Title: The null distribution of genotypic variance and the relationship between P and G matrices.

Abstract (200 words):

Key words (6): Quantitative, Genetic/s, Correlation, evolution, constrained,

Data Archival Location (DRYAD/GIthub):

Author: Alexander Baker

Affiliation: School of Biology, University of Queensland, St Lucia, QLD 4072, Australia

Introduction (dot points) (1-2 pages?):

Context:

Phenotypic and genotypic matrices used in predictive breeding, evolutionary studies.

Constraints in number of dimensions in G based on P

Work done:

Simulation, Bayesian methods, dimensionality of QG data. Random selected expression data seeing underlying pleiotropy in drosophila. Few direct comparison of P and G or done on isolated datasets. RMT predicting values and sometime structure. Known distribution for P matrix entries and eigenvalues.

What we don’t know.

Underlying distribution of G is unknown. No rigorous statistical measures to see if G eigenvalues and structure differs from some null. Explicit numerical hypothesis testing is difficult.

What I did:

Simulated random data. Looked at null distribution in even sire design. Collated data based on Pitcher’s et al to compare P and G correlation matrices. Collected and compared eigenvalues of P and G. Project P through G. (Projection captures some of the loss of variance from shifting P into G.)?

Methods:

Simulation:

Independent multivariate normal data was simulated using the MASS package in R. For each simulated dataset 250 individuals measured on 5 independent N(0,1) traits were created. These individuals were split into 50 lines with 5 replicates per line. The phenotypic (P) and genetic (G) covariance matrices were estimated. The MANOVA estimation of G was used were G=(B-W)/r were B is the between line covariance matrix, W is the within line covariance matrix and r is the number of replicated per line. 200 datasets were simulated with P and G’s eigenvalues and the projection of P’s eigenvector through G being recorded.

Extraction of data from Pitcher’s dataset.

The dataset from Pitcher’s et al 2014 was used to create a list of paired G and P matrices. Papers for each entry in Pitcher’s matrix index were retrieved by searching by searching organism, author and publication data on UQ summon2.0. If no result was retrieved the publication time was loosened and the search was repeated using google scholar. All of the retrieved papers were then searched for P matrices which matched the G matrices used by Pitcher. P matrices with no clear pairing were discarded.

Matrices from Nor1997, Con2003, Par2005 and Con1992 were removed after initial compilation. Par2005 had omitted entries in enough rows and columns to prevent reduction, Con1992 had multiple correlation >2 off of the diagonal in its G matrix and upon further inspection both Con2003 and Nor1997 did not have a set of paired matrices. Del1995 had omitted entries. However, they were all in two rows. These rows and their associated columns were removed.

Processing the sets of P and G matrices:

Trait naming in all P matrices was made consistent with that found in Pitcher’s G matrices. An index file, Pmatindex, was created to stored manually entered pair information between the P and G matrices based on information in the papers. A function was then created retrieve a set of paired matrices, check the trait names in G and retrieve the shared traits from P. This created a new P submatrix for each G in the index file.

All P submatrices shared a name with their paired G and were written into a different folder. The number of matrices of any one dimension greater than 5 was small (<20). Consequently, a sampling function randomly selected n traits from all matrices of size >n and created a new submatrix of those traits. The eigenvalues of each P and G matrix were stored and for each P its eigenvectors were projected through the paired G.

Results:

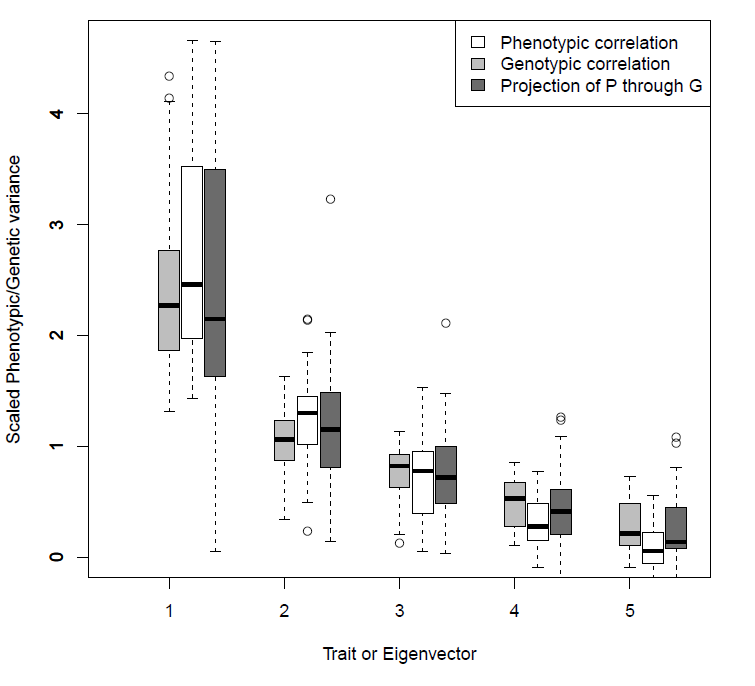
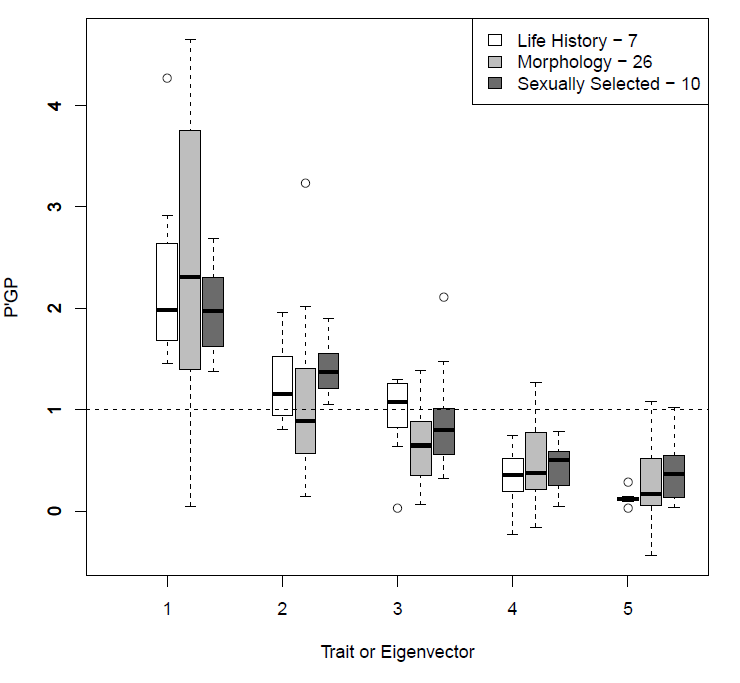
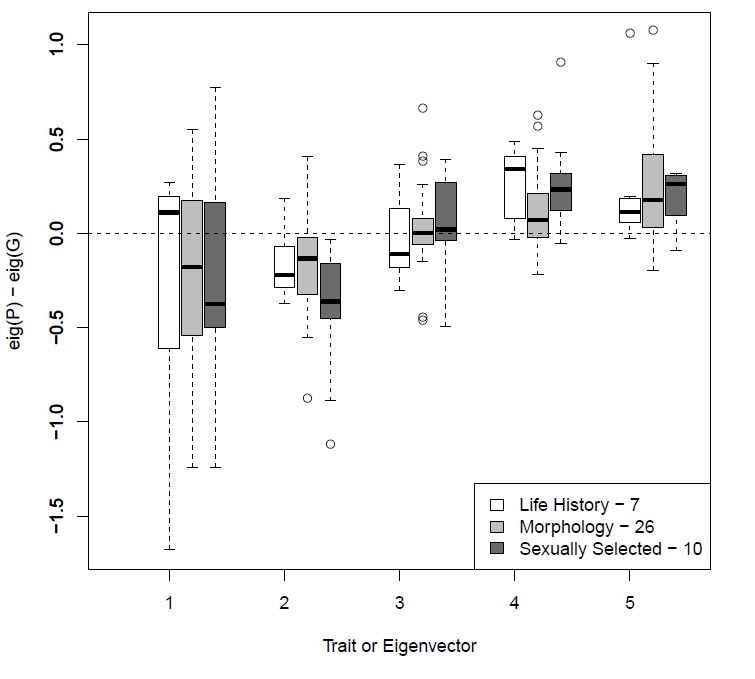
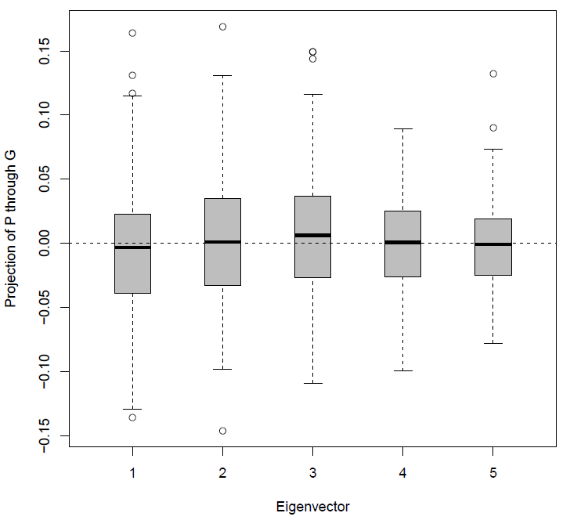


Figure 1: a) 250 individuals were simulated each with 5 N(0,1) distributed traits. The individuals were split into 50 lines with 5 replicates per line per dataset and 200 datasets were simulated. The phenotypic and genotypic covariance were calculated. The genetic covariance was estimated using an unconstrained MANOVA method. Then each of the five eigenvectors of P were projected through their paired G to get the results. B) The difference in eigenvalues of 43 paired P and G correlation matrices composed of 5 traits. G matrices have come from the Pitcher’s et al 2014 dataset and the corresponding P matrix was retrieved. Originally there were 18 matrices of size 5. Hence, larger matrix were randomly sampled to create more 5 trait matrices. The matrices were then split by trait type as defined by Pitcher’s et al and the difference of their eigenvalues was taken and placed on this graph. C) The projection as mentioned in A) of the P and G matrices from Pitcher’s after they had been split by trait type. D) The P and G matrix eigenvalues as well as the projection of P through G for the merged dataset of n=43 matrices.

Results:

Null P’GP data shows clear lack of linear trend. All values centred around 0 with large variability about it. Suggests that null P eigenvectors are close to orthogonal to G such that little of the variance in P was captured in G.

Evidence for difference in trait the median values of PGP is poor. Large Q1-Q3 range about the medians for all traits types encompasses other functions median (fig1b). However, conventional tests cannot be used due to lack of underlying distribution so a strong conclusion cannot be made. Fig1b suggests that using P’s eigenvalues to predict G will result in underestimation of the leading eigenvalues of G and overestimation of the trailing eigenvalues of G.

There was little difference in the projections of P through G between trait types except for eigenvector 5 where life history is condensed around its median compared to morphology and sexually selected traits. The projection of P showed a decreasing trend with ~40% of the variance being condensed into eigenvector 1.

Discussion:

####

Underlying distributions of G are unknown so it may be difficult to rigorously test any hypotheses posed. The distribution of the different of P and G is also unknown. Leads to comparison with simulated data, qualitative rather than quantitative comparisons.

Whole article is capped at 4500 words - maybe discussion should be 2k, methods and results 5-800 words, intro 1k, conc-200? See how it goes I guess.

####

Conclusions:

#### 15/01/15

Evolutionary rates for multivariate traits: the role of selection and genetic variation.

The distribution of genetic variance across phenotypic space and the response to selection.

Characterizing the evolution of genetic variance using genetic covariance tensors

Variation, selection and evolution of function-valued traits.

Evolutionary constraints in high-dimensional trait sets.

Determining the Effective Dimensionality of the Genetic Variance–Covariance Matrix

Characterizing the evolution of genetic variance using genetic covariance tensors

Phenotypic Divergence along Lines of Genetic Variance

Variation, selection and evolution of function-valued traits.

New Answers for Old Questions: The evolutionary Quantitative genetics of Wild Animal populations

#### 16/01/15

Comparing G: multivariate analysis of genetic variation in multiple populations

Studying phenotypic evolution using multivariate quantitative genetics.

Variance component estimation in animal breeding: a review.

A simple procedure for the comparison of covariance matrices.

An evolutionary relationship between genetic variation and phenotypic fluctuation

Comparing environmental and genetic variance as adaptive response to fluctuating selection

Phenotypic Selection in Natural Populations: What Limits Directional Selection?

Phenotypic plasticity facilitates mutational variance, genetic variance, and evolvability along the major axis of environmental variation

The quantitative genetics of phenotypic variation in animals

Patterns of quantitative genetic variation in multiple dimensions

#### 20/01/15

A REASSESSMENT OF GENETIC LIMITS TO EVOLUTIONARY CHANGE

Patterns of quantitative genetic variation in multiple dimensions

Read More: http://www.esajournals.org/doi/abs/10.1890/04-1209

####

#### Archive data on DRYAD or similar service in last week or two before end of project.