1166–1173 *Nucleic Acids Research, 2006, Vol. 34, No. 4*
doi:10.1093/nar/gkj517

Comparative whole genome transcriptome analysis of three *Plasmodium falciparum* strains

Manuel Llinás, Zbynek Bozdech[1], Edith D. Wong[2], Alex T. Adai[2] and Joseph L. DeRisi[2,*]

Department of Molecular Biology, Lewis-Sigler Institute for Integrative Genomics, Princeton University, 246 Carl Icahn Laboratory, Princeton NJ 08544, USA, [1]School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore 637551 and [2]Department of Biochemistry and Biophysics, University of California, San Francisco, 1700 4th Street, San Francisco, CA 94143-2542, USA

Received October 28, 2005; Revised December 3, 2005; Accepted January 13, 2006

ABSTRACT

Gene expression patterns have been demonstrated to be highly variable between similar cell types, for example lab strains and wild strains of *Saccharomyces cerevisiae* cultured under identical growth conditions exhibit a wide range of expression differences. We have used a genome-wide approach to characterize transcriptional differences between strains of *Plasmodium falciparum* by characterizing the transcriptome of the 48 h intraerythrocytic developmental cycle (IDC) for two strains, 3D7 and Dd2 and compared these results to our prior work using the HB3 strain. These three strains originate from geographically diverse locations and possess distinct drug sensitivity phenotypes. Our goal was to identify transcriptional differences related to phenotypic properties of these strains including immune evasion and drug sensitivity. We find that the highly streamlined transcriptome is remarkably well conserved among all three strains, and differences in gene expression occur mainly in genes coding for surface antigens involved in parasite–host interactions. Our analysis also detects several transcripts that are unique to individual strains as well as identifying large chromosomal deletions and highly polymorphic regions across strains. The majority of these genes are uncharacterized and have no homology to other species. These tractable transcriptional differences provide important phenotypes for these otherwise highly related strains of *Plasmodium*.

INTRODUCTION

Malaria is one of the most prevalent afflictions of humankind, affecting upwards of half a billion people and killing over 1.5 million persons annually. This disease is caused by the *Plasmodium* parasite, of which *Plasmodium falciparum* is the deadliest species. The lifecycle of *P.falciparum* is comprised of three major developmental stages, the mosquito, liver and intraerythrocytic stages. Upon red blood cell invasion, the parasite undergoes a complex series of morphological transitions over the next 48 h, ultimately resulting in a new population of mature merozoites, which reinitiate the intraerythrocytic developmental cycle (IDC). While there have been major global efforts to eliminate this disease over the past several decades, it continues to persist as a major health burden, especially to the poorest of nations (1).

Malaria continues to be a major worldwide health issue because resistant strains of *P.falciparum* have evolved against virtually every known anti-malarial drug (2). However, the underlying mechanisms of *Plasmodium* drug resistance remain poorly understood. In part, resistance is due to polymorphisms in a variety of genes including the plasmodium chloroquine resistance transporter—Pfcrt (chloroquine resistance) (3), dihydrofolate reductase (pyrimethamine) (4), dihydropteroate synthase (sulfa drugs) (4) and cytochrome b (atovaquone) (5). However, these polymorphisms are not sufficient to completely describe drug resistance in *P.falciparum*, because often their effects vary with strain genotype and drug concentration (6). Thus, while polymorphisms can confer drug resistance, different strains of *P.falciparum* may also rely on alternative mechanisms of gene regulation to overcome the selective pressures of anti-malarial drugs. These mechanisms may include overexpression of transporters in response to drug exposure, induction of alternative pathways to bypass the effect of the compounds or repression of genes that exacerbate deleterious effects of the drug. For example, recent data suggests that mefloquine resistance are directly modulated by an increased copy number of the *Plasmodium* multi-drug resistance gene *pfmdr1*, resulting in an amplification in the number of cellular transcripts produced (7). Therefore, regulatory responses in *P.falciparum* may act

*To whom correspondence should be addressed. Tel: +1 415 476 4132; Fax: +1 415 514 4140; Email: joe@derisilab.ucsf.edu

alone or in concert with previously identified polymorphisms to favor parasite survival under drug pressure.

The completion of the full genome sequence of the 3D7 strain of *P.falciparum* (8) has ushered in exciting new possibilities in malaria research using whole-genome gene expression profiling methods (9). We recently demonstrated for the HB3 strain of *P.falciparum* that a continuous cascade of gene expression guides *Plasmodium* development throughout the IDC, from erythrocyte invasion to merozoite release. This cascade is highly streamlined for efficient progression toward parasite replication, with most genes being turned on once and only once with peak expression correlating with the function of the gene products (10) similar to a 'just in time' production process. The apparent lack of complexity in this progression, coupled with the paucity of predicted transcription factors found in the genome (8) suggests that the IDC gene expression program in *P.falciparum* may be governed by a relatively small set of factors that ensure its fidelity and precision over many cycles. This is further supported by the identification of only 156 transcription-associated proteins in the *P.falciparum* genome (two-thirds less by genome size than in other eukaryotes), with an over-representation of proteins containing a CCCH-type zinc finger found in single strand RNA binding proteins involved in posttranscriptional control mechanisms (11).

Given the high degree of conservation in both the number and functional similarities of genes present in the genomes of related species, the observed differences in protein coding regions alone are not sufficient to explain the degree of diversity seen among species. Clearly, phenotypic diversity relies on variation in gene expression as well (12). Gene expression-based variation has been well demonstrated in organisms from yeast to humans using whole-genome approaches (13–16). In some cases, this variation has been shown to be heritable and has been exploited for use as quantitative trait loci in genetic association studies. In the budding yeast *Saccharomyces cerevisiae*, for example, it is remarkable that two strains grown under identical culture conditions result in variable gene expression for nearly 50% of the measurable genes (16). Similarly, we were interested in determining the likelihood that changes in gene expression may play a role in phenotypic variation among strains of *P.falciparum*, including the evolution of drug resistance. We hypothesized that by characterizing the IDC transcriptome of several *P.falciparum* strains with different drug resistances, we could identify strain-specific changes in gene expression and gain insight into the underlying phenotypic strain differences.

To investigate the basis for strain-specific phenotypes and regulatory mechanisms, we analyzed the complete IDC transcriptomes of the 3D7 and Dd2 strains of *P.falciparum* by DNA microarray analysis. The results are compared to our prior analysis of the HB3 strain transcriptome (10). The three isolates we chose originate from distant geographical regions and represent distinct drug sensitivity phenotypes. 3D7, HB3 and Dd2 were isolated from The Netherlands, Honduras and Indochina, respectively. These strains are resistant and/or sensitive to various common anti-malarial chemotherapeutics. HB3 and 3D7 are both sensitive to chloroquine, although 3D7 is sulfadoxine resistant and HB3 is pyrimethamine resistant (17), while Dd2 is chloroquine and mefloquine resistant. Despite these geographical and phenotypic differences, our results indicate a remarkable conservation of the transcriptional program between these three strains of *P.falciparum*. Significant transcriptional differences are detected mostly for genes encoding surface antigens and suggest either chromosomal deletions or significant polymorphism between strains. In contrast to other microorganisms, our results are consistent with a simple, highly conserved streamlined cascade of gene expression.

MATERIALS AND METHODS

Culturing, DNA microarrays and northern blotting

P.falciparum 5 l fermentor preparation and culturing, RNA isolation, amino-allyl labeled cDNA preparation and DNA microarray printing were performed as described previously (10). In brief, a highly synchronized population of late-stage schizonts was attained by culturing parasites in 150 cm^2 tissue culture flasks and by standard synchronization with 5% sorbitol. Invasion of fresh red blood cells was initiated within the fermentor by adding cells harvested from 25–30 flasks at ~15% parasitaemia and allowing these parasites to reinvade at high hematocrit (~16%) followed by dilution to a total volume of 4.5 l after 2–3 h. The end of invasion established the time for timepoint one and thereafter, hourly samples were harvested and frozen for RNA extraction. A pool of amino-allyl labeled cDNAs from each of the timepoints used in the experiment was assembled for each strain and used as a reference. For DNA microarray hybridization, the pool cDNA was always coupled to Cy3 dye as reference, while cDNA from an individual timepoint was coupled to Cy5 dye.

Northern blots were performed as described (18). To generate our 597 bp DNA probe for the MAL13P1.344 northern blot, we used the following two primers for PCR: 5′-GAATTGTCCTATAGTAAAAACAGG-3′ and 5′-GCAAATCTTTGTAATTCTCCTCCAC-3′.

The PCR product was labeled with [α-^{32}P]dCTP using the Invitrogen™ Random Primers DNA Labeling System as per the manufacturer's instructions. RNA used for the northern blot was purified from sorbitol-synchronized parasites isolated at the ring and schizont stages of development.

Data analysis

For strains 3D7 and Dd2, 60 and 58 DNA microarray hybridizations were performed, respectively. For each strain we collected as many timepoints as possible. The 3D7 dataset were comprised of 53 hourly timepoints with hours 3, 15, 27, 30, 32, 35 and 39 represented by more than one array hybridization. The Dd2 dataset were comprised of 50 timepoints and is missing timepoint 8 with hours 4, 7, 12, 14, 20, 25, 28, 30 and 47 represented by more than one array hybridization. DNA microarrays were scanned using an Axon 4000B scanner and images analyzed using Axon GenePix software (Axon Instruments, Union City, CA, USA). Microarray data were stored using our NOMAD database (http://ucsf-nomad.sourceforge.net/) and data for individual arrays were normalized by a global normalization using unflagged features with a regression correlation coefficient greater than or equal to 0.75 and median intensity value greater than zero. All unflagged spots were selected and extracted for further

analysis (Supplementary Tables S1 and S2). For timepoints with multiple microarrays, ratio values were averaged.

Fast Fourier transform (FFT) of data

To analyze the three strains simultaneously, we required that for each oligonucleotide (oligo) measured by the microarray, data must be present in at least 60% of the timepoints: for 3D7 (31 of 53), Dd2 (29 of 50), HB3 (27 of 46). These comprise the quality control datasets (Supplementary Tables S3–S5). The filtered datasets were then normalized, log2 transformed and smoothed using the LOESS function from the statistical package R (http://www.R-project.org) with weighting restricted at 30%. Missing data were estimated based on nearest neighbors using the PREDICT function. Fast Fourier analysis was performed as described previously (10) on the continuous smoothed dataset. For all strain datasets FFT phase values ranged from $-\pi$ to $+\pi$. For FFT phase comparison between strains, both the absolute value of the difference between the two phases ABS(phase1-phase2) and 2π-ABS(phase1-phase2) were calculated. The minimum difference was then used as the actual phase difference. This was necessary because expression profiles with FFT phase values that are phase shifted by 2π are mathematically identical; a phase of $-\pi$ is equivalent to a phase of $+\pi$. Two exactly identical expression profiles therefore have a phase difference of zero whereas two completely out of phase profiles have a phase difference of π.

Pearson correlations of data

Pearson correlations were determined by analyzing the smoothed data from above (FFT analysis) using the COR function from R for each pairwise combination of oligo probes. We first determined the global median shift value for a given set of strains by calculating the Pearson correlation of every probe to every other probe along a sliding timescale window starting with the first timepoint (of one strain probe) and the final timepoint (of the second strain probe) until all possible pairwise correlations were computed. The difference in timepoints resulting in the best inter-strain correlation for each probe was noted and the global median value of this difference was then used to shift the profiles in the longer dataset. Finally, the longer profiles were trimmed to match the shorter profile before calculating the pairwise Pearson correlation.

RESULTS

Overview of the data

To identify potential inter-strain variability in the transcriptional program of different *P.falciparum* strains we characterized the complete IDC transcriptome for the Dd2 and 3D7 strains and compared them to our previous study of HB3 (10). To measure the IDC transcriptomes of these strains, we used a highly synchronous large-scale culturing methodology featuring sampling every hour throughout the 48 h IDC. For each strain, gene expression was measured through one cycle of the IDC using our custom oligo DNA microarray consisting of 7470 spotted elements [described previously (10)]. Time courses were measured for 53, 50 and 48 h, respectively for 3D7, Dd2 and HB3. Gene expression profiles from individual oligos on the array were excluded from further analysis if data were lacking for more than 40% of the total time course (Supplementary Tables S3–S5). The resulting dataset consisted of 6287 gene expression profiles for 3D7, 5294 profiles for Dd2, and 6415 profiles for HB3. To directly compare the previous HB3 strain data with the new results, we re-examined all of the data by the following two different approaches.

Transcription profiles are highly conserved

Variation in the gene expression profiles among the three IDCs manifested both in changes in amplitude and in the time of maximal expression. We utilized two methodologies to ascertain the concordance of individual expression profiles for similar probes between all possible strain pairs. First, given the inherent periodicity in the data, we used a FFT to calculate the relative timing (phase) and amplitude modulation in the expression profiles for all genes in the three strains (see Materials and Methods). Most of the FFT results demonstrate very few frequency components contributing to the measured expression profile signal and we used this property of the data to calculate a percent power score for all profiles (10). This score, which is equivalent to the percentage of the total power spectrum present in the two primary FFT frequency components, reflects the global periodic nature of the expression profiles: genes with a higher percent score are expressed with a period nearing one during the timescale of the IDC ($\sim$48 h).

An overview of the highest quality data consisting of all profiles with power scores greater than 75% ordered by FFT phase are shown in Figure 1 for the three strains (Supplementary Tables S6–S8). The transcriptional program of the *P.falciparum* IDC is remarkably conserved. The majority of (>75% of all probes) gene expression profiles were measured in all three strains as depicted in Table 1. Genes known to belong to the functional categories glycolysis, cytoplasmic transcription, translation, the tricarboxylic acid cycle, the proteasome and plastid biosynthesis are tightly correlated in their gene expression profiles as well as in the timing of maximal expression during the IDC. We do note that a global phase shift in gene expression for each strain is apparent due to the initiation point of each time course. As observed in the HB3 strain, the vast majority of the gene transcripts in our datasets for the 3D7 and Dd2 strains are expressed in a highly periodic manner during the IDC (7). The maxima of gene induction occur in a logical order consistent with cellular functions relevant to the progression of the parasitic life cycle.

The FFT phase we measure for any periodic wave is always within the range of $-\pi$ to π and reflects the timing of periodic gene expression. By measuring the difference in the calculated FFT phase for each pair of oligos between the strains, an estimate of global shift and the true inter-strain variation was analyzed (see Materials and Methods). As expected, the calculated FFT phases do not match exactly, but rather are distributed about a global experimental phase shift, with outliers representing significant inter-strain differences in gene expression (Figure 2A). Between HB3 and 3D7, the median phase shift is relatively small at 0.47 or 3/20*π. If 2π represents an $\sim$48 h IDC, then this is equivalent to an average timing difference of 3.6 h. Surprisingly, between these two strains we measure a phase difference greater than $\pi/2$ ($\sim$12 h) for probes from only 69 open reading frames (ORFs). However, many of

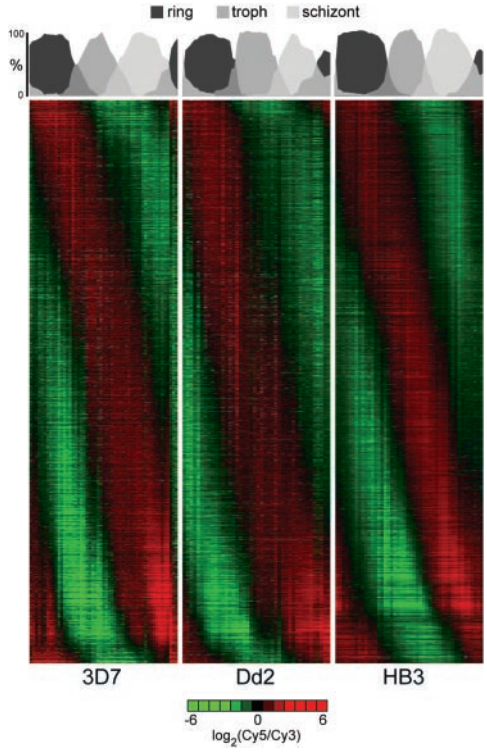

Figure 1. IDC Transcriptomes for three *P.falciparum* strains. The phaseograms depict the 48 h progression of gene expression throughout the IDC from the early ring-stage to merozoite invasion (left to right) for the 3D7, Dd2 and HB3 strains of *P.falciparum*. The profiles have been ordered (top to bottom) by the phase of gene expression during the IDC and demonstrate a highly similar cascade across the three strains. Depicted above the phaseograms (in gray scale) is the relative percentage of the three morphological stages: ring, trophozoite (troph) and schizont.

Table 1. Total number of shared and unique oligo probes measured in the three strains, with the percent of the total number of measured probes for each strain

Oligo group		3D7	Dd2	HB3
Total measured probes		6287	5294	6415
All three strains	5193	82.5%	98.1%	81.0%
HB3 and 3D7 only	739	11.8%	—	11.5%
Dd2 and HB3 only	35	—	<1.0%	<1.0%
Dd2 and 3D7 only	62	1.0%	1.2%	—
HB3 only	448	—	—	7.0%
3D7 only	296	4.7%	—	—
Dd2 only	4	—	<1.0%	—

these phase differences result from spurious FFT phase calculations due to noisy gene expression profiles that derive from low intensity data and may not reflect true differences. Overall, the majority of profiles are very similar by phase and do not suggest widespread variations in gene expression between strains. The expression profiles of the HB3 and Dd2 strains demonstrate the least difference in phase, while HB3 and 3D7 are the most distant.

Since the FFT calculation is most useful for comparing high amplitude gene expression profiles, an alternative approach was necessary to gauge profile similarity for all genes. We chose to use the Pearson correlation, since this measure is insensitive to overall amplitude differences. We calculated

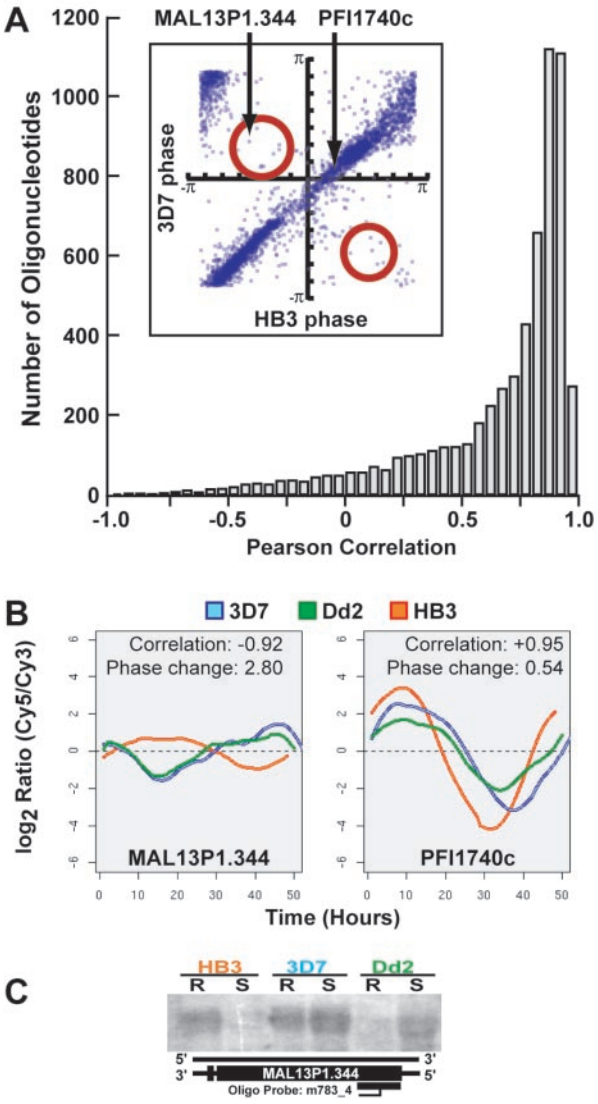

Figure 2. Inter-strain differences in gene expression are rare during the IDC. The differences in gene expression profiles as measured by Pearson correlation and the FFT phase are low between all strains measured here. (**A**) Depicts the distribution of the Pearson correlation between gene expression profiles from the 3D7 and HB3 datasets. 70% of the correlations are greater than 0.7. The inset plots the nearly linear relationship for most of the FFT phases for these two strains. The largest differences in phase are within the red circles. The arrows point to a gene with a significant change in phase between 3D7 and HB3 (MAL13P1.344) and another gene with virtually no difference (PFI1740c). The Pearson correlation and phase change between 3D7 and HB3 for these two genes are shown with the expression profiles from all three strains (**B**). Northern blot of MAL13P1.344 (**C**) shows mRNA transcript levels at the ring (R) and schizont (S) stages of development for the 3D7, Dd2 and HB3 strains of *P.falciparum*. Transcription of this gene in HB3 is out of phase with that measured in 3D7 and Dd2.

the correlation for all pairwise probe to probe comparisons between strains (see Materials and Methods) (Supplementary Table S9). The Pearson correlation for more than 70% of measured profiles is greater than 0.7. These high correlations demonstrate an excellent agreement in gene expression between strains and indicate a strong preservation of the overall timing of transcriptional regulation throughout the IDC (Figure 2A and B).

Variability in gene expression

One of the few differentially expressed genes that we observed was MAL13P1.344. Its expression profile was virtually identical in 3D7 and Dd2, but was nearly opposite in the HB3 strain (Figure 2B). This gene is a member of the superfamily of ATP-binding cassette transporters from the OABP subfamily and is an example of a gene which is expressed completely out of phase between strains. The time of maximal transcript abundance is shifted by ~22 h, with a Pearson correlation value of −0.92. The DNA microarray results for MAL13P1.344 have been confirmed by northern blot analysis using RNA from ring-stage and schizont-stage parasites (Figure 2C). Such anti-correlation is very rare in the datasets. Large changes in gene expression were also detected for PFC0110w—the cytoadherence linked asexual protein (CLAG 3.1), PF11_0512—ring-infected erythrocyte surface antigen 2 (RESA-2) and a pair of erythrocyte membrane protein 1 (PfEMP1)—PF08_0103 and PFB0010w. These genes all encode surface proteins likely to be involved in antigen presentation. In fact, we have examined the Gene Ontology (GO) parameters (19) for genes with the greatest difference in Pearson correlation between HB3 and 3D7 using the LACK tool (20), and we find a significant enrichment (P-value < 0.000005) for extracellular and plasma membrane-associated genes involved in cell adhesion, defense and immunity, antigenic variation and host-pathogen interaction all of which are properties of the *P.falciparum* cell surface (data not shown).

We also considered the cases where a gene expression profile from a particular oligo is detectable in at least one, but not all strains. Leaving basic technical microarray issues aside, this situation could only arise if there are true transcriptional differences or if deletions and significant polymorphisms are present between the strains. For every oligo, we summed the number of missing data points across the time courses and selected those whose difference in missing values was greater than 50% in a given strain pair, also requiring that the oligo passed our quality control criteria in at least one of the two strains. This analysis identified 556 such cases between HB3 and Dd2, 347 between Dd2 and 3D7 and 278 between HB3 and 3D7 (Supplementary Table S10). Over half of these cases occur in hypothetical genes. The remaining cases are almost solely comprised of surface antigens including *vars*, *var* pseudogenes, *rifins* and *stevors*.

Lack of observed expression in HB3 and Dd2 when compared to 3D7 can be explained in several instances by the presence of highly polymorphic loci within the genome. In some cases, such as merozoite surface protein 2 (MSP2, PFB0300c) and S-antigen (PF10_0343), the representative 70mer oligonucleotide lies within regions that are known to contain high levels of sequence diversity among strains (21,22). Other genes, such as knob associated histidine rich protein (KAHRP, PFB0100c) and erythrocyte membrane protein 3 (PfEMP3, PFB0095c), may be completely absent, as has been previously shown for Dd2 (23). However, while HB3 appears to contain PfEMP3 and KAHRP, these genes are likely to be highly polymorphic since previous CGH work failed to detect signal using a 70mer oligonucleotide (10).

Although the differential performance of some probes may be due to DNA polymorphisms as discussed above, 70mer oligos may tolerate slight sequence variability without compromising overall signal (24). Therefore, it is relevant to consider to what extent the inter-strain differences in expression may be explained by a high degree of polymorphism. CGH analysis of chromosomal variability between strains suggests that at most, 20% of these changes are due to deletions or polymorphisms in the genome [(10), E.D.W. and J.L.D. personal communication].

Over 1000 genes in the 3D7 genome are represented by more than one oligo on our *P.falciparum* DNA microarray. These probes are very useful for intragene probe validation. Furthermore, variability between these probes, when observed between strains, may provide evidence for deletions and/or insertions, or alternative splicing. Within a given strain, particularly for the reference strain 3D7, these differences may also simply illustrate the difficulty in annotating intron–exon boundaries. We determined the Pearson correlation for all pairs of oligos in genes represented by multiple oligos. For most genes, there is excellent intra-strain correlation (above 0.9 for over 70% of the data) in the expression profiles measured by any pair of oligos (Supplementary Table S11). We also calculated the maximal difference in Pearson correlation between all pairs of oligos for each ORF and remarkably, the most dissimilar pair for any given strain was 80% likely to be dissimilar in another strain. For oligo pairs that do not behave commonly across strains, these differences suggest either sequence variations or potential alternative splicing events.

Transcript abundance

While the DNA microarrays used here do not quantify RNA copy number within a cell directly, we can qualitatively estimate the relative abundance of each transcript throughout the IDC by examining the intensity signals for individual microarray probes. We calculated the average sum of the median intensity (AVG SOM) for each individual oligo throughout the IDC for all three strains and note a remarkable correlation (Supplementary Table S12). The 150 oligos with the highest AVG SOM from 3D7 are in the top 5% of AVG SOMs (372 oligos) of both HB3 and Dd2, which strongly suggests that the most highly abundant transcripts are common to all three strains. These results were independent of hybridization effects due to oligo probe properties as demonstrated by the lack of significant correlation between the AVG SOM values and the free energy, complexity score, G-C content or self-binding energy of the microarray probes (data not shown).

Unlike the majority of abundant transcripts that are in common across the three strains, a small subset of these high intensity genes is unique to one strain. For 3D7 these include: KAHRP, S-antigen, PfEMP3, a sexual stage specific antigen PFD0310w, MSP 2 and CLAG3.2. As described above, many of these genes are either deleted or highly polymorphic in the Dd2 and HB3 strains. There are also highly abundant transcripts unique to the HB3 strain including PFI1750c, PF11_0038 and PF14_0733 which encode predicted hypothetical proteins. In addition, a large ~20 kb region on chromosome 4 (134 000–153 500 bp) near PFD0100c encodes several over-represented transcripts in the HB3 strain, including a putative reticulocyte binding protein, PFD0110w. The presence of an established reticulocyte binding protein PFD0110w

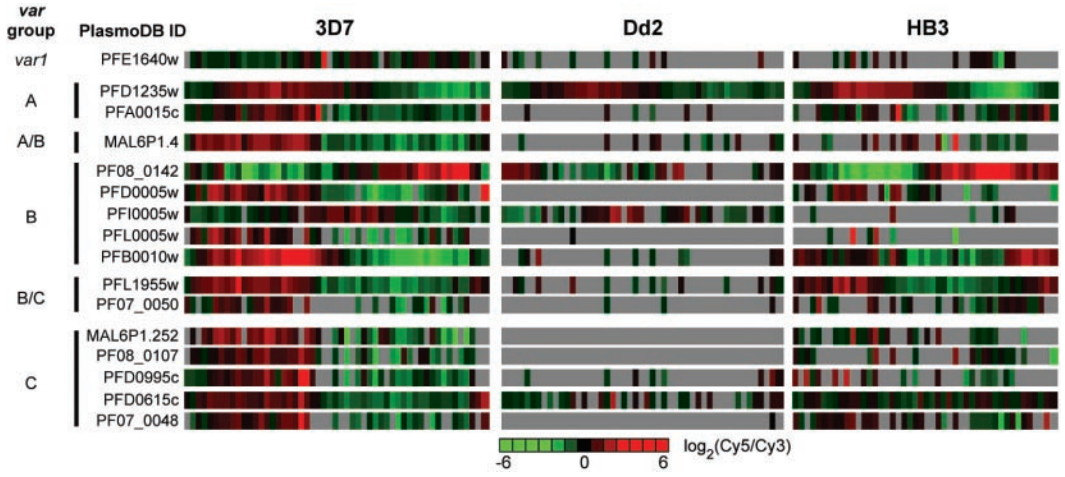

Figure 3. Multiple *var* genes are expressed during the IDC. The gene expression profiles of all *var* genes measured for the 3D7 strain along with the respective profiles from Dd2 and HB3 are displayed. The genes are organized according to *var* gene group as defined by Lavstsen *et al*. (30).

in this region suggests that HB3 may utilize different surface proteins during merozoite invasion.

To determine whether transcripts with high AVG SOM values actually represent the most abundant gene products within the cell, we compared the most abundant transcripts in our study to the most highly identified proteins in two proteomics studies performed on selected stages of the IDC (25,26). Of the 100 most abundant proteins detected by mass spectrometry (as determined by peptide count in red blood cells stages), 50% are in the 200 most abundant transcripts by our microarray analysis, while 75% are in the top 500 transcripts. Le Roch *et al*. also reported a concordance between mRNA transcription and protein expression and noted a timing lag between these two processes (27). This demonstrates that there is a strong correlation between transcription levels and translation for *P.falciparum*.

Putative deleted, polymorphic and silenced regions

Based on our transcriptome analysis, we identified several compelling chromosomal regions for further characterization. On chromosome 4 a 20 kb region containing two proteins with homology to reticulocyte binding proteins (PFD1150c and PFD1155w) and a hypothetical gene (PFD1160w) is not expressed and may be silenced in HB3 and Dd2, but contains several genes that are expressed during the 3D7 IDC. This is not due to a chromosomal deletion, since these regions are present by CGH analysis (E.D.W. and J.L.D., personal communication). Furthermore, PFD1150c (PfRh4) has recently been demonstrated to be required for switching from sialic acid dependent to independent modes of invasion (28). Thus, the observed difference in gene expression suggests that transcriptional regulation varies at this locus between strains and is important for invasion. Another chromosomal region with large detectable differences in gene expression is the first 100 kb of chromosome 2, a region known to be polymorphic and associated with cytoadherence (23). In this region, we solely detected gene expression for four rifin genes, a highly expressed PfEMP3, several hypothetical genes, and a PfEMP1 gene in the 3D7 strain.

Immune evasion: *var*, *stevor, rifin* genes

P.falciparum modulates host immune response by switching on a vast repertoire of surface proteins in successive parasite generations (29). The three major families of variant surface proteins are the *var* genes (PfEMP1 proteins), the repetitive interspersed gene family (*rifins*) and the subtelomeric variant ORFs (*stevors*). Our microarray contains probes for 50 of the 59 *vars*, 28 of the 29 *stevors* and 141 of the 154 predicted *rifins*. It is well established that *P.falciparum* expresses only a subset of *var* genes per IDC, which are replaced by other *var* genes in later generations (29). In our analysis of the 3D7 strain we detected 16 *var* genes. Overall, the maximal expression of these genes occurred during the late ring and early trophozoite stages and was synchronized with expression of KAHRP and PfEMP3, the structural elements of knob protrusions (Figure 3), with the exception of PF08_0142, which intriguingly showed maximal expression at the late trophozoite-early schizont-stage. These transcripts represent each of the five major groups of *var* genes, suggesting that at least one member of each category is expressed per IDC (30). In a prior 3D7 transcriptome study, a completely different set of *var* genes was detected indicating that even lab-adapted strains utilize different cohorts of these genes (31). We detect little to no expression for any *stevor* genes which supports evidence for a role in cytoadhesion in the gametocyte developmental stage (32). We measured gene expression for a small subset of *rifins* which peaks between 15–17 h post-invasion, supporting their previously reported role in adhesion (33). Therefore, even *in vitro* different sets of *P.falciparum* surface proteins are being expressed.

DISCUSSION

The malaria parasite *P.falciparum* has evolved and adapted to locations worldwide where the *Anopheles* mosquito survives and co-exists with humans, resulting in a wide range of inter-strain differences in virulence and drug resistance. Unfortunately, the basis for strain-specific phenotypic variability in *P.falciparum* remains poorly characterized with the exception

of a few genes such as the chloroquine resistance transporter gene *pfcrt* (3). Currently, there are few macroscopic phenotypes for differentiating between strains of *P.falciparum* and yet the identification of additional diagnostic markers could greatly aid the clinical prediction of malaria severity. This study expands our understanding of *P.falciparum* transcriptional regulation during the IDC by comparing the full transcriptomes of two commonly used laboratory strains 3D7 and Dd2 with that of HB3. The major differences between these three strains lie in their original geographical locations and divergent drug sensitivities. Our goal was to detect changes in gene expression patterns between strains and thereby identify pathways and mechanisms responsible for strain-specific differences. These changes can then be further exploited to identify regulatory regions contributing to strain adaptation.

This study demonstrates that the transcriptomes of the 3D7, Dd2 and HB3 strains of *P.falciparum* are strikingly similar in the overall gene expression patterns of the majority of genes during the IDC. This conservation implies that many key developmental processes are well preserved among strains. The streamlined nature of the transcriptome indicates that the mechanisms for control of gene expression may be quite simple for all strains.

Changes in gene expression between the strains were detected primarily at loci encoding surface antigens involved in immune evasion. These changes could be due to strain-specific deletions, polymorphisms or gene silencing. Sir2-dependent silencing has recently been demonstrated for *var* gene regulation in telomeric regions (34,35), and our data suggest that perhaps other regions of the genome may also be similarly silenced in a strain-specific manner. In addition, we found several differences in transcript abundance between oligos within genes represented by multiple oligos on our microarray, supporting the possibility for changes in gene structure or alternative splicing between strains.

We anticipated changes in expression for genes known to be involved in drug resistance. For example, chloroquine resistance is in part conferred by point mutations in the *pfcrt* transporter gene which prevents chloroquine from accumulating in the food vacuole (3). Similarly, polymorphisms in the P-glycoprotein (Pgh1) gene, *pfmdr1* or overexpression of this gene have both been shown to facilitate mefloquine resistance (36). Although up or down regulation of these genes could contribute to drug resistance, we detected no such changes. As previously observed, cells under favorable growth conditions *in vitro* may not be under any significant environmental pressure to alter the expression of these genes (37). In cultures containing chloroquine we have detected specific expression changes for the *pfcrt* transporter in 3D7 cells (chloroquine sensitive), although this response is clearly not sufficient for survival (M.L., manuscript in preparation).

It remains unclear as to the range of environmental signals to which *Plasmodium* is capable of responding. Since the parasite is sequestered within four membranes inside the red blood cell during most of the infection, it experiences a fairly homeostatic environment throughout development. In fact, a study examining *Plasmodium* transcripts from infected patients' blood found few genes that had not been previously detected by *in vitro* studies (38). The rigidity of the transcriptome is strikingly different from what has been seen for different *S.cerevisiae* strains (16). The *Plasmodium* parasite is highly adapted to its unique environment, outside of which it simply does not survive. The streamlined nature of the transcriptome gene expression cascade strongly supports this, with gene expression occurring in a 'just-in-time' fashion following a seemingly unchecked commitment to complete the IDC.

The strong conservation of the transcriptional program across various strains will be extremely useful in bioinformatic approaches that aim to strengthen the statistical significance of highly correlated gene expression profiles. The gene expression profile of any given gene during the IDC can now be treated with high confidence due to the reproducibility of the data across strains. In addition, our high resolution 3D7 transcriptome data are qualitatively in excellent agreement with the Le Roch *et al*. (31) 3D7 transcriptome study which used 6 IDC timepoint measurements. By harnessing similarities in gene expression levels, timing and uniqueness within a strain, our three strain transcriptome data will begin to allow for classification of the more than half of the *P.falciparum* genome with no homology to other species. Coupled with emerging sequence data from other *Plasmodia* species, this will certainly assist in homology searching, gene family reconstruction and regulatory motif finding. All of our strain data are publicly available at both http://malaria.ucsf.edu/comparison/ and http://PlasmoDB.org (39), the online malaria community resource center. Ultimately, these results will serve to identify desperately needed novel candidates for the development of *Plasmodium*-specific drug therapies and vaccines.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank the members of the DeRisi lab for critical reading of this manuscript. This work, J.L.D., M.L. and Z.B. were supported by a grant from the National Institute of Allergy and Infectious Disease (AI53862) and the Burroughs-Wellcome Fund. E.D.W. was supported by a fellowship from the Pharmaceutical Research and Manufacturers of America Foundation. A.T.A. was supported by a grant from the National Science Foundation. Funding to pay the Open Access publication charges for this article was provided by National Institute of Allergy and Infectious Disease (NIAID) (AI53862).

Conflict of interest statement. None declared.

REFERENCES

1. Sachs,J. and Malaney,P. (2002) The economic and social burden of malaria. *Nature*, **415**, 680–685.
2. Arav-Boger,R. and Shapiro,T.A. (2005) Molecular mechanisms of resistance in antimalarial chemotherapy: the unmet challenge. *Annu. Rev. Pharmacol. Toxicol*., **45**, 565–585.
3. Bray,P.G., Martin,R.E., Tilley,L., Ward,S.A., Kirk,K. and Fidock,D.A. (2005) Defining the role of PfCRT in *Plasmodium falciparum* chloroquine resistance. *Mol. Microbiol*., **56**, 323–333.
4. Gregson,A. and Plowe,C.V. (2005) Mechanisms of resistance of malaria parasites to antifolates. *Pharmacol. Rev*., **57**, 117–145.
5. Srivastava,I.K., Morrisey,J.M., Darrouzet,E., Daldal,F. and Vaidya,A.B. (1999) Resistance mutations reveal the atovaquone-binding

domain of cytochrome b in malaria parasites. *Mol. Microbiol*., **33**, 704–711.
6. Sidhu,A.B., Verdier-Pinard,D. and Fidock,D.A. (2002) Chloroquine resistance in *Plasmodium falciparum* malaria parasites conferred by pfcrt mutations. *Science*, **298**, 210–213.
7. Price,R.N., Uhlemann,A.C., Brockman,A., McGready,R., Ashley,E., Phaipun,L., Patel,R., Laing,K., Looareesuwan,S., White,N.J. *et al*. (2004) Mefloquine resistance in *Plasmodium falciparum* and increased pfmdr1 gene copy number. *Lancet*, **364**, 438–447.
8. Gardner,M.J., Hall,N., Fung,E., White,O., Berriman,M., Hyman,R.W., Carlton,J.M., Pain,A., Nelson,K.E., Bowman,S. *et al*. (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature*, **419**, 498–511.
9. Llinas,M. and DeRisi,J.L. (2004) Pernicious plans revealed: *Plasmodium falciparum* genome wide expression analysis. *Curr. Opin. Microbiol*., **7**, 382–387.
10. Bozdech,Z., Llinas,M., Pulliam,B.L., Wong,E.D., Zhu,J. and DeRisi,J.L. (2003) The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS Biol*., **1**, E5.
11. Coulson,R.M., Hall,N. and Ouzounis,C.A. (2004) Comparative genomics of transcriptional control in the human malaria parasite *Plasmodium falciparum*. *Genome Res*., **14**, 1548–1554.
12. King,M.C. and Wilson,A.C. (1975) Evolution at two levels in humans and chimpanzees. *Science*, **188**, 107–116.
13. Bergmann,S., Ihmels,J. and Barkai,N. (2004) Similarities and differences in genome-wide expression data of six organisms. *PLoS Biol*., **2**, E9.
14. Morley,M., Molony,C.M., Weber,T.M., Devlin,J.L., Ewens,K.G., Spielman,R.S. and Cheung,V.G. (2004) Genetic analysis of genome-wide variation in human gene expression. *Nature*, **430**, 743–747.
15. Oleksiak,M.F., Churchill,G.A. and Crawford,D.L. (2002) Variation in gene expression within and among natural populations. *Nature Genet*, **32**, 261–266.
16. Brem,R.B., Yvert,G., Clinton,R. and Kruglyak,L. (2002) Genetic dissection of transcriptional regulation in budding yeast. *Science*, **296**, 752–755.
17. Rathod,P.K., McErlean,T. and Lee,P.C. (1997) Variations in frequencies of drug resistance in *Plasmodium falciparum*. *Proc. Natl Acad. Sci. USA*, **94**, 9389–9393.
18. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. 2nd edn. Cold Spring Harbor press, Cold Spring Harbor, NY.
19. Harris,M.A., Clark,J., Ireland,A., Lomax,J., Ashburner,M., Foulger,R., Eilbeck,K., Lewis,S., Marshall,B., Mungall,C. *et al*. (2004) The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res*., **32**, D258–D261.
20. Kim,C.C. and Falkow,S. (2003) Significance analysis of lexical bias in microarray data. *BMC Bioinformatics*, **4**, 12.
21. Anders,R.F., Brown,G.V. and Edwards,A. (1983) Characterization of an S antigen synthesized by several isolates of *Plasmodium falciparum*. *Proc. Natl Acad. Sci. USA*, **80**, 6652–6656.
22. Smythe,J.A., Peterson,M.G., Coppel,R.L., Saul,A.J., Kemp,D.J. and Anders,R.F. (1990) Structural diversity in the 45-kilodalton merozoite surface antigen of *Plasmodium falciparum*. *Mol. Biochem. Parasitol*., **39**, 227–234.
23. Lanzer,M., de Bruin,D., Wertheimer,S.P. and Ravetch,J.V. (1994) Transcriptional and nucleosomal characterization of a subtelomeric gene cluster flanking a site of chromosomal rearrangements in *Plasmodium falciparum*. *Nucleic Acids Res*., **22**, 4176–4182.
24. Bozdech,Z., Zhu,J., Joachimiak,M.P., Cohen,F.E., Pulliam,B. and DeRisi,J.L. (2003) Expression profiling of the schizont and trophozoite stages of *Plasmodium falciparum* with a long-oligonucleotide microarray. *Genome Biol*., **4**, R9.
25. Florens,L., Washburn,M.P., Raine,J.D., Anthony,R.M., Grainger,M., Haynes,J.D., Moch,J.K., Muster,N., Sacci,J.B., Tabb,D.L. *et al*. (2002) A proteomic view of the *Plasmodium falciparum* life cycle. *Nature*, **419**, 520–526.
26. Lasonder,E., Ishihama,Y., Andersen,J.S., Vermunt,A.M., Pain,A., Sauerwein,R.W., Eling,W.M., Hall,N., Waters,A.P., Stunnenberg,H.G. *et al*. (2002) Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry. *Nature*, **419**, 537–542.
27. Le Roch,K.G., Johnson,J.R., Florens,L., Zhou,Y., Santrosyan,A., Grainger,M., Yan,S.F., Williamson,K.C., Holder,A.A., Carucci,D.J. *et al*. (2004) Global analysis of transcript and protein levels across the *Plasmodium falciparum* life cycle. *Genome Res*., **14**, 2308–2318.
28. Stubbs,J., Simpson,K.M., Triglia,T., Plouffe,D., Tonkin,C.J., Duraisingh,M.T., Maier,A.G., Winzeler,E.A. and Cowman,A.F. (2005) Molecular mechanism for switching of *P.falciparum* invasion pathways into human erythrocytes. *Science*, **309**, 1384–1387.
29. Kyes,S., Horrocks,P. and Newbold,C. (2001) Antigenic variation at the infected red cell surface in malaria. *Annu. Rev. Microbiol*., **55**, 673–707.
30. Lavstsen,T., Salanti,A., Jensen,A.T., Arnot,D.E. and Theander,T.G. (2003) Sub-grouping of *Plasmodium falciparum* 3D7 var genes based on sequence analysis of coding and non-coding regions. *Malar. J*., **2**, 27.
31. Le Roch,K.G., Zhou,Y., Blair,P.L., Grainger,M., Moch,J.K., Haynes,J.D., De La Vega,P., Holder,A.A., Batalov,S., Carucci,D.J. *et al*. (2003) Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science*, **301**, 1503–1508.
32. Blythe,J.E., Surentheran,T. and Preiser,P.R. (2004) STEVOR: a multifunctional protein? *Mol. Biochem. Parasitol*., **134**, 11–15.
33. Kyes,S.A., Rowe,J.A., Kriek,N. and Newbold,C.I. (1999) Rifins: a second family of clonally variant proteins expressed on the surface of red cells infected with *Plasmodium falciparum*. *Proc. Natl Acad. Sci. USA*, **96**, 9333–9338.
34. Duraisingh,M.T., Voss,T.S., Marty,A.J., Duffy,M.F., Good,R.T., Thompson,J.K., Freitas-Junior,L.H., Scherf,A., Crabb,B.S. and Cowman,A.F. (2005) Heterochromatin silencing and locus repositioning linked to regulation of virulence genes in *Plasmodium falciparum*. *Cell*, **121**, 13–24.
35. Freitas-Junior,L.H., Hernandez-Rivas,R., Ralph,S.A., Montiel-Condado,D., Ruvalcaba-Salazar,O.K., Rojas-Meza,A.P., Mancio-Silva,L., Leal-Silvestre,R.J., Gontijo,A.M., Shorte,S. *et al*. (2005) Telomeric heterochromatin propagation and histone acetylation control mutually exclusive expression of antigenic variation genes in malaria parasites. *Cell*, **121**, 25–36.
36. Volkman,S. and Wirth,D. (1998) Functional analysis of pfmdr1 gene of *Plasmodium falciparum*. *Methods Enzymol*., **292**, 174–181.
37. Myrick,A., Munasinghe,A., Patankar,S. and Wirth,D.F. (2003) Mapping of the *Plasmodium falciparum* multidrug resistance gene 5′-upstream region, and evidence of induction of transcript levels by antimalarial drugs in chloroquine sensitive parasites. *Mol. Microbiol*., **49**, 671–683.
38. Daily,J.P., Le Roch,K.G., Sarr,O., Ndiaye,D., Lukens,A., Zhou,Y., Ndir,O., Mboup,S., Sultan,A., Winzeler,E.A. *et al*. (2005) *In vivo* transcriptome of *Plasmodium falciparum* reveals overexpression of transcripts that encode surface proteins. *J. Infect. Dis*., **191**, 1196–1203.
39. Bahl,A., Brunk,B., Crabtree,J., Fraunholz,M.J., Gajria,B., Grant,G.R., Ginsburg,H., Gupta,D., Kissinger,J.C., Labo,P. *et al*. (2003) PlasmoDB: the *Plasmodium* genome resource. A database integrating experimental and computational data. *Nucleic Acids Res*., **31**, 212–215.