ADAPTIVE EVOLUTION IN PRIMATE IMMUNE RECEPTORS

by

NICOLE MARIE PATERSON

A DISSERTATION

Presented to the Department of Chemistry and Biochemistry and the
Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

June 2021

DISSERTATION APPROVAL PAGE

Student: Nicole Marie Paterson

Title: Adaptive Evolution in Primate Immune Receptors

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Chemistry degree in the Department of Chemistry and Biochemistry by:

Michael Harms	Chairperson
Matthew Barber	Advisor
Diane Hawley	Core Member
Andrew Kern	Core Member
Nelson Ting	Institutional Representative

and

Andy Karduna	Interim Vice Provost for Graduate Studies

Original approval signatures are on file with the University of Oregon Division of Graduate Studies.

Degree awarded June 2021

DISSERTATION ABSTRACT

Nicole Marie Paterson
Doctor of Philosophy
Department of Chemistry and Biochemistry
June 2021

Pathogens and parasites have evolved effective strategies to gain access to host resources. The immune system fends off these attacks, often through detection of pathogen associated molecules and clearance of infection. This results in interactions between host and pathogen that often take place at molecular interfaces of immune receptors that act as a first line of defense to infection. Such receptors must identify pathogen-specific molecules and mount an appropriate response. Due to the frequency of such high stakes interactions between immune receptors and pathogen-derived molecules, the immune system is under constant evolutionary pressure to innovate new modes of defense and detection, while the pathogen is under pressure to evade these efforts and mount offensive attacks. This dynamic, called evolutionary conflict, is the underlying evolutionary principle inspiring this work. Because proteins evolve functions through DNA modifications, we study the effects of nucleotide variation across related species and test how variation affects the dynamics of protein interactions. We show that phylogenetic relationship is not a good indicator of functional similarity in the systems we tested. In the first study in Chapter III, we found that presence of certain amino acids in ligand-binding hotspots are more likely to have an effect on whether a Staphylococcus aureus inhibitor binds to immune receptor than overall sequence homology. In the second study that comprises Chapter IV, we found a similar lack of correlation between predicted functional outcomes and familial relationship. Similar to the study in

Chapter III, we found that certain sites could have an outsized effect on function that could be translated across multiple species. Interestingly, site-level similarities at "hotspot" regions were a better indicator of function than phylogenetic relationships.

This dissertation includes previously published and unpublished co-authored material.

CURRICULUM VITAE

NAME OF AUTHOR: Nicole Marie Paterson

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene

Portland State University, Portland, Oregon

DEGREES AWARDED:

Doctor of Chemistry, Biochemistry, 2021 University of Oregon

Master of Science, Chemistry, 2019, University of Oregon

Bachelor of Arts, Biology/Liberal Studies, 2012, Portland State University

AREAS OF SPECIAL INTEREST:

Molecular Evolution Evolutionary Biochemistry Applied Bioinformatics

PROFESSIONAL EXPERIENCE:

Research Assistant II, Portland VA Research Foundation, 2013-2015

Research Assistant I, OHSU School of Dentistry, 2012-2013

GRANTS, AWARDS, AND HONORS:

NIH Developmental Biology Training grant appointee, 2018
Chemistry Department Travel Award, 2018
Lucille Adam Borgensen Memorial Scholarship, Outstanding New Student Award 2015

PUBLICATIONS:

Nicole M Paterson, Hussein Al-Zubieri, Brian Geisbrecht, Matthew F Barber. "Dynamic evolution of bacterial ligand recognition by formyl peptide receptors" bioRxiv. doi:TBD

Nicole M Paterson, Hussein Al-Zubieri, Matthew F Barber, "Diversification of CD1 Molecules Shapes Lipid Antigen Selectivity," Molecular Biology and Evolution, msab022 Feb. 2021, doi:10.1093/molbev/msab022

Ognjen Golub, Brett Wee, Rhonda A Newman, **Nicole M Paterson**, Kenneth E Prehoda, "Activation of Discs Large by aPKC Aligns the Mitotic Spindle to the Polarity Axis during Asymmetric Cell Division." eLife, vol. Nov. 2017, p. e32137. eLife, doi:10.7554/eLife.32137.

ACKNOWLEDGMENTS

I would not have completed this degree without personal guidance from Drs. Mike Harms and Diane Hawley. I would also like to acknowledge my amazing advisor Dr. Matthew Barber for his assistance in the preparation of this thesis as well as for excellent guidance in the academic, interpersonal and intellectual realms. Additionally, several professors went above and beyond namely Victoria DeRose, James Prell and Mark Lonergan. I had invaluable assistance from an exceptional undergraduate researcher whom I would also like to thank--Hussein Al-Zubieri, as well as two fantastic rotation students I had the pleasure of working with: Juan Tirado and Shannon Snyder. In addition, special thanks are due to the members of the Barber lab, whose willingness to discuss ideas and trouble-shoot experiments was tireless. I also thank the members of my committee for input on pivots required during this unusual time in research during the COVID-19 pandemic, as well as flexibility and morale boosts in response to tribulations faced over the last few years.

The investigations herein were supported in part by NIH Developmental Biology Training Grant and Public Health Service Predoctoral Fellowship, Number 6 RP BV-22, 731-03

This thesis is dedicated to my brother Miles, and to my goat Donkey.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Page

I. PREAMBLE

Summary of Chapters en bref

One question that has plagued evolutionary biologists is the degree to which evolutionary outcomes are predictable. We know that the effects of selection, which act upon mutations often results in non-random phenotypes in populations. Some studies suggest that the degree to which randomness dominates evolutionary outcomes is somewhat context-specific and linked to the strength of selection in an environment. Elegant studies on the co-evolution of phasmids with the plant hosts they feed on from the Nosil lab suggest that the influence of the host on co-evolving species can have a significant impact on the phenotype of the invertebrate species which feed on them. This study and many similar efforts, however, do not separate the effects of positive from negative selection. The studies undertaken here are focused on the role that positive selection (the process by which beneficial mutations spread in a population) plays in determining phenotypes at the molecular level of proteins. The first chapter of this thesis establishes the context of our studies, which use the mammalian immune system as a model to study the causes and consequences of rapid evolution. Chapter III is a study of how nucleotide sequence diversity affects the interaction of formyl peptide receptors with known S. aureus inhibitor and activator proteins. Surprisingly we found functional differences were not always linked with the greatest sequence divergence. Continuing to test the influence of sequence diversity in paralogous primate immune receptor families, Chapter IV is focused on a single receptor of a five-membered family, four of which show high signals of positive selection. We identified a single site predicted to have an outsized influence on ligand interactions, a primate innovation predicted to improve the function of the human receptor in silico. Lastly, the greater contributions of research such as this are offered to the reader, along with some musings regarding the application of similar studies in real-world settings outside the laboratory.

Characterizing functions of beneficial mutations in primates can reveal function of genetic variation

At the root of this work is a worthy question this thesis only partially addresses: How does genotype alter phenotype to facilitate adaptation? This work is a tiny part of a concerted effort by evolutionary biologists to identify randomly arising beneficial mutations and determine the function of such changes against a background that is largely noise. From a closer vantage, this work queries the effects of evolution through mutations. Mutations in DNA can affect the translation of codons, which modify the chemistry of proteins that may affect phenotype of the organism, thereby impacting fitness and the ability to propagate.

How can we detect mutations which will have the largest effect on phenotypes? Lucky for us we are not starting from zero: the field has already devised several clever methods of detecting molecular signatures of natural selection that increase our chances of isolating functional changes. Randomness may be built into evolution through the stochasticity of the mechanism, but randomly accumulated mutations are then acted upon by competition and/or the environment.

This work provides examples of how evolution promotes diversity of function through studies of immune receptors. Evolution in these systems occurs in a context with large influence over organismal fitness, the primate immune system, wherein fitness effects are often determined by the immune system's ability to detect and neutralize pathogens.

Through my work I observed a broad range of biological functions between species, even when genotypes are relatively similar. The differential response of human and bonobo FPR1 proteins (only six different codons) to bacterial molecules is one example of small changes resulting in relatively large effects on function. In another study, we saw that the Crab- eating (cynomolgous)

macaque CD1a receptor is predicted to perform much better at presenting ligands from the tuberculosis pathogen to T-cells than the human receptor, despite the fact that humans have a long history of association with this microorganism. In fact, we showed that a replacement of the amino acid in human with the amino acid in crab-macaque is predicted to improve antigen presentation by the human receptor. A key lesson from both of these studies is that evolution is a dynamic process, in the sense that there is no "finish line" or ultimate peak to climb. While biological systems may operate more efficiently when optimized by natural selection, the process operates in a non-linear fashion, sampling sequence space with mixed results.

Broader Impacts of this Work

Pathogens rapidly evolve immune evasion mechanisms against a background of changing environments. Globalization and climate change have altered ecological niches of a wide range of species and increased the likelihood of new zoonotic infections due to encroachment on wildlife habitats. In fact, at the time of this writing, human beings are experiencing the devastating effects of a worldwide pandemic. The etiological agent is a beta-coronavirus virus called SARS-COV-2, and while its mutation rate is not remarkable among viruses, new variants have quickly arisen that increase transmission and promote antibody evasion. Hospital-associated antibiotic resistance among other pathogens is also on the rise, due in part to the increases in hospitalizations from regions unprepared for the pandemic and resultant strain on hospital resources sequelae to the overwhelming demand for supportive care. An understanding of the evolution of microbes is essential to preparing for the next epidemic.

Part of an effectively tracing outbreaks in a public health crisis requires accurately discerning newly arising mutations of concern from those that are neutral, arising merely from drift. It may seem like a small facet, but it is an important feature of an agile public health response. Governments and

industry have begun to apply evolutionary theory on a wide scale in order to out-maneuver microbes. Work such as this can strengthen the understanding of mechanisms through the discovery of cryptic variation and identification of previously uncharacterized functions, as well as assist in practical approaches to decision-making by public health entities.

II. INTRODUCTION: RELEVANT EVOLUTIONARY CONCEPTS

Arms Race Dynamics and Van Valen's Red Queen Hypothesis

In "A New Evolutionary Law" published in 1973, Leigh Van Valen outlined what is now referred to as the Red Queen Hypothesis. Extinction of a species, he argues, is an inevitable side effect of existence that can be measured at a "stochastically constant" rate (Van Valen). His central argument lies in this seeming contradiction, how can a random process also have direction? He argues that within predator/prey dynamics adaptive zones exist where either party may have an advantage that can be subverted when a random mutation occurs giving one the advantage over the other: leading to an overall decline in fitness over time. This was against the prevailing idea that species evolve towards a fitness maxima at which point the evolutionary rate then slows (Brockhurst et al.). Evolution under Van Valen's argument does not progressing towards a fitness maxima where abiotic factors are the main force driving evolution, but towards extinction where the main force is competition with other species. Within what Van Valen calls the adaptive zones, the rate of change is negligible, which is the basis of the Red Queen analogy: neither predator nor prey is makes much headway until a large enough change occurs to shift into the next adaptive zone. In this manner, like Alice and the Red Queen in Lewis Carrol's novel, the two must maintain substantial speed just to maintain their original position in the fitness

landscape (Figure 1A).

Figure 1A. Alice and the Red Queen, accelerating at the same pace, appear unmoving although they must continue running just to stay in the same place

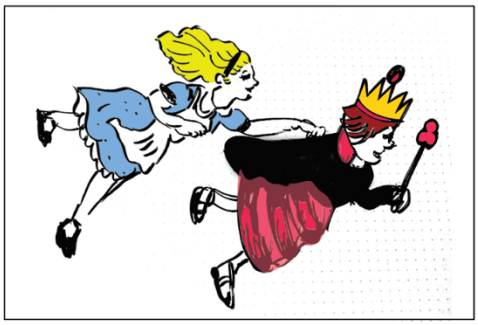


Figure modified from Lewis Carroll, 1871

Figure 1B. Fitness of species A and species RQ appear unchanging over time due to interdependent co-evolutionary relationship

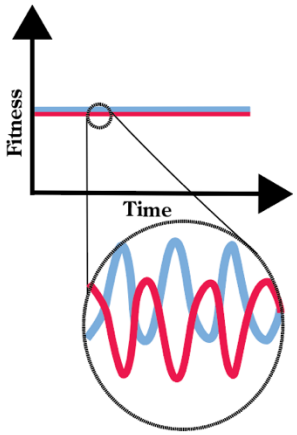


Figure modified from Strotz et al, Biol. Lett. 2018

Figure 1A, 1B. The Red Queen phenomenon. The result of natural selection to create adaptive zones wherein one species gains fitness at the expense of another (Figure 1B). The co-evolving species in this scenario regain fitness through the same means; fixation of an allele that confers some fitness benefit at a cost to the other species. This metaphor for the constant acceleration of adaptive co-evolution resulting in negligible increase in fitness benefit for either species is referred to as a "molecular arms race" at the protein level (Daugherty and Malik; Barber and Elde; McLaughlin and Malik).

Selection can be described in relationship to phenotype of the population as directional, stabilizing, or diversifying (Caballero). In directional selection, the mean fitness of the individuals shifts to the right or left of the mean, suggesting that an adaptive trait has moved through the population and shifted the phenotypic mean. Stabilizing selection occurs when extreme variants are not favored and so the mean phenotype is restricted to some central mean that is stable. Diversifying selection favors the extreme variants and results in bimodal distribution of phenotypes. What is common

to all three types of selection is the net fitness advantage at the population level. The mechanism by which selection has these effects on the phenotype of a population is through either positive or negative selection (Schultz). Broadly speaking, negative selection removes harmful alleles from the population while positive selection is the process by which beneficial alleles are selected for due to overall positive effects on fitness (Daugherty et al, Nei et al, Corona et al.). This thesis is focused on positive selection acting on the molecular components of life.

Detecting Positive Selection in a Dense Mist of Neutral Mutations

A discussion of methods for signatures indicative of positive selection in protein-coding sequences should be clarified. Methods differ for promoter and other genomic regions, and for our purposes we will just be discussing detection of positive selection at the codon level in protein coding alleles. When testing for positive selection at the gene and codon level, the null hypothesis usually tests whether a model fits the neutral theory of evolution put forward by Kimura: essentially, is this codon site or genomic region primarily subject to mutations conforming to random genetic drift or are there specific alleles or changes preferred at a site? This involves analysis of many sequences to set an accurate baseline for drift, which is defined by models as the rate of synonymous substitutions occurring at a site. The models detect positive or negative selection against the assumption that background is composed of neutral genetic drift. The random-ness of genetic drift, described as a "drunken walk" across sequence space expects that stochastically arising mutations will have one of three fates: removed if they are deleterious, maintained if there are neutral, and fixed in the population to outcompete wildtype if they are beneficial. As such when codon sites or a gene as a whole diverges from a random rate or evolution, it is implied that some form of positive or negative selection is likely to be acting on that site.

The neutral theory states:

If mutants are selectively equivalent to the pre-existing forms from which they are derived, their fate is left to chance and their frequencies increase or decrease fortuitously as time goes on. (Kimura)

Based on the assumption that neutral alleles will increase/decrease by chance, biologists seeking to detect sites under selection in the simplest case test for divergence at a codon site from randomness, where the ratio of non-synonymous to synonymous mutations per nonsynonymous or synonymous site is about equal to one. Mutations that encode different amino acids from the original sequence are referred to as nonsynonymous substitutions, versus those that encode the same amino acid called synonymous mutations. This test assumes ample time and adequate population size for natural selection to act upon allele frequency. If a change at a site is beneficial under this model, it will appear at a rate greater than expected by chance and therefore increase the value of the numerator. Synonymous mutations act as a control for evolutionary time in a sense as well, if we assume that synonymous changes are effectively neutral and selection does not act on them, we also assume that they accumulate over time where permitted and are immune to the effects of phenotypic selection, although too much evolutionary time between sites under comparison can affect analysis through the accumulation of many nonsynonymous substitutions.

This simple premise is referred to dN/dS (Ka/Ks or omega, "ω") and has been employed since the method was originally described in 1980 as a way to compare substitution rates to a molecular clock in mRNA across different genes to obtain insight into sites diverging from the null hypothesis of neutral drift (Miyata et al.). The most widely used interface for detecting sites under selection was developed in 2007 by Ziheng Yang. Phylogenetic Analysis by Maximum Likelihood (PAML), which incorporates the logic of Ka/Ks into a Bayesian framework and incorporates phylogenetic information and branch lengths to output ranked sites which are likely to have been

positively selected, with an associated p- value based on chi-squared statistical analysis.

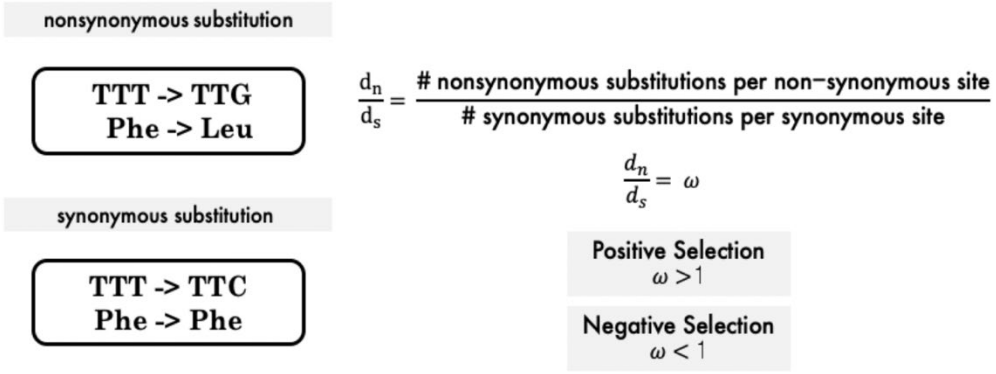


Figure 2. Explanation of omega calculations. Non-synonymous mutations are nucleotide changes that result in a different amino acid produced during translation. Synonymous substitutions, also called silent mutations, are nucleotide changes that do not result in difference. A ratio greater than one of the number of the former types to the latter type per site is one method of detecting codon positions where positive selection may be occurring.

While this method is useful to identify rare mutations that may have an outsized effect on phenotype, it does have its limitations. Many sites under selection are thought to have a subtle phenotype, and major effect mutations that drastically alter fitness are thought to occur rarely. These sites are, however, hotly pursued by scientists. Studies that identify major effect mutations with outsized effects are well-received because they can lead to promising drug targets or insight into molecular mechanisms of important biomolecules. Quantitative genetics reveals that since beneficial mutations with significant effect are so vanishingly rare in most cases (especially with regards to cumulative effect traits that quantitative genetics is expressly concerned with), even when one discovers a beneficial mutation it is most likely to have a minor effect on fitness as Kimura described:

Although a large number of mutants arise in each generation in any reasonably large population, the majority are lost by chance within a small number of generations (Fisher, 1930b; Kimura and Ohta, 1969b). It is often not realized that this is true not only for deleterious and selectively neutral mutants but also for advantageous mutants unless the advantage is very large. For example, if a mutant has a 1% selective advantage, the chance is only about 2% that it will eventually spread through the whole population (Haldane, 1927; Fisher, 1930b). In the remaining 98% of the cases, it will be lost by chance from the population without being used in evolution (Kimura).

This begs the question, how can we find these vanishingly rare phenotypes? Do they exhibit molecular signatures that we can detect, or better yet can we use existing models to simply rank sites under selection by their predicted effect on biology?

In order to approach this problem, we must first define what is missing from current models. Some more modern approaches to this problem have incorporated an intriguing binning method into detections of sites under selection: that of the conservative (substitutions that replace a codon with another of similar property, size, or other aspect) or radical (substitutions that replace a site with an amino acid that differs in these features). The larger the difference, the larger the coefficient of the radical property.

Later, in Chapter IV I will propose that we take contextual information (amino-acid level biochemistry) and apply it to the original PAML framework (individual and population- level phylogenetics) to remove sites that are functionally uninteresting or minor from the analysis. Such contextual information may assist researchers to further parse bioinformatics data into sites that are likely to have a substantial effect from those that are likely to contribute some minor and because it is so, temporary genotype unlikely to fix in the population. In the meantime, I hope to explain why the immune system is an excellent system for studying adaptive evolution.

Brief Overview of Immune System Molecular Evolution

Conceptual aspects of vertebrate immunity were put in place during the onset of multi- cellularity. Differentiating between self and non-self is a central aspect to building cooperativity in a complex organism and identifying invaders. Specific delineations for self- and non-self are required for the largescale organization of complex patterns of gene expression coordinated across numerous cell types to create and maintain integrity of intricate animals, such as primates. This foundational concept leads logically to the next question: what happens if non-self is detected? What systems are in place to protect shared resources within an organism?

In sponges, a relatively complicated immune system has been observed and described as early as 1892 with the observation of phagocytic cells that engulf foreign bacteria (Mu̇ller and Mu̇ller; Wiens et al.). Sponges possess a Toll-like receptor system which acts as a rudimentary pathogen associated pattern receptor, four caspases, and are observed to reject allografts (Mu̇ller and Mu̇ller; Wiens et al.). Animals relied on innate immunity to defend from parasites and pathogens until innovations detected in cartilaginous fishes (Suckale et al.). Innate approaches to immunity are effective, and subsequent innovations such as the complement system, Leucine rich repeat receptors, Nod like-receptors and others were adaptive inclusions to the innate immune response over time (Suckale et al).

Early adaptive immune receptor variable lymphocyte receptor (VLR) genes include some elements of innate immune receptor structure such as leucine rich repeats (LRR) receptors. A mature VLR requires combinatorial processing that results in a wide array of possible VLR sequences, which is thought to be the beginnings of how variability was first generated in the ancient immune system (Deng et al.).

The adaptive immune system is thought to have evolved as a unit in jawed fishes, since the most ancestral version of this adaptive unit was observed to have emerged altogether in the nurse shark: complete with proto-versions of RAG-genes, T-cell receptor and major histocompatibility complex (Flajnik and Kasahara). This system evolved in complexity, eventually gaining the ability to generate the large amount of sequence diversity necessary for mammalian B-cells to create highly specific antibodies tailored to an infection, a process referred to as somatic hypermutation (Di Noia and Neuberger).

The immune system in vertebrates has evolved massive complexity from the basis of a self/non-self recognition framework. The two branches of the immune system, innate and adaptive, both use receptors called pattern recognition receptors to identify invaders, but respond differently to activation by foreign molecules. Broadly speaking, the innate immune response relies on immediate and generalized methods to eliminate pathogens, while the adaptive immune response relies on the generation of highly specific antibodies for pathogen neutralization (Cooper and Alder; Rosenstiel et al.; Kasahara and Flajnik; Deng et al.).

An interesting feature of many immune-system related genes is their presence as large paralogous gene clusters. Many of the pattern recognition receptors, such as major histo- compatibility complex (MHC), the related CD1 molecules which display lipids instead of peptides, formyl peptide receptors (FPRs) which bind bacterial peptides and other metabolites, Toll-Like receptors (TLR) which identify viruses, bacteria and other ligands are just a few examples of well-known PRRs that evolved distinct ligand binding profiles through gene duplication and neo-functionalization (Cooper and Alder). Interestingly, this process is on-going and copy number and ligand binding profiles diverge widely across related species (Yang and Shi; Levin and Malik; Dietschi et al.; Hughes and Piontkivska; Flajnik and Kasahara).

How Selection Acts on Duplicated Gene Families

Susumu Ohno is credited with first describing the idea of highly similar genes resulting from duplication of existing genes. These genes are subsequently subject to neutral drift and subsequent selection based on the qualities of the stochastic mutations that arise therefrom (Ohno). The immune system has a plethora of gene families thought to have arisen through some version of this process. Some well-studied examples are the TLRs, the Nod-like receptors, the CEACAM proteins, FPRs, MHC, CD1 receptors (Adrian et al.; Rosenstiel et al.; Muto et al.; Sutton et al.; Zajonc). Many of the receptors of the immune system evolved as paralogs after sequential rounds of duplication, drift, and neofunctionalization events.

It is thought that because the first effect of gene duplication results in more copies of the gene, and as a result, often more of that gene product, the first effect of gene duplication is simply higher copy number of the original gene (Dittmar and Liberles). In fact this phase of the evolutionary trajectory of the duplicated gene is thought to be accelerated, and strangely, is thought to affect only one copy, which is determined by chance (Pegueroles et al.). The subsequent fates of a gene after duplication may be as follows: if there is some fitness advantage to having more of the molecule encoded by the gene, then evolutionary pressure may exist to maintain the original as well as the duplicated gene without altering the sequence substantially. In this case, the genes will be subject primarily to purifying selection and neutral drift. In the case of genes where increased dosage affects fitness negatively or is lethal, the gene is falls into a fitness valley and one of the genes will likely be silenced as a result of purifying natural selection, with the surviving gene finding itself in a new fitness landscape after loss of the other allele (Teufel et al.). If however, the duplicated gene has an effectively neutral effect on fitness due to dosage imbalance, the gene may first undergo drift, followed by natural selection acting to either increase diversity through

neofunctionalization when the mutations result in benefit (positive selection). If mutations resulting from drift have a negative effect on fitness the gene may be pseudogenized and silenced through purifying natural selection (Teufel et al.; Reams and Roth).

Bridge.to Chapter III

Of the gene families in the immune system which owe their origin to such an evolutionary history, the two families focused on in this thesis are the formyl peptide receptors (FPRs) and the CD1 family of receptors which will be discussed in detail presently. Firstly, a study of the how sequence diversity affects functional interactions with S. aureus inhibitor and activator across several related formyl peptide receptor proteins shows us how sequence divergence is likely not the main driver of recognition.

CHAPTER III: DETERMINING FUNCTIONAL DIVERSITY FROM SEQUENCE DIVERSITY

Nicole Paterson and Matthew Barber conceptualized the study and the experimental design. Joseph Rangona and Brian Geisbrecht contributed reagents. Juan Tirado performed the PAML analysis for FPR2 in carnivores. Nicole Paterson generated the figures and performed all other experiments and analysis. Hussein Al-Zubieri assisted with cloning and experimental design. Nicole Paterson and Matt Barber wrote and edited the manuscript.

(Formatted for submission to Journal of Molecular Biology and Evolution, under review).

Dynamic evolution of bacterial ligand recognition by formyl peptide receptors

Nicole M. Paterson[1], Hussein Al-Zubieri[1], Joseph Ragona[2], Juan Tirado[1], Brian V. Geisbrecht[2], Matthew F. Barber[1]

[1]University of Oregon Institute of Ecology and Evolution, [2]Kansas State University Department of Biochemistry and Molecular Biophysics

Intoduction

The formyl peptide receptors (FPRs) are a family of G-protein coupled receptors (GPCRs) that play crucial roles in the recruitment and activation of leukocytes. Early studies demonstrated that human cell lines migrate towards N-formylated peptides, which are present in bacterial and mitochondrial, but not eukaryotic, proteins (Schiffmann et al.). These findings led to the discovery of FPRs as a new class of pattern recognition receptor with the ability to discriminate between 'self' and 'non-self' molecules in order to activate downstream immune responses (Pike et al.; Zigmond). Since then, additional microbial and host-derived ligands have

been identified for specific FPR homologs. Of the three FPRs in humans, each has been shown to possess a unique ligand-binding profile (Karlsson et al.; Schepetkin et al.; Kretschmer et al.), with a range of outputs in response to ligand. For example, recognition of lipoxin-A by FPR2 leads to the suppression of inflammatory signaling, whereas binding of bacterial-specific formylated peptides by FPR1 results in induction of the inflammatory response and cell chemotaxis towards ligand source (John et al.; Le et al.; Schepetkin et al.).

Neutrophils and other myeloid cells play a central role in innate pathogen recognition and express high levels of FPRs in humans. When neutrophils detect foreign molecules via FPR activation, the cell migrates towards the source of the signal. Upon reaching an infection site, neutrophils contribute to pathogen clearance through phagocytosis, release of toxic granule molecules, and a rapid oxidative burst which produces high levels of antimicrobial reactive oxygen species and reactive nitrogen species (Shimizu et al.; Ö̈nnheim et al.; Bufe et al.). Neutrophils constitute roughly 50% of circulating leukocytes and are capable of detecting nanomolar concentrations of pathogen-derived peptides through activation of FPR1 and FPR2 (Fu et al.; Le et al.). Natural killer cells, monocytes, and macrophages also express high levels of formyl peptide receptors which similarly contribute to cell activation and chemotaxis (Leslie et al.; Kim et al.; Crouser et al.).

Previous studies have detected signals of positive selection in FPRs 1 and 2 in the mammalian lineage through observation of high dN/dS ratios at several codon sites (Muto et al.). This is consistent with a model of formyl peptide receptors undergoing positive selection at sites important for immune function, most likely pathogen detection. All three FPR genes are located on chromosome 19 in humans (Bao et al.) and orthologous FPR gene regions in primates have previously demonstrated evidence of accelerated evolution in promoter and other regions (Yang et al.) as well as heightened occurrences of

single nucleotide polymorphisms (Harris et al.). Many other immune system related proteins including Toll-like receptors (TLRs), TRIM5, major histocompatibility complex (MHC) family genes, and transferrin, have been subject to repeated positive selection during mammalian evolution (Brunette et al.; McLaughlin and Malik; Daugherty and Malik; Barber and Elde; Aleru and Barber; Paterson et al.; Sawyer et al.; van der Lee et al.; Hughes and Piontkivska; Hughes and Nei). Taken together, the central role of FPRs in innate immune function coupled with previously identified signatures of positive selection suggest there may be important functional consequences for sequence-level variation observed in this family of receptors.

Results

Gene Loss and Rapid Evolution of FPRs in Primates and Carnivores

We found substantial evidence that FPR1 in several primate species is contracting (Yang and Shi) as demonstrated by a lack of FPR1 detectable expression in New World monkey whole blood, brain, lung and other RNA-Seq data from AceView (Thierry-Mieg and Thierry- Mieg) (Figure 1A). Previous reports looking at FPR expression in the primate vomeronasal organ identified FPR1 pseudogene in marmosets (Yang and Shi). We identified several additional pseudogenes in other New World Monkey lineages, after observing absence of gene expression as a lack of annotated or homologous FPR1 New World monkey genes in the NCBI database. The absence of FPR1 expression suggested to us that there may be additional pseudogenes in other New World Monkeys. We scanned available New World monkey genomes using BLAT search, and found additional regions of homology to FPR1 genes that when tested for the presence of exons using MIT's GENESCAN tool (Burge) failed to identify exons at the given probability cut-off. We aligned pseudogenes identified to human FPR1 and marmoset FPR1 pseudogene, which are available as high coverage, well-annotated genes in the NCBI database. In *Sapajus apella, Cebus imitator*,

Saimiri boliviensis and *Aotus nancymaae* at the genomic loci with substantial homology to FPR1 (located to the plus-strand adjacent to the FPR2 and FPR3 genes on chromosome 19) these gene regions contain homology but lack one or more features of a functional gene (Figure 2A, Figure 2B). *Saimiri boliviensis* pseudogene is the most striking, as this region has the least homology (77.5% identity in 271bp region located on the plus-strand adjacent to the FPR2 and FPR3 genes on chromosome 19) that lacks evidence of a start or stop codon and lacks significant homology to marmoset FPR1 outside the 271bp region. A caveat to this analysis is that each of these genomes has different levels of coverage and/or one or more builds (details on the quality of genomes used in this study can be found in the Supplemental Figure 2). However, *Saimiri boliviensis, Aotus nancymaae* and *Callithrix jacchus* have multiple builds with high coverage (*Saimiri boliviensis*: 2 builds most recent 111x coverage, *Aotus nancymaae*: 4 builds, most recent 132x, *Callithrix jacchus*: 11 builds, most recent 40x coverage) and good agreement with regards to pseudogenes at these regions across builds (full details in Supplemental Figure 2).

The gene expression and genome scanning analysis suggest that FPR1 function was lost early in the New World monkey lineage. This suggests there may be some diversification along the branch leading to FPR2 in New World monkeys. Because New World monkeys also appear to down-regulate expression of FPR3 (although the gene is still intact in the genomes of all New World primates scanned), there appears to be some downregulation of gene expression in FPR3 which is notable. Like other immune receptor families including TLRs (Levin and Malik), gene duplication and loss have occurred periodically throughout the evolution of the FPR gene family in mammals as evidenced by the variable copy number across different mammalian lineages (Figure 1B), and a loss of FPR gene expression is not an unusual event in mammals.

The differential gene expression profile in New World monkeys that shows increased expression of FPR2 and decreased expression of FPR1 and FPR3 is supported by our analysis of positive selection at the codon level in primate genes. Our analysis shows good agreement with previous studied in mammals showing heightened rates of positive selection in FPR1 and FPR2 (Figure 3). We ran additional analysis testing for branch-site episodic positive selection using DataMonkey's abSREL test (Smith et al.) and found evidence of heightened selection on the branch of the phylogenetic tree leading to the New World monkeys FPR2. It is possible that FPR2 has a divergent function from FPR2 in other primate lineages.

Another lineage known to rely on FPR2 function solely, is the Carnivora clade, which encode only the single FPR2 paralog. Phylogenetic analysis of positive selection at the codon level in carnivores revealed similar patterns of positive selection in FPR2, though with one fewer site overall. The sites that appeared in carnivores mapped to the N terminus and the third and fourth extracellular domain, the exact regions that appear to be undergoing selection in primates as well. Curious to find out whether this lineage could give us insight into the evolutionary processes at play in FPR2 in mammals more broadly, we undertook a functional analysis using a microbial system likely to have interacted with most, if not all of the species we analyzed.

Aceview NIHTPR RNA seq data plotted to display expression levels across primate kingdoms for hominoid (human, chimpanzee), Old World monkey (pig-tailed macaque, crab- eating macaque, baboon, mangabey) and New World monkeys (marmoset, squirrel monkey and owl monkey). High expression of FPR1 in hominoids of FPR1 in Whole Blood, Brain and Lung. Moderate-high expression in Old World monkeys and notable absence of FPR1 and FPR3 expression in New World monkeys. FPR2 expression is less extreme, but variable across kingdoms. A) Whole blood B) Brain C) Lung.

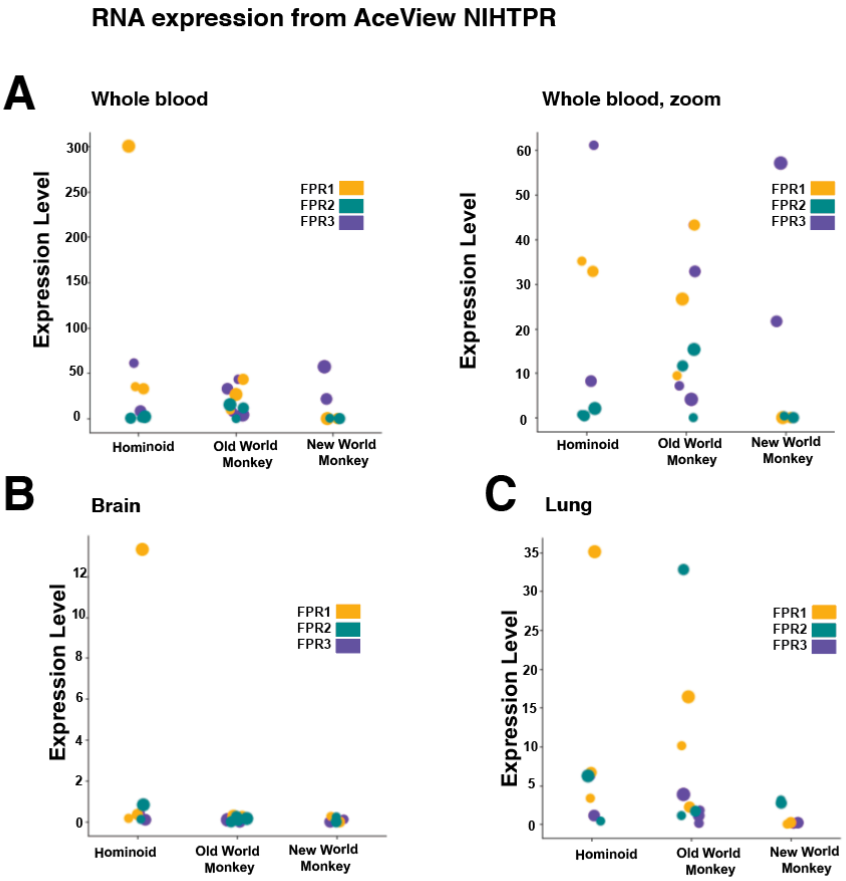


Figure 3. FPR1 gene expression among simian primates

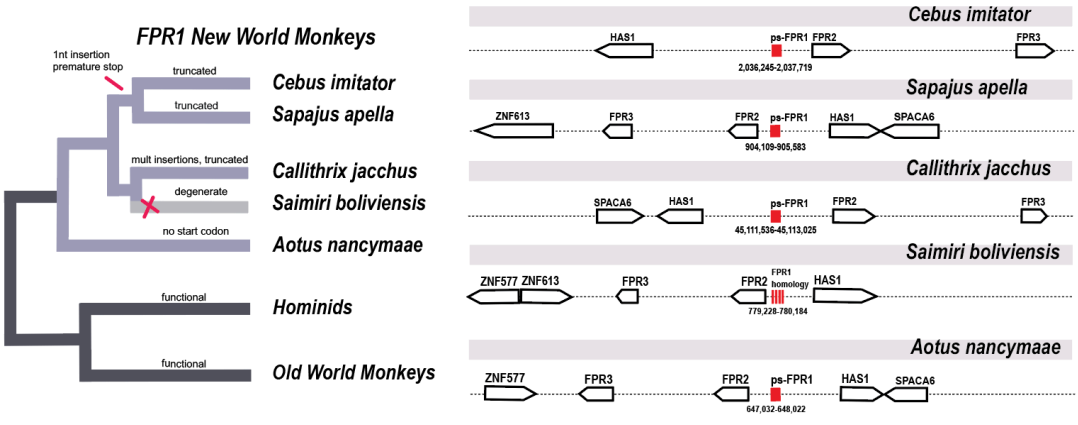


Figure 4. FPR1 pseudogenization in New World monkeys

Genome scanning for regions of homology using marmoset FPR1 as the query sequence showed evidence for loss of FPR1 functional sequence in the New World Monkey lineage. A) *Saimiri boliviensis* displayed the most significant gene loss at this region, with a short 271bp region with significant homology and no apparent start or stop codons. B) The loci where apparent pseudogenes were identified is relatively conserved across the New World monkey lineages

Recognition of bacterial ligands by mammalian FPRs

To assess the functional consequences of sequence variation in primate FPRs, we focused on Staphylococcus aureus interactions due to expression of both pathogenesis-associated inhibitors and activators shown to interact with FPRs (Annette M. Stemerding et al.; Sundqvist et al.; Prat et al.; Koymans et al.; Kretschmer et al.; Li et al.). *S. aureus* is a gram-positive bacterium is known for its multiplicity of virulence factors and rapid acquisition of antibiotic resistance across a broad mammalian host range including livestock, rodents and companion animals (Thammavongsa et al.; Koymans et al.; Haag et al.). This adaptable microbe colonizes the nares of roughly 30% of humans, but is also a major cause of skin and soft tissue infections, bacterial sepsis, pneumonia, and other life- threatening infections (Thammavongsa et al.; Haag et al.). S. aureus produces many extremely effective toxins that contribute to its virulence, including enteroxins, leukocidins, and alpha-hemolysin to name a few (Balasubramanian et al.; Lowy; Priatkin and Kuz'menko). S. aureus also produces a range of proteins that target and silence immune receptors such as TLRs, FPRs and complement receptors which are responsible for detecting and/or mounting an immune response against infection (Thammavongsa et al.; Prat et al.; Kretschmer et al.; Koymans et al.; Wright et al.) While *S. aureus* is generally believed to be human-adapted, primates, rodents and livestock are frequently

colonized by divergent strains of *S. aureus* and related staphylococci (Schaumburg et al.). Strains sampled from wild gorilla, chimpanzee, green monkey, and colobus contain the gene for the virulence factor staphylococcal enterotoxin B (SEB) as well as other virulence factors (Schaumburg et al.).

SEB has been shown to potently activate FPRs (Dürr et al.) Kretschmer et al.; Fu et al., Dürr et al.). In addition, S. aureus produces formylated peptides that can induce cell migration in neutrophils (Dürr et al.). These bacteria also produce a number of inhibitory molecules that target FPRs, such as FLIPR and FLIPR-like which have demonstrated specificity for FPR2, although FLIPR-like can bind FPR1 with lower affinity (Annette M. Stemerding et al.).

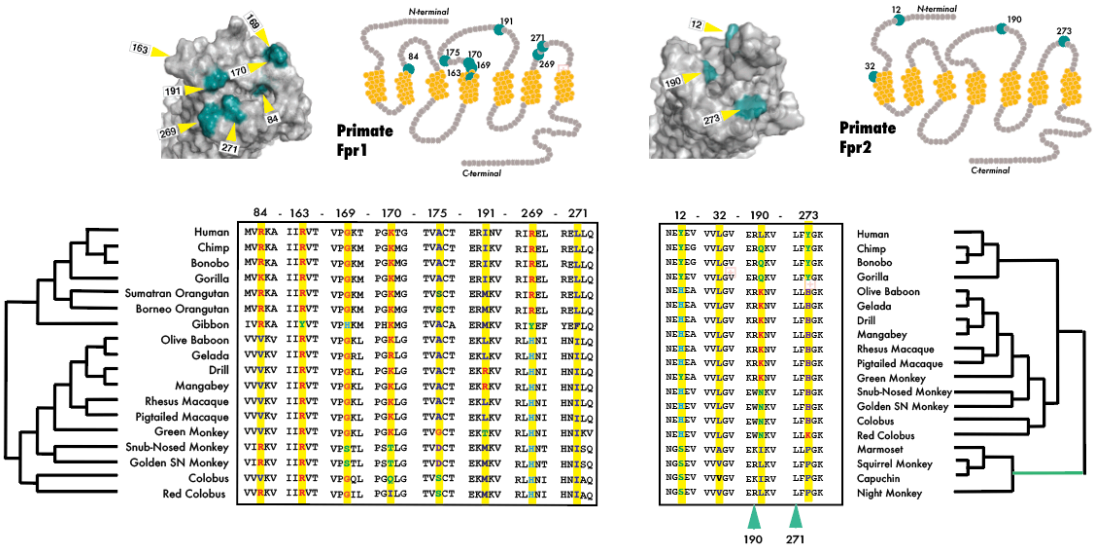


Figure 5 A,B. Evidence of repeated positive selection among primate FPRs A) Sites in primate FPR1 with elevated dN/dS as determined by PAML and HyPhy. Residues of the transmembrane domain are denoted in yellow on the protein diagram, with the majority of high dN/dS sites located in the extracellular ligand-binding loops. B) Sites in primate FPR2 with elevated dN/dS as determined by PAML and HyPhy. Residues of the transmembrane domain are denoted in yellow on the protein diagram, with all high dN/dS sites located in the extracellular ligand-binding loops. Branch tests indicate elevated dN/dS in the common ancestor of New World monkeys, which have also lost FPR1.

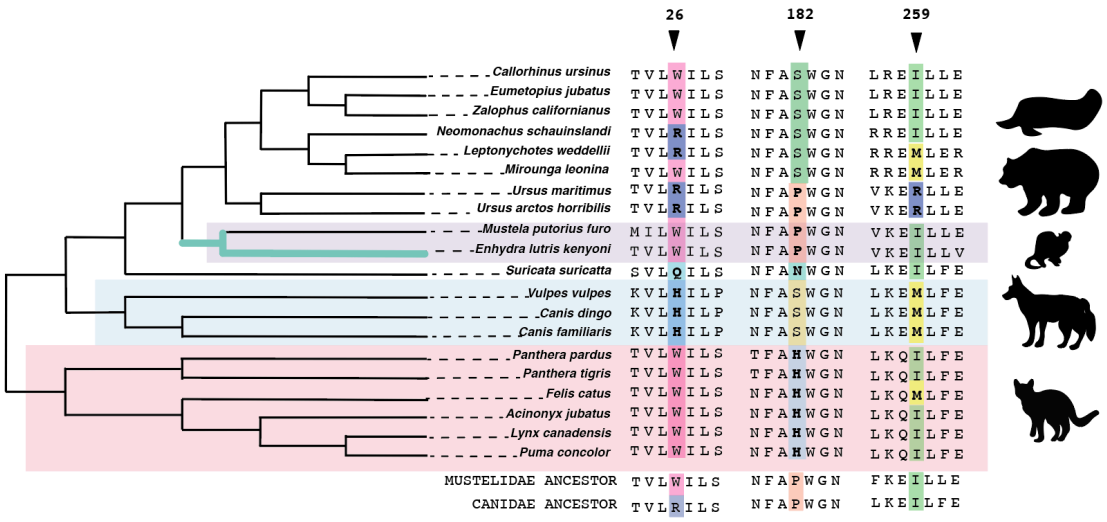


Figure 6. Evidence of repeated positive selection in carnivore FPR2

Sites in carnivore FPR2 with elevated dN/dS as determined by PAML and HyPhy.

Using HEK293T cell lines stably expressing FPR orthologs at approximately equivalent levels, we tested binding of FITC-labeled SEB and FLIPR-like proteins by flow cytometry. Most of the receptors we tested displayed some level of SEB binding, but surprisingly the level seen in bonobo FPR1 matched the binding profile of human FPR2 more than human FPR1. This was also true for FITC-labeled FliprLike protein. (Figure 4). This was unexpected given the human FPR1 receptor has very high sequence conservation with bonobo FPR1, diverging only at six amino acid positions overall, and far less sequence homology with human FPR2. Since bonobo FPR1 displayed activity more similar to human FPR2 in response to FLIPR-like binding as well, this suggests that there may be some site- specific properties that confer function between bonobo FPR1 and human FPR1 that are similar between bonobo and human FPR2.

We next considered which amino acids were the most likely to be responsible for these binding differences. Without previously published crystal structures for FPRs, we generated homology models using I-TASSER, an online

webserver that uses a threading algorithm to match primary sequence data to crystal structures in the server's database (there are many related GPCRs with sufficient homology to act as reference crystal structures). These structures were docked to the entire extracellular region using the seven amino acids that form the region of Flipr and FliprLike N-terminal peptide required for its inhibitory activity (FFSYEWK)(Annette M. Stemerding et al.) using the Schrodinger Glide molecular modeling system. The results were surprising.

Several amino acids in the first and last extracellular loops of FPR1 formed hydrogen bonds with the FliprLike peptide. One of these, site 190, is predicted to be under selection in primates (Figure 8C). Adjacent to this site in human FPR2, site 189 also forms a hydrogen bond to FFSYEWK in docking studies. Interestingly, there were several sites of overlap between the two including positions S6, N10, and R190 (Figure 8C). Several studies have shown site 190 can modify peptide binding to other microbial ligands, notably studies in the S. aureus chemotaxis inhibitor CHIPs, which shows a preference for lysine over tryptophan at position 190 (Mills et al). This same site has been shown to modify susceptibility of immune cells to recognition by the Yersinia pestis type 3 secretion system, where tryptophan at position 190 has been shown to drastically reduce cell entry versus wildtype K190 (Osei-Owusu et al.).

However, in the bonobo FPR1 docking study, the region near sites 189-190 did not interact with the FFSYEWK peptide (Figure 8B). Interestingly, it appears that the amino acid at position 170, which differs between human FPR1 and bonobo FPR1 possesses a methionine instead of a threonine at that position, which creates a ridge where the FFSYEWK peptide docks in the bonobo model. To test how this may affect the inhibitory activity of FliprLike, we performed an additional docking to f-MLF, the canonical activator of FPRs. From comparison of the docking poses of human FPR1 and bonobo

FPR1 we saw similar binding of the Fliprlike peptide across several of the extracellular loops of human FPR2 and bonobo FPR1, with the fMLF peptide buried in the central region of the helices (Figure 8A,B).

Additionally we observed several naturally occurring single nucleotide polymorphisms in the human population R190W, predicted to be involved in binding Nterm FliprLike and N192K which differs between human FPR1 and bonobo FPR1. Additionally, the site which forms the ridge in bonobo FPR1 T170M naturally occurs in the human population, as well as an additional T170P which occurs at that site as well, although at low frequency (Figure 9).

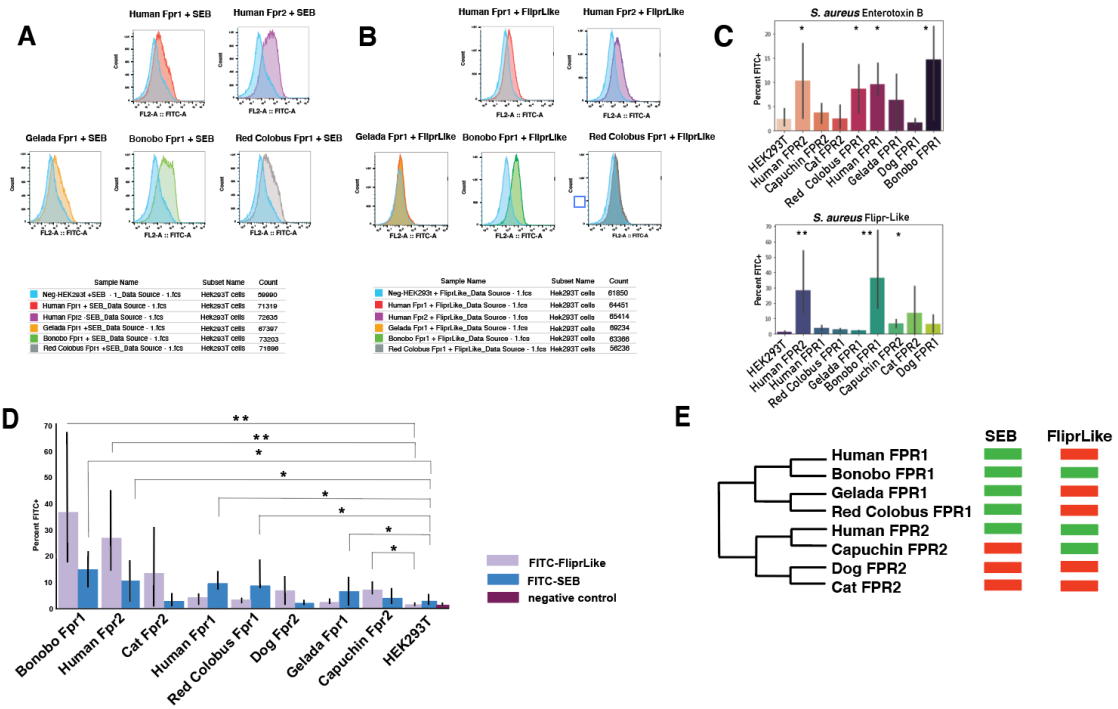


Figure 7. Recognition of S. aureus ligands by mammalian FPRs. A,B) Data from at least three flow cytometry experiments for each species. Singlet cells were gated and percent of FL2 (488+) positive cells reported from the parent. C) Percent FITC+ cells plotted from a subset of singlets, HEK293T cells incubated with FITC-labeled protein used as control to set negative levels.

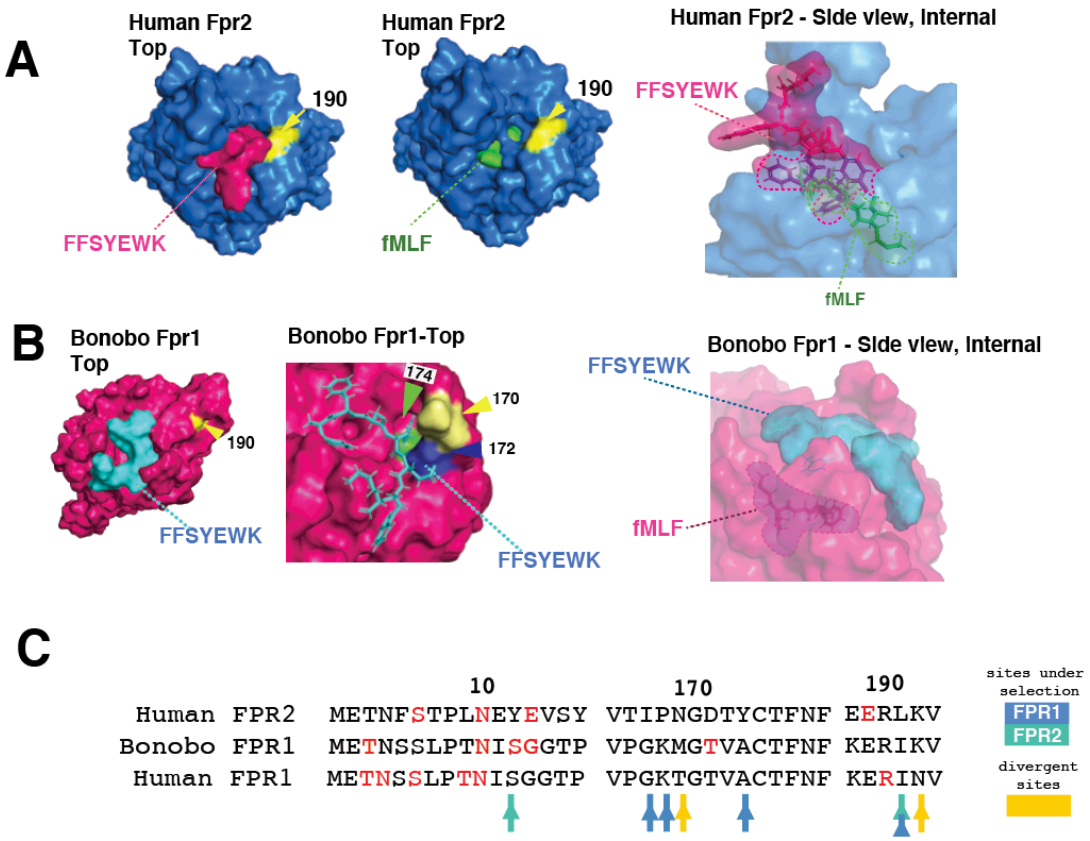


Figure 8. Sites that differ between human and bonobo FPR1 are adjacent to sites under selection in primates. A) Schrödinger Glide ligand docking prediction to the Nterm region of FliprLike, FFSYEWK-shown to be required for inhibitory function, docks in the region above the canonical FPR ligand fMLF. Site 190 labeled for orientation purposes. Side view shows the FFSYEWK docking pose internal to the alpha-helices of FPR2, aligned to the fMLF docking pose. B) FFSYEWK shows different hydrogen bonding patterns to bonobo FPR1 and docks in a slightly different region of the receptor, due to a protruding ridge formed by M170 in the homology model of bonobo FPR1. Side view of bonobo FPR1 shows similar arrangement between FFSYEWK and fMLF docks, despite different docking pose. C) Sites that differ between Bonobo FPR1 and human FPR1 are also adjacent to sites predicted to be undergoing positive selection. Sites labeled with red text form hydrogen bonds with FFSYEWK. Arrows in blue indicate sites with high dN/dS for FPR1, green for FPR2. Sites that diverge between Bonobo FPR1 and Human FPR1 in yellow. Site 192 and site 195 differ between human FPR1 and bonobo FPR1, but are the same for bonobo FPR1 and human FPR2. Sites that differ between human FPR1 and bonobo FPR1 shown at right.

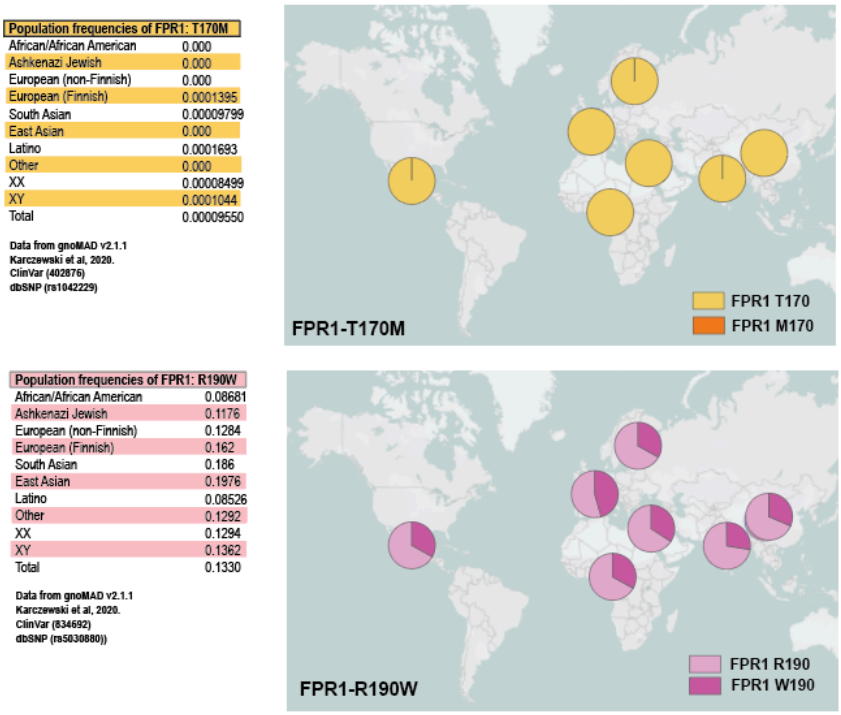


Figure 9. Single nucleotide polymorphisms in human FPR1 at sites predicted to affect binding affinity to FliprLike are relatively infrequent for the mutation found in bonobo FPR1 that forms a new hydrogen bonding interaction in docking studies. Mutations at site 190 that may interfere with binding to FliprLike are relatively frequent.

Discussion

Studies in rodents have shown remarkable functional plasticity in FPRs (Dietschi et al.). This plasticity likely explains in part why these genes have expanded independently in so many lineages. There appears to be a broad capacity for FPRs to act as environmental sensors, as evinced by aforementioned studies in rodents. The sensitivity of detection possible for FPR1 which can detect formylated peptide at remarkable low nanomolar concentrations (Mills et al) may explain why FPR1 is so highly expressed in

humans even though it is associated with inflammatory disease and cancer (T et al.; Khau et al.; Vacchelli et al.). Since much is known about why FPR genes may expand in lineages, it may be worthwhile in this stage of FPR research to seek answers for why genes in this family may contract, such as observed in this study. There have been studies in humans associating FPRs with exacerbating human disease states such as glioblastoma (Yao et al.; Cussell et al.) and breast cancer (Khau et al.), especially for FPR1 which is strongly associated with inflammation. Additionally, FPR1 has been shown to be required for *Yersinia pestis* recognition of host cells, a function which can be lost when site 190 (two amino acids upstream from site 192 shown to be relevant in this study) is modified from an arginine to a tryptophan (Osei-Owusu et al.). Loss of this gene may confer benefit in animals prone to plague, and such a scenario offers a ready explanation for why loss of a functionally important gene may be worth the sacrifice to the long-term survival of a species.

In our system, we see many components of a classical molecular arms race dynamic. Evolutionary conflict is defined by multiple rounds of evasion and attack on resources by a pathogen on its host species, and we see elements of these dynamics in this system where we probed the binding profiles of an inhibitor and an activator across several species with evidence of beneficial adaptation at specific codon sites. We discovered a region in the FPRs at sites 190/192, which may be responsible for the differences in binding that we see in this study. *S. aureus* FPR inhibitors CHIPs and Flipr/FliprLike preferentially bind to FPR1- R190, given that CHIPs has been shown to be specific for FPR1, while Flipr/FliprLike bind FPR2 with high affinity and FPR1 with low affinity. Given this knowledge it may be significant that human FPR2 encodes a charged amino acid glutamate at position 190, possibly reducing its affinity for CHIPs and increasing affinity for Fliprs. It is also possible that this region is targeted by S. aureus inhibitors due to its important role in ligand binding.

Putting together data gathered on primate FPRs in response to *Staph* activator SEB and inhibitor, FliprLike, it is noteworthy that bonobo FPR1 bound the S. aureus inhibitor FliprLike more similarly to human FPR2 than human FPR1, even though it shares far more sequence similarity with human FPR1. It is quite possible that positive selection in carnivore FPR2 has resulted in divergent functional response to ligand.

Notably, this is not the first time that this region has been implicated in interactions with the host-microbe interface with *S. aureus*. A study testing the function of various high- prevalence SNPs in human FPRs showed 10 fold difference in binding for R190/K192, R190/N192 chemotaxis inhibitor of FPR1 (CHIPs) not seen for FPR1 W190/N192 (Mills). The relatively low frequency of the M170 mutation in FPR1 may be adaptive, if that site does indeed increase the binding affinity of FliprLike for FPR1. Given the relatively few sites that differ between human and bonobo, and the difference in size between methionine and threonine, this substation is likely to have an effect, supporting this possibility. The implications of this research are dual-pronged: such studies increase our knowledge of how a genotype may affect susceptibility to disease, and suggest that inhibitors of inhibitors may be effective treatments for bacterial infection in those individuals with heightened susceptibility.

Methods

Identifying Sites of positive selection in FPRs

We inferred amino acid sites exhibiting elevated dN/dS using multiple computational methods. Our dataset included all available nucleotide coding sequences (cDNA) for 18 primate species of FPR1 (human, drill, mangabey, red colobus, snub-nosed monkey, golden snub-nosed monkey, Sumatran orangutan, Bornean orangutan, gorilla, chimpanzee, bonobo, white-cheeked gibbon, green monkey, crab-eating macaque, pig-tailed macaque, gelada, olive baboon), with areas of ambiguity and stop codons removed. A gene tree for

FPR paralogs was generated with PhyML (phylogenetics by maximum likelihood) with 1000 bootstraps (Supplemental Figure 1)[13]. Potential sites under positive selection were detected using the phylogenetic analysis by maximum likelihood (PAML) package (Yang) which detects signs of positive selection from the frequency of nonsynonymous/synonymous amino acid substitutions at each site (ω = dN/dS) based on maximum likelihood. Additional computational methods MEME (Pond and Frost) and FuBar (Murrell et al.)from the DataMonkey adaptive evolution server were cross-referenced and sites that appeared in more than one analysis with high confidence ($p< 0.01$) were included. absREL analysis which tests for branch site episodic selection was also performed (Smith et al.).

Cloning and Lentiviral transduction of FPRs in HEK293T cells

FPR1 genes for human, bonobo, gelada, and red colobus and FPR2 genes for human, capuchin, dog and cat were cloned from cDNA (Human FPR1 and FPR2) synthesized by Genewiz (gelada and red colobus) or synthesized as gBlocks by IDT (capuchin, dog, cat) including Kozak sequence and C-terminal Flag-tag and subsequently cloned into pBABE lentiviral vector using SLIC or Gibson cloning methods. Full length FPR1 orthologs were cloned into pBABE vector using the Gibson method for a FLAG epitope tag (Gibson et al., 2009). After expression was verified in cell lines by Western blot using anti Flag tag antibodies (Monoclonal ANTI-FLAG® M1, Sigma Aldrich #F3040), surface expression was verified for FPR1 using Thermo Fisher FPR1 polyclonal antibody ref number PA5-33534) and cell lines with similar percent expression on the cell surface were used in binding experiments. FITC-labeling was performed per manufacturer's instructions, and the Thermo ScientificTM PierceTM Dye Removal Columns, 22858 were used to remove excess dye per manufacturer's instructions. A titration was performed, and 4µg of FITC-labeled FliprLike or SEB proteins were incubated in 100µl sterile phosphate buffered saline + 100nM PMSF at 4°C with nutation for 1

hour, washed 3 with 1 ml ice-cold PBS and analyzed on a SONY SH800 flow cytometer.

FliprLike Peptide Docking to human FPRs

All structures used for analysis were generated using I-TASSER homology modeling server (Yang and Zhang), except for formyl-MLF and FFSYEWK, which represents the first six amino acids of Flipr and Flipr-Like proteins (Annette M. Stemerding et al.) generated in PyMol (Schrodinger). Peptide docking was run with Schrödinger Glide, SP and XP dock with sidechain protonation set to represent charged states at pH 7, highest ranking docking was used for analysis.

Genome scanning for FPR1 pseudogenes in New World Monkeys

Complete genomes for available New World Monkey species were queried for FPR1 pseudogenes by BLAST search, BLAT search of genome in UCSC Genome Browser (*Callithrix jacchus, Saimiri boliviensis*) while pseudogenes for *Cebus capucinus imitator, Sapajus apella*, and *Aotus nancymaae* were identified by BLAT search for homology in genome viewer in NCBI's Genome Data viewer. Details for genome assemblies in Supplemental Figure 2. Pseudogenes were aligned using MUSCLE. Resulting data was analyzed by alignment and exons were searched for and in some cases eliminated using GENESCAN (Burge and Karlin).

Yersinia pestis interactions with formyl peptide receptors

Yersinia pestis, etiological agent of the Black Plague evokes dread like no other. Yet this deadly pathogen can be rendered relatively benign if the receptor required for immune cell entry, formyl peptide receptor 1 (FPR1), is mutated at a single amino acid position, a site implicated in binding to formyl peptides at low concentration and in binding to S. aureus inhibitor (Osei-Owusu et al. 2019, 1) It is quite possible that binding of this site directly by

Y. pestis LcrV and S. aureus FliprLike also serves to dampen the detection of their own bacterial peptides.

Yersinia pestis is essentially a hyper-virulent derived strain of Yersinia pseudotuberculosis, diverged from the latter from approximately 10,000 years ago(Anisimov, Lindler, and Pier 2004). *Y. pestis* strains encode virulence factors on 32 Y. pestis-specific chromosomal genes, a unique pCD plasmid, and higher copy expression of Yersinia spp. virulence-associated plasmids(Perry and Fetherston 1997). The general trend in Y. pestis appears to be loss of functional metabolic genes in tandem with gain of virulence factors, as demonstrated by its incomplete pentose phosphate pathway and loss of other metabolism-associated genes(Hinnebusch, Jarrett, and Bland 2017). The three biovars of plague have different metabolic capacities. The causative agent of the ancestral Justinian plague, biovar *Antiqua*, has been shown to reduce nitrate as well as metabolize glycerol, followed by *Medievalis* which cannot reduce nitrate, and *Orientalis* which cannot metabolize glycerol. In this manner, *Yersinia* species have evolved several differing sequences for LcrV that have been reported and studied in the literature (Anisimov, Lindler, and Pier 2004). It is not known whether these changes are due to genetic drift or molecular arms races between the interfaces of host receptors and pathogen proteins.

We found that Order *Carnivora* which possesses only a single copy of FPR2 also has heightened signatures of positive selection. Interestingly, the branch in Carnivora predicted to be undergoing repeated bouts of episodic selection is also the branch which displays variable susceptibility to plague. This is particularly interesting, because as already mentioned, FPR1 can act as a plague receptor on immune cells in humans. This leads to the obvious question: in a mammal that is susceptible to plague infection but does not possess an FPR1, what then is the plague receptor?

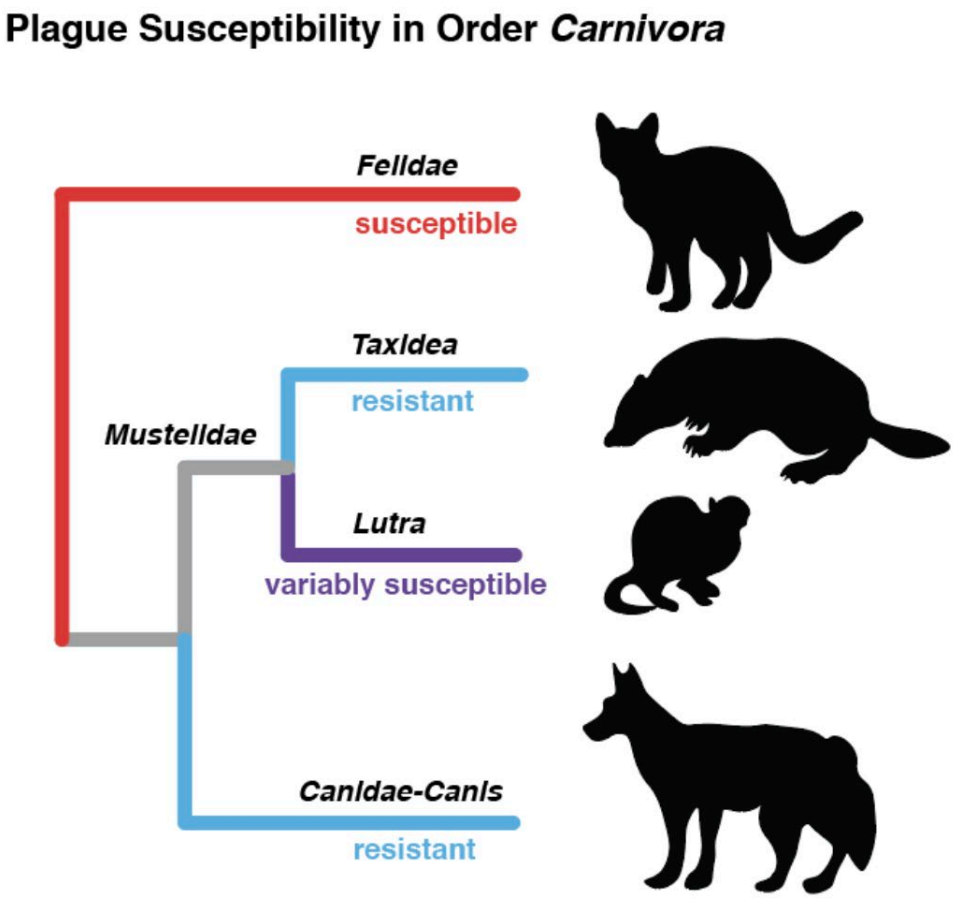


Figure 10. Susceptibility to plague varies across carnivores, who lack FPR1, required for entry in human immune cells

Some mammals have done away with FPR1 function entirely, and no longer have a working copy of FPR1 (such as Canids), here we show FPR2 may/may not be able to act as a plague receptor in species that have lost functional FPR1. Many immune receptor gene families such as MHC, CD1, TLR to name a few, expanded in number (and in many cases evolved specific functions) via gene duplication events. For the primate FPR family of genes, we show that encoding proteins with similar properties may be a drawback since pathogens can exploit similarities to gain access to the host by alternative means.

Some mammals have done away with FPR1 function entirely, and no longer have a working copy of FPR1 (such as Canids), here we show FPR2 may/may not be able to act as a plague receptor in species that have lost functional FPR1. Many immune receptor gene families such as MHC, CD1, TLR to name a few, expanded in number (and in many cases evolved specific functions) via gene duplication events. For the primate FPR family of genes, we show that encoding proteins with similar properties may be a drawback since pathogens can exploit similarities to gain access to the host by alternative means.

Some mammals have done away with FPR1 function entirely, and no longer have a working copy of FPR1 (such as Canids), here we show FPR2 may/may not be able to act as a plague receptor in species that have lost functional FPR1. Many immune receptor gene families such as MHC, CD1, TLR to name a few, expanded in number (and in many cases evolved specific functions) via gene duplication events. For the primate FPR family of genes, we show that encoding proteins with similar properties may be a drawback since pathogens can exploit similarities to gain access to the host by alternative means.

It is unclear in mammals with only one FPR such as in most carnivores which only possess a single copy of FPR2, whether FPR2 can act as a plague receptor. Presumably the FPR2 receptor was immune to plague hijacking in humans prior to its emergence, since Denisovan DNA suggests the sequence of FPR2 has not diverged at sites under selection compared to extant FPR2. This is consistent with our results, which show lack of binding to LcrV protein (Figure 5), which is required for cell injection of pathogen virulence factors.

However, felines remain susceptible to plague, even though they lack the FPR1 receptor (Bevins et al.; Castle et al.). The canine lineage, however, which also lacks and FPR1 receptor and only possesses an FPR2 receptor is

relatively insensitive to plague, with most canines experiencing only minor illness, if any (Vernati et al.; Salkeld and Stapp). Interestingly, the arginine at site 190 is fairly well-conserved in carnivore FPR2, as well as in Human FPR2 and many primates except for the Snub-nosed monkeys and the Colobus and Ugandan Red Colobus. The carnivore lineage is predicted to have several sites under selection in FPR2, and it is possible that one or more of these sites is involved in plague resistance/susceptibility, if FPR2 is indeed a viable receptor for plague entry. C5AR, the complement receptor is another immune receptor with sequence similarity to FPR1 that could be an additional candidate for plague entry to the cell.

However, as always in evolution the story is complex. The W190 mutation in humans associated with resistance to plague is also associated with an increased affinity for SARS- coronavirus-related peptides, and a decreased affinity for formylated peptides (Mills). It is possible that the recent pandemic will have a selective effect for the W190 allele with the additional benefit of increased resistance to future plague outbreaks. Given the results of our analysis in S. aureus showing the importance of the region surrounding site 190 for ligand binding, further studies testing the trade-offs in plague susceptibility and ligand recognition for single nucleotide polymorphisms at this site would be valuable to understanding how hijacking this receptor might affect host:pathogen interactions broadly.

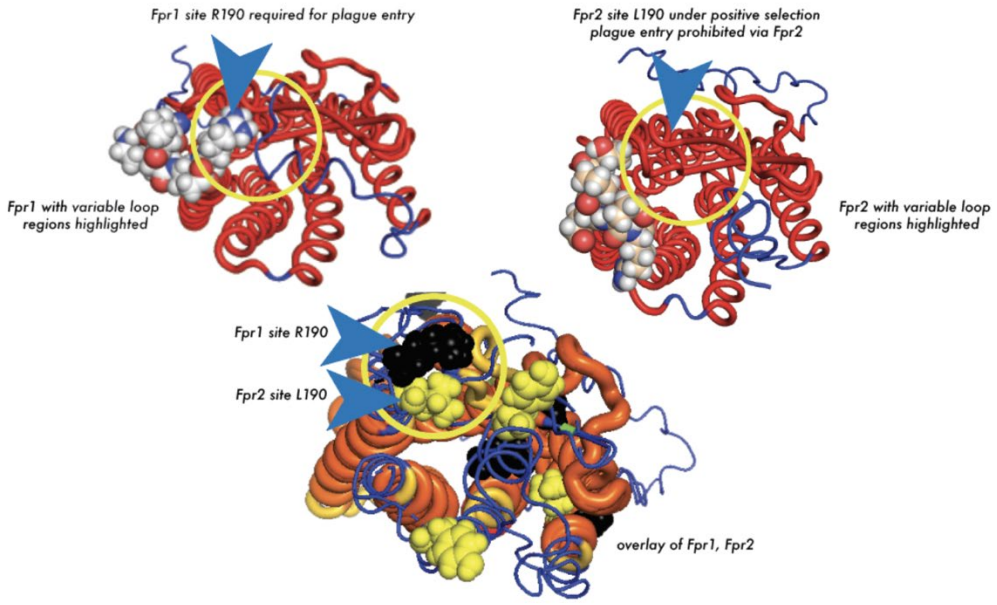


Figure 11. FPR1 site 190R is required for plague entry to the cell. FPR2 shows signatures of positive selection at this site, which appears to be an important site for S. aureus inhibitor binding

Other microbial interactions with formyl peptide receptors.

Formyl peptide receptors in both human and cat have been shown to detect viral signatures of clinically relevant human immunodeficiency virus (HIV) and feline immuno-difficiency virus (FIV) respectively (Ding et al.; Mills). This detection appears to have a preference for the R190 versus the W190 polymorphism in FPR1, suggesting that pressure to maintain the R190 allele to detect of the FIV peptide may be in direct opposition to fixing the W190 allele in the population, despite its plague resistance effects. It is possible that detecting these viral signatures confers fitness benefits resulting in selective pressure to maintain the receptor despite the vulnerability to plague infection, in addition to fitness effects conferred by innate immune detection of bacterial peptides.

Bridge to Chapter IV

Studies in formyl peptide receptors across multiple species suggest that there

are forces at play in the relationship of genotype-phenotype beyond simply sequence divergence affecting function. Three of the most divergent proteins from primate in our analysis (Bonobo FPR1, Human FPR2 and Capuchin FPR2) showed similar binding profiles to FliprLike and enterotoxin that were not shared by Old World Monkeys studied. For a receptor as promiscuous as the formyl peptide receptor, studies addressing dominant forces driving evolution will be challenging. However, we did identify a region of the protein that appears to be extremely important for tuning ligand affinity. Identification of these hotspots can direct practical applications such as drug design or horizon scanning for mutations in pathogen-associated proteins that may be associated with extreme consequences for the host. Chapter IV continues the pursuit of understanding how sequence diversity affects the function of immune receptors, using a very different system. The CD1 family studied in the next chapter is unique in that it binds and presents lipid and lipopeptide molecules to T- cells. Due to the different characteristics of the ligands it presents, CD1 molecules have slightly different biology than the related MHC proteins. However, studies of CD1 molecules can have an impact on microbial pathogens such as Mycoplasma species which produce a plethora of exotic lipids and are excellent at evading immune defenses.

CHAPTER IV. DIVERSIFICATION OF CD1 MOLECULES SHAPES LIPID ANTIGEN SELECTIVITY

(Published June 2021, reproduced here with permission from the publisher)
Nicole Paterson and Matthew Barber conceptualized the study and the experimental design. Nicole Paterson performed all experiments and data analysis. Hussein Al-Zubieri assisted with cloning and experimental design. Nicole Paterson and Matt Barber wrote and edited the manuscript.

Diversification of CD1 Molecules Shapes Lipid Antigen Selectivity

Nicole M. Paterson,[1,2] Hussein Al-Zubieri,[1,2] and Matthew F. Barber [1,3,*]

[1]Institute of Ecology and Evolution, University of Oregon, Eugene, OR 97403, USA [2]Department of Chemistry and Biochemistry, University of Oregon, Eugene, OR 97403, USA [3]Department of Biology, University of Oregon, Eugene, OR 97403, USA
*Corresponding author: E-mail: mfbarber@uoregon.edu
Associate editor : Kelley Harris

Introduction

Early detection of pathogen-specific molecules by the im- mune system can mean the difference between resistance, latency, or succumbing to infectious disease. Previous studies have illustrated that host–pathogen protein interaction sur- faces are hotspots for repeated natural selection by influenc- ing resistance or susceptibility to infection (Daugherty and Malik 2012; Enard et al. 2016; van der Lee et al. 2017). Such conflicts between hosts and pathogens can give rise to a variety of evolutionary dynamics including Red Queen arms races (Van Valen 1973; Daugherty and Malik 2012), frequency-dependent selection (Takahata and Nei 1990), and over-dominance (Hughes

and Nei 1990; Takahata and Nei 1990; Nei and Rooney 2005). Although vertebrate im- mune systems are tuned to recognize a wide variety of pathogen-associated macromolecules including DNA, RNA, lipids, and glycans, our understanding of host–pathogen evo- lutionary conflicts is largely restricted to protein–protein interactions (Sawyer et al.2005; Elde et al. 2009; Barber and Elde 2014; Choby et al. 2018). In the case of lipid and lip- opeptide antigens, the production of a functional molecule involves the synthesis of precursors that are further processed by enzymatic modifications. As such, evolutionary dynamics involving these macromolecules and their host receptors may be distinct from protein–protein interactions.

The major histocompatibility complex (MHC) superfamily comprises a variety of cell surface proteins which present self and foreign antigens to T-cells. Recognition of foreign anti- gens by T-cell receptors (TCRs) leads to T-cell activation and initiation of an adaptive immune response (Frank 2002). Multiple evolutionary forces are hypothesized to contribute to the immense diversity in MHC haplotypes, including over- dominance wherein increased heterozygosity is favored by selection and polymorphisms are maintained over time (Takahata and Nei 1990). In addition to class I and class II MHC molecules which present peptide antigens, the paralo- gous cluster of differentiation 1 (CD1) and MR1 molecules have been shown to present lipid and lipoprotein antigens to T-cells (Blumberg et al. 1995; Barral and Brenner 2007; Birkinshaw et al. 2015; Moriet al.2016; Zajonc and Flajnik 2016). CD1 molecules display rare and infrequent polymor- phism with limited genetic diversity within humans and other populations relative to class I and II MHC (Han et al. 1999; Golmogghaddam et al. 2013). The MHC and CD1 gene fam- ilies therefore appear to have experienced divergent evolutionary paths after their duplication from a common ancestor with respect to antigen recognition and population genetic variation. CD1 paralogs are divided into groups, whereby group 1 CD1 family members (including human CD1a, CD1b, and

CD1c) present antigen primarily to cyto- toxic CD8þ T-cells (Mori et al.2016). Group 2 CD1 molecules (including CD1d in humans) present antigens to invariant natural killer (iNK) T-cells (Pereira and Macedo 2016).

CD1 and MHC (fig. 1A) arose in jawed fishes with the advent of the B-cell and T-cell immune receptors during the genesis of the adaptive immune system of vertebrates (Barral and Brenner 2007). After the gene duplications that gave rise to ancestral MHC and CD1, vertebrate CD1 paralogs expanded through repeated duplication events (Dascher 2007). This initial expansion was followed by differential pseu- dogenization and expansion of CD1 paralogs to various degrees across vertebrate species (Rogers and Kaufman 2016). As such, CD1 gene content varies widely across verte- brates: primates have a single copy of the CD1A paralog, mice possess none, dogs encode six, and horses possess five (Dascher 2007). In primates, CD1D is believed to represent the most deeply conserved member of the CD1 family (Salomonsen et al.2005). CD1d receptors can display antigen to specialized iNKT-cells, which are able to mount an earlier response to infection reflecting their dual role in innate and adaptive immunity (Pereira and Macedo 2016). Current evi- dence indicates that human CD1e does not present antigen (Garcia-Alles et al. 2011) but rather assists in antigen loading onto CD1d in lysosomes and endosomes (Cala-De Paepe et al. 2012). This wide range of responsive T-cell types along with evidence that these nonclassical T-cell types mount an early immune response to infection (Godfrey et al. 2015) makes human CD1-expressing cells surprisingly flexible responders to infections even with their lack of exceptional sequence variation.

CD1 molecules possess an extracellular domain containing a subsurface hydrophobic-binding pocket used to present antigen to CD1-restricted T-cells (fig. 1B) (Barral and Brenner 2007). During the adaptive immune response, CD1 on the surface of antigen-presenting cells activates T-cells by displaying specific classes of hydrophobic ligands to TCRs (fig. 1C) (Blumberg et al.

1995; Barral and Brenner 2007; Chancellor et al.2018). According to structural studies, CD1a has the smallest of the human CD1-binding pockets with a volume of about 1,280 A$^{'}$ (Ly and Moody 2014). After the gene duplication event that gave rise to this paralog, CD1a likely evolved to present either self-lipids or small ex- ogenous lipopeptides (Mori et al.2016). Consistent with this hypothesis, CD1a has been crystallized in complex with self- lipids sphingomyelin, lysophosphatidylcholine (Birkinshaw et al. 2015), sulfatide (Zajonc et al. 2003), as well as the my- cobacterial lipopeptideanalog didehydroxy-mycobactin (Zajonc et al. 2005). The binding pocket of human CD1a is composed of a double-chambered cavity termed the A^{0} and F^{0} pockets with a single (A$^{'}$) portal that coordinates the pre- sentation of lipid antigen to the TCR (fig. 1B) (Zajonc et al. 2005). The TCR lands just above the A$^{'}$ pocket on a surface termed the A$^{'}$ roof (Zajonc et al. 2005). The diminutive size of the CD1a A$^{'}$ pocket is thought to be formed by the electro- static interaction of the two side chains belonging to the A$^{'}$ roof that also draw the two parallel alpha helices of the pocket in close proximity, whereas an amino acid sidechain blocks the base of the pocket thereby limiting size of tail groups that can be accommodated (in human CD1a this amino acid is valine 28) (Zajonc et al. 2003). Several other CD1 homologs, except for CD1c, lack this roof structure (Blumberg et al. 1995). CD1a does not feature a late endo- somal targeting element and does not require low pH for antigen binding as is the case for other CD1 proteins such as CD1b, CD1c, and CD1d (Chancelloret al.2018).

The immense diversity of the MHC family within and be- tween populations at surfaces necessary for peptide antigen recognition has made these genes classic study systems of adaptive protein evolution (Danchin and Pontarotti 2004; Castroet al. 2015; Grimholt 2016). CD1 molecules possess similar structure and function to class I and II MHC proteins, although their relative lack of diversity at the population level has been attributed to a lack of

diversity in their cognate lipid ligands. Although variation in pathogen-derived lipids has been implicated in host immune recognition and virulence (Chandler et al. 2020), the potential for lipid antigens to pro- mote evolutionary conflicts with host species is unclear. In the present study, we used the CD1 family as a system to inves- tigate the diversity and evolution of lipid antigen recognition by the vertebrate immune system.

Results

Diversification of the CD1 Gene Family in Primates

A comparison of MHC class I and CD1 protein structures illustrates the homology between these antigen presentation molecules (fig. 1A). CD1 presents antigen to the TCR with the lipid tail groups tucked into the hydrophobic pocket and head groups exposed where they are "read" by the TCR (fig. 1B). Distinct CD1 molecules present antigen to a wide variety of T-cell types (fig. 1C) (Godfrey et al. 2015). To assess patterns of genetic diversity among primate CD1 family members, we first assembled a collection of simian primate CD1 homologs from publicly available genome databases and generated a phylogenetic gene tree using PhyML (fig. 1C and supplementary fig. S1, Supplementary Material online). The five human CD1 paralogs are present in the majority of pri- mate genomes surveyed, allowing us to reliably compare structural and genetic diversity within this family. A compar- ison of the sequences between CD1 orthologs revealed a striking degree of diversification, particularly in the MHC- like domain responsible for lipid antigen presentation (fig. 1D). To assess the potential consequences of this varia- tion on CD1 function, we plotted the structural conservation among primate CD1a orthologs using color by conservation (Mura et al. 2010) (fig. 1D and E). Our analysis revealed several hotspots of high amino acid divergence among CD1 mole- cules, focused on both interior regions of the antigen-binding pocket as well as surface helices that are known to contact the TCR. Together our results indicate that, despite their limited

polymorphism within populations, CD1 paralogs exhibit a high degree of genetic divergence between simian primate species.

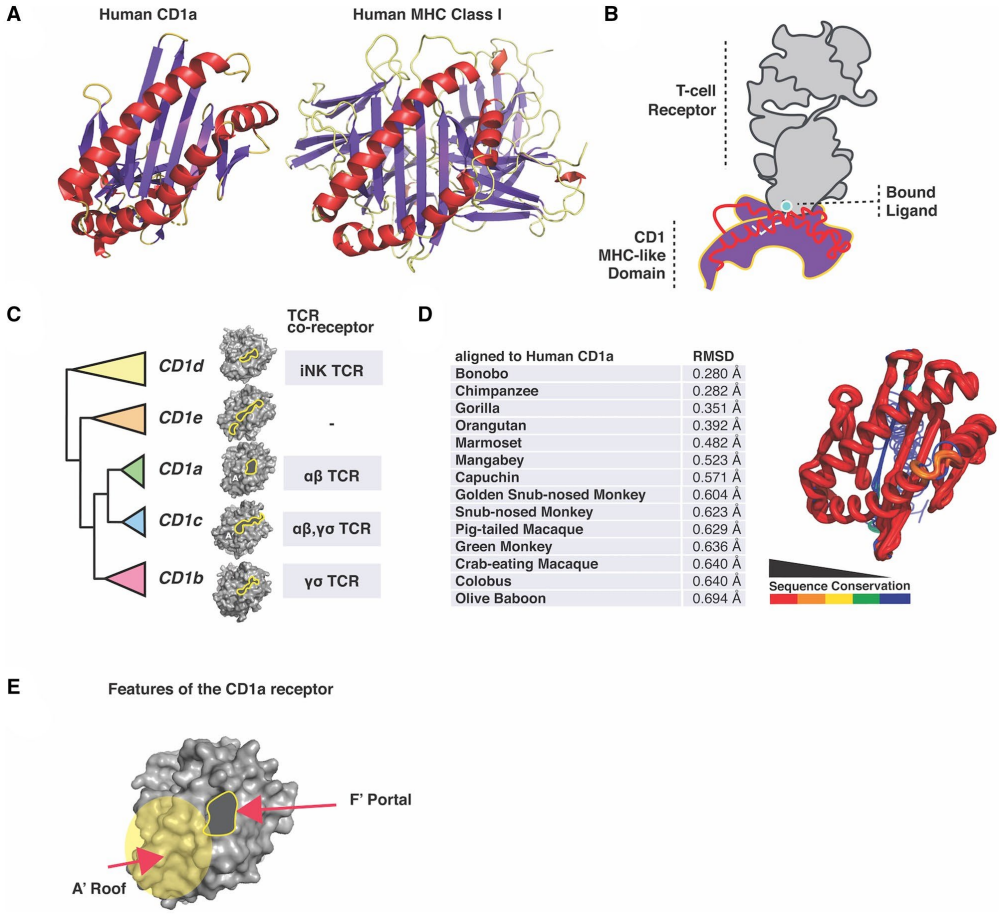


FIG. 12. Diversity of the CD1 gene family in primates. (A) Ribbon diagrams of human CD1a (PDB ID: 5J1A) and MHC (PDB ID: 1AKJ) with alpha helices highlighted in red, beta-sheet in purple, loops in yellow. (B) Illustration of the CD1-TCR interaction where CD1 bind lipid tails in a hydrophobic pocket with polar head groups typically exposed. The TCR (gray) "reads" the displayed antigen leading to T-cell activation. (C) Cladogram representing phylogenetic relationship of primate CD1a-e paralogs used in this study with surfaces generated in PyMol and antigen- binding pockets outlined in yellow (PDB IDs: 5J1A, 4ONO, 4MQ7, 3S6C). TCR types that recognize each CD1 paralog are also indicated. (D) Primate CD1a diverges most in the lipid-binding domain, which may alter pocket morphology and TCR interactions. Most of the sequence divergence in the primate CD1a proteins is predicted to exist in the beta-sheet that transects the center of the protein with some variation in the central surface region. RMSD, root mean square deviation. (E) Structural features of the CD1a receptor. Labels showing location of A' Roof, F' portal.

Signatures of Repeated Positive Selection Acting on Primate CD1 Genes

Given their relative lack of within-population diversity, we were surprised by the elevated genetic divergence between primate CD1 orthologs. Previous studies of CD1A genetic diversity in humans revealed only three low-frequency poly- morphisms (Han et al. 1999). Across primates included in this study, however, only 52.8% sites are identical. We hypothe- sized that this variation could be the result of repeated positive selection in response to diverse lipid antigen struc- tures. To identify potential amino acid sites that may have been subject to repeated positive selection, we used the codeml package from PAML (Yang 2007) in addition to MEME (Murrell et al. 2012) and FuBar (Murrell et al. 2013) algorithms from the HyPhy software package to detect ele- vated dN/dS (x) at sites in CD1 paralogs across primates. Elevated x values were detected for all CD1 family members (fig. 2A) with the exception of CD1b, consistent with elevated nonsynonymous substitution rates associated with positive selection. We noted that the majority of rapidly-evolving sites among CD1 paralogs were focused in the MHC-like domain sites between CD1a orthologs cluster to a region of the pro- tein near the center of the binding pocket and around the outer surface (fig. 2B). We grouped all of the rapidly-evolving sites we identified in this study into three categories: residues located at or near the TCR landing site (the A' roof in the human structure), residues within the binding pocket, and residues in the N-terminus for which we have no structural information. Overall, predicted structural features do not cor- relate well with phylogenetic relatedness, consistent with multiple lineages undergoing episodic selection (figs. 2C and 3A).

Accelerated Evolution of the CD1a-TCR Interface

To assess how variation in CD1a may influence immune functions in primates, we used I-TASSER to generate pre- dicted structures of several CD1a orthologs (fig. 3A). Of the hominoid structures modeled, human,

bonobo, and orangu- tan have remarkably different topologies at the TCR-CD1a interface as well as the geometry of the internal binding pocket (fig. 3A). The morphologies of the binding pockets vary widely, most notably in the crab-eating macaque which is predicted to contain one main and two accessory portals, with a narrow meandering channel (fig. 3A, bottom panel). The length and volume of the pockets limit the types of lipid tail groups that can be accommodated, whereas the size and location of the portals have effects on how well the T-cell receptor can read the antigen presented (Birkinshaw et al. 2015).

In human CD1a, the A' roof is hypothesized to aid in determining whether an antigen will elicit an immune response by supporting interactions with the TCR and assisting in display of the ligand head-group (fig. 3B) (Zajonc et al. 2005; Birkinshaw et al. 2015). The predicted orangutan CD1a structure lacks an A' roof entirely (fig. 3C), whereas bonobo CD1a possesses two portals. Additionally, it has been speculated that disruption of hydrogen bonding between R73, R76, and E154 that form the A' roof may indicate whether a given ligand will stimulate TCR activation (Birkinshaw et al. 2015). However, several of the CD1a structures are predicted to form an A^0 roof that does not depend on this particular interaction. For example, crab-eating macaque and olive baboon CD1a are predicted to form a relatively unique A^0 roof composed of an R73/153Q linkage that does not involve R76 (fig. 3C). The TCR does not recognize CD1a-bound ligands without ade- quate projection of lipid head groups, and it is likely that hydrogen bonding between the head groups of smaller ligands and residues that make up the portal are important for display. Headless ligands buried in the CD1a pocket, for example, can result in T-cell auto-reactivity (de Jong et al. 2010). Site 153, which is highly variable across primates (fig. 2D) has been shown to form a hydrogen bond in human CD1a to the head group of self- antigen lysophosphatidylcholine and sulfatide in addition to its role in forming the A^0 roof (Zajonc et al. 2003;

Birkinshaw et al. 2015). This site bears a glycine in orangutan, with no ability to form an A^0 roof or salt bridges with ligand (fig. 3C). Together, these predicted struc- tural differences suggest that natural selection may have had a significant impact on the ability of CD1a to display self or foreign lipid antigens across related primates.

Structural Remodeling of the CD1a Antigen-Binding Pocket

To determine whether the structural differences observed across CD1a paralogs are likely to have functional consequen- ces for antigen recognition, we applied a ligand-docking approach using AutoDock Vina. We were particularly interested to test affinity differences between endogenous and exoge- nous lipid ligands, since the current model for CD1a antigen presentation is a swapping mechanism wherein lower affinity endogenous ligand is replaced at the cell surface by higher affinity exogenous ligand. We used ligands previously crystal- lized in complex with human CD1a in our studies since there is a wealth of structural information available on these par- ticular binding interactions. In our docking simulations, we found that a single loop region is required to redock all ligands in the CD1a-binding pocket. We assigned flexibility to this region for all the structures tested, as well as any nonbonded sidechains in the region of the portal (supplementary fig.S2, Supplementary Material online). We believe this is likely the region responsible for conferring flexibility in the native CD1a, which must be flexible enough to accommodate ligands with a diversity of molecular weights (molecular weight of urushiol is 330 g/mol, dideoxymycobactin is 838 g/mol).

We next measured predicted CD1a-binding affinities for the panel of lipid ligands including endogenous ligands sphingo- myelin, sulfatide, lysophosphatidylcholine, and exogenous li- gand dideoxy-mycobactin (DDM) (fig. 4A and supplementary figs. S5–S7, Supplementary Material online). We chose these lipids because published structural information exists for each ligand bound to the human CD1a receptor. Given that CD1a is believed to

swap endogenous lipids for exogenous lipids based on differences in relative binding affinities, we then estimated the likelihood of a lipid-swap using our panel of ligands. We calculated the fold differences in Kd (dissociation constant) between the highest affinity endogenous lipid and compared this to the Kd of the exogenous ligands. Crab-eating macaque, snub-nosed monkey, olive baboon, capuchin, mangabey, bo- nobo, and human were predicted to swap out endogenous for DDM (fig. 4B). It is worth noting in this case that we assume the endogenous lipid with the lowest Kd is also present in abundance, which we can not know for certain in vivo. It has been shown in previous studies that human CD1a molecules bind to a diverse repertoire of lipid types in vitro (Birkinshaw et al. 2015). Since lipid profiles are not available for all cell types in the primates we studied, we chose this simplification as a rough estimate for the feasibility of lipid swapping.

A notable result from these ligand docking predictions was that binding profiles failed to group by species phylogeny, consistent with branch-site test results that detected several branches undergoing multiple bouts of episodic positive selection (supplementary fig.S5, Supplementary Material online). Unlike in humans where the largest binding pocket (CD1b) (Ly and Moody 2014) also has the most promiscuous ligand- binding profile, ligand docking predictions do not group higher affinity binding with predicted pocket volume (fig. 4C and which is responsible for lipid binding (fig. 2A). These results suggest that multiple members of the CD1 family have undergone repeated episodes of positive selection in simian primates specifically within regions important for lipid antigen presentation.

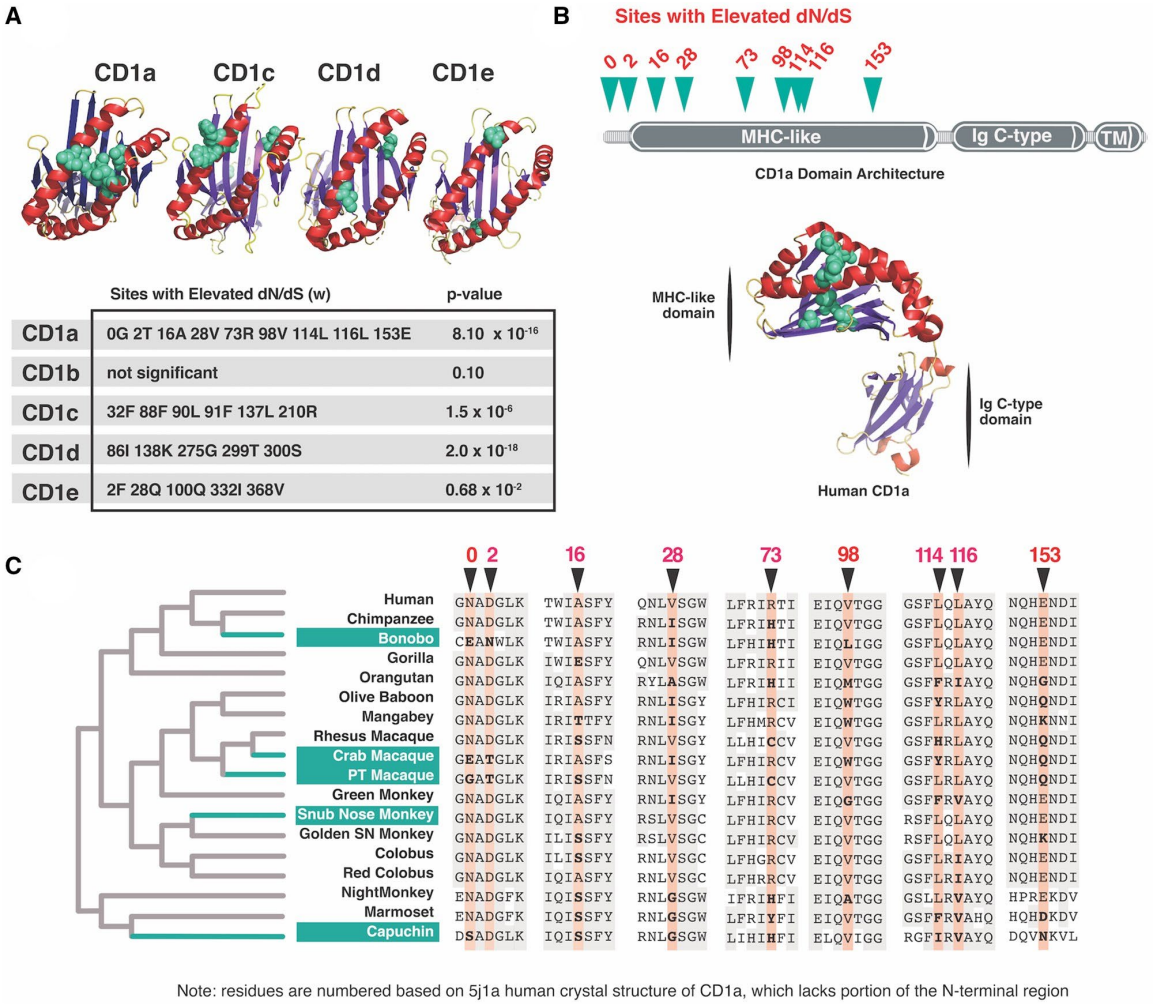


FIG. 13. Evidence of repeated positive selection among primate CD1 orthologs. (A) Amino acid sites exhibiting strong signatures of positive selection (elevated dN/dS) are highlighted in teal and mapped onto corresponding crystal structures. Alpha helices are denoted in red, beta-sheet in purple. (PDB IDs: 5J1A, 4ONO, 4MQ7, 3S6C). Table summarizes positions in CD1 paralogs contributing to signatures of positive selection as well as statistics from PAML M7-M8 model comparisons. (B) Sites with elevated dN/dS indicative of positive selection (teal) cluster in the MHC domain of CD1a protein (PDB ID: 5J1A). Alpha helices denoted in red, beta-sheet in purple. (C) Multiple sequence alignment of primate species used to calculate dN/dS ratios for CD1a paired with phylogenetic species tree highlighting the branches (teal) predicted by aBSREL to be undergoing episodic positive selection.

Having detected evidence of positive selection acting on several CD1 family members, we chose CD1a for additional in-depth analysis. CD1a has less stringent lipid loading requirements than other CD1 homologs as it does not require a reduced pH environment encountered in late endosomes, nor does it have a known adapter protein required for antigen loading (Barral and Brenner 2007). For these reasons, we an- ticipated that empirical and molecular modeling studies of antigen recognition would be less complex for CD1a than other paralogs. CD1a has been shown to present mycobac- terial antigens from the cell surface where it interacts with langerin on Langerhans cells (Mizumoto and Takashima 2004), a specialized dendritic cell type that surveys epithelial monolayers for molecular indicators of infection.

To determine what domains of CD1a are subject to pos- itive selection, we mapped the sites with high x values from our previous analysis. Results show clustering of rapidly evolv- ing sites in the MHC-like domain, similar to those observed with other CD1 paralogs (fig. 2B and supplementary figs. S2– S4, Supplementary Material online). These sites also map to regions where CD1a is likely to interface with lipid antigen or the TCR, suggesting that selection may have acted to alter lipid-binding and T-cell interactions. The majority of variable supplementary fig.S8, Supplementary Material online). These findings indicate that predicted structural alterations in the CD1a ligand-binding pocket have significant impacts on recog- nition of both endogenous and pathogen-derived antigens which collectively shape downstream T-cell activation.

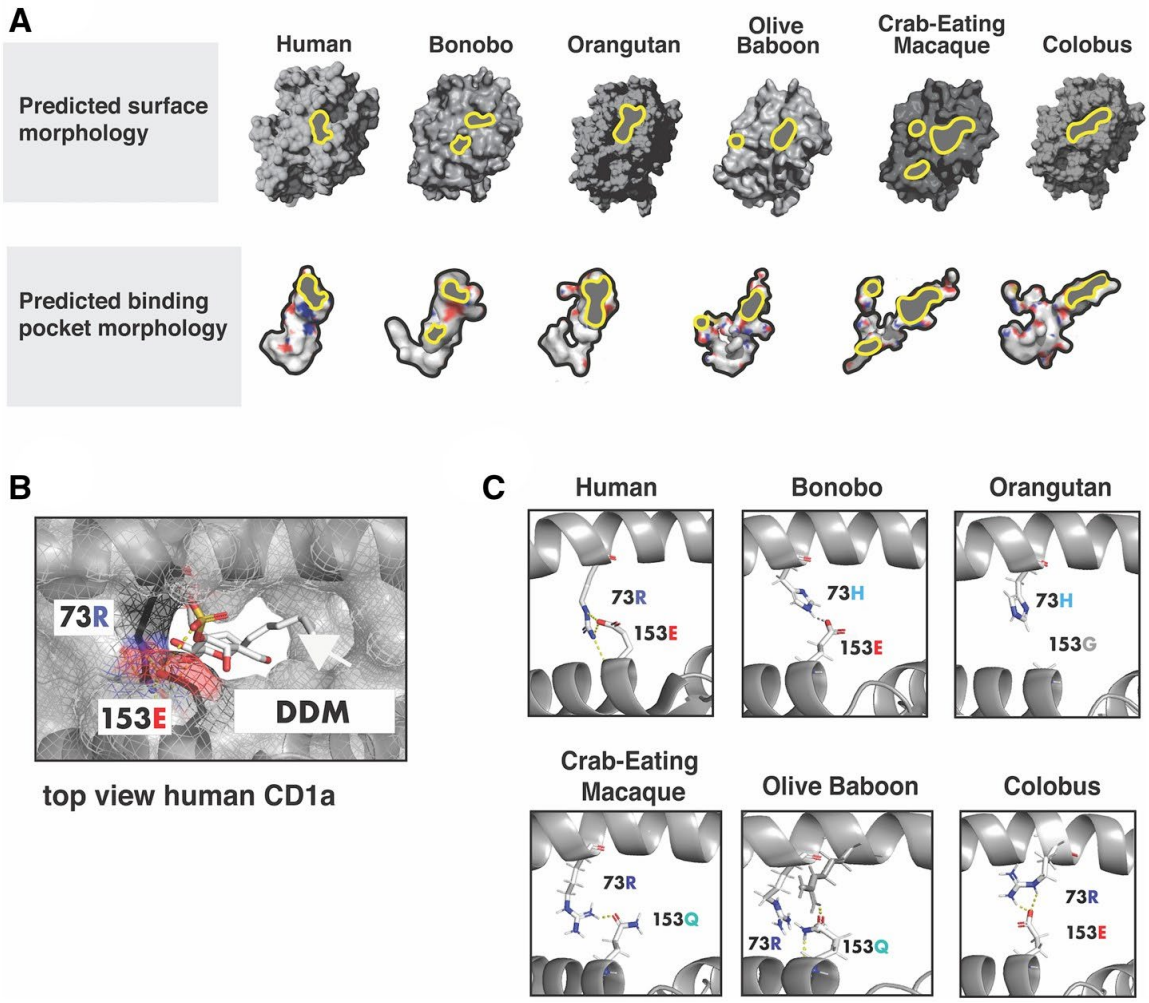


FIG. 14. Structural modeling illustrates diversity at the CD1a T-cell interaction interface. (A) Predicted attributes of various primate CD1a structures. Surface characteristics across selected primates reveal differences in portal size, number of portals, and pocket morphology. Portals where T-cell receptor "reads" head group are highlighted with gray/yellow outlines. Pocket morphologies and electrostatic properties are shown below surface models. (B) PyMol generated top-view of human CD1a bound to dideoxy-mycobactin (PDB ID: 1XZO). Rapidly-evolving positions 73 and 153 coordinate head groups of antigenic ligand. Note hydrogen bonding between head group and 153E. (C) Primate CD1a A^0 roof predicted structures where CD1a interacts with TCR. Notably, the orangutan model does not form roof structure due to mutation at site under selection. Olive baboon and crab-eating macaque form A' roof with residue of differing property at site 153.

Modeling the Effects of Rapidly-Evolving Sites on Antigen Presentation by CD1a

We next assessed how variation at single rapidly-evolving posi- tions in the CD1a-binding pocket may alter lipid antigen rec- ognition. Using PyMol, we substituted single extant amino acids for the ancestral amino acid at sites undergoing positive selec- tion (ancestral sites predicted by DataMonkey package SLAC; Pond and Frost 2005) and used these altered structures in our ligand-docking simulation. We then tested the effects of mutations in positively-selected sites on crab-eating macaque CD1a. We observed that the W98G substitution (which replaces a bulky tryptophan at the base of the pocket for the smallest residue, ancestral glycine) significantly increasedbinding affinity for endogenous lipids in crab-eating macaque, thus making it unlikely that swapping for DDM would occur (fig. 5A). This mutation appears to have similar effects in other genetic back- grounds as well, including humans (fig. 5A and B). Analysis of the binding pose in crab macaque W98G bound to lysophosphatidylcholine shows the ligand buried in the pocket without an exposed head group (de Jong et al. 2010) (fig. 6A). This provides a possible explanation for why the reduction in accessible pocket volume may be beneficial, both for lipid swap- ping and TCR ligand recognition. In the human V98W muta- tion, we noticed that the tail group accesses deeper regions of the pocket, which may partially explain the higher affinity for DDM seen in this model (fig. 6B).

To probe our system further, we used the genetic back- ground of snub-nosed monkey to simulate the effects of muta- tions since it encodes primate consensus residues at positions with elevated dN/dS. We mutated seven sites that appear at the interaction interface to the ancestral sites at all loci, resulting in a protein that is not likely to swap endogenous ligand for DDM by our predictions (supplementary fig.S6, Supplementary Material online). Smaller effect mutations were identified when introducing combinations of mutations in crab-eating macaque at po- sition 114 where tyrosine appears to

lower affinity for endogenous ligand and increases affinity for mycobacterial li- gand slightly (supplementary fig.S7, Supplementary Material online). Taken together, several species are predicted to bind DDM with relatively high affinity but may not necessarily present exogenous antigen due to equally or greater affinity for endogenous lipid. This suggests that selective pressure may exist to decrease affinity for endogenous ligand in conjunction with increased affinity for exogenous antigens, resulting in increased effectiveness of CD1a-dependent immune responses. We observed that even single substitutions in rapidly-evolving sites substantially alter both endogenous and pathogen-derived lipid antigen recognition, providing further evidence for the functional impact of divergence in CD1a.

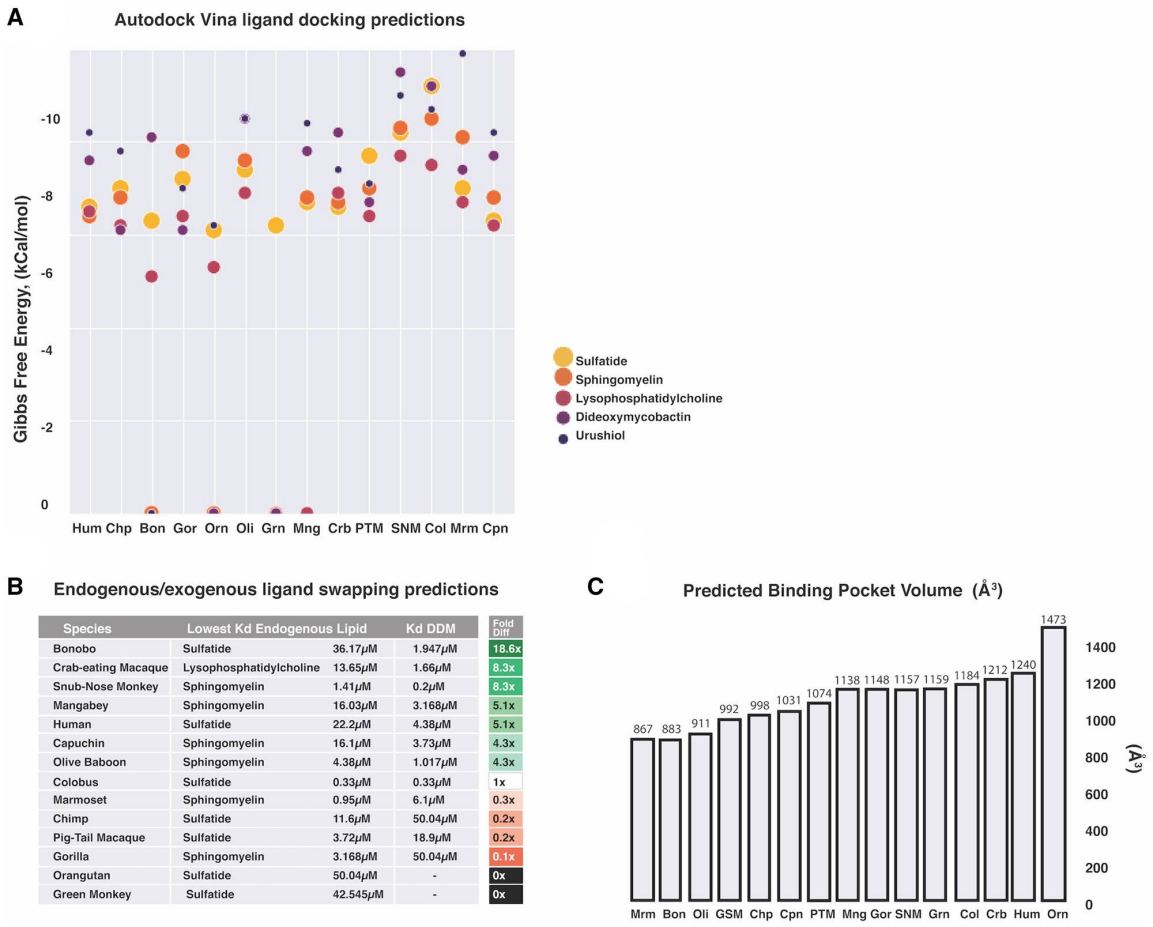


FIG. 15. Divergence of CD1a shapes predicted endogenous and exogenous lipid antigen affinities. (A) Plot of relative Gibbs free energy values for all ligands tested by ligand docking predictions using AutoDock Vina. Lowest energy values for each set are plotted. Sulfatide, sphingomyelin, lysophosphatidylcholine are endogenous lipid ligands. Dideoxymycobactin (DDM) is a synthetic lipid analog of Mycobacterium tuberculosis siderophore mycobactin. Urushiol is the etiological agent of poison ivy rash. (B) Lipid-swapping predictions based on predicted Kd (dissociation constant) from docking studies. (C) Predicted pocket volume for CD1a orthologs. Legend: Hum, Human; Chp, Chimpanzee; Bon, Bonobo; Gor, Gorilla; Orn, Orangutan; Oli, Olive Baboon; Grn, Green Monkey; Mng, Mangabey; Crb, Crab-eating macaque; SNM, Snub-nosed monkey; Col, Colobus;Mrm, Marmoset;Cpn, Capuchin.

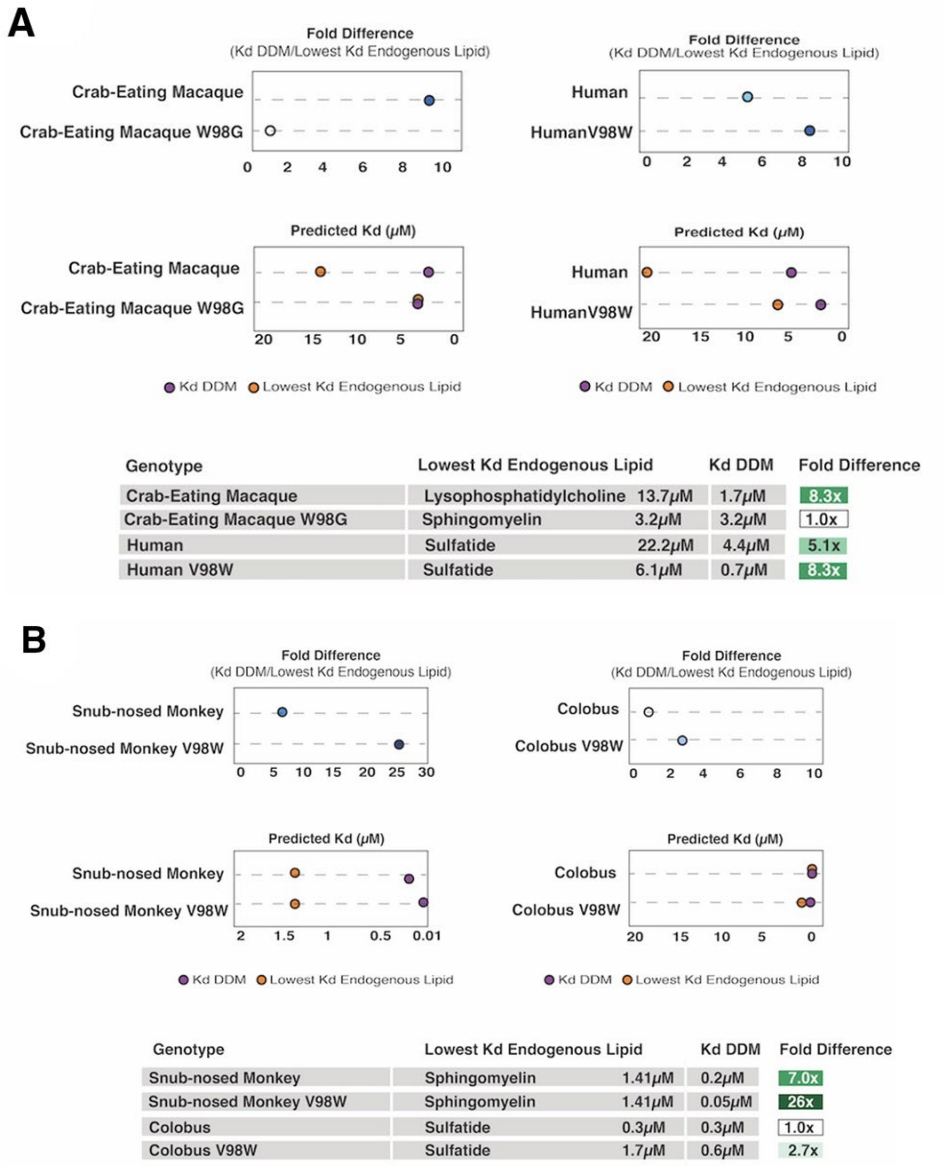


FIG. 16. Rapidly-evolving positions in CD1a are sufficient to modulate predicted affinity for lipid antigens. (A) Mutation of site 98 to tryptophan in human CD1a (olive baboon and crab-eating macaque share this amino acid at this position) results in increased predicted binding affinity to DDM, with overall fold increase between endogenous ligand and DDM. Mutation of tryptophan at site 98 in crab-eating macaque to ancestral glycine results in higher binding affinity for all endogenous ligands tested, and loss of feasible lipid-swapping and DDM presentation. (B) Mutation of site 98 to tryptophan in snub-nosed monkey CD1a results in increased predicted binding affinity to DDM. Colobus, which is not predicted to swap endogenous ligand for DDM, also increases spread between binding affinities. In colobus, however, it is a decrease in affinity for endogenous lipid rather than increase in DDM affinity that is responsible for the fold change.

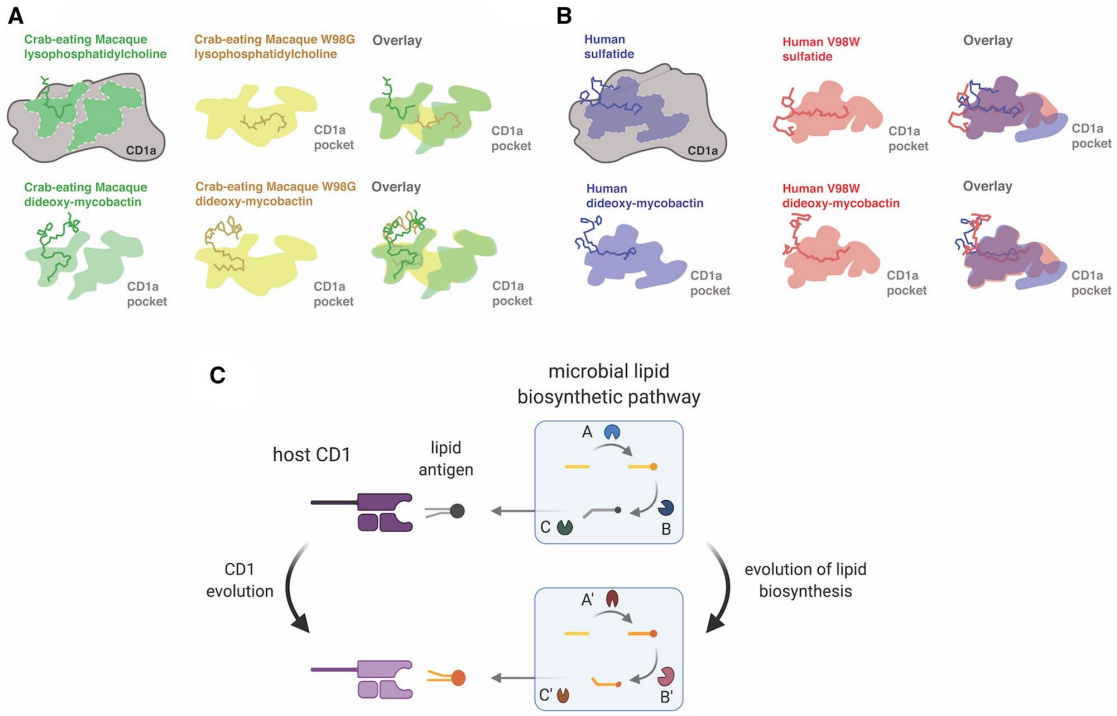


FIG. 17. Conceptual framework for lipid-driven diversification of CD1 molecules. (A) Crab-eating macaque CD1a, which encodes a tryptophan in position 98, is predicted to lose the ability to present self-lipid lysophosphatidylcholine when this position is mutated to the consensus at this site, glycine. An overlay of the differences in pocket morphology shows how the tryptophan limits access to the deeper chambers of the pocket. (B) Humans possess a valine at position 98, which has been proposed to act as a barrier limiting larger ligands access to the pocket. When this residue is mutated to a tryptophan in silico, further decreasing access to the deeper chambers of the pocket, the ability to swap out endogenous for exogenous ligand is improved, suggesting that a large hydrophobic residue in this position may be beneficial in the context of Mycobacterium tuberculosis infection in primates. Cartoons were informed by analysis of Autodock Vina docking results analyzed in PyMol. (C) Conceptual framework for lipid-driven evolution of CD1, resulting in accelerated evolution and rapid diversification of host immune receptors. Lipid biosynthesis pathways are complex and interdependent, thereby adding levels of complexity that may slow the rate at which pathogens can successfully evolve new lipid antigens. Figure created using Biorender.com.

Discussion

Overdominance has been proposed as an important force acting on MHC genes producing diversity across the gene family (Hughes and Nei 1990). CD1 genes exhibit limited sequence variation within humans (Blumberg et al. 1995), which suggests overdominance is likely not a major factor shaping the evolution of this family. Rather, our observations of elevated dN/dS be- tween CD1 orthologs and limited polymorphism within species are most consistent with a history of repeated selective sweeps

driven by positive selection. Moreover, the patterns of diver- gence in CD1, with amino acid variation enriched within the MHC-like domain, supports the hypothesis that lipid antigen recognition and presentation are the functional drivers of this divergence. These patterns are also observed in MHC genes (which are also undergoing positive selection), with elevated x at hotspots in the MHC antigen recognition groove (Hughes and Nei 1990; Manlik et al. 2019). The electrostatic property variation in lipid ligands is found almost exclusively in the head-groups, with differences in the tail groups restricted to length and ge- ometry of the hydrocarbon tails. As these tail groups have the most physical contact with CD1-binding pockets, amino acids changes affecting the length and geometry of this pocket deter- mine which hydrophobic chains can be accommodated. Patterns of evolution observed in CD1 could reflect a classical arms race in which host receptors and a subset of microbial antigens antagonistically coevolve through time. Alternatively, selection in a fluctuating environment where the fitness benefit of recognizing a particular lipid antigen changes over time could also produce elevated patterns of divergence in CD1. Coevolution between lipid antigens and host proteins would likely involve mutations in microbial genes responsible for lipid processing or modification (fig. 6C). Future studies could aid in determining how variation in lipid-modifying genes shapes CD1-dependent immune responses to specific pathogens.

Collectively our results suggest that, for species predicted to undergo lipid swapping of endogenous lipid for mycobactin, natural selection may have acted to decrease affinity for endog- enous ligand while increasing affinity for exogenous antigen by CD1a. We observed that a single substitution can significantly alter the predicted effects of ligand-binding affinity, with potential consequences for antigen presentation (fig. 5A and B). Notably, a major effect mutation identified in species undergo- ing episodic bouts of selection has the ability to reliably increase affinity for DDM and/or decrease the affinity of endogenous ligands by CD1a (fig. 5). Our analyses also indicate other residues determining binding pocket volume in human CD1a are undergoing repeated positive selection across primates (fig. 2). In particular, valine 28 has been reported to form a molecular barrier that acts as a size-limiting determinant for antigen bind- ing (Zajonc et al. 2003). Notably, New World monkeys encode a smaller residue (glycine) at this position. Replacement of valine with glycine might be expected to expand the size of the binding pocket. However, our molecular modeling indicates that the binding pocket in the New World monkey lineages is predicted to be smaller than even the crab-eating macaque or mangabey, which bear an isoleucine and a threonine, respectively, at this same site. These observations suggest that molecular determinants of binding pocket volume and morphology are complex and influenced by a combination of variable amino acid substitutions. Additionally, the size of the binding pocket does not appear to correlate with feasibility of DDM presentation. This is notable because in other CD1 molecules multiple lipids can be accommodated, negating the need for a stronger binding affinity for exogenous ligand. In fact, other CD1 molecules such as CD1b may even require “spacer” lipids (Garcia-Alles et al. 2011). These observations may reflect selection acting to produce a binding pocket that is able to swap out endogenous ligand without the need for a loading protein as seen in other CD1 paralogs. This feature enables CD1a to directly surveil the environment for pathogen-associated molecules, a potential advantage compared with the other CD1 molecules which require

lysosomal processing and accessory protein loading before antigen presentation can occur at the cell surface.

In order for a microbial pathogen to evolve alternative lipid antigen structures, mutations likely occur in genes responsible for synthesis or modification of the lipid antigen. Mutations in processing and production of lipids will most likely have effects on steps of the biosynthesis pathways that are down- stream of the mutated enzyme (fig. 6C). In the future, it would be intriguing to test whether primate CD1a orthologs have evolved to detect other lipid types or variations of mycobactin derived from other pathogen sources. According to data from NIHTPR's AceView (Thierry-Mieg and Thierry-Mieg 2006), gene expression of CD1a/c is exceptionally high in

tissues in pig-tailed macaque. Additionally, certain orthologs such as the marmoset CD1a exhibit very low gene expression (Thierry-Mieg and Thierry-Mieg 2006) and may be undergo- ing rapid birth-and-death evolution (Nei and Rooney 2005) and eventual pseudogenization. Such observations would be consistent with findings of dynamic CD1 gene duplication and loss across vertebrates (Nei and Rooney 2005). The sig- nificance of changes in endogenous lipid presentation will also be an area for important future investigation. Certain isoforms of sulfatide, for example, are associated with cancerous cells and when bound to CD1a can prime T- cells(Takahashi and Suzuki 2012). It has also been shown that presentation of endogenous ceramides by CD1d is asso- ciated with the ability to detect disease (Paget et al. 2019).

CD1 molecules possess the ability to bind and present hydro- phobic antigens from a variety of pathogens, many of which likely remain to be described. It is notable, however, that the majority of CD1 antigens identified to date are derived from pathogenic mycobacteria including Mycobacterium tuberculosis, the causative agent of tuberculosis in humans. Tuberculosis remains a devastating human public health burden, recently accounting for more

deaths due to infectious disease than any other single pathogen (Forrellad et al. 2013). It is tempting to speculate whether mycobacterial antigens have indeed imposed particularly strong selective pressure on CD1 molecules during animal evolution. Given the limited effectiveness of the current tuberculosis vaccine (Schito et al. 2015; Gonget al.2018), addition of CD1-targeted antigens in a next-generation vaccine could provide one avenue for increased efficacy (Gonget al.2018). Functional characterization of diverse CD1 orthologs beyond humans may reveal whether detection of mycobacterial antigens is a widely conserved feature in this family, as well as possible routes to enhance CD1-mediated immunity against M. tuberculosis. Alternatively, evolution-guided development of synthetic lipid antigens that confer increased activation of CD1- responsive T-cells could provide an alternative strategy to en- hance lipid-based vaccines.

Although we focused our molecular modeling and simulation studies on CD1a, comparable signatures of positive selec- tion were identified in primate CD1c, CD1d, and CD1e. Further investigation of these receptors and their cognate antigens would greatly advance our understanding of the importance for CD1 diversity in the evolution of vertebrate immunity. For this study, Autodock Vina was used because published results show strong correlation between docking and experimental values (Trott and Olson 2009) especially when iterations are increased (Jaghoori et al. 2016) as we did in this study. . Additionally, Autodock Vina has been reported to perform well with lipid ligands specifically (Gathiaka et al. 2013). However, there is improved reliability when comparing dock- ing results from the same receptor molecule bound to variable ligands (Jaghoori et al. 2016). The main caveats of this analysis exist in the uncertainties inherent in the structural prediction models. I-TASSER predictions are often very good, but rely on availability of structural information on similar molecules in the database which may not be available (Yang and Zhang 2015).

Although lipids and other pathogen-derived macromolecules have long been appreciated as critical targets for host innate and adaptive immune responses, the potential for these factors to promote evolutionary conflicts with host species has been relatively unexplored. By combining comparative genetics and molecular modeling approaches, this study illuminates how lipid antigens have shaped fundamental features of primate immunity and the detection of globally devastating pathogens.

Materials and Methods

Phylogenetic Analyses

A gene tree of primate CD1 was generated with PhyML (phy- logenetics by maximum likelihood) with Bayes selection crite- rion and 1,000 bootstraps (Yang 2007). Between 18–21 primate cDNA sequences were aligned for each CD1A-E gene using MUSCLE (supplementary fig. S8, Supplementary Material on- line), sequences were trimmed manually using the species phy- logeny as reported by Perelman et al.(2011). Our CD1A data set included all available nucleotide coding sequences (cDNA) for 19 primate species, with areas of ambiguity and stop codons removed. Positively selected sites for all CD1 genes were detected using the phylogenetic analysis by maximum likeli- hood (PAML) software package with F3X4 codon frequency model. Likelihood ratio tests compared pairs of site-specific models M1 with M2 (neutral and selection, respectively), M7 with M8 (neutral, beta distribution of $dN/dS < 1$; selection, beta distribution $dN/dS > 1$, respectively). Additional tests were per- formed which account for synonymous rate variation and re- combination, including FuBAR (Murrell et al. 2013) and MEME (Murrell et al. 2012), using the HyPhy software package (Murrell et al. 2012, 2013). We chose a stringent selection criteria for the sites we focused on in this study: PAML and FuBAR posterior probability of greater than or equivalent to 0.9, MEME P value of 0.1 or less. All sites analyzed (unless otherwise stated) fit these criteria under all three tests.

CD1a Structural Predictions

The Eukaryotic Linear Motif resource (Kumar et al. 2019) (http://elm.eu.org/, last accessed February 12, 2021) was used to identify structural motifs from the primary amino acid sequence of CD1a. Primate CD1a structures were pre- dicted with amino acid sequences submitted to I-TASSER server (Yang and Zhang 2015) (https://zhanglab.ccmb.med. umich.edu/, last accessed February 12, 2021) to generate structures for analysis using PyMol, primate structural align- ment from 14 primate structures colored by conservation based on RMSD calculations from PyMol alignment (https://pymolwiki.org/index.php/Color_by_conservation, last accessed February 12, 2021) (Mura et al. 2010), CASTp for volume predictions, and for use in ligand docking simulations. To assess confidence in our structural predictions, isoform 1 of full-length human CD1a was analyzed (there are several crystal structures available for this molecule) with a C-score of □0.36. C-score values vary from -5, 2 with positive values indicating higher confidence, and only structures with C values between -1 and 2 were used for analysis. Structures were analyzed using PyMol (The PyMol Molecular Graphics System, Version 2.0 Schro¨dinger, LLC). For binding pocket volume predictions, CASTp was used and the radius probe was set at 0.75 A˚ for each iTASSER-predicted structure sub- mitted for analysis. The predicted volumes for all species were plotted using the Seaborn package in Python.

Ligand Docking with AutoDock Vina

Redocking with human CD1a was first performed to identify flexible residues required for all known ligands to redock in the same model. We calibrated our modeling by redocking known ligands in our human CD1a iTASSER-predicted structure. According to our calculations, a comparison of CD1a crystal structures bound to the smallest and largest ligands (PDB ID 4X6D, 1XZ0) yields an RMSD of 1.23A˚. This suggests there is flexibility in the CD1a pocket, supported by a number of hydro- gen bonds between residues of

the main CD1a-binding domain alpha helices. Dorsal loop of alpha helix 2 was identified as re- quired and made flexible in all primate CD1a structures analyzed. Receptors with amino acid side chains that occluded the binding pocket were also made flexible if not engaged in hydrogen bonds, and any of these additionally flexible residues (see Supplementary Material online for details). AutoDockTools 1.5.6 (Trott and Olson 2009) was used to prepare the ligands and receptors for ligand docking. AutoDock Vina was run in the command line and docking results were analyzed in PyMol and plotted with the Python Seaborn package. A Python script was written to perform K_D calculations. Details of Vina settings in- cluding exhaustiveness, grid center, and x, y, z coordinates are available in the Supplementary Material online.

Data Availability

The data underlying this article are available in the article and in its online supplementary material.

CHAPTER V.
SUMMARY AND CONCLUDING REMARKS

Contribution to the Field of Molecular Evolution

Pathogens and parasites have evolved effective strategies to gain access to host resources. The immune system fends off these attacks, often through detection of pathogen associated molecules and clearance of infection. This results in interactions between host and pathogen that often take place at molecular interfaces of immune receptors that act as a first line of defense to infection. Such receptors must identify pathogen-specific molecules and mount an appropriate response. Due to the frequency of such high stakes interactions between immune receptors and pathogen- derived molecules, the immune system is under constant evolutionary pressure to innovate new modes of defense and detection, while the pathogen is under pressure to evade these efforts and mount offensive attacks. This dynamic evolutionary antagonism is the underlying phenomenon inspiring this work. Because proteins evolve functions through DNA modifications, we study the effects of nucleotide variation across related species and test how variation affects the dynamics of protein interactions. We find that phylogenetic relatedness is not always a predictable indicator of functional similarity in the systems we tested. In the first study in Chapter III, we found that presence of certain amino acids in ligand-binding hotspots are more likely to have an effect on whether a Staphylococcus aureus inhibitor binds to immune receptor than overall sequence homology. In the second study that comprises Chapter IV, we found a similar lack of correlation between predicted functional outcomes and familial relationship. Similar to the study in Chapter III, we found that certain sites could have an outsized effect on function that could be translated across multiple species. Interestingly, site-level similarities at "hotspot" regions were a better indicator of function than phylogenetic relationships.

APPENDIX A

SUPPLEMENTARY MATERIAL FOR CHAPTER III

Figure S1. Pre-incubation with formyl peptide prior to FliprLike binding (n=1)

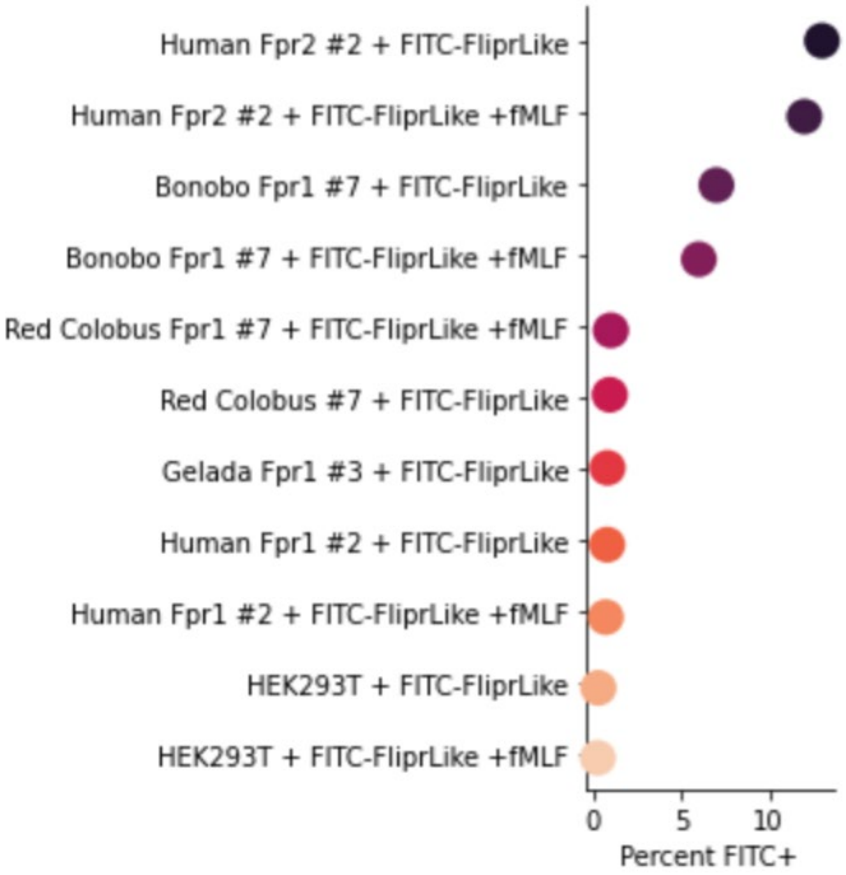


Figure S2. Pre-incubation with FliprLike before binding to S. enterotoxin B (n=1)

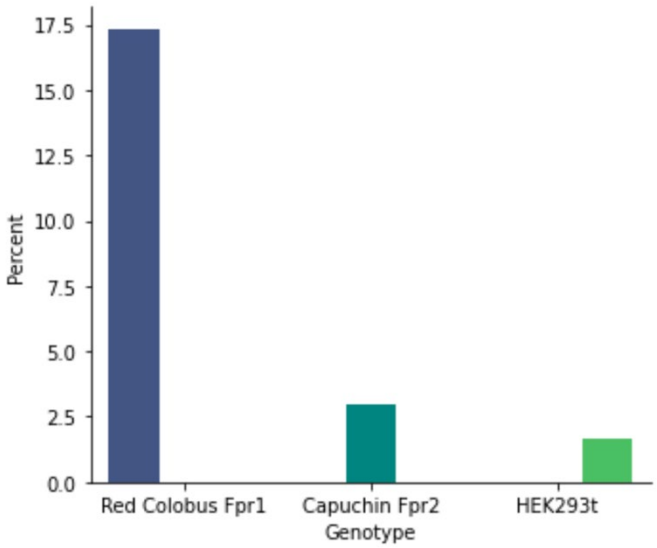


Figure S3. FPR gene tree of primates

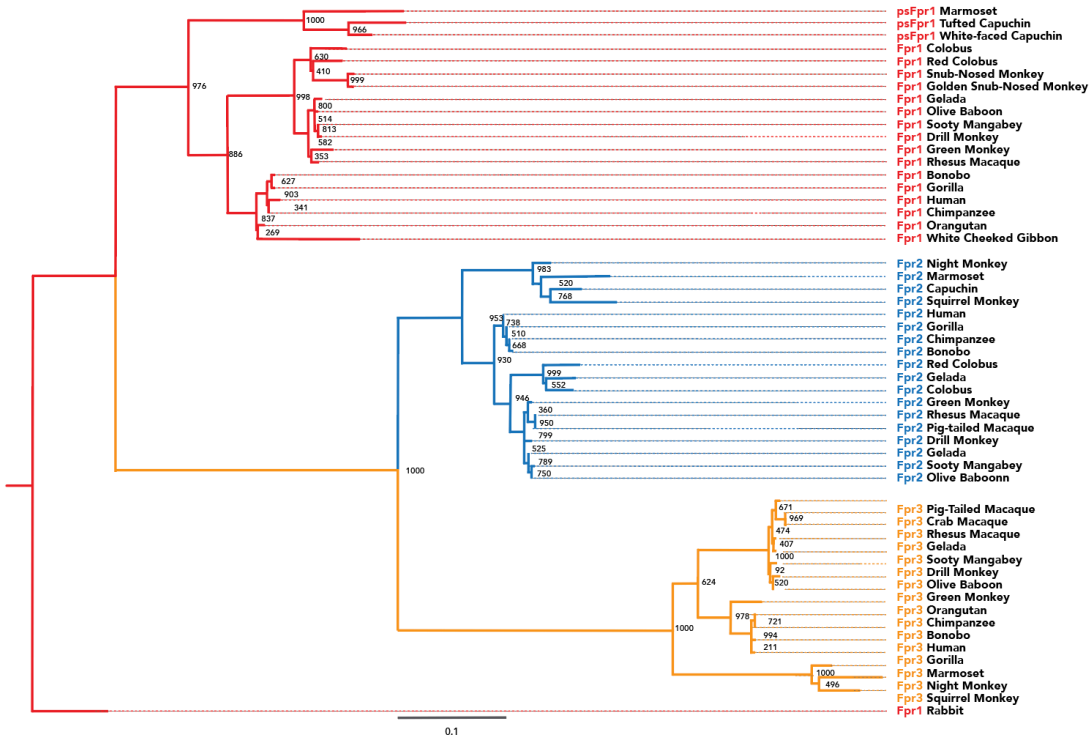


Figure S4. Few differences between Human FPR1 and Bonobo FPR1

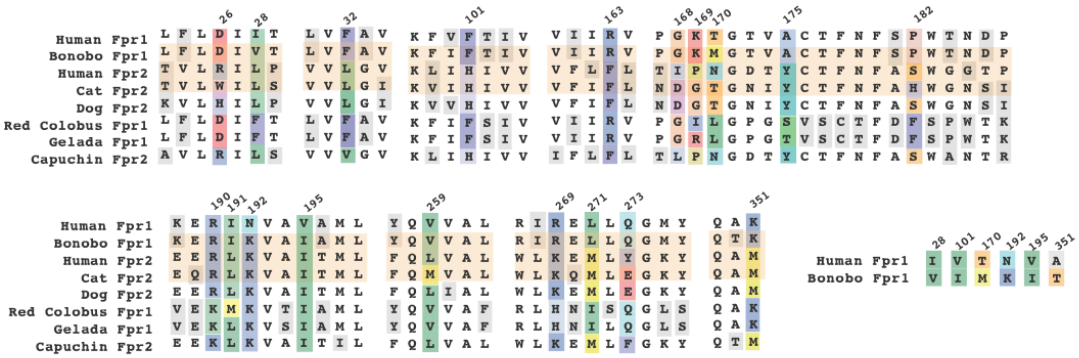


Figure S5. Denisovan and modern-day human FPR1 sites under selection
Lack of changes at positively selected sites suggests positive selection occurring in other mammalian/primate lineages in FPR1 and FPR2 in response to plague selective pressure.

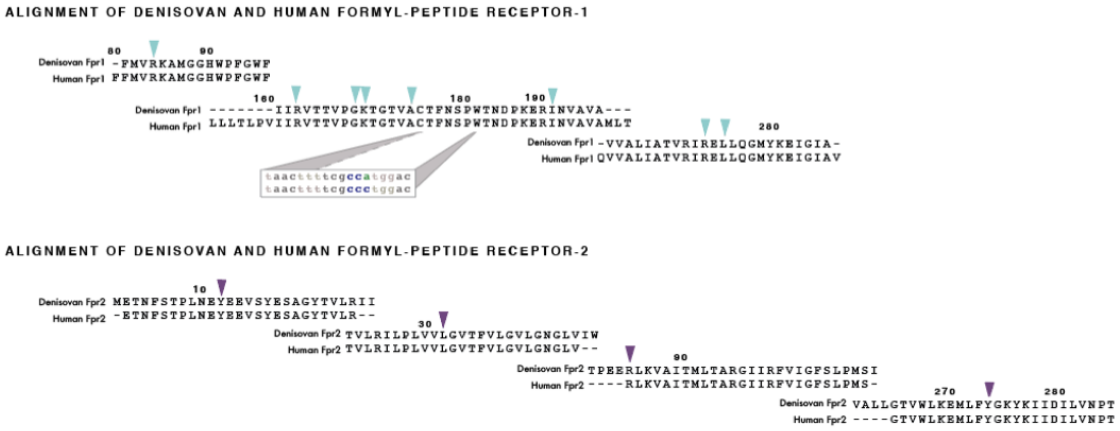


Figure S6. Indels removed in *Cebus imitator* NCBI cDNA are also present in *Sapajus apella* sequence

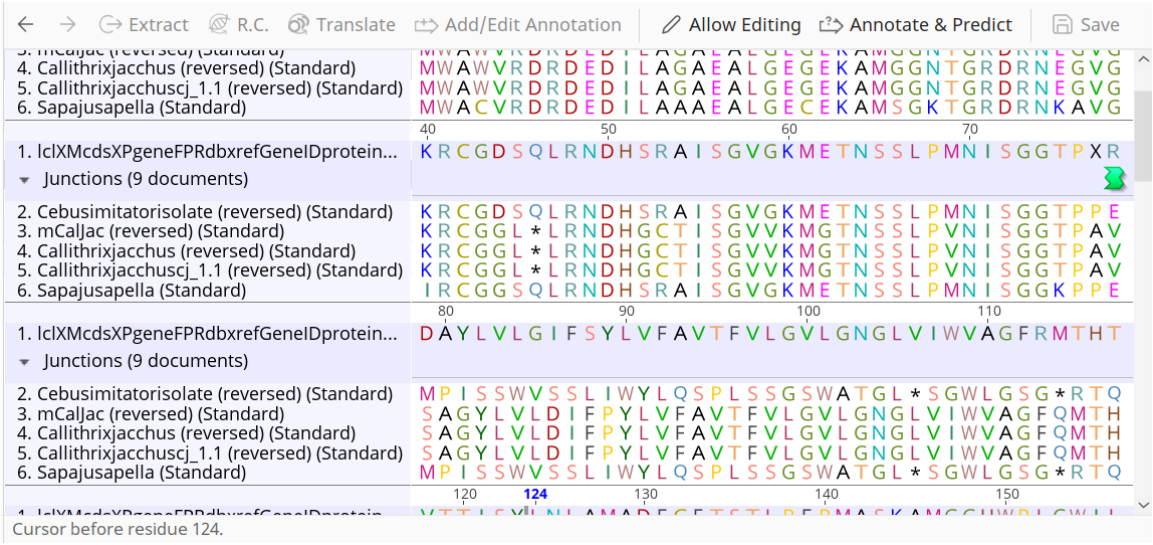


Figure S7. Insertion of three nts removed in *Cebus imitator* NCBI cDNA are also present in *Callithrix jacchus* FPR1 pseudogene sequence

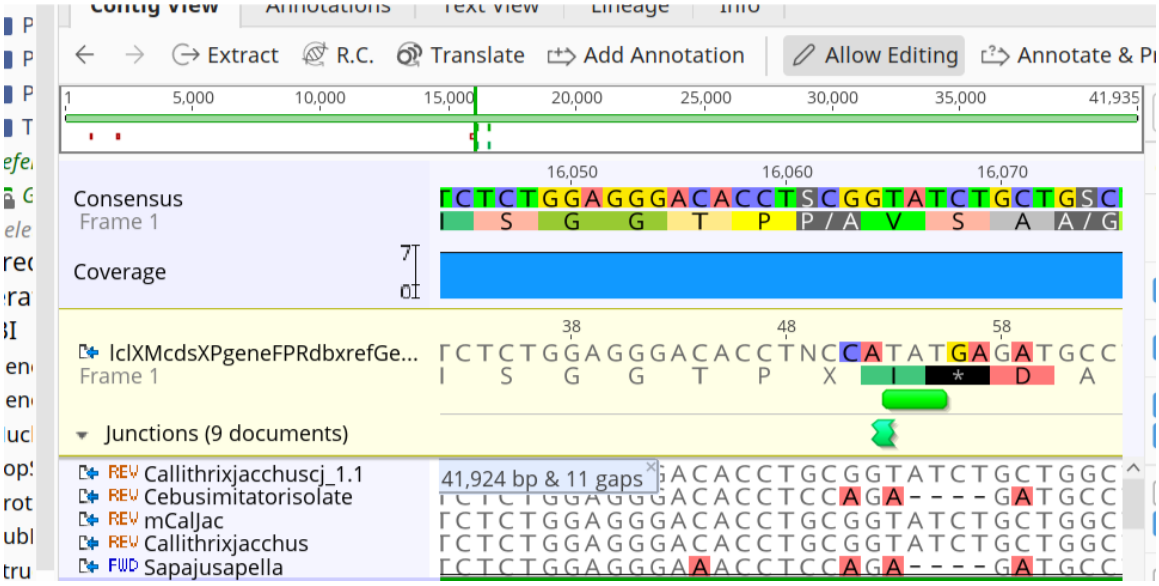


Figure S8. Alignment of region near FPR1 locus from *Aotus nancymaae* with most homology (bottom) to *Cebus imitator* FPR1 reference sequence

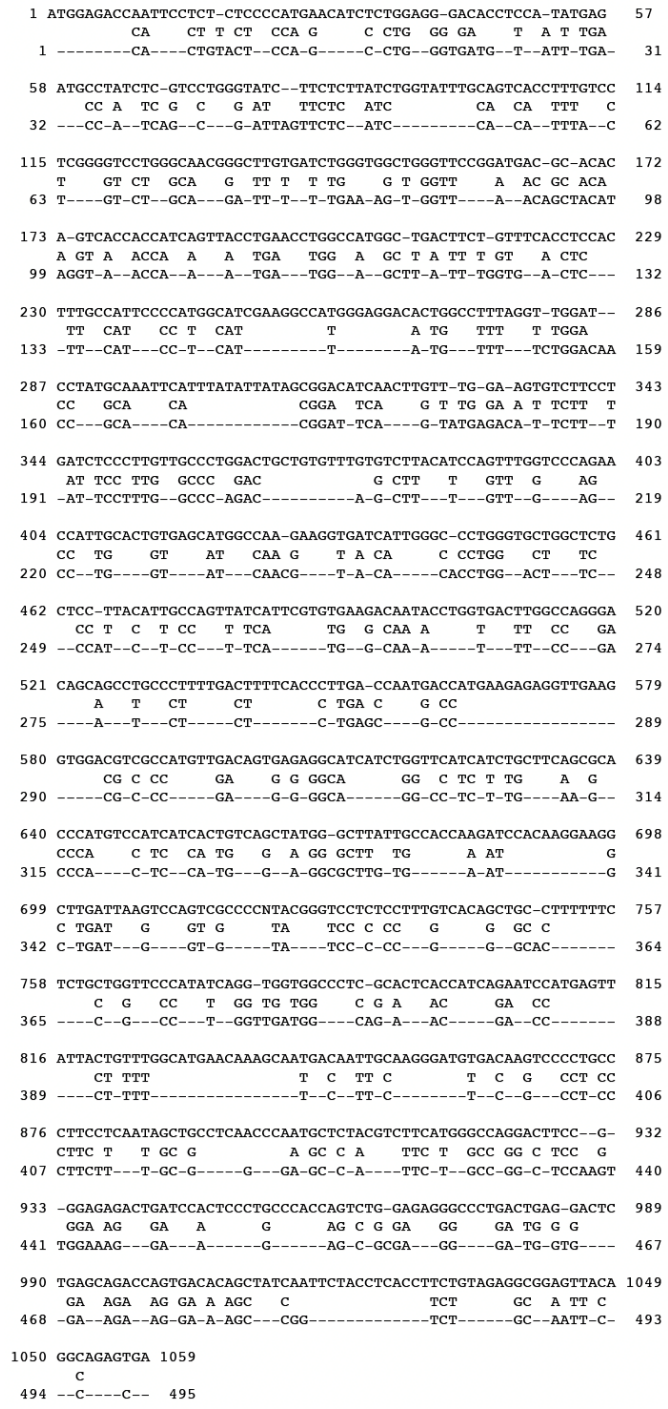
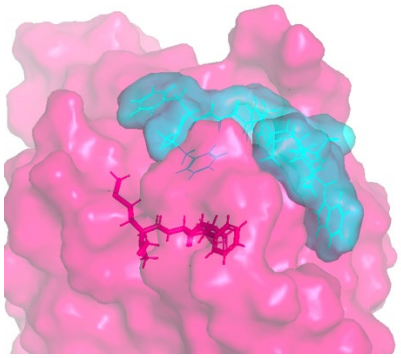
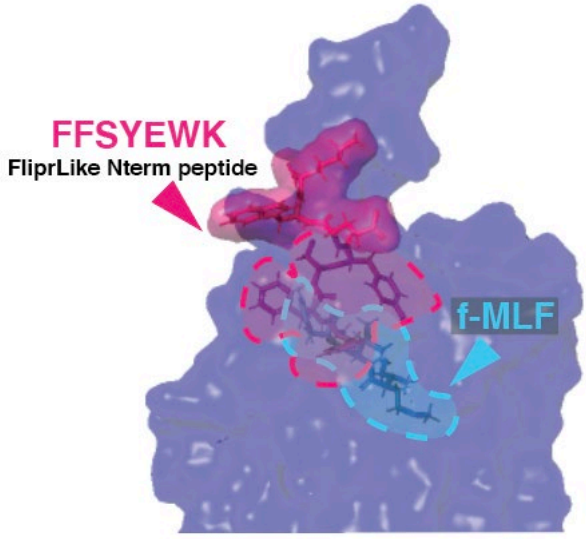


Figure S9. Binding of f-MLF to Bonobo FPR similar characteristics to human FPR2

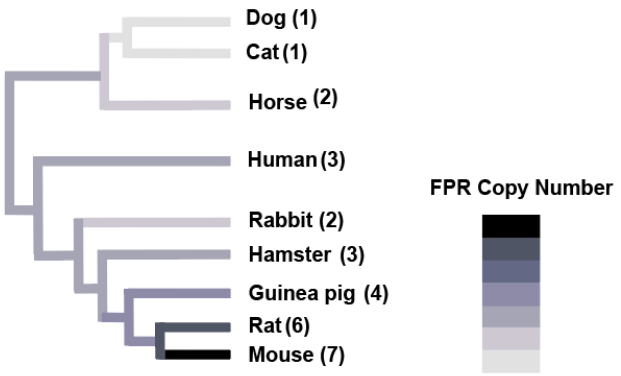


Bonobo FPR1-fMLF in dark pink, FFSYEWK in cyan



Human FPR2 docking, FFSYEWK in dark pink, fMLF in cyan

Figure S10. FPR copy number varies across mammals



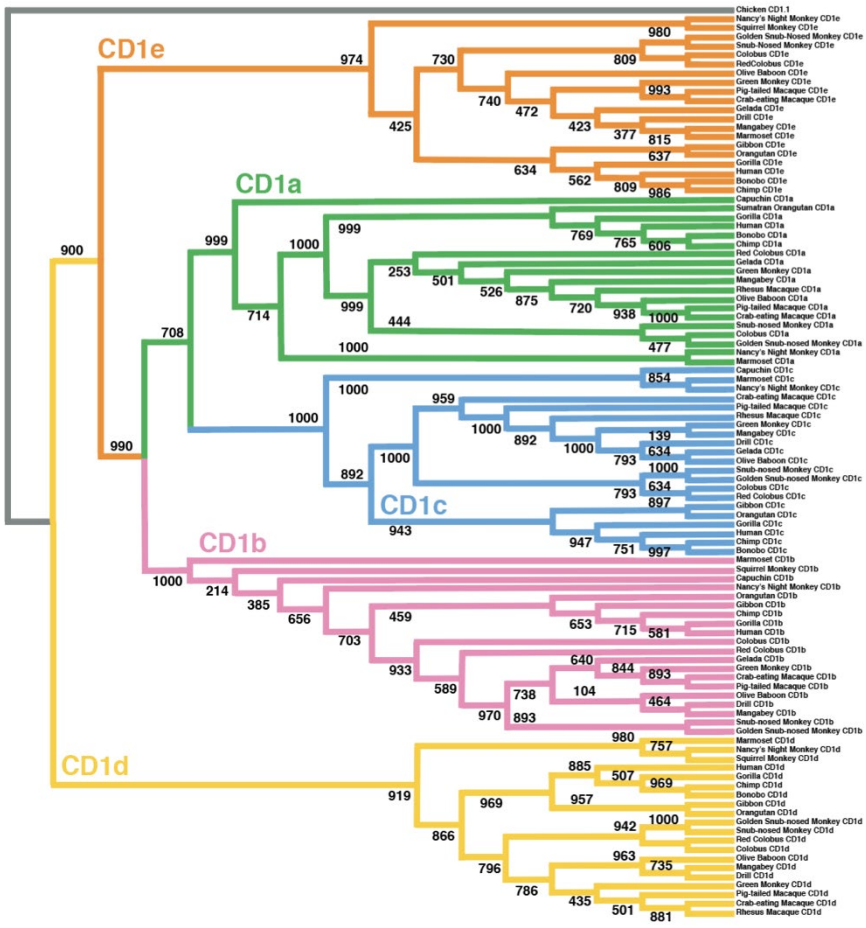
	Fpr Copy Number
Mouse	7
Rat	6
Hamster	3
Guinea pig	4
Rabbit	2
Human	3
Cat	1
Horse	2
(Dietschl et. al 2017)	-

APPENDIX B

SUPPLEMENTARY MATERIAL FOR CHAPTER

Supplemental Figures 1-8

Figure S1. CD1 Primate Family Tree



Phylogenetic relationship of primate CD1 homologs used in this study. Tree was generated in PhyML with 1000 bootstraps. Chicken CD1.1 was included as an outgroup.

Figure S2. Vina Settings: Grid box parameters and flexible residues

Species	Bounding box x	Bounding box y	Bounding box z	Flexible residues
Human	size_x = 82	size_y = 114	size_z = 126	F143-M157
Chimpanzee	size_x = 124	size_y = 76	size_z = 114	V141-N156
Bonobo	size_x = 126	size_y = 98	size_z = 124	L138-M146
Orangutan	size_x = 110	size_y = 90	size_z = 126	P150-N156
Gorilla	size_x = 118	size_y = 126	size_z = 92	L149-N156
Olive Baboon	size_x = 78	size_y = 98	size_z = 122	L149-N156
Green Monkey	size_x = 118	size_y = 82	size_z = 126	L149-N156
Mangabey	size_x = 126	size_y = 110	size_z = 108	P150-N156
Pig-Tailed Macaque	size_x = 110	size_y = 126	size_z = 90	D72-N91
Crab-eating Macaque	size_x = 84	size_y = 124	size_z = 126	T71-N92
SnubNoseMonkey	size_x = 100	size_y = 96	size_z = 126	L57-N77 E82 L83 L86 I174 L178
GoldenSnubNoseMonkey	size_x = 124	size_y = 108	size_z = 116	A150-N156 L86 I174
Colobus	size_x = 82	size_y = 118	size_z = 126	L57-N77
Marmoset	size_x = 78	size_y = 108	size_z = 108	F66-N77 L83 V97 L161 L178 V174
Capuchin	size_x = 100	size_y = 124	size_z = 96	F143-N156 L86 V170 V173 L174 R177

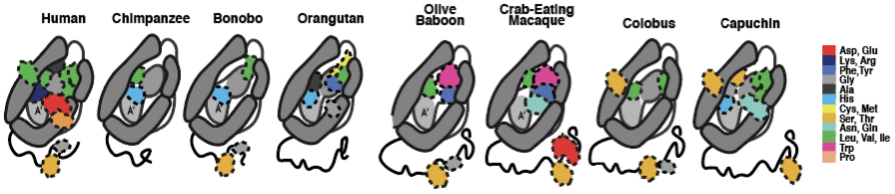
Parameter settings for config file used to generate ligand docking models with Autodock Vina, including residues set as flexible which correspond to the loop residues between alpha helices 1 and 2, and any occluding residues not engaged in hydrogen bonding at the portal entrance.

Figure S3. CD1 family members recognize a variety of mycobacteria-derived and endogenous lipids

-	CD1a	CD1b	CD1c	CD1d	CD1e
Co-receptor	αβ TCR	γδ TCR	iNKT TCR	TypeII NKT TCR	MAIT cell TCR
Mycobacterial ligand	Didehydroxymycobactin	Mycolic acids (Mycobacterium)	Mannosyl-β1-phosphomycoketides (Mycobacterium)	Phosphatidylinositol mannosides	Phosphatidylinositol mannosides (indirect)
Ligand in crystal structure	Lysophosphatidylcholine (self)	GM2 ganglioside (self)	Phosphomycoketide (Mycobacterium)	Sulfatide (self)	-
-	Sphingomyelin (self)	Phophatidylinositol (self)	Phosphatidylcholine (self)	Lysophosphatidylcholine (self)	-
-	Sulfatide (self)	Phosphatidylcholine (self)	Mannosyl-β1-phosphomycoketide (Mycobacterium)	Sphingomyelin (self)	-
-	Synthetic mycobactin lipopeptide	Phosphatidylserine (self)	-	Glycosphingolipid (self)	-
-	-	Phosphatidic acid (self)	-	Ganglioside GD3 (self)	-
-	-	Glucose monomycolate, C36 GMM 9 (Mycobacterium)	-	-	-

Multiple mycobacterial lipids and lipoproteins are recognized by CD1 receptors, suggesting that this bacterial family known for exotic lipids has been interacting with the CD1 receptors across an extended timespan.

Figure S4. Sites under positive selection that differ from consensus, illustrated in a visual cartoon. Property differences in amino acids cluster around the portal and at the TCR interaction surface, while internal residues are mainly hydrophobic residues of varying sizes



.

Figure S5. Plots of CD1a primate docking experiments by individual lipid. Dark blue-Hominoids, Grey-Old World Monkeys, Light blue-New World Monkeys.

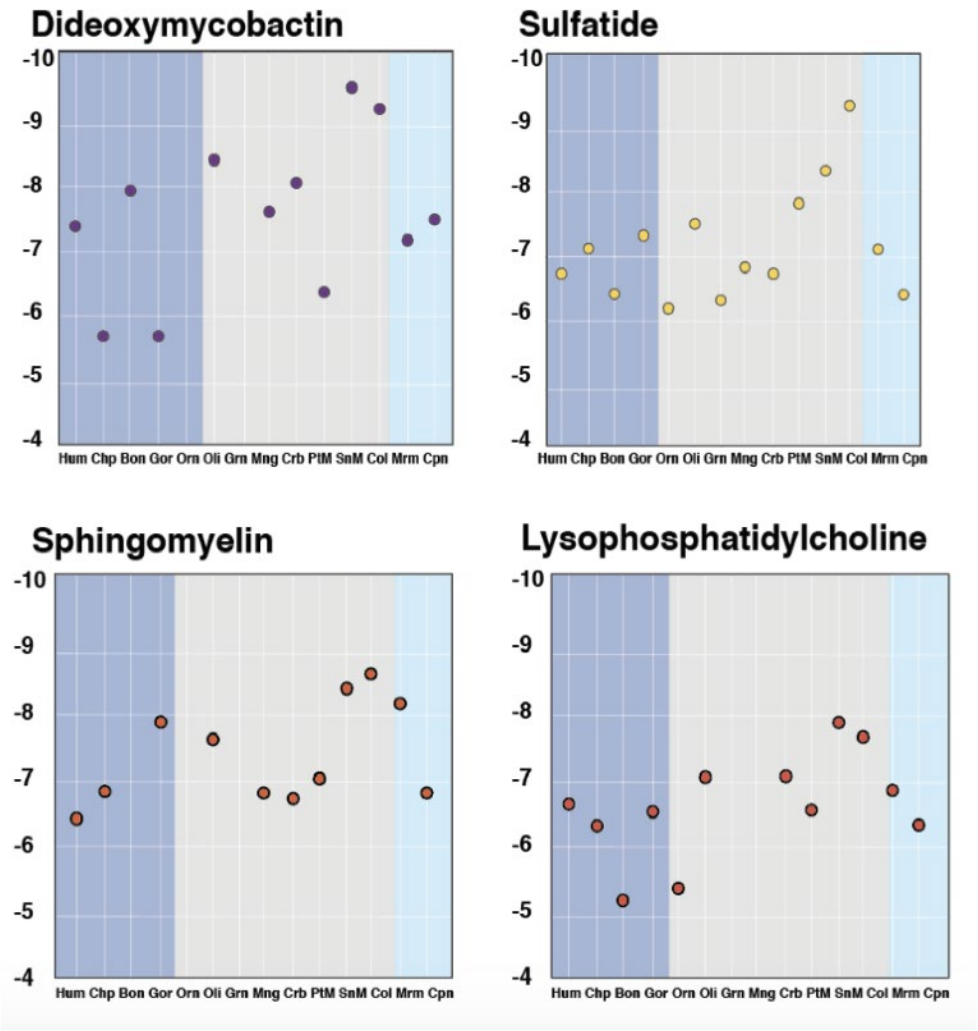


Figure S6. Snub-nosed monkey represents consensus at sites under selection. Ancestrally-predicted amino acids differ from consensus. Reversion of sites to ancestral state in snub-nosed monkey background results in lower affinity for DDM and higher affinity for endogenous lipid.

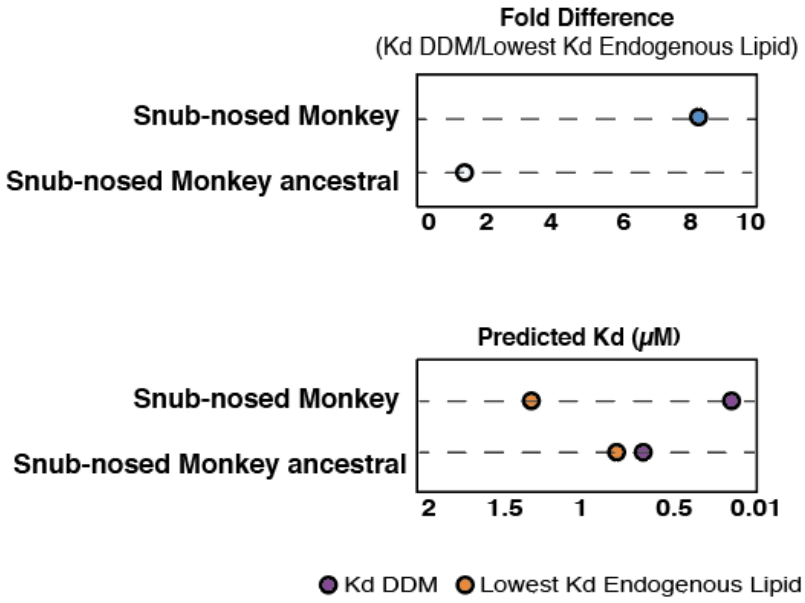


Figure S7. Additional amino acid changes confer minor increase in spread between endogenous and exogenous lipid. Small changes in affinity can be seen when other sites with high omega values are plotted in the Crab-eating macaque, such as site 114 which appears to have a similar effect to site 98, though smaller in magnitude, presumably due to the loss of bulky residue in the deeper chambers of the pocket.

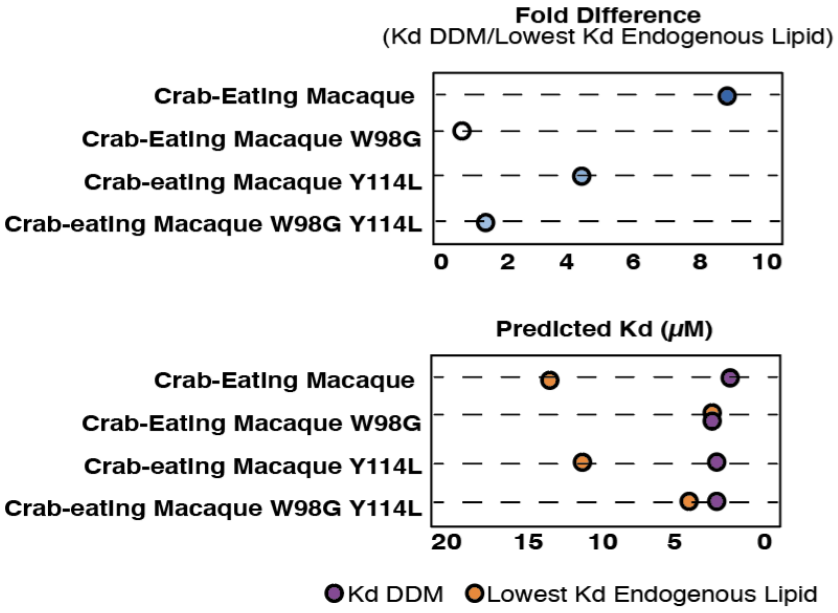


Figure S8. Pocket volume does not correlate with lipid binding affinity. Plots of docking results were converted to Kd (dissociation constant) and plotted against predicted pocket volumes. Observations did not support a case where larger pocket was associated with better binding.

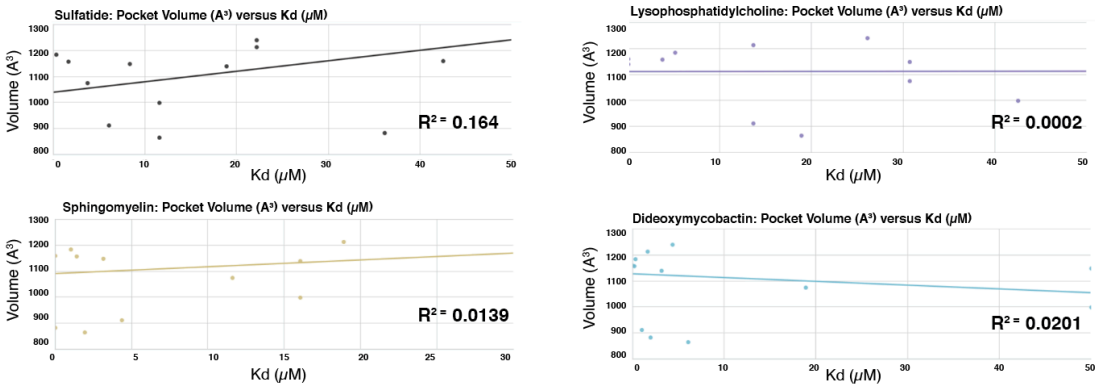


Figure S9. Summary of PAML output for CD1 paralogs

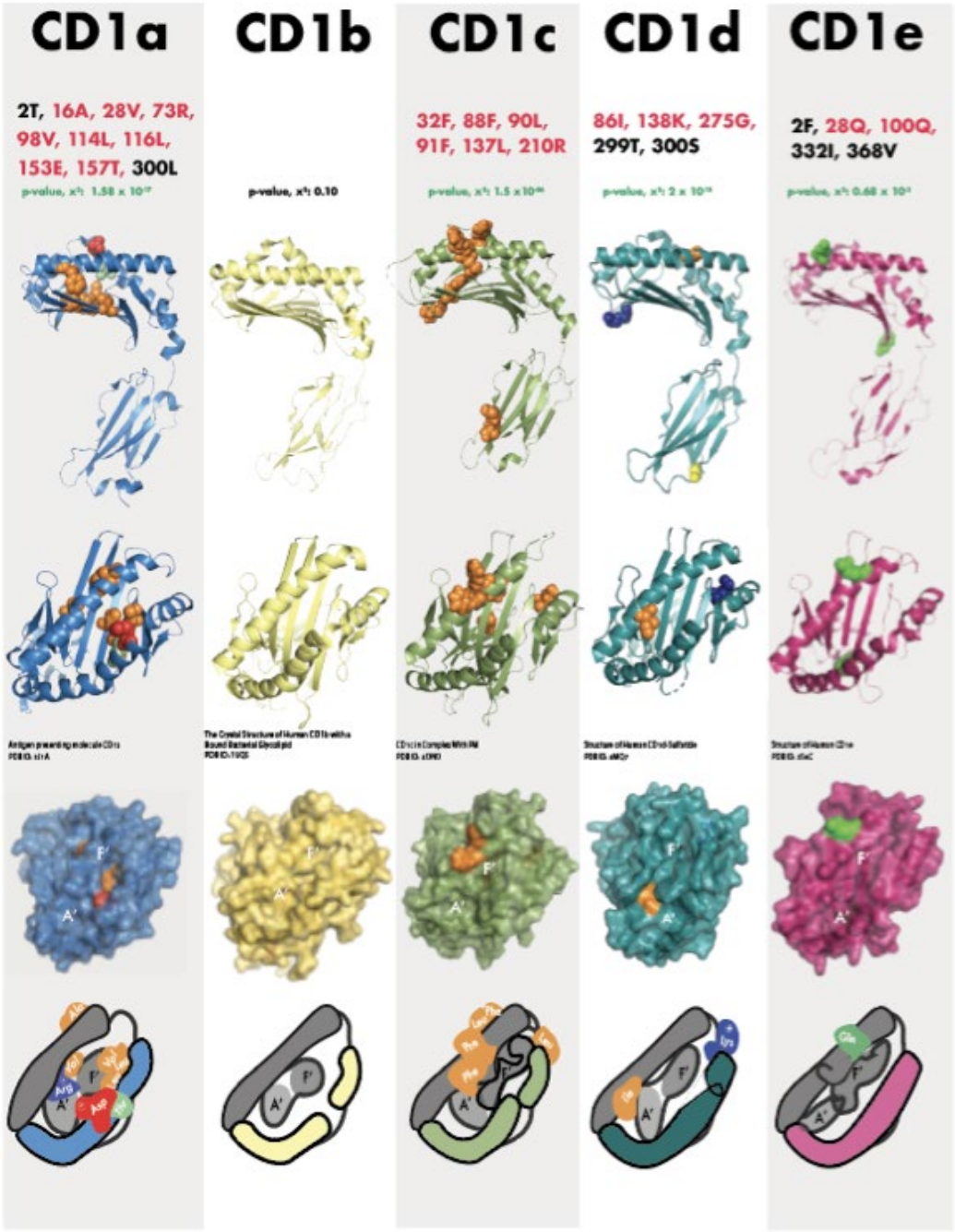


Figure S10. CD1 Sequences used for PAML codeml analysis and multiple sequence alignments

Sequences used for phylogenetic analysis with NCBI accession identification numbers.

Primate CD1a, 18 species
Homo sapiens, NCBI Reference Sequence: NM_001763.3; Macaca mulatta, NCBI Reference Sequence: NM_001145818.1; Pan troglodytes, NCBI Reference Sequence: XM_001169121.4; Macaca fascicularis, NCBI Reference Sequence: XM_005595550.2; Papio anubis, NCBI Reference Sequence: XM_003892864.5; Gorilla gorilla gorilla, NCBI Reference Sequence: XM_004027022.3; Pan paniscus, NCBI Reference Sequence: XM_024926560.2; Pongo abelii, NCBI Reference Sequence: XM_002809984.3; Aotus nancymaae, NCBI Reference Sequence: XM_012449747.1; Rhinopithecus bieti, NCBI Reference Sequence: XM_017868910.1; Rhinopithecus roxellana, NCBI Reference Sequence: XM_030935243.1; Colobus angolensis palliatus, NCBI Reference Sequence: XM_011928485.1; Piliocolobus tephrosceles, NCBI Reference Sequence: XM_023214151.2 Cebus capucinus imitator, NCBI Reference Sequence: XM_017503703.1; Callithrix jacchus, NCBI Reference Sequence: XM_002764268.4; Chlorocebus sabaeus, NCBI Reference Sequence: XM_007976629.1; Cercocebus atys, NCBI Reference Sequence: XM_012093252.1; Macaca nemestrina, NCBI Reference Sequence: XM_011769959.2

Primate CD1b, 21 Species
Homo sapiens, GenBank: AK303330.1; Papio anubis, NCBI Reference Sequence: XM_003892865.5; Theropithecus gelada, NCBI Reference Sequence: XM_025362870.1; Mandrillus leucophaeus NCBI Reference Sequence: XM_011978334.1; Cercocebus atys, NCBI Reference Sequence: XM_012093255.1; Chlorocebus sabaeus, NCBI Reference Sequence: XM_007976624.1; Macaca fascicularis, NCBI Reference Sequence:

XM_005541341.2; Macaca nemestrina, NCBI Reference Sequence: XM_011769969.2
Rhinopithecus roxellana, NCBI Reference Sequence: XM_010387232.2; Colobus angolensis palliatus, NCBI Reference Sequence: XM_011957139.1; Piliocolobus tephrosceles, NCBI Reference Sequence: XM_023214158.2; Nomascus leucogenys, NCBI Reference Sequence: XM_003258651.4; Pan troglodytes, NCBI Reference Sequence: XM_513909.5; Pongo abelii, NCBI Reference Sequence: XM_002809978.3; Gorilla gorilla gorilla, NCBI Reference Sequence: XM_004027024.3;
Pan paniscus, NCBI Reference Sequence: XM_003821008.2; Cebus capucinus imitator,
NCBI Reference Sequence: XM_017503705.1; Saimiri boliviensis boliviensis, NCBI Reference Sequence: XM_003937886.2; Aotus nancymaae, GenBank: AY605931.1; Callithrix jacchus, NCBI Reference Sequence: XM_002760142.3

Primate CD1c, 20 species
Homo sapiens, Reference Sequence: NM_001765.3; Papio anubis, NCBI Reference Sequence: XM_021925284.2; Theropithecus gelada, NCBI Reference Sequence: XM_025393420.1; Mandrillus leucophaeus, NCBI Reference Sequence: XM_011969349.1; Macaca mulatta, NCBI Reference Sequence: NM_001145533.1; Macaca fascicularis, NCBI Reference Sequence: XM_005595546.2, Macaca nemestrina, NCBI Reference Sequence: XM_011769962.2, Rhinopithecus bieti, NCBI Reference Sequence: XM_017868926.1; Rhinopithecus roxellana, NCBI Reference Sequence: XM_010381246.2; Piliocolobus tephrosceles, NCBI Reference Sequence: XM_023214153.2; Colobus angolensis palliates, NCBI Reference Sequence: XM_011928487.1; Nomascus leucogenys, NCBI Reference Sequence: XM_003258650.2; Pongo abelii, NCBI Reference Sequence: XM_002809979.3; Pan paniscus, NCBI Reference Sequence: XM_003821009.3; Gorilla gorilla gorilla, NCBI Reference Sequence: XM_019025012.2; Pan troglodytes, NCBI Reference Sequence: XM_513908.6; Cebus capucinus imitator, NCBI

Reference Sequence: XM_017503701.1; Cercocebus atys, NCBI Reference Sequence: XM_012093253.1; Callithrix jacchus, NCBI Reference Sequence: XM_035279155.1

Primate CD1d, 18 species

Homo sapiens, NCBI Reference Sequence: NM_001766.4; Aotus nancymaae, NCBI Reference Sequence: XM_012449750.2; Rhinopithecus roxellana, NCBI Reference Sequence: XM_030935246.1; Cercocebus atys, NCBI Reference Sequence: XM_012093246.1; Papio anubis, NCBI Reference Sequence: XM_017948205.3; Macaca fascicularis, NCBI Reference Sequence: XM_005541342.2; Macaca nemestrina, NCBI Reference Sequence: XM_011769953.1; Chlorocebus sabaeus, NCBI Reference Sequence: XM_007976632.1; Pongo abelii,

NCBI Reference Sequence: XM_024247050.1; Pan troglodytes, NCBI Reference Sequence: NM_001071804.1; Pan paniscus, NCBI Reference Sequence: XM_008974483.3 Macaca mulatta, NCBI Reference Sequence: NM_001033114.2; Gorilla gorilla gorilla, NCBI Reference Sequence: XM_019024988.2; Aotus nancymaae, NCBI Reference Sequence: XM_012449750.2; Saimiri boliviensis boliviensis, NCBI Reference Sequence: XM_010348509.1; Piliocolobus tephrosceles, NCBI Reference Sequence: XM_023214147.2;

Rhinopithecus bieti, NCBI Reference Sequence: XM_017868929.1; Colobus angolensis palliates, NCBI Reference Sequence: XM_011928483.1;

Primate CD1e, 20 species

Homo sapiens, NCBI Reference Sequence: NM_030893.4; Rhinopithecus roxellana, NCBI Reference Sequence: XM_030935233.1; Macaca nemestrina, NCBI Reference Sequence: XM_011769970.2; Macaca fascicularis, NCBI Reference Sequence: XM_015455159.1; Cercocebus atys, NCBI Reference Sequence: XM_012093256.1; Chlorocebus sabaeus, NCBI Reference

Sequence: XM_007976621.1; Theropithecus gelada, NCBI Reference Sequence: XM_025356228.1; Aotus nancymaae, NCBI Reference Sequence: XM_012449741.1; Pongo abelii, NCBI Reference Sequence: XM_003775532.3; Saimiri boliviensis boliviensis, NCBI Reference Sequence: XM_003937882.2; Cercocebus atys, NCBI Reference Sequence: XM_012093259.1; Gorilla gorilla gorilla,
NCBI Reference Sequence: XM_004027025.3; Pan troglodytes, NCBI Reference Sequence: XM_513910.6;
Pan paniscus, NCBI Reference Sequence: XM_003821003.4; Nomascus leucogenys, NCBI Reference Sequence: XM_003258652.4; Rhinopithecus bieti, NCBI Reference Sequence: XM_017868912.1; Piliocolobus tephrosceles, NCBI Reference Sequence: XM_023214159.1; Colobus angolensis palliatus, NCBI Reference Sequence: XM_011957140.1; Papio anubis, NCBI Reference Sequence: XM_021925303.2; Mandrillus leucophaeus, NCBI Reference Sequence: XM_011978335.1;

Figure S11. Proposing new paradigm for framing lipid-influenced adaptive evolution

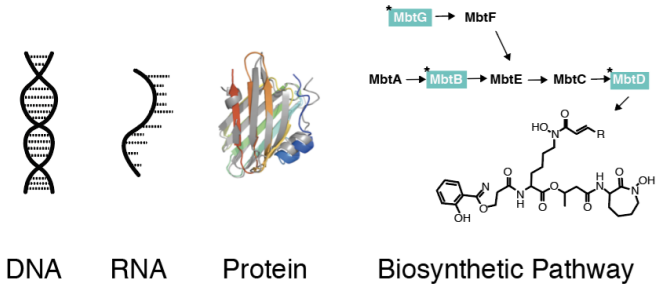


Figure S12. Ancestral reconstruction by DataMonkey SLAC for CD1a surface- associated sites that contact TCR

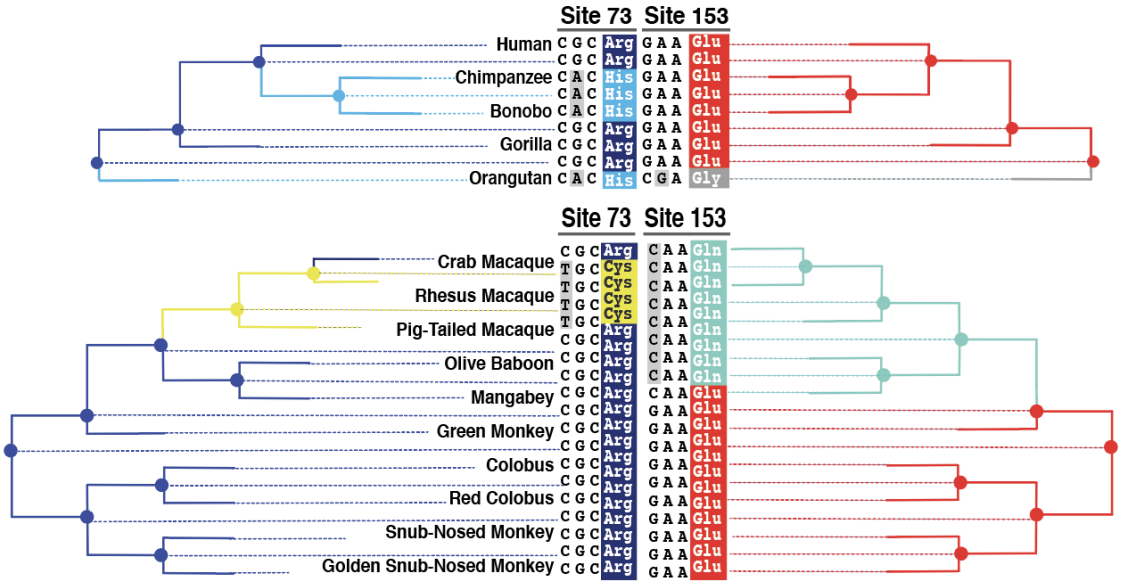
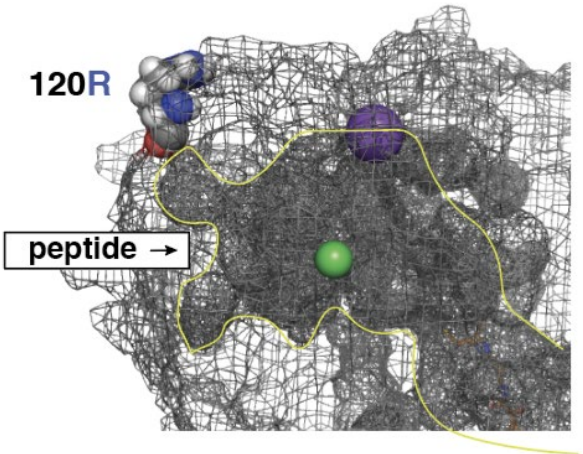


Figure S13. Genes in Mycobacterium tuberculosis involved in lipid metabolism have high dN/dS

Gene	PAML sites	p-value
MbtA	140T	0.004
MbtB	120R*	1.13537E-06
MbtC	16T 417T	2.11012E-06
MbtD	596G*	1.4357E-11
MbtG	417R 418S	2.52887E-08
MbtK	84T	0.0041
MbtL	none	none
MbtN	255S**	5.2204E-06



REFERENCES CITED

Aleru, Omoshola, and Matthew F. Barber. “Battlefronts of Evolutionary Conflict between Bacteria and Animal Hosts.” PLoS Pathogens, vol. 16, no. 9, Sept. 2020. PubMed Central, doi:10.1371/journal.ppat.1008797.

Adrian, Jonas, et al. “Adaptation to Host-Specific Bacterial Pathogens Drives Rapid Evolution of a Human Innate Immune Receptor.” Current Biology, vol. 29, no. 4, Feb. 2019, pp. 616-630.e5. ScienceDirect, doi:10.1016/j.cub.2019.01.058.

Anisimov, Andrey P., Luther E. Lindler, and Gerald B. Pier. 2004. “Intraspecific Diversity of Yersinia Pestis.” *Clinical Microbiology Reviews* 17(2): 434–64.

Barber MF, Elde NC. 2014. Escape from bacterial iron piracy through rapid evolution of transferrin. Science 346(6215):1362–1366.

Barral DC, Brenner MB. 2007. CD1 antigen presentation: how it works. Nat Rev Immunol. 7(12):929–941.

Birkinshaw RW, Pellicci DG, Cheng T-Y, Keller AN, Sandoval-Romero M, Gras S, de Jong A, Uldrich AP, Moody DB, Godfrey DI, et al. 2015. Ab T cell antigen receptor recognition of CD1a presenting self lipid ligands. Nat Immunol. 16(3):258–266.

Blumberg RS, Gerdes D, Chott A, Porcelli SA, Balk SP. 1995. Structure and function of the CD1 family of MHC-like cell surface proteins. Immunol Rev. 147(1):5–29.

Bao, L., et al. “Mapping of Genes for the Human C5a Receptor (C5AR), Human FMLP Receptor (FPR), and Two FMLP Receptor Homologue Orphan Receptors (FPRH1, FPRH2) to Chromosome 19.” Genomics, vol. 13, no. 2, June 1992, pp. 437–40. PubMed, doi:10.1016/0888-7543(92)90265-t.

Barber, Matthew F., and Nels C. Elde. “Escape from Bacterial Iron Piracy through Rapid Evolution of Transferrin.” Science (New York, N.Y.), vol. 346, no. 6215, Dec. 2014, pp. 1362–66. PubMed Central, doi:10.1126/science.1259329.

Brunette, Rebecca L., et al. “Extensive Evolutionary and Functional Diversity among Mammalian AIM2-like Receptors.” The Journal of Experimental Medicine, vol. 209, no. 11, Oct. 2012, pp. 1969–83. PubMed Central, doi:10.1084/jem.20121960.

Bufe, Bernd, et al. “Recognition of Bacterial Signal Peptides by Mammalian Formyl Peptide Receptors: A NEW MECHANISM FOR SENSING PATHOGENS*.” Journal of Biological Chemistry, vol. 290, no. 12, Mar. 2015, pp. 7369–87. ScienceDirect, doi:10.1074/jbc.M114.626747.

Burge, C., and S. Karlin. “Prediction of Complete Gene Structures in Human Genomic DNA.” Journal of Molecular Biology, vol. 268, no. 1, Apr. 1997, pp. 78–94. PubMed, doi:10.1006/jmbi.1997.0951.

Cala-De Paepe D, Layre E, Giacometti G, Garcia-Alles LF, Mori L, Hanau D, de Libero G, de la Salle H, Puzo G, Gilleron M. 2012. Deciphering the role of CD1e protein in mycobacterial phosphatidyl-myo- inositol mannosides (PIM) processing for presentation by CD1b to T lymphocytes. J Biol Chem. 287(37):31494–31502.

Castro CC, Luoma AM, Adams EJ. 2015. Coevolution of T-cell receptors with MHC and non-MHC ligands. Immunol Rev. 267(1):30–55.

Chancellor A, Gadola SD, Mansour S. 2018. The versatility of the CD1 lipid antigen presentation pathway. Immunology 154(2):196–203.

Chandler CE, Harberts EM, Pelletier MR, Thaipisuttikul I, Jones JW, Hajjar AM, Sahl JW, Goodlett DR, Pride AC, Rasko DA, et al. 2020. Early evolutionary loss of the lipid A modifying enzyme PagP resulting in innate immune evasion in Yersinia pestis. Proc Natl Acad Sci USA. 117(37):22984–22991.

Choby JE, Buechi HB, Farrand AJ, Skaar EP, Barber MF. 2018. Molecular basis for the evolution of species-specific hemoglobin capture by Staphylococcus aureus. MBio 9(6) e01524–18.

Caballero, Armando. Quantitative Genetics. Cambridge University Press, 2020. Cooper, Max D., and Matthew N. Alder. “The Evolution of Adaptive Immune Systems.” Cell, vol. 124, no. 4, Feb. 2006, pp. 815–22. ScienceDirect, doi:10.1016/j.cell.2006.02.001.

Crouser, Elliott D., et al. “Monocyte Activation by Necrotic Cells Is Promoted by Mitochondrial Proteins and Formyl Peptide Receptors.” Critical Care Medicine, vol. 37, no. 6, June 2009, pp. 2000–09. PubMed, doi:10.1097/CCM.0b013e3181a001ae.

Cussell, Peter J. G., et al. “The Formyl Peptide Receptor Agonist FPRa14 Induces Differentiation of Neuro2a Mouse Neuroblastoma Cells into Multiple Distinct Morphologies Which Can Be Specifically Inhibited with FPR Antagonists and FPR Knockdown Using SiRNA.” PLoS ONE, vol. 14, no. 6, Public Library of Science, June 2019, pp. e0217815–e0217815. go-gale- com.libproxy.uoregon.edu, doi:10.1371/journal.pone.0217815.

Daugherty, Matthew D., and Harmit S. Malik. “Rules of Engagement: Molecular Insights from Host-Virus Arms Races.” Annual Review of Genetics, vol. 46, 2012, pp. 677–700. PubMed, doi:10.1146/annurev-genet-110711-155522.

Dietschi, Quentin, et al. “Evolution of Immune Chemoreceptors into Sensors of the Outside World.” Proceedings of the National Academy of Sciences, National Academy of Sciences, June 2017. www.pnas.org, doi:10.1073/pnas.1704009114.

Dŭrr, Manuela C., et al. “Neutrophil Chemotaxis by Pathogen-Associated Molecular Patterns--Formylated Peptides Are Crucial but Not the Sole Neutrophil Attractants Produced by Staphylococcus Aureus.” Cellular Microbiology, vol. 8, no. 2, Feb. 2006, pp. 207–17. PubMed, doi:10.1111/j.1462- 5822.2005.00610.x.

Deng, Lu, et al. “Structural Insights into the Evolution of the Adaptive Immune System.” Annual Review of Biophysics, vol. 42, no. 1, Annual Reviews, May 2013, pp. 191–215. www-annualreviews-org.libproxy.uoregon.edu (Atypon), doi:10.1146/annurev-biophys-083012-130422.

Di Noia, Javier M., and Michael S. Neuberger. “Molecular Mechanisms of Antibody Somatic Hypermutation.” Annual Review of Biochemistry, vol. 76, no. 1, Annual Reviews, June 2007, pp. 1–22. www-annualreviews-org.libproxy.uoregon.edu (Atypon), doi:10.1146/annurev.biochem.76.061705.090740.

Dietschi, Quentin, et al. “Evolution of Immune Chemoreceptors into Sensors of the Outside World.” Proceedings of the National Academy of Sciences, National Academy of Sciences, June 2017. www-pnas-org.libproxy.uoregon.edu, doi:10.1073/pnas.1704009114.

Dittmar, Katharina, and David A. Liberles. Evolution after Gene Duplication. Wiley-Blackwell, 2010.

Danchin EGJ, Pontarotti P. 2004. Towards the reconstruction of the bilaterian ancestral pre-MHC region. Trends Genet. 20(12):587–591.
Dascher CC. 2007.Evolutionary biology of CD1. Curr Top Microbiol Immunol. 314:3–26.

Daugherty MD, Malik HS. 2012. Rules of engagement: molecular insights from host-virus arms races. Annu Rev Genet. 46(1):677–700.
de Jong A,Pen~a-Cruz V, Cheng T-Y, Clark RA, Van Rhijn I, Branch Moody D. 2010. CD1a-autoreactive T cells are a normal component of the human Ab T cell repertoire. Nat Immunol. 11(12):1102–1109.

Enard D, Cai L, Gwennap C, Petrov DA. 2016. Viruses are a dominant driver of protein adaptation in mammals. ELife 5:e12469.

Forrellad MA, Klepp LI, Gioffr□e A, Sabio y Garc□ıa J, Morbidoni HR, Santangelo MDLP,Cataldi AA, Bigi F. 2013. Virulence factors of the Mycobacterium tuberculosis complex. Virulence 4(1):3–66.

Frank SA. 2002. Immunology and evolution of infectious disease . Princeton (NJ): Princeton University Press.

Fu, Huamei, et al. “Ligand Recognition and Activation of Formyl Peptide Receptors in Neutrophils.” Journal of Leukocyte Biology, vol. 79, no. 2, 2006, pp. 247–56. Wiley Online Library, doi:https://doi.org/10.1189/jlb.0905498.

Elde NC, Child SJ, Geballe AP, Malik HS. 2009. Protein kinase R reveals an evolutionary model for defeating viral mimicry. Nature 457(7228):485–491.

Flajnik, Martin F., and Masanori Kasahara. “Comparative Genomics of the MHC: Glimpses into the Evolution of the Adaptive Immune System.” Immunity, vol. 15, no. 3, Elsevier, Sept. 2001, pp. 351–62. www.cell.com, doi:10.1016/S1074- 7613(01)00198-4.

Fu, Huamei, et al. “Ligand Recognition and Activation of Formyl Peptide Receptors in Neutrophils.” Journal of Leukocyte Biology, vol. 79, no. 2, 2006, pp. 247–56. Wiley Online Library, doi:https://doi.org/10.1189/jlb.0905498.

Garcia-Alles LF, Giacometti G, Versluis C, Maveyraud L, de Paepe D, Guiard J, Tranier S, Gilleron M, Prandi J, Hanau D, et al. 2011. Crystal structure of human CD1e reveals a groove suited for lipid- exchange processes. Proc Natl Acad Sci USA. 108(32):13230–13235.

Gathiaka S, Nanayakkara G, Boncher T, Acevedo O, Wyble J, Patel S, Patel A, Shane ME, Bonkowski B, Wieczorek J, et al. 2013. Design, development and evaluation of novel dual PPARd/PPARc agonists. Bioorg Med Chem Lett. 23(3):873–879.

Godfrey DI, Uldrich AP, McCluskey J, Rossjohn J, Branch Moody D. 2015. The burgeoning family of unconventional T cells. Nat Immunol. 16(11):1114–1123.

Golmogghaddam H, Arandi Abbas Ghaderi N, Doroudchi M. 2013. Polymorphism in exon 2 of CD1 genes in southwest of Iran. Iran J Public Health. 42(7):775–782.

Gong W, Liang Y, Wu X. 2018. The current status, challenges, and future developments of new tuberculosis vaccines. Hum Vaccines Immunother. 14(7):1697–1716.

Grimholt U. 2016. MHC and evolution in teleosts. Biology 5(1):6.

Han M, Hannick LI, DiBrino M, Robinson MA.1999. Polymorphism of human CD1 genes. Tissue Antigens 54(2):122–127.

Hughes A, Nei M. 1990.Evolutionary relationships of class II MHC genes in mammals. Mol Biol Evol . 7(December):491–514.

Harris, R. Alan, et al. “Unusual Sequence Characteristics of Human Chromosome 19 Are Conserved across 11 Nonhuman Primates.” BMC Evolutionary Biology, vol. 20, no. 1, 1, BioMed Central, Dec. 2020, pp. 1–12. link-springer- com.libproxy.uoregon.edu, doi:10.1186/s12862-020-1595-9.

Hughes, Austin L., and Helen Piontkivska. “Functional Diversification of the Toll- like Receptor Gene Family.” Immunogenetics, vol. 60, no. 5, May 2008, pp. 249–56. PubMed Central, doi:10.1007/s00251-008-0283-5.

Harris, R. Alan, et al. “Unusual Sequence Characteristics of Human Chromosome 19 Are Conserved across 11 Nonhuman Primates.” BMC Evolutionary Biology, vol. 20, no. 1, 1, BioMed Central, Dec. 2020, pp. 1–12. link-springer- com.libproxy.uoregon.edu, doi:10.1186/s12862-020-1595-9.

Hughes, Austin L., and Helen Piontkivska. “Functional Diversification of the Toll- like Receptor Gene Family.” Immunogenetics, vol. 60, no. 5, May 2008, pp. 249–56. PubMed Central, doi:10.1007/s00251-008-0283-5.

Hughes, Austin, and Masatoshi Nei. “Hughes AL, Nei M. Evolutionary Relationships of Class II MHC Genes in Mammals. Mol Biol Evol 7: 491.” Molecular Biology and Evolution, vol. 7, Dec. 1990, pp. 491–514.

Jaghoori MM, Bleijlevens B, Olabarriaga SD. 2016. 1001 Ways to run AutoDock Vina for virtual screening. J Comput Aided Mol Des. 30(3):237–249.

John, C. D., et al. “Formyl Peptide Receptors and the Regulation of ACTH Secretion: Targets for Annexin A1, Lipoxins, and Bacterial Peptides.” The FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology, vol. 21, no. 4, Apr. 2007, pp. 1037–46. PubMed Central, doi:10.1096/fj.06-7299com.

Hinnebusch, B. Joseph, Clayton O. Jarrett, and David M. Bland. 2017. “‘Fleaing’ the Plague: Adaptations of Yersinia Pestis to Its Insect Vector That Lead to Transmission.” *Annual review of microbiology* 71: 215–32.

Karlsson, Jennie, et al. “The FPR2-Specific Ligand MMK-1 Activates the Neutrophil NADPH-Oxidase, but Triggers No Unique Pathway for Opening of Plasma Membrane Calcium Channels.” Cell Calcium, vol. 45, no. 5, May 2009, pp. 431–38. DOI.org (Crossref), doi:10.1016/j.ceca.2009.02.002.

Karolchik, Donna, et al. “The UCSC Genome Browser.” Current Protocols in Bioinformatics / Editoral Board, Andreas D. Baxevanis ... [et Al.], vol. CHAPTER, Dec. 2009, p. Unit1.4. PubMed Central, doi:10.1002/0471250953.bi0104s28.

Khau, Thippadey, et al. “Annexin-1 Signals Mitogen-stimulated Breast Tumor Cell Proliferation by Activation of the Formyl Peptide Receptors (FPRs) 1 and 2.” The FASEB Journal, vol. 25, no. 2, WILEY, 2011, pp. 483–96. alliance- primo.com, doi:10.1096/fj.09-154096.

Kim, Sang Doo, et al. “Functional Expression of Formyl Peptide Receptor Family in Human NK Cells.” The Journal of Immunology, vol. 183, no. 9, American Association of Immunologists, Nov. 2009, pp. 5511–17. www-jimmunol- org.libproxy.uoregon.edu, doi:10.4049/jimmunol.0802986.

Koymans, Kirsten J., et al. “Staphylococcal Immune Evasion Proteins: Structure, Function, and Host Adaptation.” Staphylococcus Aureus: Microbiology, Pathology, Immunology, Therapy and Prophylaxis, edited by Fabio Bagnoli et al., Springer International Publishing, 2017, pp. 441–89. Springer Link, doi:10.1007/82_2015_5017.

Kretschmer, Dorothee, et al. “Human Formyl Peptide Receptor 2 (FPR2/ALX) Senses Highly Pathogenic Staphylococcus Aureus.” Cell Host & Microbe, vol. 7, no. 6, June 2010, pp. 463–73. PubMed Central, doi:10.1016/j.chom.2010.05.012.

Kumar M, Gouw M, Michael S, S□amano-S□anchez H, Pancsa R, Glavina J, Diakogianni A, Valverde JA, Bukirova D, C□aly□seva J, et al. 2019. ELM – the eukaryotic linear motif resource in 2020. Nucleic Acids Res. 48(D1):D296–306.

Kasahara, Masanori, and Martin F. Flajnik. “Origin and Evolution of the Adaptive Immune System: Genetic Events and Selective Pressures.” Nature Reviews. Genetics, vol. 11, no. 1, NATURE PUBLISHING GROUP, Nature Publishing Group, 2010, pp. 47–59. alliance-primo.com, doi:10.1038/nrg2703.

Kimura, Motoo. The Neutral Theory of Molecular Evolution. Cambridge University Press, 1983.

Klaver, Dominik, et al. “Peptides from Allergenic Lipocalins Bind to Formyl Peptide Receptor 3 in Human Dendritic Cells to Mediate TH2 Immunity.” The Journal of Allergy and Clinical Immunology, vol. 145, no. 2, Feb. 2020, pp. 654–65. PubMed Central, doi:10.1016/j.jaci.2019.07.008.

Kretschmer, Dorothee, et al. “Human Formyl Peptide Receptor 2 (FPR2/ALX) Senses Highly Pathogenic Staphylococcus Aureus.” Cell Host & Microbe, vol. 7, no. 6, June 2010, pp. 463–73. PubMed Central, doi:10.1016/j.chom.2010.05.012.

Levin, Tera C., and Harmit S. Malik. “Rapidly Evolving Toll-3/4 Genes Encode Male-Specific Toll-Like Receptors in Drosophila.” Molecular Biology and Evolution, vol. 34, no. 9, Sept. 2017, pp. 2307–23. PubMed, doi:10.1093/molbev/msx168.

Liberles, Stephen D., et al. “Formyl Peptide Receptors Are Candidate Chemosensory Receptors in the Vomeronasal Organ.” Proceedings of the National Academy of Sciences, National Academy of Sciences, June 2009. www-pnas- org.libproxy.uoregon.edu, doi:10.1073/pnas.0904464106.

Lee R. V D,Wiel L, van Dam TJP, Huynen MA. 2017. Genome-scale detection of positive selection in nine primates predicts human-virus evolutionary conflicts. Nucleic Acids Res. 45(18):10634–10648.

Ly D, Moody DB. 2014. The CD1 size problem: lipid antigens, ligands, and scaffolds. Cell Mol Life Sci. 71(16):3069–3079.

Le, Yingying, et al. “Formyl-Peptide Receptors Revisited.” Trends in Immunology, vol. 23, no. 11, Nov. 2002, pp. 541–48. www.cell.com, doi:10.1016/S1471- 4906(02)02316-5.

Leslie, Jack, et al. “FPR-1 Is an Important Regulator of Neutrophil Recruitment and a Tissue-Specific Driver of Pulmonary Fibrosis.” JCI Insight, vol. 5, no. 4. PubMed Central, doi:10.1172/jci.insight.125937. Accessed 14 Mar. 2021.

Levin, Tera C., and Harmit S. Malik. “Rapidly Evolving Toll-3/4 Genes Encode Male- Specific Toll-Like Receptors in Drosophila.” Molecular Biology and Evolution, vol. 34, no. 9, Sept. 2017, pp. 2307–23. PubMed, doi:10.1093/molbev/msx168.

McLaughlin, Richard N., and Harmit S. Malik. “Genetic Conflicts: The Usual Suspects and Beyond.” The Journal of Experimental Biology, vol. 220, no. 1, Jan. 2017, pp. 6–17. PubMed Central, doi:10.1242/jeb.148148.

Muto, Yoshinori, et al. “Adaptive Evolution of Formyl Peptide Receptors in Mammals.” Journal of Molecular Evolution, vol. 80, no. 2, Feb. 2015, pp. 130– 41. PubMed, doi:10.1007/s00239-015-9666-z.

Manlik O, Kru¨tzen M, Kopps AM, Mann J, Bejder L, Allen SJ, Fre`re C, Connor RC, Sherwin WB. 2019. Is MHC diversity a better marker for conservation than neutral genetic diversity? A case study of two contrasting dolphin populations. Ecol Evol. 9(12):6986–6998.

Mizumoto N, Takashima A. 2004. CD1a and Langerin: acting as more than Langerhans cell markers. J Clin Invest. 113(5):658–660.

Mori L, Lepore M, De Libero G. 2016. The immunology of CD1- and MR1-restricted T cells. Annu Rev Immunol. 34(1):479–510.

Mura C, McCrimmon CM, Vertrees J, Sawaya MR. 2010.An introduction to biomolecular graphics. PLoS Comput Biol. 6(8):e1000918.

Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, Scheffler K. 2013. FUBAR: a fast, unconstrained Bayesian approx-imation for inferring selection. Mol Biol Evol. 30(5):1196–1205.

Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. 2012. Detecting individual sites subject to episodic diversi- fying selection. PLoS Genet. 8(7):e1002764.

Maldonado, Manuel. “Choanoflagellates, Choanocytes, and Animal Multicellularity.” Invertebrate Biology, vol. 123, no. 1, May 2005, pp. 1–22. DOI.org (Crossref), doi:10.1111/j.1744-7410.2004.tb00138.x.

Mills, John S. “Peptides Derived from HIV-1, HIV-2, Ebola Virus, SARS Coronavirus and Coronavirus 229E Exhibit High Affinity Binding to the Formyl Peptide Receptor.” Biochimica et Biophysica Acta. Molecular Basis of Disease, vol. 1762, no. 7, July 2006, pp. 693–703. PubMed Central, doi:10.1016/j.bbadis.2006.05.008.

Miyata, T., et al. “Nucleotide Sequence Divergence and Functional Constraint in MRNA Evolution.” Proceedings of the National Academy of Sciences, vol. 77, no. 12, Dec. 1980, pp. 7328–32. DOI.org (Crossref), doi:10.1073/pnas.77.12.7328.

Müller, Werner E. G., and Isabel M. Müller. “Origin of the Metazoan Immune System: Identification of the Molecules and Their Functions in Sponges.” Integrative and Comparative Biology, vol. 43, no. 2, Oxford University Press, 2003, pp. 281–92.

Muto, Yoshinori, et al. “Adaptive Evolution of Formyl Peptide Receptors in Mammals.” Journal of Molecular Evolution, vol. 80, no. 2, Feb. 2015, pp. 130–41. PubMed, doi:10.1007/s00239-015-9666-z.

Nei M, Rooney AP. 2005.Concerted and birth-and-death evolution of multigene families. Annu Rev Genet. 39(1):121–152.

Ohno, Susumu. Evolution by Gene Duplication. Springer-Verlag, 1970.
Pegueroles, Cinta, et al. “Accelerated Evolution after Gene Duplication: A Time- Dependent Process Affecting Just One Copy.” Molecular Biology and Evolution, vol. 30, no. 8, Aug. 2013, pp. 1830–42. Silverchair, doi:10.1093/molbev/mst083.

Önnheim, Karin, et al. “A Novel Receptor Cross-Talk between the ATP Receptor P2Y2 and Formyl Peptide Receptors Reactivates Desensitized Neutrophils to Produce Superoxide.” Experimental Cell Research, vol. 323, no. 1, Apr. 2014, pp. 209–17. ScienceDirect, doi:10.1016/j.yexcr.2014.01.023.

Osei-Owusu, Patrick, et al. “FPR1 Is the Plague Receptor on Host Immune Cells.” Nature, vol. 574, no. 7776, 2019, pp. 57–62. PubMed, doi:10.1038/s41586-019- 1570-z.

Paterson, Nicole M., et al. “Diversification of CD1 Molecules Shapes Lipid Antigen Selectivity.” Molecular Biology and Evolution, no. msab022, Feb. 2021. Silverchair, doi:10.1093/molbev/msab022.

Pike, MC, et al. “Development of Specific Receptors for N-Formylated Chemotactic Peptides in a Human Monocyte Cell Line Stimulated with Lymphokines.” The Journal of Experimental Medicine, vol. 152, no. 1, July 1980, pp. 31–40.

Prat, Cristina, et al. “A Homolog of Formyl Peptide Receptor-like 1 (FPRL1) Inhibitor from Staphylococcus Aureus (FPRL1 Inhibitory Protein) That Inhibits FPRL1 and FPR.” Journal of Immunology (Baltimore, Md.: 1950), vol. 183, no. 10, Nov. 2009, pp. 6569–78. PubMed, doi:10.4049/jimmunol.0801523.

Paget C, Deng S, Soulard D, Priestman DA, Speca S, von Gerichten J, Speak AO, Saroha A, Pewzner-Jung Y, Futerman AH, et al. 2019. TLR9-mediated dendritic cell activation uncovers mammalian gan- glioside species with specific ceramide backbones that activate in- variant natural killer T cells. PLoS Biol. 17(3):e3000169.

Pereira CS, Macedo MF. 2016. CD1-restricted T cells at the crossroad of innate and adaptive immunity. J Immunol Res. 2016:1–11.

Perelman P, Johnson WE, Roos C, Seu□anez HN, Horvath JE, Moreira MAM, Kessing B, Pontius J, Roelke M, Rumpler Y, et al. 2011. A molecular phylogeny of living primates. PLoS Genet. 7(3):e1001342.

Perry, R. D., and J. D. Fetherston. 1997. “Yersinia Pestis--Etiologic Agent of Plague.” *Clinical Microbiology Reviews* 10(1): 35–66.

Pond SLK, Frost SDW. 2005. Datamonkey: rapid detection of selective pres- sure on individual sites of codon alignments. Bioinformatics 21(10):2531–2533.

Rogers SL, Kaufman J. 2016. Location, location, location: the evolutionary history of CD1 genes and the NKR-P1/ligand systems. Immunogenetics 68(8):499–513.

Reams, Andrew B., and John R. Roth. “Mechanisms of Gene Duplication and Amplification.” Cold Spring Harbor Perspectives in Biology, vol. 7, no. 2, Feb. 2015, p. a016592. cshperspectives.cshlp.org, doi:10.1101/cshperspect.a016592.

Rosenstiel, Philip, et al. “Evolution and Function of Innate Immune Receptors - Insights from Marine Invertebrates.” Journal of Innate Immunity, vol. 1, no. 4, Karger AG, Aug. 2009, pp. 291–300. EBSCOhost, doi:10.1159/000211193.

Schepetkin, I. A., et al. “Development of Small Molecule Non-Peptide Formyl Peptide Receptor (FPR) Ligands and Molecular Modeling of Their Recognition.” Current Medicinal Chemistry, vol. 21, no. 13, 2014, pp. 1478–504. PubMed Central, doi:10.2174/09298673216661312180955 21.

Schöneberg, Torsten, et al. “Learning from the Past: Evolution of GPCR Functions.” Trends in Pharmacological Sciences, vol. 28, no. 3, Mar. 2007, pp. 117–21. ScienceDirect, doi:10.1016/j.tips.2007.01.001.

Sogabe, Shunsuke, et al. “Pluripotency and the Origin of Animal Multicellularity.” Nature, vol. 570, no. 7762, Nature Publishing Group, June 2019, pp. 519–519. go- gale-com.libproxy.uoregon.edu, doi:10.1038/s41586-019-1290-4.

Suckale, Jakob, et al. “Evolution of Innate Immune Systems*.” Biochemistry and Molecular Biology Education, vol. 33, no. 3, 2005, pp. 177–83. Wiley Online Library, doi:https://doi.org/10.1002/bmb.2005.494033032466.

Sutton, Jolene T., et al. “Disentangling the Roles of Natural Selection and Genetic Drift in Shaping Variation at MHC Immunity Genes.” Molecular Ecology, vol. 20, no. 21, 2011, pp. 4408–20. Wiley Online Library, doi:10.1111/j.1365- 294X.2011.05292.x.

Salomonsen J, Sørensen MR, Marston DA, Rogers SL, Collen T, van Hateren A, Smith AL, Richard K, Beal K, Skjødt J, Kaufman ND. 2005. Two CD1 genes map to the chicken MHC, indicating that CD1 genes are ancient and likely to have been present in the pri- mordial MHC. Proc Natl Acad Sci U S A . 102:8668–8673.

Sawyer SL, Wu LI, Emerman M, Malik HS.2005. Positive selection of primate TRIM5aidentifies a critical species-specific retroviral restric- tion domain. Proc Natl Acad Sci U S A. 102:2832–2837.

Schepetkin, I. A., et al. “Development of Small Molecule Non-Peptide Formyl Peptide Receptor (FPR) Ligands and Molecular Modeling of Their Recognition.” Current Medicinal Chemistry, vol. 21, no. 13, 2014, pp. 1478– 504. PubMed Central, doi:10.2174/09298673216661312180955 21.

Schiffmann, E., et al. “N-Formylmethionyl Peptides as Chemoattractants for Leucocytes.” Proceedings of the National Academy of Sciences of the United States of America, vol. 72, no. 3, Mar. 1975, pp. 1059–62.

Shimizu, Nobuaki, et al. “A Formylpeptide Receptor, FPRL1, Acts as an Efficient Coreceptor for Primary Isolates of Human Immunodeficiency Virus.” Retrovirology, vol. 5, no. 1, June 2008, p. 52. BioMed Central, doi:10.1186/1742-4690-5-52.

Schito M, Migliori GB, Fletcher HA, McNerney R, Centis R, D’Ambrosio L, Bates M, Kibiki G, Kapata N, Corrah T, et al. 2015. Perspectives on advances in tuberculosis diagnostics, drugs, and vaccines. Clin Infect Dis. 61(Suppl 3):S102–S118.

Takahashi T, Suzuki T. 2012. Role of sulfatide in normal and pathological cells and tissues. J Lipid Res. 53(8):1437–1450.

Takahata N, Nei M. 1990.Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histo-compatibility complex loci. Genetics 124(4):967–978.

Thierry-Mieg D, Thierry-Mieg J. 2006. AceView: a comprehensive CDNA-supported gene and transcripts annotation. Genome Biol. 7(Suppl 1):S12.
Trott O, Olson AJ. 2009. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimi- zation and multithreading. J Comput Chem. 31(2):61.

Teufel, Ashley I., et al. “The Many Nuanced Evolutionary Consequences of Duplicated Genes.” Molecular Biology and Evolution, edited by Mary O’Connell, vol. 36, no. 2, Feb. 2019, pp. 304–14. DOI.org (Crossref), doi:10.1093/molbev/msy210. Van Valen, Leigh. A New Evolutionary Law. 1973.

Van Valen L. 1973. A new evolutionary law. Evol. Theory 1: 1-30.
Yang J, Zhang Y. 2015. Protein structure and function prediction using I-TASSER. Curr Protoc Bioinformatics 52(December):5.8.1–5.815. Yang Z. 2007.PAML 4: phylogenetic analysis by maximum likelihood. Mol

Thierry-Mieg, Danielle, and Jean Thierry-Mieg. “AceView: A Comprehensive CDNA- Supported Gene and Transcripts Annotation.” Genome Biology, vol. 7, no. 1, Aug. 2006, p. S12. BioMed Central, doi:10.1186/gb-2006-7-s1-s12.

van der Lee, Robin, et al. “Genome-Scale Detection of Positive Selection in Nine Primates Predicts Human-Virus Evolutionary Conflicts.” Nucleic Acids Research, vol. 45, no. 18, Oct. 2017, pp. 10634–48. PubMed, doi:10.1093/nar/gkx704.

Yang, Huan, et al. “Relating Gene Expression Evolution with CpG Content Changes.” BMC Genomics, vol. 15, no. 1, Aug. 2014, p. 693. BioMed Central, doi:10.1186/1471-2164-15-693.

Yang, Hui, and Peng Shi. “Molecular and Evolutionary Analyses of Formyl Peptide Receptors Suggest the Absence of VNO-Specific FPRs in Primates.” Journal of Genetics and Genomics, vol. 37, no. 12, Dec. 2010, pp. 771–78. DOI.org (Crossref), doi:10.1016/S1673-8527(09)60094-1.

Yang, Z. “PAML 4: Phylogenetic Analysis by Maximum Likelihood.” Molecular Biology and Evolution, vol. 24, no. 8, Apr. 2007, pp. 1586–91. DOI.org (Crossref), doi:10.1093/molbev/msm088.

Yao, Xiao-Hong, et al. “Production of Angiogenic Factors by Human Glioblastoma Cells Following Activation of the G-Protein Coupled Formylpeptide Receptor FPR.” Journal of Neuro-Oncology, vol. 86, no. 1, Jan. 2008, pp. 47–53. Springer Link, doi:10.1007/s11060-007-9443-y.

Wiens, Matthias, et al. “Toll-Like Receptors Are Part of the Innate Immune Defense System of Sponges (Demospongiae: Porifera).” Molecular Biology and Evolution, vol. 24, no. 3, Mar. 2007, pp. 792–804. Silverchair, doi:10.1093/molbev/msl208.

Yang, Hui, and Peng Shi. “Molecular and Evolutionary Analyses of Formyl Peptide Receptors Suggest the Absence of VNO-Specific FPRs in Primates.” Journal of Genetics and Genomics, vol. 37, no. 12, Dec. 2010, pp. 771–78. DOI.org (Crossref), doi:10.1016/S1673-8527(09)60094-1.

Zajonc, Dirk M. “The CD1 Family: Serving Lipid Antigens to T Cells since the Mesozoic Era.” Immunogenetics, vol. 68, no. 8, Aug. 2016, pp. 561–76. PubMed Central, doi:10.1007/s00251-016-0931-0.

Zajonc DM, Crispin MDM, Bowden TA, Young DC, Cheng T-Y, Hu J, Costello CE, Rudd PM, Dwek RA, Miller MJ, et al. 2005. Molecular mechanism of lipopeptide presentation by CD1a. Immunity 22(2):209–219.

Zajonc DM, Elsliger MA, Teyton L, Wilson IA. 2003. Crystal structure of CD1a in complex with a sulfatide self antigen at a resolution of 2.15 A. Nat Immunol. 4(8):808–815.

Zajonc DM, Flajnik MF.2016. CD1, MR1, NKT, and MAIT: evolution and origins of non-peptidic antigen recognition by T lymphocytes. Immunogenetics 68(8):489–490.

Zigmond, S. H. “Consequences of Chemotactic Peptide Receptor Modulation for Leukocyte Orientation.” Journal of Cell Biology, vol. 88, no. 3, Mar. 1981, pp. 644–47. Silverchair, doi:10.1083/jcb.88.3.644.