Calculating Adaptation Rates in the Human Genome

Using the McDonald-Kreitman Test

Drake Sadosky, Partnered with Minhao Chen
ECOL 346 - Bioinformatics - The University of Arizona

Abstract:

The malaria parasite, *Plasmodium falciparum,* interacts with hundreds of proteins in the human genome. *Plasmodium*-resistant alleles have developed within human populations, and only recently have researchers began to study host adaptation to the *Plasmodium* parasite. Here, we present a comparison of adaptation rates between Plasmodium Interacting Proteins and the human genome in its entirety using a modified version of the McDonald-Kreitman Test. Using polymorphism data obtained from the 1000 Genomes Project, we show that the adaptation rate of PIPs is almost equal to that of the human genome as a whole. The equal rates of adaptation do not conclusively rule out positive selection, but suggest that there are other factors related to the relative genetic resistance against the plasmodium parasite in the human genome.

Introduction:

With over 200 million cases of the disease recorded worldwide in 2017, malaria is among the world's most villainous diseases (1). Malaria is caused by multiple species in the genus *Plasmodium*, and is known to cause serious symptoms, including death, in some hosts (2). Multiple genes within the human genome have been associated with defense against the some

of the ill effects of Malaria, and host parasite interactions have been shown to be an overwhelming contributor in adaptation within the human genome (3, 4). In this paper, we present a method of analyzing the adaptation rates of Plasmodium Interacting Proteins (PIPs) within the human genome.

Positive Selection

The process in which beneficial alleles rise in frequency throughout a population, known as positive selection, has long been a topic of debate for population geneticists (5). When beneficial alleles arise in a population, they are likely to increase in frequency until fixation. In contrast, deleterious alleles (alleles detrimental to an organism's fitness) will not remain in a population at high frequencies; organism's carrying a deleterious allele will reproduce less efficiently, and not pass the allele to future generations at a high rate. However, deleterious alleles may remain in the population for a relatively extended period of time at small frequencies, presenting a potential problem for adaptation rate calculation (6, 7).

McDonald-Kreitman Test

To calculate the rates of adaptation at selected loci, population geneticists have developed a method to compare the ratio of synonymous polymorphisms to nonsynonymous polymorphisms within a species and compared to a phylogenetic outgroup (8). Known as the McDonald-Kreitman Test (MK Test), this statistical tool compares divergence within a species to the divergence between species. If positive selection is at work, the amount of fixed differences between two species will be greater than the variation within a single species **(**see supplemental material**)**.

Materials & Methods:

To perform the MK Test, polymorphism data specifying the number of nonsynonymous polymorphisms (P_n), synonymous polymorphisms (P_s), synonymous fixed differences (D_n), and nonsynonymous fixed differences (D_s) is required of an organism in relation to at least one outgroup. For our purposes, we used polymorphism data from Human African Populations provided by Dr. David Enard. All code, as well as sample input data has been uploaded to github.

MK Test: Whole Genome

To conduct the MK Test on the human genome in its entirety, a python script was developed to

first search for alleles present in the population at frequencies greater than .5000. This step is important as slightly deleterious alleles may be present in small frequencies in the population, causing an underestimate of the alpha value (6). After excluding alleles present in frequencies under .5000, the script returned an alpha value representing the average adaptation rate of specified alleles within the genome. It is important to note that this is a modified version of the McDonald-Kreitman Test, due to the selection of alleles present at frequencies above .5000.

MK Test: *Plasmodium* Interacting Proteins

To calculate the adaptation rates of PIPs within the human genome, annotation regarding alleles' interaction with the *Plasmodium* parasite was obtained from Dr. David Enard. A python script was developed to perform a modified version of the MK Test on PIPs present in the population at frequencies greater than .5000.

Results and Discussion

Our results suggest that when compared to the genome as a whole, PIPs are not undergoing adaptation at an accelerated rate. Alpha values for the the entire genome compared PIPs were relatively equal at .0914 and .0966 respectively. While this does not conclusively rule out positive selection as a dominant driver in the partial resistance to malaria within the human genome, it does not sufficiently explain the partial genetic resistance to malaria seen in some populations.

As noted by Enard et al., host-viral interactions contribute substantially to the overall adaptation rates of the human genome, with a reported alpha value of .224 for Viral Interacting Proteins (noted VIPs) (9). The difference in alpha values between PIPs and VIPs suggest that viruses contribute to the overall adaptation rate of the human genome at a higher rate compared to the *Plasmodium* parasite.

The relatively low adaptation rate of PIPs in the human genome suggests that factors other than positive selection could be contributing to the partial genetic resistance to malaria seen in some human populations. Further research into the overall level of PIP expression within the genome could be a possible route of further exploration.

A possibly novel route of further research could include neural-network assisted prediction of structural amino acid change in context of environment. Understanding at which amino acid structural mutation is occuring could help predict amino acid change in context of a parasitic rich environment environment. While the MK Test can suggest whether a loci is undergoing adaptation, it does not she insight as to what type of mutation occurred in context of amino acid.

References

1. “Fact Sheet about Malaria.” *World Health Organization*, World Health Organization, www.who.int/news-room/fact-sheets/detail/malaria.
2. Ebel, Emily R., et al. “High Rate of Adaptation of Mammalian Proteins That Interact with Plasmodium and Related Parasites.” *PLOS Genetics*, vol. 13, no. 9, 2017, doi:10.1371/journal.pgen.1007023.
3. Verra, F., et al. “Genetics of Susceptibility To Plasmodium Falciparum: from Classical Malaria Resistance Genes towards Genome-Wide Association Studies.” *Parasite Immunology*, vol. 31, no. 5, 2009, pp. 234–253., doi:10.1111/j.1365-3024.2009.01106.x.
4. Mitchell-Olds, Thomas. “Faculty of 1000 Evaluation for Evidence for Polygenic Adaptation to Pathogens in the Human Genome.” *F1000 - Post-Publication Peer Review of the Biomedical Literature*, 2013, doi:10.3410/f.718003958.793487620.
5. Enard, David, et al. “Viruses Are a Dominant Driver of Protein Adaptation in Mammals.” *ELife*, vol. 5, 2016, doi:10.7554/elife.12469.
6. Messer, P. W., and D. A. Petrov. “Frequent Adaptation and the McDonald-Kreitman Test.” *Proceedings of the National Academy of Sciences*, vol. 110, no. 21, 2013, pp. 8615–8620., doi:10.1073/pnas.1220835110.
7. Allendorf, Fred W. “Genetic Drift and the Loss of Alleles versus Heterozygosity.” *Zoo Biology*, vol. 5, no. 2, 1986, pp. 181–190., doi:10.1002/zoo.1430050212.
8. Mcdonald, John H., and Martin Kreitman. “Adaptive Protein Evolution at the Adh Locus in Drosophila.” *Nature*, vol. 351, no. 6328, 1991, pp. 652–654., doi:10.1038/351652a0.
9. Uricchio, Lawrence H, et al. “Exploiting Selection at Linked Sites to Infer the Rate and Strength of Adaptation.” 2018, doi:10.1101/427633.