Supplementary Figure Captions

**Figure S1. CLANS clustering of chemokines and related molecules sequences.** Initial identification and annotation of clusters was performed at the strict p-value of 1E-35 **(A)**. Subsequent loosening of the p-value clarified the relationships across clusters and defined bigger groups. At p-value 1E-15 **(B)** two major canonical chemokine groups are well-defined: the CCL group, which includes also XCL and X3CL; and the CXCL group. At this level of stringency, only few canonical chemokines remain isolated: CCL27/28; CXCL12; CXCL14; CXCL16; as well as CXCL17. TAFA and CYTL are also isolated. At p-value 1E-10 **(C)** the two major chemokine groups connect to each other. CCL27/28 is connected to the CCL group and CXCL12 and CXCL14 are connected to the CXCL group, while CXCL16 and CXCL17 are still isolated. At p-value 1E-6 **(D)** all chemokine groups are connected in one big cluster, except for CXCL17. TAFA and CYTL are also still isolated. Crosses indicate the few invertebrate sequences that were collected from the BLAST search, more information in Supplementary Results.

**Figure S2. CLANS clustering of CKLF Super Family sequences.** Initial identification and annotation of clusters was performed at the strict p-value of 1E-60 **(A)**. Subsequent loosening of the p-value clarified the relationships across clusters and defined bigger groups. At 1E-20 **(B)** two major clusters have formed. One, that we called CKLF group I, includes CKLF, CMTM1,2,3,5, and PLP2. The other, that we called CKLF group II, includes CMTM4/6,7,8, and other groups. At 1E-16 **(C)** more sequences have joined the two major groups that are still separate. At 1E-15 **(D)** the two major groups connect, as well as few extra sequences, see Supplementary Results for extra details. Crosses indicate invertebrate sequences.

**Figure S3. Unrooted phylogenetic tree of canonical chemokines with TBE supports.** Phylogenetic tree under the model GTR20+F+R4. Nodal support is calculated from 100 non-parametric bootstrap repeats with transfer bootstrap expectation. CCL clade is in orange, CXCL clade in blue.

**Figure S4. Unrooted phylogenetic tree of canonical chemokines with UFB supports.** Phylogenetic tree under the model GTR20+F+R4. Nodal support is calculated from 1000 ultrafast bootstrap repeats. CCL clade is in orange, CXCL clade in blue.

**Figure S5. Unrooted phylogenetic tree of TAFA with TBE supports.** Phylogenetic tree under the model JTT+R5. Nodal support is calculated from 100 non-parametric bootstrap repeats with transfer bootstrap expectation.

**Figure S6. Unrooted phylogenetic tree of TAFA with UFB supports.** Phylogenetic tree under the model JTT+R5. Nodal support is calculated from 1000 ultrafast bootstrap repeats.

**Figure S7. Unrooted phylogenetic tree of CYTL with TBE supports.** Phylogenetic tree under the model JTT+I+G4. Nodal support is calculated from 100 non-parametric bootstrap repeats with transfer bootstrap expectation.

**Figure S8. Unrooted phylogenetic tree of CYTL with UFB supports.** Phylogenetic tree under the model JTT+I+G4. Nodal support is calculated from 1000 ultrafast bootstrap repeats.

**Figure S9. Unrooted phylogenetic tree of CXCL17 with TBE supports.** Phylogenetic tree under the model JTT. Nodal support is calculated from 100 non-parametric bootstrap repeats with transfer bootstrap expectation.

**Figure S10. Unrooted phylogenetic tree of CXCL17 with UFB supports.** Phylogenetic tree under the model JTT. Nodal support is calculated from 1000 ultrafast bootstrap repeats.

**Figure S11. Unrooted phylogenetic tree of CKLFSF with TBE supports.** Phylogenetic tree under the model GTR20+F+R7. Nodal support is calculated from 100 non-parametric bootstrap repeats with transfer bootstrap expectation. Red clade = CMTM4/6; blue clade = CKLF I group; green clade = CMTM7; turquois clade = MAL/MALL/MAL2.

**Figure S12. Unrooted phylogenetic tree of CKLFSF with UFB supports.** Phylogenetic tree under the model GTR20+F+R7. Nodal support is calculated from 1000 ultrafast bootstrap repeats. Red clade = CMTM4/6; blue clade = CKLF I group; green clade = CMTM7; turquois clade = MAL/MALL/MAL2.

**Figure S13. Alignment of candidate brachiopod CCL24 sequence with mammalian CCL24s.** Our BLAST searches picked up a sequence from the brachiopod *Lingula unguis* that when re-blasted versus SwissProt returned a CCL24 as hit. Alignment of the brachiopod sequence with mammalian CCL24 sequences reveals a poor overall conservation, with the brachiopod sequence also being significantly longer than any of the other sequences. Further details about this sequence can be found in supplementary file S3 and in Supplementary Results.

**Figure S14. Alignment of candidate cnidarian CCL3 sequence with mammalian CCL3s.** Our BLAST searches picked up a sequence from the cnidarian *Clytia hemisphaerica* that when re-blasted versus SwissProt returned a CCL3 as hit. Alignment of the cnidarian sequence with mammalian CCL3 sequences reveals a poor overall conservation, with the cnidarian sequence being extremely longer than any of the other sequences. Further details about this sequence can be found in supplementary file S3 and in Supplementary Results.

**Figure S15. Alignment of candidate echinoderm CXCL10 sequence with mammalian CXCL10s**. Our BLAST searches picked up a sequence from the echinoderm *Acanthaster planci* that when re-blasted versus SwissProt returned a CXCL10 as hit. Alignment of the echinoderm sequence with mammalian CXCL10 sequences reveals a poor overall conservation, with the brachiopod sequence also being significantly longer than any of the other sequences. Further details about this sequence can be found in supplementary file S3 and in Supplementary Results.

**Figure S16. Rooted species tree reconciled gene tree for canonical chemokines.** The canonical chemokines gene tree was reconciled with the species tree using GeneRax. “S” or “D” at the node indicates a speciation or duplication event respectively. CCL clade is in orange, CXCL clade in blue.

**Figure S17. Alignment of 4 candidate urochordate TAFA sequences with vertebrate TAFAs.** Our BLAST searches picked up 4 sequences from the urochordate *Ciona intestinalis*, that connected with the TAFA cluster in the CLANS analysis. One of these sequences when blasted versus SwissProt returned a TAFA as hit. This sequence was also annotated as TAFA by InterProScan. Alignment of the urochordate sequences with vertebrate TAFA sequences reveals that only the one annotated as TAFA aligns well. While the other 3 align poorly and are also significantly longer than any of the other sequences. Further details about these sequences can be found in supplementary file S3 and in Supplementary Results.

**Figure S18. Alignment of best candidate urochordate TAFA sequence with vertebrate TAFAs.** Of the 4 urochordate candidate TAFA sequences, only one was annotated as TAFA with both SwissProt and InterProScan annotation and appeared to align well with other TAFAs with a preliminary alignment with all urochordate sequences (Figure S6). Here we removed the other 3 urochordate sequences and aligned only the best candidate with the vertebrate TAFAs. The sequence conservation is even more apparent with this alignment. Importantly, 8 of the 10 typical cysteine residues of TAFA1-4 are conserved, and the two missing cysteines are the same ones missing in TAFA5. Further discussion can be found in Supplementary Results.

**Figure S19. Rooted species tree reconciled gene tree for TAFA.** The TAFA gene tree was reconciled with the species tree using GeneRax. “S” or “D” at the node indicates a speciation or duplication event respectively.

**Figure S20. Rooted species tree reconciled gene tree for CYTL.** The CYTL gene tree was reconciled with the species tree using GeneRax. “S” or “D” at the node indicates a speciation or duplication event respectively.

**Figure S21. Rooted species tree reconciled gene tree for CXCL17.** The CXCL17 gene tree was reconciled with the species tree using GeneRax. “S” or “D” at the node indicates a speciation or duplication event respectively.

**Figure S22. Rooted species tree reconciled gene tree for CKLFSF.** The CKLFSF gene tree was reconciled with the species tree using GeneRax. “S” or “D” at the node indicates a speciation or duplication event respectively. Red clade = CMTM4/6; blue clade = CKLF I group; green clade = CMTM7; turquois clade = MAL/MALL/MAL2.

**Figure S23. CLANS clustering of receptors and related molecules sequences.** A CLANS clustering layout where shapes indicate sequences and lines are connections indicating similarity between sequences at or surpassing the p-value similarity threshold. Sequences are positioned in clusters based on similarity. Initial identification and annotation of clusters was performed using the inbuilt convex clustering at the p-value of 1E-100. **(A)** clustering was loosened till the canonical receptor annotated groups formed a cluster at 1E-65. **(B)** loosening of the p-value to 1E-60 identified relationships between clusters of interest and identified the intermediate group as connecting to both canonical and chemokine-like plus groups. All sequences connected to groups of interest are vertebrate sequences. **(C)** further loosening to p-value 1E-50 connects the vertebrate sequences of interest to a large cluster of sequences which contains vertebrate and invertebrate sequences which are annotated as opioid and somatostatin receptors and other GPCRs. Crosses indicate invertebrate sequences and Y-shape indicates the reference viral sequences included. Shapes are colour coded by group of interest: purple = canonical chemokine receptors; yellow = chemokine-like plus; green = atypical receptor 3/GPR182; blue = intermediate group; pink = relaxin receptors.

**Figure S24. Unrooted phylogenetic tree of receptors with TBE supports.** Phylogenetic tree of receptor sequences of interest and putative outgroups under the model GTR20+F+G4. Sequences used are a subset of sequences extracted from CLANs, specifically they are those in the chordate specific clade in the ultrafast bootstrap tree. Nodal support is calculated from 100 non-parametric bootstrap repeats with transfer bootstrap expectation. Branches colour coded by group of interest: purple = canonical chemokine receptors; yellow = chemokine-like plus; green = atypical receptor 3/GPR182; blue = intermediate group; pink = relaxin receptors.

**Figure S25. Unrooted phylogenetic tree of receptors with UFB supports.** Phylogenetic tree of all receptor sequences of interest and putative outgroups extracted from clans under the model GTR20+F+G4. Nodal support is calculated from 1000 ultrafast bootstrap repeats. Branches colour coded by group of interest: purple = canonical chemokine receptors; yellow = chemokine-like plus; green = atypical receptor 3/GPR182; blue = intermediate group; pink = relaxin receptors.

**Figure S26. Rooted species tree reconciled gene tree for receptors.** The ultrafast bootstrap receptor tree was modified to extract the subtree of the chordate specific clade. This gene tree was reconciled with the species tree using GeneRax. “S” or “D” at the node indicates a speciation or duplication event respectively. Branches colour coded by group of interest: purple = canonical chemokine receptors; yellow = chemokine-like plus; green = atypical receptor 3/GPR182; blue = intermediate group; pink = relaxin receptors.