ELSEVIER

Costimulatory receptors in jawed vertebrates: Conserved CD28, odd CTLA4 and multiple BTLAs

David Bernard[a], John D. Hansen[b,c], Louis Du Pasquier[d], Marie-Paule Lefranc[e], Abdenour Benmansour[a], Pierre Boudinot[a,*]

[a]*Institut National de la Recherche Agronomique, Unité de Virologie et Immunologie Moléculaires, 78352 Jouy-en-Josas cedex, France*
[b]*Western Fisheries Research Center, USGS Biological Resources Division, Seattle, WA 98115, USA*
[c]*Department of Pathobiology, University of Washington, Seattle, WA 98195, USA*
[d]*University of Basel, Institute of Zoology and Evolutionary Biology, Vesalgasse1, CH-4051 Basel, Switzerland*
[e]*IMGT®, Laboratoire d'ImmunoGénétique Moléculaire LIGM, IGH, UPR CNRS 1142 and Université Montpellier II, Montpellier, France*

Abstract

CD28 family of costimulatory receptors is comprised of molecules with a single V-type extracellular Ig domain, a transmembrane and an intracytoplasmic region with signaling motifs. CD28 and cytotoxic T lymphocyte antigen-4 (CTLA4) homologs have been recently identified in rainbow trout. Other sequences similar to mammalian CD28 family members have now been identified using teleost, *Xenopus* and chicken databases. CD28- and CTLA4 homologs were found in all vertebrate classes whereas inducible costimulatory signal (ICOS) was restricted to tetrapods, and programmed cell death-1 (PD1) was limited to mammals and chicken. Multiple B and T Lymphocyte Attenuator (BTLA) sequences were found in teleosts, but not in *Xenopus* or in avian genomes. The intron/exon structure of *btla*s was different from that of *cd28* and other members of the family. The Ig domain encoded in all the *btla* genes has features of the C-type structure, which suggests that BTLA does not belong to the CD28 family. The genomic localization of these genes in vertebrate genomes supports the split between the BTLA and CD28 families.
© 2006 Elsevier Ltd. All rights reserved.

Keywords: CD28; CTLA4; BTLA; Evolution of costimulatory receptors; Teleosts

1. Introduction

The activation of T lymphocytes is regulated by positive and negative signals delivered through different membrane receptors. Costimulatory and inhibitory receptors of the CD28 family are the best characterized, and their interaction with ligands from the B7 family initiates different pathways, that are crucial for T cell activation and tolerance. The CD28/CTLA4-B7-1/B7-2 pathway was the first to be described within this family for its importance

for immunity (reviewed in [1,2]). The costimulatory receptor inducible costimulatory signal (ICOS) [3] (also known as H4 and AILIM) was later identified as a new member of the CD28 family which recognizes a ligand named ICOSL [4] (also known as B7h, GL50, B7RP-1, LICOS and B7-H2) that is expressed by antigen presenting cells. Another member of the family, programmed cell death-1 (PD1) [5], interacts with two other homologs of B7, PD-L1 [6] (also known as B7-H1) and PD-L2 [7] (also known as B7-DC). The last receptor to be considered as a member of the CD28 family was identified as "B and T lymphocyte attenuator" (BTLA) [8], and was initially described as a ligand for an additional B7-related molecule, B7x. However, BTLA is distinct from the CD28 family for structural reasons [9]. CD28, cytotoxic T lymphocyte antigen-4 (CTLA4) and ICOS are closely linked in the genome and belong to a common immunological cluster [10]. PD1 is also encoded on the same chromosome in man and in mice, but BTLA is localized on another chromosome.

CD28 engagement is a central event for T cell activation. Ligation of CD28 with B7 ligands triggers signaling events that synergize with T cell receptor (TCR) initiated activation pathways, thereby promoting T cell survival and interleukin-2 (IL2) production [11]. In contrast, CTLA4 delivers a negative signal, which stops T cell responses. CTLA4 interferes with TCR- and CD28-dependent signaling by inhibiting the production of IL2 and the subsequent proliferation of T cells [12]. CTLA4 has a crucial importance in maintaining peripheral T cell tolerance through interactions with B7-1 (CD80) and B7-2 (CD86) as demonstrated by the fatal lymphoproliferative disease and autoimmune disorders observed in *ctla4*−/− mice [13]. ICOS was identified from activated human T cells [3]. Its expression is induced on T helper Th1 and Th2 after TCR engagement. During the differentiation of T helper cells, the expression of ICOS is initially upregulated then decreases on Th1 cells, but is sustained on Th2 subpopulations [14]. ICOS delivers positive stimulatory signals that are important for T helper differentiation. It modulates the production of several cytokines including IL10. Like ICOS, PD1 expression is induced within activated cells but is not T cell specific. It has been detected on T and B lymphocytes as well as on myeloid cells [15]. PD1 delivers negative signals that regulate B cell activation, T cell differentiation and the proliferation of myeloid cells. PD1 also plays a

role in regulating CD8+ T cell responses and peripheral tolerance as *pd-1* −/− mice develop autoimmune diseases [16,17]. BTLA was recently discovered and was considered to be a member of the CD28 family. This inhibitory receptor is expressed on both developing and mature B and T lymphocytes, macrophages and bone marrow-derived dendritic cells. The coligation of BTLA to TCR inhibits T cell activation, and *btla*−/− mice have increased incidence of autoimmune disorders [8]. The ligand of BTLA was first described as a B7 homolog, but this view has been recently challenged and the herpesvirus entry mediator (HVEM), a member of TNF/TNFR superfamily, is now considered as the true ligand of BTLA [18,19].

Costimulatory and inhibitory receptors of the CD28 family have been primarily identified and characterized in mammals. A homolog of CD28 expressed by chicken T cells [20] has also been identified and possesses similar functional capacities as observed for mammals [21,22]. However, the other members of the family have yet to be characterized in avian species. We have recently described homologs of CD28 (*Oncmyk*-CD28) and CTLA4 (*Oncmyk*-CTLA4) in rainbow trout (*Oncorhynchus mykiss*) [23]. *Oncmyk*-CD28 and *Oncmyk*-CTLA4 display the typical structure of CD28 family costimulatory receptors with an extracellular immunoglobulin superfamily (IgSF) domain (V-LIKE-domain), a connecting region, a transmembrane region and an intracytoplasmic region. A conserved ligand-binding site in the FG loop (CDR3 equivalent) of the V-LIKE-DOMAIN of both molecules suggests that these receptors recognize B7 homologs as found for their mammalian counterparts. The intracytoplasmic region of *Oncmyk*-CD28 possesses a motif that is conserved in mammalian costimulatory receptors and initiates signaling through phosphoinositid 3 kinase (PI3K) binding. In contrast, the intracytoplasmic region of *Oncmyk*-CTLA4 lacked obvious signaling motifs. Accordingly, a chimeric receptor composed of the extracellular domain of human CD28 fused to the intracytoplasmic region of *Oncmyk*-CD28 promoted TCR-induced IL2 production in a human T cell line while the chimeric receptor containing the cytoplasmic tail of *Oncmyk*-CTLA4 failed to signal. Moreover, *Oncmyk*-CD28 and *Oncmyk*-CTLA4 genes are localized on different chromosomes and do not belong to a costimulatory cluster as reported in mammals. Thus, fish CD28 and CTLA4 homologs share many structural features with their

mammalian counterparts but likely participate in different signaling for lymphocyte activation and regulation.

To gain more insight into the diversity of CD28 family members among vertebrates, we performed a systematic and comprehensive survey of CD28, CTLA4, ICOS, PD1 and BTLA homologs using available sequence databases. We identified genes similar to costimulatory receptors in most lineages including birds, amphibians and teleosts. This work therefore provides a broad picture of the evolution of CD28 family costimulatory receptors in vertebrates.

2. Material and methods

2.1. Strategy for identification of costimulatory sequences

Systematic searches were performed with available EST indices and genome databases using the TBLASTN program by using human, mouse and salmonid sequences of costimulatory receptors as the queries. Searches in EST databases were mainly performed at http://www.ncbi.nlm.nih.gov/ and http://www.tigr.gov/. For each gene, ESTs representing partial sequences were subjected to assembly using the tools of the Genetic Computer Group (Madison university, Wisconsin). Blast queries on complete genomes were sent to http://www.ensembl.org, http://www.ncbi.nlm.nih.gov/ and http://dolphin.lab.nig.ac.jp/medaka/ for medaka (Oryzias latipes). When relevant genomic regions were identified, potential exons were identified by comparison with known sequences, or from predictions made by Genscan (http://genes.mit.edu/GENSCAN.html). Relevant consensus nucleotide sequences were translated and ORF considered as candidate costimulatory sequences were used to search the non-redundant protein database with BLASTP. Sequences were kept for further analysis only when they matched CD28 family members with a significant e value ($< 1e^{-3}$). Retrieved sequences were then subjected to multiple sequence alignments using GCG (pileup) and ClustalW. Multiple alignments were edited using Boxshade software and further improved using IMGT® tools and expertise (IMGT®, the international ImMunoGeneTics information system, http://imgt.cines.fr [24]. Signal peptide and transmembrane hydrophobic regions were identified using Tmpred (http://www.chnet.org) and N-linked glycosylation sites were predicted by NetNGlyc 1.0 Server (http://www.cbs.dtu.dk/services/NetNGlyc). Secondary structural predictions and IMGT Collier de Perles representations were based on the IMGT unique numbering system for V-DOMAIN and V-LIKE-DOMAIN [25], respectively and using IMGT's unique numbering for C-DOMAIN and C-LIKE-DOMAIN [26] using IMGT® tools (http://imgt.cines.fr/) [27].

3. Results

3.1. CD28 homologs in teleosts, chicken and Xenopus

We have recently described the homologs of CD28 in rainbow trout and Atlantic salmon (Salmo salar). These sequences were used to search for CD28 sequences in other teleost species. Relevant hits were assembled and compared to the non-redundant protein database using BLASTP. When sequences were retrieved with significant e values, they were subjected to further analysis and multiple sequence alignments (Fig. 1). Relevant ESTs were found for orange-spotted grouper (Epinephelus coioides) and channel catfish (Ictalurus punctatus). Grouper CD28 sequence was based on a single EST, but the deduced amino acid sequence possessed all of the conserved residues and motifs found in the other fish CD28 sequences. The catfish sequence shared the same typical motifs but lacked a hydrophobic transmembrane region probably due to abnormal splicing, and was not included in the alignment. Sequences similar to rainbow trout CD28 were also determined from genome sequences of medaka (Oryzias latipes), fugu (Takifugu rubripes or Fugu rubripes), green spotted pufferfish (Tetraodon nigroviridis) and zebrafish (Danio rerio). In zebrafish, it was not possible to find sequences corresponding to exons 3 and 4 (encoding transmembrane and intracytoplasmic region) from the available genome assembly. The Ensembl zebrafish genome assembly (v36) had two copies of CD28 exons 1 and 2 in reverse orientation, located at close proximity on chromosome 1. Since these two regions were almost 100% identical, this duplication is most probably an assembly artifact. Mouse, human and fish sequences were then used to identify CD28 homologs in the chicken (Gallus gallus) and Xenopus (X. tropicalis) genomes, following the same strategy. The avian and amphibian sequences were also included in the multiple sequence alignment.

All V-LIKE-DOMAIN sequences displayed typical features of the CD28 V-LIKE-DOMAIN including the absence of the typically conserved

tryptophan residue (W 41) in strand C for the Ig fold and the relative conservation of the potential B7-binding site M-Y-P-P-P-I in the FG loop

```
              <--------- L-REGION-------> <--------------V-LIKE-DOMAIN-----------
                                          <----------FR1-IMGT-------> < CDR1-IMGT>
                                          ------A-------> -----B----->        BC
                                          1       10   15 16 20  23 26   30        38
                                          |.......|....| |...|....|   ...|......|
Oncmyk    MNVYWIPTILLS..LSSAANMISSNN       CKDKLRTFYVVRVSV NGIASVRCPNL TGKD...QEEMR
Salsal    MNVYWIPTILLSLCLSSAANMISSHN       YKDKLRTFNVVHVSV NGTASVRCPNL TGKD...QEEMR
Danrer          MSSMTILTLELNKSNRDLL        TESAGELPLIEYVPL NSTVKVACPKL SAP......EMD
Fugrub    MLEANEAIWLYSFRDKKSIGGYE          NANRKNRLNTVCVPV GGNVNVPCPNS SAE......DIK
Orylat          MAGFCWTLLVLMLSRTS          HTSHASQGLDTKDVT LKRVFVLCPNF TTQ......HVL
Epicoi    MRACLVFVILLGCRFSHAIES            QSVCKDQLKIACVAP GHNVSVPCPKL TTQ......DVT
Xentro                                    /DSSETQQQVVLVA VKQIEITFPSN FKDN....KESR
Galgal          MLGILVVLCLIPAADVT          ENKILVAQRPLLIVA NRTATLVCNYT YNGT...GKEFR
Homsap          M.LRLL.LALNLFPSIQVT(G)NKILVKQSPMLVAY DNAVNLSCKYS YNLF...SREFR
Musmus          MTLRLLFLALN.FFSVQVT(E)NKILVKQSPLLVVD SNEVSLSCRYS YNLL...AKEFR

              ----------------------- V-LIKE-DOMAIN------------------------
              <----FR2-IMGT----><CDR2-IMGT><--------------FR3-IMGT-----------
              ---C---> ---C'---> C'C"  ---C"---> ----D----> -----E----->
              39 41   46 47 50  55      60    66  70  74 75  80  84 85 90        96
              |.|....| |..|....| ....|.....| |...|...| |....|...| |...|......|
Oncmyk    FHLYLGL. .VEVGNHTH DSAHNHNSTE TVSPVGE.. GLGLRVNEQD HTVSFVLSGMTT
Salsal    FHLYLGL. .VEVGNHTH DNAHNHNSTE TVSPVGE.. GLGLRVNKQD HTVSFVLSGMTT
Danrer    FRLFKDG. .DEISIYVK INHT...SIP KTNTNKE.. FPAEFSVDLD NSSSFILTVVTV
Fugrub    FSLFKDG. .ECVHHHTQ ISG.....AT TPNQNTS.. NPVGVEYPSA NSPNFTITGVNA
Tetnig                          /GPPRTNT.. VGVEFKGPEN GSYSFVLPGANA
Orylat    FKLTINQ. .ELMATCEI KNS.....SQ IFESQPS.. NTDVEFKKTD ALVGFELRGAMA
Epicoi    FYLLKEE. .EVIYKLAC MRNN..ATSD CEQSNTGV. GVELQENTEN KSASFTLTGVNA
Xentro    VSLLRGV. .KKSITVCS GSFNSSYQSF ESNAYQS.. .IRCIGIPTE NKITFHLSGLTE
Galgal    ASLHKGT. .DSAVEVCF ISWN..MTKI NSNSNKE.. .FNCRGIHDK DKVIFNLWNMSA
Homsap    ASLHKGL. .DSAVEVCV VYGNY.SQQL QVYSKTG.. .FNCDGKLGN ESVTFYLQNLYV
Musmus    ASLYKGV. .NSDVEVCV GNGNF.TYQP QFRSNAE.. .FNCDGDFDN ETVTFRLWNLHV

              ---------V-LIKE-DOMAIN-----------> <------CONNECTING-REGION------>
              FR3-IMGT>< CDR3-IMGT > < FR4-IMGT>
              ---F---> FG ----G---->
              97   104 105  110 115 118 121  128
              |..|...| .....|....|.. |........|
Oncmyk    ERAGVYTC EGYPMYPPPIEKV PDEPQTLVLVE .AFQCQAG...GCVGPRVHGVP.........
Salsal    ERAGVYTC EGHPMYPPPIEKV PDEPQTLVLVE .ALQCQARKTVGCEGSGDDGVP.........
Danrer    NMTALYTC EASISYPPPFKYV DHKPQTIVFVE/
Fugrub    SSHGIYTC EVTEMFPPPLMTL YSDLRILVLIE .GHQCSAKDKNPTIRENQTCGC.........
Tetnig    CSHGIYTC EGTVMFPPPLKPL PSTPAVLLLVE .GHQCSNGTDKSPEGDGTSGS.........
Orylat    RADVAYRC DADKKFPPPFESF PNHSWFLVFHE ..ETCTNPEACQEAGGGGGQL.........
Epicoi    SSYGIYRC EVHIMFPPPLRKV PSTLSILVRVK .GHQCTVNETYIIKPAIEEAQKD........
Xentro    NHTDIYFF RKEDMYPPPYTCY QDNGTVIHVKE LVAPVKDPLPQPQW.................
Galgal    SQTDIYFC KIEAMYPPPYVYN EKSNGTVIHVR ETPIQTQEPESATS.................
Homsap    NQTDIYFC KIEVMYPPPYLDN EKSNGTIIHVK GKHLCPSPLFPGPSKP...............
Musmus    NHTDIYFC KIEFMYPPPYLDN ERSNGTIIHIK EKHLCHTQSSPEL..................

              <-----------TM----------> <-----------------IC--------------------->
Oncmyk    VWAWMLGFWVTTIYGLVVTVFAFVI WLRLRRVECSQSDYMDIKPKAPLRGHRKKQGVQHPIRMGRY*
Salsal    AWVWVLGFWVTIIYGLAVTVIAFAI RLRLRRVECSQSDYMSAKPRAPLRWPRKKQGVQHPIRMGRY*
Fugrub    GWMWIFAL.VIGSYSVTVTIGAIII WVKWRRED.SQSDYINTISKTTRYNKRK...LHIPTMHYFYAPN*
Tetnig    DWLWILTL.MVSTYGVAVTIGAVFI WVKWRKDD.SQSAYVNTNPKTTRSKGRL....YNPKPHFFYAPD*
Orylat    LWVWVTGVTLLATYGLAVTVAAFVL WXKLKDAD.SQNDYMNTKPKA....ARKRRGLQNPTPKYV*
Epicoi    LWIWILGIVVLGVYGIIVTIVAIII WAKWRQSD.SQCDYMNTKPKA.TRDRKKKRGVQIPIPRHF*
Xentro    GFPLLVAIGVMAAYSLTITAAFFYI LKKGRRTRVLQSEYINVVPRRPK....CHQPYAPAPVHCRPR*
Galgal    YWVMVAVTGLLGFYSMLITAVFIIY RQKSKRNNYRYQSDYMNMTPRHPPHQKNKGYPSYAPTRDYTAYRSWQP*
Homsap    FWVLVVVGGVLACYSLLVTVAFIIF WVRSKRSRLLHSDYMNMTPRRPGPTRKHYQPYAPPRDFAAYRS*
Musmus    FWALVVVAGVLFCYGLLVTVALCVI WTNSRRNRLLQSDYMNMTPRRPGLTRKPYQPYAPARDFAAYRP*
```

(positions 109–114 in CDR3-IMGT). The two cysteine residues (C 23 and C 114) involved in the disulfide bond of the V-LIKE-DOMAIN were conserved in all species except in *Xenopus* where these residues were replaced by phenylalanine (F), a rare but known pattern of substitution in IgSF domains. The sequences also possessed a cysteine (C) residue in the connecting region, which is involved in dimerization. This residue was absent in *Xenopus* and in chicken, which may suggest a different configuration of the receptor. In the intracytoplasmic region of CD28 sequences, a remarkable feature is a tyrosine-based motif (D/E)-Y-(M/I)-(N/D) that is conserved from teleosts to mammals. The typical PI3K-binding motif (Y-M-x-M) [28] present in mammals and birds is not conserved in cold-blooded vertebrates. In contrast, the Y-x-N motif a signature for SH2 domain binding of the adaptor growth factor receptor bound protein-2 (GRB2) [29]—is well conserved through vertebrates. The Asn residue in the above motif is either present or substituted by a residue with similar biochemical properties. Interestingly, the conserved Asparagine is critical for human CD28 signaling as human T cells expressing CD28 with a N-mutated motif are not costimulatory. They do not phosphorylate VAV1 but the binding of PI3K is not affected [30]. The two P-x-x-P motifs present in the intracytoplasmic region of human CD28 are also important for its signaling properties as they can associate with either ITK or the SH3 domain of LCK and GRB2 [30,31], respectively. These two proline-based motifs are present in *Xenopus*, but they are not conserved in teleosts. The first P-x-x-P motif is only observed in salmonids and the P-Y-A-P motif is absent from all fish sequences. Taken together, these observations suggest that the interactions mediating CD28 signaling are distinct in the different classes of vertebrates with the exception of the Y-x-N motif which mediates the binding of SH2 domains.

3.2. CTLA4 homologs in teleosts, chicken and Xenopus

Homologs of CTLA4 sequences were found both among ESTs and genomic sequences of zebrafish. The CTLA4-like gene was partially duplicated in the available assembly of the zebrafish genome as exons 3 and 4 which encode the transmembrane and intracytoplasmic regions were found twice in the same orientation and at close proximity. Since the sequences of the two copies were not identical, they may correspond to a partial and recent duplication of the region. Sequences similar to CTLA4 were also identified in medaka. In addition, TBLASTN searches with fish and mammalian CTLA4 sequences led to the identification of avian and amphibian homologs.

All fish CTLA4 sequences found in this study are significantly similar to mammalian CTLA4 (Fig. 2) and possessed typical features of the rainbow trout sequence. Among proteins retrieved by BLASTP, the highest E values were obtained for mammalian CTLA4 sequences (E values $< 1e$-9) but not for mammalian CD28 or ICOS. Therefore, teleosts possess homologs of CTLA4 similar to most of the tetrapod lineages. The proline-based motif (L/MYPPPY) which binds B7-1/B7-2 in human and mouse CTLA4 was conserved in all vertebrates including fish, as well as the G-strand G-N-G-T motif that is necessary for the high affinity of CTLA4 for B7-1/B7-2 [32,33]. A cysteine is also conserved in the connecting region that is likely involved in the dimerization of the receptor. While the V-LIKE-DOMAINs showed the typical features of CTLA4, the intracytoplasmic signaling motifs described in mammals were absent in the teleost sequences. The Y-V/E-K-M motif responsible for binding PI3K in humans and mice was conserved in chicken and *Xenopus* CTLA4. In contrast, the teleost CTLA4 intracellular regions lack this motif and some of them encode an Y-x-x-F motif in the C

```
               <------------- L-REGION------->  <-------------V-LIKE-DOMAIN------------
                                                <----------FR1-IMGT-------> <CDR1-IMGT>
                                                ------A-------> -----B---->      BC
                                                1      10    15 16 20 23 26    30      38
                                                |.......|....|  |..|.....|...  ...|......|
       Oncmyk       MTLSLLTFLCLGLCPVW NALRVTQPY.RVVSH RGEVELFCSYH HTGRN.EPEELR
       Orylat       MVVADCVFGWRMLTVLGVCLPL YSAVKVVQP.RVEST NGTARIRCTFH PEMN...LEDLR
       Danrer                                  /SMHCSFR SKLRP.EEVQVS
       Xentro       MMRIPFITGIFCLITNAS GLKVTQPDI.IVANR HGKAMLVCDYR IHAK...VEEMR
       Xentro2                          /SKSYPKV.VIASQ DGEPCLLCHHV PSTD...SKFNL
       Galgal                           /MEVTQPAI.VLANR QGVASLVCNYK HIGN...AKEIR
       Homsap  MACLGFQRHKAQLNLATRTWPCTLLFFLLFIPVFCK AMHVAQPAV.VLASS RGIASFVCEYA SPGK...ATEVR
       Musmus  MACLGLRRYKAQLQLPSRTWPFVALLTLLLFIPVFSE AIQVTQPSV.VLASS HGVASFPCEYS PSHN...TDEVR


               ------------------------ V-LIKE-DOMAIN------------------------
               <----FR2-IMGT----><CDR2-IMGT><-------------FR3-IMGT------------
               ---C---> ---C'---> C'C''    ---C''--> ----D----> -----E----->
               39 41  46 47 50    55     60    66 70  74 75  80  84 85 90      96
               |......|AB|..|...|  ....|.....  |...|...|  |...|...|  |...|......|A
       Oncmyk  ITLYRGMYGEQEE.QKVCT SSFTH.NNTA FQVEGEGER KVRCRGQLRP GKVNLTISGLRG.
       Salsal                     /VTCRGQLGP GKVNLTISGLRG.
       Orylat  VTLLRGL...HGH.QELCS FFLNL.TGQT ETHGGRP.. .GQCSGQTAD GAVEV.L.DLKA.
       Danrer  VYKALH.....GH.ERICS AYVNL.SEPY FATDGP.. .VRCRGNINF GGVDLIIAGLRG.
       Xentro  FRLLRKMG..NQV.KEICA FSYS.TNYES VTTGDA... .IQCEGEPGP NNVTLHLSGMQM.
       Xentro2 TLMRGN.....KK.QEVCM VYTDG.KNTS FYSWNDN.. .PKCNWTKSN DSSISFTLSNFDI
       Galgal  VTLLKQTG..DKF.TEICA STYTT.EFKM FSVEEV... .IQCHVSPSR NNVTLTLTGLQA.
       Homsap  VTVLRQAD..SQV.TEVCA ATYMM.GNEL TFLDDS... ..ICTGTSSG NQVNLTIQGLRA.
       Musmus  VTVLRQTN..DQM.TEVCA TTFTE.KNTV GFLDYP... ..FCSGTFNE SRVNLTIQGLRA.


               ---------V-LIKE-DOMAIN-----------> <------CONNECTING-REGION------->
               FR3-IMGT> < CDR3-IMGT > < FR4-IMGT>
               ---F---> FG       -----G---->
               97   104 105  110 115 118 121   128
               |..|...|  .....|....|..  |........|
       Oncmyk  NDTDLYRC AIEVLYPPPYLRT FGNGTLLYIT. EEPGCFTPEAQRRDDVTGETSVRLP......
       Salsal  NDTDLYRC AIEVLYPPPYLRT FGNGTLLYIT. EEPGCFTPEAQRRDDITGETSVRLP......
       Orylat  ADTDVYRC KIQTFYPPPYLQL TGNGTLIHVI. EHPDCPSVNAQRRHEVEGEKEATLA......
       Danrer  EDTDVYRC EIEILFPPPYLRA LGNGTVVYIQ. EIPNCLTAHNASTQSQRQSKSQTSDRETVEN
       Xentro  SDTGMYIC KLDIMYPPPYRTT EGNGTLIYVSD LMSECAQSIEPPEFI................
       Xentro2 KHTDNYTC EIRIFYPPPFRSI IINETYVYIH/
       Galgal  NDTGLYVC KMERMYPPPYFMN KGNGTQLYVID .PEPCPDTA.....................
       Homsap  MDTGLYIC KVELMYPPPYYLG IGNGTQIYVID .PEPCPDSD.....................
       Musmus  VDTGLYLC KVELMYPPPYFVG MGNGTQIYVID .PEPCPDSD.....................


               <----------TM----------> <---------------IC------------------------>
       Oncmyk  LAGL..AAVLIVISAIAILLVHQVL QRKRRFEAIVPMMSKNDGRFDYGNFQ*
       Salsal  LAGMTAATILIVISAIAILLVHQVL QRKRRFEAIIPMTSQNDGRFDYGNFQ*
       Orylat  AASTPVAVLVTVIICVLVIIIY..L QVVRTNIKVVDR.SNSEKVHEHLFLRGCREQQVVSNNYPTHSSV*
       Danrer  DIGPLLLLYIILIITSCSFFLQIAI CKWRTSSSMAPMLSQKE...SYVKF*
       Xentro  LDQRILLVVCLVMFLYSMFITAVLL ..CGKQRKKFTVGNYEKM..LESDQGNGFSPYYIRVN*
       Galgal  TYLWVLGATASGFFLYSIIISAIVV GKAIQRRQRLTTGVYVKMPSEKLEKKVI..PFHITV*
       Homsap  FLLWILAAVSSGLFFYSFLLTAVSL SKMLKKRSPLTTGVYVKMPPTEPECEKQFQPYFIPIN*
       Musmus  FLLWILVAVSLGLFFYSFLVTAVSL SKMLKKRSPLTTGVYVKMPPTEPECEKQFQPYFIPIN*
```

Fig. 2. Multiple alignments of CTLA4 and CTLA4 related sequences. Sequence numbering is according to the IMGT unique numbering for V-DOMAIN and V-LIKE-DOMAIN [24] and IMGT region delimitations. Residues of the ligand-binding site (V-LIKE-DOMAIN) and residues of the potential signaling sites (IC region) are bold. Relevant conserved positions are shadowed. Sequences in the alignment were computed from the following entries: *Oncorhynchus mykiss* (Oncmyk) AY789436; *Salmo salar* (Salsal): CA044588; *Danio rerio* (Danrer): Ensembl assembly v36 chromosome 9, 16772109–16789355 and ESTs bj016306, bj517516, bj495736 and bj496049; *Oryzias latipes* (Orylat): scaffold3303 and EST GOLWno4565_h05.b1; *Xenopus tropicalis* (Xentro): Ensembl v36 scaffold_203; *Gallus gallus* (Galgal): Ensembl v36 chromosome C; *Homo sapiens* (Homsap) ENSG00000163599 and *Mus musculus* (Musmus) EN-SMUSG00000026011. Xentro2 (Ensembl v36 scaffold_168) corresponds to a *Xenopus* sequence similar to both CTLA4 and ICOS. Zebrafish CTLA4 partial sequence has been submitted to the Third Party Annotation Section of the DDBJ/EMBL/GenBank databases (accession TPA: BN000923).

terminus as shown for the salmonid and zebrafish sequences. Moreover, the P-x-x-P and P-x-x-x-P motifs in the CTLA4 intracytoplasmic region in human and mouse were present in other mammals such as dog and cattle (data not shown), but were not conserved in other lineages. Taken together, these observations extend and confirm our previous observations for rainbow trout CTLA4 and indicate that CTLA4 was subjected to teleost-specific evolutionary events.

3.3. Others receptors from the CD28 family in vertebrates

In mammals, the CD28 family has three other members: the inducible positive costimulatory receptor ICOS and two inhibitory receptors, PD1 and BTLA.

Systematic searches in fish EST or genomic databases could not identify any sequence similar to mammalian ICOS or PD1 receptor. However, a typical ICOS receptor was found in the chicken genome and a sequence with a V-LIKE-DOMAIN equally similar to the IgSF domains of mammalian CTLA4 and ICOS was identified in *Xenopus* (Fig. 2, sequence Xentro2). This sequence did not contain the typical G-x-G bulge that is conserved in all CTLA4 sequences. This sequence may therefore constitute the *Xenopus* homolog of ICOS. A sequence similar to PD1 was found in the chicken, but neither typical ICOS nor PD1 were found among the ESTs or genome draft of *Xenopus*. Neither ICOS nor PD1 sequences could be identified in the teleost databases.

In contrast, several sequences homologous to BTLA were retrieved from *Xenopus* and from different teleosts. Two distinct, partial rainbow trout ESTs similar to BTLA were the first to be identified in our analysis. Two additional salmonid ESTs similar to the C terminus of BTLA were then identified in the Atlantic salmon EST gene index that span from the connecting peptide through the end of the intracytoplasmid domain. In the zebrafish genome, a gene localized on chromosome 4 apparently encodes a complete BTLA homolog. These sequences were all included in a multiple sequence alignment that includes also mammalian BTLA sequences (Fig. 3).

As observed for their mammalian counterparts, non-mammalian BTLA sequences contain an IgSF domain, a connecting region, a transmembrane region and a long intracytoplasmic region with conserved immunoreceptor tyrosine-based inhibition motifs (ITIMs). The Ig domain of BTLA was first considered to be a V-domain [18,19], but recent three-dimensional structural analysis demonstrated that it belongs to the I-set category of IgSF domains [9], which in the IMGT domain classification represents a C-LIKE-DOMAIN (see Table 3 in Ref. [25] for correspondence between the diverse domain designations). The presence of a C-LIKE-DOMAIN [25,29] was confirmed in this study for all BTLA sequences from the various vertebrates. The C-LIKE-DOMAIN from BTLA was clearly different from the V-LIKE-DOMAINs found in CD28 and CTLA4 as visualized by the IMGT Collier de Perles representation (Fig. 4). Indeed, BTLA lacks the C′ and C″ strands and the C′–C″ loop but instead has a CD transversal strand. In human, the two β sheets of the BTLA IgSF domain are linked by three disulfide bonds (cysteine residues 1–28, 23–104 (not shown in Fig. 4) and 44–45F). The C^1–C^{28} disulfide bond joins the N-terminal region of strand A to the BC loop, the conserved C^{23}–C^{104} disulfide bond connects the B and F strands, and the C^{44}–C^{45F} (IMGT C^{45F} is the cysteine found at position 6 between the C and CD strands) disulfide bond connects the C and CD strands, with an unusual β turn formed by six additional amino acids. The multiple sequence alignment demonstrates that cysteine residues 44 and 45F are highly conserved from teleosts to mammals, confirming that all BTLA domains likely share this feature. Whereas the C^{45F} is present in BTLA from C57BL/6 and other strains of mice, the C^{45F} is missing in the BALB/c sequence, which is included in the alignment [9]. However, the β turn structure (positions 45A–45F) is conserved (IMGT/3Dstructure-DB, 1xau_A) in the different murine BTLAs. The B7-binding site motif ("Y-P-P-P-Y") that is conserved in CD28, CTLA4 and ICOS is absent from all BTLA sequences. The three-dimensional structure of the HVEM–human BTLA complex established that the ligand-binding site of BTLA consists of two short regions including residues 1–8 (beginning of A strand) and residues 114–119 of the G strand [9]. Residues Q^3, R^8, R^{103}, N^{114}, Y^{124} and H^{119} are involved in hydrogen bonds and I^{116} maintains hydrophobic interactions at the HVEM/BTLA interface [9] (IMGT/3Dstructure-DB, http://imgt.cines.fr, entry: 2aw2). Interestingly, these residues (or amino acids with similar properties) were rather conserved within the different vertebrates at positions 3, 8, 103 and 119 suggesting

```
        <-------------- L-REGION-------> <---------------C-LIKE-DOMAIN-----------
                                           A             B          BC
                                         -------------->   ---------->
                                      1.11    10   15    16 20  23 26    30    36
                                       |........|....|...|...|.....|  ...|.....|
Danrer  MSFFFQIFCNVAVSLLLLSLSVSGTSNGTEPH CDVKLKVPRQTIFKARI..MTQLKINCTVT HNG....CQW
Oncmyk1                  ~LCVHGNAQDQGPT CEV..RVKRNTIWNAVP..QKALTINCPVR HCG.....EK
Oncmyk2       MDPLWSDHILLVLCLLFLVCINGNAQGP CEIEVPIKRNTVWNAVP..QKRLTINCPVK HCG.....ET
XenTro          MRGRGNFLNSLNVGCVRAQGRCKSE CSPQIHTERNTKYNESV..GESLAIQCPVS FCSK....EL
Homsap  MLGTGKLFWVFFLIPYLDIWNI.......HGKES CDVQLYIKRQSEHSILA..GDPFELECPVK YCA.....NR
Musmus  MLGTPRLFREFF.ILHLGLWSILCEKATKRNDEE CEVQLNIKRNSKHSAWT..GELFKIECPVK YCV.....HR


        ----------------------C-LIKE-DOMAIN----------------
           C             CD        D      DE        E
         ------->               ------->      ----------->
         39    45      45.1 45.7 77    84     85         96
         |....|ABC DEF |.....|   |......|     |..........|
Danrer  NPISWCKIYG DHC  KALNHSD  LIKMECRNTAQ..HEGMAFLIFQNITM
Oncmyk1 LHATWCKFVDLLNC  EQIHETE  QVTFGWRQK.....DILILDFKKISI
Oncmyk2 LRVTWCKFED ANC  KQINETE  QVKIGWKPFDRGSENRIAFLDFKKISR
Xentro  PVVSWCKFVG ETC  ESLSAGP  GRLMEWTDRK..ENTAVYLLTFVGAQL
Homsap  PHVTWCKLNG TTC  VKLE...  DRQTSWKEEK...NISFFILHFEPVLP
Musmus  PNVTWCKHNG TIW  VPLEVGP  QLYTSWEENR...SVPVFVLHFKPIHL


        ---------C-LIKE-DOMAIN----------> <--------CONNECTING-REGION------------------->
            F             FG          G
         ------->      .....|......|  ---------->
         97    104 105        117 118    128
         |.....|   .....|......|  |.........|
Danrer  ENSGSYRC THG......DTSV SHAINVTITGN .......GEDEFSHNQSNKTDDFNSVPDNGLK
Oncmyk1 NDDGLYRC GVAEN...GSTTV SHAIKVSVSDS .......NLEITSTADHNTNDSTDAKRDRNDPS
Oncmyk2 NDDGSYRC GMAGK...KSSIV SHAINVSVSDI .......NLTMTSTADHNTNNSPNAKHDEKYLS
Salmon                       ~VSVSDR .......NLKMTSTAEHNTNDSTIKKHDEKYLS
Xentro  NDTGYYRC TLTTQ...SVPVV SWCKFVGETCE SLSAGPGRLMENTVGRTVMVKIIGHKERTTAQLPSTTHIMGGCSRVE&
Homsap  NDNGSYRC SANFQ...SNLIE SHSTTLYVTDV ...........KSA......SERPSKDEMASRP.   &GSTQPGY
Musmus  SDNGSYSC STNFN...SQVIN SHSVTIHVRER TQNSSEHPLITVSDIPDATNASGPSTMEERPGRT


        <-----------TM------------> <----------------IC-------------------------------
Danrer  WLWPYVYICSGIAALVIIIITITLFVF NCQDTKSKTEELTNKNQYMETQSSDLLPPPLPHPRHDTR......
Oncmyk1 FM.PYVYICVGIAVLVVIVTAFSFFCL HAYKGESRSTGNRTKKQEMSTLTAPCPDLSFNPEETFPSSPSTPN
Oncmyk2 WL.PYAYICVGIVLLVVVVTPISFFYL YGCKGPTRSTNNGIKEQ.....AAPCPDLSSCSKG/
Salmon  WL.PYVYICVGIVLLVVVVTAISFLCI YGCKGPTRSTNNGIMEQ.....AAPCPDLSSCSKGTFPSSPSTPN
Xentro  WV.ILLRILIVIGASSLLFYCLGTITA NRRFSKTLRRENVTGINA....SKTPISAVEPLDIGSLHTPGP
Homsap  WL.LYSLLPLGGLPLLITTCFCLFCCL RRHQGKQNELSDTAGRE..INLVDAHLKSEQTEASTRQNS...QV
Musmus  WL.LYTLLPLGAL.LLLLACVCLLCFL KRIQGKEKKPSDLAGRD..TNLVDIP.......ASSRTNH...QA


        ----------------------------IC-------------------------------------------
Danrer  SLSQQRGSALDRGSEIPPGRGTSSAGRVKERSHHTVSAGRGEDDNALVYASLNHQAVPKVHRKTARQEPELS
Oncmyk1 PVQPANLYDNGLYCNRINGKNGALPFS/
Salsal  PAHSDNLYNNGPYGKHINRASAHSLLVTNGTLPSSDITARGQGSNHLVYASLNHQTPRESLKPPHIPNPQDECS
Xentro  ASANEITYNNITAEETCSSVEDSNDKEQNQTA*
Homsap  LLSETGIYDNDP..........DLCFRMQEGSEVYSNPCLEENKPGIVYASLNHSVIGLNSRLARNVKEAPT
Musmus  LFSGTGIYDNDP...........WSSMQDESELTISLQSERNNQGIVYASLNHCVIGRNPRQENNMQEAPT


        -------->
Danrer  EVAAVRFR*
Salsal  EVAAIRVY*
Homsap  EVASICVRS*
Musmus  EVASICVRS*
```

Fig. 3. Multiple alignments of BTLA and BTLA related sequences. Sequence numbering is according to the IMGT unique numbering for C-DOMAIN and C-LIKE-DOMAIN [25] and IMGT region delimitations. Residues of the ligand-binding site (C-LIKE-DOMAIN) and residues of the potential signaling sites (IC region) are bold. Relevant conserved positions are shadowed. Sequences in the alignment were computed from the following entries: *Oncorhynchus mykiss* (Oncmyk1: ca377385) (Oncmyk2: ca362415,cx253095); *Salmo salar* (Salsal): cb517927, cb514733; *Danio rerio* (Danrer): Ensembl assembly v36 chromosome 4_23583558–23585903; *Xenopus tropicalis* (Xentro): Ensembl v36 scaffold_351; *Homo sapiens* (Homsap) BC107092 and *Mus musculus* (Musmus) NM_001037719.

similar points of contact. The beginning of the G strand S^{118}–H^{119} is absolutely conserved in teleosts and tetrapods, suggesting that all BTLAs may bind the same ligand(s). The connecting region of the

BTLA sequences does not contain a cysteine residue similar to the one involved in receptor dimerization in CD28 and CTLA4. The assembled trout BTLA ESTs prematurely ended within the intracytoplasmic

region but the Atlantic salmon sequence showed relative conservation of the tyrosine-based interaction motifs that are present in rodent and human BTLA: both L-Y-D-N (a potential GRB2-binding motif) and L-V-Y-A-S-L-N-H-(X)$_{18/20}$-E-Y-A/S-I (ITIM/ITSM motifs) sequences motifs were present in teleosts and mammals. The first motif was also found in rainbow trout (L-Y-D-N) and in *Xenopus* (T-Y-N-N) sequences, but was absent from the zebrafish sequence. The second motif and neighboring residues were absolutely conserved in salmon, zebrafish and mammals. Intriguingly, the *Xenopus* sequence retrieved from both genomic and EST searches, lacked the C terminus of the intracytoplasmic region and thus did not contain an ITIM motif. Another noteworthy feature of the intracytoplasmic region of non-mammalian BTLA was the high frequency of proline residues situated as P-P, P-x-P or P-x-x-P motifs, which may have importance for signaling. Thus, teleosts possess receptors with a C-LIKE-DOMAIN similar to BTLA and conserved ITIM/ITSM motifs in their intracytoplasmic region that may constitute inhibitory receptors.

The multiplicity of BTLA-like sequences was a remarkable feature observed in teleosts. Two distinct rainbow trout sequences similar to BTLA were identified suggesting that this species may possess two different genes similar to BTLA, or that these genes possess very divergent alleles (Fig. 3). Zebrafish ESTs also grouped into two types of sequences related to BTLA (e.g. co921859 and dr723980). These sequences were used to explore the genomic diversity of zebrafish BTLA. BLAST analysis identified nine sequences with a C-LIKE-DOMAIN similar to the BTLA domain (Fig. 5). Only one gene (Chr4_1) possessed all six exons, encoding a BTLA-like receptor with the same length and structure as observed in mammals and salmonids. Apart from this gene, only one had exon 6, which encodes the intracytoplasmic region with the conserved ITIM (Chr9_1). Another gene (Chr11_2) encoded a premature STOP codon in the C-LIKE-DOMAIN thus constituting a likely pseudogene. Taken together, the diversity of C-LIKE-DOMAIN sequences related to BTLA indicated that these genes have been subjected to several duplication events, but only one complete BTLA gene with all exons was found in zebrafish genome.

3.4. Intron/exon structure of CD28 family members

The intron/exon structure of vertebrate genes is generally well-conserved inside a gene family. We therefore checked whether this applied to the members of the CD28 family.

In mammals, CD28 and CTLA4 genes have four exons. The first exon (EX1) codes for the signal peptide or L-REGION, exon 2 (EX2) codes for the V-LIKE-DOMAIN. The third exon (EX3) codes for the connecting region and the transmembrane region, and for the first three amino acids of the CTLA4 intracytoplasmic region. The fourth exon (EX4) codes for the intracytoplasmic region. We first investigated the intron/exon structure of CD28 and CTLA4 sequences identified in pufferfish, zebrafish and medaka. The position and length of the introns were determined using Genscan and direct sequence analysis and then compared to the structure of human CD28 and CTLA4 (Fig. 6). When the first intron was long (zebrafish and fugu CD28 and zebrafish CTLA4), it was not possible to identify the first exon with absolute confidence. However, this analysis showed convincingly that the intron/exon structure described for mammalian CD28 and CTLA4 is conserved in teleosts, confirming the results previously reported from salmonids. The only exception was medaka CTLA4, which displayed five exons instead of four. However, this was due to an additional intron inserted in the middle of exon 2 that encodes the V-LIKE-DOMAIN in other species but the overall structure of the gene was conserved. Interestingly, the additional intron of medaka CTLA4 was in splicing frame 0, and not in frame 1 as generally observed for introns between IgSF domain exons [24]. Intronic distances were not conserved but a general pattern emerged where the introns of fish were shorter than in humans. The same organization was observed in chicken and *Xenopus*. Taken together, these observations demonstrate that CD28 and CTLA4 genes have the same intron/exon structure in all vertebrate lineages, further substantiating the hypothesis that they are respectively true orthologs.

The intron/exon structure of BTLA was not strictly conserved among mammals, since the human BTLA lacked the short exon 3 (Fig. 6). However, the intron/exon structure of zebrafish BTLA was clearly similar to that of the mouse as the exon limits were localized at the equivalent positions, the size of exons was similar and the splicing frames of the introns were conserved. These features were observed for all exons identified in the different BTLA genes of zebrafish (Fig. 4). Furthermore, *Xenopus* BTLA showed the

same pattern of intron/exon organization. These observations therefore reinforced the hypothesis that BTLA-like sequences found in cold-blooded vertebrates were closely related to mammalian BTLA.

3.5. Multiplicity and localization of genes encoding CD28 members in vertebrates

The numbers and localization of genes detected in a representative list of species were summarized in

CD28

(A)   Omcorhynchus mykiss   Homo sapiens

CTLA4

(B)   Omcorhynchus mykiss   Homo sapiens

BTLA4

(C)   Omcorhynchus mykiss   Homo sapiens
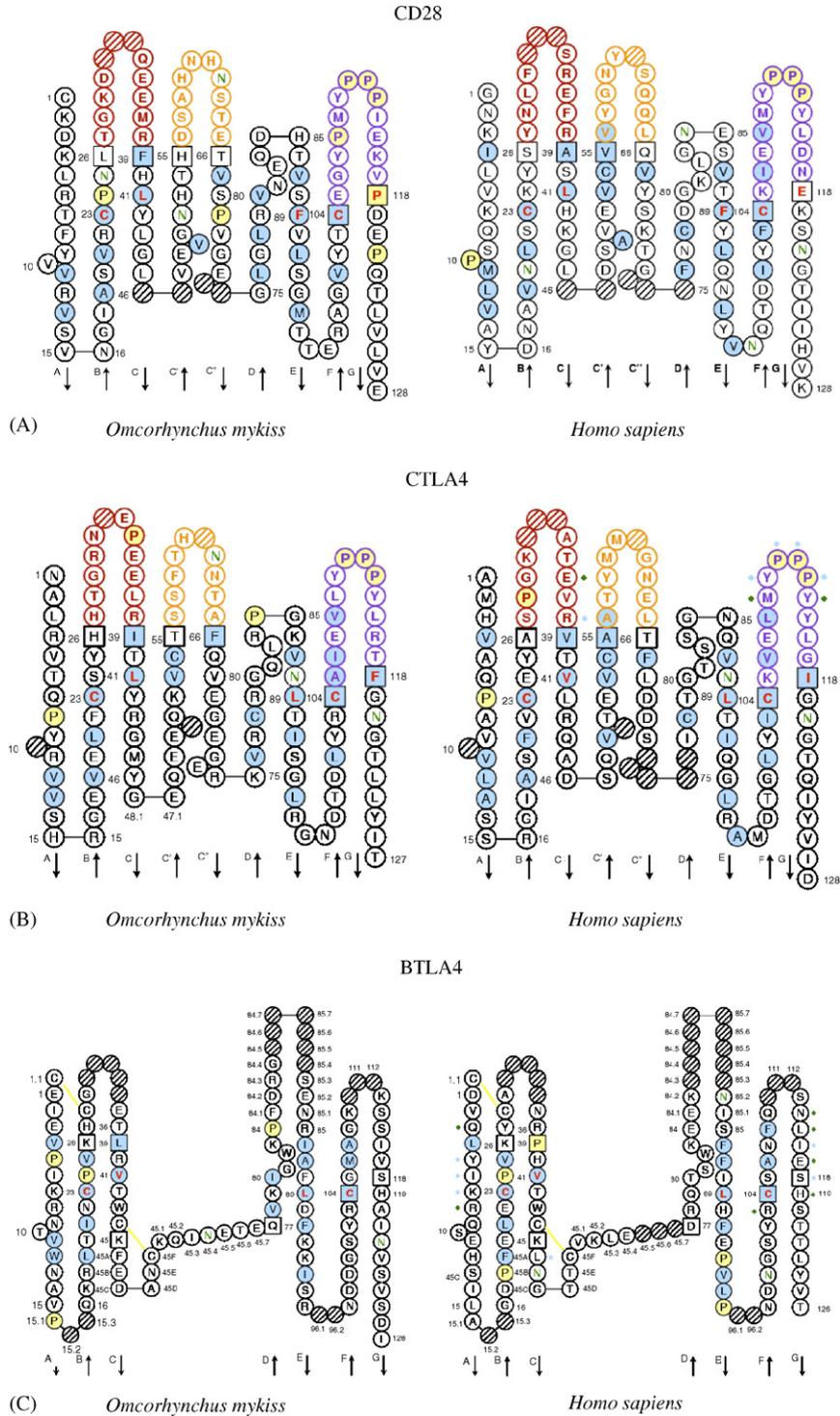
Table 1. In eutherian mammals, the genes encoding CD28, CTLA4 and ICOS were present as one copy and at close proximity on the same chromosome forming a cluster as previously reported in humans and mice. A unique copy of PD1 and BTLA genes was also identified in each eutherian genome. PD1 homologs were generally localized in the same region as the CD28/CTLA4/ICOS cluster, except in dogs where PD1 is on another chromosome. In contrast, the BTLA homologs of all species were not linked to the CD28/CTLA4/ICOS cluster. In the South American marsupial *Monodelphis*, CD28, CTLA4 and ICOS were also present in the same scaffold. A PD1 homolog, but no BTLA was detected in this species. These results indicated that the situation described in man and in mice (CD28/CTLA4/ICOS cluster) was generally well-conserved among mammals. Chicken CD28, CTLA4, ICOS and PD1 were also unique and localized in a similar fashion suggesting that these features may be common to mammals and birds. PD1 was lacking in the *Xenopus tropicalis* genome assembly while sequences highly similar to BTLA were identified. Only one gene similar to CD28 was identified in *Xenopus* and not surprisingly, a gene highly similar to CTLA4 was encoded in the same scaffold. A third CD28 family sequence was identified from another scaffold with an extracellular domain equally similar to CTLA4 and ICOS which also encoded a ligand-binding motif typical of ICOS in the FG loop of the V domain (see Fig. 2). Interestingly, several C-LIKE-DOMAIN sequences similar to the extracellular region of BTLA were identified in *Xenopus*, suggesting that this gene had

been subjected to several duplications, as described above for zebrafish. Among teleosts, unique copies of CD28 and CTLA4 were identified in several species. In zebrafish, CD28 and CTLA4 were located on distinct chromosomes as previously observed in rainbow trout. However, the detailed analysis of zebrafish BTLAs suggests that most of these homologous sequences may be pseudogenes or may encode truncated proteins of unpredictible function.

3.6. Synteny between regions encoding CD28 and CTLA4 genes in teleost genomes

In two teleost species, rainbow trout and zebrafish, CD28 and CTLA4 are not localized on the same chromosome, while they are tightly linked in mammals, birds and amphibians forming a "costimulatory gene cluster" with ICOS. Such a split of a gene complex onto different fish chromosomes may have resulted from the whole-genome duplication event that occurred in the fish lineage subsequent to its divergence from the other vertebrates [34]. Thus, a possible explanation for the lack of linkage of the CD28 and CTLA4 genes may be that each copy of the cluster would have lost either one or the other gene after the duplication. To test this hypothesis, we searched the fish genomes for the presence of conserved genes that flank the costimulatory cluster in mammals. RAPH1 is found immediately upstream of CD28 in humans and mice. Homologs of RAPH1 were present at close proximity for both CD28 and CTLA4 in the zebrafish genome (Fig. 7). No pseudogenes similar to either CD28 or CTLA4

Fig. 4. IMGT Colliers de Perles of rainbow trout (*Oncorhynchus mykiss*) and human CD28 (A), CTLA4 (B) and BTLA IgSF domain (C). The *Oncorhynchus mykiss* and *Homo sapiens* CD28 (A) and CTLA4 (B) V-LIKE-DOMAIN amino acid numerotation, and strand and loop delimitations, are based on the IMGT unique numbering for V-DOMAIN and V-LIKE DOMAIN [24] (http://imgt.cines.fr). The *Oncorhynchus mykiss* and *Homo sapiens* BTLA C-LIKE-DOMAIN amino acid numerotation, and strand and loop delimitations, are based on the IMGT unique numbering for C-DOMAIN and C-LIKE DOMAIN [25]. In (A) and (B), the loops BC, C′C″ and FG (corresponding to the CDR-IMGT) are limited by amino acids shown in squares (anchor positions), which belong to the neighboring strands (FR-IMGT in V-DOMAINs). BC loops are represented in red, C′C″ loops in orange and FG loop in purple. In (C), the positions 26, 39 and 104 are shown in squares by homology with the corresponding regions in the V-set (V-DOMAINs and V-LIKE-DOMAINs). Positions 45 and 77 which delimit the characteristic CD strand of the C-set (C-DOMAIN and C-LIKE-DOMAIN), and position 119 which corresponds structurally to J-PHE and J-TRP of the immunoglobulin and T cell receptor J-REGION are also shown in squares. The CD strand in BTLA is characterized by a beta turn with six additional amino acids (45A–45F) and a disulfide bridge between positions 44 and 45F. Amino acids are shown in the one-letter abbreviation. Positions at which hydrophobic amino acids (hydropathy index with positive value: I, V, L, F, C, M, A) and tryptophan (W) are found in more than 50% of analyzed sequences are shown in blue. All proline (P) are shown in yellow. Hatched circles correspond to missing positions according to the IMGT unique numbering. Arrows indicate the direction of the beta strands and their designation in 3D structures (from IMGT Repertoire, http://imgt.cines.fr). N from *N*-glycosylation sites are in green. EMBL/GenBank/DDBJ accession numbers: *Oncorhynchus mykiss* CD28: genbank:AY789435, CTLA4: AY789436 and BTLA: ca362415,cx253095; *Homo sapiens* CD28: M37812-M37815, CTLA4: M37243-M37245 and BTLA: AJ717664. PDB and IMGT/3Dstructure-DB entries: *Homo sapiens* CD28:1yjd_C, CTLA4: 1i8l_C and BTLA:2aw2_A. Green and blue dots indicate amino acids having hydrogen bonds and hydrophobic interactions, respectively, with amino acids of B7-1 for (B) and HVEM for (C).

Table 1
Co-stimulatory receptors in vertebrate genomes

Species	CD28	CTLA4	ICOS	PD1	BTLA
Mammals (eutherians)					
Homo sapiens	1(chr 2q33)[a,b,c]	1 (chr 2q33)[b]	1(chr 2q33)[b]	1(chr2q37.3)[b]	1 (chr 3q13.2)[b]
Canis familiaris	1 (chr 37)[b]	1 (chr 37)[b]	1(chr 37)[b]	1 (chr 25)[b]	1 (chr 33)[b]
Bos taurus	1 (chr 2)[b]	1 (chr 2)[b]	1 (chr 2)[b]	1(chr 2)[b]	1 (not placed)[b]
Mus musculus	1 (chr 1C2)[b]	1 (chr 1C2)[b]	1 (chr 1C2)[b]	1(chr 1D)[b]	1 (chr16B5)[b]
Mammals (marsupials)					
Monodelphis domestica	1(scaff39)[b]	1(scaff39)[b]	1(scaff39)[b]	1(scaff162)[b]	—
Birds					
Gallus gallus	1 (chr 7)[b]	1 (chr 7)[b]	1 (chr 7)[b]	1 (chr9)[b]	—
Amphibians					
Xenopus tropicalis	1 (scaff 203)[b]	1 (scaff 203)[b]	1(scaff 168)[b]	—	3[b,c](scaff 351, scaff 545) [d]
Teleosts (ESTs only)					
Oncorhynchus mykiss	1 (chr 7)[e]	1 (chr 34)[e]	—	—	>2[f]
Salmo salar	1[f,e]	1[f,e]	—	—	>2[f]
Ictalurus punctatus	1[f]	—	—	—	—
Epinephelus coioides	1[f]	—	—	—	—
Teleosts (complete genome)					
Danio rerio	1 (chr 1)[b,g]	1 (chr 9)[b]	—	—	9[b,f](chr 4, 9 11)
Fugu rubripes	1(scaff 32)[b]	—	—	—	—
Tetraodon nigroviridis	1 (chr 3)[b]	—	—	—	—
Oryzias latipes	1[b] (scaff 114, 299)	1[b] (scaff 468, 3303)	—	—	5[b] (scaff 203, scaff 2417)

[a]Sequences used for this table were extracted from NCBI, Ensembl, TIGR and NIG databases. Genome assembly corresponds to the Ensembl v36 release.

[b]Based on genome data.

[c]Two of three BTLA-like sequences are present on the same scaffold.

[d]Xenopus ESTs confirming the diversity of expressed BTLAs: bx752445, cf784237, al955570, cf784238, cf784237, cx479995.

[e]Ref. [23].

[f]Based on EST sequences.

[g]The same CD28 genomic sequence is present twice in close proximity and in reverse orientation, corresponding most probably to an assembly artifact.

were present in the regions. Furthermore, a homolog of RAPH1 was also found close to CD28 in the fugu genome. These observations suggest the existence of a gene cluster comprising CD28 and CTLA4 counterparts within the common ancestor of both tetrapods and teleosts.

4. Discussion

We have recently identified homologs of costimulatory receptors CD28 and CTLA4 in a teleost fish, the rainbow trout. To gain further insight into the history of the CD28 costimulatory receptors during vertebrate evolution, we performed a systematic survey of these sequences in available databases.

The comparison of sequences identified in several other teleosts including Atlantic salmon, medaka

and zebrafish confirmed the presence of specific features that were previously observed in rainbow trout CD28 and CTLA4. The B7-binding site (conserved PPP motif) was the most prominent feature conserved in all CD28 and CTLA4 sequences from all teleosts and tetrapods. Interestingly, the flanking residues are more variable in CD28 than in CTLA4, which is consistent with a strong constraint of high affinity of CTLA4 for its ligands B7-1 and B7-2 as reported in mammals. This is reminiscent of the observation that cattle CD28 binds B7 ligands through a slightly divergent motif LYPPPK motif, while cattle CTLA4 has the typical MYPPPY ligand-binding motif [35].

The quest for CD28 orthologs also led to the discovery of sequences related to mouse and human BTLA in *Xenopus* and in different teleosts. The comparison of BTLA sequences from different
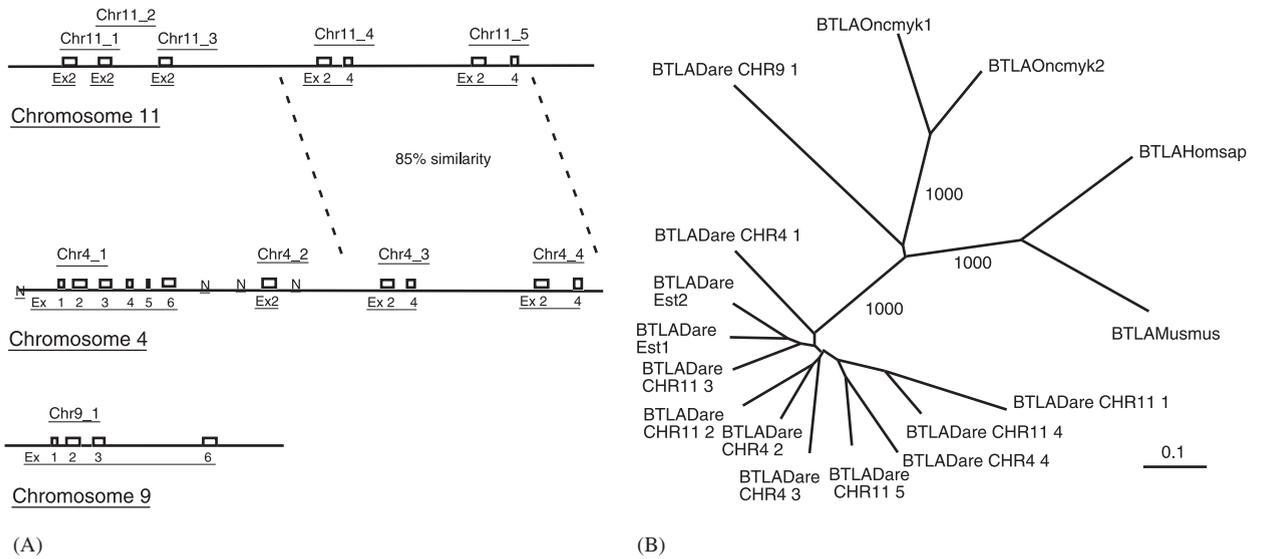
Fig. 5. Structure and diversity of zebrafish BTLA loci. (A) follows the Ensembl v36 zebrafish genome assembly. These sequences were localized on chromosome 4 (four BLAST hits, Chr4_1: 23583558–23585903; Chr4_2: 23614061–23614402; Chr4_3: 23616552–23617362; Chr4_4: 23623223–23623744), on chromosome 9 (1 BLAST hit, Chr9_1: 11207480–11212084) and on chromosome 11 (5 BLAST hits, Chr11_1: 26662805–26662805; Chr11_2: 26663597–26663908; Chr11_3: 26667634–26667969; Chr11_4: 26680331–26680894; Chr11_5: 26695333–26695908). Panel (B) represents the phylogenetic tree (NJ, clustalw) derived from a multiple alignment of the IgSF domains from selected BTLA receptors. Dare: *Danio rerio*; Oncmyk: *Oncorhynchus mykiss*; Musmus: *Mus musculus*; Homsap: *Homo sapiens*. Boostrap values higher than 900 (Boostrap:1000) are indicated at corresponding branches.

tetrapods and teleosts reveals that the BTLA IgSF domains do not correspond to V-LIKE-DOMAINs but to C-LIKE-DOMAINs in the IMGT® domain nomenclature [24,25,35]. Furthermore, the key residues localized at the HVEM/humanBTLA interface were rather well conserved, suggesting that the nature of the ligand recognition may be the same for all vertebrate BTLAs. These observations were compatible with the idea that the BTLA ligand would not belong to the B7 family but rather to the TNFR family, as demonstrated for human BTLA [18,19]. In addition to the characteristics of the C-LIKE-DOMAIN, the presence of the conserved tyrosine-based potential signaling motifs in the intracytoplasmic region of BTLAs from mammals and lower vertebrates strongly suggests that they are orthologous. It also suggests that all of these receptors induce negative (inhibitory) signals, as demonstrated for human BTLA.

Although the BTLA ligand has been identified and proven not to be B7x as reported earlier, several experiments suggested that both molecules are engaged in the control of T cell responses. BTLA deficiency and treatment with anti-B7x antibodies both result in enhanced T cell responsiveness [36]. In mammals, the two genes belong to a conserved linkage group with paralogs. CD80 and 86 are

linked to BTLA on chromosome 3q in human and 16B3-5 in mouse. B7x is present on chromosome 1p13 close to prostaglandin F2a receptor-regulatory protein (FPRP) and CD101, which are members of the PG regulatory-like protein (PGRL) family. The PGRL gene itself is in the 1q23 region, a segment paralogous of the 3q13 segment where BTLA and CD80 and 86 are located [37]. Interpreted in the context of gene duplication by polyploidization through which the paralogs are generated [38], our observation suggests that the linkage BTLA/B7 family is ancient. These linkage studies therefore confirm the observations (different exon/intron structures and an Ig domain belonging to the C-type) showing that BTLA is structurally not a member of the CD28 ([9] and Fig. 4) but belongs to another evolutionary lineage present from primitive vertebrates onwards.

A tentative evolutionary scenario emerges from the systematic analysis of sequences related to CD28, CTLA4, ICOS, PD1 and BTLA among tetrapods (Fig. 8). This view remains rather speculative since the diversity of teleosts has yet to be fully characterized and since both birds and amphibians are represented by only one species. The reptiles could not be taken into account since there are no sequence data available for these genes.
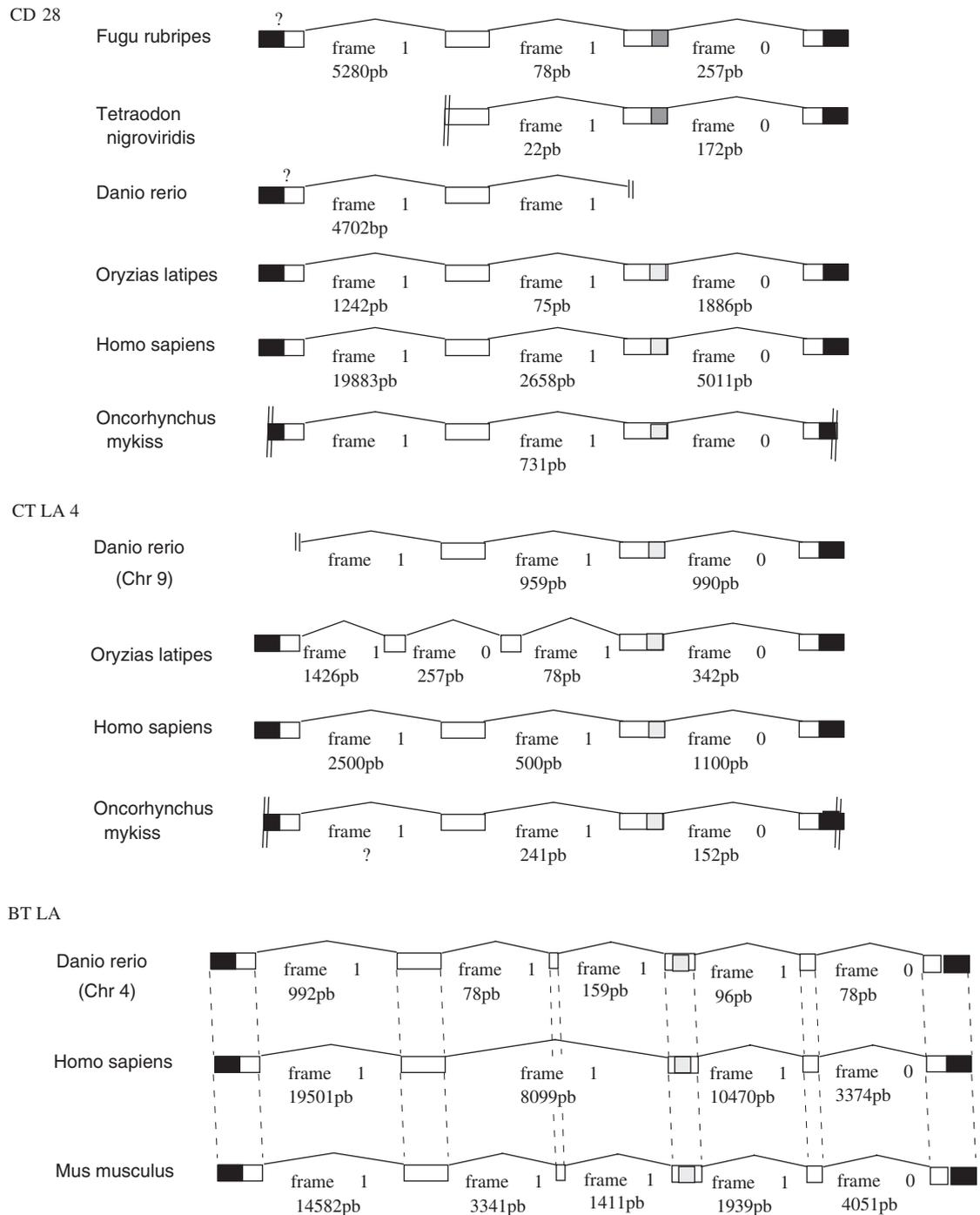
Fig. 6. Intron/exon structure of fish CD28, CTLA4 and BTLA genes compared with their human counterpart. Human CD28, CTLA4 and BTLA structures were from http://www.ensembl.org (ENSG00000178562, ENSG00000163599 and ENSG00000186265). Medaka CD28 and CTLA4 structure were deduced from genomic scaffold3303 and 114, respectively, available at http://dolphin.lab.nig.ac.jp/medaka/. Pufferfish CD28 was from emb|CAAB01000123.1| and Tetraodon CD28 from Chromosome3 6229231–6230005 Ensembl v38. For zebrafish CTLA4, the intron/exon structure was deduced from Ensembl v36 zebrafish genome assembly, chromosome 9 16772109–16789355. For zebrafish BTLA, the intron/exon structure was deduced from Ensembl v36 zebrafish genome assembly, chromosome 4 23583558–23585903. Mouse BTLA structure was from http://www.ensembl.org (ENSMUSG00000052013). UTR regions are represented in black, and transmembrane regions in light gray. Genbank accession numbers for trout CD28 and CTLA4 introns are: AY789437, AY789438, AY789439, AY789440, AY789441.
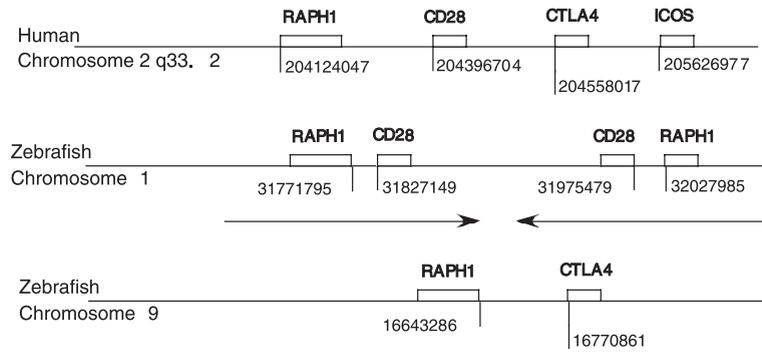
Fig. 7. Conserved association of CD28 and CTLA4 genes with RAPH1 in human and zebrafish genomes. The structure is from Ensembl v36 assemblies of human and zebrafish genomes.
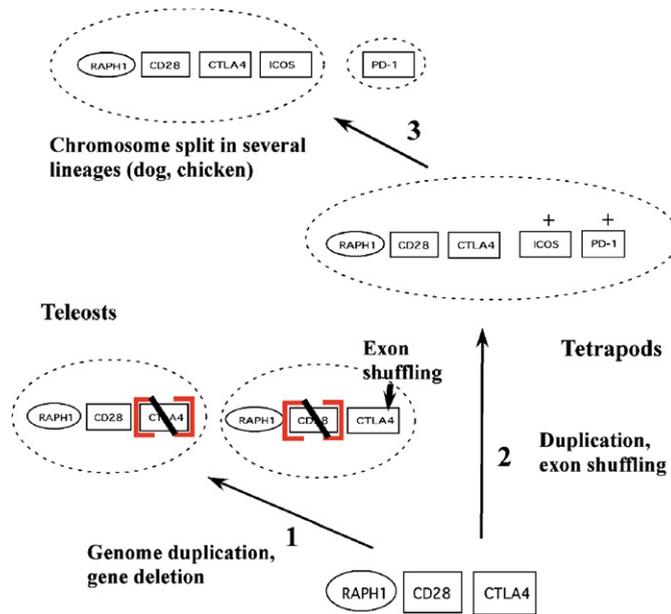
Fig. 8. Tentative evolutionary scenario of the CD28 family. The dotted lines indicate different chromosomes, and + corresponds to new genes appeared by duplication.

However, our survey of the costimulatory receptors revealed that CD28, CTLA4 and BTLA sequences were already present in the common ancestor of teleosts and tetrapods. Several duplications of BTLA in fish and amphibians would explain the multiple, slightly divergent BTLA sequences found in these lineages. In contrast, BTLA was probably not subjected to further diversification during the evolution of tetrapods. In light of this observation, another evolutionary pathway may have occurred for the CD28 family (Fig. 8). We did not find many copies of CD28, CTLA4 or other members of the family in any teleost species, but CD28 and CTLA4 were localized to different chromosomes. This may

be due to the deletion of one gene from two copies of the whole region, as suggested by the presence of the RAPH1 gene in close proximity of both fish CD28 and CTLA4. Indeed, a whole-genome duplication event likely occurred in the beginning of the evolution of teleosts [34] and therefore provides a rationale explanation for the duplication. Such a schedule has been proposed to explain the split of the MHC class I and class II loci observed in teleosts, which is unique among vertebrates. The intracytoplasmic region of fish CTLA4 is deficient of conserved signaling motifs that is likely due to exon shuffling events either in the first teleosts or in the early tetrapods. In tetrapods, members of the

CD28 family remained at close proximity but were subjected to both duplication and diversification, leading to the emergence of two new members, ICOS and PD1. Our results suggest that ICOS probably appeared before the separation of amphibians from the main stream of tetrapod evolution, while PD1 was first detected in birds. However, a broader sample of species is necessary to clarify this issue. Finally, chromosome recombination or break probably led to the separation of PD1 from the main costimulatory cluster in some sublineages. Accordingly, PD1 is localized on another chromosome in species with a high number of chromosomes such as dog and chicken.

BTLA and the members of the CD28 family respectively, constitute two ancient and distinct lineages of receptors that are represented in all groups of vertebrates studied to date. Further studies will show how the signaling properties of these receptors and their specificity for different ligands contribute to the activation and regulation of vertebrate lymphocytes and immune responses.

References

[1] Sharpe AH, Freeman GJ. The B7-CD28 superfamily. Nat Rev Immunol 2002;2:116–26.

[2] Acuto O, Michel F. CD28-mediated co-stimulation: a quantitative support for TCR signalling. Nat Rev Immunol 2003;3:939–51.

[3] Hutloff A, Dittrich AM, Beier KC, Eljaschewitsch B, Kraft R, Anagnostopoulos I, et al. ICOS is an inducible T-cell co-stimulator structurally and functionally related to CD28. Nature 1999;397:263–6.

[4] Swallow MM, Wallin JJ, Sha WC. B7h, a novel costimulatory homolog of B7.1 and B7.2, is induced by TNFα. Immunity 1999;11:423–32.

[5] Ishida Y, Agata Y, Shibahara K, Honjo T. Induced expression of PD-1, a novel member of the immunoglobulin gene superfamily, upon programmed cell death. EMBO J 1992;11:3887–95.

[6] Freeman GJ, Long AJ, Iwai Y, Bourque K, Chernova T, Nishimura H, et al. Engagement of the PD-1 immunoinhibitory receptor by a novel B7 family member leads to negative regulation of lymphocyte activation. J Exp Med 2000;192:1027–34.

[7] Latchman Y, Wood CR, Chernova T, Chaudhary D, Borde M, Chernova I, et al. PD-L2 is a second ligand for PD-1 and inhibits T cell activation. Nat Immunol 2001;2:261–8.

[8] Watanabe N, Gavrieli M, Sedy JR, Yang J, Fallarino F, Loftin SK, et al. BTLA is a lymphocyte inhibitory receptor with similarities to CTLA-4 and PD-1. Nat Immunol 2003;4:670–9.

[9] Compaan DM, Gonzalez LC, Tom I, Loyet KM, Eaton D, Hymowitz SG. Attenuating lymphocyte activity: the crystal structure of the BTLA-HVEM complex. J Biol Chem 2005; 280:39553–61.

[10] Ling V, Wu PW, Finnerty HF, Agostino MJ, Graham JR, Chen S, et al. Assembly and annotation of human chromosome 2q33 sequence containing the CD28, CTLA4, and ICOS gene cluster: analysis by computational, comparative, and microarray approaches. Genomics 2001;78: 155–68.

[11] Lanzavecchia A, Lezzi G, Viola A. From TCR engagement to T cell activation: a kinetic view of T cell behavior. Cell 1999;96:1–4.

[12] Walunas TL, Lenschow DJ, Bakker CY, Linsley PS, Freeman GJ, Green JM, et al. CTLA-4 can function as a negative regulator of T cell activation. Immunity 1994;1: 405–13.

[13] Mandelbrot DA, McAdam AJ, Sharpe AH. B7-1 or B7-2 is required to produce the lymphoproliferative phenotype in mice lacking cytotoxic T lymphocyte-associated antigen 4 (CTLA-4). J Exp Med 1999;189:435–40.

[14] Coyle AJ, Lehar S, Lloyd C, Tian J, Delaney T, Manning S, et al. The CD28-related molecule ICOS is required for effective T cell-dependent immune responses. Immunity 2000;13:95–105.

[15] Agata Y, Kawasaki A, Nishimura H, Ishida Y, Tsubata T, Yagita H, et al. Expression of the PD-1 antigen on the surface of stimulated mouse T and B lymphocytes. Int Immunol 1996;8:765–72.

[16] Nishimura H, Minato N, Nakano T, Honjo T. Immunological studies on PD-1 deficient mice: implication of PD-1 as a negative regulator for B cell responses. Int Immunol 1998;10:1563–72.

[17] Nishimura H, Nose M, Hiai H, Minato N, Honjo T. Development of lupus-like autoimmune diseases by disruption of the PD-1 gene encoding an ITIM motif-carrying immunoreceptor. Immunity 1999;11:141–51.

[18] Sedy JR, Gavrieli M, Potter KG, Hurchla MA, Lindsley RC, Hildner K, et al. B and T lymphocyte attenuator regulates T cell activation through interaction with herpesvirus entry mediator. Nat Immunol 2005;6:90–8.

[19] Gonzalez LC, Loyet KM, Calemine-Fenaux J, Chauhan V, Wranik B, Ouyang W, et al. A coreceptor interaction between the CD28 and TNF receptor family members B and T lymphocyte attenuator and herpesvirus entry mediator. Proc Natl Acad Sci USA 2005;102:1116–21.

[20] Young JR, Davison TF, Tregaskes CA, Rennie MC, Vainio O. Monomeric homologue of mammalian CD28 is expressed on chicken T cells. J Immunol 1994;152:3848–51.

[21] Arstila TP, Vainio O, Lassila O. Evolutionarily conserved function of CD28 in $\alpha\beta$ T cell activation. Scand J Immunol 1994;40:368–71.

[22] Koskela K, Arstila TP, Lassila O. Costimulatory function of CD28 in avian $\gamma\delta$ T cells is evolutionarily conserved. Scand J Immunol 1998;48:635–41.

[23] Bernard D, Riteau B, Hansen JD, Phillips RB, Michel F, Boudinot P, et al. Costimulatory receptors in a teleost fish: typical CD28, elusive CTLA4. J Immunol 2006;176:4191–200.

[24] Lefranc MP, Giudicelli V, Kaas Q, Duprat E, Jabado-Michaloud J, Scaviner D, et al. IMGT, the international ImMunoGeneTics information system. Nucleic Acids Res 2005;33:D593–7.

[25] Lefranc MP, Pommie C, Ruiz M, Giudicelli V, Foulquier E, Truong L, et al. IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. Dev Comp Immunol 2003;27:55–77.

[26] Lefranc MP, Pommie C, Kaas Q, Duprat E, Bosc N, Guiraudou D, et al. IMGT unique numbering for immunoglobulin and T cell receptor constant domains and Ig superfamily C-like domains. Dev Comp Immunol 2005;29:185–203.

[27] Kaas Q, Ruiz M, Lefranc MP. IMGT/3Dstructure-DB and IMGT/StructuralQuery, a database and a tool for immunoglobulin, T cell receptor and MHC structural data. Nucleic Acids Res 2004;32:D208–10.

[28] Prasad KV, Cai YC, Raab M, Duckworth B, Cantley L, Shoelson SE, et al. T-cell antigen CD28 interacts with the lipid kinase phosphatidylinositol 3-kinase by a cytoplasmic Tyr(P)-Met-Xaa-Met motif. Proc Natl Acad Sci USA 1994;91:2834–8.

[29] Songyang Z, Shoelson SE, Chaudhuri M, Gish G, Pawson T, Haser WG, et al. SH2 domains recognize specific phosphopeptide sequences. Cell 1993;72:767–78.

[30] Kim HH, Tharayil M, Rudd CE. Growth factor receptor-bound protein 2 SH2/SH3 domain binding to CD28 and its role in co-signaling. J Biol Chem 1998;273:296–301.

[31] Holdorf AD, Green JM, Levin SD, Denny MF, Straus DB, Link V, et al. Proline residues in CD28 and the Src homology (SH)3 domain of Lck are required for T cell costimulation. J Exp Med 1999;190:375–84.

[32] Peach RJ, Bajorath J, Brady W, Leytze G, Greene J, Naemura J, et al. Complementarity determining region 1 (CDR1)- and CDR3-analogous regions in CTLA-4 and CD28 determine the binding to B7-1. J Exp Med 1994;180:2049–58.

[33] Metzler WJ, Bajorath J, Fenderson W, Shaw SY, Constantine KL, Naemura J, et al. Solution structure of human CTLA-4 and delineation of a CD80/CD86 binding site conserved in CD28. Nat Struct Biol 1997;4:527–31.

[34] Jaillon O, Aury JM, Brunet F, Petit JL, Stange-Thomann N, Mauceli E, et al. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. Nature 2004;431:946–57.

[35] Parsons KR, Young JR, Collins BA, Howard CJ. Cattle CTLA-4, CD28 and chicken CD28 bind CD86: MYPPPY is not conserved in cattle CD28. Immunogenetics 1996;43:388–91.

[36] Carreno BM, Collins M. BTLA: a new inhibitory receptor with a B7-like ligand. Trends Immunol 2003;24:524–7.

[37] Du Pasquier L, Zucchetti I, De Santis R. Immunoglobulin superfamily receptors in protochordates: before RAG time. Immunol Rev 2004;198:233–48.

[38] Durand D. Vertebrate evolution: doubling and shuffling with a full deck. Trends Genet 2003;19:2–5.