**Main Figure Legends**

**Figure 1: Cluster Analysis and Phylogeny of Ligand groups. A)** Similarity-based clustering, using CLANS, of canonical chemokines and related molecules with sequence similarity. Canonical chemokines are an independent group from other related molecules (TAFA, CYTL and CXCL17). Canonical chemokines are composed of two large groups (CC-type and CXC-type) within which some divergent subgroups are highlighted. The clustering and connections shown are at the p-value threshold of 1E-6. Other p-values tested are shown in Supplementary Figure S1. Candidate invertebrate sequences are shown as crosses and further information regarding them can be found in Supplementary Results. **B)** Maximum-Likelihood un-rooted phylogenetic tree of canonical chemokines. CC-type and CXC-type are split into two separate clades. Supports for key nodes are indicated in boxes with Transferable Bootstrap Expectation (TBE) represented by triangles and the Ultrafast Bootstraps (UFB) as circles. A traffic light colour code is used to indicate the level of support: high (green); intermediate (yellow) and low (red). **C)** Similarity-based clustering, using CLANS, of the CKLF super family (CKLFSF). Two major clusters are formed: the smaller “CKLF Group I” and the heterogenous “CKLF group II” that also includes some invertebrate sequences (shown as crosses). Subclades, including the known members of the CKLF super family, are highlighted. The clustering and connections shown are at the p-value threshold of 1E-15, as this is the threshold at which the two major clusters connect. Other p-values tested are shown in Supplementary Figure S2. **D)** Maximum-Likelihood un-rooted phylogenetic tree of the CKLF super family (CKLFSF). The CKLF group I is monophyletic, while the CKLF group II is not. Supports for key nodes are indicated in boxes with Transferable Bootstrap Expectation (TBE) represented by triangles and the Ultrafast Bootstraps (UFB) as circles. A traffic light colour code is used to indicate the level of support: high (green); intermediate (yellow) and low (red).

**Figure 2: Distribution and duplication patterns of ligand groups. A)** Presence of all ligand groups are mapped onto a species tree. Gene trees and duplication events are based on the gene tree to species tree reconciliation analyses. The nomenclature for canonical chemokines is primarily based on known chemokines of human (or mouse). Where human and mouse chemokines do not correspond, the default name refers to the human gene and the mouse (*Mus musculus*) one is indicated with “Mm”. Chemokines that have been classically described as having either homeostatic or inflammatory function are indicated with a circle or a star respectively. The classification used here was based on Zlotnik and Yoshie 2012 (1) with the inflammatory type also including chemokines they described as plasma/platelet types. Overall, canonical chemokines originated in vertebrates and expanded a first time in jawed vertebrates and a second time in mammals. Homeostatic chemokines (e.g., CXCL12) are generally more ancient than inflammatory ones. CXCL17 and CYTL are mammal and jawed vertebrate specific respectively. TAFA originated in the common ancestor of vertebrates and urochordates, while the CKLF super family is present in invertebrates although key duplications occurred at the base of vertebrates. **B)** Number of complements for each ligand group at key species nodes are mapped onto the species tree. The number of complements in each group reflects the pattern of duplications. The major increase occurred at the level of jawed vertebrates with canonical chemokines undergoing a second significant increase within placentals.

**Figure 3: Phylogeny of Receptor groups.**

An unrooted maximum likelihood phylogeny of chemokine receptors. The tree shown is the transfer bootstrap expectation (TBE) tree including just the chordate specific clade from the ultrafast bootstrap tree (UFB). Node supports from both TBE (triangle) and UFB (circle) shown for equivalent key nodes in boxes with arrows to indicate node. A traffic light colour code is used to indicate the level of support: high (green); intermediate (yellow) and low (red). Key clades highlighted: yellow = chemokine like plus group (CMLplus); blue = intermediate group; green = atypical 3 and GPR182 (ACKR3/GPR182); purple = canonical chemokines (Canonical CKR); and pink = relaxin receptors (RL3R). Branches scaled by amino acid substitutions per site.

**Figure 4: Distribution and duplication patterns of receptor groups.**

**A)** Presence of all receptor groups are mapped onto a species tree. Gene trees and duplication events are based on the gene tree to species tree reconciliation analyses. The nomenclature for genes is primarily based on human chemokines. The canonical chemokines had 5 paralogs present in the vertebrate common ancestor. These undergo a heterogeneous pattern of duplication throughout vertebrates with different paralogs duplicating different number of times and in different groups of species. Chemokines that have been classically described as having either homeostatic or inflammatory function are indicated with a circle or a star respectively. The classification used here was based on Zlotnik and Yoshie 2012 (1). **B)** Number of complements for each receptor group at key species nodes are mapped onto the species tree. The number of complements in each group reflects the pattern of duplications. The chemokine groups diverged in the vertebrate stem group. The major expansion occurred at the level of jawed vertebrates with canonical chemokine receptors, the chemokine-like receptor plus group and intermediate groups increasing in copy number. Canonical chemokine underwent another small subsequent increase within placentals.

**Figure 5: Summary of the evolution of ligand and receptors.**

A summary diagram of the evolution of the different chemokine system components. A simplified phylogenetic tree of species is shown, calibrated to time according to Erwin et al. 2011 (2), Delsuc et al. 2018 (3)and Cohen et al. 2013 (4). Circles represent ligand groups and 7 transmembrane domain structure icon represents GPCR groups. Icons are colour coded by group and placed adjacent to the branch in the species tree where they first appear. X2 and X5 shown to indicate the number of paralogs present for CXCL ligand group and the canonical CKR groups respectively, on the branch where they first appear. Question mark refers to the uncertainty regarding the origin of the CKLF group I in jawed vertebrates or deuterostome stem group (see Figure 2). Geological column is shown along the bottom, in accordance with the ICS International Chronostratigraphic Chart (updated 2017).

**References for Main Figure Legends**

1. A. Zlotnik, O. Yoshie, The Chemokine Superfamily Revisited. *Immunity* **36**, 705–716 (2012).

2. D. H. Erwin, *et al.*, The Cambrian Conundrum: Early Divergence and Later Ecological Success in the Early History of Animals. *Science* **334**, 1091–1097 (2011).

3. F. Delsuc, *et al.*, A phylogenomic framework and timescale for comparative studies of tunicates. *BMC Biol* **16**, 39 (2018).

4. K. M. Cohen, S. C. Finney, P. L. Gibbard, J.-X. Fan, The ICS International Chronostratigraphic Chart. *Episodes Journal of International Geoscience* **36**, 199–204 (2013).