Immunogenetics (1996) 44: 459–467 © Springer-Verlag 1996

ORIGINAL PAPER

Noriyuki Kuroda · Hironori Wada · Kiyoshi Naruse
Atsuko Simada · Akihiro Shima · Makoto Sasaki
Masaru Nonaka

Molecular cloning and linkage analysis of the Japanese medaka fish complement *Bf/C2* gene

Received: 7 June 1996 / Revised: 10 July 1996

Abstract Evolutionary studies of complement factor B (Bf) and C2 in lower vertebrates have revealed the presence of the Bf/C2 common ancestor-like molecule in lamprey (cyclostome) and the Bf molecule encoded by the duplicated genes closely linked to the major histocompatibility complex (*MHC*) in *Xenopus* (amphibian). To further define when *Bf/C2* gene duplication occurred and when linkage between the *Bf/C2* gene and the *MHC* was established, we amplified the *Bf/C2* sequences in teleost, the Japanese medaka (*Oryzias latipes*), by reverse transcription – polymerase chain reaction with primers corresponding to the common amino acid sequences shared by mammalian Bf and C2. Only a single molecular species has been amplified, and the corresponding cDNA clones were isolated from the liver cDNA library. The longest insert contained 2384 nucleotides with an open reading frame of 754 residues. The deduced amino acid sequence showed 33.6% and 34.1% overall identity with the human Bf and C2 sequences, respectively, hence this clone was named medaka *Bf/C2*. The single-copy medaka *Bf/C2* gene had exactly the same exon-intron organization as the mammalian *Bf* and *C2* genes, and spanned about 8 kilobases. The *Bf/C2* locus was mapped to the close proximity (2.9 cM) of the superoxide dismutase locus on the linkage group XX by the use of a restriction site polymorphism between two inbred strains of the medaka.

The nucleotide sequence data reported in this paper have been submitted to the GenBank, DDBJ, EMBL, and NCBI nucleotide sequence databases and have been assigned the accession number D84063

N. Kuroda · M. Sasaki · M. Nonaka (✉)
Department of Biochemistry, Nagoya City University Medical School, Mizuho-Ku, Nagoya 467, Japan

H. Wada · K. Naruse · A. Simada · A. Shima
Laboratory of Radiation Biology, School of Science, University of Tokyo, Bunkyo-ku, Tokyo 113, Japan

Introduction

Accumulating evidence indicates the presence of the complement system in lower vertebrate species (Dodds and Day 1993). Although cyclostomes, the most primitive extant vertebrate, seem to have a primitive complement system consisting only of an alternative pathway and playing the role of opsonin (Nonaka et al. 1984), a multi-component system which leads to target cell lysis has been identified in cartilaginous fish (Jensen et al. 1981) and higher vertebrates (Dodds and Day 1993). Functional analysis in a bony fish, rainbow trout, indicated the presence of both the alternative and classical activation pathways as well as the cell lytic pathway (Nonaka et al. 1981). In addition, cDNA clones with definitive features of *C3* (Lambris et al. 1993) and *C9* (Stanley and Herz 1987) have been isolated from rainbow trout. These results, together with the recent finding of the *C3* gene in an invertebrate sea urchin (Smith et al. 1996), support the hypothesis that the complement system had been established by the time of vertebrate emergence as a simple system similar to the mammalian alternative pathway. By the time of the emergence of cartilaginous fish, the classical and lytic pathways seem to have been acquired. It is generally believed that the classical pathway was generated from the alternative pathway through two crucial gene duplication events between *Bf* and *C2*, and *C3* and *C4*. However, the molecular evidence confirming these gene duplication events has been identified only in mammalian species so far, except for an amphibian, *Xenopus*, from which both *C3* (Lambris et al. 1995) and *C4* (Mo et al. 1996) cDNA clones have been isolated. Non-mammalian *Bf* or *C2* clones have been isolated from *Xenopus* (Kato et al. 1994, 1995) and lamprey (Nonaka et al. 1994). The *Xenopus* clone was unambiguously identified as *Bf*, whereas the lamprey clone showed almost the same similarity to both mammalian *Bf* and *C2*, and was considered to represent the putative common ancestor molecule of Bf and C2 (Nonaka et al. 1994). Recently, a Bf-like cDNA clone has been isolated from the zebrafish, which also showed almost the same similarity to both mammalian *Bf* and *C2* (Seeger et al.

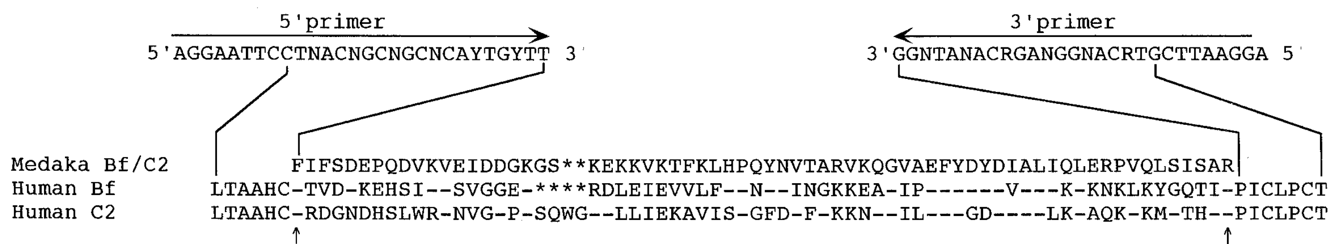

Fig. 1 RT-PCR amplification of the Bf/C2-like serine protease domain from medaka. RT-PCR primers were prepared following the amino acid sequences in the serine protease domain where human Bf and C2 show complete amino acid identity. Only single Bf/C2-like sequence was amplified and tentatively named medaka Bf/C2. Any residues in the human sequences identical with medaka Bf/C2 are indicated by *dashes*. Gaps introduced to increase similarity are shown by *asterisks*. *Arrows* indicate the Bf/C2 diagnostic residues considered to be specific to Bf/C2

1996). One interesting point about the two gene duplication events between *Bf* and *C2*, and *C3* and *C4*, is that the genes encoding three of them, *Bf*, *C2*, and *C4*, reside in the class III region of the mammalian major histocompatibility complex (*MHC*; Carroll et al. 1984; Chaplin et al. 1983). The possible physiological meaning of the linkage between these complement genes and the class I or II genes is still to be demonstrated.

Phylogenetic studies of the genes which are involved in antigen presentation and are encoded in the mammalian *MHC* indicated that most of them appeared at the cartilaginous fish stage during vertebrate evolution. Thus the class I *A* (Hashimoto et al. 1992), class II *A* (Kasahara et al. 1992), and class II *B* (Bartl and Weissman 1994) genes and the *LMP7* (Kandil et al. 1996) gene have been identified in cartilaginous fish, whereas trials to isolate these genes from cyclostomes have not yet succeeded. The same is true for some non-*MHC* genes which play essential roles in the immune system, such as immunoglobulin (Hinds and Litman 1986), T-cell antigen receptor (Rast and Litman 1994), and recombination activating genes (*RAG*; Greenhalgh and Steiner 1995), suggesting that the major components of the mammalian immune system emerged simultaneously in the main line of vertebrate evolution after the divergence of cyclostomes before the divergence of cartilaginous fish. Structural analysis of the *Xenopus MHC* has revealed a close linkage between an *LMP7* (Namikawa et al. 1995), a class I *A* (Flajnik et al. 1991; Shum et al. 1993), three class II *B* (Sato et al. 1993), two *Bf* (Kato et al. 1994; Kato et al. 1995), a *C4* (Mo et al. 1996), and a few *HSP70* genes (Salter-Cid et al. 1994), indicating that the basic structure of the *MHC* had been established by the time of amphibian divergence. However, it is still not clear whether the corresponding genes in bony and cartilaginous fishes form a similar cluster.

Medaka fish (*Oryzias latipes*) is one of the ideal lower vertebrates for molecular genetic studies, since 1) generation time is relatively short; 2) several inbred strains are available; 3) genome size is about one-third that of the human genome; 4) germ-cell mutagenesis has been well studied (Shima and Shimada 1991; Kubota et al. 1995); and 5) a genetic linkage map is present (Wada et al. 1995). To analyze the evolution of the complement system in lower vertebrates and to establish the starting point of the molecular analysis of the possible teleost *MHC*, we isolated medaka cDNA clones corresponding to those of the mammalian *Bf* or *C2*, and performed linkage analysis of the gene.

Materials and methods

Materials

Restriction enzymes were purchased from Toyobo (Osaka, Japan) and New England Biolabs (Beverly, MA). The ligation kit was from Takara (Kyoto, Japan). [α-^{32}P]dCTP, Rediprime random primer labeling kit and cDNA Synthesis Systems Plus were from Amersham Japan (Tokyo, Japan). λ Zap II, λ Dash II, and Gigapack Gold were from Stratagene (La Jolla, CA), DNA Sequencing System 373 A Analysis Software version 1.01 and primer cycle sequencing kit were from Applied Biosystems Japan (Tokyo, Japan). *Eco* RI adapter and Riboprobe Gemini System were from Promega (Madison, WI). Two inbred strains of medaka fish, AA2 and HNI, were bred and maintained at the Laboratory of Radiation Biology, Department of Biological Sciences, School of Science, University of Tokyo.

RNA extraction and cDNA library construction

RNA was isolated from livers of 30 AA2 fish using TRIzol reagent (Gibco-BRL, Tokyo, Japan), and poly(A)$^{+}$ RNA was selected by an oligo-dT cellulose column (Aviv and Leder 1972). Construction of a medaka liver cDNA library was performed as described previously (Nonaka and Takahashi 1992), and approximately 7×10^5 independent plaques of this library were screened. For northern blotting analysis, total RNA were isolated from liver, gill, ovary, and intestine of an HNI female using TRIzol reagent.

Reverse transcriptase-polymerase chain reaction (RT-PCR) amplification of an mRNA segment of medaka Bf/C2

Degenerate PCR primers were the same as those used for the amplification of lamprey (Nonaka et al. 1994) and *Xenopus Bf* (Kato et al. 1994; Fig. 1). PCR template was double-stranded cDNA synthesized from medaka liver mRNA using the cDNA Synthesis Systems Plus. Thirty cycles of amplification were carried out in an Astec PC700 thermocycler (Fukuoka, Japan) using the following parameters: 94 °C for 0.5 min, 50 °C for 1 min, and 72 °C for 1 min (Saiki et al. 1988). PCR products of expected size [250 base pairs (bp)] were gel-purified, digested with *Eco* RI, and ligated into the *Eco* RI site of the pGEM-3Zf(+) vector.

```
                              ┌→ SCR 1                                  47
Me Bf/C2       MKTSVIWSLLVFLLLCTGEV***********ECQCTADGLEIQGGTYTLTKGLETNSL
Hu C2            MGPLMV-FCL-F-YP-LAD**********SAPSCPQNVN-S---F--SH-WAPG--
Mo C2            MAPLLA-FYL-Q-GP-LA************ALFCNQNVN-T--NF--SH-WAPG--
Hu Bf         MGSNL-PQLC-MP-I-GLLSGGVTTTPWSLARPQGS-SLE-V--K--SFR-LQEG***QA
Mo Bf            ME-PQLC-VLLV-GFSSGGVSATPVLEARPQVS-SLE-V--K--SFQ-LQ-G***QA
Xe Bf                  M--LL-ITHVSVS**********LAV--DLTKVA-I--S--VSD-GNVGGK
Da Bf       MTSMECGLRLK-LI-ALICPL-AGAPS********R-GS-PEEN-D-A--SF--SN-YSDG-Y
La Bf           MPRARLTVVV-SIVVWASHQQLCD*******ARL---KQ-VS-L--NTSMPDEPAVG-V

                                                  ┌→ SCR 2          98
Me Bf/C2    LVYQCPDNFYPYPDTIRVCHPNKQWKPRPK*********KFSPQRCKPVECPDPNVLENG
Hu C2       -T-S--QGL--S-AS*-L-KSSG--QTPG*******ATRSL-KAV----R--A-VSF---
Mo C2       -I-S--LGR--S-AW*-K-QS-G--LTPRSSSHHTLRSSRMVKAV----R-LA-SSF---
Hu Bf       -E-V--SG-----VQT-T-RSTGS-STLKTQ*****DQKTVRKAE-RAIH--R-HDF---
Mo Bf       -E-L--SG-----VQT-T-RSTGS-SDLQTR*****DQKIVQKAE-RAIR--R-QDF---
Xe Bf       VE----KGK----KYT-E-QY-GF-TDQKA************KTI--D-R--R-VTF-D-
Da Bf       -Q-I----H--SISS*-R-QF*GV-T-KAS*********SRKKAE--KIT--N-R-----
La Bf       -K-R--YAMR-F-VHT---QK-GD-S-LVN*****AYNQ-ARRAS-R-MT-VG-LEF---

                                                   ┌→ SCR 3         158
Me Bf/C2    NVFPPLERYLAGNTTTFECYSGYTMRGSSSRTCLSNGKWSGSTTICSRDSGDACPDPGVP
Hu C2       IYT-R-GS-PV-GNVS---ED-FIL---PV-Q-RP--M-D-E-AV-DNGA-*H--N--IS
Mo C2       IY--R-VS-PV-SNVS---DEDF-L---PV-Y-RP--L-D-E-AV-DNGAS*H--N--IS
Hu Bf       EYW-RSPY-NVSDEIS-H--D---L---AN---QV--R---Q-A--DNGA-*Y-SN--I-
Mo Bf       EFW-RSPF-NLSDQIS-Q--D--VL---AN---QE--R-D-Q-A--DDGA-*Y--N--I-
Xe Bf       DYE-RQPF-KV-D-LY------F--K-PQN---QE-A--T-E----DDNN-*Y--N--I-
Da Bf       E-A-YQ---YINDV--YS-S-D-KF---KV-V-QP----N---P--G---D*H-------
La Bf       EYY-RKHP-NV-D-V--S----FLFY--GT---QA------TVPA-DDE-S*F-RN----

                                                         ┌→Unidentified217
Me Bf/C2    PGASRRGNTFDVGYDVTYSCNG*NLFLVGSRVRVCQENSQWTGREPACYSKFTYDTSQEV
Hu C2       L--V-T-FR-GH-DK-R-R-SS*--V-T--SE-E--G-GV-S-T--I-RQPYS--FPED-
Mo C2       V-TA-T-LN--L-DK-RYR-SSS-MV-T--AE-E--S-GV-S-S--I-RQPYS--FPED-
Hu Bf       I-TRKV-SQYRLEDS---H-SR*G-T-R--QR-T---GGS-S-T--S-QDS-M---P---
Mo Bf       I-TRKV-SQYRLEDI---H-SR*G-V-R--QK-K---GGS-S-T--S-QDSFM--SP---
Xe Bf       I---KS-SSYKMENK-S-N-QQ*G-VMF--KE-E-L-DKS-S-T--S-RQWY----PK--
Da Bf       --S--T-SI-NIDDE---H-DS*P-T-I--K--SVWMYG--S-T--Q--AD----PAM-A
La Bf       F-GRKM-VD--IEGV-SFT-SP*G-VMS-DTR-T-LSTGE---K-SD-EDIYS--NPED-

                                  ┌→von Willebrand                  273
Me Bf/C2    SEAFGSSIKDSLTTLQPTN****DTQAGRKIRISKNGTLNIYIALDISESVEEEHFKRAK
Hu C2       AP-L-T-FSHM-GATN--Q**KTKESL----Q-QRS-H--L-LL--C-Q--S-ND-LIF-
Mo C2       AS-LDT-LTNL-GATN--QN*LLTKSL----I-QRS-H--L-LL--A-Q--T-KD-DIF-
Hu Bf       AE--L--LTETIEGVDAEDGHGPGE-QK---VLDPS-SM---LV--G-D-IGASN-TG--
Mo Bf       A---L--LTETIEGADAEDGHSPGE-QK---VLDPS-SM---LV--G-D-IGSSN-TG--
Xe Bf       AKT-S--MLENVD-TNLED******RSD-SV--L-D-LM--F-V--T-K--GQNR-DE--
Da Bf       A----N-LTTT--VQ-GFE****-D-H-K--SLDRG-K-D----V-A-D-IDPKD-DK--
La Bf       -F-LS*KVTL-MGVE-SES*******MA-S-NLTSLYDTH--LVI-A-Y--GK-D-DTGL

                                                                      333
Me Bf/C2    LAIITLIKKIAAFTVSPNYEILFFSADVYEVVSIVEF**YEGKITLESAIKNLEDFQIGDKS
Hu C2       ESASLMVDR-FS-EINVSVA-IT-ASEPKVLM-VL****NDNSRDMTEV-SS--NANYK-HE
Mo C2       KSAELMVER-FS-E-NVTVA-IT-ASQPKTIM-IL****S-RSQDVTEV-TS-DSASYK-HE
Hu Bf       KCLVN--E-V-SYG-K-R-GLVTYATYPKIW-KVS****-ADSSNADWVT-Q-NEINYE-HK
Mo Bf       RCLTN--E-V-SYG-R-R-GL-TYATVPKVL-RVS****D-RSSDADWVTEK-NQISYE-HK
Xe Bf       S-S-LF-E-MSNYDIK-R-C-ISYASKAIS---LR****DPDSNNADAVMEH--E--YDRHE
Da Bf       KI-K---E--SYYE--------M-AT--DQI-KMRD-KTN-KARKILKIFED-DN-NYDK-G
La Bf       NFVKD--NR-GMYVRNIR-S-VMYATNPSLKL-VR****DSWSNDPNAV--I-D-LDYYEFD
```

DNA extraction and genomic library construction

High relative molecular mass (M_r) DNA was isolated from eight HNI fish. After anesthesis by immersion in iced water, decapitated and skinned fish were frozen in liquid nitrogen, homogenized, and suspended in Tris-buffered saline. After phenol extraction and RNase treatment, DNA was precipitated with isopropanol. High M_r DNA was digested with *Sau* 3AI incompletely, 15 to 20 kilobase (kb) fragments were purified by agarose gel electrophoresis, and ligated with the *Bam*HI-digested λ Dash II arms.

Nucleotide sequence analysis

DNA sequence analysis was performed by the dideoxy chain termination method (Sanger et al. 1977) using an Applied Biosystems 373A DNA sequencer. Sequencing primers to extend sequence readings were synthesized using an Applied Biosystems 381A DNA synthesizer. Each sequence was determined at least twice from both strands.

Northern blotting analysis

Total RNA from various medaka tissues were denatured by glyoxal, separated on a 1% agarose gel, and blotted to a nylon membrane (Hybond-N, Amersham; Thomas 1980). After pre-hybridization for 4 h at 65 °C in a solution containing 50% formamide, 50 mM sodium phosphate buffer (pH 6.5), 0.8 M sodium chloride, 1 mM ethylenediaminetetraacetic acid (EDTA), 10× Denhardt's solution (1× Denhardt's solution is 0.02% bovine serum albumin, 0.02% ficoll, and 0.02% polyvinylpyrrolidone), 0.25 mg/ml denatured salmon sperm DNA, and 2 mg/ml yeast tRNA, membranes were then hybridized with radiolabeled RNA probes (Promega, Madison, WI) at 55 °C for

```
                                                                   386
Me Bf/C2    **TGTDVNAALKKFEEGMAWIEQKTG*****DKFSEHRHVFLLFTDGAYNMGGSPLPTLA
Hu C2       NG---NTY---NSVYLM-NNQMRLL-MET**MAWQ-I--AII-L---KS------KTAVD
Mo C2       NA--ANTYEV-IRVYSM-QTQMDRL-MET**SAWK-I--TII-L---KS---D--KKAVT
Hu Bf       LKS--NTKK--QAVYSM-S-PDDVPP*EG****WNRT---II-M---LH----D-ITVID
Mo Bf       LKS--NTKR--QAVYSM-S-AGDAPP*EG****WNRT---IIIM---LH----N-VTVIQ
Xe Bf       DKQ--NTR---HAIY-HLIEQ-LAYEREGKKED-MKIHN-I--M---KF----D-REEMK
Da Bf       DR---NIAKLYL-ILDS-SLEQVQNK*ED****-LQTQ--IIV----QA----N-K-KVD
La Bf       DTP--NTAM---MVLDT--LYKVANQN*****T-KDI-QAII-L---RS-V-PP-G***K

                                                                   445
Me Bf/C2    RIKNRVYMSPTGDPGSRLDYLESYVFGIG*ANIFDDDLLPLTAGTEGELHYFRLKKETNL
Hu C2       H-REILNINQK*****-N---DI-AI-V-KLDVDWRE-NE-GSKKD--R-A-I-QDTKA-
Mo C2       --RELLSIEQN*****-D---DI-AI-V-KLDVDWKE-NE-GSKKD--R-A-I-QDAKA-
Hu Bf       E-RDLL-IGKDRK*NP-E---DV----V-*PLVNQVNINA-ASKKDN-Q-V-KV-DME--
Mo Bf       D-RALLDIGRDPK*NP-E---DV----V-*PLVDSVNINA-ASKKDN-H-V-KV-DMED-
Xe Bf       L--RFLDVGIRK-*NP-EE--DV----L-*SD-DQPEIND-ASKK-K-V-T-H-QNVEKM
Da Bf       L---L-IKNNAS****-ENK-DL----V-*KDVKKE-MNG-VSEKKD-R-F-K-PDLDEV
La Bf       FLMDNIDLDIP******KEHMDV----M-**DVYK-EIETIASQKPN-Q-S-I-RDYDD-

                   ┌→ Serine Protease                              493
Me Bf/C2    AATFDDIIDENEVIG*LCGLHRDYELTADKDGK*****RRRYPWVVFINIQGK******S
Hu C2       HQV-EHML-VSKLTDTI--VGNMSANAS-*********QE-T--H-T-KPKSQ*******
Mo C2       QQI-EHML-VSKLTDTI--VGNMSANAS-*********QE-T--Q-TFKPKS-*******
Hu Bf       EDV-YQM---SQSLS*---MVWEHRKGT-*********YHKQ--QAK-SVIRPSK****G
Mo Bf       ENV-YQM---TKSLS*---MVWEHKKGN-*********YHKQ--QAK-SVTRPLK****G
Xe Bf       KEF-ELM-I-DD-FD*T---SKYHSVEL-PKK-A****TVMF--I-K-T-THN******G
Da Bf       QN---LML-DST-V-*---MQQN-DGSNK*********-SA---LAQLS-AQS******Q
La Bf       NEV-EKMLHAD-KLFTQ--TSGTFRIPRARIAGGDPTKIELW--QAQ-SMRVHISNDHVK

                            ◆                                     545
Me Bf/C2    PKKCLGSLVSSEFVLTAAHCFIFS**DEPQDVKVEIDDGKGS******KEKKVKTFKLHP
Hu C2       *ET-R-A-I-DQW--------RDG**NDHSLWR-NVG-P-SQW****G--LLIEKAVIS-
Mo C2       *ET-Q---I-DQW--------HDIQMEDHHLWR-NVG-PTSQH****G--FL-EDVIIA-
Hu Bf       HES-M-AV--EY---------TVD**-KEHSI--SVGGE-********RDLEIEVVLF--
Mo Bf       HET-M-AV--EY---------MVD**-QKHSI--SVGGQR********RDLEIEEVLF--
Xe Bf       IQY-K-TIL-QY-I-------DLD**-KT-KIH-K--G**********--YL--D-YR--
Da Bf       ISD-M----T-RYI-------KEG**-T-DKIT-YLEKN*********TDV--EKVFI--
La Bf       -AF-G--IIAEQWI-------DEFAITDDEWWRGS--VVI--SNKLGGDKISP-QIII-E
                              ↑
                            ◆                                     600
Me Bf/C2    QYNVTARVKQGVAEFYDYDIALIQLER**PVQLSISARPICIPCTKETSDALRLPGS***
Hu C2       GFD-F-KKN--IL---GD----LK-AQ**K-KM-TH-----L---M-ANL---R-QG***
Mo C2       GF--H-KR---IS---AD----LK-S-**K-KM-TH-----L---VGANM---RSPG***
Hu Bf       N--INGKKEA-IP------V---K-KN**KLKYGQTI----L---EG-TR-----PT***
Mo Bf       K--INGKKAE-IP------V--VK-KN**KLKYGQTL----L---EG-TR-----QT***
Xe Bf       K-DPISKKDK-IKRAF---V--LE-Q-NDKIEF-EN---------QG-AQ--KQ--****
Da Bf       N-SL--KQSI-IK----F-V--L--KT**--KM-VNL----L------NR--K-SD-***
La Bf       G--RNPD*AHVQI-NL-N-----K-SK**RLTFGYTY----L------NAI-D-NSANKD
                                               ↑

                                                                   652
Me Bf/C2    **ATCRDQEELLLKNQRERLSFLTRTEPLVG******EKDVYAKLGDNRDLCIKKALKAK
Hu C2       **S----H-NE--NK-SVPAH-VALNGS*********KLNINL-M-VEWTS-AEVVSQE-
Mo C2       **S--K-H-TE--SQ-KVPAH-VALNGN*********RLNINLRT-PEWTR--QAVSQN-
Hu Bf       **T--QQ-K-E--PA-DIKAL-VSEE-KK******LTR-E--I-N--KKGS-ERD-QY-P
Mo Bf       **---KQHK-Q--PVKDVKAL-VSEQGKS******LTR-E--I-N--KKAS-ERD-T--Q
Xe Bf       **-P-SSH-KT--SEEEVKAV-IAEESNK******PMKMH-LI-R-QK-SA-LEA-K--P
Da Bf       *QG--EKH-QI--S-ELVDAA-TSKMDMEKRSP**RKIRRITV---KYL-A-VED-K---
La Bf       WTTL-NIHGKN-IDVKKNTSLTV-G-GL-E-DKKHAQQLQQATVQYAKKEV-L-DIMARF

                                               ◆                   707
Me Bf/C2    G**ITTTDPKVPVTDNFLCTGGDR***DHIACTGDSGGAVFKNYESRTIQIALVSWGTQE
Hu C2       TMFPNL--VREV---Q---S-TQE***-ESP-K-E------LERRF-FF-VG-----LYN
Mo C2       NIFPSL-NVSEV---Q---S-MEEE**-DNP-K-E------LGRRY-FF-VG-----LFD
Hu Bf       *GYDKVK-ISEV--PR------VSPYA-PNT-R-----PLIVHKR--F--VGVI---VVD
Mo Bf       *GYEKVK-ASEV--PR------VDPYA-PNT-K-----PLIVHKR--F--VGVI---VVD
Xe Bf       *ELKNV-NIEDA---Q------IVPVA-PPV-K-----PLLIQVKR-YV-VGII----VD
Da Bf       **ESKWQMRRRQLQKISCGS--NQPQR-DVS-K-E----THVDKYG-L---GV----VKN
La Bf       N**V-EEKAEKHI-E-M--AWNAT****ADT-R-----PLVLQKNR-W--VGI-AG-VAQ

                                                               754
Me Bf/C2    ICTG***GGMRETTPES*****RDFHINLFKMVPFLKSILGDDDQDDYAPLTFIN
Hu C2       P-L-SADKNS-KRA-R-KVPPP--------R-Q-W-RQH---*******V-N-LPL
Mo C2       P-H-SSNKNL-KKP-RG**VLP-----S--RLQ-W-RQH-DG*******V-D-LPL
Hu Bf       V-K***NQKRQKQV-AH****A--------QVL-W--EK-Q-**E****D-G-L
Mo Bf       V-R***DQRRQQLV-SY****A--------QVL-W--DK-K-**E****D-G-L
Xe Bf       H-D***K-TRIKQ-QKN****A---YQDI--VL-WIQKV-EEKHE****V---LPN
Da Bf       L-S***KKRNLMQFSV-***DS--Y---P-
La Bf       H-G********KNIKS-*******-YT-VA--M-WV-RQIP-********-N-GDV
```

Fig. 2 Alignment of medaka Bf/C2 (Me Bf/C2) amino acid sequence with the human C2 (Hu C2), mouse C2 (Mo C2), human Bf (Hu Bf), mouse Bf (Mo Bf), *Xenopus* Bf (Xe Bf), zebrafish Bf (Da Bf), and lamprey Bf (La Bf) sequences. The entire amino acid residues of medaka Bf/C2 are presented. *Numbers* on the *right* indicate amino acid numbers of medaka Bf/C2. Residues identical to medaka Bf/C2 in other sequences are indicated by *dashes*. Insertions and deletions introduced to increase identity are indicated by *asterisks*. N-terminal ends of each domain are indicated by L-shaped *arrows*. The residues completely conserved are indicated by *dots* below the sequences. *Upward arrows* indicate diagnostic residues for Bf/C2 identification of PCR products. Three diamonds (◆) indicate the conserved H, D, and S residues at the active center of serine proteases

```
-21                   +1
GCTCCGTCTTTGAGGTTAAGCATGAAAACTTCTGTCATCTGGAGTTTGCTGGTCTTTCTT    39
                     M  K  T  S  V  I  W  S  L  L  V  F  L      13
      ↓intron 1
CTTCTCTGCACGGGAGAGGTTGAGTGTCAGTGCACAGCAGATGGCCTGGAAATTCAGGGG    99
L  L  C  T  G  E  V  E  C  Q  C  T  A  D  G  L  E  I  Q  G      33

GGTACTTACACCCTAACAAAGGGCCTAGAGACCAACAGCTTGCTGGTCTATCAGTGCCCT   159
G  T  Y  T  L  T  K  G  L  E  T  N  S  L  L  V  Y  Q  C  P      53

GATAACTTCTACCCGTACCCAGACACAATTCGTGTGTGTCACCCCAACAAACAATGGAAA   219
D  N  F  Y  P  Y  P  D  T  I  R  V  C  H  P  N  K  Q  W  K      73
                                          ↓intron 2
CCACGACCAAAAAAGTTTTCTCCCCAGAGATGCAAGCCTGTGGAGTGTCCTGACCCTAAC   279
P  R  P  K  K  F  S  P  Q  R  C  K  P  V  E  C  P  D  P  N      93

GTCCTGGAGAACGGAAATGTGTTTCCCCCTTTGGAGAGATACCTTGCTGGCAACACCACC   339
V  L  E  N  G  N  V  F  P  P  L  E  R  Y  L  A  G  N  T  T     113

ACCTTTGAGTGCTACTCTGGTTACACAATGCGCGGCTCTTCAAGTCGTACCTGCTTATCC   399
T  F  E  C  Y  S  G  Y  T  M  R  G  S  S  S  R  T  C  L  S     133
                                                ↓intron 3
AATGGGAAGTGGAGTGGCTCCACCACTATCTGCAGTCGTGATTCCGGGGACGCATGTCCT   459
N  G  K  W  S  G  S  T  T  I  C  S  R  D  S  G  D  A  C  P     153

GATCCTGGTGTCCCACCCGGTGCCTCAAGGAGAGGAAACACATTCGATGTTGGTTATGAT   519
D  P  G  V  P  P  G  A  S  R  R  G  N  T  F  D  V  G  Y  D     173

GTCACGTACTCCTGCAATGGCAACCTGTTTTTAGTGGGGTCCAGAGTCAGAGTGTGTCAG   579
V  T  Y  S  C  N  G  N  L  F  L  V  G  S  R  V  R  V  C  Q     193
                                             ↓intron 4
GAGAACAGCCAATGGACCGGCAGAGAGCCTGCATGCTACTCCAAATTCACATATGACACT   639
E  N  S  Q  W  T  G  R  E  P  A  C  Y  S  K  F  T  Y  D  T     213

TCCCAAGAGGTCTCAGAAGCATTTGGCAGTTCCATCAAAGACAGCCTCACCACCCTGCAG   699
S  Q  E  V  S  E  A  F  G  S  S  I  K  D  S  L  T  T  L  Q     233
    ↓intron 5
CCCACAAATGACACACAAGCTGGGAGAAAAATCCGCATTTCAAAGAACGGGACGCTCAAC   759
P  T  N  D  T  Q  A  G  R  K  I  R  I  S  K  N  G  T  L  N     253

ATCTACATCGCCTTGGACATTTCAGAGAGTGTTGAAGAAGAACACTTCAAAAGAGCAAAG   819
I  Y  I  A  L  D  I  S  E  S  V  E  E  E  H  F  K  R  A  K     273

                             ↓intron 6
CTGGCCATCATTACACTTATTAAAAAGATTGCAGCCTTCACTGTGTCGCCAAACTATGAA   879
L  A  I  I  T  L  I  K  K  I  A  A  F  T  V  S  P  N  Y  E     293

ATCCTCTTTTTCTCCGCTGATGTGTATGAAGTTGTCAGCATTGTTGAATTTTATGAAGGC   939
I  L  F  F  S  A  D  V  Y  E  V  V  S  I  V  E  F  Y  E  G     313
                                                    ↓intron 7
AAAATAACTCTGGAGTCTGCCATCAAAAATTTGGAAGATTTTCAAATAGGTGACAAGAGC   999
K  I  T  L  E  S  A  I  K  N  L  E  D  F  Q  I  G  D  K  S     333

ACTGGAACAGACGTGAACGCTGCTTTAAAGAAATTTGAAGAGGGTATGGCTTGGATAGAG  1059
T  G  T  D  V  N  A  A  L  K  K  F  E  E  G  M  A  W  I  E     353
                                                         ↓intron 8
CAAAAAACTGGAGATAAATTTAGTGAACATCGTCACGTCTTTCTCCTCTTCACGGACGGT  1119
Q  K  T  G  D  K  F  S  E  H  R  H  V  F  L  L  F  T  D  G     373

GCATACAACATGGGTGGCTCCCCATTGCCAACATTAGCCAGAATAAAGAACAGGGTCTAC  1179
A  Y  N  M  G  G  S  P  L  P  T  L  A  R  I  K  N  R  V  Y     393
                                            ↓intron 9
ATGAGTCCCACAGGTGATCCCGGATCAAGATTGGATTATCTTGAGAGCTACGTCTTCGGC  1239
M  S  P  T  G  D  P  G  S  R  L  D  Y  L  E  S  Y  V  F  G     413

ATCGGGGCTAACATCTTTGATGATGACCTGCTGCCCCTCACAGCAGGAACGGAAGGAGAA  1299
I  G  A  N  I  F  D  D  D  L  L  P  L  T  A  G  T  E  G  E     433

CTGCATTACTTCAGACTGAAAAAAGAAACAAATTTAGCAGCCACCTTTGATGATATAATT  1359
L  H  Y  F  R  L  K  K  E  T  N  L  A  A  T  F  D  D  I  I     453
↓intron 10
GATGAAAATGAAGTCATTGGCCTCTGTGGTCTGCACAGAGACTATGAACTCACAGCTGAC  1419
D  E  N  E  V  I  G  L  C  G  L  H  R  D  Y  E  L  T  A  D     473
                                                  ↓intron 11
AAAGACGGCAAAAGGAGAAGATATCCATGGGTGGTGTTTATCAATATCCAGGGCAAATCC  1479
K  D  G  K  R  R  R  Y  P  W  V  V  F  I  N  I  Q  G  K  S     493

CCGAAGAAATGCTTGGGCTCTCTGGTGTCGAGTGAGTTTGTCTTGACCGCTGCCCACTGC  1539
P  K  K  C  L  G  S  L  V  S  S  E  F  V  L  T  A  A  H  C     513

TTCATCTTCAGTGATGAGCCACAAGATGTCAAAGTTGAAATTGACGATGGGAAAGGCAGC  1599
F  I  F  S  D  E  P  Q  D  V  K  V  E  I  D  D  G  K  G  S     533
   ↓intron 12
AAAGAGAAGAAAGTGAAAACATTTAAGCTACATCCACAGTACAATGTTACGGCTCGAGTG  1659
K  E  K  K  V  K  T  F  K  L  H  P  Q  Y  N  V  T  A  R  V     553

AAACAAGGCGTGGCTGAGTTTTATGACTATGATATTGCTCTCATCCAGCTGGAGCGACCT  1719
K  Q  G  V  A  E  F  Y  D  Y  D  I  A  L  I  Q  L  E  R  P     573

                           ↓intron 13
GTTCAACTCTCCATTTCTGCCAGACCCATATGCATACCTTGTACTAAGGAAACCAGTGAT  1779
V  Q  L  S  I  S  A  R  P  I  C  I  P  C  T  K  E  T  S  D     593
                                           ↓intron 14
GCTCTCAGACTCCCCGGGTCAGCAACATGCAGAGATCAAGAGGAGCTTTTATTGAAAAAC  1839
A  L  R  L  P  G  S  A  T  C  R  D  Q  E  E  L  L  L  K  N     613

CAACGAGAAAGACTCAGTTTCTTGACTCGGACAGAACCACTAGTTGGTGAAAAAGATGTC  1899
Q  R  E  R  L  S  F  L  T  R  T  E  P  L  V  G  E  K  D  V     633
                           ↓intron 15
TATGCAAAGCTTGGAGATAATAGAGATCTATGCATCAAAAAGGCATTGAAAGCAAAGGGA  1959
Y  A  K  L  G  D  N  R  D  L  C  I  K  K  A  L  K  A  K  G     653

ATAACAACAACAGACCCAAAGGTTCCTGTGACGGATAACTTTCTGTGCACTGGAGGTGAT  2019
I  T  T  T  D  P  K  V  P  V  T  D  N  F  L  C  T  G  G  D     673
                          ↓intron 16
AGAGACCACATAGCATGCACAGGTGACTCTGGAGGAGCTGTGTTTAAGAACTATGAGAGT  2079
R  D  H  I  A  C  T  G  D  S  G  G  A  V  F  K  N  Y  E  S     693
               ↓intron 17
CGCACAATACAGATTGCTTTGGTAAGCTGGGGAACCCAGGAGATTTGCACTGGAGGTGGT  2139
R  T  I  Q  I  A  L  V  S  W  G  T  Q  E  I  C  T  G  G  G     713

ATGAGAGAGACAACGCCTGAATCCAGAGATTTCCACATCAATCTTTTCAAAATGGTTCCT  2199
M  R  E  T  T  P  E  S  R  D  F  H  I  N  L  F  K  M  V  P     733

TTTCTCAAATCAATCCTTGGAGATGACGATCAGGACGATTATGCACCCCTTACATTTATA  2259
F  L  K  S  I  L  G  D  D  D  Q  D  D  Y  A  P  L  T  F  I     753

AACTAACAAACACATTCAATTAAGAAAGTCTCTTGAATAAGAGGCGACAATGCCTTTGAA  2319
N  *                                                            754

TTAAAAAAAAAGCTAAAGACCAGAAAAAAAAAAAAAAAAAAAAA                  2363
```

Fig. 3 Nucleotide and deduced amino acid sequence of medaka *Bf/C2*. The entire nucleotide sequence of the clone 10 insert is presented together with the deduced amino acid sequence (*italics*). The initiation methionine codon was assigned from alignment with human Bf and C2, and is marked +1. The nucleotide and amino acid number of the *rightmost* residues, starting from the initiation codon, are shown for each lane. Positions of intron insertion revealed by gene analysis (see *below*) are indicated by *downward arrows*

16–20 h. Membranes were washed twice for 30 min at 65 °C in 50 mM sodium chloride, 20 mM sodium phosphate buffer (pH 6.5), 1 mM EDTA, an 0.1% sodium dodecyl sulfate.

Genotyping and linkage analysis

Polymorphic enzyme sites between AA2 and HNI were searched by PCR amplification of a part of the gene using primers synthesized based on the exon sequences in combination with the following enzyme digestion. An *Rsa* I polymorphism was identified in intron 6. Native polyacrylamide gel electrophoresis were used for genotyping of each fish (Davis 1964). *Bf/C2* type was determined using DNA samples of 80 backcross prognenies from a cross, (AA2 female × HNI male)F_1 males with AA2 females. These progenies have been typed for random amplified polymorphic DNA (RAPD) markers as described previously (Wada et al. 1995).

Results

RT-PCR amplification of Bf/C2

RT-PCR amplification of the medaka liver cDNA resulted in a single DNA band of the expected size (about 250 bp). The DNA was gel purified, digested by *Eco*RI, and ligated with the *Eco*RI-digested pGEM-3Zf(+) vector. Nineteen clones were randomly selected and nucleotide sequences of the inserts were determined. All inserts had the same nucleotide sequence which predicted an amino acid se-

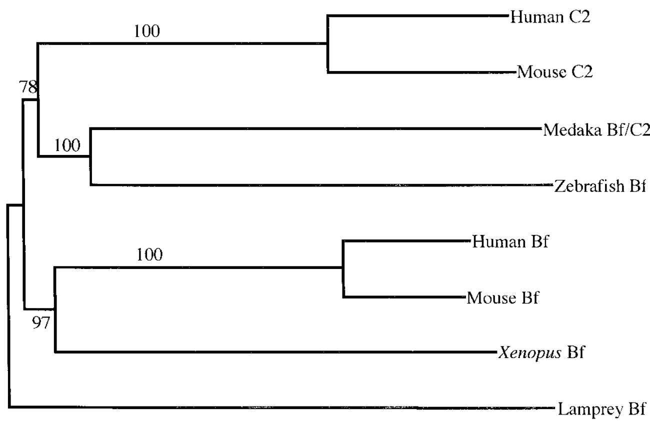

Fig. 4 Bf and C2 phylogenetic tree. The relationship among human Bf, human C2, mouse Bf, mouse C2, *Xenopus* Bf, zebrafish Bf, lamprey Bf, and medaka Bf/C2 was analyzed by the neighbor-joining method for the entire amino acid sequences based on the alignment shown in Figure 3. Numbers on branches are bootstrap percentages supporting a given partitioning

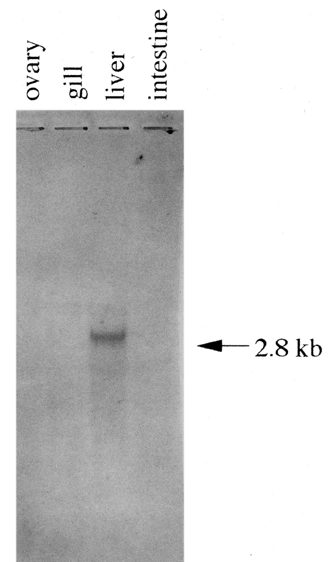

Fig. 5 Northern blotting analysis of medaka *Bf/C2*. Five micrograms of ovary, gill, liver, and intestine total RNA of the AA2 medaka were denatured by glyoxal and separated on a 1% agarose gel. After blotting on nylon membrane, hybridization was performed using the RNA probe, corresponding to the medaka *Bf* cDNA region spanning from the 5′ end to the *Pst*I site (+433) of clone 10. The washed filter was exposed to an X-ray film for 6 days at room temperature. *Lines* at the top of each lane indicate origin

quence showing 31% and 33% identity to the corresponding regions of human Bf (Horiuchi et al. 1993) and human C2 (Bentley 1986), respectively (Fig. 1). Although amino acid identity is not high enough, the presence of the diagnostic Phe and Arg residues (Fig. 1, arrows) which are present in all the Bf and C2 sequences reported so far (see Fig. 2), helped to identify the amplified sequence as *Bf* or *C2*. Therefore, this sequence was tentatively designated medaka *Bf/C2*.

Isolation and sequence analysis of medaka Bf/C2 cDNA

Using the PCR-amplified clone as a probe, we screened the medaka liver cDNA library containing 7×10^5 independent clones. Thirteen clones were isolated, and nucleotide sequence analysis of both termini of each insert indicated that three of them contained the entire protein coding sequence. The insert of the longest clone (clone 10) was 2384 bp long, and predicted a single long open reading frame of 754 amino acids (Fig. 3).

Multiple alignment and phylogenetic tree of Bf and C2 sequences from various species

The deduced amino acid sequence of medaka Bf/C2 was aligned with the amino acid sequences of human C2 (Bentley 1986), mouse C2 (Ishikawa et al. 1990), human Bf (Horiuchi et al. 1993), mouse Bf (Ishikawa et al. 1990),

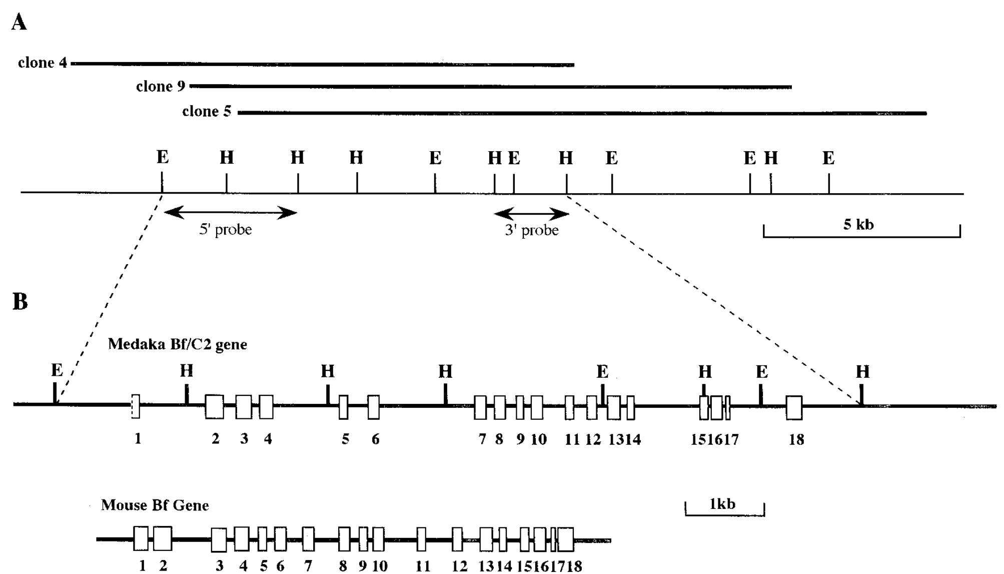

Fig. 6A, B Structure of the medaka *Bf/C2* gene. **A** Restriction enzyme map of the three overlapping medaka Bf/C2 phage clones. Three genomic clones (clones 4, 5, 9) were isolated from a genomic library. Restriction enzyme sites of *Eco*RI (E), *Hin*dIII (H) are shown. The fragments which hybridized with the probes representing the 5′ and 3′ ends of clone 10 are shown by *double-headed arrows*. **B** Genomic structure of the medaka *Bf/C2* gene and the mouse *Bf* gene as reference. *Open rectangles* indicate the size and location of exons. Exon-intron boundaries of the medaka *Bf/C2* gene were determined by nucleotide sequence analysis. The transcriptional start site of the medaka *Bf/C2* gene is yet to be determined, and hence the 5′ end of exon 1 is indicated by a *dotted line*

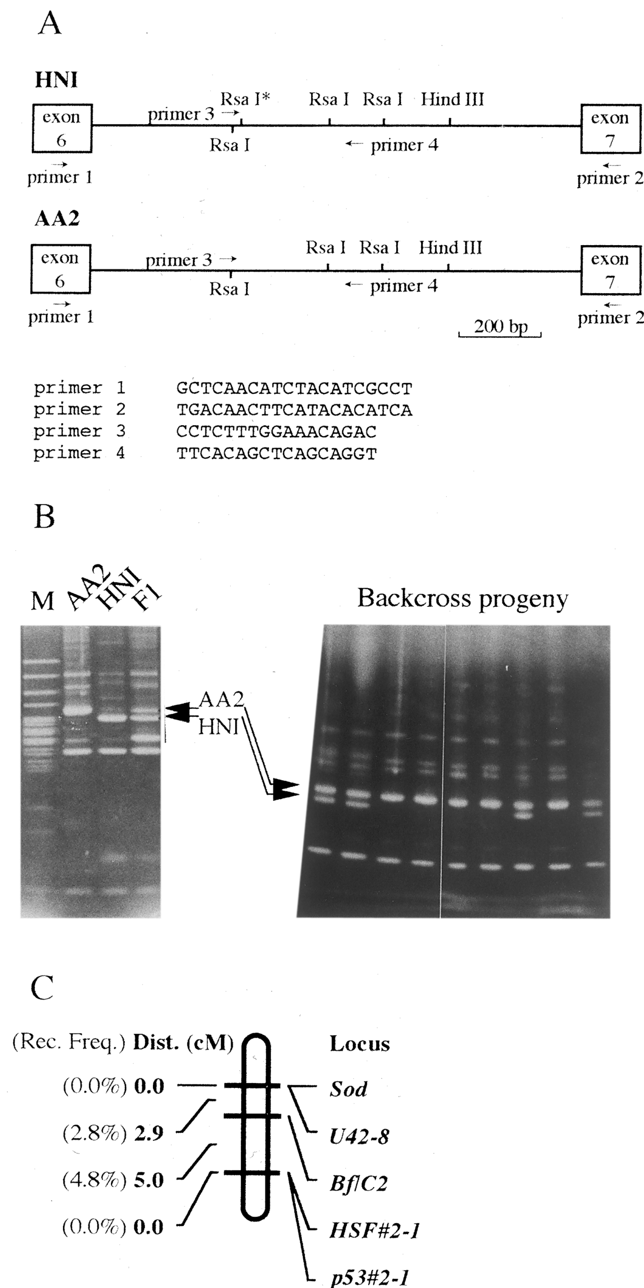

Fig. 7 A–C Linkage analysis of the medaka *Bf/C2* gene. **A** Mapping strategy. PCR primers were designed to detect the *Rsa* I site polymorphism between the HNI and AA2 strains in intron 6. The positions of PCR primers are indicated by *arrows*. The first PCR was performed using primers 1 and 2, and the second PCR with primers 3 and 4. The polymorphic *Rsa* I site is indicated by an *asterisk*(*). **B** Genotyping of backcross progeny. The *left panel* shows the bands polymorphic between AA2 and HNI (*arrows*). F_1 DNA shows both bands. The *right panel* shows a typical typing pattern of nine offspring of a crossing of AA2 females × (AA2×HNI)F_1 males. Homozygotes show only an AA2 type band. Heterozygotes show both AA2 and HNI type bands. In total 80 offspring were typed. M indicates relative molecular mass marker (pBR322 DNA *Msp* I digest marker). **C** Linkage relationships of the *Bf/C2* locus and other loci in the linkage group XX. The MAPMARKER Macintosh V.2.0 program was used for linkage analysis. Lod score of 3.0 and theta value of 4.0 were used as the default setting. Kosambi's mapping function was used to calculate map distances (Dist.). The *U42-8*, *HSF#2-1*, and *p53#2-1* were loci identified as RAPD markers. Superoxide dismutase locus was shown as *Sod*

Xenopus Bf (Kato et al. 1994), zebrafish Bf (Seeger et al. 1996), and lamprey Bf (Nonaka et al. 1994) using the Clustal W software (Thompson et al. 1994). As shown in Figure 2, the medaka Bf/C2 sequence showed a significant similarity to other Bf and C2 sequences throughout its entire length. This result indicated that medaka Bf/C2 has the same basic domain structure as Bf and C2 of other species: that is, three SCR domains, a von Willebrand domain, and a serine protease domain from the N-terminus. The calculated amino acid identities based on this alignment of medaka Bf/C2 were: 34.1% with human C2, 32.6% with mouse C2, 33.6% to human Bf, 32.5% with mouse Bf, 34.3% with *Xenopus* Bf, 37.1% with zebrafish Bf, and 27.4% with lamprey Bf. The mammalian and *Xenopus* Bf and C2 sequences showed almost the same degree of identity to the medaka sequence, making it difficult to assign Bf or C2 to the medaka sequence. The zebrafish sequence showed a slightly higher degree of identity, whereas the lamprey sequence, considered to represent the state before *Bf/C2* gene duplication (Nonaka et al. 1994), showed a lower degree of identity to the medaka Bf sequence than the mammalian and *Xenopus* sequences. The phylogenetic tree of the Bf and C2 sequences was drawn using the neighbor-joining method (Saitou and Nei 1987; Fig. 4). Medaka Bf/C2 showed clustering with zebrafish Bf, and these teleost sequences showed clustering with human and mouse C2, suggesting the possibility that the teleost sequences are actually C2. However, the bootstrap percentage supporting this branching pattern (78%) is not high enough, and more information on other lower vertebrates is required before making a conclusive assignment.

Northern blotting analysis

To analyze the tissue distribution of the medaka *Bf/C2* messages, northern blotting analyses were performed. As shown in Figure 5, when ovary, gill, liver, and intestine were analyzed for *Bf/C2* mRNA, significant hybridization was detected only with liver. The estimated size of the liver hybridizing band was 2.8 kb, indicating that an approximately 400 to 500 bp region of the 5′ end of medaka *Bf/C2* is missing from clone 10.

Isolation and structural analysis of the medaka Bf/C2 gene

A medaka genomic DNA library with 7.5×10^5 independent clones was constructed using the λ Dash II arms. Screening of this library using the medaka *Bf/C2* spanning from the 5′ end to the *Pst* I site (+433) (5′ probe) and the *Hin* dIII site (+1906) to the 3′ end of clone 10 (3′ probe) as probe resulted in isolation of three clones. Restriction mapping analysis of these clones indicated that they overlap each other (Fig. 6 A). Southern blotting analysis of these clones using the 5′ and 3′ probes indicated that the medaka *Bf/C2* gene spans about 8 kb. The exact positions of the exon-intron boundaries were determined by nucleotide sequence

analysis after subcloning DNA fragments from phage clone 4 into pGEM-3Zf(+) vector. As shown in Figure 6B, exon-intron organization is completely conserved between the mouse *Bf* or *C2* (Ishikawa et al. 1990) and medaka *Bf/C2* genes.

Linkage analysis of the medaka Bf/C2 gene

The genetic linkage map of medaka has been constructed recently, mainly using random amplified polymorphic DNA (RAPD) markers, and 28 linkage groups were identified (Wada et al. 1995). As the first step to explore the possible presence of the mammalian *MHC*-like gene organization in medaka, we performed linkage analysis of the medaka *Bf/C2* gene. Search for the restriction enzyme site polymorphism between the AA2 and HNI strains by PCR-amplification of the intron sequences resulted in identification of the *Rsa* I site polymorphism in intron 6. PCR primers were designed as shown (Fig. 7A) for nested amplification of the DNA segment encompassing this polymorphic *Rsa* I site. Genomic DNA was amplified by PCR, digested with *Rsa* I, and typed by polyacrylamide gel electrophoresis (Fig. 7B). Eighty backcross progenies were typed, and the result shown in Figure 7C indicated that the medaka *Bf/C2* locus is closely linked to the *Sod* locus (2.9 cM) on the linkage group XX.

Discussion

The medaka cDNA clone corresponding to mammalian Bf or C2 was isolated. The primary structure of the medaka molecule deduced from the nucleotide sequence showed the same domain structure as Bf and C2 from other species, clearly identifying the medaka sequence as Bf or C2. The discrimination between Bf and C2, however, was not obvious, since the deduced medaka amino acid sequence showed almost the same similarity to the Bf and C2 sequences of other species. A similar situation has been reported recently with the Bf cDNA clone isolated from another teleost species, zebrafish (Seeger et al. 1996). The phylogenetic tree constructed using the neighbor-joining method (Saitou and Nei 1987) indicated that the medaka Bf/C2 and zebrafish Bf sequences form a cluster with human and mouse C2, although a bootstrap value to support this branching pattern was not high enough to draw a final conclusion. In contrast, from the number of charged amino acid residues in the exon 15-encoded region, which differs greatly and was suggested to be involved in the functional difference between human Bf and C2 (Krumdieck and Volanakis 1995), the medaka Bf/C2 and zebrafish Bf sequences show more similarity to Bf than to C2. Thus the medaka and zebrafish sequences have 12 and 13 charged rsidues in this region, respectively, whereas numbers of charged residues in this region of human C2, mouse C2, human Bf, mouse Bf, *Xenopus* Bf, and lamprey Bf are 6, 6, 15, 12, 13, and 10, respectively. Due to the difficulty of assigning the medaka sequence to Bf or C2, we named this sequence medaka Bf/C2. The presence of an intermediate molecule between Bf and C2 in two teleost species may imply that teleost represent the state before the *Bf/C2* gene duplication. However, functional studies performed in another teleost species, rainbow trout, clearly demonstrated the presence of both the classical and alternative pathways (Nonaka et al. 1981), suggesting that *Bf/C2* gene duplication predated the emergence of teleost. We tried to identify another *Bf/C2*-like sequence from RT-PCR-amplified medaka clones, but all 19 clones we sequenced had the same insert as medaka *Bf/C2*. Since amino acid sequences corresponding to the PCR primers are perfectly conserved among the Bf and C2 sequence so far determined, except for one position where a Leu to Ile substitution was recognized with medaka Bf/C2 and *Xenopus* Bf (Kato et al. 1994; amino acid residue number 585 in Figure 3), it is unlikely that additional medaka *Bf* or *C2* sequence, if any, can not be amplified by RT-PCR using these primers. A similar attempt to find another gene in zebrafish related to zebrafish Bf was also unsuccessful (Seeger et al. 1996). Thus, there is an intriguing possibility that teleost has a single Bf/C2 molecule which can work in both the classical and alternative pathways. If this is so, the differentiation between Bf and C2 is located in the evolution history of vertebrates after divergence of teleosts, but before divergence of amphibians, since the Bf/C2-like cDNA clones isolated from *Xenopus* were clearly identified as Bf and not C2 (Kato et al. 1994, 1995). In contrast, C3/C4 differentiation seems to have occurred before the divergence of teleosts, because trout has a molecule clearly identified as C3 and not C4 (Lambris et al. 1993). Thus the gene duplications and differentiations between *Bf* and *C2*, and *C3* and *C4* could have occurred at different stages of vertebrate evolution.

The medaka *Bf/C2* gene has exactly the same intron insertion positions as those found in the mammalian *Bf* and *C2* genes, indicating that the basic structure of the *Bf/C2* genes has been conserved for more than 400 million years since teleost and mammals last shared a common ancestor. The size of the medaka *Bf/C2* gene, 8 kb, is also not so different from that of mouse *Bf* (6 kb) and *C2* (20 kb; Ishikawa et al. 1990). Although the medaka genome size is only about one-third that of the human genome, the size of each medaka gene relative to mammalian counterparts seems to be variable, and the medaka cytochrome P-450 aromatase gene is much smaller than the human counterpart (medaka, 2.6 kb vs human, at least 70 kb; Tanaka et al. 1995). The human *C2, Bf*, and duplicated *C4* genes are located in an approximately 100 kb DNA segment in the class III region of the *MHC* (Carroll et al. 1984). Moreover, the distance between the *C2* and *Bf* genes is less than 400 bp (Wu et al. 1987). It is of special interest to walk from the medaka *Bf/C2* gene to the adjacent gene to see the organization and the size of the intergenic region in this small-genome species.

Evolutionary studies of the nonmammalian species *MHC* structure have focused on the amphibian *X. laevis*, and have revealed that the basic structure of the *MHC* was

established before the mammal/amphibian divergence. Although some of the constituent genes, such as class I (Hashimoto et al. 1992), II (Kasahara et al. 1992; Bartl and Weissman 1994), and *LMP7* (Kandil et al. 1996), can be traced back to cartilaginous fish, it is still not known whether these genes form a cluster in a single chromosomal region, like the mammalian *MHC*. In teleost fish, both the class I and II genes have been isolated from carp (Hashimoto et al. 1990; Ono et al. 1993 b), zebrafish (Ono et al. 1993 a, 1992) and Atlantic salmon (Grimholt et al. 1993; Hordvik et al. 1993). The possible genetic linkage between the class I and II genes, however, is yet to be demonstrated, and there is still no evidence to show the presence of the mammalian *MHC*-like gene organization in bony fish. Here we report the linkage analysis of the medaka *Bf/C2* gene, providing entry to the molecular analysis of the teleost *MHC*. We have recently isolated cDNA clones for the medaka LMP7, HSP70, and C3/C4-like molecules. Linkage analysis of these genes in reference to the medaka *Bf/C2* gene will provide the first insight into the possible teleost *MHC* class III region.

Acknowledgments This work was supported in part by grants-in-aid from the Ministry of Education, Science and Culture of Japan, No. 07265225 to M. N. We are willing to share the cDNA and genomic DNA clones described in this paper with interested investigators.

References

Aviv, H. and Leder, P. Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. *Proc Natl Acad Sci USA 69:* 1408–1412, 1972

Bartl, S. and Weissman, I. L. Isolation and characterization of major histocompatibility complex class II B genes from the nurse shark. *Proc Natl Acad Sci USA 91:* 262–266, 1994

Bentley, D. R. Primary structure of human complement component C2. Homology to two unrelated protein families. *Biochem J 239:* 339–345, 1986

Carroll, M. C., Campbell, R. D., Bentley, D. R., and Porter, R. R. A molecular map of the human major histocompatibility complex class III region linking complement genes C4, C2 and factor B. *Nature 307*: 237–241, 1984

Chaplin, D. D., Woods, D. E., Whitehead, A. S., Goldberger, G., Colten, H. R., and Seidman, J. G. Molecular map of the murine *S* region. *Proc Natl Acad Sci USA 80:* 6947–6951, 1983

Davis, B. J. Disc electrophoresis-II. Method and application to human serum proteins. *Ann N Y Acad Sci 121:* 404–427, 1964

Dodds, A. W. and Day, A. J. The phylogeny and evolution of the complement system. *In* K. Whaley, M. Loos, and J. M. Weiler (eds.): *Immunology and Medicine*, pp. 39–88, Kluwer Academic Publishers, London, 1993

Flajnik, M. F., Canel, C., Kramer, J., and Kasahara, M. Evolution of the major histocompatibility complex: molecular cloning of major histocompatibility complex class I from the amphibian *Xenopus*. *Proc Natl Acad Sci USA 88:* 537–541, 1991

Greenhalgh, P. and Steiner, L. A. Recombination activating gene 1 (*Rag1*) in zebrafish and shark. *Immunogenetics 41:* 54–55, 1995

Grimholt, U., Hordvik, I., Fosse, V. M., Olsaker, I., Endresen, C., and Lie, Ø. Molecular cloning of major histocompatibility complex class I cDNAs from Atlantic salmon (*Salmo salar*). *Immungenetics 37:* 469–473, 1993

Hashimoto, K., Nakanishi, T., and Kurosawa, Y. Isolation of carp genes encoding major histocompatibility complex antigens. *Proc Natl Acad Sci USA 87:* 6863–6867, 1990

Hashimoto, K., Nakanishi, T., and Kurosawa, Y. Identification of a shark sequence resembling the major histocompatibility complex class I α3 domain. *Proc Natl Acad Sci USA 89:* 2209–2212, 1992

Hinds, K. R., and Litman, G. W. Major reorganization of immunoglobulin V_H segmental elements during vertebrate evolution. *Nature 320:* 546–549, 1986

Hordvik, I., Grimholt, U., Fosse, V. M., Lie, Ø., and Endresen, C. Cloning and sequence analysis of cDNAs encoding the *MHC* class II β chain in Atlantic salmon (*Salmo salar*). *Immunogenetics 37:* 437–441, 1993

Horiuchi, T., Kim, S., Matsumoto, M., Watanabe, I., Fujita, S., and Volanakis, J. E. Human complement factor B: cDNA cloning, nucleotide sequencing, phenotypic conversion by site-directed mutagenesis and expression. *Mol Immunol 30:* 1587–1592, 1993

Ishikawa, N., Nonaka, M., Wetsel, R. A., and Colten, H. R. Murine complement C2 and factor B genomic and cDNA cloning reveals different mechanisms for multiple transcripts of C2 and B. *J Biol Chem 265:* 19040–19046, 1990

Jensen, J. A., Festa, E., Smith, D. S., and Cayer, M. The complement system of the nurse shark: hemolytic and comparative characteristics. *Science 214:* 566–569, 1981

Kandil, E., Namikawa, C., Nonaka, M., Greenberg, A. S., Flajnik, M. F., Ishibashi, T., and Kasahara, M. Isolation of low molecular mass polypeptide complementary DNA clones from primitive vertebrates: implication for the origin of MHC class I-restricted antigen presentation. *J Immunol 156:* 4245–4253, 1996

Kasahara, M., Vazquez, M., Sato, K., McKinney, E. C., and Flajnik, M. F. Evolution of the major histocompatibility complex: Isolation of class II A cDNA clones from the cartilaginous fish. *Proc Natl Acad Sci USA 89:* 6688–6692, 1992

Kato, Y., Salter-Cid, L., Flajnik, M. F., Kasahara, M., Namikawa, C., Sasaki, M., and Nonaka, M. Isolation of the *Xenopus* complement factor B complementary DNA and linkage of the gene to the frog MHC. *J Immunol 153:* 4546–4554, 1994

Kato, Y., Salter-Cid, L., Flajnik, M. F., Namikawa, C., Sasaki, M., and Nonaka, M. Duplication of the MHC-linked *Xenopus* complement factor *B* gene. *Immunogenetics 42:* 196–203, 1995

Krumdieck, R. and Volanakis, J. E. Construction and expression of complement factor B/C2 chimeras of the region encoded by exon 15. *Abstract book, The 9th International Congress of Immunology,* p. 102, 1995

Kubota, Y., Shimada, A., and Shima, A. DNA alterations detected in the progeny of paternally irradiated Japanese medaka fish (*Oryzias latipes*). *Proc Natl Acad Sci USA 92:* 330–334, 1995

Lambris, J. D., Lao, Z., Pang, J., and Alsenz, J. Third component of trout complement. cDNA cloning and conservation of functional sites. *J Immunol 151:* 6123–6134, 1993

Lambris, J. D., Pappas, J., Mavroidis, M., Wang, Y., Manzone, H., Swager, J., Pasquier, L. D., and Silibovsky, R. The third component of *Xenopus* complement: cDNA cloning, structural and functional analysis, and evidence for an alternate C3 transcript. *Eur J Immunol 25:* 572–578, 1995

Mo, R., Kato, Y., Nonaka, M., Nakayama, K., and Takahashi, M. Fourth component of *Xenopus laevis* complement: cDNA cloning and linkage analysis of the frog *MHC*. *Immunogenetics 43:* 360–369, 1996

Namikawa, C., Salter-Cid, L., Flajnik, M. F., Kato, Y., Nonaka, M., and Sasaki, M. Isolation of *Xenopus LMP-7* homologues. Striking allelic diversity and linkage to MHC. *J Immunol 155:* 1964–1971, 1995

Nonaka, M., Yamaguchi, N., Natsuume-Sakai, S., and Takahashi, M. The complement system of rainbow trout (*Salmo gairdneri*). I. Identification of the serum lytic system homologous to mammalian complement. *J Immunol 126:* 1489–1494, 1981

Nonaka, M., Fujii, T., Kaidoh, T., Natsuume-Sakai, S., Nonaka, M., Yamaguchi, N., and Takahashi, M. Purification of a lamprey complement protein homologous to the third component of the mammalian complement system. *J Immunol 133:* 3242–3249, 1984

Nonaka, M. and Takahashi, M. Complete complementary DNA sequence of the third component of complement of lamprey. Implication for the evolution of thioester containing proteins. *J Immunol 148:* 3290–3295, 1992

Nonaka, M., Takahashi, M., and Sasaki, M. Molecular cloning of a lamprey homologue of the mammalian MHC class III gene, complement factor B. *J Immunol 152:* 2263–2269, 1994

Ono, H., Klein, D., Vincek, V., Figueroa, F., O'hUigin, C., Tichy, H., and Klein, J. Major histocompatibility complex class II genes of zebrafish. *Proc Natl Acad Sci USA 89:* 11 886–11 890, 1992

Ono, H., Figueroa, F., O'hUigin, C., and Klein, J. Cloning of the β_2-microglobulin gene in the zebrafish. *Immunogenetics 38:* 1–10, 1993 a

Ono, H., O'hUigin, C., Vincek, V., Stet, R. J. M., Figueroa, F., and Klein, J. New β chain-encoding *Mhc* class II genes in the carp. *Immunogenetics 38:* 146–149, 1993 b

Rast, J. P. and Litman, G. W. T-cell receptor gene homologs are present in the most primitive jawed vertebrates. *Proc Natl Acad Sci USA 91:* 9248–9252, 1994

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., and Erlich, H. A. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science 239:* 487–491, 1988

Saitou, N. and Nei, M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol 4:* 406–425, 1987

Salter-Cid, L., Kasahara, M., and Flajnik, M. F. *HSP70* genes are linked to the *Xenopus* major histocompatibility complex. *Immunogenetics 39:* 1–7, 1994

Sanger, F., Nicklen, S., and Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA 74:* 5463–5467, 1977

Sato, K., Flajnik, M. F., Pasquier, L. D., Katagiri, M., and Kasahara, M. Evolution of the MHC: Isolation of class II β-chain cDNA clones from the amphibian *Xenopus laevis*. *J Immunol 150:* 2831–2843, 1993

Seeger, A., Mayer, W. E., and Klein, J. A complement factor B like cDNA clone from the zebrafish (*Brachydanio rerio*). *Mol Immunol 32:* 511–520, 1996

Shima, A. and Shimada, A. Development of a possible nonmammalian test system for radiation-induced germ-cell mutagenesis using a fish, the Japanese medaka (*Oryzias latipes*). *Proc Natl Acad Sci USA 88:* 2545–2549, 1991

Shum, B. P., Avila, D., Pasquier, L. D., Kasahara, M., and Flajnik, M. F. Isolation of a classical MHC class I cDNA from an amphibian. Evidence for only one class I locus in the *Xenopus* MHC. *J Immunol 151:* 5376–5386, 1993

Smith, L. C., Chang, L., Britten, R. J., and Davidson, E. H. Sea urchin genes expressed in activated coelomocytes are identified by expressed sequence tags. *J Immunol 156:* 593–602, 1996

Stanley, K. K. and Herz, J. Topological mapping of complement component C9 by recombinant DNA techniques suggests a novel mechanism for its insertion into target membranes. *EMBO J 6:* 1951–1957, 1987

Tanaka, M., Fukada, S., Matsuyama, M., and Nagahama, Y. Structure and promoter analysis of the cytochrome *P*-450 aromatase gene of the teleost fish, medaka (*Oryzias latipes*). *J Biochem 117:* 719–725, 1995

Thomas, P. S. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. *Proc Natl Acad Sci USA 77:* 5201–5205, 1980

Thompson, J. D., Higgins, D. G., and Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res 22:* 4673–4680, 1994

Wada, H., Naruse, K., Shimada, A., and Shima, A. Genetic linkage map of a fish, the Japanese medaka *Oryzias latipes*. *Mol Marine Biol Biotech 4:* 269–274, 1995

Wu, L., Morley, B. J., and Campbell, R. D. Cell-specific expression of the human complement protein factor B gene: Evidence for the role of two distinct 5′-flanking elements. *Cell 48:* 331–342, 1987