Hindawi Publishing Corporation
Comparative and Functional Genomics
Volume 2008, Article ID 879023, 6 pages
doi:10.1155/2008/879023

Research Article

Protein Coding Gene Nucleotide Substitution Pattern in the Apicomplexan Protozoa *Cryptosporidium parvum* and *Cryptosporidium hominis*

Guangtao Ge,[1] Lenore Cowen,[1] Xiaochuan Feng,[2] and Giovanni Widmer[2]

[1] *Department of Computer Science, Tufts University, Medford, MA 02155, USA*
[2] *Department of Biomedical Sciences, Tufts Cummings School of Veterinary Medicine, Tufts University, North Grafton, MA 01536, USA*

Correspondence should be addressed to Giovanni Widmer, giovanni.widmer@tufts.edu

Received 18 November 2007; Accepted 29 March 2008

Recommended by Neil Hall

Cryptosporidium parvum and *C. hominis* are related protozoan pathogens which infect the intestinal epithelium of humans and other vertebrates. To explore the evolution of these parasites, and identify genes under positive selection, we performed a pairwise whole-genome comparison between all orthologous protein coding genes in *C. parvum* and *C. hominis*. Genome-wide calculation of the ratio of nonsynonymous versus synonymous nucleotide substitutions (*dN*/*dS*) was performed to detect the impact of positive and purifying selection. Of 2465 pairs of orthologous genes, a total of 27 (1.1%) showed a high ratio of nonsynonymous substitutions, consistent with positive selection. A majority of these genes were annotated as hypothetical proteins. In addition, proteins with transmembrane and signal peptide domains are significantly more frequent in the high *dN*/*dS* group.

1. Introduction

Protozoan parasites belonging to the genus *Cryptosporidium* (Phylum: *Apicomplexa*) develop in the intestinal epithelium of many vertebrate species and frequently cause diarrhea. The two main species known to infect humans are *C. parvum* and *C. hominis*. Although highly similar at the genome level, these species differ in host range. Specifically, *C. parvum* infects humans and other mammals, particularly calves, but *C. hominis* is typically only found in humans. Human infections are mainly caused by the ingestion of drinking or recreational water contaminated with oocysts, an environmentally resistant form of the parasite.

The complete sequence of the *C. parvum* and *C. hominis* genomes was recently published [1, 2]. Both genomes are about 9 Mb in size and their gene complements appear to be identical, each coding for about 3900 genes. The *C. parvum* and *C. hominis* genomes exhibit only 3–5% sequence divergence and no large insertions or deletions have been identified [1]. Genome annotation has shown that energy metabolism pathways are present, but biosynthetic ability is limited. The pathways for de novo purine, pyrimidine, and amino acid synthesis are lacking. Consistent with the reliance of the intracellular developmental stages on the host cell for many metabolites, numerous genes encoding membrane transporters have been identified [1].

The analyses of nucleotide substitution patterns and genes under positive selection are of interest for understanding the evolution of these parasites and to improve genome annotation. Whereas "housekeeping genes" are under purifying selection to conserve many metabolic functions, genes that enable the parasite to adapt to different host environments may evolve more rapidly. Evolutionary analysis based on comparative genomics has been applied to different organisms, such as mammals [3–6], drosophila [7, 8], bacteria [9], and malaria [10, 11]. Here, we describe a genome-wide analysis of nucleotide substitution patterns in *C. parvum*/*C. hominis* orthologous protein coding genes.

2. Materials and methods

C. parvum and *C. hominis* protein coding gene nucleotide and amino acid sequences were downloaded from CryptoDB

[12] (http://cryptodb.org/cryptodb/) release 3.2. There are 3806 genes from *C. parvum* and 3886 genes from *C. hominis* in total.

To identify orthologous genes between *C. parvum* and *C. hominis,* best reciprocal hits were first generated with BLASTP using an all-versus-all method and a significance threshold $e < 10^{-10}$. The alignment regions were required to cover more than 80% of sequence lengths. This procedure resulted in 2465 pairs of orthologous gene sequences. To calculate *dN*/*dS*, amino acid sequences were first aligned using ClustalW1.83 [13] (http://www.ebi.ac.uk/clustalw/) with default parameters. Then, corresponding codon sequences were aligned based on the original protein sequence alignment using an in-house Perl script. The same process was also used on the raw ORF sequences to search for possible high *dN*/*dS* fragments that were erroneously annotated as being part of genes.

A PAML package implements three popular methods for calculating *dN*/*dS* values from sequence alignments [14] (http://abacus.gene.ucl.ac.uk/software/paml.html). These are a Nei and Gojobori method [15] based on a simple model of nucleotide substitution, and two more complex methods: ML94 and YN00 [16, 17]. ML94 method uses maximum likelihood framework to estimate *dN*/*dS* value directly based on codon substitution model, while YN00 method is another approximate counting method that incorporates the codon substitution model and unequal weight substitution pathways. We report results from the algorithm of Yang and Nielsen [16] that uses the codon model and nucleotide substitution model correction F84 and HKY85, and additionally weights pathways by their evolutionary probability. When we also calculated the *dN*/*dS* values using all three methods, the results were very similar. This is not surprising; results from the three methods are only expected to diverge significantly when the evolutionary distance between the organisms being compared is much larger.

Signal peptide motifs were predicted using SignalP v3.0 [18] at http://www.cbs.dtu.dk/services/SignalP/. Its neural network and hidden Markov model output YES or NO predictions individually. We considered one sequence to have a signal peptide if both, neural network and hidden Markov model, predicted positive results. Transmembrane domains were predicted using the TMHMM program v2.0 [19] at http://www.cbs.dtu.dk/services/TMHMM/. This program predicts the number of transmembrane domains and the number of amino acids in each domain. We considered any sequence to have a TM domain if at least one transmembrane helix included more than 18 amino acids. Protein functions were assigned based on sequence similarity search against NCBI nonredundant protein database.

To analyze the expression of selected genes by RT PCR, RNA was extracted from portions of the small intestine of an experimentally infected mouse. The gut was removed on day 14 post-infection from freshly euthanized animals and immediately chilled on ice. The epithelial cells were mechanically removed from the tissue. Total RNA was extracted from the cell scraping using Trizol reagent and further purified using a Qiagen RNA extraction kit. RNA was reverse transcribed into cDNA in the presence of poly-T primers. The absence of genomic DNA contamination from the RNA samples was confirmed by including an RNA control mock reverse transcribed in the absence of reverse transcriptase in each RT PCR experiment. We randomly selected 7 sequences from the high *dN*/*dS* ($dN/dS > 1$) group and 17 from the low *dN*/*dS* ($dN/dS < 1$) group. PCR primers for these sequences were designed using the LightCycler Probe Design software version 1.0 (Roche, Applied Science Indianapolis, IN.) with minimum cross complements. Amplifications were performed in a LightCycler (Roche Diagnostics) using FastStart SYBR Green I master mix (Roche Diagnostics). Initial denaturation was performed at 95°C for 10 minutes, followed by 40 cycles of amplification: denaturation at 95°C for 1 second, annealing at 64°C for 2 seconds, and extension at 72°C for 13 seconds. After amplification, melting curve analysis was performed as follows: initial denaturation at 95°C for 0 second, followed by annealing at 65°C for 15 seconds, and melting was performed in 0.1 degree increment per second until 95°C was reached for 0 second. After these procedures, amplification products were transferred from capillaries to microcentrifuge tubes, mixed with the loading buffer, and loaded onto a 1% agarose gel in a Tris-Acetate-EDTA buffer. Expression of selected sequences was confirmed if a single amplicon of the expected size was identified on ethidium bromide stained gels.

3. Results

To identify the effect of selective pressure on the *Cryptosporidium* proteome, we started our analysis with 2465 pairs of orthologous *C. parvum* and *C. hominis* genes identified by BLASTP reciprocal best hits. Figures 1 and 2 in Supplementary Material, available online at doi:10.1155/2008/879023, show the *dS* and *dN* value distributions for *C. parvum* and *C. hominis*. Most of gene sequences have a *dS* ratio below 0.0625 and a *dN* ratio smaller than 0.025. Using the ratio of nonsynonymous over silent substitutions (*dN*/*dS*) [20], Figure 1 shows the distribution of *dN*/*dS* values among 2465 orthologous *C. parvum*/*C. hominis* gene pairs. As in two other whole-genome studies [3, 11], we found that most of the genes (2438 of 2465, 98.9%) are under negative selection ($dN/dS < 1$), while the remaining 1.1% evolve neutrally or have accelerated rates. This distribution is consistent with most annotated genes being genuine, as opposed to mispredictions or pseudogenes. The median *dN*/*dS* was 0.1484, similar to what was found in the taxonomically related malaria parasites [11, 21], where, for example, the median *dN*/*dS* values of 0.27, 0.30, and 0.17 for three pairwise comparisons between two laboratory isolates of *P. falciparum* and a *P. reichenowi* and a *P. falciparum* isolate from Ghana were found. Among the *C. parvum*/*C. hominis* orthologous gene pairs, 391 (15.8%) sequences had a *dN*/*dS* ratio less than 0.05. At the other extreme of the distribution, 27 (1.1%) genes with $dN/dS > 1$ were identified. This high *dN*/*dS* gene percentage is slightly lower than those of previous reports from human/chimpanzee orthologs comparison [6, 22].

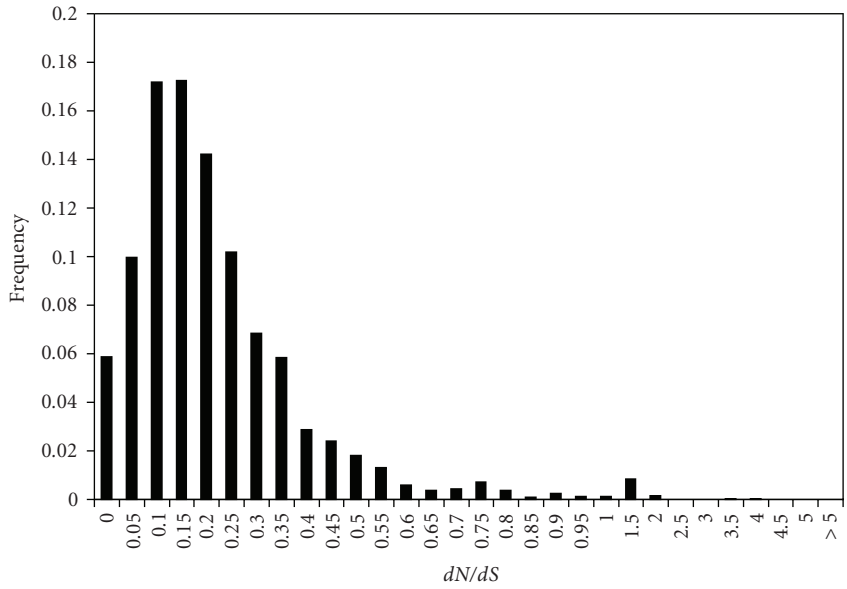

Figure 1: *dN/dS* ratio distribution of orthologous gene sequences from *C. parvum* and *C. hominis*. Black bars represent the frequency of gene sequences in each *dN/dS* category. For $dN/dS < 1$, category size is 0.05; for $dN/dS > 1$, category size is 0.5.

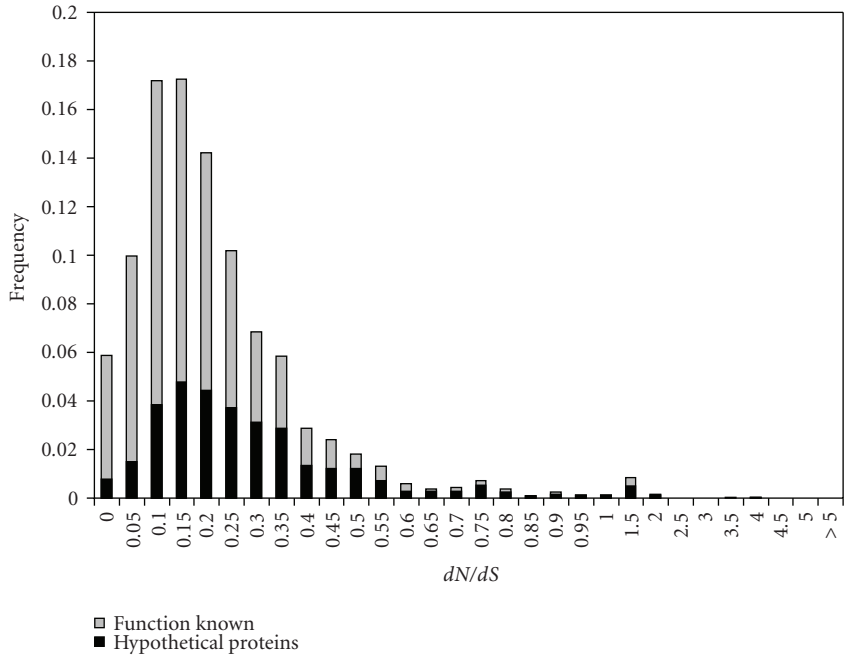

Figure 2: *dN/dS* ratio distribution of orthologous genes with or without functional annotation. Black bars represent gene with unknown function, while gray bars indicate known function.

To ensure that high *dN/dS* values did not originate from noncoding sequences erroneously annotated as genes, 7 *C. parvum* genes were randomly selected from 27 orthologs with $dN/dS > 1$ for RT-PCR analysis. Specifically, under the assumption that such genes were not a result of misannotation, we expected to detect mRNA transcripts from a similar ratio of genes with $dN/dS > 1$ as from genes with much lower *dN/dS* values. Thus, as controls, *dN/dS* transcripts with $dN/dS < 1$ were analyzed in parallel. For this analysis, *C. parvum* RNA was extracted from the intestinal epithelium of experimentally infected mice. Because such animal infections are not synchronized, we expected these RNA samples to contain transcripts of genes expressed at different stages of the intracellular life cycle. The RT PCR analysis successfully detected transcripts from 2 of 7 high *dN/dS* and 4 of 17 low *dN/dS* sequences (Table 1). The similar level of RT-PCR detection in sequences with high and low *dN/dS* indicates that high *dN/dS* genes are unlikely to have originated from

Table 1: Summary of RT-PCR results.

	RT PCR detection	
	$dN/dS > 1$	$dN/dS < 1$
Total number of genes[(1)]	7	17
Number of detected	2	4
dN/dS (mean)	1.16	0.08
TM/SP (%)[(2)]	100	12
Annotated (%)[(3)]	28	88

[(1)] Number of transcripts which were detected or not detected by RT PCR
[(2)] Percent of transcripts predicted to encode a transmembrane domain or a signal peptide
[(3)] Percent of transcripts with annotated function.

the alignment of noncoding sequences erroneously identified as genes.

With genomes as close as *C. parvum* and *C. hominis*, there is always a possibility that high dN/dS values are found by chance. This is particularly a concern for the YN00 estimation method [16], which uses multiple correction terms for distant genomes. To verify our gene sets, we also ran the Nei and Gojobori NG86 method [15] and maximum likelihood method ML94 [17] on the data, which we would expect to show similar result with closely related genomes as is the case for *C. parvum* and *C. hominis*. In fact, for example, 22 of 27 gene pairs still exhibited $dN/dS > 1$ when ML94 method were used, whereas the remaining five all had $dN/dS > 0.84$.

For 797 (32.3%) of the 2465 orthologous gene pairs included in the dN/dS analysis the function is unknown. The remaining 67.7% (1668) have been assigned functional predictions based on homology to genes in other species at NCBI nonredundant protein database. Consistent with the difficulty of annotating fast-evolving genes, for high dN/dS genes ($dN/dS > 1$) putative functional annotation was rare: 59.2% of sequences with $dN/dS > 1$ (16 of 27) are annotated as "hypothetical protein." Table 1 in supplementary material. In contrast, only 18.5% of sequences with $dN/dS < 0.1$ (151 of 814) are classified as hypothetical. The dN/dS distributions of these two groups of genes are significantly different ($P < 6.51\text{e-}32$, Komolgorov-Smirov (KS) test) as shown in Figure 2. The statistics of these distributions are summarized in Table 2.

Host-*Cryptosporidium* interaction is believed to involve many extracellular proteins with transmembrane (TM) domains or signal peptides (SP) [23]. Such proteins are likely to be exposed to the host immune response and may therefore evolve rapidly [24]. To investigate this hypothesis, we predicted TM and SP domains across all *C. parvum*/*C. hominis* orthologous gene pairs. A total of 16 of 27 sequences (59.2%) with $dN/dS > 1$ were predicted to have TM or SP domains, compared to 192 of 814 sequences (23.5%) with $dN/dS < 0.1$. We also observed that a higher percentage of sequences in the TM/SP group has $dN/dS > 1$ (2.2%), as compared to 0.6% of sequences without TM/SP. Figure 3 shows the dN/dS value distributions for these two groups. The two distributions are significantly different ($P < 1.86\text{e-}7$, KS test), consistent with many membrane proteins being under positive selection.

4. Discussion

The large proportion of *C. parvum* and *C. hominis* genes without functional annotation indicates the extent to which the genome of these parasites has been shaped by the adaptation to a complex life cycle, and to the host and outside environment. Bioinformatics contributes to our understanding of these genomes by flagging proteins showing signs of extreme selection. A computational approach is particularly relevant for studying pathogens such as *Cryptosporidium* species, which are difficult to culture, cannot be genetically manipulated, and are resistant to antiprotozoal drugs.

Cryptosporidium genomes are not intron-rich; only 5%–20% of genes are predicted to have multiple exons [1, 2]. Therefore, we also performed our dN/dS analysis using ORF sequences, so that we might find high dN/dS genes that were missed by the automatic gene annotation used in CryptoDB. We estimated selection pressures on *C. parvum* and *C. hominis* homologous ORF pairs using the same dN/dS statistics. A total of 251 out of 4279 orthologous ORFs (5.8%) showed signs of positive selection. One possible reason for the higher percentage is that some ORFs represent only parts of whole gene sequences, and averaging over multiple ORFs will lower dN/dS value of whole gene. For example, ORF CpIOWA_V_AAEE01000007-1-707587-707922 has a dN/dS value of 1.79, but further analysis showed that it is only a part of a protein which has a much lower dN/dS of 0.054. Nonetheless, it may be worth in future studies to examine these high dN/dS ORF regions more closely to elucidate if any high dN/dS genes were missed by gene finding methods.

Given the low divergence between *C. parvum* and *C. hominis*, estimation of dN/dS value may be inaccurate due to stochastic noise and a low level of substitutions. Most of the gene sequences have dS below 0.0625 (Supplemental Figure 1), which means that some of the high dN/dS values may be due to chance. Previous studies between closely related species such as human and chimpanzee showed that a large number of genes under positive selection are expected to occur by chance [6, 22]. The detection in our analyses of high dN/dS sequences with two different methods together with a significant enrichment for TM/SP sequences in these genes indicates that our analysis detects rapidly evolving genes.

As with other pathogens, sequences with TM/SP structures are expected to be directly involved in host-pathogen contact. In our study we identified a large proportion of sequences with predicted TM or SP motifs that have elevated dN/dS ratios. This observation is consistent with other studies. For instance, Jeffares et al. reported that *P. falciparum* genes mediating host cell invasion have high dN/dS ratios [11]. These authors also examined the substitution rates of genes according to cellular location. Significantly, more rapid rates of evolution were observed for predicted membrane-spanning proteins and exported proteins, as compared to proteins localized to the nucleus, cytoplasm, and mitochondrion. In another study comparing

Table 2: Summary of *dN/dS* analysis.

	Classification by structure		Classification by annotation	
	With TM/SP[2]	Without TM/SP	Hypothetical	Known function
Number of genes[1]	721	1744	797	1668
Median *dN/dS*	0.1762	0.1388	0.2124	0.1276
KS[3] test *P* value	1.86e-7		6.51e-32	

[1] Number of orthologous gene sequences between *C. parvum* and *C. hominis*
[2] Transmembrane or signal peptide
[3] Kolmogorov-smirnov test.

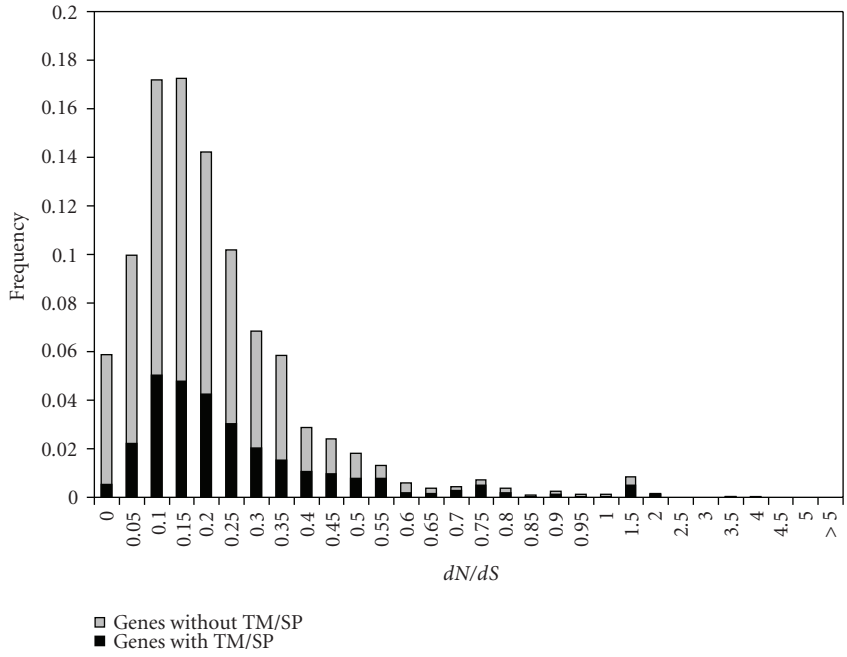

Figure 3: *dN/dS* ratio distribution of orthologous gene sequences with or without TM/SP structural domains. Black bars represent genes with TM/SP structural domain while gray bar represents the gene without TM/SP. TM: transmemebrane domain; SP: signal peptide.

the genome of the rodent malaria species *P. berghei* and *P. chabaudi* [10], a significant difference between *dN/dS* values in TM/SP-containing and non-TM/SP-containing genes was found in proteins expressed during the blood stage, but not in genes expressed in the mosquito vector. A similar trend was also reported in a comparative study of the two *Theileria* species *T. annulata* and *T. parva* [25]. Whether similar life cycle dependent differences also occur in *Cryptosporidium* species or are a characteristic of *Apicomplexa* of the blood is an interesting topic for future investigation.

Even though many of the high *dN/dS* sequences in the *C. parvum*/*C. hominis* genome comparison remain without annotation, the few annotated genes in this category are worth mentioning. In particular, the genes with the second and sixth highest *dN/dS* values, (cgd2_440 with *dN/dS* > 3.36 and cgd2_220 with *dN/dS* > 1.5, resp.) are annotated as secreted mucins. Such mucins have been shown to be involved in attachment and invasion of host intenstinal epitethelial cells by *C. parvum* sporozoites and are crucially involved in pathogenesis of cryptosporidiosis [26]. They are found on the cell surface and apical region of invasive stages and are shed in trails during gliding motility [26, 27]. Another high *dN/dS* protein is from the DHHC family of palmitoyl transferases (cgd5_1260, with *dN/dS* > 1.46), which are enzymes involved in cellular signaling and membrane trafficking. In some parasitic protozoa palmatoylated proteins have been predicted to play a role in the evasion of the host immune response [28, 29].

Acknowledgments

The authors thank Donna Slonim for many helpful discussions and suggestions and Yi-Lin Yang for assistance with bench experiments. The first and second authors were supported in part by NSF grant (ASE+NHS) (dms) 0428715. Financial support from the National Institute of Allergy and Infectious Diseases (GW) is gratefully acknowledged. The authors thank anonymous reviewers for their insightful suggestions.

References

[1] P. Xu, G. Widmer, Y. Wang, et al., "The genome of *Cryptosporidium hominis*," *Nature*, vol. 431, no. 7012, pp. 1107–1112, 2004.

[2] M. S. Abrahamsen, T. J. Templeton, S. Enomoto, et al., "Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*," *Science*, vol. 304, no. 5669, pp. 441–445, 2004.

[3] L. Arbiza, J. Dopazo, and H. Dopazo, "Positive selection, relaxation, and acceleration in the evolution of the human and chimp genome," *PLoS Computational Biology*, vol. 2, no. 4, pp. 288–300, 2006.

[4] A. G. Clark, S. Glanowski, R. Nielsen, et al., "Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios," *Science*, vol. 302, no. 5652, pp. 1960–1963, 2003.

[5] D. J. Lynn, A. R. Freeman, C. Murray, and D. G. Bradley, "A genomics approach to the detection of positive selection in cattle: adaptive evolution of the T-cell and natural killer cell-surface protein CD2," *Genetics*, vol. 170, no. 3, pp. 1189–1196, 2005.

[6] R. Nielsen, C. Bustamante, A. G. Clark, et al., "A scan for positively selected genes in the genomes of humans and chimpanzees," *PLoS Biology*, vol. 3, no. 6, pp. 976–985, 2005.

[7] J. C. Fay, G. J. Wyckoff, and C.-I. Wu, "Testing the neutral theory of molecular evolution with genomic data from *Drosophila*," *Nature*, vol. 415, no. 6875, pp. 1024–1026, 2002.

[8] N. G. C. Smith and A. Eyre-Walker, "Adaptive protein evolution in *Drosophila*," *Nature*, vol. 415, no. 6875, pp. 1022–1024, 2002.

[9] S. L. Chen, C.-S. Hung, J. Xu, et al., "Identification of genes subject to positive selection in uropathogenic strains of *Escherichia coli*: a comparative genomics approach," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 15, pp. 5977–5982, 2006.

[10] N. Hall, M. Karras, J. D. Raine, et al., "A comprehensive survey of the *Plasmodium* life cycle by genomic, transcriptomic, and proteomic analyses," *Science*, vol. 307, no. 5706, pp. 82–86, 2005.

[11] D. C. Jeffares, A. Pain, A. Berry, et al., "Genome variation and evolution of the malaria parasite *Plasmodium falciparum*," *Nature Genetics*, vol. 39, no. 1, pp. 120–125, 2007.

[12] D. Puiu, S. Enomoto, G. A. Buck, M. S. Abrahamsen, and J. C. Kissinger, "CryptoDB: the *Cryptosporidium* genome resource," *Nucleic Acids Research*, vol. 32, database issue, pp. D329–D331, 2004.

[13] J. D. Thompson, D. G. Higgins, and T. J. Gibson, "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice," *Nucleic Acids Research*, vol. 22, no. 22, pp. 4673–4680, 1994.

[14] Z. Yang, "PAML: a program package for phylogenetic analysis by maximum likelihood," *Computer Applications in the Biosciences*, vol. 13, no. 5, pp. 555–556, 1997.

[15] M. Nei and T. Gojobori, "Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions," *Molecular Biology and Evolution*, vol. 3, no. 5, pp. 418–426, 1986.

[16] Z. Yang and R. Nielsen, "Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models," *Molecular Biology and Evolution*, vol. 17, no. 1, pp. 32–43, 2000.

[17] N. Goldman and Z. Yang, "A codon-based model of nucleotide substitution for protein-coding DNA sequences," *Molecular Biology and Evolution*, vol. 11, no. 5, pp. 725–736, 1994.

[18] J. D. Bendtsen, H. Nielsen, G. von Heijne, and S. Brunak, "Improved prediction of signal peptides: signalP 3.0," *Journal of Molecular Biology*, vol. 340, no. 4, pp. 783–795, 2004.

[19] A. Krogh, B. Larsson, G. von Heijne, and E. L. L. Sonnhammer, "Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes," *Journal of Molecular Biology*, vol. 305, no. 3, pp. 567–580, 2001.

[20] M. Nei and S. Kumar, *Molecular Evolution and Phylogenetics*, Oxford University Press, Oxford, UK, 2000.

[21] M. J. Gardner, N. Hall, E. Fung, et al., "Genome sequence of the human malaria parasite *Plasmodium falciparum*," *Nature*, vol. 419, no. 6906, pp. 498–511, 2002.

[22] T. S. Mikkelsen, L. W. Hillier, E. E. Eichler, et al., "Initial sequence of the chimpanzee genome and comparison with the human genome," *Nature*, vol. 437, no. 7055, pp. 69–87, 2005.

[23] S. Tzipori and H. Ward, "Cryptosporidiosis: biology, pathogenesis and disease," *Microbes and Infection*, vol. 4, no. 10, pp. 1047–1058, 2002.

[24] M. U. Ferreira, W. L. Ribeiro, A. P. Tonon, F. Kawamoto, and S. M. Rich, "Sequence diversity and evolution of the malaria vaccine candidate merozoite surface protein-1 (MSP-1) of *Plasmodium falciparum*," *Gene*, vol. 304, no. 1-2, pp. 65–75, 2003.

[25] A. Pain, H. Renauld, M. Berriman, et al., "Genome of the host-cell transforming parasite *Theileria annulata* compared with *T. parva*," *Science*, vol. 309, no. 5731, pp. 131–133, 2005.

[26] A. M. Cevallos, N. Bhat, R. Verdon, et al., "Mediation of *Cryptosporidium parvum* infection in vitro by mucin-like glycoproteins defined by a neutralizing monoclonal antibody," *Infection and Immunity*, vol. 68, no. 9, pp. 5167–5175, 2000.

[27] L. D. Sibley, "Intracellular parasite invasion strategies," *Science*, vol. 304, no. 5668, pp. 248–253, 2004.

[28] S. Delbecq, E. Precigout, A. Vallet, B. Carcy, T. P. M. Schetters, and A. Gorenflot, "*Babesia divergens*: cloning and biochemical characterization of Bd37," *Parasitology*, vol. 125, no. 4, pp. 305–312, 2002.

[29] R. Behnia and S. Munro, "Organelle identity and the signposts for membrane traffic," *Nature*, vol. 438, no. 7068, pp. 597–604, 2005.