The measurement of coevolution in the wild

Bob Week, Scott Nuismer

Article title:

The measurement of coevolution in the wild

Contact information:

Corresponding/primary author: Bob Week, bobweek@gmail.com, 1-360-216-9074, Department of Biological Sciences, University of Idaho

Secondary author: Scott L. Nuismer, snuismer@uidaho.edu, Department of Biological Sciences, University of Idaho

Keywords:

coevolution, maximum likelihood, quantitative genetics, evolutionary ecology, coevolutionary arms race

Article type:

Letter

Word counts:

Abstract: 134

Main text: 4521

Text boxes: 0

Item counts:

References: 37

Figures: 4

Tables: 2

Text boxes: 0

Statement of authorship:

B.W. and S.L.N. conceived of the study and developed the models. B.W. performed the analyses. B.W. and S.L.N. wrote the paper.

Data accessibility statement:

All data and code used in analyses can be found at the following Github repository:

`https://github.com/bobweek/measuring.coevolution`

Abstract

Coevolution has long been thought to drive the exaggeration of traits, promote major evolutionary transitions such as the evolution of sexual reproduction, and influence epidemiological dynamics. Despite coevolution's long suspected importance, we have yet to develop a quantitative understanding of its strength and prevalence because we lack generally applicable statistical methods that yield numerical estimates for coevolution's strength and significance in the wild. Here we develop a novel method that derives maximum likelihood estimates for the strength of direct pairwise coevolution by coupling a well established coevolutionary model to spatially structured phenotypic data. Applying our method to two well-studied interactions reveals evidence for coevolution in both systems. Broad application of this approach has the potential to further resolve long-standing evolutionary debates such as the role species interactions play in the evolution of sexual reproduction and the organization of ecological communities.

Main text

Introduction:

Our current understanding of coevolution's importance rests upon methods that fall into two general classes: those that are broadly applicable but yield only qualitative evidence for coevolution and those that produce quantitative estimates for the strength of coevolution but can be applied only in a narrow range of systems. For example, one popular approach for inferring coevolution relies on measuring the spatial correlation between traits of interacting species and using significant interspecific correlations as evidence of a coevolutionary process (Berenbaum *et al.*, 1986; Hanifin *et al.*, 2008; Toju, 2008; Pauw *et al.*, 2009). Strengths of this approach include the relative ease of collecting the relevant data and its broad applicability to a wide range of species interactions. The critical weakness of this approach, however, is that significant interspecific correlations are neither necessary nor sufficient for demonstrating coevolution (Nuismer *et al.*, 2010; Janzen, 1980). Similarly, time-shift experiments have been broadly implemented in systems where experimental evolution is a tractable approach, but do not yield quantitative estimates of the strength of coevolution (Koskella, 2014; Blanquart & Gandon, 2013; Gaba & Ebert, 2009). In contrast, more quantitative approaches such as selective source analysis, a method that additively partitions selection gradients into independent components of selection (Ridenhour, 2005), require the collection of extensive trait and fitness data from interacting species and thus have proven difficult to employ in all but a few specialized systems (Brodie III & Ridenhour, 2003; Nuismer & Ridenhour, 2008; Burkhardt *et al.*, 2012). As a consequence of these trade-offs in existing approaches, rigorous quantitative estimates of the strength of coevolution in natural populations are extremely scarce.

A promising alternative to existing approaches is the development of model-based inference methods that use easily collected phenotypic data to estimate the significance of well established coevolutionary models and hence to test for the significance of coevolution. In particular, coevolutionary models now exist that predict the statistical distribution of traits across multiple populations for a pair of interacting species that evolve in response to random genetic drift, abiotic selection, and coevolution (Nuismer *et al.*, 2010). Crucially, these

models predict that the distribution of local population trait means in the interacting species across a metapopulation will approach a bivariate normal distribution entirely described by five statistical moments: the average value of the key trait in each species among populations, the variance of the key trait in each species among populations, and the spatial association (covariance) between the key traits in each species. The phenotypic data necessary to calculate these statistical moments can be visualized as a two-dimensional scatter plot. Where each axis measures the mean trait value for one of the species. Hence, each point in the scatter plot corresponds to a pair of mean traits of the two interacting species within a given population.

Because the models predict a bivariate normal distribution of traits, calculating the likelihood of observing any particular set of trait values in a pair of interacting species is straightforward. With the five statistical moments that describe the bivariate normal distribution, we can infer up to five model parameters. The five parameters our method infers includes strengths of reciprocal selection caused by the focal interaction (the strengths of biotic selection B_1, B_2), the strengths of selection due to any other source (the strengths of "abiotic" selection A_1, A_2), and the optimal offset between trait values that optimize biotic fitness (δ). The parameters quantifying selection (B_i and A_i) are proportional to the selection gradients due to the biotic and abiotic components of selection in each population (see Appendix S1.4). By maximizing the resulting likelihood with respect to these key parameters, our method can be used to rigorously test for the presence of coevolution. Specifically, for a coevolutionary hypothesis to be supported, reciprocal selection must be demonstrated (Janzen, 1980; Thompson, 1994). In our maximum likelihood framework, this long-standing and widely accepted definition of coevolution corresponds to demonstrating that both strengths of biotic selection are significantly non-zero. By performing likelihood ratio tests, support for the coevolutionary hypothesis can be compared relative to support for the null hypotheses of unilateral evolution where $B_1 = 0$ or $B_2 = 0$ (also referred to as tracking, see Figure 1). Due to the nested structure of these models, the likelihood of coevolution and the likelihoods of the null models can be directly compared via likelihood ratio tests. Figure 1 shows that each p-value p_1 and p_2 must be less than the significance threshold α (we use $\alpha = 0.05$) to support a coevolutionary hypothesis. Rejecting either null hypothesis of unilateral evo-

lution automatically implies the rejection of evolution completely absent of biotic selection ($B_1 = B_2 = 0$) since the likelihood of the this third null model will always be less than the likelihoods of tracking.

While showing both B_1 and B_2 are non-zero is necessary for demonstrating the significance of pairwise coevolution, the strength of coevolution can most easily be quantified as the geometric mean of the absolute value of the two biotic selection strengths: $\mathfrak{C} \equiv \sqrt{|B_1 B_2|}$. If either strength of biotic selection is zero, and hence coevolution is absent, then $\mathfrak{C} = 0$ as desired and if $|B_1| = |B_2|$, then $\mathfrak{C} = |B_1| = |B_2|$. However, our metric $\mathfrak{C}$ fails to capture a sense of balance in the forces of biotic selection. We therefore propose an accompanying measure based on Shannon entropy that takes this into account. Setting $b_i = |B_i|/(|B_1| + |B_2|)$ we define the balance of coevolutionary selection as

$$\mathfrak{B} \equiv \frac{(b_1 \ln b_1 + b_2 \ln b_2)}{\ln(1/2)}. \tag{1}$$

Standardizing by $\ln(1/2)$ makes $0 \leq \mathfrak{B} \leq 1$ with $\mathfrak{B} = 1$ representing perfect balance and $\mathfrak{B} = 0$ representing unilateral evolution. Though the strength and balance of coevolution can be subjectively inferred upon inspection of the biotic selection strengths, these two metrics provide a way to quantitatively compare these aspects of coevolution across systems.

Materials and methods

The coevolutionary model:

To model the coevolutionary process, we begin by considering a local population level model of pairwise coevolution. This model assumes fitness is a function of the environment, the trait of the focal individual and the trait of the individual being encountered. In particular, we assume species i has an optimal phenotype θ_i that maximizes fitness in the absence of the interaction (the abiotic phenotypic optimum). We define A_i to be the strength of abiotic selection on species i so that the abiotic component of fitness ($W_{A,i}$), as a function of the trait value z_i, is proportional to

$$W_{A,i} \propto \exp\left(-\frac{A_i}{2}(\theta_i - z_i)^2\right). \tag{2}$$

Likewise, beginning from first principles, we derive the biotic component of fitness for an individual of species i. We assume that biotic fitness is maximized when the trait value of the focal individual z_i is offset from the trait value being encountered z_j by an ideal amount δ. We refer to δ as the "optimal offset". A simple example of an optimal offset comes from considering the interaction between long-tubed flowers and the long-proboscid flies that visit them. The biotic component of fitness for the fly is maximized when its proboscis is slightly longer than the nectar tube depth of the flower, allowing the fly to easily extract its nectar reward. The difference between tube depth and proboscis length that maximizes the flies biotic fitness component is the optimal offset for the fly. Note how this differs from a "bigger is better" situation commonly referred to for the explanation of coevolutionary arms races. Under the optimal offset model, fitness is a unimodal function and therefore does not increase indefinitely with larger (or lesser) trait values. A more general model would allow different δ's for each species, but since our method can only infer up to five parameters we make the parsimonious assumption that both species have the same optimal offset. Defining B_i to be the strength of biotic selection on species i, the biotic component of fitness ($W_{B,i}$) is proportional to

$$W_{B,i} \propto \exp\left(-\frac{B_i}{2}(\bar{z}_j + \delta_i - z_i)^2\right) \tag{3}$$

when biotic selection is weak ($|B_i| \ll 1$). Here $\bar{z}_j$ is the within population average phenotype of species j. Net fitness is given by the product of the abiotic and biotic components of fitness. Since the amount by which fitness is proportional to these values is irrelevant for evolutionary dynamics, we leave them out here. Detailed derivations are provided in Appendix S1. As noted above our method infers values for B_1, B_2, A_1, A_2 and δ and can thus accommodate most coevolutionary scenarios including escalation ($\delta \neq 0$) and matching ($\delta = 0,\ B_1, B_2 > 0$).

With a functional form of fitness in hand, we employed theoretical quantitative genetics to formally derive the local population model of mean trait dynamics for the two species. From this local model we derived the dynamics of the distribution of pairs of mean traits across the metapopulation. Since our model predicts the metapopulation distribution of mean-trait-pairs will converge to a bivariate normal (a proof is given in Appendix S1.8),

we are justified in tracking only the first five moments of the metapopulation distribution. These are the metapopulation mean traits of each species (μ_1 and μ_2), the metapopulation variance of local mean traits for each species (V_1 and V_2) and the metapopulation covariance of local mean traits for the two species (C). For species i we denote the additive genetic variance by G_i and the local effective population size by n_i. Results derived in Appendix S1 demonstrate that the five moments change according to the following recursions:

$$\Delta\mu_1 = G_1\{B_1\delta + B_1(\mu_2 - \mu_1) + A_1(\theta_1 - \mu_1)\} \tag{4a}$$

$$\Delta\mu_2 = G_2\{B_2\delta + B_2(\mu_1 - \mu_2) + A_2(\theta_2 - \mu_2)\} \tag{4b}$$

$$\Delta V_1 = -2A_1G_1V_1 + 2B_2G_2(C - V_1) + G_1/n_1 \tag{4c}$$

$$\Delta V_2 = -2A_2G_2V_2 + 2B_1G_1(C - V_2) + G_2/n_2 \tag{4d}$$

$$\Delta C = B_2G_2(V_1 - C) + B_1G_1(V_2 - C) - (A_1G_1 + A_2G_2)C. \tag{4e}$$

Parameter estimation:

After solving for the equilibrium expressions of the first five moments from equations (4), we use maximum likelihood to estimate the selection strengths (A_1, A_2, B_1 and B_2) and the optimal offset (δ). However, to do so requires more than estimates of mean trait pairs from multiple populations. Background parameters of the model also need to be estimated. These include the effective population sizes n_1, n_2, the optimal phenotypes favored by abiotic stabilizing selection θ_1, θ_2 and the additive genetic variances G_1, G_2.

We show in Appendix S1.5 that if n_i has been estimated from multiple locations, these can be included by using their harmonic mean as the effective population size in our model. Likewise, if G_i has been estimated from multiple populations, these can be included by using their arithmetic mean as the effective additive genetic variance for our model. Finally, the model used in this manuscript assumes the abiotic optimum is constant across space. In the associated Mathematica notebook, we expand the model to formally account for variable θ_i. The results of this notebook demonstrate that the two models are equivalent when variation in θ_i is small and therefore implies that the average abiotic optimum across space works as the effective abiotic optimum needed to perform inference. This notebook also implies that

our method is readily adaptable for the inclusion of spatially varying optima as such data become available.

The likelihood is a routine calculation in terms of the first five moments which are in turn functions of model parameters $(n_1, n_2, \theta_1, \theta_2, G_1, G_2, \delta)$ and selection strengths (A_1, A_2, B_1, B_2). In Appendix S2 we show how to invert these expressions to obtain analytic solutions for the maximum likelihood estimates of selection strengths. Full expressions are provided in the associated Mathematica notebook. Although our focus is on finding point estimates for the strengths of biotic selection, coevolution and coevolutionary balance, we also estimated uncertainty due to error caused by sampling from the metapopulation. To do so we calculated 95% confidence intervals for each selection strength.

Estimating significance:

Denoting the likelihood of the coevolutionary model by L_c and the likelihood of null model i (for which $B_i = 0$) by L_i, we compute the log-likelihood difference statistic by

$$\Lambda_i = 2(\ln L_c - \ln L_i). \tag{5}$$

Denote by $F_j(x)$ the distribution function of a χ^2 random variable with degrees of freedom j. Wilk's theorem implies the distribution of Λ_i is approximately a χ^2 (Wilks, 1938). Since in each null model we fix just one parameter, the degrees of freedom is one for both tests. Thus, the p-value associated with testing against null hypothesis i (written p_i) has the following approximation

$$p_i \approx 1 - F_1(\Lambda_i). \tag{6}$$

If both p_1 and $p_2 < 0.05$ for a given study system then our method asserts significant evidence for coevolution exists in this system. We provide a tutorial for implementing our approach using the statistical programming language R at the following url:

`https://bobweek.github.io/measuring_coevolution.html`

Evaluation of performance:

Before applying our maximum likelihood methodology to specific study systems, we evaluated its performance when challenged with simulated data. We assessed the type-1 error

rate and statistical power of our method across a range of biotic selection strengths and metapopulation sample sizes. These analyses were performed by simulating data under the model with randomly drawn model parameters. Distributions used for each background parameter are reported in Table 1. For error rates as functions of biotic selection strengths, sample sizes were drawn at random from a Poisson distribution with a mean of 20. Draws were repeated until a sample size of at least three was obtained. For type-1 error rates as functions of unilateral selection we chose one biotic strength to be zero and set the other to the strength of unilateral selection. For type-2 error rates as functions of the strength of coevolution $\mathfrak{C}$, we drew one biotic selection strength from a uniform distribution on the interval $(\mathfrak{C}/10, 10\mathfrak{C})$ and set the other such that their geometric mean equates to $\mathfrak{C}$. When calculating type-2 error rates as functions of sample size, strengths of biotic selection were drawn independently from a uniform distribution on (0,0.01). A similar approach was taken for calculating type-1 error rate as a function of sample size, except one or both of the biotic selection strengths were set to zero at random. If either strength of biotic selection was set to zero in the simulation and reported significantly non-zero by our method, a false positive was accumulated. Likewise, if both strengths of biotic selection were set to some non-zero number and our method failed to detect coevolution, then a type-2 error was accumulated. This scheme was repeated 10,000 times for each estimated error rate.

Alongside our analyses of error rates, we investigated our methods ability to accurately infer the strength of coevolution using simulated data. For each replicate, we simulated phenotypic data using the coevolutionary model with known selection strengths and background parameters drawn from the same set of distributions as those used for the error rates as functions of sample size analysis. We then estimated the strength of coevolution as defined above using our maximum likelihood approach and compared it against its actual value via linear regression. Each regression was performed across a range of sample sizes (Figure 2). We also extended this analysis using more general simulations that relax key assumptions such as the absence of gene-flow and normality of data in Appendix S3.

Numerical analyses of our methods performance were done using the statistical programming language *R*. The scripts are publicly available at the following Github repository:

`https://github.com/bobweek/measuring.coevolution`

Measuring coevolution in the wild:

We next applied our maximum likelihood approach to two well-studied species interactions where previous work implicated coevolution as a cause of trait exaggeration and spatial variability (Pauw *et al.*, 2009; Toju, 2011): the mutualism between the long tongued fly *Moegistorhynchus longirostris* and a plant it pollinates *Lapeirousia anceps* as well as the antagonism between the camellia plant *Camellia japonica* and its seed predator, the weevil *Curculio camelliae*. In both cases, the interactions are thought to depend largely on a single key trait in each species (fly proboscis and plant floral tube lengths or weevil rostrum length and camellia pericarp thickness). This is a crucial detail as the models upon which our method is based assume interactions are mediated by a single trait in each species. Phenotypic data for these systems have been collected from several populations, providing a sample of pairs of mean trait values, the core data required by our method. In addition to the essential phenotypic data, previous work in both systems provided valuable additional information that allowed us to estimate the key background parameters required by our method: the likely trait optima in the absence of the interaction (the "abiotic" optima), the effective population sizes for each species (assumed fixed over time and space), and the effective additive genetic variances for each species (also assumed to be fixed over time and space).

The long proboscid fly, *M. longirostris*, resides in lowland habitats near the coast of South Africa and pollinates a guild of at least 20 plant species (Manning & Goldblatt, 1997). Among these species, the most widespread is *L. anceps*, a long tubed perennial whose distribution extends outside the range of *M. longirostris* (Pauw *et al.*, 2009). We were able to estimate the likely optimal tube and proboscis lengths for these species in the absence of this particular interaction. Using the phenotypic data published by Pauw *et al.* (2009), we inferred this parameter for the flower as the average mean tube length of two populations not visited by the fly. Estimating the abiotic optima for the fly was more challenging because we were unable to identify fly populations where the plant did not co-occur. However, there are data available for the proboscis lengths in three sister species of *M. Longirostris* (41.0 mm for *M. braunsi*, 11.5 mm for *M. brevirostris*, and 32.0 mm for *M. perplexus*) (Bequaert, 1935). Since these sister species do not interact with *L. anceps* (Barraclough & Slotow, 2010), their traits

represent potential evolutionary trajectories that could have been taken by *M. longirostris* in the absence of its interaction with *L. anceps*. Given that none of the three sister species underwent a similar arms race with some other flower (which appears likely based on their relatively modest proboscis lengths), we therefore take these values as rough approximations of the actual abiotic optimal phenotype for *M. longirostris*. Hence, we estimated selection strengths and significance when the abiotic optimum was set equal to each of the three trait values and the average of all three. The result presented in the main text correspond to the average of all three sister species, but we present the results for all four abiotic optima in Appendix S4.1. Effective population sizes have not been estimated for either species. We therefore relied on the biologically plausible census sizes of 1000 for *L. anceps* and 100 for *M. longirostris*, as suggested by B. Anderson (personal communications). Since heritabilities for neither of these traits have been estimated, we relied on within population phenotypic variances as a rough proxy for the additive genetic variances in this system.

We complement our analysis of this plant pollinator mutualism with an analysis of the antagonistic interaction between *C. camelliae* and *C. japonica* (Toju & Sota, 2005). Female weevils bore holes into the woody pericarps of the camellia to oviposit. Inside the fruit, weevil larvae feed on the seeds of the camellia up until the fourth instar, at which time they exit the fruit and overwinter (Toju & Sota, 2005). These two species co-occur across Japan, although camellia populations where the weevil is absent also exist (Toju & Sota, 2005). We were able to establish point estimates of each background parameter using data from previously published work (Toju *et al.*, 2011b,a) and the fact that male weevil rostrum lengths could be used as a proxy for the abiotic optimum of the female weevils since males do not interact with the camellia. Hence, our method does not inherently require estimates of abiotic optima to come from populations where the interaction is absent. However, using male traits as a surrogate for the abiotic optimum assumes that male and female trait values are either genetically uncorrelated or have reached equilibrium. The abiotic optimum for the pericarp thickness of the camellia was inferred by averaging pericarp thicknesses across populations where weevils are absent. Heritability of pericarp thickness has been estimated directly (Toju *et al.*, 2011a) and can be at least crudely inferred for weevil rostrum length via estimates of related species (Toju & Sota, 2009). We used the average of these values for

each species multiplied by the average within population phenotypic variances to estimate additive genetic variances in this system.

To assess the biological significance of the strengths of coevolution inferred, we compared the distribution of trait values we would expect in the presence vs absence of coevolution. This was accomplished by setting both B_1 and B_2 equal to zero and maximizing the resulting restricted likelihood function with the remaining free parameters (A_1, A_2 and δ). Using a multivariate generalization of effect size (see Appendix S4.3), we summarize with a single number the effect of coevolution in each system.

Results:

Evaluation of performance:

Regressions of randomly drawn strengths of coevolution onto those inferred by our method were heteroskedastic with variation proportional to the strength of coevolution (Bartlett's test: p-value $< 2.22e-16$). To rectify this we used weighted least squares. For each point in the regression we set its weight equal to the inverse of its Euclidean distance to the origin. Analysis of regression results demonstrates that at low sample sizes our method tends to overestimate the strength of coevolution, but this bias rapidly diminishes with sample size (see Figure 2).

False positive rates are greatly exaggerated for small sample sizes (e.g., < 5), modestly inflated for sample sizes between $5-10$, but approach their set value (0.05) for sample sizes > 10 (Figure 2). This behavior is attributable to two factors. First, statistical artifacts accumulate in sample moments for small sample sizes. For example, the correlation of a sample of size two will always be ± 1. Second, the distribution of our p-values may significantly diverge from a Chi-square distribution at small sample sizes (Wilks, 1938). We therefore suggest this method only be used for sample sizes of at least five. Another important caveat, however, is that as biotic selection becomes increasingly imbalanced under the null scenario when one strength is zero and the other set to some non-zero number, the false positive rate increases monotonically (see Figure 2). Hence, our method can be tricked by extreme unilateral selection.

Power to detect coevolution is reasonably high at low sample sizes (≈ 0.9) and increases

monotonically with sample size. As a function of the strength of coevolution, power is initially negligible but increases quickly and monotonically.

Measuring coevolution in the wild:

We found that the biotic selection strengths B_1 and B_2 acting on *M. longirostris* and *L. anceps* both differ significantly from zero (Table 2). Thus, our analysis supports the hypothesis of pairwise coevolution in this system. Likewise, both B_1, the strength of biotic selection on the weevil, and B_2, biotic selection on the camellia plant, significantly differed from zero. Hence, we also found evidence for pairwise coevolution between the seed-eating weevil *C. camelliae* and its host plant *C. japonica*. For numerical estimates of biotic selection strengths, p-values, and the strength and balance of coevolution, see Table 2. Cross-system comparison of biotic selection strengths is visualized in Figure 3.

In addition to providing information on the magnitude and significance of coevolution, we quantified the extent of trait exaggeration produced by coevolution by comparing the equilibrium phenotypic distribution we would expect with and without the levels of coevolution we estimated (Figure 4). This comparison reveals that although the numerical estimates of coevolutionary selection appear superficially small, for the camellia-weevil interaction coevolution results in a 111% increase in the mean rostrum length of the camellia weevil and a 66.0% increase in the pericarp thickness of the camellia fruit (Figure 4). For the fly-flower system coevolution appears to have caused a 134% increase in proboscis length and a 34.5% increase in floral tube depth compared to equilibrium estimates for these values we predict when coevolution is absent. Using a multivariate analog of effect size we calculated the effect of coevolution in each system. We found an effect size of 7.55 for the fly-flower system and an effect size of 3.07 for the camellia-weevil interaction.

Discussion:

Our results demonstrate that coupling existing coevolutionary models with a maximum likelihood approach allows the strength of coevolutionary selection to be estimated using routinely collected phenotypic data. Regression analysis shows that with sufficient sample sizes we can obtain accurate estimates of the strength and significance of coevolution. Furthermore, our method is robust to modest amounts of gene flow and weakly non-normal data (Appendix

S3).

Applying our method to two textbook examples of pairwise coevolution, we find strong evidence for significant coevolution in both systems. This qualitative result is complemented by quantitative estimates of the strength of coevolution in the wild. By applying this method to various systems, it will be possible to obtain an empirical distribution of the strength of coevolution in nature. After the appropriate transformation (analogous to standardizing selection gradients with respect to phenotypic distributions) such data will allow for a meta-analysis akin to (Kingsolver *et al.*, 2001; Siepielski *et al.*, 2009, 2013) which would provide a yardstick allowing us to further understand the biological significance of our numerical results.

In spite of the various merits of our method, there are serious limitations that must be confronted empirically. Most notable is the necessity of providing estimates of abiotic optima. Since these parameters are seldomly estimated for natural populations, we are restricted in our analysis here to two data sets in which sufficient information was provided. In particular, phenotypic measurements in populations that do not partake in the interaction (due to geographical isolation or sexual dimorphism) provide reasonable estimates of the abiotic optima, though other means of estimating these parameters exist as demonstrated above.

Alongside the empirical work necessary for estimating background parameters of our model, our results suggest that increasing the number of populations used in studies of trait matching would also substantially improve opportunities for coevolutionary inference. Specifically, we suggest sample sizes of at least five and ideally more than twenty to avoid type-1 errors. Taken together, these considerations outline a reasonably tractable set of sufficient conditions empirical data-sets must meet in order to utilize our method.

Theoretical limitations of our approach stem from its grounding in classic quantitative genetics and include the assumptions of fixed additive genetic variance and weak selection. Although we do not assume strict equilibrium for each component population, we do assume that the system as a whole has reached approximate statistical equilibrium so that the means, variances and spatial covariance have become relatively constant with respect to time. This implies that pairs of species for which this method is ideal have been interacting for a sufficiently long period of time. In reality, however, empirical systems may be far

enough from equilibrium that a significant contemporary trend in the five moments describing their distribution should be accounted for. Lastly, our method assumes the key traits mediating the interaction are univariate which may not be ubiquitous across coevolving systems. Future work that generalizes our approach to multivariate traits, strong selection and non-equilibrium (ie, time-series data) will result in a more broadly applicable method.

By providing a methodology that does not rely on extensive and system specific experimental manipulation, our approach greatly expands the range of systems for which the strength of coevolutionary selection can be estimated, paving the road for a more quantitative and critical assessment of coevolution's importance in natural systems. To add substance to this claim we provide three examples. First, with finer spatial resolution in phenotypic data this method can be applied to the same pair of species across different partitions of their range to infer the strength of selection mosaics argued to be central to the coevolutionary process by the Geographic Mosaic Theory of Coevolution (Thompson, 2005). Second, previous investigations have resulted in mixed views on the significance of pairwise coevolution in shaping various aspects of ecological communities including inter- and intraspecific diversity, demographic stability, network structure and ecosystem function (Iwao & Rausher, 1997; Roughgarden, 1979; Nuismer *et al.*, 2013; Althoff *et al.*, 2014; Yamamura *et al.*, 2001). By applying our method to each pairwise interaction in a set of interacting species, the distribution of pairwise coevolution can be inferred within a community to provide empirical insight into the degree to which coevolution molds the previously mentioned properties of ecological communities. Third, theoretical studies suggest that only very strong coevolution favors the evolution of sexual reproduction (Otto & Nuismer, 2004; Lively, 2010; Agrawal, 2006). Our method could inform this hypothesis by determining the strength of coevolution in specific systems where the evolution of sex has been attributed to interspecific interactions. Hence, when coupled with data from a broad range of empirical systems, this method and its future iterations hold the potential to settle long standing debates involving the importance of species interactions and coevolution in the evolution of various phenomena including phenotypic diversity, sexual reproduction, community structure, and epidemiological dynamics (Yoder & Nuismer, 2010; Hamilton, 1980; McPeek, 2017; Anderson & May, 1982).

References

Agrawal, A.F. (2006). Similarity selection and the evolution of sex: revisiting the red queen. *PLoS biology*, 4, e265.

Althoff, D.M., Segraves, K.A. & Johnson, M.T. (2014). Testing for coevolutionary diversification: linking pattern with process. *Trends in Ecology & Evolution*, 29, 82–89.

Anderson, R.M. & May, R. (1982). Coevolution of hosts and parasites. *Parasitology*, 85, 411–426.

Barraclough, D. & Slotow, R. (2010). The south african keystone pollinator moegistorhynchus longirostris (wiedemann, 1819)(diptera: Nemestrinidae): notes on biology, biogeography and proboscis length variation. *African Invertebrates*, 51, 397–403.

Bequaert, J. (1935). Notes on the genus moegistorhynchus and description of a new african species of nycterimyia (diptera, nemestrinidae). *Annals of the Transvaal Museum*, 15, 491–502.

Berenbaum, M., Zangerl, A. & Nitao, J. (1986). Constraints on chemical coevolution: wild parsnips and the parsnip webworm. *Evolution*, pp. 1215–1228.

Blanquart, F. & Gandon, S. (2013). Time-shift experiments and patterns of adaptation across time and space. *Ecology letters*, 16, 31–38.

Brodie III, E. & Ridenhour, B. (2003). Reciprocal selection at the phenotypic interface of coevolution. *Integrative and Comparative Biology*, 43, 408–418.

Burkhardt, A., Ridenhour, B., Delph, L. & Bernasconi, G. (2012). The contribution of a pollinating seed predator to selection on silene latifolia females. *Journal of evolutionary biology*, 25, 461–472.

Gaba, S. & Ebert, D. (2009). Time-shift experiments as a tool to study antagonistic coevolution. *Trends in Ecology & Evolution*, 24, 226–232.

Hamilton, W.D. (1980). Sex versus non-sex versus parasite. *Oikos*, pp. 282–290.

Hanifin, C.T., Brodie Jr, E.D. & Brodie III, E.D. (2008). Phenotypic mismatches reveal escape from arms-race coevolution. *PLoS biology*, 6, e60.

Iwao, K. & Rausher, M.D. (1997). Evolution of plant resistance to multiple herbivores: quantifying diffuse coevolution. *The American Naturalist*, 149, 316–335.

Janzen, D.H. (1980). When is it coevolution. *Evolution*, 34, 611–612.

Kingsolver, J.G., Hoekstra, H.E., Hoekstra, J.M., Berrigan, D., Vignieri, S.N., Hill, C., Hoang, A., Gibert, P. & Beerli, P. (2001). The strength of phenotypic selection in natural populations. *The American Naturalist*, 157, 245–261.

Koskella, B. (2014). Bacteria-phage interactions across time and space: merging local adaptation and time-shift experiments to understand phage evolution. *The American Naturalist*, 184, S9–S21.

Lively, C.M. (2010). A review of red queen models for the persistence of obligate sexual reproduction. *Journal of Heredity*, 101, S13–S20.

Manning, J.C. & Goldblatt, P. (1997). Themoegistorhynchus longirostris (diptera: Nemestrinidae) pollination guild: long-tubed flowers and a specialized long-proboscid fly pollination system in southern africa. *Plant Systematics and Evolution*, 206, 51–69.

McPeek, M.A. (2017). The ecological dynamics of natural selection: traits and the coevolution of community structure. *The American Naturalist*, 189, E91–E117.

Nuismer, S. & Ridenhour, B. (2008). The contribution of parasitism to selection on floral traits in heuchera grossulariifolia. *Journal of evolutionary biology*, 21, 958–965.

Nuismer, S.L., Gomulkiewicz, R. & Ridenhour, B.J. (2010). When is correlation coevolution? *The American Naturalist*, 175, 525–537.

Nuismer, S.L., Jordano, P. & Bascompte, J. (2013). Coevolution and the architecture of mutualistic networks. *Evolution*, 67, 338–354.

Otto, S.P. & Nuismer, S.L. (2004). Species interactions and the evolution of sex. *Science*, 304, 1018–1020.

Pauw, A., Stofberg, J. & Waterman, R.J. (2009). Flies and flowers in darwin's race. *Evolution*, 63, 268–279.

Ridenhour, B.J. (2005). Identification of selective sources: partitioning selection based on interactions. *The American Naturalist*, 166, 12–25.

Roughgarden, J. (1979). *Theory of population genetics and evolutionary ecology: an introduction*.

Siepielski, A.M., DiBattista, J.D. & Carlson, S.M. (2009). Its about time: the temporal dynamics of phenotypic selection in the wild. *Ecology letters*, 12, 1261–1276.

Siepielski, A.M., Gotanda, K.M., Morrissey, M.B., Diamond, S.E., DiBattista, J.D. & Carlson, S.M. (2013). The spatial patterns of directional phenotypic selection. *Ecology letters*, 16, 1382–1392.

Thompson, J.N. (1994). *The coevolutionary process*. University of Chicago Press.

Thompson, J.N. (2005). *The geographic mosaic of coevolution*. University of Chicago Press.

Toju, H. (2008). Fine-scale local adaptation of weevil mouthpart length and camellia pericarp thickness: altitudinal gradient of a putative arms race. *Evolution*, 62, 1086–1102.

Toju, H. (2011). Weevils and camellias in a darwins race: model system for the study of eco-evolutionary interactions between species. *Ecological research*, 26, 239–251.

Toju, H., Abe, H., Ueno, S., Miyazawa, Y., Taniguchi, F., Sota, T. & Yahara, T. (2011a). Climatic gradients of arms race coevolution. *The American Naturalist*, 177, 562–573.

Toju, H. & Sota, T. (2005). Imbalance of predator and prey armament: geographic clines in phenotypic interface and natural selection. *The American Naturalist*, 167, 105–117.

Toju, H. & Sota, T. (2009). Do arms races punctuate evolutionary stasis? unified insights from phylogeny, phylogeography and microevolutionary processes. *Molecular ecology*, 18, 3940–3954.

Toju, H., Ueno, S., Taniguchi, F. & Sota, T. (2011b). Metapopulation structure of a seed–predator weevil and its host plant in arms race coevolution. *Evolution*, 65, 1707–1722.

Wilks, S.S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, 9, 60–62.

Yamamura, N., Yachi, S. & Higashi, M. (2001). An ecosystem organization model explaining diversity at an ecosystem level: Coevolution of primary producer and decomposer. *Ecological Research*, 16, 975–982.

Yoder, J.B. & Nuismer, S.L. (2010). When does coevolution promote diversification? *The American Naturalist*, 176, 802–817.

Acknowledgements:

Funding for this project was provided by NSF grant DEB 1450653 to SLN. We dedicate this manuscript to the memory of Dr. Paul Joyce who inspired this project and initiated its progress.

Supplementary information: Is available.

Author information: The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to B.W. at `bobweek@gmail.com`.

Table 1: Distributions of background parameters used for generating error rates and regression analyses.

Parameter(s)	Description	Distribution
A_i	Strength of abiotic selection	Uniform(0,0.01)
δ	Optimal offset	Exp(0.1)
θ_i	Abiotic optima	Normal(0,10)
G_i	Additive genetic variance	Exp(1)
n_i	Effective population size	Exp(0.01)

Table 2: Biotic and abiotic selection strengths, optimal offsets, p-values, and strengths of coevolution and coevolutionary balance for each system. CW refers to the camellia-weevil system and FF refers to the fly-flower system. Units of selection strengths are all inverse square phenotypic units (mm^{-2} in this case). Optimal offsets (δ) are in phenotypic units (mm). The p-values and balances of coevolutionary selection are unitless.

	CW	FF
B_1	7.17e-04	6.40e-05
B_2	5.00e-06	1.84e-06
A_1	2.59e-04	7.04e-06
A_2	8.05e-06	3.13e-06
δ	4.51	14.2
p_1	<2.22e-16	<2.22e-16
p_2	<2.22e-16	1.19e-07
$\mathfrak{C}$	5.99e-05	1.08e-05
$\mathfrak{B}$	5.97e-02	1.84e-01

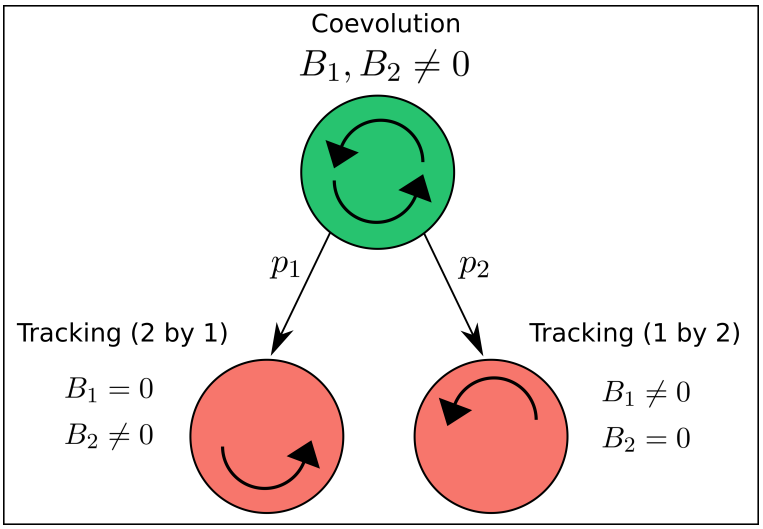

Figure 1: The network structure of hypotheses that can be distinguished using our approach. Nodes represent the three relevant hypotheses for coevolutionary inference. Edges represent comparisons labeled by their p-values. The upper node (in green) represents the coevolutionary hypothesis in which both strengths of selection induced by the interaction are non-zero. The pink colored nodes represent the hypotheses of unilateral evolution, or tracking, where one species experiences biotic selection, but the other does not. By ruling out tracking this approach automatically rejects evolution completely absent of biotic selection.

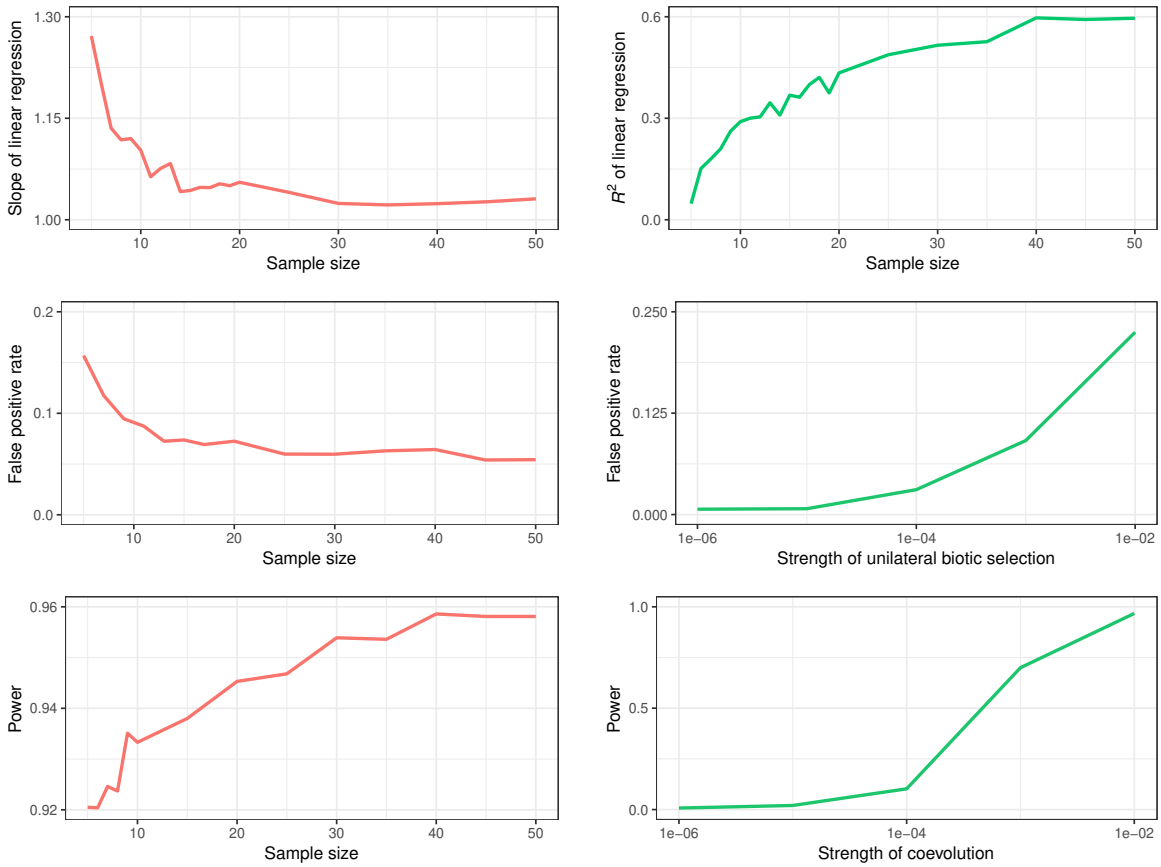

Figure 2: Top row: Performance of parameter estimation as a function of sample size. The left-hand panel shows the slope of the regressions converging near one as sample size increases. The right-hand plot shows the percent variance explained (R^2) increasing with sample size. Lower two rows: Error rates as functions of sample sizes and selection strengths. The left-hand column shows the type-1 and type-2 error rates as functions of sample size. The right-hand column shows type-1 error as a function of the strength of tracking (ie, unilateral selection where the species being tracked does not experience biotic selection) and power as a function of the strength of coevolution.

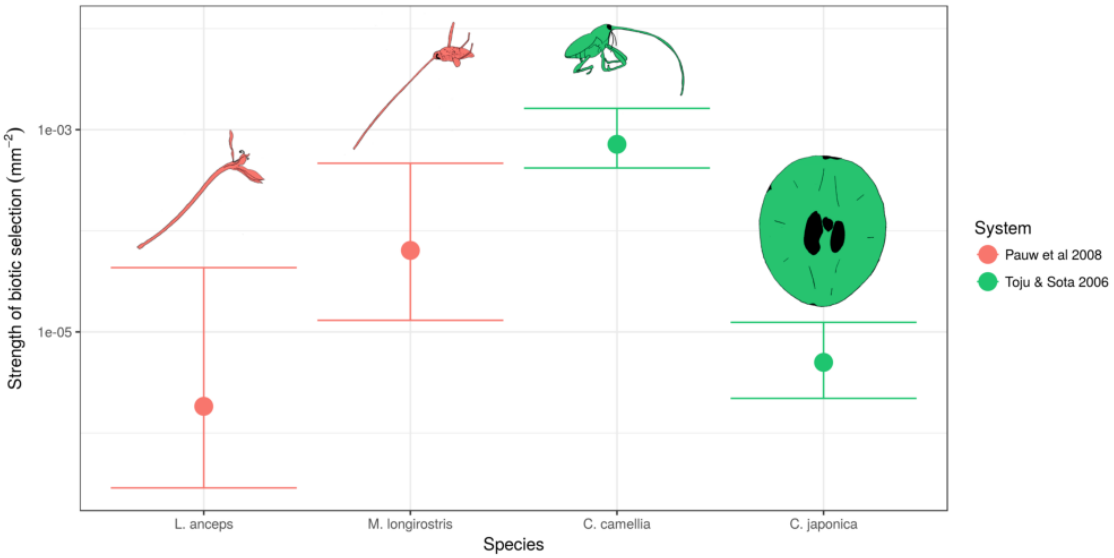

Figure 3: The estimated strength of biotic selection for the *M. longirostris-L. anceps* interaction (pink) and the *C. japonica-C. camelliae* interaction (green). Units for each strength are in mm^{-2}, the inverse of the square of the phenotypic units. 95% confidence intervals are shown around each estimate. Each selection strength was found to be statistically significant and hence coevolution was detected in both systems.

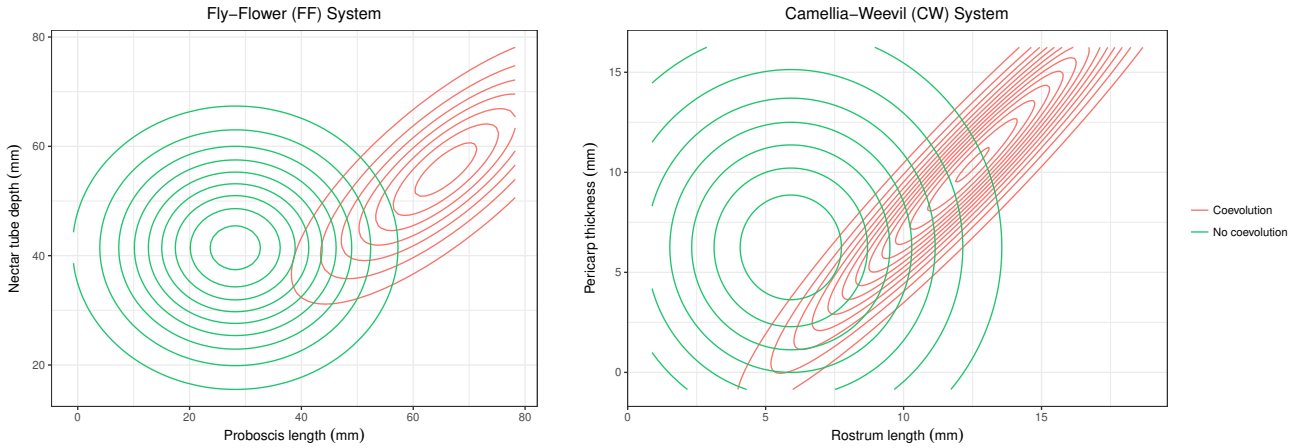

Figure 4: The effect of coevolution on the trait distributions predicted by our model. The point in the center of each contour represents the mean traits of the species involved. The green contours represent data predicted without coevolution and the pink contours represent the observed data.