

# Genome-wide scan reveals that genetic variation for transcriptional plasticity in yeast is biased towards multi-copy and dispensable genes

Christian R. Landry<sup>a</sup>, Julia Oh<sup>b,1</sup>, Daniel L. Hartl<sup>a</sup>, Duccio Cavalieri<sup>b,\*</sup>

<sup>a</sup> Department of Organismic and Evolutionary Biology, Harvard University, Cambridge MA, USA

<sup>b</sup> Bauer Center for Genomics Research, Harvard University, Cambridge MA, USA

Received 26 May 2005; received in revised form 14 August 2005; accepted 14 October 2005

Available online 20 January 2006

Received by C. Saccone

## Abstract

One of the most important aspects of the evolution of development and physiology is the interplay between gene expression and the environment, by which traits become altered in response to environmental triggers. This feature is known as phenotypic plasticity. When different genotypes show different levels of plasticity for a trait, then they show genotype-by-environment interaction, or GEI. It is now clear that gene expression plays an important role in organismic-level phenotypic plasticity, but we know very little about whether gene expression itself is subject to genetic variation for phenotypic plasticity (GEI). Given that gene regulation is likely to have evolved to respond to environmental changes, it is of central importance to understand how environmental and genetic variation interact to produce variation in gene expression. Here we investigate genetic variation for phenotypic plasticity in the yeast transcriptome for the whole genome. Six strains of *Saccharomyces cerevisiae* were grown in four different environments representing a continuum of rich and poor natural conditions. Using DNA-microarray data and an ANOVA analysis with a stringent criterion of significance, we found significant genetic variation for transcriptional plasticity (GEI) among strains for ~5% of the genes in the genome. There are about twice as many genes that show genetic variation for phenotypic plasticity as show genetic variation in transcription level independent of the environment. We also found that genes with genetic variation for plasticity were less likely to be essential and were significantly biased towards genes that have paralogs.

© 2005 Elsevier B.V. All rights reserved.

**Keywords:** Gene expression; *Saccharomyces cerevisiae*; Gene duplicates; Environment; Phenotypic plasticity

## 1. Introduction

Understanding the genetic bases of complex diseases and adaptive traits are among the fundamental goals of medical and evolutionary genetics. The environment is usually seen as the agent of natural selection, selecting the fittest genotypes in a

given population. But the environment is far from constant. Moreover, genetic and environmental effects may interact with each other and may produce complex evolutionary trajectories. For instance, phenotypic changes in response to environmental changes—known as phenotypic plasticity—can themselves be genetically variable. Also, the evolution of robustness against environmental perturbations can lead to the evolution of robustness also against genetic perturbations, consequently redrawing the maps that connect genotype and environment with phenotype and fitness (Waddington; 1942; Hartman et al., 2001; Meiklejohn and Hartl, 2002). Consideration of both genetic and environmental perturbations is therefore necessary for a more complete understanding of evolutionary processes (Levins, 2004).

Genetic variation for phenotypic plasticity emerges from the non-additive interaction between the genotypes of organisms

*Abbreviations:* GEI, genotype-by-environment interaction; YPD, yeast-peptone-dextrose rich medium; SWM, synthetic wine must; NS4, 4 h nitrogen starvation; NS24, 24 h nitrogen starvation; ANOVA, analysis of variance; KEGG, Kyoto Encyclopedia of Genes and Genomes.

\* Corresponding author. Department of Pharmacology, University of Florence, Viale Pieraccini 6, 50139 Florence, Italy. Tel.: +39 055 4271327; fax: +39 055 4271280.

E-mail address: [duccio.cavalieri@unifi.it](mailto:duccio.cavalieri@unifi.it) (D. Cavalieri).

<sup>1</sup> Present address: Department of Biological Sciences, Stanford University, Stanford CA, USA.

and variation in the environment. In the presence of such interactions, different genotypes respond differently to the same environmental change, so there is genotype-by-environment interaction (GEI). Sensitivity to therapeutic drugs in individuals with some genotypes but not others is a manifestation of GEI in pharmacogenomics. The visualization of GEI can be facilitated through the use of norms of reaction—a mapping of the phenotype onto the genotype as a function of the environment, depicted graphically as a plot of phenotypic values against environmental values (Stearns, 1992, Fig. 1). GEI refers to genetic variation in the shape of those norms of reaction. Different hypothetical norms of reaction are illustrated in Fig. 1, in which each color represents a different genotype. Panel (A) shows a trait with genetic variation but no phenotypic plasticity, and panel (B) shows a trait with phenotypic plasticity but no genetic variation. The trait in panel (C) exhibits both genetic variation and phenotypic plasticity, but there is no genetic variation in phenotypic plasticity: each genotype responds in the same way to each environment. The only trait in Fig. 1 that shows genetic variation for phenotypic plasticity (GEI) is that depicted in panel (D), where the norms of reaction of the different genotypes are not parallel.

The machinery of gene regulation can exhibit an exquisite response to the environment because it has evolved to allow transcription to be induced, repressed, or modulated in response to environmental cues (Suiter et al., 2003). The transcriptional response can be at the level of individual cells, tissues, or the whole organism. In fact, a large fraction of phenotypic diversity within species may result from phenotypic plasticity controlled at the level of differential gene expression (Pigliucci, 1996; Abouheif and Wray, 2002). Studying gene regulation is therefore key to our understanding of the genetic and environmental dissection of complex phenotypes, and this in the perspective of better understanding phenotypic evolution, which is a grand challenge facing evolutionary biology (Singh, 2003). To this end, microarray technologies have become

valuable tools because they enable the survey of gene expression at a genomic scale and this quantitatively (e.g., Gibson, 2002).

In order to quantify the extent of genetic variation for transcriptional plasticity, we studied gene expression in *Saccharomyces cerevisiae*. This unicellular fungus is particularly well suited for such an experiment because the time scale of environmental changes can be much shorter than the generation time (e.g., Gasch and Werner-Washburne, 2002), a condition known to favor the evolution of phenotypic plasticity (Schlichting and Smith, 2002). Additionally, transcriptional responses in microorganisms are at the organismic-level and do not include confounding factors of differential responses in distinct tissues or organs. Finally, the metabolism of microorganisms is relatively tightly linked to genetic regulatory gene networks whose transcriptional state can be estimated by assaying mRNA levels. In yeast, genetic variation in gene expression in a single environment has been documented (e.g. Cavalieri et al., 2000; Townsend et al., 2003), as have changes in gene expression in different environments (e.g., Gasch et al., 2000; Causton et al., 2001). There is also evidence that different strains show different responses to chemical agents (Fay et al., 2004). What is lacking, however, is a large-scale assay for GEI, which would reveal how much genetic variation is present in the transcriptional response to environmental changes and what genetic characteristics of this species favor or limit the diversity of transcriptional responses to environmental triggers. Because genetic variation for phenotypic plasticity emerges from the interaction between genotypes and environments, it can be assayed only by examining the transcriptional response of multiple genotypes across several environments. This can now be carried out on a genomic scale.

In this study, we used cDNA microarrays to examine genetic variation for phenotypic plasticity in genome-wide gene expression in six isolates of *S. cerevisiae* grown in four different environments, either in standard rich medium or any of three alternative media simulating environmental stress. The specific objective was to reveal which general classes of genes show genetic variation for phenotypic plasticity. We also hoped to identify factors that might constrain or favor the diversity of norms of reaction in the genome of *S. cerevisiae*.

## 2. Material and methods

### 2.1. Genotypes and environments

Six diploid prototrophic strains were studied: Ds288c, EM93, Sgu52, Sgu407, Sgu421, and Sg60. Ds288c is a diploid heterothallic strain derived from a cross of two S288c derivatives isolated in the 1980s. EM93 was originally isolated from a fig in Mercedes, California, in the early days of yeast genetics (Mortimer and Johnston, 1986). The other strains are from the Polsinelli collection of Italian yeast isolates: Sgu52, Sgu407, and Sgu421 were isolated from grapes in the Chianti region, and Sg60 from natural wine fermentation. Strains were grown in YPD (laboratory conditions), synthetic wine must (SWM) simulating fermentation conditions, or in low-nitrogen,

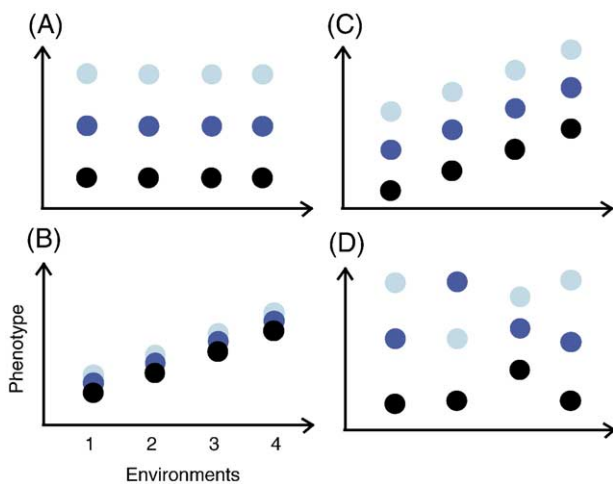


Fig. 1. Norms of reaction illustrating (A) genetic variation only, (B) phenotypic plasticity only, (C) and (D) both phenotypic plasticity and genetic variation. Only (D) shows genetic variation for phenotypic plasticity as revealed by the non-parallel norms of reaction.

proline-supplemented SWM (SWM\_NS). In order to simulate fermentation conditions, the strains were grown in flasks with an oxygen trap on the top, preventing the exchange of oxygen. The medium was however not prepared anaerobically so initially the culture had some oxygen. The standard nutrient medium for yeast cultivation (YPD) is composed of 1% yeast extract, 2% peptone, and 2% dextrose. SWM contains 0.17% yeast nitrogen base, without amino acids and  $(\text{NH}_4)_2\text{SO}_4$ , 0.15% casamino acids, 0.05%  $\text{NH}_4\text{Cl}$ , 0.60% DL-malic acid, 0.02% citric acid, 0.15% L-tartaric acid, 21% glucose, and 0.20% anaerobic factors ergosterol–Tween 80 (total 10 mg/L ergosterol, 0.5 ml/L Tween 80). The pH was adjusted to 3.3 with KOH. SWM has a total nitrogen concentration of 247.2 mg/L. The nitrogen starvation medium (SWM\_NS) was composed of 0.17% yeast nitrogen base, 0.1% L-proline, 0.60% DL-malic acid, 0.02% citric acid, 0.15% L-tartaric acid, 21% glucose, and 0.20% ergosterol–Tween 80 (10 mg/L ergosterol, 0.5 ml/L Tween 80). The pH was adjusted to 3.3. SWM\_NS has a total nitrogen concentration of 123 mg/L. All cells were grown at 28 °C. For YPD, cells were grown aerobically in a shaker at 200 rpm and harvested at an optical density (OD) of 0.8 ( $A_{600}$ ). SWM and SWM\_NS: Strains were grown aerobically overnight in SWM. Cells were then diluted to 0.2 OD in SWM and left to grow to 0.8 OD in a shaker at 50 rpm. Anaerobic conditions were established by locked fermentation flasks with  $\text{CO}_2$  outlets filled with water. For growth under nitrogen starvation, cells were then pelleted (3000  $\times$ g, 5 min), washed with sterile water, resuspended in SWM\_NS medium, and left to grow for 4 or 24 h under the same conditions. For all conditions, cells were harvested by centrifugation (3000  $\times$ g, 10 min) at then flash frozen using liquid nitrogen.

## 2.2. Gene expression analysis

The experimental data consist of the results of 49 competitive hybridizations carried out according to the design diagrammed in Fig. S1. The raw data have been submitted to Gene Expression Omnibus (accession number GSE3021). This design represents a practical trade-off between efficient use of the experimental materials and the perfectly balanced design ideal for analysis of variance. As a reference strain for the hybridizations, we used the strain BY4743, a derivative of S288c, grown in YPD. This scheme of comparisons resulted in the following number of replicates per genotype:environment combinations: EM93:YPD=6, EM93:SWM=3, EM93:NS4=2, EM93:NS24=3, Ds288c:YPD=5, Ds288c:SWM=5, Ds288c:NS4=3, Ds288c:NS24=4, Sg60:YPD=5, Sg60:SWM=3, Sg60:NS4=2, Sg60:NS24=5; Sgu407:YPD=4; Sgu407:SWM=3; Sgu407:NS4=2, Sgu407:NS24=4; Sgu421:YPD=4, Sgu421:SWM=3, Sgu421:NS4=2, Sgu421:NS24=3; Sgu52:YPD=5; Sgu52:SWM=4, Sgu52:NS4=3, Sgu52:NS24=4. The hybridization protocol and data acquisition steps are described in Townsend et al. (2003). Spots were visually inspected and flawed ones discarded from the analysis. Data were normalized to a mean ratio intensity of unity within arrays using the loess method implemented in Limma

(Smyth, 2004) in the Bioconductor package (Ihaka and Gentleman, 1996). The median of spot intensities was corrected for background intensities using the method “minimum”: any intensity which is zero or negative after background subtraction is set equal to half the minimum of the positive corrected intensities for that array. A total of 5258 genes could be analyzed.

Significant differences in gene expression among strains, environments and their interaction were obtained by applying an analysis of variance (ANOVA) to each gene individually as implemented in the MAANOVA package (Kerr and Churchill, 2001; Cui and Churchill, 2003). MAANOVA implements a two-stage ANOVA model, with the first stage normalizing among arrays. The second stage uses the residuals of the first stage to test for the effect of the strain and the growth conditions on the level of expression and their interaction. We used two ANOVA models within the MAANOVA framework. One full model tests the interaction between the effect of the genotype and the environment:  $y_{kij} = \text{Dye}_k + u + S_i + E_j + S_i E_j + \text{Array}$ , where  $\text{Dye}_k$  captures the average effect of dye  $k$ ,  $u$  the average effect of gene,  $S_i$  the average effect of Strain  $i$ ,  $E_j$  the average effect of the growth condition  $j$  and  $S_i E_j$  the average effect of the interaction of Strain  $i$  in growth condition  $E_j$ . The Array effect is a random factor accounting for technical replicates and the Dye effect is a fixed factor. We then used a simplified model:  $y_{kij} = \text{Dye}_k + u + S_i + E_j + \text{Array}$ . This model tests for the effect Strain  $i$  and growth condition  $j$ . The significance of the effects was computed using the tabulated  $F_s$  statistics. This test statistics uses an estimator of variance that can borrow information across genes. This provides a powerful and robust approach to test differential expression of genes that utilizes information not available in individual gene testing approaches and does not suffer from biases of the pooled variance approach (see Cui et al., 2005). We used  $P$ -values with a significance threshold of 0.001 to determine significant effects. At this level of stringency, we should expect only 0.1% of false positives.

Two main methods are used for the analysis of microarray data, ratio-based methods and single channel based methods. ANOVA methods rely on the single channel fluorescence estimates as a measure of gene expression. In ratio-based methods, the relative expression of two samples on a single array is assessed through the relative intensities on the same spot, which allows transitive comparison for experimental designs as the one here. In order to examine the robustness of the ANOVA approach, we compared our results with a ratio-based approach. We found that there is a good agreement between the relative gene expression level as estimated with the ANOVA model with a ratio-based method (Supplementary material, Appendix 1), so the ANOVA results only are presented here.

Norms of reaction of genes whose expression level was significantly affected by the growth conditions, but without significant GEL, were clustered using a  $k$ -means clustering algorithm. The optimal number of clusters to be used for this set of 1862 genes was determined using the number of clusters with the lowest Figure of Merit values (Yeung et al., 2001).  $K$ -means clustering needs a given number of groups, which can be

difficult to guess in most of the cases. The figure of merit of a clustering method measures the predictive power of a clustering algorithm estimated by removing one experiment from the data set and clustering the genes based on the remaining of the data, and then measuring the within-cluster similarity of expression in this experiment. A total of 5000 bootstrap iterations of the clustering were performed and consensus clusters were determined for genes that clustered more than 50% of the time in the same cluster using the methods implemented in MAANOVA (Kerr and Churchill, 2001). A total of 1383 genes could be clustered in 12 groups, a number chosen based on the Figure of Merit method (Fig. 3). Genes within each group were examined for functional enrichment using Gene Ontologies categories (GO) and KEGG metabolic pathways using hypergeometric statistics as implemented in GeneMerge (Castillo-Davis and Hartl, 2003). The most significant enrichments are presented in Fig. 3. Yeast deletion data were obtained from (<http://mips.gsf.de/projects/fungi/yeast.html>). Genes that were identified as being essential in some studies and nonessential in different studies were labeled as missing data. Fitness of gene knockouts in yeast (Steinmetz et al., 2002; Winzeler et al., 1999) was obtained from ([http://www.deletion.stanford.edu/YDPM/YDPM\\_index.html](http://www.deletion.stanford.edu/YDPM/YDPM_index.html)). Growth rate of specific mutants were expressed relative to the average growth rate of the mutants and averaged over the two experiments. Genes with paralogs were identified as defined in Davis and Petrov (2004). Briefly, the set of *S. cerevisiae* proteins was submitted to BLASTP search (Altschul et al., 1997) against the entire proteome with an *E*-value cutoff of  $1E-10$ . Proteins that were reciprocal hits and that aligned over at least 60% of the shortest protein were considered as paralogs of each other. A total of 1905 genes with at least one paralog in the genome were identified. We also performed the same analyses with more stringent criteria (*E*-value =  $10E-11$  and alignment at least 80% the size of the shortest protein), which yielded 1180 genes with at least a paralog. The results from analyzing this more stringent set were comparable with those of the larger set reported. All statistical analyses were performed using R (R Development Core Team, 2005).

### 3. Results and discussion

#### 3.1. Variation in gene expression across environments

Gene expression was estimated in replicate experiments in which each strain of yeast was grown in four environmental conditions: aerobic growth in YPD medium, a typical rich laboratory condition; anaerobic growth in synthetic wine must (SWM), which mimics wine fermentation conditions; and either 4 or 24 h of anaerobic growth with nitrogen starvation (NS) in medium in which proline was the main source of nitrogen. These environmental conditions simulate the progression of environments that yeast cells may encounter as they ferment rotting fruit. The environmental stresses include anaerobiosis, osmotic stress associated with high glucose (10–40%), acidity (pH=2.5–3.8 due to organic acid content), and nitrogen starvation. Nitrogen starvation is an ecologically relevant stress

because nitrogen is known to be limiting in nature (Bisson, 1991). Nitrogen limitation is also known to trigger putatively important ecological traits in yeast, such as filamentous growth and foraging into solid nutrients (Gimeno et al., 1992).

We estimated the contribution of genotype (strain), environment (growth condition) and their interaction on gene expression variation using an ANOVA model for each of 5258 genes assayed. Detecting significant effects of those factors on gene expression level represent respectively genetic variation for gene expression (G), phenotypic plasticity (E), and genetic variation for phenotypic plasticity (GEI, genotype-by-environment interaction). A total of 127 genes varied in expression level among the strains independently of the growth conditions ( $P_G < 0.001$ ,  $P_{GEI} > 0.001$ ). This is a set of genes whose transcription level is genetically variable but this variation is not environment-dependent. The transcriptional level of 1863 genes was affected by the growth conditions independently of the genotype ( $P_E < 0.001$ ,  $P_{GEI} > 0.001$ ), which represent roughly 1/3 of the genes surveyed. Finally, 223 genes showed significant interaction between the genotype and the growth conditions on transcription ( $P_{GEI} < 0.001$ ). This latter group represents genes that respond differently to the environments examined, depending on the strain, and therefore represent genes that are genetically variable for transcriptional plasticity.

One example for each type of variation is represented in Fig. 2. For instance, the expression level of *Prm8* (Fig. 2 (a)) is not significantly affected by the conditions but varies among the strains ( $P_G < 4E-7$ ;  $P_E = 0.13$ ;  $P_{GEI} = 0.1$ ). *Prm8* codes for a pheromone-regulated protein involved in conjugation with

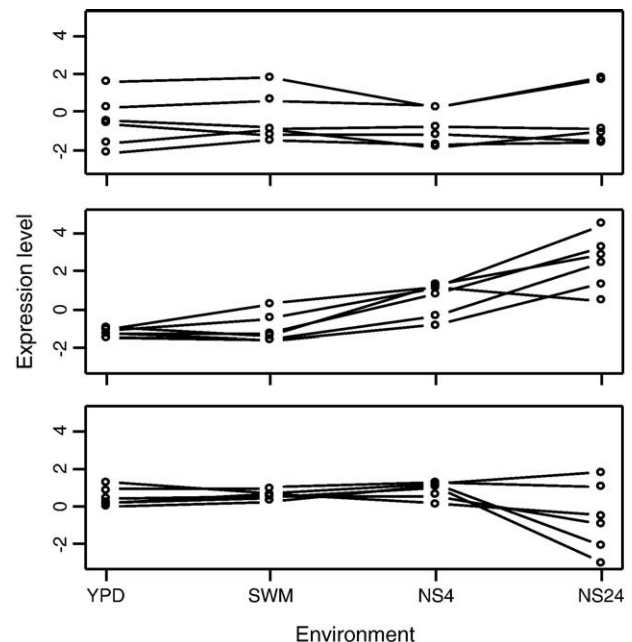


Fig. 2. Examples of genes showing (a) genetic variation in *Prm8* transcription, (b) phenotypic plasticity in transcription of *Msn4*, a major regulator of stress response in yeast, and (c) genetic variation for phenotypic plasticity of transcription of *Crh1*. Expression values represent deviation from the average expression ( $\log_2$ ) of the gene in the experiment. Along the x-axis are the four growth conditions.

cellular fusion (Heiman and Walter, 2000). On the other hand, the expression level of *Msn4* (Fig. 2 (b)), which encodes a transcription factor that is a major component of stress response in yeast (Causton et al., 2001), is highly plastic ( $P_E < 3E-11$ ;  $P_G = 0.02$ ;  $P_{GEI} = 0.06$ ). *Msn4* has been shown to be highly induced by nitrogen depletion and amino acid starvation in previous experiments, and also in other environmental response such as heat shock (Gasch et al., 2000). The pattern of expression of *Msn4* across the environmental gradient is therefore conserved among these strains. Finally, the expression of *Crh1* (Fig. 2 (c)) shows genetic variation for transcriptional plasticity ( $P_{GEI} < 5E-6$ ;  $P_E = 0.0003$ ;  $P_G = 0.001$ ). *Crh1* encodes a putative glycosidase of the cell-wall, and may have a role in cell wall architecture. It has been shown to be repressed during nitrogen depletion (>12 h) and amino acid starvation (>0.5 h) by Gasch et al. (2000) in a single-strain study.

In order to examine what are the cellular processes induced or modified by the environmental gradient, we clustered the norms of reaction of 1863 genes with significant phenotypic plasticity ( $P_E < 0.001$ ,  $P_{GEI} > 0.001$ ) into group of genes with similar patterns of expression. We

did not include genes with genetic variation for phenotypic plasticity because, by definition, these are genes whose norms of reaction depend on the genotype. Patterns of expression were grouped into twelve consensus clusters using *k*-means clustering (Kerr and Churchill, 2001) (Fig. 3). Each cluster is a group of genes with similar norms of reaction, each gene represented as the averaged environmental effect over all the strains. These are therefore transcriptional responses that are conserved among strains. We examined which biological processes are over-represented in each of the clusters using Gene Ontologies and KEGG metabolic pathways (Castillo-Davis and Hartl, 2003). For instance, genes of cluster 11 (Fig. 3, panel 11) show a decreasing expression in nitrogen limited media and are highly enriched for genes involved in ribosomal function ( $P < 2E-73$ ). The repression of ribosomal proteins is a typical response of yeast to environmental stresses (Causton et al., 2001). The most significant enrichments for all the clusters are described in the legend of Fig. 3. The conditions examined therefore induce coordinated responses at the transcriptional level that are biologically relevant for *S. cerevisiae*.

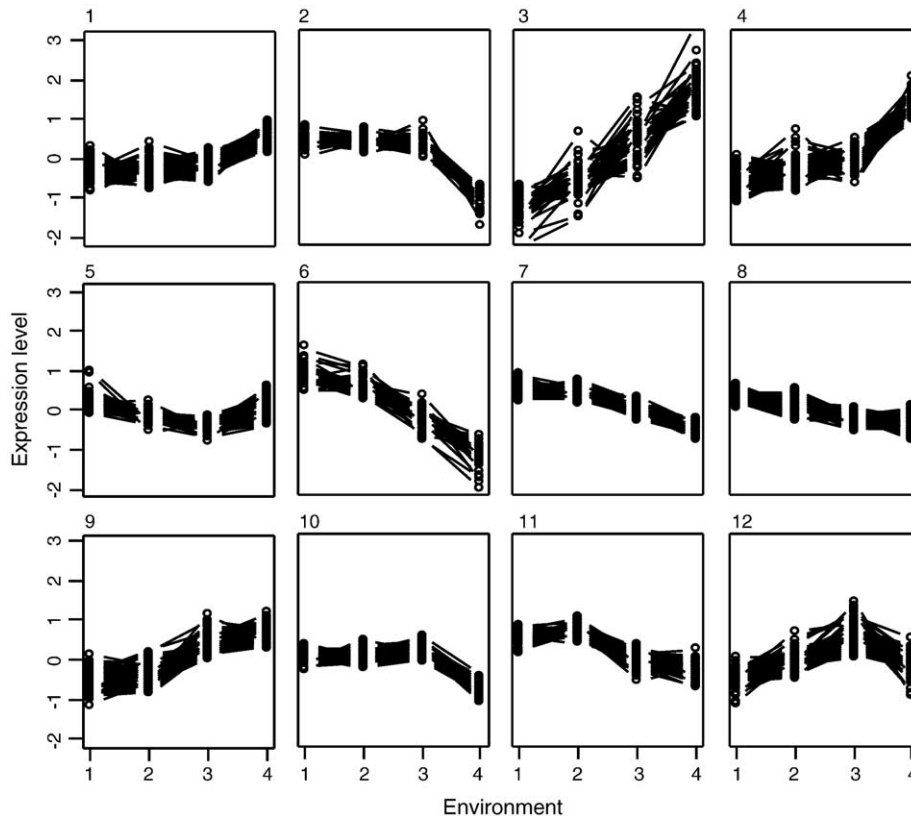


Fig. 3. Consensus clusters for genes whose transcription level is significantly affected by the environment ( $P_E < 0.001$ ) but show no significant GEI. From left to right in each cluster are the expression values in 1: YPD, 2: SWM, 3: NS4, 4: NS24 as deviations from the average level of expression of each gene across the environments. The most significant functional enrichments for each cluster are: (1) sensory perception of chemical stimulus (bp,  $P < 0.003$ ); (2) nucleolus (cc,  $P < 0.001$ ); (3) arginine and proline metabolism (path,  $P = 0.001$ ); (4) response to copper ion (bp,  $P = 0.001$ ); (5) signal transduction (bp,  $P = 0.01$ ); (6) cell wall (sensu Fungi) (cc,  $P = 0.0002$ ); (7) small nucleolar RNAs binding (SnoRNA, ribosome biogenesis) (mf,  $P = 2E-5$ ); (8) protein-nucleus import (bp,  $P = 0.0004$ ); (9) mitochondrial electron transport, succinate to ubiquinone (bp,  $P = 0.0005$ ); (10) GINS complex (DNA replication) (cc,  $P < 0.002$ ); (11) ribosome (path,  $P = 2E-73$ ); (12) spliceosome assembly (bp,  $P = 0.001$ ). In the above listing, the abbreviations are bp, biological process; cc, cellular component; mf, molecular function; path, KEGG metabolic pathway.

### 3.2. Genetic variation for transcriptional plasticity

Although GEI has been reported previously (Fay et al., 2004), here we present the first genome-wide study aiming specifically at studying genetic variation for transcriptional plasticity. We aimed at determining what features of these genes that show GEI or their evolution might favor their diversity of transcriptional responses. In other words, what factors are likely to contribute to the diversity of transcriptional response to environmental perturbations in *S. cerevisiae*? To address this question, we carried out comparisons among genes by comparing the set of genes with significant GEI to the rest of the genes studied.

Genes with significant GEI cover a variety of biological processes (Table S1). For instance, fourteen genes are involved in protein biosynthesis (e.g., *Rpl4b*, *Tpd3* and *Rpl24a*), seven in cell wall organization and biogenesis (e.g., *Fmp45*, *Exg2* and *Dse1*), six in response to stress (e.g., *Gre2*, *Tps1* and *Hsp82*) and five in aerobic respiration (e.g., *Ysc84*, *Cox4* and *Mbr1*). About a third of the genes showing GEI (76 genes) are involved in unknown biological processes.

We tested whether genes that showed significant GEI were enriched for genes involved in particular biological processes, molecular function, cellular components or metabolic pathways (GeneMerge, Castillo-Davis and Hartl, 2003). Only the pentose phosphate pathway (metabolic pathway) showed significant statistical overrepresentation after correcting for multiple tests ( $P=0.026$ ). Among the 5278 genes studied, twenty-five are in this pathway and six show significant GEI (*YGR034C*, *Tkl2*, *Sol4*, *Gnd2*, *Pgm2* and *Sol1*). We conclude from this that no particular biological process or metabolic pathway can explain entirely the patterns observed, and that other properties of those genes might explain their diversity of transcriptional response to environmental perturbations.

GEI can come about by two not exclusive ways. First, there might be a change in scale, whereby genotypes react more (or less) strongly in one environment than in the other. Second, there might be a change in ranking of the genotypes across environments (Lynch and Walsh, 1998). The two types of GEI have different impacts on the evolutionary trajectories of plastic characters. In the first case, evolution by natural selection on a character would proceed at a different pace in the two environments but would favor the same genotypes. However, in the second case, different genotypes would be favored in each environment. Thus, in a randomly fluctuating environment, this would result in balancing selection. In order to examine whether the crossing of norms of reactions contributed to GEI, we examined the two conditions at the extreme of the gradient, rich YPD and NS24. For each gene, we considered the 15 possible pair-wise comparisons between two genotypes and counted the number of those pairs that had crossing norms of reaction, i.e., whose expression values changed in rank between environments. We then compared whether GEI genes were more likely to have crossing norms of reaction than genes with significant environmental effects but no significant GEI or G effects. In these cases, the number of crossing norms of reaction would represent how much is expected due to the level of experimental

noise. On average, 8.8 pairs of genotypes had crossing norms of reactions for the GEI genes, which is significantly higher than the proportion observed for the E genes (7.8, Wilcoxon rank sum test,  $P<4E-7$ ), which suggests that change in the ranks of expression level between these two conditions contributes to GEI for a significant fraction of the genes.

Gene products do not act alone. The cellular context in which a gene product is required may contribute a large fraction to the constraints acting on it. Biological processes such as cellular metabolism and the cell cycle represent complex systems of interacting components organized as modules arranged in hierarchies (e.g., Ravasz et al., 2002). The function of a gene product, its position in one or more protein or metabolic modules, and their importance in the network of interactions will contribute to shaping the phenotypic consequences of variation in gene expression. Genes that dominate hierarchies or that represent hubs in the protein–protein interaction or metabolic network might therefore be expected to have greater effects on fitness than other genes, owing to their pleiotropic effects (e.g., Lemos et al., 2004). Such genes are expected to experience stronger constraint in their response to environmental perturbations, which would lead to stronger stabilizing selection and canalization of their norms of reaction and therefore limit the amount of genetic variation for transcriptional plasticity we observe.

We first hypothesized that genes not required for essential functions of the organism would have less constrained norms of reaction than essential genes. Lethality of deletion phenotypes can be used as a surrogate for gene essentiality. Under rich conditions, ~80% of the genes in *S. cerevisiae* are dispensable, i.e., they do not lead to lethal phenotypes upon deletion (Winzeler et al., 1999). Using the lethality of the deletion as a surrogate for gene essentiality, we tested whether genes showing significant GEI were enriched for genes that are nonessential. Among the 5258 genes studied here, 3997 are dispensable genes (76%) (excluding genes related to transposable elements). Dispensable genes, however, represent 190 of the 221 genes with significant GEI (86%), which is a much higher proportion than expected by chance (hypergeometric distribution,  $P=9E-5$ ). Genes that are dispensable to cell survival are therefore over-represented in the set of genes with genetic variation for transcriptional plasticity. On the other hand, genes with a significant response to the environment, but without GEI, show the expected proportion of dispensable genes (1390/1862, 75%); a similar result was observed for genes that show genetic variation in gene expression independently of the growth conditions (89/118, 75%). The same result regarding genes with significant GEI was obtained when comparing the growth effects of gene knockouts (Steinmetz et al., 2002): the relative fitness of knockout mutants for genes with significant GEI is on average significantly higher than genes with no significant GEI (1.01 and 0.99, respectively, Wilcoxon rank sum test  $P<0.0005$ ).

Gene essentiality is a proxy for the level of constraint acting on a gene and might reflect the aggregate effects of different constraints. Gene dispensability has, for instance, been negatively associated with the number of functions a gene is

Table 1  
Number of genes with paralogs in the genome for each phenotypic class

Significant effect	Number of genes	Number of genes with paralogs	Number expected
Genotype	118	44	35
Environment	1862	551	547
GEI	221	101	65
No effect	3099	862	910

The total number of tabulated genes is 5300 rather than 5258 because genes with both a significant G effect and a significant E effect were included in both classes and counted twice; also, genes with functions associated with transposable elements were not included.

associated with (Yu et al., 2004), the rate of protein evolution (Hirsh and Fraser, 2001; Yang et al., 2003), and the connectivity and centrality of a protein in the protein–protein interaction network (Hahn and Kern, 2005). We examined whether any of those factors could contribute to explaining why genes with GEI are enriched for nonessential genes (Supplementary material, Appendix 2). We examined whether genes with significant GEI were expressed at a lower level than the rest of the genes examined, and found that genes with significant GEI are not expressed at a lower level than other genes in the genome (Supplementary material, Appendix 2). We also tested whether genes with significant GEI were coding for proteins involved in more protein–protein interactions or metabolic pathways and found that GEI genes do not code for proteins involved in fewer protein–protein interactions on average nor involved in fewer metabolic pathways (Supplementary material, Appendix 2). Finally, using the two closely related species *S. paradoxus* and *S. bayanus*, we tested whether genes showing significant GEI were evolving faster than other genes in the genome and found no significant difference (Supplementary material, Appendix 2).

Although genes with significant GEI tend to be nonessential to *S. cerevisiae*, they are not under weaker selective constraints at the protein level. How can these seemingly contradictory results be reconciled? Gene dispensability in yeast in part reflects the capacity of the cell to compensate for their absence. For instance, it has been proposed that the products of dispensable genes act in metabolic pathways whose metabolites can be rerouted through the metabolic network (metabolic buffering) (Papp et al., 2004). It has also been proposed that dispensable genes have paralogs that can compensate for their deletion (genetic buffering) (Gu et al., 2003; Kafri et al., 2005). Our observation that dispensable genes show more genetic variation for transcriptional plasticity is compatible with both these hypotheses. The buffering of gene deletions could be an indication that dispensable genes are members of robust regulatory networks that can deal with environmental perturbations and thus allow the accumulation of environment-specific regulatory mutations. Mutations with environment-specific effects on gene expression would contribute to GEI in an analysis like the one presented here. Similarly, the compensation of gene-expression level by genes with redundant function (i.e., paralogs) could produce similar results. There is some evidence in yeast that genes with redundant functions alter their own expression under certain conditions to compensate for mutations in other genes with similar functions (Kafri et al., 2005). Such a

mechanism would allow the accumulation of environment-dependent genetic variation in genes with redundant functions. The prediction is therefore that genes with paralogs in the genome will be more likely to show GEI.

Our data provide strong evidence that genes with genetic variation for transcriptional plasticity indeed tend to have paralogs in the genome. Among the 5248 genes studied, 1544 have at least one paralogous copy in the genome. Of the 221 genes with significant GEI, 101 have at least one paralog, which is much higher than expected by chance (hypergeometric distribution,  $P < 1E-6$ ). The number of genes with paralogs for each phenotypic class is listed in Table 1. The number of genes with paralogs is not randomly distributed among these classes ( $\chi^2 = 25.17$ , d.f. = 3,  $P < 2E-5$ ). Genes that show significant G but no GEI effect are also slightly enriched for genes with paralogs, and the proportions of paralogs observed in the two classes are not significantly different ( $\chi^2 = 1.27$ ,  $df = 1$ ,  $P = 0.25$ ). However, over-representation of paralogs alone is apparently not sufficient to explain why so many genes with significant GEI are dispensable. Indeed, 101 of the 120 genes identified as having no clear paralogs are dispensable, which still represents a proportion greater than that expected by chance alone (hypergeometric distribution,  $P < 0.05$ ). Other characteristics of dispensable genes must also contribute to the diversity in their transcriptional response to environmental perturbation. The specific reasons why genes that show genetic variation in their transcriptional response to environmental perturbations tend to be nonessential and tend to have paralogs in the genome will require further investigation of the specific genes themselves. It is important to note, for instance, that gene essentiality has been assessed in rich conditions only, and that the fraction of essential genes has been proposed to be much higher under more stringent conditions (Papp et al., 2004).

### 3.3. Conclusions

Here we provide the first genome-wide study examining genetic variation for transcriptional plasticity to environmental perturbation. We find significant genetic variation for transcriptional plasticity in more than 200 genes among the six strains studied, and this set of genes is statistically biased for non-essential genes and genes with duplicates in the genome. Among the genes showing any significant difference in expression across genotypes (G or GEI or both), about two-thirds show significant GEI. Previous studies looking at GEI in gene expression also support the observation that GEI is an important factor in gene expression variation. For instance, among ten genes studied, Correa and Cheung (2004) found that two genes showed genetic variation in their transcriptional response to radiation in human cell lines. Also, Fay et al. (2004) found 56 genes with significant GEI in their expression in response to copper sulfate in *S. cerevisiae*. Genotype-by-age interaction effect on gene expression in *Drosophila* has also been investigated, but no significant effects were found (Jin et al., 2001).

The presence of GEI in gene expression has important consequences for the study of the genetic architecture of the

gene-expression network. Genetic architecture refers to the nature, number and relative contributions of the genetic factors underlying a quantitative trait. Phenotypic plasticity and genotype-by-environment interaction have important effects on the genetic architecture of quantitative traits because they can change the ranking of the genotypic values across environments and also affect the genetic variance of the traits (Figs. 1 and 2; Stearns et al., 1991). Extensive studies of the genetic architecture of gene expression in rich medium have revealed numerous epistatic interactions (Brem and Kruglyak, 2005). Our results suggest that the nature and the amplitude of these effects will differ when assessed in alternative growth conditions. Interestingly, recent large scale experiments on the identification of transcription factor interactions on the promoters of *S. cerevisiae* genes revealed that the occupancy of the binding sites is condition-dependent (Harbison et al., 2004). This condition dependence of the regulatory code should result in many regulatory mutations showing condition-dependent effects, which could be manifested as GEI in an experiment like the one reported here. This is a specific and testable hypothesis for the mechanisms by which GEI is produced in the case of gene-expression regulation. Experiments analogous to ours will be necessary to determine whether these principles generalize to transcriptional genotype-by-environment interaction in other organisms. Ultimately one would wish to understand the genetic basis of genetic variation for phenotypic plasticity, but the genetic architecture of transcriptional plasticity is likely to be at least as complex as that of other multifactorial traits, and perhaps more so owing to abundant epistasis.

## Acknowledgments

We would like to thank Nadia Aubin-Horth and Cristian Castillo-Davis for comments on an earlier version of the manuscript and members of the Hartl lab for discussions on the work. We would also like thank Ping Ma for helpful discussions about statistical issues and Hao Wu for information about the use of MAANOVA, Keith Morneau for lab assistance and Mario Polsinelli for providing the strains for this work. CRL is a Frank Knox Memorial fellow at Harvard University and is supported by graduate fellowships from the Fondation Desjardins (Québec, Canada), NATEQ (Fonds de la Recherche sur la Nature et les Technologies, Québec, Canada) and CRSNG (Conseil National de Recherche en Sciences et en Génie du Canada). This work was supported by the 2010 program, Bauer Center for Genomics Research (Harvard U.) to DC, and by NIH grants to DLH.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.gene.2005.10.042.

## References

Abouheif, E., Wray, G.A., 2002. Evolution of the gene network underlying wing polyphenism in ants. *Science* 297, 249–252.

- Altschul, S.F., et al., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402.
- Bisson, L.F., 1991. Influence of nitrogen on yeast fermentation of grapes. In: *Am. Soc. Enology and Viticulture (Ed.), Proc Int Symp Nitrogen in Grapes and Wine* Davis, CA.
- Brem, R.B., Kruglyak, L., 2005. The landscape of genetic complexity across 5700 gene expression traits in yeast. *Proc. Natl. Acad. Sci. U. S. A.* 102, 1572–1577.
- Castillo-Davis, C.I., Hartl, D.L., 2003. GeneMerge-post-genomic analysis, data mining, and hypothesis testing. *Bioinformatics* 19, 891–892.
- Causton, H.C., et al., 2001. Remodeling of yeast genome expression in response to environmental changes. *Mol. Biol. Cell* 12, 323–337.
- Cavaliere, D., Townsend, J.P., Hartl, D.L., 2000. Manifold anomalies in gene expression in a vineyard isolate of *Saccharomyces cerevisiae* revealed by DNA microarray analysis. *Proc. Natl. Acad. Sci. U. S. A.* 97, 12369–12374.
- Correa, C.R., Cheung, V.G., 2004. Genetic variation in radiation-induced expression phenotypes. *Am. J. Hum. Genet.* 75, 885–890.
- Cui, X., Churchill, G.A., 2003. Statistical tests for differential expression in cDNA microarray experiments. *Genome Biol.* 4, 210.
- Cui, X., Hwang, J.T., Qiu, J., Blades, N.J., Churchill, G.A., 2005. Improved statistical tests for differential gene expression by shrinking variance components estimates. *Biostatistics* 6, 59–75.
- Davis, J.C., Petrov, D.A., 2004. Preferential duplication of conserved proteins in eukaryotic genomes. *PLoS Biol.* 2, E55.
- Fay, J.C., McCullough, H.L., Sniegowski, P.D., Eisen, M.B., 2004. Population genetic variation in gene expression is associated with phenotypic variation in *Saccharomyces cerevisiae*. *Genome Biol.* 5, R26.
- Gasch, A.P., Werner-Washburne, M., 2002. The genomics of yeast responses to environmental stress and starvation. *Funct. Integr. Genomics* 2, 181–192.
- Gasch, A.P., et al., 2000. Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell* 11, 4241–4257.
- Gibson, G., 2002. Microarrays in ecology and evolution: a preview. *Mol. Ecol.* 11, 17–24.
- Jimeno, C.J., Ljungdahl, P.O., Styles, C.A., Fink, G.R., 1992. Unipolar cell divisions in the yeast *S. cerevisiae* lead to filamentous growth: regulation by starvation and RAS. *Cell* 68, 1077–1090.
- Gu, Z., Steinmetz, L.M., Gu, X., Scharfe, C., Davis, R.W., Li, W.H., 2003. Role of duplicate genes in genetic robustness against null mutations. *Nature* 421, 63–66.
- Hahn, M.W., Kern, A.D., 2005. Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. *Mol. Biol. Evol.* 22, 803–806.
- Harbison, C.T., et al., 2004. Transcriptional regulatory code of a eukaryotic genome. *Nature* 431, 99–104.
- Hartman, J.L.t., Garvik, B., Hartwell, L., 2001. Principles for the buffering of genetic variation. *Science* 291, 1001–1004.
- Heiman, M.G., Walter, P., 2000. Prm1p, a pheromone-regulated multispanning membrane protein, facilitates plasma membrane fusion during yeast mating. *J. Cell Biol.* 151, 719–730.
- Hirsh, A.E., Fraser, H.B., 2001. Protein dispensability and rate of evolution. *Nature* 411, 1046–1049.
- Ihaka, R., Gentleman, R., 1996. R: A language for data analysis and graphics. *J. Comput. Graph. Stat.* 5, 299–314.
- Jin, W., Riley, R.M., Wolfinger, R.D., White, K.P., Passador-Gurgel, G., Gibson, G., 2001. The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*. *Nat. Genet.* 29, 389–395.
- Kafri, R., Bar-Even, A., Pilpel, Y., 2005. Transcription control reprogramming in genetic backup circuits. *Nat. Genet.* 37, 295–299.
- Kerr, M.K., Churchill, G.A., 2001. Bootstrapping cluster analysis: assessing the reliability of conclusions from microarray experiments. *Proc. Natl. Acad. Sci. U. S. A.* 98, 8961–8965.
- Lemos, B., Meiklejohn, C.D., Hartl, D.L., 2004. Regulatory evolution across the protein interaction network. *Nat. Genet.* 36, 1059–1060.
- Levins, R., 2004. *Toward a Population Biology*, Still. Cambridge University Press.
- Lynch, M., Walsh, B., 1998. *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Sunderland, MA.

- Meiklejohn, C.D., Hartl, D.L., 2002. A single mode of canalization. *Trends Ecol. Evol.* 17, 468–473.
- Mortimer, R.K., Johnston, J.R., 1986. Genealogy of principal strains of the yeast genetic stock center. *Genetics* 113, 35–43.
- Papp, B., Pal, C., Hurst, L.D., 2004. Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. *Nature* 429, 661–664.
- Pigliucci, M., 1996. How organisms respond to environmental changes: from phenotypes to molecules (and vice versa). *Trends Ecol. Evol.* 11, 168–173.
- Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N., Barabasi, A.L., 2002. Hierarchical organization of modularity in metabolic networks. *Science* 297, 1551–1555.
- R Development Core Team R, 2005. A language and environment for statistical computing. R Foundation for statistical computing, Vienna, Austria. ISBN: 3-900051-07-0. <http://www.R-project.org>.
- Schlichting, C.D., Smith, H., 2002. Phenotypic plasticity: linking molecular mechanisms with evolutionary outcomes. *Evol. Ecol.* 16, 189–211.
- Singh, R.S., 2003. Darwin to DNA, molecules to morphology: the end of classical population genetics and the road ahead. *Genome* 46, 938–942.
- Smyth, G.K., 2004. Linear models and empirical bayes for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* 3 Article 1.
- Stearns, S.C., 1992. *The evolution of life histories*. Oxford University Press, Oxford.
- Stearns, S.C., de Jong, G., Newman, R., 1991. The effects of phenotypic plasticity on genetic correlations. *Trends Ecol. Evol.* 6, 122–126.
- Steinmetz, L.M., et al., 2002. Systematic screen for human disease genes in yeast. *Nat. Genet.* 31, 400–404.
- Suiter, A.M., Banziger, O., Dean, A.M., 2003. Fitness consequences of a regulatory polymorphism in a seasonal environment. *Proc. Natl. Acad. Sci. U. S. A.* 100, 12782–12786.
- Townsend, J.P., Cavalieri, D., Hartl, D.L., 2003. Population genetic variation in genome-wide gene expression. *Mol. Biol. Evol.* 20, 955–963.
- Waddington, C.H., 1942. Canalization of development and the inheritance of acquired characters. *Nature* 150, 563–565.
- Winzeler, E.A., et al., 1999. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* 285, 901–906.
- Yang, J., Gu, Z., Li, W.H., 2003. Rate of protein evolution versus fitness effect of gene deletion. *Mol. Biol. Evol.* 20, 772–774.
- Yeung, K.Y., Haynor, D.R., Ruzzo, W.L., 2001. Validating clustering for gene expression data. *Bioinformatics* 17, 309–318.
- Yu, H., Greenbaum, D., Xin Lu, H., Zhu, X., Gerstein, M., 2004. Genomic analysis of essentiality within protein networks. *Trends Genet.* 20, 227–231.