

University of Groningen

Reconstruction of Networks with Direct and Indirect Genetic Effects

Kruijer, Willem; Behrouzi, Pariya; Bustos-Korts, Daniela; Rodríguez-Álvarez, María Xosé; Mahmoudi, Seyed Mahdi; Yandell, Brian; Wit, Ernst; van Eeuwijk, Fred A

Published in:
 Genetics

DOI:
[10.1534/genetics.119.302949](https://doi.org/10.1534/genetics.119.302949)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Final author's version (accepted by publisher, after peer review)

Publication date:
2020

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Kruijer, W., Behrouzi, P., Bustos-Korts, D., Rodríguez-Álvarez, M. X., Mahmoudi, S. M., Yandell, B., Wit, E., & van Eeuwijk, F. A. (2020). Reconstruction of Networks with Direct and Indirect Genetic Effects. *Genetics*, 214(4), 781-807. <https://doi.org/10.1534/genetics.119.302949>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

1 January 4, 2019

2 **Reconstruction of networks with direct and indirect** 3 **genetic effects**

4 Willem Kruijer^{1,*}, Pariya Behrouzi¹, Daniela Bustos Korts¹, Maria Xose
 5 Rodriguez-Alvarez^{2,3}, Seyed Mahdi Mahmoudi^{4,†}, Brian Yandell⁵, Ernst Wit^{4,††}, Fred
 6 van Eeuwijk¹

7 **1 Biometris, Wageningen University and Research (Wageningen,**
 8 **Netherlands)**

9 **2 Basque Center for Applied Mathematics (Bilbao, Spain)**

10 **3 IKERBASQUE, Basque Foundation for Science (Bilbao, Spain)**

11 **4 Johann Bernoulli Institute, University of Groningen (Groningen,**
 12 **Netherlands)**

13 **5 University of Wisconsin-Madison (Madison, Wisconsin 53706-1510, USA)**

14 *** Corresponding author (E-mail: willem.kruijer@wur.nl)**

15 **† Current address: Faculty of Mathematics, Statistics and Computer**
 16 **Science, Semnan University (Semnan, Iran)**

17 **†† Current address: University of Lugano (Lugano, Switzerland)**

18 **Abstract:** Genetic effects and functional relationships between traits can be repre-
 19 sented by directed graphs, as proposed by Wright in 1921. Nowadays such graphs are
 20 often estimated from empirical data, typically by applying causal inference methods to
 21 multi-trait observations and a small number of QTLs. When however individual QTLs
 22 explain little genetic variance, much of the genetic signal will be missed. To overcome this
 23 limitation, Gianola and Sorensen (2004) defined structural equation models with random
 24 genetic effects. Current causal inference methods for these models treat the genetic effects
 25 as nuisance terms, that need to be eliminated by taking residuals from an unstructured
 26 multi-trait mixed model (MTM). Fitting such MTM for large numbers of traits is however
 27 computationally and statistically challenging.

28 Here we propose an alternative strategy, where genetic effects are formally included
 29 in the graph. Using theoretical results, simulations and real data we show that this has
 30 several advantages: (1) the extended graph satisfies the global Markov property (2) genetic
 31 effects can be directly incorporated in algorithms like PC, allowing for many more traits
 32 (3) we can distinguish direct and indirect effects, and use more of the causal information
 33 contained in the data. For example, we can, under certain assumptions, recover the
 34 structure $G \rightarrow Y_1 \rightarrow Y_2$ if the genetic variance of Y_2 given Y_1 is found to be zero. Finally,
 35 we show that this can be achieved with much higher power if individual plant or plot data
 36 are used, extending the results of Kruijer et al (2015) for single trait analyses. We have
 37 implemented the method in the R-package pegen, publicly available from CRAN.

1 Introduction

Structural equation models (SEM) have been proposed almost a century ago by Wright ([1]), and have been frequently used to describe causal relationships between phenotypes (for a review see [2]). SEM are called structural or functional because each variable is explicitly modeled as a function of the other variables and a noise term, and have a causal interpretation. The advantages compared to regression models are that (i) one can predict the behavior of the system when one or more of the structural equations are modified by some kind of intervention (ii) one can distinguish direct and indirect effects of one variable on another.

In some cases a specific structural model may be formulated using prior biological knowledge (see e.g. [3]), but there is an increasing number of applications (especially in system genetics) where no causal model can be specified in advance. In such cases causal inference methods ([4], [5], [6]) can be used to propose models. These methods are not a substitute for randomized experiments, but rather propose causal models that are most compatible with the observed data, which can be highly useful when having to prioritize future experiments. In genomics for example, the effect of gene-knockouts in yeast was better predicted with causal inference methods than using penalized regression [7]. Statistically, this is because of the manipulation [5] or truncated factorization theorem [4] which describes the distribution after an intervention on one or several variables, which regression methods cannot do.

Causal inference methods have also been applied to genetic data, where observations come from different genotypes. An important question is then how genotypic differences should be accounted for in the model. A popular strategy is to perform causal inference on the traits and all available markers, or QTLs found by mapping ([8], [9], [10]). When however part of the genetic variance is not explained by QTLs, genetic differences may have little added value to the reconstruction of the network. Moreover, typical model assumptions such as independence of residual errors may be violated.

An important class of models that may overcome this limitation was introduced in [11], who defined structural equation models containing random genetic effects. Causal inference for these models is however challenging, due to the correlations of the genetic effects across traits and individuals. To deal with these correlations, [12] and [13] proposed to perform causal inference after subtracting genomic predictions obtained from a multi-trait model (MTM). Similarly, [14] applied the PC-algorithm to the residuals of multi-SNP models. The difficulty with these approaches is that the MTM is limited to small numbers of traits, and that the existence of direct genetic effects cannot be tested. For example, if $Y_1 \rightarrow Y_2 \leftarrow Y_3$, with direct genetic effects on Y_1 and Y_3 , the absence of direct genetic effects on Y_2 cannot be inferred from MTM residuals. Inspired by these problems we define a class of causal graphs in which direct genetic effects are part of the graph, and a single node G represents all direct genetic effects. For each trait Y_j an arrow $G \rightarrow Y_j$ is present if and only if the direct genetic effect G_j is nonzero, i.e. has positive variance.

78 See Figure 2 below for an example.

79 Although this idea is not new, ([15], [16], [17]), this work is, to the best of our
 80 knowledge, the first that formalizes it. In particular, we show that the Markov property
 81 holds for the graph extended with genetic effects (Theorem 1), and based on this develop
 82 the pcgen algorithm. pcgen stands for PC with genetic effects, and is an adaptation of the
 83 general PC-algorithm [5] (named after its inventors Peter Spirtes and Clark Glymour).
 84 Briefly, pcgen assesses the existence of a direct genetic effect on a given trait by testing
 85 whether its genetic variance is zero, conditional on various sets of other traits. For the
 86 existence of an edge between traits Y_1 and Y_2 we test whether in a bivariate MTM the
 87 residual covariance between Y_1 and Y_2 is zero, again conditional on sets of other traits.
 88 Alternatively, this test may be based on MTM residuals (as in existing approaches, who
 89 however did not test the existence of direct genetic effects). Under the usual assumptions
 90 of independent errors, recursiveness and faithfulness, we show that pcgen can recover
 91 the underlying partially directed graph (Theorem 2); this result holds for general genetic
 92 relatedness, and regardless of the correlations between the direct genetic effects.

93 Successful network reconstruction with pcgen requires sufficient power, in the test
 94 for direct genetic effects ($G \rightarrow Y_j$) as well as for the trait to trait relations ($Y_j \rightarrow Y_k$).
 95 Given a plant or other immortal population with observations on genetically identical
 96 replicates, this power is likely to be highest when the original observations are used,
 97 instead of genotypic means and a marker based genetic relatedness matrix (see [18], for the
 98 estimation of genetic variance). Because of this, we focus on experiments with replicates,
 99 although the pcgen algorithm and most of our results are generally applicable, to any
 100 species and relatedness matrix. Moreover, the use of the models developed here is not
 101 limited to pcgen, and in the discussion we will provide directions for further applications.
 102 Because fitting the MTM for all traits simultaneously is no longer necessary, pcgen can
 103 handle a considerably larger number of traits.

104 **Overview of the paper**

105 ... Appendix D contains an overview of the notation.

106 **2 Materials and methods**

107 **2.1 Genetic Structural Equation Models**

108 **2.1.1 Structural Equation Models**

109 To introduce structural models, let us first consider a simple linear structural equation
 110 model (SEM) without genetic effects:

$$\mathbf{y}_i = \mathbf{x}_i B + \mathbf{y}_i \Lambda + \mathbf{e}_i, \quad (1)$$

111 where \mathbf{y}_i is a $1 \times p$ vector of phenotypic values for p traits measured on the i th individual,
 112 \mathbf{e}_i a vector of random errors, and Λ is a $p \times p$ matrix of structural coefficients. The $q \times p$

113 matrix $B = [\beta^{(1)} \dots \beta^{(p)}]$ contains intercepts and trait specific fixed effects of (exogenous)
 114 covariates, whose values are contained in the $1 \times q$ vector \mathbf{x}_i .

115 Defining $n \times p$ matrices $\mathbf{Y} = [\mathbf{Y}_1 \dots \mathbf{Y}_p]$ with rows \mathbf{y}_i and $\mathbf{E} = [\mathbf{E}_1 \dots \mathbf{E}_p]$ with rows
 116 \mathbf{e}_i , and a $n \times q$ design matrix \mathbf{X} , we can write

$$\mathbf{Y} = \mathbf{X}B + \mathbf{Y}\Lambda + \mathbf{E}. \quad (2)$$

117 Λ has zeros on the diagonal and defines a directed graph \mathcal{G}_y over the traits Y_1, \dots, Y_p ,
 118 containing the edge $Y_j \rightarrow Y_k$ if and only if the (j, k) th entry of Λ is nonzero¹. The
 119 columns in (2) correspond to p linear *structural equations*, one for each trait. These are
 120 determined by the *path coefficients*, the nonzero elements in Λ . For example, in Figure
 121 1, if $\mathbf{X} = \mathbf{1}_n$ is the $n \times 1$ vector of ones and $B = [\mu_1 \mu_2 \mu_3]$, the third trait has values
 122 $\mathbf{Y}_3 = \mu_3 \mathbf{1}_n + \lambda_{13} \mathbf{Y}_1 + 2\lambda_{23} \mathbf{Y}_2 + \mathbf{E}_3$. The equality sign here can be understood as an
 123 assignment, i.e. \mathbf{Y}_3 is determined by the values of \mathbf{Y}_1 and \mathbf{Y}_2 (its *parents* in the graph
 124 \mathcal{G}_y) and an error. If the directed graph does not contain any cycle (i.e. a directed path
 125 from a trait to itself), it is a *DAG* (directed acyclic graph), and the SEM is said to be
 126 *recursive*. In the notation we will distinguish the nodes Y_1, \dots, Y_p in the graph \mathcal{G}_y (normal
 127 type), and the random vectors $\mathbf{Y}_1, \dots, \mathbf{Y}_p$ that these nodes represent (bold face).

128 Assuming that the error vectors \mathbf{e}_i are independent and follow a multivariate normal
 129 $N(0, \Sigma_E)$ distribution, it follows from (1) that

$$\begin{aligned} \mathbf{y}_i &= \mathbf{x}_i B (I - \Lambda)^{-1} + \mathbf{e}_i (I - \Lambda)^{-1} = \mathbf{x}_i B \Gamma + \mathbf{e}_i \Gamma \\ &\sim N(\mathbf{x}_i B \Gamma, \Gamma^t \Sigma_E \Gamma) = N(\mathbf{x}_i B \Gamma, \Sigma), \end{aligned} \quad (3)$$

130 i.e., the covariance of the \mathbf{y}_i 's is determined by Σ_E and $\Gamma = (I - \Lambda)^{-1}$.² Given sufficiently
 131 strong assumptions on Λ and Σ_E , their non-zero elements are identifiable. In Figure 1 for
 132 instance, assuming Σ_E is diagonal, it is possible to estimate the 8 parameters (λ_{12} , λ_{13} ,
 133 λ_{23} and λ_{34} , and the variances of the 4 error variables) based on the sample covariance
 134 matrix $\hat{\Sigma}$ (which has 4 diagonal and 6 unique off-diagonal elements). For a more general
 135 discussion, see [19], [20] and [21].

136 An important property of SEM is that the effects of *interventions* can be predicted,
 137 which are changes in one or more of the structural equations. For example, suppose that
 138 in Figure 1, Y_1 , Y_2 and Y_3 are the expression levels of 3 genes, and Y_4 is plant height.
 139 Then after forcing Y_2 to be zero (e.g. by knocking out the gene), the total effect of Y_1 on
 140 Y_4 changes from $(\lambda_{13}\lambda_{34} + \lambda_{12}\lambda_{23}\lambda_{34})$ to $\lambda_{13}\lambda_{34}$. More generally, the new joint distribution
 141 of $\mathbf{Y}_1, \dots, \mathbf{Y}_p$ after intervention can be obtained from the manipulation or truncated
 142 factorization theorem ([4]), *without* needing observations from the new distribution. [16]
 143 discussed the consequences of interventions for genomic selection.

¹We put this path coefficient in the (j, k) th entry (instead of the (k, j) th), and in (2) we post-multiply with Λ . This has the advantage that the data-matrix \mathbf{Y} has the usual dimensions $n \times p$, and that the covariance of $\text{vec}(\mathbf{Y})$ has terms of the form $(\Gamma^t \Sigma_G \Gamma) \otimes K$, instead of $K \otimes (\Gamma^t \Sigma_G \Gamma)$; see e.g. equation (8)

²since we *post*-multiply \mathbf{e}_i with $\Gamma = (I - \Lambda)^{-1}$, the covariance is $\Gamma^t \Sigma_E \Gamma$ and not $\Gamma \Sigma_E \Gamma^t$

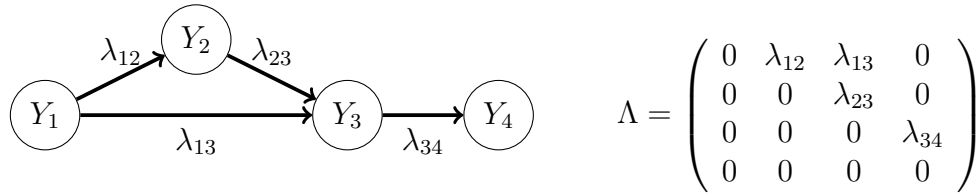


Figure 1. An example of a SEM.

144 2.1.2 GSEM: Structural Equation Models with genetic effects

145 [11] extended model (1) with random genetic effects \mathbf{g}_i : for individuals $i = 1, \dots, n$, it is
 146 then assumed that

$$\mathbf{y}_i = \mathbf{x}_i B + \mathbf{y}_i \Lambda + \mathbf{g}_i + \mathbf{e}_i, \quad (4)$$

147 where again \mathbf{y}_i is a $1 \times p$ vector of phenotypes, Λ contains the structural coefficients, and
 148 $\mathbf{e}_i \sim N(0, \Sigma_E)$ are vectors of random errors. We will refer to model (4) as a linear GSEM
 149 (genetic structural equation model), or simply GSEM. While the genetic effects introduce
 150 relatedness between individuals, there is no form of social interaction (as in e.g. [22], [23]).

151 The $1 \times p$ vectors \mathbf{g}_i contain the direct genetic effects for individuals $i = 1, \dots, n$.
 152 Each \mathbf{g}_i^t follows a $N(0, \Sigma_G)$ distribution, where Σ_G is a $p \times p$ matrix of genetic variances
 153 and covariances. The vectors \mathbf{g}_i are independent of the \mathbf{e}_i 's, but not independent among
 154 themselves. Defining a $n \times p$ matrix $\mathbf{G} = [\mathbf{G}_1 \cdots \mathbf{G}_p]$ with rows \mathbf{g}_i and columns \mathbf{G}_j
 155 ($j = 1, \dots, p$), we can extend (2) as follows:

$$\mathbf{Y} = [\mathbf{Y}_1 \cdots \mathbf{Y}_p] = \mathbf{X}B + \mathbf{Y}\Lambda + \mathbf{G} + \mathbf{E}. \quad (5)$$

156 Each vector \mathbf{G}_j is the vector of direct genetic effects on the j th trait. We make the
 157 following assumptions about the GSEM defined in (5):

- 158 1. *all traits are measured on each individual*: the rows \mathbf{y}_i of \mathbf{Y} may be either observa-
 159 tions on individual organisms or genotypic means of a number of replicates (plants,
 160 or plots in a field trial), but the original observations should always come from the
 161 same experiment. In addition, the residual errors originate from biological variation.
- 162 2. *recursiveness*: the graph \mathcal{G}_Y defined by Λ is a DAG.
- 163 3. *causal sufficiency*: the covariance matrix Σ_E of the error vectors \mathbf{e}_i is diagonal, i.e.
 164 there are no latent variables. This means that all nonzero (non-genetic) correlations
 165 between traits must be the consequence of causal relations between the traits. In
 166 the discussion we describe how this assumption may be relaxed. We also assume
 167 the diagonal elements of Σ_E to be strictly positive.
- 168 4. *Genetic relatedness among individuals*: \mathbf{G} is independent from \mathbf{E} , and has a matrix-
 169 variate normal distribution with row-covariance K and column covariance Σ_G , where

170 K is a $n \times n$ relatedness matrix, which we describe in more detail in section 2.1.5.
 171 Equivalent with this, the $np \times 1$ vector $vec(\mathbf{G}) = (\mathbf{G}_1^t, \dots, \mathbf{G}_p^t)^t$ is multivariate
 172 normal with covariance $\Sigma_G \otimes K$, where vec denotes the operation of creating a
 173 column vector by stacking the columns of a matrix. Consequently, each \mathbf{G}_j is mul-
 174 tivariate normal with covariance $\sigma_{G,j}^2 K$, where the variances $\sigma_{G,j}^2$ form the diagonal
 175 of Σ_G . Using the same notation, we can write that \mathbf{E} is matrix-variate normal with
 176 row-covariance I_n and column covariance Σ_E , and that $vec(\mathbf{E}) \sim N(0, \Sigma_E \otimes I_n)$.

177 5. *No collinear genetic effects*: the diagonal elements of Σ_G do not need to be strictly
 178 positive, but for all nonzero elements, the corresponding correlation is not 1 or -1 .

179 Assumptions 1-4 were also made in related work on structural models with random genetic
 180 effects ([12], [13]), and 1-3 are commonly made for structural models without such effects.
 181 Assumption 1 is implicit in model (5) itself, as it is assumed that the structural equations
 182 propagate all errors to traits further down in the graph. Network reconstruction without
 183 this assumption would rely completely on the genetic effects, requiring Σ_G to be diagonal,
 184 which is a rather unrealistic assumption (see the discussion, section 4.1). Assumption 1
 185 does not require traits to be measured at the same time. In particular, it is possible to
 186 include the same trait measured at different timepoints, which of course puts constraints
 187 on the causality. Such constraints can in principle be incorporated in our model, just as
 188 other biological constraints (see e.g. [24]), although we will not explore this here.

189 2.1.3 Graphical representation of GSEM: extending \mathcal{G}_y with genetic effects

190 Contrary to previous work, we will explicitly take into account the possibility that there
 191 are no direct genetic effects on some of the traits. In this case, the corresponding rows
 192 and columns in Σ_G are zero. We let $D \subseteq \{1, \dots, p\}$ denote the index set of the traits
 193 with direct genetic effects, and write $\Sigma_G[D, D]$ for the sub-matrix with rows and columns
 194 restricted to D . From assumption 3 above it follows that $\Sigma_G[D, D]$ is non-singular, i.e.
 195 there can be no perfect correlations between direct genetic effects.

196 We graphically represent model (5) by a graph \mathcal{G} with nodes Y_1, \dots, Y_p and a node G ,
 197 which represent respectively $\mathbf{Y}_1, \dots, \mathbf{Y}_p$ and the matrix $\mathbf{G} = [\mathbf{G}_1 \dots \mathbf{G}_p]$. \mathcal{G} contains an
 198 edge $Y_j \rightarrow Y_k$ if the (j, k) th entry of Λ is nonzero, and an edge $G \rightarrow Y_j$ if \mathbf{G}_j is nonzero
 199 with probability one, i.e., if $\sigma_{G,j}^2 > 0$. See Figure 2 for an example. In words, \mathcal{G} is defined
 200 as the original graph \mathcal{G}_y over the traits (defined by Λ), extended with arrows $G \rightarrow Y_j$
 201 for all traits with a direct genetic effect, i.e. for all $j \in D$. Consequently, our main
 202 objective of reconstructing trait-to-trait relationships and direct genetic effects translates
 203 as reconstructing \mathcal{G} .

204 As for the \mathbf{Y}_j 's, we distinguish the node in the graph G (normal type) and the random
 205 matrix \mathbf{G} it represents (bold face). \mathbf{G} is represented by a *single* node G , instead of
 206 multiple nodes G_1, \dots, G_p . This choice is related to our assumption that K is the same
 207 for all traits; see Appendix G.1 for a motivating example. The orientation of any edge

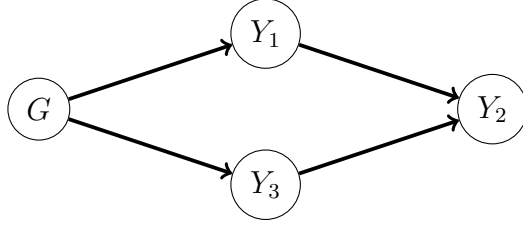


Figure 2. An example of a graph \mathcal{G} representing a genetic structural equation model (GSEM). There is no direct genetic effect on Y_2 , and therefore no edge $G \rightarrow Y_2$.

208 between G and Y_j is restricted to $G \rightarrow Y_j$, because G arises from a randomized treatment,
 209 and also because the opposite orientation would be biologically nonsensical. Because of
 210 our assumption that \mathcal{G}_Y is a DAG, it follows that \mathcal{G} is a DAG as well, as a cycle would
 211 require at least one edge pointing into G .

212 We emphasize that \mathcal{G} is just a mathematical object and not a complete visualization of
 213 all model terms and their distribution, as is common in the SEM-literature. In particular,
 214 \mathcal{G} does not contain nodes for the residual errors, path coefficients, or information about
 215 the off-diagonal elements of Σ_G . While Σ_G is usually not entirely identifiable ([11]), we
 216 will see that \mathcal{G} is identifiable in terms of its skeleton (the undirected graph obtained when
 217 removing the arrowheads) and some of the orientations.

218 2.1.4 Direct and indirect genetic effects

219 Taking the term $\mathbf{y}_i\Lambda$ from the right- to the left-hand side in equation (4), and assuming
 220 that the inverse $\Gamma = (I - \Lambda)^{-1}$ exists, it follows that

$$\mathbf{y}_i = \mathbf{x}_i B\Gamma + \mathbf{g}_i\Gamma + \mathbf{e}_i\Gamma = \mathbf{x}_i B\Gamma + \mathbf{u}_i + \mathbf{e}_i\Gamma \sim N(\mathbf{x}_i(B\Gamma), \Gamma^t \Sigma_G \Gamma + \Gamma^t \Sigma_E \Gamma), \quad (6)$$

221 where $\mathbf{u}_i = \mathbf{g}_i\Gamma$ is the *total* genetic effect. Hence, as pointed out by [16], the genetic
 222 variance of a trait is not only driven by its direct genetic effect (\mathbf{g}_i), but also by direct
 223 genetic effects on traits affecting it, i.e. its parents in the graph \mathcal{G}_Y . The *indirect* genetic
 224 effect is the difference $\mathbf{u}_i - \mathbf{g}_i$.

225 Similarly, we can distinguish the contribution of direct and indirect genetic effects to
 226 the genetic covariance. The (j, k) th element of $\Gamma^t \Sigma_G \Gamma$ in (6) is the *total* genetic covariance
 227 between \mathbf{Y}_j and \mathbf{Y}_k . This is what is usually meant with genetic covariance. When
 228 necessary, we distinguish this from the covariance between the direct genetic effects \mathbf{G}_j
 229 and \mathbf{G}_k , determined by $\Sigma_G[j, k]$. Indeed $\Sigma_G[j, k]$ affects the total genetic covariance, but
 230 the latter is also driven by causal relationships between traits, as defined by $\Gamma = (I - \Lambda)^{-1}$.
 231 Regarding the diagonal of $\Gamma^t \Sigma_G \Gamma$, we note that traits without a direct genetic effect may
 232 still have positive genetic variance.

233 2.1.5 Genetic relatedness

234 The genetic relatedness matrix K introduced in assumption 4 determines the covariance
235 between the rows of \mathbf{G} . We will assume that K is one of the following types:

- 236 • $K = ZZ^t$, Z being the $n \times m$ incidence matrix assigning $n = mr$ plants (or plots)
237 to m genotypes, in a balanced design with r replicates for each genotype. This K is
238 obtained when each genotype has an independent effect, as in the classical estimation
239 of broad-sense heritability. Since no marker-information is included, the model
240 cannot be used for genomic prediction, but we will see that for the reconstruction of
241 \mathcal{G} (using the training genotypes) it has considerable computational and statistical
242 advantages.
- 243 • K is estimated from a dense set of markers, assuming additive infinitesimal effects.
244 To keep the notation simple we will still use $\sigma_{G,j}^2$ for the diagonal elements of Σ_G ,
245 instead of $\sigma_{A,j}^2$. This type of relatedness matrix is used when there is only a single
246 individual per genotype, or when only genotypic means are available.

247 In both cases K has dimension $n \times n$. The balance required in the first case is necessary
248 in Theorems 5 and 6 below, but is not a general requirement for our models, or for the
249 pcgen algorithm.

250 2.1.6 The joint distribution implied by the GSEM

251 The sum $\mathbf{G} + \mathbf{E}$ does in general not have a matrix-variate normal distribution, but from
252 our assumption 4 it still follows that $\text{vec}(\mathbf{G} + \mathbf{E})$ is multivariate normal with covariance
253 $\Sigma_G \otimes K + \Sigma_E \otimes I_n$. We can therefore rewrite equation (5) as

$$\mathbf{Y} = \mathbf{X}B\Gamma + \mathbf{G}\Gamma + \mathbf{E}\Gamma = \mathbf{X}B\Gamma + \mathbf{U} + \mathbf{E}\Gamma, \quad (7)$$

254 and equation (6) generalizes to

$$\text{vec}(\mathbf{Y}) \sim N(\text{vec}(\mathbf{X}(B\Gamma)), (\Gamma^t \Sigma_G \Gamma) \otimes K + (\Gamma^t \Sigma_E \Gamma) \otimes I_n). \quad (8)$$

255 As pointed out in [11], [12] and [13], (8) can be re-written as

$$\text{vec}(\mathbf{Y}) \sim N(\text{vec}(\mathbf{X}\tilde{B}), V_G \otimes K + V_E \otimes I_n), \quad (9)$$

256 where $V_G = \Gamma^t \Sigma_G \Gamma$ and $V_E = \Gamma^t \Sigma_E \Gamma$, and \tilde{B} is the matrix of fixed effects transformed
257 by Γ . This is a common model for multi-trait GWAS and genomic prediction (see among
258 others [15], [25], [26]).

259 Using the results of [5] (p. 371), it turns out that Γ can be written directly in terms of
260 sums of products of path coefficients (see Appendix E.2). Consequently, there is no need

261 to invert $(I - \Lambda)$, although it still holds that $\Gamma = (I - \Lambda)^{-1}$, provided the inverse exists.
 262 Defining γ_j as the j th column of Γ , we can express the j th trait as

$$\mathbf{Y}_j = \mathbf{X}B\gamma_j + \mathbf{G}\gamma_j + \mathbf{E}\gamma_j = \mathbf{X}B\gamma_j + \mathbf{U}_j + \mathbf{E}\gamma_j, \quad (10)$$

263 i.e. equation (7), restricted to the j th column.

264 2.1.7 Causal inference without genetic effects

265 So far we have assumed that \mathcal{G} is known: given sufficient restrictions on Λ , Σ_G and Σ_E ,
 266 it may then be possible to estimate these matrices ([11]). In this work however, we aim
 267 to reconstruct an unknown \mathcal{G} , based on observations from a GSEM of the form (5). We
 268 will do this with the pegen algorithm introduced in section 2.2, but first briefly review
 269 the necessary concepts. Appendix E.1 contains a more detailed introduction.

270 Suppose for the moment we have observations generated by an acyclic SEM without
 271 latent variables, and without genetic effects. From the pioneering work of Judea Pearl
 272 and others in the 1980s it is known that we can recover the skeleton of the DAG and
 273 some of the orientations, i.e. those given by the v -structures. A v -structure is any triple
 274 of nodes Y_j, Y_k, Y_l such that $Y_j \rightarrow Y_k \leftarrow Y_l$, without any edge between Y_j and Y_l . All
 275 DAGs with the same skeleton and v -structures form an *equivalence class*, which can be
 276 represented by a completed partially directed acyclic graph (CPDAG). DAGs from the
 277 same equivalence class cannot be distinguished from observational data, at least not under
 278 the assumptions we make here. For reconstruction of the CPDAG, constraint-based and
 279 score-based methods have been developed (for an overview, see [24]).

280 Here we focus on constraint-based methods, which rely on the equivalence of condi-
 281 tional independence (a property of the distribution) and directed separation (d-separation;
 282 a property of the graph). An important property is that an edge $Y_j - Y_k$ is missing in
 283 the skeleton of the DAG if and only if Y_j and Y_k are d-separated by at least one (possibly
 284 empty) set of nodes Y_S . Such Y_S is called a *separating set* for Y_j and Y_k . Given the
 285 equivalence of d-separation and conditional independence, this means that we can infer
 286 the presence of the edge $Y_j - Y_k$ in the skeleton by testing $\mathbf{Y}_j \perp\!\!\!\perp \mathbf{Y}_k | \mathbf{Y}_S$ for all \mathbf{Y}_S . The
 287 PC- and related algorithms therefore start with a fully connected undirected graph, and
 288 remove the edge $Y_j - Y_k$ whenever \mathbf{Y}_j and \mathbf{Y}_k are found to be conditionally independent
 289 for some \mathbf{Y}_S . While the first constraint-based algorithms such as IC [27] exhaustively
 290 tested all possible subsets, the PC-algorithm ([5]) can often greatly reduce the number
 291 of subsets to be considered.

292 Although structural equations are often assumed to be linear and the noise Gaussian,
 293 these assumptions are not essential for the equivalence of conditional independence and
 294 d-separation. For example, the pcalg package [28] contains a conditional independence
 295 test for binary data. However, structural equations with additional random effects are
 296 rarely considered, and many causal inference algorithms therefore assume independent
 297 observations.

2.1.8 Existing approaches for estimating \mathcal{G}_Y , given genetic effects

To deal with the dependence introduced by the genetic effects, [12] and [13] proposed to predict the total genetic effects (i.e., the terms $\mathbf{u}_i = \mathbf{g}_i\Gamma$ in (6)), and perform causal inference on the residuals. These methods are flexible in the sense that any genomic prediction method can be used, and combined with any causal inference method.

A disadvantage however is that the presence of direct genetic effects cannot be tested. Suppose for example $G \rightarrow Y_1 \rightarrow Y_2 \rightarrow Y_3$, and we subtract the total genetic effects. Then given only the residuals, we can never know if part of the genetic variance of Y_2 was due to a direct effect $G \rightarrow Y_2$. To use the causal information contained in the genetic effects, [13] estimated 'genomic networks', based on the predictions themselves. These however seem to require additional assumptions not required for the residual networks (in particular, diagonal Σ_G), and it seems difficult to relate edges in such a network to direct genetic effects (Appendix G.2). In summary, residual and genomic networks only estimate the subgraph \mathcal{G}_Y of trait to trait relations, instead of the complete graph \mathcal{G} .

Another disadvantage is that the MTM (9) can only be fitted for a handful of traits [25], for statistical as well as computational reasons. For example, [29] showed that for general Gaussian covariance models, (residual) ML-estimation behaves like a convex optimization problem only when $n \gtrsim 14p$. Similar problems are likely to occur for Bayesian approaches. The main problem with fitting the MTM to data from GSEM model (8) is that one cannot exploit the possible sparseness of \mathcal{G} . Even for sparse graphs with few direct genetic effects, the matrices $\Gamma^t \Sigma_G \Gamma$ and $\Gamma^t \Sigma_E \Gamma$ may still be dense, requiring a total of $p(p+1)$ parameters. To overcome these limitations, we consider the presence or absence of direct genetic effects to be part of the causal structure, and develop pcgen, a causal inference approach directly on \mathcal{G} .

2.2 The pcgen algorithm

The main idea behind pcgen is that the PC-algorithm is applicable to any system in which d-separation and conditional independence are equivalent, and where conditional independence can be tested. We first describe the algorithm and propose independence tests; the equivalence is addressed in section 3.4. If we define $\mathbf{Y}_{p+1} := \mathbf{G}$ and temporarily rename the node G as Y_{p+1} , pcgen is essentially the PC-algorithm applied to Y_1, \dots, Y_{p+1} :

1. **skeleton-stage.** Start with the fully connected undirected graph over $\{Y_1, \dots, Y_{p+1}\}$, and an empty list of separation sets. Then test the conditional independence between all pairs \mathbf{Y}_j and \mathbf{Y}_k , given subsets of other variables \mathbf{Y}_S . Whenever a p-value is larger than the pre-specified significance threshold α , update the skeleton by removing the edge $Y_j - Y_k$, and add Y_S to the list of separation sets for Y_j and Y_k . This is done for conditioning sets of increasing size, starting with the empty set ($S = \emptyset$; marginal independence between \mathbf{Y}_j and \mathbf{Y}_k). Only consider S that, in the current skeleton, are adjacent to Y_j or Y_k .

336 **2. orientation-stage.** Apply the orientation rules given in Appendix A (R1-R3 in
 337 Algorithm 1) to the skeleton and separating sets found in the skeleton-stage. For
 338 example, if the skeleton is $Y_1 - Y_2 - Y_3$ and $\{Y_2\}$ is *not* a separating set for Y_1 and
 339 Y_3 , the skeleton is oriented $Y_1 \rightarrow Y_2 \leftarrow Y_3$; otherwise, none of the two edges can be
 340 oriented.

341 To get pcgen, we only need to make a few refinements to these steps. First, in the skeleton
 342 stage we need to specify *how* to test conditional independence statements. Clearly, inde-
 343 pendence between two traits requires a different test than independence between a trait
 344 and \mathbf{G} (i.e., \mathbf{Y}_{p+1}), in particular because the latter is not directly observed. Second, we
 345 need to modify the orientation rules, in order to avoid edges pointing into G . The usual
 346 rules give the correct orientations given perfect conditional independence information,
 347 but statistical errors in the tests may lead to edges of the form $Y_j \rightarrow G$. Third, statis-
 348 tical errors can also make the output of pc(gen) order-dependent. We therefore adopt
 349 the PC-stable algorithm of [30], who proposed to perform all operations in skeleton- and
 350 orientation-stage list-wise (details given in Appendix A).

351 In summary, pcgen is the PC-stable algorithm with: (1) specific conditional indepen-
 352 dence tests (described in sections 2.2.1-2.2.3 below), and (2) modified orientation rules,
 353 in order to avoid edges pointing into G (Appendix A.2). As in the original PC-algorithm,
 354 the number of type-I and type-II errors occurring in the tests is determined by the choice
 355 of the significance threshold α , which is discussed in sections 2.5 and 4.3.

356 2.2.1 Skeleton stage: conditional independence tests

357 Writing again \mathbf{G} for \mathbf{Y}_{p+1} , we can distinguish the following types of conditional indepen-
 358 dence statements in the skeleton stage:

$$359 \mathbf{Y}_j \perp\!\!\!\perp \mathbf{Y}_k | (\mathbf{G}, \mathbf{Y}_S), \quad (\text{A})$$

$$360 \mathbf{Y}_j \perp\!\!\!\perp \mathbf{G} | \mathbf{Y}_S, \quad (\text{B})$$

$$\mathbf{Y}_j \perp\!\!\!\perp \mathbf{Y}_k | \mathbf{Y}_S, \quad (\text{C})$$

361 where $j, k \in \{1, \dots, p\}$ ($j \neq k$) and $S \subseteq \{1, \dots, p\} \setminus \{j, k\}$ (or $S \subseteq \{1, \dots, p\} \setminus \{j\}$
 362 in statement (A)). In words, (B) means that the trait \mathbf{Y}_j is independent of all genetic
 363 effects (\mathbf{G}), conditional on the traits \mathbf{Y}_m ($m \in S$). If S is the empty set, this is under-
 364 stood as marginal independence of \mathbf{Y}_j and \mathbf{G} . Similarly, (A) and (C) express conditional
 365 independence of traits \mathbf{Y}_j and \mathbf{Y}_k given \mathbf{G} and \mathbf{Y}_S , or given \mathbf{Y}_S alone.

366 We now propose statistical tests for statements (A) and (B), which rely on the linearity
 367 of our GSEM (model (5)). Statement (C) can be tested using standard partial correlations
 368 and Fisher's z-transform. However, as we show in appendix F, this test is redundant, since
 369 for any set Y_S that d-separates Y_j and Y_k , the set $Y_S \cup \{G\}$ will also d-separate them. We
 370 therefore skip any test for $\mathbf{Y}_j \perp\!\!\!\perp \mathbf{Y}_k | \mathbf{Y}_S$, and instead test the corresponding statement
 371 with \mathbf{G} , i.e., $\mathbf{Y}_j \perp\!\!\!\perp \mathbf{Y}_k | \mathbf{Y}_S, \mathbf{G}$.

372 **2.2.2 Testing $\mathbf{Y}_j \perp \mathbf{Y}_k | (\mathbf{G}, \mathbf{Y}_S)$**

373 For statement (A) we consider two different tests:

- 374 • The residual covariance (RC) test. The RC-test is based on the bivariate distribution
 375 of $(\mathbf{Y}_j, \mathbf{Y}_k)$ conditional on the observed $\mathbf{Y}_S = \tilde{y}_S$, which can be written as

$$\begin{pmatrix} \mathbf{Y}_j \\ \mathbf{Y}_k \end{pmatrix} | \mathbf{Y}_S = \tilde{y}_S \sim N \left(\begin{pmatrix} \mu_{j|S} \\ \mu_{k|S} \end{pmatrix}, \Sigma_{jk|S} \right). \quad (11)$$

376 Expressions for $\mu_{j|S}$, $\mu_{k|S}$ and $\Sigma_{jk|S}$ can be derived from equation (8), and are given
 377 in appendix E.6. For testing (A) it is assumed that $\mu_{j|S}$ and $\mu_{k|S}$ are of the form

$$\mathbf{X}B\gamma_j + \tilde{y}_S\beta_S^{(j)}, \quad \mathbf{X}B\gamma_k + \tilde{y}_S\beta_S^{(k)}, \quad (12)$$

378 where $\mathbf{X}B\gamma_j$ is the marginal mean of \mathbf{Y}_j (see (10)), and $\beta_S^{(j)}$ and $\beta_S^{(k)}$ are $|S| \times 1$
 379 vectors of regression coefficients. The covariance in (11) is assumed to be of the
 380 form

$$\Sigma_{jk|S} = V_G(jk|S) \otimes K + V_E(jk|S) \otimes I_n, \quad (13)$$

381 for some 2×2 matrices $V_G(jk|S)$ and $V_E(jk|S)$. Given these assumptions, we test
 382 whether the off-diagonal element in $V_E(jk|S)$ is zero, using the likelihood ratio
 383 test (LRT) described in Appendix A.3. In words, the conditional distribution is a
 384 bivariate MTM, in which we test the residual covariance³.

385 The underlying idea is that a nonzero residual covariance must be the consequence
 386 of an edge $Y_j \rightarrow Y_k$ or $Y_k \rightarrow Y_j$, because of the assumed normality and causal
 387 sufficiency. On the other hand, a nonzero *genetic* covariance may also be due to
 388 covariance between direct genetic effects on these variables, or due to a genetic effect
 389 on a common ancestor. The RC-test therefore compares the full bivariate mixed
 390 model with the submodel with diagonal $V_E(jk|S)$, while accounting for all genetic
 391 (co)variance using $V_G(jk|S)$.

- 392 • The RG-test (*Residuals of GBLUP*), which is based on the residuals of MTM (9).
 393 Solving the mixed model equations, we obtain the BLUP \mathbf{U}^* of the total genetic
 394 effects $\mathbf{U} = \mathbf{G}\Gamma$, the BLUE \tilde{B}^* of the fixed effects, and residuals $\mathbf{Y} - \mathbf{U}^* - \mathbf{X}\tilde{B}^*$.
 395 Next, we test statement (A) using the partial correlation between the residuals of
 396 \mathbf{Y}_j and \mathbf{Y}_k , given those of \mathbf{Y}_S , and assess significance with Fisher's z-transform.
 397 This is essentially the test used by [12] and [13], who instead took a fully Bayesian
 398 approach to predict \mathbf{U} . Here we stick to the GBLUP, and consider two variants:
 399 (i) the one described above, based on the multivariate GBLUP, and (ii) based on
 400 univariate GBLUPs, i.e. all predictions \mathbf{U}_j^* are obtained from single trait mixed
 401 models. In both cases, the idea is that when \mathbf{U}^* is close enough to \mathbf{U} , it follows

³alternatively, we could test the residual correlation

402 from equations (7)-(8) that $\text{vec}(\mathbf{Y} - \mathbf{U}^*)$ approximately has covariance $(\Gamma^t \Sigma_E \Gamma) \otimes I_n$,
 403 i.e. that of independent samples, without any genetic relatedness.

404 Both tests rely on approximations: they do not directly test conditional independence
 405 statement (A), but a related statement. In some cases the approximation is exact, and
 406 the related statement equivalent. In other cases it is not entirely equivalent, which may
 407 introduce an additional source of error (on top of the statistical errors that come with
 408 the test itself). In the RG-test the prediction error $(\mathbf{U}^* - \mathbf{U})$ determines the quality of
 409 the approximation. The RC-test relies on assumptions (12)-(13) about the conditional
 410 mean and covariance of $(\mathbf{Y}_j, \mathbf{Y}_k)$ given $\mathbf{Y}_S = \tilde{y}_S$. We discuss the appropriateness of these
 411 approximations in section 3.4.3.

412 Model (13) (with $S = \emptyset$) can also provide an estimate of the total genetic covariance,
 413 i.e., the off-diagonal element of $V_G(jk|S)$. In Appendix A.4 we describe a test for zero
 414 genetic covariance, which can be useful for data exploration, but has no role in pgen.
 415 A similar test for the total genetic correlation can be obtained using the delta method
 416 ([26]).

417 2.2.3 Testing $\mathbf{Y}_j \perp\!\!\!\perp \mathbf{G} | \mathbf{Y}_S$

418 Our test for statement (B) is based on the intuition that \mathbf{Y}_j is independent of $\mathbf{G} =$
 419 $[\mathbf{G}_1 \cdots \mathbf{G}_p]$ given \mathbf{Y}_S , whenever there is no direct genetic effect on \mathbf{Y}_j (i.e. $\mathbf{G}_j = \mathbf{0}$), and
 420 all directed paths from G to Y_j are blocked by the set $\{Y_m : m \in S\}$. In particular, if S
 421 is the empty set, there should not be any directed path from G to Y_j . Because directed
 422 paths from G to Y_j will generally introduce some genetic variance in \mathbf{Y}_j , the idea is to test
 423 whether there is significant genetic variance in the conditional distribution of \mathbf{Y}_j given
 424 $\mathbf{Y}_S = \tilde{y}_S$. This is done as follows:

- 425 • When $K = ZZ^t$, we use the classical F-test in a 1-way ANOVA, with \mathbf{X} and \tilde{y}_S
 426 as covariates. Technically, this is an ANCOVA (analysis of covariance), where the
 427 treatment factor genotype is tested conditional on the covariates being in the model.
- 428 • For other K one can use the LRT, as in the RC-test for (A). The asymptotic
 429 distribution under the null-hypothesis is a mixture of a point mass at zero and a
 430 chi-square.

431 In both cases, it is assumed that the mean of the conditional distribution of \mathbf{Y}_j given
 432 $\mathbf{Y}_S = \tilde{y}_S$ is of the form (12), and the covariance of the form

$$\Sigma_{j|S} = \sigma_G^2(j|S)K + \sigma_E^2(j|S)I_n, \quad (14)$$

433 for some variance components $\sigma_G^2(j|S)$ and $\sigma_E^2(j|S)$. Again, the covariance assumption
 434 holds exactly when $K = ZZ^t$ (Theorem 6); otherwise it is an approximation. As in the
 435 RC-test for statement (A), (12) is an approximation as well, and the assumed linearity

436 is essential. Suppose for example that $\mathbf{Y}_2 := \mathbf{Y}_1^2$ and $\mathbf{Y}_1 := \mathbf{G}_1 \sim N(0, ZZ^t)$, where for
 437 the sake of the argument we assume absence of residual errors. Then the factor genotype
 438 will generally be significant in the ANCOVA with Y_1 as covariate. For example there
 439 could be 2 replicates of 3 genotypes, with genetic effects $(-1, -1, 0, 0, 1, 1)$, then clearly
 440 there is some unexplained genetic variance when regressing $\mathbf{Y}_2 = (1, 1, 0, 0, 1, 1)^t$ on \mathbf{Y}_1 .

441 2.2.4 pcRes: reconstructing only trait-to-trait relationships

442 Testing only conditional independencies of the form (A), one can reconstruct the graph
 443 \mathcal{G}_y of trait-to-trait relations. Moreover, if this is done with the RG-test, the algorithm is
 444 very similar to the residual approaches of [12] and [13]. Staying within the context of the
 445 PC-algorithm and using residuals from GBLUP, we will call this pcRes. Similar to pcgen
 446 with the RG-test, the performance of pcRes strongly depends on the prediction error of
 447 the GBLUP.

448 2.3 Causal inference based on genotypic means

449 In section 2.1.5 we assumed the genetic relatedness to be either $K = ZZ^t$ (given observa-
 450 tions on replicates), or a marker-based relatedness matrix (given a single observation on
 451 each genotype). However, in many applications both a marker-based relatedness matrix
 452 and replicates are available. Suppose we have r replicates of m genotypes in a completely
 453 randomized design, with a $m \times m$ relatedness matrix A .⁴ Reconstruction of \mathcal{G} with pcgen
 454 or \mathcal{G}_y with pcRes is then possible using:

- 455 1. both A and the replicates. The distribution of the data is assumed to be

$$vec(\mathbf{Y}) \sim N(vec(\mathbf{X}(B\Gamma)), (\Gamma^t \Sigma_G \Gamma) \otimes (ZAZ^t) + (\Gamma^t \Sigma_E \Gamma) \otimes I_n), \quad (15)$$

456 i.e. equation (8) with $K = ZAZ^t$.

- 457 2. only the replicates. The true distribution of the data could be as in (15), but
 458 ignoring the information in A , we assume that

$$vec(\mathbf{Y}) \sim N(vec(\mathbf{X}(B\Gamma)), (\Gamma^t \Sigma_G \Gamma) \otimes (ZZ^t) + (\Gamma^t \Sigma_E \Gamma) \otimes I_n). \quad (16)$$

- 459 3. genotypic means and A . Assuming that the original plant or plot data are dis-
 460 tributed as in (15), the distribution of the $m \times p$ matrix $\bar{\mathbf{Y}}$ of genotypic means is
 461 such that

$$vec(\bar{\mathbf{Y}}) \sim N\left(0, (\Gamma^t \Sigma_G \Gamma) \otimes A + \frac{1}{r} (\Gamma^t \Sigma_E \Gamma) \otimes I_m\right), \quad (17)$$

462 where it is assumed that the fixed effects have been accounted for in the estimation
 463 of genotypic means.

⁴We use the letter A to avoid confusion with the general $n \times n$ matrix K , and also because A typically models additive effects.

464 In our simulations (section 3.1 below) we compare these cases for pcRes. A similar
 465 comparison for pcgen is left for future work; here we will focus on the second case, using
 466 only replicates and assuming independent genetic effects. This is motivated by the good
 467 results for pcRes for this case, and by theoretical arguments (see Theorems 5-6 below, and
 468 the discussion (4.2)). However, we stress that in all simulations, the data are simulated
 469 as in (15), i.e. with additive genetic effects, inducing population structure.

470 A second issue in pcRes (independent of the choice of K) are the residuals, which can
 471 be obtained from single- or multi-trait models. In the latter case, we fit the multi-trait
 472 model (9), with K equal to ZAZ^t , ZZ^t or A . In the univariate case, we do the same,
 473 assuming the covariance model $\sigma_G^2 K + \sigma_E^2 I$ (where I is of dimension $n \times n$ or $m \times m$,
 474 depending on K).

475 2.4 Software, and overview of algorithms

476 We now give an overview of pcgen, pcRes, and variations on these methods. Most of these
 477 are available in our R-package pcgen, freely available at <https://cran.r-project.org/web/packages/pcgen/>
 478 pcgen is built on the pcalg package ([28], [31]), in which we modified the orientation rules
 479 and the default conditional independence test. Tables 5 and 6 in Appendix D provide a
 480 brief description, with the abbreviations and the required R-commands.

481 pcgen can be run with either the RC- or the RG-test (pcgen-RC versus pcgen-RG).
 482 In case of the RG-test, either univariate or multivariate GBLUP could be used (pcgen-
 483 RG-uni versus pcgen-RG-multi). pcRes can be based on either univariate or multivariate
 484 GBLUP (pcRes-uni versus pcRes-multi). In either case, one can use a relatedness matrix
 485 only (postscript -K, for 'kinship'), only replicates (-R), or both replicates and a relatedness
 486 matrix (-RK). Finally, one could reconstruct \mathcal{G}_y using the GBLUP itself (pc-GBLUP),
 487 similar to the approach of [13].

488 2.5 Assessing uncertainty

489 The PC-algorithm is asymptotically correct, in the sense that the underlying CPDAG is
 490 recovered if conditional independence can be tested without error [5]. In Theorem 2 below
 491 we provide a related consistency result for pcgen. In practice however, type-I or type-II
 492 errors are likely to occur, leading to incorrect edges in the graph. Depending on the
 493 significance level α used in each test, there may be more type-I errors (large α) or rather
 494 more type-II errors (small α). However, reliable control of the (expected) false positive
 495 rate or total number of false positives remains challenging; see the discussion, section 4.3.
 496 We will therefore just consider the p-values as they are, and analyze a given dataset for
 497 various significance thresholds. A rough indication of confidence for each remaining edge
 498 is given by the largest p-value found across all conditioning sets for which the edge was
 499 tested.

500 2.6 Extensions of pcgen

501 We conclude the materials and methods section with the following extensions of pcgen:

- 502 • **The RC-test with prior screening (pcgen-screening):** The RC-test for state-
503 ment (A) is computationally efficient, in the sense that only a bivariate MTM needs
504 to be fitted, instead of for all traits. At the same time however, once residuals are
505 available, the RG-test is much faster than the RC-test, as the former is based on
506 partial correlations that can be computed recursively (see e.g [32]). An appealing
507 strategy is therefore to run PC-stable on univariate residuals (i.e., a form of pcRes),
508 and use the resulting skeleton as the starting point for pcgen with the RC-test. The
509 advantage, at least for sparse graphs, is that the number of conducted RC-tests is
510 greatly reduced. In the pcgen-package this is implemented in the pcgenFast func-
511 tion. The skeleton based on univariate residuals typically contains somewhat more
512 false edges, but these may be removed later on with the RC-test.

- 513 • **Inclusion of QTLs:** apart from the random genetic effects, the GSEM considered
514 here could include fixed effect QTLs as well. Each QTL is represented by a single
515 node, which like the random effect node G is always a root node. No edges are
516 allowed between QTLs or between a QTL and G ; moreover every edge between
517 a QTL and a trait is oriented towards the trait. Since the QTLs may further
518 improve the orientation of the graph, the pcgen package provides an experimental
519 implementation of this, although a full investigation of the added value is left for
520 future work.

- 521 • **Comparing pcgen output with genetic variance estimates:** for every trait
522 \mathbf{Y}_j having positive genetic variance, there should be either a direct genetic effect
523 $G \rightarrow Y_j$ or a partially directed path from G to Y_j (with possibly undirected edges,
524 but all directed edges pointing towards Y_j). However, because of statistical errors it
525 may happen that neither of the two exist in the CPDAG obtained from pcgen, while
526 at the same time the genetic variance (considered in the independence test $\mathbf{Y}_j \perp\!\!\!\perp \mathbf{G}$)
527 is significant. In the pcgen-package, such conflicts can be detected using the checkG
528 function. If conflicts occur, we conclude that there was insufficient evidence to
529 remove the direct genetic effect from the graph, and re-run pcgen, skipping all tests
530 $\mathbf{Y}_j \perp\!\!\!\perp \mathbf{G} | \mathbf{Y}_S$ for all \mathbf{Y}_j that produced conflicts in the first run. This forces the
531 algorithm to keep the edge $G \rightarrow Y_j$.

532 3 Results

533 3.1 Simulations, with randomly drawn graphs

534 3.1.1 Simulation scheme

535 To compare the different algorithms we simulated random GSEMs, by combining a ran-
 536 domly drawn DAG over the traits (\mathcal{G}_Y) with random path coefficients, closely following
 537 the simulation scheme of [32]. Traits were simulated for 200 genotypes which (for each
 538 simulated dataset) were randomly drawn from an existing population of 254 maize hy-
 539 brids [33]. Two replicates of each genotype were simulated. For each simulated dataset we
 540 randomly sampled the set D (defining the edges $G \rightarrow Y_j$) and the covariance matrix Σ_G .
 541 Given an additive relatedness matrix A based on 50k SNPs, genetic effects were simulated
 542 as in (15), i.e. such that $\text{vec}(\mathbf{G}) \sim \Sigma_G \otimes (ZAZ^t)$. Appendix B.1 provides further details,
 543 such as the magnitude of genetic (co)variances. We focus here on the comparison of

- 544 • pcgen-RC: pcgen with the RC-test, using replicates
- 545 • pcgen-RG-uni: pcgen with the RG-test, based on univariate GBLUP, and using
 546 replicates
- 547 • pcRes-uni-R: pcRes based on univariate GBLUP, using replicates
- 548 • pcRes-multi-K: pcRes based on multivariate GBLUP, using genotypic means and
 549 the relatedness matrix A that was used to simulate the data
- 550 • pcgen-screening: pcgen with the RC-test, starting with the skeleton obtained from
 551 pcRes-uni-R (see section 2.6)

552 Results for the other algorithms are given in Appendix B.3. In all simulations the signifi-
 553 cance threshold was $\alpha = 0.01$. The effect of sample size and the trade-off between power
 554 and false positives as function of α was already investigated by [32] for the standard
 555 PC-algorithm, and is likely to be similar for pcgen.

556 We separately evaluated the reconstruction of \mathcal{G}_Y and the edges $G \rightarrow Y_j$, as the latter
 557 is only possible with pcgen. To assess the difference between estimated and true skeleton
 558 of \mathcal{G}_Y , we considered the true positive rate (TPR), the true discovery rate (TDR) and
 559 the false positive rate (FPR). Additionally we used the Structural Hamming Distance
 560 (SHD), which also takes into account the orientation of the edges. Appendix B.2 provides
 561 definitions of these criteria. Reconstruction of $G \rightarrow Y_j$ is only assessed in terms of TPR,
 562 TDR and FPR, as these edges can have only one orientation.

563 3.1.2 Results

564 We first performed simulations with $p = 4$ traits (scenario 1), with each potential edge
 565 between traits occurring in the true graph with probability $p_t = 1/3$. Hence, for any

566 given trait, the expected number of adjacent traits was $(p - 1)p_t = 1$. The edges $G \rightarrow Y_j$
 567 were included in the true graph with probability $p_g = 1/2$. In a related set of simulations
 568 (scenario 2), p_t was increased to 0.5, giving denser graphs. In both scenarios, pcgen
 569 reconstructed the edges $G \rightarrow Y_j$ with little error, the average TPR being above 0.95 and
 570 the TDR above 0.99. The FPR for these type of edges remained below 0.05. For the
 571 trait-to-trait relations (i.e. reconstruction of \mathcal{G}_Y), the TPR, FPR and TDR all increased
 572 in scenario 2. Hence, for denser graphs, more of the true edges were found, at the expense
 573 of a somewhat higher number of false edges.

574 pcRes with univariate residuals from replicates (pcRes-uni-R) outperformed pcRes-
 575 multi-K, despite the fact that only the latter used the actual relatedness matrix (from
 576 which the genetic effects were simulated). Consequently, the information contained in
 577 the replicates appears much more important than the precise form of the relatedness ma-
 578 trix, or unbiased estimation of genetic correlations. The performance of pcRes strongly
 579 depends on the prediction error of the GBLUP (see sections 2.2.2 and 2.2.4), and, in
 580 line with the results of [18], this error appeared lowest when using the replicates. pcRes-
 581 uni-R performed almost as well as the 3 pcgen approaches; motivated by the additional
 582 computational advantage, this is the only variant of pcRes that we consider in the re-
 583 mainder. For the same reasons, we only considered the RG-test in pcgen in the univariate
 584 version, based on replicates (pcgen-RG-uni). The latter appeared to give more false pos-
 585 itives than pcgen with the RC-test (pcgen-RC), although differences were small. A good
 586 compromise between computational speed and performance appears to be achieved with
 587 pcgen-screening, starting pcgen-RC with the skeleton found by pcgen-RG-uni. This will
 588 therefore be our default choice for analyzing real data below.

589 pcRes-uni-R and pcgen-RG-uni had identical performance in terms of TPR, TDR and
 590 FPR, as they use exactly the same tests for the trait-trait relations. However, pcgen-RG-
 591 uni (as other pcgen approaches) had the advantages that (i) also the edges $G \rightarrow Y_j$ could
 592 be inferred and (ii) due to these genetic effects, the orientation of the edges between traits
 593 was improved, as shown by a strongly reduced SHD.

594 To assess performance in higher dimensions, we simulated data sets with $p = 20$ traits,
 595 $p_g = 0.3$ and $p_t = 0.1$ (scenario 3) and with $p = 100$, $p_g = 0.1$ and $p_t = 0.01$ (scenario 4).
 596 Both scenarios consider sparse graphs; denser graphs can be analyzed as well, but, for p
 597 larger than 20-30, require several hours or even days, unless the size of the conditioning
 598 sets is restricted or pcgen would be parallelized. Here we limited the size of conditioning
 599 sets to 3 (scenario 3) and 2 (scenario 4). As in the first two scenarios, pcgen led to a
 600 strong reduction in SHD, and reliable reconstruction of the direct genetic effects.

	\mathcal{G}_y				$G \rightarrow Y_j$		
	TPR	FPR	TDR	SHD	TPR	FPR	TDR
Scenario 1 ($p = 4$)	$(p_t = 1/3)$				$(p_g = 0.5)$		
pcgen-RC	0.607	0.012	0.960	0.568	0.986	0.020	0.994
pcgen-RG-uni	0.607	0.036	0.906	0.670	0.984	0.022	0.994
pcgen-screening	0.603	0.008	0.982	0.548	0.986	0.020	0.994
pcRes-uni-R	0.607	0.036	0.906	1.362			
pcRes-multi-K	0.465	0.432	0.388	3.212			
Scenario 2 ($p = 4$)	$(p_t = 0.5)$				$(p_g = 0.5)$		
pcgen-RC	0.814	0.042	0.980	1.236	0.966	0.041	0.993
pcgen-RG-uni	0.820	0.074	0.951	1.320	0.965	0.043	0.992
pcgen-screening	0.805	0.034	0.990	1.238	0.966	0.042	0.992
pcRes-uni-R	0.820	0.074	0.951	2.274			
pcRes-multi-K	0.666	0.466	0.603	3.778			
Scenario 3 ($p = 20$)	$(p_t = 0.1)$				$(p_g = 0.3)$		
pcgen-RG-uni	0.911	0.004	0.961	7.020	0.968	0.018	0.991
pcgen-screening	0.895	0.002	0.985	6.800	0.969	0.018	0.991
pcRes-uni-R	0.912	0.004	0.961	9.866			
Scenario 4 ($p = 100$)	$(p_t = 0.01)$				$(p_g = 0.1)$		
pcgen-RG-uni	0.959	0.001	0.942	27.298	0.976	0.022	0.943
pcRes-uni-R	0.962	0.001	0.940	38.404			

Table 1. Performance of pcgen and residuals-based approaches, averaged over 500 simulated data-sets per scenario. For scenarios 1 and 2, performance of additional residual approaches is given in Appendix B.3.

601 3.2 Simulations, using a crop-growth model

602 We also simulated data using the popular crop growth model APSIM [34,35]. Compared
603 to the preceding simulations this represents a more challenging (and probably more real-
604 istic) scenario, as several of the underlying assumptions are violated. In particular, the
605 data-generating process introduces nonlinearities and latent variables. We simulated 12
606 traits for an existing wheat population of 199 genotypes, with 3 replicates each. The
607 traits include 7 primary traits, 4 secondary traits and yield. Appendix B.4 provides trait
608 acronyms (Table 3) and further details. Traits were simulated by running a discrete
609 dynamic model from the beginning ($t = 0$) to the end of the growing season ($t = T$).
610 Motivated by the fact that some trait measurements are destructive, observations are
611 only taken at $t = T$. Figure 3a shows the summary graph, defining the causal effects
612 from one time-step to the next ([24]). It does not directly describe the joint distribution
613 of the traits at $t = T$ (obtained by marginalizing over $t = 0, \dots, T - 1$, and typically
614 represented by an ancestral graph [36]), but for simplicity we nevertheless investigate to

615 what extent we can reconstruct the summary graph. There are direct genetic effects on
 616 all of the primary traits, which have heritability 0.9. The genetic effects originate from
 617 a large number of trait-specific QTLs, with randomly drawn effect sizes. There are no
 618 direct genetic effects on the secondary traits and yield.

619 Figure 3b shows the pcgen estimate for a single simulated data-set, with pcgen-RC
 620 and prior screening (pcgen-screening), and $\alpha = 0.01$. Similar results (not shown) were
 621 obtained for other data-sets, smaller α or without screening. Compared to the simulations
 622 above, it turned out to be much harder to detect the absence of direct genetic effects:
 623 in the pcgen reconstruction, all 12 traits had such effects (highest p-value: $1.7 \cdot 10^{-4}$).
 624 Because of the large degrees of freedom for genotype and generally small residual variances,
 625 most of the F-statistics for genotype are highly significant, but often orders of magnitude
 626 smaller than the F-ratios for the conditioning traits. For example, in the test for a direct
 627 genetic effect on flowering time (FT) conditional on SV and TFI, F-ratios for SV, TFI
 628 and genotype were respectively 3469.95, 68744.14 and 5.71. As in our example in section
 629 2.2.3, another reason for the significant genetic effects may be the nonlinearity of the
 630 underlying model.

631 The reconstructed trait-to-trait relations were mostly correct, except for the missing
 632 edge $GN \rightarrow Y$, and one incorrect orientation ($Y \rightarrow GW$). pcRes made the same errors
 633 (Figure 3c), with an additional false arrow ($MGS - SP$). The standard PC-stable algo-
 634 rithm applied to all traits and QTLs lead to many more errors (Figure 3d), for example
 635 the false edge between GW and RUE , the missing edge $TFI \rightarrow FT$ and the incorrect
 636 orientations $BM \rightarrow RUE$ and $Y \rightarrow BM$. These errors appeared to be the consequence
 637 of small effect QTLs not being detected; consequently part of the genetic variance was
 638 not taken into account.



Figure 3. Networks of 12 APSIM traits in the dry environment Emerald 1993, including 7 primary traits (grey), 4 secondary traits (green) and yield (red). Trait acronyms are given in supplementary Table 3 (Appendix B.4)

639 3.3 Real data

640 3.3.1 Maize

641 We now use pcgen-screening to analyze the field trials reported by [33], involving 254
 642 hybrids of maize (*Zea mays*). We consider a subset of 6 trials, conducted in 2013 under
 643 two different watering regimes (rain-fed (-R) and irrigated (-I)) and at three locations:
 644 Karlsruhe (Germany), Nérac (France) and Graneros (Chile). Most of the trials included 3
 645 height related traits (tassel height (*TH*), ear height (*EH*) and (total) plant height (*PH*)),
 646 2 flowering time related traits (silking (*S*) and anthesis (*A*)), and 3 yield related traits
 647 (grain number (*GN*), grain weight (*GW*) and yield itself (*Y*)). All traits were measured
 648 independently, i.e. no trait measurement was derived from other traits.

649 Each trial was laid out as an alpha-lattice design, with two replicates for the irrigated
 650 trials and three for the rain-fed trials. Spatial trends and (in)complete block effects were
 651 estimated using the mixed model of [37] (R-package SpATS), and subtracted from the
 652 original data; pcgen was then applied to the detrended data, assuming a completely
 653 randomized design. Residuals from SpATS appeared approximately Gaussian, and no
 654 further transformation was applied. For the yield related traits, prior biological knowledge
 655 (*Y* being the product of *GS* and *GN*) might suggest a log-transform, but we assume that
 656 such biological knowledge is not available to the algorithm.

657 Figure 4 shows the estimated networks for $\alpha = 0.01$. Networks obtained with $\alpha =$
 658 0.001 (Appendix H) were mostly identical, except for 1 or 2 edges missing in the Karlsruhe-
 659 I, Nérac-R and Graneros-I trials. In all 6 trials there was a clear clustering of traits
 660 according to their biological nature (height, flowering and yield related). Edges across
 661 these groups were only found for the 3 rain-fed trials (**comment on the *S – PH* and
 662 *S – TH* edges in Karlsruhe and Nérac, and the *PH – Y* edge in Graneros**). Except for the
 663 Graneros trials, direct genetic effects were present for all traits, which might again be the
 664 consequence of nonlinear relationships. As in the APSIM simulations, the test for direct
 665 genetic effects typically produced highly significant results, but often with small genetic
 666 effects. In the Karlsruhe-I trial for example, the ANCOVA for yield (*Y*) conditional
 667 on *GN* and *GS* gave an F-ratio for genotype of only 2.71, compared to 55364.24 and
 668 8036.53 for respectively *GN* and *GS*. In two of the six trials pcgen correctly identified
 669 $GN \rightarrow Y \leftarrow GS$, i.e. the expected relations between yield and its components. In
 670 two other ones the edges $GN - Y - GS$ could not be oriented because of the additional
 671 (and unexpected) edge $GN - GS$. In the remaining two trials (Graneros-R and Nérac-I)
 672 we found $GS \rightarrow GN \leftarrow Y$. While biologically improbable, this simply represents the
 673 outcomes of the tests for the given data. In particular, the test for $\mathbf{Y}_j \perp \mathbf{Y}_k | (\mathbf{G}, \mathbf{Y}_S)$ is
 674 highly significant when $\mathbf{Y}_j, \mathbf{Y}_k$ and \mathbf{Y}_S are respectively yield, grain size and grain number,
 675 but not significant when dropping grain number from the conditioning set ($p = 0.056$, for
 676 Nérac-I).

677 The trials also illustrate the distinction made in section 2.1.4, between the total genetic
 678 covariance and the covariance among direct genetic effects (as defined by Σ_G). For most

679 pairs of traits, the former was highly significant, with a typical (total) genetic correlation
680 of $\rho_g = 0.3 - 0.8$ (for better interpretability we report correlations, although we test for
681 zero genetic covariance, see Appendix A.4). Only for pairs involving *GS*, values were
682 often negative, and not always significant. In the Karlsruhe-I trial for example, we found
683 $\rho_g = -0.11$ for *GS* and *EH*, and $\rho_g = -0.41$ for *GS* and *S*. In both cases, the two
684 traits are d-separated in the graph, but only for *S* (silking) the genetic covariance is
685 significant ($p = 4.89 \cdot 10^{-9}$). While this may provide information about Σ_G , we recall that
686 usually the latter is not entirely identifiable. For example, *GN* and *Y* always had a high
687 genetic correlation, but without restrictions on the path coefficients one cannot estimate
688 the corresponding element of Σ_G .

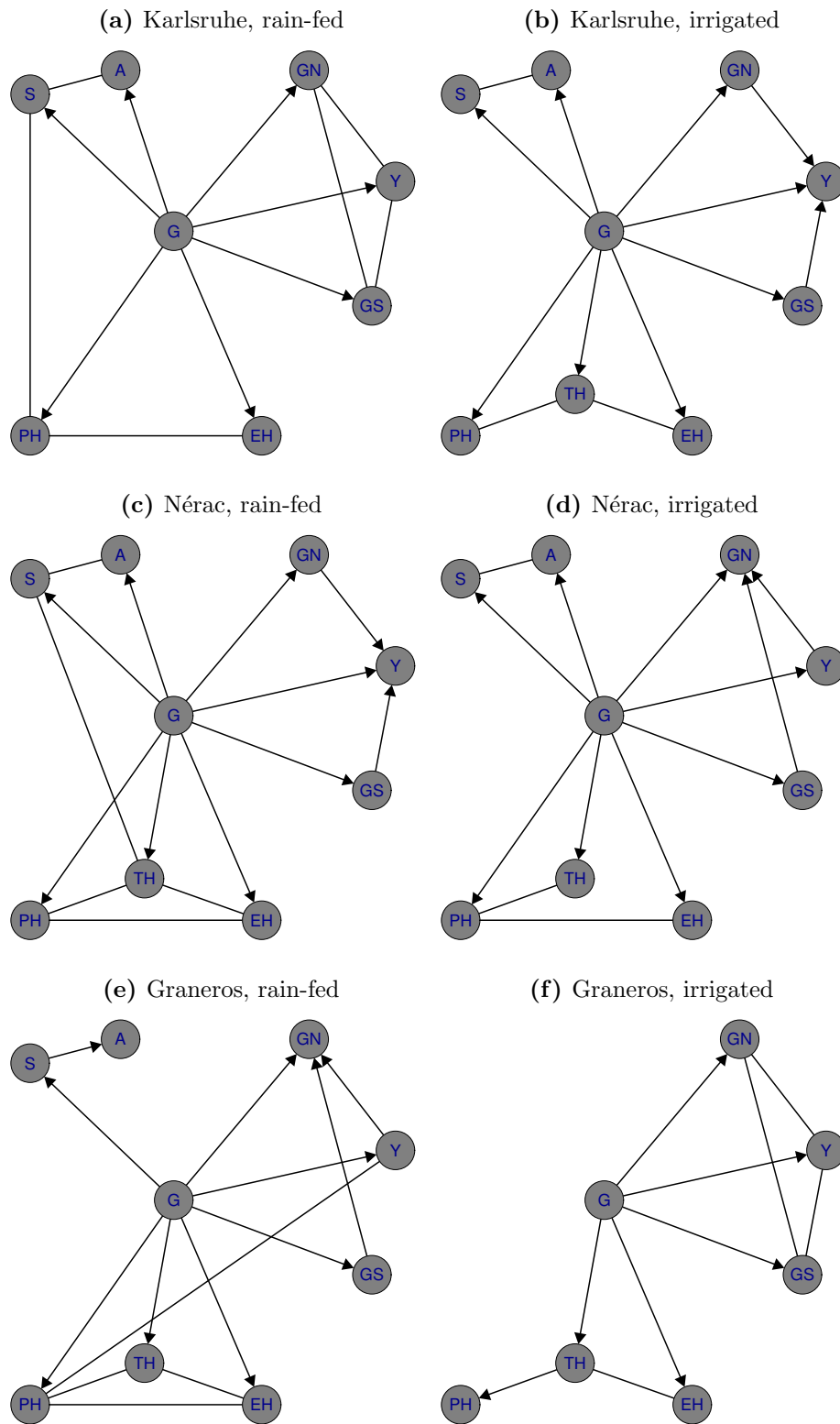


Figure 4. Estimated networks for six of the DROPS field trials in 2013, with $\alpha = 0.01$. Rows correspond to locations (Karlsruhe, Nérac, Graneros), columns to treatments (rain-fed, irrigated).

689 3.3.2 Rice

690 Using again pcgen-screening, we analyzed 34 traits measured on 274 *indica* genotypes
691 of rice (*Oryza sativa*) under water-deficit, reported by [38]. Trait acronyms are given in
692 Supplementary Table 4 in Appendix C. 3 replicates of each genotype were phenotyped in a
693 randomized complete block design, and block was included as a covariate in all conditional
694 independence tests, which were performed at significance level $\alpha = 0.01$. A first run of
695 pcgen produced many inconsistencies in the genetic effects, i.e. traits with significantly
696 positive heritability but without a partially directed path coming from the genotype node.
697 We therefore applied the correction described at the end of section 2.6, adding edges $G \rightarrow$
698 Y_j for all traits Y_j with this inconsistency, and then re-ran pcgen. The final reconstruction
699 is given in Figure 5, where traits are grouped into 4 shoot morphological traits (blue), 2
700 physiological traits (rose), 16 root morphological traits (green), 6 root anatomical traits
701 (gray) and 6 dry matter traits (orange).

702 Even after correcting the inconsistencies, there were 10 traits without a direct genetic
703 effect; for 7 of these there was a partially directed path starting from genotype, and
704 first passing through *RL1015* (root length, per plant, restricted to roots with diameter
705 between 10 and 15 mm). Relations among traits mostly clustered according to their
706 biological categories, except for the effect of TW (total weight) on CWT (cumulative
707 water transpiration), and RW and RS (root and stem weight) appearing in a cluster of
708 root morphology traits. In particular, there are no links from root morphological traits to
709 biomass-related (physiological?) traits, which might be expected under water-deficit (cite
710 ...?). Apart from errors in the test, this may be a consequence of the experimental setup,
711 where the use of pots might have restricted root growth. For the anatomical traits, the
712 lack of connections with trait categories could be the result of high measurement error,
713 violating our assumption that residual errors are driven by biological variation.

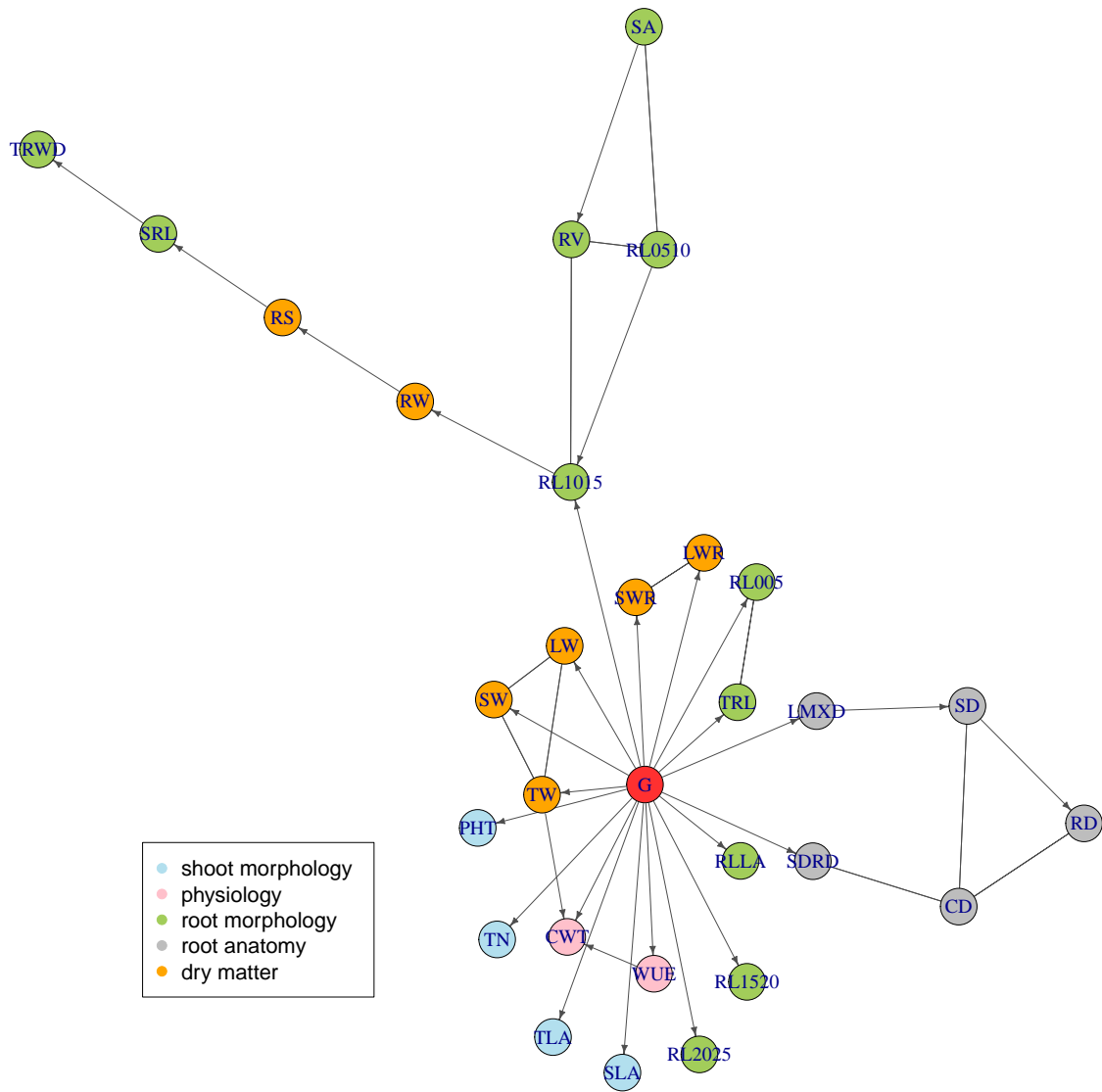


Figure 5. pcgen-reconstruction for the rice-data from [38], with $\alpha = 0.01$.

714 3.4 Statistical properties of pcgen

715 We now investigate a number of statistical issues: the assumptions required for asymptotic
 716 consistency of pcgen (section 3.4.1), the assumptions required for faithfulness (section
 717 3.4.2), and properties of the conditional independence tests (section 3.4.3). Proofs of
 718 Theorems 1-6 are given in Appendix E.

719 **3.4.1 Consistency**

720 Asymptotic consistency holds if, for increasing sample size, the probability of finding the
 721 correct network converges to 1. Correct in this context means that we recover the class of
 722 partially directed graphs (CPDAG) that contains the underlying DAG. Consistency of the
 723 PC-algorithm was shown by [5] (for low dimensions) and [32] (for high dimensions). These
 724 authors distinguished consistency of the oracle version of PC, where conditional indepen-
 725 dence information is available without error, and the sample version, where conditional
 726 independence is obtained from statistical tests. For pcgen we will focus on the oracle
 727 version and consistency of the skeleton, leaving the sample version and the correctness of
 728 the orientations for future work.

729 As for the standard PC-algorithm, consistency of pcgen requires the equivalence be-
 730 tween conditional independence and d-separation in the graph. Part of this is the Markov
 731 property, which states that d-separation of two nodes in the graph given a set of other
 732 nodes implies conditional independence of the corresponding random variables. The con-
 733 verse (conditional independence implying d-separation) is known as faithfulness. The
 734 following result provides the Markov property for SEM with genetic effects. The proof
 735 (Appendix E.8) is a straightforward adaptation of Pearl’s proof for general SEM ([4]).

736 **Theorem 1** *Suppose we have a GSEM as defined by equation (5), with a graph \mathcal{G} as*
 737 *defined in section 2.1.3, and satisfying assumptions 1-4 given in section 2.1.2. Then*
 738 *the global Markov condition holds for \mathcal{G} and the joint distribution of $\mathbf{G}, \mathbf{Y}_1, \dots, \mathbf{Y}_p$. In*
 739 *particular, d-separation of Y_j and G given Y_S implies $\mathbf{Y}_j \perp\!\!\!\perp \mathbf{G} \mid \{\mathbf{Y}_S\}$, and d-separation of*
 740 *Y_j and Y_k given $\{Y_S, G\}$ implies $\mathbf{Y}_j \perp\!\!\!\perp \mathbf{Y}_k \mid \{\mathbf{Y}_S, \mathbf{G}\}$, for all traits \mathbf{Y}_j and \mathbf{Y}_k and subsets*
 741 *\mathbf{Y}_S .*

742 If we now assume faithfulness, the preceding result directly gives the equivalence be-
 743 tween conditional independence and d-separation. This in turn implies that pcgen will
 744 recover the correct skeleton, at least if conditional independence can be tested without
 745 error:

746 **Theorem 2** *Let $dsep(\mathcal{G})$ denote d-separation in the graph \mathcal{G} . Suppose we have a GSEM*
 747 *as in Theorem 1, and we make the additional assumptions of faithfulness:*

$$\mathbf{Y}_j \perp\!\!\!\perp \mathbf{Y}_k \mid \{\mathbf{Y}_S, \mathbf{G}\} \implies Y_j \text{ dsep}(\mathcal{G}) Y_k \mid \{Y_S, G\} \quad (18)$$

748

$$\mathbf{Y}_j \perp\!\!\!\perp \mathbf{G} \mid \{\mathbf{Y}_S\} \implies Y_j \text{ dsep}(\mathcal{G}) G \mid \{Y_S\}, \quad (19)$$

749 *for all traits \mathbf{Y}_j and \mathbf{Y}_k and subsets \mathbf{Y}_S . Then the oracle version of pcgen gives the correct*
 750 *skeleton.*

751 **3.4.2 Faithfulness**

752 For our first faithfulness condition (equation (18)), it suffices to have faithfulness for the
 753 network without genetic effects. A necessary (but not sufficient) condition for this is that
 754 contributions from different paths do not cancel out (Appendix E.4).

755 **Theorem 3** *Let $P_{Y|G\Gamma}$ denote the joint distribution of $\mathbf{Y}_1, \dots, \mathbf{Y}_p$ conditional on $\mathbf{U} =$
 756 $\mathbf{G}\Gamma$, the matrix of total genetic effects. Then $\mathbf{Y}_j \perp\!\!\!\perp \mathbf{G}|\{\mathbf{Y}_S\}$ is equivalent with $\mathbf{Y}_j \perp\!\!\!\perp_{P_{Y|G\Gamma}}$
 757 \mathbf{Y}_k , and $Y_j \text{dsep}(\mathcal{G}) G|\{Y_S\}$ is equivalent with $Y_j \text{dsep}(\mathcal{G}_Y) Y_k|Y_S$. Consequently, (18) holds
 758 if*

$$\mathbf{Y}_j \perp\!\!\!\perp_{P_{Y|G\Gamma}} \mathbf{Y}_k|\{\mathbf{Y}_S\} \implies Y_j \text{dsep}(\mathcal{G}_Y) Y_k|Y_S. \quad (20)$$

759 The second faithfulness statement (equation (19)) involves d-separation of Y_j and G ,
 760 and requires that the genetic effects are not collinear. If for example we have $\mathbf{Y}_3 =$
 761 $\mathbf{Y}_1 + \mathbf{Y}_2 + \mathbf{E}_3$, with $\mathbf{Y}_1 = \mathbf{G}_1 + \mathbf{E}_1$, $\mathbf{Y}_2 = \mathbf{G}_2 + \mathbf{E}_2$, and $\mathbf{G}_2 = -\mathbf{G}_1 = \mathbf{G}$, it follows
 762 that $\mathbf{Y}_3 = \mathbf{E}_1 + \mathbf{E}_2 + \mathbf{E}_3$. Consequently, because $\mathbf{G}_3 = (0, \dots, 0)^t$, we find that \mathbf{Y}_3 and
 763 $\mathbf{G} = [\mathbf{G}_1 \ \mathbf{G}_2 \ \mathbf{G}_3]$ are marginally independent, but in the graph \mathcal{G} , the nodes Y_j and G
 764 are not d-separated by the empty set, as there are directed paths $G \rightarrow Y_2 \rightarrow Y_3$ and
 765 $G \rightarrow Y_1 \rightarrow Y_3$. Conversely, if \mathbf{G}_1 and \mathbf{G}_2 are not perfectly correlated this violation
 766 of faithfulness cannot occur. The following theorem shows that marginal independence
 767 always implies d-separation. We conjecture (but could not prove) that (19) also holds for
 768 non-empty conditioning sets.

769 **Theorem 4** *Suppose we have a GSEM satisfying Assumptions 1-5, and faithfulness for
 770 the network without genetic effects, given by (20). Then (19) holds for $S = \emptyset$, i.e.,
 771 marginal independence implies d-separation of Y_j and G_j .*

772 **3.4.3 Properties of the tests**

773 Theorem 2 provides consistency of the oracle version of pcgen, where conditional indepen-
 774 dence information is available without error. Proving consistency of the sample version
 775 is challenging for two reasons. First, the conditional independence tests often rely on
 776 approximations (as described in section 2.2.2). In those cases, the tests are based on
 777 misspecified models, and instead of $\mathbf{Y}_j \perp\!\!\!\perp \mathbf{Y}_k|\{\mathbf{Y}_S, \mathbf{G}\}$ or $\mathbf{Y}_j \perp\!\!\!\perp \mathbf{G}|\{\mathbf{Y}_S\}$, we test a some-
 778 what different statement. Consequently, a necessary condition for consistency is that the
 779 approximation error converges to zero. Second, even without approximation errors, the
 780 probability on type-I and type-II errors still needs to converges to zero, with increasing
 781 sample size. This is well known for the PC-algorithm with independent Gaussian data
 782 ([32]), but more difficult to establish in the presence of genetic effects.

783 Here we address the first issue, leaving the second for future work. We will focus on

784 the RC-test ⁵ for $\mathbf{Y}_j \perp\!\!\!\perp \mathbf{Y}_k | \{\mathbf{Y}_S, \mathbf{G}\}$ and the test for $\mathbf{Y}_j \perp\!\!\!\perp \mathbf{G} | \{\mathbf{Y}_S\}$, assuming $K = ZZ^t$.
 785 These tests rely on the conditional distributions $\mathbf{Y}_j | \mathbf{Y}_S$ and $(\mathbf{Y}_j, \mathbf{Y}_k) | \mathbf{Y}_S$, for which we
 786 made several assumptions regarding their means and covariances. It turns out that the
 787 covariances assumptions seem to hold, at least for $K = ZZ^t$. However, the assumption on
 788 the conditional means is often violated, also for $K = ZZ^t$. We will now first summarize
 789 our results for the covariance, then compute the conditional mean in two examples, and
 790 conclude with a discussion on how a better approximation of the conditional mean may
 791 be obtained.

792 Conditional covariance

793 **Theorem 5** *When $K = ZZ^t$, the distribution of $(\mathbf{Y}_j, \mathbf{Y}_k) | \mathbf{Y}_S$ has covariance of the form*
 794 *given by (13), i.e., that of a bivariate MTM. Moreover, under faithfulness condition (18),*
 795 *the residual covariance in the MTM is zero if and only if $\mathbf{Y}_j \perp\!\!\!\perp \mathbf{Y}_k | \{\mathbf{Y}_S, \mathbf{G}\}$.*

796 The idea behind our test for $\mathbf{Y}_j \perp\!\!\!\perp \mathbf{G} | \mathbf{Y}_S$ was that the conditional distribution of
 797 \mathbf{Y}_j given \mathbf{Y}_S is of the form $\sigma_G^2(j|S)K + \sigma_E^2(j|S)I_n$, and that $\sigma_G^2(j|S) = 0$ if and only if
 798 $\mathbf{Y}_j \perp\!\!\!\perp \mathbf{G} | \{\mathbf{Y}_S\}$. While the former statement always holds, we could prove the latter only
 799 for the empty conditioning set. This is because faithfulness is required, which we also
 800 established only for $S = \emptyset$ (see Theorem 4)

801 **Theorem 6** *Suppose we have a GSEM as described in Theorem 1, with $K = ZZ^t$ and*
 802 *$\Sigma_G[D, D]$ of full rank (Assumption 5). Then the distribution of $\mathbf{Y}_j | \mathbf{Y}_S$ is of the form*
 803 *$\sigma_G^2(j|S)K + \sigma_E^2(j|S)I_n$, for any conditioning set S . Moreover, under faithfulness condition*
 804 *(19), $\sigma_G^2(j) = \sigma_G^2(j|\emptyset)$ is zero if and only if $\mathbf{Y}_j \perp\!\!\!\perp \mathbf{G}$.*

805 Conditional mean

806 In addition to the covariance assumptions, our tests for $\mathbf{Y}_j \perp\!\!\!\perp \mathbf{Y}_k | \{\mathbf{Y}_S, \mathbf{G}\}$ and $\mathbf{Y}_j \perp\!\!\!\perp$
 807 $\mathbf{G} | \{\mathbf{Y}_S\}$ relied on the assumption that the conditional mean for the j th trait was $\mathbf{X}B\gamma_j +$
 808 $\tilde{y}_S\beta_S^{(j)}$, i.e., the sum of the fixed effects on \mathbf{Y}_j and a linear regression on the traits in
 809 the conditioning set (see equation (12)). This seems correct in case Y_S contains only 1
 810 trait, which, in the true graph, is the only parent of Y_j . In other situations however, the
 811 regression is only an approximation, as shown in the following example.

812 Suppose that $\mathbf{Y}_1 = \mathbf{G}_1 + \mathbf{E}_1$ and $\mathbf{Y}_2 = \lambda\mathbf{Y}_1 + \mathbf{E}_2$, with independent vectors $\mathbf{G}_1 \sim$
 813 $N(0, \sigma_{G,1}^2 K)$, $\mathbf{E}_1 \sim N(\sigma_{E,1}^2 I_n)$ and $\mathbf{E}_2 \sim N(\sigma_{E,2}^2 I_n)$. Then the graph \mathcal{G} is given by

⁵The RG-test for $\mathbf{Y}_j \perp\!\!\!\perp \mathbf{Y}_k | \{\mathbf{Y}_S, \mathbf{G}\}$ requires that the GBLUP \mathbf{U}^* is close to the true matrix of genetic effects (\mathbf{U}). Apart from the difficulty of obtaining good estimates of genetic and residual covariances, the quality of this approximation can be easily assessed using expressions for the prediction error variance (see e.g. [39]).

814 $G \rightarrow Y_1 \rightarrow Y_2$. There is no edge $G \rightarrow Y_2$, although this is not essential for the example.
 815 The distributions are given by

$$\begin{aligned} \mathbf{Y}_1 &\sim N(0, \Sigma_1) = N(0, \sigma_{G,1}^2 K + \sigma_{E,1}^2 I_n), \\ \mathbf{Y}_2 &\sim N(0, \Sigma_2) = N(0, \lambda^2 \sigma_{G,1}^2 K + (\lambda^2 \sigma_{E,1}^2 + \sigma_{E,2}^2) I_n), \\ \text{Cov}(\mathbf{Y}_1, \mathbf{Y}_2) &= \Sigma_{12} = \lambda(\sigma_{G,1}^2 K + \sigma_{E,1}^2 I_n) = \lambda \Sigma_1. \end{aligned}$$

816 The conditional mean of \mathbf{Y}_2 given $\mathbf{Y}_1 = y_1$ is $\mu_{2|1} = \Sigma_{12} \Sigma_1^{-1} y_1 = \lambda y_1$. As expected
 817 given the graph, the conditional mean is a simple linear regression on Y_1 . However, the
 818 conditional mean of \mathbf{Y}_1 given $\mathbf{Y}_2 = y_2$ equals

$$\mu_{1|2} = \lambda \Sigma_1 (\lambda^2 \Sigma_1 + \sigma_E^2(2) I_n)^{-1} y_2,$$

819 which is a linear transformation, but not a multiple of y_2 . Consequently, our models
 820 for $\mathbf{Y}_j | \mathbf{Y}_S$ and $(\mathbf{Y}_j, \mathbf{Y}_k) | \mathbf{Y}_S$ are sometimes misspecified in terms of the mean, although
 821 still correct in terms of covariance, provided $K = ZZ^t$ (Theorems 5 and 6). In these
 822 cases, (RE)ML estimates will minimize the Kullback-Leibler divergence with the true
 823 distribution, giving the right estimates of genetic variance (or residual covariance), but
 824 not necessarily the correct p-values in hypotheses testing.

825 **The conditional mean: improving the approximation**

826 More generally, the conditional mean is a function of the genetic and residual covari-
 827 ances among \mathbf{Y}_j and \mathbf{Y}_S . In Appendix E.6 (equation (28)) we derive that $\mathbf{Y}_j | \mathbf{Y}_S = \tilde{y}_S$ has
 828 mean $\mu_{j|S} = \mathbf{X}B\gamma_j + \Sigma_{j,S} \Sigma_S^{-1} \text{vec}(\tilde{y}_S - \mathbf{X}B\Gamma_S)$. Defining $\eta_{j|S} = 0$ for $S = \emptyset$, we can write
 829 $\mu_{j|S} = \mathbf{X}B\gamma_j + \eta_{j|S}$. Consequently, our approximation of the conditional mean effectively
 830 models $\eta_{j|S}$ as a linear regression on \tilde{y}_S .

831 This approximation could probably be improved if instead we could obtain good esti-
 832 mates of $\hat{\Sigma}_{j,S}$ and $\hat{\Sigma}_S^{-1}$, and set $\hat{\eta}_{j|S} = \hat{\Sigma}_{j,S} \hat{\Sigma}_S^{-1} \text{vec}(\tilde{y}_S - \mathbf{X}\hat{B}\Gamma_S)$. Such estimates however
 833 require fitting a MTM for $|S| + 1$ traits, which for large S is statistically and compu-
 834 tationally challenging, unless pairwise or other approximations are applied ([40], [41]).
 835 Moreover, it seems unclear how this $\hat{\eta}_{j|S}$ should be incorporated in the tests.

836 **4 Discussion**

837 Causal inference for data with random genetic effects is challenging because of the covari-
 838 ance between these effects, and because the usual assumption of independent observations
 839 is violated. To address these problems we proposed a model where random genetic effects
 840 are part of the causal graph, rather than a nuisance factor that first needs to be elimi-
 841 nated. The resulting distributions and graphs were shown to satisfy the Markov property.
 842 This lead us to develop the pcgen algorithm, which tests conditional independence be-
 843 tween traits in the presence of genetic effects, and also conditional independence between

844 a trait and the genetic effects. We showed that the presence of a direct genetic effect
 845 can in principle be tested, just like the direct effect of a QTL can be tested. This is of
 846 course relative to the observed traits, i.e. for any effect $G \rightarrow Y_j$ there may always be an
 847 unmeasured trait Z such that $G \rightarrow Z \rightarrow Y_j$.

848 In our simulations pcgen outperformed existing approaches. Part of this improvement
 849 is due to phenotypic information on replicates, reducing the number of errors in the tests.
 850 Another part is due to the improved orientation, which is a consequence of the additional
 851 edges $G \rightarrow Y_j$. Compared to previous algorithms, pcgen also appeared computationally
 852 more efficient: depending on the choice of independence tests and the sparseness of the
 853 network, it can analyze around 20-100 traits on a single core, and many more if we limit the
 854 maximum size of the conditioning sets, or would parallelize the conditional independence
 855 tests.

856 To a considerable extent this efficiency is due to the use of univariate GBLUP, in
 857 the RG-test or the RC-test with prior screening. However, even without using univariate
 858 GBLUP, and with direct genetic effects on all traits, the RC-test still has an advantage
 859 over existing approaches: by incorporating the genetic effects in the PC-algorithm, we
 860 do not need to fit a MTM for all traits simultaneously, but only bivariate models. Our
 861 approach also makes genetic network reconstruction feasible with just two traits, and in
 862 absence of QTLs, or even no genotypic data at all.

863 As any causal inference method, pcgen only suggests causal models that are in some
 864 sense compatible with the data, and cannot validate the existence of a functional rela-
 865 tionship, which is only possible through additional experiments. Because of the required
 866 assumptions, the identifiability issues and the uncertainty in the estimated networks, it
 867 may be better to speak of algorithms for causal *exploration* than causal *discovery*. At
 868 the same time, analysis of one trait given another (e.g. yield given flowering time) is
 869 a common and natural thing to do ([15]). From that perspective, pcgen could be seen
 870 simply as a tool that performs such analyses systematically, compares them and visualizes
 871 the results. pcgen results for different significance levels could then be reported alongside
 872 other 'descriptive' statistics like heritability estimates and genetic correlations, suggesting
 873 functional hypotheses interesting for future research.

874 4.1 Data from different experiments

875 We assumed traits to be measured on the same individuals in the same experiment,
 876 with residuals errors arising from biological variation (Assumption 1 in section 2.1.2).
 877 In certain applications this assumption can indeed be restrictive, but it seems necessary
 878 for any constraint-based causal inference approach. Suppose traits were measured in
 879 different experiments, or residual errors would mostly come from measurement errors.
 880 Then our model (5) would be replaced by $\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{Y}_{\mathbf{G}}\mathbf{\Lambda} + \mathbf{G} + \mathbf{E}$, where $\mathbf{Y}_{\mathbf{G}} =$
 881 $\mathbf{X}\mathbf{B} + \mathbf{Y}\mathbf{\Lambda} + \mathbf{G}$ are the trait values without errors. At first sight this may appear attractive,
 882 as appropriate design of the experiments will ensure independent errors, and assumption

883 3 will be guaranteed. However, the residual covariance of the observed traits will be
 884 diagonal as well, instead of the matrix $\Gamma^t \Sigma_E \Gamma$ obtained under assumption 1 (see equations
 885 (3), (6) and (8)). However, the residual covariances contained in the latter matrix turned
 886 out to be essential for network reconstruction (e.g. Theorem 5). Without assumption 1
 887 we would therefore need to rely completely on the genetic effects. This in turn would
 888 require Σ_G to be diagonal, which seems even more restrictive. A relevant alternative
 889 approach here is that of invariance causal prediction [42], which infers causal effects that
 890 are consistent across several experiments, but still requires all traits to be measured in
 891 each experiment.

892 4.2 Genetically identical replicates and marker-based related- 893 ness matrices

894 In principle pcgen allows for any type of genetic relatedness. We have however focused
 895 on the case of independent genetic effects, for the following reasons:

- 896 • Higher power: estimates of (total) genetic variance based on replicates are typically
 897 more accurate than marker-based estimates based on genotypic means ([18], [43]).
 898 Similarly, our simulations suggest that replicates give better tests for (zero) genetic
 899 variance and residual covariance.
- 900 • Theoretical: when $K = ZZ^t$, the conditional independence statement considered in
 901 the RC-test is completely equivalent with $\mathbf{Y}_j \perp\!\!\!\perp \mathbf{Y}_k | \{\mathbf{Y}_S, \mathbf{G}\}$ (Theorem 5), while
 902 for other K it is only an approximation. Another argument is the robustness under
 903 misspecification: (univariate) broad-sense heritability estimates capture any type of
 904 genetic effect, while a model assuming only additive effects can produce strongly
 905 biased heritability estimates, if the actual genetic effects are for example partly
 906 epistatic. This can be formally shown by computing the Kullback-Leibler divergence
 907 between the true distribution and the model under consideration [44]. It seems
 908 plausible that this robustness extends to the multivariate models considered here,
 909 in particular when direct genetic effects are driven by different sets of QTLs, leading
 910 to trait-specific relatedness matrices.
- 911 • Computational: the test for $\mathbf{Y}_j \perp\!\!\!\perp \mathbf{G} | \mathbf{Y}_S$ can be based on standard ANOVA, which
 912 is many times faster than the LRT for a mixed model. Also the tests for $\mathbf{Y}_j \perp\!\!\!\perp$
 913 $\mathbf{Y}_k | \{\mathbf{Y}_S, \mathbf{G}\}$ are faster when $K = ZZ^t$.

914 The contributions of different types of genetic effects could in theory be incorporated
 915 in the network by introducing multiple genetic nodes, and conditional independence tests
 916 based on models that can distinguish these effects. This seems however difficult in practice
 917 due to the computational requirements and lack of statistical power.

918 Finally, we have not investigated the performance of pcgen for unbalanced designs, but
 919 it seems likely that small unbalancedness has only a minor effect. A more fundamental
 920 challenge seems to be the presence of incomplete blocks or spatial trends ([37], [45]).

921 4.3 Assessing uncertainty

922 If one mistakenly rejects the null-hypothesis of conditional independence (type-I error),
 923 pcgen leaves the corresponding edge, although it may still be removed at a later stage, with
 924 a different conditioning set. If the null-hypothesis is mistakenly accepted (type-II error),
 925 a true edge is removed, and will not be recovered. Moreover, it may affect the remaining
 926 tests, since d-separation of Y_j and Y_k is only tested given conditioning sets contained in
 927 $adj(Y_j)$ or $adj(Y_k)$, where the adjacency sets are defined relative to the current skeleton.
 928 This is correct in the oracle version, where some tests may indeed be skipped, but in the
 929 sample version pc(gen) may mistakenly skip an essential independence test. See [30] for
 930 examples.

931 Consequently, assessing uncertainty for constraint-based algorithms is difficult, and
 932 cannot be achieved by just applying some multiple testing correction to the p-values. To
 933 obtain bounds on the expected number of false edges in the skeleton, several authors
 934 have used stability selection [46], [47] or other sample-splitting techniques [13], but these
 935 are typically over- conservative and require an additional exchangeability assumption
 936 ([48], [49]). Alternatively, uncertainty may be assessed using Bayesian methods, which are
 937 however computationally very demanding and outside the scope of this work. Moreover,
 938 despite the recent progress in Bayesian asymptotics (...), there do not seem to be results
 939 yet regarding the correct coverage of posteriors in these models.

940 4.4 Scope for improving genomic prediction

941 pcgen can select traits with direct genetic effects, which are the most relevant in genomic
 942 selection. More generally, the usefulness of structural models for genomic selection de-
 943 pends on whether there are particular interventions of interest ([16], [17]). Informally
 944 speaking, an intervention is an external manipulation that forces some of the traits to
 945 have a particular distribution. For example, with a so-called hard intervention on the j th
 946 trait, \mathbf{Y}_j is forced to a constant level c , e.g. $c = 0$, when \mathbf{Y}_j is the expression of a gene
 947 that is knocked out. The manipulation or truncated factorization theorem [4], [5] can
 948 then predict the joint distribution of the system *after* the intervention:

$$p_{\mathbf{Y}_j:=c}(\mathbf{G}, \mathbf{Y}_{-j}) = p(\mathbf{G}) \prod_{j' \neq j} p(\mathbf{Y}_{j'} | pa(\mathbf{Y}_{j'}), \mathbf{G}_{j'}). \quad (21)$$

949 This is generally different from the distribution

$$p(\mathbf{G}, \mathbf{Y}_{-j} | \mathbf{Y}_j = c), \quad (22)$$

950 obtained from conditioning on $\mathbf{Y}_j = c$, *prior* to the intervention (see e.g. [24]). In other
 951 words, *conditioning* is not the same as *doing* (intervening). An exception occurs however
 952 when we intervene on a root node, in which case (21) and (22) are the same.

953 In absence of interventions on the traits, we can think of genomic prediction in terms
 954 of an intervention on the node genotype. Because the latter is a root node by definition,
 955 standard genomic prediction methods can in principle have optimal performance ([16]).
 956 More specifically, genomic prediction usually involves a regression of a target trait on a
 957 number of markers, having either fixed or random effects. In either case, it is only the
 958 total effect of genotype on the target trait that matters, not through which other traits
 959 this effect passes.

960 (Near) optimal prediction accuracy however requires that the regression model con-
 961 tains the true distribution (or a good approximation), and a sufficiently accurate estimate
 962 of this distribution. We therefore believe that structural models may sometimes be an
 963 appealing alternative, especially if the underlying model is highly nonlinear, or when prior
 964 physiological knowledge can be incorporated. The extent to which this can really improve
 965 accuracy remains to be investigated.

966 4.5 Open questions and extensions

967 Although we have shown the Markov property for our model and studied consistency of
 968 pcgen, there are a number of open questions left for future work. First, the behavior of
 969 our conditional independence tests is not completely understood, and it may be possible
 970 to construct better tests, especially for nonlinear structural models. The recent work
 971 of [50] seems particularly relevant here. The RC-test proposed here for the conditional
 972 independence (A) is exact for $K = ZZ^t$ and certain conditioning sets, but in other cases
 973 is misspecified.

974 A second issue is the consistency of the orientations: while we have shown pcgen’s
 975 consistency in reconstructing the skeleton, we did not address this for the final CPDAG.
 976 This is well known for the PC-algorithm without genetic effects ([5], [32]), but more
 977 difficult to establish here, as the class of CPDAGs needs to be restricted to those without
 978 errors pointing to G . More generally, orientation constraints seem to be of interest for
 979 trait-to-trait relationships as well, e.g. one may require that, *if there is an edge*, the
 980 expression of a gene can only affect a metabolite and not the other way round. To the
 981 best of our knowledge, current methodology and theory has only considered the forced
 982 absence/presence of an edge, leaving the orientation to the algorithm⁶. A final question
 983 for future work is whether Theorems 4 and 6 hold for general conditioning sets.

984 Apart from these open questions, we believe that the idea of explicitly modeling direct
 985 genetic effects can be applied more generally. A first generalization would be to replace

⁶The pcalg-package [28] has the addBgKnowledge option to add orientations (‘background knowledge’) in the estimated CPDAG. This is however only done *after* running PC or a related algorithm, and is only allowed if compatible with the CPDAG.

986 the PC-algorithm with other constraint-based algorithms, in particular FCI and RFCI
987 ([5], [51]). These have the advantage that the causal sufficiency assumption (no latent
988 variables) can be dropped or considerably weakened. Finally, the presence or absence of
989 direct genetic effects may also be incorporated in invariant causal prediction ([42]), or in
990 Bayesian approaches for genetic network reconstruction.

991 **4.6 Author contributions**

992 WK developed the pcgen algorithm. PB developed the pcgen package, based on code
993 written by WK and the EM-algorithm contributed by MR. WK wrote the paper, with
994 input from FvE, DBK, EW, BY, PB and MR. DBK simulated data with APSIM. PB
995 visualized the estimated networks for the rice, maize and APSIM data. WK proved
996 Theorem 1-2 and WK, EW and MM proved Theorems 3-6.

997 **4.7 Acknowledgements**

998 WK was funded by the Learning from Nature project of the Dutch Technology Foundation
999 (STW), which is part of the Netherlands Organisation for Scientific Research (NWO). We
1000 thank Niteen Kadam for providing the rice data, and Xinyou Yin for useful discussions
1001 on the interpretation of the resulting networks. Emilie Millet and François Tardieu are
1002 acknowledged for their interpretation of the maize data.

1003 A pcgen: implementation details

1004 A.1 The PC-stable algorithm

1005 We first state the PC-stable algorithm of [30], which forms the basis of pcgen. As the orig-
 1006 inal PC-algorithm, PC-stable has a skeleton and orientation stage; the former is described
 1007 separately as Algorithm 2 below.

1008 Instead of updating the skeleton directly after each conditional independence test,
 1009 PC-stable only updates the skeleton list-wise, after doing all tests with Y_S of a given size
 1010 $|S| = s$. More specifically, lines 7-9 in Algorithm 2 make an inventory of the current
 1011 adjacency sets, which determines which tests of a given size s are to be conducted. In the
 1012 original PC-algorithm, the skeleton (and hence the adjacency sets) were updated after
 1013 each individual test, introducing an undesirable order-dependence. Since edges are not
 1014 directly removed after finding conditional independence, multiple separation sets may
 1015 be found for a given pair of variables. These may lead to conflicts in the orientation,
 1016 for example when there are conflicting v-structures $Y_j \rightarrow Y_k \leftarrow Y_l$ and $Y_k \rightarrow Y_l \leftarrow Y_m$.
 1017 Whenever possible these conflicts are resolved using the majority rule (line 6 of Algorithm
 1018 1). Unresolved conflicts are represented with an undirected edge⁷. In some cases this may
 1019 lead to partially directed graphs that are not a CPDAG, but a considerable advantage
 1020 of PC-stable is that it can be parallelized. Finally, also the orientation rules R1-R3 in
 1021 Algorithm 1 are applied listwise.

⁷Alternatively, conflicts can be represented with a bi-directed edge (\leftrightarrow), but the arrowheads do not have a causal interpretation

Algorithm 1 The PC-stable algorithm (taken from [30], and adapted to our notation)

- 1: **INPUT:** A set of variables V , and an ordering $order(V)$. Conditional independence information among all variables in V (perfect information in the oracle version of the algorithm; estimated from data in the sample version)
 - 2: **OUTPUT:** a graph \mathcal{G} (in the oracle version \mathcal{G} is always a CPDAG; in the sample version it may be only a PDAG, due to conflicts)
 - 3: For all Y_j, Y_k , find the skeleton \mathcal{C} and collections $Sep(Y_j, Y_k)$ of separating sets. To this end, we use PC-skeleton (Algorithm 2 given below).
 - 4: Determine which unshielded triples in the skeleton \mathcal{C} are unambiguous, and orient them using the separating sets:
 - 5: **for all** pairs of nonadjacent variables Y_j, Y_k with common neighbour Y_l (such that (Y_j, Y_l, Y_k) is unambiguous) **do**
 - 6: **if** Y_l is contained in less than half of the separating sets in $Sep(Y_j, Y_k)$ **then**
 - 7: Replace $Y_j - Y_l - Y_k$ in \mathcal{C} by $Y_j \rightarrow Y_l \leftarrow Y_k$
 - 8: **end if**
 - 9: **end for**
 - 10: In the resulting PDAG, try to orient as many undirected edges as possible by repeated application of the following rules:
 - 11: **R1** Orient $Y_k - Y_l$ into $Y_k \rightarrow Y_l$ whenever there is an arrow $Y_j \rightarrow Y_k$ such that Y_j and Y_l are nonadjacent (otherwise a new v-structure is created).
 - 12: **R2** Orient $Y_j - Y_k$ into $Y_j \rightarrow Y_k$ whenever there is a chain $Y_j \rightarrow Y_l \rightarrow Y_k$ (otherwise a directed cycle is created).
 - 13: **R3** Orient $Y_j - Y_k$ into $Y_j \rightarrow Y_k$ whenever there are two chains $Y_j - Y_l \rightarrow Y_k$ and $Y_j - Y_l \rightarrow Y_k$ such that Y_l and Y_l are nonadjacent (otherwise a new v-structure or a directed cycle is created).
-

Algorithm 2 PC-skeleton (taken from [30], and adapted to our notation)

- 1: **INPUT:** A set of variables (nodes) V , and an ordering $order(V)$. Conditional independence information among all variables in V (perfect information in the oracle version of the algorithm; estimated from data in the sample version)
 - 2: **OUTPUT:** Estimated skeleton \mathcal{C} , collections $Sep(Y_j, Y_k)$ of separating sets (only needed when directing the skeleton afterwards)
 - 3: Form the complete undirected graph $\tilde{\mathcal{C}}$ on the set of variables V .
 - 4: Let $s = -1$; $\mathcal{C} = \tilde{\mathcal{C}}$
 - 5: **repeat**
 - 6: Let $s = s + 1$
 - 7: **for all** variables Y_j in \mathcal{C} **do**
 - 8: Let $a(Y_j) = adj(\mathcal{C}, Y_j)$
 - 9: **end for**
 - 10: **repeat**
 - 11: Select a (new) ordered pair of variables (Y_j, Y_k) that are adjacent in \mathcal{C} and satisfy $|a(Y_j) \setminus \{Y_k\}| \geq s$, using $order(V)$
 - 12: **repeat**
 - 13: Choose a (new) set $S \subseteq a(Y_j) \setminus \{Y_k\}$ with $|S| = s$, using $order(V)$;
 - 14: **if** Y_j and Y_k are conditionally independent given S **then**
 - 15: Delete the edge $Y_j - Y_k$ from \mathcal{C}
 - 16: Save S in $Sep(Y_j, Y_k)$ and $Sep(Y_k, Y_j)$
 - 17: **end if**
 - 18: **until** edge $Y_j - Y_k$ is deleted from \mathcal{C} , or all $S \subseteq a(Y_j) \setminus \{Y_k\}$ with $|S| = s$ have been considered
 - 19: **until** all ordered pairs of adjacent variables (Y_j, Y_k) in \mathcal{C} with $|a(Y_j) \setminus \{Y_k\}| \geq s$ have been tested for conditional independence
 - 20: **until** all pairs of adjacent variables (Y_j, Y_k) in \mathcal{C} satisfy $|a(Y_j) \setminus \{Y_k\}| < s$.
-

1022 A.2 Modified orientation rules

1023 pcgen follows the usual orientation rules of the PC-stable algorithm (lines 11-13 in Algo-
 1024 rithm 1), except for the following modifications, which are required to avoid arrows point-
 1025 ing into G . Note that Algorithm 1 is written in generic notation with nodes Y_1, \dots, Y_p ;
 1026 in pcgen we have Y_1, \dots, Y_{p+1} , corresponding to G, Y_1, \dots, Y_p .

- 1027 • in line 5 in Algorithm 1, we skip those triples where Y_k turns out to be G
- 1028 • after line 9, we orient all remaining undirected edges $G - Y_j$ as $G \rightarrow Y_j$.

1029 These changes appear to be necessary, as edges $G \leftarrow Y_j$ cannot be avoided with the
 1030 `fixedEdges` argument in the `pc`-function of the `pcalg`-package ([28]), where one can only

1031 enforce the presence of an edge in the skeleton, but not its orientation. pcalg also has the
 1032 addBgKnowledge option to add orientations ('background knowledge') in the estimated
 1033 CPDAG, but this is only done *after* running the PC-algorithm, and is only allowed if
 1034 compatible with the CPDAG. Here we intend to always enforce the orientation $G \rightarrow Y_j$,
 1035 and include it in the causal inference algorithm.

1036 A.3 The RC-test for $\mathbf{Y}_j \perp\!\!\!\perp \mathbf{Y}_k | \mathbf{G}, \mathbf{Y}_S$

1037 In the RC-test for $\mathbf{Y}_j \perp\!\!\!\perp \mathbf{Y}_k | \mathbf{G}, \mathbf{Y}_S$, the null-hypothesis is that the residual covariance is
 1038 zero, where the residual covariance is the off-diagonal element of $V_G(jk|S)$, in equations
 1039 (11) and (13) in the main text (section 2.2.2). The residual likelihood ratio test (RLRT)
 1040 statistic for this hypothesis is defined as twice the difference in residual log-likelihood
 1041 between the full and reduced model. As the null-hypothesis is not on the boundary of
 1042 the parameter space, the distribution of the RLRT is approximately chi-square with 1
 1043 degree of freedom. REML-estimates for the full and reduced model are obtained using an
 1044 EM-algorithm [52]. To improve efficiency, we use the following computational shortcuts:

- 1045 1. For the reduced model, we compute starting values for the genetic (co) variances and
 1046 residual variances, using multivariate analysis of variance (MANOVA), and then fit
 1047 the model using the EM-algorithm described before.
- 1048 2. For the full model, we take as starting values the estimates found for the reduced
 1049 model. At each iteration of the EM-algorithm, we compute a preliminary RLRT
 1050 p-value on the basis of the current restricted log-likelihood and the restricted log-
 1051 likelihood of the reduced model, and stop the EM-algorithm if the p-value is below
 1052 the significance threshold. This is possible because the EM-algorithm always in-
 1053 creases the likelihood at every iteration, while pcalg only requires an accept/reject
 1054 decision. Of course, if one wants to know the exact p-value (e.g. to obtain a rough
 1055 indication of the strength of the causal relationships), the EM-algorithm needs to
 1056 be run until convergence.
- 1057 3. We set a maximum number of 50 EM-iterations for the full model and 5 for the re-
 1058 duced model. Given the good starting values this is often sufficient, but occasionally
 1059 EM would otherwise take very many iterations until convergence. Usually, in these
 1060 cases, the RLRT statistic obtained with an unrestricted number of EM-iterations is
 1061 not significant, meaning that stopping EM earlier rarely affects the outcome of the
 1062 RLRT.

1063 A.4 Testing the genetic covariance

1064 Assuming that the joint distribution of $\text{vec}([\mathbf{Y}_j \mathbf{Y}_k])$ has covariance $V_G(jk) \otimes K + V_E(jk) \otimes$
 1065 I_n (i.e. (13) in the main text, with empty conditioning set), it is sometimes of interest

1066 to test for absence of genetic covariance, i.e., whether the off-diagonal element of $V_G(jk)$
1067 is zero. Although not part of `pcgen`, this can be tested similar to the RC-test described
1068 above.

1069 Again we define a LRT statistic as twice the difference in residual log-likelihood be-
1070 tween the full and reduced model, where the latter is restricted to have diagonal $V_G(jk)$.
1071 As before the distribution of this LRT is approximately chi-square with 1 degree of free-
1072 dom, and REML-estimates for the full and reduced model are obtained with the same
1073 EM-algorithm used earlier. In the `pcgen`-package this is implemented in the `gencorTest`
1074 function.

1075 B Simulations

1076 B.1 Simulation setup

1077 We simulate data from model (5) using the following steps, closely resembling the simu-
1078 lations of [32]:

- 1079 • Given the required number of traits (p) and sparseness of the graph (defined by
1080 the parameter p_t below), we first generate the $p \times p$ matrix Λ (see equation (5)),
1081 which determines the structural relations among the traits. Λ is simulated using
1082 the `randomDAG` function from the R-package `pcalg` ([28]), where edges (i.e. the
1083 nonzero elements of Λ) occur with probability $p_t \in (0, 1)$ (for details see the `pcalg`-
1084 documentation and [32], p. 621). The expected number of neighbors of each node
1085 is then $p_t(p - 1)$. The values of the nonzero coefficients are drawn independently
1086 from the uniform distribution on $[0.5, 1]$ and then given a random sign.
- 1087 • The DAG defined by Λ is now extended with a genetic node G . For a proportion of
1088 $p_g \in (0, 1)$ of the traits, we add an edge $G \rightarrow Y_j$. The subset of traits D for which
1089 there is a direct genetic effect then contains $p \cdot p_g$ traits. These are always chosen
1090 to be the traits of highest topological order (in the initial DAG defined by Λ). For
1091 example, if $p = 4$ and $p_g = 0.5$, and the initial DAG is $Y_1 \rightarrow Y_2 \rightarrow Y_3 \rightarrow Y_4$, D will
1092 consist of Y_1 and Y_2 .
- 1093 • Next, the corresponding genetic variances and covariances in Σ_G are simulated as fol-
1094 lows. The genetic variances are drawn independently from a uniform distribution on
1095 $[1, 2]$ and random covariances are introduced through random eigenvectors as in [53],
1096 using the `genPositiveDefMat` function from the R-package `clusterGeneration`.
- 1097 • Given the relatedness matrix (K) and the required numbers of genotypes (n) and
1098 replicates (r), the direct genetic effects (\mathbf{G}) are drawn from the matrix-variate nor-
1099 mal distribution with column covariance ZKZ^t and row-covariance Σ_G .
- 1100 • Similarly, the residual effects (\mathbf{E}) are drawn from the matrix-variate normal distri-
1101 bution with column covariance I_{nr} and row-covariance I_p . Although the residual
1102 variance is 1 for all traits, the heritability of the traits still varies, as the variances
1103 of the direct genetic effects are between 1 and 2. Traits without a direct genetic
1104 effect typically have heritability below 0.5.
- 1105 • Given the matrices \mathbf{G} , \mathbf{E} and Λ obtained in the previous steps, \mathbf{Y} is computed using
1106 (5). This is done recursively, following the topological ordering of the DAG ([32],
1107 p. 621).

1108 B.2 Performance criteria

1109 Following again [32] we compare the true (simulated) CPDAGs and the estimated CPDAGs.
 1110 Instead of considering the complete CPDAG, we evaluate separately the subgraph defined
 1111 by Λ (relations among traits) and the direct genetic effects (the subgraph with edges
 1112 $G_j \rightarrow Y_j$, whenever $j \in D$). For both subgraphs we consider the following criteria, again
 1113 following [32]:

- 1114 • True Positive Rate (TPR): the number of correct edges (in the estimated skeleton)
 1115 divided by the total number of true edges (in the true skeleton).
- 1116 • False Positive Rate (FPR): the number of incorrect edges (in the estimated skeleton)
 1117 divided by the total number of true gaps (in the true skeleton).
- 1118 • True Discovery Rate (TDR): the number of edges in the estimated graph that are
 1119 correct (i.e. exist in the true skeleton) divided by the total number of edges in the
 1120 estimated graph.
- 1121 • Structural Hamming Distance (SHD): the number of edge deletions, additions and
 1122 flips required to transform the estimated CPDAG into the true (simulated) CPDAG.
 1123 See also [54].

1124 All criteria were computed using functions from the `pcalg`-package (TPR and FPR using
 1125 the `compareGraphs` function, and the SHD using the function `shd`).

1126 B.3 Simulation results for alternative methods

1127 B.4 Simulation with APSIM

1128 We used the crop-growth model APSIM to simulate 12 traits ($\mathbf{Y}_1, \dots, \mathbf{Y}_{12}$) for an exist-
 1129 ing population of 199 wheat genotypes, characterized with 3,035 SNPs with minor allele
 1130 frequency larger than 0.05 (details in [55]). APSIM simulations were carried out in Emer-
 1131 ald, during 1993, which corresponds to a severe drought environment. Simulation settings
 1132 were the same as in [56] and genotype-specific parameters had the ranges specified in [55].
 1133 The wheat panel was assumed to segregate for 7 of the APSIM parameters, which we refer
 1134 to as the primary traits. These are the only ones for which there are direct genetic effects.
 1135 The 7 primary traits ($\mathbf{Y}_1, \dots, \mathbf{Y}_7$) are a subset of those used in [55], and were chosen
 1136 because they have an important impact on grain yield, as shown by global sensitivity
 1137 analysis [56]. Apart from the primary traits there are 4 intermediate traits, each of them
 1138 depending on some of the primary traits, and sometimes some of the other intermediate
 1139 traits. The final trait is yield, which depends on 3 of the intermediate traits. Acronyms
 1140 for all 12 traits are given in Table 3. For each genotype three replicates were simulated.

1141 The direct genetic effects on the 7 primary traits were simulated as the sum of 300
 1142 additive QTL-effects. Different samples of 300 SNPs were used for each trait, and each

	\mathcal{G}_y				$G \rightarrow Y_j$		
	TPR	FPR	TDR	SHD	TPR	FPR	TDR
Scenario 1 ($p = 4$)	$(p_t = 1/3)$				$(p_g = 0.5)$		
pcgen-RC	0.607	0.012	0.960	0.568	0.986	0.020	0.994
pcgen-RG-uni	0.607	0.036	0.906	0.670	0.984	0.022	0.994
pcgen-screening	0.603	0.008	0.982	0.548	0.986	0.020	0.994
pcRes-multi-R	0.605	0.161	0.626	2.052			
pcRes-uni-R	0.607	0.036	0.906	1.362			
pcRes-multi-RK	0.570	0.192	0.624	2.060			
pcRes-uni-RK	0.608	0.029	0.920	1.330			
pcRes-multi-K	0.465	0.432	0.388	3.212			
pcRes-uni-K	0.595	0.088	0.747	1.668			
pc-GBLUP	0.536	0.490	0.375	3.472			
Scenario 2 ($p = 4$)	$(p_t = 0.5)$				$(p_g = 0.5)$		
pcgen-RC	0.814	0.042	0.980	1.236	0.966	0.041	0.993
pcgen-RG-uni	0.820	0.074	0.951	1.320	0.965	0.043	0.992
pcgen-screening	0.805	0.034	0.990	1.238	0.966	0.042	0.992
pcRes-multi-R	0.824	0.149	0.835	2.640			
pcRes-uni-R	0.820	0.074	0.951	2.274			
pcRes-multi-RK	0.786	0.242	0.787	2.892			
pcRes-uni-RK	0.821	0.063	0.957	2.230			
pcRes-multi-K	0.666	0.466	0.603	3.778			
pcRes-uni-K	0.786	0.098	0.893	2.522			
pc-GBLUP	0.752	0.560	0.586	4.124			

Table 2. Performance of pcgen and residuals-based approaches, averaged over 500 simulated data-sets, for scenarios 1 and 2.

1143 effect was sampled from a trait-specific Gamma distribution. The shape and rate of this
1144 distribution were obtained by fitting a Gamma distribution to empirical additive effects
1145 estimated in a GWAS analysis of real phenotypes observed for this population in the
1146 Australian wheat belt. For the phenology-related traits SV and SP we set $k = 0.7$ and
1147 $b = 13.6$, and $k = 1.3$ and $b = 13.6$ for the other primary traits. We then added Gaussian
1148 noise, to get a heritability of 0.9 for all primary traits.

1149 The secondary traits ($\mathbf{Y}_8, \dots, \mathbf{Y}_{11}$) and yield (\mathbf{Y}_{12}) were simulated by running a
1150 dynamic model from time zero to time T , the time-point at which all traits are observed:

$$\mathbf{Y}_j(T) = \int_0^T f_z(pa(\mathbf{Y}_j(t)))dt, \quad (23)$$

1151 where $pa(\mathbf{Y}_j(t))$ are the values of the 'parent traits' at time-point t , and z represents a
1152 set of fixed parameters, specific for the environment under consideration. The form of f_z

trait name / category	acronym
Primary traits	
Radiation use efficiency	RUE
Number of grains per gram of stem at flowering	NGF
Maximum grain size	MGS
Lower limit for water uptake	LL
Sensitivity to photoperiod	SP
Sensitivity to vernalization	SV
Thermal time required to reach floral initiation	TFI
Secondary traits and yield	
flowering time	FT
grain weight	GW
grain number	GN
biomass	BM
yield	Y

Table 3. Acronyms for 12 traits simulated using APSIM.

1153 is detailed in The sets of parental traits stay the same over time, and therefore define
1154 the summary graph given in Figure 3 (main text).

1155 **C Rice trait acronyms and networks for different sig-**
1156 **nificance thresholds**

Traits	Trait	acronym	Unit
(A) Shoot morphological traits			
	Plant height	PHT	cm
	Tiller number	TN	$plant^{-1}$
	Total leaf area	TLA	$m^2 plant^{-1}$
	Specific leaf area	SLA	$m^2 g^{-1}$
(B) Physiological traits			
	Cumulative water transpiration	CWT	$kg plant^{-1}$
	Water use efficiency	WUE	$g kg^{-1}$
(C) Root anatomical traits			
	Total root length	TRL	$m plant^{-1}$
	Root length (RL) with diameter class		
	RL_0-0.5	RL005	$m plant^{-1}$
	RL_0.5-1.0	RL0510	$m plant^{-1}$
	RL_1.0-1.5	RL1015	$m plant^{-1}$
	RL_1.5-2.0	RL1520	$m plant^{-1}$
	RL_2.0-2.5	RL2025	$m plant^{-1}$
	RL_2.5-3.0	RL2530	$m plant^{-1}$
	RL_3.0-3.5	RL3035	$m plant^{-1}$
	RL_3.5	RL35	$m plant^{-1}$
	Maximum root length	MRL	cm
	Surface area	SA	$cm^2 plant^{-1}$
	Root volume	RV	$cm^3 plant^{-1}$
	Average root thickness	ART	mm
	Specific root length	SRL	$m g^{-1}$
	Total root weight density	TRWD	$g cm^{-3}$
	Root length per unit leaf area	RLLA	$m m^{-2}$
(D) Root anatomical traits			
	Root diameter	RD	μm
	Cortex diameter	CD	μm
	Stele diameter	SD	μm
	Late metaxylem	LMXD	μm
	Late metaxylem	LMXN	μm
	Stele diameter in proportion of root diameter	SD:RD	%
(E) Dry matter traits			
	Leaf weight	LW	$g plant^{-1}$
	Stem weight	SW	$g plant^{-1}$
	Root weight	RW	$g plant^{-1}$
	Total weight	TW	$g plant^{-1}$
	Root:shoot ratio	RS	–
	Leaf weight ratio	LWR	–
	Stem weight ratio	SWR	–

Table 4. Trait acronyms used in Figure 5; table taken from [38]. Leaf weight (LW) was removed prior to our analysis, as it appeared to be an exact linear combination of 3 of the other traits (TW - RW - SW, i.e. total weight minus root weight minus shoot weight).

1157 **D Overview of notation, acronyms and commands in**
 1158 **pcgen**

Symbol	Meaning / Category	Introduced in section
	Other (exogenous) fixed effects in SEM	
B		
X		
	Path coefficients in linear (G)SEM	
Λ		
Γ		
γ_j		
	Graphs	
\mathcal{G}		
G		
Y_j		
\mathcal{G}_y		
d-separation		
$dsep(\mathcal{G})$		
	Genetic effects	
\mathbf{G}		
\mathbf{g}_i		
\mathbf{G}_j		
\mathbf{U}		
\mathbf{u}_i		
\mathbf{U}_j		
	Residual effects	
\mathbf{E}		
\mathbf{e}_i		
\mathbf{E}_j		
	Traits	
	Genetic relatedness	

1160 $\Sigma_G, \sigma_{G,j}^2, \Sigma_E, \mathbf{n}, \mathbf{m}, \mathbf{r}, \mathbf{p}, \mathbf{q}, \mu_{j|S}$
 1161 GBLUP, RC-test, MTM, SEM, GSEM, DAG, CPDAG

Abbreviation	Description	Target
pcgen-RC	pcgen with the RC-test	\mathcal{G}
pcgen-RG-uni	pcgen with the RG-test, based on univariate GBLUP	\mathcal{G}
pcgen-RG-multi	pcgen with the RG-test, based on multivariate GBLUP	\mathcal{G}
pcgen-screening	pcgen with the RC-test, starting with the skeleton obtained from pcRes-uni-R (see section 2.6)	\mathcal{G}
pcRes-uni-R	pcRes based on univariate GBLUP, using replicates	\mathcal{G}_y
pcRes-uni-RK	pcRes based on univariate GBLUP, using replicates + GRM	\mathcal{G}_y
pcRes-uni-K	pcRes based on univariate GBLUP, using GRM	\mathcal{G}_y
pcRes-multi-R	pcRes based on multivariate GBLUP, using replicates	\mathcal{G}_y
pcRes-multi-RK	pcRes based on multivariate GBLUP, using replicates + GRM	\mathcal{G}_y
pcRes-multi-K	pcRes based on multivariate GBLUP, using GRM	\mathcal{G}_y
pc-GBLUP	pc(stable) applied to multivariate GBLUP, similar to [13]	\mathcal{G}_y

Table 5. Overview of the algorithms available in the pcgen package, for reconstructing either \mathcal{G} (the complete graph) or \mathcal{G}_y (the subgraph of trait-to-trait relations). The required commands in the pcgen-package are given in Table 6 in Appendix D.

E Faithfulness, conditional distributions and proofs of Theorems 1-6

E.1 Overview of graph theoretic definitions

The following definitions can be found in a large number of books and articles on graph theory, graphical models and causal inference; see for example [57], [4], [5] and [32].

- Given different nodes Y_j and Y_k , a *path* from Y_j to Y_k is a sequence of edges connecting Y_j and Y_k . When all edges are directed and pointing towards Y_k , we have a *directed path*. An *undirected path* or *non-directed path* is a path that is not directed.
- A *cycle* is a path from Y_j to Y_k with an additional edge between Y_j and Y_k . A *directed cycle* is a directed path from Y_j to Y_k together with a directed edge $k \rightarrow j$.
- A *directed acyclic graph* (DAG) is a directed graph without cycles. When a graph underlying a SEM is a DAG, the SEM is said to be recursive.
- $pa(j)$ is the set of nodes Y_k for which is a directed edge $k \rightarrow j$; in this case Y_j is a *child* of Y_k and Y_k is a *parent* of Y_j . The nodes Y_j and Y_k are *adjacent* if there is an edge between them.
- If in a DAG \mathcal{G} there is a directed path from Y_j to Y_k , Y_j is an ancestor of Y_k , and Y_k is a descendant of Y_j .

Abbreviation	Commands in pcgen-package
pcgen-RC	pcgen(d,..., use.res=F)
pcgen-RG-uni	$C \leftarrow \text{cor}(\text{getResiduals}(d,\dots, \text{cov.method}='uni'))$ pcgen(d,..., use.res=T, res.cor = C) equivalently: pcgenFast(..., use.res=T)
pcgen-RG-multi	$C \leftarrow \text{cor}(\text{getResiduals}(d,\dots, \text{cov.method}='us'))$ pcgen(d,..., use.res=T, res.cor = C)
pcgen-screening	pcgenFast(d,..., use.res=F)
pcRes-uni-R	pcRes(d,..., cov.method='uni', use.GBLUP=F)
pcRes-uni-RK	pcRes(d,..., cov.method='uni', use.GBLUP=F, K=A)
pcRes-uni-K	pcRes(m,..., cov.method='uni', use.GBLUP=F, K=A)
pcRes-multi-R	pcRes(d,..., cov.method='us', use.GBLUP=F)
pcRes-multi-RK	pcRes(d,..., cov.method='us', use.GBLUP=F, K=A)
pcRes-multi-K	pcRes(m,..., cov.method='us', use.GBLUP=F, K=A)
pc-GBLUP	pcRes(m,..., cov.method='us', use.GBLUP=T, K=A)

Table 6. R-commands needed to run the different algorithms, with the package pcgen. T stands for TRUE, F for FALSE. The first argument is the required phenotypic data-frame (suffStat = d (replicates) or suffStat = m (genotypic means)). The dots represent generic arguments (e.g. alpha and m.max, which define the significance threshold and the maximum size of the conditioning sets). cov.method determines whether univariate (uni) or multivariate (us) GBLUP is to be used ('us' stands for unstructured, as opposed to e.g. factor analytic models, which have not yet been implemented). All algorithms involving GBLUP use the residuals (use.GBLUP = F), except the genomic network similar to [13] (pc-GBLUP, with use.GBLUP = T). Finally, A is a genetic relatedness matrix, which can be included by putting K=A; otherwise the default is used (K = NULL, in which case replicates are required).

- 1179 • If, for a given path, two directed edges point into the same node, the latter is a
1180 *collider*. For example, given the DAG $A \rightarrow C \leftarrow B$, C is a collider on the (only)
1181 path between A and B . In all other cases ($A \leftarrow C \rightarrow B$, $A \rightarrow C \rightarrow B$ and
1182 $A \leftarrow C \leftarrow B$), C is a *non-collider*. Several different paths can pass through a node,
1183 and being a (non-)collider is always relative to the path.

- 1184 • In a DAG, a *v-structure* or *immorality* is a collection of three nodes (say A , B and
1185 C), such that there are directed edges $A \rightarrow B$ and $C \rightarrow B$ but no edge between A
1186 and C . In this case B is an *unshielded* collider; otherwise it is a *shielded* collider.
1187 Similarly, in an undirected graph, A , B and C form an *unshielded triple* if there are
1188 edges $A - B$ and $C - B$ but no edge $A - C$.

- 1189 • The *skeleton* of a (partially) directed graph is the undirected graph obtained after
1190 removing all arrowheads.

- 1191 • Given a directed graph \mathcal{G} , two nodes A and B , and a (possibly empty) subset of
1192 nodes S not containing A and B , a path between A and B is *blocked* by S if at
1193 least one of the following two conditions holds: (i) there exists a collider on the
1194 path which is not in S , and also none of its descendants are in S . (ii) there exists a
1195 non-collider on the path that is in S .

- 1196 • Nodes A and B are *d-separated* by a set S if S blocks all paths from A to B .

- 1197 • Given disjoint sets U , V and S (U and V should be non-empty), U and V are
1198 *d-separated* by S if S blocks all paths from Y_j to Y_k , for all nodes $j \in U$ and $k \in V$.

- 1199 • Two DAGs are *equivalent* if they have the same skeleton and the same v-structures.

- 1200 • An equivalence class of DAGs is a set containing all DAGs that are equivalent to
1201 one another. Any DAG in the class can be used to represent the class; however an
1202 equivalence class can also be represented by a *completed partially directed acyclic*
1203 *graph* (CPDAG). Following the formulation of [32], a *partially directed acyclic graph*
1204 (PDAG) is 'a graph where some edges are directed and some are undirected and one
1205 cannot trace a cycle by following the direction of directed edges and any direction
1206 for undirected edges'. A PDAG is a CPDAG if (a) every directed edge in the PDAG
1207 exists in all DAGs in the equivalence class it represents (b) for every undirected edge
1208 $j - k$ in the PDAG, the equivalence class contains at least one DAG with $j \rightarrow k$
1209 and at least one with with $k \rightarrow j$. Chickering [58] showed that CPDAGs represent
1210 equivalence classes uniquely. For example, given a skeleton $A - B - C$, there is
1211 one equivalence class containing the three DAGs $A \rightarrow B \rightarrow C$, $C \rightarrow B \rightarrow A$ and
1212 $A \leftarrow B \rightarrow C$, and one equivalence class with only one DAG ($A \rightarrow B \leftarrow C$).

1213 E.2 The matrix Γ expressed as a function of path coefficients

1214 Let \mathcal{G}_y denote the DAG over the nodes Y_1, \dots, Y_p , with edges defined by Λ . For each
 1215 $j \in \{1, \dots, p\}$, let U_j denote the union of the set $\{Y_j\}$ and the set of root traits (i.e.
 1216 those without parents in \mathcal{G}_y) for which there is a directed path towards Y_j . For all
 1217 $j, k \in \{1, \dots, p\}$, let Π_{jk} denote the set of all directed paths from Y_j to Y_k . For $k = j$,
 1218 Π_{jj} contains only the empty path from Y_j to itself. For any directed path π from Y_j to
 1219 Y_k , let $L(\pi)$ denote the product of the corresponding path coefficients as given by Λ ; for
 1220 the empty path we define $L(\pi) = 1$.

1221 Using these definitions, we can decompose the variance of a trait into contributions
 1222 from different ancestors, as well as its own error variance. To this end, we follow [5] and
 1223 define the $p \times 1$ column vector γ_j with elements ($l = 1, \dots, p$)

$$\begin{aligned} \gamma_{j,l} &= \sum_{\pi \in \Pi_{lj}} L(\pi) \quad \text{if } Y_l \in U_j \\ &= 0 \quad \text{otherwise.} \end{aligned} \quad (24)$$

1224 E.3 The covariance between \mathbf{Y}_j and \mathbf{Y}_k as function of path co- 1225 efficient

1226 Since $\mathbf{Y}_j = \mathbf{X}B\gamma_j + \mathbf{G}\gamma_j + \mathbf{E}\gamma_j$ (equation (10) in the main text), the covariance between
 1227 the $n \times 1$ vectors \mathbf{Y}_j and \mathbf{Y}_k can be written in terms of γ_j and γ_k :

$$\text{Cov}(\mathbf{Y}_j, \mathbf{Y}_k^t) = E[(\mathbf{Y}_j - \mathbf{X}B\gamma_j)(\mathbf{Y}_k - \mathbf{X}B\gamma_k)^t] = (\gamma_j^t \Sigma_G \gamma_k) K + (\gamma_j^t \Sigma_E \gamma_k) I_n, \quad (25)$$

1228 for all $j, k \in \{1, \dots, p\}$. Consequently, we can express the genetic and residual covariance
 1229 between traits in terms of quadratic forms, involving Σ_G , Σ_E and the path coefficients.

As a special case of (25), it follows that without random genetic effects,

$$\text{Cov}(\mathbf{Y}[i, j], \mathbf{Y}[i, k]) = \gamma_j^t \Sigma_E \gamma_k$$

1230 is the covariance between the j th and k th trait, for each individual i . See also [5] (Lemma
 1231 3.1.6), or [59] (Appendix 2). Using standard expressions for multivariate Gaussian distri-
 1232 butions, this implies that

$$\text{Cov}(\mathbf{Y}[i, j], \mathbf{Y}[i, k] \mid \mathbf{Y}[i, S]) = \gamma_j^t \Sigma_E \gamma_k - (\gamma_j^t \Sigma_E \Gamma_S)(\Gamma_S^t \Sigma_E \Gamma_S)^{-1}(\Gamma_S^t \Sigma_E \gamma_k). \quad (26)$$

1233 E.4 The path coefficients condition

1234 It is well known that faithfulness is violated when contributions from different paths cancel
 1235 out. For example, in the SEM defined by $Y_1 \rightarrow Y_2$, $Y_1 \rightarrow Y_3$ and $Y_2 \rightarrow Y_3$, with respective
 1236 path coefficients 1, 1 and -1 , Y_1 and Y_3 are marginally independent but not d-separated.
 1237 Conversely, when faithfulness holds, we know that such cancellations cannot occur, and
 1238 that the sum in (24) is never zero, i.e. $\gamma_{j,l} = 0$ only for $Y_l \notin U_j$. We will refer to this as
 1239 the **path coefficients condition**.

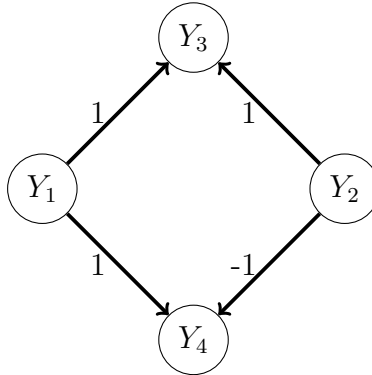


Figure 6. An example of a SEM where faithfulness does not hold, because the contributions to the covariance from the treks $Y_3 \leftarrow Y_1 \rightarrow Y_4$ and $Y_3 \leftarrow Y_2 \rightarrow Y_4$ cancel out. If Y_1 and Y_2 are Gaussian with equal (error) variances, it follows that $Cov(Y_3, Y_4) = Cov(Y_1 + Y_2 + E_3, Y_1 - Y_2 + E_4) = Cov(Y_1 + Y_2, Y_1 - Y_2) = 0$. Consequently, Y_3 and Y_4 are marginally independent, but not d-separated by the empty set.

1240 E.5 The path coefficients condition and faithfulness

1241 The path coefficients condition is a necessary but not a sufficient condition for faithfulness.
 1242 First, faithfulness can also be violated when contributions from different paths cancel out
 1243 when summing over a subset of all directed paths; see Example 2.10 in [60]. Second, it
 1244 is not only the contributions of directed paths that should not cancel out, but also those
 1245 of treks. A trek between Y_j and Y_k is any path between these nodes without a collider
 1246 ([5]). Every trek consists of 2 directed paths, starting at the *source* of the trek, and going
 1247 towards Y_j and Y_k . One of these can be the empty path; hence each directed path is also
 1248 a trek. Figure 6 provides an example where contributions from different treks cancel out,
 1249 leading to non-faithfulness.

1250 Another necessary condition for faithfulness is that all error variances are strictly
 1251 positive. Figure 7 provides an example of non-faithfulness due to a zero error variance.
 1252 An extended version of the path coefficients condition (involving sums over subset of
 1253 treks) together with strictly positive error variances may be sufficient for faithfulness, but
 1254 we could not find such a result in the literature. However, from (26) it follows that for
 1255 Gaussian linear SEM, faithfulness is equivalent with

$$\gamma_j^t \Sigma_E \gamma_k - (\gamma_j^t \Sigma_E \Gamma_S)(\Gamma_S^t \Sigma_E \Gamma_S)^{-1}(\Gamma_S^t \Sigma_E \gamma_k) = 0 \implies Y_j \text{ and } Y_k \text{ d-separated by } Y_S. \quad (27)$$

1256 E.6 Conditional means and covariances

1257 Using the notation $[, S]$ to select the columns corresponding to S , and $[S_1, S_2]$ to select
 1258 both rows and columns, it follows from (8) that $\mathbf{Y}_j | \mathbf{Y}_S = \tilde{y}_S$ is multivariate normal with

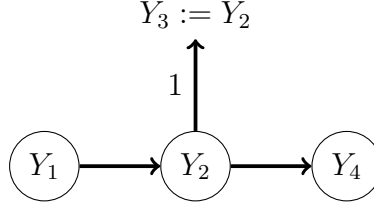


Figure 7. An example of a SEM where faithfulness does not hold, because the variance of the error E_2 is zero. The random variables Y_4 and Y_1 are conditionally independent given Y_3 , but in the graph, the nodes Y_4 and Y_1 are not d-separated by Y_3 .

1259 mean and covariance

$$\begin{aligned}\mu_{j|S} &= (\mathbf{X}B\Gamma)[, j] + \Sigma_{j,S}\Sigma_S^{-1}\text{vec}(\tilde{y}_S - (\mathbf{X}B\Gamma)[, S]) \\ &= \mathbf{X}B\gamma_j + \Sigma_{j,S}\Sigma_S^{-1}\text{vec}(\tilde{y}_S - \mathbf{X}B\Gamma_S),\end{aligned}\quad (28)$$

1260

$$\Sigma_{j|S} = \Sigma_j - \Sigma_{j,S}\Sigma_S^{-1}\Sigma_{j,S}^t, \quad (29)$$

1261 where

$$\Sigma_{j,S} = (\Gamma^t\Sigma_G\Gamma)[j, S] \otimes K + (\Gamma^t\Sigma_E\Gamma)[j, S] \otimes I_n \quad (30)$$

$$= (\gamma_j^t\Sigma_G\Gamma_S) \otimes K + (\gamma_j^t\Sigma_E\Gamma_S) \otimes I_n, \quad (31)$$

$$\Sigma_S = (\Gamma_S^t\Sigma_G\Gamma_S) \otimes K + (\Gamma_S^t\Sigma_E\Gamma_S) \otimes I_n, \quad (32)$$

$$\Sigma_j = (\Gamma^t\Sigma_G\Gamma)[j, j]K + (\Gamma^t\Sigma_E\Gamma)[j, j]I_n = (\gamma_j^t\Sigma_G\gamma_j)K + (\gamma_j^t\Sigma_E\gamma_j)I_n. \quad (33)$$

1262 The matrices Σ_j , Σ_S and $\Sigma_{j,S}$ are the variance-covariance matrix of respectively $\text{vec}(\mathbf{Y}_j) =$
1263 \mathbf{Y}_j and $\text{vec}(\mathbf{Y}_S)$, and the covariance between \mathbf{Y}_j and $\text{vec}(\mathbf{Y}_S)$.

1264 From equation (8) we also obtain the conditional distribution

$$\begin{aligned}\text{vec}([\mathbf{Y}_j \ \mathbf{Y}_k]) \mid \mathbf{Y}_S = \tilde{y}_S &\sim N\left(\begin{pmatrix} \mu_{j|S} \\ \mu_{k|S} \end{pmatrix}, \Sigma_{jk|S}\right) \\ &= N\left(\begin{pmatrix} \mu_{j|S} \\ \mu_{k|S} \end{pmatrix}, \Sigma_{jk} - \Sigma_{jk,S}\Sigma_S^{-1}\Sigma_{jk,S}^t\right),\end{aligned}\quad (34)$$

1265 where $\mu_{j|S}$ and $\mu_{k|S}$ are as in equation (28), and Σ_{jk} is the $2n \times 2n$ block matrix with
1266 diagonal blocks Σ_j and Σ_k (defined as in (33)), and off-diagonal blocks $(\gamma_j^t\Sigma_G\gamma_k)K +$
1267 $(\gamma_j^t\Sigma_E\gamma_k)I_n$. Similarly, given the $p \times 2$ matrix Γ_{jk} with columns γ_j and γ_k , it follows that

$$\Sigma_{jk,S} = (\Gamma_{jk}^t\Sigma_G\Gamma_S) \otimes K + (\Gamma_{jk}^t\Sigma_E\Gamma_S) \otimes I_n$$

1268 is the $2n \times |S|n$ covariance between $\text{vec}([\mathbf{Y}_j \ \mathbf{Y}_k])$ and $\text{vec}(\mathbf{Y}_S)$.

1269 E.7 Covariance structure of the conditional distributions

1270 When $K = ZZ^t$ is block-diagonal, with $r \times r$ blocks of ones on the diagonal, then for any
 1271 nonnegative constants c and d ,

$$(cI_n + dK)^{-1} = c^{-1}I_n - \frac{1}{c^2(1/d + r/c)}K.$$

1272 Hence, the inverse of $(cI_n + dK)$ is again a linear combination of I_n and K . This follows
 1273 from the Woodbury identity ([61] and [62])

$$(A + CBC^t)^{-1} = A^{-1} - A^{-1}C(B^{-1} + C^tA^{-1}C)^{-1}C^tA^{-1}, \quad (35)$$

1274 with $A = cI_n$, $B = dI_m$ and $C = Z$. In addition we have $Z^tZ = rI_m$, and therefore
 1275 $K^2 = rK$. Consequently, any product of matrices of the form $(cI_n + dK)$ or their inverse
 1276 is a linear combination of I_n and K .

From this it follows that when $K = ZZ^t$, $\Sigma_{j|S}$ in (29) is of the form $\sigma_G^2(j|S)K + \sigma_E^2(j|S)I_n$. Similarly, it follows that $\Sigma_{jk|S}$ in (11) is of the form

$$V_G(jk|S) \otimes K + V_E(jk|S) \otimes I_n,$$

1277 for some 2×2 matrices $V_G(jk|S)$ and $V_E(jk|S)$.

1278 E.8 Proof of Theorem 1

1279 Pearl ([4], p. 51) showed that under quite general assumptions, structural equation
 1280 models satisfy the global Markov property, which means that d-separation in the graph
 1281 implies conditional independence. It turns out that in our case, the required assumption
 1282 of independent errors applies to the p error variables and not to G . The intuition behind
 1283 this is that G is not just an additional error node, but part of the causal graph, and we
 1284 can always distinguish between residual (co)variance and genetic (co)variance. We now
 1285 give the proof of Theorem 1, which only requires minor modifications of the proof given
 1286 by Pearl for the case without the genetic effects.

1287 Let \mathcal{G}_E denote the extended graph, obtained by adding the error variables, i.e. for
 1288 traits $j = 1, \dots, p$ we add the node E_j and an edge $E_j \rightarrow Y_j$. We first show that the local
 1289 Markov property holds for \mathcal{G}_E , i.e. for any variable $Z \in \{G, Y_1, \dots, Y_p, E_1, \dots, E_p\}$, Z is
 1290 conditionally independent of its non-descendants given its parents. This is obvious for $Z \in$
 1291 $\{G, E_1, \dots, E_p\}$; we now consider Y_j . In \mathcal{G}_E , the set of parents of Y_j is $pa(Y_j) \cup \{E_j\}$, where
 1292 $pa(Y_j)$ contains G if $j \in D$. By construction, Y_j is entirely determined by $pa(Y_j) \cup \{E_j\}$,
 1293 and constant conditional on these variables. Consequently, given $pa(Y_j) \cup \{E_j\}$, it is
 1294 independent of any E_k ($k \neq j$), and of any Y_k that it is a non-descendant of Y_j (Note
 1295 that if $G \notin pa(Y_j)$, Y_j is indeed conditionally independent of any non-descendant; if
 1296 $G \in pa(Y_j)$, G cannot be the non-descendant because it is already in the conditioning

1297 set). Therefore the local Markov property holds for \mathcal{G}_E . By Lemma E.1 below, we find
 1298 that also the Markov factorization property holds for \mathcal{G}_E , since for any distribution having
 1299 a density it is equivalent with the local and global Markov properties. Given the Markov
 1300 factorization property for \mathcal{G}_E and the fact that $f(e_1, \dots, e_p) = \prod_{j=1}^p f_j(e_j)$, we can just
 1301 integrate out the e_j , and obtain the Markov factorization property for \mathcal{G} .

1302 Markov properties

1303 The following lemma is taken from [57] (p. 51), and reformulated with somewhat less
 1304 general conditions, which however suffice for our purpose.

1305 **Lemma E.1** *Let P be the joint distribution of random variables (Y_1, \dots, Y_p) , having a*
 1306 *density f , and let \mathcal{G} be a DAG on these variables. The following properties are equivalent:*

- 1307 • *The Markov factorization property: given the parents pa_j of each x_j , the joint density*
 1308 *(f) can be decomposed as*

$$f(y_1, \dots, y_p) = \prod_{j=1}^p f_j(y_j | pa_j),$$

1309 *where the f_j are the conditional densities.*

- 1310 • *The local Markov property: any variable is conditionally independent of its non-*
 1311 *descendants, given its parents.*
- 1312 • *The global Markov property: for all disjoint sets $U, V, S \subset \{Y_1, \dots, Y_p\}$, d -separation*
 1313 *of U and V by S in the graph \mathcal{G} implies conditional independence of U and V given*
 1314 *S . In contrast to U and V , the conditioning set S may be empty here. A definition*
 1315 *of d -separation is given in S1.*

1316 E.9 Proof of Theorems 3 and 5

1317 We first prove Theorem 3, by showing the equivalence of the left- and right hand sides of
 1318 (18) and (20). The d -separation statements on the right hand sides are equivalent, as G
 1319 can never be a (descendant of a) collider. Also the left hand sides ($\mathbf{Y}_j \perp\!\!\!\perp \mathbf{Y}_k | \{\mathbf{Y}_S, \mathbf{G}\}$
 1320 and $\mathbf{Y}_j \perp\!\!\!\perp_{P_{Y|G\Gamma}} \mathbf{Y}_k | \{\mathbf{Y}_S\}$) are equivalent, since

$$p_{Y|G\Gamma}(y_j, y_k | y_S) = p(y_j, y_k | y_S, G\Gamma) = p(y_j | y_S, G\Gamma) p(y_k | y_S, G\Gamma) = p_{Y|G\Gamma}(y_j | y_S) p_{Y|G}(y_k | y_S).$$

1321 For Theorem 5 we make the additional assumption that $K = ZZ^t$, $Z = I_m \otimes (1, \dots, 1)^t$
 1322 being the $mr \times m$ design matrix for r replicates of m genotypes in a balanced design
 1323 (with $mr = n$). The first part of Theorem 5 then follows from the results in Appendix
 1324 E.7. For the second part, first recall the equivalence of $\mathbf{Y}_j \perp\!\!\!\perp \mathbf{Y}_k | \{\mathbf{Y}_S, \mathbf{G}\}$ and $\mathbf{Y}_j \perp\!\!\!\perp_{P_{Y|G\Gamma}}$

1325 $\mathbf{Y}_k | \{\mathbf{Y}_S\}$. Because of the Gaussianity and the assumed faithfulness, the latter conditional
1326 independence is equivalent with

$$\gamma_j^t \Sigma_E \gamma_k - (\gamma_j^t \Sigma_E \Gamma_S) (\Gamma_S^t \Sigma_E \Gamma_S)^{-1} (\Gamma_S^t \Sigma_E \gamma_k) = 0, \quad (36)$$

1327 where we used (26).

1328 Next we consider the conditional distribution of $\text{vec}([\mathbf{Y}_j \ \mathbf{Y}_k]) | \mathbf{Y}_S = \tilde{y}_S$ given in (11),
1329 whose covariance is the $2n \times 2n$ block matrix $\Sigma_{jk} - \Sigma_{jk,S} \Sigma_S^{-1} \Sigma_{jk,S}^t$. All $n \times n$ blocks are a
1330 linear combinations of K and I_n , and it suffices to show that the coefficient of I_n in the
1331 off-diagonal blocks is zero if and only if (36) holds. We recall from (32) that

$$\Sigma_S = (\Gamma^t \Sigma_G \Gamma) [S, S] \otimes K + (\Gamma^t \Sigma_E \Gamma) [S, S] \otimes I_n = (\Gamma_S^t \Sigma_G \Gamma_S) \otimes K + (\Gamma_S^t \Sigma_E \Gamma_S) \otimes I_n.$$

1332 Using the Woodbury identity (equation (35)) with $A = V_E \otimes I_n$, $B = V_G \otimes I_m$ and
1333 $C = I_p \otimes Z$, it follows that for any positive (semi) definite $p \times p$ matrices V_G and V_E , we
1334 have

$$(V_G \otimes K + V_E \otimes I_n)^{-1} = (V_E^{-1} \otimes I_n) - (V_E^{-1} (V_G^{-1} r V_E^{-1})^{-1} V_E^{-1}) \otimes K. \quad (37)$$

1335 Setting $V_G = \Gamma_S^t \Sigma_G \Gamma_S$, $V_E = \Gamma_S^t \Sigma_E \Gamma_S$ and $A = V_E^{-1} (V_G^{-1} r V_E^{-1})^{-1} V_E^{-1}$, it follows that

$$\Sigma_S^{-1} = (\Gamma_S^t \Sigma_E \Gamma_S)^{-1} \otimes I_n - A \otimes K, \quad (38)$$

Combining this with the expressions for Σ_{jk} and $\Sigma_{jk,S}$ given in section E.6, we find that
 $\Sigma_{jk} - \Sigma_{jk,S} \Sigma_S^{-1} \Sigma_{jk,S}^t$ has off-diagonal blocks

$$\begin{aligned} & (\gamma_j^t \Sigma_G \gamma_k) \otimes K + (\gamma_j^t \Sigma_E \gamma_k) \otimes I_n \\ & - ((\gamma_j^t \Sigma_E \Gamma_S) \otimes I_n + B_j \otimes K) ((\Gamma_S^t \Sigma_E \Gamma_S)^{-1} \otimes I_n - A \otimes K) (\Gamma_S^t \Sigma_E \gamma_k \otimes I_n + B_k^t \otimes K), \end{aligned}$$

1336 for $B_j = \gamma_j \Sigma_G \Gamma_S$ and $B_k = \gamma_k \Sigma_G \Gamma_S$.

1337 Finally, working out the products in the last display (using that $K^2 = rK$), we find
1338 that all terms involving a kronecker product with I_n correspond exactly to the left-hand-
1339 side of (36). Consequently, the residual covariance in the distribution $(\mathbf{Y}_j, \mathbf{Y}_k) | \mathbf{Y}_S = \tilde{y}_S$
1340 is zero if and only if $\mathbf{Y}_j \perp\!\!\!\perp \mathbf{Y}_k | \{\mathbf{Y}_S, \mathbf{G}\}$.

1341 E.10 Proof of Theorem 4

1342 To obtain faithfulness for $S = \emptyset$, we need to prove that $\mathbf{Y}_j \perp\!\!\!\perp \mathbf{G}$ implies d-separation of Y_j
1343 and G in the graph \mathcal{G} . Because the conditioning set is empty, it suffices to show that there
1344 are no directed paths from G to Y_j , where we can assume that $j \notin D$ (otherwise \mathbf{G}_j would
1345 be nonzero, and because of the non-collinearity, \mathbf{Y}_j and \mathbf{G} would not be independent).

1346 Because of the assumed Gaussianity, the independence of \mathbf{Y}_j and \mathbf{G} implies that

$$\text{Cov}(\mathbf{Y}_j^t, \mathbf{G}) = \text{Cov}(\gamma_j^t \mathbf{G}^t, \mathbf{G}) = \text{trace}(K) \gamma_j^t \Sigma_G = (0, \dots, 0), \quad (39)$$

1347 where we used that $\text{vec}(\mathbf{G}) \sim N(0, \Sigma_G \otimes K)$, and therefore $E(\mathbf{G}[i, j]\mathbf{G}[i, k]) = \Sigma_G[j, k]K[i, i]$,
 1348 for all $i \in \{1, \dots, n\}$ and $j, k \in \{1, \dots, p\}$. Since $\text{trace}(K)$ is strictly positive and the sub-
 1349 matrix $\Sigma_G[D, D]$ has full rank, equation (39) implies that $\gamma_{j,l} = 0$ for all $l \in D$.

1350 Finally, we use that the assumed faithfulness implies the path coefficients condition
 1351 (see sections E.2-E.5). Consequently, it follows from $\gamma_{j,l} = 0$ that there is no directed
 1352 path from Y_l to Y_j . Since this is the case for all $l \in D$, there can neither be a directed
 1353 path from G to Y_j .

1354 E.11 Proof of Theorem 6

1355 Assuming $K = ZZ^t$, the first part of theorem follows from the results in Appendix E.7.
 1356 For the second part, we use that \mathbf{Y}_j has genetic variance $\sigma_j^2(G) = \gamma_j^t \Sigma_G \gamma_j$ (see equation
 1357 (25)). Because for traits without a direct genetic effect, rows and columns in Σ_G are zero,
 1358 we can rewrite this as $\gamma_j^t[D] \Sigma_G[D, D] \gamma_j[D]$. Hence, $\sigma_j^2(G) = 0$ is equivalent with $\gamma_{j,l} = 0$
 1359 for all $l \in D$, where we used that $\Sigma_G[D, D]$ is of full rank. Using the arguments from the
 1360 proof of Theorem 4 and the assumed faithfulness, it follows that this is equivalent with
 1361 independence of \mathbf{Y}_j and \mathbf{G} .

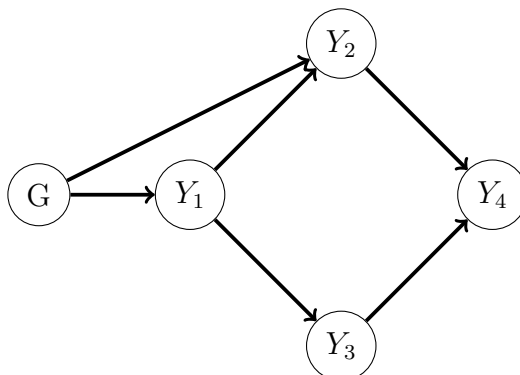


Figure 8. A genetic DAG with independent genetic effects on Y_1 and Y_2 , and a direct effect $Y_1 \rightarrow Y_2$

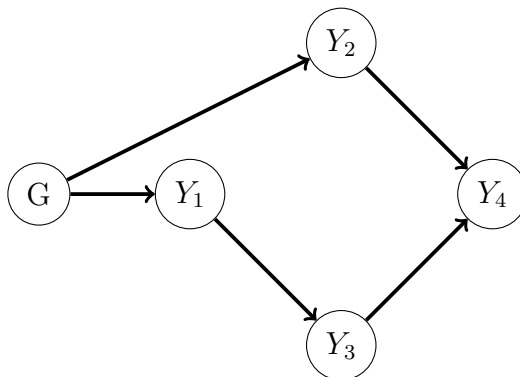


Figure 9. A typical output of PC-gen (without the modifications in section F.1), based on observations of Y_1, Y_2, Y_3, Y_4 generated by the genetic DAG of Figure 8, for 400 genotypes and 2 replicates. The edge between Y_1 and Y_2 is missing because the test for conditional independence between Y_1 and Y_2 given Y_4 has too little power.

1362 F Skipping independence tests that do not involve G

1363 F.1 Motivation for skipping the test for $Y_j \perp\!\!\!\perp Y_k | \{Y_S\}$

1364 Although PC-gen can be shown to be consistent, its finite sample performance can be
 1365 improved if we skip some of the tests in the skeleton-stage. Differences between the
 1366 population and sample version of PC-skeleton can occur everywhere in the graph, but are
 1367 most likely for conditioning sets not containing G . This is illustrated in the example in
 1368 Figure 8, in which there are genetic effects on traits Y_j and Y_k , as well as a direct effect
 1369 of Y_j on Y_k .

1370 Then given a large number of observations of Y_1, Y_2, Y_3, Y_4 and assuming faithfulness,
 1371 we will recover the true skeleton. However, with a small or moderate sample size, the test

1372 for conditional independence of Y_1 and Y_2 given Y_4 has very little power. The test for
 1373 conditional independence of Y_1 and Y_2 given $\{Y_4, G\}$ is a lot more powerful here; however
 1374 the standard PC-skeleton algorithm (and neither PC-stable) does not perform this test
 1375 anymore after the null-hypothesis of conditional independence of Y_1 and Y_2 given Y_4 alone
 1376 has been accepted. Therefore, a typical output of PC-gen looks like Figure 9.

1377 In order to make PC-gen more powerful we therefore propose to perform only those
 1378 conditional independence tests where G is contained in the conditioning set, at least when
 1379 both variables whose conditional independence is tested have positive genetic variance⁸.
 1380 Leaving out all conditional independence tests where G is not in the conditioning set still
 1381 gives a valid algorithm, in the sense that in the population version of PC-skeleton, no false
 1382 positives are obtained. Intuitively this is obvious, as G is a root node, and everything can
 1383 be done conditionally on G . We formally show this in appendix F.

1384 F.2 A characterization of the skeleton

1385 If a distribution P is faithful with respect to a DAG \mathcal{G} , we have the following result for
 1386 the skeleton of \mathcal{G} :

$$\begin{aligned} & \text{there is an edge between nodes } A \text{ and } B \text{ in the skeleton of a DAG } \mathcal{G} \\ \iff & \forall S \subseteq V \setminus \{A, B\}, A \text{ and } B \text{ are conditionally dependent given } S. \end{aligned} \quad (40)$$

1387 This was shown in [5] (Theorem 3.4); here we adopt the formulation of [32] (p.616). It
 1388 is important to note that in general the skeleton is not equal to the so-called conditional
 1389 independence graph (CIG), which is the undirected graph associated with the inverse
 1390 covariance or precision matrix. The latter is characterized by an equivalence statement
 1391 similar to (40), but with on the right-hand side only $S = V \setminus \{A, B\}$. Hence, if data
 1392 are generated by a DAG \mathcal{G} and we assume faithfulness, the skeleton of \mathcal{G} is typically a
 1393 subgraph of the CIG. In case $A \rightarrow B \leftarrow C$ for example, the CIG also contains an edge
 1394 $A - C$ (because $S = \emptyset$ d-separates A and C , but $S = B$ does not, B being a collider on
 1395 the path $A \rightarrow B \leftarrow C$).

1396 F.3 Skipping the test for $Y_j \perp\!\!\!\perp Y_k | \{Y_S\}$ does not affect pcgen 1397 (oracle)

1398 In view of (40), our modification is correct in the sense that the population version of
 1399 PC-skeleton still recovers the true skeleton. This correctness follows from the facts that

⁸In the true graph, we can partition $V = \{Y_1, \dots, Y_p\}$ in a set V^G containing all variables having positive genetic variance and a set $(V^G)^c$ with variables without genetic variance. A variable Y_k is in V^G when there exists at least one directed path $G \rightarrow \dots \rightarrow Y_k$. Any edge between $Y_j \in V^G$ and $Y_k \in (V^G)^c$ must be directed $Y_k \rightarrow Y_j$. When estimating the true graph (CPDAG) from data, we start PC-skeleton by testing marginal independence between each trait and G , i.e. testing genetic variance. Based on these tests we obtain estimates V^G and $(V^G)^c$, used in the remainder.

- 1400 1. the PC-skeleton algorithm starts with the complete undirected graph, and then
1401 tests conditional independencies, removing edges when a conditional independence
1402 relation is found. When in fact there *is* an edge $Y_j - Y_k$ in the true genetic DAG,
1403 then removing some of these tests clearly still produces the correct result (and in
1404 the sample version even with higher probability, which is the motivation of doing
1405 this...)

- 1406 2. if there *is no* edge $Y_j - Y_k$ and a set S not containing G is blocking a path between
1407 Y_j and Y_k , then $\{G\} \cup S$ is also blocking it (since G can never be a collider, and
1408 neither be a descendant of any node). In other words, it can not happen that a set
1409 S not containing G is blocking all paths between Y_j and Y_k , and that the addition
1410 of G would 'unblock' one of these paths. Consequently, it can not happen that
1411 such a set S is the *only* set separating Y_j and Y_k , and that we would miss the only
1412 opportunity to remove the edge between Y_j and Y_k .

1413 G Miscellaneous

1414 G.1 Representing G by a single node: a motivating example

1415 The main reason for representing $\mathbf{G}_1, \dots, \mathbf{G}_p$ with a single node (instead of G_1, \dots, G_p ,
 1416 sometimes used in the literature) is that the relatedness matrix K is the same for all
 1417 traits. If for example $G_1 \rightarrow Y_1 \rightarrow Y_2 \leftarrow Y_3 \leftarrow G_3$ (as in Figure 2), it follows from (8) that
 1418 the marginal distribution of \mathbf{Y}_2 has covariance $c_1 K + c_2 I_n$, for some nonnegative constants
 1419 c_1, c_2 . However, based on this distribution alone, we cannot distinguish the contributions
 1420 of \mathbf{G}_1 and \mathbf{G}_3 . This differs from the scenario $QTL_A \rightarrow Y_1 \rightarrow Y_2 \leftarrow Y_3 \leftarrow QTL_B$, where
 1421 the total (fixed) effects of QTLs A and B on Y_2 can be estimated from the marginal
 1422 distribution of Y_2 . If we condition on \mathbf{Y}_1 and \mathbf{Y}_3 , \mathbf{Y}_2 becomes independent of \mathbf{G}_1 and \mathbf{G}_3 .
 1423 In fact it is also independent of \mathbf{G}_2 , since the latter is zero. Consequently, given \mathbf{G}_3 , \mathbf{Y}_2
 1424 is independent of $\mathbf{G} = [\mathbf{G}_1 \mathbf{G}_2 \mathbf{G}_3]$, which illustrates that the conditional independencies
 1425 correspond to a property of the graph \mathcal{G} , with $[\mathbf{G}_1 \mathbf{G}_2 \mathbf{G}_3]$ represented by a single node
 1426 G .

1427 G.2 The limitations of genomic networks

1428 [13] recently proposed to estimate a directed network based on the predicted genetic
 1429 effects themselves, rather than the residuals. Compared to residual-based estimation or
 1430 pcgen, this however seems to require stronger assumptions.

1431 As an example, consider the graph $Y_1 \rightarrow Y_3 \leftarrow Y_2$, with direct genetic effects on all
 1432 3 traits, i.e. $\text{rank}(\Sigma_G) = 3$. For the sake of the argument, assume also that the total
 1433 genetic effects can be predicted without error, i.e. we can observe the matrix $\mathbf{U} := \mathbf{G}\Gamma$
 1434 (see equation (7)); because Y_1 and Y_2 are root nodes in \mathcal{G}_y , it turns out that $\mathbf{U}_1 := \mathbf{G}_1$
 1435 and $\mathbf{U}_2 := \mathbf{G}_2$.

1436 In order not to get an incorrect edge between Y_1 and Y_2 the PC-algorithm must
 1437 find that these traits are marginally independent. Using residuals, we indeed find that
 1438 $\mathbf{Y}_1 - \mathbf{U}_1$ and $\mathbf{Y}_2 - \mathbf{U}_2$ are independent. However, \mathbf{U}_1 and \mathbf{U}_2 are only independent if
 1439 Σ_G is diagonal, which is a rather strong and restrictive additional assumption.

1440 The only advantage of genomic networks over the residual-based ones is that they
 1441 may infer some of the edges $G \rightarrow Y_j$. This is however only indirectly, through comparison
 1442 with the trait-to-trait network estimated by the residuals-based network, and requires
 1443 additional testing if the \mathbf{U}_j are zero (i.e. the tests for $\mathbf{Y}_j \perp\!\!\!\perp \mathbf{G} | \emptyset$ in pcgen, which were not
 1444 considered in [13]). Even then, the edges $G \rightarrow Y_j$ can be inferred only partially. In the
 1445 above example, if we conclude that $\mathbf{U}_j = 0$ for some j , this clearly excludes $G \rightarrow Y_j$. But
 1446 if $\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3$ are all nonzero, we can only conclude (by comparing with the already known
 1447 trait-to-trait network $Y_1 \rightarrow Y_3 \leftarrow Y_2$) that $G \rightarrow Y_1$ and $G \rightarrow Y_2$. But it is impossible to
 1448 make any inference about $G \rightarrow Y_2$, which pcgen can in principle do.

1449 **H** Maize data: reconstructions with $\alpha = 0.001$

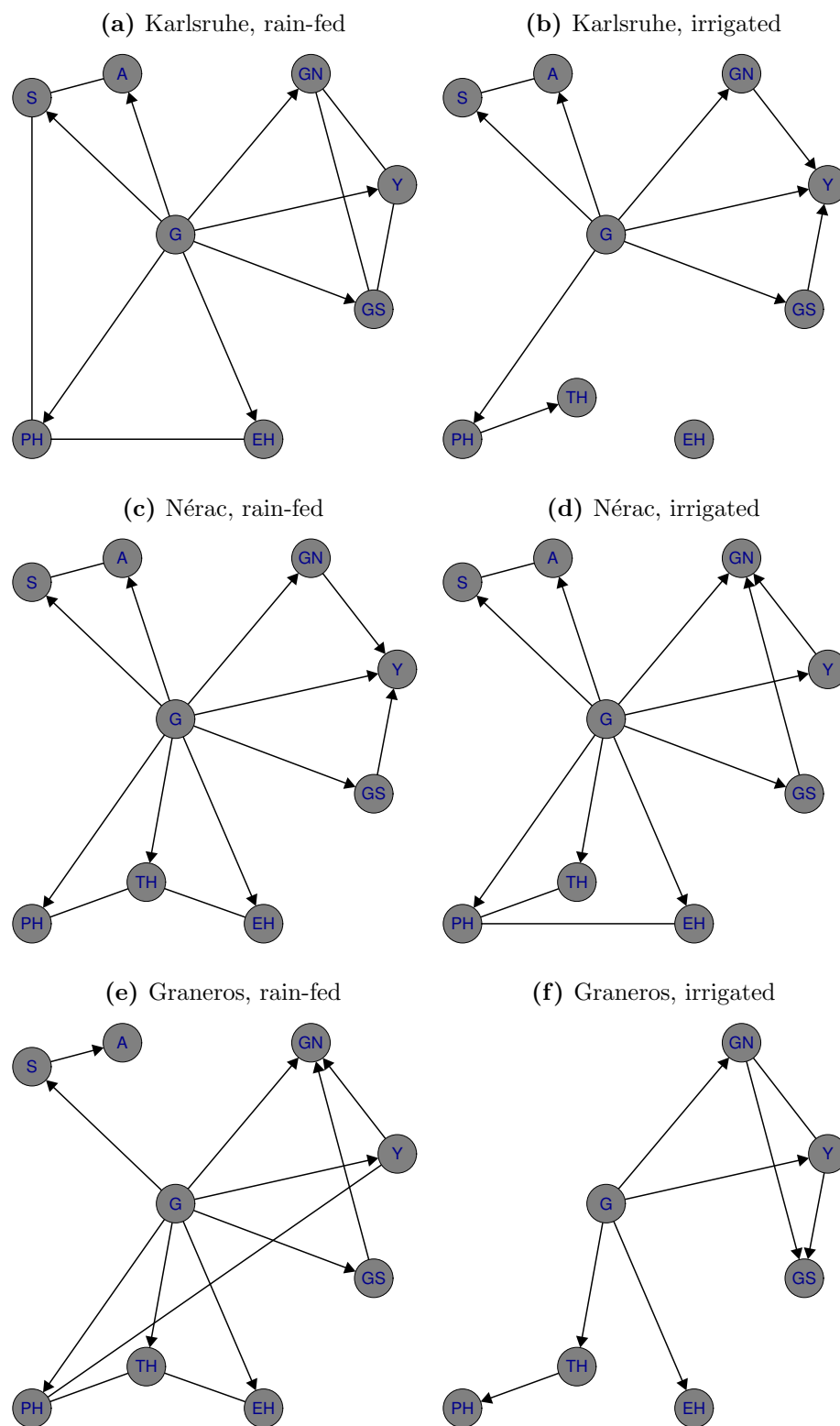


Figure 10. Estimated networks for six of the DROPS field trials in 2013, with $\alpha = 0.001$. Rows correspond to locations (Karlsruhe, Nérac, Graneros), columns to treatments (rain-fed, irrigated).

References

1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482

1. Wright S (1921) Correlation and causation. *Journal of Agricultural Research* : 557-585.
2. Wu XL, Heringstad B, Gianola D (2010) Bayesian structural equation models for inferring relationships between phenotypes: a review of methodology, identifiability, and applications. *Journal of Animal Breeding and Genetics* 127: 3–15.
3. Onogi A, Ideta O, Yoshioka T, Ebanu K, Yamasaki M, et al. (2016) Uncovering a nuisance influence of a phenological trait of plants using a nonlinear structural equation: Application to days to heading and culm length in asian cultivated rice. *PLoS ONE* 11: 1-17.
4. Pearl J (2000) Causality. *Causality*, by Judea Pearl, pp.~400.~ISBN 0521773628.~Cambridge, UK: Cambridge University Press, March 2000. URL http://adsabs.harvard.edu/cgi-bin/nph-bib_query?bibcode=2000caus.book.....P.
5. Spirtes P, Glymour C, Scheines R (2001) Causation, Prediction, and Search, Second Edition (Adaptive Computation and Machine Learning). A Bradford Book, second edition edition. URL <http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0262194406>.
6. Maathuis MH, Nandy P (2016) A review of some recent advances in causal inference. *Handbook of Big Data* : 387.
7. Maathuis MH, Colombo D, Kalisch M, Bühlmann P (2010) Predicting causal effects in large-scale systems from observational data. *Nature Methods* 7: 247–248.
8. Chaibub Neto E, Ferrara CT, Attie AD, Yandell BS (2008) Inferring causal phenotype networks from segregating populations. *Genetics* 179: 1089–1100.
9. Chaibub Neto E, Broman AT, Keller MP, Attie AD, Zhang B, et al. (2013) Modeling causality for pairs of phenotypes in system genetics. *Genetics* .
10. Scutari M, Howell P, Balding DJ, Mackay I (2014) Multiple quantitative trait analysis using bayesian networks. *Genetics* 198: 129–137.
11. Gianola D, Sorensen D (2004) Quantitative genetic models for describing simultaneous and recursive relationships between phenotypes. *Genetics* 167: 1407–1424.
12. Valente BD, Rosa GJM, de los Campos G, Gianola D, Silva MA (2010) Searching for recursive causal structures in multivariate quantitative genetics mixed models. *Genetics* 185: 633–644.

- 1483 13. Töpner K, Rosa GJM, Gianola D, Schön CC (2017) Bayesian networks illustrate
1484 genomic and residual trait connections in maize (*zea mays* l.). *G3: Genes, Genomes,*
1485 *Genetics* 7: 2779–2789.
- 1486 14. Gao B, Cui Y (2015) Learning directed acyclic graphical structures with genetical
1487 genomics data. *Bioinformatics* : btv513.
- 1488 15. Stephens M (2013) A unified framework for association analysis with multiple re-
1489 lated phenotypes. *PLoS ONE* 8: e65245.
- 1490 16. Valente BD, Rosa GJM, Gianola D, Wu XL, Weigel K (2013) Is structural equation
1491 modeling advantageous for the genetic improvement of multiple traits? *Genetics*
1492 194: 561–572.
- 1493 17. Valente BD, Morota G, Peñagaricano F, Gianola D, Weigel K, et al. (2015) The
1494 causal meaning of genomic predictors and how it affects construction and compar-
1495 ison of genome-enabled selection models. *Genetics* 200: 483–494.
- 1496 18. Kruijer W, Boer MP, Malosetti M, Flood PJ, Engel B, et al. (2015) Marker-based
1497 estimation of heritability in immortal populations. *Genetics* 199: 379-398.
- 1498 19. Drton M, Foygel R, Sullivant S (2011) Global identifiability of linear structural
1499 equation models. *Ann Statist* 39: 865–886.
- 1500 20. Shipley B (2016) Cause and correlation in biology: A user’s guide to path analysis,
1501 structural equations and causal inference. Cambridge, USA: Cambridge University
1502 Press, 2 edition.
- 1503 21. Bollen KA (1989) Structural Equations with Latent Variables. Wiley-Interscience,
1504 1 edition. URL <http://www.worldcat.org/isbn/0471011711>.
- 1505 22. Moore AJ, Brodie III ED, Wolf JB (1997) Interacting phenotypes and the evolution-
1506 ary process: I. direct and indirect genetic effects of social interactions. *Evolution*
1507 51: 1352–1362.
- 1508 23. Bijma P (2014) The quantitative genetics of indirect genetic effects: a selective
1509 review of modelling issues. *Heredity* 112: 61.
- 1510 24. Peters J, Janzing D, Schölkopf B (2017) Elements of causal inference: foundations
1511 and learning algorithms. MIT press.
- 1512 25. Zhou X, Stephens M (2014) Efficient multivariate linear mixed model algorithms
1513 for genome-wide association studies. *Nat Meth* 11: 407–409.

- 1514 26. Korte A, Vilhjalmsson BJ, Segura V, Platt A, Long Q, et al. (2012) A mixed-model
1515 approach for genome-wide association studies of correlated traits in structured pop-
1516 ulations. *Nat Genet* 44: 1066–1071.
- 1517 27. Pearl J (1989) Probabilistic reasoning in intelligent systems: networks of plausible
1518 inference. San Mateo etc.: Morgan Kaufmann Publishers, xiv + 552 pp.
- 1519 28. Kalisch M, Mächler M, Colombo D, Maathuis MH, Bühlmann P (2012) Causal
1520 inference using graphical models with the R package pcalg. *Journal of Statistical
1521 Software* 47: 1–26.
- 1522 29. Zwiernik P, Uhler C, Richards D (2014). Maximum likelihood estimation for linear
1523 gaussian covariance models. [arXiv:1408.5604](https://arxiv.org/abs/1408.5604).
- 1524 30. Colombo D, Maathuis MH (2014) Order-independent constraint-based causal struc-
1525 ture learning. *Journal of Machine Learning Research* 15: 3741–3782.
- 1526 31. Hauser A, Bühlmann P (2012) Characterization and greedy learning of interven-
1527 tional Markov equivalence classes of directed acyclic graphs. *Journal of Machine
1528 Learning Research* 13: 2409–2464.
- 1529 32. Kalisch M, Bühlmann P (2007) Estimating high-dimensional directed acyclic graphs
1530 with the PC-algorithm. *J Mach Learn Res* 8: 613–636.
- 1531 33. Millet E, Welcker C, Kruijer W, Negro S, Nicolas S, et al. (2016) Genome-wide
1532 analysis of yield in europe: allelic effects as functions of drought and heat scenarios.
1533 *Plant Physiology* : pp–00621.
- 1534 34. Keating BA, Carberry PS, Hammer GL, Probert ME, Robertson MJ, et al. (2003)
1535 An overview of APSIM, a model designed for farming systems simulation. *Eur J
1536 Agron* 18: 267–288.
- 1537 35. Holzworth DP, Huth NI, DeVoil PG, Zurcher EJ, Herrmann NI, et al. (2014) Apsim
1538 - evolution towards a new generation of agricultural systems simulation. *Env Model
1539 Soft* .
- 1540 36. Richardson T, Spirtes P (2002) Ancestral graph markov models. *Annals of Statistics*
1541 : 962–1030.
- 1542 37. Rodríguez-Álvarez MX, Boer MP, van Eeuwijk FA, Eilers PH (2017) Correcting for
1543 spatial heterogeneity in plant breeding experiments with p-splines. *Spatial Statistics*
1544 23: 52 - 71.
- 1545 38. Kadam N, Tamilselvan A, Lawas LMF, Quinones C, Bahuguna R, et al. (2017)
1546 Genetic control of plasticity in root morphology and anatomy of rice in response to
1547 water-deficit. *Plant Physiology* .

- 1548 39. Henderson CR (1975) Best Linear Unbiased Estimation and Prediction under a
1549 Selection Model. *Biometrics* 31.
- 1550 40. Furlotte NA, Eskin E (2015) Efficient multiple-trait association and estimation of
1551 genetic correlation using the matrix-variate linear mixed model. *Genetics* 200: 59–
1552 68.
- 1553 41. Joo JWJ, Kang EY, Org E, Furlotte N, Parks B, et al. (2016) Efficient and accu-
1554 rate multiple-phenotype regression method for high dimensional data considering
1555 population structure. *Genetics* 204: 1379–1390.
- 1556 42. Peters J, Bühlmann P, Meinshausen N (2016) Causal inference by using invariant
1557 prediction: identification and confidence intervals. *Journal of the Royal Statistical*
1558 *Society: Series B (Statistical Methodology)* 78: 947–1012.
- 1559 43. Visscher PM, Goddard ME (2014) A general unified framework to assess the sam-
1560 pling variance of heritability estimates using pedigree or marker-based relationships.
1561 *Genetics* .
- 1562 44. Kruijer W (2016) Misspecification in mixed-model based association analysis. *Ge-*
1563 *netics* 202: 363-366.
- 1564 45. Flaxman SR, Neill DB, Smola AJ (2015) Gaussian processes for indepen-
1565 dence tests with non-iid data in causal inference. Provisional acceptance at
1566 ACM Transactions on Intelligent Systems and Technology (TIST), 2015b URL
1567 <http://wwwsethrfcom/files/gp-dependpdf> .
- 1568 46. Meinshausen N, Bühlmann P (2010) Stability selection. *Journal of the Royal Sta-*
1569 *tistical Society: Series B (Statistical Methodology)* 72: 417–473.
- 1570 47. Stekhoven DJ, Moraes I, Sveinbjörnsson G, Hennig L, Maathuis MH, et al. (2012)
1571 Causal stability ranking. *Bioinformatics* 28: 2819-2823.
- 1572 48. Bühlmann P, Rütimann P, Kalisch M (2011) Controlling false positive selections in
1573 high-dimensional regression and causal inference. *Statistical Methods in Medical*
1574 *Research* .
- 1575 49. Meinshausen N, Hauser A, Mooij JM, Peters J, Versteeg P, et al. (2016) Methods for
1576 causal inference from gene perturbation experiments and validation. *Proceedings*
1577 *of the National Academy of Sciences* 113: 7361–7368.
- 1578 50. Pfister N, Bühlmann P, Schölkopf B, Peters J (2018) Kernel-based tests for joint in-
1579 dependence. *Journal of the Royal Statistical Society: Series B (Statistical Method-*
1580 *ology)* 80: 5-31.

- 1581 51. Colombo D, Maathuis MH, Kalisch M, Richardson TS (2012) Learning high-
1582 dimensional directed acyclic graphs with latent and selection variables. *Ann Statist*
1583 40: 294–321.
- 1584 52. Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete
1585 data via the em algorithm. *Journal of the royal statistical society Series B (method-*
1586 *ological)* : 1–38.
- 1587 53. Joe H (2006) Generating random correlation matrices based on partial correlations.
1588 *Journal of Multivariate Analysis* 97: 2177–2189.
- 1589 54. Tsamardinos I, Brown LE, Aliferis CF (2006) The max-min hill-climbing bayesian
1590 network structure learning algorithm. *Machine learning* 65: 31–78.
- 1591 55. Bustos-Korts D, Malosetti M, Chapman SC, Chenu K, Boer M, et al. (2017) A
1592 protocol combining statistical and crop growth modelling to evaluate phenotyp-
1593 ing strategies useful for selection under different drought patterns. Ph.D. thesis,
1594 Wageningen University, Wageningen. doi:10.18174/421321.
- 1595 56. Casadebaig P, Zheng B, Chapman S, Huth N, Faivre R, et al. (2016) Assessment
1596 of the potential impacts of plant traits across environments by combining global
1597 sensitivity analysis and dynamic modeling in wheat. *PLoS One* .
- 1598 57. Lauritzen SL (1996) *Graphical Models*. Oxford Statistical Science Series. New
1599 York, USA: Oxford University Press. URL [http://www.worldcat.org/isbn/](http://www.worldcat.org/isbn/0198522193)
1600 [0198522193](http://www.worldcat.org/isbn/0198522193).
- 1601 58. Chickering DM (2002) Learning equivalence classes of bayesian-network structures.
1602 *J Mach Learn Res* 2: 445–498.
- 1603 59. Lynch M, Walsh B (1998) *Genetics and Analysis of Quantitative Traits*. Sinauer As-
1604 sociates, 1 edition. URL [http://www.amazon.com/exec/obidos/redirect?tag=](http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0878934812)
1605 [citeulike07-20&path=ASIN/0878934812](http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0878934812).
- 1606 60. Peters J (2012) *Restricted Structural Equation Models for Causal Inference*. Ph.D.
1607 thesis, ETH Zurich and MPI for Intelligent Systems. [http://dx.doi.org/10.](http://dx.doi.org/10.3929/ethz-a-007597940)
1608 [3929/ethz-a-007597940](http://dx.doi.org/10.3929/ethz-a-007597940).
- 1609 61. Petersen KB, Pedersen MS, et al. (2008) *The matrix cookbook*. Technical University
1610 of Denmark 7: 510.
- 1611 62. Golub GH, Van Loan CF (2012) *Matrix computations*, volume 3. JHU Press.