January 2014

Evidence For Selection Acting On Immune-Related Iron Transport Genes In Wild Rodents

Dylan Adam Duchen
Yale University, dylan.duchen@yale.edu

Evidence for Selection Acting on Immune-Related Iron Transport Genes in Wild Rodents

Thesis submitted in accordance with the requirements of the

Yale School of Public Health for the degree of Master of Public Health

by

Dylan Adam Duchen

2014

Abstract

Every organism requires certain resources that they cannot endogenously produce; one such dietary component, iron, is essential for many basic cellular functions and the host-immune response. By studying patterns of genetic diversity within the genes involved in immune-related iron transport we can achieve a greater understanding of the evolutionary mechanisms which drive these processes, including those underlying host-pathogen interactions.

This thesis describes further evidence of patterns indicative of positive selection within apical regions of the transferrin receptor complex (*Tfr1*), sites known to be involved in an evolutionary arms race with pathogens, within two rodent species endemic to Europe and Asia, the bank vole, *Myodes glareolus,* and the field vole, *Microtus agrestis*. Other immune-related iron transport genes, including lipocalin-2 (*Lcn2*), haptoglobin, (*Hp*), solute carrier family 11 member 1 (*Slc11a1*), lactotransferrin (*Ltf*), lipocalin-12 (*Lcn12*), and mitochondrial ferritin (*Ftmt*) were sequenced and analyzed in the hopes of identifying regions under positive selection.

This is the first study to date describing the genetic variation and site-specific non-neutral selection pressures for several immune-related iron transport proteins in natural populations of bank vole and field vole. We identified 7 SNPs in *Lcn2* across both species and identified two sites displaying patterns indicative of positive selection within the functional loci of the lipocalin-2 siderophore binding calyx. This work contributes to the growing body of literature describing the diversity and the non-neutral selection pressures acting on immune-related iron transport proteins in natural populations.

Acknowledgments

First and foremost, I would like to thank the incredible professors at Yale University and at the Yale School of Public Health for reinforcing my love of public health, scientific research, and my goals of becoming a passionate generalist. To my first thesis reader, Dr. James Childs, your enthusiasm for infectious disease research has successfully been passed on- thank you for your time and effort inside the classroom and outside while advising me on this thesis. To Professor Stephen Paterson at the University of Liverpool, your dedication, support, and guidance in this research project has meant an incredible amount to me. Dr. Turner, thank you for your time and patience throughout the past year, you most of all had to deal with my "bollocks". Professor Stephen Stearns, Durland Fish, Paul Turner, and guests of the Studies in Evolutionary Medicine courses - thank you for providing the inspiration and forum without which this project would not have been possible. Dr. Fish and the Yale Institute for Biospheric Studies, thank you especially for the research grant which funded this work.

I am also incredibly fortunate to have met so many wonderful people here at Yale. I thank my YSPH family for keeping me sane the past two years and pretending to be as excited about iron transport proteins as I am.

To my parents, thank you for your constant support throughout my education and always. To my brother and sister, I hope you find something to be as passionate about as I have, and you one day find yourself in an environment like Yale in which I know you will thrive.

Introduction

Infectious Disease and Evolution

Forces of natural selection play a fundamental role in creating or maintaining the genetic differences between and within species. Such selection is fueled by co-evolutionary dynamics, such as the relationship between hosts and infectious diseases first proposed by Haldane[4]. Since Haldane's initial description in 1949, we have come to understand in incredible detail how this relationship has shaped human evolution. The host-pathogen relationship's effect on driving genetic diversity has been established across many species. Some of the strongest evidence for this lies in the observation that immune genes and genes encoding proteins which lie at the host-pathogen interface undergo greater rates of adaptive evolution compared with the genome average[5-7].

When individuals are best suited to thrive and reproduce within certain environments due to variation in their genetic makeup, the gene variants conferring those advantages will increase in prevalence within that population over time. Such genetic variants are under positive selection. The identification of genes or proteins that have been under positive selection throughout our evolutionary past can reveal functionally useful information about human biology, such as disease susceptibility or resistance.

By studying patterns of genetic variation we can achieve a greater understanding of the evolutionary mechanisms that drive such diversity, including underlying host-pathogen interactions. Exploring the genetics of wild populations rather than laboratory-

bred populations is sometimes best suited for drawing such inferences, as using wild populations allow for the inclusion of ecologically relevant selective forces such as natural pathogen exposures, the effects of seasonality, resource variability, and other variables involved in the development of a mature immune system.

There are many factors that contribute to whether a microbe can successfully invade a host, propagate, and become infectious. Not all microbes are pathogenic; many microbes have a mutually beneficial relationship with humans. Indeed, there are 10 times as many bacteria living within and on the human gut as there are human cells[8]. This intimate relationship between humans and the microbial world can be disrupted by numerous factors, including diet, antibiotic use, and many other traits.

This commensal relationship is also not limited to humans. Each animal, plant, and bacteria requires certain resources that they cannot endogenously create. One such dietary component, iron, is vital for many basic cellular functions.

The Importance of Iron

Incredibly abundant in nature, iron is by mass the commonest element in the earth's crust. It is therefore not coincidental that this nutrient is a requirement for most organisms. Iron utilization is especially important during oxygen transport, electron transport, and DNA synthesis in all types of biological systems[9]. Indeed, its role as an electron donor and acceptor, coupled with its redox properties, allow it to facilitate oxygen-based chemistry[10].

Iron exists within living organisms in two forms; ferrous iron (Fe^{2+}) which is soluble in anoxic and neutral pH environments, and ferric iron (Fe^{3+}) which is more

soluble in acidic environments such as the stomach[11]. Various proteins have evolved to transport ferrous iron across cellular membranes, which are evolutionarily conserved across the six kingdoms. Bacterially produced siderophores and siderophore-transport proteins transport ferric iron. In mammals transferrin, endogenously produced siderophores, and various other iron transport proteins are involved with ferric iron transport. These proteins are essential, as unbound iron can be toxic when in circulation. This is due to the electron transfer properties of iron which can catalyze the conversion of hydrogen peroxide into free radicals that damage cellular architecture and can themselves act as pro-inflammatory signals. Therefore, iron levels are tightly regulated. There are many clinical pathologies associated with inappropriate levels of free iron which are avoided via the formation of iron-protein complexes. These complexes play integral parts to iron absorption, metabolism, and organismal iron homeostasis.

Briefly, in humans the process involves acquiring iron from the diet, whereby iron is transported across the duodenal lining into developing enterocytes. Membrane bound iron-import proteins such as the divalent metal transporter 1 protein (DMT1) deliver reduced Fe^{3+} and Fe^{2+} to the iron exporter protein ferroportin-1 (SLC40A1) which mediates the delivery of Fe^{2+} to circulating proteins such as transferrin (TF) once the Fe^{2+} is converted to Fe^{3+} by copper-containing oxidases. Circulating transferrin then delivers its Fe^{3+} cargo by forming complexes with membrane-bound transferrin-receptor proteins (TfR) which are then internalized forming low-pH endosomes containing a membrane-bound reductase and DMT1. The endosomal environment subsequently allows for Fe^{3+} reduction to Fe^{2+} and DMT1-mediated transport of ferrous iron into the

cellular cytosol[11]. Most transferrin-bound iron is destined for hemoglobin synthesis in erythroid cells, where roughly 60% of the available iron within the host is harbored[9]. The majority of the iron within the body is retained, recycled from the hemoglobin of aged red blood cells by the reticuloendothelial system[12]. However, some iron is lost via bleeding, sweating, shedding of skin epithelial cells, or excessive secretion of the mucous membranes and must be replenished via diet. When serum iron levels drop too low the iron responsive/regulatory element binding protein (IRE-BP) upregulates transferrin receptor production via another regulator protein, hepcidin (HAMP)[13].

These processes are far more complex and the intricacies of iron absorption, distribution, metabolism, and excretion can be found elsewhere in the literature[14,15]. These processes are relevant, however, as they play a role within the hosts immune response.

Iron levels are vital for the development of a robust immune response as iron is essential for T-cell clonal expansion and the generation of antimicrobial oxygen radicals[16]. A recent large retrospective cohort study found that patients who were administered iron-supplementation during hemodialysis had far greater rates of bacterial infections[17]. Other studies have similarly identified an increased risk of infection and pathogen-mediated disease morbidity in the setting of iron supplementation[18].

Testament to this notion is the phenomenon of iron-withholding, which many vertebrates undergo while fighting an infection. While most iron is reserved within hemoglobin or tightly bound inside ferritin, there remains a trace amount of bioavailable unbound iron within intra- and inter-cellular spaces. Both invading bacteria and viruses

require iron for DNA synthesis and other processes with which to thrive. Thus, pathogens sequester iron either from these intra- and inter-cellular spaces or directly from host proteins. In response, the host downregulates the expression of these iron-transport proteins and decreases the uptake of dietary iron. Chronic infections can even lead to a state of prolonged iron-withholding, termed anemia of chronic disease[19]. Additionally, many iron transport proteins have been shown to have direct antimicrobial properties. Therefore, TfR1 and other iron-transport proteins involved in iron-withholding play an important role in the innate-immune response and should be characterized as immune-related iron transport proteins.

For these reasons viruses and bacteria have evolved complex methods to sequester iron from the host. Indeed, many viruses have exploited steps within iron metabolism pathways to mediate cellular entry. For example, many microbial pathogens target the membrane-bound Transferrin receptor binding protein 1 (TfR1) for cellular entry or iron sequestration. This is a particularly ideal target for invasive pathogens as *TfR1* expression is directly related to host iron levels, it is nearly ubiquitously expressed, and it is involved in the endocytosis pathway for the cellular internalization of iron making it a prime target for viruses to invade host cells.

Transferrin Receptor as a Viral Entry Mechanism

Indeed, of the five New World Arenaviruses known to cause hemorrhagic fevers in humans, four implement TfR1 as a means for cellular entry[4,20,21]. TfR1 has also been identified as the initial binding receptor for other important pathogens such as the

mouse mammary tumor virus[22], panleukopenia feline parvoviruses, and canine parvoviruses[23].

These zoonotic New World Arenaviruses are of particular concern to humans, as they cause severe disease in humans who are dead-end hosts (mortality rates average ~20%"). Three of these viruses: Junin, Machupo, and Guanarito, along with mouse mammary tumor virus, have been shown to play a role in driving positive selection in apical regions of TfR1.

While TfR1 domains required for normal cellular function were conserved, apical TfR1 residues not involved in iron transport were identified as the targeted sites for binding and cellular entry. These same sites were found to be under positive selection (Figure 2[11]). While the specific amino acid residues within TfR1 that mediate New World Arenavirus binding seem to vary experimentally[2,24], the residues identified to be under positive selection by this study corroborate Demogines *et al*.'s findings.

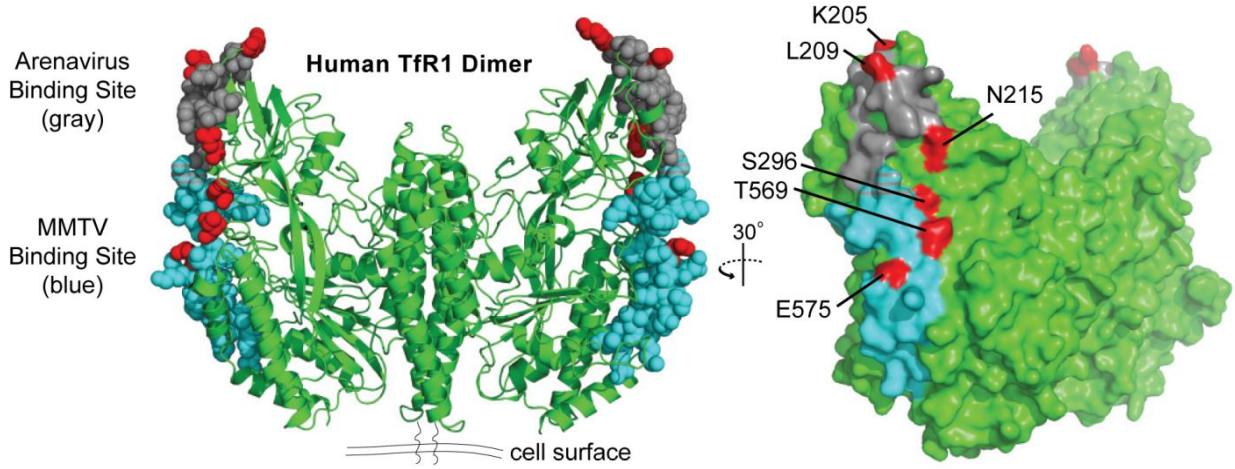

Figure 1: Human TFR1 is shown from two angles. Highlighted regions correspond to the indicated viral binding sites. Individual residues shown in red were determined to be under positive selection due to an evolutionary arms race, from Demogines, *et al*.[2]

Lipocalin-2, Bacteria, and the Fight Over Iron

The targeting of iron transport proteins can occur throughout the course of an infection and not solely for cellular entry. Immune-related iron transport protein targeting is also not limited to viruses. Bacteria secrete proteins, termed siderophores, which enable bacterial pathogens to chelate the trace amounts of free ferric iron from within the intra-host environment and directly from host iron transport proteins similar to TfR1. A wide range of bacteria produces siderophores, which have an incredibly high affinity for ferric iron. Indeed, over 500 unique siderophores have been identified and 270 have been well-characterized chemically[25].

As a response to the production of bacterial siderophores, many vertebrate hosts, including humans, secrete the protein Siderocalin, encoded by Lipocalin-2 (*LCN2*), a siderophore-binding lipocalin. This siderocalin-siderophore binding inhibits the transport of iron-siderophore complexes into bacterial cells[26]. While only a few siderocalin homologs have been identified, they are able to bind hundreds of diverse siderophores. Furthermore, these siderocalins have important biological functions apart from their role in the innate immune defense against iron-sequestering pathogens, such as during embryonic development via the use of host-derived siderophores which have been identified in the literature[27].

Siderocalin appears to have evolved to counter the sequestration of host iron reserves by bacteria[27]. However, the only evidence for this claim includes an *in vitro* study of human and mouse primary macrophages which were analyzed for the conservation and divergence of gene transcription, where *LCN2* was one of the genes

with significant evolutionary divergence[28]. The authors postulated such a difference might be due to an evolutionary arms race between siderocalin and bacterial siderophores. However, a cross-species analysis for evidence of positive selection hinting at such an arms race similar to Demogines et al's *TfR1* analysis of *Lcn2* had yet to be conducted.

The Aims of This Study

This study aimed to provide additional evidence of positive selection in the apical regions of *TfR1* within two rodent species endemic to Europe and Asia whose genomes have recently been sequenced, the bank vole *Myodes glareolus* (Figure 2a) and the field vole *Microtus agrestis* (Figure 2b). Other immune-related iron transport genes, including siderocalin (*Lcn2*), were sequenced and analyzed in both field vole and bank vole specimens in the hopes of identifying regions undergoing positive selection (Table 1).

Figure 2a Wild bank vole[1]

Figure 2b: Wild field vole[3]

Methods

Gene Selection

The bank vole and field vole genes sequenced include haptoglobin (*Hp*), polymorphisms of which have been associated with various pathologies including atherosclerosis, tuberculosis infections, inflammatory and autoimmune disorders[29], transplant rejection[30], and reduced mortality in sepsis[31]; solute carrier family 11 member

1 (*Slc11a1*), also known as the natural resistance-associated macrophage protein 1, a key H+/metal transmembrane iron importer which is associated with resistance to various intracellular pathogens[10,32]; lactotransferrin (*Ltf*) which in addition to its anticancer and anti-inflammatory properties has been shown to modulate host immune activity against many microorganisms[33,34]; and sites immediately flanking the apical residues of interest identified to be under postitve selection by Demogines, *et al*.[2] for transferrin receptor protein 1 (*TfR1*). Lipocalin-12 (*Lcn12*), a relatively unstudied lipocalin ~29% identical in DNA sequence with a predicted identical binding calyx to *Lcn2*[35], and mitochondrial ferritin (*Ftmt*), a nuclear-transcribed ferritin-like protein expressed in mitochondria associated with various inflammatory neurological disorders were sequenced for field vole only.

Bank Vole and Field Vole DNA Extraction and Sequencing

Using the Qiagen genomic-tip 20/G kit genomic DNA was extracted from a single field vole (Kielder Forest, UK) and bank vole (Jyväskylä, Finland) individual. Integrity and size of DNA were checked on an agarose gel, purity assessed using a NanoDrop 1000 spectrophotometer (Thermo Scientific) and concentration measured with a Qubit fluorometer (Life Technologies). DNA was then submitted to the Centre for Genomic Research (CGR) at the University of Liverpool and sequenced on the Illumina HiSeq2000 platform. The MAKER pipeline[36] was used to align all known *Mus musculus* protein sequences to the two draft genomes, in order to identify and annotate orthologous genes within the two vole species. Candidate field vole and bank vole genes for this study were identified by running a BLASTP search of known *Mus musculus* sequences against the set of MAKER-predicted vole proteins.

Primer Design and Amplification

Primer sequences were created from the immediate non-coding DNA regions flanking the exons of interest identified via BLASTP searches against the MAKER-predicted vole protein sequences. Primer design was conducted using the online primer design tool Primer3[37] and ordered from Eurofins MWG Operon (http://www.eurofinsgenomics.eu/). Individual primer sequences, optimized annealing temperatures, and predicted product lengths can be found in Supplementary Table 1.

Purification of genomic DNA from 10 frozen bank vole and 10 frozen field vole (3 additional field voles were used for *LCN2* purification, isolation, amplification, etc.) whole-liver samples were done using Qiagen's DNeasy Blood & Tissue Kit.
DNA integrity and size were checked on an agarose gel, purity was assessed using a NanoDrop 1000 spectrophotometer (Thermo Scientific) and the concentration measured with a Qubit fluorometer (Life Technologies). PCR isolating and amplifying of the regions of interest were conducted using Bioline's BioMix™ Red PCR kit. PCR products were confirmed by running an agarose gel indicating both product size and specificity. Product was also confirmed through the eventual use of the DNA sequence homology tool BLAST.

Sequencing and Single Nucleotide Polymorphism (SNP) Identification

PCR products were cleaned using the ExoSAP-IT enzymatic mixture (USB Corporation) and both forward and reverse stands of DNA were sequenced via BigDye® Terminator v3.1 Cycle Sequencing Kit. Products were subsequently purified with BigDye® XTerminator Purification Kit. DNA Sequencing was conducted using

capillary electrophoresis in the Applied Biosystems 3130x Genetic Analyzer (Life Technologies). Sequences were aligned using the bioinformatics software package Geneious[38], and single nucleotide polymorphisms (SNPs) were manually identified from the front-end user interface of Geneious which can display forward and reverse sequence traces generated by the Applied Biosystems 3130x Genetic Analyzer. Heterozygous individuals were also identified manually via Geneious' front-end user interface, where heterozygous peaks can be readily seen in both forward and reverse sequences traces.

SNP designation as either synonymous or non-synonymous was determined by aligning the sequenced field vole and bank vole exons with known orthologous house mouse (*Mus musculus*) cDNA sequences obtained from GenBank. This method assumes no frame-shift mutations have occurred since the divergence of bank vole or field vole and *Mus musculus*.

Evolutionary Analysis of Sequences

Statistical analyses were carried out using DnaSP version 5.10.01[39] and the online HyPhy package available through datamonkey.org[40]. Estimates of nucleotide diversity were performed using the average number of pairwise nucleotide differences between paired sequences, π[41], an estimate of the neutral population parameter θ based on the total number of segregating sites, Watterson's θ_W[42], and haplotype diversity (H_d), the probability that two haplotypes chosen at random are different within the sample where N is the number of gene copies, k is haplotype number, and x_i is the frequency of the i^{th} haplotype[41].

$$H = \frac{N}{N-1}(1 - \sum_i x_i^2)$$

Within and between species tests using DnaSP v. 5.10.01 identified several significant departures from neutrality under the infinite sites model whereby the number of nucleotide sites is assumed to be large enough so that a new mutation is assumed not to occur over a previously mutated site[43]. In a neutrally evolving diploid population, the expected nucleotide variation is equal to $\theta = 4N_e\mu$, where μ is the mutation rate and N_e is the effective population size. Tajima's *D* estimates the difference between any two of θ, θ_W, and π which in a perfect neutrally evolving population is zero. Values of *D* significantly away from zero result either from selective or demographic effects.

For tests requiring an outgroup, orthologous house mouse (*mus musculus*) sequences were implemented for all analyses. One set of tests, Fu and Li's *D* and *F* tests[44], similarly implement two estimators of θ. The *D* statistic is based on the differences between mutation count in external branches of the genealogy and the total mutation count. The *F* statistic is the normalized difference between the mutation count in external branches of the genealogy and π. Negative or positive selection is likely to cause excess recent mutations, rather than older mutations, either through the low frequency of deleterious alleles or the rapid fixation of advantageous ones respectively[44]. Another test, Fay and Wu's *H*-test[45], identifies patterns indicative of positive selection based on differences between θ_h (homozygosity of derived variant weighted estimator) and π, and on the effect selective sweeps have on relative levels of high-frequency-derived alleles. An excess indicates genetic hitchhiking and positive selection[45]. The McDonald-Kreitman test[46], or MK test, compares rates of synonymous and non-synonymous substitutions both within and between species.

Sequence diversity estimates and tests for non-neutral selection can be found in Tables 4a, 4b and 5. These values were generated using DnaSP v5.10.01's neutral coalescent simulations assuming no recombination for 16,000 replicates based on the observed number of segregating sites. The reported significant *p*-values are an average of 10 simulations. As these tests were conducted with a one sided p-value, statistically significant values at $p<0.05$ for the observed test statistic included those either below the 2.5th percentile or those above the 97.5th percentile.

Other tests capable of identifying evidence of positive selection were carried out using the HyPhy package available via the online Datamonkey webserver (http://www.datamonkey.org, accessed April 2014). The single likelihood ancestry counting (SLAC) method implements counts of dS (synonymous mutations) and dN (non-synonymous mutations) for each codon position generated by reconstructing the ancestral sequences using a joint likelihood reconstruction method to detect evidence of positive selection[47,48]. The random-effects likelihood (REL) method was used to detect evidence of positive or negative selection at the codon level by modeling variation in dN and dS rates at individual sites using a Bayesian approach within a predefined distribution[48]. The Fast Unconstrained Bayesian AppRoximation (FUBAR) method identifies both positive and purifying selection with a Markov Chain Monte Carlo routine by taking averages over many predefined site classes resulting in an unconstrained distribution of selection parameters[49]. The mixed-effects model of evolution (MEME) method was implemented to identify sites undergoing positive selection which varies throughout the predicted tree (episodic selection)[50]. These analyses differ from tests

carried out with DnaSP in that they are not limited by homologous coverage, i.e. they are able to analyze sequences with missing sites.

Results

Predicted exons from 7 field vole and 5 bank vole immune-related iron transport genes were successfully amplified and sequenced from between 9 and 13 individuals (18 - 26 haplotypes), of which at the least 10 haplotypes and at the most 24 haplotypes were of good enough quality for DnaSP analyses (adequate coverage across all individuals). In total, 6,627 basepairs (bp) (3,600 bp for field vole, 3,027 bp for bank vole) of good quality coding sequence was obtained. Fifty nine SNPs across both species were identified, with bank vole yielding an average of one SNP per 124 bp and field vole yielding an average of one SNP per 145 bp of coding sequence (Table 2a, Table 2b).

Across both vole species the immune-related iron transport gene sequences covered between 59.3% - 95.6% of the total length of the orthologous mouse coding sequence; this coverage estimate is not including the *Tfr1* sequences as only the genetic sequences immediately flanking specific Tfr1 apical sites of interest were amplified for sequencing. The average coverage, again excluding *Tfr1* sequences, across both species was 71%. Sequence length across all within-species genes, the total coverage length of each gene and its respected coverage compared to orthologous mouse coding sequence, as well as individual species coverage averages can be seen in Table 2a and Table 2b.

Table 2a Summary of sequenced field vole immune-related iron transport genes

Gene	2*n*[a]	Length[b]	Length & Coverage[c]	SNPs	Bp per SNP
Hp	18	663	729 (69.83)	5	133
Lcn12	20	141	345 (59.28)	1	141
Ftmt	20	384	477 (66.81)	7	55
Tfr1	20	327	617 (26.92)	2	164
Slc11a1	12	636	1084 (65.82)	2	318
Ltf	10	1077	1515 (71.33)	13	83
Lcn2	24	372	576 (95.52)	3	124
	Total	**3600**	**5343 (71.43)**	**33**	**145**

[a] Number of haplotypes sequenced

[b] Length sequenced for each locus excluding sites/sequences with gaps when aligned

[c] Total length sequenced for each locus and coverage as estimated from length of orthologous mouse coding sequence

Table 2b Summary of sequenced bank vole immune-related iron transport genes

Gene	2*n*[a]	Length[b]	Length & Coverage[c]	SNPs	Bp per SNP
Hp	19	597	678 (64.94)	5	119
Tfr1	17	132	210 (9.16)	1	132
Slc11a1	17	669	1076 (65.33)	4	167
Ltf	15	1224	1662 (78.25)	12	102
Lcn2	20	405	441 (73.13)	4	101
	Total	**3027**	**4067 (70.41)**	**26**	**124**

[a] Number of haplotypes sequenced

[b] Length sequenced for each locus excluding sites/sequences with gaps when aligned

[c] Total length sequenced for each locus and coverage as estimated from length of orthologous mouse coding sequence

Single nucleotide polymorphisms were discovered in every immune-related iron transport gene sequenced. Frequency of SNPs ranged from 1 - 13 per gene and from one SNP per 55 - 318 bp of sequence (refers to sequence length in loci excluding sites/sequences with gaps when aligned) (Table 2a, 2b). All discovered SNPs found within the sequenced iron transport genes can be found in Table 3, with indicated SNP positions relative to the orthologous mouse reference sequence. Twenty four of the 59 total SNPs were nonsynonymous mutations resulting in an amino acid change of the translated protein.

Significantly high levels of diversity, either through significantly high haplotype diversity (H_d) or number of haplotypes (N_h), was found within field vole transferrin receptor (N_h =4 p<0.05), field vole haptoglobin (N_h=7 p<0.05), field vole mitochondrial ferritin (N_h=10 p<0.05, H_d=0.879 p<0.1), Field vole lactotransferrin (N_h=10 p<0.05, H_d=1.0 p<0.05), bank vole lactotransferrin (N_h =11 p<0.05) and borderline non-significance in bank vole lipocalin-2 (N_h =6 p<0.1). π and Watterson's estimator, θ_w (see methods) can be seen along with the significant estimates of diversity in Table 4a, 4b for field vole and bank vole respectively.

Field Vole - Sequence Diversity and Tests of Neutrality

					Tajima's	Fu and Li's[e]		Fay and Wu's[e]	*MK*[f]
Gene	*Nh*[a]	*Hd*[b]	π[c]	θ_π[d]	*D*	*D*	*F*	*H*	*p*
Hp	**7***	0.791	19.6	21.9	-0.333	1.150	1.002	**-3.268***	0.153
Lcn12	2	0.189	13.4	20	-0.592	0.627	0.362	**-1.516***	1
Ftmt	**10***	**0.879†**	50.8	51.4	-0.034	-0.717	-0.614	-1.253	1
TfR1	**4***	0.284	11.6	17.2	-0.739	0.627	0.362	0.168	1
Slc11a1	3	0.530	9.1	10.4	-0.382	-0.483	-0.535	0.424	0.189
Ltf	**10***	**1.0***	59.6	49.2	0.973	0.334	0.576	2.400	1
Lcn2	5	0.743	29.9	21.6	0.967	0.985	1.148	-1.167	0.238

a Number of haplotypes
b Haplotype Diversity
c Mean number of pairwise differences: per site (x10⁴)
d Watterson's Estimator (x10⁴). Significance values in bold
e For tests requiring an outgroup, mus musculus sequence was used
f McDonald-Kreitman Test p-values determined via Fisher's Exact Test
* = p <0.05
† = 0.05 < p < 0.10

Table 4a: Sequence diversity and tests of neutrality for sequenced field vole immune-related iron transport proteins.

Bank Vole - Sequence Diversity and Tests of Neutrality

Gene	Nh[a]	Hd[b]	π[c]	θ_π[d]	Tajima's D	Fu and Li's[e] D	Fu and Li's[e] F	Fay and Wu's[e] H	MK[f] p
Hp	3	0.368	12.9	24	-1.432	-0.853	-1.129	-1.333	0.1669
TfR1	2	0.221	16.7	22.4	-0.491	0.648	0.417	0.191	1
Slc11a1	4	0.419	10.1	17.7	-1.311	-0.800	-1.109	0.603	**0.03497†** ***G=5.263****
Ltf	**11***	0.933	37	30.2	0.890	**1.564***	**1.756†**	-1.457	0.7637
Lcn2	**6†**	0.768	26.6	27.8	-0.125	0.128	0.067	0.368	0.6128

a Number of haplotypes
b Haplotype Diversity
c Mean number of pairwise differences: per site ($x10^4$)
d Watterson's Estimator ($x10^4$). Significance values in bold
e For tests requiring an outgroup, mus musculus sequence was used
f McDonald-Kreitman Test p-values determined via Fisher's Exact Test
* = $p < 0.05$
† = $0.05 < p < 0.10$

Table 4b: Sequence diversity and tests of neutrality for sequenced bank vole immune-related iron transport proteins.

Various tests for non-neutral evolution yielded significant results for every immune-related iron transport protein in either species explored in this study. Fu and Li's *D* and *F* were significant and borderline significant, respectively, for *Ltf* within bank vole (*D* = 1.564, $p<0.05$, *F*=1.756 $p<0.1$). *Slc11a1* was determined in bank vole to depart from neutrality by the McDonald-Kreitman test (one-tailed $p=0.03497$), and patterns of diversity consistent with the action of positive selection were detected in field voles for both *Hp* and *Lcn12* by Fay and Wu's H-test (*H* = -3.268, $p<0.05$, *H* = -1.516, $p<0.05$ respectively). No departures from neutrality were detected with the HKA test.

Individual codons predicted to be under positive or negative selection were detected for every immune-related iron transport protein. Across all genes analyzed, 20 codons were detected to be under positive or negative selection using the single likelihood ancestry counting (SLAC) method, 56 codons under positive or negative selection using the random-effects likelihood (REL) method, 31 codons under positive selection using mixed-effects model of evolution (MEME) method, and 118 codons predicted to be under positive or purifying selection using the fast unconstrained

Alignment			Diversity Estimate	Sites Under Selection					
Locus	Taxa	bp (AA)	dn/ds[a]	SLAC	REL	Either[b]	MEME (+)	FUBAR (diversifying)	FUBAR (purifying)
Hp	4	1044 (348)	0.179	5	10	13 (3.74%)	1	1	0
Lcn12	3	744 (248)	0.654	0	6	6 (2.42%)	1	0	1
Ftmt	3	714 (238)	0.142	2	0	2 (0.84%)	1	0	8
FTH1-like	3	729 (243)	0.406	1	4	4 (1.65%)	2	1	7
TfR1	4	2295 (765)	0.352	0	11	11 (1.44%)	0	1	13
Slc11a1	4	1647 (549)	0.155	0	0	0 (0%)	11	0	46
Ltf	4	2124 (708)	0.290	12	14	20 (2.82%)	12	4	29
Lcn2	4	603 (201)	0.419	0	11	11 (5.47%)	3	2	5

a dn/ds estimate derived from SLAC procedure
b Either refers to the number (and percentage) of codons at which selection was detected with either the SLAC or REL methods

Table 5: Individual codons predicted to be under positive or negative selection using SLAC, REL, MEME, and FUBAR methods See Materials and Methods)

bayesian approximation (FUBAR) method (Table 5).

Across all analyzed genes, 2.29% of codons were under either positive or negative selection according to REL or SLAC analyses. Detailed information regarding SLAC, REL, MEME, and FUBAR results can be found in Table 5 and the raw datamonkey output in Supplementary Table 2.

While there were many sites identified to be under positive or negative selection, these analyses identified several previously described functionally relevant apical sites of the transferrin receptor complex as under positive selection (Figure 4[2]). Not included within the figure is another amino acid previously found to be under positive selection (apical residue 109[2]), which the REL analysis performed in this study corroborates (Table 6).

Two non-synonymous mutations, corresponding to areas near predicted binding sites 128 and 147, were identified to be under positive selection by both REL and MEME methods (Table 6) for lipocalin-2.

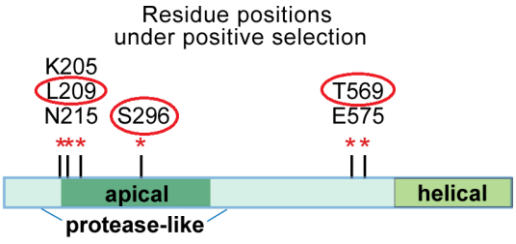

Figure 4: The circled residues were identified in this study to be under positive selection. All listed residues were previously found to be under positive selection by Demogines, *et al.*[2]

		Previously Identified Sites of Selection						
Gene	Species	Residue 109	Residue 205	Residue 209	Residue 215	Residue 296	Residue 569	Residue 575
Tfr1	Human	LAGTESPVRE	IIVDKNGRL	IVDKNGRLV	YLVENPGG	AELSFFGH	YLGTTMD	MDTYKEL
	House Mouse	LAETEETDKS	TIVQSNGNL	TIVQSNGNL	DPVESPEG	ADLALFGH	YLGTRLD	LDTYEAL
	Bank Vole	N/A	TIINTSGGG	VTIINTSGGG	YLLENPEG	ADLSLFGH	N/A	N/A
	Field Vole	LAET-ETGYS	TIINTNGGG	VTIINTNGGG	YLLENPEG	ADLPLFGH	YLGT[P/S]LD	LDTYEKL
Significant Results of Tests for Positive Selection		REL		REL		REL	REL	

		Sites of Interest[a]		
Gene	Species	Residues 124, 125, 128[b]	Residues 144, 147[b]	Residue 156[b]
Lcn2	Human	SYPGLTSYLV	AMVFFKKVSQ	NREYFKITLY
	House Mouse	RYPQVQSYNV	AMVFFRKTSE	NKQYFKITLY
	Bank vole	SYP[R/W]LRSYNV	AMV[F/Y]FKKTSG	NKQYFKVTLY
	Field vole	SYPQIRSYNV	AMVFFKRTSG	NKQYFKVTLY
Significant Results of Tests for Positive Selection		REL MEME	REL MEME	

[a] sites of interest refer to predicting binding sites (green), as well as residues determined to be under positive selection (yellow)

[b] Residue position number reflects the aligned position, UniProt P80188 has these residues located at 126, 145, 154 respectively

Table 6: A) Tfr1 sites previously found to be under selection[2] aligned with bank vole and field vole protein sequences. If the residue of interest was predicted to be under selection the test method is listed below the aligned sequences. B) Aligned lipocalin-2 protein sequences for human, mouse, bank vole, and field vole with binding residues and residues predicted to be under selection highlighted.

DISCUSSION

This is the first study to date describing the genetic variation and site-specific non-neutral selection pressures for several immune-related iron transport proteins in natural populations of bank vole and field vole. Fifty-nine SNPs, 24 of which are non-synonymous, have been described between the two species included in this study. Several of these SNPs support previous evidence for an ongoing evolutionary arms race between *TfR1* and mammalian viruses, some of which have serious potential for emerging into human populations.

The identification of codons detected to display patterns indicative of positive selection within this study which have previously been identified as sites of positive selection should lend confidence to our statistical methods and internal validity as they relate to other previously unstudied genes. Furthermore, the identification of sites undergoing positive selection which corroborate previous findings further support the theory of an evolutionary arms race between *TfR1* and mammalian viruses.

This study also describes putative evidence of positive selection in *Lcn2*. While *Lcn2* has previously been found to be under negative selection[51], and indeed many residues within *Lcn2* were identified to be under negative selection within this study (Supplementary Table 2), this work is the first to provide evidence of patterns indicative of positive selection within sites adjacent to functional loci of the lipocalin-2 siderophore binding calyx. Seen in Figure 5 is human LCN2 bound to the siderophore Fluvibactin[52] obtained from the RCSB Protein Data Bank (DOI:10.2210/pdb4k19/pdb) with residues of interest shown. It should be noted that the non-synonymous mutations in bank vole were within a five-residue range of the predicted human binding site when aligned with the orthologous mouse sequence. Shown here are those predicted sites aligned to match the human protein sequence (Figure 5).

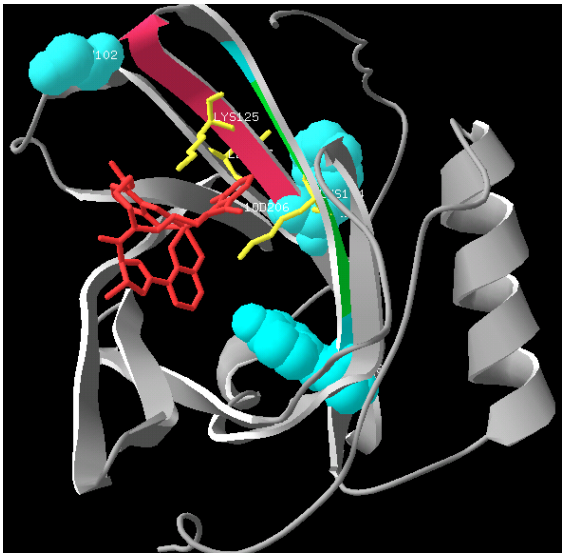

Figure 5: Human LCN2 bound to the siderophore Fluvibactin. In red, the siderophore can be seen within the calyx, the three residues predicted to form the binding calyx are in yellow, and the blue residues are the bank vole residues where SNPs were identified (2 non-synonymous, 1 synonymous).

It is important to mention that these proteins are all members of different yet overlapping pathways. Thus, functional changes in one protein undoubtedly impact the activity of the others. For example, Slc11a1 has been shown to mediate macrophage activity and the successful killing of salmonella enterica serovar typhimurium via the upregulation of lipocalin-2[53]. Whether or not the *Slc11a1* variation described in this study impacts its interaction with *Lcn2*, or vice versa, is unknown and should be addressed in future research.

This study adds to the body of literature indicating immune-related biological pathways undergo extensive non-neutral selection pressures. This work also contributes to our growing understanding of the diversity within iron-transport proteins in natural populations. There are, however, several limitations to this study.

The lack of certain genetic regions and entire genes confounds the conclusions we have drawn. Certain *TfR1* sites as well as the entire *Lcn12* protein were not predicted by the bioinformatics software to be present within bank vole or could not be optimized during the primer-design stage and are thus not included. Similarly, certain coding segments for *Slc11a1*, *Ltf*, *Hp*, and *Lcn12* were unsuccessfully sequenced and thus affect coverage and the power of our statistical analyses. This flaw could be attributable to initial high-throughput sequencing errors, errors within the bioinformatics software used to predict exons of the sequenced genes or proteins, primer non-specificity, or are not errors and merely reflect the biological differences between bank vole and field vole.

Finally, whether or not the genetic variation described here confer advantages in iron binding or transport, in disease resistance or susceptibility, or whether they play a role in an evolutionary arms race similar to *TfR1* cannot be drawn from host-genetics alone. Future research should therefore be conducted to determine the relationship between these proteins and the pathogens encountered by natural populations of field vole and bank vole. Iron-binding and iron-transport assays should also be conducted to determine if the mutations described in this paper affect the purported primary function of these proteins.

Conclusion

This study aimed to broaden research on immune-related genetics within natural populations by focusing on a suite of iron-transport proteins shown to be directly or indirectly involved in the host immune response. Describing both the genetic diversity and providing evidence of patterns indicative of non-neutral selection pressures acting upon these iron transport proteins was accomplished. Importantly, this study describes such genetic variation for several immune-related iron transport proteins within functionally important loci. Such descriptions allow for the formulation of novel hypotheses as to why such variations at important loci have occurred. Whether or not such diversity effects iron transport or immune function *in vivo*, or whether this variation is due to evolutionary arms races over iron with pathogens have yet to be elucidated and should be the focus of future research.

Supplementary Table 2: Summary of SNPs

Organism	Gene	SNP[a]	cDNA site[b]	AA site[c]	Codon (amino acid) change	Syn/Nonsyn[d]
Field Vole	*Ftmt*	Ftmt 309 G/A	309	103	GTG (Val)→GTA(Val)	Syn
		Ftmt 316 T/C	316	106	TAC (Tyr)→CAC(His)	Nonsyn
		Ftmt 354 G/A	354	118	GAG (Glu)→GAA(Glu)	Syn
		Ftmt 369 G/A	369	123	GCG(Ala)→GCA(Ala)	Syn
		Ftmt 375 G/A	375	125	AAG(Lys)→AAA(Lys)	Syn
		Ftmt 475 T/A	475	159	TGT (Cys)→AGT(Ser)	Nonsyn
		Ftmt 663 C/T	663	221	GGC (Gly)→GGT(Gly)	Syn
	Hp	Hp 456 T/C	456	152	GAT (Asp) → GAC (Asp)	Syn
		Hp 474 T/C	474	158	ATT(Ile)→ATC (Ile)	Syn
		Hp 639 C/T	639	213	TAC(Tyr)→TAT(Tyr)	Syn
		Hp 717 T/C	717	239	TAT(Tyr)→TAC(Tyr)	Syn
		Hp 834 T/C	834	278	CAT(His)→CAC(His)	Syn
	Lcn12	Lcn12 208 A/C	208	70	ATG (Met)→CTG (Leu)	Nonsyn
	Lcn2	Lcn2 180 C/G	180	60	GCC (Ala)→GCG(Ala)	Syn
		Lcn2 228 C/G	228	76	TAC(Tyr)→TAT(Tyr)	Syn
		Lcn2 540 T/C	540	180	TCT(Ser)→TCC(Ser)	Syn
	Ltf	Ltf 197 C/A	197	66	GCC(Ala)→GAC(Asp)	Nonsyn
		Ltf 198 C/T	198	66	GCC(Ala)→GCT(Ala)	Syn
		Ltf 200 T/C	200	67	ATT(Ile)→ACT(Thr)	Nonsyn
		Ltf 850 G/A	850	284	GCC(Ala)→ACC(Thr)	Nonsyn
		Ltf 1072 G/T	1072	358	GCC(Ala)→TCC(Ser)	Nonsyn
		Ltf 1122 C/T	1122	374	CGC(Arg) →CGT(Arg)	Syn
		Ltf 1287 G/T	1287	429	CTG(Leu)→CTT(Leu)	Syn
		Ltf 1293 G/A	1293	431	GAG(Glu)→GAA(Glu)	Syn
		Ltf 1337 G/A	1337	446	AGA(Arg)→AAA(Lys)	Nonsyn
		Ltf 1344 G/A	1344	448	GCG(Ala)→GCA(Ala)	Syn
		Ltf 1351 T/G & 1353 C/A	1351 & 1353	451	TAC (Tyr)→GAA(Glu)	Nonsyn
		Ltf 1354 T/A	1354	452	TTT (Phe)→ATT(Ile)	Nonsyn
		Ltf 1355 T/A	1355	452	TTT (Phe)→AAT(Asn)	Nonsyn
		Ltf 1542 C/T	1542	514	GAC(Asp)→GAT(Asp)	Syn
		Ltf 1551 C/T	1551	517	TCC(Ser)→TCT (Ser)	Syn
	Slc11a1	Slc11a1 581 C/G	581	194	GCT(Ala)→GGT(Gly)	Nonsyn
		Slc11a1 1128 C/T	1128	376	GGC(Gly)→GGT(Gly)	Syn
	Tfr1	TfR1 636 C/A	636	212	GGC (Gly)→GGA (Gly)	Syn
		TfR1 1717 C/T	1717	573	CCT (Pro)→TCT(Ser)	Nonsyn

Bank Vole	*Hp*	Hp 474 C/T	474	158	ATC(Ile)→ATT(Ile)	Syn
		Hp 708 C/T	708	236	CGC(Arg)→CGT(Arg)	Syn
		Hp 724 C/T	724	242	CTG(Leu)→TTG(Leu)	Syn
		Hp 768 C/A	768	256	GGC(Gly)→GGA(Gly)	Syn
		Hp 846 T/C	846	282	GCT(Ala)→GCC(Ala)	Syn
	Lcn2	Lcn2 124 T/A	124	42	TTC(Phe)→ATC(Ile)	Nonsyn
		Lcn2 370 C/T	370	124	CGG(Arg)→TGG(Trp)	Nonsyn
		Lcn2 431 T/A	431	144	TTT(Phe)→TAT(Tyr)	Nonsyn
		Lcn2 480 T/C	480	160	TAT(Tyr)→TAC(Tyr)	Syn
	Ltf	Ltf 375 C/T	375	125	AAC(Asn)→AAT(Asn)	Syn
		Ltf 483 G/A	483	161	GCG(Ala)→GCA(Ala)	Syn
		Ltf 660 T/C	660	220	AAT(Asn)→AAC(Asn)	Syn
		Ltf 721 G/C	721	241	GAA(Glu)→CAA(Gln)	Nonsyn
		Ltf 823 G/A	823	275	GCC(Ala)→ACC(Thr)	Nonsyn
		Ltf 1753-4 GA/CG	1753-4	585	GAG(Glu)→CGG(Arg)	Nonsyn
		Ltf 1764 G/A	1764	588	GAG(Glu)→GAA(Glu)	Syn
		Ltf 1768 T/C	1768	590	TTG(Leu)→CTG(Leu)	Syn
		Ltf 1782	1782	594	GAC(Asp)→GAT(Asp)	Syn
		Ltf 1881 G/T	1881	627	CAG(Gln)→CAT(His)	Nonsyn
		Ltf 1933 G/A	1933	645	GGG(Gly)→AGG(Arg)	Nonsyn
	Slc11a1	Slc11a1 308 G/T	308	103	TGC(Cys)→TTC(Phe)	Nonsyn
		Slc11a1 964 A/C & 965 G/A	964-5	321	AGC(Ser)→CAC(His)	Nonsyn
		Slc11a1 1185 C/T	1185	395	TTC(Phe)→TTT(Phe)	Syn
	Tfr1	TfR1 658 G/A	658	220	GAG(Glu)→AAG(Lys)	Nonsyn

Table 1

Gene	Gene Symbol	Summary
Haptoglobin	Hp/ Bp/ Hpa1s	A plasma protein which binds free hemoglobin. There are three major subtypes. Case control studies have associated Haptoglobin type with various cancers, infectious diseases, liver disease, diabetes, cardiovascular disease, neurological disorders, anemia, and haemochromatosis[54].
Lipocalin 12	Lcn12	Member of the Lipocalin family, roughly 28-29% identical to *Lcn2*. Very little is known regarding the function of this lipocalin, however it has been shown to be a component of seminal fluid[51] and has a positively charged triad of side-chains identical to the binding site of *LCN2*[35].
Mitochondrial Ferritin	Ftmt/ Mtf	A protein expressed in the brain and testis and is localized to mitochondria. FtMt has been suggested to play a neuroprotective role in neurodegenerative diseases, preventing cell death in the context of oxidative stress and neurotoxic proteins[55] .
Transferrin Receptor Protein 1	Tfrc/ TfR1/ CD71/ p90	A ubiquitously expressed transmembrane protein that regulates the delivery of iron from transferrin into cells. TfR1 has also been shown to be the target of cellular entry for multiple mammalian viruses[2].
Solute carrier family 11, member 1	Slc11a1/ Lsh/ NRAMP1	A membrane protein belonging to the proton-coupled divalent metal ion transporter family. Plays important roles in macrophage activation, the regulation of chemokine KC, interleukin-1β, inducible nitric oxide synthase and nitric oxide release, major histocompatibility complex class II molecules, and tumor necrosis α. Also associated with viral, bacterial, and protozoal pathogens[56].
Lactoferrin / Lactotransferrin	Ltf/ Lf/ Gig12/ Hlf2	A major iron-binding protein secreted in neutrophil granules and various bodily fluids. Shown to induce iron-deficiency as a form of host defense have direct antiviral, antifungal, and antibacterial activity; Ltf can promote wound healing, bone growth, and modulate various aspects of the host immune response. Can bind lipopolysaccharides such as endotoxin which can modulate the immune system[57].
Lipocalin-2/ Neutrophil Gelatinase-Associated Lipocalin	Lcn2/ Ngal/ 24p3/ Msfi	A member of the lipocalin family which binds to iron-sequestering siderophores produced by bacteria. Found in humans, mice, chickens, quail, and various mammals. Siderocalin has also been proposed to play roles in embryogenesis, differentiation, inflammation, and cancer[27].

Sources

1 (ed Rötelmaus.jpg) (Wikipedia Commons, http://commons.wikimedia.org/wiki/File:R%C3%B6telmaus.jpg).

2 Demogines, A., Abraham, J., Choe, H., Farzan, M. & Sawyer, S. L. Dual host-virus arms races shape an essential housekeeping protein. *PLoS biology* **11**, e1001571, doi:10.1371/journal.pbio.1001571 (2013).

3 Turner, A. K. *Immunogenetics and polymorphism in a natural population of field voles (Microtus agrestis)* Doctor in Philosophy thesis, University of Liverpool, (2010).

4 Haldane, J. B. S. Disease and Evolution. *La Ricerca Scientifica* **Supplement A.**, 68-76 (1949).

5 Fumagalli, M. *et al.* Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. *PLoS genetics* **7**, e1002355, doi:10.1371/journal.pgen.1002355 (2011).

6 Nielsen, R. *et al.* A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS biology* **3**, e170 (2005).

7 McTaggart, S. J., Obbard, D. J., Conlon, C. & Little, T. J. Immune genes undergo more adaptive evolution than non-immune system genes in Daphnia pulex. *BMC evolutionary biology* **12**, 63 (2012).

8 Johansson, M. E. V. *et al.* The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria. *Proceedings of the National Academy of Sciences* **105**, 15064-15069, doi:10.1073/pnas.0803124105 (2008).

9 Ponka, P., Beaumont, C. & Richardson, D. R. Function and regulation of transferrin and ferritin. *Seminars in hematology* **35**, 35-54 (1998).

10 Kaplan, J. & Ward, D. M. The essential nature of iron usage and regulation. *Current Biology* **23**, R642-R646 (2013).

11 Kaplan, C. D. & Kaplan, J. Iron acquisition and transcriptional regulation. *Chemical reviews* **109**, 4536-4552, doi:10.1021/cr9001676 (2009).

12 Knutson, M. & Wessling-Resnick, M. Iron metabolism in the reticuloendothelial system. *Critical reviews in biochemistry and molecular biology* **38**, 61-88, doi:10.1080/713609210 (2003).

13 Ganz, T. & Nemeth, E. Hepcidin and iron homeostasis. *Biochimica et biophysica acta* **1823**, 1434-1443, doi:10.1016/j.bbamcr.2012.01.014 (2012).

14 Wang, J. & Pantopoulos, K. Regulation of cellular iron metabolism. *The Biochemical journal* **434**, 365-381, doi:10.1042/bj20101825 (2011).

15 Anderson, G. J., Frazer, D. M. & McLaren, G. D. Iron absorption and metabolism. *Current opinion in gastroenterology* **25**, 129-135, doi:10.1097/MOG.0b013e32831ef1f7 (2009).

16 Nairz, M., Schroll, A., Sonnweber, T. & Weiss, G. The struggle for iron – a metal at the host–pathogen interface. *Cellular Microbiology* **12**, 1691-1702, doi:10.1111/j.1462-5822.2010.01529.x (2010).

17 Brookhart, M. A. *et al.* Infection risk with bolus versus maintenance iron supplementation in hemodialysis patients. *Journal of the American Society of Nephrology : JASN* **24**, 1151-1158, doi:10.1681/asn.2012121164 (2013).

18 Oppenheimer, S. J. Iron and Its Relation to Immunity and Infectious Disease. *The Journal of Nutrition* **131**, 616S-635S (2001).

19 Weiss, G. & Goodnough, L. T. Anemia of chronic disease. *New England Journal of Medicine* **352**, 1011-1023 (2005).

20 Radoshitzky, S. R. *et al.* Transferrin receptor 1 is a cellular receptor for New World haemorrhagic fever arenaviruses. *Nature* **446**, 92-96, doi:10.1038/nature05539 (2007).

21 O'Connell, D. Viral pathogenesis - An iron-clad interaction. *Nat Rev Microbiol* **5**, 169-169, doi:Doi 10.1038/Nrmicro1628 (2007).

22 Ross, S. R., Schofield, J. J., Farr, C. J. & Bucan, M. Mouse transferrin receptor 1 is the cell entry receptor for mouse mammary tumor virus. *Proceedings of the National Academy of Sciences* **99**, 12386-12390, doi:10.1073/pnas.192360099 (2002).

23 Parker, J. S. L., Murphy, W. J., Wang, D., O'Brien, S. J. & Parrish, C. R. Canine and Feline Parvoviruses Can Use Human or Feline Transferrin Receptors To Bind, Enter, and Infect Cells. *Journal of Virology* **75**, 3896-3902, doi:10.1128/jvi.75.8.3896-3902.2001 (2001).

24 Radoshitzky, S. R. *et al.* Machupo virus glycoprotein determinants for human transferrin receptor 1 binding and cell entry. *PloS one* **6**, e21398, doi:10.1371/journal.pone.0021398 (2011).

25 Hider, R. C. & Kong, X. Chemistry and biology of siderophores. *Natural product reports* **27**, 637-657, doi:10.1039/b906679a (2010).

26 Flo, T. H. *et al.* Lipocalin 2 mediates an innate immune response to bacterial infection by sequestrating iron. *Nature* **432**, 917-921, doi:10.1038/nature03104 (2004).

27 Sia, A. K., Allred, B. E. & Raymond, K. N. Siderocalins: Siderophore binding proteins evolved for primary pathogen host defense. *Current opinion in chemical biology* **17**, 150-157, doi:10.1016/j.cbpa.2012.11.014 (2013).

28 Schroder, K. *et al.* Conservation and divergence in Toll-like receptor 4-regulated gene expression in primary human versus mouse macrophages. *Proceedings of the National Academy of Sciences of the United States of America* **109**, E944-953, doi:10.1073/pnas.1110156109 (2012).

29 Langlois, M. R. & Delanghe, J. R. Biological and clinical significance of haptoglobin polymorphism in humans. *Clinical Chemistry* **42**, 1589-1600 (1996).

30 Shen, H. *et al.* Haptoglobin activates innate immunity to enhance acute transplant rejection in mice. *The Journal of clinical investigation* **122**, 383 (2012).

31 Bastarache, J., Wickersham, N., Sills, G., Ware, L. & Janz, D. Increased Haptoglobin Levels Are Associated With Reduced Mortality In Sepsis. *Am J Respir Crit Care Med* **187**, A1011 (2013).

32 Gruenheid, S., Pinner, E., Desjardins, M. & Gros, P. Natural resistance to infection with intracellular pathogens: the Nramp1 protein is recruited to the membrane of the phagosome. *The Journal of experimental medicine* **185**, 717-730 (1997).

33 Spik, G. & Montreuil, J. The role of lactotransferrin in the molecular mechanisms of antibacterial defense. *Bulletin europeen de physiopathologie respiratoire* **19**, 123-130 (1983).

34 García-Montoya, I. A., Cendón, T. S., Arévalo-Gallegos, S. & Rascón-Cruz, Q. Lactoferrin a multiple bioactive protein: an overview. *Biochimica et Biophysica Acta (BBA)-General Subjects* **1820**, 226-236 (2012).

35 Holmes, M. A., Paulsene, W., Jide, X., Ratledge, C. & Strong, R. K. Siderocalin (Lcn 2) also binds carboxymycobactins, potentially defending against mycobacterial infections through iron sequestration. *Structure (London, England : 1993)* **13**, 29-41, doi:10.1016/j.str.2004.10.009 (2005).

36 Cantarel, B. L. *et al.* MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res* **18**, 188-196, doi:Doi 10.1101/Gr.6743907 (2008).

37 Rozen, S. & Skaletsky, H. in *T Bioinformatics Methods and Protocols* Vol. 132 *Methods in Molecular Biology* 365-386 (1999).

38 Kearse, M. *et al.* Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-1649, doi:10.1093/bioinformatics/bts199 (2012).

39 Rozas, J., Librado, P., Sanchez-DelBarrio, J., Messeguer, X. & Rozas, R. DnaSP 5.10. 00. *Universitat de Barcelona, Spain* (2009).

40 Delport, W., Poon, A. F., Frost, S. D. & Kosakovsky Pond, S. L. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* **26**, 2455-2457, doi:10.1093/bioinformatics/btq429 (2010).

41 Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**, 406-425 (1987).

42 Watterson, G. A. On the number of segregating sites in genetical models without recombination. *Theoretical population biology* **7**, 256-276 (1975).

43 Kimura, M. THE NUMBER OF HETEROZYGOUS NUCLEOTIDE SITES MAINTAINED IN A FINITE POPULATION DUE TO STEADY FLUX OF MUTATIONS. *Genetics* **61**, 893-903 (1969).

44 Fu, Y. X. & Li, W. H. Statistical tests of neutrality of mutations. *Genetics* **133**, 693-709 (1993).

45 Fay, J. C. & Wu, C. I. Hitchhiking under positive Darwinian selection. *Genetics* **155**, 1405-1413 (2000).

46 McDonald, J. H. & Kreitman, M. Adaptive protein evolution at the Adh locus in Drosophila. *Nature* **351**, 652-654, doi:10.1038/351652a0 (1991).

47 Poon, A. F., Frost, S. D. & Pond, S. L. K. in *Bioinformatics for DNA Sequence Analysis* 163-183 (Springer, 2009).

48 Kosakovsky Pond, S. L. & Frost, S. D. W. Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection. *Molecular Biology and Evolution* **22**, 1208-1222, doi:10.1093/molbev/msi105 (2005).

49 Murrell, B. *et al.* FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Mol Biol Evol* **30**, 1196-1205, doi:10.1093/molbev/mst030 (2013).

50 Murrell, B. *et al.* Detecting individual sites subject to episodic diversifying selection. *PLoS genetics* **8**, e1002764, doi:10.1371/journal.pgen.1002764 (2012).

51 Suzuki, K. *et al.* Molecular evolution of epididymal lipocalin genes localized on mouse chromosome 2. *Gene* **339**, 49-59, doi:http://dx.doi.org/10.1016/j.gene.2004.06.027 (2004).

52 Allred, B. E., Correnti, C., Clifton, M. C., Strong, R. K. & Raymond, K. N. Siderocalin outwits the coordination chemistry of vibriobactin, a siderophore of Vibrio cholerae. *ACS chemical biology* **8**, 1882-1887 (2013).

53 Fritsche, G., Nairz, M., Libby, S. J., Fang, F. C. & Weiss, G. Slc11a1 (Nramp1) impairs growth of Salmonella enterica serovar typhimurium in macrophages via stimulation of lipocalin-2 expression. *Journal of Leukocyte Biology* **92**, 353-359, doi:10.1189/jlb.1111554 (2012).

54 Carter, K. & Worwood, M. Haptoglobin: a review of the major allele frequencies worldwide and their association with diseases. *International journal of laboratory hematology* **29**, 92-110, doi:10.1111/j.1751-553X.2007.00898.x (2007).

55 Yang, H. *et al.* Mitochondrial ferritin in neurodegenerative diseases. *Neuroscience research* **77**, 1-7, doi:10.1016/j.neures.2013.07.005 (2013).

56 Blackwell, J. M. *et al.* SLC11A1 (formerly NRAMP1) and disease resistance. *Cell Microbiol* **3**, 773-784 (2001).

57 Vogel, H. J. Lactoferrin, a bird's eye view. *Biochemistry and Cell Biology* **90**, 233-244, doi:10.1139/o2012-016 (2012).