**Authors:** Heath Blackmon1,2 and Jeffery P. Demuth1,3

1Department of Biology, University of Texas at Arlington

2hblackmon@uta.edu

3jpdemuth@uta.edu

**Title:** An information-theoretic approach to estimating genetic architectures: moving beyond the joint-scaling test for line cross analysis

**Running Title:** I-T approach to line cross analysis

**key words:** line cross analysis, joint-scaling test, composite genetic effects, genetic architecture

**words:** 5422

**tables:** 2

**figures:** 5

**archiving:** R package will be archived on the CRAN repository

**Title**

An information-theoretic approach to estimating genetic architectures: moving beyond the joint-scaling test for line cross analysis

Heath Blackmon1,2 and Jeffery P. Demuth1

1Department of Biology, University of Texas at Arlington

2hblackmon@uta.edu

**Abstract**

The genetic architecture of a trait is important because it fundamentally determines the pace and direction of evolution. Unfortunately, widely implemented methods for estimating genetic architecture in line cross analysis (LCA) experiments may provide a biased view because there has been no way to quantify the uncertainty in selecting the “right” genetic model from an often vast model space. Genetic parameters estimated from LCA are thus conditional on the chosen model, which is often biased toward simpler architectures, even when there are many models with similar fits. We develop and demonstrate a full information-theoretic approach to line cross analysis implemented in the R package SAGA. The software automatically determines and comprehensively tests the model space defined by a user specified set of genetic crosses. By quantifying model selection uncertainty and using model weighted averaging, SAGA is able to accurately estimate composite genetic effect contributions to complex genetic architectures even when there is substantial uncertainty in which is the best overall model. Future line cross analyses will benefit from the I-T approach implemented in SAGA because it is able to more accurately define the components of complex genetic architectures than previous approaches.

**Introduction**

Two of the fundamental questions of evolutionary biology are how species adapt to their environments and how continuous variation among populations becomes discontinuous variation among species. The answers to these broad questions and many other, more specific ones rely fundamentally on the genetic architecture of the focal trait (Demuth and Wade 2007a). For instance, it is well known that adaptation proceeds most efficiently when genetic architectures are simply additive (Fisher 1941; Lande and Arnold 1983), but many phenomena (e.g. speciation, the origin of sex, mating systems) are more easily explained by complex architectures that account for gene interactions (i.e. epistasis; Wright 1931; Wade 2000, 2002). Evolutionary biologists have disagreed about the prevalence of simple vs. complex architectures in natural systems for decades (e.g. Wright 1931; Fisher 1958; Coyne et al. 1997; Wade and Goodnight 1998), in part because efforts to measure genetic architectures are often laborious and suffer from methodological biases against finding gene interactions (Demuth and Wade 2006).

Line cross analysis (LCA) is a widely used quantitative genetics method for estimating genetic architecture by partitioning the mean phenotypes among relatives into their composite genetic effects (CGEs). Originally developed in efforts to understand the genetic architecture of traits important to human agriculture (Mather and Jinks 1982), LCA has also provided insights into the genetic basis of adaptation and speciation (Armbruster et al. 1997; Lair et al. 1997; Edmands 1999; Galloway and Fenster 2001; Miller et al. 2003; Demuth and Wade 2005, 2007b, a; Tymchuk et al. 2007; Fuller 2008; Demuth et al. 2014). The LCA approach typically uses two parental strains that have diverged in a phenotype of interest. These parents are crossed, producing an F1, and subsequent crosses (e.g. F2, backcross, reciprocals) are made to generate groups that have different combinations of parental genes. We refer to each of these groups as cohorts. Using a weighted least squares regression with weights inversely proportional to the variance of the cohort means, the degree to which a phenotype is determined by different CGEs (e.g. additive, dominance, and epistatic gene action) may be estimated (Cavalli 1952; Hayman 1958).

The most widely implemented statistical approach to testing among LCA models with different CGEs is called the joint-scaling test (essentially a forward variable selection weighted least squares regression). Using this approach a simple additive model is first fit and then additional higher order CGEs are added until, based on the results of a likelihood ratio test, no significant improvement in the model is achieved (Mather and Jinks 1982; Lynch and Walsh 1998). This approach is common in studies of both plants and animals (Edmands 1999; Schiffer et al. 2006; van Heerwaarden et al. 2008; Bentz et al. 2011). Due to concerns that this approach may not find the best model, a number of alternative variable selection approaches have been implemented including backward and stepwise variable selection (Gilchrist and Partridge 1999; Demuth and Wade 2007b). Additionally, hybrid approaches have been implemented using Akaike information criterion (AIC) to choose a most parsimonious model from an *a priori* chosen subset of potential models followed by the use of significance tests to add or remove CGEs and arrive at a final “best” model (Bieri and Kawecki 2003; Fritz et al. 2006; Fox et al. 2011). Finally, an iterative approach using AIC to compare multiple small candidate sets of models has been used to choose a single “best” model (Fox et al. 2004).

There are several documented issues with the variable selection processes implemented in the joint-scaling test and existing *ad hoc* approaches (Whittingham et al. 2006): 1) Different variable selection approaches (forward, backward, and stepwise) do not consistently identify the same variables as important (Derksen and Keselman 1992). 2) Parameter estimates under these approaches are biased away from zero (Burnham and Anderson 2002). 3) Calculating valid p-values is difficult due to multiple comparisons (Wilkinson 1979). 4) Hypothesis testing approaches place an inappropriate focus on a single model ignoring the degree of model uncertainty implied by the data (Burnham and Anderson 2002). Finally, because the number of cohorts is often small relative to the number of CGEs being considered, the use of AIC as a metric is inappropriate for line cross data and should be replaced with AICc (Hurvich and Tsai 1989). More broadly, the often repeated aphorism that “all models are wrong, but some are useful” suggests that the search for a single best model may often be misguided (Box and Draper 1987). A more prudent approach would be to evaluate all models that the data allow and identify the useful one(s).

A full information-theoretic (I-T) approach to model selection and parameter estimation alleviates the difficulties associated with previous approaches and provides additional understanding that is not possible under a joint-scaling, hybrid, or ad hoc approach. Here we present a full I-T approach that leverages the finite sample size corrected version of the Akaike information criterion (AICc) to explore all possible models and make unbiased and, when appropriate, model averaged estimates of the contribution of CGEs to cohort means. We have developed an R package: Software for the Analysis of Genetic Architecture (SAGA) that makes this approach straightforward to implement. The SAGA software is available from the R package repository CRAN. We describe our approach and its performance on simulated datasets as well as contrasting the results of this method to those from the joint-scaling test using 17 previously published datasets.

**Method and Interpretation**

The first step in LCA is the development of a C-matrix that describes the potential contribution of CGEs to cohort means. Two versions of C-matrices have been widely used and have been shown to have a linear relationship (Basford and De Lacy 1979). We provide users with a C-matrix that is scaled to the midparent mean (equivalent to F∞), and includes 23 potential CGEs (Table S1). By maximizing the number of CGEs (many of which are normally not explored), we reduce the risk that results are biased by the selection of CGEs tested. For each CGE we have calculated coefficients for 23 potential crosses; each of which is divided into male, female, or mixed sex cohorts. Our software also allows users to modify or supply custom C-matrices.

Depending on the identity of the cohorts supplied, some CGEs may not vary or may be perfectly correlated with one another. Therefore, the first step in our approach is to reduce the C-matrix to include only the CGEs that can be partitioned with the available cohorts. Next, we generate all possible models that have at most two fewer CGEs than the number of cohorts being analyzed (number of cohorts – intercept – 1). We use the existing implementation of weighted least squares regression in the base R package in the function GLM (R Development Core Team 2013). Although SAGA will automatically remove and warn the user if there are perfectly correlated CGEs, some may still be either highly correlated or contain a linear dependency (when one CGE can be described as a linear combination of one or more other CGEs). The GLM function will drop the highest order CGE from a model if it is highly correlated with a lower order CGE or if its inclusion creates a linear dependency. When this occurs it would effectively create a duplicate lower order model in the set being evaluated. If this occurs, we remove the equation with the confounded CGEs from the set being evaluated allowing the CGEs to be estimated independently by lower order equations that are otherwise equivalent.

The function GLM returns the parameter and standard error estimates conditional on the model as well as the AIC value for the model. We convert AIC to AICc using equation 1, where n is the number of cohorts and K is the number of parameters being estimated.

(1)

Once all models have been evaluated, we calculate differences () using equation 2.

(2)

Where is the minimum score calculated across all possible models and is the calculated for a specific model. allows models to be ranked and is used in generating Akaike weights () using equation 3. The denominator in equation 3 is the summation of the numerator across all possible models being evaluated (*R*). The generated in this way will sum to one and can be evaluated as evidence for whether a model is correct.

(3)

Under the default settings, if of the best model is 0.95 or greater then SAGA will perform parameter estimation under a single model. If no model reaches this threshold then we construct a 95% confidence set of models that contains the minimum number of models whose sum to 0.95. The software then computes model-averaged results for the 95% confidence set. To calculate model averaged parameter estimates and unconditional standard errors we recalculate for each model performing the summation in the denominator of equation 3 across all models in the confidence set. The model weighted parameter estimates () are then calculated using equation 4 where is the recalculated model weight and is the parameter estimate from the model; the product of these values is summed across all models (R) in the confidence set.

(4)

Standard error estimates that are not conditional on any one model are calculated using equation 5.

(5)

The term represents the conditional variance of a parameter estimate under an individual model while is simply the squared deviation of the parameter estimate under a given model from the model weighted average for that parameter.

Our I-T approach also provides estimates of variable importance calculated by summing of all models (R) in which a CGE occurs (Eq. 6).

(6)

The score can provide evidence that a CGE is important even if its contribution is small or poorly defined. We also include a function in SAGA that provides a graphical representation of the model space that is being studied by plotting pixels for all models examined and coloring them based on their . This allows users to graphically examine the distribution of fits between all possible models and the available data.

For the remainder of the paper we denote CGEs using a capital letter for the source of an effect: autosomal (A), cytotype (C), maternal effect (M), X chromosome (X), Y chromosome (Y); and we use lower case to denote the type of effect: additive (a) or dominance (d). Two-locus epistatic CGEs are denoted by joining the single locus notations (e.g. AaAd is autosomal additive by autosomal dominance epistasis).

*Interpreting results of an I-T approach to LCA*

Interpreting LCA results using an I-T approach involves consideration of a number of factors not present in the traditional joint-scaling approach. In our experience, few LCA datasets strongly support a single model of genetic architecture, yet goodness-of-fit statistics and parameter estimates are typically only reported for the best model. A major advance of SAGA, and the I-T approach more generally, is the ability to quantify the degree of model selection uncertainty. In general the lower the Akaike weight of the best model and or the more models contained in the 95% confidence set the greater the degree of model selection uncertainty. Using the function (VisModSpace) the user can plot the distribution of Akaike weights across all possible models to allow a simple visual interpretation of model selection uncertainty.

In cases where there is considerable model selection uncertainty, scores can often provide a strong indication of the importance of a specific CGE’s role in the genetic architecture of a trait. This occurs because even when a 95% confidence set contains many models, one or a handful of CGEs may be present in the majority of them. In these situations, we can infer that the CGE is likely to be important, but we may have little confidence in estimating its true magnitude or sign because it depends on the other components in the model. By default SAGA will return scores and color the plot of parameter estimates to reflect .

After assessing model selection uncertainty and scores, we can evaluate the CGE estimates and their unconditional error estimates. Due to the inherent biases in previous joint-scaling approaches noted above, our model averaged parameter estimates will often be of lower magnitude and since the error now properly includes model selection uncertainty it will often be higher. SAGA returns a table containing the model averaged parameter estimate for all CGEs as well as the unconditional standard error for each.

Finally, we find that in some datasets there are a small number of models that are much better than all other models. If the user wishes to explore each model individually, they can use the function (EvalModel), to specify the model for which SAGA will return parameter estimates and standard errors conditional on that model. Despite the inclusion of this function, we strongly recommend reporting parameter estimates of a single model only if its Akaike weight is greater than 0.95 (i.e. the data suggest that there is a single best model).

*Validation Testing of the I-T Approach*

To validate our approach to LCA we created five simulated datasets based on empirical variation in sample size, standard deviation of cohort means, and mean phenotype for sperm receptacle length in *Drosophila* *mojavensis* (Miller et al. 2003). We generated the first four simulated datasets based on a simple genetic architecture where mean = 4.58 and three CGEs (Aa, Ma, AaAd) each with equal magnitude. In datasets 1-4 the magnitudes of CGE’s were 0.25, 0.5, 1, and 2 respectively. We also generated a fifth dataset with a more complex genetic architecture, again with a mean = 4.58, but with six CGEs (Aa, Ad, Ma, AaAa, AaAd, AdAd) that were all given a magnitude of 1. For each simulated dataset, we included the following mixed sex cohorts: P1, P2, F1 (P1xP2), rF1 (P2xP1), (P1xF1), (rF1xP1), (P2xF1), (rF1xP2); parents are indicated as sire x dam. We introduced sampling error to the simulated cohort means by randomly sampling 30 values from a normal distribution centered on the known cohort mean (specified by model parameters), and a standard deviation equal to 0.087 (the highest standard deviation recorded in the *D. mojavensis* dataset). For each of the five simulation conditions we generated 250 replicate datasets. The CGEs that we can evaluate with these cohorts are: Aa, Ad, Ca, Ma, Md, AaAa, AaAd, AdAd, CaAa, and CaAd. These 10 effects allow for more than 800 possible underlying models.

To demonstrate the performance of the I-T approach with empirical data we analyzed data from two previously published works (Miller et al. 2003; Demuth 2004). Miller *et al.* (2003) investigated the genetic architecture of male sperm length and female sperm receptacle length that have coevolved in subpopulations of *Drosophila mojavensis.* This study was chosen because is one of the only studies where the authors published the underlying data for sexed cohorts, and thus allows the investigation of sex chromosome effects. From this study, we include data for female sperm receptacle length and data for male sperm length*,* empirical datasets 1 and 2 respectively. Demuth (2004; see also Demuth and Wade 2007a, b) set out to describe the genetic architecture of divergence among a cosmopolitan sample of *Tribolium castaneum* populations. Here we reanalyze 15 datasets that we expect to harbor a wide range of genetic architectures. These are referred to as empirical datasets 3-17. The original analyses both implemented a joint- scaling approach using *X*2 to asses model fit and likelihood ratio tests to add CGEs to a more simple initial model; we contrast these to results with the I-T approach implemented in SAGA.

All analyses were computed using our software package SAGA version 1.0 loaded with RStudio version 0.98.976 running R version 3.02 on a MacBookPro with 4GB of 2600MHz RAM and a 2.5GHz processor (RStudio 2012; R Development Core Team 2013). Our R package is available from the CRAN repository and includes a vignette that guides users through an analysis of two empirical datasets.

**Results**

*Simulations Studies*

To validate the performance of an I-T approach to LCA using simulated data we report the accuracy of parameter estimates, as well as performance in identification of CGEs included in the generating model. In simulated datasets 1 - 4 we found that the magnitude of CGEs was accurately estimated in all cases across a range of effect sizes (Figure 1A-1D). In fact, even in dataset 1 where the magnitude of the CGEs was only 5% of the mean, the estimate of the magnitude of the CGEs was within 2% of the true value (Figure 1A). A key benefit of SAGA is that the parameter estimates are accurate despite substantial overall model selection uncertainty. The true model used to generate the simulated data was only identified as the best model in 73, 71, 73, and 69% of the replicates among simulated datasets 1 - 4 respectively. The fact that SAGA very accurately estimates parameter values while simultaneously failing to always identify the correct model as best, is likely due to CGEs that have small parameter estimates, large standard errors, and low but by chance explain enough stochastic variation in cohort means to be included in a subset of high scoring models.

Since the analysis of dataset 4 had the lowest success in identifying the generating model as best, we used it to investigate the impact of model misspecification on parameter estimates. We parsed the results from the 250 replicates based on whether the true generating model was identified as the best model (figure 1E) or not (figure 1F). Even in cases where the generating model was not identified as the best model, the individual parameter estimates remain accurate. There were no cases where the estimates for CGEs included in the generating model overlapped with the estimates for CGEs excluded from the generating model.

Dataset 5 offers the opportunity to evaluate a more complex genetic architecture where the number of CGEs generating the cohort means is approaching the maximum number of estimable parameters (8 cohorts – [6 CGEs + intercept] = 1 degree of freedom). Our analysis of this dataset shows that we can clearly distinguish the CGEs generating the line means based on scores. The minimum for a CGE included in the generating model was 0.659 while the maximum for a CGE excluded from the generating model was 0.020 (Table 1). We find that most CGEs are accurately estimated. In particular, Aa, Ma, AaAd have mean estimates of 0.994, 1.000, and 0.998 respectively, and scores of 0.993, 0.997, and 0.996 respectively. The three other CGEs included in the generating model (Ad, AaAa, AdAd) have more variable estimates ranging from 0.663 to 1.173 due to a linear dependency between the CGEs that precludes their joint estimation with the cohorts simulated. Thus, the parameter estimates for Ad, AaAa, AdAd are generated by fewer models (those that are missing at least one of the CGEs). Despite this, the parameter estimates and the scores for these three CGEs are all at least one order of magnitude greater than the estimates for CGEs that were not included in the generating model (Table 1).

*Analysis of Empirical Data:*

Analysis of the 17 empirical datasets in SAGA revealed the anticipated advantages of using an I-T approach: 1) successfully finding complex models where joint-scaling fails 2) providing a clear signal of model selection uncertainty 3) identifying CGEs that are often ignored; and 4) providing parameter and error estimates unconditional on any single model. The results for all empirical datasets are reported in Table 2. Below we briefly report results from the analyses of three datasets that illustrate the range of model selection uncertainty we have found in empirical datasets.

Analysis of empirical dataset 1 provides an example with low model selection uncertainty. Our analysis of all possible models indicated that a single complex model was far better than any alternatives. This model included 6 CGEs: Aa, Ma, AaAa, XaAa, CaAd, CaXa and had a of 0.976. The second best model had = 0.010; producing very strong evidence that the parameter estimates for the CGEs do not need to be model weighted (evidence ratio = 93.8; (Jeffreys 1948; Evett and Weir 1998)). Epistatic interactions are the most common (4 of 6), and have the largest magnitude AaAa (-0.540±0.018) and AaXa (0.634±0.020) of all CGEs included in the best model (Figure 2). Other CGEs had parameter estimates less than half this magnitude.

Dataset 2 provides an example of intermediate model selection uncertainty. In this case, no single model was found to be best (95% confidence set includes 302 models with the highest = 0.154). However, a clear signal from scores suggests that Aa and CaYa are important contributors to line means (0.91 and 0.79 scores respectively). These two CGEs also had the largest estimated magnitudes of 0.0495±0.0133 and 0.0420±0.0184 (figure 3). Most other CGEs estimated for this dataset had low scores, or in the case of Xa, though the score was 0.72 the magnitude of the estimate was comparatively small -0.0123±0.0077. This intermediate level of uncertainty where no single model is defined as far superior (Table 2) but a clear signal as to which CGEs are important was the most common result in our empirical analyses, found in datasets 2, 3-10, 11, 12, 14, and 15.

Finally, dataset 13 is an example where we find high model selection uncertainty. A 95% confidence set of models required the inclusion of 363 models and the highest = 0.034 (Table 2). The scores were also low, with the highest being assigned to Ca and CaAd both with scores of 0.43. Furthermore, the standard error of the estimates for all CGEs overlap zero (figure 4). This dataset illustrates that in some cases the degree of model selection uncertainty can be so high that LCA is not able to recover the genetic architecture of the trait of interest with the cohorts available. The only other empirical dataset in our analyses that showed a similar level of uncertainty was dataset 11, which contained 310 models in the 95% confidence set and all CGE estimates failed to exclude zero.

*Software Performance*

All analyses attempted in our study were completed in reasonable times on a standard laptop described in the methods section. Evaluation of the 250 replicates of all of the simulated datasets required a total of 12 minutes. Due to the nature of memory usage in R, performance on the empirical datasets 1 and 2 that contain independent data for males and females was initially a concern because it allows for a far greater number of models. For instance, the analysis of sperm receptacle length in *Drosophila mojavensis* females (dataset 1) required evaluation of approximately 13,000 models while the analysis of sperm length in *D. mojavensis* males (dataset 2) required evaluation of 63,000 models. However, these analyses required only 2 and 5 minutes respectively to complete.

**Discussion**

Existing approaches to LCA share two common shortcomings: 1) there is no framework to adequately describe model selection uncertainty and 2) there is no way to quantify the impact of model uncertainty on the estimated contributions of individual CGEs. The importance of model selection uncertainty is highlighted by our analysis of empirical data where 16 of 17 datasets showed non-trivial model selection uncertainty. This ability to quantify model selection uncertainty is perhaps one of the most important benefits of turning to an I-T approach. Previous analyses (even those that implemented AIC to choose a model) have presented only results conditional on specific models, and have largely ignored uncertainty in model selection. Furthermore, hypothesis-testing approaches do not provide us with a way to rank models relative to one another. For instance the result of the joint-scaling approach cannot tell us if one or many models are almost as good as the best model identified. Akaike weights and evidence ratios offer a natural way to do this.

The maximum of all models tested as well as the number of models required to produce a 95% confidence set are two simple metrics that quantify the degree of uncertainty in model selection. The maximum we recorded ranged from 0.03 to 0.98 with a mean maximum = 0.26. The number of models required to construct a 95% confidence model set varied accordingly, ranging from just a single model in the case of sperm receptacle length in *Drosophila mojavensis* (dataset 1) to over 300 models of productivity in *Tribolium castaneum* (datasets 2, 11, and 13) (Table 2). We illustrate examples where model selection uncertainty is low (dataset 4; Figure 5A) and high (dataset 6; Figure 5B). The model uncertainty metrics and visual depiction of model space allow for a more realistic interpretation of LCA experiments than previous approaches.

By implementing an I-T approach and examining all models possible given the data, we also resolve the issue of finding the best possible model. The potential of failing to find the best model was illustrated in our analysis of dataset 1 where we found a model that outperformed all other possible models that the joint-scaling test had failed to find. However, the ultimate goal of LCA is to find the composite genetic effects responsible for a phenotype. Previous methods depend on identifying the best model and interpreting the CGEs that are included in that model (conditional effects). With SAGA we get accurate estimates of the CGEs that are not dependent on the ability to specify one overall model as best, and our analysis of simulated datasets indicates that even when we are unable to identify the generating model our I-T approach is still able to identify the generating CGEs.

There has been concern that comparing the large number of possible models in LCA experiments may lead to spurious results (Bieri and Kawecki 2003). This concern seems to trace back to discussions of “data dredging” (Burnham and Andersen 1998; Burnham and Anderson 2002). Described in the context of ecological studies, data dredging is the process of measuring and searching for significance among a great many variables without a clear *a priori* decision of what variables may be biologically important. Burnham and Anderson encourage careful selection of a reduced set of variables based on a sound understanding of the biology involved and by doing this reducing the total number of models that must be evaluated (2002). In LCA, the variables are known CGEs, and each one describes a biologically plausible component of the genetic architecture underlying the phenotypes of the observed cohorts. The goal of LCA, finding the set of CGEs that best explains the observed data, can best be accomplished if we examine all possible combinations of CGEs. Assuming the necessary cohorts are available, the I-T approach accomplishes this goal.

Despite the improvements provided by implementing an I-T approach in SAGA, we would caution users to carefully consider a number of issues in interpreting results. First among these would be the inclusion of spurious variables, as we found in our analysis of simulated datasets 1-4 approximately 29% of iterations identified a model as best which included a CGE that was not in the generating model. These spurious variables are easily identified by small parameter estimates with standard errors overlapping zero.

Some previous studies have used AIC scores to choose a most parsimonious model (Bieri and Kawecki 2003; Fox et al. 2004; Fox et al. 2011). However, If the ratio of the number of cohorts (n) to the number of CGEs being evaluated (K) is less than 40, which will almost always be the case in LCA, then AICc is preferable(Burnham and Andersen 1998). AICc provides an appropriate trade off between model complexity and goodness of fit, and as sample size increases it converges on AIC (McQuarrie and Tsai 1998). The higher penalty assessed for additional parameters under AICc should help to reduce the risk of including spurious variables and overfitting.

Another issue that users should consider is the presence of linear dependencies in C-matrices. In our simulated dataset 5 the generating model included the CGEs: Aa, Ad, Ma, AaAa, AaAd, AdAd, and by the choice of our eight cohorts three CGEs (Ad, AaAa, and AdAd) became linearly dependent. SAGA deals with this by dropping any model that includes all three of the variables. Consequently, the importance and magnitude of each particular CGE involved in the linear dependency is then estimated only by the subset of models where it still appears. In our experience a strong signal of the importance of CGEs remains, but not surprisingly the parameter estimates become less accurate. When variables with high scores do not appear jointly in the equations included in the confidence set, it is a strong indication that there is a linear dependency, and warrants additional investigation. Often the only solution to this problem will be a careful examination of the C-matrix to determine what type of additional cohort(s) could be measured to allow the joint estimation of the variables of interest.

Examining the results from of all 17 empirical datasets suggests that an I-T approach is also more successful at identifying higher order CGEs. For example, 7 of the datasets had one or more epistatic CGEs (9 in total) not identified with the joint-scaling test that had a > 0.5 in the I-T analysis. Two datasets had maternal effects not identified with the joint-scaling test that had > 0.5 in the I-T analysis. Only one dataset had a non-epistatic autosomal CGE not identified with the joint-scaling test identified that had > 0.5 in the I-T analysis. These results indicate that the traditional forward variable selection version of the joint-scaling test may underestimate the contribution of epistatic interactions in determining phenotypes. Finding a larger role for epistasis in empirical datasets is particularly important since it affects our perception of how adaptation and speciation are likely to occur in nature. For instance, if complex architectures are common, processes such as conversion from epistatic to additive variation, may result in sub populations rapidly evolving different phenotypic outcomes to the same selective pressure due to variation in initial allele frequencies (Goodnight 1987, 1988; Wade and Goodnight 1998). Perhaps where our method will be most useful is in studies of speciation, where being able to finely parse the epistatic components of hybrid breakdown are important (Demuth and Wade 2005). Although LCA requires the production of crosses and backcrosses between divergent lines, if two species are in the so called “Goldilocks zone” where viable hybrids can still be produced, it is possible to use LCA to investigate the architecture of divergence in traits between species and the evolution of reproductive isolation (e.g.(Demuth et al. 2014).

The I-T approach to LCA we have presented eliminates issues in existing approaches and offers a more powerful and nuanced examination of the genetic architecture of quantitative traits. Furthermore, estimates of CGEs are unbiased and confidence intervals incorporate model selection uncertainty, a characteristic impossible under previous approaches. Finally, the ability to visualize the distribution of Akaike weights of all possible models can provide a strong indication of whether LCA of the phenotype of interest is informative. We recommend that future studies assess model uncertainty and shift away from making estimates that are conditional on a single model.

**Table 1.** Variable importance and parameter estimation for simulated dataset 5. Results are the mean of 250 replicate datasets. Bolded lines are variables included in the generating model.

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Variable Importance** | **Estimate / True Value** | **Unc. SE** |
| **Mean** | **1.000** | **4.746/4.58** | **0.773** |
| **Aa** | **0.993** | **0.994/1.00** | **0.016** |
| **Ad** | **0.659** | **0.662/1.00** | **1.545** |
| Ca | 0.005 | 0.007/0.00 | 0.007 |
| **Ma** | **0.997** | **1.000/1.00** | **0.013** |
| Md | 0.005 | 0.000/0.00 | 0.005 |
| **AaAa** | **0.66** | **0.827/1.00** | **0.774** |
| **AaAd** | **0.996** | **0.998/1.00** | **0.035** |
| **AdAd** | **0.67** | **1.173/1.00** | **0.774** |
| CaAa | 0.02 | 0.007/0.00 | 0.004 |
| CaAd | 0.014 | -0.007/0.00 | 0.003 |

**Table 2** Parameter estimates for models of the genetic architecture from empirical datasets. The analysis method is indicated as either I-T or J-S for information-theoretic or joint-scaling approach. The CSS column indicates the number of models included in the 95% confidence set. Only CGEs that had a of greater than 0.50 in at least one analysis are shown.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Dataset |  | max. | CSS | M | Aa | Ad | Ma | Md | Ca | Xa | AaAa | AaAd | AdAd | CaAa | CaAd | CaXa | CaYa | XaAa |
| 1 | I-T | 0.976 | 1 | 4.58 | **0.22** | - | **0.11** | - | - | - | **-0.54** | - | - | - | **-0.09** | **-0.25** | † | **0.63** |
|  | J-S |  |  | nr | 0.31 | - | - | - | - | - | - | - | - | † | † | † | † | - |
| 2 | I-T | 0.154 | 302 | 1.88 | **0.05** | 0.01 | 2.15 | -4.63 | -0.21 | **-0.01** | **-0.08** | -0.00 | -0.01 | - | 0.00 | 0.00 | **0.04** | -0.00 |
|  | J-S |  |  | 1.73 | 0.05 | 0.30 | - | - | - | -0.02 | 0.07 | - | -0.20 | † | † | † | † | - |
| 3 | I-T | 0.192 | 135 | 48.71 | -2.98 | 1.86 | 2.16 | **-4.64** | -0.21 | † | **-22.21** | **-4.49** | -27.23 | 0.11 | 2.37 | † | † | † |
|  | J-S |  |  | 58.99 | -1.88 | -38.12 | - | - | - | † | -30.50 | - | - | - | - | † | † | † |
| 4 | I-T | 0.508 | 110 | 33.85 | **-18.27** | 6.08 | **10.27** | **-10.87** | -0.11 | † | **-15.67** | **24.67** | -2.20 | 0.02 | 0.21 | † | † | † |
|  | J-S |  |  | 21.74 | -23.29 | 17.75 | 11.69 | -7.41 | - | † | - | 32.94 | - | - | - | † | † | † |
| 5 | I-T | 0.101 | 280 | 63.92 | 0.51 | **-38.64** | 0.34 | **-7.42** | -0.48 | † | -30.62 | 8.45 | 6.19 | 0.93 | -0.85 | † | † | † |
|  | J-S |  |  | 78.47 | - | -51.65 | - | -12.95 | - | † | -45.21 | 14.02 | - | - | - | † | † | † |
| 6 | I-T | 0.072 | 145 | 81.26 | 5.10 | -4.28 | -1.32 | 0.87 | 0.08 | † | 1.61 | -1.30 | -3.32 | **-25.04** | -0.05 | † | † | † |
|  | J-S |  |  | 36.69 | 22.18 | 71.82 | -7.76 | 16.14 | -1.79 | † | 23.11 | -18.53 | -39.72 | - | - | † | † | † |
| 7 | I-T | 0.06 | 216 | 41.92 | -1.96 | 9.90 | -1.64 | **25.50** | -1.76 | † | -5.31 | **-21.72** | 4.70 | 5.53 | -0.08 | † | † | † |
|  | J-S |  |  | 42.62 | - | - | -6.32 | 25.17 | - | † | - | -26.47 | 18.03 | - | - | † | † | † |
| 8 | I-T | 0.381 | 136 | 61.26 | **-40.07** | -0.71 | **12.00** | 2.06 | **18.95** | † | **-8.95** | **106.4** | -1.28 | -0.94 | **-36.17** | † | † | † |
|  | J-S |  |  | 65.88 | -21.81 | -1.25 | 7.68 | 3.11 | 1.07 | † | -17.67 | 77.54 | -17.31 | - | - | † | † | † |
| 9 | I-T | 0.052 | 219 | 21.46 | -1.69 | 7.91 | 0.62 | 0.96 | 0.12 | † | **14.06** | **49.19** | 6.43 | -0.28 | 1.59 | † | † | † |
|  | J-S |  |  | -13.41 | -13.65 | 56.29 | 8.89 | 11.74 | - | † | 52.06 | 42.42 | - | - | - | † | † | † |
| 10 | I-T | 0.432 | 58 | 3.66 | **20.25** | **51.85** | **-11.96** | -0.24 | 1.88 | † | **16.61** | **-44.70** | **-29.04** | -0.12 | -3.19 | † | † | † |
|  | J-S |  |  | - | 22.55 | 63.23 | -12.78 | - | - | † | 19.41 | -47.89 | -37.25 | - | - | † | † | † |
| 11 | I-T | 0.029 | 310 | 31.88 | -2.28 | 13.51 | 0.59 | 2.00 | -0.07 | † | 8.09 | -0.51 | 0.91 | -0.56 | 2.41 | † | † | † |
|  | J-S |  |  | - | - | 54.30 | - | 10.74 | - | † | 39.62 | - | - | - | - | † | † | † |
| 12 | I-T | 0.32 | 119 | -41.88 | -0.52 | **97.15** | **11.55** | **16.45** | -0.09 | † | **80.93** | 7.05 | -4.73 | -1.79 | -0.13 | † | † | † |
|  | J-S |  |  | -69.54 | -18.07 | 151.43 | 14.11 | 23.02 | -1.54 | † | 114.0 | 45.78 | -33.29 | - | - | † | † | † |
| 13 | I-T | 0.034 | 363 | 27.91 | -4.38 | 4.97 | 1.03 | 1.36 | 2.60 | † | -1.93 | 4.44 | 4.52 | -1.09 | 2.22 | † | † | † |
|  | J-S |  |  | 25.51 | -14.01 | - | - | 5.77 | 6.40 | † | - | 17.59 | 14.56 | - | - | † | † | † |
| 14 | I-T | 0.378 | 20 | 97.79 | -0.07 | **-181.63** | 0.04 | **-16.67** | 0.01 | † | **-38.94** | -0.24 | **158.1** | 0.02 | -1.54 | † | † | † |
|  | J-S |  |  | 98.57 | -0.66 | -183.39 | 2.34 | -16.32 | -2.06 | † | -39.44 | -3.74 | 159.1 | - | - | † | † | † |
| 15 | I-T | 0.23 | 25 | 63.62 | 4.72 | **-123.45** | -0.83 | 0.06 | **-4.40** | † | -4.19 | **61.34** | **131.3** | -0.01 | **-2.35** | † | † | † |
|  | J-S |  |  | 78.56 | 8.31 | -154.18 | -3.56 | - | -5.39 | † | -19.16 | 57.24 | 147.1 | - | - | † | † | † |
| 16 | I-T | 0.193 | 16 | 73.56 | **34.75** | **-148.68** | **-17.62** | -0.25 | -0.42 | † | -0.15 | -0.13 | **142.2** | 2.52 | -2.26 | † | † | † |
|  | J-S |  |  | 76.81 | - | -147.04 | -15.72 | -3.75 | -3.02 | † | - | 65.02 | 136.7 | - | - | † | † | † |
| 17 | I-T | 0.089 | 184 | 40.63 | **-7.77** | -0.10 | 2.10 | 1.20 | 1.20 | † | **-16.77** | -10.22 | -1.13 | **14.39** | 2.18 | † | † | † |
|  | J-S |  |  | 39.07 | -9.75 | - | - | - | 5.66 | † | - | -24.41 | - | - | - | † | † | † |

† composite effect not included in analysis; - composite effect not included in the estimating model(s); nr: not reported; bold indicates > 0.50

* + - 1. Armbruster, P., W. E. Bradshaw, and C. M. Holzapfel. 1997. Evolution of the genetic architecture underlying fitness in the pitcher-plant mosquito, *Wyeomyia smithii*. Evolution 51:451-458.
      2. Basford, K. E. and I. H. De Lacy. 1979. The use of matrix specifications in defining gene action in genotypic value models and generation mean analysis. Theoretical and Applied Genetics 55:225-229.
      3. Bentz, B. J., R. R. Bracewell, K. E. Mock, and M. E. Pfrender. 2011. Genetic architecture and phenotypic plasticity of thermally-regulated traits in an eruptive species, Dendroctonus ponderosae. Evolutionary Ecology 25:1269-1288.
      4. Bieri, J. and T. J. Kawecki. 2003. Genetic Architecture of Differences between Populations of Cowpea Weevil (Callosobruchus Maculatus) Evolved in the Same Environment. Evolution 57:274.
      5. Box, G. E. and N. R. Draper. 1987. Empirical model-building and response surfaces. John Wiley & Sons, New York.
      6. Burnham, K. P. and S. B. Andersen. 1998. Model Selection and Inference: a practical information-theoretic approach. Springer-Verlag, New York.
      7. Burnham, K. P. and D. R. Anderson. 2002. Model selection and multimodel inference: a practical information-theoretic approach. Springer, New York.
      8. Cavalli, L. L. 1952. An analysis of linkage in quantitative inheritance. Papers read at a colloquium held at the Institute of Animal Genetics Edinburgh University under the auspices of the Agricultural Research Council. HM Stationery Office.
      9. Coyne, J. A., N. H. Barton, and M. Turelli. 1997. Perspective: A critique of Sewall Wright's shifting balance theory of evolution. Evolution 51:643-671.
      10. Demuth, J. P. 2004. Evolution of Hybrid Incompatibility in the beetle Tribolium Castaneum. Pp. 152. Biology. Indiana University, Bloomington.
      11. Demuth, J. P., R. J. Flanagan, and L. F. Delph. 2014. Genetic architecture of isolation between two species of Silene with sex chromosomes and Haldane's rule. Evolution 68:332-342.
      12. Demuth, J. P. and M. J. Wade. 2005. On the theoretical and empirical framework for studying genetic interactions within and among species. American Naturalist 165:524-536.
      13. Demuth, J. P. and M. J. Wade. 2006. Experimental Methods for Measuring Gene Interactions. Annual Review of Ecology, Evolution, and Systematics 37:289-316.
      14. Demuth, J. P. and M. J. Wade. 2007a. Population differentiation in the beetle Tribolium castaneum. II. Haldane'S rule and incipient speciation. Evolution 61:694-699.
      15. Demuth, J. P. and M. J. Wade. 2007b. Population differentiation in the beetleTribolium castaneum. I. Genetic architecture. Evolution 61:494-509.
      16. Derksen, S. and H. J. Keselman. 1992. Backward, forward and stepwise automated subset selection algorithms; Frequency of obtaining authentic and noise variables. The British jounal of mathematical and statistical psychology 45:265-282.
      17. Edmands, S. 1999. Heterosis and outbreeding depression in interpopulation crosses spanning a wide range of divergence. Evolution 53:1757-1768.
      18. Evett, I. W. and B. S. Weir. 1998. Interpreting DNA evidence: statistical genetics for forensic scientists. Sinauer Associates, Sunderland.
      19. Fisher, R. A. 1941. Average excess and average effect of a gene substitution. Annals of Eugenics 11:53-63.
      20. Fisher, R. A. 1958. The genetical theory of natural selection. Dover Publications, New York.
      21. Fox, C. W., M. E. Czesak, and W. G. Wallin. 2004. Complex genetic architecture of population differences in adult lifespan of a beetle: nonadditive inheritance, gender differences, body size and a large maternal effect. J Evol Biol 17:1007-1017.
      22. Fox, C. W., J. D. Wagner, S. Cline, F. A. Thomas, and F. J. Messina. 2011. Rapid Evolution of Lifespan in a Novel Environment: Sex-Specific Responses and Underlying Genetic Architecture. Evolutionary Biology 38:182-196.
      23. Fritz, R. S., C. G. Hochwender, B. R. Albrectsen, and M. E. Czesak. 2006. Fitness and Genetic Architecture of Parent and Hybrid Willows in Common Gardens. Evolution 60:1215.
      24. Fuller, R. C. 2008. Genetic incompatibilities in killifish and the role of environment. Evolution 62:3056-3068.
      25. Galloway, L. F. and C. B. Fenster. 2001. Nuclear and cytoplasmic contributions to intraspecific divergence in an annual legume. Evolution 55:488-497.
      26. Gilchrist, A. S. and L. Partridge. 1999. A comparison of the genetic basis of wing size divergence in three parallel body size clines of Drosophila melanogaster. Genetics 153:1775-1787.
      27. Goodnight, C. J. 1987. On the effect of founder events on epistatic genetic variance. Evolution 41:80-91.
      28. Goodnight, C. J. 1988. Epistasis and the effect of founder events on the additive genetic variance. Evolution 42:441-454.
      29. Hayman, B. I. 1958. The seperation of epistatic from additive and dominance variation in generation means. Heredity 12:371-390.
      30. Hurvich, C. M. and C.-L. Tsai. 1989. Regression and time series model selection in small samples. Biometrika 76:297-307.
      31. Jeffreys, H. 1948. Theory of Probability. Oxford University, Oxford, UK.
      32. Lair, K. P., W. E. Bradshaw, and C. M. Holzapfel. 1997. Evolutionary divergence of the genetic architecture underlying photoperiodism in the pitcher-plant mosquito, Wyeomyia smithii. Genetics 147:1873-1883.
      33. Lande, R. and S. J. Arnold. 1983. The Measurement of Selection on Correlated Characters. Evolution 37:1210-1226.
      34. Lynch, M. and B. Walsh. 1998. Lynch, M., & Walsh, B. (1998). Genetics and analysis of quantitative traits. Sinauer Associates, Inc., Sunderland, Massachusetts.
      35. Mather, K. and J. L. Jinks. 1982. Biometrical genetics: The study of continuous variation. Chapman and Hall, London.
      36. McQuarrie, A. D. and C.-L. Tsai. 1998. Regression and time series model selection. World Scientific Publishing Company, Singapore.
      37. Miller, G. T., W. T. Starmer, and S. Pitnick. 2003. Quantitative genetic analysis of among-population variation in sperm and female sperm-storage organ length in Drosophila mojavensis. Genetical Research 81:213-220.
      38. R Development Core Team. 2013. R: A Language and Environment for Statistical Computing, Vienna, Austria.
      39. RStudio. 2012. RStudio: Integrated development environment for R (Version 0.98.976), Boston, MA.
      40. Schiffer, M., A. S. Gilchrist, and A. A. Hoffmann. 2006. The Contrasting Genetic Architecture of Wing Size, Viability, and Development Time in a Rainforest Species and Its More Widely Distributed Relative. Evolution 60:106.
      41. Tymchuk, W. E., L. F. Sundstrom, and R. H. Devlin. 2007. Growth and survival trade-offs and outbreeding depression in rainbow trout (Oncorhynchus mykiss). Evolution 61:1225-1237.
      42. van Heerwaarden, B., Y. Willi, T. N. Kristensen, and A. A. Hoffmann. 2008. Population bottlenecks increase additive genetic variance but do not break a selection limit in rain forest Drosophila. Genetics 179:2135-2146.
      43. Wade, M. J. 2000. Epistasis as a genetic constraint within popualtions and an accelerant of adaptive divergence among them. Pp. xiii, 330 p. *in* J. B. Wolf, E. D. Brodie, and M. J. Wade, eds. Epistasis and the evolutionary process. Oxford University Press, Oxford England ; New York.
      44. Wade, M. J. 2002. A gene's eye view of epistasis, selection and speciation. Journal of Evolutionary Biology 15:337-346.
      45. Wade, M. J. and C. J. Goodnight. 1998. Perspective: the theories of Fisher and Wright in the context of metapopulations: when nature does many small experiments. Evolution 52:1537-1553.
      46. Whittingham, M. J., P. A. Stephens, R. B. Bradbury, and R. P. Freckleton. 2006. Why do we still use stepwise modelling in ecology and behaviour? The Journal of animal ecology 75:1182-1189.
      47. Wilkinson, L. 1979. Tests of Significance in Stepwise Regression. Psychological Bulletin 86:168-174.
      48. Wright, S. 1931. Evolution in Mendelian populations. Genetics 16:97-159.

**Figure 1. Analysis of 4 simulated datasets with varying magnitude of CGE contribution.** Each plot shows the distribution of model weighted parameter estimates from 250 replicates. In each plot curve 1 is for CGEs with a true value of 0, curve 2 for CGEs that generated the line means and varied from 0.25 to 2, curve 3 is for the estimate of the mean. In all cases, the dashed red line indicates the generating parameter values. **A**) Dataset 1 effect size of 0.25. **B**) Dataset 2 effect size of 0.5. **C**) Dataset 3 effect size of 1 **D**) Dataset 4 effect size of 2 **E**) Dataset 4 Iterations that correctly identified the true model **F**) Dataset 4 Iterations that failed to identify the true model.

**Figure 2. Parameter estimates for sperm receptacle length in *Drosophila mojavensis*.** Bars indicate the magnitude of the genetic effect indicated on the X axis and whiskers provide the conditional standard error.

**Figure 3.** Model weighted parameter estimates for the genetic architecture underlying sperm length in *Drosophila mojavensis*. Bars are colored based on scores and indicate the magnitude of the genetic effects indicated on the X axis. Whiskers indicate the unconditional standard errors. Only CGEs with of at least 0.15 are included.

**Figure 4.** Model weighted parameter estimates for the genetic architecture underlying reproductive isolation in *Tribolium castaneum*. Bars are colored based on scores and indicate the magnitude of the genetic effects indicated on the X axis. Whiskers indicate the unconditional standard errors.

**Figure 5.** Visual depictions of model space for the genetic architecture of reproductive isolation in *Tribolium castaneum*. Each box represents a genetic architecture model and is colored to reflect its . The color scale ranges from white for models with of zero to red for those with the maximum value for the analysis. A) Results from the analysis of dataset 4 showing low model selection uncertainty. B) Results from the analysis of dataset 6 showing considerable model selection uncertainty.