Available online at www.sciencedirect.com
ScienceDirect
Genomics 90 (2007) 213 – 224
GENOMICS
www.elsevier.com/locate/ygeno

Gene conversion between mammalian *CCR2* and *CCR5* chemokine receptor genes: A potential mechanism for receptor dimerization

Núria Vàzquez-Salat [a,b], Naoya Yuhki [a], Thomas Beck [c], Stephen J. O'Brien [a], William J. Murphy [c,d,*]

[a] *Laboratory of Genomic Diversity, National Cancer Institute, Frederick, MD 21702, USA*
[b] *Hood College, Frederick, MD 21702, USA*
[c] *SAIC-Frederick, Inc. Laboratory of Genomic Diversity, National Cancer Institute, Frederick, MD 21702, USA*
[d] *Department of Veterinary Integrative Biosciences, College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, TX 77843-4458, USA*

Received 2 October 2006; accepted 26 April 2007
Available online 1 June 2007

Abstract

The chemokine receptor genes of the CCR cluster on human chromosome 3p21 play important roles in humoral and cellular immune responses. Several of these receptors have been shown to influence human immunodeficiency virus infection and progression to AIDS, and their homologues may play a role in feline immunodeficiency virus infection. We report the isolation and sequencing of a 150-kb domestic cat BAC clone containing the feline CCR genes *CCR1, CCR2, CCR3,* and *CCR5* to further analyze these four receptor genes within the family Felidae. Comparative and phylogenetic analyses reveal evidence for historic gene conversion between the adjacent *CCR2* and *CCR5* genes in the Felidae and in three independent mammalian orders (Primates, Cetartiodactyla, and Rodentia), resulting in higher than expected levels of sequence similarity between the two paralogous genes within each order. The gene conversion was restricted to the structural (transmembrane) domains of the *CCR2* and *CCR5* genes. We also discovered a recent gene conversion event between the third extracellular loop of *CCR2* and *CCR5* genes that was fixed in Asian lions and found at low frequency in African lions (*Panthera leo*), suggesting that this domain may have an important functional role. Our results suggest that ongoing parallel gene conversion between *CCR2* and *CCR5* promotes receptor heterodimerization in independent evolutionary lineages and offers an effective adaptive strategy for gene editing and coevolution among interactive immune response genes in mammals.
© 2007 Elsevier Inc. All rights reserved.

Keywords: Chemokine receptor; Evolution; Mammals; Gene conversion; Lions

Chemokine receptors are 7-transmembrane G-protein-coupled receptors expressed mainly in leukocytes. They are generally 350 amino acids in length with a short extracellular amino terminus and an intracellular carboxyl terminus. The transmembrane domains are α-helices connected by three intracellular and three extracellular loops. Their ligands are chemokines, which comprise a superfamily of low-molecular-weight (8-to 15-kDa) cytokines known for their role in the immune system where they function primarily as regulators of the immune response [1,2]. As is the case for all cytokines, chemokines are functionally redundant, have synergetic effects, and can behave as antagonists. These characteristics enable the creation of a complex and precise network that activates the immune response.

The human CC chemokine receptor family includes 10 genes (*CCR1* through *CCR10*), and several homologous genes have been described for other placental mammals (i.e., rat, mouse, pig, rabbit). With the exception of *CCR7*, most of the human CC chemokine receptor genes are concentrated on the short arm of chromosome 3, within a syntenic block of conserved gene order with other mammalian species. Receptors *CCR1* through *CCR5* are very similar in gene structure, suggesting that all five receptors share a recent common ancestor and originated through gene duplications [3]. Two pairs of loci, *CCR1-CCR3* and *CCR2-CCR5* are more closely related based on sequence similiarity and physical distance than to the other members of the cluster.

* Corresponding author. Department of Veterinary Integrative Biosciences, College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, TX 77843-4458, USA. Fax: +1 979 845 9972.
E-mail address: wmurphy@cvm.tamu.edu (W.J. Murphy).

0888-7543/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2007.04.009

Chemokine receptors, particularly those that comprise the CCR cluster, have been shown to serve as cell entry receptors for human immunodeficiency virus (HIV) life cycle, and polymorphic variants in their genes influence HIV infection and progression to AIDS [4–9]. Because the domestic cat is endemic with feline immunodeficiency virus (FIV), it is a good naturally occurring model for AIDS [10,11]. FIV is prevalent in several wild cat species [12,13] and is hypothesized to have infected cats earlier than HIV in humans, giving domestic cats and their wild relatives more time for host adaptation to the virus [14]. Thus, knowledge of the feline CCR cluster might highlight some aspects of CCR gene evolution that will aid in elucidating host-virus coevolution.

In this study, we sequenced and characterized part of the domestic cat CCR gene cluster on feline chromosome A2, which contains the *CCR1, CCR2, CCR3,* and *CCR5* genes. Comparative sequence analysis revealed a conserved gene structure and orientation in all examined placental mammals. Phylogenetic and recombination analyses provide evidence for ongoing gene conversion between the structural domains of CCR2 and CCR5 proteins in several mammalian genomes (mouse, rat, human, pig, and cat) but not between CCR1 and CCR3. Further analysis of *CCR2* and *CCR5* sequences in 171 lions revealed a local, directed case of gene conversion in all Asian and some African lions that effectively substitutes the third extracellular loop of CCR5 with the corresponding region from CCR2. We discuss the potential physiological and adaptive features of these gene conversion events and their role in the immune response.

Results

Characterization and sequencing of feline CCR cluster genes

A domestic cat BAC library (RPCI-86: http://bacpac.chori.org) was screened using human PCR probes for *CCR1, CCR2, CCR3, CCR5,* and *CCR9*. DNA fingerprints generated from hybridization-positive BAC clone DNAs yielded an overlapping contig of 12 BAC clones spanning the five CCR genes and adjacent genomic regions (Supplemental Fig. 1). The hybridization results indicated that one clone, 522e10, contained *CCR1*, *CCR3*, and two *CCR5*-related genes, which was selected for full sequencing.

A total of 3272 shotgun reads (bidirectional sequencing of 1636 clones) were produced from BAC clone 522e10, giving 10X coverage of the full clone, which resulted in a finished 151,416-bp BAC sequence. Four genes were identified by GenScan and verified by BLAST: *CCR1, CCR2, CCR3,* and *CCR5* (Fig. 1). The genomic structure of the four chemokine receptor genes is either a single exon with a 3′UTR region (*CCR3*, *CCR5,* and *CCR2*) or two exons and a 3′ UTR region (*CCR1*). All four genes have similar gene structure in human, mouse, and cat, with the orientation and distance among genes being largely conserved among these three species (Fig. 1, Supplemental Fig. 2). In these three species *CCR1* has an orientation opposite to those of the remaining CCR genes. With the exception of a few noncoding regions in human, only those regions corresponding to the CCR-coding sequence show significant nucleotide sequence identity (Supplemental Fig. 2).

Numerous genomic repeats were detected by RepeatMasker in BAC clone 522e10. As a whole, repetitive DNA comprises 37% of the full BAC sequence, mostly SINEs and LINEs (33%). No satellites and 58 short tandem repeats (2% of the total sequence) were found within the full clone. Repeats were more frequent between *CCR1* and *CCR3* than between *CCR2* and *CCR5* in all three species (Supplemental Figs. 2 and 3).

Examination of the nucleotide similarity of the four CCR genes across each species reveals that, while *CCR1* and *CCR3* genes have an average pairwise nucleotide similarity among the four species of 66%, *CCR2* and *CCR5* average 80% (Table 1). The same distinction is observed when the genes are translated into amino acid sequences; GC content is similar across the four genes and is fairly conserved in mouse, rat, human, and pig

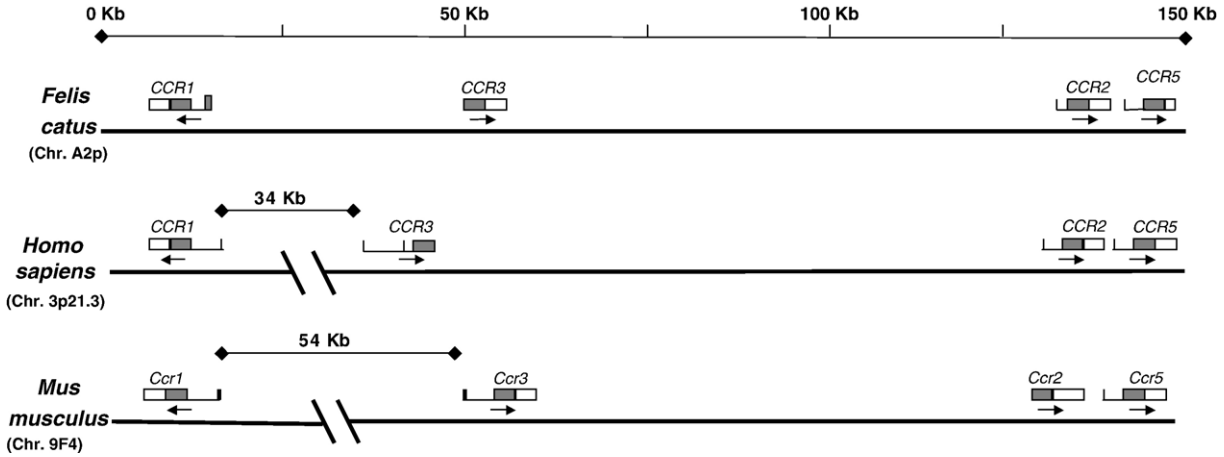


Fig. 1. Comparative physical map of the CC-chemokine receptor complex in domestic cat (*Felis catus*) from BAC clone 522e10, human (*Homo sapiens*), and mouse (*Mus musculus*). Chromosomal location is indicated below the species name. All four genes have a conserved orientation and order in human and mouse throughout the cluster. 5′UTR regions are marked in black, 3′UTR regions are white boxes, and coding regions are shaded boxes. Arrows indicate transcriptional orientation. Distances and gene structure for the human *CCR* cluster were obtained from the UCSC browser (genome.ucsc.edu) and for the mouse *Ccr* cluster from the Mouse Genome Server (www.ensembl.org/Mus_musculus).

Table 1
Nucleotide and amino acid similarities in cat *CCR1*, *CCR2*, *CCR3*, and *CCR5* genes compared to human, mouse, and pig

		CCR1	CCR3	CCR2	CCR5
Nucleotide	Cat vs Human	68%	68%	80%	82%
	Cat vs Mouse	64%	65%	75%	77%
	Cat vs Pig	n/a	n/a	82%	84%
Amino acid	Cat vs Human	60%	59%	75%	79%
	Cat vs Mouse	59%	58%	73%	76%
	Cat vs Pig	n/a	n/a	79%	83%

(Supplemental Table 1). Each gene alignment shows an intercalation of conserved segments followed by segments of high variability (both with similar length) (Fig. 2).

Phylogenetic analyses

The results comparing nucleotide and amino acid sequences of all four chemokine genes described above are corroborated by phylogenetic analyses of the amino acid aligment of the feline CCR sequences obtained in contig 522e10 and other mammalian homologues available in GenBank (Fig. 3). Phylogenetic results revealed that, while *CCR1* and *CCR3* loci were reciprocally monophyletic and diverged in accordance with the established evolutionary relationships of the different species, by contrast *CCR2* and *CCR5* loci clustered together within each lineage (i.e., they were sister taxa), producing four paraphyletic groups, one for each mammalian lineage (Fig. 3).

To further investigate the paraphyly of *CCR2* and *CCR5* genes across mammalian orders, we examined the amino acid alignment (Fig. 2) obtained from the domestic cat, primates, rodents, and cetartiodactyls for evidence supporting the phylogenetic results. By demarcating the transmembrane (nucleotide positions 90–173, 204–267, 306–371, 423–497, 594–663, 705–809, and 831–899 bp), extracellular (positions 268–305, 450–593, and 810–830 bp), intracellular (positions 174–203, 372–422 and 664–704 bp), amino terminus (1–90 bp), and carboxy terminus (900–1106 bp) domains, it is clear that a different pattern of conservation exists in the transmembrane and intracellular domains versus the amino terminus, carboxy terminus, and extracellular loops. While CCR5 and CCR2 can clearly be distinguished in the amino/carboxy terminus and extracellular loop areas, most notably by a four-amino-acid deletion in extracellular loop 2 of all CCR2 sequences (E2 in Fig. 2), the opposite was observed in the transmembrane and intracellular areas, where many substitutions are shared by both homologues of a given species (e.g., TM VI in Fig. 2).

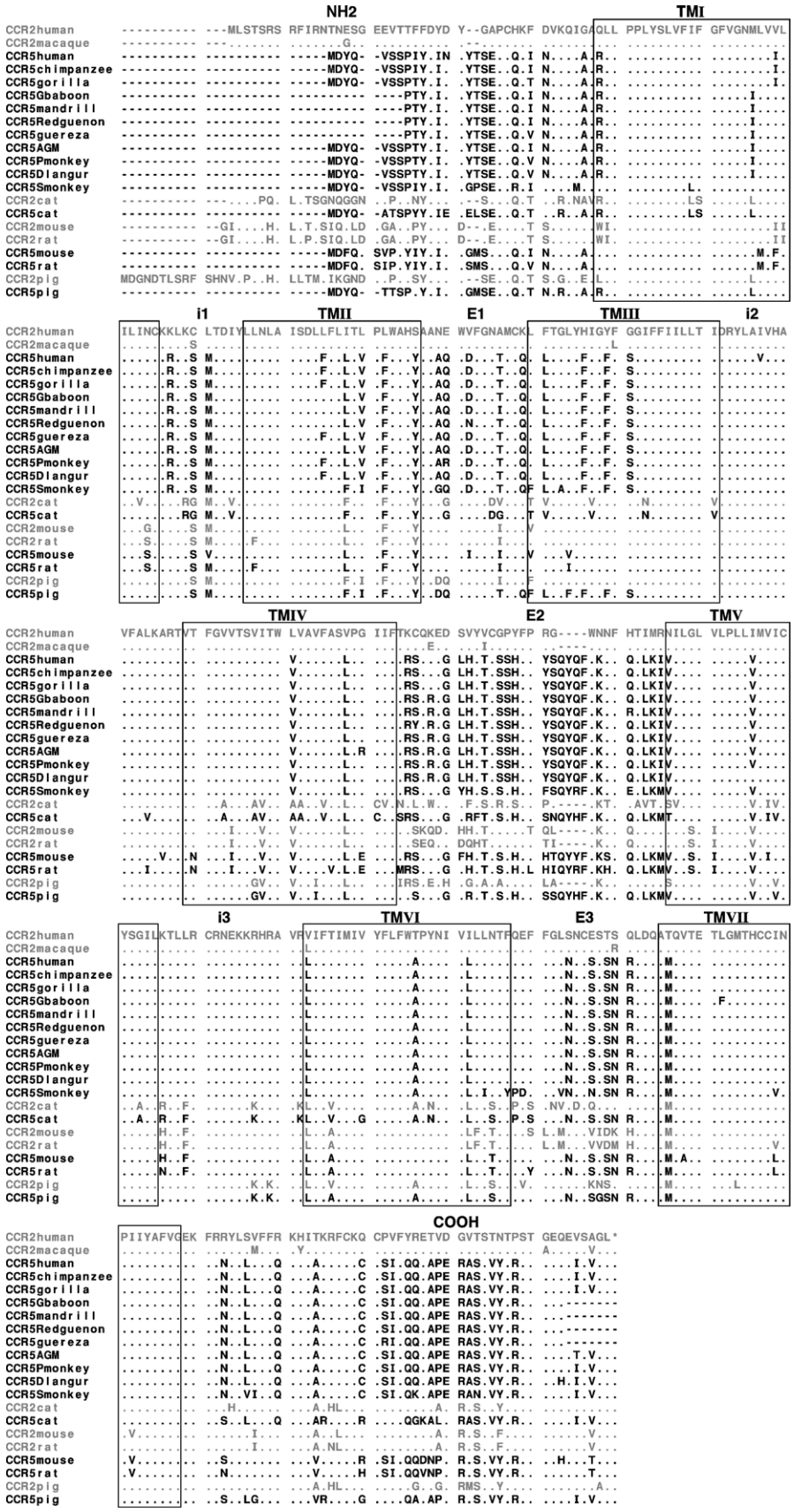
This pattern was further evident after performing additional phylogenetic analyses dividing the full alignment into two partitions: (i) functional domains containing the amino terminus, carboxy terminus, and extracellular loops and (ii) structural domains containing the transmembrane regions and intracellular loops. Whereas the maximum likelihood (ML) tree derived from the functional domains clearly separated *CCR2* and *CCR5* loci into two monophyletic groups (Fig. 4A), the structural domain ML tree grouped the *CCR2* and *CCR5* genes from each mammalian order as sister taxa (Fig. 4B).

Fig. 5A displays the domestic cat CCR3 protein sequence, with residues that differ between feline CCR1 and CCR3 highlighted in black. Although there are a few notably conserved regions (transmembranes III and VII), amino acid substitutions are dispersed throughout the remaining protein domains. However, a similar comparison between domestic cat CCR2 and CCR5 shows a different pattern: the vast majority of amino acid substitutions are found in the functional domains (Fig. 5B). Given these results we hypothesized that some form of recombination was occurring between *CCR2* and *CCR5* genes, although restricted to specific regions of each gene.

Gene conversion analysis of CCR5 and CCR2

Because the comparative alignment of the full feline BAC clone with human and mouse sequences showed no evidence of independent duplications of the *CCR2* and *CCR5* genes and because the gene order of the *CCR* cluster is conserved and likely the ancestral mammalian form (Fig. 1), we hypothesized that the observed homogenization of the two genes was due to recombination events, most probably gene conversion. Specifically, an ancient duplication likely produced *CCR2* and *CCR5* paralogues in an ancestral mammal (or vertebrate), followed by multiple independent recombination events between *CCR5* and *CCR2* loci in mammalian ordinal lineages. Gene conversion is a common process of recombination between physically close genes with high levels of sequence similarity. Because *CCR5* and *CCR2* fit both criteria, we searched for evidence of gene conversion between *CCR2* and *CCR5* using the program Likewind. This program examines phylogenetic signals of gene homologues in local regions (windows) of the alignment and searches for regions of discord from the index tree based on the whole alignment as evidence of recombination or gene conversion (see Materials and methods). We tested two scenarios: the first is based on the maximum likelihood topology derived from the analysis of the full gene alignment (which produces *CCR2* monophyly and *CCR5* monophyly, also observed in the functional domain ML tree in Fig. 4A), while the second assumes a tree where *CCR2* and *CCR5* clustered as monophyletic groups within each mammalian order (i.e., the structural domain ML tree in Fig. 4B).

The Likewind analysis identified four areas within the multispecies gene alignment that departed significantly from each of the two scenarios being tested. In the scenario that assumes paraphyly of *CCR2* and *CCR5* (i.e., the independent gene duplication hypothesis), four regions of the alignment that are discordant with this evolutionary hypothesis were identified. These gene regions correspond to the functional domains (extracellular loops, amino and carboxy termini) in addition to the seventh transmembrane domain (Fig. 6A, blue lines). In the scenario that assumes monophyly of *CCR2* and *CCR5*, the structural domains were instead found to differ significantly from the original tree (Fig. 6A, pink lines). The first transmembrane domain was found to be significant in both analyses. However high levels of conservation make this region uninformative for both evolutionary hypotheses.


NH2
TMI
i1
TMII
E1
TMIII
i2
TMIV
E2
TMV
i3
TMVI
E3
TMVII
COOH
CCR2human
CCR2macaque
CCR5human
CCR5chimpanzee
CCR5gorilla
CCR5Gbaboon
CCR5mandrill
CCR5Redguenon
CCR5guereza
CCR5AGM
CCR5Pmonkey
CCR5Dlangur
CCR5Smonkey
CCR2cat
CCR5cat
CCR2mouse
CCR2rat
CCR5mouse
CCR5rat
CCR2pig
CCR5pig

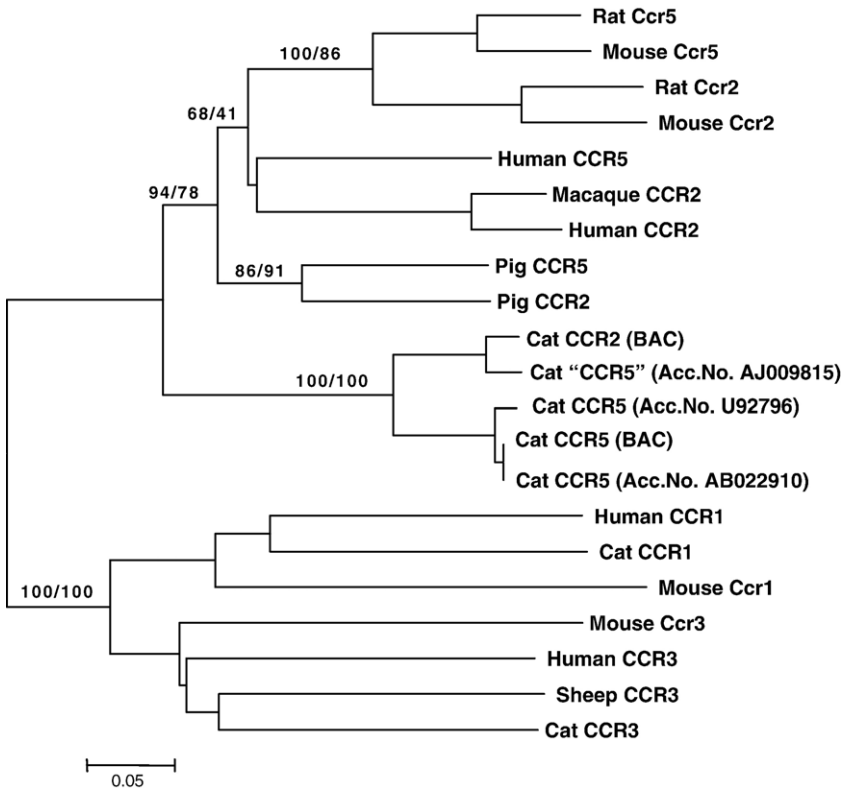


Fig. 3. Neighbor-joining tree based on the amino acid alignment of CCR1, CCR2, CCR3, and CCR5. Only the fragments considered structural were added for the alignment. Branch lengths are proportional to percentage substitutions; see scale. Bootstrap values for 100 replicates for maximum parsimony and neighbor-joining are indicated (MP/NJ). GenBank accession numbers for previously published feline sequences are shown in parentheses. Note that Accession No. AJ009815, identified in GenBank as *CCR5*, is most similar to feline *CCR2* and was thus likely misidentified.

An additional hallmark of gene conversion is increased GC content at third positions in converted regions [15–17]. Our analysis contrasting GC content at third positions between functional and structural domains supports a significant increase in third position GC content, consistent with gene conversion (Figs. 6B and 6C). Thus both the Likewind and the G-C content analyses are complementary and support the hypothesis of gene conversion occurring only in the structural regions.

Recent gene conversion of CCR5 and CCR2 in lions

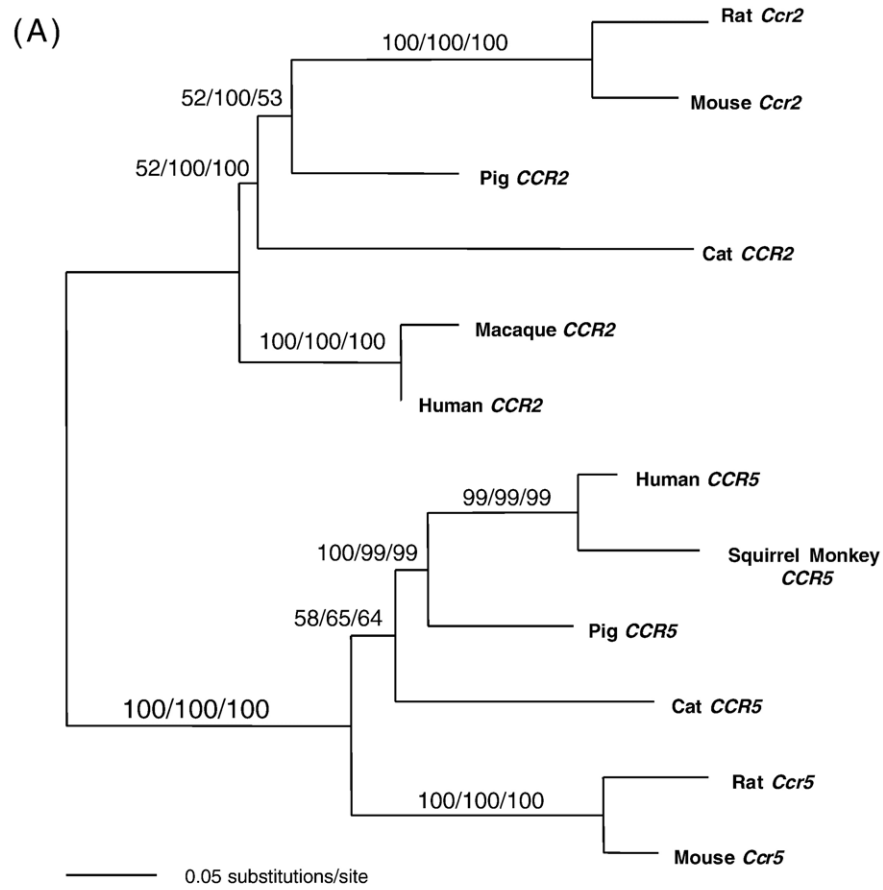
We further characterized the *CCR2* and *CCR5* genes in a related feline species, the lion (*Panthera leo*). We also obtained sequence for these genes from an additional domestic cat to rule out the possibility that the individual cat used to generate the BAC library may have carried specific mutation(s) that could bias the results. Finally, we generated *CCR2* and *CCR5* sequences from another divergent felid, a puma (*Puma concolor*), to establish broader trends of sequence conservation within the Felidae.

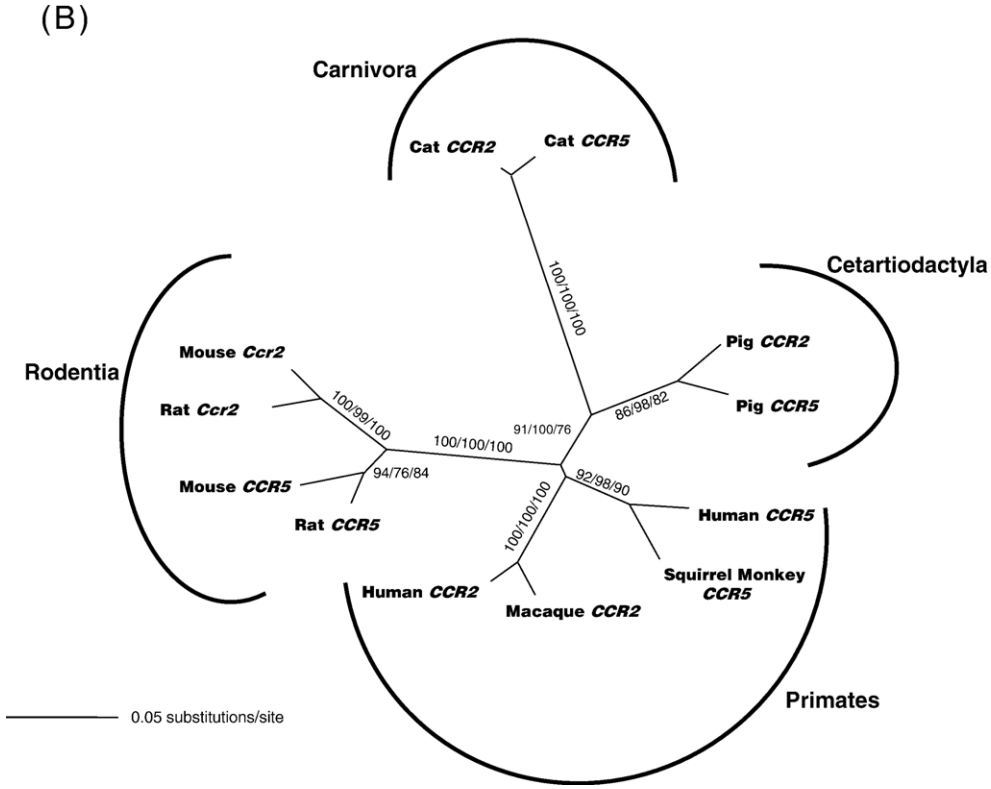
Lions are subdivided in two continental populations: African lions (*P. leo leo*) and Asian lions (*P. leo persica*) [18]. Asian lions are presently restricted to the Gir Forest of Gujarat state in western India and show reduced genetic variation [18,19]. Divergence between the populations is estimated to have occurred between 55,000 and 200,000 years ago [20,18]. Populations of lions examined are from the Serengeti ecosystem (Tanzania), Ngorongoro Crater (Tanzania), Etosha National Park (Namibia), Kalahari Gemsbok (Namibia), Kruger National Park (Republic of South Africa), and Gir forest (India).

CCR2 and CCR5 amino acid sequences showed extremely high levels of similarity, diverging primarily in the extracellular loops and the C terminus, while all other domains were very similar among the three felid species (Fig. 7A), with puma and domestic cat sequences more similar to each other than to the lion sequences. No indels or nonsense mutations were detected in the coding regions of the individuals sampled.

Several lion *CCR5* sequences showed a cluster of multiple polymorphic sites in a 30-bp region of the third extracellular loop (Fig. 7B). Upon further analysis it was discovered that

Fig. 2. Amino acid alignment of the full *CCR5* and *CCR2* genes in species from four different mammalian orders. CCR2 sequences are shown in gray, while CCR5 sequences are black. Structural domains are boxed. While the functional domains (extracellular loops) show conservation within the *CCR2* and *CCR5* gene families across all four mammalian orders, the structural (transmembrane and intracellular loops) are more similar among orders than within gene families. Thus, within each of the four mammalian orders *CCR2* and *CCR5* have very similar amino acid sequences in the transmembrane regions. This pattern is particularly marked in cat and pig *CCR2/CCR5* sequences. AGM, African green monkey; Dlangur, Doug langur; Gbaboon, Guinea baboon. NH2 (N terminus), TM (transmembrane domains), I (intracellular loop), E (extracellular loop), COOH (carboxyl terminus).


(A)
Rat Ccr2
100/100/100
Mouse Ccr2
52/100/53
52/100/100
Pig CCR2
Cat CCR2
Macaque CCR2
100/100/100
Human CCR2
Human CCR5
99/99/99
Squirrel Monkey CCR5
100/99/99
58/65/64
Pig CCR5
Cat CCR5
100/100/100
Rat Ccr5
100/100/100
Mouse Ccr5
0.05 substitutions/site


(B)
Carnivora
Cat CCR2
Cat CCR5
100/100/100
Cetartiodactyla
Pig CCR2
Pig CCR5
86/98/82
Rodentia
Mouse Ccr2
Rat Ccr2
100/99/100
91/100/76
100/100/100
94/76/84
Mouse CCR5
Rat CCR5
92/98/90
Human CCR5
100/100/100
Squirrel Monkey CCR5
Human CCR2
Macaque CCR2
Primates
0.05 substitutions/site

Asian lions and African lions had different sequences in this stretch of nucleotides, and only a subset of African lions contained a stretch of heterozygous sites (Fig. 7B). To further investigate this observation, a total of 171 lions were sequenced for this specific region (approximately 340 bp). All Asian lions ($n=11$) share a single distinct haplotype (FNVSDCQSTSQ in positions 226–236), and 14 of 160 African lions were heterozygous for the African haplotype (FGLNNCSSSNR) in this stretch of amino acids. Heterozygous individuals contain copies of both the "Asian" allele and the "African" allele (Fig. 7B), which was confirmed by cloning and sequencing PCR products from a subset of individuals. No African lions were found to be homozygous for the Asian *CCR5* allele. Sequences from heterozygous lions were devoid of adjoining heterozygous residues flanking this gene region. Both "African" and "Asian" haplotypes are in-frame but translate into eight different amino acids in this 11-residue stretch (Fig. 7B). The Asian lion CCR5 third extracellular loop amino acid sequences were identical to CCR2. This pattern was not observed in puma or domestic cat sequences, suggesting that it occurred in an ancestor of lions subsequent to their divergence from the common ancestor of lion, puma, and domestic cat.

Discussion

Two hypotheses can explain the observation that transmembrane and intracellular domains for *CCR2* and *CCR5* share more similarity in each species than with their corresponding orthologues in other mammalian orders (Fig. 2): 1) the *CCR5* and *CCR2* gene pair found in species of each mammalian order are the product of independent gene duplication events or (2) *CCR2* and *CCR5* derive from an ancestral duplication and have subsequently undergone recombination within each mammalian order. Our analysis showed no evidence for gene duplications in the CCR cluster, as we observed high levels of gene order, orientation, and gene structure identity among all species examined. Instead our data support a scenario of ancestral duplication followed by recombination, where the observed similarities between each species' *CCR2* and *CCR5* genes are due to species-specific homogenization of structural domains via gene conversion coupled with receptor-specific divergence in functional ligand-interacting domains. Homogenization via recombination is not unexpected because both genes are within 3–13 kb of each other in the species examined.

Gene conversion is a common homogenization mechanism within gene families that originate through gene duplication [21]. It is found more frequently in rapidly evolving genes (i.e., immune-response-related genes) such as VDJ rearrangements in B cells [22] or the major histocompatibility complex [23] and has an important role in evolution [24]. Following gene duplication, the resulting copies are prone to gene conversion events, which decrease in frequency as the copies acquire more nucleotide differences. Processes such as methylation or insertion of repetitive elements that generate greater sequence divergence prevent gene conversion [25]. We observed these features in the CCR cluster, where although the physical distance between *CCR1* and *CCR3* genes is similar to the intergenic distance of *CCR2* and *CCR5* genes, the higher concentration of repetitive elements within the *CCR1-CCR3* (Supplemental Fig. 3) region may inhibit gene conversion events between these two genes.

A similar pattern of gene conversion concentrated in structural regions was found in the C4a and C4b complement proteins [26]. Phylogenetic analysis of the full nucleotide sequence of both genes within various primate species produced a tree in which the genes group by primate species rather than by locus. Only when the functional domains of the proteins were analyzed separately was monophyly recovered. The authors concluded that gene conversion was the most plausible explanation for the observed results [26]. Other transmembrane receptors of the G-protein-coupled receptor superfamily, such as the *MOR28* olfactory receptor gene cluster [27], show similar evidence of gene conversion events.

A further biochemical explanation for the gene conversion events may be found in light of protein structure and function. Many members of the G-protein-coupled receptors family, such as GABA receptors and beta-adrenergic receptors, are known to form homodimers, heterodimers, or oligomers for signal activation and/or transduction [28–30]. The mechanism that involves dimerization is different among these proteins, suggesting independent origins for each event and highlighting the importance of this dimerization for correct receptor function. *CCR2* has previously been shown to form homodimers [31] but also forms heterodimers with *CCR5*, producing synergistic effects that are able to activate a cellular response in the presence of very reduced (10-to 100-fold) concentrations of their respective ligands, producing novel signal transduction pathways that delay the activation of IP3 [32,33]. Because heterodimerization requires that both receptors share a high degree of similarity, it is plausible that the gene conversion documented here between the *CCR2* and the *CCR5* genes maintains a similar transmembrane structure (where dimerization is known to occur) that would facilitate recognition between CCR2 and CCR5 proteins and heterodimerization, while the functional areas of the genes are free to evolve novel functions.

Our examination of *CCR5* in lions, and comparison with the other felid *CCR5* and *CCR2* sequences, showed the Asian lion *CCR5* haplotype had acquired a segment of the lion *CCR2* third extracellular loop. All Asian lions were homozygous for this *CCR2*-to-*CCR5* third extracellular loop gene conversion allele, while none of the 170 African lions tested for this area were

Fig. 4. (A) Maximum likelihood tree (K+80 model) derived from nucleotide analysis of the functional domains of *CCR2* and *CCR5* (amino terminus, extracellular loops, and carboxyl terminus). The tree is divided into monophyletic gene families for both *CCR2* and *CCR5* genes. Bootstrap values are shown for maximum parsimony/neighbor-joining/maximum likelihood. (B) Maximum likelihood tree (HKY+G model) derived from nucleotide analysis of the structural domains of *CCR2* and *CCR5* (transmembrane and intracellular loops). Note that gene families are paraphyletic. Bootstrap values are for maximum parsimony/neighbor-joining/maximum likelihood.

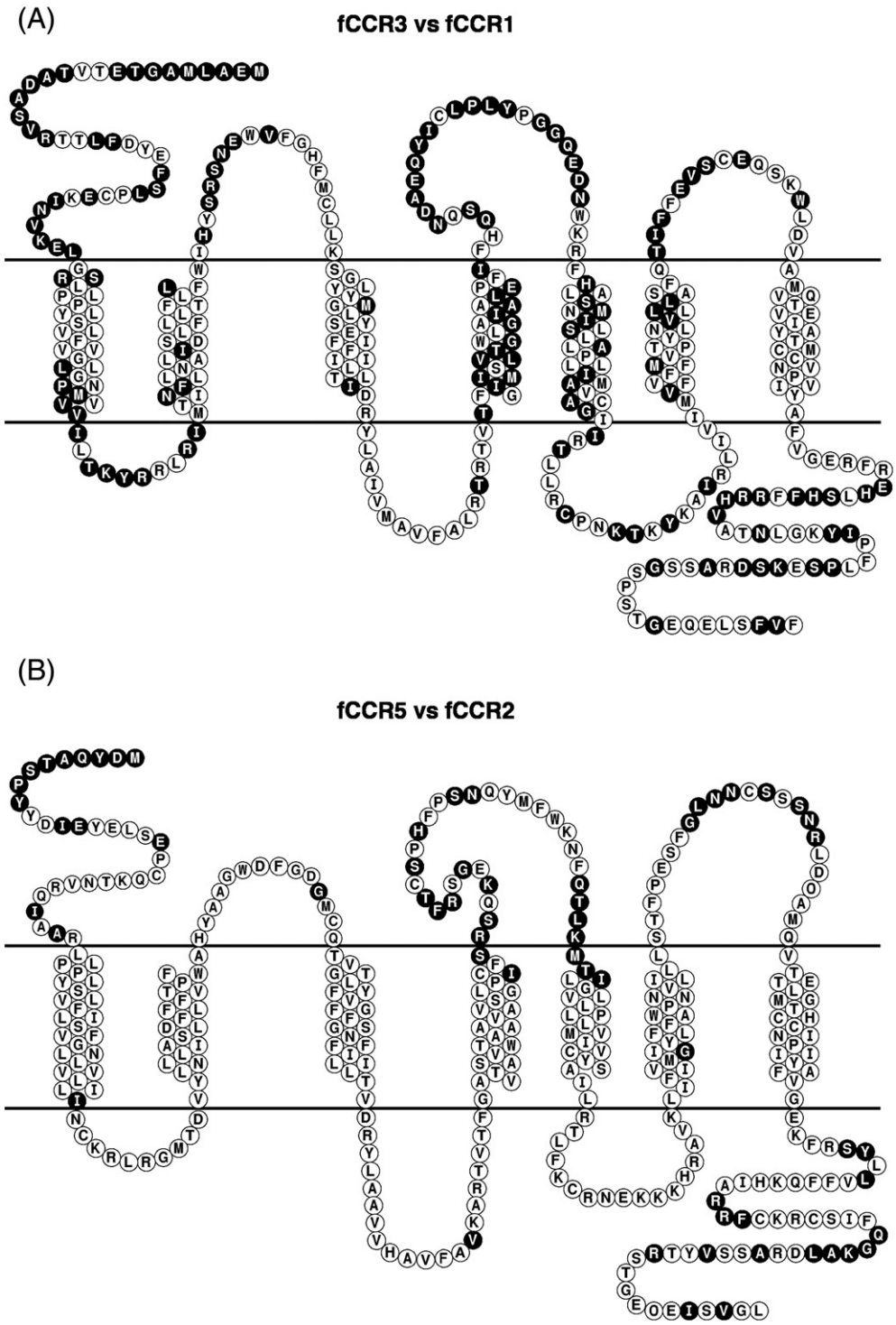


Fig. 5. Feline chemokine receptor amino acid sequence comparisons. (A) CCR3 versus CCR1, where the given amino acid sequence corresponds to CCR3. All conserved amino acids between CCR1 and CCR3 are shown in white. Substitutions (in black) are equally distributed throughout the sequence, except for a higher concentration in the extracellular domains. (B) CCR5 versus CCR2, where the amino acid sequence corresponds to CCR5 and amino acid changes between CCR2 and CCR5 are shown in black. In this case, amino acid differences are concentrated on the N terminus, extracellular loops, and C terminus. Both CCR1 and CCR2 have longer N termini than CCR3 and CCR5, respectively. Sequence comparisons in this figure starts with the first amino acids of CCR3 and CCR5.

homozygous for this allele. The lack of observed homozygosity for the gene conversion allele within the African lion population is likely due to its low frequency (0.04), whereas the most probable explanation for fixation of this rare allele in the Asian lions is their well-documented severe bottleneck [20].

Several studies have examined the structure and ligand-specificity of *CCR2* and/or *CCR5* [34–36]. All establish the

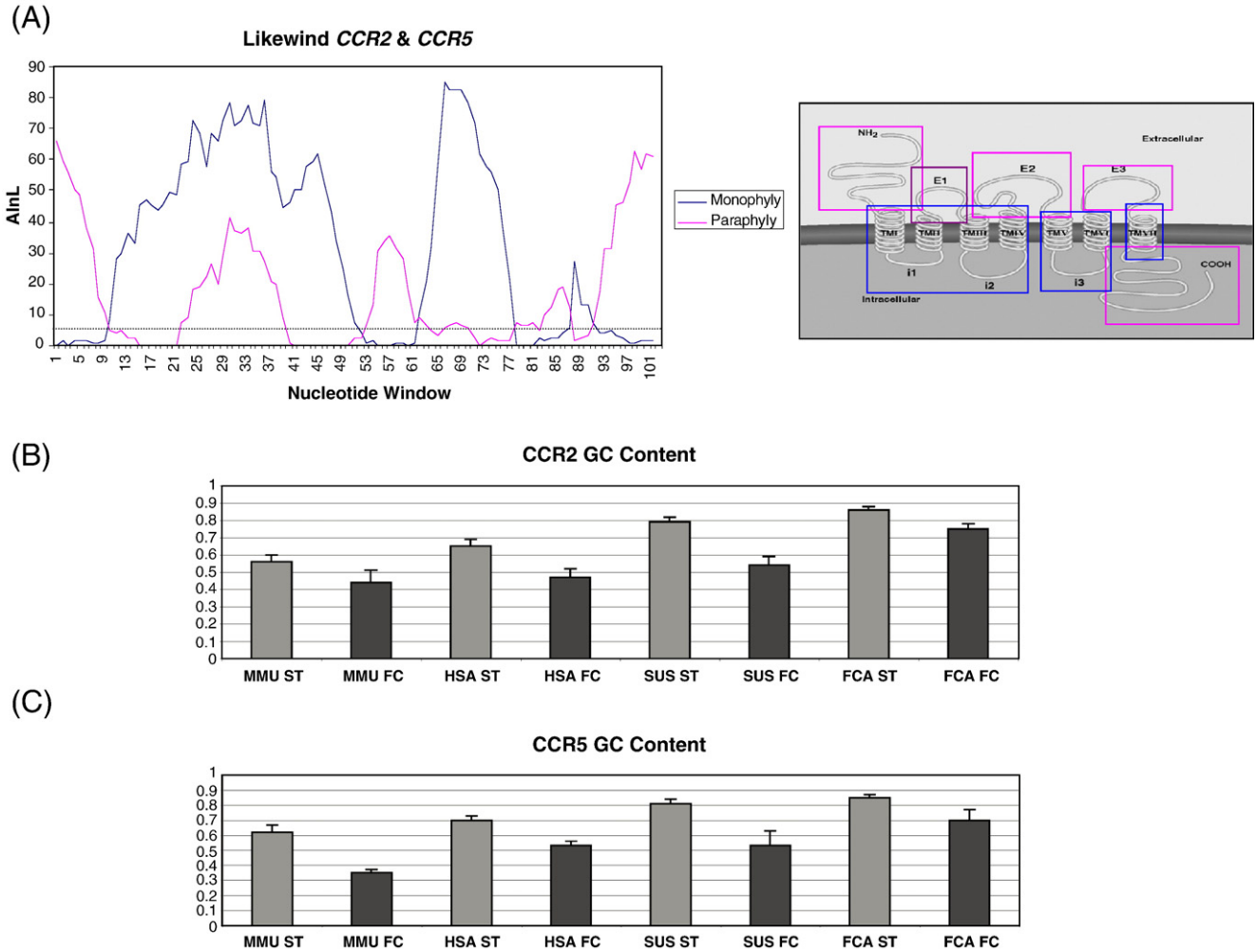


Fig. 6. (A) Maximum likelihood sliding window analysis of *CCR2* and *CCR5* from domestic cat, mouse, human, and pig sequences. Domains where no recombination is detected are highlighted. NH2 (N terminus), E (extracellular loop), TM (transmembrane), and COOH (C terminus). Both runs are color coded according to their original hypothesis: pink lines indicate paraphyly of gene families and blue lines monophyly of gene families. (B) *CCR2* third position GC content (and standard errors) calculated for structural (ST) and functional (FC) domains for each species. (C) *CCR5* third position GC content (and standard errors) calculated for structural (ST) and functional (FC) domains for each species. MMU, mouse; HSA, human; FCA, cat; SUS, pig.

importance of the N terminus and C terminus in ligand recognition and signal transduction, respectively. The first and second extracellular loops have important roles in ligand recognition; however the importance of the third extracellular loop is still not clear. Our results from lions suggest that the third extracellular loop may be involved in ligand recognition due to the large sequence divergence observed between *CCR2* and *CCR5* genes. However, further in vitro studies are needed to test whether the two distinct lion haplotypes differ in ligand binding and/or cellular activity.

In conclusion, we find evidence for gene conversion acting as a mechanism to homogenize transmembrane regions of mammalian CCR2 and CCR5 proteins. This similarity is driven to promote heterodimerization via the structural domains. Furthermore, gene conversion is also not uniformly strong across lineages, as feline and porcine CCR2-CCR5 similarity is higher than that in other lineages and in some species (i.e., lions) it appears to be an ongoing process.

Materials and methods

Sampling scheme and DNA extraction

DNA was extracted using either the phenol-chloroform method [37] or the QIAMP tissue kit (Qiagen) following the manufacturer's protocol with some slight variations, as follows: digest tissue sample with 200 µl of ATL Buffer, 20 µl of Proteinaise K at 56 °C for the overnight incubation and 15 min at 70 °C for the final incubation with AL Buffer. The remaining steps follow the protocol, except the last step which uses 300 µl of AE Buffer during a 5-min centrifugation.

Isolation of the CCR cluster-containing BAC clone

BAC clone 522e10 was selected from the RPCI-86 feline BAC library (http://bacpac.chori.org) based on hybridization experiments using human PCR products as probes [38]. A restriction map of the overlapping clones revealed that clone 522e10 contained the four CCR genes of interest, *CCR2, CCR5, CCR3*, and *CCR1* (Supplemental Material). Large-scale BAC DNA isolation was performed with a Qiagen Large-Construct Kit. The clone was sequenced to 10X coverage using a random subcloning approach [39]. A total of 1636

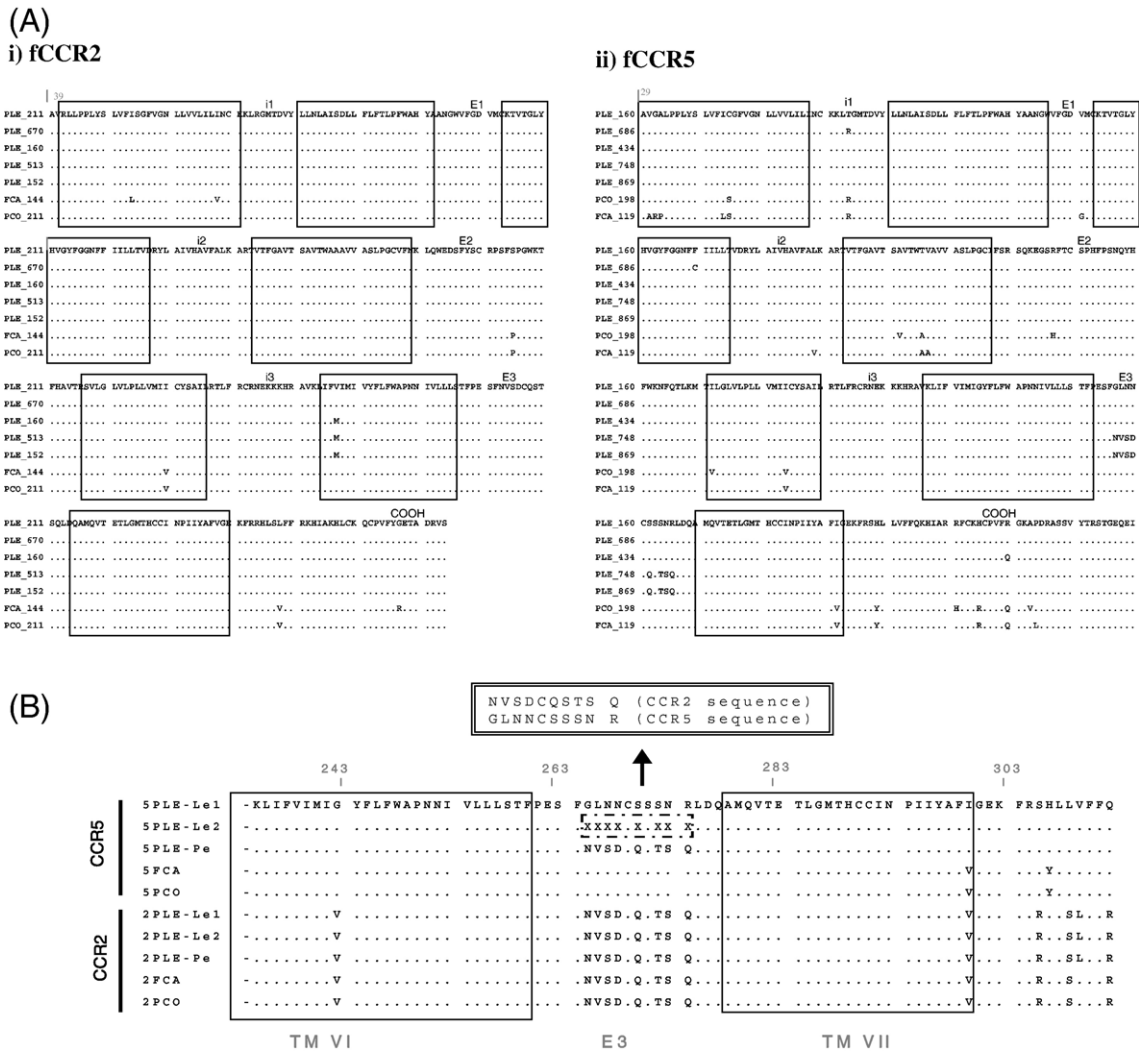


Fig. 7. (A) Amino acid alignment for *CCR2 (i)* and *CCR5 (ii)* of domestic cat (FCA), puma (PCO), and lion (PLE) individuals (indicated by species prefix and numerical identity code). All transmembrane domains are boxed. E, extracellular loops; I, intracellular loops; COOH, carboxyl terminus. Sequences start at transmembrane domain I. Dots indicate residues identical to the reference. (B) Amino acid alignment of CCR5 and CCR2 of lion (PLE), puma (PCO), and domestic cat (FCA). The area shown covers the sixth and seventh transmembrane domains (TM VI and TM VIII), the third extracellular loop (E3), and the carboxyl terminus (COOH). PLE-Le indicates lions of African origin; PLE-Pe represents Asian lions; 5 and 2 at the beginning of the individual names stand for CCR5 and CCR2 sequences, respectively. The E3 region translated into amino acids code for very distinct sequences. While the African lion, PLE-Le1, contains the CCR5 expected sequence based upon comparison to domestic cats (FCA) and pumas (PCO), the Asian lion, PLE-Pe, contains a CCR2-like sequence within E3, apparently due to gene conversion between CCR2 and CCR5. PLE-Le2 is a heterozygous individual (FXXXXCXSXXX in this region) having both Asian and African CCR5 alleles.

colonies were selected and grown individually in 1.25 ml of Superbroth (Quality Biologicals) overnight at 37 °C. Plasmid DNA was extracted using the protocol modified from [40] and described by [39]. Clones were sequenced using BigDye Terminator chemistry (ABI) and purified with G-50 Sephadex plates (Amersham Pharmacia). Sequences were generated on an ABI 3730 automated sequencer.

All sequences were analyzed using a Biolims (ABI) enviroment (PE Informatics), including base calling, quality assessment, elimination of vector and *Escherechia coli* sequences, and assembly into contigs using Phred, Phrap, and Consed [41,42]. The Autofinish program was used to combine contigs and finish sequencing until a Phred quality value of 30 for each base and a Phrap quality value of 40 were obtained for the BAC sequence. Automated sequence analysis was carried out using the e3500, DEC alpha, and SGI high-speed computers at the Advanced Biomedical Computing Center (Frederick, MD). BLASTN [43], Genscan [44], and BLASTP [45] were used, respectively, for nucleotide sequence homology searches in databases, for prediction of protein-coding genes, and for homology search of the Genscan-predicted peptide sequences for all genes in the singleton nucleotide sequences and sequence contigs. Analyses comparing orthologous regions in human, mouse, and feline were done using PipMaker Dot Plot [46].

PCR-based resequencing of CCR5 and CCR2 in lions and domestic cats

Due to the high sequence similarity observed between *CCR5* and *CCR2*, gene-specific primers were designed in divergent areas observed between the two feline genes obtained from the aligned sequences of *CCR2* and *CCR5* from the domestic cat BAC clone 522e10 sequence. Only the exonic and UTR areas were obtained (Fig. 2, Supplemental Tables 2 and 3). All primers were designed using Primer3 (http://www-genome.wi.mit.edu/cgibin/primer/primer3_www.cgi).

DNA was amplified using Taq Gold (Applied Biosystems, Inc.) DNA polymerase and the following mix: 1.5 mM $MgCl_2$, 1X Taq Gold PCR Buffer, 2 mM dNTP blend (Roche Molecular Biochemicals), and 0.4 μM each foward

and reverse primer. PCR profiles optimized for each primer set are described in Supplemental Material.

PCR products were purified with either Exonuclease I (0.38 μl per 25-μl reaction) and shrimp alkaline phosphatase (0.74 μl per 25-μl reaction) (Amersham) or the Psi-clone 96-well reaction clean-up kit (Princeton Separations Inc.). DNA sequencing reactions were performed with Big Dye Terminator enzyme mix (Applied Biosystems, Inc.) and 0.2 μM PCR primers. Sequencing products were purified with Sephadex G-50 plates (Amersham Pharmacia) and analyzed with an ABI 3730 automated capillary sequencer.

Sequences were manually inspected, the full gene was reconstructed, and variants were identified using SEQUENCER (Gene Codes). All sequences were verified via BLAST searches [43] and aligned with ClustalX [47] and SeAl (http://evolve.zoo.ox.ac.uk/).

Phylogenetic analyses

Phylogenetic analyses were performed on the combined *CCR1, CCR2, CCR3,* and *CCR5* amino acid alignment. MEGA 2.0 [48] was used for maximum parsimony and neighbor-joining amino acid searches. Bootstrap searches used 1000 heuristic replicates. Other domain-specific phylogenetic analyses using PAUP 4.0*b10 [49] were performed using the amino acid alignments for *CCR5* and *CCR2* genes with maximum parsimony, neighbor-joining, and maximum likelihood. Substitution models were selected by Modeltest [50]: HKY+G for the structural domains and K80+G for the functional domains of the receptors. One thousand bootstrap replicates were performed for each analytical method (except ML). MEGA 2.0 [48] was used to determine the GC content and nucleotide divergence for genes in all three species.

Tests for gene conversion

The Likewind program [51] was used to detect possible recombination between *CCR2* and *CCR5*. The program is based on three scripts: likewind.pl, getlikes.pl, and simblock.pl. Likewind.pl creates a file that commands PAUP to create a 100-bp sliding window every 10 bp through the full alignment. For each 100-bp window a phylogenetic tree is produced, and its ML score is calculated. For the same 100-bp window the ML score of either a user input tree or the tree based on the full alignment is also calculated. Getlikes.pl calculates the difference in likelihood score (ΔLn) for both trees from each window, and the resulting ΔLns are then plotted to show a distribution throughout the sequence alignment. A confidence level is created using simblock.pl, which uses SeqGen [52] to create 100 simulated datasets using the original dataset parameters. PAUP then creates an independent distribution of ΔLns for each of the 100 datasets produced. Getlikes.pl is used to obtain a list of the highest ML values of each of the 100 datasets to obtain a 95% confidence interval. Two analyses were performed with a dataset containing an alignment of *CCR2* and *CCR5* genes for human, mouse, cat, and pig: one where the starting tree assumed a monophyletic origin for both genes and another one that assumed paraphyly between *CCR2* and *CCR5* genes.

Acknowledgments

We thank Nicole Crumpler and Evir Goldman for technical support. We thank our colleagues Colm O'huigin, Mike Dean, Agostinho Antunes, Gila Kahila Bar-Gal, Al Roca, Eduardo Eizirik, and Mike Smith for discussions and valuable insight into this study.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ygeno.2007.04.009.

References

[1] M. Baggiolini, C.A. Dahinden, CC chemokines in allergic inflammation, Immunol. Today 15 (1994) 127–133.

[2] M.D. Miller, M.S. Krangel, Biology and biochemistry of the chemokines: a family of chemotactic and inflammatory cytokines, Crit. Rev. Immunol. 12 (1992) 17–46.

[3] P. Lio, M. Vannucci, Investigating the evolution and structure of chemokine receptors, Gene 317 (2003) 29–37.

[4] P. An, G.W. Nelson, L. Wang, S. Donfield, J.J. Goedert, J. Phair, D. Vlahov, S. Buchbinder, W.L. Farrar, W. Modi, S.J. O'Brien, C.A. Winkler, Modulating influence on HIV/AIDS by interacting RANTES gene variants, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 10002–10007.

[5] H. Choe, M. Farzan, Y. Sun, N. Sullivan, B. Rollins, P.D. Ponath, L. Wu, C.R. Mackay, G. LaRosa, W. Newman, N. Gerard, C. Gerard, J. Sodroski, The beta-chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates, Cell 85 (1996) 1135–1148.

[6] M. Dean, M. Carrington, C. Winkler, G.A. Huttley, M.W. Smith, R. Allikmets, J.J. Goedert, S.P. Buchbinder, E. Vittinghoff, E. Gomperts, S. Donfield, D. Vlahov, R. Kaslow, A. Saah, C. Rinaldo, R. Detels, S.J. O'Brien, 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 273 (1996) 1856–1862.

[7] B.J. Doranz, J. Rucker, Y. Yi, R.J. Smyth, M. Samson, S.C. Peiper, M. Parmentier, R.G. Collman, R.W. Doms, A dual-tropic primary HIV-1 isolate that uses fusin and the beta-chemokine receptors CKR-5, CKR-3, and CKR-2b as fusion cofactors, Cell 85 (1996) 1149–1158.

[8] P. Duggal, P. An, T.H. Beaty, S.A. Strathdee, H. Farzadegan, R.B. Markham, L. Johnson, S.J. O'Brien, D. Vlahov, C.A. Winkler, Genetic influence of CXCR6 chemokine receptor alleles on PCP-mediated AIDS progression among African Americans, Genes Immun. 4 (2003) 245–250.

[9] M.W. Smith, M. Dean, M. Carrington, C. Winkler, G.A. Huttley, D.A. Lomb, J.J. Goedert, T.R. O'Brien, L.P. Jacobson, R. Kaslow, S. Buchbinder, E. Vittinghoff, D. Vlahov, K. Hoots, M.W. Hilgartner, S.J. O'Brien, Contrasting genetic influence of CCR2 and CCR5 variants on HIV-1 infection and disease progression. Hemophilia Growth and Development Study (HGDS), Multicenter AIDS Cohort Study (MACS), Multicenter Hemophilia Cohort Study (MHCS), San Francisco City Cohort (SFCC), ALIVE Study, Science 277 (1997) 959–965.

[10] N.C. Pedersen, E.W. Ho, M.L. Brown, J.K. Yamamoto, Isolation of a T-lymphotropic virus from domestic cats with an immunodeficiency-like syndrome, Science 235 (1987) 790–793.

[11] B.J. Willett, J.N. Flynn, M.J. Hosie, FIV infection of the domestic cat: an animal model for AIDS, Immunol. Today 18 (1997) 182–189.

[12] E.W. Brown, N. Yuhki, C. Packer, S.J. O'Brien, A lion lentivirus related to feline immunodeficiency virus: epidemiologic and phylogenetic aspects, J. Virol. 68 (1994) 5953–5968.

[13] J.L. Troyer, J. Pecon-Slattery, M.E. Roelke, L. Black, C. Packer, S.J. O'Brien, Patterns of feline immunodeficiency virus multiple infection and genome divergence in a free-ranging population of African lions, J. Virol. 78 (2004) 3777–3791.

[14] M.A. Carpenter, S.J. O'Brien, Coadaptation and immunodeficiency virus: lessons from the Felidae, Curr. Opin. Genet. Dev. 5 (1995) 739–745.

[15] N. Galtier, Gene conversion drives GC content evolution in mammalian histones, Trends Genet. 19 (2003) 65–68.

[16] N. Galtier, G. Piganeau, D. Mouchiroud, L. Duret, GC-content evolution in mammalian genomes: the biased gene conversion hypothesis, Genetics 159 (2001) 907–911.

[17] J.P. Noonan, J. Grimwood, J. Schmutz, M. Dickson, R.M. Myers, Gene conversion and the evolution of protocadherin gene cluster diversity, Genome Res. 14 (2004) 354–366.

[18] S.J. O'Brien, J.S. Marterson, C. Packer, L. Herbst, V. de Vos, P. Joslin, J. Ott-Joslin, D.E. Wildt, M. Bush, Biochemical genetic variation in geographic isolates of African and Asian lions, Natl. Geog. Res. 3 (1987) 114–124.

[19] D.E. Wildt, M. Bush, K.L. Goodrowe, C. Packer, A. Pusey, J.L. Brown, P. Joslin, S.J. O'Brien, Reproductive and genetic consequences of founding isolated lion populations, Nature 329 (1987) 328–331.
[20] C.A. Driscoll, M. Menotti-Raymond, G. Nelson, D. Goldstein, S.J. O'Brien, Genomic microsatellites as evolutionary chronometers: a test in wild cats, Genome Res. 12 (2002) 414–423.
[21] D.C. Shields, Gene conversion among chemokine receptors, Gene 246 (2000) 239–245.
[22] M. Diaz, M.F. Flajnik, Evolution of somatic hypermutation and gene conversion in adaptive immunity, Immunol. Rev. 162 (1998) 13–24.
[23] T. Ohta, Effect of gene conversion on polymorphic patterns at major histocompatibility complex loci, Immunol. Rev. 167 (1999) 319–325.
[24] L. Frisse, R.R. Hudson, A. Bartoszewicz, J.D. Wall, J. Donfack, A. Di, Rienzo, Gene conversion and different population histories may explain the contrast between polymorphism and linkage disequilibrium levels, Am. J. Hum. Genet. 69 (2001) 831–843.
[25] J.C. Schimenti, Gene conversion and the evolution of gene families in mammals, Soc. Gen. Physiol. Ser. 49 (1994) 85–91.
[26] H. Kawaguchi, Z. Zaleska-Rutczynska, F. Figueroa, H.C. O'Uigin, J. Klein, C4 genes of the chimpanzee, gorilla, and orang-utan: evidence for extensive homogenization, Immunogenetics 35 (1992) 16–23.
[27] F. Nagawa, S. Yoshihara, A. Tsuboi, S. Serizawa, K. Itoh, H. Sakano, Genomic analysis of the murine odorant receptor MOR28 cluster: a possible role of gene conversion in maintaining the olfactory map, Gene 292 (2002) 73–80.
[28] B.A. Jordan, L.A. Devi, G-protein-coupled receptor heterodimerization modulates receptor function, Nature 399 (1999) 697–700.
[29] K. Kaupmann, B. Malitschek, V. Schuler, J. Heid, W. Froestl, P. Beck, J. Mosbacher, S. Bischoff, A. Kulik, R. Shigemoto, A. Karschin, B. Bettler, GABA(B)-receptor subtypes assemble into functional heteromeric complexes, Nature 396 (1998) 683–687.
[30] J.H. White, A. Wise, M.J. Main, A. Green, N.J. Fraser, G.H. Disney, A.A. Barnes, P. Emson, S.M. Foord, F.H. Marshall, Heterodimerization is required for the formation of a functional GABA(B) receptor, Nature 396 (1998) 679–682.
[31] J.M. Rodriguez-Frade, A.J. Vila-Coro, A.M. de Ana, J.P. Albar, A.C. Martinez, M. Mellado, The chemokine monocyte chemoattractant protein-1 induces functional responses through dimerization of its receptor CCR2, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 3628–3633.
[32] M. Mellado, J.M. Rodriguez-Frade, A.J. Vila-Coro, S. Fernandez, A. Martin de Ana, D.R. Jones, J.L. Toran, A.C. Martinez, Chemokine receptor homo-or heterodimerization activates distinct signaling pathways, EMBO J. 20 (2001) 2497–2507.
[33] M. Mellado, A.J. Vila-Coro, C. Martinez, J.M. Rodriguez-Frade, Receptor dimerization: a key step in chemokine signaling, Cell Mol. Biol. (Noisy-le-grand) 47 (2001) 575–582.
[34] G. Alkhatib, E.A. Berger, P.M. Murphy, J.E. Pease, Determinants of HIV-1 coreceptor function on CC chemokine receptor 3. Importance of both extracellular and transmembrane/cytoplasmic regions, J. Biol. Chem. 272 (1997) 20420–20426.
[35] F.S. Monteclaro, I.F. Charo, The amino-terminal extracellular domain of the MCP-1 receptor, but not the RANTES/MIP-1alpha receptor, confers chemokine selectivity. Evidence for a two-step mechanism for MCP-1 receptor activation, J. Biol. Chem. 271 (1996) 19084–19092.
[36] M.G. Paterlini, Structure modeling of the chemokine receptor CCR5: implications for ligand binding and selectivity, Biophys. J. 83 (2002) 3012–3031.
[37] J. Sambrook, D. Russell, Molecular cloning, A laboratory manual, Cold Spring Harbour Laboratory Press, New York, 2001, p. 999.
[38] T.W. Beck, J. Menninger, G. Voigt, K. Newmann, Y. Nishigaki, W.G. Nash, R.M. Stephens, Y. Wang, P.J. de Jong, S.J. O'Brien, N. Yuhki, Comparative feline genomics: a BAC/PAC contig map of the major histocompatibility complex class II region, Genomics 71 (2001) 282–295.
[39] N. Yuhki, T. Beck, R.M. Stephens, Y. Nishigaki, K. Newmann, S.J. O'Brien, Comparative genome organization of human, murine, and feline MHC class II region, Genome Res. 13 (2003) 1169–1179.
[40] W.L. Ng, M. Schummer, F.D. Cirisano, R.L. Baldwin, B.Y. Karlan, L. Hood, High-throughput plasmid mini preparations facilitated by micro-mixing, Nucleic. Acids Res. 24 (1996) 5045–5047.
[41] B. Ewing, L. Hillier, M.C. Wendl, P. Green, Base-calling of automated sequencer traces using phred. I. Accuracy assessment, Genome Res. 8 (1998) 175–185.
[42] D. Gordon, C. Abajian, P. Green, Consed: a graphical tool for sequence finishing, Genome Res. 8 (1998) 195–202.
[43] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (1990) 403–410.
[44] C. Burge, S. Karlin, Prediction of complete gene structures in human genomic DNA, J. Mol. Biol. 268 (1997) 78–94.
[45] W. Gish, D.J. States, Identification of protein coding regions by database similarity search, Nat. Genet. 3 (1993) 266–272.
[46] S. Schwartz, Z. Zhang, K.A. Frazer, A. Smit, C. Riemer, J. Bouck, R. Gibbs, R. Hardison, W. Miller, PipMaker-a web server for aligning two genomic DNA sequences, Genome Res. 10 (2000) 577–586.
[47] J.D. Thompson, T.J. Gibson, F. Plewniak, F. Jeanmougin, D.G. Higgins, The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools, Nucleic. Acids Res.. 25 (1997) 4876–4882.
[48] S. Kumar, K. Tamura, I.B. Jakobsen, M. Nei, MEGA2: molecular evolutionary genetics analysis software, Bioinformatics 17 (2001) 1244–1245.
[49] D.L. Swofford, PAUP* Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts. (2003).
[50] D. Posada, K.A. Crandall, MODELTEST: testing the model of DNA substitution, Bioinformatics 14 (1998) 817–818.
[51] J.M. Archibald, A.J. Roger, Gene conversion and the evolution of euryarchaeal chaperonins: a maximum likelihood-based method for detecting conflicting phylogenetic signals, J. Mol. Evol. 55 (2002) 232–245.
[52] A. Rambaut, N.C. Grassly, Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees, Comput. Appl. Biosci. 13 (1997) 235–238.