Available online at www.sciencedirect.com
SCIENCE DIRECT®
Genomics 83 (2004) 735–738
GENOMICS
www.elsevier.com/locate/ygeno

Short Communication

CCL3L1 and *CCL4L1* chemokine genes are located in a segmental duplication at chromosome 17q12☆,☆☆

William S. Modi*

Basic Research Program, SAIC Frederick, National Cancer Institute at Frederick, Frederick, MD 21702-1201, USA

Received 6 June 2003; accepted 10 September 2003

Abstract

Sixteen CC chemokine genes localize to a 2.06-Mb interval at 17q11.2–q12 on genomic contig NT_010799.13. Four of these genes comprise two closely related paralogous pairs: *CCL3–CCL3L1* and *CCL4–CCL4L1*. Members within each pair share 95% sequence identity at both the genomic and the amino acid levels. One BAC clone (AC131056.5) on the contig with substantial internal sequence duplication contains two complete copies of *CCL3L1* and *CCL4L1* and one truncated copy of *CCL3L1*, while a partially overlapping clone (AC003976.1) contains one copy each of *CCL3* and *CCL4*. Dot-matrix comparison of the regions of AC131056.5 with those of AC003976.1 containing the four genes reveals 90% sequence similarity over 37 kb. These observations support the idea that the multiple copies of *CCL3L1* and *CCL4L1* present in a single diploid genome are the result of segmental duplication.

Chemokines are a family of low-molecular-weight polypeptides that signal through seven-transmembrane, G-protein-coupled receptors. These cytokines regulate a vast array of inflammatory and homeostatic processes, including leukocyte chemotaxis and degranulation, lymphopoiesis, angiogenesis, and tumor growth [1–3]. About 50 chemokine genes are now known and are assigned to four subfamilies based upon the numbers and positions of conserved cysteine residues. The two major subfamilies include the CXC or α subfamily, in which the first two cysteines are separated by an intervening amino acid. The 15 genes in this group (with the exception of *CXCL12*) are clustered at chromosome 4q21. Members of the CC or β subfamily contain two adjacent cysteines, and 16 CC genes have been assigned to a 2.06-Mb interval on contig NT_010799.13 at 17q11.2–q12 while the remaining 8 are dispersed to chromosomes 2, 7, 9, and 16 [2]. The high degree of structural and functional similarity among these genes argues for a series of duplication events in their evolutionary histories.

Included among this group of 16 genes on chromosome 17 are two very closely related pairs, *CCL3–CCL3L1* and *CCL4–CCL4L1*. Members of each pair exhibit 95% nucleotide similarity at the genomic level and encode very similar proteins (*CCL3* and *CCL3L1* are identical at 88 of 93 residues [4], *CCL4* and *CCL4L1* are identical at 89 of 92 amino acids [5]). Functionally, these four proteins are known to bind the CCR5 receptor and limit in vitro infection of cells by HIV-1 [6–9]. Thus, the exact delineation and precise localization of these four genes are necessary to evaluate them in genetic studies aimed at identifying mutations involved with disease.

During the discovery and validation of single-nucleotide polymorphisms (SNPs) in and around these four genes, it became apparent that certain primer pairs were amplifying more than one genomic template [5]. This prompted an analysis of the genomic clones and genomic assembly for these genes. A genomic reference sequence for each gene *CCL3*, D90144 [4]; *CCL3L1*, D90145 [4]; *CCL4*, AC003976 (AC003976 was identified from the original *CCL4* cDNA sequence J04130 [5]); and *CCL4L1*, X53682 [10]) was used to query the NCBI, Celera, and UCSC human genome databases using BLAST or BLAT. Each gene may be distinguished based upon nonsynonymous substitutions

☆ Supplementary data associated with this article can be found in the online version, at doi:10.1016/j.ygeno.2003.09.019.

☆☆ The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

* Fax: +1-301-846-1909.
E-mail address: modi@ncifcrf.gov.

0888-7543/$ - see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2003.09.019

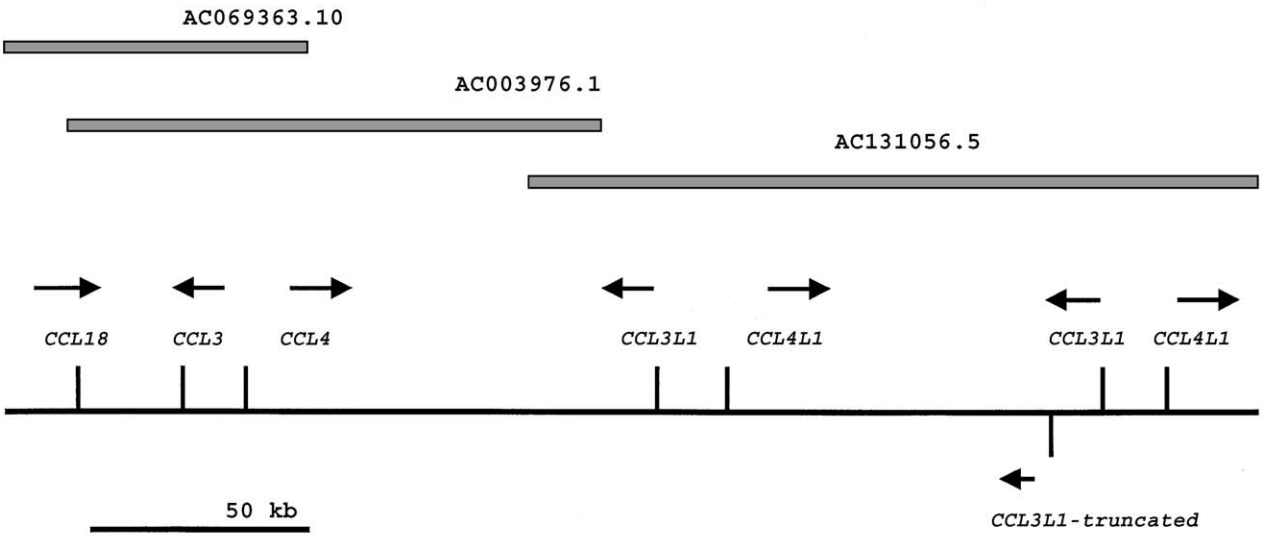
([4,5] and Supplement 1). Multiple alignments were prepared for all four genes (Supplement 1). This is particularly important for *CCL4L1* since it is not recognized as a distinct gene by the UCSC browser.

Nucleotide variation within a gene ranged from 0 to 1.0% (Supplement 2). The nonsynonymous substitutions distinguishing *CCL3* from *CCL3L1* and *CCL4* from *CCL4L1* are fixed in all sequences with the exception of one nucleotide in *CCL3* (Supplement 1). These findings reinforce an earlier conclusion that *CCL4* and *CCL4L1* are distinct genes having no other close relatives in the human genome [5]. Intragenic variation is presumably attributable to sequencing error, single-nucleotide polymorphism, and copy number differences.

CCL3 and *CCL4* were found together on clones AC069363.10 and AC003976.1, being 11,891 bp apart on AC003976.1. One or more copies of *CCL3L1* and *CCL4L1* were found on four clones, AC131056.5, AC036181.3, AC139367.2, and AC130293.3, being 12,552 and 12,557 bp apart on AC131056.5. *CCL4* and proximal *CCL3L1* are 90 kb apart on contig NT_010799.13 (Fig. 1). Of the latter four clones only AC131056.5 has been completely sequenced; the others contain draft sequence of unordered pieces. AC131056.5 was analyzed in further detail.

The July 2003 UCSC browser indicates that AC131056.5 contains two complete copies of *CCL3L1*, and the GenScan predictions NT_010799.243 and NT_010799.247 (Fig. 1). Correct splicing and translation of NT_010799.243 and NT_010799.247 reveals that they encode the functional CCL4L1 protein (Supplement 1) and that the gene name *CCL4L1* should apply to these sequences. Further analysis of AC131056.5 also identified a truncated copy of *CCL3L1* (Fig. 1, Supplement 1). Since AC131056.5 contains multiple copies of *CCL3L1* and *CCL4L1*, different pieces of this clone were aligned with itself to look for internal duplication. Nearly half of this clone exhibits a high degree of internal similarity (Fig. 2). This pattern of internal similarity was also apparent when the unordered pieces of AC036181.3 were compared with themselves and with AC131056.5 (not shown).

Next, since the section of AC003976.1 containing *CCL3* and *CCL4* and that of AC131056.5 containing *CCL3L1* and *CCL4L1* are known to share sequence similarity, dot-matrix analysis was used to compare the entire sequences of AC131056.5 (194 kb) and AC003976.1 (124 kb). Two regions of AC131056.5 (approximately 19–64 kb and 133–175 kb from the beginning of the clone) were similar to a region of AC003976.1 (20–60 kb) with at least 90%




```
CCL3     MQVSTAALAVLLCTMALCNQf.SAsLAADTPTACCFSYTSRQIPQNFIADYFETSSQCSKPgVIFLTKRsRQVCADPSEEWVQKYVSDLELSA
CCL3L1   MQVSTAALAVLLCTMALCNQvlSApLAADTPTACCFSYTSRQIPQNFIADYFETSSQCSKPsVIFLTKRgRQVCADPSEEWVQKYVSDLELSA

CCL4     MKLCVTVLSLLmLVAAFCSpALSAPMGSDPPTACCFSYTARKLPRNFVVDYYETSSLCSQPAVVFQTKRsKQVCADPSESWVQEYVYDLELN
CCL4L1   MKLCVTVLSLLvLVAAFCSlALSAPMGSDPPTACCFSYTARKLPRNFVVDYYETSSLCSQPAVVFQTKRgKQVCADPSESWVQEYVYDLELN
```

Fig. 1. Molecular map of partial NT_010799.13 contig from the July 2003 UCSC browser. Upper diagram illustrates overlap between three completely sequenced BAC clones. Middle section positions CC chemokine genes, with orientations indicated by arrows. The UCSC browser does not list *CCL4L1* as a RefSeq Gene, but rather as NT_010799.243 and NT_010799.247 GenScan predictions. The truncated copy of *CCL3L1* (not in the UCSC browser) starts 39 bp downstream of exon 2 in intron 2 and continues through the 3′ UTR. Lower section presents amino acid alignments with differences between paralogous genes in lowercase and underlined.

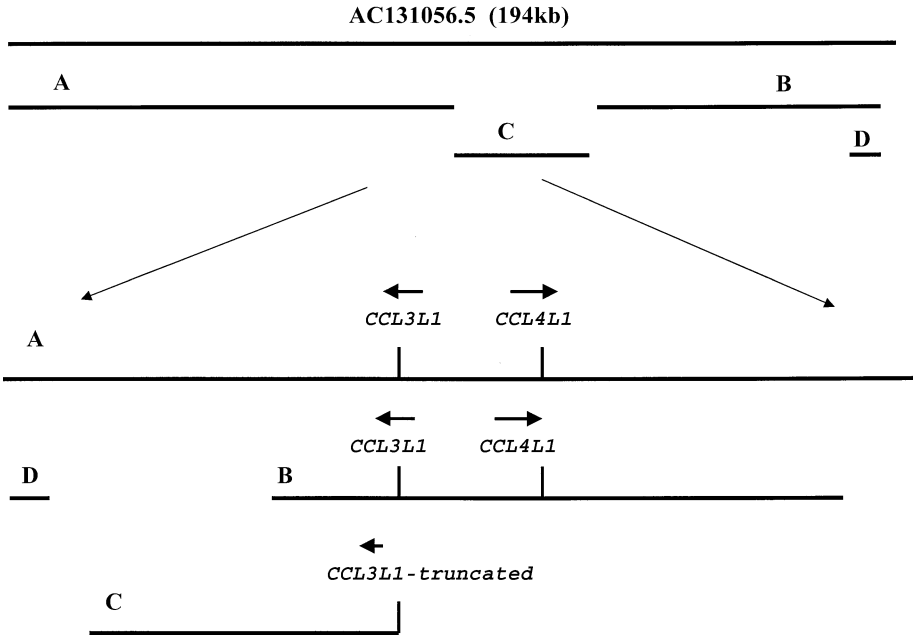


Fig. 2. Internal duplication within clone AC131056.5 (analyzed using Sequencher Software, Gene Codes, Inc.). Map above the arrows shows the positions of four regions labeled A–D. Below the arrows is a multiple sequence alignment in which comparisons A vs B, A vs C, and A vs D each present 99% sequence identity, exclusive of length variation in microsatellites. Region A spans AC131056.5 from bp 1 to 96,800, region B from 129,183 to 193,640, region C from 96,800 to 129,000 and region D from 189,281 to 193,640.

similarity (Fig. 3). This similarity includes all four genes and intervening DNA.

The high level of similarity between AC003976.1 and AC131056.5, common gene strand orientations, comparable intergenic distances, and amino acid similarities between *CCL3–CCL4* and *CCL3L1–CCL4L1* strongly argue for segmental duplication as a mechanism responsible for the origin of one of these gene pairs [11]. Using Southern blotting and quantitative PCR, it was recently shown that the copy number

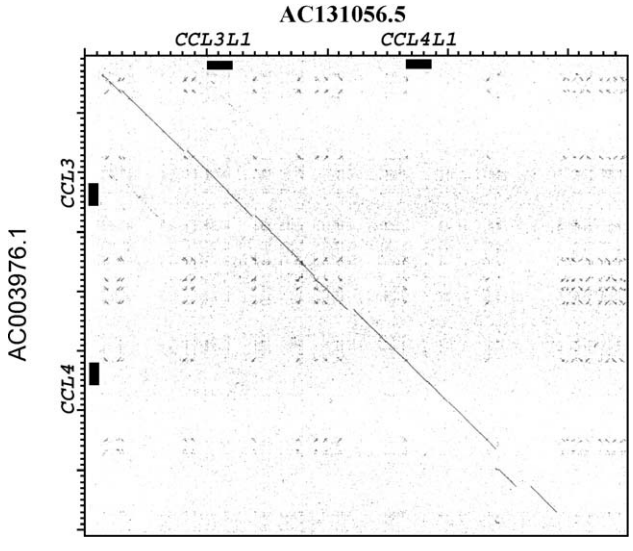


Fig. 3. Dot-matrix analysis (from DOTER [13]) of clones AC131056.5 (bp 18,620 to 63,626) and AC003976.1 (bp 20,325 to 60,378) with average similarity about 90% over the entire interval.

per diploid genome of *CCL3L1* and *CCL4L1* varies between 0 and 6 in a population of Europeans [12]. The haplotype arrangements or chromosomal locations of the duplicate copies were not reported. This observed copy number variation is supported by the present study, which found that two complete copies of each gene and one truncated copy of *CCL3L1* are present on the AC131056.5 haplotype. Of the other three clones containing both genes, only AC130293.3 has more than one copy of either (it has two copies of *CCL3L1*). AC131056.5 and AC130293.3 are both from library RP11 but exhibit minor sequence divergence (Supplements 1 and 2) so that they may represent different haplotypes. Further, preliminary SNP genotyping results (Modi et al.) suggest that different copy numbers exist between European Americans and African Americans, indicating that the overall pattern of duplication in the human species may be quite complex. To understand fully the genomic organization of these duplicate genes, it will be necessary to carry out genomic mapping/sequencing of additional haplotypes containing different gene arrangements/copy numbers.

Databases referred to in this article are at the following URLs: University of California at Santa Cruz, http://genome.ucsc.edu/index.html?org=human; National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov/; and Celera Genomics, Inc., http://www.celera.com/.

Acknowledgments

This project has been funded in whole or in part with federal funds from the National Cancer Institute, National

Institutes of Health, under Contract No. NO1-CO-124000. Dr. K. Scott assisted with computer software, and Dr. B. Gold in querying the Celera database.

References

[1] J.J. Oppenheim, C.O. Zachariae, N. Mukaida, K. Matsushima, Properties of the novel proinflammatory supergene "intercrine" cytokine family, Annu. Rev. Immunol. 9 (1991) 617–648.

[2] A. Zlotnik, O. Yoshie, Chemokines: a new classification system and their role in immunity, Immunity 12 (2000) 121–127.

[3] M. Baggiolini, B. Dewald, B. Moser, Interleukin-8 and related chemotactic cytokines—CXC and CC chemokines, Adv. Immunol. 55 (1994) 97–179.

[4] M. Nakao, H. Nomiyama, K. Shimada, Structures of human genes coding for cytokine LD78 and their expression, Mol. Cell. Biol. 10 (1990) 3646–3658.

[5] W.S. Modi, J. Bergeron, M. Sanford, The human MIP-1beta chemokine is encoded by two paralogous genes, ACT-2 and LAG-1, Immunogenetics 53 (2001) 543–549.

[6] F. Cocchi, A.L. DeVico, A. Garzino-Demo, S.K. Arya, R.C. Gallo, P. Lusso, Identification of RANTES, MIP-1 alpha, and MIP-1 beta as the major HIV-suppressive factors produced by $CD8^+$ T cells, Science 270 (1995) 1811–1815.

[7] T. Miyakawa, K. Obaru, K. Maeda, S. Harada, H. Mitsuya, Identification of amino acid residues critical for LD78beta, a variant of human macrophage inflammatory protein-1alpha, binding to CCR5 and inhibition of R5 human immunodeficiency virus type 1 replication, J. Biol. Chem. 277 (2002) 4649–4655.

[8] R.J. Nibbs, J. Yang, N.R. Landau, J.H. Mao, G.J. Graham, LD78beta, a non-allelic variant of human MIP-1alpha (LD78alpha), has enhanced receptor interactions and potent HIV suppressive activity, J. Biol. Chem. 274 (1999) 17478–17483.

[9] X. Xin, T. Shioda, A. Kato, H. Liu, Y. Sakai, Y. Nagai, Enhanced anti-HIV-1 activity of CC-chemokine LD78beta, a non-allelic variant of MIP-1alpha/LD78alpha, FEBS Lett. 457 (1999) 219–222.

[10] E. Baixeras, S. Roman-Roman, S. Jitsukawa, C. Genevee, S. Mechiche, E. Viegas-Pequignot, T. Hercend, F. Triebel, Cloning and expression of a lymphocyte activation gene (LAG-1), Mol. Immunol. 27 (1990) 1091–1102.

[11] S.G. Irving, P.F. Zipfel, J. Balke, O.W. McBride, C.C. Morton, P.R. Burd, U. Siebenlist, K. Kelly, Two inflammatory mediator cytokine genes are closely linked and variably amplified on chromosome 17q, Nucleic Acids Res. 18 (1990) 3261–3270.

[12] J.R. Townson, L.F. Barcellos, R.J. Nibbs, Gene copy number regulates the production of the human chemokine CCL3-L1, Eur. J. Immunol. 32 (2002) 3016–3026.

[13] E.L. Sonnhammer, R. Durbin, A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis, Gene 167 (1995) GC1–GC10.