# Metabolic Adaptation after Whole Genome Duplication

*Milan J. A. van Hoek*[1] *and Paulien Hogeweg*
Theoretical Biology/Bioinformatics Group, Utrecht University, Utrecht, The Netherlands

Whole genome duplications (WGDs) have been hypothesized to be responsible for major transitions in evolution. However, the effects of WGD and subsequent gene loss on cellular behavior and metabolism are still poorly understood. Here we develop a genome scale evolutionary model to study the dynamics of gene loss and metabolic adaptation after WGD. Using the metabolic network of *Saccharomyces cerevisiae* as an example, we primarily study the outcome of WGD on yeast as it currently is. However, similar results were obtained using a recontructed hypothetical metabolic network of the pre-WGD ancestor. We show that the retention of genes in duplicate in the model, corresponds nicely with those retained in duplicate after the ancestral WGD in *S. cerevisiae*. Also, we observe that transporter and glycolytic genes have a higher probability to be retained in duplicate after WGD and subsequent gene loss, both in the model as in *S. cerevisiae*, which leads to an increase in glycolytic flux after WGD. Furthermore, the model shows that WGD leads to better adaptation than small-scale duplications, in environments for which duplication of a whole pathway instead of single reactions is needed to increase fitness. This is indeed the case for adaptation to high glucose levels. Thus, our model confirms the hypothesis that WGD has been important in the adaptation of yeast to the new, glucose-rich environment that arose after the appearance of angiosperms. Moreover, the model shows that WGD is almost always detrimental on the short term in environments to which the lineage is preadapted, but can have immediate fitness benefits in "new" environments. This explains why WGD, while pivotal in the evolution of many lineages and an apparent "easy" genetic operator, occurs relatively rarely.

## Introduction

The occurrence of whole genome duplications (WGDs) was already proposed by Ohno (1970). Nowadays, there is conclusive evidence for WGDs in several lineages, such as yeast (Wolfe and Shields 1997; Kellis et al. 2004), plants (Arabidopsis Genome Initiative 2000; Bowers et al. 2003), and teleost fishes (Amores et al. 1998; Taylor et al. 2001; Semon and Wolfe 2007). It has been speculated that these WGDs caused major transitions in evolution (Huerta-Cepas et al. 2007), but as yet it is unknown what the precise effects of a WGD are on cellular behavior.

In this paper, we study the effects of WGD on cellular metabolism using an evolutionary systems biology approach, combining genome-scale, quantitative metabolic modeling with individual-based evolutionary simulations. Besides that this gives us more insight into the evolution of metabolic networks after WGD, it shows that detailed, quantitative modeling can be easily combined with evolutionary simulations, yielding new insights and predictive power. To our knowledge, there have been very few attempts to combine evolutionary simulations with detailed, quantitative modeling (exceptions are Gatenby and Vincent 2003, Pal et al. 2006, and Van Hoek and Hogeweg 2006, 2007). Such an evolutionary systems biology approach allows us to go beyond a posteriori plausible explanations of observed evolutionary outcomes, by exploring which evolutionary outcomes are likely, or even necessary.

One would expect that cellular behavior, and specifically its metabolism, will change after WGD. After WGD, it is common that most duplicate genes are lost and only a minority of the genes is retained in duplicate (e.g., approximately 500 gene pairs in *Saccharomyces cerevisiae*). In this way, if we assume a gene dosage effect, the relative abundance of gene products will alter and therefore cellular metabolism can change extensively. Because in the yeast lineage a WGD occurred and metabolism of *S. cerevisiae* is very well studied, we use *S. cerevisiae* as a model system to study the possible effects of WGD and subsequent gene loss on cellular metabolism.

We use a fully compartmentalized, genome-scale metabolic model (Duarte et al. 2004a), which is publicly available. Because such a model is not available for the pre-WGD ancestor of yeast, we study the effect that a WGD would have if it would occur now in *S. cerevisiae*. In this way, we hope to get a better understanding of how WGD can change a metabolic network.

The fluxes through such a metabolic network can be calculated using flux balance analysis (FBA). FBA is a constraint-based modeling approach which, instead of modeling the dynamics of a metabolic network, focuses on determining the steady-state fluxes. Using stoichiometric (and possibly other) constraints, FBA tries to find a flux distribution that optimizes for a certain objective function, for example, the growth rate of the cell. Optimizing for the growth rate of a cell has successfully been used to reproduce the growth rates for gene deletion studies (Edwards and Palsson 2000; Famili et al. 2003) and by-product secretion of cells (Famili et al. 2003; Duarte et al. 2004b).

Here we develop an individual-based, evolutionary model to describe metabolic network evolution after WGD and subsequent gene loss. The fitness of a metabolic network is determined using FBA. This model assumes that cell size is correlated with genome size and that the maximal flux through the reactions depends on gene dosage. In this way, the flux constraints of the network, and therewith cellular metabolism and fitness, change during evolution.

Indeed, it has been observed that haploid yeast cells are smaller than diploid cells, which are again smaller than tetraploid cells (Hennaut et al. 1970; Galitski et al.

[1]Present address: NCSB-NISB, Science Park 904, 1098 XH & CWI, Science Park 123, 1098 XG, Amsterdam, The Netherlands.

Key words: evolutionary systems biology, whole genome duplication, metabolic network, flux balance analysis, *Saccharomyces cerevisiae*.

E-mail: milan.van.hoek@cwi.nl.

*Mol. Biol. Evol.* 26(11):2441–2453. 2009
doi:10.1093/molbev/msp160
Advance Access publication July 22, 2009

1999). Furthermore, even between organisms, a correlation between cell size and genome size is observed (Cavalier-Smith 1978). Therefore, we assume that after a WGD, the cell size increases and gradually decreases when genes are lost. This changes the surface area to volume ratio, which is an important factor determining the metabolic capacities of a cell (see, e.g., Kooijman et al. 1991).

It has been shown previously that gene retention of many genes after WGD is for gene dosage reasons (Aury et al. 2006) (in *Paramecium tetraurelia*). It must also be noted that genetic regulation of most genes in yeast is not affected by polyploidy. Only 17 yeast genes significantly change their gene expression after polyploidy (Galitski et al. 1999). Therefore, for almost all genes, it is reasonable to assume that gene expression increases together with gene dosage.

Recently it has been proposed that the WGD that occurred in yeast led to an increase in glycolytic flux (Conant and Wolfe 2007). Using a kinetic model of glycolysis and assuming a dosage effect for duplicated genes, it was shown that these duplicated genes could indeed enhance glycolytic flux. It was also shown that increased glycolytic flux could generate a fitness advantage in a glucose-rich environment. Using a comparative genomics approach, it has also been established that in general yeast species that underwent WGD are Crabtree positive, which means that they produce ethanol in the presence of oxygen, whereas most species that did not undergo WGD are Crabtree negative (Merico et al. 2007).

These studies indicate that the WGD caused yeast to be able to rapidly consume glucose. However, whether this is the "expected" outcome of WGD and subsequent gene loss in yeast is a completely different matter. WGD may open many different evolutionary possibilities, from which increase in glycolytic flux is only one. In contrast, here we take a different approach. We study the effect of a WGD on the whole metabolic network, using evolutionary modeling. By performing evolutionary simulations, we study the WGD in yeast and the subsequent loss of genes. Instead of a posteriori interpretation of the evolutionary outcome, we study which evolutionary outcome(s) is/are to be expected.

The goal of this research is 2-fold. First, we are interested in the effects of a WGD and subsequent gene loss on cellular metabolism. Second, we want to know under which circumstances a WGD is adaptive. An important question then is whether a WGD opens up new possibilities for evolution that are not available with small scale duplications (SSDs). This we study by performing evolutionary simulations with single gene duplications.

We find that our model can describe the essential features of the WGD in *S. cerevisiae*. The WGD is followed by massive gene loss, during which the relative abundance of genes changes. The model satisfactorily predicts which genes are retained in duplicate after the WGD that occurred in the yeast lineage. Furthermore, we find that the WGD leads to an increase in glycolytic flux, as has been proposed previously (Conant and Wolfe 2007).

We also find that if cells are perfectly adapted to the environment, WGD leads to a fitness decrease, whereas it can increase fitness in new environments. Furthermore, if duplication of a whole pathway is needed to increase fitness, WGD is favored over SSDs. It appears that the ability to grow anaerobically evolved just before the WGD (Moller et al. 2001; Gojkovic et al. 2004). Therefore, there was a new anaerobic environment available for the yeast lineage to adapt on. All these findings are in line with the idea that the WGD in yeast helped to adapt to the newly arisen environment of glucose-rich fruits.

Our results indicate that network evolution after WGD is surprisingly consistent: Different evolutionary simulations give rise to very similar metabolic networks. Furthermore, these *in silico* evolved networks seem to have changed in a similar way as the network of *S. cerevisiae* after WGD. This is surprising, as WGD opens many different evolutionary paths, due to the increase in genetic redundancy. Here we show that even in a changing environment, potentially allowing for different ways of adaptation, consistently the same evolutionary path is chosen.

## Materials and Methods

To study the evolution of yeast after a WGD, we used a previously developed genome-scale model of yeast metabolism (Duarte et al. 2004a). This model consists of 1061 metabolites and 1266 reactions. Therefore, the model can be described by a $1266 \times 1061$ stoichiometric matrix.

All reactions are linked with genes of *S. cerevisiae*. Some reactions have no corresponding genes, other reactions have more than one corresponding gene. For simplicity, it has been assumed that genes work together in a Boolean way. For example, reaction X is performed by GENE A OR GENE B, reaction Y is performed by (GENE C AND GENE D) OR (GENE C AND GENE E), etc.

We now use this model as the basis for an evolutionary model that can describe the effect of a WGD and subsequent gene loss on the metabolic network. For this, we use a flux balance approach. FBA maximizes the growth rate of the cell, given the stoichiometric constraints and possible constraints on the flux values of each reaction. The optimum that is found with FBA is global, though not necessarily unique.

Our approach to model the WGD is to change the constraints on the network and in this way the growth rate, rather than changing the network itself. During evolution, some genes will be kept in duplicate, some as singleton, and some will be entirely deleted. We assume that the maximal flux value of a certain reaction is determined by the number of genes present to catalyze that particular reaction. Therefore, we assume dosage dependence, such that the maximal flux through each reaction depends on the copy number of the associated genes. When genes are deleted during evolution, these maximal fluxes change and in this way the constraints on the cells metabolism change and cells can adapt to the environment.

We used the Matlab COBRA Toolbox (Becker et al. 2007) to perform FBA and all evolutionary simulations were performed using Matlab.

### A Model to Determine Changes in Flux Constraints

Below we describe the model we used to study metabolic network evolution after WGD and its assumptions.

For a more detailed description of the model, we refer to the Supplementary Material online.

To model the evolution of *S. cerevisiae* after a WGD, we need a formalism to describe the changes in flux constraints during evolution. Here we develop a model that describes the changes in flux constraints of each reaction, dependent on the copy number of all genes, the cell volume and surface area. We assume that gene duplications only change gene dosage, so we assume that duplicate genes retain their function.

To check our assumption of dosage dependence, we studied whether yeast genes that are retained in duplicate after the WGD were associated with higher fluxes. Indeed, we found that duplicated genes had significantly higher fluxes, in particular in an anaerobic, glucose-rich environment. This confirms the results of Papp et al. (2004), who already showed that the average *in silico* flux for duplicated enzymes in yeast is higher than of single-copy enzymes. Furthermore, Kuepfer et al. (2005) showed that gene dosage explains the retention of some duplicate genes in yeast because they were associated with reactions with high (experimentally measured) carbon fluxes. Finally, genes that are retained in duplicate after WGD have on average higher expression rates (Seoighe and Wolfe 1999). All these results indicate that selection for gene dosage is an important reason for gene retention of duplicated genes.

We assume that, after a WGD, the cell size changes. This appears to be a natural assumption, given the good correlation between DNA content and cell size that is ubiquitous in nature (Cavalier-Smith 1978; Gregory 2001; Cavalier-Smith 2005). Furthermore, it has been observed that polyploid yeast cells are larger than haploid yeast cells. Hennaut et al. (1970) observed that a diploid cell is 1.87 as large as a haploid cell. The volume ratio between tetraploid and diploid cells was very similar (1.91). Furthermore, the surface area increased with a factor 1.56. Using these data, we now assume for our model that $V \sim N^{0.9}$ and $A \sim V^{0.7}$, where $V$ is the volume of a cell; $N$, the number of genes; and $A$, the surface area. Accordingly, after a WGD the volume increases by a factor 1.87. We also assume that cells cannot become smaller than 1, that is, not smaller than before WGD.

We distinguish two type of reactions, exchange reactions and intracellular reactions. First we explain how we model the changes in constraints of intracellular reactions. For intracellular reactions that are not associated with a gene, we assume that the maximum flux (measured in mmol/(gram dry weight hour)) remains constant. For intracellular reactions that are associated with a gene, we assume that the maximum flux scales with gene dosage divided by volume.

$$F_{\max}(i) = F_{\max,0}(i)\frac{\gamma(i)}{\beta} \tag{1}$$

where $\gamma(i)$ refers to the relative change in gene dosage of reaction $i$ and $\beta$ to the relative change in cell volume.

For exchange reactions, calculating how the maximal fluxes change is less straightforward than for intracellular reactions. It is often assumed that the surface area to volume ratio is a crucial factor that determines the uptake of nutrients. Indeed, this factor can be important, but another factor is the amount of transporter proteins that are available. It has been shown previously that in yeast both the surface area to volume ratio and the amount of transporter proteins can be the rate-limiting factor for the amount of nutrient uptake (Hennaut et al. 1970). It was shown that some uptake reactions are not (or hardly) different between haploid and diploid yeast cells, whereas other reactions (almost) change by the surface area to volume ratio between haploid and diploid cells. The authors conclude that some uptake reactions are limited by the surface area of the cell and others by the amount of transporter protein. In the Supplementary Material (online) we derive a general formula that describes both these situations, the saturation for surface area and number of proteins. Suppose the surface area of a cell changes with a factor $\alpha$, the volume with a factor $\beta$ and gene dosage with a factor $\gamma$, we then find

$$F_{\max}(i) = F_{\max,0}(i)\frac{\alpha\gamma(i)}{\beta}\frac{1+x(i)}{\gamma(i)x(i)+\alpha} \tag{2}$$

where $x(i)$ is a parameter indicating whether exchange reaction $i$ is limited by the surface area of the cell or by the amount of transporter protein. When $x(i) = 0$, the exchange reaction is only dependent on the number of transporter proteins; when $x(i) = \infty$, the exchange reaction only depends on the surface area to volume ratio. This formula can be compared with equation 1, which describes how intracellular fluxes change after WGD. Indeed, the limit for $x(i)$ goes to 0 gives equation 1. Finally, uptake and excretion fluxes that are not associated with genes (such as oxygen uptake) are modeled using the limit for $x(i)$ to infinity of equation 2, such that these fluxes are only determined by the surface area to volume ratio.

Unfortunately, it is not possible to know the $x(i)$ values for all exchange reactions that are associated with genes in yeast. We choose $x(i) = 1$ for all these reactions, such that both the surface area to volume ratio and the amount of transporter proteins are important in determining these exchange fluxes.

### Initialization

In the previous section we described how the flux constraints change during the course of evolution. However, we also need a way to establish the initial constraints on the fluxes. As there is no experimental data for all reactions, we approach this problem from a different point of view. We assume that the initial maximal flux for each reaction is determined by evolution. Genes that correspond to reactions that need high fluxes, will have evolved high transcription rates and vice versa. Therefore, we performed FBA in 1000 of the environments that are given during the evolutionary simulations with the original yeast model and used the maximal fluxes that we found in these runs as initial constraints.

### Different Environments

Papp et al. (2004) showed that an important reason that so many enzymes are not essential for viability in *S. cerevisiae* is that some enzymes are essential in environments different from laboratory conditions. This was

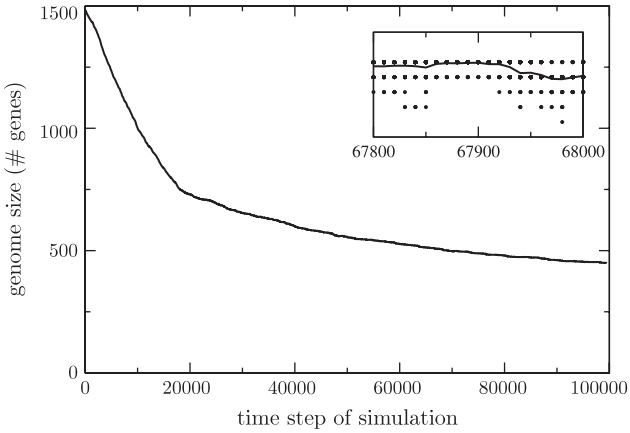


FIG. 1.—Running average of genome size. The inset shows the within population variation and average genome size. Note the selective sweep, eliminating shorter genomes, increasing average genome size temporarily despite the fact that only gene deletions occur.

done using a genome-scale model of yeast (Forster et al. 2003), the precursor of the model we use (Duarte et al. 2004a). They tested enzyme essentiality in 9 different environment types that might have been important for the evolution of yeast. We now use nine types of environment, similar to the ones used by Papp et al. (2004). These environment types were also used in the initialization procedure. We used one rich medium (environment 1, taken from Bilu et al. 2006) and eight minimal media (2, minimal glucose, low $O_2$; 3, minimal glucose, anaerobic; 4, minimal ethanol, low $O_2$; 5, minimal acetate, low $O_2$; 6, minimal glucose, carbon limited; 7, minimal glucose, nitrogen limited, anaerobic; 8, minimal glucose, phosphate limited, anaerobic; 9, minimal glucose, sulfate limited, anaerobic). For each metabolite in the environment we choose an input flux from a uniform distribution between a certain minimum and a maximum flux. In this way, there are nine types of environment, but every time an environment is drawn, the concentrations are different. A precise description of these media is given in the Supplementary Material (online).

### Evolutionary Algorithm

We use a standard evolutionary algorithm to simulate evolution after WGD. Again, details can be found in Supplementary Material (online). In the basic simulations, we start with 100 cells that have a duplicated genome. After this, gene deletions are the only mutational operator. Point mutations are not taken into account for two reasons. One is that the effect of point mutations are not easy to predict. Second, when point mutations would for example change the maximal flux through reactions, all reactions would evolve to very high maximal flux values. Only allowing for gene deletion is a baseline assumption, with which we circumvent this problem. After a mutation, the new constraints on all fluxes are recalculated. We also performed simulations in which we start with the normal, nonduplicated genome and allow for gene duplication and deletion. In these simulations the copy number of a gene cannot exceed 2.

## Results

### Evolution of a metabolic network after WGD

In this section, we study the effects of WGD and subsequent gene loss on the metabolic network of *S. cerevisiae*, using the evolutionary model as described in the Materials and Methods and the Supplementary Material online. As explained there, we evolve a population of 100 cells in a changing environment, after WGD. Because this is a small population size, we expect much evolutionary drift. Therefore, we performed 20 simulations using different random seeds. Then we performed competition experiments, where five individual cells of each of the 20 simulations (yielding 100 cells) were competed against each other. The mutation rate was set to 0 in these simulations. Here we report the findings of the evolutionary simulation yielding the population that most often won the competition experiments. In Van Hoek and Hogeweg (2006), we used a similar method to cope with evolutionary drift.

In our model, a fitness increase can be gained by decreasing the genome size and therewith the cell size. When the cell size decreases due to the deletion of a gene, the maximal fluxes of all other reactions increase. This leads to a selection pressure for a small genome apart from the expected neutral loss of redundant genes after WGD. Indeed massive gene loss has been observed after all WGDs (see, e.g., Wolfe and Shields 1997; Aury et al. 2006).

In figure 1, the genome size of the population is shown. Initially, the model consists of 743 genes (and 7 mitochondrial genes). We observe that after WGD the genome sizes indeed decreases even under 743 and stabilizes around approximately 450 genes. This is because many genes in the model are dispensable, as has already been studied elsewhere (Papp et al. 2004). This indicates that gene loss in our model is sometimes neutral. However, the faster decrease in genome size when there are more than 750 genes is partly because there are still more dispensable genes in the genome and partly because the selection advantage of volume decrease. The population variation over a small timeframe is shown in the inset of the figure. Note that the average genome size can increase, despite the fact that only gene deletions occur, because individuals with larger genome size can replace less fit individuals with smaller genome size.

In figure 2, we show how the fitness of cells changes during the evolutionary simulations in each of the nine different environment types that are explained in the Supplementary Material online, using maximal and intermediate metabolite concentrations. Because we observed that population heterogeneity is small, we only focus on these individual cells. We observe that the growth rate in all environments initially (this means at the WGD) decreases, both for maximal and intermediate concentrations. For intermediate metabolite concentrations however, this decrease is markedly lower. Indeed, it is easier to adapt to lower metabolite concentrations because then the internal fluxes also tend to be smaller. The decrease in fitness is caused by the approximately 2-fold volume increase, decreasing the surface area to volume ratio, which again makes uptake of nutrients more difficult. Therefore, it appears that in these

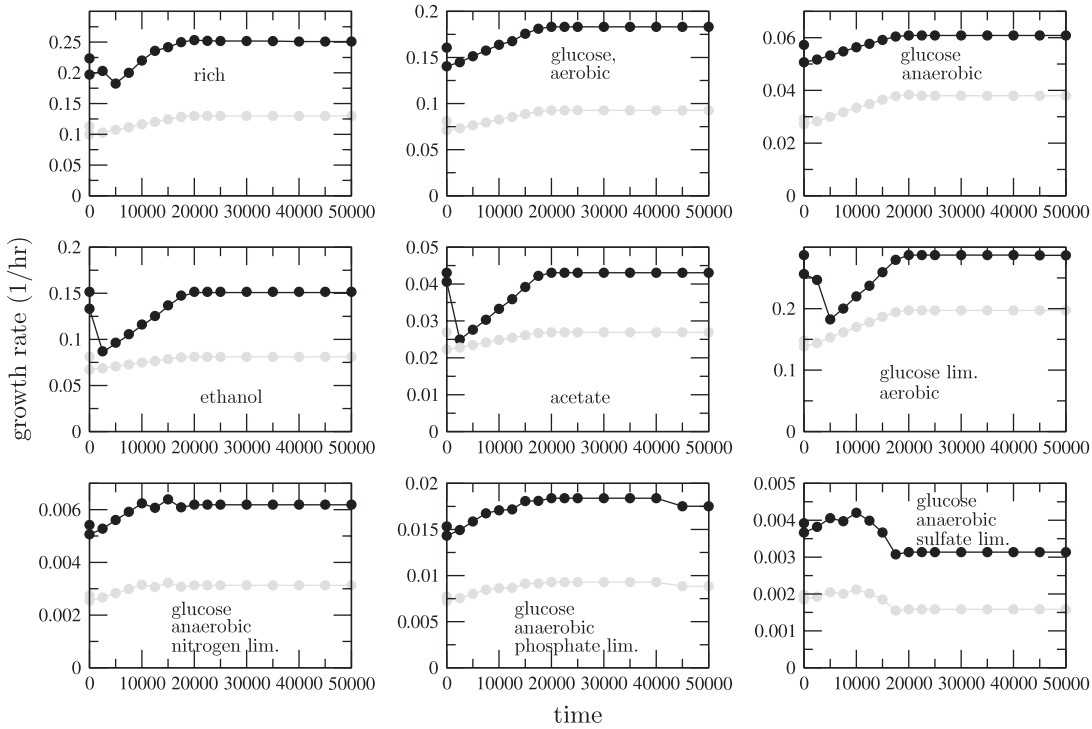


FIG. 2.—Growth rate as a function of time in all nine environment types, with maximum (black) and intermediate (gray) metabolite concentrations. Note that we zoomed in on the first half of the evolutionary simulation, during which the genome shrinks most.

circumstances WGDs are generally unfavorable for a cell (we will come back to this later).

However, during evolution, the growth rate increases in most environment types both for intermediate and maximal concentrations. In some environment types, the growth rate becomes larger than the initial growth rate, whereas the volume of the cells at the end of the simulations equals the volume at $t = 0$. This increase in growth rate is accomplished by retention of certain genes in duplicate, for example, the hexose transporters. In this way, in the long run a WGD can be adaptive. In the the sulfate-limited environment, fitness is lost over time. This is not a general result. In most simulations, fitness is also increased in this environment. Because growth rates in the other eight environments are high in this simulation, the population resulting from this simulation won the competition experiments nevertheless. Note that the initial growth rate in this environment is low compared with other environments, therefore this environment type contributes less to overall fitness.

### *Duplicated Genes*

At the end of the simulations, the genome size of the cells is much reduced. However, some genes are retained in duplicate, on average approximately 50. Some genes are retained in each of the 20 simulations we performed, others only in a few. We determined whether the genes retained in the model correspond to the genes retained in duplicate in yeast after WGD in its ancestor. This comparison is compromised by the fact that we do not have the metabolic network of the pre-WGD ancestor, only of yeast itself.

In figure 3, the fraction of genes that are retained in duplicate after the WGD in yeast is plotted against the number of times that these genes are retained in duplicate in our 20 simulations, ranging from 0 to 20 times. A similar approach of measuring the predictive power of a metabolic model was used in Pal et al. (2006).

We observe that genes that are never retained in duplicate in our model are less often retained in duplicate in *S. cerevisiae*, also lower than on average (the dashed line). These results are significant. The *P* values are given in table 1. The null hypothesis for calculating these *P* values was that every gene has an equal retention probability,

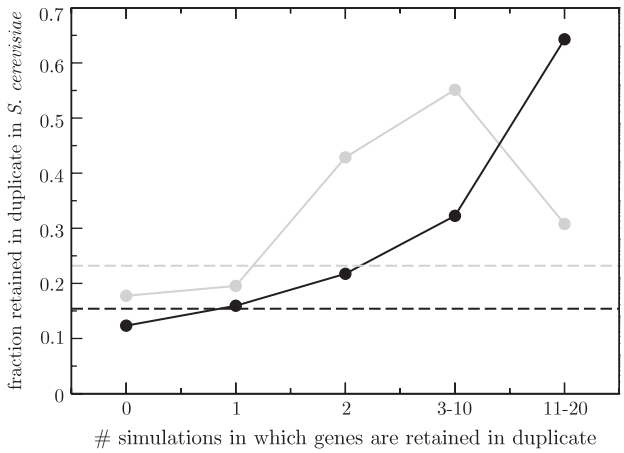


FIG. 3.—Fraction of genes that are retained in duplicate after the WGD in *S. cerevisiae*, for five categories of genes (genes that are 0, 1, 2, 3–10, or 11–20 times retained in the simulations). The average retention fraction is given by the dashed line. The gray lines give the result for the network of *S. cerevisiae*, the black lines for the reconstructed network of the pre-WGD ancestor.

**Table 1**
**Probabilities that Retention Rates are More Extreme than When Calculated with the Null-Hypothesis that Every Gene has an Equal Retention Probability Using a One-Tailed Binomial Test**

| | *Saccharomyces cerevisiae* | | | Ancestor | | |
|---|---|---|---|---|---|---|
| Bin | Total | Duplicated | *P*-Value | Total | Duplicated | *P*-Value |
| 0 | 524 | 93 | 0.0014 | 446 | 55 | 0.039 |
| 1 | 87 | 17 | 0.25 | 69 | 11 | 0.37 |
| 2 | 28 | 12 | 0.0058 | 23 | 5 | 0.13 |
| 3–10 | 78 | 43 | 2.8e-10 | 31 | 10 | 0.0049 |
| 11–20 | 26 | 8 | 0.13 | 14 | 9 | 4.1e−06 |

which is obviously rejected for some bins. The drop in the retention frequency for genes that are in more than 50% of the simulations retained in duplicate is an artifact of using *S. cerevisiae* instead of a pre-WGD ancestor as explained (and tested) below.

More than 80% of the ohnolog pairs (duplicated genes that arose from WGD) in *S. cerevisiae* function in the same reaction in an "OR"-like way. This means that reactions that are retained in duplicate after the WGD in *S. cerevisiae* are now performed by at least two genes. Therefore, if such a reaction would again remain duplicated in our simulations each of the two ohnologs can remain duplicated, whereas the other can be deleted. Therefore, in only 50% of the cases (or less, if more genes code for this reaction) a certain gene that codes for this reaction will be retained in duplicate. This means that genes that are *always* retained in duplicate, either belong to reactions for which all genes belonging to that reaction are retained in duplicate (which happens very seldom) or code for a reaction on their own. If genes code for a reaction on their own, they most most often do not belong to an ohnolog pair. This explains the fact that for larger probabilities to be retained in our model, the probability to belong to an ohnolog pair decreases.

To check this explanation, we reconstructed a hypothetical metabolic network of the pre-WGD ancestor of *S. cerevisiae*, using the Yeast Gene Order Browser (YGOB), version 2.0 (Byrne and Wolfe 2005) as follows. We deleted all genes that are present in *S. cerevisiae* but not in the ancestor (according to YGOB) from the metabolic network. Furthermore, for every ohnolog pair, we deleted one ohnolog (choosing the ohnolog that had the smallest probability to be deleted in the evolutionary simulations). As was to be expected, this metabolic network did not function. Note that this network only did not function in anaerobic environments. It turns out that the ancestor network regains its functionality by adding only two genes of yeast (AUS1 and URA1). We selected these genes because they were retained in all our simulations. Knockout experiments show that both these genes are needed for efficient anaerobic growth, both in our model and in yeast. AUS1 is involved in sterol transport (Wilcox et al. 2002). Because sterol biosynthesis is only possible under aerobic conditions (Andreasen & Stier 1953), sterol transporters are needed for anaerobic growth. Indeed, knockout of AUS1 together with PDR11, another sterol transporter that was also not present in the pre-WGD ancestor, leads to a growth defect under anaerobic conditions (Wilcox et al. 2002). However, PDR11 is not included in the metabolic network and therefore deletion of AUS1 alone leads to a growth defect in the model. The second gene, URA1, is involved in pyrimidine biosynthesis in anaerobic conditions in yeast (Nagy et al. 1992). Strikingly, it has been shown that this gene came from horizontal gene transfer of a bacterial gene around the same time as the WGD in yeast, enabling it to grow anaerobically. Knockout of this gene also leads to a growth defect in anaerobic environments (Gojkovic et al. 2004).

It has been shown however that *Saccharomyces kluyveri*, a yeast species that diverged before the WGD, is also able to grow anaerobically. Indeed, this species also has URA1 (D15268), according to YGOB, although it is not included in the ancestor (in YGOB). Furthermore, *S. kluyveri* also has a homologue of PDR11 (F04312), although it is not recognized as an ohnolog by YGOB because it has lost synteny. This confirms that our reconstruction of the metabolic network of the pre-WGD ancestor is sensible. Moreover, it suggests that the pre-WGD ancestor did have these two genes and therefore that anaerobic growth has evolved before the WGD (see also (Moller et al. 2001)). Therefore, it might be that the ability for anaerobic growth has rendered the WGD selective.

When we perform the simulations with this reconstructed network of the pre-WGD ancestor, we indeed find a better correlation between the probability that a certain gene is retained in duplicate in our simulations and in yeast (fig. 3, black lines). For genes that are very often retained in duplicate in the model, we thus find approximately a 4-fold higher probability to be retained in duplicate in the data than expected by chance (dashed line) and the result is highly significant (table 1). This proves that the model is capable of explaining the process of gene retention after WGD well and therefore that gene dosage is an important factor in this process. Because the outcome of these simulations are qualitatively very much comparable with the simulations using the metabolic network of *S. cerevisiae* (e.g., in terms of what kind of genes are retained in duplicate, namely glycolysis and transporter genes), it follows that the results with the metabolic network of *S. cerevisiae* are not due to any circular logic. Glycolytic genes are not retained in duplicate because *S. cerevisiae* itself is already specialized on high glycolytic fluxes. From now on we again focus on the simulations using the original metabolic network of *S. cerevisiae*.

To find out what kind of reactions are retained in duplicate in the simulations, we looked at reactions that have a higher reaction multiplicity at the end of the simulations than at the beginning, in at least 4 of 20 simulations. There are 43 of such reactions; 20 of these reactions belong to the glycolysis/gluconeogenesis, 19 are extracellular transport reactions, 2 belong to amino acid metabolism, 1 to pyruvate metabolism, and 1 to fatty acid biosynthesis.

Our model therefore establishes that genes that function in transport reactions and glycolysis have a higher probability to be kept in duplicate after WGD because of metabolic optimization in the variable environments. For example, all of the 17 glucose transporters genes that are described in the model are retained in duplicate in the simulations. In *S. cerevisiae*, 6 of 17 glucose transporter

**Table 2**
**Correlation between Transport Reactions and Glycolysis Reactions and WGD**

| | Transport | | | Glycolysis | | |
|---|---|---|---|---|---|---|
| | Yes | No | Percentage | Yes | No | Percentage |
| WGD | 67 | 256 | 21 | 18 | 305 | 5.6 |
| Non-WGD | 138 | 895 | 13 | 25 | 1008 | 2.4 |
| Percentage | 33 | 22 | | 42 | 23 | |
| *P* value | | 0.0012 | | | 0.0045 | |

genes described in the model form ohnolog pairs. This appears to be a contradiction between the model and the data. However, all 11 glucose transporters that do not belong to an ohnolog pair are the result of gene duplications that occurred after the WGD (as can be seen in the YGOB (Byrne and Wolfe, 2005)). Therefore, after WGD, *all* glucose transporters were retained in duplicate, exactly as in our model.

Is there evidence that glycolysis and transport reactions are also generally more often retained in duplicate in *S. cerevisiae*? Each gene belongs to one or more reactions and for every reaction the pathway or function is given in the model. Therefore, each gene belongs to at least one pathway/function. We now counted how often genes are assigned to glycolysis and transport reactions, both for genes that are known to be retained in duplicate after WGD (see Byrne and Wolfe 2005) and genes that are not. Genes that are assigned to different reactions were counted more than once. The result is shown in table 2. We calculated the *P* values in this table using a $\chi^2$ test.

We observe for example that 21% of the ohnologs are functional in a transport reaction, whereas of the other genes this is only 13%. For glycolysis genes, these percentages are 5.6 and 2.4, respectively. The fact that genes belonging to glycolysis reactions are often retained in duplicate has already been noted by Conant and Wolfe (2007).

After the WGD in yeast, many more genes are retained in duplicate and only a minority of these genes have a function in glycolysis or transport (as can also be seen from table 1. These other genes are kept in duplicate for reasons our model does not explain. It could of course be that retention of some of these genes gives a fitness benefit that is, in our model, too small for these genes to be fixed in the population, because of our small population size. However, because we performed 20 different simulations, we believe that this effect will not be large. Finally, many ohnologs are not incorporated in the model (only 750 of over 6000 yeast genes are incorporated in the model), partially because they are not metabolic genes.

*Pathway Usage During Evolution*

During evolution, we expect that pathways are activated and inactivated because genes are deleted or metabolic capacities change. To study how the fluxes through reactions can change during evolution, we followed the fluxes through reactions during the evolutionary simulations. To avoid futile fluxes, we also minimized the total flux through the network.

In figure 4, we have selected the behavior of reactions that change significantly over time in the anaerobic environment (env. 3), for three different random seeds of the evolutionary simulations. Surprisingly, most reactions change in a similar way for all random seeds. For example, reactions 278, 1148, 1149, and 1186 (marked yellow), which are involved in methionine metabolism, all become active after a certain time. It appears that this pathway takes over the function of a different pathway. Indeed, each time reactions 763 and 112 (marked yellow), which are also involved in methionine metabolism, become inactive, this pathway becomes active.

Deletions of reactions are also very consistent. There are several reactions that are deleted in all three simulations, although they are initially active in this environment. From the pattern of deletions, it is very clear that deletion of one reaction can inactivate whole pathways and thereby activate others. For example, deletion of reaction 180 or 169 in the first and the second simulation, respectively, deactivates a whole pathway (marked orange).

Some changes in flux are easy to explain. For example, reaction 78 (marked *) is activated after a certain time, in all three simulations. Reaction 78 is responsible for acetate transport. Therefore, because of the increase in glycolytic flux, cells start to also excrete acetate, instead of only ethanol. However, most changes in the network are not so easy to interpret.

Another interesting behavior is exhibited by reactions 183, 187 and 12, 1225 and 1229 (marked **). Both these pathways convert $\alpha$-ketoglutarate into glutamate, an important step in amino acid metabolism. Clearly, small changes in the network lead to different optimal patterns through these reactions, which causes the switching between these two optima.

All these examples show that the response of a metabolic network to certain changes in uptake rates can be very nonlinear. Indeed, it has previously been shown that gene expression and therewith the fluxes through the network can change drastically during short-term evolution of yeast, by only relatively small events (Ferea et al. 1999; Dunham et al. 2002). Such abrupt changes in network dynamics can also be seen in figure 4. This could also partly explain why so many yeast genes are dispensable: maybe they are only used under certain evolutionary circumstances.

We observe that the fate of the metabolic network after WGD is more or less fixed. We might have expected that in several evolutionary simulations, the outcome would be different, because WGD might open many different possibilities. This is however not what we observe. Given a certain environment, the outcome of network evolution after WGD appears to be fixed. Of course, this observation is important for our method to be valid. If WGD could have totally different outcomes for a certain environment, we could never predict the outcome of WGD.

*Under which Circumstances can a WGD be Adaptive?*

*The Effect of WGD on Adaptation to New Environments.* Up to now we have seen that the effects of a WGD in our

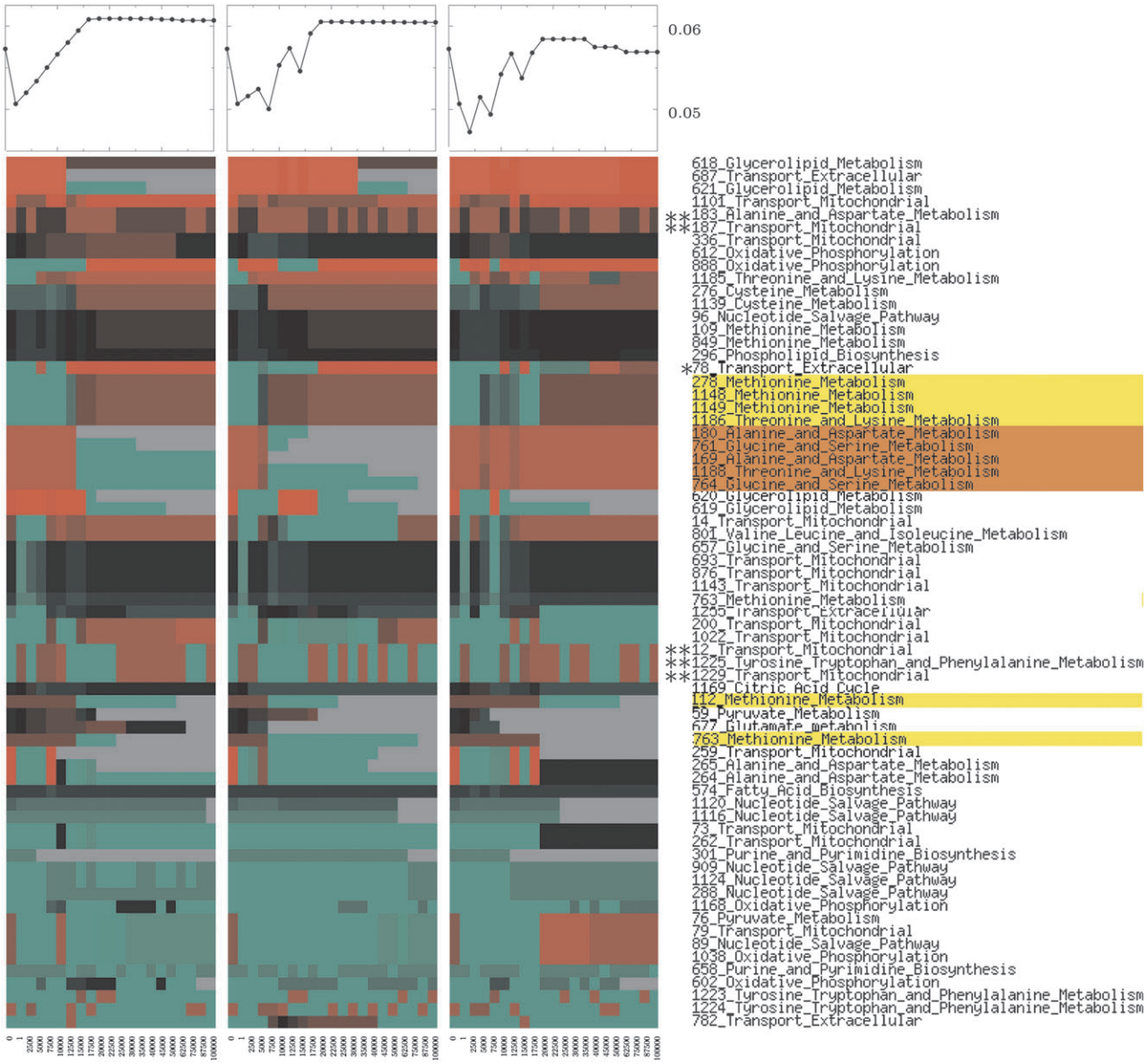


FIG. 4.—Flux through reactions in environment 3, for three different random seeds. Green indicates low flux, black intermediate flux, and red high flux. Gray indicates that the reaction is deleted. The genes are ordered according to a clustering, grouping genes with similar expression patterns together. The clustering is performed on data of all three simulations together on the indicated timepoints, using EPCLUST (http://ep.ebi.ac.uk/), with a linear correlation-based distance and average linkage. The first timepoint indicates pre-WGD. The pathway/function for each reaction is indicated at the right. At the top of the panel, we have shown the growth rate over time for the different cases. Reactions that are discussed in the text are marked.

simulations are comparable with the effect of the WGD in *S. cerevisiae*. The WGD is followed by massive gene loss, transporter genes and glycolysis genes are more often retained, which leads to an increase in glycolytic flux. Furthermore, we have seen that on the long term, the fitness can increase compared with before the WGD. However, it appears that a WGD cannot increase the growth rate instantaneously (fig. 2). In this section, we focus on the question how WGD mutants can take over the population.

In the simulations discussed in the previous section, we assumed that cells were already perfectly adapted to the environment because we used the maximal flux through each reaction experienced in 1000 environments as the initial constraints on the network. Therefore, because of the decrease in surface area to volume ratio, the fitness is initially decreased after WGD.

We wondered whether a WGD could lead to a fitness increase if cells were not previously adapted to the environment. Therefore, we performed simulations for which cells were only adapted to eight of nine different environment types. During the simulations, we let cells evolve mostly in the previously unknown environment type.

The results of these nine simulations, using the maximal metabolite concentration as in figure 2, are shown in figure 5 (gray lines). For every simulation, the growth rate in the "new" environment type is shown. For most environment types, we still observe a drop in the fitness. For the ethanol and acetate environments however, we observe

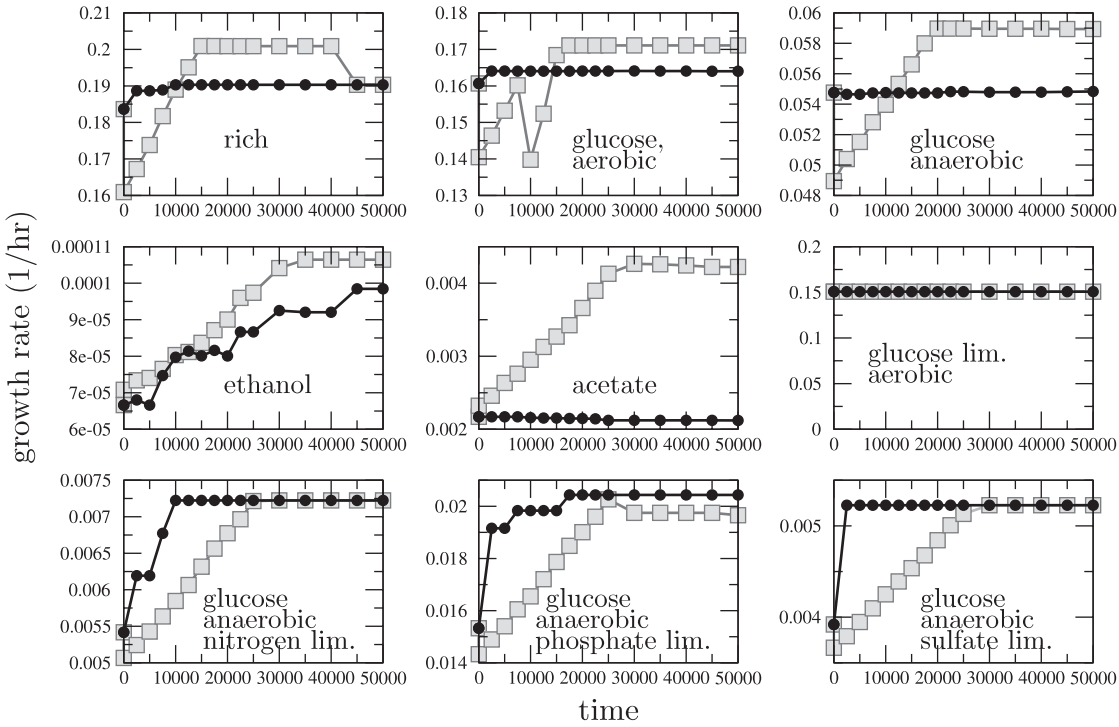


FIG. 5.—Evolution of growth rate in "new" environment types using maximum metabolite concentrations. Cells were fully adapted to all environment types *except* the environment type for which we report the growth rate. In the simulation for environment 4, a "temperature" of 1e−05 was used (see Materials and Methods) because the growth rate was so low in this environment. Again we zoomed in on the first half of the evolutionary simulation. Gray, evolution using WGD and subsequent massive gene loss; Blacks, evolution using SSDs.

that the fitness increases instantly at the WGD and keeps increasing continuously, until it reaches a certain level.

Why is it that in the ethanol and acetate environments fitness immediately increases? All other seven environment types are glucose based. Therefore, these environment types mostly differ in the amounts of other metabolites (e.g., oxygen, ammonia, or phosphate). This means that adapting to one of those environments also gives a pretty good adaptation to the others. However, preadapting to the other eight environments does not give good adaptation to acetate or ethanol, as growth on acetate or ethanol is very different from growth on glucose.Indeed, we observe that only in these environments the fitness is much lower in figure 5 than in figure 2.

Given the parameters we use, a WGD will always decrease the maximal exchange fluxes because the surface area to volume ratio becomes lower. The maximal intracellular fluxes however *do* increase because the increase in dosage is larger than the increase in volume, which is somewhat less than a factor 2 after WGD. Therefore, if the intracellular fluxes are perfectly adjusted, cells do not benefit from the increase of these maximal fluxes and cells only experience the negative effect of the lower surface area to volume ratio. If these intracellular fluxes are not perfectly adjusted, the cells can gain a benefit by the increase in the maximal flux of these reactions and therefore cells can increase their fitness. For this result, the fact that the volume increase is less than a factor 2 is crucial. Note that Galitski et al. (1999) measured a volume increase between tetraploid and diploid cells of more than a factor 2 (2.60), in contrast to Hennaut et al. (1970). The volume increase between diploid and haploid yeast cells was however smaller than 2 (1.54).

*Comparison between WGD and SSDs.* Instead of WGD, SSDs can also help to adapt to new environment types. We are interested whether SSDs can lead to similar adaptations as WGD. It has been proposed that WGD can lead to the duplication of entire pathways or protein complexes, whereas SSDs cannot. Indeed, it has been shown that gene duplicates that arose from the WGD in yeast are more likely part of a protein complex than gene duplicates that arose from SSDs (Hakes et al. 2007). This is even more clearly the case for ribosomal proteins (Scannell et al. 2006). It has also been proposed that SSDs cannot increase glycolytic flux as well as WGD (Conant and Wolfe 2007). To test these ideas, we performed simulations with which we start with the original metabolic network of *S. cerevisiae* and let cells adapt using single gene duplications and deletions. The results are also shown in figure 5 (black lines).

First note that the black lines do not experience the drop in fitness at time equals 0 because we do not start with a WGD. In the first three environment types, the rich, aerobic, and anaerobic glucose environment, we observe that, whereas WGD can lead to a large fitness increase, SSDs can only lead to a minor fitness increase. In these three environments, the glucose metabolism is an important determinant for the fitness of the cells. In the simulations with WGD, all glycolysis reactions are retained in duplicate at the end of the simulations. In the simulations with SSDs, only approximately half of the

glycolysis reactions have been duplicated at the end of the simulations. This leads to an increase in glycolytic flux, but less so than after WGD. This increase in glycolytic flux for SSDs is possible because one reaction after the other becomes rate limiting in the glycolysis and in this way the pathway can be duplicated. However, it is very difficult to duplicate the whole pathway, particularly the reaction phosphofructokinase, which is coded by a protein complex of two proteins (PFK1 and PFK2).

In the ethanol and acetate environment, WGD also performs better than SSDs. In the case of acetate, SSDs cannot increase fitness altogether, whereas WGD leads to an instant fitness increase. In the aerobic, glucose-limited environment, no fitness increase is observed for either mutational mechanisms.

Finally, in the anaerobic nitrogen-, phosphate- and sulfate-limited environments, we observe that SSDs lead to a faster fitness increase than WGD. Only after a long time can WGD lead to a similar fitness increase than SSDs. The difference is that, because these three environment types are limited by the amount of minerals, cells need only to duplicate the right transporter genes to gain a fitness increase. Indeed, all fitness increases in these environments can be attributed to a duplication of a transporter gene. Therefore, when cells need to duplicate only single genes, instead of whole pathways, SSDs lead to much faster adaptation than WGD. In summary, WGD performs better if a whole pathway needs to be duplicated, whereas SSDs perform better if only single genes need to be duplicated.

## Discussion

WGDs have been extensively studied using bioinformatic pattern analysis. In this way, much has been learned about genome evolution after WGD. It has been shown that WGD is followed by massive gene loss and fixation of genes because of gene dosage, subfunctionalization, or neofunctionalization. Here we extend this work using an evolutionary systems biology approach.

We have studied the evolution of a metabolic network after WGD. Using very simple assumptions (gene dosage and a correlation between cell size and genome size), we have been able to formulate a model that can describe genome and network evolution after WGD. This model captures the essential features of this process. It is important to note that, although the model is quite complex, the model has surprisingly few parameters because of the way we modeled the preadaptation of the cells.

One of the few parameters in the model is the volume increase after WGD. We used a value of 1.87, taken from literature (Hennaut et al. 1970). However, as mentioned earlier, higher values have also been measured (2.60 by Galitski et al. 1999). To test whether the results change for different values of this factor, we performed additional simulations for values of 1.5, 2.0, and 2.15. We found that the results did not change in essentials. Only the immediate effect of WGD changed. When this factor is small, WGD can become instantly adaptive in some environments, even if cells are perfectly adapted because the increase in dosage of transporter proteins becomes more important than the decrease in surface area to volume ratio (see equation 2). If this factor becomes larger than 2.0, we never observe an immediate increase in fitness for "new" environments anymore because both the surface area to volume ratio and gene dosage decrease. The end results of the evolutionary simulations however did not change when varying this factor.

Our model only focuses on the metabolic consequences of WGD. Other effects of polyploidy are not taken into account in the model. For example, it has been proposed that heterosis is stronger in polyploid cells. Furthermore, gene redundancy is increased, which could also be advantageous (Comai 2005). However, it is also known that polyploids suffer from genome instability (Storchova and Pellman 2004; Storchova et al. 2006), which will have a negative effect on fitness after WGD. Interestingly, it has been shown that metabolic genes were more retained in duplicate than other genes after the most recent WGD in *P. tetraurelia* due to the higher expression levels of metabolic genes (Gout et al. 2009).

Apart from selection for a smaller genome, neutral loss of genes also plays an important role. Because polyploidy is followed by genome instability, we expect that after WGD, gene deletion rates are higher than gene duplication rates. Without any selection pressure, this would already lead to a decrease in genome size. In the process of genome shrinkage, many duplicate genes are lost, which changes, in contrast to the WGD itself, the relative protein abundances of each gene. In this way, the constraints on the metabolic network change.

Our model satisfactorily predicts gene retention, particularly when we use a reconstructed metabolic network of the pre-WGD ancestor. We also find that the WGD in yeast leads to an increase in glycolytic flux because both glycolysis and transporter genes are retained more than on average.

This last result is in correspondence with the results of Conant and Wolfe (2007). They showed that glycolysis genes are retained in duplicate more often than expected by chance. Furthermore they found, using a kinetic model for glycolysis and assuming gene dosage, that the retention of duplicated genes leads to an increase in glycolytic flux. Instead of a posteriori interpretation of the outcome of WGD in the yeast lineage, we here study how the metabolic network of yeast evolves after WGD. In this way, we found that an increase in glycolytic flux is not only a possible but also the expected outcome of WGD in yeast, which is surprising considering that WGD opens a whole range of evolutionary possibilities.

In our model, the only pathway that is entirely retained in duplicate is glycolysis. Reactions coming from pyruvate, the endpoint of glycolysis, are not retained in duplicate. Pyruvate can be regarded as a "hub" in the metabolic network because many different pathways have pyruvate as a starting point. Therefore, in the model, an increased glycolytic flux can be shuttled to different pathways after pyruvate, which could explain why in the model reactions coming from pyruvate have not been duplicated. In contrast, the glycolysis is a very linear pathway, in which it is much more difficult to reroute fluxes. Whether this explanation also holds for the ancestral WGD in *S. cerevisiae*

is however unclear as several reactions coming from pyruvate have been duplicated.

Functional divergence through subfunctionalization (Force et al. 1999) or neofunctionalization (Byrne and Wolfe 2007) is a possible evolutionary outcome of gene duplication, which we did not include. However, at least for metabolic genes, neofunctionalization does not occur often, as we observed that more than 80% of the ohnolog pairs, for which both genes are present in the model, function in the same reaction in an "OR"-like way. This confirms that most genes retain their function (or possibly subfunctionalize). Functional divergence of duplicate genes through subfunctionalization can be partly accounted for in our model. For example, the glucose transporter gene HXT1 of *S. cerevisiae* is expressed at high glucose concentrations, whereas HXT2 and 4 are expressed at low glucose concentrations (Ozcan and Johnston 1999). By gene duplication, it has been possible for *S. cerevisiae* to specialize genes for different glucose concentrations and hence different fluxes. If there were only one glucose transporter gene, this gene would probably be expressed under all glucose concentrations. Because this gene would not be specialized for high glucose concentrations, the maximum attainable flux would be lower. Generally, we expect that if duplicated genes undergo subfunctionalization, both duplicates will perform their function better and we expect a higher possible flux for these reactions.

We did not take genetic regulation into account in our model, in contrast to, for example, Covert et al. (2001), Herrgard et al. (2006) and Shlomi et al. (2007). If we would incorporate genetic regulation in the model, we would also need to incorporate the *evolution* of genetic regulation during network evolution. Instead, we assume that evolution of genetic regulation is faster than network evolution, such that genetic regulation is optimized during evolution to maximize the growth rate.

Previously, it has been found that ribosomal proteins are retained in duplicate after WGD more often than expected by chance (Wolfe 2004). In our model, we find a strong correlation between growth rate (in a certain environment) and the total flux through the network. Therefore, if the growth rate would increase after WGD, which we expect if WGD is fixed in the population, the total flux through the network should also increase. To be able to reach this higher total flux, it might be necessary to have more ribosomes, which could explain, again assuming dosage dependence of ribosomal genes, the retention of ribosomal genes.

Pal et al. (2006) have adopted a similar approach to model genome reduction in *Buchnera aphidicola*. Performing sequential gene deletions in the metabolic network of *Escherichia coli*, they arrived to a minimal genome that could still function in a nutrient-rich environment. Their algorithm is much simpler as ours because we study genome evolution after WGD. Therefore, genes can also be present twice in our model and we needed a way to describe the difference in fluxes for reactions that are present once or twice. Furthermore, we used evolutionary simulations instead of the greedy search that was used in Pal et al. (2006). Pal et al. (2006) indeed showed for the first time that FBA can also be used to study metabolic network evolution and they were able to predict the evolutionary outcome of genome shrinkage in *B. aphidicola* with more than 80% accuracy. Just as in Pal et al. (2006), we find that metabolic network evolution is surprisingly predictable, given the environments to which cells are exposed, in contrast with the idea that WGD opens up many different evolutionary opportunities.

With predictability, however, we do not mean that in other organisms WGD would have the same outcome as in yeast, only that in yeast WGD is not expected to lead to other outcomes. We believe that this method could also be used for studying the effect of WGD in other single-cell eukaryotes (such as *P. tetraurelia*), but not for multicellular organisms. The outcome of such simulations would then of course depend on the imposed environment, which in our case were mostly glucose-rich. Such a glucose-rich environment is relevant for *S. cerevisiae*, but might not be relevant for other organisms.

Concluding, we have developed a model based on FBA, with which it is possible to study the effects of WGD and subsequent gene loss on a metabolic network. Our model can satisfactorily predict the evolutionary outcome of the WGD in *S. cerevisiae*. There is good agreement between genes that are retained in duplicate in our model and genes that have been retained in duplicate in *S. cerevisiae*. We also found that during evolution, metabolic pathways are activated and inactivated consistently between different evolutionary simulations, indicating that gene dispensability changes during evolution, which could for some part explain why so many genes are dispensable in *S. cerevisiae* (Papp et al. 2004).

Furthermore, we have shown that WGDs can be selective in environments, for which a cell was not yet perfectly adapted. Finally, we have shown that WGDs are favored over SSDs if entire pathways need to be duplicated, whereas the opposite is true if only a few genes need to be duplicated.

These results are in nice correspondence with the hypothesis that the WGD helped yeast to adapt to the newly arisen environment of glucose-rich fruits (Conant and Wolfe 2007; Merico et al. 2007) because we show that WGD is likely to decrease fitness in known environments but can increase fitness in new environments. Furthermore, it appears that just before WGD, yeast evolved the ability to grow anaerobically, which also gave rise to a "new" environment. For adaptation to such a, possibly anaerobic, glucose-rich environment, an increased glycolytic flux is needed, which is easy to achieve via WGD, but much more difficult via SSDs.

## Supplementary Material

Supplementary material is available at *Molecular Biology and Evolution* online (http://mbe.oxfordjournals.org/).

## Literature Cited

Amores A, Force A, Yan YL, et al. (12 co-authors). 1998. Zebrafish hox clusters and vertebrate genome evolution. Science. 282:1711–1714.

Andreasen AA, Stier TJ. 1953. Anaerobic nutrition of *Saccharomyces cerevisiae*. I. Ergosterol requirement for growth in a defined medium. J Cell Physiol. 41:23–36.

Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature. 408:796–815.

Aury JM, Jaillon O, Duret L, et al. (43 co-authors). 2006. Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. Nature. 444:171–178.

Becker SA, Feist AM, Mo ML, Hannum G, Palsson BO, Herrgard MJ. 2007. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA toolbox. Nat Protoc. 2:727–738.

Bilu Y, Shlomi T, Barkai N, Ruppin E. 2006. Conservation of expression and sequence of metabolic genes is reflected by activity across metabolic states. PLoS Comput Biol. 2:e106.

Bowers JE, Chapman BA, Rong J, Paterson AH. 2003. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature. 422:433–438.

Byrne KP, Wolfe KH. 2005. The yeast gene order browser: combining curated homology and syntenic context reveals gene fate in polyploid species. Genome Res. 15:1456–1461.

Byrne KP, Wolfe KH. 2007. Consistent patterns of rate asymmetry and gene loss indicate widespread neofunctionalization of yeast genes after whole-genome duplication. Genetics. 175:1341–1350.

Cavalier-Smith T. 1978. Nuclear volume control by nucleoskeletal DNA, selection for cell volume and cell growth rate, and the solution of the DNA C-value paradox. J Cell Sci. 34:247–278.

Cavalier-Smith T. 2005. Economy, speed and size matter: evolutionary forces driving nuclear genome miniaturization and expansion. Ann Bot (Lond). 95:147–175.

Comai L. 2005. The advantages and disadvantages of being polyploid. Nat Rev Genet. 6:836–846.

Conant GC, Wolfe KH. 2007. Increased glycolytic flux as an outcome of whole-genome duplication in yeast. Mol Syst Biol. 3:129.

Covert MW, Schilling CH, Palsson B. 2001. Regulation of gene expression in flux balance models of metabolism. J Theor Biol. 213:73–88.

Duarte NC, Herrgard MJ, Palsson BO. 2004a. Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. Genome Res. 14:1298–1309.

Duarte NC, Palsson BO, Fu P. 2004b. Integrated analysis of metabolic phenotypes in *Saccharomyces cerevisiae*. BMC Genomics. 5:63.

Dunham MJ, Badrane H, Ferea T, Adams J, Brown PO, Rosenzweig F, Botstein D. 2002. Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae*. Proc Natl Acad Sci USA. 99:16144–16149.

Edwards JS, Palsson BO. 2000. The *Escherichia coli* MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. Proc Natl Acad Sci USA. 97:5528–5533.

Famili I, Forster J, Nielsen J, Palsson BO. 2003. *Saccharomyces cerevisiae* phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. Proc Natl Acad Sci USA. 100:13134–13139.

Ferea TL, Botstein D, Brown PO, Rosenzweig RF. 1999. Systematic changes in gene expression patterns following adaptive evolution in yeast. Proc Natl Acad Sci USA. 96:9721–9726.

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics. 151:1531–1545.

Forster J, Famili I, Fu P, Palsson BO, Nielsen J. 2003. Genome-scale reconstruction of the Saccharomyces cerevisiae metabolic network. Genome Res. 13:244–253.

Galitski T, Saldanha AJ, Styles CA, Lander ES, Fink GR. 1999. Ploidy regulation of gene expression. Science. 285:251–254.

Gatenby RA, Vincent TL. 2003. An evolutionary model of carcinogenesis. Cancer Res. 63:6212–6220.

Gojkovic Z, Knecht W, Zameitat E, Warneboldt J, Coutelis JB, Pynyaha Y, Neuveglise C, Moller K, Loffler M, Piskur J. 2004. Horizontal gene transfer promoted evolution of the ability to propagate under anaerobic conditions in yeasts. Mol Genet Genomics. 271:387–393.

Gout JF, Duret L, Kahn D. 2009. Differential retention of metabolic genes following whole-genome duplication. Mol Biol Evol. 26:1067–1072.

Gregory TR. 2001. Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. Biol Rev Camb Philos Soc. 76:65–101.

Hakes L, Pinney JW, Lovell SC, Oliver SG, Robertson DL. 2007. All duplicates are not equal: the difference between small-scale and genome duplication. Genome Biol. 8:R209.

Hennaut C, Hilger F, Grenson M. 1970. Space limitation for permease insertion in the cytoplasmic membrane of *Saccharomyces cerevisiae*. Biochem Biophys Res Commun. 39:666–671.

Herrgard MJ, Lee BS, Portnoy V, Palsson BO. 2006. Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in *Saccharomyces cerevisiae*. Genome Res. 16:627–635.

Huerta-Cepas J, Dopazo H, Dopazo J, Gabaldon T. 2007. The human phylome. Genome Biol. 8:R109.

Kellis M, Birren BW, Lander ES. 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. Nature. 428:617–624.

Kooijman SA, Muller EB, Stouthamer AH. 1991. Microbial growth dynamics on the basis of individual budgets. Antonie Van Leeuwenhoek. 60:159–174.

Kuepfer L, Sauer U, Blank LM. 2005. Metabolic functions of duplicate genes in *Saccharomyces cerevisiae*. Genome Res. 15:1421–1430.

Merico A, Sulo P, Piskur J, Compagno C. 2007. Fermentative lifestyle in yeasts belonging to the *Saccharomyces complex*. FEBS J. 274:976–989.

Moller K, Olsson L, Piskur J. 2001. Ability for anaerobic growth is not sufficient for development of the petite phenotype in *Saccharomyces kluyveri.* J Bacteriol. 183:2485–2489.

Nagy M, Lacroute F, Thomas D. 1992. Divergent evolution of pyrimidine biosynthesis between anaerobic and aerobic yeasts. Proc Natl Acad Sci USA. 89:8966–8970.

Ohno S. 1970. Evolution by gene duplication. Heidelberg (Germany): Springer-Verlag.

Ozcan S, Johnston M. 1999. Function and regulation of yeast hexose transporters. Microbiol Mol Biol Rev. 63:554–569.

Pal C, Papp B, Lercher MJ, Csermely P, Oliver SG, Hurst LD. 2006. Chance and necessity in the evolution of minimal metabolic networks. Nature. 440:667–670.

Papp B, Pal C, Hurst LD. 2004. Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. Nature. 429:661–664.

Scannell DR, Byrne KP, Gordon JL, Wong S, Wolfe KH. 2006. Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. Nature. 440:341–345.

Semon M, Wolfe KH. 2007. Reciprocal gene loss between etraodon and zebrafish after whole genome duplication in their ancestor. Trends Genet. 23:108–112.

Seoighe C, Wolfe KH. 1999. Yeast genome evolution in the post-genome era. Curr Opin Microbiol. 2:548–554.

Shlomi T, Eisenberg Y, Sharan R, Ruppin E. 2007. A genome-scale computational study of the interplay between transcriptional regulation and metabolism. Mol Syst Biol. 3:101.

Storchova Z, Breneman A, Cande J, Dunn J, Burbank K, O'Toole E, Pellman D. 2006. Genome-wide genetic analysis of polyploidy in yeast. Nature. 443:541–547.

Storchova Z, Pellman D. 2004. From polyploidy to aneuploidy, genome instability and cancer. Nat Rev Mol Cell Biol. 5:45–54.

Taylor JS, Van de Peer Y, Braasch I, Meyer A. 2001. Comparative genomics provides evidence for an ancient genome duplication event in fish. Philos Trans R Soc Lond B Biol Sci. 356:1661–1679.

Van Hoek MJ, Hogeweg P. 2006. In silico evolved lac operons exhibit bistability for artificial inducers, but not for lactose. Biophys J. 91:2833–2843.

Van Hoek MJ, Hogeweg P. 2007. The effect of stochasticity on the lac operon: an evolutionary perspective. PLoS Comput Biol. 3:e111.

Wilcox LJ, Balderes DA, Wharton B, Tinkelenberg AH, Rao G, Sturley SL. 2002. Transcriptional profiling identifies two members of the ATP-binding cassette transporter superfamily required for sterol uptake in yeast. J Biol Chem. 277:32466–32472.

Wolfe K. 2004. Evolutionary genomics: yeasts accelerate beyond BLAST. Curr Biol. 14:R392–R394.

Wolfe KH, Shields DC. 1997. Molecular evidence for an ancient duplication of the entire yeast genome.Nature. 387:708–713.

Aoife McLysaght, Associate Editor

Accepted July 7, 2009