TELKOMNIKA, Vol 16, No 1: February 2018, pp. 415~423
ISSN: 1693-6930, accredited **A** by DIKTI, Decree No: 58/DIKTI/Kep/2013
DOI: 10.12928/TELKOMNIKA.v16i1.6898

Similarities of Antimalarial Resistance Genes in Plasmodium Falciparum Based on Ontology

Dinar Munggaran Akhmad*[1], Yeni Herdiyeni[2], Etih Sudarnika[3]
[1]Department of Computer Science, Universitas Pakuan, Jl. Pakuan, Bogor, Jawa Barat, Indonesia
[2]Department of Computer Science, Bogor Agricultural University, Jl. Raya Dramaga, Dramaga Campus, Bogor 16680, Indonesia
[3]Department of Animal Infectious Diseases and Veterinary Public Health, Jl. Raya Dramaga, Dramaga Campus, Bogor 16680, Indonesia
*Corresponding author, e-mail: dinar.munggaran@unpak.ac.id

Abstract

The finding of P. falciparum chloroquine resistance (pfcrt) and P. falciparum multidrug resistance 1(pfmdr1) gene in P. falciparum has become an obstacle in treating malaria. The polymorphism between the two genes may result in molecular functions, in cellular components, or in biological processes. The objective of this research is to find similarities between the two genes in 3 components; cellular components, molecular functions and biological processes, based on Gene Ontology. the similarity will be counted semantically by path length approach with Wang method. The range of similarity values is 0-1. After the similarity value examined; in Molecular Function the similarity is 1 due to the same drug transmembrane transporter activity, in Cellular Component is 0,714, the similarity only at the same vacuole food cells, and in Biological Processes is 1 due to the same proces in responding to drug. Therefore, this research proves both genes have similarities based on gene ontology.

Keywords*: chloroquine, ontology, resistance, similarity, wang*

1. Introduction

Malaria is a main health problem in isolated islands, spreaded in tropical and subtropical countries in the world [1]-[2]. Malaria is a disease caused by parasites that transferred from one carrier to another through the bite of infected Plasmodium anopheles mosquitoes. Malaria is still a health major problem in countries in the world, about 214 million malaria cases happened all over the world in 2015 [3]. There are four kinds of malaria parasites, Plasmodium vivax, Plasmodium ovale, Plasmodium Malariae and Plasmodium Falciparum. From which about Plasmodium Falciparum is the most dangerous and lethal [4], [5] .

There are two kinds of anti malarial that normaly use to cure Falciparum Malaria, they are chloroquine dan sulfadoxine-pyrimethamin. Chloroquine has been decades used as main chemotherapy and control [6]. The effort to control malaria has been off-progress, due to the finding of gene in Plasmodium falcifparum that resistant to falciparum treatment [3]. The emergence of Plasmodium falciparum that resistence to chloroquin has failed the goal of controlling Malaria [7]. In research of molecular approach using PCR (polymerase chain reaction) technic found the existence of polymorphism on P. falciparum chloroquine resistance transporter (pfcrt) gene and P. falciparum multidrug resistance 1 (pfmdr1) gene that influence the chloroquine [8]. Research in molecular is conducted to locate the mutation position in various genes related to the resistance [9]. By comparing the sequence of DNA or protein from individual that has different evolution. Changes in molecular level can be observed as well as resistance. The more differentiation owned by two sequences, the farer the relatives between them [10].

However, it is not easy to conduct molecular research due to the cost and timeconsuming. The presence of knowledge based on ontology can help researcher or public to gain prior information of a gene through gen ontology technology. One of the techniques is by tracing information in GO (Gene Ontology) to know the characteristic of two genes that related to the resistance (GO) is information management of Gene ontology that developed by the Gen Ontology Consortium. GO consists of three components, molecular function, biological

Received Jul 27, 2017; Revised January 4, 2018; Accepted January 15, 2018

processes and cellular components [11]. GO is built by using the Web Ontology Language (OWL), Gene OWL is one of techniques to create semantic web [12]. Polymorphism that causes two genes resistant to antimalarial chloroquine may shows similarities between genes whether in biological processes, cellular components or in molecular function. This finding will ease to indentify other genes which show indications of resistance.

To measure the similarities between genes above can be done by counting the similarities between concepts in ontology. There are two kinds of semantic similarity counting in Ontology, one based on Information Content [13], [14] and the other based on Path [15]. Prior research conducted by [16] to predict the protein functions from human species by using 3 groups of GO, Molecular function, Cellular component dan Biological processes. [17] Research to predict the protein functions involved in mitosis by using annotation data from GO, 2 methods used are Bayesian Network dan SVM. [18] Research to predict interaction among proteins in yeast by counting semantic similarities based on path WU between two terms in document of GO. [19] Analyzed interaction between protein by counting semantic similarities between concepts in document of GO by using IC and Path. For Malaria, [20] research to predict functions of genes in Plasmodium falcifarum by counting the semantic similarity values based on patch to cluster the genes as well as to conduct its enrichment analysis. While research on antimalarial resistance has been done by [8]. Based on the explanation above on semantic similarity counting, many of which to predict function of genes in some species, one of which is malaria. Therefore this research is conducted to count the semantic similarities between parasite genes in Plasmodium falciparum that caused resistance to antimalarial choloroquine based on documents GO (Gene Ontology) by using Wang methods based on Path. The research is hoped to ease identification of other genes that indicated resistance.

2. Research Method

Stages to conduct in this research refer to Figure 1. Stages begin with collecting documents, matching gene query to final process of evaluation.

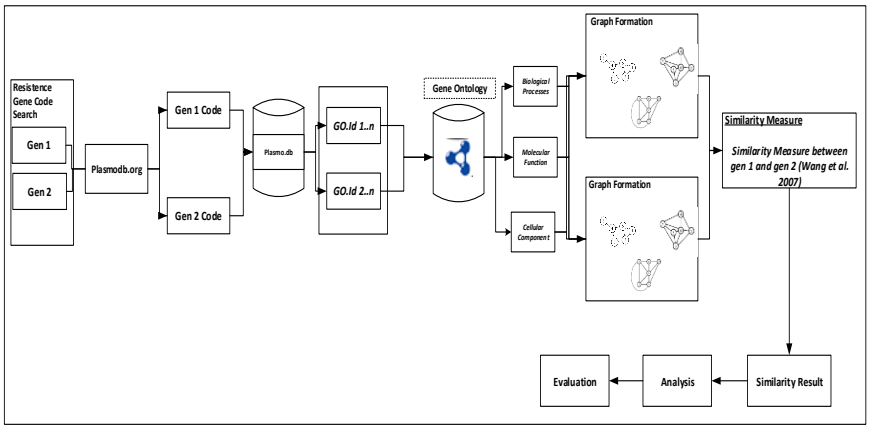

Figure 1. Research method

Data required in this research are gene ontology documents. Gene Ontology consists of 3 components, namely: biological process, molecular function, and cellular component. Biological processes show series of processes accomplished by one or more molecular function organization. Molecular function shows activity occurs on molecular level such as “catalytic activity” or “binding activity”. Cellular component shows component from cell as part of bigger object, like an anatomy structure (e.g. crude endoplasm reticulum or core) or group of gene products (e.g. ribosome, proteasome or dimmer protein) [11].

Other data required to this research are database of plasmodium falciparum gene species. Apair of Gene 1 and gene 2, Pfcrt and Pfmdr1, by looking up at the database procured

GO id which represent each gene with its components in gene ontology. Both genes refer to the prior research conducted by [8].

After having had the GO id in each gene ontology component, the next stage is to form Directed Acyclic Graph (DAG) as linier to Wang method [15]. DAG has been formed then the next stage is to count the similarity between two genes semantically. Gene ontology arranged as directed acyclic (DAG) where each term defines as having associated to one and another in the same domain and sometimes to other domains with relation "is-a" and "part-of". The dot lines shows relation "part-of" and the bold line shows relation "is-a". To measure semantic value similarity in GO term, first we have to change the semantic into numeric format; therefore [15] define the contribution as S-Value from GO t that related to term A. For so many terms t in DAGA (A,TA, EA), S-Value associated to term A, SA(t) define as:

$$\begin{cases} S_A(A) = 1 \\ S_A(t) = \max\{W_e * S_A(t') | t' \in \text{childrenof}(t)\} \text{if } t \neq A \end{cases} \quad (1)$$

After defining S-Values from all terms in DAG_A, then we count semantic value from term A in GO, SV(A) as:

$$SV(A) = \sum_{t \in T_A} S_A(t) \quad (2)$$

Let us assume that semantic contribution factor for relation "is-a" and "part-of" is 0,8 and 0,6 [15] then we use equation 1 and 2 to count S-value in GO term in each DAG. Given DAGA=(A,TA, EA) and DAGB= (B,TB, EA) for each term A and B in GO, therefore semantic similarity between 2 GO terms, SGO(A, B), defined as Equation 3.

$$S_{GO}(A,B) = \frac{\sum_{t \in T_A \cap T_B} (S_A(t) + S_B(t))}{SV(A) + SV(B)} \quad (3)$$

Where *S(t)* is value from S-value from term t GO that associated to term A and $S_B(t)$ is the value of *S-value* that associated to term B. Usually a single gene is explained by many GO terms. Semantics similarity between a set of GO terms and Go term is GO={ go1, go2,..., gok }, Sim (go, GO), defined as maximum semantics similarity among go terms and many more terms in a set/series of GO terms with Equation 4.

$$Sim(G1,G2) = \frac{\sum_{1 \leq i \leq m} Sim(go_{1i}, GO_2) + \sum_{1 \leq j \leq n} Sim(go_{2j}, GO_1)}{m+n} \quad (4)$$

The similarity result will be analyzed and evaluated by reviewing any related literature.

3. Results and Analysis

In this research, there are two genes from *Plasmodium falciparum* which caused resistant to *antimalarial chloroquine,* they are *P. Falciparum Chloroquine Resistant* (*Pfcrt*) and *P. Falciparum Multidrug 1* (*Pfmdr1*) which have been part of this research [8]. Those 2 genes will be counted semantically their similarity based on ontology and *graph* method by Wang method [15]. Based on the previous explanation on methodology of research1 gene usually represented by many terms/GO terms, whether they are in *biological process*, *molecular function* or in *cellular component*. Table1 contains GO terms information that explains *Pfcrt* dan *Pfmdr1* genes on GO *molecular* function component. Table 2 contains GO terms information that explains *Pfcrt* dan *Pfmdr1* genes on GO *Cellular* Component and Table 3 on *Biological* Processes component.

Table 1. Information on GO Terms Associated to Genes Pfcrt and Pfmdr1 in *Molecular* Function

Pfcrt	
GO:0015238	*Drug transmembrane transporter activity*
Pfmdr1	
GO:0015238	*Drug transmembrane transporter activity*
GO:0042626	*ATPase activity, coupled to transmembrane movement of substances*

Table 2. Information on GO Terms Associated to Genes Pfcrt and Pfmdr1 in *Cellular* Component

Pfcrt	
GO:0005783	*Endoplasmic Reticulum*
GO:0016020	*Membrane*
GO:0020020	*Food vacuole*
Pfmdr1	
GO:0016021	*Integral component of membrane*
GO:0020020	*Food vacuole*

Table 3. Information on GO Terms Associated to Genes Pfcrt and Pfmdr1 in *Biological* Processes

Pfcrt	
GO:0034635	*Glutathione transport*
GO:0042493	*Response to drug*
Pfmdr1	
GO:0042493	*Response to drug*

3.1. Dag Formation

3.1.1. DAG GO:0015238 Formation

Based on Wang method [15], after knowing a pair of genes which will be counted the similarity of it, then the next stage is to form Directed Acyclic Graph (DAG) for terms that represent genes based on Gene Ontology as shown on Figure 2 below.

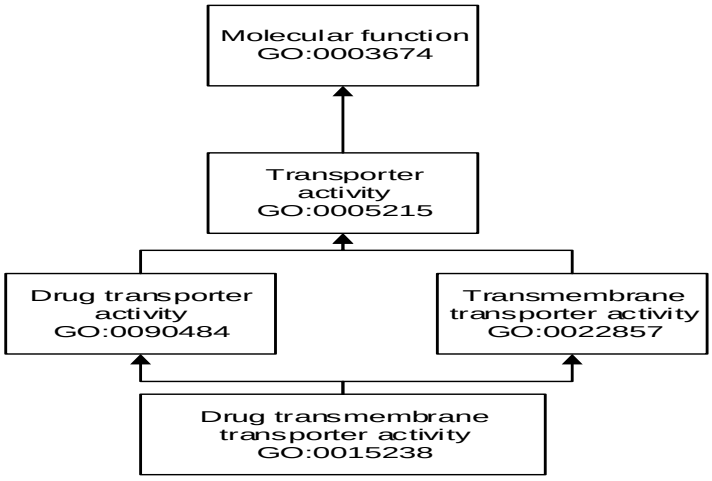

Figure 2. DAG for GO term drug transmembrane transporter activity:0015238

In this paper, one of the GO component will be discussed is *Molecular function,* due to the terms that represent 2 genes resistances to chloroquine, *Pfcrt* and *Pfmdr1* are not so many, therefore the form of DAG is more simple. Based Table 1 there will be 2 DAG formations, those are GO id GO:0015238 as seen in Figure 2 and GO:0042626 as seen in Figure 3. After DAG formed then the next stage is to count S-values in each term in DAG by using Equation 1 and 2. Table 4 shows S-values grade for DAG GO:0015238.

Table 4. S-values for GO terms in DAG for term Drug Transmembrane Transporter Activity:0015238

GO terms	GO:0015238	GO:0090484	GO:0022857	GO:0005215	GO:0003674
S-Value	1.0	0.8	0.8	0.64	0.512

3.1.2. DAG GO:0042626 Formation

Figure 3 shows DAG for GO term ATPase activity, coupled to transmembrane movement of substances:0042626. Table 5 shows S-values for DAG GO:0042626.

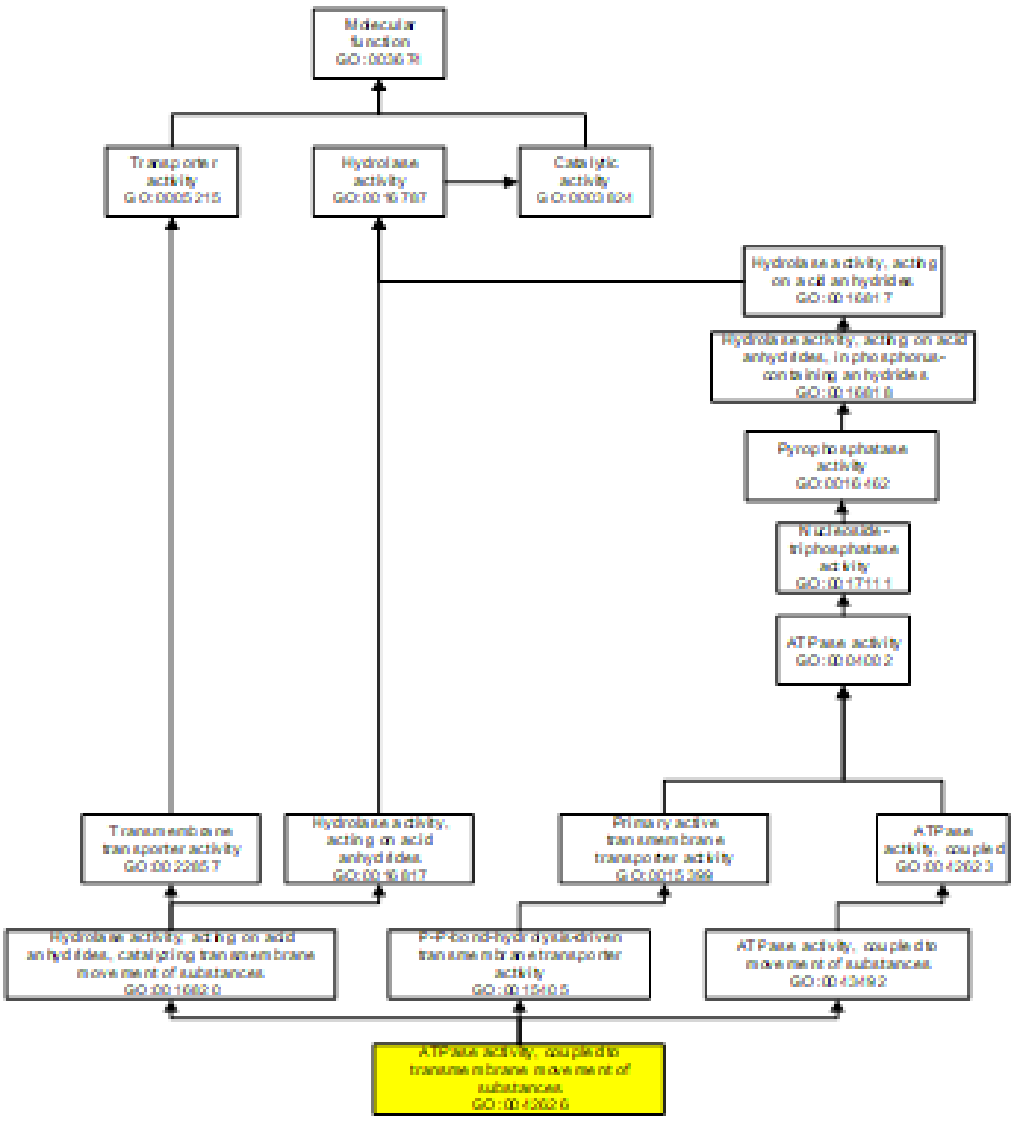

Figure 3. DAG for GO term ATPase activity, coupled to transmembrane movement of substances:0042626

Table 5. S-values for GO Terms in DAG for Term ATPase Activity, Coupled to Transmembran Movement of Substances:0042626

GO terms	GO:0042626	GO:0016820	GO:0015405	GO:0043492	GO:0022857
S-Value	1.0	0.8	0.8	0.8	0.64
GO terms	GO:0016817	GO:0015399	GO:0042623	GO:0005215	GO:000402
S-Value	0.64	0.64	0.64	0.512	0.512
GO terms	GO:0017111	GO:0016462	GO:0016818	GO:0016817	GO:0016787
S-Value	0.409	0.327	0.262	0.209	0.168
GO terms	GO:0003824	GO:0003674			
S-Value	0.134	0.409			

3.2. Gen Similarity

After DAG formation and S-value calculation by Equation 1 and 2 have been done, then the next stage is to count the semantics similarity values among terms by using equation 3. The result shows in Table 6.

Table 6. Similarities between Molecular Function Terms That Annotate Genes *Pfcrt* and *Pfmdr1*, Respectively

		pfmdr1	
	GO_ID (MF)	GO:0015238	GO:0042626
Pfcrt	GO:0015238	1	0.259

Based on Table 6 both genes for pfcrt and pfmdr1 have the same characteristics with similarity value of 1 by id go GO:0015238. Due to having the same Drug Transmembrane Transporter Activity or having resistances to antimalarial drugs surely both genes have the same molecule functions. Following by similarity value of 0,259 for id go GO:0015238 and GO:0042626, it is happened because GO:0042626 only has part of the same chart as GO:015238 those are transmembrane transporter activity and transporter activity. By using Equation 4 and 5 acquired similarity result between Pfcrt and Pfmdr1 gene in Molecular Function component of 1. With the same stages, the similarity results obtained between Pfcrt and Pfmdr1 genes in other GO components, namely Cellular Component and Biological Processes, are shown in Table 7 which includes the results of the entire study.

Table 7. Similarities between Resistance Genes *Pfcrt* and *Pfmdr1* in 3 GO Components

Gene Ontology Component	Gen Resistensi		
	Pfcrt	‖	Pfmdr1
Molecular function	1		
Cellular component	0.714		
Biological processes	1		

Based on Table 7 using equations 4 and 5 we get the similarity values between the pfcrt and pfmdr1 genes in the other 2 GO components: Cellular component and Biological processes. Both genes for pfcrt and pfmdr1 have similarity values in the Cellular component of 0.714 because both genes have similarities in some components such as Food vacuole where every parasite such as P.falciparum uses vacuole as its host, Membrane containing all protein and protein complex and Integral component of membranes containing complex gene and protein products. The similarity value in Biological processes is 1 because both genes have Response

to drug which can produce many process changes in cell or organism activity (Plasmodium falciparum) by producing drug stimulus.

Based on the research explanation that relevant to the introduction paragraph, semantic similarity measurements predicted more gene functions. However, this study tried to apply semantic similarity measurements to measure the similarity between the two genes causing P. falciparum resistance to chloroquine. Furthermore, this study can be used to obtain initial information about resistance to other P. falciparum genes. The results obtained are only graphic images (DAG) and tables.

Based on the results obtained, this evaluation refers to the literature [21]. There are several intrinsic/innate mechanisms:

a. The parasite is capable of producing enzymes capable of destroying an antimicrobial agent before reaching a target or modifying the drug so that it is not recognized by the target.
b. Parasitic walls become impermeable against antimicrobial agents.
c. The target position is altered by mutation so that it can no longer related to antimicrobial agents.
d. The parasite performs an efflux pump so that it removes the antimicrobial agent from inside the cell before the agent reaches the target.
e. Specific metabolic pathways in the parasite are genetically altered so that antibiotic genes can not produce the expected effect. The response to drugs is slow [22].

While the mechanism of resistance acquire/acquire one of them can be through mutation. Evaluation results as shown in Table 8.

Table 8 Evaluation

Resistance Mechanism	Gene Ontology Information		
	Compatibility	GO id	Definition
a. The parasite is capable of producing enzymes capable of destroying an antimicrobial agent before reaching a target or modifying the drug so that it is not recognized by the target.	√	GO:0042493 (Response to drug)	Any process that results in a change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a drug stimulus. A drug is a substance used in the diagnosis, treatment or prevention of a disease. The resulting enzyme may affect the response to the drug *Biological Processes*
b. The parasitic wall becomes impermeable to antimicrobial agents.	√	GO:0016020 (Membrane)	A lipid bilayer along with all the proteins and protein complexes embedded in it an attached to it. The parasite of P. falciparum certainly has a membrane *Cellular Component*
c. The target position is altered by mutation so that it can no longer bind to antimicrobial agents.	√	GO:0042493 (Response to drug)	Any process that results in a change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a drug stimulus. A drug is a substance used in the diagnosis, treatment or prevention of a disease. Mutations can alter the position of the target thus affecting the response to the drug *Biological Processes*
d. The parasite performs an efflux pump so that it removes the antimicrobial agent from inside the cell before the agent reaches the target.	√	GO:0015238 (Drug transmembranetransporter activity)	Enables the directed movement of a drug from one side of a membrane to the other. A drug is any naturally occurring or synthetic substance, other than a nutrient, that, when administered or applied to an organism, affects the structure or functioning of the organism; in particular, any such substance used in the diagnosis, prevention, or treatment of disease. *Molecular Function*

Table 8. Continued

Resistance Mechanism	Gene Ontology Information		
	Compatibility	GO id	Definition
	√	GO:0016021 (Integral component of membrane)	The component of a membrane consisting of the gene products and protein complexes having at least some part of their peptide sequence embedded in the hydrophobic region of the membrane. *Cellular Component*
e. Specific metabolic pathways in the parasite are genetically altered so that antibiotic genes can not produce the expected effect.	√	GO:0042493 (Response to drug)	Any process that results in a change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a drug stimulus. A drug is a substance used in the diagnosis, treatment or prevention of a disease. Changed paths can also affect the response to the drug. *Biological Processes*
f. Mutation	√	GO:0042493 (Response to drug)	Any process that results in a change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a drug stimulus. A drug is a substance used in the diagnosis, treatment or prevention of a disease. The existence of mutations can certainly affect/slow down the response to the drug. *Biological Processes*
	√	GO:0015238 (ATPase activity, coupled to transmembrane movement of substances)	Catalysis of the reaction: ATP + H2O=ADP + phosphate, to directly drive the active transport of a substance across a membrane. Mutations occur in many ATPase codon [8] *Molecular Function*
	√	GO:0020020 (Food Vacuole)	Vacuole within a parasite used for digestion of the host cell cytoplasm. An example of this component is found in the Apicomplexa. The parasite P. falciparum lives in the vacuole as its host cell. Mutations occur in parasitic vacuoles[8]. *Cellular Component*

4. Conclusion

Based on the result of similarity calculation using Wang method [15] and evaluation result, Gene Ontology can be used to get initial information related to resistance on Plasmodium falciparum genes specially. So it can be used also to detect other genes that indicated resistance before molecular research.

References

[1] Sillehu S, Arwati H, Dachlan YP, Keman S. Sensitivity and Specificity of Rapid Diagnostic Test with Microscopic Gold Standard to Identify Plasmodium Species. *International Journal of Public Health Science (IJPHS)*. 2016; 5(4): 354-8.
[2] Musa MI. Malaria disease distribution in Sudan using time series ARIMA model. *International Journal of Public Health Science (IJPHS)*. 2015; 4(1): 7-16.
[3] Organization WH. World malaria report 2015: World Health Organization. 2016.
[4] Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, et al. Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature*. 2002; 419(6906).

[5] Petersen I, Eastman R, Lanzer M. Drug-resistant malaria: Molecular mechanisms and implications for public health. *FEBS letters*. 2011; 585(11): 1551-62.
[6] Sidhu ABS, Verdier-Pinard D, Fidock DA. Chloroquine resistance in Plasmodium falciparum malaria parasites conferred by pfcrt mutations. *Science*. 2002; 298(5591): 210-3.
[7] Foote SJ, Thompson JK, Cowman AF, Kemp DJ. Amplification of the multidrug resistance gene in some chloroquine-resistant isolates of P. falciparum. *Cell*. 1989; 57(6): 921-30.
[8] Syafruddin D, Asih PB, Casey GJ, Maguire J, Baird JK, Nagesha HS, et al. Molecular epidemiology of Plasmodium falciparum resistance to antimalarial drugs in Indonesia. *The American journal of tropical medicine and hygiene*. 2005; 72(2): 174-81.
[9] Suwandi JF. Gen PfATP6 dan Resistensi Plasmodium falciparum Terhadap Golongan Artemisinin. *Jurnal Kedokteran*. 2015; 5(9): 141-6.
[10] Elrod S, Stansfield W. Genetika edisi keempat. *Penerbit Erlangga Hal*. 2007: 1-15.
[11] Consortium GO. Gene ontology consortium: going forward. *Nucleic acids research*. 2015; 43(D1): D1049-D56.
[12] Suma T, Swamy K. Email Classification Using Adaptive Ontologies Learning. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*. 2016; 14(4): 1472-9.
[13] Lin D. *An information-theoretic definition of similarity*. In fifteenth International Conference on Machine Learning. 1998;98: 296-304.
[14] Resnik P. Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language. *J Artif Intell Res(JAIR)*. 1999; 11: 95-130.
[15] Wang JZ, Du Z, Payattakool R, Yu PS, Chen C-F. A new method to measure the semantic similarity of GO terms. *Bioinformatics*. 2007; 23(10): 1274-81.
[16] Jensen LJ, Gupta R, Staerfeldt H-H, Brunak S. Prediction of human protein function according to Gene Ontology categories. *Bioinformatics*. 2003; 19(5): 635-42.
[17] Barutcuoglu Z, Schapire RE, Troyanskaya OG. Hierarchical multi-label prediction of gene function. *Bioinformatics*. 2006; 22(7): 830-6.
[18] Wu X, Zhu L, Guo J, Zhang D-Y, Lin K. Prediction of yeast protein–protein interaction network: insights from the Gene Ontology and annotations. *Nucleic acids research*. 2006; 34(7): 2137-50.
[19] Guzzi PH, Mina M, Guerra C, Cannataro M. Semantic similarity analysis of protein data: assessment with biological features and issues. *Briefings in bioinformatics*. 2011; 13(5): 569-85.
[20] Tedder PM, Bradford JR, Needham CJ, McConkey GA, Bulpitt AJ, Westhead DR. Gene function prediction using semantic similarity clustering and enrichment analysis in the malaria parasite Plasmodium falciparum. *Bioinformatics*. 2010; 26(19): 2431-7.
[21] Cavaleri J, Rankin D, Harbeck J, Sautter L, McCarter S, Sharp S, et al. Manual of antimicrobial susceptibility testing. *American Society for Microbiology, Seattle, Washington*. 2005; 12: 53-42.
[22] White NJ. Antimalarial drug resistance. *Journal of clinical investigation*. 2004; 113(8): 1084.