Differential Retention of Metabolic Genes Following Whole-Genome Duplication

Jean-François Gout, Laurent Duret, and Daniel Kahn

Université de Lyon; Université Lyon 1; CNRS; INRIA; UMR5558; Laboratoire de Biométrie et Biologie Evolutive,Villeurbanne, France

Classical studies in Metabolic Control Theory have shown that metabolic fluxes usually exhibit little sensitivity to changes in individual enzyme activity, yet remain sensitive to global changes of all enzymes in a pathway. Therefore, little selective pressure is expected on the dosage or expression of individual metabolic genes, yet entire pathways should still be constrained. However, a direct estimate of this selective pressure had not been evaluated. Whole-genome duplications (WGDs) offer a good opportunity to address this question by analyzing the fates of metabolic genes during the massive gene losses that follow. Here, we take advantage of the successive rounds of WGD that occurred in the *Paramecium* lineage. We show that metabolic genes exhibit different gene retention patterns than nonmetabolic genes. Contrary to what was expected for individual genes, metabolic genes appeared more retained than other genes after the recent WGD, which was best explained by selection for gene expression operating on entire pathways. Metabolic genes also tend to be less retained when present at high copy number before WGD, contrary to other genes that show a positive correlation between gene retention and preduplication copy number. This is rationalized on the basis of the classical concave relationship relating metabolic fluxes with enzyme expression.

Introduction

What are the selective forces shaping metabolic activities in an organism? On the evolutionary timescale, enzyme activities may change as a result of changes of catalytic efficiency, gene expression, or protein stability. Maximal attainable enzyme expression may change as a result of changes in promoter strength, mRNA stability, translation efficiency, or gene dosage through localized amplification, gene duplication, or changes in ploidy (see for instance Kondrashov and Kondrashov 2006 for a review). The resulting changes in enzyme activities will be constrained by the sustainability of metabolic fluxes, that is, by the possibility to reach steady-state fluxes and concentrations compatible with physiological requirements. Classical studies in Metabolic Control Theory (MCT) have shown that metabolic fluxes are not directly proportional to activities of individual enzymes. Instead, fluxes typically show a hyperbolic dependency with respect to enzyme concentrations (see for instance Small and Kacser 1993; Fiévet et al. 2006). Moreover, most enzymes tend to have a relatively low individual control on the fluxes they support. This has been rationalized in the framework of MCT because the sum of the control coefficients of all enzymes upon a flux always sums up to 1, which is one of the classical summation theorems of MCT (see Fell 1997 for an introduction). Because the control of flux is usually distributed among several enzymatic steps, each enzyme tends to have little control over the flux it carries in a large cellular metabolic network. As a consequence, metabolic fluxes are in general insensitive to gene dosage of individual enzymatic steps, which explains why metabolic mutations are typically recessive with respect to their wild-type alleles (Kacser and Burns 1981). For similar reasons, one would expect little selective pressure for maintaining individual duplicated enzyme genes. However, the situation should be different for entire duplicated pathways because fluxes are sensitive to a global change of all enzymes in a pathway.

Up to now, very few data have been available to analyze the selective forces acting on enzyme gene dosage. Here, we reinvestigate this question taking advantage of the large-scale changes in gene dosage that can be monitored following whole-genome duplications (WGDs) in the lineage of *Paramecium tetraurelia*. Indeed the massive gene losses that follow WGDs represent an extreme case that can be exploited to study the sensitivity of metabolism to gene dosage. The *P. tetraurelia* genome is particularly well suited for this analysis because three successive rounds of WGD can be clearly deduced from its genome structure, with only limited genome rearrangements (Aury et al. 2006). These three rounds of WGD will be referred to as "old" "intermediary" and "recent." Paralogous genes that are related by a WGD event are called "ohnologues," in honor of the pioneering ideas of Susumu Ohno on the role of WGDs in genome evolution (Ohno 1970; Wolfe 2001). We also name "ohnologon" a maximal set of mutually ohnologous genes. An ohnologon may thus contain any number from one to eight genes as we consider three WGD events in the history of the *P. tetraurelia* genome. The time elapsed between the old and the intermediary WGDs, as well as between the intermediary and the recent WGDs, was long enough that most gene pairs (respectively 92% and 76%) eventually returned to a single gene. On the other hand, the last WGD must be relatively young because only 49% of the resulting duplicated genes were lost. Therefore, we are witnessing an active phase of massive pseudogenization, which is confirmed by the direct observation of numerous recent pseudogenes. It thus becomes possible to evaluate selective pressure on gene dosage by direct measurements of the rates of gene loss. Contrary to the above expectation, Aury et al. (2006) found that metabolic genes tend to be more retained than other genes after the recent WGD. We present here a detailed analysis of this retention pattern and interpret it in the broad framework of MCT.

Key words: metabolism, gene dosage, whole-genome duplication, *Paramecium*, expression.

E-mail: kahn@biomserv.univ-lyon1.fr.

Mol. Biol. Evol. 26(5):1067–1072. 2009
doi:10.1093/molbev/msp026
Advance Access publication February 13, 2009

Methods

PRIAM-Based Inference of *P. tetraurelia* Metabolic Network

The complete annotated sequence of the *P. tetraurelia* macronuclear genome was downloaded from http://paramecium.cgm.cnrs-gif.fr./download/gff/ (Arnaiz et al. 2007) and integrated into a local ACNUC database (Gouy et al. 1985) to extract sequences of all encoded proteins. We used PRIAM to assign Enzyme Commission (EC) numbers to annotated genes (Claudel-Renard et al. 2003). PRIAM uses position-specific scoring matrices ("profiles") to detect specific enzyme modules with Reverse Position Specific BLAST (RPS-BLAST) (Marchler-Bauer et al. 2002). We retained only matches for which the *E*-value was below 10^{-3} and for which more than 70% of the profile span was aligned. EC numbers were assigned on the basis of lowest *E*-values (93% of the selected matches had an *E*-value below 10^{-10}). In a second step, we used the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al. 2008) to select metabolic enzymes from the EC data set, which was achieved by filtering for "Metabolism" as the first CLASS field of KEGG orthology. This resulted in excluding nonmetabolic enzymes such as protein kinases that are typically encoded by large gene families. After this filtering, enzymes encoded by more than 30 genes were also excluded in order to avoid distorting the analysis by a few very atypical enzymes. This resulted in the exclusion of 280 genes corresponding to only three distinct EC codes: 2.5.1.18 (glutathione transferase, 33 genes); 3.6.3.14 (H^+-transporting two-sector ATPase, 38 genes); 3.1.4.12 (sphingomyelin phosphodiesterase, 209 genes). The complete listing of 1,144 metabolic genes and assigned EC numbers is provided in the Supplementary Material online.

Ohnologous Gene Sets

We used ohnology relationships published previously (supplementary tables 17—19 from Aury et al. 2006). In some cases, the annotation of metabolic genes appeared to be inconsistent among sets of ohnologues: some genes were assigned an EC number, whereas some of their ohnologues were not. Inspection of these cases showed that these ohnologues that are not annotated as metabolic correspond mainly to truncated genes either because of annotation artefacts or because of pseudogenization. Indeed about 1,500 of the 40,000 genes annotated in the *Paramecium* genome are truncated and probably correspond to pseudogenes (Aury et al. 2006). It is not possible by sequence analysis alone to determine whether these truncated genes are functional or pseudogenes. However, the fact that they are still recognizable as ohnologues indicates that if they are pseudogenes, the pseudogenization event must be recent. Hence, these genes have been retained as functional duplicates at least up to a recent past. Such truncated ohnologues were therefore assigned the same EC number as their annotated ohnologues and were not counted as gene losses.

Correction for Gene Expression Levels

Expression data were obtained from single channel microarrays in several growth conditions. Microarrays were designed by Nimblegen, with six 50-mer probes per gene. Signals from the 45 arrays were simultaneously normalized using the normalizeBetweenArrays function from the Limma package (Smyth and Speed 2003). The expression of each gene in each condition was taken as the median of the six individual 50-mer signals. We calculated a global expression level of each gene as the $\log_2$-transformed median value across all 45 arrays.

Correction for gene expression levels aimed at obtaining two subsets with an identical distribution of expression from two sets of ohnologons. Expression levels of ohnologons were taken as that of a randomly chosen gene within each ohnologon, rounded at the first decimal. For each bin of expression level (from 0.0 to 16.0, with a 0.1 interval), we randomly picked an equal number of ohnologons from both data sets. *P* values were obtained on the basis of 1,000 randomizations.

Table 1
Gene Loss Frequencies after WGD

WGD Event	Metabolic Genes	Nonmetabolic	*P* Value
Recent	42%	49%	<0.001
After correction for gene expression levels	42%	40%	NS*
Intermediary	77%	76%	NS
Old	91%	92%	NS

Gene loss frequencies are expressed as the frequencies of ohnologous gene pairs that returned to a single gene after WGD. Significance was assessed with the χ^2 test (NS: not significant at the 5% level). *Significance after correction for gene expression was calculated on the basis of 1,000 random sets of nonmetabolic ohnologons with the same distribution of expression levels (see Methods).

Results

Retention of Metabolic Genes Following WGD is Best Explained by Selection for Gene Expression

Metabolic genes were identified in the *P. tetraurelia* genome in two steps. The first step consisted in mapping genes to EC numbers using EC-specific profiles from PRIAM (Claudel-Renard et al. 2003). The second step selected all enzymes involved in metabolism, excluding enzymes acting on macromolecules on the basis of the KEGG orthology classification (Kanehisa et al. 2008). This resulted in 1,144 predicted metabolic genes, corresponding to 346 distinct EC numbers including 305 enzymes involved in central metabolism (see Supplementary Material online for a complete listing). In order to study the selective pressure acting on the dosage of these metabolic genes, we exploited the relationships between ohnologues, that is, paralogous genes that are related by a WGD event.

In a previous work, Aury et al. (2006) were able to precisely match ohnologous genes deriving from three rounds of WGD, taking advantage of the low rearrangement frequency of the *P. tetraurelia* genome. This makes it possible to dissect gene retention for various gene classes following WGD. Moreover, the last WGD of the *P. tetraurelia* genome is sufficiently recent that it becomes possible to capture dynamic features of gene retention. The *P. tetraurelia* genome is thus particularly well suited for the analysis of selective forces acting on gene dosage.

We compared frequencies of gene loss for metabolic versus nonmetabolic genes after WGD (table 1). We found

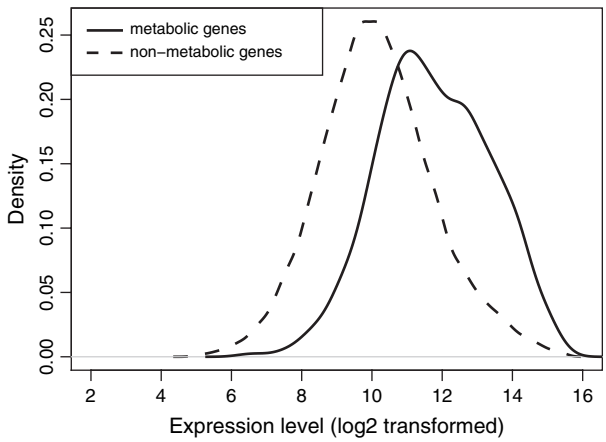

FIG. 1.—Expression of metabolic genes in *Paramecium tetraurelia*. The density distributions of expression levels are shown for the 38,498 nonmetabolic genes (dashed line) and the 1,144 metabolic genes (solid line). The mean expression level is significantly higher for metabolic genes (11.8 vs. 10.1; P value $< 10^{-6}$, Welch two sample t test).

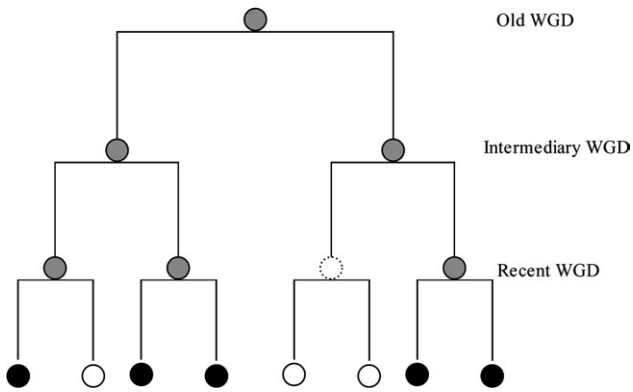

FIG. 2.—Schematic evolution of an ohnologon. The evolution of an ohnologon is inferred by parsimony analysis from the contemporary ohnologon. Full black and gray circles refer to present and past ohnologues, respectively. In this example, gene loss frequencies are 33% and 50% for the recent and intermediary WGD, respectively (calculated as in table 1).

that metabolic genes were more retained than other genes following the recent WGD, in agreement with previous observations (Aury et al. 2006). This selective pressure on gene copy number suggests selection for metabolic gene expression, because the concentration of an enzyme should be proportional to relative gene copy number after WGD. Only on a longer timescale are compensatory mechanisms expected to arise that may compensate for a decrease in relative gene copy number. Indeed there is a positive correlation between gene retention and gene expression, as noted previously both on the yeast genome duplication (Seoighe and Wolfe 1999) and on *P. tetraurelia* (Aury et al. 2006). We therefore compared expression patterns of metabolic genes with that of other genes, on the basis of microarray data obtained in different growth conditions (see Methods) and found that metabolic genes are significantly more expressed than other genes (fig. 1). In order to compare metabolic genes with nonmetabolic genes with similar expression levels, we extracted random subsets of nonmetabolic genes with the same distribution of expression levels (see Methods). We found that metabolic and nonmetabolic genes were retained at a similar level after the recent WGD when correcting for similar expression (table 1). Therefore, the overretention of metabolic genes can be explained primarily by a selective pressure on gene expression.

Distribution of the Number of Ohnologous Genes after Three Rounds of WGD

In order to better characterize metabolic ohnologons, we compared the size distribution of metabolic and nonmetabolic ohnologons. Indeed the present state of an ohnologon is the result of a succession of duplications and gene losses over the three rounds of WGD, which is more informative than gene retention after a single WGD (fig. 2).

The distribution of the number of metabolic genes per ohnologon was narrower than that of nonmetabolic genes, with an excess of ohnologons with two or three genes (fig. 3). Conversely, there was a deficit of metabolic singletons and of large ohnologons with four metabolic genes or more. This deficit was even more pronounced when correcting for similar expression levels, showing that metabolic genes behave indeed differently from other genes with respect to gene dosage. Metabolic singletons were slightly underrepresented after correction for expression levels, which suggests a stronger selective pressure for metabolic gene expression than for nonmetabolic genes at low gene dosage.

Relationship between Gene Retention and Gene Copy Number before WGD

Generally, genes that were retained after a WGD tend to be also retained preferentially after the WGD that follows. For instance, gene loss frequency decreased from 54% to 42% when comparing duplicated genes that were present at single copy or multiple ohnologous copies,

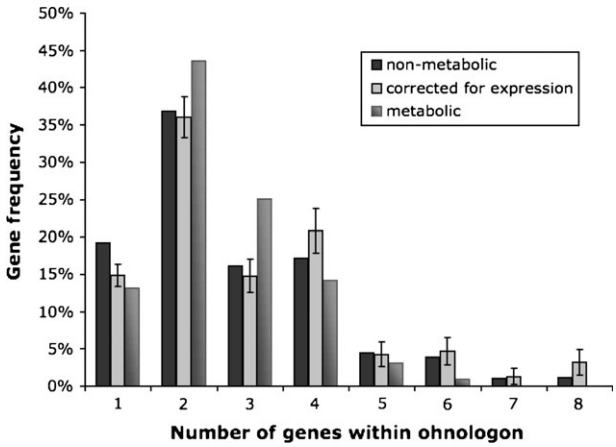

FIG. 3.—Distribution of gene numbers in ohnologons. The histogram reports the distribution of gene numbers for metabolic genes and nonmetabolic genes with dark gray and black bars, respectively. Correction for gene expression was obtained by drawing 1,000 random sets of nonmetabolic genes with the same distribution of expression as metabolic genes (light gray; error bars indicate standard deviations).

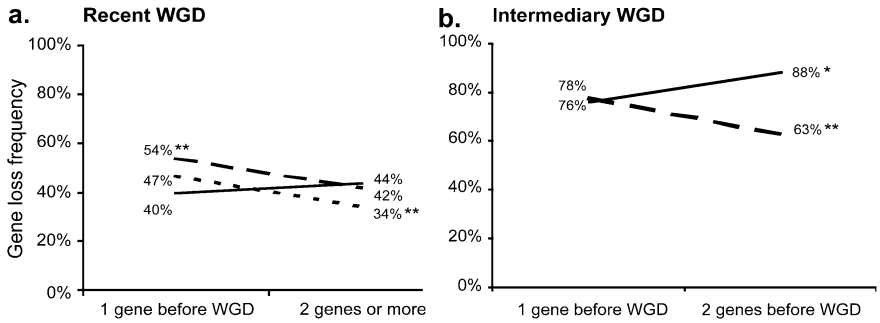

FIG. 4.—Relationship between gene loss frequency and number of genes before WGD. Gene loss frequencies were calculated as in table 1 and compared, depending on the number of ohnologous genes before WGD. Full line: metabolic genes. Long dashes: nonmetabolic genes. Significance of differential gene loss depending on gene copy number was assessed with a χ^2 test. When correcting for expression levels of nonmetabolic genes (short dashes), significance was assessed on the basis of the fraction of randomized sets that contradicted the main trend. **, P value $< 10^{-3}$; *, P value $= 6.6 \times 10^{-2}$.

respectively before the recent WGD (fig. 4*a*). Similarly, gene loss frequency decreased from 78% to 63% when comparing genes that had been retained as single copy versus two copies before the intermediary WGD (fig. 4*b*). Therefore, there is a positive correlation between gene retention propensity at successive WGDs, which can be understood as resulting from a continuous selective pressure on gene expression, hence on gene dosage, consistently exerted both before and after the WGD event.

The retention pattern appeared quite different for metabolic genes, which did not show any positive correlation between retention and copy number before WGD. Their average loss rate even increased from 76% to 88% when comparing genes that had been retained as single copy versus two copies before the intermediary WGD (fig. 4*b*). This difference in retention patterns was not related to differences in gene expression. Indeed nonmetabolic genes drawn randomly with the same distribution of gene expression still exhibited a positive correlation between gene retention and copy number (fig. 4*a*). Moreover, metabolic genes were more retained than nonmetabolic genes with similar expression levels when there was only one gene per ohnologon before the recent WGD (40% vs. 47% gene loss, P value $= 9 \times 10^{-3}$, fig. 4*a*). Therefore, there is a preferential loss of metabolic genes at high gene copy number, which makes their retention pattern atypical.

Discussion

According to classical MCT, metabolic fluxes are usually quite insensitive to changes in expression of an individual enzyme in the large metabolic network encountered in the cellular context. This inherent robustness of metabolic networks explains why most metabolic mutations are recessive (Kacser and Burns 1981). By analogy, the theory predicts that duplicated metabolic genes are generally dispensable, which would make their loss neutral and result in rapid gene decay. A priori, one may expect two exceptions to that rule. First, the retention of duplicated metabolic genes may be favored by natural selection when one or both copies evolved toward different functions (by neo or subfunctionalization) (Force et al. 1999). Although some cases of functional divergence have been described, the fact that most duplicated genes are massively lost following WGDs suggests that the vast majority of ohnologous genes perform redundant functions during this phase of genome reduction. Therefore, neofunctionalization or subfunctionalization concerns primarily the minority (8%) of ohnologous genes that are retained over long evolutionary time frames and only marginally the more recent ohnologues that we study here. Thus, we can conclude that the dynamics of gene loss on the short term (i.e., after the recent or intermediary WGD) is mainly governed by dosage constraints, not by functional divergence. Second, some metabolic genes may be retained as duplicates because they encode enzymes with particularly high flux control coefficients (between 0.5 and 1). An example of this type concerns the hexose transporter that exerts high control over glycolysis in a number of organisms (e.g., Pritchard and Kell 2002). For these enzymes with exceptionally high control on flux, high gene dosage can be interpreted directly as resulting from a selection for high flux. However, such high flux control coefficients are the exception, not the rule (see for instance Chapter 6 of Fell 1997). Therefore, high flux coefficients cannot explain the global observation that metabolic genes have been preferentially retained following the recent WGD in the *Paramecium* lineage (Aury et al. 2006).

Selective Pressure for Metabolic Gene Expression

The preferential retention of metabolic genes suggests a selective pressure for metabolic gene expression. Indeed gene dosage directly influences the global expression level of a protein, so that highly expressed genes tend to be retained preferentially after WGD (Seoighe and Wolfe 1999). Metabolic genes were expressed at a higher level than nonmetabolic genes (fig. 1), and they were retained at a similar level as other genes after correction for this higher gene expression (table 1). Therefore, we attribute the higher retention of metabolic genes to selection for gene expression, which seems at odds with the reasoning above, based on MCT, concluding that duplicated metabolic genes should be generally dispensable because of a lack of control on flux. Note however that this reasoning considered the effect of the loss of a single duplicated metabolic gene, which

differs from the situation experienced after WGD in which a large number of duplicated genes are gradually lost.

Consider for instance a linear metabolic pathway with n distinct steps. After WGD, each of the n enzymes is expressed from two ohnologous genes, one of which may be lost randomly. Let p be the number of steps for which one of the two ohnologues has been lost, numbered from 1 to p without loss of generality. Following Small and Kacser (1993) (their eq. 24), the change in flux J in the pathway is well approximated by the following relationship:

$$\frac{J}{J_0} = \frac{1}{1 + \sum_{i=1}^{p} C_i^{J_0}} \quad (1)$$

where J_0 is the metabolic flux achieved immediately after WGD and $C_i^{J_0}$ are flux control coefficients relating the steady-state flux J and the reaction rates v_i around the initial flux J_0: $C_i^{J_0} = \frac{v_i}{J_0}\frac{\partial J}{\partial v_i}$. The metabolic flux will decrease gradually as a hyperbolic function of the group control coefficient $\sum_{i=1}^{p} C_i^{J_0}$ for each additional loss of an ohnologue. Ultimately, the flux will be divided by 2 if all ohnologues are lost because of the summation theorem:

$$\sum_{i=1}^{n} C_i^{J_0} = 1 \quad (2)$$

Such a 2-fold decrease in flux will have a very strong impact on fitness when it concerns catabolism or central metabolism (Dykhuizen and Dean 1990), and indeed 88% of the metabolic genes considered here concern central metabolism. Therefore, although the loss of individual ohnologues is expected to be generally neutral, the eventual loss of all ohnologous duplications in a metabolic pathway will be strongly counterselected in the absence of compensating changes in relative gene expression. This explains the global pattern of high retention of metabolic genes after the recent WGD, which is related to their expression levels. Potentially, two types of compensating changes could relieve the selective pressure on enzyme gene dosage. The first type concerns changes in promoter activities so that individual genes are more expressed. The second type of compensating change results from the decrease of the genome size. Indeed even in the absence of changes in promoter activity, relative gene expression will change in proportion to the relative gene dosage l/L, where l is the dosage of the gene of interest and L is the total number of genes in the genome. As a result, the flux normalized per total protein amount, hence per cellular unit volume, will scale as J/L and increase upon the gradual decrease of genome size. Ultimately, the normalized flux will return to its original preduplication level when all gene duplications have been lost.

Retention of Ohnologous Genes across Successive WGDs

Metabolic genes behave differently from other genes for one aspect: their retention across successive WGDs. Indeed, nonmetabolic genes were retained preferentially when they had been retained from the previous WGD, which manifests a continuous selection on gene expression. Metabolic genes did not show this trend or showed even the inverse trend after the intermediary WGD (fig. 4). We propose that this difference reflects a different relationship between fitness and gene dosage (fig. 5).

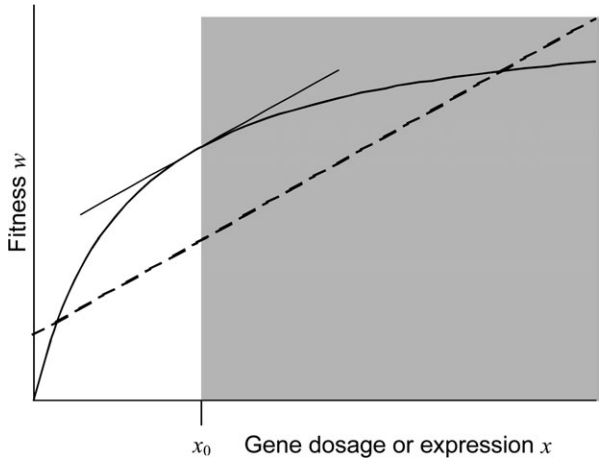

FIG. 5.—General relationship between fitness and gene dosage or expression. The model posits a concave hyperbolic relationship between fitness w and gene dosage or expression x for metabolic genes (full line), versus a general affine relationship for other genes (dashed line). The selective pressure on gene dosage is the same for both models at a particular value x_0 of total ohnologue expression such that $\partial w/\partial x$ is equal under both models. Selection on metabolic genes is lower than on other genes above this value (shaded area). The converse holds true below x_0. This model explains why retention of metabolic genes is inversely related to gene dosage.

Consider the same linear pathway as above, immediately after the recent WGD: Each of the n enzymes is expressed from $2k$ ohnologous genes, depending on the number k of genes in the ohnologon before WGD ($k = 1,\ldots,4$). As above, we can approximate the flux J remaining after retaining only lp genes out of the $2kp$ genes initially involved in p reactions (Small and Kacser 1993):

$$\frac{J}{J_0} = \frac{1}{1 + \sum_{i=1}^{p} C_i^{J_0} \frac{2k-l}{l}} \quad (3)$$

This concave hyperbolic relationship between metabolic flux J and gene dosage l implies that the initial losses of flux are small upon loss of metabolic ohnologues: $J/J_0 > l/2k$ will apply, provided $\sum_{i=1}^{p} C_i^{J_0} < 1$. Only if all n metabolic ohnologons are reduced from $2k$ to l genes will the flux scale in proportion to gene dosage with $J/J_0 = l/2k$. We translate these relationships into relative fitness w under a simple model in which the normalized flux contributes linearly to fitness above a basal fitness w_0 that remains when the pathway is disrupted:

$$w = w_0 + (1 - w_0)\frac{J}{J_0}\frac{L_0}{L} \quad (4)$$

where L is the current number of genes in the genome and L_0 the total number of genes immediately after WGD. Because most metabolic pathways are essential, $w_0 \approx 0$ so that fitness scales linearly with flux (Dykhuizen and Dean 1990):

$$w \approx \frac{J}{J_0}\frac{L_0}{L} > \frac{l}{2k}\frac{L_0}{L} \quad (5)$$

Therefore, metabolic ohnologons are intrinsically robust to initial gene losses particularly when k is high. Our analysis shows that this is not the case for nonmetabolic genes, because they are globally more retained when $k \geq 2$. Therefore, the general relationship between fitness and gene dosage must be different. The simplest model is to consider an affine relationship between relative fitness w and gene dosage l:

$$w = w_0 + (1 - w_0)\frac{l}{2k}\frac{L_0}{L} \quad (6)$$

which entails a constant fitness penalty $\{[1 - w_0]/2k\}\{L_0/L\}$ for each loss of an ohnologue. This model explains the lower retention of metabolic genes compared with nonmetabolic genes with similar expression levels when k is high and the converse when k is low. Indeed we may plot fitness as a function of total ohnologue expression under both models (fig. 5). Here "total ohnologue expression" refers to the cumulated expression from all isofunctional ohnologues: It results both from the relative gene dosage $\{l/2k\}\{L_0/L\}$ and from individual promoter activities. One may define two regions of total ohnologue expression: Above a certain level, the selective pressure is below average because of the concave relationship relating fitness and expression. The converse is true at low dosage or expression, when the selective pressure on metabolic ohnologues is higher than average. These opposite trends tend to stabilize metabolic ohnologue expression, hence dosage, around a neutral point at which the selective pressure on metabolic gene expression is equal to the average. Therefore, in the long term, the relative dosage of metabolic genes will be buffered by selection, as we could indeed observe following the old WGD. In the short term, however, metabolic genes will be more retained, so that their relative dosage will increase as various other genes are lost at a faster rate. As a consequence, they will experience less and less selective pressure (shaded area, fig. 5), so that their dosage will eventually tend to return to their equilibrium level.

In conclusion, the pattern of retention of metabolic genes after WGD in the genome of *Paramecium* is consistent with the predictions of MCT. Interestingly, the predictions of this model are different in case of individual gene duplications compared with WGDs: In the first case MCT predicts a rapid loss of duplicated genes, whereas in the latter, it predicts a global selective pressure to retain gene duplicates, because the pressure on a given gene depends on the fate of the other duplicated genes in the metabolic pathway. With its three full rounds of WGD, the *Paramecium* genome has thus offered a unique opportunity to study selective forces acting on metabolic gene dosage on a wide range of evolutionary timescales.

Supplementary Material

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

Acknowledgments

We thank Mireille Bétermier, Jean Cohen, Eric Meyer, Linda Sperling, and their colleagues for generating the expression data. We also thank Vicente Acuña, Ludovic Cottret, and Marie-France Sagot for helpful discussions. This work was supported by the Agence Nationale de la Recherche (DUPLIGEN: NT05-2_41522).

Literature Cited

Arnaiz O, Cain S, Cohen J, Sperling L. 2007. ParameciumDB: a community resource that integrates the *Paramecium tetraurelia* genome sequence with genetic data. Nucleic Acids Res. 35:D439–D444.

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.

Claudel-Renard C, Chevalet C, Faraut T, Kahn D. 2003. Enzyme-specific profiles for genome annotation: PRIAM. Nucleic Acids Res. 31:6633–6639.

Dykhuizen DE, Dean AM. 1990. Enzyme activity and fitness: evolution in solution. Trends Ecol Evol. 5:257–262.

Fell D. 1997. Understanding the control of metabolism. London: Portland Press.

Fiévet JB, Dillmann C, Curien G, de Vienne D. 2006. Simplified modelling of metabolic pathways for flux prediction and optimization: lessons from an in vitro reconstruction of the upper part of glycolysis. Biochem J. 396:317–326.

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.

Gouy M, Gautier C, Attimonelli M, Lanave C, di Paola G. 1985. ACNUC – a portable retrieval system for nucleic acid sequence databases: logical and physical designs and usage. Comput Appl Biosci. 1:167–172.

Kacser H, Burns JA. 1981. The molecular basis of dominance. Genetics. 97:639–666.

Kanehisa M, Araki M, Goto S, et al. (11 co-authors). 2008. KEGG for linking genomes to life and the environment. Nucleic Acids Res. 36:D480–D484.

Kondrashov FA, Kondrashov AS. 2006. Role of selection in fixation of gene duplications. J Theor Biol. 239:141–151.

Marchler-Bauer A, Panchenko AR, Shoemaker BA, Thiessen PA, Geer LY, Bryant SH. 2002. CDD: a database of conserved domain alignments with links to domain three-dimensional structure. Nucleic Acids Res. 30:281–283.

Ohno S. 1970. Evolution by gene duplication. London: George Allen and Unwin.

Pritchard L, Kell DB. 2002. Schemes of flux control in a model of *Saccharomyces cerevisiae* glycolysis. Eur J Biochem. 269:3894–3904.

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

Small JR, Kacser H. 1993. Responses of metabolic systems to large changes in enzyme activities and effectors. 1. The linear treatment of unbranched chains. Eur J Biochem. 213:613–624.

Smyth GK, Speed T. 2003. Normalization of cDNA microarray data. Methods. 31:265–273.

Wolfe KH. 2001. Yesterday's polyploids and the mystery of diploidization. Nat Rev Genet. 2:333–341.

Kenneth Wolfe, Associate Editor

Accepted January 26, 2009