

Accelerated Rates of Intron Gain/Loss and Protein Evolution in Duplicate Genes in Human and Mouse Malaria Parasites

Cristian I. Castillo-Davis,¹ Trevor B. C. Bedford,¹ and Daniel L. Hartl

Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts

Very little is known about molecular evolution in the human malaria parasite *Plasmodium falciparum*. Given the potentially important role that introns play in directing transcription and the posttranscriptional control of gene expression, we compare rates of intron/gain loss and intronic substitution in *P. falciparum* and the rodent malaria *P. y. yoelii* in both orthologous and duplicate genes. Specifically, we test the hypothesis that intron gain/loss and protein evolution is accelerated in duplicate genes versus orthologous genes in both parasites using the genome sequence of both species. We find that duplicate genes in both *P. falciparum* and *P. y. yoelii* exhibit a dramatic acceleration of both intron gain/loss and protein evolution in comparison with orthologs, suggesting increased directional and/or relaxed selection in duplicate genes. Further, we find that rates of intron gain/loss and protein evolution are weakly coupled in orthologs but not paralogs, supporting the hypothesis that selection acts on genes as functionally integrated units after speciation but not necessarily after gene duplication. In contrast, we find that rates of nucleotide substitution do not differ significantly between intronic sites and synonymous sites among duplicate genes, implying that a large fraction of intronic sites in *Plasmodium* evolve under little or no selective constraint.

Introduction

It has been suggested that regulatory control of gene expression in the human malaria parasite *Plasmodium falciparum* is unique. Recent microarray experiments have shown that transcriptional control of asexual development in *P. falciparum* follows a rigid clocklike scheme, distinct from any eukaryote known so far (Bozdech et al. 2003). Studies using SAGE have also revealed potentially novel mechanisms of gene regulation at the posttranscriptional level in *P. falciparum* involving antisense transcripts across a significant portion of the genome (Patankar et al. 2001). Additionally, known enhancers in *P. falciparum* lack homology to enhancers in any other eukaryote, leading to speculation that *P. falciparum* has developed a unique set of transcription factors different from yeast and higher eukaryotes (Horrocks, Dechering, and Lanzer 1998). Finally, expression and differential silencing among different members of the *var* antigenic gene family have been shown to involve a novel cooperative interaction between introns and upstream elements (Calderwood et al. 2003), suggesting an important role for introns in directing gene regulation in this organism. However, very little is known about intron evolution in *P. falciparum*, although it has been recently suggested that polymorphism in intronic regions may be much lower than in protein-coding synonymous sites because of intense purifying selection (Jongwutiwes et al. 2002).

In species so far examined, intron positions are remarkably conserved over long intervals of evolutionary time (Moriyama, Petrov, and Hartl 1998; Kent and Zahler 2000; Roy, Fedorov, and Gilbert 2003), although there is mounting evidence that lineage-specific intron loss and gain may occur (Rogozin et al. 2003). Mechanistically, intron loss is thought to take place both by partial DNA deletion (Llopart et al. 2002) and by gene conversion

events with reverse transcribed pre-mRNA (Roy et al. 2003). Intron gain is thought to occur by reverse splicing of a preexisting nuclear intron into a pre-mRNA, followed by reverse transcription and gene conversion (Tarrío, Rodríguez-Trelles, and Ayala 1998).

Even in highly expressed genes where selection may act to reduce the size or presence of introns because of transcriptional cost, short introns, but not the loss of introns, appear to be favored (Castillo-Davis et al. 2002). It has, therefore, been suggested that functional constraints on introns at the level of gene regulation may be responsible for their maintenance (Castillo-Davis et al. 2002). For example, it is known that spliceosomal introns play a critical role in eukaryotic gene regulation, both stimulating and repressing transcription (Fedorova and Fedorov 2003) and controlling the nucleocytoplasmic transport of mRNAs from the nucleus (Zhou et al. 2000; Maniatis and Reed 2002).

Given the unique nature of gene regulation in *P. falciparum*, in particular the potentially important role that introns may play in directing transcription and post-transcriptional control of gene expression, we compare rates of intron/gain loss and intronic substitution as well as protein evolution between *P. falciparum* and the rodent malaria parasite *P. y. yoelii*. Additionally, because gene duplication is thought to be central to the evolution of novel molecular functions, adaptation, and the generation of genetic diversity (Ohno 1970; Lynch and Conery 2000), we further examine these evolutionary parameters among duplicate genes in each species. In particular, we test the hypothesis that intron gain/loss, intronic substitution, and protein evolution are accelerated in duplicate versus orthologous genes in both parasites using the genome sequence of each species (Carlton et al. 2002; Gardner et al. 2002).

We find that duplicate genes in both *P. falciparum* and *P. y. yoelii* exhibit a dramatic acceleration of both intron gain/loss and protein evolution in comparison with orthologs, suggesting increased directional selection and/or relaxed selection in duplicate genes. At the same time, we find that rates of nucleotide substitution do not differ significantly between introns and fourfold degenerate

¹ These authors contributed equally to this work.

Key words: gene duplication, genome evolution, intron gain/loss, malaria.

E-mail: dhartl@oeb.harvard.edu.

Mol. Biol. Evol. 21(7):1422–1427. 2004

doi:10.1093/molbev/msh143

Advance Access publication April 14, 2004

synonymous sites among duplicate genes, suggesting that a large fraction of intronic sites evolve under little or no selective constraint.

Methods

Protein Orthology, Duplication, and Evolutionary Analysis

Nucleotide sequences for 5,409 mapped and annotated genes of *P. falciparum* were obtained from PlasmoDB release 4.0 (<http://www.plasmodb.org>). Nucleotide sequences for 7,861 annotated genes of *P. y. yoelii* were obtained from the TIGR *Plasmodium yoelii* Genome Database (<http://www.tigr.org/tdb/e2k1/pya1/>), which contained the draft 5x shotgun genome assembly. Sequences that did not begin with ATG, that did not end with a stop codon, that possessed internal stop codons, that contained ambiguous bases, or that were less than 100 amino acids in length, were removed, yielding 5,054 and 4,106 genes for *P. falciparum* and *P. y. yoelii*, respectively.

Orthologous genes between *P. falciparum* and *P. y. yoelii* were obtained from the TIGR *Plasmodium yoelii* Genome Database as identified by Carlton et al. (2002) using the criterion of reciprocal best hits (Tatusov, Koonin, and Lipman 1997) with BlastP scores of $E < 1 \times 10^{-15}$. Only alignments with greater than 80% similarity in length were retained, yielding 1,822 orthologs.

Duplicate genes within the *P. falciparum* and *P. y. yoelii* genomes were obtained by alignment of each protein against every other in the proteome using BlastP version 2.26 (Altschul et al. 1997). Alignments with greater than 80% similarity in length and with $E < 1 \times 10^{-10}$ were considered significant. Following Lynch and Conery (2000), in an effort to avoid biases caused by the differing evolution of large gene families (including antigenic genes), we eliminated genes which had six or more significant BlastP alignments within a genome. After such screening, 927 and 497 pairs of duplicate genes remained for *P. falciparum* and *P. y. yoelii*, respectively. Next, all coding sequence pairs were globally aligned with ClustalW version 1.82 (Thompson, Higgins, and Gibson 1994) (default parameters) using amino acid sequences followed by back-translation into nucleotides using the original nucleotide sequence.

Maximum-likelihood estimates of rates of nonsynonymous substitution (dN) and synonymous substitution (dS) between pairwise alignments were obtained with PAML version 3.13d (Yang 2000) using a codon-based model of sequence evolution (Goldman and Yang 1994, Yang et al. 2000) with dN and dS as free parameters and average nucleotide frequencies estimated from the data at each codon position (F3 × 4 MG model [Muse and Gaut 1994]); transition/transversion bias (κ) was estimated from unsaturated ($dS < 0.4$) paralogous genes in *P. falciparum* and *P. y. yoelii* and found to be similar in both genomes ($\kappa = 1.535$). It was, therefore, held constant in all analyses (Yang 2000). Based on simulations using random sequence pairs, pairs of sequences with $dS > 3$ were excluded from analysis because these sequences are likely misidentified as orthologs or paralogs (more than 90% of random gene pairs have $dS > 3$; data not shown), yielding 1,490 valid

orthologs and 717 and 378 paralogs in *P. falciparum* and *P. y. yoelii*, respectively. Furthermore, because estimates of $dS > 1.5$ are prone to error, only genes with $dS < 1.5$ were used for statistical calculations, yielding 1,095 valid orthologs and 250 and 110 paralogs in *P. falciparum* and *P. y. yoelii*, respectively.

To facilitate comparison of genes of a similar age/mutational class, we compared duplicate-gene pairs with a dS centered around the mode of the distribution of dS between orthologs ($dS = 1.15$) unless otherwise stated ($dS = 0.9-1.4$, $n = 184$). Duplicate genes were identified as tandemly duplicated on the basis of gene annotations if no intervening gene was present between a given duplicate pair.

Intron Gain/Loss and Substitution

Intron gain/loss was determined in both orthologous and duplicate-gene pairs by comparing annotation information between genes. For duplicate genes that are part of larger gene families (three to five members), a gain or loss may be counted more than once by this method. Therefore, we obtained a subset of duplicate-gene pairs that were each others closest relatives by the method of reciprocal best hits (Tatusov, Koonin, and Lipman 1997) within each genome, where a gain/loss could be counted only once. We repeated all analyses with this smaller data set.

Intron sequences of paralogous genes were obtained from PlasmoDB and the TIGR *Plasmodium yoelii* Genome Database and aligned using ClustalW under default parameters. Because intronic nucleotide substitutions are saturated in orthologous genes, we compared rates of intronic nucleotide substitution with rates of fourfold synonymous substitution in recent duplicate genes ($dS < 1.0$). Substitutions per intronic site were counted directly from intronic nucleotide alignments without correcting for multiple hits. Substitutions per fourfold synonymous site were similarly calculated to facilitate a direct comparison between intron and coding sequence substitution. Comparisons using corrections for multiple hits did not change the results (data not shown).

Control for Errors in Gene Prediction Using Expression Data

To test the possibility that the correlation observed between dN and intron gain/loss was an artifact of poor gene prediction, we examined this relationship using only those genes known to be expressed in *P. falciparum*. Unfortunately, genome-wide expression data is not yet available for *P. y. yoelii*. We considered a gene expressed if it (1) significantly matched a known expressed sequence tag (EST) in PlasmoDB (>500 bp match) and (2) was detected as expressed according to Le Roch et al. (2003) based on Affymetrix microarray expression data.

Results and Discussion

We observe substantially accelerated rates of nonsynonymous substitution (dN) in duplicate genes in both *P. falciparum* and *P. y. yoelii* ($n = 250$ and $n = 110$, respectively) compared with orthologous genes ($n = 1,490$) ($P \ll 10^{-4}$; Mann-Whitney U test) (fig. 1). Note that in

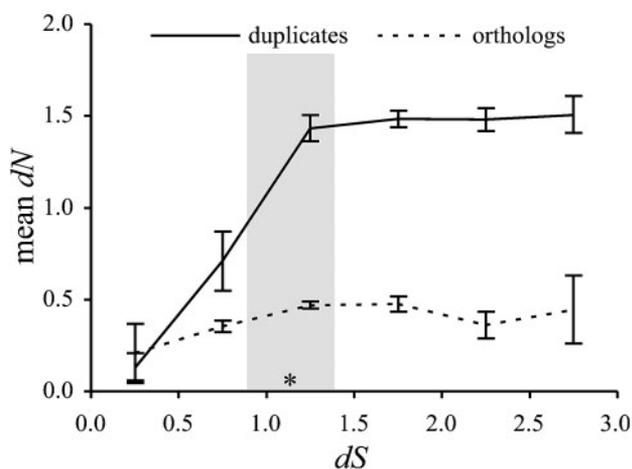


FIG. 1.—Duplicate genes exhibit accelerated rates of nonsynonymous substitution (dN) in comparison with orthologous genes at almost all levels of synonymous divergence (dS). Mean values of dN for each bin are given and error bars show 95% confidence intervals as determined by nonparametric bootstrap replication with 1,000 replicates. The mode of dS of orthologous genes is shown (asterisk) as well as the range of dS used in ortholog-duplicate comparisons (shaded area). Note that accelerated rates of nonsynonymous substitution (dN) are also observed for duplicate genes in the *P. falciparum* and *P. y. yoelii* genomes analyzed separately.

orthologous genes, the spread in dS represents stochastic variation in substitution rate among genes, because all gene pairs are by definition the same age (the time of species divergence). In duplicate genes, dS is affected by both stochastic factors and the amount of time since duplication. Assuming speciation of *P. falciparum* and *P. y. yoelii* occurred 80 to 100 MYA, coinciding with the speciation of the primate-rodent lineage (Perkins and Schall 2002), the average rate of synonymous substitution is approximately 5.75 to 7.19 substitutions per synonymous site per 10^9 years.

Mean rates of protein evolution (dN) are also substantially accelerated in duplicate genes in both the *P. falciparum* and *P. y. yoelii* genomes in comparison with orthologs of approximately the same age (see *Methods*) ($dup_{fal} = 1.48$, $n = 151$ and $dup_{yoe} = 0.98$, $n = 33$ versus $orth = 0.43$, $n = 1095$; $P \ll 10^{-4}$ for each test [fig. 1]). A similar pattern has been observed in the protein-coding regions of duplicate genes in other eukaryotic species (Kondrashov et al. 2002; Nembaware et al. 2002; Castillo-Davis et al. 2004) and for upstream regulatory sequences in *C. elegans/C. briggsae* (Castillo-Davis et al. 2004). New to this study is the observation that intron gain/loss in duplicate genes in the genomes of both *Plasmodium* species is dramatically accelerated compared with orthologs, ($dup_{fal} = 1.15$ and $dup_{yoe} = 1.42$ versus $orth = 0.39$, $P \ll 10^{-4}$ for each test; Mann-Whitney U test [fig. 2]). Overall, twice as many amino acid substitutions occur and twice as many introns are gained or lost between duplicate-gene pairs compared with between orthologs scaled by the same amount of time/mutation. Results did not change when using data where intron gain/loss was estimated from terminal duplicate pairs only (see *Methods*). Because intron gain/loss increases with increasing dS in duplicates, it is likely that intron gain/loss is not caused by duplication

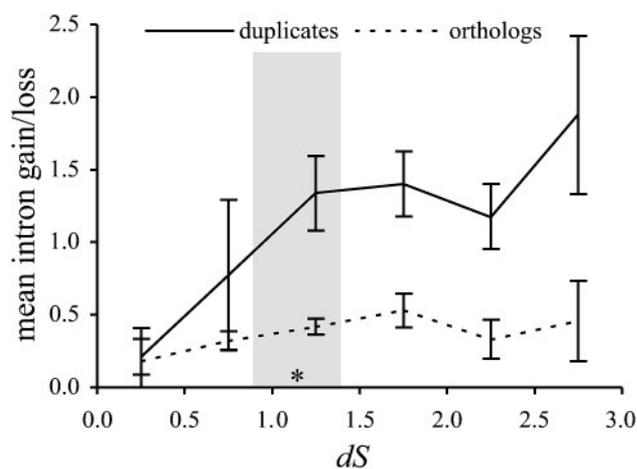


FIG. 2.—Duplicate genes exhibit accelerated rates of intron gain/loss in comparison with orthologous genes at almost all levels of synonymous divergence (dS). Mean values of intron gain/loss for each bin are given. Error bars show 95% confidence intervals as determined by nonparametric bootstrap replication with 1,000 replicates. The mode of dS of orthologous genes is shown (asterisk) as well as the range of dS used in ortholog-duplicate comparisons (shaded area). Note that accelerated rates of intron gain/loss are also observed for duplicate genes in the *P. falciparum* and *P. y. yoelii* genomes analyzed separately.

by retrotransposition but by another molecular mechanism such as nonhomologous recombination.

Interestingly, the pattern of accelerated evolution observed in duplicates was different for tandem and nontandem duplicate genes, with tandem duplicate genes showing a lower mean rate of protein evolution (dN) than nontandem duplicates ($tandem_{fal} = 0.32$, $nontandem_{fal} = 1.33$, $P < 0.001$). Tandem duplicates also show fewer (although not significant) intron gains/losses ($tandem_{fal} = 0.286$, $nontandem_{fal} = 1.177$, $P = 0.12$). Given that dS is also significantly reduced in tandem pairs ($tandem_{fal} = 0.753$, $nontandem_{fal} = 1.148$, $P = 0.04$), it is likely that gene conversion between, and/or a recent origin of, tandem duplicate genes, is responsible for this pattern.

Two non-mutually exclusive scenarios can be envisaged to explain the accelerated evolution of duplicate genes. First, duplicate genes could experience weaker purifying selection than orthologs (i.e., relaxed selection). Second, duplicated genes could experience greater positive selection than orthologs. Although there is still much debate concerning the process by which initially identical duplicate genes come to diverge in sequence and function, it is certain that after duplication, the resulting genes are subject to either one of two fates: silencing of one copy by degenerative mutations or preservation of both copies via natural selection. Classically, preservation is thought to occur by one of the copies acquiring a beneficial mutation and novel function (neofunctionalization) (Ohno 1970; Ohta 1987; Walsh 1995). More recently, it has been suggested that preservation of duplicate genes could be achieved by degenerative yet complementary mutations in both copies (subfunctionalization), with the organism subsequently requiring both genes (Hughes 1994; Force et al. 1999). Yet another possibility is maintenance of duplicates through a beneficial increase in gene dosage (Kondrashov et al. 2002).

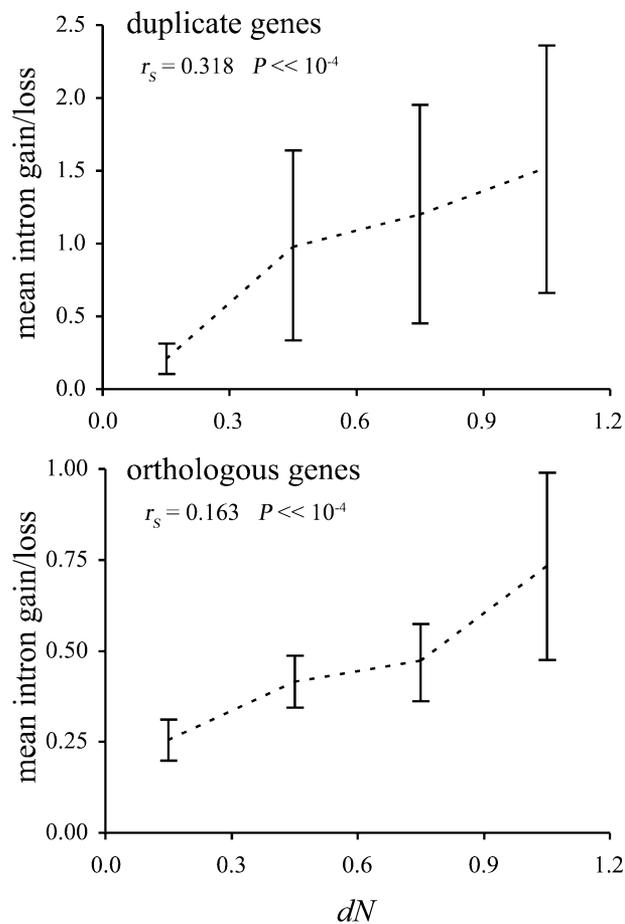


FIG. 3.—Positive correlation between protein evolution (dN) and intron gain/loss in orthologous genes and duplicate genes. Note that orthologous genes show a significant correlation between protein and intron gain/loss change even after correcting for the effect of age/local mutation rate, but duplicate genes do not (table 1).

Although it is not possible to differentiate among these models here, we note that rates of protein evolution and intron evolution both exhibit an approximate twofold increase after gene duplication. This result suggests that rates of protein evolution and intron evolution are related, such that a relaxation of selective constraint and/or positive selection acts on both aspects of gene structure. Indeed, among both duplicate and orthologous genes, the rate of intron gain/loss in a given gene is significantly correlated with its rate of protein evolution ($r_s = 0.163$, $r_s = 0.318$, $P \ll 10^{-4}$ for orthologs and duplicates, respectively; Spearman rank correlation, corrected for ties [fig. 3]). Thus, genes that evolve slowly are more likely to show low rates of intron gain and loss. Conversely, genes that evolve quickly in protein sequence are more likely to have higher rates of intron gain/loss. Notably, this result holds both for orthologous genes between *P. falciparum* and *P. y. yoelii* and for duplicate genes within each *Plasmodium* genome ($r_s = 0.132$ and $r_s = 0.264$, respectively; $P \ll 10^{-4}$ for both versus orthologs).

Because similarities in local mutation rates, or similar divergence times in the case of duplicates, may lead to the observed correlation between protein coding and intron

Table 1
Multiple Regression Analysis of dN on Intron Gain/Loss and dS , Among Orthologs and Duplicate Genes

	Variable	Coefficient	Standard Error	t -value	$Pr(> t)$
Orthologs	dS	0.24385	0.03443	7.083	2.52e-12***
	intron gain/loss	0.06157	0.01124	5.479	5.32e-8***
Duplicates	dS	1.085473	0.061151	17.751	2e-16***
	intron gain/loss	0.009647	0.015147	0.637	0.525

NOTE.—Formula: $dN \sim dS + \text{intron gain/loss}$.

gain/loss, we carried out multiple regressions involving dN , intron gain/loss, and dS using dS as a simple measure of age/mutation rate in both orthologs and duplicates. In duplicates, we found that the correlation between protein (dN) evolution and intron gain/loss was a result of their correlation with dS alone; dN and intron gain/loss increase together over time but are not themselves related (table 1). In contrast, orthologs continue to exhibit a significant correlation between protein and regulatory evolution even after controlling for the possibility that this correlation is a consequence of dS —a similarity in local mutation rates (table 1). A similar result has been found among orthologous and duplicate genes in nematodes for coupling between protein and upstream regulatory change (Castillo-Davis et al. 2004).

The observation that protein change and intron gain/loss in duplicates is not coupled in duplicates implies that these aspects of gene structure may evolve independently. Such independence is not unexpected, because both the neofunctionalization and the subfunctionalization hypotheses predict changes in duplicate-gene protein function, regulatory control, or both. It is possible that intron gain/loss and coding sequence change occur asymmetrically between duplicate genes; for example, accelerated intron gain/loss in one copy but no protein change or accelerated protein change in the other copy but no intron gain/loss. Although there is some evidence for differences in rates of functional diversification and protein change among young duplicate pairs in yeast and human, respectively (Wagner 2002; Zhang, Gu, and Li 2003), the proportion of functional divergence events among duplicate genes that occurs because of changes in different aspects of duplicate-gene structure is currently not known. In contrast, there is no evidence that evolution proceeds asymmetrically among orthologous genes.

A correlation between intron gain/loss and protein evolution in orthologs is not entirely unexpected, as it has been recently shown that rates of upstream *cis*-regulatory evolution and protein evolution are similarly weakly coupled in nematodes (Castillo-Davis et al. 2004). Because many spliceosomal introns play critical roles in eukaryotic gene regulation, for example, acting as transcriptional enhancers or silencers (Fedorova and Fedorov 2003) or controlling posttranscriptional mRNA export from the nucleus (Zhou et al. 2000; Maniatis and Reed 2002), their gain or loss, presumably resulting in a change in regulation, may be similarly coupled to protein change.

Because errors in gene prediction may result in a spurious relationship between dN and intron/gain loss, we reanalyzed the data using only genes for which there was evidence of transcriptional expression in *P. falciparum* as assessed by significant matches to ESTs and significant expression based on Affymetrix microarray data (Le Roch et al. 2003 [see *Methods*]). Using only these genes in our data set, we found the relationship between dN and intron/gain loss in orthologs and paralogs did not change ($r_s = 0.130$ and $r_s = 0.327$ for orthologs and *P. falciparum* duplicates, respectively; $P < 0.0005$ for both).

Given that evolutionary changes do not occur strictly asymmetrically among orthologs, the observed relationship between exon-intron structure and protein sequence over evolutionary time in orthologous genes suggests a functional linkage between these two aspects of gene structure. If relaxed selection is responsible for this pattern, we may deduce that the degradation of gene function by changes in amino acid sequence and intron gain/loss have similar fitness consequences, because they proceed similarly over time. On the other hand, if positive selection is driving protein change and intron gain/loss evolution, then it would appear that, in some cases, changes in gene function vis-a-vis protein divergence require (or are enhanced by) changes in intron gain or loss or vice-versa. In either case, the observation that multiple aspects of gene structure and function are evolutionarily related lends support to the hypothesis that selection acts on genes as integrated units (Castillo-Davis et al. 2004).

In contrast, a genome-wide comparison of rates of intronic and synonymous codon substitution in duplicate genes in both genomes indicates that intronic and synonymous codon substitution rates are not significantly different from each other (slope for combined data = 0.93, 95% CI [0.77, 1.10], $n = 67$; $slope_{fal} = 0.93$, 95% CI [0.78, 1.07], $n = 33$; and $slope_{yoel} = 0.95$, 95% CI [0.62, 1.27], $n = 34$; $P \ll 10^{-4}$ for all [fig. 4]). Further, after correcting for duplicate age (dS) by multiple regression, we observe no correlation between rates of intronic nucleotide substitution and rates of intron gain/loss in duplicate genes in either the *P. falciparum* or *P. y. yoelii* genomes or between intron nucleotide substitution rates and protein change (data not shown). Thus, whereas intron gain/loss is accelerated in duplicate genes, intronic nucleotide substitution is not, suggesting that most intronic sites are selectively neutral and not subject to either functional deterioration or adaptive evolution.

This result stands in contrast to those of Jongwutiwes et al. (2002), in which large differences in the level of polymorphism of intronic and synonymous sites were found in the genes *MSP4* and *MSP5* in *P. falciparum*. The low, population-level intronic site diversity and high synonymous site diversity in these genes was interpreted as evidence that introns in *P. falciparum* are under selection related to AT content. However, it is likely that this result represents differences unique to *MSP4* and *MSP5*, as it is not observed across the genome as a whole. Our results suggest that, for the purposes of population genetic studies of *P. falciparum*, intronic sequences and fourfold synonymous sites may be treated as approximately neutrally evolving.

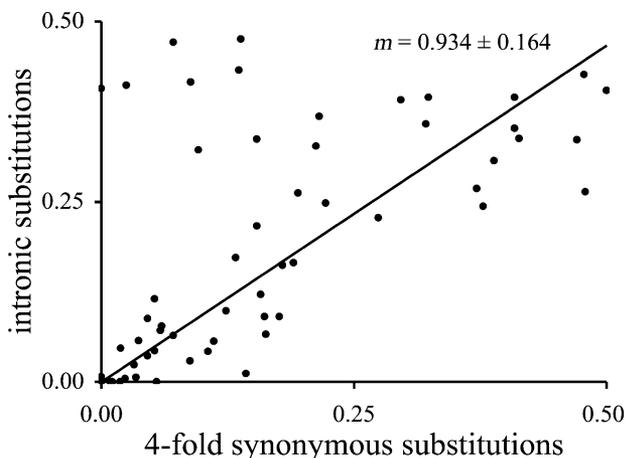


FIG. 4.—Nucleotide substitution counts in introns and fourfold synonymous sites in unsaturated ($dS < 1.0$) duplicate genes in both *P. falciparum* and *P. yoelii*. The ratio of intronic divergence to fourfold synonymous divergence does not differ from 1 in both species (slope for combined data = 0.93, 95% CI [0.77, 1.10], $n = 67$; $slope_{fal} = 0.93$, 95% CI [0.78, 1.07], $n = 33$; and $slope_{yoel} = 0.95$, 95% CI [0.62, 1.27], $n = 34$; $P \ll 10^{-4}$ for all).

Conclusion

In summary, intron gain/loss and protein evolution is dramatically accelerated in duplicate genes in both *P. falciparum* and *P. y. yoelii* because of either relaxed selection or positive selection or both. Additionally, rates of protein divergence and intron gain/loss are correlated over evolutionary time after speciation but not necessarily gene duplication. This suggests a functional linkage between these two aspects of gene structure that may have important implications for how adaptation proceeds in *Plasmodium*. Although it remains to be seen whether the acceleration of intron gain/loss in duplicate genes is unique to *Plasmodium*, it seems likely that selection on coding sequences, intron-exon structure, and upstream regulatory sequences are closely related in eukaryotes. It remains to be seen how far this emerging picture of genes as integrated selective units will extend.

Acknowledgments

We would like to thank all members of the Hartl lab for lively discussion and the Bauer Center for Genomics Research at Harvard University for computational resources. This work was supported by NIH grant GM61351 and by grants from the Ellison Medical Foundation. DLH is an Ellison Medical Foundation Senior Scholar in Global Infectious Disease.

Literature Cited

- Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**:3389–3402.
- Bozdech, Z., M. Llinás, B. L. Pulliam, E. D. Wong, J. Zhu, and J. L. DeRisi. 2003. The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS Biol.* **1**:001–016.

- Calderwood, M. S., L. Gannoun-Zaki, T. E. Wellems, and K. W. Deitsch. 2003. *Plasmodium falciparum* var genes are regulated by two regions with separate promoters, one upstream of the coding region and a second within the intron. *J. Biol. Chem.* **278**:34125–34132.
- Carlton, J. M., S. V. Angiuoli, B. B. Suh et al. (41 co-authors). 2002. Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii*. *Nature* **419**:512–519.
- Castillo-Davis, C. I., D. L. Hartl, and G. Achaz. 2004. *cis*-regulatory and protein evolution in orthologous and duplicate genes (submitted).
- Castillo-Davis, C. I., S. L. Mekhedov, D. L. Hartl, E. V. Koonin, and F. A. Kondrashov. 2002. Selection for short introns in highly expressed genes. *Nat. Genet.* **31**:415–418.
- Fedorova, L., and A. Fedorov. 2003. Introns in gene evolution. *Genetica* **118**:123–131.
- Gardner, M. J., N. Hall, E. Fung et al. (42 co-authors). 2002. Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* **419**:498–511.
- Goldman, N., and Z. Yang. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* **11**:725–726.
- Horrocks, P., K. Dechering, and M. Lanzer. 1998. Control of gene expression in *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **95**:171–181.
- Hughes, A. L. 1994. The evolution of functionally novel proteins after gene duplication. *Proc. R. Soc. Lond. B Biol. Sci.* **256**:119–124.
- Force, A., M. Lynch, F. B. Pickett, A. Amores, Y. L. Yan, and J. Postlethwait. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**:1531–1545.
- Jongwutiwes, S., C. Putaporntip, R. Friedman, and A. L. Hughes. 2002. The extent of nucleotide polymorphism is highly variable across a 3-kb region on *Plasmodium falciparum* chromosome 2. *Mol. Biol. Evol.* **19**:1585–1590.
- Kent, W. J., and A. M. Zahler. 2000. Conservation, regulation, synteny, and introns in a large-scale *C. briggsae*-*C. elegans* genomic alignment. *Genome Res.* **10**:1115–1125.
- Kondrashov, F. A., I. B. Rogozin, Y. I. Wolf, and E. V. Koonin. 2002. Selection in the evolution of gene duplications. *Genome Biol.* **3**:research 0008.1–0008.9.
- Le Roch, K. G., Y. Zhou, P. L. Blair et al. (8 co-authors). 2003. Discover of gene function by expression profiling of the malaria parasite life cycle. *Science* **301**:1503–1508.
- Llopart, A., J. M. Comeron, F. G. Brunet, D. Lachaise, and M. Long. 2002. Intron presence-absence polymorphism in *Drosophila* driven by positive Darwinian selection. *Proc. Natl. Acad. Sci. USA* **99**:8121–8126.
- Lynch, M., and J. S. Conery. 2000. The evolutionary fate and consequences of duplicate genes. *Science* **290**:1151–1155.
- Maniatis, T., and R. Reed. 2002. An extensive network of coupling among gene expression machines. *Nature* **416**:499–506.
- Moriyama, E. N., D. A. Petrov, and D. L. Hartl. 1998. Genome size and intron size in *Drosophila*. *Mol. Biol. Evol.* **15**:770–773.
- Muse, S. V., and B. S. Gaut. 1994. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Mol. Biol. Evol.* **11**:715–724.
- Nembaware, V., K. Crum, J. Kelso, and C. Seoighe. 2002. Impact of the presence of paralogs on sequence divergence in a set of mouse-human orthologs. *Genome Res.* **12**:1370–1376.
- Ohno, S. 1970. *Evolution by gene duplication*. Springer-Verlag, Heidelberg.
- Ohta, T. 1987. Simulating evolution by gene duplication. *Genetics* **115**:207–213.
- Patankar, S., A. Munasinghe, A. Shoaibi, L. M. Cummings, and D. F. Wirth. 2001. Serial analysis of gene expression in *Plasmodium falciparum* reveals the global expression profile of erythrocytic stages and the presence of anti-sense transcripts in the malaria parasite. *Mol. Cell* **12**:3114–3125.
- Perkins, S. L., and J. J. Schall. 2002. A molecular phylogeny of malarial parasites recovered from cytochrome b gene sequences. *J. Parasitol.* **88**:972–978.
- Rogozin I. B., Y. I. Wolf, A. V. Sorokin, B. G. Mirkin, and E. V. Koonin. 2003. Remarkable interkingdom conservation of intron positions and massive, lineage-specific intron loss and gain in eukaryotic evolution. *Curr Biol.* **13**:1512–1517.
- Roy, S. W., A. Fedorov, and W. Gilbert. 2003. Large-scale comparison of intron positions in mammalian genes show intron loss but no gain. *Proc. Natl. Acad. Sci. USA* **99**:984–989.
- Tarrío, R., F. Rodríguez-Trelles, and F. J. Ayala. 1998. New *Drosophila* introns originate by duplication. *Proc. Natl. Acad. Sci. USA* **95**:1658–1662.
- Tatusov, R. L., E. V. Koonin, and D. J. Lipman. 1997. A genomic perspective on protein families. *Science* **278**:631–637.
- Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**:4673–4680.
- Wagner, A. 2002. Asymmetric functional divergence of duplicate genes in yeast. *Mol. Biol. Evol.* **19**:1760–1768.
- Walsh, J. B. 1995. How often do duplicated genes evolve new functions? *Genetics* **139**:421–428.
- Yang, Z. 2000. Maximum likelihood estimation on large phylogenies and analysis of adaptive evolution in human influenza virus. *A. J. Mol. Evol.* **51**:423–432.
- Yang, Z., R. Nielsen, N. Goldman, and A.-M. K. Pedersen. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155**:431–449.
- Zhang, P., Z. Gu, and W.-H. Li. 2003. Different evolutionary patterns between young duplicate genes in the human genome. *Genome Biol.* **4**:R56.
- Zhou, Z., M. J. Luo, K. Straesser, J. Katahira, E. Hurt, and R. Reed. 2000. The protein Aly links pre-messenger-RNA splicing to nuclear export in metazoans. *Nature* **407**:401–405.

Geoffrey McFadden, Associate Editor

Accepted March 22, 2004