Genetic Characterization and Evolutionary Inference of TNF-α Through Computational Analysis

Gauri Awasthi, Suchita Singh, A.P. Dash and Aparup Das
Evolutionary Genomics and Bioinformatics Laboratory, National Institute of Malaria Research; Delhi, India

TNF-α is an important human cytokine that imparts dualism in malaria pathogenicity. At high dosages, TNF-α is believed to provoke pathogenicity in cerebral malaria; while at lower dosages TNF-α is protective against severe human malaria. In order to understand the human TNF-α gene and to ascertain evolutionary aspects of its dualistic nature for malaria pathogenicity, we characterized this gene in detail in six different mammalian taxa. The avian taxon, *Gallus gallus* was included in our study, as TNF-α is not present in birds; therefore, a tandemly placed duplicate of TNF-α (LT-α or TNF-β) was included. A comparative study was made of nucleotide length variations, intron and exon sizes and number variations, differential compositions of coding to non-coding bases, etc., to look for similarities/dissimilarities in the TNF-α gene across all seven taxa. A phylogenetic analysis revealed the pattern found in other genes, as humans, chimpanzees and rhesus monkeys were placed in a single clade, and rats and mice in another; the chicken was in a clearly separate branch. We further focused on these three taxa and aligned the amino acid sequences; there were small differences between humans and chimpanzees; both were more different from the rhesus monkey. Further, comparison of coding and non-coding nucleotide length variations and coding to non-coding nucleotide ratio between TNF-α and TNF-β among these three mammalian taxa provided a first-hand indication of the role of the TNF-α gene, but not of TNF-β in the dualistic nature of TNF-α in malaria pathogenicity.
Key-Words. Malaria, TNF-α, TNF-β, evolution, computational analyses.

Evolutionary study of genes and genomes helps in deciphering the role of various evolutionary forces, including the signatures of natural selection. While most of the structural and compositional aspects of a gene are inherited by the organism from its ancestors, different evolutionary forces affect the genes in changing compositional and conformational aspects for the best adaptation to a particular environment. Thus, evolutionary studies would not only reveal these kinds of changes in genes and correlate their function in a particular organism, but also in closely related taxa.

Human malaria is a complex infectious disease caused by protozoan parasites (*Plasmodium* species). Infection in humans is followed by repeated cycles of growth of the parasite in the erythrocytes [1]. As a result, patients may suffer systemic, single, or multi-organ involvement, including acute respiratory distress, coagulopathy, shock, metabolic acidosis, hypoglycemia, renal failure, pulmonary edema, and cerebral involvement, including seizures and coma, leading to complicated cerebral malaria [2]. Basic mechanisms controlling these processes are thought to include the site-specific localization of malaria parasites by adherence to vascular endothelial markers, such as the intracellular cell adhesion molecule-1 (ICAM-1) and both local, and systemic inflammatory responses arising from the action of cytokines. Thus, cytokines play a major role in imparting pathogenicity to malaria [2]. One of the most prominent cytokines is tumor necrosis factor alpha (TNF-α). TNF-α is thought to play an important role in cerebral malaria (CM) physiopathology by inducing changes in cerebral endothelial cells, leading to the surface expression of adhesion molecules, thereby enhancing parasitized erythrocyte sequestration [3,4]. When TNF-α is excessively produced, malaria patients suffer from fever, hypoglycemia, bone marrow, depression, coagulopathy, hypergammaglobulinemia, hypotension and rise in serum levels [4]. In a large case control study of Gambian children, a genetic variant of the TNF-α gene promoter region was found to have a relative risk of seven for death or severe neurological sequels due to severe malaria, suggesting that regulatory polymorphisms of cytokine genes affect the outcome of severe infection [5]. However, TNF-α and another cytokine, interferon gamma (IFN-γ) have been shown to govern a new type of protective response against infectious agents in malaria, in addition to causing disease when generated in excess [6]. This dual role (protective and pathogenic) of TNF-α depends on the quantity of TNF-α produced, the time-period over which its production is sustained, the tissue in which it is produced, and on other cytokines that regulate its production [6,7]. Patients infected with the malaria parasite *Plasmodium falciparum* have higher levels of TNF-α; these levels are directly proportional to the severity of disease [7]. Apart from malaria, the dual role of TNF-α also extends to other infectious diseases, such as leishmaniasis; it is well documented that TNF-α, along with IFN-γ, are over produced in cutaneous and mucosal leishmaniasis [8].

The dual nature of TNF-α in humans thus raises a challenge in evolutionary research to understand whether this dual nature is an intrinsic property of human TNF-α or if a similar genetic nature also occurs in related taxa. The whole genome sequence database in the public domain (www.ensembl.org) provides an opportunity to look for the genetic architecture of this gene. To initiate studies on evolutionary inference, we first must take a comparative

genetic approach to find interrelationships among TNF-α genes across different taxa. With a phylogenetic approach, we could look for differential genetic composition patterns of TNF-α genes. Such studies would not only help us understand the genetic nature of the TNF-α gene in general, but would also unravel new genetic features of this gene that could provide new directions for further research on this gene in general, and its relation to malaria, in particular.

Material and Methods

Nucleotide sequences of the TNF-α gene from seven different taxa (*Homo sapiens, Macaca mulatta, Mus musculus, Rattus norvegicus, Pan troglodytes, Canis familiaris and Gallus gallus)* were downloaded with the Ensembl genome browser (www.ensembl.org) version e48. This database was well annotated; frequent repetitive elements were not observed in any of the sequences. No annotations were imported from any other source. The computer software DNASTAR (DNASTAR Inc., Madison, USA, www.dnastar.com) was used to align the nucleotide and amino acid sequences following the incorporated Clustal W algorithm. These DNA sequences were subjected to rigorous computational and statistical analyses. Total gene lengths, coding and non-coding sequences of this gene in different taxa were calculated and compared using the Ensembl web interface. Short and long introns were defined according to the description provided by Gazave et al., 2006 [9], who suggested that introns of more than 1,029 nucleotide base pairs (bp) may be considered as long and less than 1,029 bp as short. A similar guideline was followed in our analyses. Statistical analyses were run and correlation coefficients among the different variables (total gene length, exon length and intron length) were compared across taxa, using the free version of the computer software "Analyze it", an add-on to Microsoft Excel. Chi square values were calculated for exon and intron lengths, using online statistical tools; $p < 0.05$ was considered as the level of statistical significance. In order to infer evolutionary interrelationships among various taxa at the TNF-α gene, nucleotide sequences from different taxa were aligned using the computer software 'MegAlign', a part of the DNASTAR software package. A phylogenetic tree was constructed using the same alignments. Further, pair and root distances were calculated for the reconstructed tree, and bootstrapped and branch length values were calculated using the phylogeny option of the VEGAZZ software (www.vegazz.net).

Results and Discussion

We characterized the TNF-α gene from six mammalian taxa and inferred phylogenetic positions of these taxa based on the TNF-α gene sequence. Furthermore, since LT-α (TNF-β) and TNF-α are considered to be very closely related genes, are products of duplication events, and are placed tandemly [10,11], evolutionary changes in these two genes were also studied. Since our goal was to infer the evolutionary status of the TNF-α gene, and it is known that TNF-α is absent in the

avian lineage [12], we included the TNF-β sequence of an avian species, *Gallus gallus*, in our study. Detailed information on the chromosome locations, accession numbers, gene orientations, starting and ending sequences, and length in nucleotide base pairs of the TNF gene in different taxa are shown in Table 1. Fine-scale characterization of the TNF-α gene across all the taxa that we studied revealed a clear picture of the coding positions (exons) and non-coding regions (introns), along with the non-coding exons or the untranslated regions (UTRs) (Figure 1). The lengths of UTRs in each taxon were quite variable at both the 3' and 5' ends. However, both the 3' and 5' UTRs were entirely absent in *C. familiaris*. We found the TNF-α gene to have four exons and three introns in all the mammalian taxa. However, in *G. gallus*, the TNF-β gene contains four introns and five exons. Like the UTRs, the sizes of coding exons and introns were very variable across different mammalian taxa; in *G. gallus*, exon 1 was found to be unexpectedly large (Figure 2). The total size of the TNF-α gene varied from 1,722 (*C. familiaris*) to 6,818 (*G. gallus*) nucleotide base pair lengths (Figure 3). The *H. sapiens* (2,779 bp) and *M. mulatta* (2,755bp) genes had almost the same sizes. Similarity in gene size was also seen between *R. norvegicus* (2,619bp) and *M. musculus* (2,627bp). Interestingly, the TNF-α genes in *H. sapiens* and *P. troglodytes* are not the same size, not even in the coding regions. The total coding nucleotide contents (exons) of four mammalian taxa (*H. sapiens, M. mulatta, M. musculus* and *R. norvegicus*) were found to be quite similar (Figure 3), whereas the coding nucleotide contents of *P. troglodytes* was most similar to *C. familiaris* and *G. gallus* (Figure 3). Here, three taxa of different evolutionary lineages contain almost same-sized coding nucleotides. However, since the distribution of each exon and intron was not uniform, we tested the randomness in the distribution of length of exons in six mammalian taxa with the chi-square test and found non-random distributions of length variations of exons across the different taxa ($\chi^2 = 101.4$, df=15, $p < 0.0001$), indicating significant differences in length of each individual exon. A similar test was also performed for introns; there was a random distribution ($\chi^2 = 6.976$, df=10, p=0.72). This indicates that the coding parts of the gene have undergone drastic modifications in size across different lineages, while conserving the total number of exons. On the other hand, we found the length of the introns in the TNF-α gene to be conserved across these taxa. This is somewhat surprising, as most of the conserved nucleotide stretches in orthologous genes have been found to be non-coding nucleotides [13]. Evolutionary significance of the conserved non-coding sequences (CNSs) relies on assumptions that some function of regulatory elements are conserved across taxa and that this functional conservation is reflected through nucleotide sequence similarity, as functional sequences tend to evolve at a slower rate than non-functional ones [14,15]. About 0.3 to 1% of the human genome was found to have such CNSs shared among mammals [15]. However, since our study included only non-coding sequences of the TNF gene, whether CNSs in the non-coding

Table 1. Detail information on the chromosomal location, accession number, gene orientation, starting and ending sequences, and length in nucleotide base pairs of the TNF gene in different taxa.

Species	Chromosome No.	Accession number	Gene orientation	Gene		Sequence Length(bp)
				Start	End	
Human (*Homo sapiens*)	06	NC_000006.10	Forward	31,651,329	31,654,091	2,779
Chimpanzee (*Pan troglodytes*)	20	NC_006473.2	Forward	32,137,447	32,140,065	2,618
Rhesus monkey (*Macaca mulatta*)	04	NC_007861.1	Forward	31,210,446	31,213,201	2,755
Mouse (*Mus musculus*)	17	NC_000083.5	Reverse	35,336,326	335,338,952	2,626
Rat (*Rattus norvegicus*)	20	NC_005119.2	Forward	3,661,000	3,663,618	2,619
Chicken (*Gallus gallus*)	19	NC_006594.2	Forward	923,095	929,913	6,818
Dog (*Canis familiaris*)	12	NC_006101.2	Forward	4,077,724	4,079,445	1721

Figure 1. Fine-scale characterization of TNF-α gene among six mammalian taxa with coding exons (red) and un-translated region (UTR) or non-coding exons (blue). Information on the TNF-β of *Gallus gallus* has been included. The lengths of non-coding exons, coding exons and introns are not to scale.
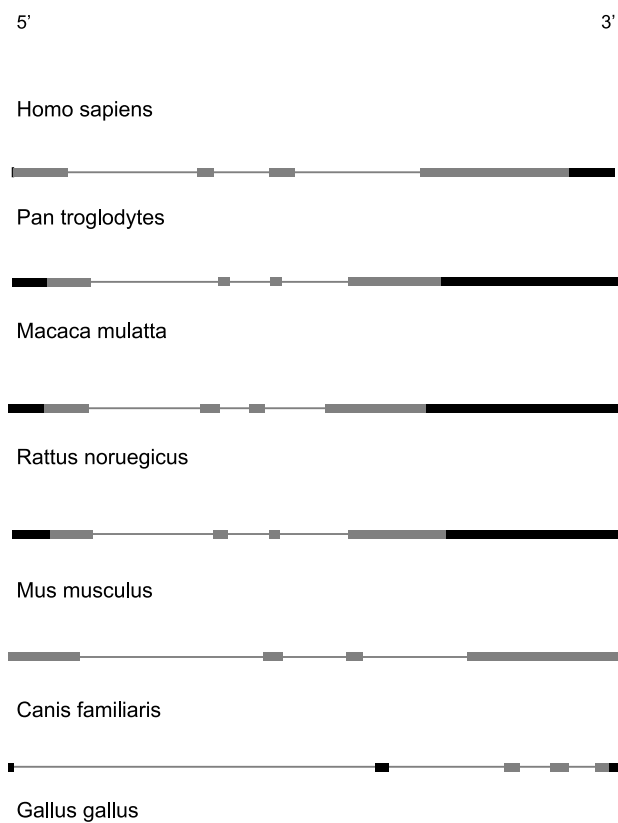

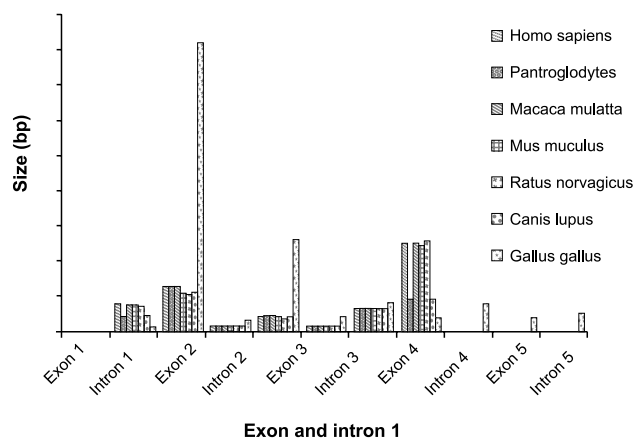**Figure 2.** Size distribution of exons and introns across all seven taxa.

Figure 3. Proportion of total coding and non-coding nucleotides in the TNF-α gene across different taxa.
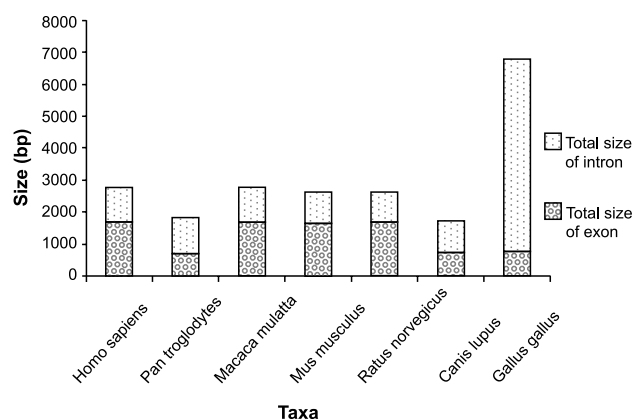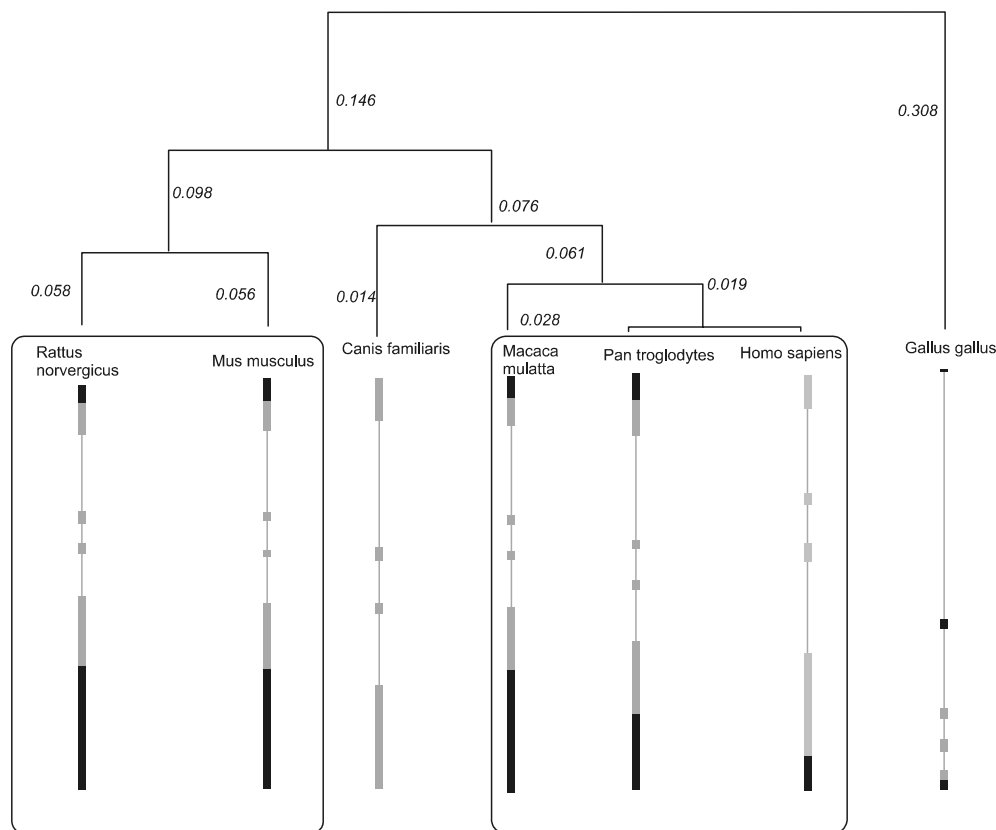
Figure 4. Taxonomic position of different taxa in an un-rooted neighbor-joining tree at the TNF-α gene. The branch length values are shown at each branch leading to taxa. In the case of *Gallus gallus*, the TNF- β gene was considered, as the TNF-α gene is not present in birds.

regulatory elements are conserved could not be ascertained. We also found a significant positive correlation between the intron size and total TNF-α gene size (r=0.97, p< 0.001). This observation corroborates the hypothesis that the sizes of genes and genomes have increased across taxa, due to accumulation of non-coding nucleotides [13,16].

In order to infer the phylogenetic position of each individual taxon, we constructed an un-rooted phylogenetic tree for the TNF-α gene, considering all seven taxa (Figure 4). The length of each branch was obtained through randomization of bootstrapped trees and shown at each branch leading to a taxon (Figure 4). The strength of each internal node was found to be absolute in all cases after 1,000 comparisons of bootstrapped trees through simulation (values not shown). It is clear from Figure 4 that *P. troglodytes* and *H. sapiens* are closely related to each other at the TNF-α gene, as are *R. norvegicus* and *M. musculus*. Further, it is clear that *M. mulatta* recently diverged from the *H. sapiens-P. troglodyte's* clade and that the avian species *G. gallus* falls in an entirely separate branch (Figure 4), acting as an out-group. Thus, phylogenetic positions of different taxa based on the TNF-α gene adhere to tree topologies for most of the other genes and also based on a human immune system gene (CD36) involved in malaria pathogenicity [16].

Given the importance of human TNF-α for disease pathogenicity (especially its dual nature), and the evolutionary closeness with two other mammalian taxa (*P. troglodytes* and *M. mulatta*) that we included in our study (see above), we aligned the amino acid sequences of this gene in all three taxa to locate variations in amino acid sequences (Table 2). Amino acid changes in the TNF-α polypeptide chain were detected at only four positions (41, 42, 45 and 46), when we compared *H. sapiens* and *P. troglodytes* (Table 2), whereas major differences were seen in the *M. mulatta* lineage (Table 2). Although the role of TNF-α in disease pathogenesis is little known in *P. troglodytes* and *M. mulatta*, large differences in amino acid compositions in *M. mulatta* and the fact that there are only four amino acid substitutions from *P. troglodytes* might hold the key to understanding the dual nature of TNF-α in human malaria pathogenicity. An earlier study involving sequencing of the FOXP2 protein in the chimpanzee, gorilla, orangutan, rhesus monkey and mouse, in comparison with human cDNA, showed that human FOXP2 contains changes in amino acid coding and a pattern of nucleotide polymorphism that strongly suggests that this gene has been the target of selection during recent human evolution, involving the evolution of speaking ability in the

Table 2. Amino acid alignment of TNF-α protein across three different taxa, *Homo sapiens, Pan troglodytes* and *Macaca mulatta*. The red color denotes an amino acid substitution at that particular position and lineage.

Taxa	11	14	21	36	41	42	43	45	46	47	72	88	128	129	130	131	132
Pan troglodytes	K	G	-	R	G	S	S	R	R	-	R	T	R	D	Y	D	A
Homo sapiens	K	G	-	R	V	R	-	-	S	R	R	T	R	D	Y	D	A
Macaca mulatta	R	A	W	K	-	V	R	-	S	R	T	N	D	Y	D	A	-

A-alanine, D-aspartic acid, G-glycine, K-lysine, N-aspargine, R-arginine, S-serine, T-threonine, V-valine, W-tryptophan, -Y-tyrosine.

Table 3. Comparison of the total coding and non-coding nucleotides in TNF-α and TNF-β among *Homo sapiens, Pan troglodytes* and *Macaca mulatta.*

Taxa	Gene	Total coding nucleotides (in bases)	Total non-coding nucleotides (in bases)
H. sapiens	TNF-α	1,685	1,094
	TNF-β	1,419	620
P. troglodytes	TNF-α	699	1,099
	TNF-β	1,386	618
M. mulatta	TNF-α	1,662	1,093
	TNF-β	1,385	612

human lineage [17]. Thus, it is apparent from molecular evolutionary studies in mammals that small changes in amino acid compositions between taxa can result in large phenotypic differences. However, how the finding of only four amino acid substitutions between *H. sapiens* and *P. troglodytes* relates to the origin of the dual nature of TNF-α is still unknown. Further studies that include population genetic and gene expression analyses are needed to confirm this hypothesis.

In order to determine if the above-mentioned characteristics were also possessed by the TNF-β gene [10, 11], we measured and compared length variations in the total coding and non-coding regions of these two genes (TNF-α and TNF-β) in all three taxa (*H. sapiens, P. troglodytes* and *M. mulatta*, Table 3). While significant differences in lengths of both the coding and non-coding sequences were found between these two genes (TNF-α and TNF-β), remarkable similarities were seen in the length of the coding nucleotide contents across taxa for TNF-β, except that *H. sapiens* contains about 35 bases of coding nucleotides more than the other two mammalian taxa (Table 3). This is in dramatic contrast to the TNF-α gene, where the total coding nucleotides are more than double in number in *H. sapiens* compared to *P. troglodytes* (Table 3). Thus, while the length of coding contents of the TNF-β gene remained conserved throughout the evolutionary time scale, that of TNF-α varies significantly between *H. sapiens* and *P. troglodytes*. Further, the coding to the non-coding nucleotide ratio was found to be more than double in *H. sapiens* (1.54) compared to *P. troglodytes* (0.64, Figure 3). Thus, TNF-α seems to have arisen

independently in humans and chimpanzees and is more polymorphic than TNF-β. It has been suggested that such polymorphic genes can easily carry mutations for hereditary diseases [18]. Since TNF-α plays a dual role for malaria in humans [7], and no report of this phenotype has been made for chimpanzees, it can thus be argued, though perhaps prematurely, based on the polymorphic nature of TNF-α, that this gene holds the key for this dualism in human malaria. However, this assumption is very hypothetical, pending confirmation from empirical evidence, including functional studies for both genes, with comparative analyses in *H. sapiens* and *P. troglodytes* (also see above). Furthermore, considerable length similarities in non-coding contents in both the genes across these three taxa remind us that these genes are independently evolving and are very recent in origin, as it is known that genes accumulate non-coding nucleotides more rapidly [19]. Thus, significant changes in the coding nucleotides across closely related taxa provide evidence for operation of natural selection in the coding regions in the TNF-α gene (not to the same extent in TNF-β), when only length variation is taken into consideration. However, as indicated above, further empirical studies are needed to confirm such hypotheses.

In addition to its role in malaria pathogenesis, TNF also plays significant roles in leishmaniasis, listeriosis, and other infectious diseases. High circulating levels of TNF have been found in patients with cerebral malaria, particularly in those whose disease runs a lethal course. The mechanism by which TNF acts is well known. Among its many activities, TNF up-

regulates endothelial adhesion molecules and thereby increases the tendency of infected red cells to stick to the walls of blood vessels and interrupt blood flow. The molecular genetic mechanisms of this have been well investigated, and it has been found that three polymorphisms in the promoter region of the TNF gene contribute to malarial pathogenesis and poor outcome, through a complex system of increased risk and counterbalancing protective effects. One such polymorphism, TNF_{-376A}, increases the secretion of TNF, most likely by causing the helix-turn-helix transcription factor OCT-1 to bind to a novel region of complex protein-DNA interactions, altering gene expression in human monocytes [20]. Further, a single nucleotide G→A polymorphism at position 376 upstream of the TNF transcriptional initiation site (the OCT-1 binding phenotype), has been linked to unfavorable outcome from malarial infection in two ethnically distinct groups, one from Gambia and the other from Kenya [20]. These results underscore the potential to link molecular events to clinical outcomes using the emerging knowledge of whole organism genomics, which is likely to influence future research on the prevention and treatment of malaria and many other diseases [21].

In conclusion, we examined the genetic architecture and evolutionary aspects of a major human cytokine that is highly relevant to human malaria pathogenicity, using computational approaches. Since the unique dual nature of this cytokine in human malaria could be a key to deciphering how evolutionary forces act at selected genes in the genome, such studies should begin with a comparative genetic approach involving closely-related taxa. Such approaches have proven to be useful in an array of animal taxa, *e.g.*, in *Candida* species [22], influenza B virus [23], in human mental retardation [24], Duffy [25], and CD36 [16] genes, and in detecting novel biochemical pathways in the human intestinal bacterium, *Escherichia coli* [26].

Acknowledgements

References

1. Greenwood B., Mutabingwa T., Malaria in 2002. Nature **2002**;415:670-2.
2. Brennan F.M., Feldmann M., Cytokines in autoimmunity. Curr Opin Immunol **1996**;8:872-7.
3. Clark IA., Alleva L.M., Mills A.C., Cowden W.B., Pathogenesis of malaria and clinically similar conditions. Clin. Microbiol Rev **2004**;17:509-39.
4. Kwiatkowski D., Hill A.V., Sambou I., et al. TNF concentration in fatal cerebral, non-fatal cerebral and uncomplicated *Plasmodium falciparum* malaria. Lancet **1990**;336:1201-4.
5. McGuire W., Hill A.V., Allsopp C.E., et al. Variation in the TNF-alpha promoter region associated with susceptibility to cerebral malaria. Nature **1994**;371:508-10.
6. Grau G.E., Taylor T.E., Molyneux M.E., et al. Tumor necrosis factor and disease severity in children with falciparum malaria. New England Journal of Medicine **1989**;320:1586-91.
7. Brandts C.H., Mordmüller B.G., Lehman L.G., Kremsner P.G., The dual role of tumor necrosis factor (TNF) during a malarial attack. Sante **1997**;4:271-4.
8. Bacellar O., Lessa H., Schriefer A., et al. Up-regulation of Th1-type responses in mucosal leishmaniasis patients. Infection and Immunity **2002**;70:6734-40.
9. Gazave E., Bonet T.M., Fernando O., et al. Patterns and rates of intron divergence between humans and chimpanzees. Genome Biology **2006**;8:1-13.
10. Sainz J., Pérez E., Hassan L., et al. Variable number of tandem repeats of TNF receptor type 2 promoter as genetic biomarker of susceptibility to develop invasive pulmonary aspergillosis. Human Immunology **2007**;68:41-50.
11. Stanualla M., Schrauder A., Welte K., Schrappe M. Tumor necrosis factor and lymphotoxin-alpha genetic polymorphisms and risk of relapse in childhood B-cell precursor acute lymphoblastic leukemia: a case-control study of patients treated with BFM therapy. BMC Blood Disorder **2001**;1:2.
12. Deakin J.E., Papenfuss A.T., Belov K., et al. Evolution and comparative analysis of the MHC Class III inflammatory region. BMC Genomics **2006**;7:1-14.
13. Sironi M., Menozzi G., Comi G.P., et al. Fixation of conserved sequences shapes human intron size and influences transposon-insertion dynamics. Trends in Genetics **2005**;21:484-8.
14. Keightley P. D., Kryukov G. V., Sunyaev S., et al. Evolutionary constraints in conserved nongenic sequences of mammals. Genome Research **2005**;15:1373-8.
15. Dermitzakis E.T., Reymond A., Scamuffa N., et al. Evolutionary discrimination of mammalian conserved non-genic sequences (CNGs). Science **2003**;302:1033-5.
16. Awasthi G., Dash A.P., Das A., Characterization and evolutionary analysis of human CD36 gene. Indian Journal of Medical Research **2008**; in press.
17. Enard W., Przeworski M., Fisher S. E., et al. Molecular evolution of FOXP2, a gene involved in speech and language. Nature **2002**;418:869-72.
18. Lavoie H., Debeane F., Trinh Q-D., et al. Polymorphism, shared functions and convergent evolution of genes with sequences coding for polyalanine domains. Human Molecular Genetics **2003**;12:2967-79.
19. Knibbe C., Coulon A., Mazet M., et al. A long-term evolutionary pressure on the amount of non-coding DNA. Molecular Biology and Evolution **2007**;24:2344-53.
20. Knight J.C., Udalova I., Hill A.V., et al. A polymorphism that affects OCT-1 binding to the TNF promoter region is associated with severe malaria. Nature Genetics **1999**;22:145-50.
21. Broder S., Venter J. C. Sequencing the entire genomes of free-living organisms: The foundation of pharmacology in the new millennium. Annual Review of Pharmacology and Toxicology **2000**;40:97-132.
22. Massey S.E., Moura G., Beltrão P., et al. Comparative evolutionary genomics unveils the molecular mechanism of reassignment of the CTG codon in Candida spp. Genome Research **2003**;13:544-7.
23. Lindstrom S.E., Hiromoto Y., Nishimura H., et al. Comparative analysis of evolutionary mechanisms of the hemagglutinin and three internal protein genes of influenza B virus: Multiple co-circulating lineages and frequent reassortment of the NP, M, and NS genes. Journal of Virology **1999**;73:4413-26.
24. Inlow J.K., Restifo L.L. Molecular and comparative genetics of mental retardation Genetics **2004**;166:835-81.
25. Awasthi G., Dash A.P., Das A. Evolutionary insights into Duffy gene in mammalian taxa with comparative genetic analyses. [Under review].
26. Piskur J., Schnackerz K.D., Andersen G., Bjornberg O. Comparative genomics reveals novel biochemical pathways. Trends in Genetics **2007**;23:369-72.