Review

The chemokine and chemokine receptor superfamilies and their molecular evolution

Albert Zlotnik*, Osamu Yoshie† and Hisayuki Nomiyama‡

Addresses: *Neurocrine Biosciences, Inc., Department of Molecular Medicine, 12790 El Camino Real, San Diego, CA 92130, USA.
†Department of Microbiology, Kinki University School of Medicine, Osaka-Sayama, Osaka 589-8511, Japan.
‡Department of Biochemistry, Kumamoto University Medical School, Kumamoto 860-0811, Japan.

Correspondence: Albert Zlotnik. Email: albertzlotnik@gmail.com

Abstract

The human chemokine superfamily currently includes at least 46 ligands, which bind to 18 functionally signaling G-protein-coupled receptors and two decoy or scavenger receptors. The chemokine ligands probably comprise one of the first completely known molecular superfamilies. The genomic organization of the chemokine ligand genes and a comparison of their sequences between species shows that tandem gene duplication has taken place independently in the mouse and human lineages of some chemokine families. This means that care needs to be taken when extrapolating experimental results on some chemokines from mouse to human.

The chemokine superfamily includes a large number of ligands that bind to a smaller number of receptors [1,2]. The best known function of the chemokines is the regulation of migration of various cells in the body, hence their name (from 'chemotactic cytokines'). The importance of the chemokines has grown in recent years, as it has become recognized that they are key players in many disease processes, including inflammation, autoimmune disease, infectious diseases (such as HIV/AIDS), and more recently, cancer (in particular in regulating metastasis) [3]. Multiple chemokine ligands can bind to the same receptor; the perceived complexity and promiscuity of receptor binding has often made this field a challenge to understand and given the impression that chemokines lack specific effects. We have now, however, probably identified most human chemokine ligands. The chemokines are small peptides, whereas their receptors are class A G-protein-coupled receptors. They are best known from mammals, but chemokine genes have also been found in chicken, zebrafish, shark and jawless fish genomes, and possible homologs of chemokine receptors have been reported in nematodes. Careful analysis of the members of the superfamily and their receptors shows a logical order to its genomic organization and function, which in turn is the result of evolutionary pressures. Here, we provide a global view of the chemokine and chemokine receptor superfamilies, focusing particularly on the relationship between their evolution and their functions.

The chemokine ligand and receptor superfamilies

As shown in Table 1, there are at least 46 chemokine ligands in humans. There are also 18 functionally signaling chemokine receptors (plus one, CXCR7, which has been recently reported as a potential chemokine receptor) and two 'decoy' or 'scavenger' receptors, DARC and D6, which are known to bind several chemokines but do not signal; their function may be to modulate inflammatory responses through their ability to remove chemokine ligands from inflammatory sites. In the second half of the 1990s, a large number of new ligands were discovered following the growth of expressed sequence tag (EST) databases. The chemokines were easy to recognize from their characteristic structure, containing several (usually four) cysteines in conserved positions, as well as from their relatively small size (8-14 kDa) and from the fact that they are produced in very large amounts by the cells that produce them. Their high expression levels may be due to the way they function, by establishing concentration gradients along which the responding cells migrate. The

Table 1

The chemokine superfamily

Human	Other names	Chromo- some	Function	Cluster	Mouse	Other names	Chromo- some	Function	Cluster	Receptor
CXC family										
CXCL1	Groα	4q13.3	I	GRO	Cxcl1*	Gro/KC	5qE2	I	GRO	CXCR2, CXCR1
CXCL2	Groβ	4q13.3	I	GRO	Cxcl2*	MIP-2	5qE2	I	GRO	CXCR2
CXCL3	Groγ	4q13.3	I	GRO	Gm1960*	Dcip1	5qE2	I	GRO	CXCR2
CXCL4	PF4	4q13.3	U	GRO	Cxcl4*	PF4	5qE2	U	GRO	CXCR3B†
CXCL4V1		4q13.3	U	GRO						
CXCL5	ENA-78	4q13.3	I	GRO	Cxcl5*	LIX	5qE2	I	GRO	CXCR2
CXCL6	GCP-2	4q13.3	I	GRO						CXCR1, CXCR2
CXCL7	NAP-2	4q13.3	I	GRO	Cxcl7	Ppbp	5qE2	I	GRO	Unknown
CXCL8	IL-8	4q13.3	I	GRO	Unknown					CXCR1, CXCR2
CXCL9	MIG	4q21.1	I	IP10	Cxcl9	MIG	5qE3	I	IP10	CXCR3, CXCR3B
CXCL10	IP-10	4q21.1	I	IP10	Cxcl10	IP-10	5qE3	I	IP10	CXCR3, CXCR3B
CXCL11	I-TAC	4q21.1	I	IP10	Cxcl11	I-TAC	5qE3	I	IP10	CXCR3, CXCR3B, CXCR7‡
CXCL12	SDF-1α/β	10q11.21	H		Cxcl12	SDF-1α/β	6qF1	H		CXCR4, CXCR7‡
CXCL13	BLC, BCA-1	4q21.1	H	IP10	Cxcl13	BLC, BCA-1	5qE3	H	IP10	CXCR5
CXCL14	BRAK, Bolekine	5q31.1	I		Cxcl14	BRAK	13qB2	I		Unknown
Unknown					Cxcl15	Lungkine, Weche	5qE2	U		Unknown
CXCL16		17p13.2	I		Cxcl16	Cxcl16	11qB4	I		CXCR6
CXCL17	DMC	19q13.2	U		Cxcl17	DMC	7qA3	U		Unknown
CC family										
CCL1	I-309	17q11.2	I	MCP	Ccl1	TCA-3	11qB5	I	MCP	CCR8
CCL2	MCP-1	17q11.2	I	MCP	Ccl2	JE	11qB5	I	MCP	CCR2
CCL3	MIP-1α, LD78α	17q11.2	I	MIP	Ccl3*	MIP-1α	11qB5	I	MIP	CCR1, CCR5
CCL3L1	LD78β	17q12	I	MIP						
CCL3L3	LD78β	17q12	I	MIP						
CCL4	MIP-1β	17q12	I	MIP	Ccl4*	MIP-1β	11qB5	I	MIP	CCR5
CCL4L1	AT744.2	17q12	I	MIP						
CCL4L2		17q12	I	MIP						
CCL5	RANTES	17q12	I		Ccl5	RANTES	11qB5	I		CCR1, CCR3, CCR5
CCL7	MCP-3	17q11.2	I	MCP	Ccl7	MARC	11qB5		MCP	CCR1, CCR2, CCR3
CCL8	MCP-2	17q11.2	I	MCP	Ccl8*, Ccl12*	MCP-2, MCP-5	11qB5	I	MCP	CCR1, CCR2, CCR3, CCR5
CCL11	Eotaxin	17q11.2	I	MCP	Ccl11	Eotaxin	11qB5	I	MCP	CCR3

Continued on the next page

Table 1 *(continued from the prevoiuus page)*

The chemokine superfamily

| Human	Other names	Chromo-some	Function	Cluster	Mouse	Other names	Chromo-some	Function	Cluster	Receptor
CCL13	MCP-4	17q11.2	I	MCP	Unknown				MCP	CCR1, CCR2, CCR3
CCL14	HCC-1	17q12	H	MIP	Unknown					CCR1
CCL15	HCC-2	17q12	H	MIP	Ccl9	MMRP2, CCF18, MIP-1γ	11qB5	H	MIP	CCR1, CCR3
CCL16	HCC-4 LEC	17q12	H	MIP	Pseudogene		11qB5		MIP	CCR1, CCR2, CCR5, HRH4§
CCL17	TARC	16q13	D	MIP	Ccl17	TARC	8qC5	D		CCR4
CCL18	PARC	17q12	H		Pseudogene					Unknown
CCL19	MIP3β, ELC	9p13.3	H		Ccl19	MIP3β	4qB1	H		CCR7
CCL20	MIP3α, LARC	2q36.3	D		Ccl20	MIP3α, LARC	1qC5	D		CCR6
CCL21	SLC	9p13.3	D		Ccl21a, Ccl21b, Ccl21c*	SLC	4qB1	D		CCR7
CCL22	MDC	16q13	D		Ccl22	ABCD-1	8qC5	D		CCR4
CCL23	MPIF-1	17q12	I	MIP	Ccl6	C10	11qB5	I	MIP	CCR1, FPRL-1¶
CCL24	Eotaxin 2	7q11.23	I		Ccl24	Eotaxin 2	5qG1	I		CCR3
CCL25	TECK	19p13.2	H		Ccl25	TECK	8qA1.2	H		CCR9
CCL26	Eotaxin 3	7q11.23	I		Ccl26l	Eotaxin 3-like	5qG1	I		CCR3
CCL27	CTACK, ILC,	9p13.3	H		Ccl27a,b*	CTACK, ILC	4qB1	H		CCR10
CCL28	MEC	5p12	U		Ccl28	MEC	13	U		CCR10,CCR3
Other classes										
XCL1	Lymphotactin, SCM-1α	1q24.2	D		Xcl1*	Lymphotactin	1qH2	D		XCR1
XCL2	SCM-1β	1q24.2	D							
CX3CL1	Fractalkine	16q13	I		Cx3cl1	Fractalkine	8qC5	I		CX3CR1

Functions are as follows: I, inflammatory; H, homeostatic; D, dual (homeostatic and inflammatory); U, unknown. The lists of alternative names are not comprehensive. Chromosomal location data are derived from the Ensembl [39] or Mouse Genome Informatics [40] databases. GRO, GRO region of the CXC major gene cluster; IP10, IP10 region of the CXC major gene cluster; MCP, MCP region of the CC major gene cluster; MIP, MIP region of the CC major gene cluster. *See also Figure 2. †An alternatively spliced variant of CXCR3 that has been reported to mediate the ability of CXCL4, CXCL9, CXCL10 and CXCL11 to control angiogenesis. ‡Binding has been reported, but signalling is still controversial. §CCL16 has been reported to bind and signal through histamine receptor type 4. ¶A splice variant of CCL23 has been reported to bind to and signal through formyl peptide receptor like-1 (FPRL-1).

most recent human chemokine ligand to be reported (CXCL17, also called dendritic and monocyte chemokine-like protein, DMC) was found by fold-recognition methods [4].

The members of the human and mouse chemokine super-family are listed in Table 1, together with their receptors, and shown in schematic form in Figure 1; phylogenetic trees for the two superfamilies are shown in Figure 2. The two main chemokine ligand superfamiles are named according to the arrangement of the (typically four) cytokines within them: in the CC family, the first two cys-teines near the amino terminus are adjacent, whereas in the CXC family there is one amino acid between them. The human molecules are represented using capital letters, whereas the mouse molecules use lower case, and an L or R is added to indicate ligand or receptor, respectively. For example, CCL5 is the human ortholog of a chemokine pre-viously known as RANTES, Ccl5 is its mouse ortholog and CCR5 is a human receptor for several CCL ligands. Ligands encoded at a given chromosomal location, shown in the same color in Figure 1, usually bind the same receptor.

Some chemokines are produced in very large amounts by many different cell types (for example, CCL2, CCL3 and CCL5), whereas others can have very high specificity for par-ticular tissues or cell types, such as CCL25 (thymus and
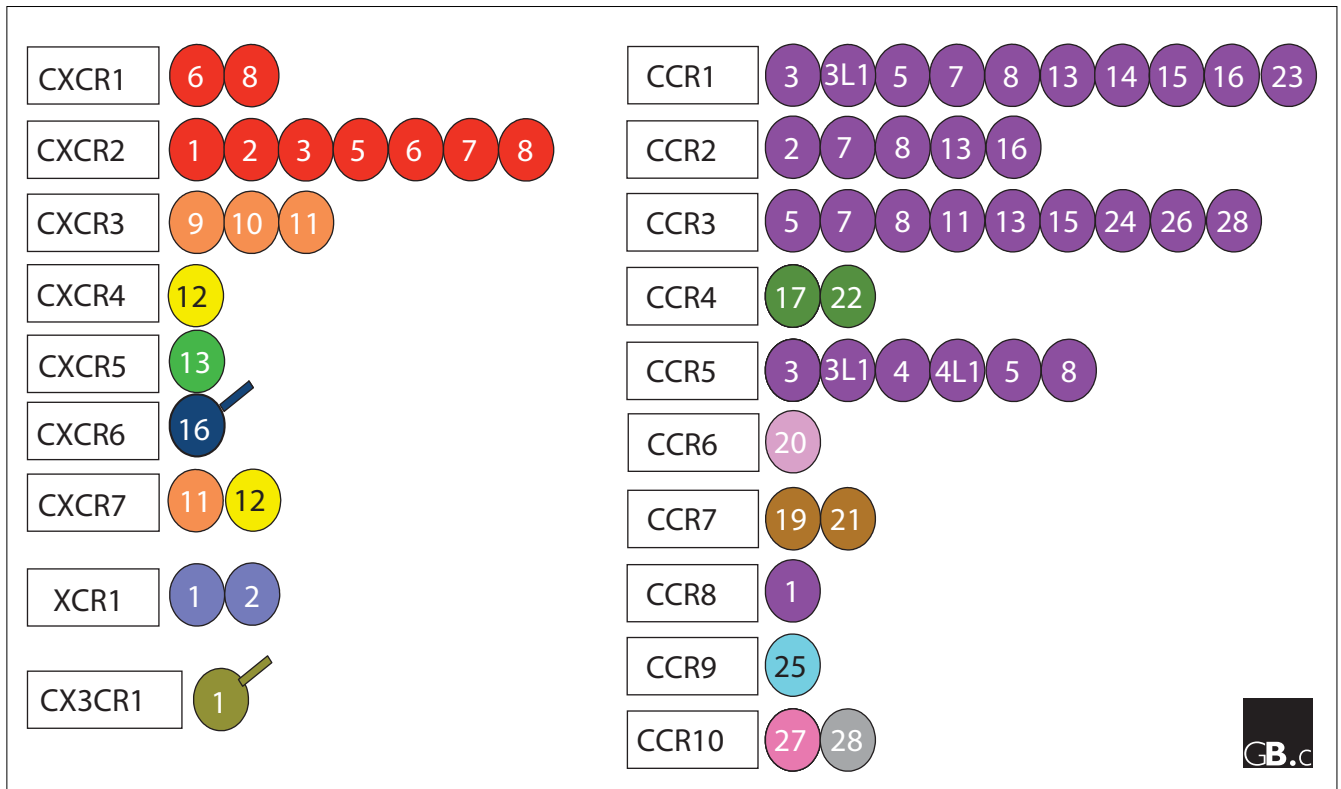
Figure 1
A simplified diagram of the human chemokine superfamily, arranged by the receptors they bind to. Chemokines are represented by only their ligand number, and the receptor name also indicates whether each ligand is a CC or CXC; for example, the '6' adjacent to 'CXCR1' represents CXC6. The colors represent the chromosomal location of the ligands: the genes encoding the ligands shown in the same color are at the same chromosomal location. It can be seen that ligands whose genes are located in the same chromosomal location tend to bind to the same receptor. The extra lines attached to CXCL16 and CX3CL1 mean that these proteins exist as transmembrane proteins.

intestine), CCL27 (skin keratinocytes), CCL28 (certain mucosal epithelial cells) or CXCL17 (stomach and trachea). Other important aspects that differ between chemokines include their biological activities, the regulation of their expression, their receptor-binding specificities and the chromosomal locations of the genes that encode them. These features of the chemokine superfamily have been determined by the forces that have shaped their molecular evolution.

Linking the evolution and function of chemokines
Classification, clustering and gene duplication
The chemokines have been divided into two major groups based on their expression patterns and functions - a useful division, though oversimplified. Those that are expressed by cells of the immune system (leukocytes) or related cells (epithelial and endothelial cells, fibroblasts and so on) only upon activation belong to the 'inflammatory' class, whereas those that are expressed in discrete locations in the absence of apparent activating stimuli have been classified as 'homeostatic' (Table 1). The genomic organization of chemokines (Table 1, Figure 3) also enables us, however, to divide

chemokines into two alternative groups: those whose genes are located in large clusters at particular chromosomal locations (the 'major-cluster' chemokines; Figure 3a) and the 'non-cluster' or 'mini-cluster' chemokines whose genes are located separately in unique chromosomal locations (Figure 3b,c) [2]. There are two major clusters of CC chemokine genes and two of CXC genes, plus numerous non-clustered or mini-cluster genes of both types, in both the mouse and human genomes (Figure 3).

An explanation for this chromosomal arrangement is found in the evolutionary forces that have shaped the genome into gene superfamilies [5]. Over the course of evolution, gene duplication has been a common event, affecting most gene families [6]. Once a duplication occurs, the two copies can evolve independently and develop specialized functions. This explains the origin of the cluster chemokines, which show two other characteristics that do not apply to the non-cluster or mini-cluster chemokines: first, the members of a given gene cluster usually bind to multiple receptors and vice versa (the complex and promiscuous ligand-receptor relationships; Figure 1); and second, cluster chemokines
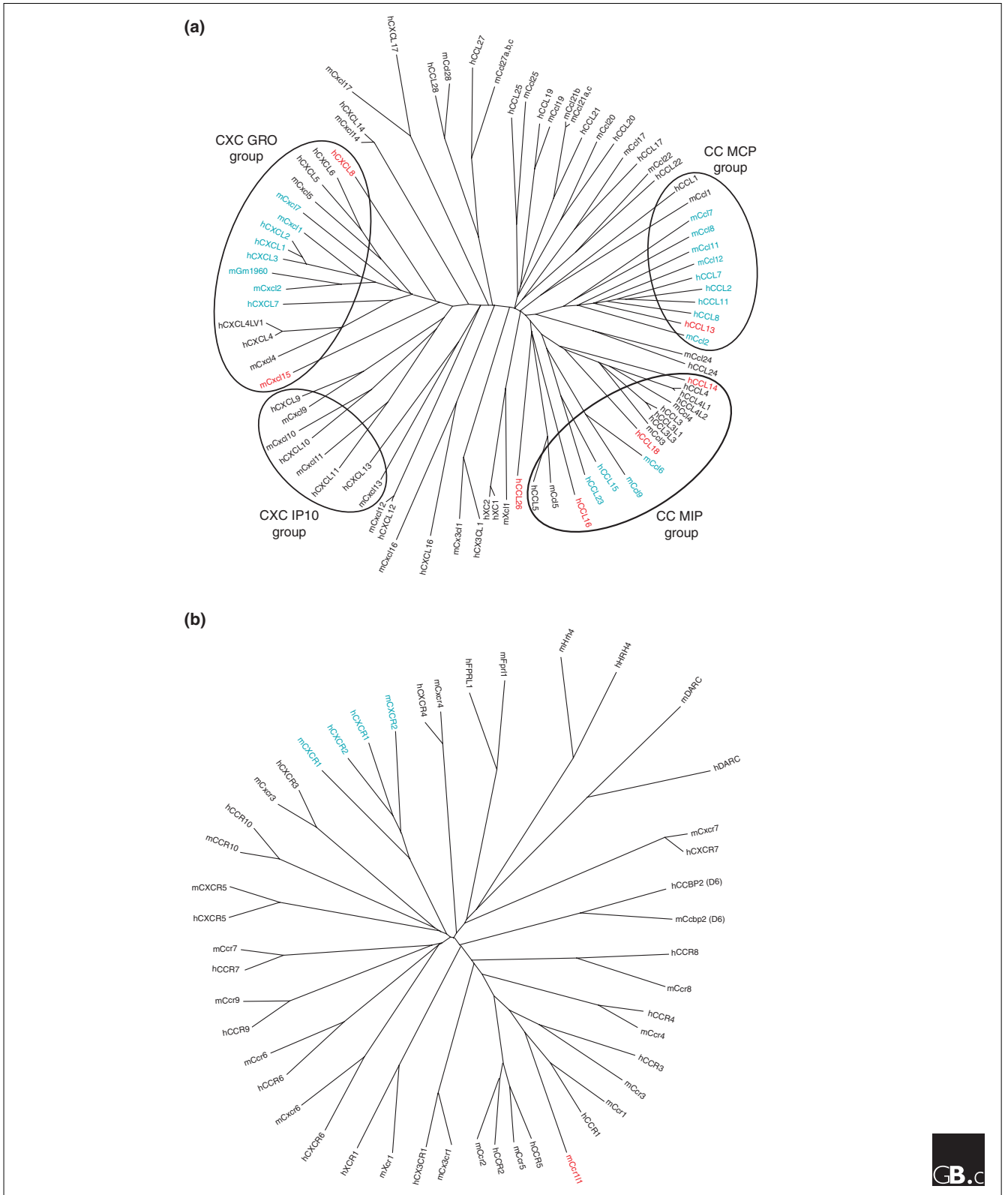
Figure 2
Sequence relationship analysis of the human (h) and mouse (m) **(a)** chemokines and **(b)** chemokine receptors. Phylogenetic trees were constructed using amino acid sequences with Clustal X and PAUP* (the neighbor joining method) programs [37]. In (a), the GRO and IP10 groups of CXC chemokines and the MCP and MIP groups of CC chemokines (see also Figure 3) are circled. Red letters indicate proteins that are found in only mouse or human but not the other. Blue letters indicate proteins for which the relationships are uncertain.
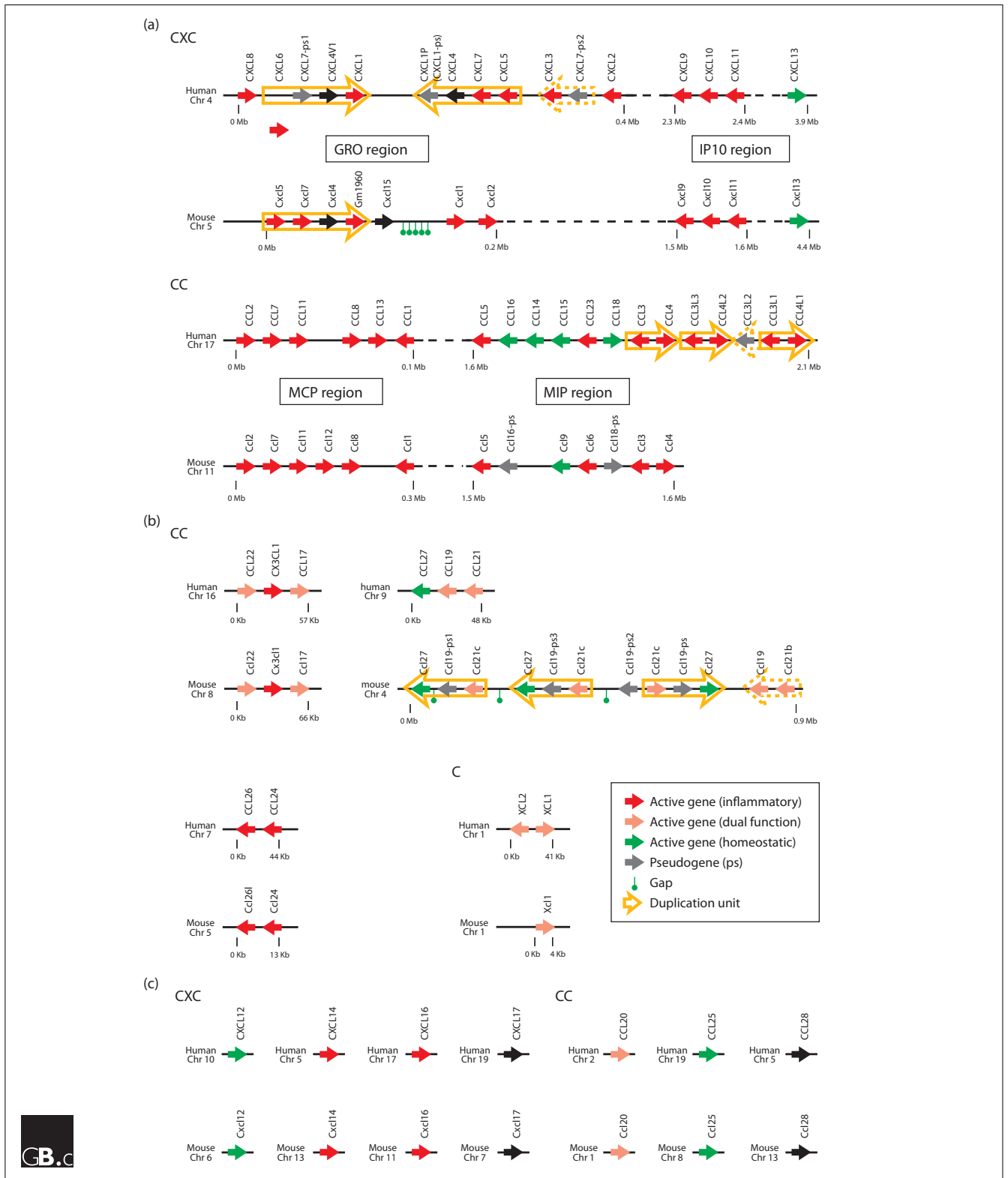
Figure 3
Schematic genomic organization of the human and mouse chemokine superfamily. **(a)** Major-cluster chemokines; **(b)** mini-cluster chemokines; **(c)** non-cluster chemokines. Solid arrows indicate chemokine genes and their transcriptional orientation; red, green and pink arrows indicate inflammatory, homeostatic and dual function chemokine genes, respectively, and gray arrows indicate pseudogenes. Duplication units in the major clusters are indicated by open yellow arrows. This figure is based on the NCBI 36 and 35 assemblies of the human and mouse genomes [38]. A gap indicates a region not yet covered by the genome sequencing consortiums, while a dashed line denotes a similar region of more than 1 Mb.

often do not correspond well between species (for example, between human and mouse) [2].

These two characteristics can be explained as follows: the cluster chemokines and their receptors multiplied from their ancestral genes by a series of tandem gene-duplication events that occurred relatively recently in evolutionary terms, that is, even after the branching of human and mouse [2]. This is apparent from the phylogenetic tree shown in Figure 2, in which the cluster chemokines form compact clusters termed groups: the monocyte chemotactic protein (MCP) group, the macrophage inflammatory protein (MIP) group (both of CC chemokines), and the GRO group and the IP-10 group (both of CXC chemokines). This common evolutionary origin suggests that the cluster chemokines are a group of proteins sharing a common primary function. In the case of the chemokines encoded by the CXC GRO cluster on chromosome 4, which in human includes CXCL1-CXCL8, the primary function is the regulation of neutrophil recruitment to inflammatory sites [7]. The chemokines in this cluster do this through interaction with CXCR1 and CXCR2 (Table 1, Figure 1). Similarly, the main function of the cytokines encoded in the MIP and MCP clusters of CC chemokines in human chromosome 17, which includes CCL1-CCL16, CCL18 and CCL23, is the recruitment of monocytes, subsets of T cells, eosinophils, and so on, to sites where inflammation is developing, through their interaction with CCR1, CCR2, CCR3 and/or CCR5 (Table 1, Figure 1).

Functional reasons for clustering

An explanation for the large number of ligands for these receptors is that, during inflammation, multiple chemokines can be needed to induce a robust leukocyte response [2]. Furthermore, differential expression of these chemokines among different tissues may finely orchestrate the recruitment of leukocytes to the tissues and could enable a 'customization' of the inflammatory responses. Accordingly, most cluster chemokines belong to the inflammatory category [2].

Clustering and its consequences could provide a critical survival advantage to a species faced with a particular infectious agent. For example, CCR5 expression has recently been shown to be pivotal in resistance to infection with the West Nile virus in humans [8]. The protective mechanism of CCR5 may involve directing leukocytes to the brain, where they can fight the infection more effectively [9]. Another hypothesis, however, involves 'viral' chemokines, believed to be mammalian genes that were at some point 'hijacked' by viruses. To cope with the proliferation of such viral chemokines, mammals may have increased the numbers of their own endogenous chemokines to circumvent the effects of the viral molecules. For example, humans have *CCL3L1* and *CCL4L* genes, which are homologs of *CCL3* and *CCL4* [10] and are found in a unit of zero to three copies depending on the individual (Figure 3a); CCL3L1 has an affinity for CCR5

ten times higher than that of CCL3 [11]. This higher affinity ligand would give an evolutionary advantage for an organism when coping with viral infections.

These hypotheses also explain the lack of correspondence between cluster chemokine ligands in mouse and human, which may reflect the 'infectious experience' of the two species after they separated. This effect is shown graphically in the separation of the human and mouse chemokine clusters in the phylogenetic tree shown in Figure 2: in the groups of chemokines there is often no one-to-one correspondence between human and mouse genes or the relationships between them may be uncertain. This evolution is ongoing, and it is therefore possible that variations in these genes will be documented even among relatively close species.

The only CC cluster chemokine that has a one-to-one ligand/ receptor relationship (with CCR8) is CCL1 (Figure 1, Table 1). Its specific receptor, CCR8, is expressed by monocytes, activated helper Th2 cells and natural killer T cells, CD4+ thymocytes [12], regulatory T cells [13], normal skin-homing T cells [14], skin-homing γδ T cells and CD56+ CD16- natural killer cells [15]. The *CCL1* gene is located in the MCP subregion (Figure 3a) but is rather distantly related to other members of the MCP group (Figure 2a), suggesting that it was generated much earlier than the rest of the cluster chemokines in this region. In fact, CCL1 may represent an early chemokine that branched before the CC cluster chemokines in the phylogenetic tree (Figure 2a). It is therefore possible that this chemokine-receptor pair has specific roles in shaping the immune system [16] and, in this context, its expression by T regulatory cells [13] is intriguing.

Non-cluster and mini-cluster chemokines

By contrast, the non- cluster or mini-cluster chemokines are relatively conserved between species and tend not to act on multiple receptors (Table 1, Figure 1). Indeed, several of these have a single ligand-receptor relationship, such as CCL25-CCR9 or CXCL13-CXCR5. The evolutionary model described above predicts that these particular chemokine ligand-receptor pairs probably have pivotal roles in the development of the organism or in the function of physiological systems necessary for the organism's survival to reproductive age (in other words, they are under evolutionary pressure). In support of this hypothesis, the genes for most homeostatic chemokines are found in non-cluster chromosomal locations (Table 1, Figure 3b,c). For example, CXCR4-deficient and CXCL12-deficient mice both have a lethal phenotype, and their embryos have various defects in critical organs, such as the heart, brain or bone marrow [17]. Therefore, throughout evolution, several non-cluster chemokines have participated in organogenesis, and their critical functions must be conserved in order for the species to survive. Another example is the CXCL13-CXCR5 pair, which is pivotal for successful B cell homing and, because it regulates T cell-B cell interactions, for the production of antibodies

[18]. Thus, evolutionary pressure selects against changes in these genes by preventing them from diverging from their original function.

Early chemokines

In contrast to the cluster chemokines, the non-cluster and mini-cluster chemokines have been conserved throughout evolution and are therefore thought to be more 'ancestral' genes. This prediction is also supported by the phylogenetic tree shown in Figure 2, in which non-cluster and mini-cluster chemokines branch much earlier than the major-cluster chemokines and each human chemokine of this type has a clearly identifiable mouse counterpart [2]. There are data to support this model. Two groups have reported that, in the zebrafish, the CXCL12-CXCR4 pair regulates the homing of primordial germ cells to the gonads, where they differentiate into gametes [19,20]. Importantly, the G-protein-coupled receptor Odysseus is readily recognizable as the zebrafish ortholog of CXCR4; 61% of the amino acid residues are identical between the zebrafish and human sequences (Figure 4). Similarly, the zebrafish ortholog of CXCL12 (with a remarkable 47% of residues in the coding region being identical; Figure 4) is also easy to identify.

The zebrafish genome contains many other chemokine genes, including those with the GenBank accession numbers NM131627 and NM131062 [21], yet, in contrast to CXCL12, the correspondence of these molecules with human chemokines is not easy to establish. These observations underscore the importance of the CXCR4-CXCL12 pair throughout vertebrate evolution. GenBank now includes many chemokine gene entries from various genomes, including many mammals, shark, fish (including zebrafish) and even what may be homologs of chemokine receptor genes in *Caenorhabditis elegans* [22]. Another notable example is the chemokine LFCA-1 identified from the genome of the river lamprey (a jawless fish), which shows 46-49% identity to the chicken orthologs of CXCL8, K60 and 9E3 [23], and also has homology with human CXCL8 (Figure 4).

This interspecies genomic analysis will eventually help us understand the evolutionary history of the chemokine super-family and may even allow us to identify a 'primordial' chemokine gene. It should be interesting to identify what the original function of this ancestral chemokine gene could have been. The function of the CXCR4-CXCL12 pair in the zebrafish in primordial germ cell homing suggests that chemokines and their receptors first arose as molecules controlling the transit of various cells within organisms simpler than mammals, and suggests that chemokines and their receptors have key roles in cellular transit *in vivo* during embryogenesis and/or in the adult organism. Another area of intense research is the function of chemokines in the development and function of the central nervous system [24]. This primary function in cellular traffic *in vivo* also supports a role for chemokines in cancer metastasis [25].

Recently, Balabanian *et al.* [26] reported the identification of a second human receptor (RDC-1) that binds CXCL12, the characterization of this receptor is on going, but it may also bind CXCL11. The sequence and characteristics of this receptor indicates that it belongs to the CXC receptor family and, as such, it should be named CXCR7. Its expression is more restricted than that of CXCR4, and it will be interesting to characterize its function in detail. RDC-1 may have another ligand [27], however, and it might, therefore, not be specific for CXCL12. Its capacity to bind CXCL12 suggests that it may represent another receptor (besides CXCR4) with important functions even in simpler organisms.

Mini-cluster chemokines and gene translocations

The evolution of the chemokines is an ongoing process, and there are examples of ligands forming 'mini-clusters' as well as major clusters (Figure 2b). One of these includes the *CXCL9*, *CXCL10* and *CXCL11* genes, which are located in the CXC IP-10 inflammatory cluster (4q21.21). The chemokines they encode function in T-cell recruitment through CXCR3 [28] and also in the negative control of angiogenesis through CXCR3B, an alternatively spliced variant of CXCR3 [29]. Another mini-cluster includes *CCL19* and *CCL21*, which are located in close proximity (9p13 in human) and whose encoded chemokines share a receptor, CCR7. Likewise, human *CCL17* and *CCL22* are located in close proximity (16q13 in human) and their chemokines share a receptor (CCR4). Interestingly, another protein encoded in the same mini-cluster as *CCL17* and *CCL22*, CX3CL1 (previously called fractalkine) is totally different from them: it is a transmembrane-type chemokine with the CX3C motif (two cysteines separated by three amino acids) instead of the CC motif and interacts specifically with CX3CR1 (Figure 1, Table 1). The position of *CX3CL1* is probably due to its translocation from elsewhere to between *CCL17* and *CCL22* (Figure 3b).

Another example of a translocation is *CCL27*, which maps in close vicinity to *CCL19* and *CCL21* (Figure 3b) but does not share CCR7 with the encoded chemokines (Table 1). Instead, CCL27 is most similar to CCL28, and they share CCR10 (Table 1). Thus, it is possible that *CCL27* was originally located in chromosome 5p12 and may have translocated to its present site. Alternatively, the location of the *CCL27* gene could be explained by the fact that the gene for the α chain of the interleukin 11 receptor is located on this site but in opposite orientation [30], indicating that this locus has been subjected to multiple evolutionary forces. Further evidence that chemokine evolution is ongoing is provided by XCL1 and XCL2 (previously called lymphotactin), which are the result of a recent gene duplication as they only differ by one amino acid [31] and they share the receptor XCR1 [32] (Figure 3b, Table 1). Another example (in the mouse) is Ccl21, which is encoded by three different genes that differ in one amino acid codon and are expressed in distinct anatomical locations [33].
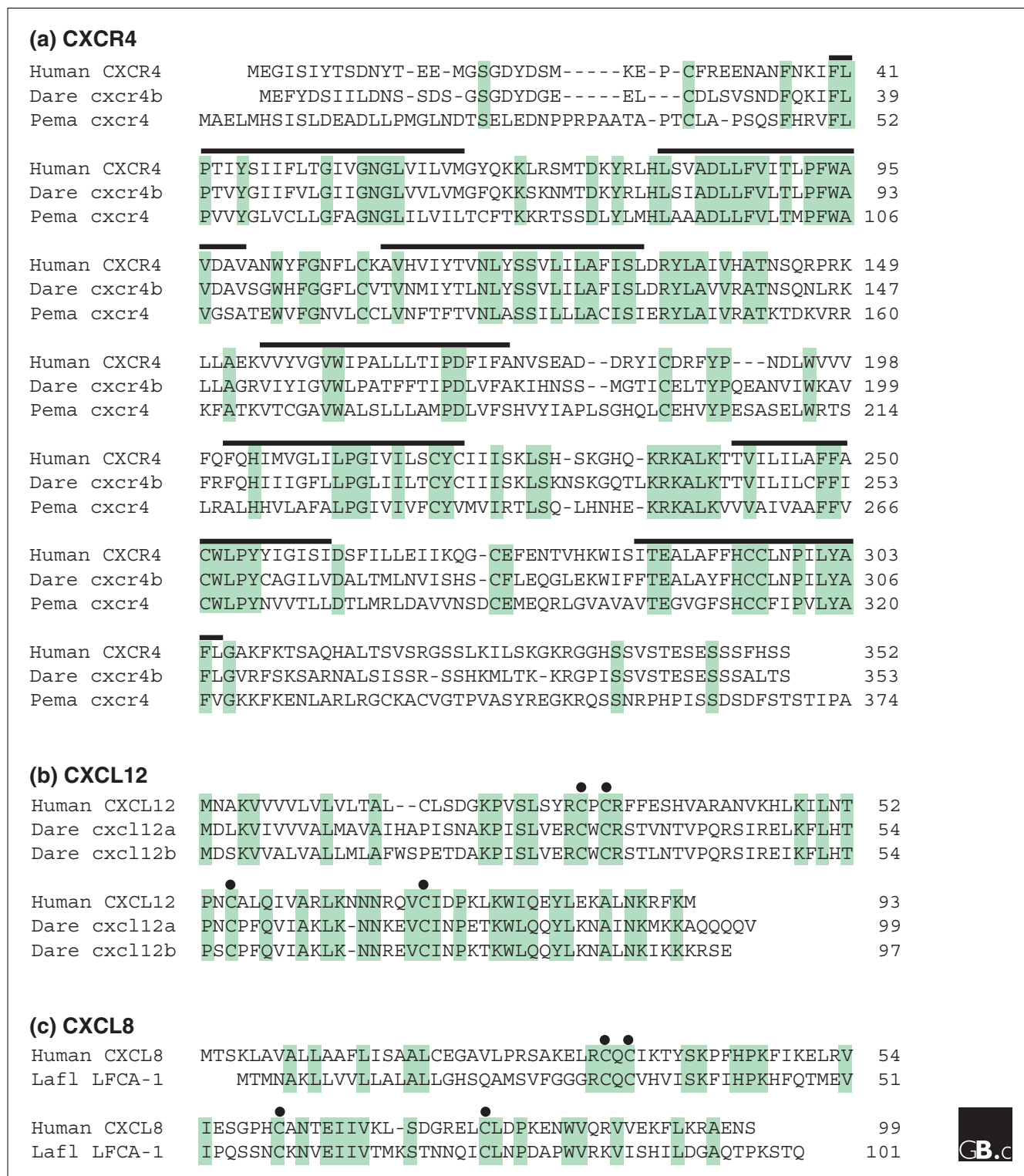
(a) CXCR4

```
Human CXCR4      MEGISIYTSDNYT-EE-MGSGDYDSM-----KE-P-CFREENANFNKIFL  41
Dare cxcr4b       MEFYDSIILDNS-SDS-GSGDYDGE-----EL---CDLSVSNDFQKIFL  39
Pema cxcr4       MAELMHSISLDEADLLPMGLNDTSELEDNPPRPAATA-PTCLA-PSQSFHRVFL 52

Human CXCR4      PTIYSIIFLTGIVGNGLVILVMGYQKKLRSMTDKYRLHLSVADLLFVITLPFWA  95
Dare cxcr4b      PTVYGIIFVLGIIGNGLVVLVMGFQKKSKNMTDKYRLHLSIADLLFVLTLPFWA  93
Pema cxcr4       PVVYGLVCLLGFAGNGLILVILTCFTKKRTSSDLYLMHLAAADLLFVLTMPFWA 106

Human CXCR4      VDAVANWYFGNFLCKAVHVIYTVNLYSSVLILAFISLDRYLAIVHATNSQRPRK 149
Dare cxcr4b      VDAVSGWHFGGFLCVTVNMIYTLNLYSSVLILAFISLDRYLAVVRATNSQNLRK 147
Pema cxcr4       VGSATEWVFGNVLCCLVNFTFTVNLASSILLLACISIERYLAIVRATKTDKVRR 160

Human CXCR4      LLAEKVVYVGVWIPALLLTIPDFIFANVSEAD--DRYICDRFYP---NDLWVVV 198
Dare cxcr4b      LLAGRVIYIGVWLPATFFTIPDLVFAKIHNSS--MGTICELTYPQEANVIWKAV 199
Pema cxcr4       KFATKVTCGAVWALSLLLAMPDLVFSHVYIAPLSGHQLCEHVYPESASELWRTS 214

Human CXCR4      FQFQHIMVGLILPGIVILSCYCIIISKLSH-SKGHQ-KRKALKTTVILILAFFA 250
Dare cxcr4b      FRFQHIIIGFLLPGLIILTCYCIIISKLSKNSKGQTLKRKALKTTVILILCFFI 253
Pema cxcr4       LRALHHVLAFALPGIVIVFCYVMVIRTLSQ-LHNHE-KRKALKVVVAIVAAFFV 266

Human CXCR4      CWLPYYIGISIDSFILLEIIKQG-CEFENTVHKWISITEALAFFHCCLNPILYA 303
Dare cxcr4b      CWLPYCAGILVDALTMLNVISHS-CFLEQGLEKWIFFTEALAYFHCCLNPILYA 306
Pema cxcr4       CWLPYNVVTLLDTLMRLDAVVNSDCEMEQRLGVAVAVTEGVGFSHCCFIPVLYA 320

Human CXCR4      FLGAKFKTSAQHALTSVSRGSSLKILSKGKRGGHSSVSTESESSSFHSS      352
Dare cxcr4b      FLGVRFSKSARNALSISSR-SSHKMLTK-KRGPISSVSTESESSSALTS      353
Pema cxcr4       FVGKKFKENLARLRGCKACVGTPVASYREGKRQSSNRPHPISSDSDFSTSTIPA 374
```

(b) CXCL12

```
Human CXCL12     MNAKVVVVLVLVLTAL--CLSDGKPVSLSYRCPCRFFESHVARANVKHLKILNT  52
Dare cxcl12a     MDLKVIVVVALMAVAIHAPISNAKPISLVERCWCRSTVNTVPQRSIRELKFLHT  54
Dare cxcl12b     MDSKVVALVALLMLAFWSPETDAKPISLVERCWCRSTLNTVPQRSIREIKFLHT  54

Human CXCL12     PNCALQIVARLKNNNRQVCIDPKLKWIQEYLEKALNKRFKM              93
Dare cxcl12a     PNCPFQVIAKLK-NNKEVCINPETKWLQQYLKNAINKMKKAQQQQV          99
Dare cxcl12b     PSCPFQVIAKLK-NNREVCINPKTKWLQQYLKNALNKIKKKRSE            97
```

(c) CXCL8

```
Human CXCL8      MTSKLAVALLAAFLISAALCEGAVLPRSAKELRCQCIKTYSKPFHPKFIKELRV  54
Lafl LFCA-1       MTMNAKLLVVLLALALLGHSQAMSVFGGGRCQCVHVISKFIHPKHFQTMEV     51

Human CXCL8      IESGPHCANTEIIVKL-SDGRELCLDPKENWVQRVVEKFLKRAENS          99
Lafl LFCA-1      IPQSSNCKNVEIIVTMKSTNNQICLNPDAPWVRKVISHILDGAQTPKSTQ     101
```

GB.c

Figure 4
Chemokine and chemokine receptor sequences, such as **(a)** CXCR4, **(b)** CXCL12 and **(c)** CXCL8, are highly conserved throughout evolution, from jawless fish to humans. Identical amino acid residues are highlighted in green; the seven transmembrane regions of the receptors are indicated by black lines; the four conserved cysteine residues are indicated by dots above the sequences. Species abbreviations: dare, *Danio rerio* (zebrafish); pema, *Petromyzon marinus* (sea lamprey); lafl, *Lampetra fluviatilis* (European river lamprey). Accession numbers (from GenBank) are as follows: human CXCR4, NM_003467; zebrafish cxcr4b, NM_131834; sea lamprey cxcr4, AY178969; human CXCL12, NM_000609; zebrafish cxcl12a, NM_178307; zebrafish cxcl12b, NM_198068; human IL-8, NM_000584; river lamprey CXCL8, AJ231072.

Of mice and men

The mouse is generally considered a valuable model for human diseases. The completion of the mouse genome supports this view, because it seems to be remarkably similar to the human genome [34]. Analysis of the human and mouse genomes has revealed that the genes involved in immune and host defense roles are under positive selection pressure, accumulating amino acid changes more rapidly than other genes. Chemokines are listed as one of the eight most rapidly changing proteins and domains [35]. Examination of the gene organization of human and mouse chemokine clusters also shows great divergence (Figure 3) [36]. The following are three important differences.

First, some chemokine genes exist in one species but not the other. This is the most dramatic example of lack of correlation between species and applies specifically to the inflammatory/cluster chemokines. Table 1 and Figure 3a show that, in the CXC subfamily, CXCL8 does not have a mouse counterpart, whereas Cxcl15 exists in the mouse but not in human. Among the CC subfamily (Figure 3b), CCL13 and CCL14 exist in the human but not in the mouse. Alternatively, a given gene in one species (for example, CCL16 and CCL18) may be represented by a pseudogene in the other.

Second, a given chemokine may be related to (or represented) by more than one ortholog in the other species (Table 1). This is due to independent duplication events that have occurred in one of the species. Human *XCL1* and *XCL2* and the varying number copies of human CCL3 and CCL4 and of mouse Ccl27, Ccl19 and Ccl21 described above are examples of this.

Third, there can be similar genes in the two species but they may not be 'exact' structural or functional equivalents. One of the best examples of the latter is the MCP group. Structurally, it is difficult to assign a human counterpart unambiguously to each mouse gene, because they are all closely related molecules that probably arose independently in each species (Figure 2a).

Differences like these may result in important differences in the function of chemokines between species. These potential differences do not, however, exclude the mouse as a valid model for human disease. But they do mean that there are limitations to the extrapolations we can make when using mouse models to understand human disease. It is worth emphasizing that these differences may be particularly important in studies of inflammatory diseases, which involve the inflammatory chemokines (most of which are major-cluster cytokines), and less so in experiments designed to understand the function of homeostatic chemokines, which, because they are generally noncluster cytokines and thus more conserved between species, should be more readily applicable to the human system.

The progress in the discovery and characterization of chemokines has been remarkable, and we are approaching the completion of the discovery phase of many other molecular superfamilies. The sudden availability of so many new molecules is an excellent opportunity for understanding the roles of chemokines, not only in the immune system, but also in development and general physiology. Analysis of the syntenic genomic regions between mouse and human has enabled investigation of the relationships between the chemokines of these species. The mouse is a popular model for investigating gene function, but it is important that the significant differences in the chemokine ligand superfamily between mouse and human are taken into account, especially as the ability to extrapolate mouse data to human disease depends on the gene under study. This type of analysis should be applicable to other molecular superfamilies. It is our hope that the issues we have discussed here will facilitate understanding of the biology of the chemokine superfamily.

References

1. Zlotnik A, Yoshie O: **Chemokines: a new classification system and their role in immunity.** *Immunity* 2000, **12**:121-127.
2. Yoshie O, Imai T, Nomiyama H: **Chemokines in immunity.** *Adv Immunol* 2001, **78**:57-110.
3. Muller A, Homey B, Soto H, Ge N, Catron D, Buchanan M, McClanahan T, Murphy E, Yuan Y, Wagner S, *et al.*: **Involvement of chemokine receptors in breast cancer metastasis.** *Nature* 2001, **410**:50-56.
4. Pisabarro MT, Leung B, Kwong M, Corpuz R, Frantz GD, Chiang N, Vandlen R, Diehl LJ, Skelton N, Kim HS, *et al.*: **Cutting edge: novel human dendritic cell- and monocyte-attracting chemokine-like protein identified by fold recognition methods.** *J Immunol* 2006, **176**:2069-2073.
5. Thornton JW, DeSalle R: **Gene family evolution and homology: genomics meets phylogenetics.** *Annu Rev Genomics Hum Genet* 2000, **1**:41-73.
6. Wagner A: **Birth and death of duplicated genes in completely sequenced eukaryotes.** *Trends Genet* 2001, **17**:237-239.
7. Patel L, Charlton SJ, Chambers JK, Macphee CH: **Expression and functional analysis of chemokine receptors in human peripheral blood leukocyte populations.** *Cytokine* 2001, **14**:27-36.
8. Glass WG, McDermott DH, Lim JK, Lekhong S, Yu SF, Frank WA, Pape J, Cheshier RC, Murphy PM: **CCR5 deficiency increases risk of symptomatic West Nile virus infection.** *J Exp Med* 2006, **203**:35-40.
9. Glass WG, Lim JK, Cholera R, Pletnev AG, Gao JL, Murphy PM: **Chemokine receptor CCR5 promotes leukocyte trafficking to the brain and survival in West Nile virus infection.** *J Exp Med* 2005, **202**:1087-1098.
10. Irving SG, Zipfel PF, Balke J, McBride OW, Morton CC, Burd PR, Siebenlist U, Kelly K: **Two inflammatory mediator cytokine genes are closely linked and variably amplified on chromosome 17q.** *Nucleic Acids Res* 1990, **18**:3261-3270.
11. Nibbs RJ, Yang J, Landau NR, Mao JH, Graham GJ: **LD78beta, a non-allelic variant of human MIP-1alpha (LD78alpha), has enhanced receptor interactions and potent HIV suppressive activity.** *J Biol Chem* 1999, **274**:17478-17483.
12. Zingoni A, Soto H, Hedrick JA, Stoppacciaro A, Storlazzi CT, Sinigaglia F, D'Ambrosio D, O'Garra A, Robinson D, Rocchi M, *et al.*: **The chemokine receptor CCR8 is preferentially expressed in Th2 but not Th1 cells.** *J Immunol* 1998, **161**:547-551.
13. Iellem A, Mariani M, Lang R, Recalde H, Panina-Bordignon P, Sinigaglia F, D'Ambrosio D: **Unique chemotactic response profile and specific expression of chemokine receptors CCR4 and

CCR8 by CD4(+)CD25(+) regulatory T cells. *J Exp Med* 2001, **194:**847-853.

14. Schaerli P, Ebert L, Willimann K, Blaser A, Roos RS, Loetscher P, Moser B: **A skin-selective homing mechanism for human immune surveillance T cells.** *J Exp Med* 2004, **199:**1265-1275.

15. Ebert LM, Meuter S, Moser B: **Homing and function of human skin gammadelta T cells and NK cells: relevance for tumor surveillance.** *J Immunol* 2006, **176:**4331-4336.

16. Chensue SW, Lukacs NW, Yang TY, Shang X, Frait KA, Kunkel SL, Kung T, Wiekowski MT, Hedrick JA, Cook DN, *et al.*: **Aberrant *in vivo* T helper type 2 cell response and impaired eosinophil recruitment in CC chemokine receptor 8 knockout mice.** *J Exp Med* 2001, **193:**573-584.

17. Zou Y, Kottmann A, Kuroda M, Taniuchi I, Littman D: **Function of the chemokine receptor CXCR4 in haematopoiesis and in cerebellar development.** *Nature* 1998, **393:**595-599.

18. Breitfeld D, Ohl L, Kremmer R, Ellwart J, Sallusto F, Lipp M, Forster R: **Follicular B helper T cells express CXC chemokine receptor 5, localize to B cell follicles and support immunoglobulin production.** *J Exp Med* 2000, **192:**1545-1552.

19. Doitsidou M, Reichman-Fried M, Stebler J, Koprunner M, Dorries J, Meyer D, Esguerra C, Leung T, Raz E: **Guidance of primordial germ cell migration by the chemokine SDF-1.** *Cell* 2002, **111:**647-659.

20. Knaut H, Werz C, Geisler R, Nusslein-Volhard C: **A zebrafish homologue of the chemokine receptor CXCR4 is a germ cell guidance receptor.** *Nature* 2003, **421:**279-282.

21. DeVries ME, Kelvin AA, Xu L, Ran L, Robinson J, Kelvin DJ: **Defining the origins and evolution of the chemokine/chemokine receptor system.** *J Immunol* 2006, **176:**401-415.

22. Robertson H: **Two large families of chemoreceptor genes in the nematodes *Caenorhabditis elegans* and *Caenorhabditis briggsae* reveal extensive gene duplication, diversification, movement, and intron loss.** *Genome Res* 1998, **8:**449-463.

23. Najakshin AM, Mechetina LV, Alabyev BY, Taranin AV: **Identification of an IL-8 homolog in lamprey (*Lampetra fluviatilis*): early evolutionary divergence of chemokines.** *Eur J Immunol* 1999, **29:**375-382.

24. Tran PB, Miller RJ: **Chemokine receptors: signposts to brain development and disease.** *Nat Rev Neurosci* 2003, **4:**444-455.

25. Zlotnik A: **Chemokines in neoplastic progression.** *Semin Cancer Biol* 2004, **14:**181-185.

26. Balabanian K, Lagane B, Infantino S, Chow KY, Harriague J, Moepps B, Arenzana-Seisdedos F, Thelen M, Bachelerie F: **The chemokine SDF-1/CXCL12 binds to and signals through the orphan receptor RDC1 in T lymphocytes.** *J Biol Chem* 2005, **280:**35760-35766.

27. Infantino S, Moepps B, Thelen M: **Expression and regulation of the orphan receptor RDC1 and its putative ligand in human dendritic and B cells.** *J Immunol* 2006, **176:**2197-2207.

28. Belperio JA, Keane MP, Burdick MD, Lynch JP 3rd, Xue YY, Li K, Ross DJ, Strieter RM: **Critical role for CXCR3 chemokine biology in the pathogenesis of bronchiolitis obliterans syndrome.** *J Immunol* 2002, **169:**1037-1049.

29. Lasagni L, Francalanci M, Annunziato F, Lazzeri E, Giannini S, Cosmi L, Sagrinati C, Mazzinghi B, Orlando C, Maggi E, *et al.*: **An alternatively spliced variant of CXCR3 mediates the inhibition of endothelial cell growth induced by IP-10, Mig, and I-TAC, and acts as functional receptor for platelet factor 4.** *J Exp Med* 2003, **197:**1537-1549.

30. Morales J, Homey B, Vicari AP, Hudak S, Oldham E, Hedrick J, Orozco R, Copeland NG, Jenkins NA, McEvoy LM, *et al.*: **CTACK, a skin-associated chemokine that preferentially attracts skin-homing memory T cells.** *Proc Natl Acad Sci USA* 1999, **96:**14470-14475.

31. Yoshida T, Imai T, Takagi S, Nishimura M, Ishikawa I, Yaoi T, Yoshie O: **Structure and expression of two highly related genes encoding SCM-1/human lymphotactin.** *FEBS Lett* 1996, **395:**82-88.

32. Yoshida T, Imai T, Kakizaki M, Nishimura M, Takagi S, Yoshie O: **Identification of single C motif-1/lymphotactin receptor XCR1.** *J Biol Chem* 1998, **273:**16551-16554.

33. Vassileva G, Soto H, Zlotnik A, Nakano H, Kakiuchi T, Hedrick JA, Lira SA: **The reduced expression of 6Ckine in the PLT mouse results from the deletion of one of two 6Ckine genes.** *J Exp Med* 1999, **190:**1183-1188.

34. Guigo R, Dermitzakis E, Agarwal P, Ponting C, Parra G, Reymond A, Abril J, Keibler E, Lyle R, Ucla C, *et al.*: **Comparison of mouse and human genomes followed by experimental verification yields an estimated 1,019 additional genes.** *Proc Nat Acad Sci USA* 2003, **100:**1140-1145.

35. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, *et al.*: **Initial sequencing and comparative analysis of the mouse genome.** *Nature* 2002, **420:**520-562.

36. Nomiyama H, Egami K, Tanase S, Miura R, Hirakawa H, Kuhara S, Ogasawara J, Morishita S, Yoshie O, Kusuda J, *et al.*: **Comparative DNA sequence analysis of mouse and human CC chemokine gene clusters.** *J Interferon Cytokine* Res 2003, **23:**37-45.

37. Swofford DL, Waddell PJ, Huelsenbeck JP, Foster PG, Lewis PO, Rogers JS: **Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods.** *Syst Biol* 2001, **50:**525-539.

38. **NCBI Genomic Biology** [http://www.ncbi.nlm.nih.gov/Genomes/]

39. **Ensembl** [http://www.ensembl.org]

40. **Mouse Genome Informatics** [http://www.informatics.jax.org/]