bioRxiv preprint doi: https://doi.org/10.1101/2022.05.19.492732; this version posted May 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

The limitations of phenotype prediction in metabolism

Pablo Yubero, Alvar A. Lavin, Juan F. Poyatos[1]

[1]Logic of Genomic Systems Lab (CNB-CSIC) Madrid. Email: jpoyatos@cnb.csic.es

ABSTRACT

Phenotype prediction is at the core of many questions in biology. Prediction is frequently attained by determining statistical associations between genetic and phenotypic variation, ignoring the exact processes that lead to the phenotype. Here, we present a framework based on genome-scale metabolic reconstructions to reveal the mechanisms behind the associations. We compute a polygenic score (PGS) that identifies a set of enzymes as predictors of growth, the phenotype. This set arises from the synergy of the functional mode of metabolism in a particular environment and its evolutionary history, and is transportable to anticipate the phenotype across a range of environments. We also find that there exists an optimal genetic variability for predictability and demonstrate how the linear PGS can yet explain phenotypes generated by the underlying nonlinear biochemistry. Thus, the explicit model interprets the black-box statistical associations of the genotype-to-phenotype map and uncovers the limits of prediction in metabolism.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.19.492732; this version posted May 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

INTRODUCTION

By understanding the factors that specify the phenotype, we aim to recognize how heritable biological information eventually maps onto action and how this mapping evolves (Waddington 2015). These issues illustrate the more general question of the emergence of function in complex systems, and the inherent attributes of these systems that help them elude function prediction (Orrell 2007). We will discuss here the properly biological case of metabolism.

The variational method of quantitative genetics symbolizes a traditional approach to this question. Its goal is to establish statistical associations between the genetic and phenotypic variation observed within a certain population (Lynch and Walsh 1998). When this *genotype-to-phenotype* (GP) map becomes determined in a supervised situation, it is then possible to develop tools that anticipate the phenotype of individuals based solely on their genetic sequence (Dudbridge 2013). What are the limits of this approach?

Valid as they are, the statistical associations of quantitative genetics depend very much on the features of the trait, the population context, and the environmental conditions under which they are identified (Zaidi and Mathieson 2020). They thus represent a kind of "black-box" expectation that does not provide any insights into the processes leading to a particular phenotype (Cannon and Mohlke 2018). This absence of mechanism has both basic and applied implications.

From the fundamental point of view, many features that define the genetic architecture of phenotypes (dominance, epistasis, etc.), while having a clear variational definition, present a less clear mechanistic interpretation (Keightley and Kacser 1987; Omholt et al. 2000). An interpretation that should also help explain how the nonlinearity that seems dominant in many biological systems does not limit the power of the –linear– statistical procedures (Feldman and Lewontin 1975).

From the applied point of view, consider, for instance, the case of genome-wide association studies (GWAS) in humans. The original purpose of GWAS was to identify the causal genetic determinants of complex phenotypes, including diseases. This plan turned out to

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.19.492732; this version posted May 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

be more complicated than expected (Visscher et al. 2017), with recent studies emphasizing the complex pleiotropic regulation of most human traits (Boyle et al. 2017; Wray et al. 2018). Similarly, while specific prediction tools are indeed available, e.g., the development of polygenic risk scores to indicate a predisposition to disease (Torkamani et al. 2018), we are still far from unraveling which are the biological foundations behind their successes and failures.

Indeed, discovering mechanistic insights behind these GP associations has proven to be a significant challenge, owing partly to the large quantity of accepted causal elements distinguished for most phenotypes. For instance, human quantitative traits were linked to only a few strong-effect determinants not long ago; a hypothesis that is now abandoned (Manolio et al. 2009; Boyle et al. 2017; Wray et al. 2018). A second factor is that natural selection weakens the impact of the *a priori* strongest statistical predictors (O'Connor et al. 2019). Most significant of all is the absence of an underlying developmental or physiological model explaining the emergence of the phenotype (Cannon and Mohlke 2018). Therefore, it is interesting to examine situations in which an explicit model replaces the black box and, in this way, one can better explain the causal characteristics.

There have been several attempts in this respect. Plant biology has pioneered works to connect gene network modeling with quantitative genetics, for example, on the prediction of flowering time (Welch et al. 2005). Other computational efforts to relate explicit phenotypic models and genetic variation include the cases of foliate-mediated one-carbon metabolism (Nijhout et al. 2017), single heart cells (Wang et al. 2012), or tooth development (Milocco and Salazar-Ciudad 2020).

In this manuscript, we follow the approach of quantitative genetics to study the emergence of statistical associations and the potential of phenotypic prediction in metabolism. While metabolism has historically been utilized to examine questions on the link between genetic variability and system-level organization since early discussions on aspects of the architecture of biological systems, see (Burns 1970; Kacser and Burns 1981; Clark 1991) to name a few,

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.19.492732; this version posted May 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

these works employed toy models and could not benefit from the current accessibility of genome-scale metabolic network reconstructions.

Here, we explicitly use a reconstruction of *Saccharomyces cerevisiae* (Duarte 2004) to develop a variational approach in which we can study how phenotype variability emerges. Genome-scale models contain all the known metabolic reactions in an organism and the genes encoding each enzyme and enable the prediction of metabolic phenotypes, e.g., biomass, under situations in which genetics and environment can be controlled. Within this framework, we first discuss the concept of polygenic score to then examine the underlying biology behind its operation. This goal will make us explore the influence of the systems' architecture, genetic variability, and gene-environment interactions. We demonstrate that the balance between functional mode and evolutionary history is crucial in revealing the limits of phenotype prediction with many implications.

RESULTS

Engineering genetic variation in metabolism

The genetic variation that exists in natural populations represents a multifactorial perturbation that enables us to understand biological processes (Rockman 2008). Quantitative genetics employs this perturbation to quantify statistical associations with phenotypes with the use of a reference, or training, population of known phenotype. The approach leads to the so-called polygenic scores (PGSs, Fig. 1A), which can predict the manifestation of a particular trait given detailed genetic factors.

Our first objective will be to *engineer* the variational approach in the *in silico* metabolic framework. We consider whole-genome metabolic reconstructions and generate variability in gene dosages as a result of the genetic variability in the population. Dosages are relative to a maximal reference value resulting from the history of yeast metabolism. This reduced enzymatic performance is in line with earlier works (Kacser and Burns 1981; Keightley and Kacser 1987; Clark 1991) and are later interpreted quantitatively in the model by gene-reaction rules. These rules define which (and how) genes participate in reactions (Fig. 1B,

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.19.492732; this version posted May 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Fig. S1 and Methods).

The engineered variability induces individual differences in any potential metabolic trait, but we focus on the biomass production rate corresponding to the growth rate that is computed through flux balance analysis (FBA, Fig. 1B; Palsson 2006). Therefore, the entire procedure generates a data set of both genetic and phenotypic variations in the context of a metabolic model, which we can dissect to explain how the system works as a whole (Fig. 1C).

A small subset of genes anticipates growth within a metabolic polygenic score

We next derive a multidimensional PGS employing data of a population of yeast metabolisms that exhibit variability in their gene dosages as described before (growing in the standard medium, Methods). Such variability induces variation in growth and in many metabolic fluxes (Figs. 2AB and Fig. S2). We obtain a PGS aimed at estimating the individual growth rate (Methods). Figure 2C compares the growth rate prediction to the values computed with FBA for this training data set. The PGS successfully anticipates the phenotype with an $R^2_{\text{train}}\sim 0.27$, from here on abbreviated R^2.

Although this situation differs from those typically observed in association studies –where the number of predictors is several orders of magnitude larger than the training population size–, it is exemplary in that we can easily overcome data shortcomings and hence the exact fit to the training data (statistical overfitting). That said, PGSs tend to lose predictive power when applied to a different *test* population. To evaluate this, we systematically generated 10^4 independent test data, each with the same size and mutational distribution as the training population. For each test we computed the R^2_{test} to obtain a mean $\langle R^2_{\text{test}}\rangle = 0.24 \pm 0.01$ for the full set. This reveals that the computed PGS slightly overfitted the training data, and that we should expect a small loss of predictability when applied to different test populations ($R^2 > \langle R^2_{\text{test}}\rangle$ with $p < 4\text{x}10^{-3}$).

The derived PGS includes all metabolic genes and their corresponding effect sizes, β, coupled to the prediction (Methods). We identify 85 genes of non-null effect, 32 of which are comparatively large ($|\beta| > 0.01$; Fig. 2DE). The latter impacts 61 metabolic reactions (out of

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.19.492732; this version posted May 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

1148). The large number of genes with null effect size is surely related to the fact that most of the fluxes are inactivated in the considered standard medium (individual metabolisms within the population typically show ∼73% of null fluxes) something that characterizes a distinct metabolic "functional mode" (a relevant notion in what follows). From now on, we focus on the subset of genes with larger effect sizes.

Few metabolic functions limit growth

What type of functions implement the predictor genes? One could think that predictors are distributed across all metabolic activities in the sense of universal pleiotropy (Kacser and Burns 1981). However, we only find a few metabolic subsystems enriched by predictors (Methods; Fig. S3). While these include the metabolisms of a variety of amino acids (valine, lysine, histidine), fatty acids, and phospholipids, it is surprising the absence of other subsystems central to metabolism like glycolysis or the citric acid cycle. These results are substantiated with a separate GO enrichment analysis (Table S1).

We observed that the detected subsystems involve specifically the production of biomass precursors. This group of metabolites fuels the biomass reaction, which defines the architecture of growth –as the trait of interest– in metabolic reconstructions (Fig. S4 shows its stoichiometry; this incorporates, for instance, the crucial role of amino acids and phospholipids for protein synthesis and the cell membrane, respectively, etc.). Consequently, we next hypothesized that the relevance of genetic predictors stems from their direct contribution to the pool of biomass precursors (Fig. 3A).

Figure 3B shows the mean aggregate metabolite production (or consumption, if negative) associated with all predictors in the population of metabolisms (Methods; Fig. S2B and S5). The strongest predictors only contribute significantly to a subset of precursors (11 out of 43), and in some cases, e.g., valine, lysine, etc., this represents the total production that is required for growth. Therefore, the functional mode, active on the standard medium, effectively selects a domain within the entire architecture of growth. Of note, while the contribution to these factors is *direct* for only 9 genes –that are precisely producing these biomass precursors–,

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.19.492732; this version posted May 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

the rest of the predictors have an effect on growth in a somehow distributed manner.

The case of histidine is exemplary (Fig. 3C). While it is produced only by *his4*, all histidine-related genes are crucial to providing intermediary precursors and thus are strongly involved in its overall production. This explains the case of *pmi40*, *sec53*, *dpm1* and *psa1*, which are predictor genes found upstream of the production of mannan, another important biomass precursor, while *pmt1-6* that ultimately produce said metabolite have null effect sizes. The same occurs with *erg4* and the production of sterol.

Pleiotropy is not a good measure of growth predictor character

The results before underline that the top predictors include genes that directly alter the availability of the limiting biomass precursors and also genes whose impact comes through other upstream reactions. Could the systemic properties of metabolism capture this second aspect? We consider here first the pleiotropic character of a gene. One quantifies pleiotropy in metabolic models as the number of biomass precursors whose maximal production becomes reduced by changing the dosage of a gene (Shlomi et al. 2007). The score thus includes system-wide phenomena like metabolic compensation, rewiring, redundancy, etc.

Within the highly pleiotropic genes, only some display large effect sizes on the PGS (Fig. 4A). This outcome might indicate that not all biomass precursors incorporated in the pleiotropic score limit growth (they do not all contribute to the operational functional mode in the standard medium). Indeed, we confirm that our set of predictors are all especially pleiotropic towards the group of biomass precursors already identified in the previous section, namely a few amino acids, phospholipids, mannan and sterol (Fig. 4B). This reflects that pleiotropy fails at pinpointing relevant genes for growth prediction since it is an aggregate measure that includes the effect of a mutation across *all* biomass precursors, while only a limiting few ones matter. On the contrary, separating the individual contributions of mutations to different functions results in a valid list of metabolites that potentially limit growth.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.19.492732; this version posted May 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Growth predictors display either large additive or epistatic effects

A second systemic measure is epistasis (indicating gene-gene interactions). We present an approach based on *global* sensitivity analysis that allows quantifying the additive and nonadditive effect of individual gene dosages on growth variability. Two indices, S_0 and ϵ_T, quantify such additive and nonadditive ("total order" epistasis) contributions, respectively (Methods and Supplementary Material).

Figure 4C shows that some predictor genes have large additive effects, S_0, and small total epistasis, ϵ_T (*ino1*, *his3*, ...), whereas others display the opposite pattern (*cho1*, *his1*). Importantly, the sum of all effects, additive and epistatic, shows the maximum correlation with the effect sizes obtained in the PGS (Pearson's $\rho > 0.97$, Fig. 4D and Fig. S6). That the fraction of genes displaying $S_0 > \epsilon_T$ and $S_0 < \epsilon_T$ is comparable highlights that the large effect sizes we obtain are associated with genes enriched by additive effects (something to be expected from a linear statistical formalism) but also with those with strong epistatic effects (which appears paradoxical, see the Discussion; see also Figs. S7 and S8, and Supplementary Material for an alternative sensitivity analysis where the growth response coefficients are analyzed).

Is there a structural basis for large S_0 or ϵ_T? We investigate their relationship with several measures for each gene: the number of (active) reactions involved, the amount of flux they control, and the number of (precursor) metabolites they utilize. Among these, we find the largest correlations of S_0 (and ϵ_T) with the number of reactions they control, $\rho = -0.19$ ($\rho = 0.19$), and the log of the summed absolute flux through their reactions, $\rho = 0.23$ ($\rho = -0.22$). Therefore, one expects large additive effects to stem from genes that regulate a small number of reactions of larger flux. In opposition, genes that control a larger number of reactions with less flux result in larger epistatic effects.

Optimal genetic variability for predictability

Beyond the previous systemic attributes, we now examine how the genetic variability accessible within a population, measured by its standard deviation σ_G, modifies the capacity

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.19.492732; this version posted May 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

to predict the phenotype. We use ten populations with equal mean dosage and increasing σ_G (Fig. 5AB; Methods) to compute the corresponding growth rates with FBA and also to train a PGS in each case.

Notably, the output of the PGS (coefficient of determination R^2) reaches a maximal optimal value for a given σ_G (Fig. 5C). This value results from the association of stronger effect sizes to the same genetic predictors found previously (Fig. 2D). Yet a better R^2 comes at the cost of a decreased mean population fitness. Besides, in populations with large genetic variability, $\sigma_G > 0.14$, the PGS' performance declines owing to an increased number of large effect predictors resulting from additional growth limiting reactions (Fig. 5D).

Next, we discover that although the main genetic predictors identified across all PGSs are identical, the effect sizes change quantitatively with genetic variability. Thus, the estimated impact of a mutation, e.g., reduction of 20% of a specific enzyme, will have a differential effect depending on the genetic background of the individual, which limits the suitability of applying a PGS on individuals if the test and training populations widely differ.

These results reveal overall a trade-off between genetic variability, population fitness, and predictability. While it is desirable to increase the performance of a PGS through sampling a population with high genetic variability, negative selection is likely to prevent scenarios of optimal predictability (O'Connor et al. 2019). Thus, selection would pose severe limitations to phenotype prediction.

Prediction is transportable across environments but also experiences extreme gene-environment interactions.

In our last analysis, we ask to what extent specific growing conditions influence the ability to predict growth, modifying, in turn, the collection of the predictors. Therefore, we randomly generated $>10^3$ (nutrient) environments of increasing richness (fixing the genetic variability as before, Methods). Then we train a separate PGS (for growth in every medium; Fig. 6A) to focus on the genes with the largest effect sizes ($|\beta| > 0.01$, as previously).

Figure 6B shows the effect sizes of each gene depending on medium richness. The top

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.19.492732; this version posted May 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

predictors identified previously appear recurrently in most media with increasing effect sizes in richer environments. Thus, they constitute a *core* set of predictors valid for most environments. Could this reflect that the growth medium has a limited impact on which reactions are active? Indeed, there exist between 70% and 100% of shared active fluxes across all random environments, and >99% if we consider reactions controlled by our initial set of top predictors. This explains the identification of similar genetic predictors across environments. Nevertheless, we observe a second general trend in poor media, where a novel set of predictors related to the mitochondrial respiratory chain –a different functional mode– becomes relevant.

One could still argue that the differences in predictability are due to subtle differences in metabolic solutions. We thus controlled for environmental richness to quantify this. Differences in metabolic solutions do not correlate with predictability (Pearson's $\rho = 0.06$, using partial correlations to control for environmental richness), further pointing towards environmental richness as a valid measure that recapitulates metabolic activity and similarity, anticipating predictability (Fig. 6C).

Next, we observe that other genes recurrently appear as strong predictors in specific –typically poor– media (Fig. 6B), and whose occurrence leads to particularly strong PGS performance with up to $R^2 = 0.56$ (Fig. 6C). Therefore, while growth prediction usually relies on a core set of genes largely "independent" of the growing medium, strong gene-environment interactions can sizeably improve the performance of a PGS.

Finally, the fact that the effect sizes change continuously with increasing environmental richness (Fig. 6B-D) ensures the "portability" of a singular PGS (trained in a reference medium) to predict the growth rate of the same population in another medium of similar richness (Fig. 6E). Figure 6F shows the performance of a PGS trained with data of the standard medium to predict the growth rates obtained in different random media (as test populations). Indeed, we observe that beyond a certain environmental similarity (Methods) the portability of $\mathrm{PGS_{std}}$ falls sharply together with the number of overlapping predictors

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.19.492732; this version posted May 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

between PGS_{std} and PGS_i (trained in the i-eth random environment). This enables us to distinguish regimes of high and low portability.

DISCUSSION

We developed a framework to generate genetic and phenotypic variability in a population of *in silico* yeast metabolisms to study how and why a polygenic score (PGS) can anticipate the phenotype (growth rate) from a specific enzymatic profile. Metabolic reconstructions are arguably the best-suited models to inspect these questions as they incorporate genetic and environmental information into an explicit GP map that renders the PGS accountable for its predictions (Kavvas et al. 2020). The quantitative interpretation of gene reaction rules (GRRs) is a fundamental layer contributing to an accurate representation of this map.

In this way, if we reveal how the associations that define the PGS arise, we can begin recognizing general limitations of predictability, which have important implications for fundamental (Burns 1970; Boyle et al. 2017), and applied biology (Dudbridge 2013; Torkamani et al. 2018; Kavvas et al. 2020). Specifically, we evaluated the limitations originated by i) biological and evolutionary constraints of the metabolism, ii) knowledge of the trait architecture, iii) non-linearities in the GP map, iv) influence of the environment, and v) genetic makeup of the training population.

Computing a case study, we obtain an $R^2 = 0.27$ and 32 genetic predictors with notably large effect sizes (Fig. 2DE). Which genes act as predictors result from the combination of two factors: the quantitative flux *required* in a certain environment and the flux constraints *derived* from the corresponding genetic variation (Fig. S9A). The former represents the "functional mode" of metabolism in that environment, while the latter results from the combination of GRRs and what we define as the "historical" reference flux bounds. These bounds represent the consequence of the adjustment of the yeast metabolism to onetime experienced evolutionary history.

Therefore, both GRRs and the reference bounds can act as a sieve of genetic variation causing part of it to be cryptic (Richardson et al. 2013; Paaby and Rockman 2014; Poy-

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.19.492732; this version posted May 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

atos 2020). GRRs could symbolize, for instance, the presence of an isozyme which would prevent such enzyme from becoming a phenotypic predictor. Moreover, reference bounds might largely differ from the required flux in a given environment, which again silences the functional impact of the genetic variability present in the population. We confirmed the latter effect in two experiments where we manipulated the reference fluxes (Fig. S9B).

The presence of the 32 distinct predictors, and their enrichment in a few metabolic functions, is thus contingent on both working regime and evolutionary history. Note, however, that these functions should be necessarily associated with biomass precursors, either directly or in upstream reactions, since the biomass reaction represents the architecture of the trait/phenotype of interest (Fig. 3AB).

Shifting our focus on systemic features, we can report several results. Pleiotropy, as an aggregate score of the impact of mutations on all biomass precursors, is a poor measure of the predictive character of a gene, with the disaggregate information nevertheless partially revealing the composition of the biomass reaction (Fig. 4B). Therefore, the PGS provides a sound but partial understanding of the architecture of the trait, exclusively the domain associated with the required precursors in a given environment.

We also discussed epistasis. Those genes whose dosages are eventually confining growth induce the non-linearities in the metabolic GP map, a combination of the working regime, i.e., functional mode, and evolutionary history as discussed above. On this basis, individuals are instances where only one, or few, dosages are particularly limiting, the exact ones varying among them (Fig. S8). But the mixture of individuals at the population level generates functional cross-dependencies, increasing the number of limiting enzymes and consequently reducing predictability (more inadequate predictions correlate with predictor number, linear $\rho = -0.94$, Fig. S9C).

This reasoning helps clarify two added phenomena. First, it describes the beneficial effect of sufficiently large gene-by-environment interactions on predictability (fewer predictors, better predictability, Fig. 6C). Second, it anticipates a conjecture where a metabolism "dis-

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.19.492732; this version posted May 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

313 abled" on its capacity to react to genetic variation would paradoxically be coupled to better
314 prediction. This premise, we proved (impairment causes simpler functional modes that lead
315 too to fewer predictors, Fig. S9D).

316 Moreover, the validity of linear methods to predict the outcome of highly non-linear
317 GP maps has intrigued the genetics community. Global sensitivity analysis confirms the
318 suitability of these approaches (effect sizes capture the sum of both additive and epistatic
319 contributions, Fig. 4D). Indeed, a PGS should capture non-linearities, since the minimization
320 of error due to a linear regression incorporates all data points, including those coupled to
321 non-linear regimes. Still and despite the presence of epistatic effects we notice that the sum
322 of additive terms accounts for over 75% of the total phenotype variability ($\sum S_0 > 0.75$,
323 Fig. 4C), certainly owing to order-preservation of gene-dosage responses (Fig. S8; Gjuvsland
324 et al. 2011). In sum, the linearity of the PGS *a priori* imposes a fundamental constraint on
325 anticipating non-linear effects but it can nevertheless partially integrate them. This trade-off
326 illustrates to what extent the GP map can be described linearly (Hill et al. 2008).

327 Finally, we asked about the portability of the predictions across populations concerning
328 differences in genetic and environmental situations. Populations experiencing an interme-
329 diate genetic variability ensure maximum predictability, in line with previous results on
330 extreme allele frequencies (Hill et al. 2008). However, such an increase in R^2 comes at the
331 cost of population growth (Fig. 5B). Achieving this maximum could then be unattainable
332 due to negative selection (O'Connor et al. 2019), which could be interpreted as yet another
333 constraint on predictive power. In GP contexts in which genetic variability is less likely to
334 cause loss of function and more forms of gain of function are possible, this constraint will be
335 less apparent, as variability will not elicit negative selection.

336 Notably, the list of genetic predictors remains largely the same, independent of the
337 amount of genetic variability in the training population (Fig. 5D). This is also observed
338 when one determines predictors associated with populations growing in different media of
339 similar richness, with two consequences (Fig. 6D). First, these results ensure the portability

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.19.492732; this version posted May 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

of the PGSs. Second, and as stated before, the predictors are specific to the evolutionary history of the metabolism. As a consequence, portability might not necessarily hold for more fixed, and maybe more realistic, histories.

Perhaps the most unifying research program of all questions currently being addressed in biology is that of phenotype prediction. Here, we have learned that the combination of functional mode, evolutionary history, and phenotypic architecture determines the limits of prediction. For this, we benefited from an *in silico* approach whose capabilities to examine the emergence of phenotypic variation are beyond current experimental setups. The debate on the need to understand our predictions is sure to spark many interesting future discussions.

MATERIALS AND METHODS

Metabolic models

Whole-genome metabolic models integrate the stoichiometry of the reactions in the metabolism of a model species, and together with computational methods they enable the estimation of an optimal network solution given an objective function where fluxes are stable. Among all fluxes, we focus on the prediction of biomass production, an analogue of growth rate and fitness. We used the genome-scale metabolic reconstruction of *Saccharomyces cerevisiae* iND750 (Duarte 2004) together with the Cobra toolbox for Python (Ebrahim et al. 2013) to compute the fitness of numerous mutants in either a standard medium or random media and the Escher package to depict the central carbon metabolism (King et al. 2015). Metabolic subsystems are typically assigned to reactions, hence we imputed a specific subsystem to a gene only if all reactions in which it participates belong to the same subsystem. Among all 750 genes present in the model, 42 of them had either none or multiple subsystems associated, which we discarded. Note that our choice of model tries to balance the presence of sufficient biological details with accessible computational time. Still, our results and the mechanisms underlying phenotype prediction are robust when using the most recent yeast metabolic reconstructions iMM904 and yeast8 ($R^2 = 0.18$ and $R^2 = 0.17$, respectively).

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.19.492732; this version posted May 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

366 **Quantitative mutations**

367 We compute the effect of a quantitative reduction in gene dosage, or equivalently enzyme
368 efficiency, in two steps (Fig. S1). First, we compute the wild type "reference" bounds of
369 each reaction. These bounds are constituted by the maximum (and minimum if reversible)
370 reaction fluxes observed in $2x10^4$ optimal solutions of metabolisms exposed to random en-
371 vironments, and random genetic backgrounds (Methods). In the latter case, we randomly
372 sampled flux bounds from a uniform distribution in the range [0,100]. In this way, the wild
373 type bounds integrate the history of yeast metabolisms, which have adapted to different
374 environmental and genetic contexts (Discussion).

375 Second, we interpret quantitatively the gene reaction rules (GRRs) to find how reducing
376 the dosage of an enzyme translates into a reduced flux through its reactions with respect
377 to the wild type. This is necessary because some reactions may require several subunits
378 or just one of several isozymes. The GRRs may contain `AND` and `OR` operators acting on
379 pairs of genes, we consider these equivalent to `min` and `sum`, respectively, acting on relative
380 gene dosages. This approach is similar to those used in noise propagation (Wang and Zhang
381 2011), or by the Escher package (King et al. 2015). In all cases, the upper/lower bounds are
382 always computed and set according to the reactions' reversibility and the bounds of ATP
383 maintenance, biomass production and exchange reactions are kept unaltered.

384 This procedure is comparable to a previous approach in which genetic variability was
385 also mapped to flux constraints (Kavvas et al. 2020). The authors construct the allele-to-
386 flux constraint map coupled to the performance of a novel objective function to classify
387 antibiotic resistance in a fixed medium. Our approach, however, assumes flux constraints
388 imposed by a history of past genetic and environmental adaptations and, in this sense it is
389 more comprehensive, together with explicit information about which genes are involved in
390 exact reactions (GRRs).

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.19.492732; this version posted May 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Genetic variability

We generate genetic variability by sampling gene dosages from a probability distribution. Unless otherwise stated, we use a normal distribution with unit mean and standard deviation $\sigma = 0.1$. In this way, σ directly reflects the variability in the population's gene dosages. This distribution follows from either Fisher's original infinitesimal model, or from the Gaussian Descendants derivation where different levels of parenthood result in different σ under neutral evolution (Barton et al. 2017; Turelli 2017). We define the wild type genotype as having all gene dosages equal to the unit. This procedure generates populations that are in linkage equilibrium.

We also engineered genetic variability based on gamma distributions with shape parameters $0.5 < k < 200$ and scale adjusted such that all distributions had equal mean. Importantly, note that gene dosages >1 are not beneficial, as wild type bounds are the extreme values observed (statistically), thus to avoid including additional cryptic genetic variation, gene dosages >1 were clipped to the unit. In this way, the resulting genetic variability in our standard population is $\sigma_G = 0.05$ (Fig. 5).

Growth media and environmental variability

The minimal medium is defined by unbounded import and export of H_2O, CO_2, ammonia, phosphate, sulphate, sodium and potassium and it is aerobic with 2 mmol/gDW/h import rate of O_2. The standard medium is additionally composed of 20mmol/gDW/h import rate of glucose. Random environments are generated following a previous protocol (Wang and Zhang 2009). Briefly, we supplement the minimal medium with an additional number of components such that the probability of including any component follows an exponential distribution with mean $m = 0.10$ (other values produce similar results). Then, for every component, we obtain their maximum import rates from a uniform distribution between 0 and 20 mmol/gDW/h.

We define the richness of a medium as the growth rate of the wild type, and the environmental similarity between two media as the ratio of their richness. To avoid including

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.19.492732; this version posted May 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

arbitrarily rich media, we consider those with richness inferior, or equal to that of the standard medium. Also, we discard media that support biomass production rates <70% of that of the standard medium to avoid possible natural or model artifacts related to our implementation of quantitative mutations (see Results). This is an alternative approach to Constrained Allocation FBA (CAFBA), which limits the growth rate of metabolic models based on resource allocation principles by fixing a medium and tuning a parameter related to proteome fractioning (Mori et al. 2016).

Polygenic Score

We used a high-dimensional regression framework for polygenic modeling and prediction:

$$\vec{y}_{N\times 1} = \mathbf{G}_{N\times M}\vec{\beta}_{M\times 1} + \vec{\epsilon}_{N\times 1}, \tag{1}$$

where N is the sample size, M is the number of genes, $\vec{y}$ is the vector of phenotypes (typically growth rate), $\mathbf{G}$ is the genotype matrix, $\vec{\beta}$ is the vector of effect sizes of the genes and $\vec{\epsilon}$ is some noise assumed normal with unknown variance. The generated data was fit using Least Absolute Shrinkage and Selection Operator (LASSO) a type of regression that under bayesian statistics assumes prior Laplace distributions in each coefficient, instead of uniform distributions as in the case of Ordinary Least Squares. Consequently, with LASSO some parameters are automatically zero (Tibshirani 1996), hence making it a remarkable alternative to pruning and thresholding (P+T) or other regularization methods (Dudbridge 2013; Wray et al. 2013). In addition, we compute the best value of the shrinkage parameter with five-fold cross validation. That effect sizes show a bimodal distribution makes our results robust to the application of other regularization, or feature selection methods (Fig. 2E).

Sensitivity analysis and total epistasis

We computed local sensitivity indices Z_i to monitor the changes in the output variable, i.e., growth rate, when every single input variable, i.e., each gene dosage, is altered (Kacser and Burns 1981; Keightley 1989). However, local sensitivity analysis (results are available in

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.19.492732; this version posted May 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

the Supplementary Material) does not explore the entire parameter space and is unable to isolate the effect of (non-linear) variable interactions.

Global sensitivity analysis, however, is ideal for this task, as it decomposes the variability of the output of a model into different terms when all variables fluctuate simultaneously. We used the method first proposed by Sobol for its easy implementation and interpretation (Sobol 1993; Saltelli et al. 2008). Note that this differs from previous flux-based applications (Nguyen Quang et al. 2019; Nobile et al. 2021). Briefly, we focus on two indices for the i-eth gene, the first order index S_0^i and the total effect index S_T^i. The former quantifies the additive part of the variability associated to a gene while the latter quantifies its total contribution, additive and all non-additive effects. From these, we derive the total epistasis which accounts for all, and only, non-additive effects as $\epsilon_T^i = S_T^i - S_0^i$ and its error $(\Delta\epsilon_T^i)^2 = (\Delta S_0^i)^2 + (\Delta S_T^i)^2$.

We computed all indices and their errors with monte carlo estimators using over 10^6 samples (Saltelli et al. 2008; Saltelli et al. 2010). We carried out these computations with pairs of genotypes sampled from the original population growing in standard medium. A detailed description of the protocol and equations are available in the Supplementary Material.

Note that we do not show negative values of both S_0 and ϵ_T as they are unrealistic and should be considered null in agreement with their error bounds.

Pleiotropy

In metabolic models, the pleiotropy of a mutation is generally computed as the number of biomass precursors whose maximum production is limited by the mutation, following a previous protocol (He and Zhang 2006; Shlomi et al. 2007). Briefly, we simulated the excretion of a given metabolite by adding an exchange reaction to the model and maximizing the flux through this reaction. Then we consider that a gene limits the production of a metabolite if, when knocked-down by 90%, its excretion rate decreases. As pleiotropy is strongly dependent on the genetic context, we computed the mean value across 10^3 individuals of the population due to the large computational load. We used a 90% decrease in dosage to avoid

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.19.492732; this version posted May 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

artifacts derived from gene essentiality, but our results are robust when using other values.

DATA AVAILABILITY

Data and code for this work is available at Zenodo (Yubero 2022). The main code used to generate quantitative mutations is available at GitHub (https://github.com/pyubero/quantitative_mutations).

ACKNOWLEDGEMENTS

This work was supported by grant PID2019-106116RB-I00 (P.Y. and J.F.P.), partially by Ph.D. fellowship BES-2016-079127 (P.Y.), and the program Severo Ochoa Center of Excellence (A.A.L.) from the Spanish Ministerio de Ciencia e Innovación and the European Social Fund.

AUTHOR CONTRIBUTIONS

P.Y., A.A.L., and J.F.P. conceived and designed the study. P.Y. conducted all analysis with contributions from A.A.L. and J.F.P. P.Y and J.F.P. discussed the results and wrote the manuscript. All authors reviewed and approved the final manuscript.

COMPETING INTERESTS

The authors declare no competing interests

MATERIALS CORRESPONDENCE

Correspondence and request for materials should be addressed to J.F.P.

REFERENCES

Barton, N., Etheridge, A., and Véber, A. (2017). "The infinitesimal model: Definition, derivation, and implications." *Theoretical Population Biology*, 118, 50–73.

Boyle, E. A., Li, Y. I., and Pritchard, J. K. (2017). "An Expanded View of Complex Traits: From Polygenic to Omnigenic." *Cell*, 169(7), 1177–1186.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.19.492732; this version posted May 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

493 Burns, J. (1970). "The synthetic problem and the genotype-phenotype relation in
494 metabolism." *Organization, Stability and Process. Towards a Theoretical Biology, C.H.*
495 *Waddington editor*, 3, 47–51.

496 Cannon, M. E. and Mohlke, K. L. (2018). "Deciphering the Emerging Complexities of Molec-
497 ular Mechanisms at GWAS Loci." *The American Journal of Human Genetics*, 103(5),
498 637–653.

499 Clark, A. G. (1991). "Mutation-selection balance and metabolic control theory.." *Genetics*,
500 129(3), 909–923.

501 Duarte, N. C. (2004). "Reconstruction and validation of saccharomyces cerevisiae iND750, a
502 fully compartmentalized genome-scale metabolic model." *Genome Research*, 14(7), 1298–
503 1309.

504 Dudbridge, F. (2013). "Power and Predictive Accuracy of Polygenic Risk Scores." *PLoS*
505 *Genetics*, 9(3), e1003348.

506 Ebrahim, A., Lerman, J. A., Palsson, B. O., and Hyduke, D. R. (2013). "Cobrapy:
507 Constraints-based reconstruction and analysis for python." *BMC Systems Biology*, 7(1),
508 74.

509 Feldman, M. W. and Lewontin, R. C. (1975). "The Heritability Hang-Up: The role of variance
510 analysis in human genetics is discussed.." *Science*, 190(4220), 1163–1168.

511 Gjuvsland, A. B., Vik, J. O., Woolliams, J. A., and Omholt, S. W. (2011). "Order-preserving
512 principles underlying genotype–phenotype maps ensure high additive proportions of ge-
513 netic variance." *Journal of Evolutionary Biology*, 24(10), 2269–2279.

514 He, X. and Zhang, J. (2006). "Toward a Molecular Understanding of Pleiotropy." *Genetics*,
515 173(4), 1885–1891.

516 Hill, W. G., Goddard, M. E., and Visscher, P. M. (2008). "Data and theory point to mainly
517 additive genetic variance for complex traits." *PLOS Genetics*, 4(2), 1–10.

518 Kacser, H. and Burns, J. A. (1981). "The molecular basis of dominance." *Genetics*, 97(3-4),
519 639–666.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.19.492732; this version posted May 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Kavvas, E. S., Yang, L., Monk, J. M., Heckmann, D., and Palsson, B. O. (2020). "A biochemically-interpretable machine learning classifier for microbial gwas." *Nature Communications*, 11(1), 2580.

Keightley, P. D. (1989). "Models of quantitative variation of flux in metabolic pathways." *Genetics*, 121(4), 869–876.

Keightley, P. D. and Kacser, H. (1987). "Dominance, pleiotropy and metabolic structure." *Genetics*, 117(2), 319–329.

King, Z. A., Dräger, A., Ebrahim, A., Sonnenschein, N., Lewis, N. E., and Palsson, B. O. (2015). "Escher: A web application for building, sharing, and embedding data-rich visualizations of biological pathways." *PLOS Computational Biology*, 11(8), e1004321.

Lynch, M. and Walsh, B. (1998). *Genetics and analysis of quantitative traits*. Sinauer, Sunderland, Mass.

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., McCarthy, M. I., Ramos, E. M., Cardon, L. R., Chakravarti, A., Cho, J. H., Guttmacher, A. E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C. N., Slatkin, M., Valle, D., Whittemore, A. S., Boehnke, M., Clark, A. G., Eichler, E. E., Gibson, G., Haines, J. L., Mackay, T. F. C., McCarroll, S. A., and Visscher, P. M. (2009). "Finding the missing heritability of complex diseases." *Nature*, 461(7265), 747–753.

Milocco, L. and Salazar-Ciudad, I. (2020). "Is evolution predictable? Quantitative genetics under complex genotype-phenotype maps." *Evolution*, 74(2), 230–244.

Mori, M., Hwa, T., Martin, O. C., De Martino, A., and Marinari, E. (2016). "Constrained allocation flux balance analysis." *PLOS Computational Biology*, 12(6), 1–24.

Nguyen Quang, M., Rogers, T., Hofman, J., and Lanham, A. B. (2019). "Global sensitivity analysis of metabolic models for phosphorus accumulating organisms in enhanced biological phosphorus removal." *Front. Bioeng. Biotechnol.*, 7, 234.

Nijhout, H. F., Sadre-Marandi, F., Best, J., and Reed, M. C. (2017). "Systems Biology of Phenotypic Robustness and Plasticity." *Integrative and Comparative Biology*, 57(2), 171–

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.19.492732; this version posted May 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

184.

Nobile, M. S., Coelho, V., Pescini, D., and Damiani, C. (2021). "Accelerated global sensitivity analysis of genome-wide constraint-based metabolic models." *BMC bioinformatics*, 22(Suppl 2), 78–78.

O'Connor, L. J., Schoech, A. P., Hormozdiari, F., Gazal, S., Patterson, N., and Price, A. L. (2019). "Extreme Polygenicity of Complex Traits Is Explained by Negative Selection." *The American Journal of Human Genetics*, 105(3), 456–476.

Omholt, S. W., Plahte, E., Øyehaug, L., and Xiang, K. (2000). "Gene Regulatory Networks Generating the Phenomena of Additivity, Dominance and Epistasis." *Genetics*, 155(2), 969–980.

Orrell, D. (2007). *The Future of Everything. The Science of Prediction*. Basic Books.

Paaby, A. B. and Rockman, M. V. (2014). "Cryptic genetic variation: evolution's hidden substrate." *Nature Reviews. Genetics*, 15(4), 247–258.

Palsson, B. (2006). *Systems biology: properties of reconstructed networks*. Cambridge University Press, Cambridge ; New York OCLC: ocm62421240.

Poyatos, J. F. (2020). "Genetic buffering and potentiation in metabolism." *PLOS Computational Biology*, 16(9), 1–15.

Richardson, J. B., Uppendahl, L. D., Traficante, M. K., Levy, S. F., and Siegal, M. L. (2013). "Histone Variant HTZ1 Shows Extensive Epistasis with, but Does Not Increase Robustness to, New Mutations." *PLoS Genetics*, 9(8), e1003733.

Rockman, M. V. (2008). "Reverse engineering the genotype–phenotype map with natural genetic variation." *Nature*, 456(7223), 738–744.

Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., and Tarantola, S. (2010). "Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index." *Computer Physics Communications*, 181(2), 259–270.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S. (2008). *Global Sensitivity Analysis: The Primer*. Wiley.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.19.492732; this version posted May 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

574 Shlomi, T., Herrgard, M., Portnoy, V., Naim, E., Palsson, B. Ø., Sharan, R., and Rup-
575 pin, E. (2007). "Systematic condition-dependent annotation of metabolic genes." *Genome*
576 *Research*, 17(11), 1626–1633.

577 Sobol, I. M. (1993). "Sensitivity analysis for non-linear mathematical models." *Mathematical*
578 *modelling and computational experiment*, 1, 407–414.

579 Tibshirani, R. (1996). "Regression shrinkage and selection via the lasso." *Journal of the Royal*
580 *Statistical Society: Series B (Methodological)*, 58(1), 267–288.

581 Torkamani, A., Wineinger, N. E., and Topol, E. J. (2018). "The personal and clinical utility
582 of polygenic risk scores." *Nature Reviews Genetics*, 19(9), 581–590.

583 Turelli, M. (2017). "Commentary: Fisher's infinitesimal model: A story for the ages." *Theo-*
584 *retical Population Biology*, 118, 46–49.

585 Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., and
586 Yang, J. (2017). "10 Years of GWAS Discovery: Biology, Function, and Translation." *The*
587 *American Journal of Human Genetics*, 101(1), 5–22.

588 Waddington, C. H. (2015). *Strategy of the Genes*. Taylor & Francis, Place of publication not
589 identified OCLC: 960838218.

590 Wang, Y., Gjuvsland, A. B., Vik, J. O., Smith, N. P., Hunter, P. J., and Omholt, S. W.
591 (2012). "Parameters in Dynamic Models of Complex Traits are Containers of Missing
592 Heritability." *PLoS Computational Biology*, 8(4), e1002459.

593 Wang, Z. and Zhang, J. (2009). "Why Is the Correlation between Gene Importance and Gene
594 Evolutionary Rate So Weak?." *PLoS Genetics*, 5(1), e1000329.

595 Wang, Z. and Zhang, J. (2011). "Impact of gene expression noise on organismal fitness and
596 the efficacy of natural selection." *Proceedings of the National Academy of Sciences*, 108(16).

597 Welch, S. M., Dong, Z., Roe, J. L., and Das, S. (2005). "Flowering time control: gene net-
598 work modelling and the link to quantitative genetics." *Australian Journal of Agricultural*
599 *Research*, 56(9), 919.

600 Wray, N. R., Wijmenga, C., Sullivan, P. F., Yang, J., and Visscher, P. M. (2018). "Common

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.19.492732; this version posted May 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

601 Disease Is More Complex Than Implied by the Core Gene Omnigenic Model." *Cell*, 173(7),
602 1573–1580.
603 Wray, N. R., Yang, J., Hayes, B. J., Price, A. L., Goddard, M. E., and Visscher, P. M.
604 (2013). "Pitfalls of predicting complex traits from snps." *Nature Reviews Genetics*, 14(7),
605 507–515.
606 Yubero, P. (2022). "Data and code of The limitations of phenotype prediction in metabolism.
607 Zaidi, A. A. and Mathieson, I. (2020). "Demographic history mediates the effect of stratifi-
608 cation on polygenic scores." *eLife*, 9, e61548.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.19.492732; this version posted May 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

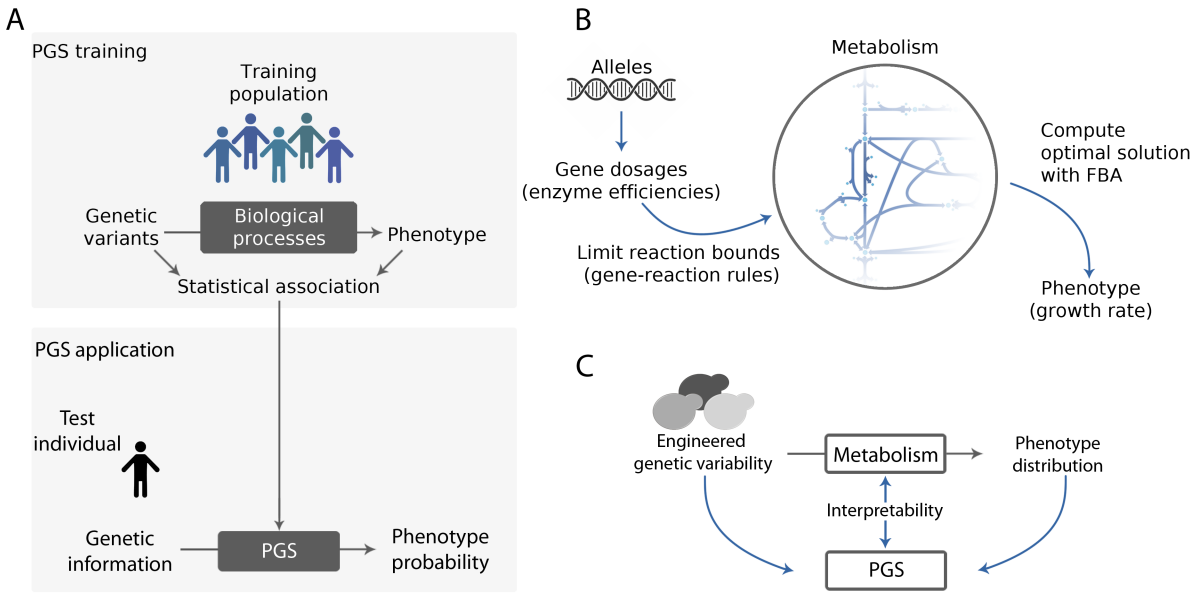

Fig. 1. Metabolic reconstructions provide an explicit GP map that allows us to open the black box of phenotype prediction. (**A**) Phenotype prediction is typically based on the statistical association of genetic and phenotype variants in a training population. This approach defines a black box that bypasses all underlying biological processes. With the genetic information of a test individual, the model calculates a *polygenic score* (PGS) with the probability of observing the specific trait. (**B**) We benefit from the metabolic reconstruction of *S. cerevisiae* to generate an *in silico* population of yeast metabolisms. Genetic variability is modeled by the effect of alleles on gene dosages, which limits the maximum flux through their reactions according to the gene-reaction rules. Given these constraints, flux balance analysis (FBA) computes the growth rate of each individual in the population. (**C**) We quantify the statistical associations to elaborate a PGS for growth rate. Together with the availability of the underlying model, this enables us to "open" the black box of phenotype prediction to investigate its limitations.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.19.492732; this version posted May 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

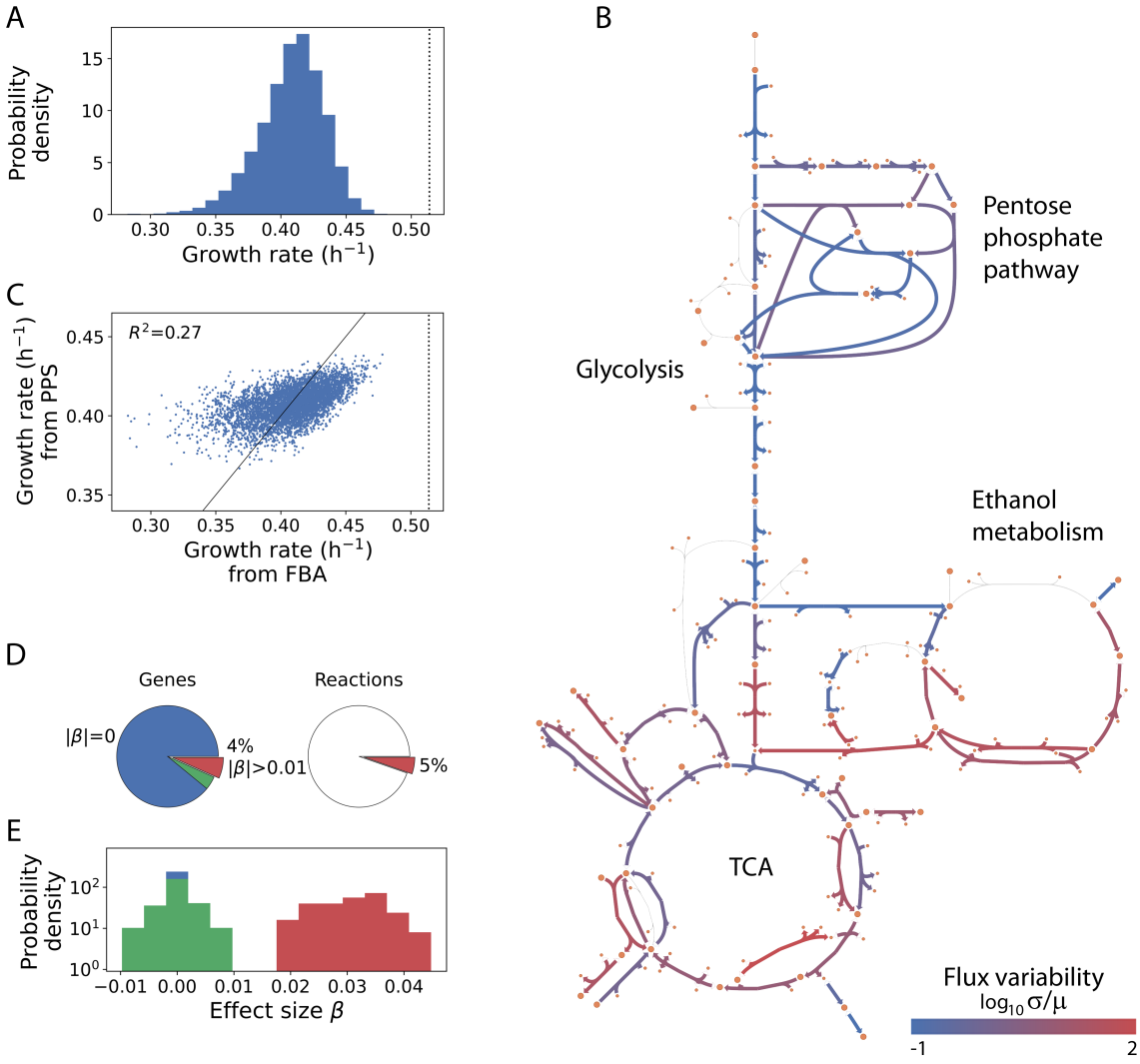

Fig. 2. The PGS reveals that a small number of genes explains 27% of growth variability in yeast metabolism. (**A**) We generate a synthetic population of $5 \cdot 10^3$ yeast metabolisms with relative gene dosages sampled from a normal distribution (Methods). For each individual, we compute its growth rate with flux balance analysis. Such genetic variability induces a distribution of growth rates with mean and deviation $\mu = 0.41 \pm 0.03$. The vertical dotted line shows the wild-type growth rate ($\mu_{wt} = 0.51$). (**B**) Central carbon metabolism cartoon emphasizing the variability in flux solutions across the population. (**C**) We trained a polygenic score (PGS) to anticipate growth rates from gene dosage data. The PGS explains 27% of the growth rate variability observed in the training population. (**D,E**) The PGS computes a specific effect size for each gene, β. Most of the genes, 88.7%, have null effect sizes (blue), while only 4.3% of the genes (red) are strongly associated with growth rate with $|\beta| > 0.01$. The latter control just 5% of the metabolic reactions.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.19.492732; this version posted May 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

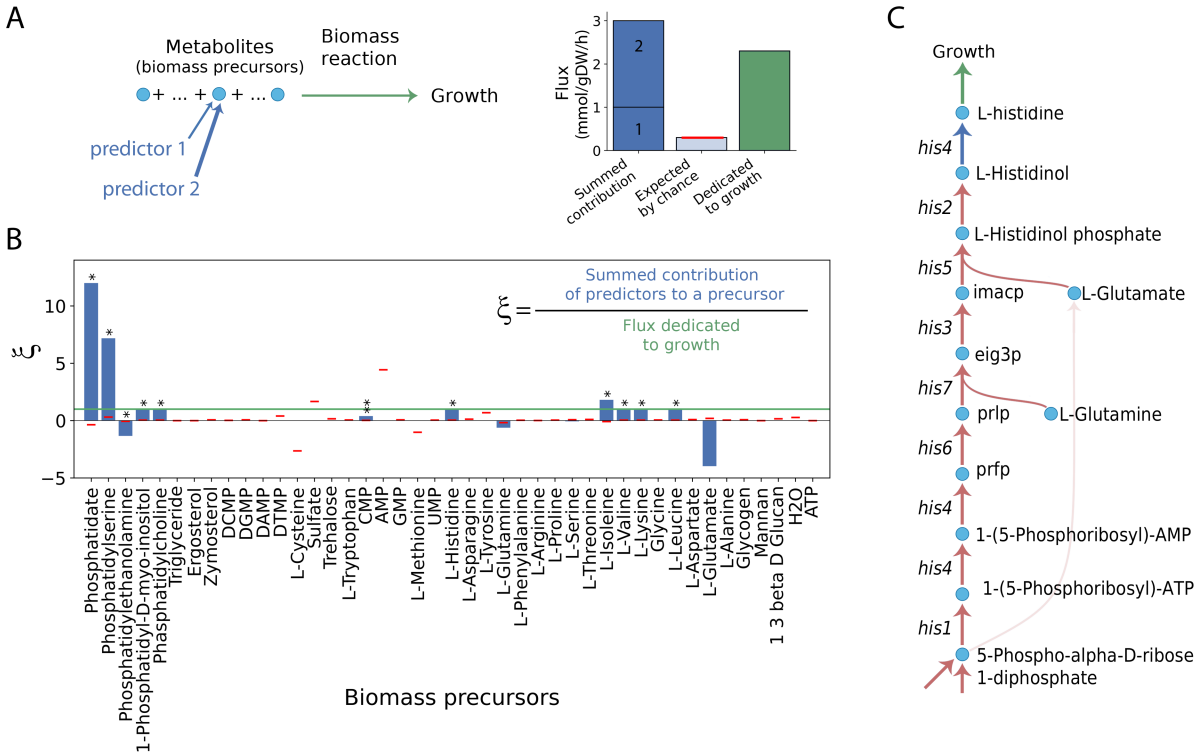

Fig. 3. Top gene predictors control the production of a few metabolites required for growth. (**A**) Biomass precursors in metabolic reconstructions are metabolites that ultimately fuel the biomass reaction, which simulates growth. We compute the contribution of several genes (e.g., predictors 1 and 2) to the production of a specific precursor and compare it with the expected contribution of randomly selected genes (red horizontal line), or with the flux devoted to growth (green). (**B**) Mean values of ξ, the aggregate contribution of the predictor genes to the production of each precursor relative to biomass consumption, across the population of yeast metabolisms. For example, we observe that the predictor genes produce 100% of the L-histidine consumed by growth (horizontal green line). We tested significant contributions after $5x10^3$ gene randomizations controlling for subset size (mean, red horizontal lines; * $p<0.05$, ** $p< 0.01$). The case of L-glutamate is not significant due to a large variance (not shown for clarity). (**C**) Part of the metabolic pathway leading to L-histidine production. Although it is produced directly by *his4*, its production is influenced by upstream histidine-related genes, which also appear as important growth predictors.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.19.492732; this version posted May 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

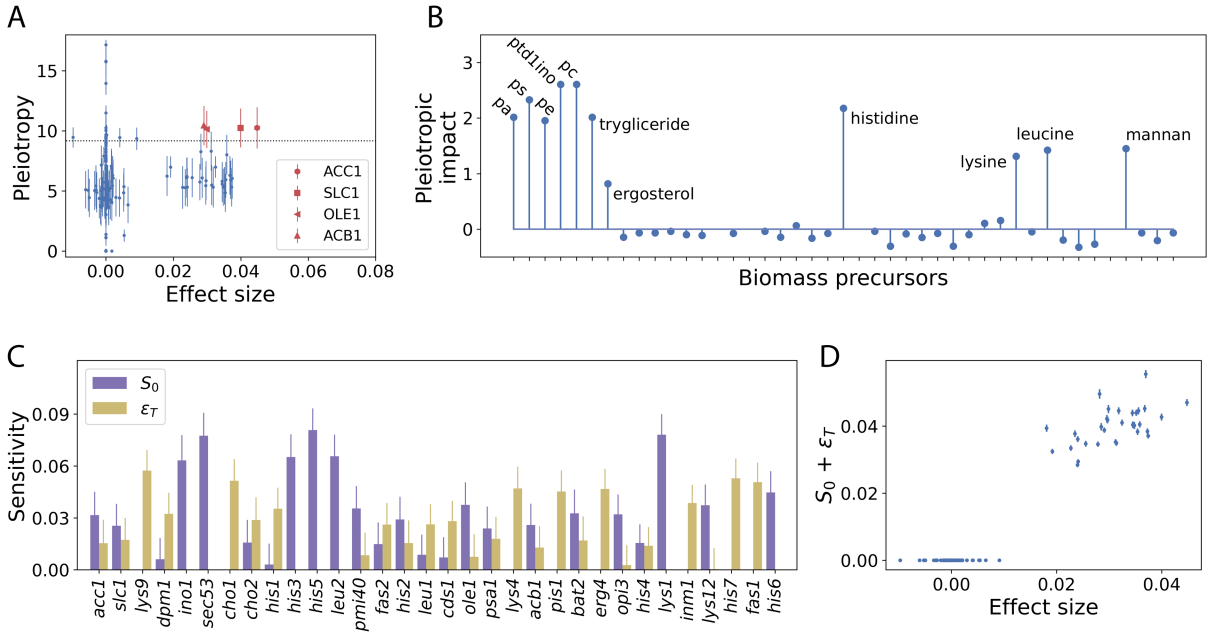

Fig. 4. System-wide effects of top predictors. (**A**) Pleiotropy of each gene after a 90% dosage reduction (Methods). Only a few top predictors have significant values of pleiotropy (over the 95th percentile, horizontal dotted line). Mean values (dots) and one standard deviation (error bars) across 10^3 individuals. (**B**) The pleiotropic impact of genes with large effect sizes focuses on a subset of biomass precursors. We show the Z-score of the sum of predictors (genes) that are pleiotropic w.r.t each metabolite against the mean and deviation found across all genes. We use abbreviations for phosphatidate (pa), phosphatidylcholine (pc), phosphatidylserine (ps), and phosphatidyl 1D myo inositol (ptd1ino). (**C**) Global sensitivity analysis allows us to quantify both the additive impact of genes on the growth rate, S_0 (purple), and the total epistatic effects, ϵ_T (yellow), which include all 2nd and higher-order gene-gene interactions. Bars and vertical lines represent mean values and a standard deviation, respectively, of $> 10^6$ simulations (Methods). (**D**) The sum of the additive and epistatic effects correlates well with the effect sizes of the polygenic score (Pearson's $\rho > 0.97$).

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.19.492732; this version posted May 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

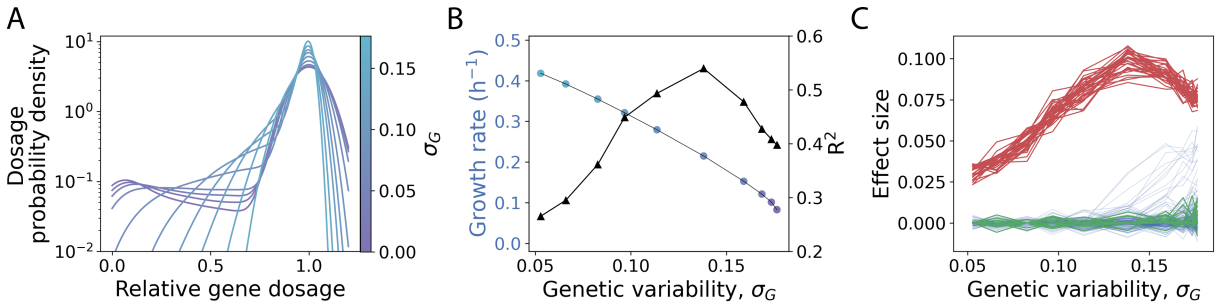

Fig. 5. Populations with different genetic variability reveal a common PGS architecture but also its predictive limit. (**A**) Probability density function estimate (kernel, bw=0.3) of relative gene dosages of 10 populations with increasing genetic variability (σ_G). (**B**) With FBA, we compute the growth rates (mean values on the left y-axis, colored circles as in panel A) and trained a separate PGS for every population (coefficient of determination R^2 on the right y-axis, black triangles). (**C**) Effect sizes of all genes as a function of σ_G. Lines are colored as in Fig. 2E.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.19.492732; this version posted May 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

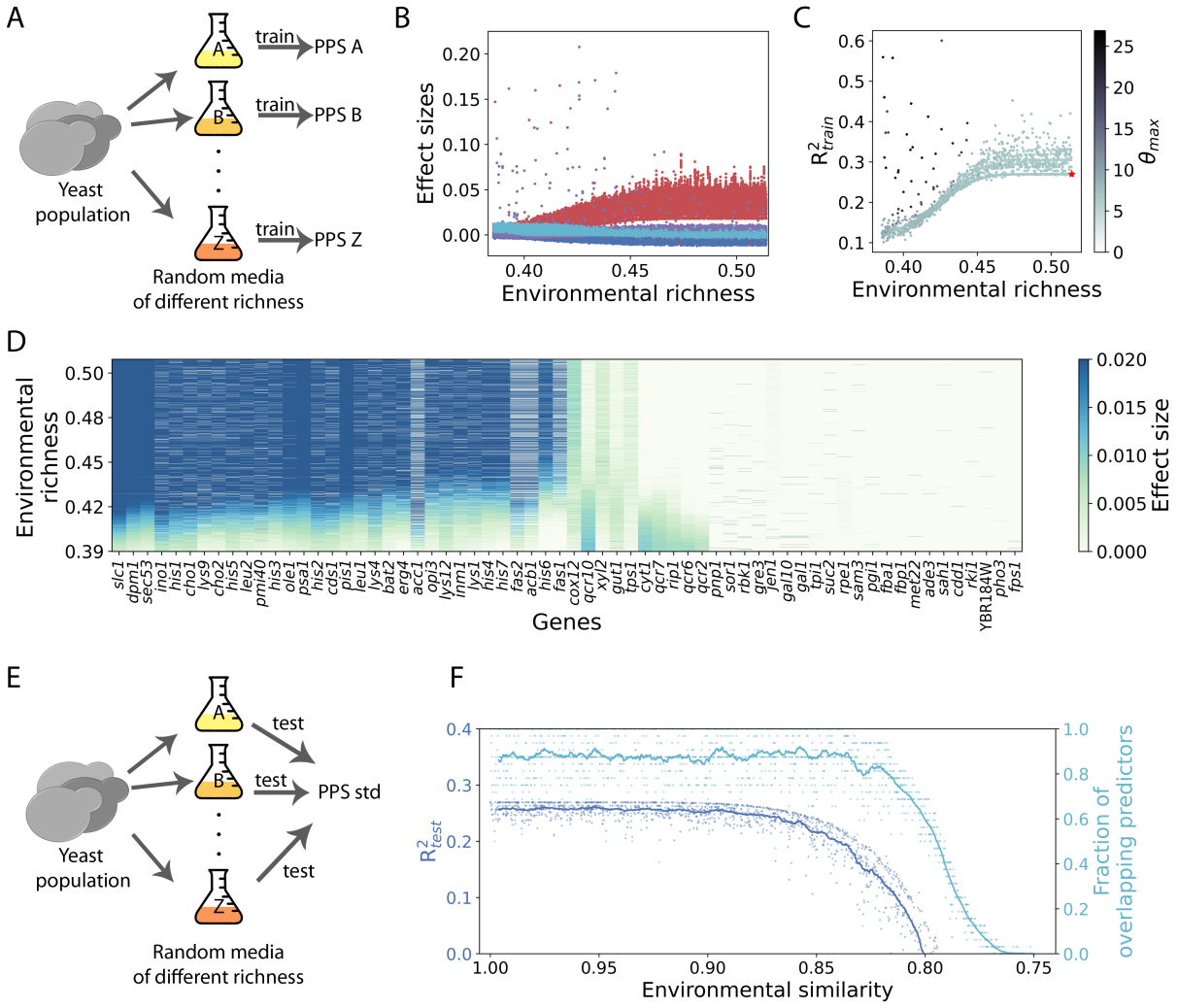

Fig. 6. Environmental effects on phenotype predictability and the portability of a PGS. (**A**) We compute the growth rates of a fixed population in 10^3 random environments with different richness (Methods) to train a PGS in each of these. (**B**) Effect sizes of all PGSs as a function of environmental richness. We highlight previously identified top predictors (red, as in Fig. 2), a novel set of predictors recurrent in poorer media that are related to the mitochondrial respiratory chain (cyan), and genes that show large effects in only specific media (purple). (**C**) The predictability of a PGS typically increases with environmental richness. However, in some media, predictability improves up to $R^2 = 0.6$ due to strong gene-environment interactions identified by outliers in the effect sizes (θ_{max} is the Z-score of the maximum effect size found in each PGS). (**D**) Effect sizes of genetic predictors follow a clear trend as a function of environmental richness. We show explicitly the values for all genes that have an effect size β>0.01 in any PGS. (**E**) Next, we test the "portability" of the PGS computed in the standard medium, $\mathrm{PGS_{std}}$, that is, its ability to predict growth rates in different environments. (**F**) The portability of $\mathrm{PGS_{std}}$ (left y-axis) holds within a certain environmental similarity, measured as the ratio of the random and the standard medium richness. The fall in "portability" is linked to the decreased overlap of predictors between $\mathrm{PGS_{std}}$ and the corresponding PGS of the medium (right y-axis). The dots and lines correspond to individual media and a running average ($n = 50$), respectively.