Genetic Architecture of Growth Traits Revealed by Global Epistatic Interactions

Lin Xu[1], Huifeng Jiang[2], Hong Chen[2], and Zhenglong Gu*[,2]

[1]Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York

[2]Division of Nutritional Sciences, Cornell University, Ithaca, New York

*Corresponding author: E-mail: zg27@cornell.edu.

Accepted: 30 June 2011

Abstract

Epistasis has long been recognized as fundamentally important in understanding the structure, function, and evolutionary dynamics of biological systems. Yet, little is known about how it is distributed underlying specific traits. Based on a global map of epistatic interactions in baker's yeast, *Saccharomyces cerevisiae*, we show that epistasis is prevalent (~13% increase from random expectation) and displays modular architecture among genes that underlie the same growth traits. More interestingly, our results indicate that hub genes responsible for the same growth traits tend to link epistatically with each other more frequently than random expectation. Our results provide a genome-wide perspective on the genetic architecture of growth traits in a eukaryotic organism.

Key words: epistasis, hub genes, growth trait.

Complex traits that vary in populations of human and other organisms are determined by multiple genetic factors. An individual genetic factor might only contribute a modest amount to the total variation observed in a trait over the entire population (Carlborg and Haley 2004; Visscher et al. 2008; Manolio et al. 2009). Genetic factors contributing to the same traits usually affect each other's phenotypic outcome, a phenomenon called epistasis (Legare et al. 2000; Manolio et al. 2009). How epistatic interactions among genetic factors are distributed underlying the same complex trait remains largely unknown (Phillips 2008). Here, we use growth traits in yeast as models to study this issue.

It is also well established that epistasis is important for the evolution of sex (Kondrashov 1982; Azevedo et al. 2006; Otto 2007), speciation (Presgraves 2007), mutational load (Hansen and Wagner 2001), ploidy (Kondrashov and Crow 1991; Musso et al. 2008), genetic drift (Perez-Figueroa et al. 2009), genomic complexity (Sanjuan and Nebot 2008), drug resistance (Trindade et al. 2009), and human disease (Phillips 2008). In model organisms, illustrating epistatic interactions also enables dissection of functional relationship between genes (Avery and Wasserman 1992; Hartman et al. 2001; Kelley and Ideker 2005; Ma et al. 2008; Brady et al. 2009). Understanding the distribution of epistasis underlying complex traits is therefore important for various fields.

Individual studies pointed out a prominent role for epistasis in genetic control of complex traits (Remold and Lenski 2004; Carlborg et al. 2006; Ehrenreich et al. 2007; Shao et al. 2008). However, a comprehensive understanding of epistasis underlying complex traits can only be achieved by reconstructing a global map of epistasis. Yeast provides a great model system to address this issue due to its abundant functional genomic data. Here, we examined the distribution and prevalence of epistasis underlying growth traits of yeast in different conditions. We firstly identified genes which contribute to growth under each of 354 conditions (Hillenmeyer et al. 2008). We then extracted subnetworks of epistasis among the contributing genes in each of the 354 conditions from the genome-wide epistatic network (Costanzo et al. 2010). Novel characteristics for the genetic architecture of growth traits are described. Although the epistasis used in this study was generated from yeast gene deletion mutants and the complex traits used were measured from yeast growth in specific laboratory conditions, both of which might be different from the real scenario in nature, our results provide the first glimpse on the genome-wide organization of epistasis underlying complex traits. The implication of our results on gene pleiotropy is also discussed.

Materials and Methods

Data Resource

This study is mainly based on the integration of two high-throughput experimental data sets: a genome-wide screen for the fitness effects of gene deletion mutants under 354 conditions (Hillenmeyer et al. 2008) and a global survey for the epistatic interactions among more than 5 million gene pairs in *Saccharomyces cerevisiae* (Costanzo et al. 2010).

In Hillenmeyer et al. (2008), ~6,000 heterozygous gene deletion mutants were screened in a total of 354 unique conditions (e.g., drugs approved by the U.S. Food and Drug Administration, well-characterized chemical probes, and compounds with uncertain biological activity). Genes whose heterozygous deletions significantly affect organism growth in a specific condition were defined as genes that contribute to organism growth in that condition. The authors defined significant growth defect with correction for multiple comparisons by controlling the false discovery rate to 0.1. Growth conditions with the same chemical compound but different concentrations were regarded as the same condition, and all genes identified in different concentrations of the same compound were regarded as contributing genes under that condition. On average, there are 368 genes in each subnetwork, and the relevant data were downloaded from http://chemogenomics.stanford.edu:16080/supplements/global/download.html.

In the synthetic genetic array (SGA) study (Costanzo et al. 2010), the authors screened 1,712 *S. cerevisiae* query genes, including 334 conditional or hypomorphic alleles of essential genes, against 3,885 array genes to generate a total of more than 5 million gene pairs spanning all biological processes. These queries were selected randomly with respect to function, while the array genes represented the whole collection of nonessential genes. In each gene pair, the epistasis value is calculated based on the equation: $\epsilon = W_{xy} - W_x W_y$, in which W_{xy} is the fitness of an organism with mutations in both genes X and Y, whereas W_x refers to the organism with the mutation in gene X but not gene Y (and vice versa for W_y). In addition, a statistical confidence measure (P value) was assigned to each interaction based on a combination of the observed variation of each double mutant across four experimental replicates and estimates of background log-normal error distributions for the corresponding query and array mutants. Finally, a defined confidence threshold ($|\epsilon| > 0.08$, $P < 0.05$) was applied to identify epistatic interactions (Costanzo et al. 2010). The gene pairs with epistatic interactions were downloaded from http://drygin.ccbr.utoronto.ca/~costanzo2009/.

Calculation of Clustering Coefficient for Epistatic Subnetworks

The clustering coefficient is a measure of the degree to which nodes in a network tend to be clustered together. For the node j with the connectivity i ($i > 1$) in a network, its clustering coefficient C_j is defined as the following:

$$C_j = \frac{2n_j}{i(i-1)},$$

where n_j is the total number of links connecting all the neighbors of the node j (Barabási and Oltvai 2004). The average clustering coefficients for each of the 354 studied traits were calculated using clustering coefficients of contributing genes in the corresponding epistasis subnetworks (Li et al. 2010).

Statistical Fitting for the Scale-Free Distribution

Scale-free topology means that the distribution of degree in the network, $P(K)$, approximates a power law:

$$P(K) = K^{-\nu},$$

where K is the degree and ν is the degree exponent, which is usually a constant for a specific network (Barabási and Oltvai 2004). In our analyses, the degree (K) was calculated as the number of epistatic interactions for each contributing gene in each of the 354 epistasis subnetworks. We then calculated the average frequency of each degree value among all 354 traits and plotted the frequency distribution of the network degree in figure 2*A*. MATLAB (Mathworks) was used to fit the regression.

Results and Discussions

Prevalent and Modular Epistasis among Genes Underlying the Same Growth Traits

In order to study the genetic architecture of growth traits, we firstly identified genes that are responsible for growth traits. Based on a genome-wide screen for growth defects of ~6,000 *S. cerevisiae* gene deletion mutants in 354 distinct growth conditions, genes that contribute to growth in each condition were defined as those genes heterozygous deletion of which significantly affect organism growth in that condition (Hillenmeyer et al. 2008). To ensure that these 354 conditions represent independent growth traits, we calculated the overlap of contributing genes between any 2 of the 354 conditions. As shown in supplementary fig. 1 (Supplementary Material online), 96% comparisons between any two conditions have less than 10% overlap of contributing genes and 99% comparisons have less than 20% overlap, indicating that most of the 354 conditions are functionally independent. In addition, we took advantage of epistatic interaction data in yeast from a recent study (Costanzo et al. 2010), in which epistatic interactions are examined among more than 5 million gene pairs in *S. cerevisiae*. Subnetworks with epistatic interactions among contributing genes for each of the 354 growth conditions were reconstructed.

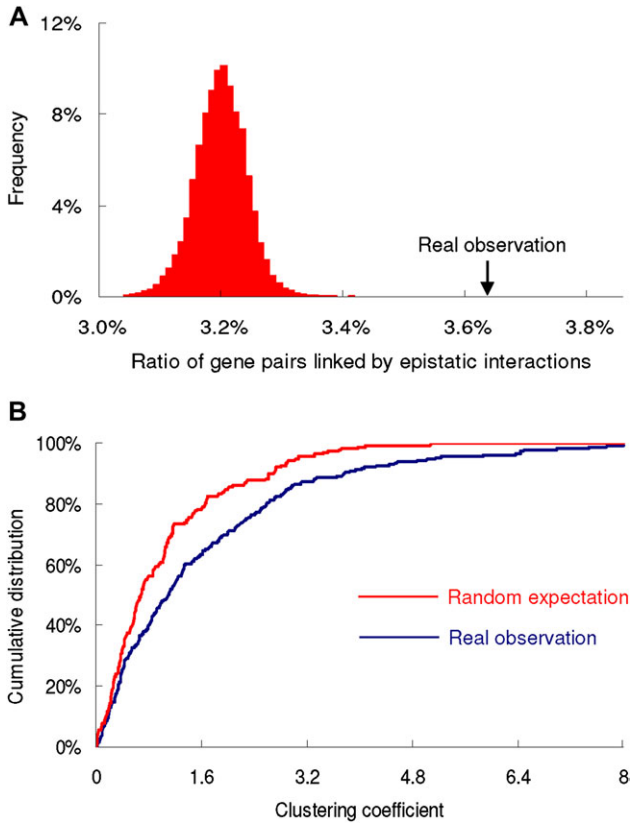

Fig. 1.—Prevalent and modular epistasis in the genetic architecture of growth traits. (*A*) The distribution (red color) represents the average ratio of contributing gene pairs that are linked by epistatic interactions in the 354 traits based on random simulations (repeated 100,000 times). The arrow indicates the average ratio of contributing gene pairs that are linked by epistatic interactions in the 354 traits based on real experimental data. (*B*) The empirical cumulative distribution of the clustering coefficients for experimental observations (all 354 traits, blue curve) and random simulations (repeated 100,000 times for the 354 traits, red curve). The Kolmogorov–Smirnov test indicates that the two distributions are significantly different ($P = 2 \times 10^{-31}$).

While most people agree that epistasis plays an important role in the genetic architecture of complex traits, there is a disagreement about how common epistatic interactions are within genes that contribute to the same trait (Moore 2003; Phillips 2008). Using the reconstructed 354 epistasis subnetworks, we found that when two genes are responsible for the same growth trait, on average, 3.6% of them are linked by an epistatic interaction. We then conducted a simulation by keeping the number of genes responsible for each trait as a constant, but randomly choosing genes to be responsible for each trait (repeated 100,000 times). In each iteration, we also calculated the fraction of gene pairs connected by epistatic interactions. As shown in figure 1*A*, a significantly higher ratio was observed for the real experimental data than that of random expectation (~3.2%, fig. 1*A*, $P < 10^{-5}$), indicating that epistasis is enriched among genes responsible for the same biological traits. It is also noteworthy that the increase of epistasis (~13% more than random expectation) among contributing genes underlying the same growth trait in yeast is not dramatic.

Previous studies proposed that gene pairs linked by epistatic interactions would be predictive of participation in common cellular functions (Tong et al. 2004; Costanzo et al. 2010). However, although our above result is consistent with this expectation, it has never been shown before that the contributing genes underlying the same traits are also enriched with epistatic interactions. To further understand whether genes that contribute to the same growth traits are closely connected by epistasis, we calculated the average clustering coefficient, a network parameter that reflects the tightness of connection for a group of genes by immediate interactions (Barabási and Oltvai 2004), for genes that underlie each growth trait. The larger the clustering coefficient is, the more interlinked the group of genes are. For each of the 354 biological traits, we calculated the average clustering coefficient among its contributing genes. For comparison, we also calculated the average clustering coefficient for each trait in each of the above 100,000 simulations. Figure 1*B* shows the cumulative distributions of the clustering coefficients for real observation and random simulations. Our result indicates that genes underlying the same biological traits tend to be closely interconnected by epistatic interactions (Kolmogorov–Smirnov test, $P = 2 \times 10^{-31}$).

Assortative Characteristic of Epistatic Interactions for Growth Traits

We further investigated how epistasis is distributed among the contributing genes for each growth trait. Most biological networks are scale-free, meaning that the network consists of a small number of highly connected "hub" genes and a majority of genes with few interactions (Barabási and Oltvai 2004). The degree (connectivity in the network) in a scale-free network usually follows a so-called "power-law" distribution. To examine whether the epistatic interactions among the contributing genes that underlie biological traits also display the scale-free characteristic, we calculated the connectivity for all contributing genes in each of the 354 subnetworks, respectively. We then investigated the distribution of degrees that were averaged over all 354 subnetworks. Figure 2*A* confirms that epistatic interactions underlying growth traits follow the power-law distribution. Contributing genes underlying most individual traits also show the similar pattern (supplementary fig. 2, Supplementary Material online).

How is epistasis distributed among the contributing genes with different connectivity? To answer this question, for each trait, we first computed the number of epistatic interactions (N_{ko}) among the contributing genes that have more than k interactions in each of the observed epistasis subnetworks. Randomized versions of the epistasis subnetwork were also generated for that trait, in which all the

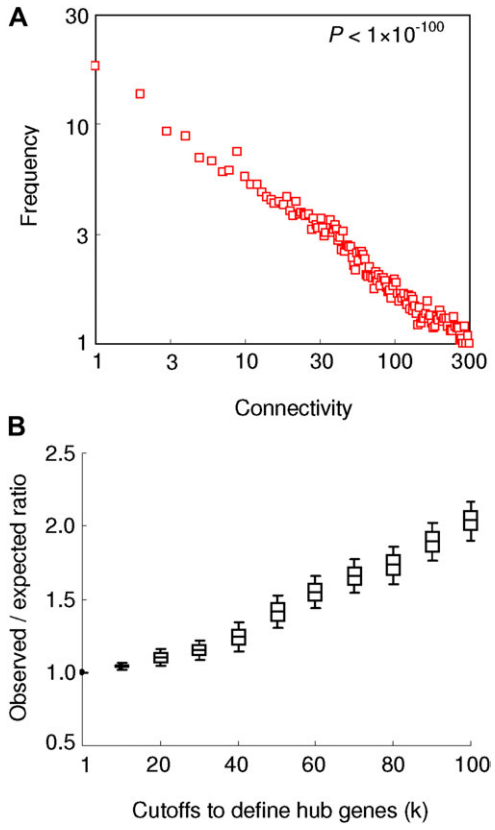

Fig. 2.—Assortative genetic architecture of growth traits. (*A*) The degree distribution of epistatic networks over 354 biological traits. MATLAB (Mathworks) was used to fit the regression and the small *P* value indicates that the network degree displays the scale-free characteristic. (*B*) Average ratio of observed/expected number of epistatic interactions among the 354 traits. For each epistasis subnetwork, the number of epistatic interactions among all contributing genes that have more than *k* epistatic interactions was calculated (the observed numbers). The epistatic interaction in the subnetwork was randomized and the average number of epistatic interactions among all contributing genes that have more than *k* epistatic interactions was also calculated from 1,000 random simulations (the expected numbers). The bands, boxes, and whiskers represent the means, ±1 standard errors, and ±95% confidence intervals, respectively.

contributing genes have the same degree as the real epistasis subnetwork, but the epistatic interactions between the contributing genes are randomly connected. For each trait, we calculated the average number of interactions (N_{ks}) among the contributing genes that have more than *k* interactions from 1,000 randomly generated networks. For each value *k*, we computed the average ratio of N_{ko}/N_{ks} among all 354 traits. As depicted in figure 2*B*, the ratio of N_{ko}/N_{ks} increases with *k*, indicating that epistasis is enriched among the contributing genes that are highly connected in the epistatic networks. When the ribosomal proteins and chaperones, which might represent universal hubs in the epistatic interaction network, are excluded, the pattern still holds (supplementary fig. 3, Supplementary Material online).

When epistatic interactions among genes that affect organism growth in a particular condition are investigated, the contributing hub genes for this trait are by definition more likely than the nonhub genes to be linked by epistatic interactions. However, this increase in connectivity for the hub genes could be due to increased interactions either linking to other hub genes or linking to nonhub genes. Indeed, the unique network architecture of enriched epistatic interactions among the hub genes revealed in this study, which is termed as "assortative" in social networks (Newman 2002), is surprising because all previous analyses of available cellular networks, including protein–protein interaction networks, transcriptional regulatory interaction networks, and metabolic interaction networks, display disassortative topologies in which the connections among hub genes are systematically suppressed, and the high connectivity for the hub genes in these networks are caused by enriched interactions between hub and nonhub genes (Maslov and Sneppen 2002; Newman 2002).

Implication for Pleiotropy, Epistasis, and Complex Traits

Why do highly connected hub genes tend to epistatically interact with each other more frequently than expected? It was shown that the hub genes in the epistasis network, when mutated, tend to display impacts on more phenotypes than the nonhub genes and thus are more likely to have a higher level of pleiotropy (Costanzo et al. 2010). It might be true that highly pleiotropic genes would have higher chances of developing functional overlaps among themselves in the fixed functional space of a cellular system. In addition, previous studies showed that two genes with overlapping functions tend to be linked by epistatic interactions (Tong et al. 2004; Costanzo et al. 2010). As a result, highly pleiotropic hub genes would have higher chance to develop epistatic interactions among themselves. Our observation in figure 2*B* is consistent with this scenario, indicating that pleiotropy might play an important role in shaping the genetic architecture of complex traits (Wagner and Zhang 2011).

Although we found several novel characteristics for the genetic architecture of growth traits, several caveats need to be addressed. First, epistatic interactions used here were inferred from high-throughput experiments, which were mostly based on double gene deletion mutants. These mutations are likely to be different from most epistatically interacting mutations that underlie organism phenotypic differences in nature. Second, growth under environmental perturbations was used to represent biological traits (Hillenmeyer et al. 2008), which are also different from naturally occurring phenotypic traits. Third, the epistatic interactions, which are deduced from single and double mutants, are

incomplete because real epistasis underlying the growth traits could exist among more than two genes. Future high-throughput dissections, if possible, on the phenotypic consequences of naturally occurring genetic variations will help illustrate the genetic architecture of growth traits. For the moment, the approach in this study, which was used in recent studies (e.g., Dowell et al. 2010), represent excellent tools to investigate this issue. We also need to point out that epistasis among genes could be condition specific, as shown in a recent study (Bandyopadhyay et al. 2010). However, using the same data set from the paper, we were able to show that the majority of the sign of epistases are shared between two conditions (supplementary fig. 4, Supplementary Material online). With these limitations in mind, our observations identified several important features of the genetic architecture of growth traits and indicate the importance of future effort for addressing the architecture of epistatic interaction networks in illustrating the genetic basis of complex traits, including human diseases.

Supplementary Material

Supplementary figures 1–4 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

We thank Balazs Papp, Patrick Stover for discussions, and Brandon Barker, Nathan Clark, Anthony Greenberg, Andy Clark and Allan Eaglesham for reading the manuscript, and three anonymous reviewers for their helpful comments. The authors are grateful to Drs Michael Costanzo and Charlie Boone for sharing data. This work was supported by startup fund from Cornell University, National Science Foundation (DEB-0949556), and National Institutes of Health (1R01AI085286-01) awarded to Z.G.

Literature Cited

Avery L, Wasserman S. 1992. Ordering gene function: the interpretation of epistasis in regulatory hierarchies. Trends Genet. 8:312–316.

Azevedo RB, Lohaus R, Srinivasan S, Dang KK, Burch CL. 2006. Sexual reproduction selects for robustness and negative epistasis in artificial gene networks. Nature 440:87–90.

Bandyopadhyay S, et al. 2010. Rewiring of genetic networks in response to DNA damage. Science 330:1385–1389.

Barabási AL, Oltvai ZN. 2004. Network biology: understanding the cell's functional organization. Nat Rev Genet. 5:101–113.

Brady A, Maxwell K, Daniels N, Cowen LJ. 2009. Fault tolerance in protein interaction networks: stable bipartite subgraphs and redundant pathways. PLoS One. 4:e5364.

Carlborg O, Haley CS. 2004. Epistasis: too often neglected in complex trait studies? Nat Rev Genet. 5:618–625.

Carlborg O, Jacobsson L, Ahgren P, Siegel P, Andersson L. 2006. Epistasis and the release of genetic variation during long-term selection. Nat Genet. 38:418–420.

Costanzo M, et al. 2010. The genetic landscape of a cell. Science 327:425–431.

Dowell RD, et al. 2010. Genotype to phenotype: a complex problem. Science 328:469.

Ehrenreich IM, Stafford PA, Purugganan MD. 2007. The genetic architecture of shoot branching in Arabidopsis thaliana: a comparative assessment of candidate gene associations vs. quantitative trait locus mapping. Genetics 176:1223–1236.

Hansen TF, Wagner GP. 2001. Epistasis and the mutation load: a measurement theoretical approach. Genetics 158:477–485.

Hartman JL, Garvik B, Hartwell L. 2001. Principles for the buffering of genetic variation. Science 291:1001–1004.

Hillenmeyer ME, et al. 2008. The chemical genomic portrait of yeast: uncovering a phenotype for all genes. Science 320:362–365.

Kelley R, Ideker T. 2005. Systematic interpretation of genetic interactions using protein networks. Nat Biotechnol. 23:561–566.

Kondrashov AS. 1982. Selection against harmful mutations in large sexual and asexual populations. Genet Res. 40:325–332.

Kondrashov AS, Crow JF. 1991. Haploidy or diploidy: which is better? Nature 351:314–315.

Legare ME, Bartlett FS, Frankel WN. 2000. A major effect QTL determined by multiple genes in epileptic EL mice. Genome Res. 10:42–48.

Li J, et al. 2010. Exploiting the determinants of stochastic gene expression in *Saccharomyces cerevisiae* for genome-wide prediction of expression noise. Proc Natl Acad Sci U S A. 107:10472–10477.

Ma X, Tarone AM, Li W. 2008. Mapping genetically compensatory pathways from synthetic lethal interactions in yeast. PLoS One. 3:e1922.

Manolio TA, et al. 2009. Finding the missing heritability of complex diseases. Nature 461:747–753.

Maslov S, Sneppen K. 2002. Specificity and stability in topology of protein networks. Science 296:910–913.

Moore JH. 2003. The ubiquitous nature of epistasis in determining susceptibility to common human diseases. Hum Hered. 56:73–82.

Musso G, et al. 2008. The extensive and condition-dependent nature of epistasis among whole-genome duplicates in yeast. Genome Res. 18:1092–1099.

Newman MEJ. 2002. Assortative mixing in networks. Phys Rev Lett. 89:208701.

Otto SP. 2007. Unraveling the evolutionary advantage of sex. Genet Res. 89:447–449.

Perez-Figueroa A, Caballero A, Garcia-Dorado A, Lopez-Fanjul C. 2009. The action of purifying selection, mutation and drift on fitness epistatic systems. Genetics 1831:299–313.

Phillips PC. 2008. Epistasis: the essential role of gene interactions in the structure and evolution of genetic systems. Nat Rev Genet. 9:855–867.

Presgraves DC. 2007. Speciation genetics: epistasis, conflict and the origin of species. Curr Biol. 17:R125–R127.

Remold SK, Lenski RE. 2004. Pervasive joint influence of epistasis and plasticity on mutational effects in Escherichia coli. Nat Genet. 36:423–426.

Sanjuan R, Nebot MR. 2008. A network model for the correlation between epistasis and genomic complexity. PLoS One. 37:e2663.

Shao H, et al. 2008. Genetic architecture of complex traits: large phenotypic effects and pervasive epistasis. Proc Natl Acad Sci U S A. 105:19910–19914.

Tong AH, et al. 2004. Global mapping of the yeast genetic interaction network. Science. 303:808–813.

Trindade S, et al. 2009. Positive epistasis drives the acquisition of multidrug resistance. PLoS Genet. 5:e1000578.
Visscher PM, Hill WG, Wray NR. 2008. Heritability in the genomics era: concepts and misconceptions. Nat Rev Genet. 9:255–266.
Wagner GP, Zhang J. 2011. The pleiotropic structure of the genotype-phenotype map: the evolvability of complex adaptations. Nat Rev Genet. 12:204–213.

Associate editor: Esther Betran