundel et al. 2008). We consider a plant population structured by two life stages (seeds and reproductive plants) with two endophyte states in each stage: endophyte-symbiotic (E+) and non-symbiotic (E-). Let the vector $N_t$ represent the densities of E- seeds, E- plants, E+ seeds, and E+ plants at time $t$. Change in population size and composition is governed by a discrete-time transition matrix ($A$) according to $N_{t+1} = AN_t$. We consider one-year time steps, a relevant interval for life history transitions of plants in temperate environments. Time steps of differing intervals can be substituted for various host organisms as is necessary. For simple transition matrices, eigenanalysis provides all of the information necessary to predict (1) whether or not symbionts will persist (“persistence”) in the long term and, (2) if so, the frequency at which they will occur (“prevalence”) (Caswell 2001, Gundel et al. 2008); all of our analyses focus on these two components of symbiont dynamics. We assume that the environment is constant, that above-ground plants are semelparous, and that symbiont transmission is exclusively vertical, such that transitions from the E- state to the E+ state have probability zero. We present three model variants of increasing complexity, representing host life histories that vary in flexibility of entry to and exit from the storage stage.
1.1.1. Obligate seed dormancy

We begin with the assumption that seed banking is obligate and of fixed duration. All seeds enter the seed bank; if they survive, they recruit into the plant stage in the following year. Transition from E+ to E- may occur via imperfect transmission from maternal plants to seeds or via imperfect retention from the seed bank to reproductive plants. The transition matrix is thus:

\[
A = \begin{bmatrix}
0 & f^- & 0 & f^*(1 - \tau) \\
0 & s^- & 0 & 0 \\
0 & 0 & s^+ & 0 \\
0 & 0 & s^+\phi & 0 \\
\end{bmatrix}
\]

Equation 1-1 Obligate seed dormancy model,

where each element \(i,j\) represents the transition from life stage / endophyte status \(j\) to \(i\). Parameter \(f\) indicates per-capita seed production by plants and \(s\) indicates the probability of seed survival in the seed bank. Both demographic processes may be unique to endophyte status, indicated with - (E-) and + (E+) superscripts. The endophyte loss parameters are vertical transmission rate \(\tau\) (probability that a seed from an E+ maternal plant is also E+) and retention rate \(\phi\) (probability that a reproductive plant grown from an originally E+ seed is also E+). We refer to E- seeds arising from E+ seeds that lose their endophyte as being “converted”. Whether or not converted seeds maintain their original E+ survival rate will depend on the mechanism of endophyte effects (e.g. maternal provisioning effects should be retained in coverted seeds) and the timing of the effect (e.g., if the endophyte dies near the end of the annual time step, the original E+ value would be the effective survival rate). With very little known about these processes, the
model in Eq.1-1 assumes the more likely scenario that converted seeds retain their original E+ seed survival rate as endophyte death would not likely be immediately after seed dispersal from a maternal plant. The alternative scenario is explored in appendix A, whereby converted seed take on the E- seed survival rate, and discussed below; condition for symbiont persistence were unaffected by this assumption although symbiont prevalence was sensitive to it.

To reduce dimensionality of the model and simplify the analysis, we define $F$ as the ratio of reproductive rates of E+ to E- plants and $S$ as the ratio of survival rates of E+ to E- seeds:

$$F = \frac{f^+}{f^-},$$

**Equation 1-2 Ratio of reproductive rates**

and

$$S = \frac{s^+}{s^-}.$$

**Equation 1-3 Ratio of seed survival**

Long-term endophyte persistence requires that the E+ component of the population increases at a rate that is equal to or greater than the E- component. For consistency with prior theory, we solved for endophyte persistence conditions in terms of the reproductive ratio $F$ (Gundel et al. 2008, Rudgers et al. 2012).
1.1.1.1. Obligate seed dormancy: Persistence

We found that endophytes persistence requires that:

$$ F > \frac{1}{S \tau \phi}.$$ 

**Equation 1-4 Conditions for symbiont persistence (Obligate seed dormancy)**

The reproductive ratio must exceed the inverse of the product of the survival ratio, the vertical transmission rate, and the seed bank retention rate. Fig. 1-1 depicts the boundaries of symbiont persistence / extinction in F-S space. We can glean several biological insights from Eq. 1-4 and Fig. 1-1. First, if endophytes are retained perfectly and have no effects on seed survival ($\phi = 1, S = 1$), Eq. 1-4 reduces to, $F > 1/ \tau$, the established condition for endophyte persistence in an annual, un-structured host: fertility benefits must exceed the inverse of the transmission rate (Gundel et al. 2008). This tells us that demographic complexity *per se* does not modify host-symbiont dynamics; seed banks matter only when they provide new avenues for symbiont loss or effects on hosts. Second, if endophytes have no effect on seed survival ($S = 1$), the potential for imperfect endophyte retention ($\phi < 1$) means that the existence of a seed bank would only destabilize endophyte persistence, all else equal, as we hypothesized. However, if endophytes enhance seed survival ($S > 1$), then seed banks may stabilize endophyte persistence but only if $S > 1/\phi$; that is, seed survival benefits must balance imperfect retention in the same way that fertility benefits must balance imperfect vertical transmission in models lacking seed banks. Third, both pathways of endophyte loss, either in reproductive ($\tau$) or storage ($\phi$) stages, play exactly the same biological role. Inefficiencies in transmission at
either stage have the same effect on persistence of endophytes in the population and their combined effect is multiplicative: all that matters is their product. Finally, Eq. 1-4 can be satisfied even if endophytes are costly somewhere in the life cycle ($F < 1$ or $S < 1$). Costs in one vital rate must be balanced by benefits in the other, and imperfect vertical transmission and retention amplify the benefits required to offset a given cost.

**Figure 1-1 Phase plot depicting regions of endophyte persistence (above the isoclines) or extinction (below the isoclines) for the obligate seed dormancy model (Eq. 4.1).** Axes represent the effects of endophytes on seed survival in the seed bank ($S = s^*/s^*$) and host-plant seed production ($F = f^*/f$). Thin gray lines indicate neutral demographic effects. Isoclines show different levels of endophyte loss via imperfect vertical transmission ($\tau$) or retention ($\phi$), which have interchangeable effects (solid: $\tau$ or $\phi = 1$, dash: $\tau$ or $\phi = 0.75$, dotted: $\tau$ or $\phi = 0.25$).
1.1.1.2. Obligate seed dormancy: Prevalence

Given that endophytes persist, we determined their equilibrium frequency in the host population. We focus on the fraction of plants that are endophyte-symbiotic \( \hat{E}^+ \), a more relevant quantity for field studies of endophyte prevalence than the \( E^+ \) fraction of the total (seeds + plants) population. With eigenanalysis we found that

\[
\hat{E}^+ = \begin{cases} 
\frac{FS\tau\phi - 1}{F - F\tau + FS\tau - 1} & \text{if } F > \frac{1}{S\tau\phi} \\
0 & \text{otherwise.}
\end{cases}
\]

Equation 1-5 Equilibrium symbiont frequency of host plants (Obligate seed dormancy)

Increasing the fertility benefits, seed survival benefits, vertical transmission and retention rates all increase the long-run prevalence of endophyte-positive plants.

In contrast to the condition for persistence (Eq. 1-4), vertical transmission and retention influence endophyte prevalence in different ways that interact with seed survival effects (Fig. 1-2). If effects of endophytes on seed survival are negative \( (S < 1) \), a reduction in the vertical transmission rate has a more negative effect on endophyte prevalence than the same reduction in retention rate (Fig. 1-2A). In part this is caused by the fact that the retention rate does not change the survival response of the seed, only the vertical transmission rate. Conversely, if endophyte effects on seed survival are beneficial \( (S > 1) \), a reduction in retention has a more negative effect on endophyte prevalence than the same reduction in vertical transmission (Fig. 1-2C). Only in the absence of any endophyte effects on seed survival \( (S = 1) \) are the two pathways of loss interchangeable.
(Fig. 1-2B); in this case, seed mortality equally removes E- and E+ seeds such that endophyte loss at either stage has the same dynamical effects. The different consequences of endophyte loss via imperfect transmission or retention reflect the fact that converted E-seeds maintain an E+ seed survival rate. If there is an E+ seed survival advantage that is retained in the converted seeds, symbiont loss via imperfect retention has the more negative effect on prevalence because recruitment of E- plants is enhanced (Fig. 1-2C). On the other hand, negative effects on seed survival depress recruitment of converted E-seeds, making imperfect retention less bad for endophyte prevalence than imperfect transmission. Under the alternative assumption (converted seeds take on E- survival), imperfect transmission and retention have inter-changeable effects on endophyte prevalence (Appendix A).
Figure 1-2 Equilibrium endophyte frequencies ($E^+$) predicted by the obligate seed dormancy model (Eq. 5.1) in relation to the effects of endophytes on host-plant seed production ($F = f^+ / f$). Panels show effects of endophytes on seed survival that are negative (A, $S = s^+/s^- = 0.5$), neutral (B, $S = 1$), and positive (C, $S = 2$). Within each panel, line shading and type represent different levels of endophyte loss from imperfect vertical transmission ($\tau$, gray) or retention ($\phi$, black): dashed, $\tau$ or $\phi = 0.75$; dotted, $\tau$ or $\phi = 0.25$. Solid black lines at $E^+ = 1$ show perfect transmission and retention ($\tau = \phi = 1$).
Fig. 1-2 indicates that imperfect transmission and retention have different consequences depending on how symbionts influence the demography of reproductive and storage stages. We can more formally quantify this dependence by taking the limit of Eq. 1-5 as symbiont benefits to fertility ($F$) or storage survival ($S$) become large. We find that:

$$
\lim_{F \to \infty} E^{+} = \frac{S \phi r}{1 - r + S r}
$$

**Equation 1-6 Limit of the equilibrium symbiont frequency as $F \to \infty$**

and

$$
\lim_{S \to \infty} E^{+} = \phi
$$

**Equation 1-7 Limit of the equilibrium symbiont frequency as $S \to \infty$**

The asymmetry between Eq. 1-6 and 1-7 arises from the fact that seed dormancy is obligate, which means that survival in the seed bank is a very strong filter on endophyte prevalence in the plant population. When the E+ seed survival advantage is large (Eq. 1-7), retention in the seed bank is the sole determinant of endophyte prevalence, with no influence of above-ground processes. With a survival advantage promoting E-recruitment (via converted seeds maintaining an E+ seed survival advantage), retention governs prevalence, with no influence of above-ground process. However, when the E+ fertility advantage is large (Eq. 1-6), below-ground processes continue to play a role in endophyte prevalence because obligate passage through the seed bank can still modify symbiont prevalence in plants. Interestingly, when vertical transmission is perfect ($\tau = 1$),
Eq. 1-6 reduces to Eq. 1-7. This makes biological sense, as seed bank retention is then the only constraint on endophyte prevalence in plants. This asymmetry between endophyte effects on plant fertility vs. seed survival is a general result that is not sensitive to assumptions about the fate of converted seeds (Appendix A).

1.1.1.3. Obligate seed dormancy summary

Accounting for obligate passage through demographic storage can modify expectations for symbiont dynamics, but only if the symbiont is lost from or has effects on the storage stage. The potential for symbiont death during host storage increases the probability of symbiont extinction, as we hypothesized, but this consequence of imperfect retention may be offset by positive effects on host survival during storage i.e. demographic rescue. Similarly, costs of symbionts in one host life stage may be balanced by benefits in the other. Thus ignoring a storage stage of the host life cycle can yield incomplete inferences regarding the fitness effects of symbiosis, and hence the ecological and evolutionary trajectories of the interaction. We find that the storage stage is a very strong filter on symbiont prevalence among reproductive hosts, so strong that, under some conditions, the relative reproductive rates of E+ and E- hosts can have little or no influence on symbiotic plant prevalence. The consequences of imperfect symbiont retention in the seed bank may depend on whether or not converted seed retain E+ demographic rates. Under the most likely scenario where seeds retain the E+ survival rate, symbiont retention in the seed bank is the key driver of symbiont prevalence in the plant population.
1.1.2. Facultative seed dormancy

The above model applies to populations where entry into a seed bank or storage stages is obligate. A more flexible life-history strategy is one where a fraction of the individuals transition to a storage stage while the other fraction recruits directly into the reproductive stage. Here we relax the assumption of obligate dormancy. The transition matrix for a facultative seed bank model is:

\[
A = \begin{bmatrix}
0 & f^-b & 0 & f^+b(1-\tau) \\
f^- (1-b) & s^- (1-\phi) & f^+(1-b)(1-\tau) & 0 \\
0 & 0 & 0 & f^+b\tau \\
0 & 0 & s^\phi & f^+(1-b)\tau
\end{bmatrix}.
\]

Equation 1-8 Facultative seed dormancy model

Parameter \( b \) indicates the probability that a seed will enter the bank, where it remains for one year, versus recruit as a reproductive plant (1-\( b \)). Note that \( b = 1 \) corresponds to the previous, obligate seed bank model and \( b = 0 \) corresponds to an unstructured annual model. We assume that the seed-banking parameter is a species-level trait and does not differ with respect to endophyte status. All previous assumptions still apply.

This model, while complicated by only one additional parameter, is analytically intractable. Therefore, symbiont persistence and prevalence were assessed by numerical simulation. We determined whether and how flexible entry into the seed bank (\( b \)) modified the inferences drawn from the obligate seed bank model. As above, we partitioned the effects of imperfect transmission and retention, and effects of symbionts during reproductive and storage stages. We assessed symbiont persistence and prevalence
after 1,000 generations of simulated population dynamics. Population matrices were scaled with a negative exponential to impose density dependence and ensure population stability. In our simulation symbiont persistence and prevalence were sensitive to the absolute values of fecundity and seed survival parameters. We describe the general trends that are consistent and additionally describe how the absolute value of parameters affects our results.

1.1.2.1. Facultative seed dormancy: Persistence

We found that facultative entry to the seed bank substantially weakened the effect of host demographic storage on symbiont persistence (Fig. 1-3). As the probability of banking decreased, symbiont persistence became insensitive to seed survival and symbiont retention; isoclines that separated regions of symbiont extinction and persistence were nearly vertical and overlapping ($b=0.3$; Fig. 1-3A,B). This indicates that effects on seed survival ($S$) have virtually no influence and instead, persistence of symbionts is determined entirely by their effects on host reproduction. Similarly, while imperfect retention was inconsequential at low seed-banking, imperfect vertical transmission continued to play an important role. That is with low rates of vertical transmission an increasingly large fertility benefit is necessary to offset symbiont loss (Fig 1-3A), as in models lacking seed banks (Gundel et al. 2008). Even when the probability of seed banking was relatively high ($b=0.8$), symbiont effects on host reproduction continued to dominate effects on seed survival (Fig. 1-3C,D), though seed survival benefits could rescue symbionts when fertility benefits were low and vertical transmission was imperfect (Fig. 1-3C). Under high seed banking, imperfect transmission (Fig. 1-3C) still had a greater influence on symbiont persistence than did imperfect
retention (Fig. 1-3D), as was the case when seed banking was low. Only when storage was obligate or nearly so \((b = 1.0)\) were symbiont effects on host fertility and seed survival, and symbiont loss via transmission and retention, equally important (Fig. 1-3E,F). Biologically, this is a result of the host life cycle reverting back toward an obligate seed dormancy model where the filter on endophytes are both the rate of vertical transmission and retention.

![Graphs showing imperfect transmission and retention](image-url)
Figure 1-3 Grid of simulated phase plots depicting regions of endophyte persistence (above the isoclines) or extinction (below the isoclines) for the facultative seed bank model. Thin gray lines indicate neutral demographic effects, as in Fig. 1. Within each plot, line shading represents different types of endophyte loss: imperfect vertical transmission ($\tau$, gray) or retention ($\phi$, black). Line type represents different levels of endophyte loss: Solid lines show perfect transmission and retention, $\tau$ or $\phi = 1$; dashed, $\tau$ or $\phi = 0.75$; dotted, $\tau$ or $\phi = 0.25$. Each row of panel corresponds to different levels of seed bank probability: row 1 (A & B), $b = 0.3$; row 2 (C & D) $b = 0.8$, row 3 (E & F) $b = 1$.

The trends describe above are general however, for intermediate to high levels of seed banking ($0.3 \leq b < 1$), the degree of the effect of demographic storage on symbiont persistence was dependent on the absolute values of host fecundity and seed survival (not shown). For example, higher absolute values of E+ fecundity, all else equal, exaggerated the weakening effect of facultative entry into the storage stage. Isoclines for symbiont persistence became overlapping and vertical more so than if the absolute value of E+ fecundity was reduced. The opposite effect was seen with higher absolute values of E+ seed survival, all else equal. This resulted in demographic storage having a stronger effect on symbiont persistence as isoclines became less overlapping and vertical than if the absolute value of E+ seed survival were lower. These results indicate that the true affect of facultative entry into the seed bank on symbiont persistence will also depend on the magnitude of host reproduction relative to the degree of seed survival within the seed bank.

1.1.2.2. Facultative seed dormancy: Prevalence

Like symbiont persistence, the effects of seed banking on symbiont prevalence in reproductive plants, were conditional on symbiont loss and effects on host demography,
but were generally weak (Fig. 1-4). Long-run symbiont frequency increased with increasing transmission/retention rates and demographic effects on hosts, as expected, but was relatively unresponsive to seed-banking ($b$). For example, when there was a beneficial effect of symbiosis on above-ground plant fertility, seed bank processes (seed survival and retention) had virtually no influence on symbiont prevalence except when seed-banking was nearly obligate (Fig. 1-4C,F, I). When seed-banking did influence symbiont prevalence, effects were generally negative. For example, when symbiont effects on seed survival were neutral or costly ($S \leq 1$), seed banking could only decrease symbiont prevalence (Fig. 1-4A-F); a low probability of seed-banking was required for imperfect retention to drive symbionts extinct when fertility effects were neutral (Fig. 1-4B, E) and a high probability ($b > 0.7$) was required when fertility effects were beneficial (Fig. 1-4C,F). Only when symbionts positively affected seed survival did increasing the probability of seed-banking increase symbiont prevalence (Fig. 1-4G, H, I). Interestingly, seed banking could depress symbiont prevalence when symbiont retention in the seed bank was imperfect, even under strong seed survival benefits (Fig. 1-4H, I). This result is due to the cost of losing endophytes in the seed bank and is not sensitive to the assumption of how converted seeds maintain an E+ seed survival rate (Appendix B). Surprisingly, at a high probability of seed banking, imperfect retention allowed symbionts to persist at higher frequencies when seed survival effects were negative ($S < 1$) compared to when effects were positive ($S > 1$) (Fig. 1-4C,I). Biologically, this is a result of converted seeds maintaining E+ survival rates. When symbionts had a negative effect on seed survival, mortality of converted E- seeds was amplified, boosting representation of E+ hosts emerging from the seed bank. Under the
alternative assumption that converted seeds survive with an $E$- rate, this result no longer holds (figure not shown).

Figure 1-4 Equilibrium endophyte frequencies ($E^*$) are shown in relation to the probability of banking ($b$) for the facultative seed bank model. Panels show endophyte effects on fertility that are neutral, center column ($F = f^+/f^- = 1$, B & E & H), negative, left column ($F = f^+/f^- = 0.5$, A & D & G) and positive, right column ($F = f^+/f^- = 2$, C & I & F). Rows show endophyte effects on seed survival that are neutral (row 1, $S = s^+/s^- = 1$), negative (row 2, $S = s^+/s^- = 0.5$), and positive (row 3, $S = s^+/s^- = 5$). Line shading represents endophyte loss pathways for imperfect retention ($\phi$, black) or
imperfect vertical transmission ($\tau$, grey). Line types represent different levels of loss through transmission pathways: solid lines, $\tau$ or $\phi = 1$; dashed, $\tau$ or $\phi = 0.75$; dotted, $\tau$ or $\phi = 0.25$.

The effects of seed banking describe above are general trends however, the degree of that effect of seed-banking depends on the absolute values of host fecundity ($f$) and seed survival ($s$), as noted for symbiont persistence. Under a scenario where the absolute value of the fecundity advantage is larger but the ratio of $F$ remains the same (not shown), seed-banking still has a generally negative effect on symbiont prevalence when symbiont effects on seed survival were neutral or costly ($S \leq 1$) but a higher probability of seed banking was required for imperfect retention to reduce symbiont prevalence. When the absolute value of the seed survival advantage increased but the ratio of $S$ remained the same (not shown), the dynamics were opposite in effect and a lower probability of seed banking was required for imperfect retention to reduce symbiont prevalence. This suggests the effect of seed banking on symbiont prevalence will depend on the magnitude of host fecundity relative to the degree of seed survival within the seed bank, as with symbiont persistence.

1.1.2.3. Facultative seed dormancy summary

In contrast to the obligate storage model, where seed banks were as or more important than reproductive stages, facultative entry to the seed bank substantially reduced its influence on symbiont dynamics. The once symmetrical relationship between seed survival ($S$) and host reproduction ($F$) that predicts symbiont persistence became insensitive to effects on seed survival at realistically low levels of seed-banking. Only when the probability of seed-banking was high did imperfect retention and effects of
symbionts on seed survival influence the conditions for symbiont persistence. The generality of this statement will depend on the magnitude of host fecundity relative to the degree of seed survival e.g. low fecundity and high seed survival can result in a noticeable effect of moderate levels of seed banking. When symbionts did persist, seed banking generally had negative effects on long-run prevalence except when symbiosis enhanced survival in the seed bank and symbiont retention was high. As in the obligate storage model, the demographic fate of converted seeds can lead to surprising and non-intuitive effects of seed-banking on symbiont prevalence.

1.1.3. Long-term facultative seed dormancy with variable germination

Up to this point, all model forms have assumed dormancy is of fixed duration (any banked seed germinates after one time step of dormancy, given that it survives). Here we relax this assumption to allow for longer-term dormancy and effects of symbiosis on an additional demographic process in the seed bank: germination. The transition matrix for a long-term, facultative seed bank with variable germination is:

\[
A = \begin{bmatrix}
    s^{-}(1-g^-) & f^-b & s^+(1-g^+)(1-\phi) & f^+b(1-\tau) \\
    s^-g^- & f^-(1-b) & s^+g^+(1-\phi) & f^+(1-b)(1-\tau) \\
    0 & 0 & s^+(1-g^+)(1-\tau) & f^+b\tau \\
    0 & 0 & s^+g^+\phi & f^+(1-b)\tau
\end{bmatrix}
\]

**Equation 1-9 Long-term facultative seed dormancy with variable germination model.**

We introduced two new parameters \(g^+\) and \(g^-\), which represent the germination rates of infected and uninfected seeds, respectively. Note that \(g^+ = g^- = 1.0\) corresponds to the
previous, fixed-duration seed bank model (Eq. 1-9). After seeds have entered and survived in the bank for one year, they germinate into above ground plants with probability $g$. Those seed that do not germinate ($1-g$) but survive ($s$), remain dormant until either death or germination occurs in following time steps. Seeds that do not bank ($1-b$) germinate directly into above ground plants. E+ seeds and plants can produce E- seeds and plants as described in previous transition matrices. As with fertility and seed survival, we define the influence of symbionts on seed germination probability as $G = g^+ / g^-$. By relaxing the assumption of dormancy for a fixed duration we have added biological realism and another avenue through which symbiont dynamics can be affected. All previous assumptions of the facultative seed bank model apply. In addition, the projection matrix in Eq. 1-9 assumes that converted seeds maintain E+ vital rates, as before, but only for one time step. Converted seeds that persist in the seed bank eventually take on E- vital rates (because the model does not track seed history for more than one time step). We use numerical simulation to explore whether and how the influence of symbionts on seed survival and germination differ in their effects on symbiont persistence and prevalence.

1.1.3.1. Long-term facultative seed dormancy with variable germination:

Persistence

As before, we generated isoclines that define the combinations of fertility effects ($F$) and seed survival effects ($S$) ($F$ vs. $S$, black lines) necessary for symbiont persistence in the host population (Fig. 1-5). To the same plot we added the isocline between fertility effects ($F$) and seed germination effects ($G$) ($F$ vs. $G$, grey lines; Fig. 1-5). Because we know that seed bank processes only affect symbiont persistence when seed-banking is
high (Fig. 1-3), we compare the effects of seed survival and germination when $b = 1.0$. We consider each pairwise isocline ($F$ vs. $S$ or $F$ vs. $G$) with the “third” vital rate held constant at negative (Fig. 1-5A,B), neutral (Fig. 1-5C,D), and positive (Fig. 1-5E,F) effects of symbionts. For example, in Fig. 1-5C, the $F$-$S$ isocline (black lines) holds constant $G = 1$ while the $F$-$G$ isocline (grey lines) holds constant $S = 1$. Against this backdrop of demographic effects, we compare the consequences of symbiont loss via imperfect transmission (Fig. 1-5A,C,E) and retention (Fig. 1-5B,D,F). With this approach, we can assess whether host seed survival or germination differ in their effects on symbiont persistence.

Effects of symbionts on seed survival vs. germination were generally consistent in their influence on symbiont persistence (Fig. 1-5). The differences between loss from retention (right column) and loss from vertical transmission (left column) were generally slight, regardless of the levels of symbiont loss. However, there were some differences between the isoclines defined by survival and germination. Specifically, the $F$-$S$ isocline was lower than the $F$-$G$ isocline when the third vital rate (germination and survival, respectively) exhibited a cost of symbiosis and especially when transmission or retention was imperfect (Fig. 1-5A,B). Biologically, this means that the germination benefits required to offset a seed survival cost are greater than the survival benefits required to offset the same cost in germination. Put another way, a positive effect on seed survival is a more potent pathway to maintain symbionts in the host population than positive effects on germination. As before, germination costs of symbiosis can be offset by benefits to fertility and / or seed survival to maintain symbionts in the population, and vice versa.
Figure 1-5 Grid of simulated phase plots depicting regions of endophyte persistence (above the isoclines) or extinction (below the isoclines) for the full seed bank model. Within each plot, line shading represents the relationship between $F$ & $S$ (black) or $F$ & $G$ (grey). Rows represent different levels of our third parameter, depicting neutral effects, benefits, or cost of symbionts ($G$ or $S$). Row 1, $G$ or $S = 1$; row 2, $G$ or $S > 1$; row 3, $G$ or $S < 1$. Line type represents different levels of endophyte loss: dashed, $\tau$ or $\phi = 0.75$; dotted, $\tau$ or $\phi = 0.25$.
These persistence conditions were not sensitive to the demographic fate of converted seed (not shown). In addition, the absolute value of the “third” vital rate for our demographic background had insignificant effects on our results (not shown).

1.1.3.2. Long-term facultative seed dormancy with variable germination:

Prevalence

Given that symbionts can persist, we asked whether seed survival and germination differed in their effects on long-run symbiont prevalence. As above, we found that effects of symbionts on the two seed bank vital rates were virtually inter-changeable. A cost or benefit to seed germination had the same influence on symbiont prevalence as a cost or benefit to seed survival (not shown).

1.1.3.3. Variable germination summary

Accounting for longer-term storage did not modify predictions for the role of seed banks. While long-term storage adds a demographic process (seed germination), effects of symbionts on germination mirrored their effects on seed survival. The two pathways of demographic effects in the seed bank similarly affected symbiont persistence and prevalence, though there is some evidence that symbiont effects on survival in the seed bank were more consequential than germination out of the seed bank.

1.2. Discussion

Heritable symbioses are broadly distributed and ecologically consequential. Understanding the processes that promote or constrain symbiosis is a key challenge in
many basic and applied contexts. Previous theory has taken a simplistic view of host
demography, providing only one avenue (reproduction) for symbionts to influence and be
influenced by their hosts. We present the first theoretical model to account for multiple
stages of a host life cycle, including potential for long-term demographic storage, thereby
accounting for realistic demographic complexity to understand mechanisms affecting
symbiont persistence and prevalence. Our results reveal when and how storage stages
influence symbiont dynamics. Below, we describe the general significance of our
findings and what they suggest for future studies.

We explored a range of life history strategies covering realistic variation of
symbiont entry and exit into host storage stages. For life histories with obligate seed
dormancy, imperfect retention of symbionts has the same dynamical effect on symbiont
persistence as imperfect vertical transmission, and loss through either pathway at high
enough rates can lead to symbiont extinction. While our model explores the role of
dormancy, or seed banks, the seed stage itself can be considered an obligate storage stage
and influence host-symbiont dynamics regardless of dormancy. Extending or contracting
the unit of time for our obligate seed dormancy model would account for this as well as
other scenarios. For example, within the grass-endophyte literature four life history
transitions have been identified as points where symbiont transmission can be interrupted
(Gundel et al. 2011), and there is empirical evidence to demonstrate loss for three of
those four transitions (Afkhami and Rudgers 2008). Failing to account for all forms of
symbiont loss could lead to inaccurate predications about host-symbiont trajectories.

Another example of a vertically transmitted symbiont that our obligate seed bank
model can be applicable to is Wolbachia, which occur in up to 60% of insect species
Arthropods can also harbor other vertically transmitted facultative symbionts, such as members of Enterobacteriaceae, however other modes of transmission should be considered with modeling these host-symbiont dynamics i.e. biparental vertical transmission (Moran and Dunbar 2006). For Wolbachia, previous work has indicated three important factors that have a demographic basis and mediate prevalence: survival and fecundity of symbiotic and non-symbiotic types, transmission, and levels of cytoplasmic incompatibility expression (Werren 1997). Our obligate seed dormancy model suggests the integration of imperfect transmission across all life stage transitions (metamorphic stages of insect host) into this framework can facilitate accurate predictions regarding symbiont prevalence and persistence for this interaction.

In contrast to a model for obligate seed dormancy, facultative seed dormancy result in the effects of imperfect retention on symbiont persistence and prevalence dependent on the frequency that transitions to dormancy occur. As an empirical example, seed dormancy in the annual sunflower, Helianthus annuus, is estimated to occur in upwards of 40% of produced seeds (Moody-Weis & Alexander 2007). Dormancy in a pooid grass species, Agrostis hymenalis, that host the vertically transmitted endophyte, Epichloë amarillans, is estimated to occur in up to 80% of produced seed (Bibian et al., Unpublished data). Should these estimates be representative of the upper boundary for rates of dormancy in semelparous host, then our facultative seed dormancy model indicates that accounting for imperfect retention and demographic rates of facultative storage stages may significantly alter predictions for host-symbiont dynamics.

Another biological grounding for our facultative seed dormancy model is for organisms with dormant propagule pools, or egg banks, as they can substantially
influence population persistence (Hariston Jr & De Stasio Jr 1988). Long-term demographic storage can substantially vary among diapausing crustaceans, ranging from 0-16 years (Hariston Jr and Cáceres, 1996). Although no data were presented on rates that dormant offspring are produced, this variation among life-history strategies suggest the effect of dormancy on population persistence varies and should be explored.

Although diapausing host may differ biologically from scenarios considered here, being iteroparous rather than semelparous, host can harbor vertically transmitted endosymbionts such as microsporidia (Dunn & Smith 2001). Despite being considered parasitic, positive effects of microsporidia on host reproduction have been documented (Haine et al. 2004) and imperfect transmission among host has been documented (Mangin et al. 1995). In the study documenting imperfect transmission, experimental measurements of the rate of vertical transmission were 100%, suggesting other potential pathways for symbiont loss i.e. imperfect retention. Whether this and other vertically transmitted host-symbiont systems would yield different predictions about population outcomes by accounting for egg bank dynamics is an open question.

Exploration of empirical studies estimating the effects of endophytes on host seed dynamics reveal studies are sparse and variable (Table 1.1). This variation suggests future studies should be more consistent, focused, and directed. Empiricist should give consideration to the process of generating non-symbiotic seeds or using naturally occurring non-symbiotic host. We suggest the former because experimentally removing symbionts eliminates the potential of having genetic background as a confounding factor in subsequent analysis. Future studies should focus on determining the effects of endophytes on both the germination and survival rate simultaneously. Doing so can help
assess whether symbiont effects on host vital rates are consistent across two ecologically distinct processes. In addition, standardizing how germination and seed survival rates are estimated will help eliminate any bias associated with using particular methods or proxies for estimates. The current literature (Table 1-1) suggest an effort should be directed toward estimating the rates of imperfect retention, both from seeds that are dormant and make up the seed bank and those that are inherently not dormant, as few studies have attempted to do so.

From an evolutionarily perspective, the vital rates associated with the demographic storage stage of a host should reflect selection pressures that either maintain or drive symbionts to extinction. The fact that imperfect retention can result from endophyte death in seeds (Rolston et al. 1986) suggest that there should be an expectation for selection to increase germination rates of symbiotic seeds and decrease the rate that dormancy occurs (Afkhami & Rudgers 2008). Similarly, selection could increase the rate of seed dormancy for symbiotic seed and symbiotic seed survival rates if endophytes can maintain viability for long durations. Current empirical data fails to support the first hypothesis, as endophyte effects on germination are variable. Some of this variation may be attributed to the experimental treatments and/or the effect of genetic background. This coarse assessment can be more rigorous and assessing empirical conformity to the second hypothesis would be possible with additional information. What is needed are studies that estimate seed bank vital rates through time, measure the rate of symbiont transitions to dormancy, and estimate symbiont loss through imperfect retention. Nearly all former studies of endophyte effects on seed vital rates do not generate estimates across the time
scale that seed banks can affect host-symbiont dynamics or measure the rate that transition to dormancy occur; further research is needed in this area.

The inclusion of realistic complexity into models of host-symbiont dynamics can result in parameters having an unexpected influence on symbiont persistence in the population (Yule et al. 2013; Rudgers et al. 2012) and generate hypotheses for future empirical work. Confronted with the issue of assigning symbiotic or non-symbiotic vital rates to a host when symbionts are lost through imperfect retention and we see that this distinction is non-trivial (see Appendix A). Our interpretation of this dilemma is related to various mechanisms that can affect the demographic fate of converted non-reproductive host. One mechanism that would negate all components of the timing of symbiont death in our system is a direct effect of the concentration of fungal hyphae in the seed relative to its biomass. As an example, Gundel et al. (2009) hypothesize that the concentration of mycelium in the seed can alter the distribution of water content thus affecting seed survival and rendering endophyte death negligible in determining the demographic fate of converted seeds. The same can be said of the case where the demographics of converted seeds are governed by maternal effects imposed through endophyte presence. Yet, an alternative mechanism that can affect the demographics of converted seeds and where the timing of symbiont death would matter, is a scenario where viable endophytes exert a measurable influence via biological pathways. Quantifying the viability of symbionts and timing of their death however, may be particularly difficult in nature and thus an alternative would be to model these processes with variation regarding the influence of symbionts on the vital rates in non-reproductive stages. Ultimately, a mechanistic understanding of how symbionts affect host, either
through physical presences, maternal effects, or demography is needed to better understand host-symbiont dynamics.

In conclusion, demographic storage can alter host-symbiont dynamics when symbiont transitions in non-reproductive stages are obligate or in the case of a dormant seed bank, transitions to dormancy occur at a high frequency. The magnitude of host reproduction relative to symbiont survival in the storage stage can also play a role in determining symbiont prevalence and persistence. Future works should focus on empirically estimating symbiont loss during host ontogeny, determining, if any, the stage specific fitness effects symbionts confer (or not) to host. Validating the predictions of this model will require acquire estimation of host vital rates and an understanding of the mechanism responsible for symbiont effects on host non-reproductive stages.
**Table 1-1 Documented results for endophyte effects on seed viability and germination.** Benefit (+) or cost (-) refers to the endophyte effect on the specified seed bank vital rate. Symbiont removal refers to whether or not endophytes were experimentally removed from host plants. Treatment refers to experimental treatments imposed by each study. An “*” indicates the resulting endophyte effects were dependent on other factors (treatment or time).

<table>
<thead>
<tr>
<th>Seed Vital Rate</th>
<th>Benefit (+) or Cost (-) to Host</th>
<th>Treatment</th>
<th>Symbiont Removal</th>
<th>Host Spp.</th>
<th>Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Germination</td>
<td>(+)</td>
<td>N/A</td>
<td>No</td>
<td><em>Lolium perenne</em></td>
<td><em>Clay 1987</em></td>
</tr>
<tr>
<td>Germination</td>
<td>(+)</td>
<td>N/A</td>
<td>No</td>
<td><em>Lolium multiflorum</em></td>
<td><em>Clay 1987</em></td>
</tr>
<tr>
<td>Germination</td>
<td>(+)</td>
<td>N/A</td>
<td>No</td>
<td><em>Bromus setifolius</em></td>
<td>Novas et al. 2003</td>
</tr>
<tr>
<td>Germination</td>
<td>Neutral</td>
<td>N/A</td>
<td>No</td>
<td><em>Festuca arizonica</em></td>
<td>Faeth et al. 2004</td>
</tr>
<tr>
<td>Germination</td>
<td>(-)*Treatment</td>
<td>Temperature</td>
<td>Yes</td>
<td><em>Lolium multiflorum</em></td>
<td>Vila-Aiub et al. 2005</td>
</tr>
<tr>
<td>Germination</td>
<td>Neutral</td>
<td>Water Availability</td>
<td>Yes</td>
<td><em>Festuca arizonica</em></td>
<td>Neil et al. 2006</td>
</tr>
<tr>
<td>Germination</td>
<td>(-)*Light</td>
<td>Light &amp; Temperature</td>
<td>Yes</td>
<td><em>Lolium multiflorum</em></td>
<td>Gundel et al. 2006a</td>
</tr>
<tr>
<td>Viability</td>
<td>(-)<em>Treatment</em>Time (+)</td>
<td>Water Availability</td>
<td>Yes</td>
<td><em>Lolium multiflorum</em></td>
<td>Gundel et al. 2006b</td>
</tr>
<tr>
<td>Viability</td>
<td>(-)*Hum &amp; (-)*Temp</td>
<td>Humidity &amp; Temperature</td>
<td>Yes</td>
<td><em>Lolium multiflorum</em></td>
<td>Gundel et al. 2009</td>
</tr>
<tr>
<td>Viability</td>
<td>(-)*Time</td>
<td>Experimental/ Field Aging</td>
<td>Yes</td>
<td><em>Lolium multiflorum</em></td>
<td>Gundel et al. 2010</td>
</tr>
</tbody>
</table>
Chapter 2

Population-level effects of heritable symbionts during host dormancy: Bayesian estimation of latent host-symbiont interactions

Many multicellular organisms host symbiotic microbes that can alter their ecological and evolutionary trajectories (Gilbert et al. 2012). Symbiotic microbes can affect the fitness and population dynamics of their host (Oliver et al. 2008, Yule et al. 2014), and induce effects that scale up to higher levels of biological organization, altering communities (Knowlton & Rohwer 2003, Hay et al. 2004, Ferrari and Vavre 2011) and ecosystem properties (Rudgers & Clay 2007). In addition, symbioses can promote evolutionary adaptation and lead to the diversification of lineages (Moran 2007, Joy 2013). Therefore, understanding factors affecting host-symbiont dynamics and the long-term persistence of symbionts in host populations is an important line of inquiry.
The mode of symbiont transmission among host is expected to shape the ecological role of symbionts as parasites or mutualists. Vertical transmission, or the transfer of symbionts from parent to offspring, is a strategy that is expected to evolve toward a mutualism, as the fitness of the symbiont is dependent on the fitness of the host (Ewald 1987, Sachs et al. 2004). This transmission strategy is a common feature of host-symbiont interactions including arthropods and \textit{Wolbachia} (Werren 1997), grasses and fungal endophytes (Clay 1993), marine invertebrates and microbes (Sharp et al. 2007, Dubilier et al. 2008), and many components of the human microbiome (Ley et al. 2008). The fitness feedback inherent to vertical transmission suggests that symbionts should increase host fitness and become fixed in host populations (Clay 1993). However, empirical evidence demonstrates vertically transmitted symbionts persist at intermediate frequencies in host populations (Schulthess & Faeth 1998, Granath et al. 2007, Hilgenboeker et al. 2008, Rudgers et al. 2009, Semmartin et al. 2015). A more mechanistic understanding of host-symbiont interactions is needed to shed light on this discrepancy.

Imperfect vertical transmission, or the failure of offspring to inherit the symbiont, has been suggested as a mechanism that can lead to intermediate symbiont frequencies (Turelli 1994, Ravel 1997, Gundel et al. 2008). With symbiotic hosts continually supplying an influx of non-symbiotic host into the population via imperfect vertical transmission, intermediate frequencies can be maintained even if symbiotic hosts have a strong fitness advantage over non-symbiotic ones. Indeed, empirical evidence demonstrates imperfect vertical transmission of heritable symbionts is a common feature among a diversity of host organisms (Hurst et al. 2001, Moran & Dubnar 2006, Gibert &
Hazard 2013, Yule et al. 2013). However, this mechanism for explaining intermediate symbiont frequencies is derived from theory that focuses solely on reproductive life stages of the host (but see Hancock et al 2012). This means that the benefits and loss of symbiosis are assumed to occur at a single demographic transition, from adult to offspring.

Recent empirical and theoretical work implicates an important role for symbiont loss within, as well as between, host generations. We refer to the loss of a symbiont during the ontogenetic development of an individual host as “imperfect retention”, in contrast to imperfect transmission from parent to offspring. In plants, heritable symbionts can be lost during seed storage in a seed bank if symbionts lose viability at a faster rate than seeds (Rolston 1986, Gundel 2009). Additionally, plant symbionts may be lost when transitioning through pre-reproductive host life stages, such as transitions from seed to seedling and seedling to adult (Afkhami & Rudgers 2008). Similarly, arthropods can lose heritable symbionts such as *Wolbachia* when hosts metamorphose through pre-reproductive life stages (Geib et al. 2009). These examples of additional symbiont loss, introduced when accounting for the stage structure of a host, should be expected to reduce symbiont frequency and increase the likelihood of symbiont extinction, all else equal, because they provide new opportunities for symbiont loss. Recent theory suggests that imperfect retention can play a significant role in host-symbiont dynamics, but that this additional pathway of symbiont loss can be balanced by positive effects of symbionts on demographic vital rates at the seed stage. These expectations have yet to be tested empirically.
In addition to new pathways of symbiont loss, considering host stage structure also introduces new contexts in which symbionts can affect hosts, beyond the effects on host reproductive output considered in most theoretical models. Empirical evidence demonstrates that symbionts can affect vital rates associated with growth and survival of pre-reproductive life stages in a variety of ways (positive, negative, or neutral effects) in both plants (Yule et al. 2013) and arthropods (Fry et al. 2004, Himler et al. 2011). Only rarely have the component effects of symbionts on particular life stages or vital rates been integrated over the life cycle to yield the net effects on host fitness and hence long-term expectations for the persistence and prevalence of the symbiosis. Available studies indicate that negative effects on one host vital rate may be counter-balanced by positive effects on another, with an ultimate net benefit (Rudgers et al. 2012, Yule et al. 2013, Miller and Rudgers 2014). A fully demographic perspective, including demographic transitions at pre-reproductive life stages, holds potential to resolve puzzling patterns of high symbiont prevalence in host populations despite negative effects on plant reproduction (Faeth and Sullivan 2003, Rudgers et al. 2008). “Pre-reproductive” stages include potentially long-lived dormant stages, such as seed banks. To our knowledge, no previous studies have empirically estimated the influence of a host dormant stage on host-symbiont dynamics.

Here we use symbioses between grasses and vertically transmitted fungal endophytes as a model system to examine the effects of symbiosis at the seed stage and the influence of seed-banking on persistence and prevalence of symbionts in host populations. Grass life cycles often include potentially long-lived seed banks (Bakker et al. 1996) and up to 30% of all grass species harbor epichloid fungi that live intercellularly
within above-ground host tissues (Leuchtmann 1993). Most epichloid fungal endophytes are exclusively or predominantly vertically transmitted. Studies of “above-ground” host demography have shown that endophyte symbiosis can positively affect host fitness through increased drought tolerance (Morse et al. 2002, Kannadan and Rudgers 2008, Davitt et al. 2011), herbivore resistance (Muller and Krauss 2005, Rudgers and Clay 2007), and competitive ability (Clay et al. 1993, Rudger & Clay 2007), though these effects may be context-dependent. Studies have also documented that endophytes typically lose viability in the seed at a faster rate than the seed itself (Rolston et al. 1986, Gundel et al. 2009), providing evidence of imperfect retention from the seed stage. Furthermore, endophyte symbiosis can positively or negatively affect seed germination (Clay 1987, Novas et al. 2003, Gundel et al. 2006a, Clement et al. 2008, Wäli et al. 2009) and viability (Gundel et al. 2006b, Gundel et al. 2009, Gundel et al. 2010). Thus, available evidence suggests that consideration of the seed bank may be an important component of grass-endophyte population dynamics – due to imperfect retention and demographic effects of endophytes at the seed stage – and may lead to different conclusions than if only “above-ground” life history transitions were considered.

Estimating the vital rates of symbiotic (endophyte-positive, or E+) and non-symbiotic (endophyte-negative, or E-) seeds and the rates of symbiont retention during the seed stage poses a non-trivial statistical challenge. The root of the challenge stems from the fact that seeds must be destroyed to determine their endophyte status, and so understanding the effects of symbiosis on seeds requires indirect inference. Given a mixed pool of E+ and E- seeds, the endophyte frequency among seedlings that recruit from this seed pool will reflect viability and germination rates that may differ between E+
and E- seeds, as well as the conversion of E+ seeds to E- seedlings via imperfect retention. Thus, a simple contrast of endophyte frequencies between seeds and seedlings suffers from a parameter identifiability problem, since many combinations of vital rate estimates could give the same result. The imperfect retention process makes parameter inference especially tricky because it creates two, non-exclusive ways to observe an E+ seedling frequency that is lower than expected based on the E+ seed frequency: (1) negative effects of endophytes on seed viability and/or germination, and (2) imperfect retention of endophytes in E+ seeds. This challenge can be met through hierarchical Bayesian methods (Clark 2005), which generate parameter estimates based on unobservable processes (“state-space models”), as we demonstrate here.

We combined experimental seed bank studies, hierarchical Bayesian statistical methods, and stage-structured demographic models to address two major objectives. The first objective was to quantify the effects of endophyte symbiosis on demographic vital rates of seeds (viability, germination, and how these change with storage in the seed bank) and imperfect retention of endophytes during the seed stage, and to demonstrate the utility of hierarchical Bayesian methods for this purpose. We replicated these measurements across three grass host species that harbor heritable fungal endophytes. Second, for one species that showed potential for long-term seed-banking, we linked seed bank dynamics to above-ground processes in a full demographic model. This allowed us to understand how consideration of the seed bank modifies population-level predictions for host-symbiont dynamics, including the net fitness effects of symbiosis for the host-plant and the long-run prevalence of symbionts in the host population. Our expectation was that if retention of symbionts in the seed bank was low, then host seed-banking
should reduce endophyte prevalence and increase symbiont extinction risk. However, symbiont benefits during both seed bank and reproductive life stages may outweigh the negative effects of symbiont loss. Thus, our goal in the second part of the paper was to quantify the balance of costs, benefits, and loss of symbiosis over a host’s complex life cycle.

2.1. Methods

2.1.1. Study system and plant material.

Our work focused on three cool-season (C3) grass host species and their vertically transmitted fungal endophytes. For each species, we used symbiotic (E+) seeds and experimental non-symbiotic (E-) seeds whose endophytes had been eliminated. *Agrostis hyemalis* is native to eastern North America and can exhibit an annual, biennial or short-lived perennial life cycle. *A. hyemalis* hosts the endophyte *Epichloë amarillans* (White 1994), which is exclusively vertically transmitted (Yule et al. 2013, Miller and Rudgers 2014). Our E+ seed stock originated from a natural population in the Davey Crockett National Forest near Lufkin, Texas. We generated an experimental E- seed stock that was derived from another, nearby (ca. 28 km) natural population in east Texas. Endophytes were eliminated through the natural process of ‘storage death’ and grown through one generation at experimental plots in Houston, Texas. Thus, E+ and E- seed lineages were derived from the same general region but, because they were not from the same genetic lineage (a consequence of greater-than-intended efficacy of endophyte elimination), may have had some differences in host genetic background. The second species was *Lolium multiflorum*, an annual grass from southern Europe that has become naturalized in mesic
areas of South and North America (Omacini et al. 2004). This species hosts the endophyte *Neotyphodium occultans*, which is strictly vertically transmitted. E+ seeds were provided by Marina Omacini’s lab at the University of Buenos Aires and were derived from a naturalized population in Argentina. We generated E- seeds with a fungicide treatment applied to E+ seeds. Specifically, 2.5 mg of active ingredient (Prochloraz) was combined with 0.25 mL water and applied to 0.5 g of E+ seeds for 24 hrs. Seeds were then reared in the greenhouse, and seeds of the next (E-) generation were used in our experiment. The third species was the perennial agricultural cultivar *Lolium arundinaceum*, or Kentucky 31 tall fescue. This species host the vertically transmitted endophyte *Acremonium coenophialum* (White & Cole 1985). Tall fescue seeds of each host type (E+ and E-) were procured from the University of Kentucky.

### 2.1.2. Overview of seed bank experiments.

We created artificial seed banks in a field setting for each of the three host-symbiont pairs to quantify: (1) how endophyte symbiosis and seed bank storage affected seed viability and germination rate and (2) the imperfect retention of fungal endophytes from seeds to seedlings and whether it increased (poorer retention) with seed bank storage. The seed bank experiments were conducted in a recently tilled open-field habitat at the USDA East Texas Plant Materials Center located adjacent to the Stephen F. Austin Experimental Forest near Nacogdoches, Texas.

Figure 2-1 shows a schematic overview of our methods, which included viability and germination assays conducted with fresh seeds and replicated after one year in an experimental seed bank. We began by testing the initial viability and germination of E+
and E- seed stocks. The “E+” stocks were not pure (100%) E+ due to imperfect vertical transmission and intermediate frequencies in the natural populations from which they were derived. In fact, the “E+” seed stock of *L. multiflorum* was predominantly E-(Results). There was also the possibility that “E-“ seed stocks were not 100% E- due to incomplete endophyte elimination. We therefore also estimated the initial endophyte frequency in the seed stocks, an important component of our parameter estimation methods (below). We then buried E+ and E- seed bags, recovered them one year later, and repeated the viability and germination assays. These two measures provide similar but not identical information, because not all viable seeds may germinate. Differentiating these processes is important for understanding seed bank dynamics. Finally, we determined the endophyte status of seedlings that recruited in our germination trials, allowing us to infer endophyte-specific germination rates and imperfect retention of endophytes from seed to seedling

![Flow chart representation of the methods used to estimate seed bank demography](image)

**Figure 2-1 Flow chart representation of the methods used to estimate seed bank demography.** Ovals represent seeds stocks, squares represent the data generation process.
in the lab or the field, and arrows connect the origin of information to a give process. For example, from the initial seed stocks, seeds were assessed for viability and endophyte status, seeds were also used for germination trials, and finally seeds were also used for creating an artificial seed bank.

2.1.3. Initial endophyte prevalence and seed viability.

We simultaneously estimated the endophyte prevalence and initial seed viability of E+ seed stocks by splitting individual seeds in half and screening half for endophyte presence/absence and the other half for seed viability. This allowed us to assign viability and endophyte status to individual seeds. We did this for 50, 49, and 50 seeds from the E+ stocks of *A. hyemalis*, *L. multiflorum*, and *L. arundinaceum*, respectively. For endophyte scoring, seed halves were soaked in NaOH solution overnight (14-16 hours) and then squashed in aniline blue stain with a phenol fixative. Endophyte status can be assessed based on the presence absence of fungal hyphae in the aleurone layer (Bacon and White 1994). We examined seed squashes under a compound microscope (Zeiss Axiostar Plus) at 200-400x magnification. For the other halves of the seeds, viability scores were based on a tetrazolium chloride (TZ) assay (Peters 2000). Seed halves were soaked in water on moist tissue paper overnight. Seeds were then laterally bisected and placed in 0.1% TZ solution for 14-16 hours. Assessment of viability was based on the staining of embryonic tissue as viewed under a dissecting microscope (Peters 2000).

Following the methods above, a screening of E- seed stocks for all three host species indicated that these seeds were 100% E-. We conducted TZ assays and endophyte scores for 50 seeds from the E- stocks of *A. hyemalis* and *L. arundinaceum*, respectively. The original TZ assays conducted on seeds halves of *L. multiflorum* (E- stocks) were
unusable due to human error. That is seeds were incorrectly bisected and the viability assessment of individual embryos was not possible. Due to seed limitations for this stock the viability assessments for E- seeds were derived from E- seeds in the E+ stock (42 seeds). Although the data on the original TZ assays for L. multiflorum were unusable, endophyte scores were not effected and indicated that seeds were 100% E- (48 seeds).

For both E+ and E- stocks, seeds that were unfilled (lacked an embryo) were scored as inviable and not tested with TZ. We could not determine the endophyte status of seeds that were unfilled. However, these data were used as a measure of how many seeds were unfilled during germination trials and seed bank experiments (see below).

2.1.4. Initial germination and endophyte status of seedlings.

We conducted initial germination trials of E+ and E- seeds in fall of 2012, the natural time of seedling recruitment for all three species. Trials were conducted in circular plots (ca. 22 cm in diameter) to which we added 100 seeds from the E+ or E- stocks. Plots were open on top but bordered by a plastic barrier that was 3-4 cm high, to prevent displacement by wind and water. Plots were lightly disturbed and watered at the time of seed addition. Each species / endophyte combination was replicated across 40 plots (8 replicates for each type of A. hyemalis and L. arundinaceum and 4 replicates for each type of L. multiflorum) that were arranged in a grid with ca. 0.5 m spacing. We also included 20 control plots, where no seeds were sown, to account for germination from sources other than our seed additions (we maintained other experiments with these species nearby). Seeds were added on October 11, 2012. Seedlings were counted and
collected weekly until no additional germination was observed. Collected seedlings were transferred into vials containing 95% ethanol.

In the lab, we determined endophyte status for a subset of seedlings that were collected from the germination plots. These seedlings were collected from plots once they had developed a minimum of two leaves. We followed the method from Gundel et al. (2010) to score seedlings for endophytes with the modification that seedlings had been stored in 95% ethanol for a period of two months to one year. After removing seedlings from ethanol, leaf sheaths were taken from the seedling and Rose Bengal stain was applied to the leaf sheaths for approximately ten minutes on a glass slide. After ten minutes, slide covers were placed on top of the slides and endophytes status was assessed with light microscopy (Zeiss Axiostar Plus at 200-400x magnification). Assessment of endophyte status was based on the presence of fungal hyphae found in the first 1-2mm of the leaf sheath. We did this for 15 (88% of seedlings), 44 (48% of seedlings), and 114 seedlings (34% of seedlings) of *A. hyemalis*, *L. multiflorum*, and *L. arundinaceum*, respectively.

### 2.1.5. Seed bank experiments.

We buried E+ and E- seeds of each species in an experimental seed bank. Seeds were buried in mesh bags that were permeable to moisture and microbes but not to seeds. For *L. multiflorum* and *L. arundinaceum*, we used a mildew resistant shade cloth. The mesh size of this material was too large to contain the small seeds of *A. hyemalis*. For this species, seed bags were constructed from nylon mesh (manufactured by Nitex) with a 10µm mesh size.
Seed bags were filled with 100-125 seeds from each seed stock (E+ or E-); the number of seeds that went into each bag was recorded. Due to the impurity of E+ seed stocks, bags identified as “E+” also contained E- seeds. Due to the limited number of seeds available, replication of seed bags varied for each species. For A. hymealis and L. arundinaceum, there were 8 replicate bags of each seed stock (E+ and E-) that were harvested (32 bags total). For L. multiflorum, 4 replicate bags of each seed stock were filled and harvested (8 bags total). The artificial bank was initiated on October 15, 2012, when each bag was buried under 3-5cm of soil and covered with a weed barrier to reduce light penetrance. Bags were randomly distributed with 0.5 m spacing, adjacent to where germination trials and above-ground demography experiments occurred.

Seed bags were harvested from the experimental seed bank approximately ten months after the initiation of the bank, on September 13, 2013. Harvested seed bags were brought into the lab and seeds that had not decayed or germinated were separated and counted. A subset of these seeds were processed for endophyte status (E+ seeds only) and viability (all seeds), following the methods described above. For A. hymealis we processed 304 seeds from E- stocks and 281 seeds from E+ stocks. The remaining seeds recovered for this species were used for germination trials (305 seeds from E- stocks and 282 seeds from E+ stocks) beginning on October 23, 2013, followed the methods described above. For L. arundinaceum and L. multiflorum there were no seeds that had not decayed after ten months in the seed bank. As in the first year, we collected seedlings weekly until no further germination was observed, and we scored 15 seedlings (56% of seedlings) for endophyte status.
2.1.6. Overview of statistical analysis

Complete details of our statistical analysis, including application of hierarchical Bayesian “state-space” modeling, can be found in Appendix A. Here, we provide a brief overview. The premise of the approach is that we can use multiple, independent data sources to infer the values of unobserved or unobservable (latent) states. Through a sequential series of inferences, we ultimately arrive at our parameters of interest: endophyte-specific seed survival, germination, and endophyte retention from seed to seedling. First, we used data on whether seeds contained embryos to generate an expectation for the number of seeds in E+ and E- stocks that were filled and added to each germination plot. Next, we use the data on the initial viability and endophyte prevalence in our seed stocks to generate an expectation for the number of viable E+ and E- seeds that were added to each germination plot. Next, the germination observed in the E- plots (which contained exclusively E- seeds) provided an estimate for the E- germination rate that was conditional on E- seed viability. Based on this estimate and the expected number of viable E- seeds included in E+ germination plots, we generated an inference for the expected number of E- seedlings that should be observed due solely to the germination of E- seeds. Any additional seedlings that were observed must then derive from the germination of viable E+ seeds (an inferred quantity), thus providing an estimate of the E+ germination rate. Any fraction of the seedlings derived from E+ seeds that were observed as E- seedlings must then represent the probability of imperfect retention.

In addition to these basic elements, our analyses included several steps that enhanced the rigor of our estimates (Appendix C). For example, data from control plots
in the germination trials allowed us to account for seed contamination that may have occurred due to movement of seeds between plots or from a nearby experiment involving *L. arundinaceaum*. In addition, we quantified observer error in the designation of seedling endophyte status, and incorporated this source of error into our analysis.

We generated parameter estimates, following the same statistical protocols, for the initial and post-seed bank germination experiments. Our methods generated 95% credible intervals (CI) based on posterior probability distributions for each parameter. Conclusions about the effects of endophyte symbiosis on seed demographic rates were based on overlap of the CIs. We conducted a simulation experiment to verify that our approach yields accurate parameter estimates; these results are shown in Appendix D.

2.1.7. Population model for *Agrostis hyemalis*.

For one host species (*A. hyemalis*), we collected data that allowed us to connect seed bank processes to above-ground plant demography, thus closing the life cycle loop and allowing us to build a stage-structured population model. *A. hyemalis* was also the one host species that showed potential for long-term seed-banking (Results). Data on *A. hyemalis* reproduction and survival, including the rate of vertical transmission from plant to seed, were collected from E+ and E- plants in a field experiment that occurred adjacent to the seed bank and germination experiments. Details regarding the methods for data collection are provided in Appendix E.

The data allowed us to build a matrix-based population projection model (Caswell 2001) for *A. hyemalis*. The population projection is given by $\mathbf{n}_{t+1} = \mathbf{A}\mathbf{n}_t$, where the population vector $\mathbf{n}_t$ consists of E- seeds, E- plants, E+ seeds, and E+ plants in year $t$. The
population vector corresponds to population structure in spring, when above-ground plants are flowering and the seed stages represent seeds in the seed bank that were produced the preceding spring and did not germinate in fall.

The projection matrix $A$ consists of vital rate parameters that govern transition into and out of the seed bank and plant stages. These vital rate parameters can be unique to endophyte-positive (+) and –negative (−) seeds and plants. In the following, we represent endophyte status (+ or −) with a place-holder (‘) for convenience. Each year, plants produce $f^−$ seeds that are viable at $s_0^−$ rate and germinate immediately (at rate $g_0^−$) to become a reproductive plant in the following year. For seeds that do not germinate immediately and enter the seed bank at a rate $1 - g_0^−$, they survive until the next census period (when they are “observed” in the seed bank) at rate $s_b^−$. Plants survive to the next year at a rate $s_p^−$. Finally, seeds in the seed bank germinate to the above-ground plant stage at a rate $g_1^−$. We assume that seeds cannot persist in the seed bank for more than one time step (i.e. one year), consistent with our empirical results.

We allow for two pathways of endophyte loss: imperfect vertical transmission (τ: production of E- seeds from E+ plants) and imperfect retention (φ: production of E- plants from E+ seeds). Further, imperfect retention can occur at the germination of 0-year old seeds (i.e., seeds that by-pass the seed bank) or 1-year old seeds (i.e. seeds that germinate out of the seed bank). We allow these retention rates to differ ($φ_0, φ_1$) since endophytes may lose viability over time ($φ_0 > φ_1$).

Given these symbiont-specific vital rates and symbiont loss parameters (summarized in Table 2-1, results section), the projection matrix is given by
The first eigenvector associated with the leading eigenvalue of $A$ gives the stable population structure (Caswell 2001). We were primarily interested in the equilibrium frequency of symbiosis among above-ground plants, since this is the focal stage of nearly all empirical studies on endophyte frequency. To determine the role of seed banks in the endophyte frequency of above-ground plants, we compared scenarios in which the seed bank operated according to our parameter estimates and a hypothetical scenario in which no seeds survive in the seed bank ($s^+_b = s^-_b = 0$). In addition to point estimates for each parameter, our estimation methods provided a joint posterior probability distribution, allowing us to propagate uncertainty in our demographic parameters to uncertainty in our population-level predictions.

### 2.2. Results

#### 2.2.1. Parameter estimates for seed demographic rates.

Figure 2-2 shows the model estimates for seed survival, seed germination, and symbiont retention before (denoted “Initial”) and after (denoted “Post Seed Bank”) the storage of seeds in the artificial seed bank for all species, with 95% credible intervals.
Initial seed demographics showed no significant effect of symbionts on host vital rate for all species, as measured with a test of overlapping credible intervals (Fig. 2-2A-F; see Table 2-1 for all parameter estimates and 95% credible intervals). Across species initial seed survival was variable (Fig. 2-2A-C), but lowest for the native grass, *A. hyemalis*. Additionally, across species initial seed germination was low to moderate (Fig. 2-2D-F) with the lowest estimates also for the native, *A. hyemalis*.

Post seed banks demographics were estimated only for the native, *A. hyemalis*, as this species was the only one with seeds that did not decay in the seed bank after ten months of storage (Fig. 2-2A,D, G). Endophytes had a significantly positive effect on seed germination post seed bank \( (g_1^+ = 0.45, g_1^- = 0.15) \). However, endophytes did not have a significant effect on seed survival, post seed bank (Fig. 2-2A, Table 1).

Estimates for the initial rates of symbiont retention were highest for *L. arundinaceum* (77%, Fig. 2-2H) and lowest for *A. hyemalis* (43%, Fig. 2-2G). For *A. hyemalis*, storage did not significantly affect retention rates between the initial and post seed bank estimates \( (\phi_0 = 0.43, \phi_1 = 0.53) \).
Figure 2-2 Figure panel showing the means and 95% credible intervals of seed demographics from the initial and post seed bank data, as estimated from our model. Columns are specific to each species (Fig. 1A,D, E: *Agrostis hyemalis*; Fig. 1B,F, H: *Lolium arundinaceum*; Fig. 1C,F, I: *Lolium multiflorum*). Rows corresponds groups of demographic parameters, where row 1 (Fig.2A,B, C), show estimates of seed viability ($s_0^+/-, s_1^+/-$). Row 2 shows estimates of seed germination ($g_0^+/-, g_1^+/-$). Row 3 shows estimates of imperfect retention ($\phi_0, \phi_1$). Colors distinguish endophyte status (black= E+, grey= E-).
Table 2-1 Parameter definitions for the population model, estimated means and 95% credible intervals for all species and host types. Not all vital rates were estimated for every species. Bolded estimates indicate non-overlapping credible intervals and a significant difference between E+ and E- vital rates. An asterisk indicates the quantity was not used in the population model.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Definition</th>
<th>Mean &amp; CI</th>
<th>Mean &amp; CI</th>
<th>Mean &amp; CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>$g_{0}^{+/-}$</td>
<td>The probability of seeds germinating initially after seed rain.</td>
<td>E+ 0.16, 0.06 - 0.46</td>
<td>E+ 0.49, 0.22 - 0.90</td>
<td>E+ 0.44, 0.40 - 0.49</td>
</tr>
<tr>
<td>$g_{1}^{+/-}$</td>
<td>The probability of seeds germinating from the seed bank (after 1 year).</td>
<td>E+ 0.45, 0.33 - 0.61</td>
<td>E+ 0.15, 0.08 - 0.22</td>
<td></td>
</tr>
<tr>
<td>$s_{0}^{+/-}$</td>
<td>The probability of seeds surviving in the seed bank (for 1 year).</td>
<td>E+ 0.19, 0.13 - 0.23</td>
<td>E- 0.23, 0.19 - 0.26</td>
<td></td>
</tr>
<tr>
<td>$s_{1}^{+/-}$</td>
<td>The probability of seeds being viable immediately after reproduction.</td>
<td>E+ 0.19, 0.04 - 0.38</td>
<td>E+ 0.89, 0.66 - 0.99</td>
<td>E+ 0.98, 0.93 - 0.99</td>
</tr>
<tr>
<td>$\phi_{0}$</td>
<td>The probability of seeds retaining endophytes immediately after reproduction.</td>
<td>E+ 0.43, 0.17 - 0.79</td>
<td>E+ 0.56, 0.20 - 0.94</td>
<td>E+ 0.77, 0.66 - 0.86</td>
</tr>
<tr>
<td>$\phi_{1}$</td>
<td>The probability of seeds retaining endophytes from the seed bank (after 1 year).</td>
<td>E+ 0.53, 0.23 - 0.80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tau$</td>
<td>The vertical transmission rate of endophytes (estimated from seed data).</td>
<td>E+ 0.89, 0.84 - 0.93*</td>
<td>E+ 0.18, 0.09 - 0.32</td>
<td>E+ 0.98, 0.93 - 0.99</td>
</tr>
<tr>
<td>$f^{+/-}_{p}$</td>
<td>The fertility rate of plants (weighted by the probability of flowering).</td>
<td>E+ 28.8, 17.4 - 45.1</td>
<td>E- 21.9, 12.0 - 37.7</td>
<td></td>
</tr>
<tr>
<td>$s_{p}^{+/-}$</td>
<td>The probability of survival for plants (flowering and vegetative).</td>
<td>E+ 0.66, 0.37 - 0.83</td>
<td>E- 0.75, 0.56 - 0.85</td>
<td></td>
</tr>
</tbody>
</table>

2.2.2. Parameter estimates for plant demographic rates.

There was support for endophytes effects on the vital rate components that makes up plant fertility and plant survival (see Appendix E), tested with a likelihood ratio test against a null model (see Table 2-1 for vital rate estimates and 95% credible intervals).

Seed production by flowering plants, a component of plant fertility ($f^{+/-}_{p}$), was positively affected by endophytes ($df=1, \chi^2(1)= 62.398, p < 0.0001$). Additionally, the survival probability of flowering plants, a component of plant survival ($s_{p}^{+/-}$; vegetative and flowering plant survival), was negatively affected by endophytes ($df=1, \chi^2(1)= 103.18, p < 0.0001$).
2.2.3. Population-level predictions for *A. hyemalis*

Figure 2-3 shows the joint posterior probability distributions for the equilibrium frequency of symbiosis among above-ground plants under two alternative scenarios. The first scenario is where seed bank dynamics are accounted for with our parameter estimates (solid black line, Fig. 2-3) and the second scenario is under a hypothetical situation where no seeds survive in the seed bank \((s^+ = s^- = 0)\) (solid grey line, Fig. 3). There are two measures of interest from these joint distributions, the probability of symbiont persistence and given persistence, the frequency of symbiont in host populations. Under the first scenario where seed bank dynamics are accounted for the posterior probability of symbiont persistence is 55.4%. Given persistence, the mean equilibrium frequency of symbionts in above-ground plants is 21.8%. Under the alternative scenarios where no seeds survive in the bank, the posterior probability of symbiont persistence is 19.7%. Given persistence under this alternative condition, the mean equilibrium frequency of symbionts in above-ground plants is 22.8%.
Figure 2-3 Joint posterior probability distributions for the equilibrium endophyte frequency in above ground plants under two scenarios. Scenario one, where seed bank dynamics are unaltered (solid black line) and scenario two, the theoretical condition where no seeds survive in the seed bank (solid grey lines). Dashed lines correspond to means of the joint distributions.

2.3. Discussion

Imperfect retention, or the loss of a symbiont during the ontogenetic development of an individual host, is a mechanism that can affect the persistence and prevalence of heritable symbionts. Previous studies have provided evidence for imperfect retention of symbionts (Rolston et al. 1986, Geib et al. 2009, Gundel et al. 2009) however, this is the first study to rigorously estimate rates of symbiont retention from the host storage stage and integrate these results into a demographic framework. We show for the first time that integrating the effects of symbiont loss and the influence of symbionts on a host dormant
stage increases the probability of symbiont persistence compared to scenarios where dynamics from host storage are ignored.

For all host, including those that do not have a dormant stage, it is necessary to account for symbiont loss through imperfect retention when making predictions about symbiont persistence and prevalence. For example, in a demographic model of a symbiotic plant population, the plant-to-plant transitions of symbiotic host types are typically the product of the symbiotic host fecundity rate and the rate of vertical transmission (symbiont retention can be included as a multiplicative term). For any retention rate less than one, failure to include the rate of symbiont retention into the plant-to-plant demographic transition will result in an estimate of symbiont frequencies in host populations that are biased too high. This point is emphasized within our study, where we found evidence that two species, _L. multiflorum_ and _L. arundinaceum_, do not utilize the seed bank. Additionally, estimates of the initial rate of symbiont retention for both species, 56% and 77% respectively, suggest these rates must be accounted for in predictive demographic models. A failure to properly account for symbiont loss will result in biased predictions and overestimates of the probability of symbiont persistence.

Measuring the rates of symbiont retention and the effects of symbionts on individual components of host fitness requires accounting for apparent uncertainty. For example, numerous studies have attempted to estimate symbiont specific germination rates, with results varying over a range from negative to positive (see Clay 1987, Novas et al. 2003, Faeth et al. 2004, Vila-Aiub et al. 2004, Neil et al. 2006, Gundel et al. 2006a, Gundel et al. 2006b, Cleem et al. 2008, Wäli et al. 2009), however none have done so while accounting for the uncertainty of the symbiont frequency within seeds (as
demonstrated here). Had we not accounted for this uncertainty in our estimates of post seed bank germination, we would not have discovered a significant effect of symbionts for *A. hyemalis*. This issue of uncertainty not only deals with incorrect inference but also extends to introducing bias. For example, not accounting for the uncertainty in the starting state of an E- seedling that germinated from an E+ seed stock (because E+ seeds can lose symbionts), can bias estimates for rates of retention. A simple point estimate of retention stemming from the frequency of symbionts in seedlings for *L. multiflorum* yields an 18% estimate. This rate retention is considerably lower than our reported estimate of 56% that accounts for apparent uncertainty. Integrating realistic complexity into predictive models of host-symbiont dynamics will undoubtedly enhance our understanding of these interactions but this requires embracing uncertainty to ensure proper inference and unbiased results.

The expectation that accounting for the loss of symbionts from imperfect retention would promote symbiont extinction was not met in this study. This prediction stems from evidence suggesting that the viability of symbionts decreases faster that the viability of seeds (i.e. storage death, Rolston et al. 1986, Gundel et al. 2009). In our study, for the one species that utilized the seed bank (*A. hyemalis*), we found no evidence that the rate of symbiont retention differed between initial and post seed bank intervals. This result is partially consistent with other findings regarding endophyte viability in seeds. Canals et al. 2008, showed the viability of endophytes in seeds, placed on the soil surface and buried under the soil, did not differ between these two storage conditions, however this was only tested for a period of three months. Another study simulating seed dispersal under field conditions where seeds were maintained above the soil surface
(mimicking reproduction) and on the soil surface (mimicking seed rain), showed a sharp decline in the viability of endophytes in seeds over a period of 150 days (Gundel et al. 2010). Within our study, endophytes survived and transmitted to seedlings after being buried under the soil surface for a period of ten months. Variation in temperature and moisture conditions for seeds on and below the soil surface have been proposed as factors that affect the rate of symbiont retention, but no empirical work has been done to assess this hypothesis (Gundel et al. 2011). A more mechanistic understanding of the abiotic factors that affect rates of symbiont retention can be important for determining under what conditions will imperfect retention from the seed bank negatively affect symbiont persistence and prevalence.

The results from our population model of *A. hyemalis* add to existing studies that demonstrate heritable symbionts function as a net mutualist (Rudgers et al. 2012, Yule et al. 2013). This was demonstrated in our study where host received a net benefit from symbionts, which persisted in the population through the relative balance of positive and negative effects imposed on host fitness components. In our study, the significantly positive effects endophytes had on host seed germination from the seed bank counteracted the significantly negative effect endophytes had on the survival probability of flowering plants (a 10% difference). Although simulating host-symbiont dynamics without seed banks resulted in the persistence of symbionts at similar equilibrium frequencies, we attribute this result to the difference between fertility rates of host types, where E+ plants reproduced on average 28.8 seeds and E- plant produced 21.9 seeds. As seed bank dynamics are removed, the projection matrix resembles a grass life cycle without a dormant stage and with symbionts in the population, where host fecundity,
transmission efficiency, and plant survival determine symbiont prevalence and persistence. Given the result of symbiont persistence without accounting for seed bank dynamics, the criteria for a net mutualist would have already been met however, incorporating loss and symbiont effects on host fitness from dormancy increased the overall net mutualist effect, doubling the probability of symbiont persistence. We suggest a concerted effort be made to integrate realistic demographic complexity into models of host-symbiont dynamics, as we have demonstrated here.

A caveat to our deterministic model is that seed banking is a bet hedging strategy most advantageous in variable environments (Cohen 1966, Ellner 1985). Under stochastic conditions, selection should optimize life history traits that promote the persistence of the population directly experiencing the environment (i.e. the host population). It is possible that in a stochastic model of symbiont persistence, where year effects vary randomly for A. hyemalis, the positive effects of symbionts on host seed germination from the seed bank would not match the selected optimum germination rate of the host. This mismatch between the optimal germination rate of the host and the realized germination rate in the presence of the symbiont could lead to a negative effect on host population growth in the long-term. As an example, for an annual plant population with a seed bank, an evolutionarily stable strategy to ensure population persistence is that germination from the seed bank should reflect the probability of whether a host is experiencing favorable conditions (Cohen 1966, Gremer and Venable 2014). This strategy for the persistence of the host population is in conflict with conditions for the persistence of symbionts under an annual host life cycle and where symbiont viability in the dormant stage is low (although we found no evidence of this). Symbiont persistence should be promoted
through the suppression of symbiont-specific host dormancy (decreasing symbiont-specific seed banking, Afkhami & Rudgers 2008) or by increasing symbiont recruitment out of the dormant stage (i.e. increasing symbiont-specific germination from the seed bank). We found evidence for that latter, where germination fractions were influenced by symbionts. Given that *A. hyemalis* is known to exhibit an annual life cycle (although we included plant survival in our population model), the estimated rate of germination from the seed bank, while beneficial for symbionts persistence, may not necessarily reflect the optimal germination rate from the seed bank for the persistence of the host population itself (in a stochastic environment). More research is needed to address this potential conflict underlying symbiont persistence in a stochastic environment.

### 2.4. Conclusion

Many studies have attributed imperfect vertical transmission as the mechanism maintaining the persistence of symbionts at intermediate frequencies in host populations. However, we argue that imperfect vertical transmission fails to account for loss of symbionts from other demographic transitions within the life cycle of a host and not accounting for these dynamics biases predictions of long-term symbiont persistence. Viewing the transmission process of symbionts through the lens of demography can hold great potential for understanding mechanisms that affect the persistence and prevalence of symbionts in host populations. We introduced the term ‘imperfect retention’ to depict the loss of symbionts from the seed stage of plant population however, there exists many more opportunities for symbiont loss from host not accounted for in our study; areas fruitful for future research. We highlight the importance of understanding biological
factors affecting the transmission process of symbionts if we are to know when and how imperfect retention can alter host-symbiont dynamics. Our results emphasize the importance of integrating a fully demographic framework into the understanding of host-symbiont dynamics while accounting for apparent uncertainty. Emerging from our study is a need for a theoretical treatment that integrates bet-hedging, heritable symbionts, and life history evolution as we have shown potential conflict between host and symbionts can arise by accounting for the dormant stage of symbiotic host populations.
References


Bibian, A.B., R. Patterson, A. Shadow, J.A. Rudgers, T.E.X. Miller, unpublished data.


Appendix A: Exploring the converted seed assumption for an obligate seed dormancy model

A key assumption of our series of models relates to the demographic fate (survival and germination) of seeds that fail to retain their symbiont. We presented a model that depicts either symbiont effects driven by maternal provision or direct physiological effects with endophyte loss occurring at the end of a time step: in either case, “converted” seeds survive with the E+ rate right up to the census period. An alternative scenario is one where endophyte loss occurs at the beginning of a time step, in which case the previously infected seed would survive with an E- rate until the next census period. This latter scenario requires a different parameterization. To understand how the assumption about converted seeds alters our results, we performed a complimentary analysis.

i. Obligate seed dormancy model

The transition matrix that depicts a seed surviving with an E- rate is as follows:

\[
A = \begin{bmatrix}
0 & f^- & 0 & f^+(1 - \tau) \\
0 & 0 & s^-(1 - \phi) & 0 \\
0 & 0 & s^+\phi & 0 \\
0 & 0 & 0 & 0
\end{bmatrix}
\]  \hspace{1cm} (Eq. A1).

The dynamics of this model differ only in the assumption that converted seeds now take on an E- survival rate. We found that under this new parameterization, endophyte persistence is unaffected and equivalent to Eq.4.

Given persistence we can analytically determine their equilibrium frequency in the host population. As before we report the fraction of plants only, as opposed to the fraction of the total population (seeds + plants). With eigenanalysis we found that
\[ E_\mathcal{E} = \begin{cases} \frac{1-F\tau_\phi}{1-F\tau_\phi-F_r\phi} & \text{if } F > \frac{1}{S_r\phi} \\ 0 & \text{otherwise} \end{cases} \] (Eq. A2).

As before, increasing the fertility benefits, seed survival benefits, vertical transmission and retention rates all increase the long-run prevalence of endophyte-positive plants, though the expression differs from Eq. 5.

In contrast to the alternative parameterization, vertical transmission and retention influence endophyte prevalence in the same manner (Fig. A1). Regardless of whether seed survival is negative (Fig. A1a), neutral (Fig. A1b), or positive (Fig. A1c), the same reduction in either vertical transmission or retention rate would have the same effect on endophyte prevalence. This is a result of having both vertical transmission and retention rates altering the survival response of the seed.
Figure A1. Equilibrium endophyte frequencies ($\hat{E}$) predicted by the obligate seed dormancy model (Eq. A2) in relation to the effects of endophytes on host-plant seed production ($F = f^p/f$). Here endophytes lost through imperfect retention now take on an E- seed survival rate. Panels show effects of endophytes on seed survival that are negative (A, $S = s^T/s^* = 0.5$), neutral (B, $S = 1$), and positive (C, $S = 2$). Within each panel, line shading and type represent different levels of endophyte loss from imperfect vertical transmission ($\tau$, gray) or retention ($\phi$, black): dashed, $\tau$ or $\phi = 0.75$; dotted, $\tau$ or $\phi = 0.25$. Solid black lines at $\hat{E} = 1$ show perfect transmission and retention ($\tau = \phi = 1$). Gray and black lines overlap completely for all panels.
Taking the limit of Eq. A2 can quantify how host demography affects endophyte prevalence in adult plants. When symbiont benefits to fertility \((F)\) and seed survival \((S)\) become large we find that:

\[
\lim_{F \to \infty} E^+ = \frac{S\phi\tau}{S\phi\tau + 1 - \phi\tau} \quad \text{(Eq. A3)}
\]

and

\[
\lim_{S \to \infty} E^+ = 1 \quad \text{(Eq. A4)}.
\]

Here, when \(E^+\) seed production advantage is large (Eq. A3), below-ground processes continue to play a role in endophyte prevalence. As before, this can be explained by the obligate passage of seeds through the seed bank, modifying prevalence in above-ground plants. Interestingly, as the \(E^+\) seed survival advantage becomes very large (Eq. A4), endophytes become fixed in the population. This result differs from the alternative parameterization (Eq. 1-6). This difference is the result of retention now altering the survival response of the seed. Under the previous assumption, \(E^+\) seeds that lost their endophyte maintained an \(E^+\) survival rate. Therefore when there was any \(E^+\) seed survival advantage this meant there was an advantage for those \(E^+\) seeds that didn’t retain the endophyte, \((1 - \phi)\), and would recruit as \(E^-\) plants. Retention was limiting the endophyte population from fixation. Now, any \(E^+\) seed survival advantage does not result in an advantage to converted seeds and the endophyte population can reach fixation, even if retention is imperfect.

The demographic fate of converted seeds can have significant affects on symbiont prevalence in host populations. When converted seed take on an \(E^-\) survival rate, retention still plays a role in symbiont prevalence, however, benefits to host seed survival
become a more potent pathway to increase symbiont prevalence. There is a need for understanding the mechanisms, maternal or physiological effects, that operate in natural systems if we want to accurately predict host-symbiont dynamics.
Appendix B: Exploring the converted seed assumption for a facultative seed dormancy model

As stated before a key assumption of our series of models relates to the demographic fate (survival and germination) of seeds that fail to retain their symbiont. We explored an alternative scenario where symbiont loss occurs at the beginning of a time step, and converted seed survive with an E-rate until the next census period for an obligate seed dormancy model (Appendix A). Here we extend this scenario to include facultative entry into the storage stage. To understand how the assumption about converted seeds alters our results regarding the effects of seed-banking on symbiont persistence and prevalence, we performed a complimentary analysis.

i. Facultative seed dormancy model

The transition matrix that depicts a seed surviving with an E-rate is as follows:

\[
A = \begin{bmatrix}
0 & f^- b & 0 & f^+ b (1 - \tau) \\
0 & f^- (1 - b) & s^- (1 - \phi) & f^+ (1 - b) (1 - \tau) \\
0 & 0 & 0 & f^+ b \tau \\
0 & 0 & s^+ \phi & f^+ (1 - b) \tau
\end{bmatrix} \quad \text{(Eq. B1)}.
\]

This model differs only with respect to the demographic fate of converted seeds. For clarity we will use the term “E+ assumption” to refer to case where converted seeds maintain an E+ seed survival rate (Eq. 9) and “E- assumption” to refer to this new parameterization where converted seeds survive with an E-rate (Eq.B1). This model is analytically intractable and we assessed symbiont persistence and prevalence after 1000 generations of simulated population dynamics. We found that this parameterization does not alter the condition for persistence and this appendix only regards symbiont prevalence.
As before, most effects of seed-banking on symbiont prevalence were still negative. Under this new parameterization, when loss is through vertical transmission, symbiont prevalence remains unchanged for all scenarios considered. Changes in symbiont prevalence attributed to the E- assumption are only seen when loss is through retention (Fig. B1). For example, a cost to seed survival ($S < 1$), results in seed-banking decreasing the prevalence of symbionts relative to the E+ assumption i.e. grey lines depicting the E- assumption are lower than the black lines depicting the alternative. This effect, while small, holds when there are both neutral (Fig. B1B) and positive (Fig.B1C) effects on host fertility. The opposite is true when symbionts positively affected seed survival ($S > 1$) and prevalence increased dramatically relative to the E+ assumption (Fig. B1). Again, these effects were present when fertility effects were negative (Fig.B1D), neutral (Fig. B1E), or positive (Fig. B1F). Biologically, this is the result of retention altering the survival response of the seed. Under the E+ assumption, any (dis-) advantage to E+ seeds resulted in an (dis-) advantage to converted E- seeds that recruited as E- plants. Now, under the E- assumption, any E+ (dis-) advantage will result in altered symbiont prevalence because there is no corresponding (dis-) advantage to converted E- seeds that recruit as E- plants.

Having converted seeds take on an E- survival rate ensures that when symbiont loss occurs through retention, no carry over effects between symbiotic and non-symbiotic demographic rates will affect symbiont prevalence. Whether maternal provisioning or physiological effects and the timing of symbiont loss from seeds govern these assumptions in natural populations remains to be explored.
Figure B1. Comparing how the demographic fate of converted seed affects symbiont prevalence we shown the equilibrium endophyte frequencies ($E^\pm$) in relation to the probability of banking ($b$) under the “E- assumption” (grey lines) and the “E+ assumption” (black lines). Panels show endophyte effects on fertility that are negative, left column ($F = f^+/f^- = 0.5$, A,D), neutral, center column ($F = f^+/f^- = 1$, B,E), and positive, right column ($F = f^+/f^- = 2$, C,G). Rows represents endophyte effects on seed survival that are negative, top row, ($S = s^+/s^- = 0.5$) and positive, bottom, row ($S = s^+/s^- = 5$). Line types represent different levels of loss through imperfect retention: solid lines, $\phi = 1$; dashed, $\phi = 0.75$; dotted, $\phi = 0.25$. 
Appendix C: Overview of the statistical approach for Bayesian estimation of latent processes

State-space Models

To lay the foundation for our statistical approach of Bayesian estimation of latent processes we must first develop the concept of a state-space model. A state-space model is characterized as two time series running in parallel. One series, referred to as the observation process, describes realizations of a true ecological state. For example, abundance data on organisms are observations indicative of the true ecological state of the system i.e. the true abundance. The second series, referred to as the state process, describes the true ecological state that can never actually be observed (i.e. latent processes) (Parent, 2013). In a model, these two states interact such that the unobservable state process is inferred with observation level data. This interaction allows population level parameters to be estimated using both observations and inferred ecological states. The state-space framework has been used to model the dynamics of wild animal populations (Buckland et al. 2004), and their flexibility renders them useful for the analysis of experimental data (De Valpine 2003). Note, all analyses were performed using Bayesian MCMC methods and analyzed with the package R2jags, where uninformative priors were used throughout the model.

Utility of state-space models

The state-space model of our research proves useful for two reasons that we emphasize here. The first is we can derive conditional probabilities of vital rates to
populate the elements of our projection matrix (see manuscript). For example, germination from a seed bank at time, \( t \), is conditional on seeds surviving in the seed bank from time, \( t-1 \) to \( t \). Therefore, a direct calculation of germination probabilities without accounting for the conditional state of a seed surviving in the seed bank will generate biased estimates of vital rates and propagate to error in population dynamics.

The second reason our approach is useful, is that in order to derive a subset of probabilities of vital rates we must infer latent processes specific to symbiotic host. For example, when germinating seeds from an E+ stock, which contain both E+ and E- seeds, we must infer the true number of E+ seeds germinating so that the rate of germination, specific to E+ host, is not biased by the presence of germinating E- seeds in the E+ stock. These two requirements necessary for understanding the role of seed banks in host-symbiont dynamics can be met through the use of state-space models. It should be noted that parameter estimation through latent inference is only necessary when E+ seed stocks are not pure (i.e. 100% endophyte frequency). Should E+ seed stocks be pure, as with \( L. arundinaceum \), then the state-space framework is unnecessary. However, for \( A. hyemalis \) and \( L. multiflorum \) we required the use of state-space models and in the following paragraphs we demonstrate the utility of this approach for estimating conditional probabilities through latent inference, including a statistical treatment of our methods.

\textit{Inference, latent states, and conditional probabilities}

The estimation of a conditional probability through latent inference can be understood by considering the process by which seeds germinate into seedlings. For context, consider this process specific to E- seed stocks, where there is absolute certainty
that all seeds are of one host type (see Methods). The germination process can be decomposed into a series of observable states, two of which are the initial state (i.e. the known number of seeds added to germination plots) and the final state (i.e. the numbers of seedlings that germinated). Estimating a germination rate from these two states, alone, would yield an unconditional germination probability (i.e. it does not account for other processes). However, we can account for additional processes that actually occur in nature and are necessary for the germination of seeds. For example, seeds that are produced in nature must be intact (i.e. have a developed embryo) and also be viable in order to germinate into seedlings. These states of being intact and viable are observable from independent observations of seed data (these data were collected from our initial seed stocks and from our seed bags). By using the information about the initial state of seeds and all independent observation states, inferences about the number of intact and viable seeds from the initial state can be made. This inference is used to calculate the rate at which seedlings emerged from plots i.e. the conditional germination probability (conditional on intactness and viability).

*Application of state-space model to data*

Figure C1 serves as a visual guide to modeling the process of seeds from E-stocks germinating into E-seedlings. In this figure, squares indicate observation processes, ovals represent state processes (inferences), and solid arrows dictates causality. The green coloring indicates this is specific to the E-seed stocks. For clarity, we will systematically decompose this figure and describe the observation processes, the
state processes, finishing with a statistical formulation for parameter estimation. See table C1 for definition of all parameters that follow.

**Figure C1.** This figure is a visual representation of the state-space model describing the process of seeds from E- stocks germinating into E- seedlings. Squares are observation processes (observed quantities or probabilities) and ovals are state processes (inferences). Green coloration indicates this process is specific to E- host. Solid arrows dictate causality. State variable definitions are included beneath the variable notation.

In figure C1 there are two types of observation process. One type represents directly observed quantities i.e. \(y^I^-\), the initial known number of E- seeds added to germination plots and \(y^F^-\) the final number of E- seedlings that germinated. The other type of observation process represents estimated probabilities from independently observed data i.e. \(i^-\), the probability of a seed from an E- stock being intact and \(s^-\), the probability of a seed from an E- stock being viable. These estimated probabilities are derived from independent data where screening of individual seeds occurred (see 2.1.3 Methods), and where probabilities can be estimated with the follow equations,

\[y^I^- \sim Bernoulli(i^-)\]  
Eq. C1
Here, $y^I_j$, is the number of individual E-seeds ($j$) that were screened for intactness ($i$) where, $i^I$, is the estimated probability that a seed from an E-stock will be intact (i.e. have a developed embryo). Similarly, $y^S_j$, is the number of intact individual E-seeds ($j$) that were screened for viability ($s$) where, $s^I$, is the estimated probability that a seed from an E-stock will be viable. Note, all other observation processes that represent estimated probabilities within this framework were generated with a similar statistical treatment. In aggregate both types of observation processes, observed quantities and estimated probabilities, facilitate inference for state processes (latent states) and parameter estimation.

Regarding state processes, in Figure C1 there are two types. The first type of state process represents inferred latent states i.e. $n^I$, the inferred number of intact seeds given the initial number of E-seeds added to germination plots and $n^S$, the inferred number of viable seeds given the inferred number of intact seeds. The second type of state process represents an inferred parameter i.e. $g^I$, the estimated probability of germination from an E-stock conditional on E-seeds being intact and viable. Given that state processes represent binary states (e.g. intact or not), the causal relations (arrows in Figure C1) are a function of the binomial probability distribution such that state processes can be determined from the following expressions,

\[ n^I \sim Binomial(i^I, y^I) \quad \text{Eq. C4} \]

\[ n^S \sim Binomial(s^I, n^I) \quad \text{Eq. C5} \]
\[ y^F - \sim Binomial(g^-, n^s^-) \quad Eq. C6. \]

Here, the observation processes \((i^-, s^-)\) generate inferences from the initial state of seeds put in a plot \((y^I^-)\) and the final state of the number of seedlings emerged \((y^F^-)\) to model the entire process of germination from an E- stock, where, \(g^-\) is an estimated parameter.

This framework outlined was expanded to encompass estimation of latent host-symbiont interactions. To estimate the conditional germination rate for E+ host, we integrated observation states from E- seed stocks into the state-space model for E+ host. This integration of states is an underlying assumption of our model. It assumes that there is no difference between the seed survival rate of E- seeds from E+ stocks and E- seeds from E- stocks. For \(A.\ hyemalis\) we tested this assumption using a likelihood ratio test, assessing differences in seed survival of E- seeds from E- stocks and E+ stocks. We found no significant difference in seed survival of E- seeds from these stocks \((df=2, \chi^2(2)=0.51, p=0.78)\), validating our model assumption. For \(L.\ multiflorum\) the seed viability assumption could not be tested, as the only data from the initial screening of seed viability of E- stocks for this species were unusable due to human error (seeds were bisected incorrectly; see 2.1.3 Methods) and initial seed viability estimates had to come from E+ stocks. Additionally, this integration of state space also assumes there is no difference between germination of E- seeds from E- stocks and E+ stocks, an assumption that could not be tested with this framework outlined. This assumption of germination enables an inference for E- seed germination in E+ stocks, a requirement necessary to infer subsequent state processes.
Figure C2 serves as a visual guide to model the process of seeds from E+ stocks germinating into seedlings. Again, squares represent observation processes and ovals represent state processes (inferences). Purple coloring indicates states relating to E+ seed stocks, solid arrows dictate causality based on probability distributions, and dashed arrows represent deterministic functions. Green coloring indicates states relating to E- seed stocks. Lastly, grey ovals indicate state processes for which E+ parameters will be estimated i.e. retention ($\phi$) and E+ germination ($g^+$). Inference in this model follows the form of the expressions outlined above however the statistical formulation is not included for the sake of brevity. See table C1 for definitions of all states in Figure C2 as well as definitions for causal links among state processes.

Figure C2. This figure is a visual representation of the state-space model describing the process of seeds from E+ stock germinating into seedlings. Squares are observation processes (observed quantities or probabilities) and ovals are state processes (inferences). Colors are specific to E+ (purple) and E- (green) host processes. Grey ovals represent state processes that E+ parameters of germination ($g^+$) and retention ($\phi$) will be estimated. Solid arrows dictate causality and dashed arrows represent deterministic functions. Boxes indicate how many host types are in the encompassed states. See Table A1 for state variable definitions as well as defined causal links.
Table C1. Variable and causal link definitions for estimating seed bank parameters within the state-space model. Seed stock column indicates which seed stock observation and state processes are specific to. See figure C2 for the corresponding state-space model structure.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Process Definition</th>
<th>Causal link</th>
<th>Seed Stock</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y^E$</td>
<td>Observation process</td>
<td>N/A</td>
<td>E- Seed Stock</td>
</tr>
<tr>
<td>$Y^F$</td>
<td>Observation process</td>
<td>N/A</td>
<td>E- Seed Stock</td>
</tr>
<tr>
<td>$i^E$</td>
<td>Observation process</td>
<td>N/A</td>
<td>E- Seed Stock</td>
</tr>
<tr>
<td>$i^F$</td>
<td>Observation process</td>
<td>N/A</td>
<td>E- Seed Stock</td>
</tr>
<tr>
<td>$S^E$</td>
<td>Observation process</td>
<td>N/A</td>
<td>E- Seed Stock</td>
</tr>
<tr>
<td>$S^F$</td>
<td>Observation process</td>
<td>N/A</td>
<td>E- Seed Stock</td>
</tr>
<tr>
<td>$n^E$</td>
<td>Inference for the number of intact E- seeds added to E- plots</td>
<td>Binomial</td>
<td>E- Seed Stock</td>
</tr>
<tr>
<td>$n^F$</td>
<td>Inference for the number of viable E- seeds that are intact in E- plots</td>
<td>Binomial</td>
<td>E- Seed Stock</td>
</tr>
<tr>
<td>$g^E$</td>
<td>Parameter estimate for E- host germination</td>
<td>Binomial</td>
<td>E- Seed Stock</td>
</tr>
<tr>
<td>$I^E$</td>
<td>Observation process</td>
<td>N/A</td>
<td>E+ Seed Stock</td>
</tr>
<tr>
<td>$I^F$</td>
<td>Observation process</td>
<td>N/A</td>
<td>E+ Seed Stock</td>
</tr>
<tr>
<td>$I^*$</td>
<td>Observation process</td>
<td>N/A</td>
<td>E+ Seed Stock</td>
</tr>
<tr>
<td>$S^*$</td>
<td>Observation process</td>
<td>N/A</td>
<td>E+ Seed Stock</td>
</tr>
<tr>
<td>$t^E$</td>
<td>Probability of a seed from an E- stock being intact</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t^F$</td>
<td>Probability of a seed from an E- stock being intact</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$s^E$</td>
<td>Probability of a seed from an E- stock being viable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$s^F$</td>
<td>Probability of a seed from an E- stock being viable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n^E$</td>
<td>Inference for the number of intact seeds added to E- plots</td>
<td>Binomial</td>
<td>E+ Seed Stock</td>
</tr>
<tr>
<td>$n^F$</td>
<td>Inference for the number of viable E- seeds that are intact in E- plots</td>
<td>Binomial</td>
<td>E+ Seed Stock</td>
</tr>
<tr>
<td>$g^E$</td>
<td>Parameter estimate for E- host germination</td>
<td>Binomial</td>
<td>E+ Seed Stock</td>
</tr>
<tr>
<td>$I^*$</td>
<td>Observation process</td>
<td>N/A</td>
<td>E+ Seed Stock</td>
</tr>
<tr>
<td>$I^*$</td>
<td>Observation process</td>
<td>N/A</td>
<td>E+ Seed Stock</td>
</tr>
<tr>
<td>$I^*$</td>
<td>Observation process</td>
<td>N/A</td>
<td>E+ Seed Stock</td>
</tr>
<tr>
<td>$S^*$</td>
<td>Observation process</td>
<td>N/A</td>
<td>E+ Seed Stock</td>
</tr>
<tr>
<td>$t^E$</td>
<td>Probability of a seed from an E+ stock being intact</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t^F$</td>
<td>Probability of a seed from an E+ stock being intact</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$s^E$</td>
<td>Probability of a seed from an E+ stock having an endophyte</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$s^F$</td>
<td>Probability of a seed from an E+ stock having an endophyte</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n^E$</td>
<td>Inference for the number of intact E- seeds that are intact in E+ plots</td>
<td>Binomial</td>
<td>E+ Seed Stock</td>
</tr>
<tr>
<td>$n^F$</td>
<td>Inference for the number of viable E- seeds that are intact in E+ plots</td>
<td>Binomial</td>
<td>E+ Seed Stock</td>
</tr>
<tr>
<td>$g^E$</td>
<td>Parameter estimate for E- host germination</td>
<td>Binomial</td>
<td>E+ Seed Stock</td>
</tr>
<tr>
<td>$φ$</td>
<td>Parameter estimate for retention of endophytes</td>
<td>Binomial</td>
<td>E+ Seed Stock</td>
</tr>
</tbody>
</table>
Figures C1 and C2 and the series of equations above, describe the state-space model to estimate seed bank parameters for host types of \textit{A. hyemalis}. This state-space framework was implemented on data from the initial seed stocks and data from the seeds, post seed bank storage. In doing this we generated estimates for the following parameters of our matrix population model: initial and post seed bank seed survival \((s_0^{+/-}, s_1^{+/-})\), seed germination \((g_0^{+/-}, g_1^{+/-})\), and retention \((\phi_0, \phi_1)\). Note, the accuracy and precision of parameter estimation through latent-inference of host-symbiont dynamics have been determined with stochastic simulations (see appendix D).

To estimate the final parameter of our population model associated with the storage stage of our hosts, we had to incorporate another set of data relating the quantity of seeds placed in seed bank bags and the quantity of seeds recovered from the bags post storage. The integration of this data into the state-space framework allows us to estimate a rate of seed decay within the seed bank, specific to each host type. The estimate of the rate of seed decay is needed to generate an inference for the probability of seeds surviving within the seed bank (denoted in our population model as \(s_b^{+/-}\)). Figure C3 shows a visual representation for the state-space model used to estimate seed decay within the seed bank \((\psi^{+/-})\). Again, squares represent observation processes and ovals respresent state processes (inferences). Purple coloring indicates states relating to E+ seed stocks, solid arrows dictate causality based on probability distributions, and dashed arrows represent deterministic functions. Green coloring indicates states relating to E- seed stocks. Lastly, grey ovals indicate state processes from which our rate of seed decay was estimated. See table C2 for definitions of state variables and causal links.
Figure C3. This figure is a visual representation of the state-space model to estimate the probability of seed decay in the seed bank, for both host types ($\psi^+/\psi^-$). Squares are observation processes (observed quantities or probabilities) and ovals are state processes (inferences). Colors are specific to E+ (purple) and E- (green) host processes. Grey ovals represent state processes that parameters ($\psi^+/\psi^-$) will be estimated from. Solid arrows dictate causality. See Table A2 for state variable definitions as well as defined causal links.
Table C2. Variable and causal link definitions for the state-space model used to estimate seed decay within the seed bank. Seed stock column indicates which seed stock observation and state processes are specific to. See figure A4 for the corresponding state-space model structure.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Process Definition</th>
<th>Causal link</th>
<th>Seed Stock</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y^E )</td>
<td>Observation process</td>
<td>N/A</td>
<td>E- Seed Stock</td>
</tr>
<tr>
<td>( y^R )</td>
<td>Observation process</td>
<td>N/A</td>
<td>E- Seed Stock</td>
</tr>
<tr>
<td>( \hat{y} )</td>
<td>Final number of E- seeds recovered from E- seed bags</td>
<td>Observation process</td>
<td>N/A</td>
</tr>
<tr>
<td>( \hat{i}^{(E)}_{t=0} )</td>
<td>Probability of a seed from an E- stock being intact during the initial screening</td>
<td>Observation process</td>
<td>N/A</td>
</tr>
<tr>
<td>( \hat{i}^{(E)}_{t=1} )</td>
<td>Inference for the number of intact E- seeds added to E- seed bags</td>
<td>State Process</td>
<td>Binomial</td>
</tr>
<tr>
<td>( \hat{n}^{(E)}_{t=0} )</td>
<td>Inference for the number of intact E- seeds recovered from E- seed bags</td>
<td>State Process</td>
<td>Binomial</td>
</tr>
<tr>
<td>( \psi^E )</td>
<td>Estimate for the probability of an E- seed decaying in the seed bank</td>
<td>Parameter</td>
<td>Binomial</td>
</tr>
<tr>
<td>( y^+ )</td>
<td>Observation process</td>
<td>N/A</td>
<td>E+ Seed Stock</td>
</tr>
<tr>
<td>( \hat{y}^+ )</td>
<td>Initial number of seeds from an E+ stock added to E+ seed bags</td>
<td>Observation process</td>
<td>N/A</td>
</tr>
<tr>
<td>( \hat{y}^+ )</td>
<td>Final number of seeds from an E+ stock recovered from E+ seed bags</td>
<td>Observation process</td>
<td>N/A</td>
</tr>
<tr>
<td>( \hat{i}^+_{t=0} )</td>
<td>Probability of a seed from an E+ stock being intact during the initial screening</td>
<td>Observation process</td>
<td>N/A</td>
</tr>
<tr>
<td>( \hat{i}^+_{t=1} )</td>
<td>Probability of a seed from an E+ stock being intact during the screening post seed bank</td>
<td>Observation process</td>
<td>N/A</td>
</tr>
<tr>
<td>( \tau^{seed}_{t=0} )</td>
<td>Probability of a seed from an E+ stock having an endophyte during the initial screening</td>
<td>Observation process</td>
<td>N/A</td>
</tr>
<tr>
<td>( \tau^{seed}_{t=1} )</td>
<td>Probability of a seed from an E+ stock having an endophyte during the screening post seed bank</td>
<td>Observation process</td>
<td>N/A</td>
</tr>
<tr>
<td>( \hat{n}^{(E)}_{t=0} )</td>
<td>Inference for the number of intact seeds from an E+ stock added to E+ seed bags</td>
<td>State Process</td>
<td>Binomial</td>
</tr>
<tr>
<td>( \hat{n}^{(E)}_{t=1} )</td>
<td>Inference for the number of intact seeds from an E+ stock recovered from E+ seed bags</td>
<td>State Process</td>
<td>Binomial</td>
</tr>
<tr>
<td>( \tau^{(E)} )</td>
<td>Inference for the number of E+ seeds in the E+ stock added to E+ seed bags</td>
<td>State Process</td>
<td>Binomial</td>
</tr>
<tr>
<td>( \tau^{(R)} )</td>
<td>Inference for the number of E+ seeds in the E+ stock recovered from E+ seed bags</td>
<td>State Process</td>
<td>Binomial</td>
</tr>
<tr>
<td>( \psi^+ )</td>
<td>Estimate for the probability of an E+ seed decaying in the seed bank</td>
<td>Parameter</td>
<td>Binomial</td>
</tr>
</tbody>
</table>
After quantifying seed decay within the bank the derived probability of seeds surviving within the bank was determined as follows,

\[ s_b^+ = \psi^+ \times s_1^+ \quad \text{Eq. C7} \]
\[ s_b^- = \psi^- \times s_1^- \quad \text{Eq. C8}. \]

Here seed bank survival \((s_b^{+/−})\) is a product of seed decay \((\psi^{+/−})\) and post seed bank seed viability \((s_1^{+/−})\). It is this derived quantity of seed bank survival \((s_b^{+/−})\) that was used to parameterize our population model.

In addition to the above process, we performed a correction on our observed seedling data to account for seedlings that emerged in the control plots. Note, seedling emergence in control plots only occurred during the initial germination trials when simultaneous experiments were being performed on \textit{L. arundinaceum} adjacent to our field germination trials. Subsequently, this correction only applies to \textit{L. arundinaceum} and \textit{L. multiflorum} as we could not distinguish between these species from their seedling morphology. These data in the control plots were overdispersed, as indicated by model comparison and posterior-predictive checks. Thus, we used a negative binomial distribution to describe this process and incorporate it into a correction for our observations (implanted in R2jags as a Gamma-Poisson, but presented as a Negative-Binomial for clarity)(Hilbe 2011). The following equation was used to model these data,

\[ y_f^c \sim \text{Negative Binomial}(\lambda, \alpha) \quad \text{Eq. C9}. \]
Here, $y_f^c$, is the number of seedlings emerged within every replicated control plot ($j$) during the initial germination trial. The mean seedling emergence in the control plots ($\lambda$) is estimated from the data, as well as the overdispersion parameter ($\alpha$). Uninformative priors were specified. Using the parameters of the negative binomial, realization for every replicate germination plot for the initial germination period were simulated and this simulated quantity was used to correct germination data prior to utilizing the state-space framework.
Appendix D: Assessing the accuracy and precision of Bayesian estimation for seed bank demographics

Our parameter estimates were based on the specification of a state-space model that describes the process of a batch of seeds germinating into seedlings. The accuracy and precision of this method for estimating parameters will be shown using stochastic simulations. For clarity, our simulations are a proof of concept and do not mimic the full experimental design as stated in the corresponding manuscript.

To convey our simulation we start with clarification of what is necessary to simulate the process of seeds germinating into seedlings. Demographic parameters must be defined for seed survival, seed germination, vertical transmission, and symbiont retention with seed survival and germination specific to each host type. Once these parameters are defined, we can stochastically simulate the germination process with a series of expressions that will generate replicated data used to recover parameters estimates, following our novel method. The demographic parameters that were hard coded in the simulation below (i.e. did not change) are as follow: E+ seed survival \( s^+ = 0.20 \), E- seed survival \( s^- = 0.20 \), E- germination \( g^- = 0.25 \), and vertical transmission \( \tau = 0.90 \). Given the state-space framework was implemented to estimate E+ germination \( g^+ \) and retention \( \phi \), these parameters vary from simulation to simulation to demonstrates the ability to recover these parameters over a range of values (i.e. assess accuracy and precision).
In the following paragraphs we give the expressions necessary for the simulating all data emphasizing what can and can not be observed in field and lab experiments. We start with the expressions to generate data for seed germination and methods to produce data on the frequency of symbionts in seedlings. We then proceed to give expressions to generate data on the frequency of symbionts in seeds and finally, we include the expression to generate seed viability data for each host type. The results section conveys the ability of the state-space framework to estimate the seed demographics of retention and E+ germination.

**Data set for seed germination & the symbiont frequency in seedlings**

We assume the germination process follows a set of probabilities that decomposes germination into a series of individual states. For example, seeds produced by an E+ plant have an endophyte frequency equivalent to the rate of vertical transmission, $\tau$. These E+ seeds go on to survive at a specified rate, $s^+$, where surviving seeds germinate at another specified rate, $g^+$, and finally germinated seedling retain endophytes with a specified probability, $\phi$. This sequence describing germination is a first order Markov process that is simulated with the following equations:

\[
N_i^\tau \sim \text{Binomial} (\tau N_i) \quad \text{Eq. D1}
\]

\[
N_i^{s^+} \sim \text{Binomial} (s^+, N_i^\tau) \quad \text{Eq. D2}
\]

\[
N_i^{g^+} \sim \text{Binomial} (g^+, N_i^{s^+}) \quad \text{Eq. D3}
\]

\[
N_i^\phi \sim \text{Binomial} (\phi N_i^{g^+}) \quad \text{Eq. D4}.
\]
Note, bolded variables indicate the quantity is observable with experimental data. Here, $N_i$, represents the number of seeds set to germinate, where $N=100$ and $i$, is the number of replicated seed sets ($i=8$ in the simulation, as with our experimental design).

Superscripts correspond to the state of the batch of seeds: $N_i^T$, is the number of E+ seeds is in the E+ batch; $N_i^{s+}$, is the number of surviving E+ seeds; $N_i^{g+}$, is the number of germinated E+ seedlings; $N_i^\phi$, is the number of E+ seedlings that retained the endophyte.

If the endophyte frequency in seeds is less than one ($\tau < 1$) then we must account for the E- seeds in the starting batch of E+ seeds simulated to germinate. These E- seeds produced by E+ plants are indicated by a “-/+” superscript. The quantity for the number of E- seeds produced by E+ plants was simulated with the following expression,

$$N_i^{-/+} = N_i - N_i^T$$  Eq. D5

$$N_i^{s-/+} \sim Binomial (s^-, N_i^{-/+})$$  Eq. D6

$$N_i^{g-/+} \sim Binomial (g^-, N_i^{s-/+})$$  Eq. D7.

Here, $N_i^{-/+}$, is the number of E- seeds that were present in the batch to germinate but did not have an endophyte vertically transmitted. The surviving E- seeds ($N_i^{s-/+}$) and germinating E- seedlings ($N_i^{g-/+}$) were simulated from a binomial distribution with E-specific probability of seed survival ($s^-$) and germination ($g^-$).

The germination process results in two numerical values that can be observed in the field. The first value, is the total number of E+ seedlings that retained their endophyte
(\(N_i^\phi\)) and the total number of E- seedlings that started out as E- seed or lost their endophyte through imperfect retention (the latter are termed converted seedlings). By scoring seedlings for endophytes, observations of these quantities can be obtained. The two processes that generate the total number of E- seedlings is simulated following the expression,

\[ N_i^{c,e} = N_i^{e+} - N_i^\phi \quad \text{Eq. D8} \]

\[ N_i^{l+/-} = N_i^{c,e} + N_i^{e/-} \quad \text{Eq. D9}. \]

Where, \(N_i^{c,e}\), is the total number of seedlings that did not retain their endophyte (i.e. they were converted from E+ to E- states) and \(N_i^{l+/-}\), is the total number of E- seedling that started out as E- or were converted. Again, the bolded values indicate observable quantities.

To generate data regarding the frequency of endophytes in seedlings, we expanded the simulated quantities of the number of E- seedlings (\(N_i^{l+/-}\)) and E+ seedlings that retained their endophyte (\(N_i^\phi\)) into individual objects on a per replicate basis (\(i\)) and randomly drew samples to generate retention data relating the frequency of symbionts in E+ seedlings, on a per replicate basis.

As with our experimental design, we accounted for control E- seed batches by simulating germination data setting \(\tau = 0\) following equations D1-D7. These simulated quantities that are observable in the field were generated by simulated germination
processes for E- seeds where the number of E- seeds set to germination, $N_i^-$, was set at 100. This simulation yields data reflecting E- host germination rates.

*Data set for the symbiont frequency in seeds*

The endophyte frequency in seeds is directly observable in the lab and can be simulated using a Binomial distribution. To mimic the experimental design, we simulated the endophyte frequency in seeds with the following equation:

$$N^+ \sim \text{Binomial} (\tau, N)$$  \hspace{1cm} \text{Eq. D10.}

Where, $N$, is the number of individual trials ($N=50$) that were performed to assess endophyte frequency and, $N^+$, is the number of individual trials that resulted in a success i.e. endophytes were vertically transmitted with probability ($\tau$)

*Data set for seed viability*

Seed viability of E+ and E- seed is directly observable in lab as well and can be simulated using a Binomial distribution. We simulated seed viability following the equations:

$$N^{s+} \sim \text{Binomial} (s^+, N^+)$$  \hspace{1cm} \text{Eq. D11}

$$N^{s-} \sim \text{Binomial} (s^-, N^-)$$  \hspace{1cm} \text{Eq. D12.}

Here, $N^+$, is the number of E+ seeds assessed for viability (the result of Eq. D10) and $N^{s+}$, is the number of viable E+ seeds simulated under the E+ seed survival rate, $s^+$. The
same process is applied to a number of E- seeds ($N^-$, set to 50 seeds) simulated through the viability process resulting in an observation of viable E- seeds ($N^s$).

Assessing accuracy and precision of the state-space model

With Eq. D1-D12, we simulated multiple data sets varying retention ($\phi$) or E+ germination ($g^+$) and assessed the accuracy and precision of the state space method to recover these parameters simultaneously. The state-space model used to recover the parameter can be gleaned from the corresponding R script for this appendix.

Results

Figures D1 and D2 shows the accuracy and precision of the state-space model for estimating E+ germination ($g^+$) and endophyte retention ($\phi$). Each panel represents the ability of the model to recover parameter estimate from the simulated data. Results are shown where either E+ germination ($g^+$, Fig.D1) or endophyte retention ($\phi$, Fig.D2) was fixed (at one value, noted in panel titles), while the simulation generated data for 100 values of the other parameter. For example, in Fig. D1A & D1B, the parameter for E+ germination was fixed at $g^+ = 0.2$, while data was simulated for 100 values of retention ($\phi$). All parameters were then recovered simultaneously, using the generated data sets, to estimate subsequent means (red dot) and credible intervals (black lines). Note, points in the second column, where one value was re-estimated 100 times ($g^+$ in Fig. D1B), are jittered horizontally to maximize the display of information.

The state-space framework can lead to accurate estimation of parameter for most combinations of parameter values. However, we have show a consistent downward bias
associated with estimating high rates of E+ germination (see Fig. B2F). Recovering estiamtes for symbiont retention when germination is extremely low can additionally lead to bias in parameter estimates (Fig. B2F).

Figure D1. Recovery of parameter estimates from simulated data using the state-space framework. Two parameters, other than the hard coded ones specified in appendix D, were varied to simulate whole data sets. Here, E+ germination, \( g^+ \), was varied three different times, corresponding to rows in the figure (row 1, A: \( g^+ = 0.2 \), row 2, C: \( g^+ = 0.5 \), row 3, E: \( g^+ = 0.9 \)). After these values were specified in the simulation, data sets were simulated with values of retention (\( \phi \)) from, 0.01-1.0, one hundred times. Column 1
shows the means (red dots) and 95% credible intervals (black lines) for the ability of the model to recover retention estimate over the parameters space from 0.01-1.0. Column 2 shows the mean (red dots) and 95% credible intervals (vertical black lines) for the ability of the model to simultaneously recover E+ germination estimates (B: $g^+ = 0.2$, D: $g^+ = 0.5$, F: $g^+ = 0.9$); note, points are horizontally jittered in column 2.

![Figure D2. Recovery of parameter estimates from simulated data using the state-space framework.](image-url)

Two parameters, other than the hard coded ones specified in the appendix, were varied to simulate whole data sets. Here, retention, $\phi$, was varied three different times, corresponding to rows in the figure (row 1, A: $\phi = 0.2$, row 2, C: $\phi = 0.5$, row 3, E: $\phi = 0.9$).
row 3, E: $\phi=0.9$). After these values were specified in the simulation, data sets were simulated with values of E+ germination ($g^+$) from 0.01-1.0, one hundred times. Column 1 shows the means (red dots) and 95% credible intervals (black lines) for the ability of the model to recover E+ germination estimate over the parameters space from 0.01-1.0. Column 2 shows the mean (red dots) and 95% credible intervals (vertical black lines) for the ability of the model to simultaneously recover retention estimates (B: $\phi=0.2$, D: $\phi=0.5$, F: $\phi=0.9$); note, points are horizontally jittered in column 2.
Appendix E: Methods and data collection for above-ground demographics

All above-ground parameters for our *Agrostis hyemalis* population model were estimated from data collected in the control plots of a separate field experiment on this species (details in Miller et al. in prep). Here, control plots refer to those that did not receive a water addition or transmission reduction treatment. In total, 12 control plots were established in the winter of 2013 in a recently tilled open-field habitat at the USDA East Texas Plant Materials Center, adjacent to where the seed bank experiment was established. Each control plot was initiated with 20 plants at initial endophyte frequencies of 0.05, 0.35, 0.65, and 0.95, replicated 3 times each. Source plants were germinated from seed and reared in the greenhouse the year prior and are native to the Stephen F. Austin Experimental Forest adjacent to the field site.

The data used to estimate plant demographics were collected on recruits that established within control plots in the year following the plot establishment, winter of 2014. Within each control plot, three sub-plots (0.25 m x 0.25 m) were created with similar orientation among plots. Within each sub-plot we counted the number of vegetative and flowering *A. hyemalis* plants, randomly tagging four recruits, now referred to as demography plants. For each demography plants we recorded size (number of tillers) and the number of inflorescences per individual (inflorescences were collected). In the lab we weighed seeds, pooled across individual, while screening an initial 4 seeds per plant for endophyte status, using light microscopy. For plants that had any E+ seeds during the initial screening, an additional 16 seeds were scored (or the total number
available). This method gave us a total of 20 seeds for vertical transmission estimates. Survival of demography plants was determined from a winter census in January of 2015.

There were some constraints with this data collected in that endophyte status could only be assigned to plants that flowered (because their seeds were scored). Thus, the endophyte status of plants that did not flower is unknown, and we combined the probability of flowering and survival of non-flowering plants across host types (E+ & E-). To make the most of our data, we decomposed the fertility \( f^{+/-} \) and plant survival \( s_p^{+/-} \) parameters into lower-level parameters:

\[
 f^{+/-} = p_f \times n^{+/-} \quad \text{Eq. E1}
\]

\[
 s_p^{+/-} = (1 - p_f) \times s_{vp} + p_f \times s_{fp}^{+/-} \quad \text{Eq. E2}
\]

Where \( p_f \) is the probability of flowering (not endophyte specific) and \( n^{+/-} \) is the number of seeds produced per flowering plant (unique to E+ and E-). For survival, we use the weighted average of the survival probability of vegetative plants \( s_{vp} \), not endophyte specific) and survival probability of flowering plants \( s_{fp}^{+/-} \), unique to E+ and E-). The vertical transmission rate from above-ground data \( \tau \) was used in our population model.