

Analysis of Transcriptomic Profiles of Malaria Parasite using Wavelet Decomposition

Hong Cai¹, Sos Aгаian¹, Maribel Sanchez², Jianying Gu³, Yufeng Wang²

Keywords: microarray, wavelet, time series, gene networks, malaria

1 Introduction.

The development of new antimalarial interventions is urgently needed due to the growing mortality and morbidity caused by malaria and the increasing prevalence of drug-resistance in *Plasmodium* parasites. Recent developments in systems biology have set the stage for a quantum leap in our understanding of the fundamental processes of the parasite life cycle and mechanisms of drug resistance and immune evasion [1-4]. Particularly, microarray technology has become a powerful tool in malaria research since it provides a transcriptional profile of parasites at various developmental stages (temporal profiles) and subcellular locations (spatial profiles).

Bozdech et al. [2] examined the expression profiles of *P. falciparum* every hour for the entire duration of the blood stage (48 hours). They extracted phase information using Fast Fourier Transform (FFT), which captures cellular signals in the frequency domain. However, interesting biological signals often contain numerous transitory characteristics that indicate trends, abrupt changes, and beginnings and ends of regulatory events. Such features are hidden from Fourier analysis, but by applying wavelet analysis to the same data set, we were able to uncover characteristics that will serve to better understand the dynamics of gene expression in such time course studies.

2 Methods and Results.

In this study, we employed a time-frequency wavelet analysis on expression profiles of malaria parasites at one-hour time intervals during the 48-hour blood cycle [2] (<http://malaria.ucsf.edu/SupplementalData.php>). Each chip consisted of a total of 7,462 probes that represent over 4,488 genes. The final data set after quality control filtering and normalization contained the profiles of 46 time points excluding 23rd-hour and 29th-hour time points.

The wavelet transform with multilevel structures can be viewed as decomposition by high-pass and low-pass filter banks. After wavelet decomposition, two higher-order statistics, skewness and kurtosis, were used to analyze wavelet coefficients and to build feature vectors. We derived feature vectors for 12 functional classes of proteins presented in Bozdech et al. [2]. For example, as shown in Figure 1a and 1b, two of these functional classes, transcriptional machinery, which is essential for parasite development, and merozoite invasion, which is crucial for parasite infection, displayed distinct wavelet properties. Moreover, the dimensionality of the feature vectors (Figure 1c and 1d) was reduced from 46 (time points) to 8, which made computation more cost-effective. The

¹ Department of Electrical Engineering, University of Texas at San Antonio, San Antonio, TX, USA. E-mail: hongc@lonestar.utsa.edu, sagain@utsa.edu

² Department of Biology, University of Texas at San Antonio, TX, USA. E-mail: marble03@yahoo.com, ywang@utsa.edu

³ Department of Ecology and Evolution, University of Chicago, Chicago, IL, USA. Email: jgu@uchicago.edu

subsequent Kernel Fisher Discriminant (KFD) classification based on these feature vectors separated 23 and 87 genes in these two functional classes with 99% accuracy, suggesting that the time-dependent gene expression properties are well represented by the low-dimensional feature vectors.

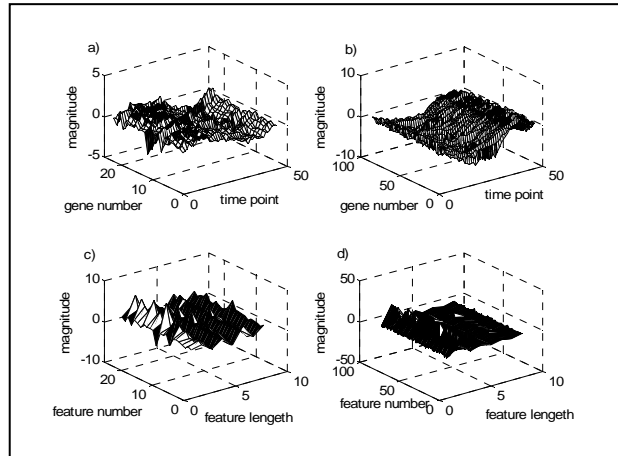


Figure 1. Gene properties captured by wavelet coefficients and feature vectors: (a) Wavelet coefficients of “translation” functional class. (b) Wavelet coefficients of “merozoite invasion” functional class. (c) Feature vectors of “translation” class (dimension = 8). (d) Feature vectors of “merozoite invasion” class (dimension = 8).

Our preliminary analysis based on wavelet higher order statistics and KFD also identified hypothetical proteins whose peak-to-trough amplitude oscillations indicate these proteins have potential roles in parasite development and infection. The increase in sensitivity and computational efficiency permitted by wavelet analysis should make it a useful tool for further annotation of the malaria genome and fill in knowledge gaps that hinder our understanding of gene networks.

3. References

- [2] Bozdech Z, Llinas M, Pulliam BL, Wong ED, Zhu J, DeRisi JL. 2003. The Transcriptome of the Intraerythrocytic Developmental Cycle of *Plasmodium falciparum*. *PLoS Biol* 1: E5
- [4] Florens L, et al. 2002. A proteomic view of the *Plasmodium falciparum* life cycle. *Nature* 419: 520-526
- [1] Gardner MJ, et al. 2002. Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* 419: 498-511.
- [3] Le Roch KG, et al. 2003. Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science* 301: 1503-1508