**Introduction (Version 1 – September 2021 – 1073 words)**

Chemokine signalling plays a fundamental role in a myriad of biological processes in vertebrates. Best known for the chemoattraction of leukocytes during host defence (Wong et al 2003, Blanchet et al 2012); chemokine signalling is also implicated in homeostasis, development (Zlotnik et al 2000, Tran et al 2003, López-Cotarelo et al 2017), and neuronal communication (Tran et al 2003, Rostène et al 2007, Haas et al 2007). Failure of the system can lead to various diseases (Tran et al 2003, Blanchet et al 2012), including cancer (Nagarsheth et al 2017).

Despite this broad diversity of functions, the chemokine system appears to be a vertebrate novelty (DeVries et al. 2006, Bajoghli 2013). Consequently making the understanding of the origin and evolution of this versatile system even more intriguing.

The chemokine system is composed of chemokine receptors (CKRs) and their chemokine ligands (CKLs). The receptors are G-protein coupled receptors (GPCRs) belonging to the gamma rhodopsin-type (class A) subfamily: so exhibit the typical 7 transmembrane domain structure of GPCRs. The ligands are small cytokines that are conventionally secreted outside the cell where they chemoattract the cells bearing corresponding receptors. Canonical chemokines have been classified into four categories based on the pattern of cysteine residues at the N-terminal: CC-type; CXC-type; XC-type and CX3C-type (Zlotnik et al. 2000). Chemokine receptors derive their naming conventions from the type of ligands they are activated by (Nomiyama et al. 2011, Bachelerie et al. 2014). Naming conventions can be misleading because both ligands and receptors are promiscuous: a receptor can be activated by more than one ligand and vice versa (Zlotnik et al. 2006, Nagarsheth er al. 2017, Hughes et al. 2018). Furthermore, several other molecules may contribute to chemokine signalling (Zhang et al. 2018).

Alternative functional ligands for chemokine receptors include chemokine-like factor (CKLF) that binds CCR4 (Wang et al 2006, Wang et al 2008, Zhang et al 2011, Liu et al 2018) and cytokine-like 1 (CYTL1) that binds CCR2 (Wang et al 2016). Moreover, the members of the cytokine family FAM19A4 (or TAFA) have been described as “brain chemokines” (Tang et al. 2004, Park et al. 2017) and have structural similarity to canonical chemokines such as MIP-1α (now referred to as CCL3) (Tang et al 2004, Zhang et al 2017, Sarver et al 2021), although their receptors are not canonical CKRs.

For example, TAFA4 is a ligand for formyl peptide receptor 1 (FPR1) (Wang et al 2014) and TAFA 5 is a likely ligand for formyl peptide receptor 2 (FPR2) (Park et al. 2017). Formyl-peptide receptor-like 1 (FPRL1) can be bound by a variant produced by peptide cleaving of the canonical ligand CCL23 (Miao et al. 2007) reinforcing the interest for this group of receptors in chemokine signalling.

Moreover, a family of “atypical” chemokine receptors (ACKR) bind canonical chemokine ligands, however, they are incapable of binding G proteins and rather than inducing cell migration they act as “scavengers” of chemokines by limiting their availability (Bonecchi and Graham 2016, Ulvmar et al. 2014). Members of this family are ACKR1/DARC, ACKR2/D6, ACKR3/CXCR7 and ACKR4/CCRL1 (Bachelerie et al 2014, Bonecchi and Graham 2016).

Finally, additional receptors that have been implicated in chemokine-like functions, such as regulation of leukocyte migration during immune response, are the chemokine-like receptor 1 (CMKLR1) and chemokine (C-C motif) receptor-like 2 (CCRL2) (Yoshimura and Oppenheim 2011). Both can bind to the non-chemokine chemoattractant chemerin (Yoshimura and Oppenheim 2011, Monnier et al. 2012), and CCRL2 was thought to bind also canonical chemokines therefore being proposed as a potential fifth member of the ACKR family (Yoshimura and Oppenheim 2011, Schioppa et al. 2020, Lokeshwar et al. 2020). Chemerin has also been shown to bind to the receptor GPR1 (Kennedy and Davenport 2018, Fischer et al. 2021) that, interestingly, has as an additional ligand TAFA 1 (Zheng et al 2018). Although a sixth member of the ACKR family has been proposed, PITPNM3, this is not even belonging to the family of GPCRs (Bachelerie et al. 2014, Lokeshwar et al. 2020), unlike all the previously mentioned receptors that are all GPCRs of class A, just like the canonical CKRs.

To date all studies support the origin of chemokine signalling in the ancestor to vertebrates followed by a great expansion within the jawed vertebrates lineage (DeVries et al. 2006, Zlotnik et al 2006, Nomiyama et al. 2011, Bajoghli 2013, Nomiyama et al 2015). De Vries et al. 2006 failed to find a member of the entire subfamily to which chemokine receptors belong (including angiotensin, bradykinin fMLP, C5a, leukotriene, and ADMRs) outside vertebrates. As for the ligands, the same authors conclude that they too are found exclusively in vertebrates, albeit their study was based on a small subset of non-vertebrate models. Zlotnik et al. 2006 agrees with DeVries though they focused on the relationships within vertebrate chemokines rather than exploring a possible presence outside of vertebrates. Further studies on the evolutionary origin of chemokine ligands are likely lacking due to the difficulty of performing phylogenetic reconstructions on short sequences. Also, to our knowledge, no studies have been conducted on the evolution of non-canonical ligands such as the CKLF superfamily or the TAFA chemokines.

For a comprehensive investigation into the origin and evolution of the chemokine system all components and their relatives need to be considered. As does the diversity of organisms in which they may exist. Here our goal is to re-evaluate the vertebrate exclusivity of the canonical chemokine system and investigate the evolution of non-canonical players involved in chemokine signalling. This will allow us to reconstruct the ancestral state of chemokine signalling. We systematically sampled all major groups representative of animal diversity and all potential components of the chemokine system. Our results were analysed using state of the art phylogenetic methodologies.

With this comprehensive approach, we report several findings: continued support for the canonical chemokine system as a vertebrate innovation; better understanding of the relationships between different types of chemokine receptors and their relatives; reconstruction of the ancestral chemokine complement; identification of multiple independent origins of chemokine receptor functioning. Cytokines with chemokine-like function such as CKLF and other members of the CKLF Superfamily, TAFA-chemokine-like proteins, CYTL and likely also CXCL17 are unrelated to canonical chemokines. CKLF Superfamily is distinct into two large groups; while the group containing CKLF (CKLF Group I) is monophyletic and vertebrate-specific, the other (CKLF Group II) is of more complex composition and widespread outside of vertebrates. Finally, we provide evidence that the TAFA family may be chordate rather than vertebrate specific.

**Introduction (Version 2 – January 2022 – 423 words)**

Cell migration is fundamental for a variety of biological processes from immunity to development. One type of cell migration is based on chemoattraction: where cells move along a chemical gradient. In vertebrates the chemokine system uses this mechanism and acts in host defence (Wong et al 2003, Blanchet et al 2012), homeostasis (López-Cotarelo et al 2017), neuronal communication (Tran et al 2003, Rostène et al 2007, Haas et al 2007) and more (Zlotnik and Yoshie 2000). This system has two components, a ligand and a receptor. The ligand is secreted to generate an extracellular gradient that is recognized by target cells sporting associated receptors that bind the ligands: initiating cell movement (REF).

The human genome encodes for 18 (plus 4 atypical) chemokine receptors, all GPCRs of rhodopsin type, and 48 chemokines, small cytokines with recognizable cysteine patterns in the N-terminal (Hughes and Nibbs 2018, Bachelerie et al. 2020); this large repertoire allows the cell to fine tune responses to different external stimuli. The diversity and interaction of these canonical components has been well characterized. However, in addition to this classical repertoire, several other “chemokine-like” molecules (Table 1) have been shown to function similarly to chemokine ligands or receptors. This causes confusion as their precise relation to the canonical chemokine system is unknown. Such molecules include, for example: chemokine-like factor (CKLF) that binds chemokine receptor CCR4 (Wang et al 2006, Wang et al 2008, Zhang et al 2011, Liu et al 2018) and drives cell migration in vivo (ref); The TAFA ‘chemokines’ that (Tang et al. 2004, Park et al. 2017) share structural similarities to canonical chemokines (Tang et al 2004, Zhang et al 2017, Sarver et al 2021); and the atypical chemokine receptors which bind canonical ligands without inducing a response.

While the evolution of canonical chemokine components has been explored, little attention has been paid to the evolution of the non-canonical components. This leaves many questions open, for example, does the diversity of components have a common origin and when in the history of life did these molecules first evolve?

To address these questions in this work we investigate the molecular relationships between the different chemokine and chemokine-like components in metazoans. We used 64 proteomes covering 19 animal phyla (Table 2) and using phylogenomic methods we reconstructed the origin and pattern of duplication of the chemokine genes.

Our results indicate that unrelated proteins evolved chemokine ligand functions multiple times independently. This pattern contrasts the evolutionary history of the receptors which primarily descend from the same gene duplication in the ancestors of vertebrates.

**Introduction (V3 – October 2022 – 377 words)**

Cell migration is fundamental for a variety of biological processes, from immunity to development. One type of cell migration is based on chemoattraction: where cells move along a chemical gradient. Vertebrates use the chemokine system for chemoattraction in host defence (Wong et al 2003, Blanchet et al 2012), homeostasis (López-Cotarelo et al 2017) and neuronal communication (Tran et al 2003, Rostène et al 2007, Haas et al 2007, Zlotnik and Yoshie 2000). This system has two components, a ligand, and a receptor. The ligand is secreted to generate an extracellular gradient that is recognized by receptors on target cells, and this triggers cell movement (REF).

The human genome encodes for 18 (plus four atypical) chemokine receptors, all GPCRs of rhodopsin type, and 48 chemokines, small cytokines with recognizable cysteine patterns in the N-terminal (Hughes and Nibbs 2018, Bachelerie et al. 2020); this large repertoire allows the cell to fine-tune responses to different external stimuli (REF). The function and interaction of canonical components has been well characterized (REF). However, in addition to this classical repertoire, several other non-canonical “chemokine-like” molecules (Table 1) have been shown to function similarly to chemokine ligands or receptors. These molecules include, for example, chemokine-like factor (CKLF) that binds to chemokine receptor CCR4 (Wang et al 2006, Wang et al 2008, Zhang et al 2011, Liu et al 2018) and drives cell migration in vivo (ref); The TAFA ‘chemokines’ that (Tang et al. 2004, Park et al. 2017) share structural similarities to canonical chemokines (Tang et al 2004, Zhang et al 2017, Sarver et al 2021); and the atypical chemokine receptors which bind canonical ligands without inducing a response (REF).

Currently, the canonical chemokine system has been well described only in vertebrates (REF). However, the evolutionary history of the system and its relationship with the non-canonical components remains unclear, including whether the canonical and non-canonical components are homologous or analogous.

Therefore, we use phylogenetic methods and genomic data from 64 species covering 19 animal phyla (Table S1) to reconstruct the origin and pattern of duplication of the chemokine genes.

Our results indicate that unrelated proteins evolved “chemokine-like” ligand function multiple times independently. This pattern contrasts with the evolutionary history of most receptors which descend from the same gene duplication in the ancestors of vertebrates.

**INTRODUCTION (V7 – January 2023 – 635 words)**

The chemokine system modulates many biological processes, from host defence to homeostasis and neuronal communication (ADD REF TO FEW REVIEW). It has two components, a ligand, usually a small cytokine named chemokine, and a receptor, typically a GPCR class A. The ligands are organized into four categories, the XC type, CC type, CXC type and CX3C, which differ in the pattern of cysteine residues in the N-terminal portion of the protein (REF). In humans, there are 28 CC, 17 CXC, 2 XC and 1 CX3C chemokines. Likewise, there are four canonical receptor paralog groups, the XCR, CCR, CXCR, and CX3CR (REF). These receptors bind different chemokines, the CXCR bind CXC chemokines, the CCR bind CC chemokines, CX3CR1 binds the sole CX3C chemokine (CX3CL1), and the XCR1 binds the two XC chemokines (XCL1 and XCL2) (REF). In addition to these canonical components, several other non-canonical “chemokine-like” molecules (Table 1) have been shown to function similarly to chemokine ligands or receptors. These include ligands such as the chemokine-like factor (CKLF) that binds to chemokine receptor CCR4 (REF) and drives cell migration *in vivo* (REF); The TAFA chemokines (REF) that are mainly expressed in the nervous system, share structural similarities to canonical chemokines (REF) and in some cases bind to formyl peptide receptors (FPR), a group of GPCRs related to chemokine receptors (doi: [10.1038/s41598-017-15586-0](https://doi.org/10.1038%2Fs41598-017-15586-0)); The Cytokine-like (Cytl) that binds CCR2 (REF DOI: 10.4049/jimmunol.1501908) and has been suggested to be related to CC ligands based on the presence of IL8-like chemokine fold (https://doi.org/10.1002/prot.22963). Likewise, there are also non-canonical receptors known as atypical chemokine receptors (10.3389/fimmu.2016.00224). Like canonical chemokine receptors, they are GPCR class A, but differently from them, they cannot initiate classical chemokine signaling after ligand binding (<https://doi.org/10.1038/s41467-020-16664-0>). There are four atypical chemokine receptors encoded in the human genome: the ACKR1 (also known as DARC), ACKR2 (also known as D6), ACKR3 (also known as CXCR7) and ACKR4 (https://doi.org/10.1038/nri3544).

The chemokine system's molecular function is well studied (there are over 320,000 papers on the chemokine system on PubMed), however, several aspects of the evolution of ligands and receptors are unclear. For example, the homology between canonical and non-canonical ligands is unclear and supported by weak evidence, such as shared specific motifs (REF). Likewise, the evolution of the gene complement for canonical and non-canonical components remains poorly understood outside a few key model systems (REF). Likewise, the evolutionary history of receptors remains unclear, with uncertainty regarding the complements in early vertebrates (REF) and the relationships between atypical and canonical receptors (REF). In addition, two competitive hypotheses have been proposed regarding the origin of the receptors, the first suggesting that they evolved from angiotensin receptors (REF) and the second that they evolved from adrenomedullin receptors (REF). These uncertainties in both ligands' and receptors' evolutionary histories have common underlining reasons i) the use of inadequate inference methods (e.g., relying on sequence similarities) and ii) the use of limited species sampling (e.g., humans, mice, and zebrafish). Finally, resolving the phylogenetic relationships for short molecules such as chemokine ligands is particularly challenging due to the lack of strong phylogenetic signals (REF).

Here, to clarify these outstanding questions, differently from previous works, we use state-of-the-art phylogenetic methods (including methods designed for single-gene phylogenies), a large taxonomical sampling composed both of vertebrate and invertebrate genomes, and the entire complement of canonical and non-canonical-components both for receptors and ligands. Our results substantially clarify the phylogenetic affinity of canonical and non-canonical ligands and receptors and we suggest that unrelated proteins evolved “chemokine-like” ligand function multiple times independently. Furthermore, we identified that all the know receptors evolved from a single duplication in the vertebrate stem group that gave rise to many GPCRs, including all the atypical receptors; but Atypical 1. Finally, we identify several other ligands and receptors that might have chemokine-related functions and can be deployed for functional work.