Intragenus variation in a Chemokine Receptor Gene (*CCR5*) in *Homo*

Kara C. Hoover[1,2]

[1]Department of Anthropology, University of Alaska Fairbanks, Fairbanks Alaska

[2]Biochemistry and Molecular Biology, University of Alaska Fairbanks, Fairbanks Alaska

*Corresponding author

E-mail: kchoover@alaska.edu

Abstract

In comparison to non-human primates, humans have a notable increase in polymorphisms of the regulatory regions of the *CCR5* gene, an immune system gene with both general and specific functions. This particular pattern of variation is interpreted as allowing flexibility and diversity of gene expression in response to varying disease loads. A broad expression repertoire is useful to humans, the only globally distributed primate, due to unique cultural interventions that increase pathogen exposure and disease loads (e.g., sedentism, subsistence practices). The main objective of the study is to determine if the pattern of variation observed in humans extends to other members of the genus. The data for this study are mined from the published genomes of extinct hominins (Altai Neandertal, Denisova), an ancient human (Ust'-Ishim), and modern humans (1000 Genomes) using a variant call bioinformatics pipeline. No strong evidence for introgression was found to explain shared patterns of variation. Of the 21 polymorphisms identified in paleogenomes, 20 were previously found in modern humans and one is novel in Denisova. Despite introgression of haplotypes with variation in regulatory regions from extinct hominins, the pattern is not upheld in *CCR5*. While humans (ancient and modern) tend to have more polymorphisms in regulatory regions of the gene, neither the average number of polymorphisms per individual nor the distribution of polymorphisms across gene regions are not significantly different across members of the genus *Homo*. Thus, humans are not unique with regards to the increased frequency of regulatory polymorphisms and the evolution of variation patterns across *CCR5* gene appears to have originated within the genus. A broader evolutionary perspective on regulatory flexibility may be that it provided an advantage during the transition to confrontational foraging (and later hunting) that altered human-environment interaction as well as during migration to Eurasia and encounters with novel pathogens.

Introduction

The chemokine receptor *CCR5* is best known for its adaptive immune system role in binding the M-tropic human immunodeficiency virus (HIV) and creating a gateway to host cell infection (1-5). When expressed on white blood cell surfaces, *CCR5* also plays an innate immune system role by participating in the recruitment of T cells to infection sites. *CCR5* gene structure, the open reading frame, and amino acid identity are evolutionary highly conserved in human and non-human primates (6-9).

Within the primate order, the *CCR5* gene is highly homologous (9), which is to be expected given the evolutionary importance of gene structure (6). Gene evolution has been slow since the emergence of primates as distinct from rodents (9), with major trajectory differences across various parts of the gene (8). In primates, polymorphisms tend to be highest in Platyrrhini (New World Monkeys) and that diversity is linked to resistance to lentiviral infection (10). Primate polymorphisms outside conserved regions are common with non-synonymous specific polymorphisms occurring more often and with minimal deleterious effect; perhaps the redundancy of the chemokine family tolerates these polymorphisms, with the ligands of one receptor binding to a related receptor (7). In humans, polymorphisms in the open reading frame (ORF) are rare (11) with most occurring in regulatory (*cis*-acting) and untranslated regions (UTR); indeed, compared to other primates, humans exhibit high sequence variation in the *cis*-regulatory region (8).

Human *CCR5* is particularly interesting from an evolutionary perspective. First, it is characterized by complex splicing events and the use of two promoters which may have been functionally useful in adaptation to disease loads by allowing flexibility and diversity in the regulation of gene expression (6). Second and in light of the first point, the case of the genus *Homo* is particularly interesting because *Homo* has one of the broadest adaptive ranges of any species (12) and we carry in our immune systems evidence of the challenges we encountered during that adaptive radiation. This is important because understanding

shared derived variation and specific variation enables potential differentiation of challenges we overcame as a genus from challenges we overcame as a species. For instance, members of the genus *Homo* share an immunological adaptation to a limited extent in Europe. Evidence for gene introgression from extinct hominin species to modern humans is clustered (among other domains) in the immune system (13-16); in particular, the *OAS* anti-viral gene cluster on Chromosome 12 shows signatures of positive selection (15, 16), which suggests that adaptation to Eurasia was partly facilitated by pre-existing hominin adaptation to local viral loads. At a minimum, the environmental challenge faced by non-human members of *Homo* facilitated human adaptation to a new environment—a shared challenge with a similar solution. In the *CCR5* gene, introgression has not been reported but polymorphisms have been attributed to selection for disease resistance (11, 17, 18), neutral evolution or demographic factors (19-23), and redundancy of the *CCR5* product with other chemokine receptors (9). Ultimately, most polymorphisms have not been deleterious to health and are tolerated rather than subjected to purifying selective pressure—similar to the general pattern discussed above in primates more broadly (23).

CCR5 Gene Structure

CCR5 is a chemokine receptor, a cell surface protein in the subfamily of G protein-coupled receptors. Chemokine receptors facilitate communication between cells and the environment (7, 24) and mediate the activity of chemokines, proteins secreted by the immune system to chemically recruit immune cells to infection sites via chemotaxis (7, 25). Chemokine receptors are integrated into the membrane of immune cells, looping through the membrane seven times with extra- and intra-cellular components (7, 24). The majority of endogenous (and viral) chemokine binding takes place in the second extracellular loop of *CCR5* (26). A chief structural characteristic of chemokine receptors is the conservation of four cysteine amino acids that control correct conformation (folding) of the final protein; this conservation is critical because misconformation most often results in chemokine receptor dysfunction and immunodeficiency

(27). Other highly conserved features of CC genes in general are gene structure (6) and motifs involved in ligand binding and signaling (7).

The chemokine receptor *CCR5* is best known for its role in M-tropic HIV infection in humans (1-5). *CCR5* binds the chemokines *RANTES*, *MIP-1α* and *MIP-1β* (25) and acts as a co-receptor with the glycoprotein *CD4* on macrophages for the *gp120* of HIV-1 (T-tropic HIV-1 strains use co-receptors *CD4* and *CXCR4 on* T-Helper cells) (1, 3, 28-31). *CCR5* is located on Chromosome 3 (3p21) and is 6,065 bases long with an open reading frame (region of protein translation) of 1,056 bases that produces a protein consisting of 352 amino acids.

CCR5 has two common transcripts (A and B). Transcript A has 4 exons (the extra exon is a 235bp stretch of noncoding DNA in Transcript B) and Transcript B has 3 exons. Based on model biophysical data, Transcript A is considered the more stable of the pair (6). In both transcripts, there are two promoter regions. The *cis*-acting promoter (P_U) lies upstream of Exon 1 in both A and B transcripts (6) and the downstream promoter (P_D) lies upstream of Exon 3 and includes all or portions of Exon 2 in Transcript A. P_U is functionally weak compared to P_D and both promoters are hard to identify due to the absence of typical promoter markers (e.g., TATA box, CCAAT motif) (6). The open reading frame lies in Exon 4 and exhibits no variation between transcripts. The transcripts may result from upstream promoter polymorphisms causing alternate splicing—differential inclusion or exclusion of particular exons— in the production of messenger RNA (8)

CCR5 Evolution and Polymorphisms in Humans

CCR5 is the site of HIV1 infection in humans and polymorphisms in the gene (e.g., *CCR5Δ32*) mitigate viral infection and disease progression. Viruses have played a central role in shaping human immune system genetic variation; indeed, humans retain signatures of positive selection for areas of the genome regulating antiviral gene expression that arose via introgression from the Altai Neandertal.

Despite the high variability outside the open reading frame at this locus and evidence for balancing selection in humans, strong levels of introgression have not been reported, as they have for other loci in similar scenarios (e.g., HLA).

Due to its association with HIV infection and AIDs progression in humans, the evolutionary history of the *CCR5* gene in living primates and modern humans is well-documented. Gene structure is similar across species which suggests that structure serves a significant adaptive function (6). In addition, primate *CCR5* sequences are functionally similar and exhibit high conservation/low genetic diversity in the ORFs and in amino acid identity (8). Humans, however, exhibit high variation in the *cis*-regulatory regions of the gene (8).

There is significant variation across primate species in non-coding areas that regulate receptor expression levels on cell surfaces; these derive from messenger RNA and variants in the untranslated regions upstream of the gene (8, 23) and may be tolerated due to the high redundancy in chemokine receptors (e.g., *CCR5* ligands also bind to *CCR1* and *CCR4)* (9). A key difference between nonhuman and human primates is viral binding without a co-receptor in all but humans (who require *CD4* to bind HIV-1). *CD4* is commonly expressed on human macrophages and T Helper Cells, a trait that renders these critical immune system cells more vulnerable to infection and eventual destruction with viral progression; the overall pattern of common CD4 expression thus makes human infection with HIV-1 more prevalent and harmful than primate infection with SIV (9).

HIV-1 resistance has been linked to a 32bp deletion on the *CCR5* gene called *CCR5Δ32* that occurs in Exon 3 (32, 33) and affects the second extracellular loop, a common viral chemokine binding site (13, 34); the result is protein conformation that lacks the final three transmembrane domains (32, 33, 35, 36) and a clinical phenotype with slower T-cell population decrease and longer AIDS-free survival in heterozygotes (37). Conversely, *CCR5Δ32* has more recently been linked to West Nile Virus susceptibility (38-40).

The geographic distribution of *CCR5Δ32* is heavily concentrated in northern European populations (41, 42) which renders those immunological naïve and genetically susceptible populations more vulnerable to West Nile Virus while also conferring resistance to HIV infection. Data for West Nile Virus suggest a northward viral expansion into higher latitudes of Europe and North America which brings it into focus as a potential future health concern for the region (43) and adds a new element to the genetic epidemiology of contemporary populations relative to the *CCR5* gene.

Dates for the origins of *CCR5Δ32* in humans range from between 275 to 1,875 years ago based on a coalescent model and 127,500 years ago based on current frequency (18). Reported signatures of positive selection in *CCR5* have been linked to disease-specific (e.g., bubonic plague, smallpox) selection events in European human history (11, 18, 21, 44) but other analyses found no evidence for disease (either plague or smallpox) as a selective pressure with neutral evolution having the most explanatory power (19, 20, 22).

While the *CCR5* gene has been carefully studied due to its role in HIV infection, no work has explored variation within the human genus, *Homo*. The assumption that increased frequency of polymorphisms observed in human *cis*-regulatory regions are specific to humans is challenged by expanding the focus to include extinct members of the genus *Homo* in analysis. The plethora of research on the evolution of receptors binding HIV was conducted prior to technological and instrumental advances in the field of genomics that allowed the generation of deep coverage, high quality paleogenomes for long extinct species (including members of the genus *Homo*, Neandertal and Denisova). Publicly available paleogenomic data for extinct members of *Homo* has opened new lines of investigation into the more recent evolutionary history of *CCR5* polymorphisms in a genus with the widest geographic distribution among primates. While paleogenomic sample sizes are not robust to make statements on selection or add to a discussion of other evolutionary forces acting on variation, they provide an evolutionary dimension to understanding the patterns of variation characterizing our genus and

adaptation to new environments, potentially providing insights into past human-environment interactions. Thus, the overall aim of this research is to determine if there is shared variation in distribution of polymorphisms across *CCR5* within the genus *Homo*.

Modern human variation in *CCR5* was examined alongside that in the paleogenomes of extinct hominins (Altai Neandertal, Denisova) and ancient humans (Ust'-Ishim) to create a picture of variation in *Homo* and explore two major questions. First, are polymorphisms in humans attributable to introgression? Despite high levels of variation at this locus and evidence for balancing selection in humans at this locus, strong levels of introgression from inter-breeding with Neandertals in Eurasia have not been reported at this locus, as they have for other immune system loci in similar scenarios (13, 15, 45-49). Based on previous studies of the immune system and Neandertal introgression into modern humans, the expectation is that there is no introgression. Second are there shared patterns of variation across the genus for polymorphism frequency and distribution of polymorphisms that suggest a common evolutionary trajectory? Based on previous studies on primate gene structure and human variation as attributable to a more flexible gene expression on *CCR5*, the expectation is that there is a shared pattern of variation that aided adaptation for members of the geographically and ecological dispersed genus *Homo.* Both expectations were met.

Materials

Modern human data is from the 1000 Genomes Project (50) which ran from 2008 to 2015 and constitutes the largest public catalogue of human variation and genotype data. The project data are maintained and in the process of expansion via The International Genome Sample Resource (IGSR) which is under the aegis of the Data Coordination Centre at EMBL-EBI. The final original data set contains data for 2,504 individuals from 26 populations (Supplementary Table 1). While coverage is low

per individual, the data are robust enough to identify the majority of polymorphisms at a frequency of at least 1% in the populations studied, which is suitable for the current study.

Denisova and Altai Neandertal are Pleistocene Eurasian hominins with Denisova representing an eastern Eurasian Pleistocene population dated to 125,000-180,000 years ago and Neandertal a western Eurasian population (with some overlap with Denisova in Siberia). Genomic analysis indicates Denisova branched from Neandertal one million years before the split with humans and probably represents an earlier migration out of Africa (51) with some genomic signatures of interbreeding with ancestors of modern-day Melanesians (51). Altai Neandertal is related to Neandertals from Croatia, Germany, and the Caucasus (52). The Ust'-Ishim individual is from western Siberia, north of the settlement of Ust'-Ishim along the River Irtysh (Omsk Oblast, Russian Federation). Dated to the Late Pleistocene (50,000–30,000 years old), Ust'-Ishim likely represents an early modern human radiation into Europe and Central Asia; based on genomic comparisons to modern populations, Ust'-Ishim failed to leave descendants (53).

The Denisovan genome was generated from DNA extracted from a finger bone and has an average autosomal coverage of 30x (54). The Altai Neandertal genome was generated from DNA extracted from a toe bone and has an average autosomal coverage of 50x (52). The Ust'-Ishim ancient human genome was generated from DNA extracted from the distal end of femur and has an average autosomal coverage of 42x (53).

All genomes were sequenced by the Max Planck Institute for Evolutionary Anthropology and aligned to the human reference genome (hg19/GRCh37) using the common Burrow-Wheeler algorithm (BWA). The aligned and mapped sequences for Denisova and Altai Neandertal are available for download as BAM format files (a binary version of Sequenced Alignment/Map format typically used with next generation sequence data) via ftp. The aligned and mapped sequence data for Ust'-Ishim are available as one 238gb BAM file at the European Nucleotide Archive. All three genomes are high coverage, meaning that each base of the mappable genome is independently sequenced multiple times.

Contamination of Denisova and Altai Neandertal with modern human DNA is estimated to be less than 1% (52, 54) and of Ust'-Ishim with modern humans as less than 0.05% (53).

Methods

The human reference sequences for two common transcript variants for the *CCR5* gene (NM_000579 and NM_001100168) were downloaded from the National Center for Biotechnology Information (NCBI). The GRCh37/hg19 version of the genome is used to be consistent with the paleogenome data alignments and introgression data. The modern human variation data for single nucleotide polymorphisms (SNPs) in the region of interest (*CCR5* and *cis*-acting elements) and previously identified in the 1000 Genomes populations were downloaded via the 1000 Genomes ftp site as variant call format (VCF) files (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/). For the sake of comparison to the paleogenomes (n=1), modern human variation was recorded as present/absent (present if 1 or 2 variant alleles and absent if none) and then averaged across sample size. Paleogenome data were downloaded from the Max Planck Institute Leipzig data repository (http://cdna.eva.mpg.de/neandertal/altai/AltaiNeandertal/bam/ and http://cdna.eva.mpg.de/denisova/) and the European Nucleotide Archive (www.ebi.ac.uk/ena/data/view/PRJEB6622) to the (now retired) Pacman cluster at the University of Alaska Research Computing Systems (previously known as the Arctic Region Supercomputing Center) in BAM format.

Altai Neandertal BAM files were already split by chromosome. For Denisova and Ust'-Ishim, Samtools (version 1.3.1) view command was used to split BAM files into separate chromosomes. BAM files for all three genomes were sorted and indexed. The Samtools mpileup command was used to call variants and BCFtools (version 1.3.1) generated a variant call file using GR37 (3:46411633-46417697. The VCF file was zipped and indexed in Samtools. VCFtools (version 0.1.14) was used to call variants from the gene of interest using flags that marked the region of interest relative to the reference genome.

All output was zipped and indexed before creation of a FASTA format consensus sequence with variants in Samtools (See S1 Table for the paleogenome bioinformatics pipeline).

Distribution of SNPs across gene structure was guided by Mummidi et al. (6). The area for P_U targeted in this study was the most inclusive range (-1976 to +33) tested by Mummidi et al. because P_U was only weakly influential and little difference was noted between putative P_U regions studied. Further selection of this range avoided overlap with P_D. The area for P_D targeted in this study was the most productive range (+119 to +828) tested by Mummidi et al. Other gene structure regions (introns, UTR5' and UTR3') were guided by gene annotations. Significant difference in distribution of SNPs across gene structure for the four samples was tested using Monte Carlo methods for the exact test.

Introgression data for areas of interest on Chromosome 3 were mined from the datasets publicly available on the Reich lab homepage (https://reich.hms.harvard.edu/datasets) (16). The European-Asian population set data values were used for average marginal probability of Neandertal ancestry and number of alleles at the SNP that have Neandertal ancestry (>=0.9). Introgression analysis was based on the Conditional Random Field method which draws from three genomic features indicating Neandertal ancestry: allelic pattern at a single nucleotide polymorphism, high sequence divergence of the non-African haplotype to all YRI haplotypes but low divergence to Neanderthal; haplotype length consistent with interbreeding 37–86-thousand years ago. Introgression of Neandertal haplotypes into present-day human genomes was assessed using 1000 Genomes Project Phase 1 data, n=1,004; Altai Neanderthal genome, 52-fold average coverage; a six-primate consensus to determine ancestral alleles; and 176 Yoruban genomes assumed to contain no Neanderthal ancestry.

Results and Discussion

The goal of the research was to establish if *CCR5* variation has a genus-wide pattern or one unique to humans and if any shared variation patterns are attributable to Neandertal introgression. Based on gene

structure in primates and evidence for introgression of extinct hominin immune haplotypes, the expectation was that there would be a shared pattern of variation not attributable to introgression. The results meet the expectations.

All SNPs in ancient samples (n=21) called by the pipeline have been previously observed in modern humans except one in Denisova (Table 1 summarizes all variation found in paleogenomes; S2 Table summarizes all variation across samples). The novel SNP in Denisova lies in the UTR3' region; given previous studies on *CCR5* structure and the impact of SNPs in this area on proteins (6, 8), the novel Denisova SNP did not likely impact regulation or expression.

Table 1: Variants

Function	RefSeq	Humans	Ust'-Ishim	Altai	Denisova
Promoter	rs7637813	○	◉	⊙	●
Promoter	rs2227010	○	●	⊙	●
Intron	rs1799987	○	●	⊙	●
UTR-5	rs1799988	○	●	⊙	●
UTR-3	rs746492	○	●	⊙	●
Intron	rs2734648	○	●		●
Promoter	rs2856757	○	●		
Promoter	rs2734225	○	●		
UTR-5	rs1800023	○	●		
Intron	rs1800024	○	●		
Intron	rs2254089	○	●		
Intron	rs2856764	○	●		
Intron	rs2856765	○	●		
Intron	rs41515644	○	●		
UTR-3	rs1800874	○	●		
Promoter	rs41499550	○			●
Promoter	rs41395049	○			●
Promoter	rs41475349	○			●
Promoter	rs41412948	○		⊙	
Intron	rs2856762	○		⊙	
UTR-3	NOVEL	○			●

Can any observed ancient SNPs be artifacts of sequencing and degradation? Ancient DNA often contains C-to-T deaminations at the end of reads (for a review see 55). The lack of variation identified from paleogenomic sequence reads is unlikely to be a result of typical problems associated with ancient DNA sequence reads since chemical processes like deamination would increase SNPs (whether false or not). More significantly, the paleogenomes were generated using protocols (typically as reported in 56) that largely eliminates this error.

Is there equivalency in variation across the four samples? In the modern human sample, 260 known SNPs were observed (including 116 singletons—SNPs occurring in only one individual) with a total of 28,987 individual instances across those SNPs (S3 Table). SNP frequency per individual (total SNPs in a population/total number of individuals) within the 26 populations ranged from 9 to 13, with Asians exhibiting the highest variation and Europeans the least (Table 2).

Table 2: Average mutations per individual in 1000 Genomes by geographic region

Region	Population	Average per Individual
Africa	Luhya in Webuye, Kenya	10.94
	Gambian in Western Divisions in the Gambia	11.22
	Esan in Nigeria	11.36
	Yoruba in Ibadan, Nigeria	11.49
	Mende in Sierra Leone	12.29
Americas	Peruvians from Lima, Peru	9.15
	Puerto Ricans from Puerto Rico	10.99
	Colombians from Medellin, Colombia	11.15
	Mexican Ancestry from Los Angeles USA	11.39
	African Caribbeans in Barbados	11.97
	Americans of African Ancestry in SW USA	12.15
East Asia	Japanese in Tokyo, Japan	11.68
	Southern Han Chinese	12.33
	Chinese Dai in Xishuangbanna, China	12.71
	Kinh in Ho Chi Minh City, Vietnam	13.20
	Han Chinese in Bejing, China	13.32
Europe	British in England and Scotland	9.97
	Finnish in Finland	10.35
	Iberian Population in Spain	10.81
	Utah Residents (NW European Ancestry)	11.15

	Toscani in Italia	12.30
South Asia	Punjabi from Lahore, Pakistan	11.39
	Sri Lankan Tamil from the UK	11.50
	Gujarati Indian from Houston, Texas	11.78
	Indian Telugu from the UK	11.82
	Bengali from Bangladesh	12.23

Is there equivalency in the distribution of SNPs across gene structure? The ancient SNP counts are comparable to the range exhibited by modern humans (Table 3). Thus, extinct hominins and the ancient human had similar numbers of polymorphisms across the gene compared to modern humans suggesting a shared pattern of variation across *Homo*.

Table 3: Variation in gene structure

	P_D	ORF	P_U	UTR3'	UTR5'	Introns	Sum
Human	7294	259	9265	4623	4	7542	28987
Human Average	3	0	4	2	0	3	12
Ust'-Ishim	2	0	4	2	0	7	15
Altai	1	0	3	1	0	2	7
Denisova	1	0	5	2	0	2	10

Despite the high levels of variation and evidence for balancing selection in humans at this locus—which have been reported for other immune system loci in similar situations—strong levels of introgression from the archaic hominins into modern Eurasian populations have previously not been reported. Can any of the shared variants be the result of genomic introgression from the Altai Neandertal? Data mined from the Reich lab (16) indicate that there is no Neandertal ancestry present in the locus of interest; average probability of Neandertal ancestry in Eurasian samples is 0.064 and number of Neandertal haplotypes is 0.059 (S4 Table). Thus, variants in each of the species examined (humans, ancient and modern; Altai Neandertal, and Denisova) are assumed to represent a shared genus evolutionary trajectory rather than artifacts of inter-breeding.

What about the distribution of polymorphisms across different regions of the gene? Data are reported for the most significant areas in gene expression and regulation (the ORF, downstream promoter, and

upstream promoter) and non-coding regions (introns, UTRs). Modern humans were highly variable in each of these regions (with the least variants in the ORF and the most in the upstream promoter region); there were no polymorphisms in the ORF in any ancient sample, which reflects the conservation trend noted in primates (Table 2) and reflects the pattern seen in humans. Overall, both modern humans and Ust'-Ishim have higher numbers of SNPs in all gene regions, including higher SNP frequency in the *cis*-regulatory regions of the gene compared to extinct members of the genus—this echoes previous findings for humans in comparison to other primates (8) and might reflect a human specific pattern but the data are too limited to contribute to robust interpretation. A structural analysis of the distribution of SNPs via an Exact Test indicated no significant statistical differences among all four samples, between modern humans versus all three ancient samples, and between pooled humans versus extinct hominins (results not shown).

The distribution of polymorphisms across gene regions suggests plasticity in gene regulation and expression in response to viral loads, as noted in previous studies, and this trend is more notable in humans (both ancient and modern). The pattern of immune gene introgression, particularly regulatory haplotypes in the antiviral OAS gene cluster (45), has suggested that selective forces in our close relatives operated on expression, not protein variation—same as we see in non-human primate *CCR5* variation (6, 8)—and those adaptations were also useful to humans; thus, an increase in polymorphisms that allowed plasticity in regulation and expression in *CCR5* makes sense even if it is not due to introgression. Of course, without more paleo-genomes to compare, we cannot know if the variation in these genomes represents true species variation.

The expectation that extinct hominins and modern humans would share this pattern of increased variation in the regulatory areas of the gene is met in the current study and may be explained by genus specific shift in subsistence activities that altered human-environment interaction, bringing hominins into greater and regular contact with animal carcasses. By the Plio-Pleistocene roughly 2 million years ago,

early hominins were shifting from opportunistic non-confrontational scavenging to confrontational scavenging and, later, top predatory behaviors (57-62). Neandertals in Europe have also been shown to be active hunters and foragers (63). These human-environment interactions increase pathogen exposure and possible disease load as varied by ecology. Humans and European Neandertals would have shared similar ecological adaptive pressures in Europe—broad and varied—whereas Altai Neandertal (related to European Neandertals) and Denisova would have shared similar ecological adaptive pressures in Siberia with Ust'-Ishim—less varied. Key pathogens year-round in tropical to temperate zones are more likely to be viral (vector-borne) or bacterial (zoonotic) with transmission via interaction with the environment (64); high latitude pathogens year-round are more likely parasitic due to the reliance on marine mammals (65) and the short season for viral vector reproductive cycles to transmit infection from insects to hominins (43).

While previous studies have examined variation in *CCR5*, particularly *CCR5Δ32* which has a more recent origin (11, 18-22, 44), as a product of more recent human-disease interaction, the widespread pattern of increased variation in the gene across the genus Homo identified in this study suggests a deeper evolutionary adaptation. A key event distinguishing members of the genus *Homo* from the last common ancestor with Australopithecus was the shift to confrontational scavenging and, later, hunting; this alteration to human-environment interaction added a new point of disease contact as evidenced by modern data showing hunting bushmeat (which ancient hominins did too (61)) alters disease exposure via introduction of retroviruses and other pathogens (66-70). Given *CCR5*'s role in both innate and adaptive immune system functioning, it's plasticity may have provided an advantage to members of *Homo* across these varied disease ecologies and its potentially greater than normal interaction with the environment in foraging and hunting activities.

Acknowledgments

None.

References

1. Choe H, Farzan M, Sun Y, Sullivan N, Rollins B, Ponath PD, et al. The beta-chemokine receptors *CCR3* and *CCR5* facilitate infection by primary HIV-1 isolates. Cell. 1996;85(7):1135-48.
2. Dragic T, Litwin V, Allaway GP, Martin SR, Huang Y, Nagashima KA, et al. *HIV-1* entry into *CD4+* cells is mediated by the chemokine receptor *CC-CKR-5*. Nature. 1996;381:667-73.
3. Alkhatib G, Combadiere C, Broder CC, Feng Y, Kennedy PE, Murphy PM, et al. *CC CKR5*: a *RANTES, MIP-1alpha, MIP-1beta* receptor as a fusion cofactor for macrophage-tropic HIV-1. Science. 1996;272(5270):1955-8.
4. Doranz BJ, Rucker J, Yi Y, Smyth RJ, Samson M, Peiper SC, et al. A dual-tropic primary *HIV-1* isolate that uses fusin and the β-chemokine receptors *CKR-5, CKR-3,* and *CKR-2b* as fusion cofactors. Cell. 1996;85(7):1149-58.
5. Deng H, Liu R, Ellmeier W, Choe S, Unutmaz D, Burkhart M, et al. Identification of a major co-receptor for primary isolates of *HIV-1*. Nature. 1996;381:661-6.
6. Mummidi S, Ahuja SS, McDaniel BL, Ahuja SK. The human CC chemokine receptor 5 (*CCR5*) gene. Multiple transcripts with 5'-end heterogeneity, dual promoter usage, and evidence for polymorphisms within the regulatory regions and noncoding exons. J Biol Chem. 1997;272(49):30662-71.
7. Baggiolini M, Dewald B, Moser B. Interleukin-8 and related chemotactic cytokines--*CXC* and *CC* chemokines. Adv Immunol. 1994;55:97-179.
8. Mummidi S, Bamshad M, Ahuja SS, Gonzalez E, Feuillet PM, Begum K, et al. Evolution of human and non-human primate CC chemokine receptor 5 gene and mRNA. Potential roles for haplotype and mRNA diversity, differential haplotype-specific transcriptional activity, and altered transcription factor binding to polymorphic nucleotides in the pathogenesis of *HIV-1* and simian immunodeficiency virus. J Biol Chem. 2000;275(25):18946-61.
9. Zhang YW, Ryder OA, Zhang YP. Sequence evolution of the *CCR5* chemokine receptor gene in primates. Mol Biol Evol. 1999;16(9):1145-54.
10. Ribeiro IP, Schrago CG, Soares EA, Pissinatti A, Seuanez HN, Russo CAM, et al. *CCR5* chemokine receptor gene evolution in New World monkeys (*Platyrrini*, Primates): implication on resistance to lentiviruses. Infect Genet Evol. 2005;5:271-80.
11. Carrington M, Kissner T, Gerrard B, Ivanov S, O'Brien SJ, Dean M. Novel alleles of the chemokine-receptor gene *CCR5*. Am J Hum Genet. 1997;61(6):1261-7.
12. Winder IC, Devès MH, King GCP, Bailey GN, Inglis RH, Meredith-Williams M. Evolution and dispersal of the genus *Homo*: A landscape approach. J Hum Evol. 2015;87:48-65.
13. Abi-Rached L, Jobin MJ, Kulkarni S, McWhinnie A, Dalva K, Gragert L, et al. The shaping of modern human immune systems by multiregional admixture with archaic humans. Science. 2011;334(6052):89-94.
14. Mendez Fernando L, Watkins Joseph C, Hammer Michael F. A haplotype at *STAT2* introgressed from Neanderthals and serves as a candidate of positive selection in Papua New Guinea. Am J Hum Genet. 2012;91(2):265-74.

15. Mendez FL, Watkins JC, Hammer MF. Neandertal origin of genetic variation at the cluster of *OAS* immunity genes. Mol Biol Evol. 2013;30(4):798-801.
16. Sankararaman S, Mallick S, Dannemann M, Prufer K, Kelso J, Paabo S, et al. The genomic landscape of Neanderthal ancestry in present-day humans. Nature. 2014;507:354-7.
17. Galvani AP, Slatkin M. Evaluating plague and smallpox as historical selective pressures for the *CCR5-Δ32 HIV*-resistance allele. Proc Natl Acad Sci U S A. 2003;100(25):15276-9.
18. Stephens JC, Reich DE, Goldstein DB, Shin HD, Smith MW, Carrington M, et al. Dating the origin of the *CCR5-Δ32* AIDS-resistance allele by the coalescence of haplotypes. Am J Hum Genet. 1998;62(6):1507-15.
19. Mecsas J, Franklin G, Kuziel WA, Brubaker RR, Falkow S, Mosier DE. Evolutionary genetics: CCR5 mutation and plague protection. Nature. 2004;427:606.
20. Hummel S, Schmidt D, Kremeyer B, Herrmann B, Oppermann M. Detection of the *CCR5-Δ32* HIV resistance gene in Bronze Age skeletons. Genes Immun. 2005;6(4):371-4.
21. Limborska SA, Balanovsky OP, Balanovskaya EV, Slominsky PA, Schadrina MI, Livshits LA, et al. Analysis of *CCR5-Δ32* geographic distribution and its correlation with some climatic and geographic factors. Hum Hered. 2002;53(1):49-54.
22. Sabeti PC, Walsh E, Schaffner SF, Varilly P, Fry B, Hutcheson HB, et al. The case for selection at *CCR5-Δ32*. PLoS Biol. 2005;3(11):e378.
23. Zhang Y-W, Ryder OA, Zhang Y-P. Intra- and interspecific variation of the *CCR5* gene in higher primates. Mol Biol Evol. 2003;20(10):1722-9.
24. Allen SJ, Crown SE, Handel TM. Chemokine: receptor structure, interactions, and antagonism. Annu Rev Immunol. 2007;25(1):787-820.
25. Samson M, Labbe O, Mollereau C, Vassart G, Parmentier M. Molecular cloning and functional expression of a new human CC-chemokine receptor gene. Biochemistry. 1996;35(11):3362-7.
26. Wu L, LaRosa G, Kassam N, Gordon CJ, Heath H, Ruffing N, et al. Interaction of chemokine receptor *CCR5* with its ligands: multiple domains for *HIV-1 gp120* binding and a single domain for chemokine binding. J Exp Med. 1997;186(8):1373-81.
27. Garzino-Demo A, DeVico AL, Conant KE, Gallo RC. The role of chemokines in human immunodeficiency virus infection. Immunol Rev. 2000;177:79-87.
28. Farzan M, Choe H, Martin K, Marcon L, Hofmann W, Karlsson G, et al. Two orphan seven-transmembrane segment receptors which are expressed in *CD4*-positive cells support Simian Immunodeficiency Virus infection. J Exp Medicine. 1997;186(3):405.
29. Liao F, Alkhatib G, Peden KWC, Sharma G, Berger EA, Farber JM. *STRL33*, A novel chemokine receptor–like protein, functions as a fusion cofactor for both macrophage-tropic and T Cell line–tropic HIV-1. J Exp Med. 1997;185(11):2015.
30. Alkhatib G, Liao F, Berger EA, Farber JM, Peden KWC. A new SIV co-receptor, *STRL33*. Nature. 1997;388:238.
31. Deng H, Unutmaz D, KewalRamani VN, Littman DR. Expression cloning of new receptors used by simian and human immunodeficiency viruses. Nature. 1997;388:296-300.
32. Dean M, Carrington M, Winkler C, Huttley GA, Smith MW, Allikmets R, et al. Genetic restriction of *HIV-1* infection and progression to AIDS by a deletion allele of the *CKR5* structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study,

Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study. Science. 1996;273:1856-62.
33. Samson M, Libert F, Doranz BJ, Rucker J, Liesnard C, Farber CM, et al. Resistance to *HIV-1* infection in caucasian individuals bearing mutant alleles of the *CCR-5* chemokine receptor gene. Nature. 1996;382:722-5.
34. Temme S, Zacharias M, Neumann J, Wohlfromm S, König A, Temme N, et al. A novel family of human leukocyte antigen class II receptors may have its origin in archaic human species. J Biol Chem. 2014;289(2):639-53.
35. Huang Y, Paxton WA, Wolinsky SM, Neumann AU, Zhang L, He T, et al. The role of a mutant CCR5 allele in *HIV-1* transmission and disease progression. Nature Medicine. 1996;2(11):1240-3.
36. Zimmerman PA, Buckler-White A, Alkhatib G, Spalding T, Kubofcik J, Combadiere C, et al. Inherited resistance to *HIV-1* conferred by an inactivating mutation in CC chemokine receptor 5: studies in populations with contrasting clinical phenotypes, defined racial background, and quantified risk. Mol Med. 1997;3(1):23-36.
37. Eugen-Olsen J, Iversen AK, Garred P, Koppelhus U, Pedersen C, Benfield TL, et al. Heterozygosity for a deletion in the *CKR-5* gene leads to prolonged AIDS-free survival and slower *CD4* T-cell decline in a cohort of HIV-seropositive individuals. AIDS (London, England). 1997;11(3):305-10.
38. Glass WG, McDermott DH, Lim JK, Lekhong S, Yu SF, Frank WA, et al. *CCR5* deficiency increases risk of symptomatic West Nile Virus infection. J Exp Med. 2006;203(1):35-40.
39. Lim JK, Louie CY, Glaser C, Jean C, Johnson B, Johnson H, et al. Genetic deficiency of chemokine receptor *CCR5* is a strong risk factor for symptomatic West Nile Virus infection: a meta-analysis of 4 cohorts in the US epidemic. J Infec Dis. 2008;197(2):262-5.
40. Diamond MS, Klein RS. A genetic basis for human susceptibility of West Nile Virus. Trends Microbiol. 2006;14:287-89.
41. Novembre J, Galvani AP, Slatkin M. The geographic spread of the CCR5 Delta32 HIV-resistance allele. PLoS Biol. 2005;3(11):e339.
42. Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, Auton A, et al. Genes mirror geography within Europe. Nature. 2008;456:98-101.
43. Hoover KC, Barker CM. West Nile virus, climate change, and circumpolar vulnerability. WIREs Clim Change. 2016;7(2):283-300.
44. Galvani AP, Slatkin M. Evaluating plague and smallpox as historical selective pressures for the CCR5-Delta 32 HIV-resistance allele. Proc Natl Acad Sci U S A. 2003;100(25):15276-9.
45. Gittelman Rachel M, Schraiber Joshua G, Vernot B, Mikacenic C, Wurfel Mark M, Akey Joshua M. Archaic hominin admixture facilitated adaptation to Out-of-Africa environments. Curr Biol. 2016;26(24):3375-82.
46. Pimenoff VN, Mendes de Oliveira C, Bravo IG. Transmission between archaic and modern human ancestors during the evolution of the oncogenic human papillomavirus 16. Mol Biol Evol. 2016.

47. Vernot B, Tucci S, Kelso J, Schraiber JG, Wolf AB, Gittelman RM, et al. Excavating Neandertal and Denisovan DNA from the genomes of Melanesian individuals. Science. 2016;352(6282):235-9.
48. Racimo F, Sankararaman S, Nielsen R, Huerta-Sanchez E. Evidence for archaic adaptive introgression in humans. Nat Rev Genet. 2015;16(6):359-71.
49. Hu Y, Ding Q, Wang Y, Xu S, He Y, Wang M, et al. Investigating the Evolutionary Importance of Denisovan Introgressions in Papua New Guineans and Australians. bioRxiv. 2015; doi: https://doi.org/10.1101/022632.
50. Consortium TGP. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012;491:56-65.
51. Krause J, Qiaomei F, Good JM, Viola B, Shunkov MV, Derevianko AP, et al. The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. Nature. 2010;464:894-7.
52. Prufer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, et al. The complete genome sequence of a Neanderthal from the Altai Mountains. Nature. 2014;505(7481):43-9.
53. Fu Q, Li H, Moorjani P, Jay F, Slepchenko SM, Bondarev AA, et al. Genome sequence of a 45,000-year-old modern human from western Siberia. Nature. 2014;514(7523):445-9.
54. Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, et al. Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature. 2010;468(7327):1053-60.
55. Dabney J, Meyer M, Pääbo S. Ancient DNA damage. Cold Spring Harb Perspect Biol. 2013;5(7):a012567.
56. Briggs AW, Stenzel U, Meyer M, Krause J, Kircher M, Pääbo S. Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. Nucleic Acids Res. 2010;38(6):e87-e.
57. Brantingham PJ. Hominid–carnivore coevolution and invasion of the predatory guild. J Anthropol Archaeol. 1998;17(4):327-53.
58. de Heinzelin J, Clark JD, White T, Hart W, Renne P, WoldeGabriel G, et al. Environment and behavior of 2.5-million-year-old Bouri hominids. Science. 1999;284(5414):625-9.
59. Domínguez-Rodrigo M. Hunting and ccavenging by early humans: the state of the debate. J World Prehist. 2002;16(1):1-54.
60. Domínguez-Rodrigo M, Pickering TR. Early hominid hunting and scavenging: a zooarcheological review. Evol Anthropol. 2003;12(6):275-82.
61. Shipman P, Bosler W, Davis KL, Behrensmeyer AK, Dunbar RIM, Groves CP, et al. Butchering of giant geladas at an Acheulian site. Curr Anthropol. 1981;22(3):257-68.
62. Moleón M, Sánchez-Zapata JA, Margalida A, Carrete M, Owen-Smith N, Donázar JA. Humans and Scavengers: The Evolution of Interactions and Ecosystem Services. BioScience. 2014;64(5):394-403.
63. Richards MP, Pettitt PB, Trinkaus E, Smith FH, Paunović M, Karavanić I. Neanderthal diet at Vindija and Neanderthal predation: The evidence from stable isotopes. Proc Natl Acad Sci U S A. 2000;97(13):7663-6.
64. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, et al. Global trends in emerging infectious diseases. Nature. 2008;451(7181):990-3.

65. Dudley JP, Hoberg EP, Jenkins EJ, Parkinson AJ. Climate change in the North American Arctic: a one health perspective. EcoHealth. 2015;12(4):713-25.
66. Steve A-M, Ahidjo A, Placide M-K, Caroline F, Mukulumanya M, Simon-Pierre N-K, et al. High prevalences and a wide genetic diversity of simian retroviruses in non-human primate bushmeat in rural areas of the Democratic Republic of Congo. EcoHealth. 2017;14(1):100-14.
67. van Vliet N, Moreno J, Gomez J, Zhou W, Fa JE, Golden C, et al. Bushmeat and human health: assessing the evidence in tropical and sub-tropical forests. Ethnobiol Conserv 2017;6.
68. Kurpiers LA, Schulte-Herbrüggen B, Ejotre I, Reeder DM. Bushmeat and emerging infectious diseases: lessons from Africa. In: Angelici FM, editor. Problematic Wildlife: A Cross-Disciplinary Approach. Cham, Switzerland: Springer International Publishing; 2016. p. 507-51.
69. Mossoun A, Calvignac-Spencer S, Anoh AE, Pauly MS, Driscoll DA, Michel AO, et al. Bushmeat hunting and zoonotic transmission of Simian T-Lymphotropic Virus 1 in tropical West and Central Africa. J Virol. 2017;91(10):e 02479-16.
70. Chen M, Reed RR, Lane AP. Acute inflammation regulates neuroregeneration through the *NF-κB* pathway in olfactory epithelium. Proc Natl Acad Sci U S A. 2017;114(30):8089-94.

Supporting Information

S1 Table: Bioinformatics Pipeline

S2 Table: 1000 Genomes Samples

S3 Table: Polymorphisms in 1000 Genomes Sample

S4 Table: Altai Neandertal Introgression Data