universität wien

MASTERARBEIT | MASTER'S THESIS

Titel | Title

From the pattern backwards to the process
What correlations of traits can tell us about their underlying genetic architecture and co-development

verfasst von | submitted by

Philipp Hummer BSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 220
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Joint-Masterstudium Evolutionary Systems Biology
Betreut von | Supervisor:	Univ.-Prof. Mag. Dr. Mihaela Pavlicev

Abstract (deutsch)

Evolution hängt von vererbbarer phänotypischer Variation ab, welche wiederum davon beeinflusst wird, wie der Genotyp zum Phänotyp führt. Dabei ist die Pleiotropie ein wichtiger Aspekt dieser Genotyp-Phänotyp-Karte, welche beschreibt wie Mutationen einen Effekt auf mehrere Merkmale haben. Die pleiotrope Struktur verursacht Einschränkungen in der Variation, welche zu einem großen Teil die Struktur der Entwicklung (welche Gene einen Einfluss auf welche Merkmale während der Entwicklung und in der Physiologie haben) widerspiegeln. Die Selektionsantworten von Merkmalen, die viele Gene teilen, müssen daher auch korrelieren. Je nach Richtung der Selektion kann dies dabei auch nachteilig sein. Um solch eine pleiotrope Struktur darzustellen hat Wagner (1984) die **B**-Matrix als eine mathematische Beziehung zwischen der Variation innerhalb der Loci und phänotypischen Kovarianzmatrizen, wie auch der Mutations-Kovarianzmatrix **M**, definiert. Mithilfe dieser Beziehung kann man untersuchen, wie sich die Struktur von **B** auf die von **M** auswirkt, welche wiederum mithilfe von Eigen-Dekomposition und hierarchischer Clusteranalyse charakterisiert werden kann. Des Weiteren beschäftigt sich diese Arbeit mit der umgekehrten Frage, was einem die Muster in der Korrelationsmatrix **rM** über die ihr zugrunde liegende pleiotrope Struktur sagen können. Dabei offenbart diese Studie, dass dieser Ansatz sehr sensibel gegenüber der Menge von eingeführter stochastischer Variation ist. Ist diese zu groß, löscht sie die Information aus, welche unterscheidbare Muster hätte aufzeigen können. Der zweite Teil dieser Arbeit beschäftigt sich mit dem analytischen Zusammenhang von Pleiotropie und Korrelation. Wenn es keine pleiotrope Assoziation zwischen zwei Merkmalen gibt, können diese auch nicht korrelieren (in **rM**), allerdings kann ein Fehlen von Korrelation auch daher kommen, dass sich positive und negative pleiotrope Effekt auslöschen, was man als „versteckte Pleiotropie“ bezeichnet. Jedoch haben die Analyse von konstruierten **B**-Matrizen und numerische Simulationen gezeigt, dass dieses Phänomen vor allem bei vielen Merkmalen sehr unwahrscheinlich ist. Außerdem hat sie auch gezeigt, dass man deshalb bei einem Fehlen von Korrelationen zwischen Merkmalen auch eher ein Fehlen von Pleiotropie in der (modularen) G-P-Karte erwarten sollte. Dabei hat sich bei dem Versuch genau zu errechnen, wie wahrscheinlich es ist eine versteckt-pleiotrope **B**-Matrix zufällig zu bekommen, auch gezeigt, dass die Art und Weise, wie zwei Merkmale über Pleiotropie miteinander verknüpft sind, jenseits von ihrer eigenen Korrelation bestimmt wie wahrscheinlich sie wie stark mit einem dritten Merkmal korrelieren.

Abstract

Evolution depends on heritable phenotypic variation, which is influenced by how the genotype maps onto the phenotype. One potent aspect of the "G-P map" is pleiotropy, which describes mutations having effects on multiple traits. Pleiotropic structure causes variational constraints, which in large part reflect the developmental structure (which genes affect which traits during development and physiology). In other words, the selection response of traits with many shared genes will be correlated. Depending on the direction of selection, this can be disadvantageous. To describe such a pleiotropic structure, Wagner (1984) has formalized the **B** matrix as a mathematical relation between the variation occurring at the loci and the phenotypic covariance matrix, such as the mutational covariance matrix **M**. Using this relation, one can explore how the structure of **B** affects the structure of **M**, which can be characterized by eigen-decomposition and hierarchical clustering. Moreover, this thesis asks the reverse question, namely to what extent the patterns in the correlation matrix **rM** can reveal the underlying pleiotropic structure. The study reveals that this is sensitive to the amount of introduced stochastic variation, which will remove the information that could reveal distinct patterns if it is too high. This thesis 'second part then analytically addresses the connection between pleiotropy and correlation. No pleiotropy between traits leads to lack of trait correlation (and thus lack of constraint), however, a lack of correlation can also be generated by pleiotropic effects canceling each other out, which is referred to as "hidden pleiotropy". Yet, the analysis of constructed **B** matrices and numerical simulations show that hidden pleiotropy – particularly for many traits – is very improbable and that a lack of pleiotropy in the (modular) G-P map is a more likely explanation for encountering a lack of trait correlations. Additionally, attempting to calculate the exact probabilities for achieving hidden pleiotropy from randomness has revealed that the way two traits are pleiotropically associated determines the probability of their correlation with a third trait beyond just their own correlation.

Contents

Introduction

Evolutionary processes depend on the amount and distribution of heritable phenotypic variation, which is itself strongly influenced by the processes that map the genotypic (genetic) to the phenotypic (trait) variation, i.e. the genotype-phenotype map. Such a map is summarizing variational contributions across the developmental and physiological processes at different levels of organization. The patterns of gene effects of single genes that contribute to the general structure of the G-P map (e.g., pleiotropy, polygeny or epistasis), however, are hard to come by experimentally. The goal of this thesis is to explore the question of what one can learn about the G-P map – particularly in terms of pleiotropy – from more easily available information on the distribution of mutational variation and covariation among traits.

One of the central structural components of G-P mapping is pleiotropy, which describes the phenomenon of a gene contributing variation to more than one trait. This is important because the pleiotropy between traits essentially affects traits' evolvability. The theoretically possible degree of pleiotropy is on the spectrum between a simple one-to-one mapping and "universal pleiotropy". While the one-to-one mapping between genes and traits has been used historically, population genetic theory often uses the assumption of universal pleiotropy. In such a case, in which all genes would affect all traits, mutational effects would be strongly limited, as was suggested by Ronald A. Fisher in his "geometric model" (Fisher 1930; reviewed in Pavlicev & Wagner 2012). Orr (2000) later expanded this view on multiple traits and termed the relationship the "cost of complexity" (Orr 2000). Empirical reports on degree of pleiotropy later revealed that pleiotropy is not universal at all but rather very restricted (Wagner & Zhang 2011) and e.g., Wagner et al. (2008) report a median degree of pleiotropy (meaning how many traits a mutation affects) of 8.6% of observed traits in their large-scale QTL study on the bony skeleton of inbred mice.

Apart from the numbers of traits affected, the aspect of pleiotropy which is relevant to evolvability is which traits share genes. This is relevant, because traits might find themselves in the same or in different selection regimes. For example, if mutations affect both a trait under directional selection and others that are under stabilizing selection, the first trait´s response to that selection will be significantly hindered (Baatz & Wagner 1997). Interestingly, in real organisms, distribution of pleiotropic effects across traits appears to not be random. This is manifested in the distribution of covariances, which is far from what may be expected for random G-P map structures. Instead, real empirical structures are more specific (Wagner 1989): some parts of the organism are more connected among one another by shared genes than they are with other parts. Such a lack of connection between parts is typical of variational modularity (Wagner & Altenberg 1996). This naturally suggests that also the pleiotropy in G-P maps of real biological systems is modular. Pleiotropy is thus restricted, such that only relatively

few traits are affected by any one gene. These distinct modules tend to contain functionally linked traits, which thus not only covary but are under common selection pressure and consequently coevolve, whereas the variational modules can respond to conflicting selection pressures separately from one another. For example, the eye is not affected by evolutionary changes in the hindlimb in some tetrapod lineage. Modularity is thus considered to have a positive effect on evolvability (Wagner & Altenberg 1996; Pavlicev & Wagner 2012). Theory has also shown that modularity can increase the rate of adaptation, even though a minimum cost of complexity will always apply (Welch & Waxman 2003).

To reveal the pleiotropic structure of the G-P map in practice is nonetheless difficult, as the available information, the genetic covariance matrix, summarizes the (co)variance across loci. In order to reflect on how the structure of the G-P map affects the variational distribution in the population, Wagner (1984) developed the concept of the **B** matrix, representing the effects of single genes on single traits (See Fig. 1). The **B** matrix will serve here as a useful framework. It is a model of mapping a genotype to its phenotype, with *n*X*k* dimensions, where *n* denotes the number of traits and *k* denotes the number of polymorphic loci (capital X will be used to denote matrix dimensions). In our use, $B_{x,y}$ is the additive genetic effect that the allele *y* has on the trait *x* which can be positive, negative or absent. This simple system makes the **B** matrix a very intuitive tool to encode pleiotropic structure, including modular arrangements (Pavlicev & Hansen 2011).

The **B** matrix after Wagner (1984)

	locus 1	**locus 2**	**...**	**locus *k***
trait 1	$b_{1,1}$	$b_{1,2}$	...	$b_{1,k}$
trait 2	$b_{2,1}$	$b_{2,2}$	...	$b_{2,k}$
:	:	:		:
trait *n*	$b_{n,1}$	$b_{n,2}$	...	$b_{n,k}$

Figure 1: A representation of the **B** matrix (Wagner 1984) that models the pleiotropic structure of the G-P map. The *n* rows signify the traits and the *k* columns the polymorphic loci. Each entry $b_{x,y}$ is then the additive genetic effect that locus *y* has on trait *x*, which can be positive, negative or absent.

An allele *y* is pleiotropic if the column $B_{1,y}$, $B_{2,y}$, … $B_{k-1,y}$, $B_{k,y}$ contains more than one non-zero value (Wagner 1989). Thus, if two or more traits share a pleiotropic allele, they can be considered pleiotropically connected and sets of traits with shared loci can then form modules, dependent on other genes. Of course, the **B** matrix is a model and an abstraction, which is hard to measure directly. While genome-wide genetic effects of loci are assessed in mapping studies, these are often focusing on only single or few traits. Additive heritable phenotypic variance of a population is captured as the covariance matrix of the genetic effects **G**, summarizing contributions of all polymorphic loci on all

included traits (the multivariate generalization of the univariate additive genetic variance V_G). **G** can be used to predict one-generation multivariate response to selection (Lande 1979). The relation of the **B** matrix to the **G** matrix is the following: **B** X **V** X $\mathbf{B}^t$ = **G**, in which **V** is the variance-covariance matrix of the alleles (and the uppercase t stands for transposed) (Wagner 1989; Pavlicev & Hansen 2011). **V** is dependent on population structure, allele frequencies and linkage. It is empirically measurable, but constructing it requires a lot of sequencing, good knowledge of the population and most importantly, knowledge of the variation of which genes are involved in the variation of which focal traits. In principle, this is possible but requires a different kind of setting than intended here. However, here we are interested in the effect of the G-P map independently of the additional population-level aspects. To study this effect, we set **V** equal to the identity matrix **I**, thereby also assuming no linkage between the loci. Additive genetic variance in a population is maintained by a balance of mutations that generate variation and drift-related decrease, which itself is dependent on the current additive genetic variance and inversely scales with population size. If one reduces the additive genetic variance to zero by working on a completely inbred population while also enforcing a uniform environment, the mutational contribution to additive genetic variance can be measured as the total variance arising in this setting. Such multivariate covariance matrix is referred to as the **M** matrix.

The **M** matrix describes the mutational contribution to heritable trait variances and covariances before selection per one generation. It has the benefit that covariances of alleles can be ignored due to the random nature and low rate of mutations per generation. For the same reason also epistasis effects do not ensue and thus can be ignored. The more generations pass, the less justified these assumptions become. However, it might be necessary to construct **M** matrices after multiple generations and then to divide the entries by the number of generations in order to have enough signal within the random noise that comes with measurements (assuming **M** grows linearly with the number of generations). Constructing it requires precise phenotypic measurements on many individuals (= phenomics), ideally in the F1 generation of a completely inbred population. Such phenotypic assays are commonly generated in mutation accumulation (MA) experiments. Starting with one inbred population, mutations are free to accumulate in separate inbred lines in absence of selection, and different mutations might drift to fixation in these lines. The **M** matrix is then constructed across the individuals of different lines. Often this is done without any mutagen and such experiments have been a very fruitful source for learning about the influence of mutations on phenotypic variation and the nature of spontaneous mutations, such as their genomic rate and distribution of effect size. Combined with molecular data, such MA experiments allow for even more precise estimation of mutation rates on a nucleotide level (Halligan & Keightley 2009).

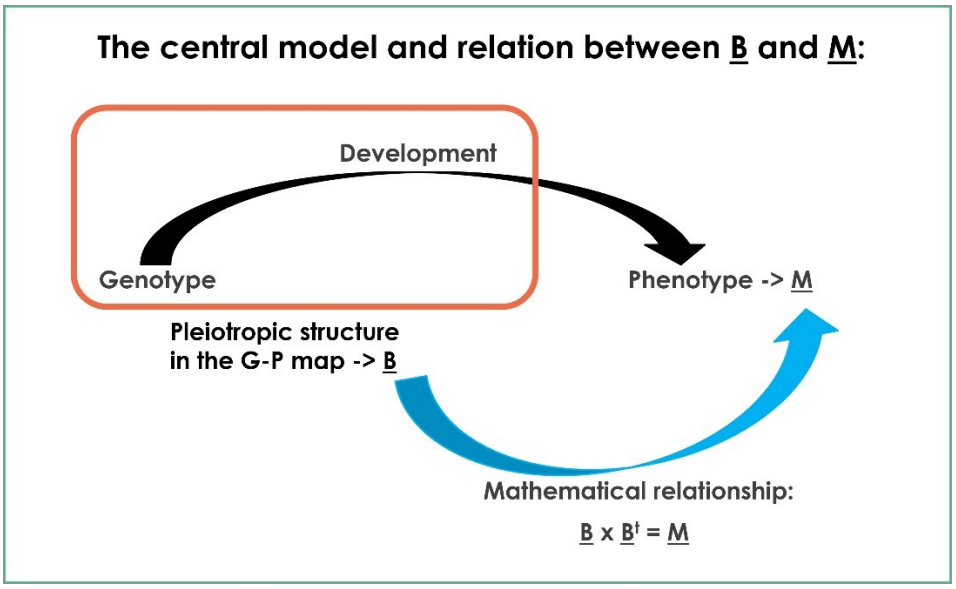

Figure 2: The central model of the genotype-phenotype mapping withing the context of variational mutational pleiotropy. The genotype is mapped onto the phenotype via development and the **B** matrix describes the pleiotropic structure of both the genotype and development. The mutational effect covariance matrix **M** is an aspect of the phenotype, and it mathematically relates to the **B** matrix via a matrix multiplication.

The model of the G-P map relates to the **M** matrix, as follows: **B** X $\mathbf{B}^t$ = **M**, and this model will be used for this thesis. In the limited context of mutational variation and its underlying (variational) pleiotropy, it represents an abstracted but direct mathematical connection between the genotype-phenotype mapping structure (**B**) and the ensuing distribution of variational effects (**M**) in empirically measured sets of biological traits. Further, both the map as well as the outcome of the mapping can be modular, in the sense that they can reflect strong association between some and a lack of association to other traits. Modularity can refer to the structure of the G-P map (a lack of pleiotropy) and the phenotype (a lack of covariance/correlation). The distinction matters, as the two aspects are not fully congruent. Specifically, this is relevant for the speculative case of hidden pleiotropy (See Fig. 3). A non-pleiotropic **B** will necessarily always result in a lack of correlations in **M** and correlations in **M** can only be derived from a pleiotropic **B**. However, it is in theory possible that the negative and positive contributions of single loci to the covariance between a pair of traits in a pleiotropic **B** cancel out and result in no (or very low) covariance. The pleiotropy of an allele would thus be invisible (hidden) to someone who only measures it through the variances and covariances of the traits in the phenotype because they do not contribute to genetic correlation (Turelli 1985; reviewed in Baatz & Wagner 1997). The case of hidden pleiotropy is particularly relevant to the aforementioned idea of “universal pleiotropy” (Fisher 1930; reviewed in Pavlicev & Wagner 2012) and the omnigenic model, which assumes that through “network pleiotropy” all genes have effects on all traits (Boyle et al. 2017). Such a model explicitly rejects the idea of modularity at a developmental level. Theoretically, hidden pleiotropy could explain empirical findings of variational modularity despite structural pleiotropy.

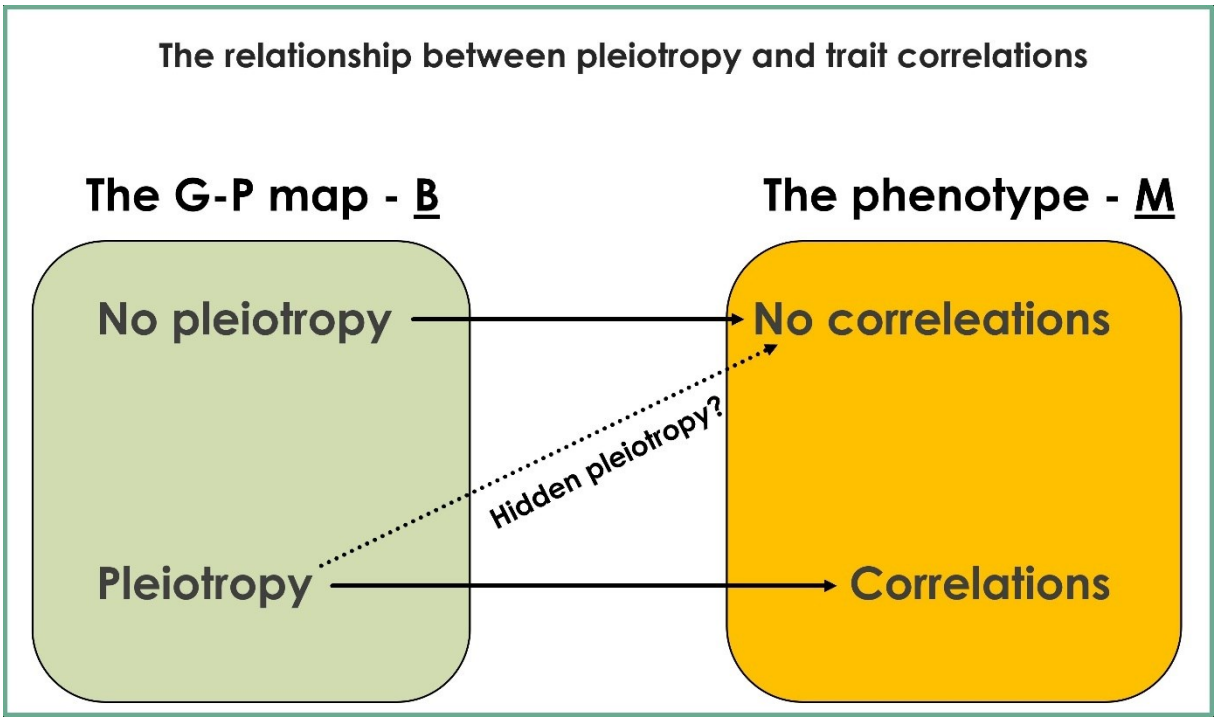

Figure 3: The relationship between pleiotropy at the level of the genotype-phenotype mapping (as covered by **B**) and trait correlations on the phenotypic one (as measured in **M**). A lack of pleiotropic trait interdependence in **B** can only result in having no correlations between these traits in **M**. Conversely, a lack of correlations in **M** can only result from no pleiotropy in **B**. However, it is theoretically possible that the pleiotropic effects within **B** cancel each other out, such that there is no correlation between the traits. This hypothetical case is referred to as hidden pleiotropy and it is sometimes proposed to claim that modularity does not exist in the G-P map and that pleiotropy is universal.

Goals

The first main task of this project is to find out what we can learn from mutational covariance (**M**) or correlation (**rM**) matrices about the underlying pleiotropic structure of the corresponding **B** matrix. Empirical **M** matrices will be taken from the published literature (See Table 1) from various studies on MA experiments and their levels of trait integration will be discussed in their biological context. Additionally, **B** matrices with different interesting biological implications will be generated. Through the equation $\mathbf{B} \times \mathbf{B}^t = \mathbf{M}$ (and by dividing by the variances to get from the covariance matrix **M** to the correlation matrix **rM**), the phenotypic consequences of hypothetical pleiotropic structures can be explored. This is also done with "stochastic **B** matrices", by which I mean **B** matrices where the entries are drawn independently from one another. They are viewed as a combination of a structural matrix $\mathbf{B}_s$ (representing the underlying structure of the G-P map) and a variational aspect $\boldsymbol{\mu}()$ (unrelated to the mutation rate which is also commonly denoted with μ), which represents the random effects of mutations. The realism – how close they can get to empirical patterns - of specific combinations as well as their usefulness as models will also be assessed. Further, for both the cases of constructed and

stochastic **B** matrices, an explicit effort will be made to include matrices with a random structure that may serve as null models against which any findings can be contrasted.

All **rM** will be eigen-deconstructed into values and vectors, which are analyzed with multiple methods. In the case of the stochastic **B** matrices, the generation of large numbers of replicates allow for the exploration of the possibility space of a certain combination $\boldsymbol{\mu(B_s)}$ through contrasting the different outcomes with one another. Such an approach could further allow for an estimation of posterior probabilities, which may serve to facilitate the use of maximum likelihood methods to attribute hypothetical $\mathbf{B_s}$ matrices to the patterns of real empirical **M**.

The second part of this thesis concerns the connection between pleiotropy (on the level of **B**) and trait correlations (in **rM**), which coincides with the question of how easy it is for a lack of correlation to evolve from randomness. As mentioned before, it is mathematically entirely possible that the effects of loci (**B**) cancel out in such a way that there is no correlation in the phenotype (**rM**) which is then referred to as hidden pleiotropy. Yet, it is a different question how likely such an outcome is. To investigate this question, at first, I use analytical considerations on simple **B** matrices with entries (+1, -1) to calculate the probabilities for achieving hidden pleiotropy for different numbers of loci and traits. Then I consider how the probability of a lack of trait correlations is affected by the normal distributions of the variational aspect $\boldsymbol{\mu()}$ for stochastic **B** matrices. This is combined with asking the question of if it is more probable to evolve a lack of correlations from a hidden pleiotropic or a random pleiotropic structure $\mathbf{B_s}$. Lastly, a large-scale generation of stochastic **B** matrices with these different $\mathbf{B_s}$ is conducted to determine posterior probabilities for achieving a lack of correlation for different numbers of loci and traits.

Methods

Eigen-decomposition (principal component analysis)

I use eigen-decomposition at the phenotypic level to explore the relation between the heritable phenotypic variation and the structure of the genotype-phenotype map. Rather than on genetic covariance matrix, it is performed on the mutational correlation matrix **rM**. These are in some cases the results of empirical measurements in mutation-accumulation experiments, and in other cases are derived from a theoretically constructed or stochastically generated G-P map in the form of a **B** matrix, given a relationship $\mathbf{B} \times \mathbf{B}^t = \mathbf{M}$ (Wagner 1984). The choice to focus on the correlation, rather than on the covariance matrix was made because it contains – regarding the goals of this thesis – the same

information and is generally easier to interpret.

In a first step, the $\mathbf{rM}_{n\text{x}n}$ matrices (n = the number of traits) are decomposed into their eigenvectors $\mathbf{E}_{n\text{x}n}$ and eigenvalues **L** (= diagonal vector; $\underline{\mathbf{L}} = (\lambda_1, \lambda_2, \ldots, \lambda_n)$). Eigenvalues are the scaling factors applied to the eigenvectors in order to arrive at the original matrix. Geometrically, an eigenvector is a direction in which a transformation causes no directional change whereas the corresponding eigenvalue is the amount of variation along that vector. Since covariance and correlation matrices are symmetric, these eigenvectors must also be orthonormal (Strang 2003). However, in practice, numerical imprecisions in their calculation will lead to slight deviations in the orthogonality of the eigenvectors.

Eigenvectors of covariance matrices (including variance normalized correlation matrices) are principal components and thus, the eigenvalues represent the variance magnitudes in the directions of the largest spread of the data. In the case of a correlation matrix, these variances are rescaled such that the sum of all eigenvalues **L** is equal to $n \cdot 1$ (with 1 being the uniform diagonal values and "variances" of a correlation matrix). The proportion of an eigenvalue to this total variance = n can thus be viewed as the contribution of a corresponding eigenvector to the total variance of the correlation matrix. As principal components, eigenvectors contain a lot of information because their loadings show in which direction a specific relative part of the variance (the eigenvalue) lies. Eigenvalues in turn must be ordered from highest to lowest, which makes the order of eigenvectors meaningful too. Thus, the first eigenvector will always explain the largest proportion of the data.

The distribution of eigenvalues has previously been used to describe the level of integration between traits, which can be related to their modularity (Wagner 1984; Pavlicev et al. 2009). Wagner (1984) also shows that the eigenvalue distribution of empirical correlation matrices deviates from that of a random pleiotropy model by a larger leading eigenvalue that is associated with a "size factor", a factor shared across all morphological traits. The distribution of eigenvalues serves as one of the measures of correlatedness of traits of this thesis. To capture the correlatednes between traits in a single value, I used in this thesis the relative standard deviation of the eigenvalues $SD_{rel}(\lambda)$ (Pavlicev et al. 2009). This measure is calculated here from the eigenvalues of **rM** matrix using the following expression:

$$SD_{rel}(\lambda) = \frac{\sqrt{Var(\lambda)}}{\sqrt{n-1}}$$

Where λ is the set of eigenvalues $\underline{\mathbf{L}} = (\lambda_1, \lambda_2, \ldots, \lambda_n)$ and n is the number of traits in a correlation matrix **rM**. Since the eigenvalues are ranked in decreasing order by convention, this measure also captures the rate of decrease independently of the number of traits. The smallest value that it could take is 0, corresponding to equal eigenvalues, such as when there are no correlations at all. The largest value is realized if all correlations are 1 and then it would be a little over 1. The exact maximum value depends

on the number of traits, and it is higher the fewer traits there are (n = 2 -> 1.41; n = 3 -> 1.22; n = 10 -> 1.05, n = 100 -> 1.005).

Constructed **B** matrices

I constructed a set of seven **B** matrices with dimensionality 4X16. All entries are integers, chosen such that the sum of all values in a row is exactly 16. This was done for better comparability and easier interpretability, as this will cause the trait variances to be approximately equal across traits. This also means that there is rather little difference between the results obtained from the **M** and **rM** matrices. The **B** matrices have the following structures:

1. A fully pleiotropic matrix with each gene affecting each trait equally, and thus each entry is the same (**pleiotropic**).

2. A strictly modular matrix in which each locus only affects one of the traits, with no loci shared between pairs of traits (**modular**).

3. A "**two big modules**"-matrix with two pairs of traits building one big module each. In such a matrix each module is affected by one half of the loci, the traits within a module sharing all loci, and the modules sharing none.

4. A matrix with few fully pleiotropic loci while the others affect either of two distinct modules of the same sizes (two traits), without sharing loci, and without trait-specific loci. The effects are thus **2-level hierarchical**.

5. A matrix with some loci affecting all traits, some loci affecting half of the traits and some affecting only a single trait. The effects of loci with different numbers of affected genes have a nested structure and are **3-level hierarchical.**

In addition to the above constructed matrices with exclusively positive effects, I built two that also incorporate negative effects. Negative effects were generally avoided for the other matrices as the interpretation of results becomes less intuitive because the combination of different positive and negative contributions of single loci to trait correlations can cancel out. Instead of modular structures in which (sets of) traits lack correlation due to a lack of shared loci, the modular correlational structure can result from the effects of shared single loci which cancel out, the phenomenon of so-called "hidden pleiotropy". To include such cases, I constructed two matrices:

6. In the first, each locus has one strong positive effect on a single trait and smaller negative ones on the others. Thus, these positive effects are distributed over all four traits, and we arrive at a matrix that is similar to the completely modular matrix mentioned above (**antagonistically modular**).

7. In the other matrix involving negative effects, the positive effects are distributed into two distinct modules, similar to the "two big modules"-matrix. The matrix differs in that the loci do not have no effect on the traits outside its module, but negative ones. Thus, we can consider modularity to be even further enhanced by these negative effects of "**two antagonistic modules**".

In addition to these constructed matrices, I also generated a set of three "**random**" matrices to include structures that are not clearly interpretable. To achieve this, I randomly drew vectors of 16 numbers. These numbers could be any integer from -5 to 5 with equal chances. I then checked their sum and if it was 16, I kept it. Then I stuck together four of these vectors as rows of a 4X16 matrix. It is important to note that these matrices are only random in so far that they do not contain any intentional information. Yet, they cannot be considered properly random in the sense of the field of random matrix theory.

Stochastic **B** matrices

Generating matrices

By "stochastic **B** matrices", I mean that all elements of a **B** matrix are drawn independently from each other from normal distributions with specific means and variances. Depending on the case, all elements may share either their mean or variance, whereas the other differs according to the predetermined structure of a matrix. For this approach, I consider the final **B** matrix as a combination of two aspects, a structure $\mathbf{B_s}$ (representing only the distribution of effects in terms of presence/absence) and variational aspect $\boldsymbol{\mu()}$ that is imposed on it and which represents the types of distribution from which the single entries are drawn.

$$\mathbf{B} = \boldsymbol{\mu}(\mathbf{B_s})$$

It allows me to generate large numbers of stochastic matrices for the combination of a certain structure $\mathbf{B_s}$ and variational aspect $\boldsymbol{\mu()}$. Using the relation between **B** and the phenotypic covariance distribution

(**B** X $\mathbf{B}^t$ = **M**), the phenotypic consequences of different combinations of structure and stochastic regimes $\boldsymbol{\mu(B_s)}$ can be explored.

The structural matrices $\mathbf{B_s}$ are *n*X*k* sized matrices with only a few possible entries (e.g., +1, 0 and -1), which describe the presence and phenotypic direction of effects, but not their size. Choosing specific theoretically compelling and biologically meaningful structures allows us to demonstrate the effect of a G-P map structure on the variational distributions, under different variational regimes. For this chapter, the primary focus is on six 4X16 matrix structures with entries that are either +1 or 0, because they are easily interpretable and compatible with the three types of stochasticity in use. Five of them are based on the first five constructed **B** matrices. All entries that are equal to 0 were kept and all other entries were replaced with 1. This leads to $\mathbf{B_s}$ for the following constructed **B**: pleiotropic, modular, two modules, two-level hierarchical and three-level hierarchical. Additionally, random matrices are used in which each entry is in each iteration drawn to be either +1 or 0.

I define three types of stochastically generated **B** matrix elements, which are all drawn from normal distributions:

- Type-1a stochasticity: the entries of **B** represent standard deviations of a normal distribution with mean = 0. For entries $Bs_{i,j} = 0$, this means that there is no effect, while the other entries in the structure $\mathbf{B_s}$ must be positive ($N(0, Bs_{i,j})$).
- Type-1b is once again focused on different SDs while the mean is constant and 0. It is a variant on type-1 in which the entries $Bs_{i,j}$ that are 0 are drawn with an SD of 1 ($Bs_{i,j} = 0 \rightarrow N(0,1)$) and those that are 1 are drawn with an SD of 2 ($Bs_{i,j} = 1 \rightarrow N(0,2)$).
- Type-2 stochasticity is focused on **B** matrix elements drawn from normal distributions with different means and a fixed SD of 1 ($N(Bs_{i,j}, 1)$). $\mathbf{B_s}$ thus represents the expected values of all entries for $\boldsymbol{\mu(B_s)}$ of that $\mathbf{B_s}$.

In addition to these $6 \cdot 3 = 18$ combinations of variational regime and the structure of **B**, I also define two additional combinations $\boldsymbol{\mu(B_s)}$, which serve to test specific hypotheses. First, I use a variant of the type-2-pleiotropic $\mathbf{B_s}$ in which all entries are randomly drawn (in each iteration) to be either +1 or -1 ("both signs pleiotropic"). This aims at checking whether the trait integration resulting from the type-2-pleiotropic **B** is caused by the full pleiotropy that stems from the absence of entries that are 0 in the $\mathbf{B_s}$, or if the large first eigenvalue can be attributed to the strong tendency of all entries in **B** to be positive (all entries that are larger than the mean minus 1 SD are positive). Further, this structure is also directly tied to models that will come up in the chapter on hidden pleiotropy.

Second, I attempt to construct a combination $\boldsymbol{\mu(B_s)}$ that is loosely based on Wang et al. (2010) who find that: (1) The mean of mutation effects is in most cases close to 0 and that there are a few genes with large SDs and many with very small ones. And that (2) the genes that have the highest SDs also tend to affect the most traits. Incorporating (2) is not possible with the used format of 4X16 $\mathbf{B_s}$ matrices, as I want to keep *k* constant to ensure comparability. To incorporate (1), I use a structural $\mathbf{B_s}$ consisting of integer entries from 1 to 5 in combination with a type-1a stochasticity. The idea of this "type-1a-different-SDs" combination is that there is a larger range for the SDs, which is achieved by defining rows that contain entries of one 5, two 4s, three 3s, four 2s and six 1s (once again summing up to 16 loci). The order of these entries is then randomly drawn for each of the four rows in each iteration. By not having constant loci, I aim to purely focus on having differently large SDs.

For all 20 combinations $\boldsymbol{\mu(B_s)}$, at first 200 stochastic matrices were generated. This number is limited due to the exponentially growing computational time in some of the downstream analyses. The stochastic matrices were then used to generate the correlation matrices **rM.** To study the variational structure associated with various G-P map structures, I used eigen-decomposition to generate eigenvalues and -vectors. The eigenvalue distributions are plotted as above, with the exception that not all can be visualized in one graph. Instead, a subset of 20 are randomly selected to be plotted together. A separate plot displays the means and variances of the four eigenvalues to obtain a generalized picture of the eigenvalue distributions. The correlatedness measure $SD_{rel}(\lambda)$ was also obtained for each of the 200 iterations and its mean and variance were computed. Additionally, a boxplot that displays the quantiles was also generated.

Analysis of eigenvectors

While the constructed **B** and empirical **M** correspond to a single correlation matrix **rM**, which results through decomposition (for four traits) in one set of eigenvectors $\mathbf{E}_{nXn}$, stochastic matrices will result in many unique **rM** and $\mathbf{E}_{nXn}$ for a particular structural matrix. For the exploration of the space of results of each combination $\boldsymbol{\mu(B_s)}$, I will describe the patterns in the diversity of eigenvectors $\mathbf{E}_{nXn}$ by developing methods for summing up their information. Geometrically, eigenvectors represent directions in phenotypic space that describe a certain proportion of the variance, which is given by the corresponding eigenvalue that has the same position in the order. The main question is then how two sets of eigenvectors that are derived from two distinct stochastic **B** matrices with the same $\boldsymbol{\mu(B_s)}$ can be compared. To answer this question, I decided to attempt to determine the corresponding vectors across matrices and compute the angles between each pair of corresponding vectors. I presumed that correspondence occurs in pairs between exactly two vectors from different sets and that each vector

has exactly one correspondent in each other set. Yet, how well two vectors may correspond and how meaningful this is, strongly depends on the structure $\mathbf{B}_s$.

For determining which two vectors correspond, one might consider correspondence according to the order (e.g., the first eigenvectors of two sets correspond to one another because they both describe the most variance). However, if two **rM** are very different, the eigenvectors with most variance might point in very different directions. Instead, we will use minimal angle between the eigenvectors to find the correspondence between them. The angle between two vectors is calculated as their dot product, resulting in the smaller angle (< 180°). For angles between eigenvectors, it must be considered that the inverses of eigenvectors also point in the same direction, just with reversed signs. An angle of 180° degrees between two vectors should thus rather be counted as 0° and one of 150° as 30°. To account for this, any angle > 90° (in the second circle quadrant) was replaced with 180° minus that angle.

However, establishing correspondences between vectors from different sets, using "pairwise minimum angles" alone can also be misleading. Calculating the angles between each vector from the first matrix and each vector from the second matrix ($4 \cdot 4$ pairs) and then choosing the four pairings that have the smallest summed up angles, neglects the relative variance contribution of the respective eigenvalues. For example, in an extreme case, a first eigenvector that explains a large proportion of the variance may correspond to a fourth vector which explains next to nothing. In such a case it would not make much sense to refer to those vectors as being caused by the same underlying pleiotropic trait association that is determined by $\mathbf{B}_s$. Only relying on minimizing the summed up angles would imply that even substantial differences between the eigenvalues are caused purely by the stochastic effects $\boldsymbol{\mu}()$ of mutational effect sizes.

To find a middle ground to this dilemma we attempted to incorporate the order of the eigenvectors while still relying on pairwise minimal angles:

1. When two sets of eigenvectors are compared, one set is randomly determined to be the "choosing" set (while the vectors of correspondent set are "chosen" by it).
2. The angles between the first eigenvector of the "choosing" set and all vectors of the other are calculated.
3. The first eigenvector of the "choosing" set is assigned its correspondent vector in "chosen" set by the criterium of it having the minimal angle with the "chooser".
4. The procedure is repeated with subsequent vectors of the "choosing" set in the order of their rank and they get to "choose" their correspondent vector from the remaining ones of the "chosen" set.

For this thesis I call this algorithm for determining the minimum pairwise angles between eigenvectors the “Damenwahl” algorithm as an homage to the ball house culture of Vienna and because it reminds me of how partners were usually found in my dancing school. It is important to note that because the vectors within each set are orthogonal to one another, that the minimum angles between the four pairings should be identical, apart from the numerical imprecision in R.

All four minimum pairwise angles between all pairs of 200 sets were computed and then plotted together in four boxplots. This was done for all 20 combinations in order to compare them. To validate the algorithm and contrast the results, random sets of four 4D eigenvectors E_{4X4} that were all orthogonal to one another were generated.

Next, I attempt to use the idea of corresponding vectors to find something resembling mean vectors, along with measures of variance around them. This could be particularly helpful for investigating the possibility spaces of $\boldsymbol{\mu(B_s)}$ because it might reveal their general tendency which may then be contrasted with those of other combinations. Hypothetically, this may then be used to compute posterior probabilities for a combination $\boldsymbol{\mu(B_s)}$ realizing a particular pattern in its eigenvalues and -vectors. In turn, these posterior probabilities may then be used with a maximum likelihood approach to associate the eigen-patterns of empirical **rM** matrices with combinations of $\boldsymbol{\mu(B_s)}$. This principle could then be fleshed-out to constitute a powerful mapping of a measurable **M** matrix to a most likely **B** matrix.

Finding mean eigenvectors is once again not as straightforward as just computing the mean vectors of all vectors that share a position. As mentioned before, we cannot expect that the vectors that correspond to one another also share the same position in the order. It is therefore necessary, to reorder the assigned positions of the vectors. This can be done by slightly modifying the “Damenwahl” algorithm to find pairs of vectors that have a minimal angle. However, this algorithm requires the distinction into a “choosing” and a “chosen” set of vectors. Thus, we must first determine the order of which set is to be used for the reordering process. Of course, one could decide to just pick any random set E_{4X4} and use it to reorder all others. However, if the sets differ greatly, it is possible that this randomly chosen set is not very representative of the whole which could then drastically impair the accuracy of the mean vector

Instead, I decided to at first find a set of eigenvectors that is most representative of the whole (“the president”) and then use it to reorder all others. This is done by running the “Damenwahl” algorithm on each pair of sets twice, with both sets serving once as the chooser and once as the chosen. Then, all the angles between combinations that involve a certain set are summed up for all sets. The set that has the lowest sum is then chosen to be the “president”. It is used in a slightly modified “Damenwahl” algorithm as the “choosing” set and the order that the vectors from the chosen set are paired up

becomes their new order. Once all sets are reordered, the mean vectors and variances around them for the four positions can easily be computed. Lastly, the mean angles – along with the variance around it – between the mean vectors and all other reordered ones are calculated as additional measures of spread.

Tree topology analysis

Another way to explore the possibility space of a combination $\boldsymbol{\mu(B_s)}$ is by performing hierarchical clustering on the correlation matrices **rM**. Like other methods of cluster analysis, it is a method for data exploration which aims to visualize the similarity and distance of elements. Specifically, it works by subsequently merging two-by-two the most similar sub-sets until an all-containing root is found. For this, one needs to define a distance metric and a linkage distance. The result is a binary merge tree which can be graphically represented as a dendrogram. Such a dendrogram is an intuitive way of visually displaying the similarity of elements (Nielsen 2016). A tree itself can on the other hand also be translated into Newick format, which is a conventional standard for encoding the topology of relatedness in phylogenetic trees (Felsenstein 2004; Olsen 1990). For the purposes of this thesis, we only focus on the topology of trait correlatedness and dismiss the branch lengths.

Performing hierarchical clustering on correlation matrix will lead to a single topology. This has also been done for the correlation matrices from the constructed **B** matrices as well as on the empirical **rM** matrices. In the case of combinations $\boldsymbol{\mu(B_s)}$, this approach is even more informative, because the same combination may achieve different topologies due to introduced variation. The relative frequencies of the different topologies and how they differ between different combinations $\boldsymbol{\mu(B_s)}$ informs about the degree to which $\mathbf{B_s}$ constrains the possibility space, given the specific level of stochastic variation.

The “ward.D2” algorithm within the hclust()-function in R – which implements Ward´s clustering criterion (hclust() documentation in R (R Core Team 2022); Murtagh et al. 2014; Ward 1963) – was used for the hierarchical clustering. For four traits there are a total of 15 possible rooted topologies, and these are always plotted in the same order (regardless of if the frequency of some may be 0) which is loosely based on the position of the first trait relative to the others.

Compiling empirical **M** matrices

I have compiled a set of published mutation accumulation (MA) studies that have each provided us with one or more workable **M** matrices (Table 1). These studies were conducted on a variety of model organisms and traits. Further **M** matrices from other publications have initially been considered but were excluded because their correlation matrices included correlations > 1, suggesting that their data had not been normalized. The included **M** matrices are the following:

Table 1: List of publications with included **M** matrices, ordered by year of publication (see references). The population size N refers to the numbers of individuals from which separate MA lines were derived.

Authors:	**Year:**	**Species:**	**Category of traits:**	**Number of traits:**	**Population size (N):**
Fernández, J. & López-Fanjul, C.	1996	*D. melanogaster*	life history	3	200
Shaw, R.G. et al.	2000	*A. thaliana*	reproductive traits	3	120
Bégin, M. & Schoen, D.J.	2006	*C. elegans*	life history/ body size	4	75 each
Dugand, R.J. et al.	2021	*D. serrata*	wing shape and size	5	102
Mallard, F. et al.	2023	*C. elegans*	locomotion behaviour	6	54 (N2) & 62 (PB)

And in the following is some biological background on these studies:

- Fernández & López-Fanjul (1996) have conducted their experiment on *Drosophila melanogaster* and the measured traits are intended to capture the females´ fitness. They recorded the number of eggs per female (**fecundity**), as well as how many pupae and then adults were still alive after certain numbers of days (**egg-pupa viability**, **pupa-adult viability**). They report that the egg-pupa viability may be underestimated because some eggs could fail to hatch due to low paternal fertility.
- Shaw et al (2000) have worked on *Arabidopsis thaliana* and they measured number of seeds per fruit (**seeds/fruit**), the total number of **fruits** produced and the dry mass of the infructescence (**reproductive mass**). It is not clear to me whether the authors have normalized the raw measurements, because the correlation matrix that I computed from their original covariance matrix includes a correlation of ≈ 1.03. I have chosen to treat this as a measurement error and still include it.
- Bégin & Schoen (2006) conducted their work on three separate strains of *Caenorhabditis*

elegans. They were interested in the effect of germline transposition of transposable elements on the **M** matrix and thus conducted their experiment on two transpositionally active mutator strains (**mut-7** and **mut-14**) as well as on a control strain (**N2**) which lacks germline transposition. The mutations in the first two strains can therefore not be considered spontaneous. They recorded how many eggs were produced over a lifetime (**fecundity**), how many days the worms lived (**longevity**), as well as the **length** and **width** of fixated individuals from the same lines. They report that their measure of fecundity may be inflated due to non-viable eggs.

- Dugand et al. (2021) have described traits on the wing morphology of *D. serrata*. They have set nine landmarks and computed the integrated landmark distance (**ILD**) between all of them. Of these, they only focused on a biologically meaningful subset. Additionally, they also computed the wing **size** as the square root of the sum of the squared distances between each landmark and their centroid. It is noteworthy that this experiment was conducted on outbred lines with a middle-class neighbourhood breeding design.
- Mallard et al. (2023) worked on the locomotion behaviour of two different strains (**N2** and **PB**) of *C. elegans*. They tracked the movements of adults and defined three states, **1** = forward, **2** = still, **3** = backwards. From the observed transition rates between these states, they computed posterior probabilities, which they defined as their traits.

Probability of hidden pleiotropy in simple **B** matrices

For answering the question of how probable it is for hidden pleiotropy to occur, we will at first consider simple **B** matrices in which entries can, with equal chance, only be either +1 or -1. This thus assumes full pleiotropy, and only considers the gene to have a unitary effect. It is straightforward to predict which subset of possible **B** matrices of this type leads to a hidden pleiotropic **M** matrix, considering the algebraic relationship between **B** and **M**, i.e., $\mathbf{B \times B^t = M}$. A matrix **B** is hidden-pleiotropic (hp) when the resulting trait covariances equal zero, i.e., all off-diagonal values in **M** are 0. The probability of a hidden-pleiotropic genotype phenotype map is then simply equivalent to their proportion out of all possible fully pleiotropic maps.

The smallest meaningful **B** matrix for our question is a 2X2 matrix, i.e., a matrix with 4 elements, each of them with a possible value +1 or -1. This constitutes a total $2^4 = 16$ possible matrices. Of these, there is a subset of eight **B** matrices that result in no covariance/correlation. The fraction of hidden-pleiotropic matrices out of all possible ones is therefore $8/16 = 0.5$, equal to the probability to randomly generate a 2X2 hp **B** matrix ($P_{hp}(\mathbf{B}_{2X2}) = 0.5$).

To generalize this principle to larger matrices, we use the relationship **B X B^t = M**, to express each entry of the upper or lower off-diagonal matrix $m_{x,y}$ as the dot product of the rows $b_{x,*}$ and $b_{y,*}$. The problem to solve is to find a set of **B** matrices for which all the off-diagonal entries are 0.

⟨1⟩ $$\sum_{i=1}^{k} b_{x,i} \cdot b_{y,i} = b_{x,y} = b_{y,x}$$

By successively increasing the dimensionality, this principle allows us to explore how the probability of hidden pleiotropy changes with the number of traits and loci. A larger hp matrix can be built from a smaller one by adding a row which has dot products of zero with each existing row because only the entries of this row contribute to the new covariances. For large **B** matrices, we will eventually resort to an algorithm which uses the expand.grid() function in R to generate all 2^k possible rows. By applying the principles that are found, this allows for a direct way to estimate the probabilities of hidden pleiotropy for larger matrices (which is done here for dimensionality of up to 20X20). All probabilities $P_{hp}(\mathbf{B}_{nXk})$ are also confirmed using numerical simulations.

Probability of hidden pleiotropy in stochastic **B** matrices

In the second approach, I ask, how the insights we gained from the chosen subset of **B** matrices above, generalize to stochastic matrices, which better represent real mutational effects. I start out with a $\mathbf{B}_s$ matrix which consists of +1s and -1s and then use these original entries as means to draw numbers from a normal distribution with SD = 1 around them. This is equivalent to the generation of stochastic matrices with type-2 stochasticity from the earlier chapter. In an additional question, I ask whether it is more probable to evolve uncorrelated traits from a hidden-pleiotropic or a random structure when a variational aspect is imposed on it. All correlations in **rM** that are nearly equal to 0 can be generated by hidden pleiotropy, or by a lack of pleiotropy in the **B** matrix.

We must then consider – for this case of a stochastic matrices with a type-2 stochasticity – what it means for a structure $\mathbf{B}_s$ to be hidden-pleiotropic or random. Just like in the first approach, a hp $\mathbf{B}_s$ is one, which leads to zero correlation in **M** (**Bs X B$_s^t$ = M**). For the random $\mathbf{B}_s$ matrix, I have utilized matrices in which each entry is randomly drawn to be +1 or -1. This is not only equivalent to the matrices discussed before, but also the “both signs pleiotropic” $\mathbf{B}_s$ from the chapter on stochastic matrices.

At first, I will consider what the variational aspect means for the lack of trait correlation and how the drawn values affect it. Similarly to the previous chapter, this is done by thinking about how one entry being drawn necessitates the value of another, to generate zero correlation. This requires an examination of the probability using normal distributions (with a mean of +1 or -1 and an SD of 1) instead of the binomial distribution used for the **B** matrices with entries +1 and -1. Last, a large number of stochastic matrices is generated in numerical simulation.

For the simulations, it is necessary to generate hidden-pleiotropic $\mathbf{B_s}$ matrices. This was done by at first generating 4X*k* matrices with random entries (limited to (+1, -1) with equal probability) until one was found where all off-diagonal values of **M** are = 0. Next, rows were added one by one in the same fashion. If the row addition did not introduce covariance in **M**, it was kept. This was done until a 10X*k* (*k* = 12, 16, 24, 40) hp $\mathbf{B_s}$ was found and how long it took for matrices to be expanded was recorded. The resulting 10X*k* matrices serve as templates for all *n*X*k* matrices with the same *k* and an *n* between 2 and 10 because any submatrix of rows is by necessity also hp. Omitting any trait (row) just leads to **M** lacking that trait (the corresponding row and column). All smaller $\mathbf{B_s}$ matrices were then constructed by leaving out the lower rows and their order was not reshuffled. For the case of *k* = 20, the 6X20 hp $\mathbf{B_s}$ that was constructed by hand by Hansen et al. (2019) and used in Hummer (unpubl.) was used. It was then extended to a 10X20 matrix in the same fashion as the others.

For each *n*X*k* category I generated a total of 1e8 stochastic matrices. Simulations with random structures in which each entry is redrawn in each simulation (+1 or -1 with equal chances) were also conducted for the exact same combinations of numbers of traits (*n* = 2 to 10 and *k* = 12, 16, 20, 24, 40). For each stochastic matrix, the matrix multiplication $\mathbf{B\ X\ B^t = M}$ is performed, and the covariance matrix **M** is converted to a correlation matrix **rM**. It is then checked if all off-diagonal values are somewhere in a range (0.1 and 0.01) around 0. If they are, the run is counted as a hit and finally the fraction of hits to total runs is calculated.

In total, I conducted 90 simulations, by exploring effects on trait correlations of 3 variables: 5 (different numbers of loci) · 9 (different numbers of traits) · 2 (hidden-pleiotropic and randomly drawn $\mathbf{B_s}$). The simulations were performed on the Life Science Compute Cluster of the University of Vienna.

Results

Eigen-patterns of the constructed **B** matrices

The following is a summary of the noteworthy patterns that have been revealed by the eigen-decomposition of the eleven constructed **B** matrices. The directions of the eigenvectors that correspond to eigenvalues that are close to being 0 are not reported on.

1. The completely universally **pleiotropic** B matrix generates a correlation matrix with all traits highly correlated, resulting in decomposition in one eigenvalue λ_1 which represents all the variance and the corresponding eigenvector $\mathbf{E_1}$ loads +/- 0.5 uniformly on all traits. The correlatedness is the highest amongst all constructed matrices ($SD_{rel}(\lambda) = 1.15$).

2. The completely **modular B** matrix where each trait is affected by a separate set of loci generates a matrix of uncorrelated traits. In this case, all four eigenvalues (λ_1, λ_2, λ_3 and λ_4) are equally big and each eigenvalue loads on exactly one trait. The correlatedness is the lowest of all ($SD_{rel}(\lambda)) = 0$). Both extreme matrices are completely unable to form clusters, as there just is no differential structure between sets of traits.

3. The **B** matrix where the effects of loci on four traits are arranged into **two big modules** results in a correlation matrix, which, upon decomposition, manifests two equal eigenvalues ($\lambda_1 = \lambda_2$) one representing each module of fully correlated pairs of traits, and on which therefore all the variance falls on, as the two modules share no variance. The corresponding vectors ($\mathbf{E_1}$, $\mathbf{E_2}$) each load ≈ +/- 0.7 onto the two traits that compose one of the two modules. The correlatedness is fairly high ($SD_{rel}(\lambda) = 0.67$) and we see a clear clustering into the two modules.

4. The correlation matrix arising from the "**2-level hierarchical**" **B** matrix decomposes into only two non-zero eigenvalues (λ_1 and λ_2) and almost 80% of the variance falls into the first, as the two modules are correlated. Both their vectors ($\mathbf{E_1}$, $\mathbf{E_2}$) load ≈ +/- 0.5 uniformly on all traits. There is clear clustering into the two modules and the correlatedness is very high ($SD_{rel}(\lambda) \approx 0.86$).

5. The similar "**3-level hierarchical**" **B** matrix generates a correlation matrix that decomposes into eigenvalues of which the distribution mirrors the previous one. The difference is that the eigenvalues are a little bit more similar to one another: More than half of the variance falls into λ_1, while almost a quarter falls into λ_2. The final quarter is shared exactly equally by λ_3 and

λ_4. All vectors $\mathbf{E_{1\text{-}4}}$ load ≈ +/- 0.5 uniformly on all traits and there is clustering in two separate modules. There correlatednesses is intermediate ($SD_{rel}(\lambda)$ = 0.46).

6. The "**antagonistically modular**" **B** matrix leads to a correlation matrix that decomposes into three eigenvalues $\lambda_{1\text{-}3}$ that are all equally high, while the last one is much smaller. The loadings are the most complicated out of all constructed **B** matrices: The first vector $\mathbf{E_1}$ loads on all traits, of which z_1 gets a high loading of ≈ +/- 0.87 and the others get ≈ +/- 0.29. Next, this trait is not loaded on in $\mathbf{E_2}$ while the second gets ≈ +/- 0.82 and $z_{3\text{-}4}$ get ≈ +/- 0.41. Then, the third eigendirection $\mathbf{E_3}$ loads nothing on the first two traits and ≈ +/- 0.71 on the other two. Finally, the loadings of the last eigendirection $\mathbf{E_4}$ are uniformly ≈ +/- 0.5. As with the modular matrix, there is no clustering. The correlatedness is rather low ($SD_{rel}(\lambda)$ ≈ 0.26).

7. For the correlation matrix that is generated from the "**two antagonistic modules**" **B** matrix, 90% of the variance falls into λ_1. The rest goes into λ_2 and both their vectors ($\mathbf{E_1}$, $\mathbf{E_2}$) load ≈ +/- 0.5 uniformly on all traits. There is clear clustering into the two modules and the correlatedness is a very high ($SD_{rel}(\lambda)$ ≈ 1). Overall, the patterns that are generated from this matrix are very similar to those of the "**2-level hierarchical**" one.

The three randomly generated matrices all show very similar patterns: The eigenvalue distribution curves are approximating a diagonal line, and the loadings of the eigenvectors are not interpretable. Their measures of correlatedness are fairly low (0.22 < $SD_{rel}(\lambda)$ < -0.34). and they all have different clustering.

Patterns of stochastic **B** matrices

To summarize the observed pattern across all (18) combinations of $\mathbf{B_s}$ (6) and **μ()** (3) as well as the two additional simulations, I categorized the resulting patterns of eigenvalue distributions (both of the mean eigenvalues and the subsamples of distributions) and $SD_{rel}(\lambda)$, and associated them with a corresponding constructed **B**. These categories also correspond with the distinct patterns created with the methods that are based on eigenvectors. In total, I grouped patterns into four qualitatively different categories, which is in contrast to the diversity of patterns observable for the constructed **B** matrices. Further, of all four pattern categories, only one can be found in more than one specific combination of $\mathbf{B_s}$ and **μ()**, which consequently means that it is observable for the vast majority (17/20) of all combinations (Fig. 4):

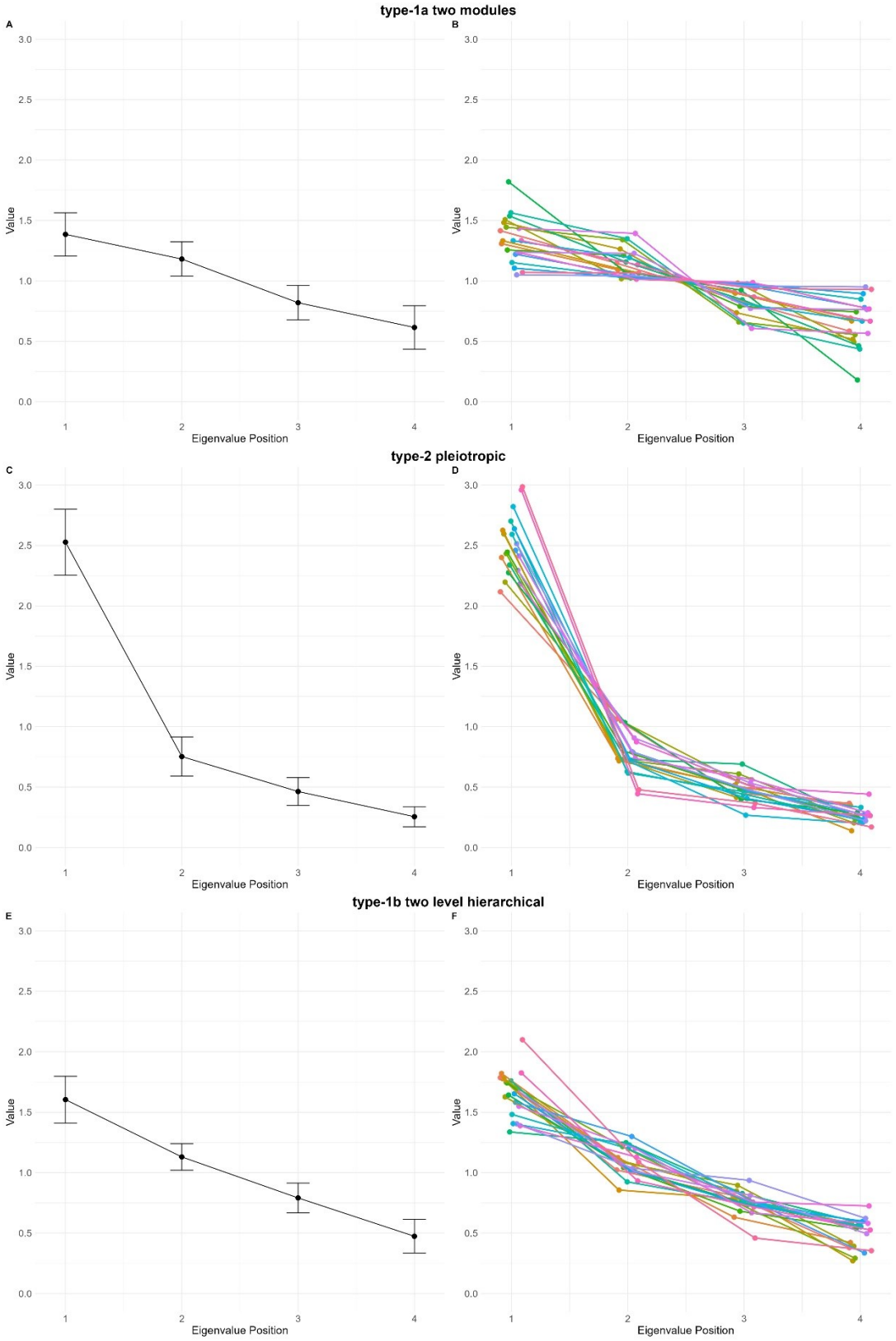

Figure 4: Three cases of eigenvalue distributions for the categories of **A&B** "two distinct modules, **C&D** "overall trait integration" and **E&F** "random trait integration". The plots on the left display the mean (N = 200) eigenvalues for each position with the variances around them. The plots on the right display a randomly chosen subset of 20 actually realized eigenvalue distributions. The eigenvalues of the fourth "no correlations" category are all exactly one. **A** shows that on average the two first and two last eigenvalues are more similar to one another, and **B** shows that there is a large diversity between the actual distributions. **C&D** show a large first eigenvalue which is a sign of trait integration and **E&F** show practically diagonal lines. Both **C&E** show that the variance around the first eigenvalue is the highest in their respective categories which cover 18 out of the 20 combinations that were tested.

1. **“No correlations”**:

   This pattern is only found for the combination of a fully modular $\mathbf{B_s}$ and the hard structure (type-1a). By the nature of correlation matrices lacking the stochastically determined different trait variances, the **rM** of all stochastic matrices must be the same which causes all eigenvalues to also be exactly 1, $SD_{rel}(\lambda)$ to be 0 and all other measures that indicate the diversity between the stochastic matrices to be 0 too. Also, each vector loads on exactly one trait. This category directly corresponds to the completely modular constructed **B** matrix.

2. **“Two distinct modules”:**

   This pattern occurs with the combination of type-1a and the two modules $\mathbf{B_s}$ and expectedly corresponds to the two modules constructed **B** matrix, even though the pattern is less pronounced. The distribution of mean eigenvalues shows that those of the first two eigenvalues are similarly big and that those of the second two are similarly small (Fig. 4A). The biggest drop therefore occurs between λ_2 and λ_3. However, the actual eigenvalue distributions that are observable in a subset of just 20 randomly chosen ones are very diverse and fall on a gradient between a horizontal and a diagonal line. Further, having them plotted together resembles the shape of a tilted hourglass because the connecting lines between λ_2 and λ_3 meet very narrowly (Fig. 4B). The mean $SD_{rel}(\lambda)$ is small with 0.21 and has a small variance of 0.009.

3. **“Overall trait integration”:**

   This category occurs only with the combination of type-2 and the fully pleiotropic $\mathbf{B_s}$. It corresponds to the pleiotropic constructed **B** and has a very large λ_1 in both eigenvalue distributions (Fig. 4C&D), which causes the mean $SD_{rel}(\lambda)$ to be the highest with 0.60 and it has a small variance of 0.010. However, given that this pattern is not observed for the special case of the “type-2 both signs pleiotropic” $\mathbf{B_s}$, it seems that this trait integration is the result of a large, shared factor associated with a strong tendency towards having a positive sign in all entries of the stochastic **B**.

4. **“Random trait integration”**:

   This pattern is found for all 17 other combinations and is thus by far the most frequent one. All eigenvalue distributions are approximating diagonal line (Fig. 4E&F), which means that it most closely resembles the ones from the randomly generated constructed **B** matrices. This is naturally followed by a low mean $SD_{rel}(\lambda)$ of between 0.26 and 0.38 (larger ones are produced in the type-2) which is also close to the measures of the randomly constructed **B** matrices. The spread of $SD_{rel}(\lambda)$ as measured by its variance is low for all and ranges between 0.005 and 0.011

It is still noteworthy that the connecting line between the first and second eigenvalue (λ_1 and λ_2) is often a little steeper than the others indicating that there is some trait integration. The mean of the first eigenvalue λ_1 also consistently has the highest variance around it.

The patterns of the pairwise minimum angles also broadly fall into the same categories: Most of the combinations that show "random trait integration" have means of around 40° where half of all measurements fall somewhere between 30° and 45° for all four angles between eigenvectors (Fig. 5A). However, there are some combinations that have a smaller first angle, such as most noticeably "type-2 two level hierarchical". While the four angles are on average very similar, it can be seen across all plots that the later ones have some more angles that are larger which may be caused by the deviations from orthogonality in R. For the "type-2 pleiotropic" combination that shows trait integration, the first angle is much smaller with a mean below 10° and very little spread around it. The other three angles look like those of the "random trait integration" ones except that their means are slightly smaller (Fig. 5B). In stark contrast, the angles of the "two distinct modules" category are so small that they are nearly 0°, while those of "no correlations" are exactly 0°.

Very little information is captured by the mean vectors, as they all always point in all four directions to an almost equal extent, with the only obvious exception being the combination that has no pleiotropy (each vector points in one direction with 1). For the "two distinct modules" category, the loadings are 0.34 and 0.37 with a relatively high variance of 0.13 around it. The remaining combinations, which fall into the categories of "random trait integration" and "overall trait integration" have loadings that range from 0.4 to 0.5 with variances that range from 0.3 to 0.7. For these combinations, the mean angles between the mean vectors and all others range from 15° to 30° with variances between 60 and 120. The only exception to this occurs for the one case of "overall trait integration", as the angle and its variance between the mean first eigenvector with the others are much lower and 6° and 14. Lastly, the "two distinct modules" has the highest mean angle of uniformly 45°, yet with the smallest variance of 8.

The eigenvector analysis was also performed on a set of randomly generated sets of four 4D vectors which are orthogonal to each other, and this serves as a null model. The resulting patterns do not differ from those that fit in the category of "random trait integration" which shows that the direction of eigenvectors that come from the stochastic matrices do not differ from total randomness.

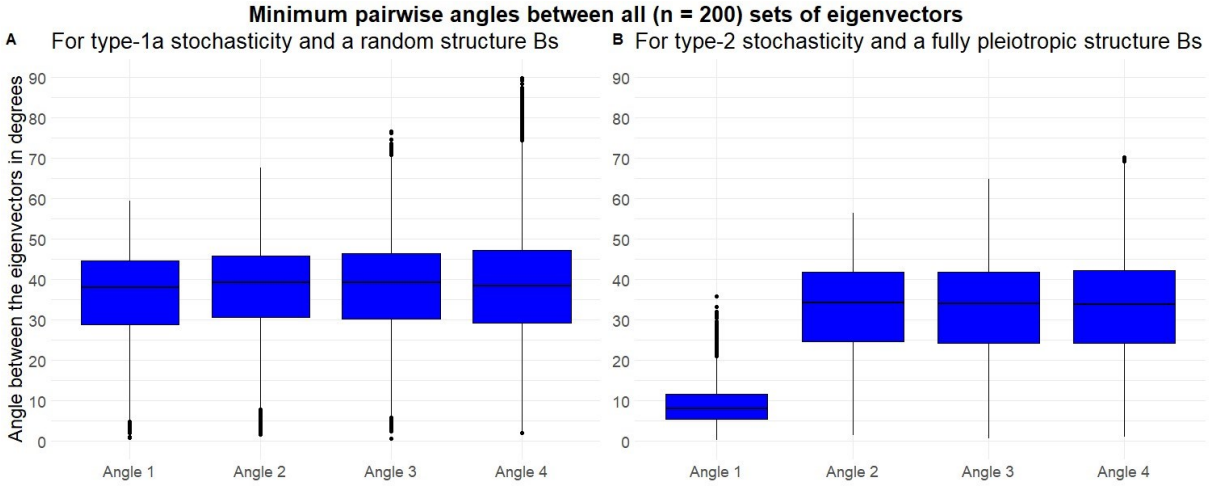

Figure 5: The boxplots of minimum pairwise angles between all (N = 200) generated sets of eigenvectors for the category of **A** "random trait integration" (here the combination "type-1a random") and **B** "overall trait integration" (the combination "type-2 pleiotropic") for all four angles. Except for the first angle in **B**, all mean angles are rather similar to one another. In **A** they are on average approximately 40° with 50% of the cases falling into a fairly small range around it. In **B** the first angle is around 10° with a small range around it yet some outliers. In both **A&B,** the number of angles that are larger increases with the number of the angle.

The topology frequency analysis shows that the above categorization based on eigenvalue distribution is reflected also in topologies, even though this method is not based on eigen-decomposition: For the "type-1a modular" matrices, there is only one (arbitrary) topology that is achievable because all **rM** are the same (Fig. 6A). The "two distinct modules" matrix group results in 6 out of the 15 possible topologies and the one that would be expected by its $\mathbf{B_s}$ has the highest frequency. The remaining five less frequently occurring topologies include the two alternative topologies with two modules –with ((A, B), (C, D)) topology – and some of the topologies that maintain one of the two modules in the $\mathbf{B_s}$ (Fig. 6B). All 18 remaining combinations have across the 200 rounds of stochastic generations of **B** matrices achieved all the 15 possible rooted topologies, showing little tendency towards specific topology, even though some topologies did occur in only very few cases.

Across these 18 topology frequency distributions (corresponding to 18 $\boldsymbol{\mu(B_s)}$ combinations), which include those with random structures, there is an increased frequency for the three two-module topologies (Fig. 6C). Such an organization has a 1-level hierarchy of trait correlatedness, which likely requires only minimal order to arise. It appears intuitively plausible that occurrence of stronger level hierarchy or some kind of specific order may be less likely than achieving the lack of it. However, there is still some diversity between these 18 frequency distributions, which mainly comes down to the underlying type of stochasticity. For those of type-1a and type-1b, there is little difference between those with a random or a non-random $\mathbf{B_s}$, which indicates that for them, the pleiotropic structure $\mathbf{B_s}$ makes practically no difference (Fig. 6C) as the variation likely overwhelms the structure. This is however different for type-2: The pleiotropic $\mathbf{B_s}$ leads to a frequency distribution that is more even than those of the random ones (Fig. 6D), whereas the three $\mathbf{B_s}$ that have an organization into two

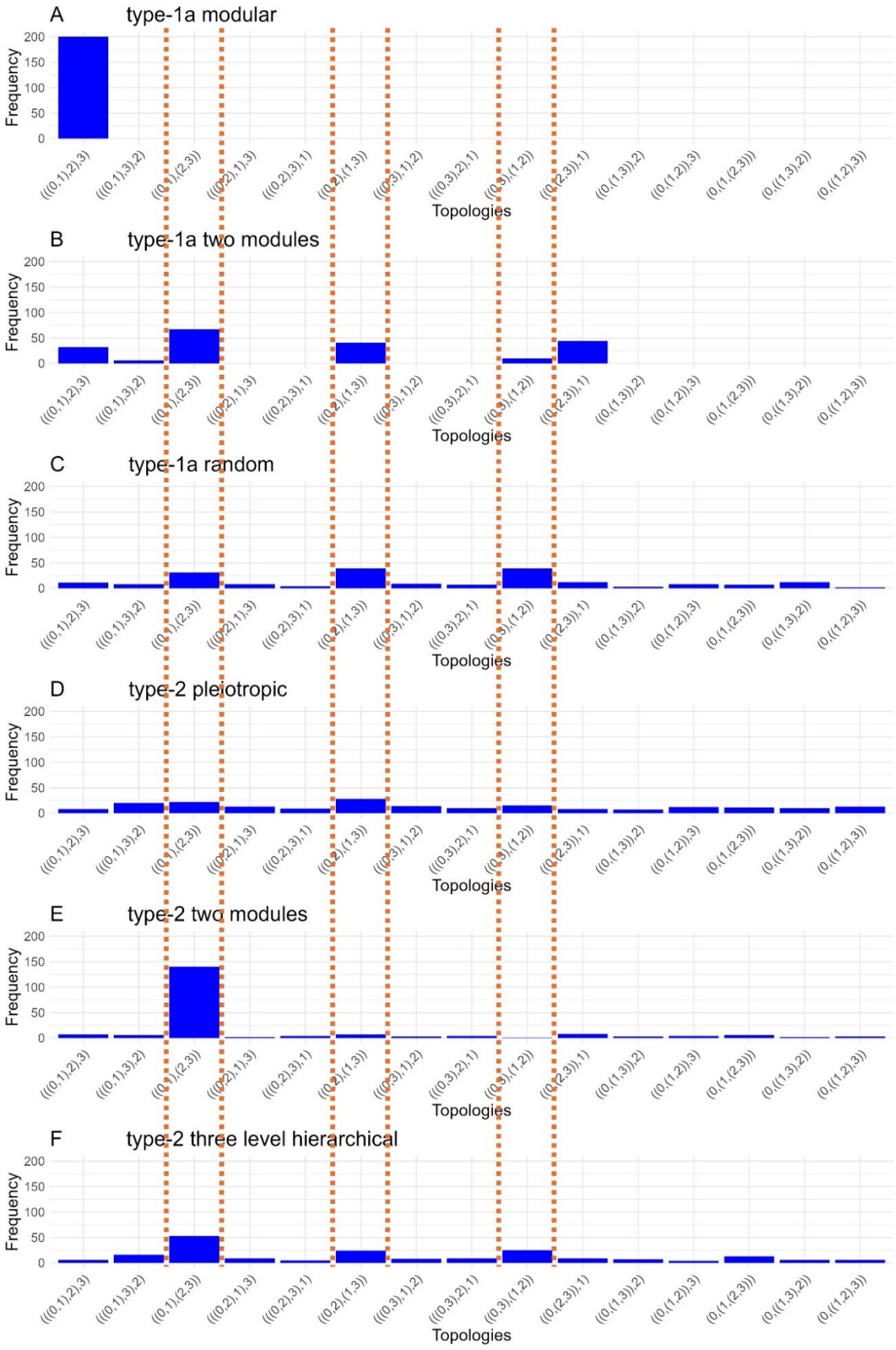

Figure 6: Topology frequency distribution from (N = 200) rounds of stochastic **B** generation for some of the different combinations. Only for **A&B** not all 15 possible rooted topologies have been achieved. The topologies that form two modules (marked with the orange dotted lines) have only one level of hierarchy and are often some of the most frequent topologies (here in all except **A&E**). The structures that have a distinction into two modules built into them (**B**, **E**, **F**) also tend to achieve this specific topology (the first that is marked by the dotted lines) more frequently.

modules built into them (two modules, two-level hierarchical and three-level hierarchical) achieve this exact topology (= ((0, 1), (2, 3))) much more frequently than the other combinations (Fig. 6E&F). The latter is also noticeable for the combination “type-1a two modules” (Fig. 6B). The degree to which this happens also directly corresponds to the strength by which this organization is built into the $\mathbf{B}_s$, which is determined by how many of the loci directly contribute to the organization into the two modules. For the “two modules” structure, all loci contribute to it and the resulting frequency distribution shows the highest number for the corresponding topology.

Eigen-patterns of empirical **rM** matrices

Importantly, these empirical correlation matrices **rM** naturally display positive and negative correlations, which is different from the correlation matrices based on the simpler constructed **B** matrices used in this work. For the process of eigen-decomposition this does not really matter much, as a vector´s loading on a trait can point in either direction. Instead, it is the absolute values of the correlations that matter for the eigenvalue distributions and measures that can be derived from it. However, for the attempt of this thesis to explore the relation of **rM** with **B**, this introduces a lot of complexity. Understanding why a particular **B** would lead to an **rM** with specific positive and negative correlations by looking at it (like we have done with the constructed **B** and the pleiotropic structures $\mathbf{B}_s$) is much less intuitive than it would be if there are only positive correlations. Yet more importantly, coming up with a **B** that is biologically interpretable while leading to such an **rM** would be very difficult.

The distribution patterns of eigenvalues across all investigated **M** matrices are similar, as can be seen when all eigenvalues are plotted together on the same scale (See Fig. 7). As would be expected, the observed patterns differ from each of the extreme patterns of eigenvalue distribution derived from the theoretical constructed and stochastic **B** matrices. There is no obvious indication of hierarchy or the formation of two or more distinct modules. In a few cases, the distribution describes a line which looks similar to the distributions of the stochastic **B** matrices that fall into the category of “random pleiotropy” of the random constructed **B** matrices. However, in most cases, more than half of the total variance is represented by the first eigenvalue λ_1 (See Table 2), which is also seen in how pronounced they are. This indicates that there is considerable trait integration. It is then noteworthy that particularly the first eigenvectors which correspond to the relatively largest eigenvalues tend to have near equal absolute loadings for all traits. This is equivalent to a “size factor” and perfectly recapitulates the findings of Wagner (1984) that this is how empirical correlation matrices deviate from random ones.

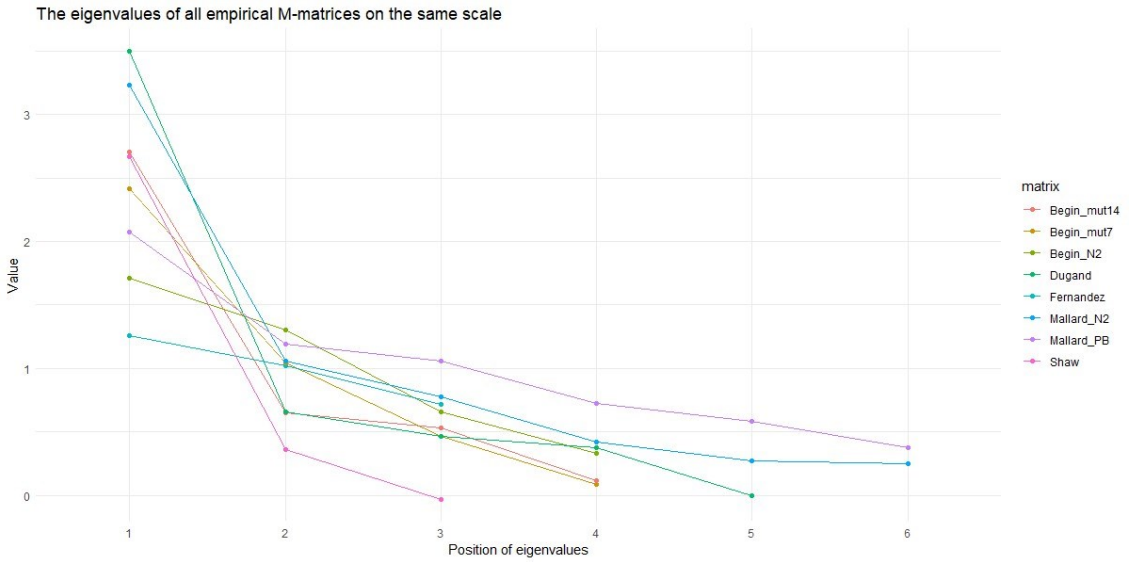

Figure 7: The eigenvalue distributions of all empirical **M** matrices from the publications in Table 1 on the same scale.

The relative eigenvalue standard deviations $SD_{rel}(\lambda)$ derived from the **rM** matrices range between 0.2 and 1 (See Table 2). The ratios between the first eigenvalue and the total variance (which, for correlation matrices, is equal to the number of traits *n*) are similar to the $SD_{rel}(\lambda)$ for the intermediary values. If eigen-decomposition is performed on the original covariance matrices **M** (instead of the correlation matrices **rM**), the relative contribution of λ_1 grows in all cases, although the extent to which it does depends on the differences between the trait variances in the different empirical **M**. This in turn suggests that the traits which are more integrated also tend to have more variance.

Table 2: Measures of integration calculated from the eigenvalues of the correlation matrices from the publications in Table 1. The eight matrices are order by $SD_{rel}(\lambda)$.

Matrix	**Traits (n)**	**$SD_{rel}(\lambda)$**	**λ_1 / n**	**Description of traits:**
Shaw	3	1.03	0.89	reproductive traits of the plant *A. thaliana*
Dugand	5	0.71	0.70	wing shape and size of the fly *D. serrata*
Begin_mut14	4	0.67	0.68	life history and body size of the nematode *C. elegans*
Begin_mut7	4	0.59	0.60	life history and body size of the nematode *C. elegans*
Mallard_N2	6	0.51	0.54	locomotion behaviour of the nematode *C. elegans*
Begin_N2	4	0.36	0.43	life history and body size of the nematode *C. elegans*
Mallard_PB	6	0.27	0.35	locomotion behaviour of the nematode *C. elegans*
Fernandez	3	0.19	0.42	life history of the fly *D. melanogaster*

It is noteworthy that the matrices that are from the same publications and that are measuring the same set of traits in different strains of the same species, are quite different. For example, in the data from Bégin & Schoen (2006), the **mut-14** strain of *C. elegans* has a very pronounced λ_1, while the eigenvalue distributions of strains **mut-7** and **N2** show less overall trait integration that can be focused into a single dimension. When instead of a visual assessment of eigenvalue distributions we use the summary measures in Table 2, **mut7** and **mut14** are much more similar, while **N2** still shows far less trait integration. This is particularly interesting due to the context of that experimental study. It is fully expected that mutator strains would have larger trait (co)variances due to mutational effects (represented by the **M** matrix), as the mutagenic treatment is explicitly designed to increase the number of mutations and the authors conclude that it has. Yet, the described patterns are based on the correlation matrices **rM** which are normalized in respects to the expectedly greater variances. Therefore, the correlations of the mutational effects on the traits are stronger in the mutagenic lines than they are in the **N2** lines. This means that the mutations that happen in them tend to affect more traits. Further, this suggests that this may be due to a difference between mutations that naturally occur and those that are mutagenically caused, in this case by the transposition of selfish elements. However, a direct comparison between the two strains from Mallard et al. (2023) reveals a similar picture, as there seems to be much stronger trait integration in **N2** than in **PB**. Yet in this case, there is no mutagenesis in either line which means that it is also possible for different strains of the same species to vary greatly in respects to the correlatedness of mutational effects.

Analysis of the constructed **B** matrix with entries {1, -1}

The k terms that make up the equation ⟨1⟩ can all either be +1 if the entries in the same column are both the same ($1^2 = (-1)^2 = 1$) or -1 if they are different ($-1 \cdot 1 = -1$). Full hidden pleiotropy – a lack of covariances – is achieved when all off-diagonal entries in **M** are 0. For each entry $m_{x,y}$ this means that there must be the same number of summands +1s and -1s. I will refer to rows that fulfill this as being *compatible* (in the context of hidden pleiotropy). Note that for **B** matrices with entries +1 and -1, the model is limited to matrices with an even number of columns (loci). For the sake of demonstrating the trends analytically, this limitation is not of concern.

First, let us consider how number of loci affects the likelihood of encountering a hidden pleiotropic G-P map, given two traits (number of columns = k, number of rows = n = 2). The probability that the two entries in the same column have the same sign is ½ (their product resulting in +1) and so is the probability that they are different (their product resulting in -1). For the sum of products to cancel out, half of the columns must give the product +1 and half of them -1. The probability for +1 and -1 to

coincide, is $(1/2)^2$. As there are two ways to achieve this (+1, -1 or -1, +1), the likelihood to generate hidden pleiotropy in our fully pleiotropic 2X2 matrix is $((1/2)^2 \cdot 2 =) 1/2$. Thus, there are $½ \cdot 2^{2 \cdot 2} = 8$ ways to obtain a hidden-pleiotropic matrix.

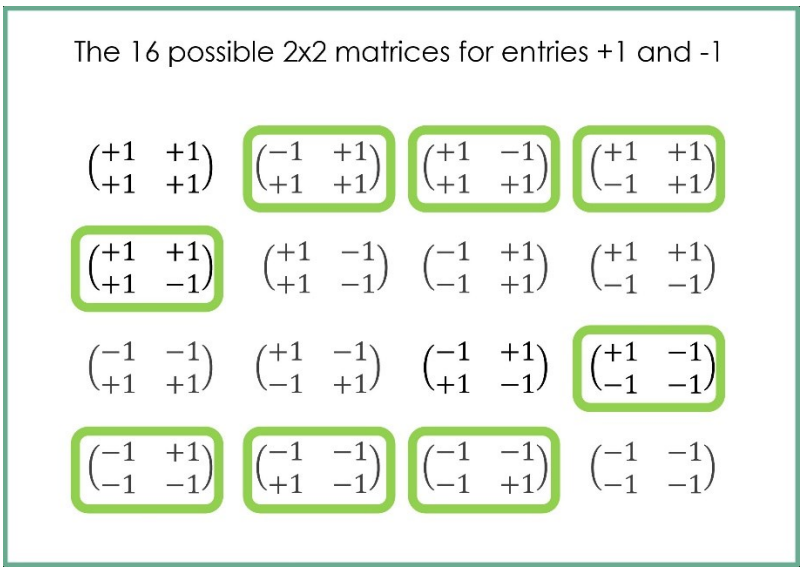

Figure 8: The 16 possible 2X2 matrices with entries +1 and -1. The covariance of the two traits is calculated with the dot product of the two rows. This leads to a sum of two products which may either be +1 (if the signs of the two entries in the column are the same) or -1 (if the signs of these two entries are different). For a covariance (and further correlation) of 0, there must be equal numbers (one and one) of summands +1 and -1. In total, 8 matrices fulfill this condition and are thus hidden-pleitropic. These are notably the ones that have three entries with one sign and one with the other. If one further considers all 16 matrices to be equally probable, than the probability for achieving hidden pleiotropy for a random 2X2 matrix is 0.5.

Next, let us increase the number of loci to a 2X4 **B** matrix, which can be realized in $2^{2 \cdot 4} = 256$ possible ways. However, if using the equation of the dot product ⟨1⟩ of the two rows we only focus on the probabilities of the individual products to be summed, there are only $2^4 = 16$ different possible sequences of +1 and -1. Of these, 6 have exactly two of each. It follows:

$$P_{hp}(\mathbf{B}_{2X4}) = 6/16 = 0.375$$

Thus, 96 out of a total of 256 possible rows are compatible. We can formalize this principle by modifying the formula for distinguishable permutations to get the number of compatible rows $c_{n,k}$ for $n = 2$ and any even k:

⟨2⟩ $$for\ n = 2, k \in 2\mathbb{N} \qquad c_{n,k} = \frac{k!}{\left(\frac{k}{2}!\right)^2}$$

The fraction of compatible rows out of all possible rows is then simultaneously also the probability of randomly drawing a hidden-pleiotropic 2Xk matrix:

⟨3⟩ $$for\ n = 2, k \in 2\mathbb{N} \qquad P_{hp}(B_{n,k}) = \frac{c_{n,k}}{2^k}$$

We can use this equation to calculate the probabilities for different number of loci and plot them (See Fig. 9). The curve shows that the probability of randomly encountering a hidden pleiotropic **B** matrix declines exponentially with the increasing number of loci. The obtained probabilities are consistent with the fractions of hidden pleiotropic matrices out of 10e5 randomly generated matrices with random entries +1 and -1. The deviations of these numerical simulations from the analytical solution above were small and without systematic tendency.

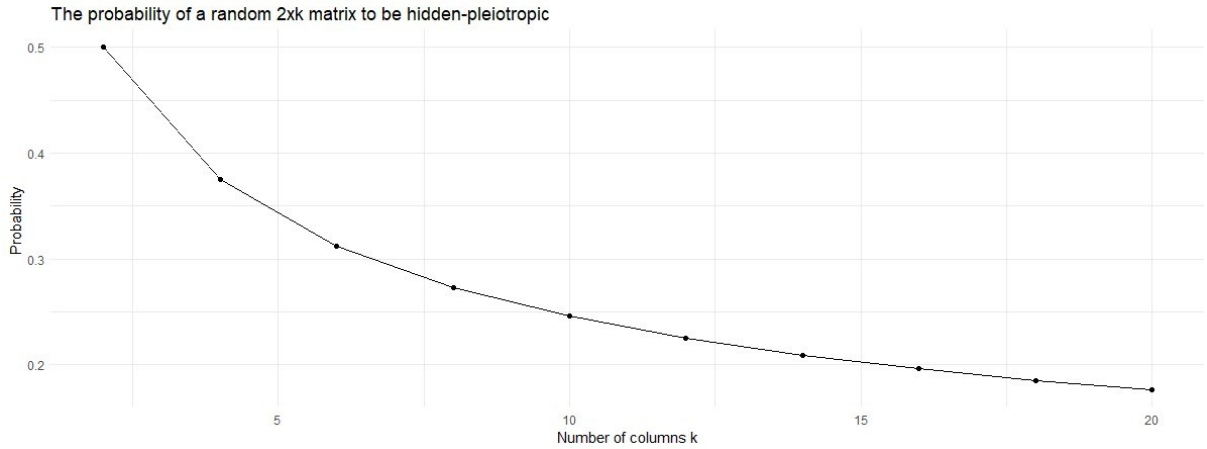

Figure 9: The probabilities of achieving hidden pleiotropy for a matrix with n = 2 rows and k = 2 to 40 plotted as a curve. The probabilities can also be understood as the fraction of all 2^{n*k} possible matrices that are hidden-pleiotropic. They consistently decrease in what appears to be negative exponential growth.

Next, we consider how the number of rows *n* impacts the probability of a hidden pleiotropic matrix. We start with increasing *n* from 2 to 3. Zero pairwise covariances mean that dot products between all combinations of two row vectors of a **B** matrix are 0. Only a subset of possible row vectors gives such result for any two vectors. Introducing a further vector with which to pairwise match the existing row vectors will reduce this subset further, while at the same time increasing the overall number of possible **B** matrices. Already from this consideration, it is expected that increasing the number of traits will make hidden pleiotropy less likely. By adding traits, the number of covariances also increases:

$$\langle 4 \rangle \qquad C(n) = \frac{n^2 - n}{2}$$

Because each individual row in a **B** matrix is involved in n-1 covariances and must have dot products equal to 0 with each other row of a **B** matrix for it to be hidden pleiotropic, it means that removing any number of rows from a larger hp matrix always maintains a hp (sub)matrix. Conversely, the possible hp solutions for a matrix of the size (n+1)X*k* are a subset of the hp solutions of a matrix of the size *n*X*k*.

Note that a hidden pleiotropic 3X2 **B** matrix with elements (+1, -1) is not possible (See the appendix for an explanation). Therefore, we start with the probability of a hidden pleiotropic 3X4 matrix. Using the above rule, we can use the set of the solutions from 2X4 **B** matrix to explore which entries in the

third row (2^k = 16 possible combinations) will maintain hidden pleiotropy. Thereby we may subset the new matrix into three 2X4 submatrices, each for a combination of two traits. Of these, one is equivalent to the original 2X4 matrix whereas the other two include the newly added row. For these two new submatrices, the 6/16 probability from considering pairs of entries in a 2X4 matrix above applies. The question becomes, how many (if any) of these 6 are shared among submatrix solutions.

It quickly becomes clear that the compatibility between two rows mainly depends on if their numbers of +1s and -1s are even (including = 0) or odd. A row with an even number of +1s and -1s can never be compatible with one that has odd numbers and vice versa. Since always exactly half of all possible rows will have even/odd numbers of the two signs (for an even *k*), we already can exclude the half that does not share the first row´s even/oddness from being considered for hidden pleiotropy. Yet, of the eight rows that have even/odd numbers, only six are also compatible. For pairing up two rows to a 2X4 matrix by adding a compatible row to the first, there are therefore always exactly two that do not result in dot products = 0 with the first row and these are the row itself (the dot product is the trait´s variance $m_{1,1}$) and its inverse ($-1 \cdot m_{1,1}$).

Adding a row onto a hidden pleiotropic 2X4 matrix, out of the 6/16 compatible rows for each of the two new 2-row submatrices, there is an overlap of four, while two are not shared. These two are the row from the original 2X4 matrix that is not present in the respective 2X4 submatrix and its inverse. Expressed differently, from the 8/16 rows that share their even/oddness of numbers of signs with the original hp 2X4 matrix, the two rows that are present in it and their inverses are not compatible. The probability for randomly encountering a 3X4 hp matrix is therefore:

$$P_{hp}(\mathbf{B}_{3X4}) = (16/16) \cdot 6/16 \cdot 4/16 = 0.09375$$

Which is less than $P_{hp}(\mathbf{B}_{2X4})$. In numbers, out of 4096 possible matrices, only 384 are hp. Increasing the number of traits, i.e., adding a row, thus decreases the probability of randomly generating a hidden pleiotropic G-P map, as the corresponding probability based on a focal matrix is multiplied by a fraction <1. Following the same reasoning as above, one can consider a 4X4 hp matrix. The overlap of the 6/16 possible row combinations for each of the now three new submatrices is exactly two. The probability is then:

$$P_{hp}(\mathbf{B}_{4X4}) = (16/16\cdot) \, 6/16 \cdot 4/16 \cdot 2/16 = 0.01171875$$

In numbers, this means that out of 65536 possible 4X4 matrices, 768 **B** matrices are hp. We can now formalize these two steps for 4-loci matrices into an expression:

$\langle 5 \rangle$ $\quad for\ n \in (\{2,3,4\}), k = 4 \quad$ $P_{hp}(B_{n,k}) = \prod_{x=2}^{n} \frac{c_{2,k} - 2 \cdot (x-2)}{2^k}$

with *n* = number of traits, *k* = number of loci and $c_{2,k}$ = number of compatible rows for *n* = 2 traits and *k* loci. This expression can also be used to show that no 5X4 **B** matrix with elements (+1, -1) can be hidden pleiotropic:

$$c_{x,k} = c_{2,4} - 2 \cdot (x-2)$$

Setting *x* = 5 shows that for extending a hp 4X4 matrix with another row, there are no compatible rows, and the probability is therefore 0:

$$for\ x = 5, k = 4 \qquad c_{5,4} = \frac{4!}{(2!)^2} - 2 \cdot (5-2) = 0$$

$$P_{hp}(\mathbf{B}_{5X4}) = P_{hp}(\mathbf{B}_{4X4}) \cdot 0/16 = 0$$

Expression $\langle 5 \rangle$ also will have features in common with a potential general formula for any *n* and *k*, which was not derived here. We know that probability $P_{hp}(\mathbf{B}_{nXk})$ must be the product of *n* fractions, which in turn depend on *k*. Each individual fraction will be the number of compatible rows $c_{x,k}$ for the extension of the matrix from *x*-1 to *x* rows, divided by the total number of possible rows 2^k. The first fraction will always be equal to 1 because we may start with any of the 2^k possible rows. The number of compatible rows for extending the matrix to 2 rows $c_{2,k}$ is then given by the modification of the formula for distinguishable permutations in formula $\langle 2 \rangle$. This much must be true for any dimensionality. The only thing that changes then, is the number by which $c_{2,k}$ is reduced for each *x* > 2. A formula that gives the probability of hidden pleiotropy for any dimensionality therefor only differs from formula $\langle 5 \rangle$ in that the number 2 with which (x – 2) is multiplied (and with which $c_{2,k}$ is reduced to get to the specific $c_{x,k}$) is replaced by a generalized function *r*(*x*,*k*).

For even larger dimensionalities I resorted to using an algorithmic approach which generates all 2^k possible rows for all even $k \leq 20$. Instead of constructing all the possible submatrices, the number of compatible rows was sequentially lowered by checking how many (if any) of the rows would be compatible to the first one. This is akin to having added one of the compatible rows and then checking each other row´s compatibility with it to create a new and smaller subset. The use of such an algorithm is justified by what we have learned about the overlap of compatible rows.

Generating the numbers of compatible rows for extending a hidden-pleiotropic matrix like this we see that not all compatible rows are equivalent regarding how they reduce the subset from $c_{x-1,k}$ to $c_{x,k}$ and some (e.g., rows that only consist of +1 or -1) will even preemptively reduce $c_{x,k}$ to 0. This is

remarkable as the above-described idea of calculating the probability of hidden pleiotropy by multiplying the fractions of compatible rows out of all possible rows in a row-by-row manner relies on the yet not explicitly stated principle that regarding the probability, that it must not matter which rows are added in what order. It should not matter which of all possible rows is the first row and all subsequentially added rows out of the subset of compatible rows are interchangeable in so far as to how large the overlap with the other rows´ subset will be. However, while these principles are certainly true for *k* = 4, they are not true for all larger *k* (in particular, this has been shown with *k* = 12 or 20).

Table 3: The numbers $c_{x,k}$ of compatible rows available for extending a hidden-pleiotropic matrix from x-1 to x traits (first row), given k loci (first column). The numbers were obtained from a reshuffled table and those based on an unshuffled one can be viewed in the supplementary material. The numbers with a green background increased with the reshuffling, whereas the ones with a red one decreased. By dividing entries with the total number of rows 2^k one can get the fraction of how many rows are compatible for each extension. Within a row (k), these fractions can then be multiplied with each other up to the specific n in order to find the probability of drawing a hidden-pleiotropic nXk matrix. This was done in Table 4.

k**x	**1 (2^k)**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
2	4	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	16	6	4	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	64	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	256	70	36	18	16	6	4	2	0	0	0	0	0	0	0	0	0	0	0	0
10	1024	252	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	4096	924	400	164	64	26	12	10	8	6	4	2	0	0	0	0	0	0	0	0
14	16384	3432	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	65536	12870	4900	1810	648	230	76	42	16	14	12	10	8	6	4	2	0	0	0	0
18	262144	48620	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	1048576	184756	63504	21252	6912	2202	676	212	72	38	36	18	16	14	12	10	8	6	4	2

Methodologically, this issue is partly resolved by simply reshuffling the order of all possible rows which has been done in Table 3. From what we see in the non-zero $c_{x,k}$ of the reshuffled rows, we can make the following statements about n_{max} which are also in accordance with what different values for *k* have been possible while generating 4X*k* hp $\mathbf{B_s}$ matrices for the stochastic simulations:

- (If *k* is odd, hidden pleiotropy is not possible at all -> n_{max} = 1)
- If *k* is two times an odd number (6, 10, 14, 18), n_{max} = 2.
- If *k* is divisible by 4 (2, 4, 8, 12, 16, 20), hidden pleiotropy is possible for n_{max} = *k*.

Unfortunately, the fact that not all compatible rows are interchangeable in regards to their effect on $c_{x,k}$ prevents me from formulating a generalized formula for $P_{hp}(\mathbf{B}_{n\mathrm{X}k})$. However, we can still use the numbers of compatible rows to obtain approximated probabilities. This has the advantage over numerical simulations that it can more easily detect smaller probabilities. All numbers $c_{x,k}$ from Table 3 can be divided by their respective 2^k and these fractions can then be multiplied up to any n to obtain $P_{hp}(\mathbf{B}_{n\mathrm{X}k})$. The resulting probabilities are given in Table 4. If we now compare the obtained numbers with the fraction of hidden-pleiotropic matrices out of 1E+6 randomly generated ones, we see that they are very similar. The number of achieved cases of hidden pleiotropy differed at most a few hundred from the number that would be expect from Table 4 and the difference even shrinks further with increasing n. Thus, we can confirm that this method for approximating the probabilities $P_{hp}(\mathbf{B}_{n\mathrm{X}k})$ produces numbers that are in the same magnitude as the actual probabilities. It is therefore a good tool to approximate probabilities for larger dimensionalities that would otherwise be too small to be captured in reasonably extensive numerical simulations.

Table 4: The probabilities $P_{hp}(\mathbf{B}_{n\mathrm{X}k})$ for drawing a hidden-pleiotropic **B** matrix with different numbers of loci k (first column, grey) and traits n (first row, grey). They were calculated by first dividing the number of compatible rows $c_{x,k}$ in Table 3 with the total number of rows 2^k for each k to obtain the fractions. These fractions can then be multiplied with each other up to the specific n to calculate the probability of drawing a hidden-pleiotropic **B** matrix with dimensions nXk.

k**n	**1 = 2^k**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
2	1	5.00E-01	0	0	0	0	0	0	0	0
4	1	3.75E-01	9.38E-02	1.17E-02	0	0	0	0	0	0
6	1	3.13E-01	0	0	0	0	0	0	0	0
8	1	2.73E-01	3.85E-02	2.70E-03	8.45E-05	1.98E-06	3.09E-08	2.42E-10	0	0
10	1	2.46E-01	0	0	0	0	0	0	0	0
12	1	2.26E-01	2.20E-02	8.82E-04	1.38E-05	8.75E-08	2.56E-10	6.26E-13	1.22E-15	1.79E-18
14	1	2.09E-01	0	0	0	0	0	0	0	0
16	1	1.96E-01	1.47E-02	4.06E-04	4.01E-06	1.41E-08	1.63E-11	1.05E-14	2.55E-18	5.45E-22
18	1	1.85E-01	0	0	0	0	0	0	0	0
20	1	1.76E-01	1.07E-02	2.16E-04	1.43E-06	2.99E-09	1.93E-12	3.90E-16	2.68E-20	9.71E-25
k**n	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
12	1.75E-21	8.54E-25	0	0	0	0	0	0	0	0
16	9.99E-26	1.52E-29	1.86E-33	1.70E-37	1.04E-41	3.17E-46	0	0	0	0
20	3.33E-29	5.72E-34	8.73E-39	1.17E-43	1.33E-48	1.27E-53	9.71E-59	5.55E-64	2.12E-69	4.04E-75

Regarding the effects of the matrix dimensions on $P_{hp}(\mathbf{B}_{n\times k})$, Table 4 tells us that the probability of generating zero trait correlations in the face of full pleiotropy is reduced with both *n* and *k*, just as we have seen with the dimensionalities that we covered before. However, we can now also see a synergistic interaction between *n* and *k*. If we compare the decrease of (non-zero) probabilities for going from *n*-1 to *n* traits for different number of loci *k*, we see that the probabilities decrease stronger the higher *k* is. Thus, the probability $P_{hp}(\mathbf{B}_{n\times k})$ shrinks non-linearly with both *n* and *k*. This also means, that it is not possible to just calculate $P_{hp}(\mathbf{B}_{n\times k})$ for any possible *n* by starting out with the probability $P_{hp}(\mathbf{B}_{2\times k})$ for $n = 2$ and then apply a uniform reduction in probability to get from 2 to *n* rows. The necessary reduction is not uniform, but is a function of both *n* and *k*.

Table 4 also shows how fast the reduction in $P_{hp}(\mathbf{B}_{n\times k})$ (for possible dimensions with $n \leq n_{max}$) becomes with an increase in the number of traits. The shown regularities give no reason to believe that probabilities would increase in larger matrices, therefore further simulations are not necessary. Indeed, we are quickly left with probabilities that are so low that testing them against randomly generated matrices is no longer feasible, as it requires a large number of iterations to find a single hidden pleiotropic matrix. Including larger *k* would in principle be possible but would similarly increase computational resources to first generate all possible rows and then calculate their dot products of as the formula $\langle 2 \rangle$ for $c_{2,k}$ includes factorials. Yet, doing so would not be particularly useful anyways.

Analysis of hidden pleiotropy in stochastic matrices

In addition to the designed matrices of +1s and -1s, we are here considering stochastically generated matrices. We explore, how the likelihood of a hidden pleiotropic matrix is affected by expanding the choice of potential matrix entries from two integer numbers [+1, -1] to normal distributions and real numbers. The stochastic matrices are combinations $\boldsymbol{\mu(B_s)}$ of a hidden-pleiotropic or random structure and the type-2 variational aspect which treats its entries $b_{sx,y}$ as means of normal distributions around which the stochastic entries are drawn with a SD of 1. For the hp $\mathbf{B_s}$, this means that the numbers are either drawn around a mean of +1 or of -1. This respective expected value, which is also at the peak of the probability distribution, is by definition of the hp $\mathbf{B_s}$ also able to result in a lack of correlation along with all other means. If out of all the normal distributions for all entries of the stochastic **B** the exact mean would be drawn, it would be identical to $\mathbf{B_s}$ and generate a lack of genetic correlations. However, since we are here dealing with real numbers, it is very unlikely to draw the exact means in all elements, or to generate an **rM** matrix with all correlations exactly 0. We would thus need to introduce an arbitrary interval around 0 as a cut-off point that defines what amount of correlation still counts as

“negligible”. The more or less stringent this interval is, the less or more cases of “no” trait correlation will be detectable.

Ultimately, there are two not mutually exclusive ways which we can imagine resulting in a lack of considerable correlation between traits in a stochastic **B** based on a hidden-pleiotropic structure $\mathbf{B_s}$. It is possible that all realizations in **B** that contribute to a particular entry of **rM** fall very close to the mean. In this case, the stochastic **B** would strongly resemble the $\mathbf{B_s}$ which we already know to be hidden-pleiotropic. The other way is that the realized entries in **B** more strongly deviate from the means of their distribution, but that these deviations cancel each other out. However, the larger a deviation is, the lower is the chance of an oppositely signed deviation in a different row to be able to cancel it out because the mean is at the peak of the normal distribution. Thus, if we revisit thinking of probabilities in terms of expanding matrices with compatible rows as we did with the matrices with entries +1 or -1, we can think of a draw that has a large deviation from its mean as making it all the less likely for a compatible row to be drawn. It is thus easiest to achieve a lack of genetic correlation if all draws are as close to the mean as possible. In this case we may calculate the probability of drawing a value out of a small range around the mean by calculating the area under the peak with an integral. Depending on the range that we decide on (which is not the same as the range around 0 for the correlations), drawing an entry in it will be more or less likely. However, since the mean is at the peak, no other area below the distribution curve with the same length on the x-axis could be larger than it. A consideration of the integral below the peak then also tells us that while increasing the range around the mean for entries $b_{x,y}$ also increases the probability, it does not grow linearly with it. This is due to the shape of the peak. If we were to double the total range around the mean on the x-axis, the probability would not double, as the area below the bell curve is not twice the size as before. Given then how the covariances $m_{x,y}$ are calculated with equation ⟨1⟩, we may reasonably assume that the same principle also applies to the range that serves as a cut-off point for a correlation of nearly 0.

Let us now consider how all of this relates to stochastic matrices with a random $\mathbf{B_s}$ and contrast it with the hidden-pleiotropic one: Each entry is randomly drawn to be either +1 or -1 and the variational aspect then draws entries around this drawn value as a mean with a SD of 1. Thus, if we consider the probability distributions of both random draws together for any entry $b_{x,y}$, we arrive at a single bimodal distribution which is composed of the two normal distributions around the two original means of +1 and -1. In this case, the mean and peak of the combined probability distribution is 0. This is significant because it means that, if for all entries the mean would be drawn, there would be no correlations since there being no effects at all. From this we learn that when lack of correlation is achieved from a random $\mathbf{B_s}$, this will most often be because the mutational effects on the traits are small. However, it would also be permissible for a total lack of correlations that per locus there is at maximum one non-zero entry

which then affects one particular trait, resembling the non-pleiotropic (modular) constructed **B** and $\mathbf{B_s}$ that have been covered before.

For a more realistic consideration of real numbers, it is once again more useful to think about ranges of deviations around the mean. Again, we can calculate the probability of an entry being drawn within a small range around the mean of 0 by calculating the integral below the peak. However, if we do this for the same small range around the means for both the random and hidden-pleiotropic $\mathbf{B_s}$, we will get a lower probability for the random one, i.e., the area below the peak covers a smaller proportion of the total area. Therefore, we can say that – at least for this particular theoretical model of stochastic **B** matrices – it is more likely to observe zero correlation when starting from a hidden-pleiotropic overall structure than from a random one. In other words, it is easier to maintain than to generate uncorrelated traits.

Hidden pleiotropy in numerical simulations of stochastic matrices

For the numerical simulations with hidden-pleiotropic stochastic **B** matrices, all *n*X*k* matrices were constructed by omitting rows from 10X*k* template $\mathbf{B_s}$ matrices. These in turn were constructed by expanding either randomly generated 4X*k* hp matrices for *k* = (12, 16, 24, 40) or a manually constructed 6X20 hidden-pleiotropic **B** matrix row-by-row. How many runs it took for the 4X*k* matrices to be expanded to 10X*k* was recorded and is plotted in Fig. 10:

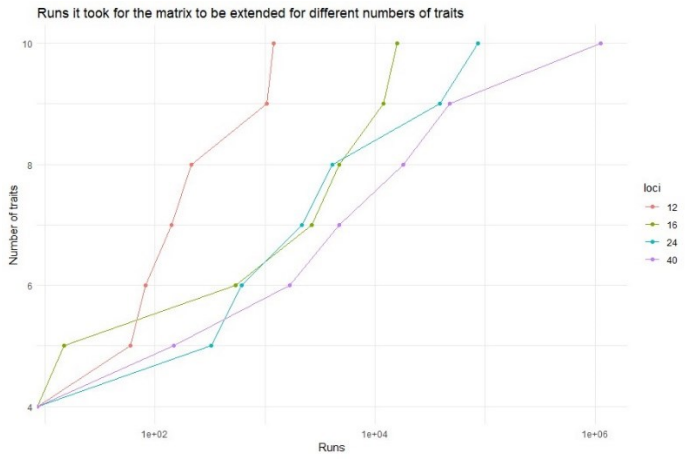

Figure 10: How many runs it took for the row-by-row expansion of hidden-pleiotropic matrices with *n* = 4 rows and different *k* (12, 16, 24, 40) to *n* = 10. The higher the number of loci k is, the longer it generally takes for a hidden-pleiotropic matrix to be extended from *n*-1 to *n* traits.

Fig. 10 is in general agreement with the analytical results on the constructed matrices with entries (+1, -1): An increase in the number of loci, as well as an increase in the number of traits lowers the number of compatible rows $c_{x,k}$ out of the 2^k possible rows. A lower probability of the solution results in on average more trials until a compatible row is found. As a compatible row is searched by testing of random rows, there can be considerable fluctuations in the actual number of rounds that it takes to find one. This may explain why the curves in Fig. 10 appear erratic and why e.g., the matrix with 16 loci grew faster to 5 rows than the one with 12.

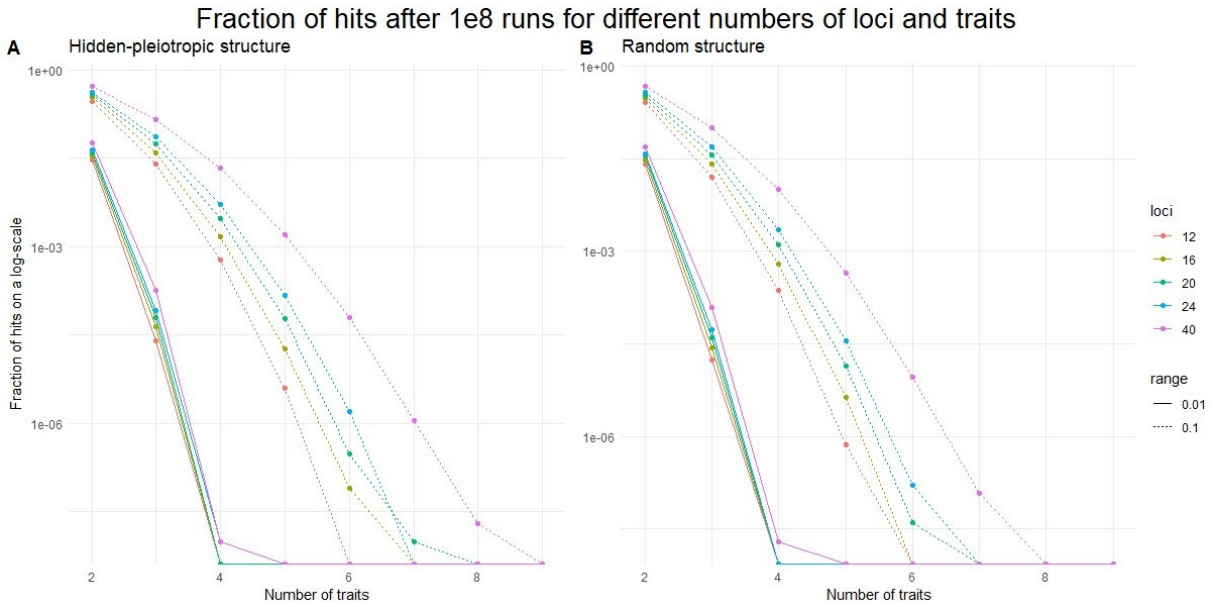

Figure 11: The fraction of stochastic B matrices that resulted in "practically" no correlations after 1E8 runs for different loci k and traits n. For counting as hidden pleiotropy, all correlations of **rM** were within a specified interval around 0 (either 0.01 or 0.1). **(A)** displays the fractions of the runs with a hidden-pleiotropic $\mathbf{B_s}$ structure and **(B)** the ones of the runs with a randomly redrawn $\mathbf{B_s}$. Both graphs are on a log10 scale.

The fractions of stochastic **B** matrices that resulted in correlation matrices **rM** in which all correlations were within a small interval around 0 are shown in Fig. 11. Both kinds of simulations, the one based on a hidden-pleiotropic structure (A) or a randomly redrawn one (B), show very similar patterns. An increase in traits and a more stringent interval around 0 both decrease the probability of finding a hidden pleiotropic matrix, i.e., a correlation matrix with next to no correlations. For higher numbers of traits, such an event becomes very unlikely very fast. This is very much expected and replicates what we have learned from the probabilities that resulted from the analysis of the simple matrices with elements (+1, -1). There, increasing the size of a **B** matrix by one row would lead to the probability of it being hidden-pleiotropic having to be multiplied with an ever-decreasing fraction of compatible rows. Yet, the number of hits slightly increases with the number of loci, which contrasts with what we have seen for the probabilities of the simple matrices with elements (+1, -1). This increase appears to be roughly the same regardless of the $\mathbf{B_s}$ structure. However, it becomes more pronounced with the number of traits which suggests that the law of large numbers is responsible for this phenomenon.

For better visibility of the different $\mathbf{B}_s$ in Fig. 11, subsets of both data sets (A and B) that only contain $n < 6$ and the range of 0.1 was plotted separately in Fig. 12. In general, we can see that the stochastic **B** matrices that had a hidden-pleiotropic $\mathbf{B}_s$ were more often phenotypically modular as well, which confirms the previous theoretical considerations. However, the difference between the runs with the two different structures also grows considerable with *n* yet does not at all depend on either *k* or the range. Whereas the runs with the hidden-pleiotropic structure were only 1.2 times as likely to lead to lack of correlation for $n = 2$, they were 10 times more likely for $n = 7$ (the highest *n* at which any hits were made for the stochastic matrices based on the random structure).

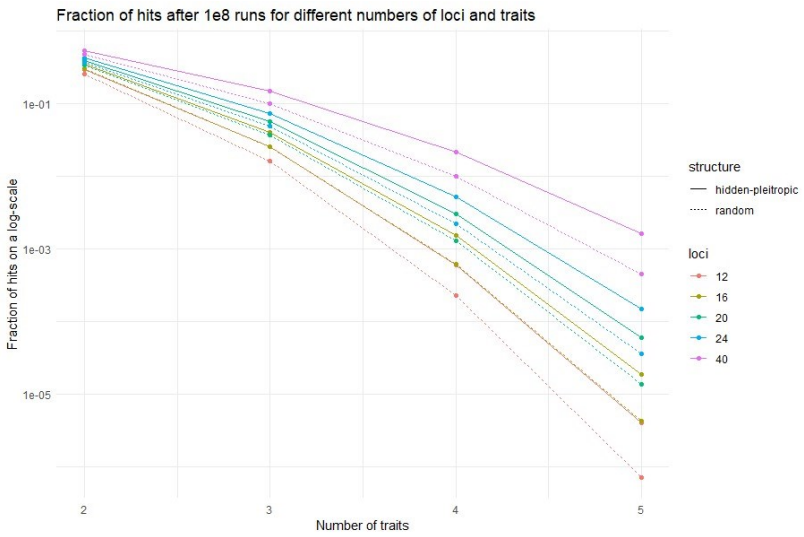

Figure 12: A subset of both sets of fractions displayed in Fig. z A and B. The displayed data points have a range of 0.1 and numbers of traits $n > 6$ were omitted. The graph is on a log10 scale.

Data accessibility

Due to the large amount of data that was produced, it is not feasible to display everything within the text of this thesis. Instead, all data sets (both tables and plots) from all chapters are accessible in the form of XLSX files in the supplementary material. To optimize readability and accessibility, they are curated in such a way that everything that was produced for e.g., a constructed **B** or a stochastic $\boldsymbol{\mu}(\mathbf{B}_s)$, is displayed on one sheet. Similarly, all relevant R scripts are shared upon request.

Discussion

Trait integration across artificial and empirical **rM**

The selection of eight empirical **M** matrices was able to cover a large range or trait correlatedness (0.2 < $SD_{rel}(\lambda)$ < 1.1) and this suggests that the strength of pleiotropic connections or linkage in mutation effects varies greatly among them (Estes & Phillips 2006). As the effect of population structure is temporary and erodes over evolutionary timescales, this variability in trait integration also reflects the broader heritable variability that evolution may act on in the form of the genetic effect covariance matrix **G** (Lande 1980; reviewed in Estes & Phillips 2006). However, a certain fraction of positive correlations between traits may not reflect mechanistic trait interdependence and could instead be attributable to the design of MA experiments. Deleterious mutations may cause generalized effects in many traits (i.e., the organism is "sick") which would be a kind of mutational pleiotropy that does not reflect actual functional associations between traits (Estes & Phillips 2006).

How integrated traits are will greatly depend on the type of traits measured, how these develop, or what kinds of traits tend to be measured on specific organismal structures. For example, Wagner (1984) has made a distinction between "mammal-like" and "insect-like" trait integration. He argues that this distinction does not necessarily come from a fundamental genetic difference between the taxonomical groups. Rather, he argues that the traits that researchers tend to measure on different taxa, themselves differ. In mammals, these are often expectedly tightly integrated ones e.g., skull morphology. In insects, they are often less directly functionally or developmentally connected traits, such as wings and legs. In contrast, when functionally connected traits are measured in an insect, these may be equally strongly integrated. This prediction was confirmed by the **rM** matrix from Dugand et al. (2021) on fly morphology in *D. serrata*. Considering the level of trait integration for a given set of traits can thus be biologically (and technically) informative.

For example, the largest $SD_{rel}(\lambda)$ results from an **M** matrix that measures reproductive traits in *A. thaliana* (Shaw et al. 2000). Specifically, the correlation of around 1 between the number of fruits and the dry weight of the infructescence is responsible for this high level of trait integration. This correlation very likely reflects that the mutations that occurred could only change the number but not the size/ weight of the individual fruits. If mutations cause new variance in how many fruits grow on the plants in a population without affecting the fruit size, then this must necessarily also directly affect their combined dry weight. One could argue that these two measured traits are two aspects of only one trait. Yet if Shaw et al. (2000) would have treated them as such prior to conducting their experiment, they might just have measured one of them. This would have led to a smaller **M** with much lower

measured overall trait integration. This example shows that the measured trait integration in **M** matrices is sensitive to the choice of phenotypic measurements. As Wagner (1984) pointed out, traits that are functionally related will tend to have a higher level of integration. The functionality of a trait itself is commonly defined by it affecting an organism´s performance (survival, development or reproduction) which is similar to it being "adaptive" and functional traits may also be distinguished from purely descriptive traits (Kearney et al. 2021). A strong functional relation between two or more traits which then also is reflected in a high level of measured trait integration can thus be considered a strong indication of modularity in the pleiotropic structure. The three reproductive traits in the aforementioned **M** matrix from Shaw et al. (2000) can thus reasonably be considered to be part of the same module.

As there is no general consensus on what constitutes a trait, in QTL studies that take a lot of phenotypic measurements this issue is often solved by calculating a number of effective traits from the correlations between them (Wagner & Zhang 2011). However, traits are not defined statistically and a large correlation between two biologically meaningful traits only indicates that they cannot vary independently. Further, the correlation patterns between traits in the mutational variation that is captured in **rM** matrices may not be intuitive. An example for this are the **rM** matrices that come from Bégin & Schoen (2006) which (among others) report on the length and width of *C. elegans*. These are traits that one may view as only aspects of a more general "size" and for both the mutator lines, they also correlate very strongly. Yet, in the non-mutagenically treated strain, the correlation between these two traits is much lower. If the mutations that occurred in both strains are representative, this suggests that naturally occurring mutations have tended to more frequently affect only one of these two traits. Further still, one can also measure more abstract traits for which there may not be an intuitive expectation, such as for e.g., the probabilities for the transition rates in locomotion behavior (Mallard et al. 2023) and still find that there is considerable trait integration.

Yet, despite the large range of $SD_{rel}(\lambda)$ that was achieved in only eight **M** matrices that were compiled without consideration of representativeness, their diversity of distribution shapes that have been realized by the eigenvalues of the empirical **rM** is very small and their distributions only vary by how far they deviate from a diagonal line by a large first eigenvalue. In comparison, the eigenvalue distributions of the constructed **B** matrices show that in principle, a large diversity of patterns that also covers the whole range of possible $SD_{rel}(\lambda)$ is possible. That the distribution shapes of the empirical **rM** matrices are so similar therefore suggests that the extreme cases of the constructed **B** really are not realistic. However, while the stochastic matrices that have been produced as combinations of a pleiotropic structure $\mathbf{B_s}$ and a type of stochasticity $\boldsymbol{\mu()}$ have succeeded in producing eigenvalue distributions that have shapes that are more similar, they all produce a relatively narrow range of small

$SD_{rel}(\lambda)$. This means that this approach does not identify a particular constructed combination of structure and variation that can be considered more realistic than others.

When considering the eigenvalue distributions, the "hourglass-like" distributions that have been realized by the combination of the type-1a stochasticity and "two modules" $\mathbf{B_s}$ are arguably the most interesting, as they show a large diversity of eigenvalue distributions. Yet, it does not really capture the large first eigenvalue as it is found in the distributions of the more strongly generally integrated **M** matrices. In this respect, the eigenvalues decomposed from the correlation matrix of the "type-2 pleiotropic" **B** gets the closest because it similarly to the empirical ones has a large, shared factor (i.e., a large first eigenvalue). Yet, this shared factor is expected due to the tendency of effects in our matrices to be positive, which is likely not general for most real **M**. This means that there is no unambiguous correspondence between empirical patterns and particular structures, as represented by the stochastic matrices, given the amount of variation. We must therefore conclude that none of the covered $\boldsymbol{\mu(B_s)}$ is particularly similar to the observed mutational pleiotropy, at least not for the common cases of strong overall trait integration. This may either be because real mutational pleiotropy is generally more ordered than can be captured with $\mathbf{B_s}$ or because the a variational aspect $\boldsymbol{\mu()}$ overwhelms the structural patterns in $\mathbf{B_s}$ in the specific combinations that have been used. If the latter is the case, the issue is not with using stochastic **B** matrices in itself and it may still be possible to find combinations that are more suitable models for realistic trait integration. For such an attempt I suggest that one may construct $\mathbf{B_s}$ matrices that more directly incorporate the findings of Wang et al. (2010) that the loci that have the largest SD also affect the most traits. This is something that the "different SDs" $\mathbf{B_s}$ was not able to do because it would have required a larger number of loci which I wanted to keep constant between the different $\mathbf{B_s}$.

Backwards mapping from the pattern to the process

The vast majority of all combinations $\boldsymbol{\mu(B_s)}$ of stochastic **B** matrices lead to very similar eigen-patterns that fall in the category of "random trait integration" because they do not differ from the patterns achieved with the randomly drawn $\mathbf{B_s}$. Further, all combinations (except for the one particularly artificial case of full modularity leading to no correlations) have resulted in eigenvectors that point in all directions almost equally. With the further exception of the one "two distinct modules" case, they also do so in a way that is indistinguishable from random orthogonal vectors. In practice, this means that the variational aspect that is imposed on the pleiotropic structure by the used $\boldsymbol{\mu()}$ is likely too strong, erasing the structural information given by the $\mathbf{B_s}$. Therefore, regarding the observed eigen-directions, every constellation is equally obtainable with every structure $\mathbf{B_s}$ which thus contains practically no

information on how stochastic matrices **μ(B_s)** may be realized. Even though the processes are very different, the patterns are at least on average identical. This shows that the proposed goal of using eigen-directions for backwards identifying the processual structure (**B**) from the pattern (**rM**), cannot be achieved with an imposed variational aspect that is so strong.

The other attempt to establish such a connection was made by counting the frequencies of specific topologies in the hierarchical clusters that were built from the correlation matrices **rM**. Again, this did result in an explicit model that connects some measurement of **rM** pattern, and $\mathbf{B_s}$. This approach also did not reveal a clear association, but it offers additional insights into the problem. All but two of the combinations (across the 200 different stochastic realizations) can result in all 15 possible topologies. This means that also with respect to this read-out, the underlying structure $\mathbf{B_s}$ did not affect the outcome much and that every topology was accessed almost regardless of the pleiotropic constraints encoded in the $\mathbf{B_s}$. Only when constraints are very explicit, such as in the case of type-1a stochasticity in combination with a very modular $\mathbf{B_s}$, was the space of topologies constrained. Within the chosen parameters of this model, we thus did not detect constraints on possible correlation patterns, due to the structure of the G-P map.

However, even though every topology was accessible for almost every pleiotropic structure, they are not accessed with the same frequencies. Firstly, we see that, if the structure $\mathbf{B_s}$ is itself random or the realizations of **μ(B_s)** are close to random, not every topology is equally likely. Instead, the topologies with a low level of hierarchy are more probable. This means in turn, that many levels of nested hierarchy embedded in **rM** are likely associated with a nonrandom underlying pleiotropic structure. Secondly, we do see that the topology that is suggested from the pleiotropy in $\mathbf{B_s}$ is achieved more frequently than alternative topologies. This suggests that even strong used variation does not remove all the information from $\mathbf{B_s}$, and that some general tendency is shared between stochastic **B** matrices based on the same $\mathbf{B_s}$. This approach thus appears more sensitive than the previous for finding structure of the G-P map reflected in the pattern of trait correlations. I suggest that future attempts to predict a most likely **B** matrix for an empirically measured **rM** matrix may be most successful if focusing on hierarchical clustering. This thesis only focused on the topologies rather than the branch lengths, yet perhaps including them could show differences between different $\mathbf{B_s}$ that are meaningful enough that describing posterior probabilities for specific outcomes could be done. However, such an attempt should also be paired with at first finding more realistic stochastic **B** matrices that can serve as models.

What the probability of hidden pleiotropy tells us about the pleiotropic association of traits

The close consideration of the constructed **B** matrices with entries (+1, -1) has shown that different combinations of loci effects on a trait differently constrain further expansions of hidden pleiotropic matrices. This observation has not been tested for continuous effects but is presumably still true for real numbers. Seen more generally, this phenomenon is a consequence of the non-independence of pairwise trait correlations, particularly in cases where correlations depend on a finite, or even small number of loci. For example, if a trait A is highly correlated with B, and B is correlated with C, then – assuming that there is only a limited number of factors that cause these correlations – there is a constrained space of correlations between A and C. This is the same for **M** which does not only have limited loci, but also the constraint of full yet hidden pleiotropy. The details of how the correlations between traits cancel out determines how many of all possible correlations are 0.

This means that calculating the exact probability of randomly drawing a hidden pleiotropic **B** matrix is much more complicated than the approximative approach that was laid out and might require checking the compatibility of each possible pair of rows. Yet, as the approximative approach works reasonably well, it is questionable that doing so would lead to much insight. Further, applying this to the more realistic stochastic **B** matrices would require checking the compatibilities of stochastic rows, which would be far more computationally extensive still. However, this has meaningful implications outside of the very artificial context of calculating purely theoretical probabilities of achieving hidden pleiotropy in matrices that "grow" row-by-row. In practice, "growing a **B** matrix by adding a row" is equivalent to a decision on which traits to measure before a (mutation accumulation) experiment is conducted. Further, the term "hidden pleiotropy" just describes features of the G-P map (**B**) which lead to a lack of correlations in the phenotypic variation (**rM**). From knowing that different configuration of how different loci affect a specific trait can differentially affect hidden pleiotropy in a set of traits, we may reasonably assume that they also differentially affect how correlated sets of traits are as a whole. This means that how exactly two traits are pleiotropically associated with one another (on the level of **B**) beyond just how they correlate (**rM**) determines the chance of how correlated both of them would be with an additional third trait. We can thus conclude that the mechanistic relationship between two traits affects their correlative relationship with a third trait.

How probable is it that a lack of correlation is due to hidden pleiotropy

The analysis of the constructed **B** matrices with entries (+1, -1), as well as the numerical simulations of stochastic **B** matrices show very similar general results: For larger numbers of traits, hidden pleiotropy quickly becomes very improbable. While the total of possible matrices increases dramatically with the number of rows, the number of compatible configurations for each additional trait can only be a subset of the previous. Even though the used formulas are only estimations, the probability of achieving full hidden pleiotropy (with practically no correlations) can still be envisioned as a product of decreasingly small fractions. Achieving hidden pleiotropy by randomly drawing a matrix is already very unlikely for smaller dimensionalities, yet for larger ones the probabilities become absurdly small. Therefore, if a lack of correlations is encountered in empirical experiments, it should not be assumed that it is caused by hidden pleiotropy. Instead, it is far more likely that there is a lack of pleiotropy in the underlying G-P map.

However, the analysis on constructed **B** matrices and the numerical simulations of stochastic **B** matrices differ regarding the effect of the number of loci. For the first, the probability for a randomly drawn **B** to be hidden pleiotropic decreases exponentially with the number of loci. In contrast, in the numerical simulations the probability increases, and this effect is relatively stronger the larger the number of traits is (which itself decrease the probability). It is therefore likely that this effect is caused by the law of large numbers. One may thus argue that hidden pleiotropy could be likely if the dimensionality is just big enough. If the trend continuous and becomes more pronounced, a large number of (potentially all the) loci might – as it may be suggested by for example the omnigenic model – result in a decently large probability for achieving hidden pleiotropy. However, it appears very unlikely that this is the case. The observed increase is only a relatively minor effect, whereas an increase in the number of traits leads to a stark steep decline in probability. There is no reason to assume that the small positive effect of the number of loci would scale in such a way that it could counteract the much larger negative effect of the number of traits. Therefore, all observed results continue to point to hidden pleiotropy being very unlikely.

This is also supported by the finding that it is more likely to achieve a lack of correlations by imposing stochastic variation on an underlying genotype-phenotype structure that is hidden-pleiotropic, than when starting from a random structure. This means that it is easier to maintain a lack of correlations than to generate it from random correlations. However, as considering the two different probability distributions has shown, if a lack of correlations arises from a random underlying structure, it is more likely that this occurs due to pleiotropic effects approaching zero and thus the lack of pleiotropy in the

G-P map. If we envision a continuous selection regime for a lack of correlations, it is unlikely that the G-P map would evolve from no pleiotropy to hidden pleiotropy. Thus, from considering how a lack of correlations could evolve from a random structure, we can also conclude that a lack of pleiotropy in the G-P map is a much more probable explanation for a lack of trait correlations than hidden pleiotropy. Although maintaining only single pairwise correlations that are close to 0 despite pleiotropy is of course more likely than maintaining no correlations for a whole set of traits, as has been addressed here.

Acknowledgements

I am extremely grateful to Mihaela Pavlicev for her mentorship, guidance and invaluable support throughout this project, as well as for giving me the opportunity to do it at all. Further, I would like to express my deepest gratitude to Günter Wagner for relentless support and scientific input which contributed to shaping this project. I am grateful to everyone in the Pavlicev group, as well as in the whole Department of Evolutionary Biology for their continued support and feedback. Finally, I would like to thank Severin Swoboda for his feedback on my mathematical argumentation and Julia Mayer and Mario Zugec for proofreading.

References

- Baatz, M. & Wagner, G.P. 1997. Adaptive Inertia Caused by Hidden Pleiotropic Effects. Theoretical Population Biology 51:49-66. doi:10.1111/j.1558-5646.2008.00472.x.
- Bégin, M. & Schoen, D.J. 2006. Transposable elements, mutational correlations, and population divergences in *Caenorhabditis elegans*. Evolution 61(5):1062-1070. doi:10.1111/j.1558-5646.2007.00097.x.
- Boyle, E.A., Li, Y.I. & Pritchard, J.K. 2017. An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell 169(7):1177-1186. doi:10.1016/j.cell.2017.05.038.
- Dugand, R.J., Aguirre, J.D., Hine, E., Blows, M.W. & McGuigan, K. 2021. The contribution of mutation and selection to multivariate quantitative genetic variance in an outbred population of *Drosophila serrata*. PNAS 118(31):e2026217118. doi:10.1073/pnas.2026217118.
- Estes, S. & Phillips. P.C. 2006. Variation in Pleiotropy and the mutational underpinnings of the **G**-Matrix. Evolution 60(12):2655-2660. doi:10.1111/j.0014-3820.2006.tb01897.x.
- Fisher, R.A. 1930. The genetical theory of natural selection. 2nd edition. New York: Dover.

- Felsenstein, J. 2004. Inferring Phylogenies. © Sinauer Associates Sunderland, MA, USA.
- Fernández, J. & López-Fanjul, C. 1996. Spontaneous Mutational Variances and Covariances for Fitness-Related Traits in *Drosophila melanogaster*. Genetics 143(2):829-37. doi:10.1093/genetics/143.2.829.
- Halligan, D.L. & Keigthley, P.D. 2009. Spontaneous Mutation Accumulation Studies in Evolutionary Genetics. Annual Review of Ecology, Evolution, and Systematics 40:151-172. doi:10.1146/annurev.ecolsys.39.110707.173437.
- Hansen, T.F., Solvin, T.M. & Pavlicev, M. 2019. Predicting evolutionary potential: A numerical test of evolvability measures. Evolution 73(4):689-703. doi:10.1111/evo.13705.
- Hummer, P. Unpublished. Quantitative-genetic simulations suggest that hidden pleiotropy affects constrained evolvability in multivariate traits. Rotation report 2022.
- Kearney, M.R., Jusup, M., McGeoch, M.A., Kooijman, S.A.L.M. & Chown, S.L. 2021. Where do functional traits come from? The role of theory and models. Functional Ecology 35:1385-1396. doi:10.1111/1365-2435.13829.
- Lande, R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. Evolution 33(1): 402-416. doi:10.2307/2407630.
- Lande, R. 1980. The Genetic Covariance between Characters Maintained by Pleiotropic Mutations. Genetics 94(1)203-215. doi:10.1093/genetics/94.1.203.
- Mallard, F., Noble, L., Baer, C.F. & Teotónio, H. 2023. Variation in mutational (co)variances. G3 13(2). doi:10.1093/g3journal/jkac335.
- Murtagh, F. & Legendre, P. 2014. Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? Journal of Classification 31:274–295. doi:10.1007/s00357-014-9161-z.
- Nielsen, F. 2016. Introduction to HPC with MPI for Data Science. Undergraduate Topics in Computer Science. Chapter 8: Hierarchical Clustering. pp. 195-211 © Springer International Publishing Switzerland. doi:10.1007/978-3-319-21903-5.
- Olsen, G. 1990. Gary Olsen's Interpretation of the "Newick's 8:45" Tree Format Standard. Link: https://phylipweb.github.io/phylip/newick_doc.html.
- Orr, A. 2000. Adaptation and the Cost of Complexity. Evolution 54(1):13-20. doi:10.1111/j.0014-3820.2000.tb00002.x.
- Pavlicev, M., Cheverud, J.M. & Wagner, G.P. 2009. Measuring Morphological Integration Using Eigenvalue Variance. Evolutionary Biology 36:157-170. doi:10.1007/s11692-008-9042-7.
- Pavlicev, M. & Hansen, T.F. 2011. Genotype-Phenotype Maps Maximizing Evolvability: Modularity Revisited. Evolutionary Biology 38:371–389. doi:10.1007/s11692-011-9136-5.

- Pavlicev, M. & Wagner, G.P. 2012. Coming to grips with Evolvability. Evolution: Education and Outreach 5: 231-244. doi:10.1007/s12052-012-0430-1.
- R Core Team 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/. [Version 4.2.2 (2022-10-31 ucrt)].
- Shaw, R.G., Byers, D.L. & Darmo, E. 2000. Spontaneous Mutational Effects on Reproductive Traits of *Arabidopsis thaliana*. Genetics 155:369-378. doi:10.1093/genetics/155.1.369.
- Strang, G. 2003. Introduction to Linear Algebra, 3rd Edition. © Wellesley-Cambridge Press, MA, USA.
- Turelli, M. 1985. Effects of pleiotropy on predictions concerning mutation-selection balance for polygenic traits. Genetics 111: 165-195. doi:10.1093/genetics/111.1.165.
- Wagner, G.P. 1984. On the eigenvalue distribution of genetic and phenotypic dispersion matrices: Evidence for a nonrandom organization of quantitative character variation. Journal of Mathematical Biology 21:77-95. doi:10.1007/BF00275224.
- Wagner, G.P. 1989. Multivariate Mutation-Selection Balance With Constrained Pleiotropic Effects. Genetic 122:223-234. doi:10.1093/genetics/122.1.223.
- Wagner, G.P. & Altenberg, L. 1996. Complex adaptations and the evolution of evolvability. Evolution 50(3):967-976. doi:10.1111/j.1558-5646.1996.tb02339.x.
- Wagner, G.P., Kenney-Hunt, J.P., Pavlicev, M., Peck, J.R., Waxman, D. & Cheverud, J.M. 2008. Pleiotropic scaling of gene effects and the 'cost of complexity'. Nature 452:470-472. doi:10.1038/nature06756.
- Wagner, G.P. & Zhang, J. 2011. The pleiotropic structure of the genotype-phenotype map: The evolvability of complex organisms. Nature Reviews Genetics 12(3):204-213. doi:10.1038/nrg2949.
- Wang, Z., Liao, B. & Zhang, J. 2010. Genomic patterns of pleiotropy and the evolution of complexity. National Academy of Sciences of the United States of America 107(42):18034-18039. doi:10.1073/pnas.1004666107.
- Ward, J.H. Jr. 1963. Hierarchical Grouping to Optimize an Objective Function. Journal of the American Statistical Association 58:236–244. doi:10.1080/01621459.1963.10500845.
- Welch, J.J. & Waxman, D. 2003. Modularity and the Cost of Complexity. Evolution 57(8):1723-1734. doi:10.1111/j.0014-3820.2003.tb00581.x.

Appendix

Finding hidden pleiotropic 3X2 **B** matrices with entries {1, -1}

In the attempt to expand a 2X2 B matrix by adding a row that is compatible with both preexisting rows of a hidden-pleiotropic 2X2 **B** matrix, we can see that every possible new row would destroy the hidden pleiotropy. This can be explained with the use of the 16 possible 2X2 submatrices: A 3X2 matrix has three possible 2X2 submatrices. In the case of attempting to expand a hp 2X2 matrix with a new 1X2 row, one of these submatrices is equal to the original matrix and therefore hp. The other two (the new ones) would also have to be hp independently of each other. Of the two rows in the original matrix, we know that one of them has to contain +1 and -1, whereas the other has to have two entries that are the same. The submatrix which contains from the original matrix the row that has +1 and -1 in either order is thus hp if the new row contains two entries with the same sign. However, because the entries of the other row of the original matrix will also have to share a sign, the last submatrix can only be hp if the new row contain entries with different signs. To form hidden-pleiotropic submatrices with both original rows, a row would thus have to fulfil two mutually exclusive conditions.