Metabolic Pathway Reconstruction for Malaria Parasite *Plasmodium falciparum*

Vachiranee Limviphuvadh[1]
limvip@kuicr.kyoto-u.ac.jp

Yasushi Okuno[1]
okuno@kuicr.kyoto-u.ac.jp

Toshiaki Katayama[2]
ktym@hgc.jp

Susumu Goto[1]
goto@kuicr.kyoto-u.ac.jp

Akiyasu C. Yoshizawa[1]
acyshzw@kuicr.kyoto-u.ac.jp

Minoru Kanehisa[1]
kanehisa@kuicr.kyoto-u.ac.jp

[1] Bioinformatics Center, Institute for Chemical Research, Kyoto University, Uji, Kyoto 611-0011, Japan
[2] Human Genome Center, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokane-dai, Minato-ku, Tokyo 108-8639, Japan

Keywords: malaria parasite, metabolic pathway, V-zyme, ortholog clusters

1 Introduction

Human malaria is caused by infection with four species of the intracellular parasitic protozoan genus *Plasmodium* that are transmitted by *Anopheles* mosquitoes. Of these four species, *Plasmodium falciparum* is the most lethal. At present, at least 300 million people are affected by malaria globally and accounts for 0.7-2.7 million annual deaths. The development of resistance in the parasite to antimalarial drugs, the lack of any licensed malaria vaccine and the fundamental complexity inherent in the malaria parasite, mean there is an urgent need to better understand the function of *P. falciparum* genes and their biological role to support the development of effective antimalarial strategies. *P. falciparum* genome indicates the presence of 5,432 genes spread across 14 chromosomes, a mitochondrial genome and a circular plastid genome. Notably, more than 60% are hypothetical proteins [1]. This fact emphasizes the need to elucidate gene function by somehow, new strategies.

In this research, we focus on KEGG metabolic pathways [2] in *P. falciparum*. Currently, there are gaps in some paths in *P. falciparum* metabolic pathways partly because of the insufficient annotation. We try to resolve such missing enzymes by using the virtual enzyme (V-zyme) system [4] and KEGG ortholog clusters (OCs).

2 Method

We used *P. falciparum* gene entries from the KEGG/GENES database and performed further annotation of EC numbers to those genes without EC numbers (originally taken from the NCBI RefSeq database), by using the information about the EC and GO assignments in PlasmoDB [3].

V-zyme constructs metabolic networks by predicting whether a pair of chemical compounds has a reactive connection with each other. When a chemical compound pair (reactant-product) is given, V-zyme outputs the EC numbers of the enzymes that possibly catalyze them. In this work, we first created a list of chemical compounds from the collection of possible *P. falciparum* enzymes with assigned EC numbers. Then, each of the all-againt-all compound pairs in the list is checked by V-zyme.

Ortholog clusters (OCs) are computationally defined as clusters of orthologs in KEGG/GENES using the graph analysis of the sequence similarity network stored in KEGG/SSDB. Each cluster is discriminated by a unique identifier called the OC number. The latest OCs are made from GENES version 27.0.

3 Results and Discussion

Of 5,342 *P. falciparum* genes, we annotated 591 (11.1%) genes as enzymes with the information on EC numbers. We predicted enzymatic reactions for 3,508 chemical compound pairs with V-zyme and we mapped them to KEGG metabolic pathways. We found many cases where the reaction steps of missing enzymes can be catalyzed by other *P. falciparum* enzymes. For example, thiamine-phosphate kinase (EC 2.7.4.16) appears to be absent in the thiamine metabolism of *P. falciparum* but V-zyme indicates that the reaction between thiamine phosphate and thiamine diphosphate may be catalyzed by 49 enzymes that the *P. falciparum* genome encodes and one of which is 2.7.4.14 (Fig. 1). The problem of V-zyme is that it cannot predict the direction of chemical reaction. It is difficult to check manually which enzyme in the predicted lists is most reliable to catalyze the reaction.

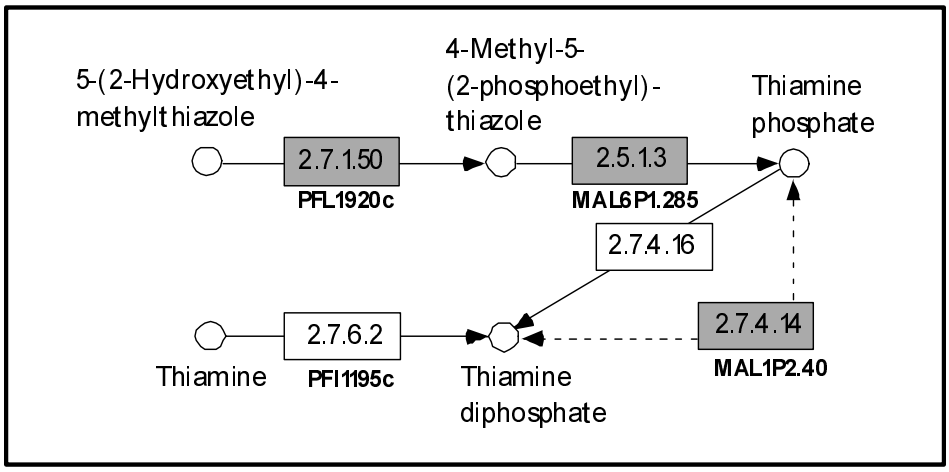

Figure 1: A part of the Thiamine Metabolism. The dark rectangles indicate enzymes that *P. falciparum* has.

We found some hypothetical proteins of *P. falciparum* are in the same OC number with annotated EC gene groups. In Fig. 1, we found PFI1195c gene was in the ortholog cluster where most of the genes were annotated as thiamin pyrophosphokinase (EC 2.7.6.2). Another example is PF11_0436 gene that was in the ortholog cluster where most of the genes were annotated as 1.3.3.3 (35 genes out of total 51). This enzyme is one of the missing enzymes in porphyrin metabolism of *P. falciparum*. Using such strategies, we can find more enzymes that may replace missing enzymes. Validation of the results by using additional information, such as microarray and proteomic data, is one of our future works.

Acknowledgments

This research was supported by the grants from the Ministry of Education, Cultures Sports, Science and Technology of Japan, the Japan Society for the Promotion of Science, and the Japan Science and Technology Corporation.

References

[1] Gardner, M.J. *et al.*, Genome sequence of the human malaria parasite *Plasmodium falciparum*, *Nature*, 419:498–511, 2002.

[2] Kanehisa, M., Goto, S., Kawashima, S., and Nakaya, A., The KEGG databases at GenomeNet, *Nucleic Acids Res.*, 30:42–46, 2002.

[3] Kissinger, J.C. *et al.*, The plasmodium genome database, *Nature*, 419:490–492, 2002.

[4] Okuno, Y., Hattori, M., Kotera, M., Igarashi, Y., Goto, S., and Kanehisa, M., Vzyme: a template-based method to predict reactions between two chemical compounds, *Genome Informatics*, 13:355–356, 2002.