Patterns of Diversity in HIV-Related Loci among Subspecies of Chimpanzee: Concordance at CCR5 and Differences at CXCR4 and CX3CRI

T.S. MacFie,*† E. Nerrienet,* N.G. de Groot,** R.E. Bontrop,**† and N.I. Mundy*

*Department of Zoology, University of Cambridge, Cambridge, UK; †Centre Pasteur du Cameroun, Yaoundé, Cameroon; and **Department of Comparative Genetics and Refined, Biomedical Primate Research Center, Rijswijk, Netherlands

Human immunodeficiency virus type 1 (HIV-1) arose in humans via zoonotic transmissions of simian immunodeficiency viruses (SIVcpz) from common chimpanzees, Pan troglodytes. Despite the close relatedness of the two viruses and their hosts, we do not yet understand what causes SIVcpz to be nonpathogenic in chimpanzees, and HIV/AIDS to be one of the most devastating infectious diseases to have emerged in humans. There have been a number of genes identified in humans that confer disease resistance/susceptibility toward HIV-1, but little is known about the evolution and diversity of most of these chemokine receptor genes in chimpanzees. Here we show that genetic variation in chimpanzees differs across the various loci related to HIV-1, and that the pattern of variation differs among the chimpanzee subspecies. For all three subspecies, low diversity at CCR5 is confined to a small area of chromosome 3, suggesting that a selective sweep at this locus may have predated sub-speciation. In contrast, diversity and neutrality tests suggest differing evolutionary forces among subspecies at CXCR4 and CX3CRI, with directional selection (in Pan troglodytes vellerosus) and demographic expansion (Pan troglodytes troglodytes) offering the most likely scenarios. These are some of the first data demonstrating differentiation in functional loci among chimpanzee subspecies.

Introduction

Co-evolutionary arms races between hosts and pathogens are expected to leave strong signatures of selection in their respective genomes, and many examples have been documented in humans (e.g., Livingstone 1985; Hill et al. 1991; Morral et al. 1994; Ruwende et al. 1995; Covacci et al. 1999). The recently emerged virus HIV-1 (human immunodeficiency virus type 1) in humans is the major causative agent of AIDS and originated from simian immunodeficiency viruses (SIVs) infecting chimpanzees (SIVcpz) (reviewed in Heeney et al. 2006). The HIV-1 virus and SIVcpz virus were found to be identical in genomic organization (Huet et al. 1990), and HIV-1 viral strains were subsequently shown to be phylogenetically closer to SIVcpz than to HIV-2 or any other SIV strain (Gao et al. 1999).

Without treatment, HIV-1 infection of a human host eventually leads to immunological failure and an array of clinical manifestations. In contrast to the severe immunodeficiency caused by HIV-1 in humans, naturally occurring SIVcpz infection in chimpanzees does not appear to lead to immunodeficiency or AIDS-like disease, and this resistance to disease may have arisen during a period of co-evolution (Gordon et al. 2005; VandeWoude and Apetrei 2006; Silvestri et al. 2007).

Despite both the hosts and viruses being closely related, the basis for the difference in pathogenicity of SIVcpz in chimpanzees and HIV-1 in humans is poorly understood. We know from several studies of SIVcpz isolates that, like HIV-1, they also use CD4 and CCR5 as their main receptor and coreceptor for entry into target cells (Muller-Trutwin et al. 2000; Onda et al. 2001; Bibollet-Ruche et al. 2004). Other parallels between pathogenic HIV-1 and non-progressive SIVcpz within the host include equivalent high levels of viral replication during primary infection, a similar degree of CD4+ T-cell depletion in the blood during acute infection, a cellular immune response of the same magnitude, and analogous in vivo biology of the viruses (Pandrea et al. 2008). Beyond these similarities between pathogenic HIV-1 and nonprogressive SIVcpz, however, there are also many differences that lead to the widely differing pathogenicities. For example, natural SIVcpz hosts preserve the functionality of various immune cells; they are able to expand coreceptor use without inducing disease progression; they have normal levels of immune activation, apoptosis, and cell proliferation during chronic infection; are able to establish a rapid anti-inflammatory milieu to prevent chronic aberrant immune activation (Pandrea et al. 2008), all of which contribute to a lack of disease progression in natural hosts.

There is strong evidence for genetic variation in susceptibility to HIV-1 in humans (reviewed in Arenzana-Seisdedos and Parmentier 2006). As in chimpanzees, there is evidence that some humans remain asymptomatic even after being repeatedly exposed to HIV. Furthermore, some HIV-positive individuals exhibit mild or no significant immunodeficiency for relatively long periods without developing AIDS (Easterbrook 1994). Comparison of human long-term nonprogressors and progressors revealed that multiple genetic factors such as chemokine genes, chemokine receptor genes, and the Major Histocompatibility Complex (MHC) influence the rate of disease progression (Rowland-Jones et al. 2001).

HIV exploits cell surface receptors to attach and gain entry into cells. Infection of human (and nonhuman pri-mate) cells by HIV (or SIV) requires not only the presence of the main virus receptor CD4 but also a coreceptor that can be one of a variety of chemokine receptors (Dragic et al. 1996; Feng et al. 1996; Chen et al. 1997; Moore et al. 1997; Kristiansen et al. 1998; Owen et al. 2000). Natural ligands of these chemokine receptors compete with HIV/SIV in binding to the receptor, and clinical studies in humans have
demonstrated that mutations involving chemokine receptors (CCR5, CCR2, CXCR4, and CX3CR1) or their natural ligands (SDF1 and RANTES) influence the susceptibility to HIV infection and/or the rate of progression toward AIDS (Samson et al. 1996; Smith et al. 1997; Winkler et al. 1998; Faure et al. 2000; McDermott et al. 2000).

If a period of coevolution led to genetic resistance to SIVcpz-induced infection/disease, this may be observable in the chimpanzee genome. Based on mitochondrial DNA (mtDNA) variation, common chimpanzees are currently classified into four subspecies: the western chimpanzee, Pan troglodytes verus; the Nigerian/NE Cameroonian chimpanzee, Pan troglodytes vellerosus; the western equatorial chimpanzee, Pan troglodytes troglodytes; and the eastern chimpanzee, Pan troglodytes schweinfurthii (Groves 1993; Morin et al. 1994; Gonder et al. 1997, 2006) (fig. 1).

To date, SIVcpz has been isolated as a naturally occurring infection from only two subspecies, Pan troglodytes and Pan troglodytes schweinfurthii. Estimates of SIVcpz prevalence rates in wild populations of Pan troglodytes vary from 0% to 35% (Keele et al. 2006) and those of Pan troglodytes schweinfurthii from 0% to 30% (Santiago et al. 2003). SIVcpz infection has yet to be detected as a naturally occurring infection in Pan troglodytes verus and Pan troglodytes vellerosus despite screening over 1,500 captive and 28 wild-living Pan troglodytes individuals and 50 captive Pan troglodytes (Prince et al. 2002; Santiago et al. 2002; Switzer et al. 2005). Previous studies of genetic variation in chimpanzees have largely concentrated on neutral markers, including mtDNA (Morin et al. 1994), nuclear microsatellites (Becquet et al. 2007), and noncoding autosomal (Fischer et al. 2004) and sex-linked regions (Kaessman et al. 1999). However, one study examined variation in the 5’ untranslated region (UTR) of CCR5 concluding that there was evidence for a selective sweep (Wooding et al. 2005).

In this study, we assay genetic variation in three chimpanzee subspecies (Pan troglodytes, Pan troglodytes verus, and Pan troglodytes vellerosus) at six loci related to HIV infection in humans (CCR5, CCR2, CXCR4, CX3CR1, SDF, and RANTES) as well as a series of control loci in order to address the following questions: Is genetic variation evenly distributed among loci? Are patterns of variation across loci similar among the three subspecies? Do any loci show evidence of natural selection, such as selective sweeps or balancing selection? Are the variants known to affect human resistance/susceptibility present in chimpanzees?

Materials and Methods

Samples

Blood samples were obtained from 36 wild-born orphaned chimpanzees of unknown geographic origin within Cameroon. Genomic DNA was extracted using standard procedures. DNA samples were also obtained from 19 Pan troglodytes (from Sierra Leone) and 2 Pan troglodytes (unknown origin) previously held at the Biomedical Primate Research Centre in the Netherlands.

Subspecies Assignment

A combination of mtDNA and nuclear DNA (Gonder et al. 1997; MacFie et al. submitted) was used to assign the Cameroon chimpanzees to subspecies. Phylogenetic analyses of HVRI sequences, using reference sequences from Gonder et al. (1997), provided preliminary subspecies assignment. Principal component analyses of genotypic data from >650 autosomal single nucleotide polymorphisms in these samples led to highly discrete subspecies clusters and reassignment of two chimpanzees with Pan troglodytes-like mtDNA to Pan troglodytes vellerosus (MacFie et al. submitted). This resulted in total sample sizes of 21 Pan troglodytes, 17 Pan troglodytes verus, and 19 Pan troglodytes vellerosus.
Polymerase Chain Reaction (PCR) and Sequencing

Genomic DNA was amplified (GenomiPhi, GE Healthcare, Uppsala, Sweden) from all samples prior to PCR. In addition to the six candidate loci related to HIV (CCR5, CCR2, CX3CR1, CXCR4, SDF, and RANTES), some control genes were selected for comparison with candidate genes. Because three of the loci (CCR5, CCR2, and CX3CR1) are located on the same chromosome (chimpanzee chromosome 3) and because of findings of low variation at CCR5, a series of loci (SEC22L3, ZNF445, CCRL2, and PTPN23) were selected at varying physical distances from the focal loci to investigate linkage disequilibrium and average levels of variation on this chromosome. The most recent chimpanzee genome assembly (2.1) shows the following arrangement: CX3CR1—3.35 Mb–SEC22L3—1.84 Mb–ZNF445—1.94 Mb–CCR2—0.01 Mb–CCR5—0.03 Mb–CCRL2—0.99 Mb–PTPN23. Thus, CCRL2 is useful for assaying close physical linkage to CCR2 and CCR5, whereas the other loci assay linkage at greater distances from the three focal loci (CCR5, CCR2, and CX3CR1) and give the general level of variation on chromosome 3. All loci studied had easily identifiable one-to-one orthologues in the chimpanzee genome. The mtDNA HVR1 region was chosen to assist in assigning individuals to subspecies and to enable comparisons of genetic diversity with other published studies.

For this broad-scale survey, we chose to study ~1-kb fragments of each autosomal locus for straightforward analysis by PCR and direct sequencing, as well as a 534-bp segment of HVR1 that has been commonly used in previous studies. For the target loci, we amplified those regions of the gene where the major human HIV susceptibility variants lie: 1,054 bp of coding sequence of CCR5 (exon 3), 926 bp of 3' untranslated region of SDF, 1,081 bp of coding sequence of CXCR4 (exon 2), 1,193 bp of coding sequence of CX3CR1 (exon 2), 1,096 bp of coding sequence of CCR2 (exon 2), and 988 bp of RANTES promoter. As we sequenced ~1-kb coding regions for the major region of focal loci, we studied ~1-kb coding regions for the control loci where the intron–exon structure permitted. We sequenced 1,077 bp of coding sequence of ZNF445 (exon 6), 1,016 bp of coding region of PTPN23 (exon 20), 1,035 bp of coding region of CCRL2 (exon 2), and 545 bp of coding sequence and 472 bp of noncoding sequence of SEC22L3 (the largest exon, exon 7, and intron 6).

Primers (see supplementary table 1, Supplementary Material online) were designed using human and chimpanzee sequences available at GenBank. HVR1 primers were chosen to ensure nuclear inserts of mtDNA were not amplified (Thalmann et al. 2004). PCR reaction mixtures (25 μl) contained 0.5 U of BioTaq DNA polymerase, 0.4 μM each primer, 0.1 mM each deoxynucleotide triphosphate, 1.5 mM MgCl2, 1× reaction buffer, and 50 ng template DNA. PCR reactions were carried out in an MJ Research thermocycler (PTC-200) with the following cycling parameters: an initial 94 °C denaturation for 2 min and a 30–35 cycle amplification profile (94 °C denaturation for 30 s, 56–69 °C annealing for 45 s, 72 °C extension for 30 s to 1.5 min) followed by a final 72 °C extension for 5 min. PCR products were visualized on a 1% agarose gel under UV light. Successful PCR products were purified with the QIAquick PCR purification Kit (Qiagen, Valencia, CA) prior to direct sequencing on both strands. Sequencing primers were designed to anneal approximately every 400 bp in both forward and reverse orientations. PCR products from the RANTES locus had length variation and hence were cloned using TOPO TA Cloning Kit (Version 1) (Invitrogen Life Technologies, Foster City, CA) and purified with QIAprep spin miniprep kit (Qiagen) prior to sequencing. Each 10 μl sequencing reaction consisted of: 10 mM primer, approximately 100 ng purified PCR product or plasmid, 1× sequencing buffer (200 mM Tris HCl pH 9.0, 10 mM MgCl2), and 1.0 μl BigDye V3.1 (ABI PRISM, PE Applied Biosystems, Foster City, CA). Sequences were run on an ABI Prism 3700 (PerkinElmer Biosystems) and edited using SEQUEN (DNAStar, Inc) and aligned using MegAlign (DNAstar, Inc, Madison, WI). Sequences have been submitted to Genbank (Accession nos. FJ642022–FJ642645).

Population Genetic Analyses

Haplotypes were inferred using PHASE (version 2.1) (Stephens et al. 2001; Stephens and Scheet 2005), and all inferred haplotypes had a high level of statistical support (all with probabilities >0.95). ARLEQUIN (Schneider et al. 2000) was used to determine standard nucleotide diversity (θ), Tajima’s D (Tajima 1989), and Fu’s Fs (Fu 1997) statistics and perform analysis of molecular variance (ΦST) analyses. Haplotype networks were constructed using NETWORK (Bandelt et al. 1999). Nucleotide diversity–divergence ratios among chimpanzees and humans were calculated against a reference human sequence.

Results

We analyzed six loci associated with HIV-1/AIDS susceptibility in humans (CCR5, CCR2, CXCR4, CX3CR1, SDF, and RANTES) in 21 P. t. troglodytes, 17 P. t. vellerosus, and 19 P. t. verus. We found variable patterns of nucleotide diversity and haplotype diversity among the different loci both within and between subspecies (table 1, fig. 2). Pan troglodytes troglodytes showed no nucleotide variation at CCR5 and moderate variation at the other loci, P. t. vellerosus exhibited low nucleotide diversity at CCR5 (0.027%) and CXCR4 (0.005%), whereas P. t. verus had moderate nucleotide diversity in SDF and RANTES but no nucleotide variation at CCR5, CXCR4, and CX3CR1. At CCR2, all subspecies possessed similar levels of moderate diversity. Patterns of haplotype diversity generally mirror nucleotide diversity, with the greatest difference among subspecies occurring at CX3CR1 where P. t. verus has a single haplotype and P. t. troglodytes has 12 haplotypes. The nucleotide diversity in the coding sequence of CCR5 in the three subspecies of chimpanzees is strikingly lower than that for the same region in a worldwide sample of humans (0.11%) (Ansari-Lari et al. 1997) especially because variation at chimpanzee nuclear loci is generally 2- to 3-fold higher than that in humans (Wise et al. 1997). None of the specific mutations conferring AIDS resistance in
Table 1
Nucleotide Diversity (% ±SE), Number of Haplotypes, and ΦST

<table>
<thead>
<tr>
<th>Loci</th>
<th>Pan troglodytes* (n = 57)</th>
<th>P. troglodytes (n = 21)</th>
<th>P. t. vellerosus (n = 17)</th>
<th>P. t. verus (n = 19)</th>
<th>ΦST</th>
</tr>
</thead>
<tbody>
<tr>
<td>CX2R1</td>
<td>0.029 ± 0.033</td>
<td>15</td>
<td>0.047 ± 0.045</td>
<td>12</td>
<td>0.005 ± 0.014</td>
</tr>
<tr>
<td>SEC22L3</td>
<td>0.108 ± 0.079</td>
<td>8</td>
<td>0.115 ± 0.084</td>
<td>5</td>
<td>0.090 ± 0.071</td>
</tr>
<tr>
<td>ZNF445</td>
<td>0.015 ± 0.024</td>
<td>5</td>
<td>0.041 ± 0.041</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>CCR2</td>
<td>0.092 ± 0.069</td>
<td>10</td>
<td>0.074 ± 0.061</td>
<td>4</td>
<td>0.098 ± 0.074</td>
</tr>
<tr>
<td>CCR3</td>
<td>0.009 ± 0.069</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0.027 ± 0.031</td>
</tr>
<tr>
<td>CCR2L2</td>
<td>0.175 ± 0.058</td>
<td>15</td>
<td>0.133 ± 0.089</td>
<td>9</td>
<td>0.089 ± 0.056</td>
</tr>
<tr>
<td>PTNP23</td>
<td>0.183 ± 0.120</td>
<td>14</td>
<td>0.205 ± 0.133</td>
<td>12</td>
<td>0.143 ± 0.102</td>
</tr>
<tr>
<td>CXC8R4</td>
<td>0.227 ± 0.118</td>
<td>3</td>
<td>0.105 ± 0.078</td>
<td>2</td>
<td>0.155 ± 0.104</td>
</tr>
<tr>
<td>RANTES</td>
<td>0.075 ± 0.062</td>
<td>11</td>
<td>0.047 ± 0.046</td>
<td>8</td>
<td>0.083 ± 0.068</td>
</tr>
<tr>
<td>SDF</td>
<td>0.065 ± 0.057</td>
<td>13</td>
<td>0.077 ± 0.065</td>
<td>9</td>
<td>0.076 ± 0.065</td>
</tr>
<tr>
<td>HVR1</td>
<td>0.719 ± 3.041</td>
<td>3.646 ± 1.887</td>
<td>3.142 ± 1.783</td>
<td>4.380 ± 2.252</td>
<td>0.547**</td>
</tr>
</tbody>
</table>

* Value calculated using all chimpanzee samples.
P < 0.05, **P < 0.01 significantly different from 0.

humans were present in the chimpanzee data (CCR5-A32; CCR2-V64I; CX3CR1-249I, T280M; SDF1-G801A; RANTES, G-403A, C-28G), confirming that these variants are all specifically derived on the human lineage. Due to low genetic variation observed at CCR5 in all three subspecies, further investigation of the surrounding region on chromosome three was undertaken in order to examine potential hitchhiking effects. The following loci were chosen: SEC22L3 (3.79 Mb distal to CCR5), ZNF445 (1.95 Mb distal to CCR5), and CCRL2 (0.03 Mb proximal to CCR5), and PTNP23 (1.02 Mb proximal to CCR5). Overall, SEC22L3 shows a fairly high level of variation, with P. t. troglodytes having the greatest amount (0.115%), and P. t. vellerosus and P. t. verus exhibiting similar levels (0.090% and 0.072%, respectively). ZNF445 is conspicuous among the control loci for containing very similar levels (0.090% and 0.073%, respectively), and is tightly confined on chromosome three was undertaken in order to examine potential hitchhiking effects. The following loci were chosen: SEC22L3 (3.79 Mb distal to CCR5), ZNF445 (1.95 Mb distal to CCR5), and CCRL2 (0.03 Mb proximal to CCR5), and PTNP23 (1.02 Mb proximal to CCR5). Overall, SEC22L3 shows a fairly high level of variation, with P. t. troglodytes having the greatest amount (0.115%), and P. t. vellerosus and P. t. verus exhibiting similar levels (0.090% and 0.072%, respectively). ZNF445 is conspicuous among the control loci for containing very little variation—P. t. troglodytes has a nucleotide diversity of 0.041%, whereas P. t. vellerosus and P. t. verus do not contain any variation. At PTNP23, the levels of nucleotide diversity are high in P. t. troglodytes (0.205%), relatively high in P. t. vellerosus (0.143%), and intermediate in P. t. verus (0.050%). Overall levels of diversity are high at CCRL2 with P. t. verus having the largest amount of variation (0.201%), followed by P. t. vellerosus (0.155%) then P. t. troglodytes (0.105%). In summary, the most notable feature is that the low variation at CCR5 is tightly confined on chromosome 3 and does not extend more than 100 kb distally (CCR2) or 300 kb proximally (CCRL2).

We analyzed nucleotide variation at mtDNA (HVR1 region of mtDNA) to assess whether our subspecies sampling was representative, because this is the region most frequently studied in chimpanzee populations (table 1). MtDNA variation in samples of all subspecies was substantial. Notably, the P. t. verus sample possesses higher nucleotide diversity at HVR1 (4.38%) than P. t. vellerosus (3.41%) and P. t. troglodytes (3.65%). The results are somewhat lower than estimates of nucleotide diversity of the same region in previous studies (e.g., 5.59% for P. t. troglodytes and 5.30% in P. t. verus; Deinard and Kidd, 2000). There are no previous estimates of mitochondrial diversity in P. t. vellerosus.

In order to make more meaningful comparisons in levels of nucleotide diversity among loci, we calculated the ratio of nucleotide diversity in chimpanzees at a locus to the chimpanzee–human genetic distance, which controls for differences in mutation rate among loci (fig. 2b). These diversity–divergence ratios enhance the pattern seen in the raw nucleotide diversity data. In particular, the ratios are low at CCR5 in all subspecies, and this low variation is the chimpanzee–human genetic distance, which controls for differences in mutation rate among loci (fig. 2b). These diversity–divergence ratios enhance the pattern seen in the raw nucleotide diversity data. In particular, the ratios are low at CCR5 in all subspecies, and this low variation is
localized to a 0.04-Mb segment of chromosome three (not found in CCR2 or CCRL2). Furthermore, there are large differences between P. t. troglodytes and P. t. verus at CX3CR1, with P. t. vellerosus intermediate. The six lowest diversity–divergence ratios (CCR5 in all subspecies, CCR4 in P. t. vellerosus and P. t. verus, and CX3CR1 in P. t. verus) (range 0–0.034) are considerably lower than average ratios at SDF (0.094) and RANTES (0.136) and those calculated from other nuclear genes in P. t. troglodytes (range 0.063–0.379, average 0.139).

Tajima’s D (Tajima 1989) and Fu’s Fs (Fu 1997) statistics were used to test for deviations from the standard neutral model (table 2). For both statistics, there was a fairly consistent pattern across loci of negative values for P. t. troglodytes, with one locus (RANTES) showing a significantly negative value for Tajima’s D, whereas 6 out of 10 loci for which the test could be applied showed a significantly negative Fu’s Fs. Population expansion and/or finescale population subdivision are the most likely explanations for the results in P. t. troglodytes (Fischer et al. 2004). In contrast, there was no consistent pattern across loci for the neutrality statistics in P. t. vellerosus or P. t. verus, suggesting an absence of population expansion. The only significant results are a negative value of Fu’s Fs for HVR1 in P. t. verus and, interestingly, negative values of Fu’s Fs for CX3CR1 and CCR4 in P. t. vellerosus.

Reconstructed haplotype networks for CCR5, CCR4, and CX3CR1 (fig. 3) show an absence of high frequency–derived haplotypes. The network for CX3CR1 shows high differentiation between P. t. verus and the other species, and this is also reflected in the high Φst of 0.403 at this locus (table 1).

The numbers of synonymous (S) and nonsynonymous (NS) substitutions in the CCR5, CCR4, and CX3CR1 data are as follows (P. t. verus is fixed for all of these loci). CCR5: P. t. troglodytes—0S/0NS, P. t. vellerosus—0S/1NS, P. t. troglodytes—0S/1NS. CCR4: P. t. troglodytes—1S/0 NS, P. t. vellerosus—0S/1NS, P. t. troglodytes—1S/1NS. CX3CR1: P. t. troglodytes—5S/2NS, P. t. vellerosus—5S/2NS, P. t. troglodytes—7S/3NS. None of the haplotypes defined by nonsynonymous substitutions are at high frequency, and overall there is little indication that the coding sequences of these loci are under positive selection in any of the subspecies.

### Discussion

We have documented variation in diversity of AIDS-related loci in chimpanzees, and found some similarities as well as differences in the patterns of diversity among chimpanzee subspecies. In particular, whereas levels of variation were substantial at CCR2, SDF, and RANTES and similar among subspecies, we found very low or zero nucleotide and haplotype diversity at CCR5 among all three subspecies, CCR4 in P. t. vellerosus and P. t. verus, and CX3CR1 in P. t. verus. Overall low genetic variation in the chimpanzee samples studied can be ruled out as a major contributor to the results, because the control locus (HVR1) and several nuclear loci showed substantial variation in all three subspecies, similar to that found in previous studies including regions sampled at HOXB6, DRD4, ND2, NRY, and XQ13.3 (Wise et al. 1997; Deinard and Kidd 1999; Kaessman et al. 1999; Stone et al. 2002). Low locus-specific mutation rates could also be excluded as a cause of low variation, as similar patterns were found when mutation rates were controlled for by comparing diversity values with human–chimpanzee nucleotide divergence (fig. 2B). However, we cannot exclude a contribution of sampling effects to the pattern of nucleotide diversity seen among subspecies, as evident in the size of the standard errors around some of the nucleotide diversity estimates.

Apart from sampling effects, other potential explanations for low diversity are demographic factors or directional selection. To examine the possibility of directional selection at CCR5 affecting variation at linked sites, we examined variation at linked loci along chromosome 3, and the results showed that low variation was tightly centered at CCR5 itself in all three subspecies, because the flanking loci, which are only 0.1 Mb (CCR2) and 0.3 Mb away (CCRL2) showed normal variation in all subspecies. Neutrality tests (Tajima’s D and Fu’s Fs) at CCR5 were not very informative because of the absence of variation in P. t. troglodytes and P. t. verus. Evidence for directional selection in the 5’ UTR of CCR5 was previously obtained in a sample of chimpanzees consisting mostly of P. t. verus with some P. t. troglodytes (Woody et al. 2005). The present results suggest that a sweep at CCR5 may be present in all three subspecies studied here. The most parsimonious

### Table 2

<table>
<thead>
<tr>
<th>Neutrality Tests</th>
<th>Tajima’s D</th>
<th>Fu’s Fs</th>
<th>P. t. troglodytes</th>
<th>P. t. vellerosus</th>
<th>P. t. verus</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CX3CR1</td>
<td>0.705</td>
<td>1.201</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SEC22L3</td>
<td>0.029</td>
<td>0.582</td>
<td>0.053</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ZNF445</td>
<td>1.245</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CCR2</td>
<td>0.365</td>
<td>1.057</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CCR5</td>
<td>0</td>
<td>0.272</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CCR2</td>
<td>0.164</td>
<td>0.820</td>
<td>0.268</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PTPN23</td>
<td>0.390</td>
<td>0.169</td>
<td>0.100</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CX3CR4</td>
<td>1.667</td>
<td>1.137</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RANTES</td>
<td>0.170</td>
<td>0.410</td>
<td>0.568</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SDF</td>
<td>1.327</td>
<td>0.682</td>
<td>0.227</td>
<td></td>
</tr>
<tr>
<td></td>
<td>HVR1</td>
<td>0.144</td>
<td>0</td>
<td>1.063</td>
<td></td>
</tr>
</tbody>
</table>

* P < 0.05 value significantly different from 0; **P < 0.01 value significantly different from 0.
explanation for these findings is that a selective sweep at CCR5 predated the separation of the subspecies and that there is a high local recombination rate in this region, isolating the effects of the sweep to a small segment of chromosome 3. Alternatively, there could have been separate sweeps in different lineages.

As well as CCR5, the data also provide suggestive evidence for directional selection acting at CXCR4 and CX3CR1. In P.t. vellerosus, CXCR4 showed low nucleotide diversity, whereas both CXCR4 and CX3CR1 had significantly negative Fu’s Fs statistics. Common explanations for a negative Fu’s Fs are population expansion and directional selection. There is little evidence for population expansion in the P.t. vellerosus data, as of 10 loci for which the statistics could be calculated, similar numbers of loci gave positive and negative values for both Tajima’s D and Fu’s Fs (five positive vs. five negative for Tajima’s D and four positive vs. six negative for Fu’s Fs) with CXCR4 and CX3CR1 being the only two loci showing significant values. These results are therefore most consistent with locus-specific directional selection at CXCR4 and CX3CR1 in P.t. vellerosus. For P.t. verus, the interpretation is more complicated. The absence of variation at CXCR4 and CX3CR1 provides no information for the neutrality tests, and although mitochondrial (HVR1) variation is substantial, there is a possibility that this sample of P.t. verus has generally low variation at autosomal loci, as many loci show low nucleotide variation. Finally, it is interesting that CXCR4 and CX3CR1 have two of the three highest $\Phi_{st}$ values in the data set, which is again consistent with directional selection, occurring independently in the subspecies. In summary, there are several pieces of evidence suggesting that directional selection is acting on these two loci in one or more chimpanzee subspecies, and it would be

Fig. 3.—Median-spanning networks for the three candidate loci with low diversity: (A) CCR5, (B) CX3CR1, and (C) CXCR4. Numbers along branches refer to specific mutations (numbered from 5’–3’).
interesting to investigate this further, for example, by examining variation around these loci in more detail.

Are the potential selective sweeps at HIV-related loci in wild chimpanzee populations related to coevolution with SIV<sub>cpz</sub>? For <i>P. t. troglodytes</i>, selection at CCR5 is consistent with naturally occurring infection by SIV<sub>cpz</sub>, and a plausible mechanism is that the 5’ UTR provides disease resistance, because the common haplotype in chimpanzees is the same as the human haplotype showing the lowest promoter activity and greatest resistance to AIDS progression in humans (Gonzalez et al. 1999; Mummidi et al. 2000). However, evidence of selection is also present at CCR5 in <i>P. t. verus</i> and <i>P. t. vellerosus</i>, in which SIV<sub>cpz</sub> has not been documented despite reasonably extensive sampling. One possibility is that SIV<sub>cpz</sub> infection in these subspecies occurred in the past and is currently extremely rare or absent (de Groot et al. 2002; Wooding et al. 2005), or, as suggested by the similar pattern of variation around CCR5 in the three subspecies, there was infection of the ancestral chimpanzee lineage with SIV<sub>cpz</sub>. For the better studied <i>P. t. verus</i>, it is interesting to note there is also low genomic diversity at MHC (de Groot et al. 2002), with the lowest variation at MHC Class I loci and MHC Class I chain–related loci (MIC), which are involved in resistance to intracellular pathogens including viruses (de Groot et al. 2005, 2008). More recently, the CD4 receptor was shown to be conserved in <i>P. t. verus</i> individuals and divergent from the other three subspecies which harbored highly variable CD4 receptors (Hvilsom et al. 2008). Also, there is an interesting coincidence that the HIV-2/SIV<sub>sm</sub> virus, of West African origin like <i>P. t. verus</i>, has a broader initial coreceptor specificity than HIV-1, including CX<sub>CR4</sub> and CX<sub>CR1</sub>, and these are the loci showing low variation in this subspecies. More generally, the SIVs isolated from chimpanzees have complex origins, suggesting that chimpanzees have encountered different SIV strains during evolution (Bailes et al. 2003). This is likely due to predation by chimpanzees on different types of Old World monkeys, many of which all have their unique strains of SIV.

Equally, other factors, including other pathogens, may have contributed to the patterns of variation seen. An interesting recent finding is that there was a widespread infiltration of the chimpanzee genome with endogenous retroviruses (Yohn et al. 2005). The major spread is thought to have occurred 3–4 Ma, which probably predates the divergence of chimpanzee subspecies, so it could be related to the common patterns of variation seen around CCR5 but is unlikely to account for the subspecies differences seen here.

In conclusion, we find no evidence for deviation from a neutral pattern of evolution at CCR2, SDF, and RANTES in any chimpanzee subspecies. In addition, we find evidence consistent with a similar action of selection on CCR5 among chimpanzee subspecies, despite the very different current distribution of SIV<sub>cpz</sub> among them. In contrast, the combination of nucleotide diversity and neutrality tests strongly suggests different selective forces acting on CXCR4 and CX<sub>CR1</sub> among the three subspecies studied here. Whatever the evolutionary forces involved, this is one of the first studies to examine variation at functional loci among subspecies of chimpanzee and suggests that local adaptation among subspecies will be an interesting avenue for future research.

Supplementary Material

Supplementary table 1 is available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank Chris Mitchell for samples of Cameroonian chimpanzees and Brenda Bradley and four anonymous reviewers for comments on the manuscript. T.S.M. was supported by the Medical Research Council and N.I.M. by the Leverhulme Trust.

Literature Cited


Anne Stone, Associate Editor

Accepted December 20, 2008