nature genetics

A mixed-model approach for genome-wide association studies of correlated traits in structured populations

Arthur Korte[1,4], Bjarni J Vilhjálmsson[1,2,4], Vincent Segura[1,3,4], Alexander Platt[1,2], Quan Long[1] & Magnus Nordborg[1,2]

Genome-wide association studies (GWAS) are a standard approach for studying the genetics of natural variation. A major concern in GWAS is the need to account for the complicated dependence structure of the data, both between loci as well as between individuals. Mixed models have emerged as a general and flexible approach for correcting for population structure in GWAS. Here, we extend this linear mixed-model approach to carry out GWAS of correlated phenotypes, deriving a fully parameterized multi-trait mixed model (MTMM) that considers both the within-trait and between-trait variance components simultaneously for multiple traits. We apply this to data from a human cohort for correlated blood lipid traits from the Northern Finland Birth Cohort 1966 and show greatly increased power to detect pleiotropic loci that affect more than one blood lipid trait. We also apply this approach to an *Arabidopsis thaliana* data set for flowering measurements in two different locations, identifying loci whose effect depends on the environment.

Most GWAS to date have been conducted using the simplest possible statistical model: a single-locus test of association between a binary SNP genotype and a single phenotype. Given that most traits of interest are multifactorial, this clearly amounts to model misspecification, and the resulting danger of biased results whenever there is a lack of independent (linkage disequilibrium) between causal loci (for example, due to population structure) is well known[1–3]. Much less attention has been devoted to the fact that phenotypes may also be correlated. Whenever multiple measurements are taken from individuals, the resulting phenotypes will be correlated because of pleiotropy, which is of direct interest, as well as shared environment and linkage disequilibrium, which are usually confounding factors. Taking these correlations into account is important, not only because of the importance of understanding pleiotropy, but also because we may expect increased power compared to marginal analyses. Intuitively, correlated traits amount to a form of replication. The importance of correlated phenotypes becomes even clearer when we consider measurements across environments. The canonical example here is an agricultural field experiment using inbred lines, a setting in which no one would consider analyzing phenotypes from different environments independently of each other because the whole point of the study is to separate genetic from environmental effects and identify genotype-environment interactions. In human genetics, disentangling genetic and environmental effects is also of obvious interest, although much more challenging, as the environment usually cannot be experimentally manipulated[4].

There is a long history of multi-trait models in quantitative genetics[5–9], but these methods have rarely been applied to GWAS. In this paper, we show how a standard linear mixed model from animal breeding[10] may be used to model correlated traits, while at the same time correcting for dependence among loci (for example, due to population structure). As designs like cohort studies become more prevalent, the need for modeling correlated traits as well as population structure will grow[2,11,12], and the same is true for the increasing number of nonhuman GWAS[13–17].

The mixed model, which handles population structure by estimating the phenotypic covariance that is due to genetic relatedness—or kinship—between individuals, has previously been shown to perform well in GWAS[2,13,18–22]. Here, we extend this approach to handle correlated phenotypes by deriving a fully parameterized multi-trait mixed model (MTMM) that considers both the within-trait and between-trait variance components simultaneously for multiple traits (Online Methods), implementing it for GWAS. The idea is not new[23–27], but it has never been applied for association mapping on a genome-wide scale. Alternative approaches for GWAS analysis at multiple traits exist, but they generally are unable to control for population structure[28,29], and often are not applicable to genome-wide data.

We validate our approach using extensive simulations based on available SNP data from *A. thaliana*[30], showing that our model increases power to detect associations while controlling the false discovery rate. We then demonstrate its usefulness by considering correlated blood lipid traits from the Northern Finland Birth Cohort 1966 (NFBC1966)[31] and environmental plasticity in an *A. thaliana* data set that contains flowering measurements for two simulated growth seasons in two different locations[32]. Finally, we discuss the usefulness of this approach, not only in terms of increasing power to detect associations, but also in terms of understanding the basic genetic architecture of the phenotypes.

[1]Gregor Mendel Institute, Austrian Academy of Sciences, Vienna, Austria. [2]Department of Molecular and Computational Biology, University of Southern California, Los Angeles, California, USA. [3]Institut National de la Recherche Agronomique (INRA), UR0588, Orléans, France. [4]These authors contributed equally to this work. Correspondence should be addressed to M.N. (magnus.nordborg@gmi.oeaw.ac.at).

Received 17 January; accepted 5 July; published online 19 August 2012; doi:10.1038/ng.2376

RESULTS

Simulations

Pairs of correlated phenotypes were simulated by adding phenotypic effects to genome-wide SNP data from *A. thaliana*[30]. A single randomly selected SNP was set to account for up to 2% of the phenotypic variance, but with the possibility of different effects in each of the two phenotypes. In addition, 10,000 SNPs were given much smaller effects to simulate the genetic background. A randomly chosen fraction of these background SNPs was shared between the two phenotypes, allowing for variation in the degree of phenotypic correlation (Online Methods and **Supplementary Fig. 1**).

We compared our ability to identify the focal locus using MTMM and marginal, single-trait analyses (using the smallest *P* value from the latter to ensure a fair comparison). Three different tests were used: a 'full test' that compares the full model, including the effect of the marker genotype and its interaction, with a model that includes neither; an 'interaction effect test' that compares the full model to one that does not include interaction; and a 'common effect test' that compares a model with a marker genotype to one without (see Online Methods for details). As expected, the results depended greatly on the effect of the focal polymorphism (**Fig. 1**). When this polymorphism had the same effect in both phenotypes (positive pleiotropy or a common effect across environments; **Fig. 1a**), MTMM performed slightly better than the single-trait mixed model, regardless of whether we tested for full model fit or just for a common effect (**Fig. 1e**). The reason for this is the increased power that results from analyzing the traits together. There is no rationale to testing for an interaction effect, as no interaction exists.

When the effect of the polymorphism is slightly weaker in one trait or environment (**Fig. 1b**), testing for a full model fit using MTMM again outperformed single-trait analyses (**Fig. 1f**). Testing only for a common or interaction effect using MTMM is also less effective. Although an interaction effect now exists, it is too weak to be detected. However, as the strength of the interaction effect increases (**Fig. 1c**,**d**), it becomes possible to detect directly, and the relative advantage of using MTMM increases markedly (**Fig. 1g**,**h**).

An alternative to carrying out two marginal single-trait analyses might be to combine the phenotypes, for example, by fitting the principal components of the traits or their sum or difference. We tested the latter, and, as might be expected, this approach worked very well when the focal SNP had exactly the same (or the opposite, when using the difference) effect on the phenotype (**Fig. 1a**,**d**). However, if the effect of the SNPs differs between the two traits, MTMM outperforms these approaches (**Fig. 1b**,**c**).

It should be noted that, because the background SNPs are correlated due to population structure, simple single-locus tests of association are strongly biased toward false positives, just as in the original data[14]. The mixed model effectively removes this bias, regardless of whether we analyze one phenotype at a time using a single-trait mixed model or both simultaneously using MTMM (**Supplementary Fig. 2**). However, analyzing these data with methods that do not take population structure into account is clearly not a realistic option (**Supplementary Fig. 3**).

In addition to the model just described, we simulated an oligogenic scenario in which each phenotype was determined by 20 loci, each of which could, with equal probability, affect (i) that phenotype only, (ii) both phenotypes in the same way or (iii) both phenotypes but in opposite ways. The behavior of each locus was chosen independently, and the resulting distribution of correlations between the phenotypes was thus centered on zero (**Supplementary Fig. 4**), which is very different from the positively correlated phenotypes generated under the first simulation scenario (**Supplementary Fig. 1**). MTMM is intended for correlated phenotypes and is expected to perform less well when phenotypes are weakly correlated. The oligogenic simulation results supported this intuition. For weakly correlated pairs of phenotypes, single-trait analysis often outperformed MTMM (especially in detecting SNPs with effect in one phenotype only); however, for more strongly correlated phenotypes, the results agreed with those presented above, in that MTMM always outperformed marginal analyses (**Supplementary Fig. 5**). Note that the correlation does not have to be positive: for negatively correlated phenotypes, MTMM has relatively higher power to detect SNPs with the same effect in both phenotypes, whereas for positively correlated phenotypes, it performs best for SNPs that have opposing effects (sometimes it may make sense to simply change the direction of correlation by negating one of the phenotypes when analyzing real data).

As noted, an advantage of MTMM is that it can be used for correlated phenotypes regardless of whether the phenotypes represent different measurements (and the correlations are due to pleiotropy) or the same trait measured in different environments (**Fig. 1a**–**d**). However, the simulations above assume that the phenotypic correlations are solely due to genetics and not environment, and this is only likely to be true for studies involving inbred lines in controlled environments. Certainly, correlations between pleiotropic traits will reflect environment as well as genotype. To verify that MTMM is able to separate these effects, we simulated another 5,000 pairs of correlated traits using the 10,000-loci model, but now with correlations reflecting environmental as well as genetic covariance (Online Methods). Both the environmental and genetic correlations were well estimated (**Supplementary Fig. 6**), although it should be noted that the residuals of the genetic and environmental correlation estimates

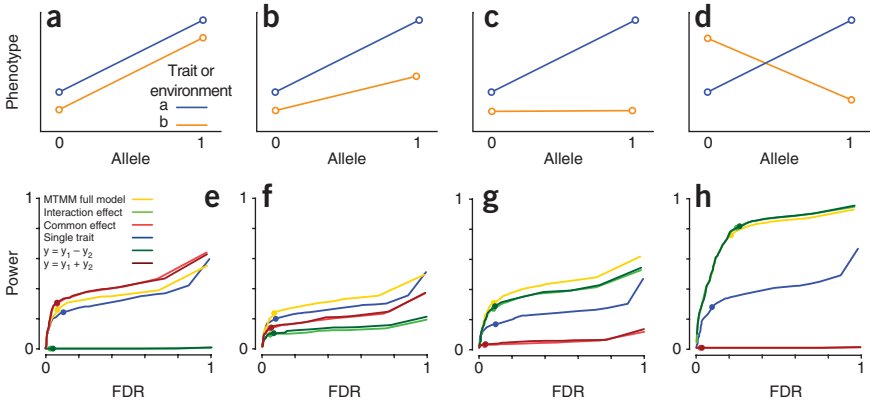

Figure 1 Simulation results. (**a**–**d**) Scenarios simulated—with positive pleiotropy, alternative common effect across environments (**a**); positive pleiotropy, alternative common effect across environments, with size of effect differing between traits and environments (**b**); effect only on one trait, alternative only in one environment (**c**); and negative pleiotropy, alternative opposite effect across environments (**d**). (**e**) Estimated relationship between power and false discovery rate (FDR) using six different statistical tests for the scenario in **a**. (**f**) Estimated relationship between power and FDR for the scenario in **b**. (**g**) Estimated relationship between power and FDR for the scenario in **c**. (**h**) Estimated relationship between power and FDR for the scenario in **d**. Dots on curves denote nominal Bonferroni-corrected 5% significance thresholds. Both power and FDR were calculated with respect to the single focal locus only.

Table 1 MTMM estimates of correlation and heritability in NFBC1966 data

		Genetic			Environmental			
	Phenotypic[a]	Corr.	s.e.m.	*P* value	Corr.	s.e.m.	*P* value	Heritability[b]
HDL-TG	–0.37	–0.42	0.14	0.024	–0.36	0.06	1.58×10^{-8}	0.38/0.18
HDL-LDL	–0.13	–0.19	0.11	0.085	–0.09	0.08	0.26	0.39/0.45
HDL-CRP	–0.19	0.24	0.23	0.25	–0.34	0.06	1.50×10^{-7}	0.39/0.14
TG-LDL	0.32	0.31	0.14	0.062	0.35	0.06	9.64×10^{-7}	0.19/0.44
TG-CRP	0.21	–0.50	0.39	0.115	0.34	0.05	3.19×10^{-9}	0.18/0.13
LDL-CRP	0.09	0.08	0.19	0.65	0.10	0.06	0.12	0.45/0.13

Corr., correlation.
[a]Direct estimates of the Pearson correlation are identical to the precision given. [b]The s.e.m. of all heritability estimates is between 0.05 and 0.06. Single-trait estimates are 0.38 (HDL), 0.18 (triglycerides, TG), 0.45 (LDL) and 0.13 (CRP).

are negatively correlated (**Supplementary Fig. 6d**). The accuracy of these estimates does affect the performance of GWAS, but the effect seems to be relatively minor (**Supplementary Fig. 7**).

Pleiotropy in human data

To show the usefulness of MTMM for traits that are correlated because they are part of the same biological system, we reanalyzed data from NFBC1966 (ref. 31) (see Online Methods for details). We focused on measurements of four blood metabolites that are strongly involved in cardiovascular heart disease[33], namely triglycerides, low-density lipoprotein (LDL), high-density lipoprotein (HDL) and C-reactive protein (CRP). These metabolites are significantly correlated, and MTMM analysis indicated that the correlations are caused by genetics as well as environment (**Table 1**), supporting the notion that these traits are mechanistically related and/or have linked causal loci. For HDL-CRP and triglycerides-CRP, the correlations of the genetic effects were in the opposite direction of the environmental correlations. However, in these cases, the genetic correlations are not significantly different from zero, and it is likely that the phenotypic correlations are driven primarily by the shared environment.

In terms of associations, the results from the joint analysis of triglycerides and LDL suffice to show how two of our main predictions were borne out. First, almost all SNPs that were found to be significantly associated in the marginal analysis of either LDL or triglycerides also had significant associations in the joint analysis (**Table 2**). However, MTMM arguably provides greater insight into the nature of the associations, as it reveals interaction effects. Second, MTMM finds associations that the marginal analyses do not. In particular, for positively correlated phenotypes such as triglyceride and LDL measures, we expect MTMM to have much greater power to detect polymorphisms whose effects differ greatly between the phenotypes. A good example of this is the *FADS1-FADS2* locus, which was not significantly associated in either marginal analysis but showed highly significant association using MTMM because of a very strong interaction effect (**Fig. 2** and **Table 2**). These genes are excellent candidates and were mentioned in the previous analysis of the NFBC1996 data[31]. Notably, they were also identified in a massive meta-analysis involving more than 100,000 individuals[34], which furthermore reported opposite effects on

Table 2 SNPs detected in the analysis of LDL and triglycerides using a genome-wide significance of 0.05

		MTMM (*P* value)[a]			Single-trait mixed model (*P* value)[a]	
SNP	Position	Full test	Interaction	Common	LDL	TG___
CELSR2 region, chromosome 1						
rs611917	109616775	**6.42×10^{-8}**	3.19×10^{-3}	7.72×10^{-7}	**1.80×10^{-8}**	0.46
rs646776	109620053	**2.48×10^{-15}**	1.42×10^{-6}	**3.28×10^{-11}**	**3.92×10^{-15}**	0.77
APOB region, chromosome 2						
rs10198175	20997364	6.32×10^{-7}	0.02	1.33×10^{-6}	**9.48×10^{-8}**	0.29
rs3923037	21011755	**6.39×10^{-9}**	0.13	**2.64×10^{-9}**	2.72×10^{-7}	7.17×10^{-7}
rs6728178	21047434	**9.57×10^{-10}**	0.11	**4.37×10^{-10}**	**7.95×10^{-8}**	1.81×10^{-7}
rs6754295	21059688	**1.31×10^{-9}**	0.14	**4.97×10^{-10}**	**7.10×10^{-8}**	4.12×10^{-7}
rs676210	21085029	**2.43×10^{-9}**	0.04	**2.56×10^{-9}**	7.23×10^{-7}	**9.21×10^{-8}**
rs693	21085700	**1.80×10^{-10}**	0.19	**5.00×10^{-11}**	**2.84×10^{-11}**	2.79×10^{-3}
rs673548	21091049	**1.63×10^{-9}**	0.04	**1.85×10^{-9}**	5.97×10^{-7}	**6.43×10^{-8}**
rs1429974	21154275	4.85×10^{-7}	0.02	1.06×10^{-6}	**7.69×10^{-8}**	0.24
rs754524	21165046	**5.30×10^{-8}**	0.02	**1.39×10^{-7}**	**7.83×10^{-9}**	0.17
rs754523	21165196	4.51×10^{-7}	0.02	1.01×10^{-6}	**7.15×10^{-8}**	0.24
GCKR region, chromosome 2						
rs1260326	27584444	**5.33×10^{-10}**	**2.10×10^{-8}**	7.73×10^{-3}	0.21	**1.87×10^{-10}**
rs780094	27594741	**5.98×10^{-9}**	**4.22×10^{-8}**	0.01	0.44	**3.15×10^{-9}**
LPL region, chromosome 8						
rs10096633	19875201	**2.42×10^{-8}**	**2.04×10^{-8}**	0.06	0.97	**1.93×10^{-8}**
FADS1 region, chromosome 11						
rs174537	61309256	**1.60×10^{-9}**	**9.02×10^{-9}**	0.01	6.82×10^{-6}	3.81×10^{-3}
rs102275	61314379	**8.79×10^{-10}**	**6.20×10^{-9}**	4.86×10^{-3}	4.13×10^{-6}	3.82×10^{-3}
rs174546	61326406	**5.52×10^{-10}**	**3.83×10^{-9}**	4.88×10^{-3}	3.69×10^{-6}	3.12×10^{-3}
rs174556	61337211	**2.56×10^{-9}**	**4.43×10^{-8}**	1.93×10^{-3}	2.03×10^{-6}	0.01
rs1535	61354548	**2.08×10^{-9}**	**1.35×10^{-8}**	0.01	6.04×10^{-6}	4.96×10^{-3}
rs2072114	61361791	**8.77×10^{-8}**	7.31×10^{-7}	4.77×10^{-3}	1.59×10^{-5}	0.03
LDLR region, chromosome 19						
rs11668477	11056030	**3.16×10^{-8}**	0.15	**1.18×10^{-8}**	**3.89×10^{-9}**	0.02
rs2228671	11071912	**7.20×10^{-8}**	5.30×10^{-4}	4.87×10^{-6}	**4.47×10^{-8}**	0.96

[a]*P* values below the Bonferroni-corrected 5% cutoff of 1.5×10^{-7} are highlighted in bold.

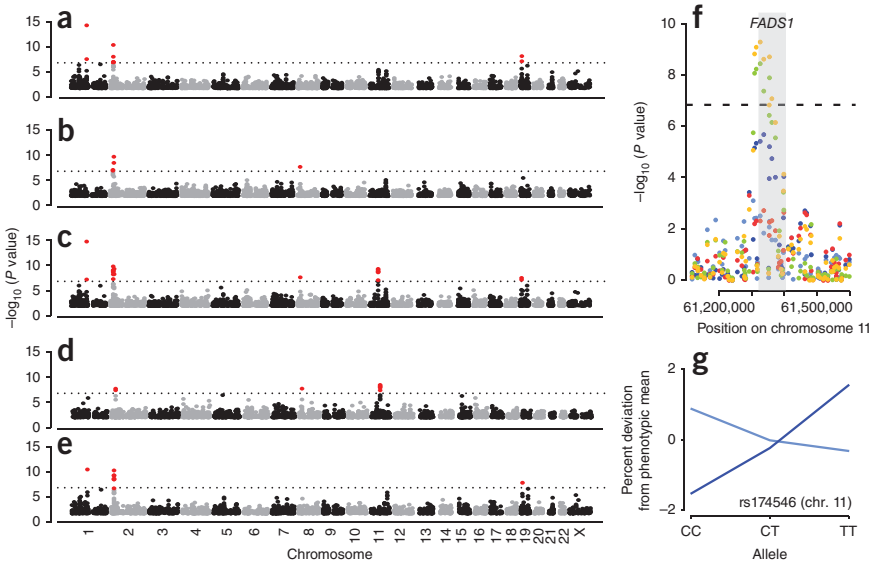

Figure 2 GWAS of LDL and triglycerides. (**a**,**b**) Manhattan plots for the marginal, single-trait analyses of LDL (**a**) and triglycerides (TG) (**b**). (**c**–**e**) Manhattan plots for the joint MTMM analyses with the full model (**c**), interaction effect (**d**) and common effect (**e**). The dotted horizontal lines denote the 5% Bonferroni-adjusted genome-wide significance level. (**f**) Enlarged view of the *FADS1-FADS2* region on chromosome 11. The points for the single-trait analyses are shown in light blue (TG) and dark blue (LDL), and the point for MTMM are shown in orange (full test), light green (interaction effect) and red (common effect). Gray shading denotes the *FADS1* gene region. (**g**) Estimated phenotypic effect of the rs174546 SNP is shown in light blue (TG) and dark blue (LDL).

triglycerides and LDL, in agreement with the strong interaction effect we observe using a sample of only 5,000 individuals (**Fig. 2g**).

The other five trait combinations gave similar results (**Supplementary Tables 1**–**5**). Almost all SNPs that were identified using single-trait analysis were also detected using MTMM, either due to a strong common or strong interaction effect. In addition, MTMM detected two more regions that were not identified using marginal analyses. First, the *PPP1R3B* gene was identified, due to strong common effects in the joint analyses of HDL-CRP and HDL-triglycerides. These pairs of phenotypes are negatively correlated (**Table 1**), so we expect MTMM to have increased power to detect common effects. The association with *PPP1R3B* was not found in previous analyses of these data[31] but was reported (and confirmed) in the much larger meta-analysis of blood lipid traits[34]. Second, the *TOM40-APOE* region was identified in the triglycerides-CRP joint analysis, this time due to an interaction effect (triglycerides and CRP are positively correlated). Albeit not quite significant, this association was noted in the previous analysis of these data[31] and was also found in the meta-analysis[34].

Genotype-environment interactions in *A. thaliana* data

The other natural application for MTMM is when phenotypes are correlated because they represent the same trait measured in different environments. In such a setting, one is often directly interested in finding genes that are involved in the differential response to the environment, that is, genotype-by-environment ($G \times E$) interactions. We tested this application using a data set from *A. thaliana* in which flowering time was measured (for a global collection of naturally occurring inbred lines) in environmental control chambers for two simulated seasons ('spring' and 'summer') and two simulated locations ('Spain' and 'Sweden')[32]. Flowering time varies in a clinal manner and is generally thought to be important in local adaptation. It is thus both natural and of interest to try identifying genes that are responsible for the differential flowering response to different environments[32].

We analyzed the *A. thaliana* data using a full 2 × 2 factorial model: in addition to estimating the effect of genotype, season and location, we have two pairwise interaction terms (Online Methods and **Supplementary Note**). The results are summarized in **Figure 3** (for details, see **Supplementary Figs. 8** and **9** and **Supplementary Tables 6** and **7**). Perhaps unexpectedly, we found very few interaction effects. Out of a total of 41 significant SNP associations, only 3 seemed to be caused by interactions. A rare allele (minor allele frequency (MAF) = 4%) on chromosome 5 was identified as having a significant genotype-by-season effect, but it does not correspond to any obvious candidate gene (**Supplementary Fig. 10**). A more convincing example was provided by the two tightly linked and perfectly correlated SNPs on chromosome 1. These were identified by comparing the full model to one without interaction terms, although the interaction with the simulated season seemed to be strongest (**Fig. 4**). The minor allele (MAF = 3%) was associated with delayed flowering (**Fig. 4b**), but the effects depend strongly on the season and are much more pronounced in the (simulated) summer. Notably, both SNPs are in the coding region of the *FRS6* gene, which is known to be involved in the phyA-mediated response to far-red light[35]. Knockout lines of this locus have an early-flowering phenotype, the magnitude of which depends on day length (one of the factors that varies between the simulated seasons).

Of the remaining 38 SNPs, 28 were found by both marginal and joint analysis (as common effects), and 10 were found only by marginal analysis. Although our simulations would seem to suggest that MTMM should always have higher power than marginal tests, even for detecting common or unique effects, this is clearly not always the case. The phenotypes analyzed here are extremely highly correlated as well as heritable (all coefficients are typically well above 0.9; **Supplementary Table 6**). In such cases, the advantage of increasing the sample size through joint analysis does not necessarily outweigh the cost of a more complex model with more degrees of freedom.

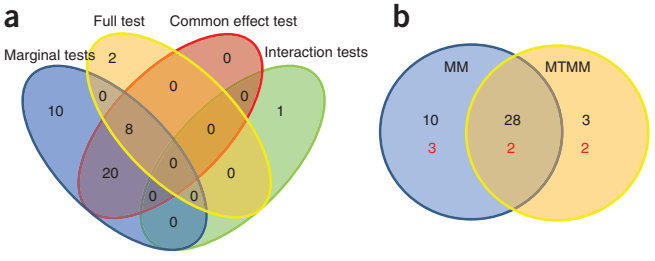

Figure 3 Venn diagrams summarizing the GWAS of *A. thaliana* flowering data[32]. (**a**) Classification of the 41 significantly associated SNPs according to the test(s) in which they were significant. (**b**) Classification of the 41 SNPs (black) and corresponding gene regions (red) according to whether they were found using marginal (mixed model, MM) or joint (MTMM) analysis.

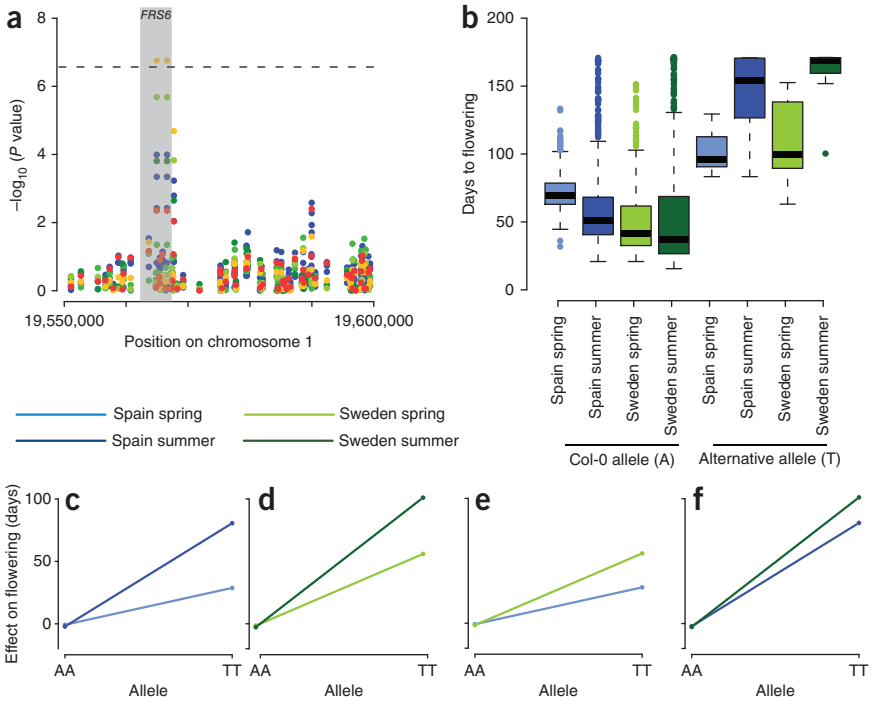

Figure 4 Summary of *FRS6* results. (**a**) Enlarged view of a 50-kb region on chromosome 1 showing significant genotype-environment associations. The *FRS6* gene is highlighted in gray. The results for the four marginal analyses (using a single trait, MM) are shown in blue, and the MTMM results are shown in orange (full test), light green (three-way interaction), green (genotype by location), dark green (genotype by season) and red (common effects). The horizontal line represents the 5% Bonferroni-corrected genome-wide significance threshold. (**b**) Phenotypic distribution shown as box plots, when partitioned by the genotype of the significantly associated SNP in **a** and by the environment. The colored boxes denote the first quantile (bottom), median value (thick black line) and the third quantile (top). The four colors represent the four different environmental conditions. Col-0 represents the *A. thaliana* reference strain. (**c**–**f**) Plots contrasting the allelic effect in different comparisons: the effect of the season in 'Spain' (**c**), the effect of the season in 'Sweden' (**d**), the effect of the location in 'spring' (**e**) and the effect of the location in 'summer' (**f**). The effect depends strongly on the season within each location (**c**,**d**) and less strongly on location with season (**e**,**f**).

DISCUSSION

We have shown how the classical mixed model from breeding may be used for GWAS of correlated phenotypes in structured populations, often providing greater statistical power than marginal analyses. However, we emphasize that our approach is much more than an *ad hoc* method for increasing power. The model we use effectively dates back to Fisher[36], and can be derived from basic genetic principles, under the assumption that heritable phenotypic variation is due to very large numbers of genes of very small effect (Online Methods). Assuming that this is a reasonable approximation (and it seems to be, for a growing number of traits), we can disentangle genetic correlations from environmental correlations, whenever these are uncorrelated. This allows us to address fundamental questions about the nature of variation.

When applied to traits that may be biologically related, the resulting variance component estimates allow us to assess the level of pleiotropy without estimating effects of individual loci. Using data from different human blood lipid measures, we demonstrated how the phenotype covariance can be decomposed into genetic and environmental terms, suggesting that most of these traits are indeed correlated due to shared genetics (they are pleiotropic or due to causal sites in linkage disequilibrium). A similar approach was recently used[37] to assess the heritability of RNA expression levels within and across human cell tissues.

Irrespective of this, we also showed increased power, detecting several interesting loci affecting human blood lipid level that were not significant in single-trait analysis but that have all been replicated in GWAS using much larger sample sizes. This finding alone strongly argues for routine application of our method to correlated phenotypes.

As an example of how the method can be used to detect environmental interactions, we applied our method to an *A. thaliana* flowering-time data set, where the plants had been phenotyped under four different environmental conditions (in a classical 2 × 2 factorial design). These phenotypes are highly correlated as well as highly heritable, and the estimated variance components suggest that there is in fact very little difference between the environments at the genetic level (**Supplementary Table 6**). Hence, it is arguably not unexpected that we detected little in terms of interaction effects. Although it is of course possible that we simply do not have the power to detect interactions, it is notable that analogous studies in maize have also been unable to detect large genome-environment interaction effects[38]. The results from *A. thaliana* and maize are strikingly different from what has been reported for mouse[39], yeast[40] and even humans[4], but the reasons for these differences are far from clear, given the dramatically different study designs.

Full factorial designs with replicated genotypes are of course not possible in most organisms; however, we note that MTMM does not require this. Indeed, a mixed-model approach has previously been proposed for estimating genome-environment variance components in humans[25] (using a special case of our model in which heritabilities are assumed to be equal across environments; Online Methods). Either approach is directly applicable to human data.

Although we have focused on relatively simple pairwise correlations in this paper, it is easy to model more than two phenotypes using MTMM. Conceptually, we believe that extending this approach to larger multi-trait experiments should allow for greater benefits in estimating error terms and elucidating functional relationships between suites of traits. However, for such complex models, the computational complexity grows fast, and the results become increasingly difficult to interpret compared to sequential two-trait analyses.

This is a well-known problem in statistics and quantitative genetics, but MTMM has the additional caveat that it assumes that the increasingly complex covariance structure, which is estimated in the absence of fixed effects, remains constant as these are added. Various intermediate approaches are possible, for example, variance components might be estimated using a full model once, followed by GWAS using submodels; more work in this area is clearly desirable.

Finally, when the phenotypes are not correlated or if the correlation is not due to genetics (something that can be deduced from the variance component estimates), a single-trait mixed model will generally have greater power to detect causal loci that are phenotype specific. When this will be the case precisely is hard to predict; however, we suggest using

the MTMM approach as a complement to, rather than replacement for, marginal GWAS. The advantages are clear: it allows the detection of both interactions and pleiotropic loci in a rigorous statistical framework while simultaneously accounting for population structure.

URLs. MTMM has been implemented in a set of R scripts (MTMM) for carrying out GWAS. These scripts rely on the software ASREML[41] for the estimation of the variance components. The scripts can be obtained https://cynin.gmi.oeaw.ac.at/home/resources/mtmm.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

We thank the NFBC1966 Study Investigators for allowing us to use their phenotype and genotype data in our study. The NFBC1966 Study is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with the Broad Institute, the University of California, Los Angeles (UCLA), the University of Oulu and the National Institute for Health and Welfare in Finland. This manuscript was not prepared in collaboration with investigators of the NFBC1966 Study and does not necessarily reflect the opinions or views of the NFBC1966 Study Investigators, the Broad Institute, UCLA, the University of Oulu, the National Institute for Health and Welfare in Finland or the NHLBI. We furthermore thank N.B. Freimer and S.K. Service for their help in preprocessing the NFBC1966 data. We would also like to thank P. Forai for excellent IT and cluster support at the Gregor Mendel Institute, the INRA MIGALE bioinformatics platform for further computational resources and J. Dekkers, P. Donnelly, E. Eskin, C. Niango and A. Price for comments on the manuscript and/or helpful discussions. This work was supported by grants to M.N. from the US National Institutes of Health (P50 HG002790) and the European Union Framework Programme 7 (TransPLANT, grant agreement 283496), as well as by grants from the Deutsche Forschungsgemeinschaft (DFG) (A.K., KO4184/1-1) and the Ecologie des Forêts, Prairies et milieux Aquatiques (EFPA) department of INRA (V.S.).

AUTHOR CONTRIBUTIONS

All authors helped design the study. A.K., B.J.V. and V.S. developed the theory and implemented the simulations. A.K., B.J.V. and M.N. wrote the paper with input from V.S., A.P. and Q.L.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/doifinder/10.1038/ng.2376.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Platt, A., Vilhjálmsson, B.J. & Nordborg, M. Conditions under which genome-wide association studies will be positively misleading. *Genetics* **186**, 1045–1052 (2010).
2. Kang, H.M. *et al.* Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* **42**, 348–354 (2010).
3. Price, A.L., Zaitlen, N.A., Reich, D. & Patterson, N. New approaches to population stratification in genome-wide association studies. *Nat. Rev. Genet.* **11**, 459–463 (2010).
4. Hamza, T.H. *et al.* Genome-wide gene environment study identifies glutamate receptor gene *GRIN2A* as a Parkinson's disease modifier gene via interaction with coffee. *PLoS Genet.* **7**, e1002237 (2011).
5. Lynch, M. & Walsh, B. *Genetics and Analysis of Quantitative Traits* (Sinauer Associates, Sunderland, Massachusetts, 1997).
6. Jiang, C. & Zeng, Z.B. Multiple trait analysis of genetic mapping for quantitative trait loci. *Genetics* **140**, 1111–1127 (1995).
7. Ferreira, M.A. & Purcell, S.M. A multivariate test of association. *Bioinformatics* **25**, 132–133 (2009).
8. Zhang, L., Pei, Y.F., Li, J., Papasian, C.J. & Deng, H.W. Univariate/multivariate genome-wide association scans using data from families and unrelated samples. *PLoS ONE* **4**, e6502 (2009).
9. Knott, S.A. & Haley, C.S. Multitrait least squares for quantitative trait loci detection. *Genetics* **156**, 899–911 (2000).
10. Henderson, C.R. *Application of Linear Models in Animal Breeding* (University of Guelph, Guelph, Canada, 1984).
11. Thomas, D. Gene–environment-wide association studies: emerging approaches. *Nat. Rev. Genet.* **11**, 259–272 (2010).
12. Ober, C. & Vercelli, D. Gene-environment interactions in human disease: nuisance or opportunity? *Trends Genet.* **27**, 107–115 (2011).
13. Yu, J. *et al.* A unified mixed model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* **38**, 203–208 (2006).
14. Atwell, S. *et al.* Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* **465**, 627–631 (2010).
15. Huang, X. *et al.* Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* **42**, 961–967 (2010).
16. Olsen, H.G. *et al.* Genome-wide association mapping in Norwegian Red cattle identifies quantitative trait loci for fertility and milk production on BTA12. *Anim. Genet.* **42**, 466–474 (2011).
17. Tian, F. *et al.* Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nat. Genet.* **43**, 159–162 (2011).
18. Zhao, K. *et al.* An *Arabidopsis* example of association mapping in structured samples. *PLoS Genet.* **3**, e4 (2007).
19. Kang, H.M. *et al.* Efficient control of population structure in model organism association mapping. *Genetics* **178**, 1709–1723 (2008).
20. Zhang, Z. *et al.* Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* **42**, 355–360 (2010).
21. Idaghdour, Y. *et al.* Geographical genomics of human leukocyte gene expression variation in southern Morocco. *Nat. Genet.* **42**, 62–67 (2010).
22. International Multiple Sclerosis Genetics Consortium and Wellcome Trust Case Control Consortium 2. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* **476**, 214–219 (2011).
23. Stich, B., Piepho, H.P., Schulz, B. & Melchinger, A.E. Multitrait association mapping in sugar beet (*Beta vulgaris* L.). *Theor. Appl. Genet.* **117**, 947–954 (2008).
24. Lee, S.H., Wray, N.R., Goddard, M.E. & Visscher, P.M. Estimating missing heritability for disease from genome-wide association studies. *Am. J. Hum. Genet.* **88**, 294–305 (2011).
25. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
26. Lee, S.H. *et al.* Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nat. Genet.* **44**, 247–250 (2012).
27. Deary, I.J. *et al.* Genetic contributions to stability and change in intelligence from childhood to old age. *Nature* **482**, 212–215 (2012).
28. Kim, S. & Xing, E.P. Statistical estimation of correlated genome associations to a quantitative trait network. *PLoS Genet.* **5**, e1000587 (2009).
29. Manning, A.K. *et al.* Meta-analysis of gene-environment interaction: joint estimation of SNP and SNP × environment regression coefficients. *Genet. Epidemiol.* **35**, 11–18 (2011).
30. Horton, M.W. *et al.* Genome-wide patterns of genetic variation in worldwide *Arabidopsis thaliana* accessions from the RegMap panel. *Nat. Genet.* **44**, 212–216 (2012).
31. Sabatti, C. *et al.* Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nat. Genet.* **41**, 35–46 (2009).
32. Li, Y., Huang, Y., Bergelson, J., Nordborg, M. & Borevitz, J.O. Association mapping of local climate-sensitive quantitative trait loci in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **107**, 21199–21204 (2010).
33. Kathiresan, S. *et al.* A genome-wide association study for blood lipid phenotypes in the Framingham Heart Study. *BMC Med. Genet.* **8** (suppl. 1) S17 (2007).
34. Teslovich, T.M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).
35. Lin, R. & Wang, H. *Arabidopsis FHY3/FAR1* gene family and distinct roles of its members in light control of *Arabidopsis* development. *Plant Physiol.* **136**, 4010–4022 (2004).
36. Fisher, R. The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edinburgh* **52**, 399–433 (1918).
37. Price, A.L. *et al.* Single-tissue and cross-tissue heritability of gene expression via identity-by-descent in related or unrelated individuals. *PLoS Genet.* **7**, e1001317 (2011).
38. Buckler, E.S. *et al.* The genetic architecture of maize flowering time. *Science* **325**, 714–718 (2009).
39. Valdar, W. *et al.* Genetic and environmental effects on complex traits in mice. *Genetics* **174**, 959–984 (2006).
40. Smith, E.N. & Kruglyak, L. Gene-environment interaction in yeast gene expression. *PLoS Biol.* **6**, e83 (2008).
41. Gilmour, A., Gogel, B., Cullis, B., Welham, S.J. & Thompson, R. *ASReml User Guide Release 1.0* (VSN International, Hemel Hempstead, UK, 2002).

ONLINE METHODS

Theory *Multiple-traits mixed model*. Following Henderson[10], we can write the mixed model for the phenotypes of n individuals as

$$\mathbf{y} = X\beta + \mathbf{g} + \mathbf{e} \tag{1}$$

where $\mathbf{y}$ is a vector of n phenotype values. In this notation, the trait mean is included, together with other fixed effects, in the design matrix X. β represents the effect size of the fixed effects, $\mathbf{g} \sim N\left(0, \sigma_g^2 K\right)$ is a random effect and $\mathbf{e} \sim N\left(0, \sigma_e^2 I\right)$. It follows that the covariance matrix for the trait values $\mathbf{y}$ is

$$\mathrm{var}\left(\mathbf{y}\right) = \sigma_g^2 K + \sigma_e^2 I \tag{2}$$

where K is an $n \times n$ kinship or relatedness matrix. If we consider two traits, $\mathbf{y}_1$ and $\mathbf{y}_2$, measured on the same set of individuals, then under the mixed model for the kth phenotype follows the partitions of the variance accordingly: $\mathrm{var}\left(\mathbf{y}_k\right) = \sigma_{gk}^2 K + \sigma_{ek}^2 I$. However, for the covariance matrix between the two phenotypes, it is not obvious what the appropriate model is. Henderson[42] suggests the covariance model

$$\mathrm{cov}\left(\mathbf{y}_1, \mathbf{y}_2\right) = \sigma_{g1}\sigma_{g2}\rho_g K + \sigma_{e1}\sigma_{e2}\rho_e I \tag{3}$$

Where ρ_g captures the genetic correlation between two phenotypes and the term ρ_e captures the correlation caused by shared environment and other nongenetic sources of correlation.

We can generalize this for phenotypes that have been measured for different sets of individuals (**Supplementary Note**).

Estimating the variance parameters. The estimation procedure for the variance components is described in the **Supplementary Note**.

Application to GWAS. As in EMMAX[2] or P3D[20], we estimate the covariance matrix only once to re-estimate a scalar in front of it for every marker. This fixes five degrees of freedom out of six in total (maximum number of variance components for two traits). For a pair of traits (the ith and jth traits), the proposed approximation effectively assumes that the three variance ratios (σ_{gi}/σ_{ei}, σ_{gj}/σ_{ej} and σ_{gi}/σ_{gj} and the two correlations ρ_{gi} and ρ_{gj} are fixed with and without the marker in the model.

With multiple traits, we can search for causal loci with common effects (across all traits) as well as trait-specific loci or loci with opposite effects for different traits. Depending on what we are interested in, a generalized least squares (GLS) F test can be constructed to compare two models. For two traits, we can write the single marker model as

$$\mathbf{y} = \begin{bmatrix}\mathbf{y}_1 \\ \mathbf{y}_2\end{bmatrix} = \mathbf{s}_1\mu_1 + \mathbf{s}_2\mu_2 + \mathbf{x}\beta + \left(\mathbf{x} \times \mathbf{s}_1\right)\alpha + \boldsymbol{\upsilon} \tag{4}$$

where $\mathbf{x}$ is the marker and $\mathbf{s}_i$ is a vector of 1 for all values belonging to the ith trait and 0 otherwise. $\boldsymbol{\upsilon} \sim N\left(0, \mathrm{cov}\left(\mathbf{y}\right)\right)$ is a random variable capturing both the error and genetic random effects. Depending on what kind of loci we are interested in, we propose three different F tests.

1. The full model tested against a null model where $\beta = 0$ and $\alpha = 0$. This identifies both loci with common and differing effects in one model but suffers in power from the extra degree of freedom.
2. To identify common genetic effects, we propose to test the genetic model ($\alpha = 0$) against a null model where $\beta = 0$ and $\alpha = 0$.
3. Finally, to identify differing genetic effects between the traits, we propose to test the full model against a null model where $\alpha = 0$.

As both the interaction test and the common effect test are sensitive to scaling of the phenotype values, we propose to normalize them either by the total variance or the genetic variance (as obtained in marginal trait analysis). To minimize multiple-testing problems, one could, for example, carry out GWAS using the full model and then use the other tests to analyze associated loci further.

Extending this model for an arbitrary number of traits is straightforward (one example for the analysis of four traits is described in the **Supplementary Note**). However, when there are more than two traits in the model, the number of possible tests grows quickly. A noteworthy special case is when there are several environmental variables in a factorial study design, in which case each environmental variable can be included in the model instead of the term

$$\sum_{i=1}^{t} \mathbf{s}_i\mu_i$$

and their interactions with the genotype could replace the term

$$\mathbf{x}\left(\sum_{i=1}^{t} \mathbf{s}_i\alpha\right)$$

This can result in a simpler and a more tractable model than if all possible combinations of environments were treated as independent.

Genotype-environment interactions. Given two measured phenotype vectors, $\mathbf{y}_1$ and $\mathbf{y}_2$, Yang *et al.*[25] include a $G \times E$ random effect in a mixed model as

$$\mathbf{y} = \begin{bmatrix}\mathbf{y}_1 \\ \mathbf{y}_2\end{bmatrix} = X\beta + u_G + u_{G\times E} + \mathbf{e} \tag{5}$$

Where μ_G and $\mu_{G\times E}$ are random effects and have covariance matrices as follows:

$$\begin{aligned}\mathrm{cov}\left(\begin{bmatrix}\mathbf{y}_1 \\ \mathbf{y}_2\end{bmatrix}\right) &= \mathrm{cov}\left(u_G\right) + \mathrm{cov}\left(u_{G\times E}\right) + \mathrm{cov}\left(\mathbf{e}\right) \\ &= \sigma_G^2\begin{bmatrix}K_{11} & K_{21} \\ K_{12} & K_{22}\end{bmatrix} + \sigma_{G\times E}^2\begin{bmatrix}K_{11} & 0 \\ 0 & K_{22}\end{bmatrix} + I\end{aligned} \tag{6}$$

Compared to the model proposed in equation (4), this model implicitly assumes two things: (i) that there are no environmental correlations and (ii) that the heritabilities are the same in each environment, such that $h_1^2 = h_2^2$. As the individuals are different in each environment, the first assumption is appropriate. However, the second assumption is not guaranteed to hold in general, and we therefore propose relaxing it.

Simulations. *10,000-loci model*. We simulated 2,000 pairs of correlated phenotypes using a model under which the phenotypes consisted of one randomly chosen SNP with a 'large' (additive) effect, accounting for up to 2% of the total phenotypic variance, and 10,000 randomly chosen SNPs with small additive effects. The effects sizes were drawn from a normal distribution, and a random error was added to fix the trait heritability to 0.95. To ensure variation in trait correlations, all trait pairs shared a random fraction of the 10,000 causal loci, with the fraction drawn from a uniform distribution. The four phenotypic models were distinguished by different effect correlations at the major locus (**Fig. 1**).

In addition, we simulated 5,000 pairs of correlated traits with environmental correlations. We fixed the heritability to 0.5 and allowed the genetic correlation to vary from –1 to 1. Additionally, we added a shared environmental term to the model, mimicking scenarios for both negative and positive environmental correlation.

20-loci model. We also simulated 1,000 pairs of correlated phenotypes using a 20-loci model. Each phenotype was determined by 20 SNPs, where each SNP was randomly assigned to one of the following three categories with equal probabilities: (i) SNPs with same effect in both phenotypes, (ii) SNPs with opposite effect in the two phenotypes and (iii) SNPs with effect in one trait only. The SNPs had additive effects drawn from an exponential distribution. Finally, a random error was added to fix the heritabilities to 0.95. To obtain a single P value for two traits, the smaller of the two P values for each SNP from the marginal mixed model analysis was retained.

Power calculations. For calculation of the power and FDR, any significantly associated SNP within 50 kb of any (or the) causal SNP was classified as a true positive; otherwise, it was classified as a false positive. The results were almost independent of the window size used (**Supplementary Fig. 11**). More important is the effect of the causal SNP(s). The nearly twofold increase of power observed at an FDR of 0.1 in **Figure 1** depended on the effect size of the simulated SNP (**Supplementary Fig. 12**). Throughout this paper, we used the single-analysis Bonferroni-corrected 5% significance threshold.

Human data. We used results from the NFBC1966, which consist of phenotypic and genotypic data for 5,402 individuals[31]. Using the exact same data set as was used in ref. 2, after filtering, the data set consisted of

doi:10.1038/NG.2376

5,326 individuals and 331,475 SNPs. To expedite mapping, unknown genotypes (<1% of the data set) were imputed by replacing missing values with the average genotypic value. Neither the marginal mixed model analysis nor the MTMM tests showed evidence of confounding due to population structure (**Supplementary Fig. 13**).

Analysis of A. thaliana data. The genotype data for *A. thaliana* consisted of 1,307 individuals genotyped at 214,051 SNPs using a custom Affymetrix SNP chip[30]. The phenotypes used were measurements of flowering time for 459 accessions[32]. Flowering time was measured in plants grown in four different environments, a factorial setting with two simulated seasons (spring and summer) and two simulated locations (Spain and Sweden). Analyzing the four phenotype vectors together, we can derive five different *F* tests (**Supplementary Note**). None of these tests showed evidence of confounding due to population structure (**Supplementary Fig. 14**).

42. Henderson, C. & Quaas, R.L. Multiple trait evaluation using relatives' records. *J. Anim. Sci.* **43**, 1188–1197 (1976).