Supplemental Data

CHARACTERIZATION OF C1Q IN TELEOSTS: INSIGHT INTO THE MOLECULAR AND FUNCTIONAL EVOLUTION OF C1Q FAMILY AND CLASSICAL PATHWAY

Yu-Lan Hu[1, 2], Xin-Min Pan[1, 2], Li-Xin Xiang[1, 2], and Jian-Zhong Shao[1, 2, 3]
From College of Life Sciences, Zhejiang University, Hangzhou 310058, P. R. China[1]; Key Laboratory for Cell and Gene Engineering of Zhejiang Province, P. R. China [2]; Key Laboratory of Animal Epidemic Etiology and Immunology Prevention of Ministry of Agriculture, P. R. China[3]

Running head: Characterization of C1q in teleosts

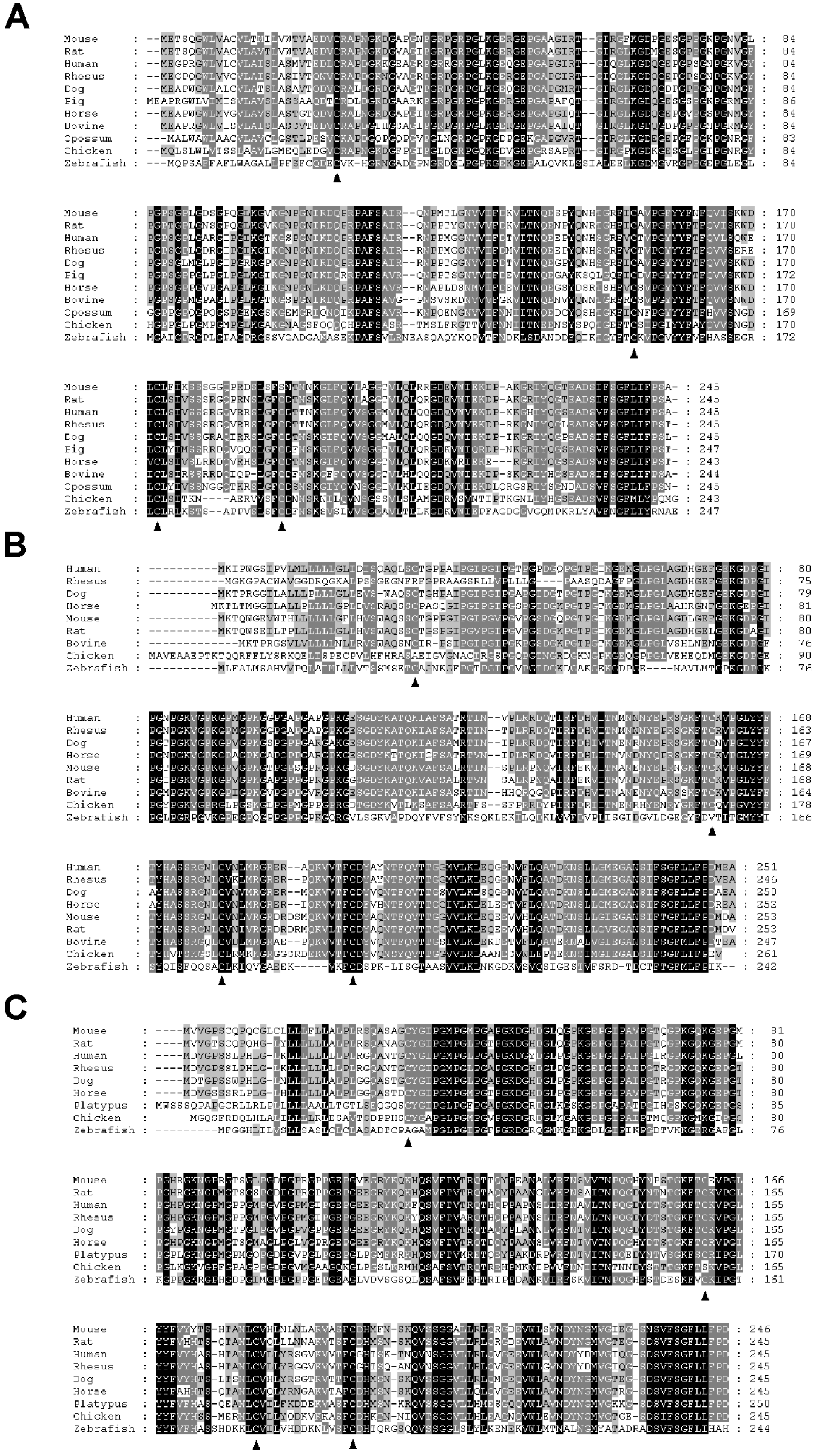
Address correspondence to: Jian-Zhong Shao, PhD, College of Life Sciences, Zhejiang University, Hangzhou 310058, P. R. China. Fax: +86-571-88206582; E-mail: shaojz@zju.edu.cn

SUPPLEMENT TABLE S1. Oligonucleotide primers used to amplify Zebrafish C1qA, C1qB, C1qC genes and for expression analysis

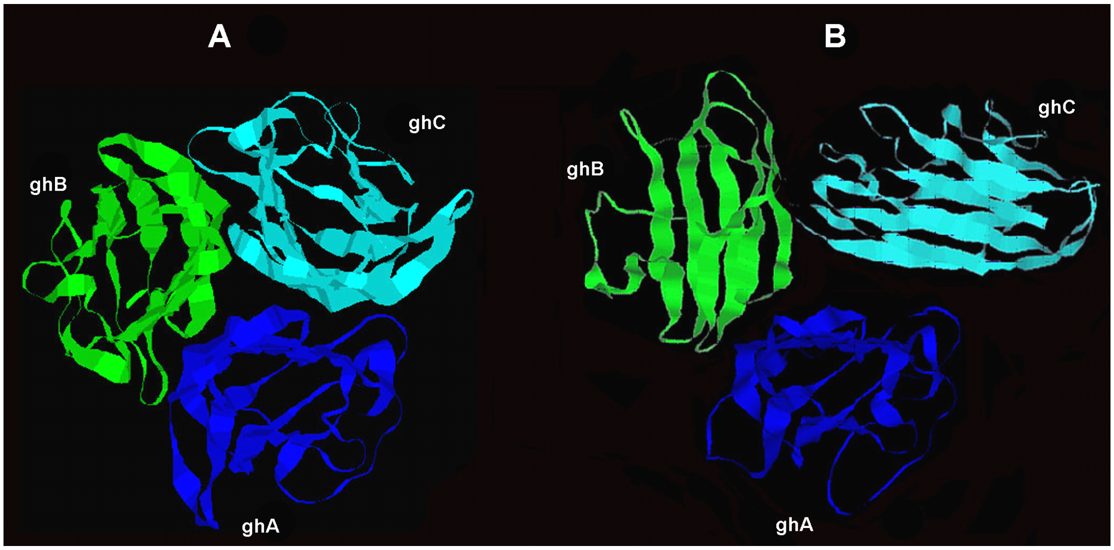
Primers	Sequence (5′–3′)	Use
3AD	CTGATCTAGAGGTACCGGATCC	3’RACE
5RT	GGTTTTGCAATTCTC	5′RACE
C1qA-F1	GCACAAAATTGACAGCGAGACAC	3’RACE
C1qA-F2	CAGCGAGACACTGATGTTTTAGG	3’RACE, gene expression
C1qA-5F1	ATCAGCACCAACACTGGATCC	5’RACE
C1qA-5F2	AGCTGGCTCTCCTTTCTC	5’RACE
C1qB-F1	CACTTTAAGGCCACCATCTG	3’RACE
C1qB-F2	CTGTTTTCCAACAGCCTTCTTG	3’RACE, gene expression
C1qB-R	GGGTAAATACAACAGAGTGACAG	ORF
C1qB-5F1	CACATCAAAGTAACCCTCACCATC	5’RACE
C1qB-5F2	CCTTCACACCTGGCCTGCCTGGGA	5’RACE
C1qC-F1	GTCTGAAGGATTGAGTGGCT	3’RACE
C1qC-F2	GTTTGGCGGTCACCTCATTT	3’RACE, gene expression
C1qC-5F1	GAGACAGTCCACCAGAGCTCA	5’RACE
C1qC-5F2	TCTACAAGGCCTGCTTCGCCG	5’RACE
C1qA-ER	CAACCATCAGCAGGGTCCAGTC	Gene expression
C1qB-ER	GGGTAAATACAACAGAGTGACAG	Gene expression
C1qC-ER	GAAGCGTATCACCTTATTGG	Gene expression
β-actin-F	AGGTCATCACCATCGGCAAT	Gene expression
β-actin-R	GATGTCCACGTCGCACTTCA	Gene expression
ghA-Ex-F	CGGAATTCAAAGCCAGTGAGAAACC	Protein expression
ghA-Ex-R	CCGCTCGAGATTGCGATAGATTAAA	Protein expression
ghB-Ex-F	CGGAATTCGTTTTCTCTTACAAAA	Protein expression
ghB-Ex-R	CCGCTCGAGGAGCATGAAGCCTGTA	Protein expression
ghC-Ex-F	CCCAAGCTTTCTCAACTGCAGTCTGC	Protein expression
ghC-Ex-R	CCGCTCGAGATGAATCAGGAAACCTG	Protein expression

SUPPLEMENT TABLE S2. Percent amino acid similarity of C1A、C1qB and C1qC in different species.

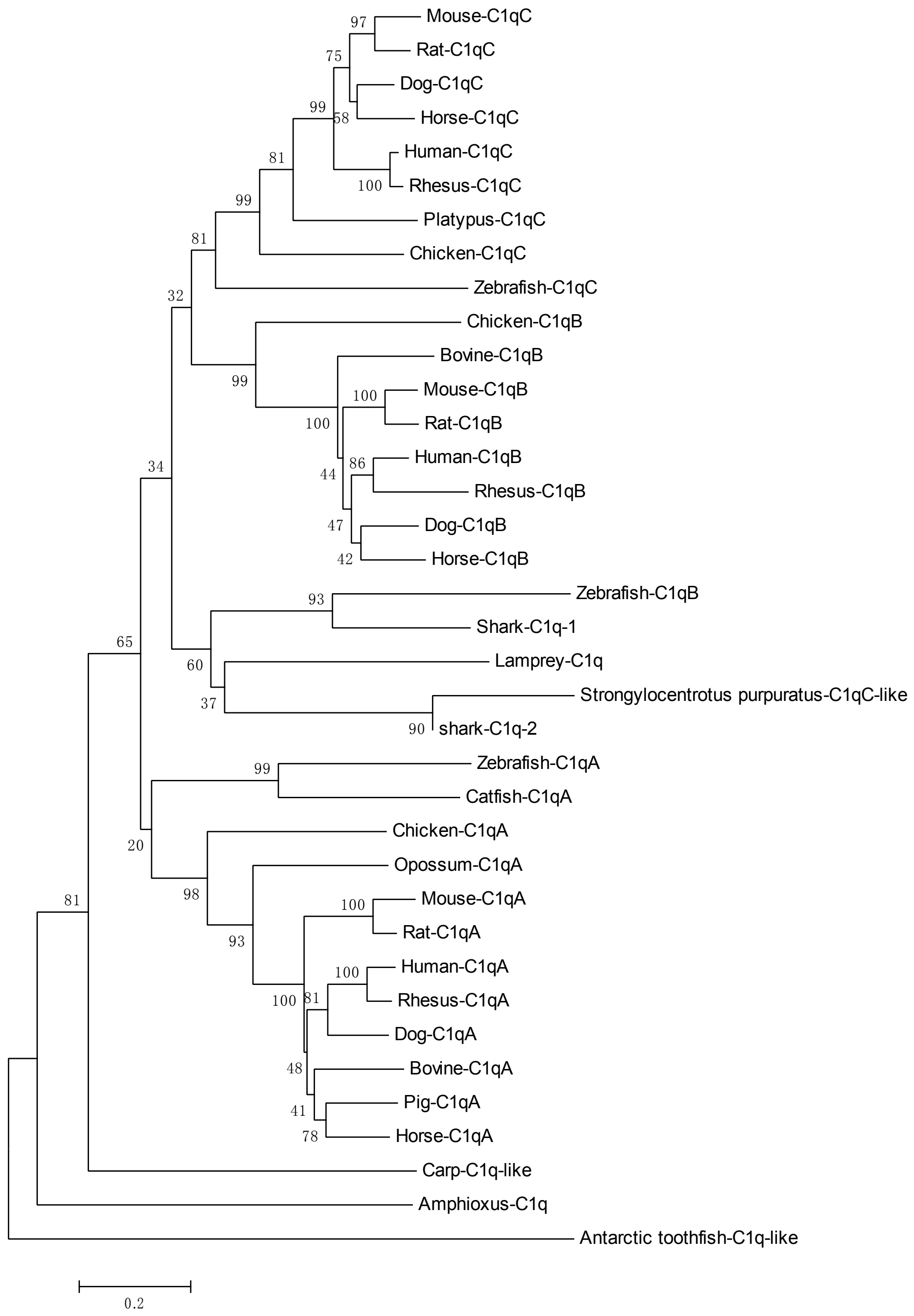
		Human	Chimpanzee	Rhesus	Dog	Bovine	Pig	Horse	Mouse	Rat	Opossum	Chicken	Zebrafish	Human	Rhesus	Dog	Horse	Mouse	Rat	Bovine	Chicken	Zebrafish	Human	Rhesus	Mouse	Rat	Dog	Horse	Platypus	Chicken	Zebrafish
C1qA	Human		98.9	91.7	81.6	77.1	68.8	72.2	72.9	75.6	62	49.2	38.7	39.8	37.2	38	38.7	38	38	38.3	32	29.7	37.7	36.6	36.2	37	37.7	37.4	36.2	40	37
	Chimpanzee			92.9	81.2	77.1	69.5	73.3	73.7	76.3	62	49.2	38.7	39.5	36.8	37.6	39.1	37.6	37.6	39.1	32.3	30.1	37.7	36.6	37	37.4	37.7	37.4	36.2	40	37.7
	Rhesus				81.2	75.2	70.7	74.8	74.1	76.7	62.8	49.6	39.1	39.8	37.2	37.2	39.1	38	38	39.5	32.3	29.7	37.7	36.6	37	37.4	37.4	37	36.6	41.1	37.7
	Dog					77.4	73.7	76.3	73.7	77.1	65	50.4	39.8	39.8	36.8	37.6	39.1	38.7	38.7	38.7	33.1	29.3	37.7	36.6	35.8	36.6	38.5	38.1	35.8	40	36.6
	Bovine						73.3	74.1	69.2	73.7	62	53	40.2	40.2	37.6	38.3	39.1	37.6	38.3	38.3	32	31.2	35.5	34.3	34.3	35.5	36.6	35.5	34.7	39.6	35.5
	Pig							76.7	67.3	70.7	60.9	51.1	39.8	40.2	37.6	38.3	40.2	38	39.1	38	31.6	32	37.7	36.6	36.6	37.4	38.1	38.1	36.6	37.7	37
	Horse								68.8	72.9	62.4	54.5	39.5	39.5	36.1	37.2	38.3	36.8	38.7	40.2	31.6	31.2	38.1	37.7	37	37.7	39.2	37.7	35.8	39.2	38.5
	Mouse									89.8	61.3	50.4	40.2	37.2	33.5	36.5	37.6	37.2	36.5	37.2	32.3	30.5	36.6	35.5	35.1	35.1	34.7	34.3	33.2	37.4	34.7
	Rat										63.5	53	41	39.1	35.3	37.6	38.7	38.7	38.7	38.7	33.1	31.6	37.7	36.6	35.1	35.5	36.2	35.5	33.6	38.1	35.8
	Opossum											54.9	41	39.8	37.2	39.5	41.7	39.8	39.5	39.8	33.1	30.8	40.8	39.6	37	38.9	40.8	38.5	37	40	39.2
	Chicken												35.6	40.8	37.1	40.4	37.8	37.1	37.5	38.2	31.6	34.6	38.5	37.7	37.7	37.4	38.5	37.7	35.5	38.1	37
	Zebrafish													32.2	30	31.8	33.7	33.3	34.5	36	28.9	29.3	32.5	32.1	32.1	33.6	33.2	31.7	30.9	34.3	33.6
C1qB	Human														78.7	84	79.5	80.6	79.1	73.1	49.6	35.3	50.6	50.2	47.5	46.8	50.2	47.2	46.4	48.7	40
	Rhesus															71.6	67.9	67.5	65.3	64.2	45.9	30.5	43.8	43.4	41.5	40.4	43.4	40.8	40.8	43	34.7
	Dog																79.1	78	77.6	72	50.4	36.5	49.8	49.8	47.2	45.3	49.1	47.5	44.9	47.5	39.6
	Horse																	74.3	75.4	73.5	48.5	35.3	49.4	49.1	47.2	46.4	49.4	46.4	45.3	47.9	40
	Mouse																		89.6	72.4	50.4	34.2	49.4	49.4	46.8	45.3	48.7	45.7	42.6	47.2	40.4
	Rat																			70.9	49.2	35.3	48.7	48.7	46.4	45.3	49.1	46	43.4	47.2	39.6
	Bovine																				48.9	35.7	46.4	46.4	44.2	43	45.7	43.4	41.5	44.9	40.4
	Chicken																					25.9	38.9	39.6	38.1	37.7	39.6	37	35.8	38.9	34
	Zebrafish																						35.1	35.1	33.6	34.3	34.7	32.8	33.6	37	30.9
C1qC	Human																							96.6	74.3	78.1	83.8	79.2	62.6	59.2	48.3
	Rhesus																								74.7	78.1	83.8	79.6	61.9	58.9	47.9
	Mouse																									86.4	79.6	78.9	63	58.1	46.4
	Rat																										84.2	83.8	65.7	61.9	46.8
	Dog																											85.7	64.2	60.8	48.7
	Horse																												63.4	58.9	46
	Platypus																													62.6	46.8
	Chicken																														45.7
	Zebrafish																														


A
B
C

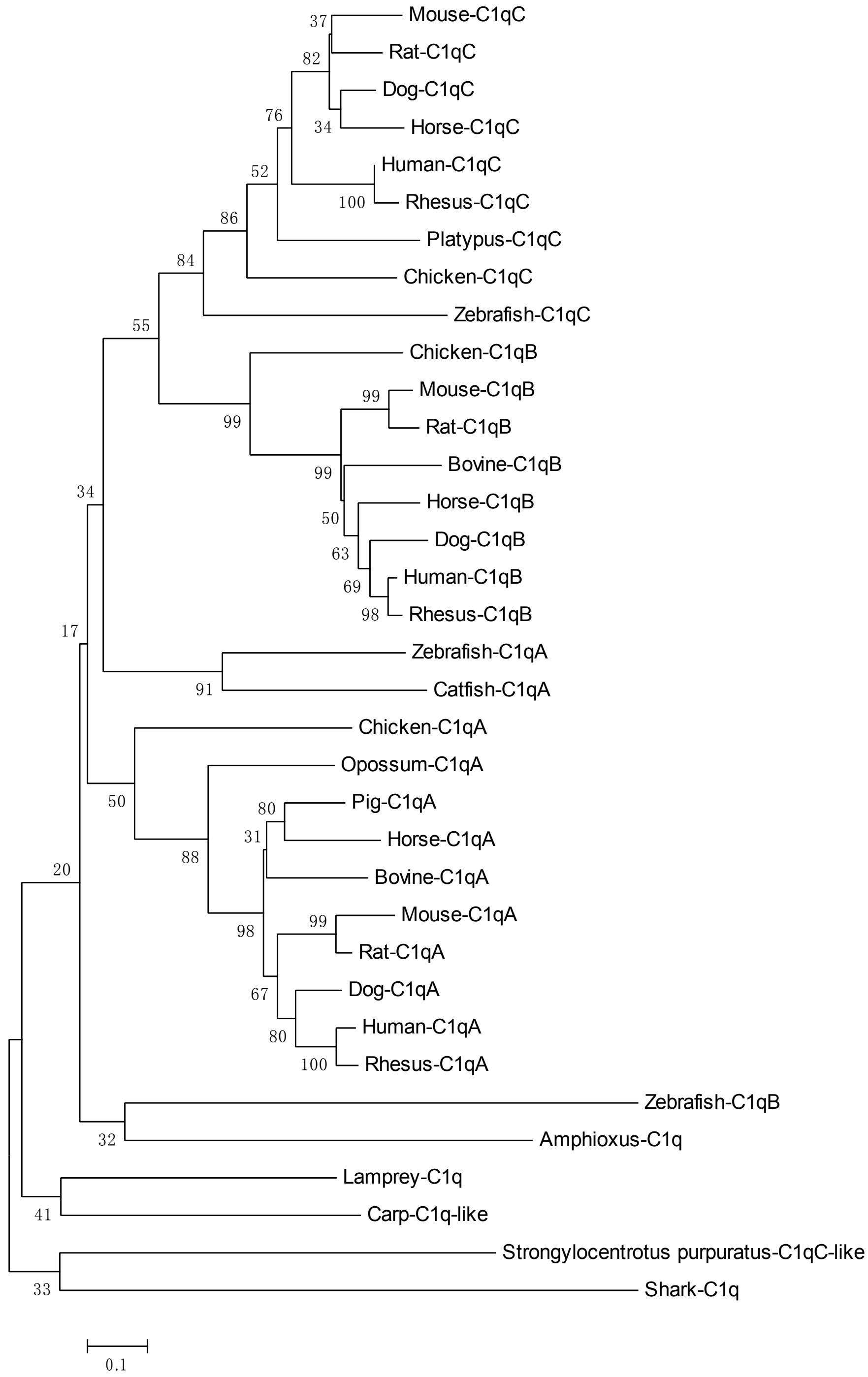
SUPPLEMENTAL FIGURE S1. Multiple alignment of the predicted amino acid sequence of zebrafish (A) C1qA, (B) C1qB, and (C) C1qC with other known C1qA, C1qB, or C1qC. Residues shaded in black are completely conserved across all species aligned, and residues shaded in grey are similar with respect to side chains. Dashes in the amino acid sequences indicate gaps introduced to maximize alignment. Upward arrows indicate the four conserved cysteine residues in different species. GenBank accession numbers used in the alignment are as follows: C1qA genes: human, P02745; rat, P31720; dog, XP_535367; horse, XP_001504311; pig, Q69DL0; bovine, Q5E9E3; opossum, XP_001376435; chicken, XP_417654; rhesus, XP_001101837; mouse, P98086. C1qB genes: human, P02746; rat, P31721; dog, XP_544507; horse, XP_001501545; bovine, Q2KIV9; chicken, XP_425756; rhesus, XP_001110783; mouse, P14106. C1qC genes: human, P02747; rat, P31722; dog, XP_544508; horse, XP_001504308; chicken, XP_417653; rhesus, XP_001102196; platypus, Contig16630:880-2400; mouse, Q02105.



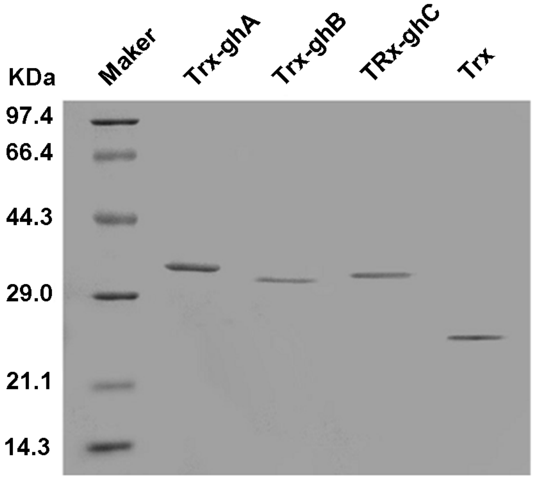
SUPPLEMENTAL FIGURE S2. Structural features of C-terminal globular region (gC1q) signature domains. (A) Superimposed structure of modules A, B, and C of human C1q (Protein Data Bank (PDB) code 1pk6). (B) Modeled 3D structure of zebrafish ghA, ghB, and ghC using human ghA, ghB, or ghC as template, respectively.


Mouse-C1qC
Rat-C1qC
Dog-C1qC
Horse-C1qC
Human-C1qC
Rhesus-C1qC
Platypus-C1qC
Chicken-C1qC
Zebrafish-C1qC
Chicken-C1qB
Bovine-C1qB
Mouse-C1qB
Rat-C1qB
Human-C1qB
Rhesus-C1qB
Dog-C1qB
Horse-C1qB
Zebrafish-C1qB
Shark-C1q-1
Lamprey-C1q
Strongylocentrotus purpuratus-C1qC-like
shark-C1q-2
Zebrafish-C1qA
Catfish-C1qA
Chicken-C1qA
Opossum-C1qA
Mouse-C1qA
Rat-C1qA
Human-C1qA
Rhesus-C1qA
Dog-C1qA
Bovine-C1qA
Pig-C1qA
Horse-C1qA
Carp-C1q-like
Amphioxus-C1q
Antarctic toothfish-C1q-like
97
75
58
99
100
81
99
81
32
99
100
100
44
86
47
42
34
93
65
60
37
90
99
20
98
93
100
100
100
81
48
41
78
81
0.2

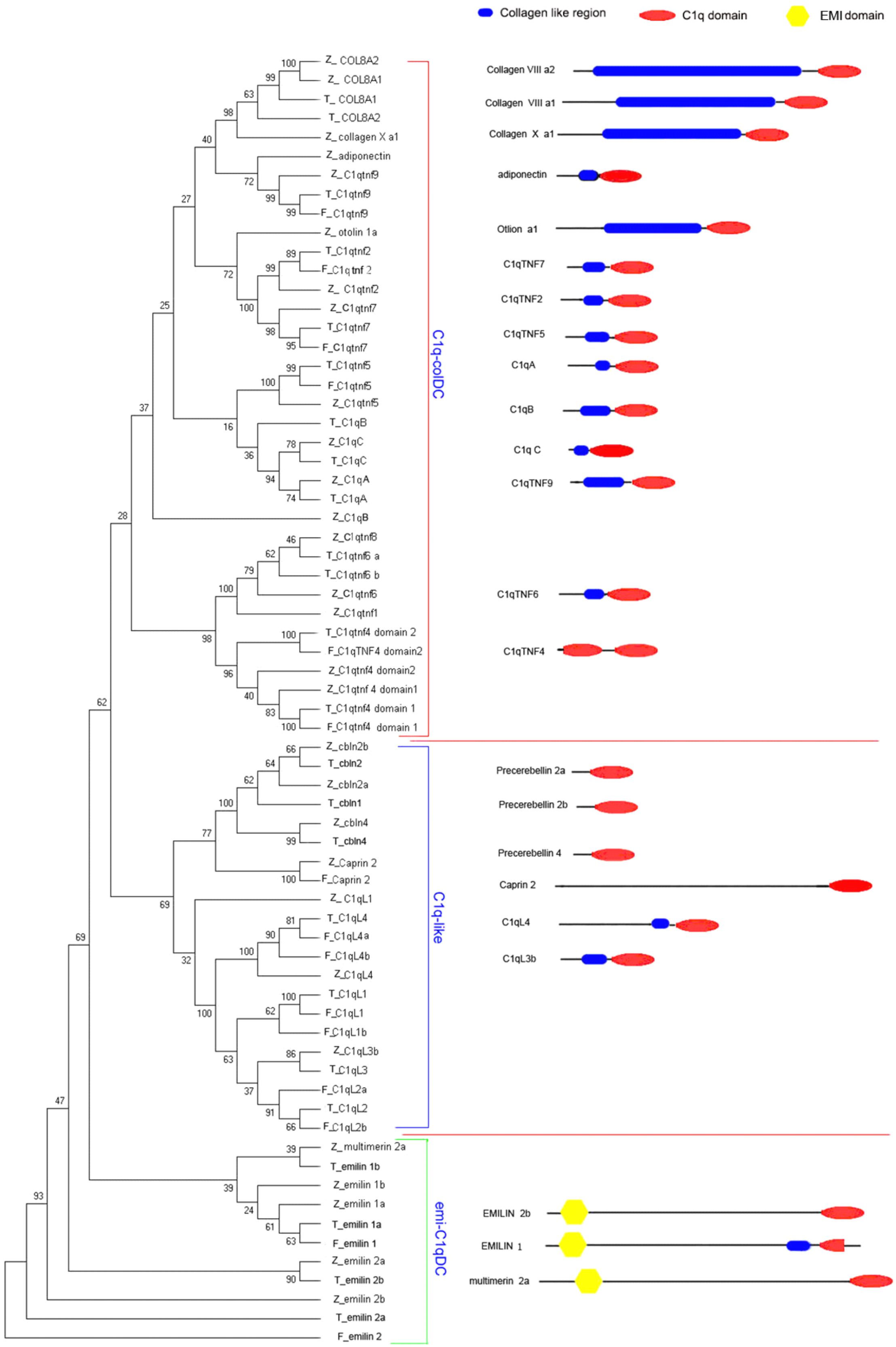
SUPPLEMENTAL FIGURE S3. Phylogenetic tree constructed with neighbor-joining method from Clustal W-generated alignments of C1q sequences. Node values represent percent bootstrap confidence derived from 2000 replicates. The accession numbers are as follows: lamprey C1q, BAD22833.1; Chinese perch C1q-like-1, ABV57765.1; Chinese perch C1q-like-2, ABV57766.1; catfish C1qA, BM425331.1; carp C1q-like, BAD22535.1; amphioxus C1q, FJ154079.1; Antarctic toothfish, ABN45966.1; strongylocentrotus purpuratus, XP_782700.1.The sequences of shark C1q1 and C1q2 were from Smith *et al.* (*39*). The other sequence accession numbers are shown in Fig. S6.


Mouse-C1qC
Rat-C1qC
Dog-C1qC
Horse-C1qC
Human-C1qC
Rhesus-C1qC
Platypus-C1qC
Chicken-C1qC
Zebrafish-C1qC
Chicken-C1qB
Mouse-C1qB
Rat-C1qB
Bovine-C1qB
Horse-C1qB
Dog-C1qB
Human-C1qB
Rhesus-C1qB
Zebrafish-C1qA
Catfish-C1qA
Chicken-C1qA
Opossum-C1qA
Pig-C1qA
Horse-C1qA
Bovine-C1qA
Mouse-C1qA
Rat-C1qA
Dog-C1qA
Human-C1qA
Rhesus-C1qA
Zebrafish-C1qB
Amphioxus-C1q
Lamprey-C1q
Carp-C1q-like
Strongylocentrotus purpuratus-C1qC-like
Shark-C1q
37
82
34
76
100
52
86
84
55
99
99
99
50
63
69
98
34
91
17
50
80
31
88
99
98
67
80
100
20
32
41
33
0.1

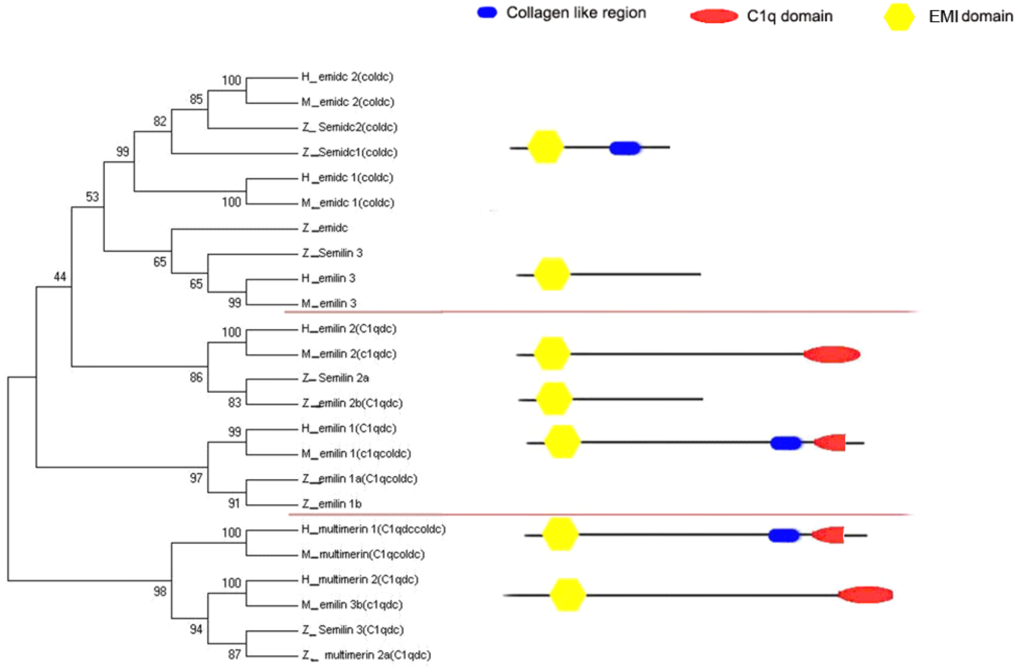
SUPPLEMENTAL FIGURE S4. Phylogenetic tree showing the relationship between zebrafish C1q amino acid sequences and those from other species. It was constructed using C1q globular head domain sequences from different species. Bootstrap values were calculated from 2000 repetitions. Node values represent percent bootstrap confidence. GenBank accession numbers used in the alignment are as described above.



SUPPLEMENTAL FIGURE S5. SDS-PAGE analysis of purified fusion proteins.


Collagen like region
C1q domain
EMI domain
Z_ COL8A2
Z_ COL8A1
T_ COL8A1
T_ COL8A2
Z_collagen X a1
Z_adiponectin
Z_C1qtnf9
T_C1qtnf9
F_C1qtnf9
Z_ otolin 1a
T_C1qtnf2
F_C1qtnf 2
Z_ C1qtnf2
Z_C1qtnf7
T_C1qtnf7
F_C1qtnf7
T_C1qtnf5
F_C1qtnf5
Z_C1qtnf5
T_C1qB
Z_C1qC
T_C1qC
Z_C1qA
T_C1qA
Z_C1qB
Z_C1qtnf8
T_C1qtnf6 a
T_C1qtnf6 b
Z_C1qtnf6
Z_C1qtnf1
T_C1qtnf4 domain 2
F_C1qTNF4 domain2
Z_C1qtnf4 domain2
Z_C1qtnf 4 domain1
T_C1qtnf4 domain 1
F_C1qtnf4 domain 1
Z_cbln2b
T_cbln2
Z_cbln2a
T_cbln1
Z_cbln4
T_cbln4
Z_Caprin 2
F_Caprin 2
Z_C1qL1
T_C1qL4
F_C1qL4a
F_C1qL4b
Z_C1qL4
T_C1qL1
F_C1qL1
F_C1qL1b
Z_C1qL3b
T_C1qL3
F_C1qL2a
T_C1qL2
F_C1qL2b
Z_multimerin 2a
T_emilin 1b
Z_emilin 1b
Z_emilin 1a
T_emilin 1a
F_emilin 1
Z_emilin 2a
T_emilin 2b
Z_emilin 2b
T_emilin 2a
F_emilin 2
100
99
63
98
40
72
27
99
99
89
99
72
25
100
98
95
99
100
37
16
78
36
94
74
28
46
62
79
100
100
98
96
40
62
83
100
66
64
62
100
77
99
100
69
81
90
69
100
32
100
62
100
63
86
37
47
91
66
39
39
93
24
61
63
90
C1q-colDC
C1q-like
emi-C1qDC
Collagen VIII a2
Collagen VIII a1
Collagen X a1
adiponectin
Otlion a1
C1qTNF7
C1qTNF2
C1qTNF5
C1qA
C1qB
C1q C
C1qTNF9
C1qTNF6
C1qTNF4
Precerebellin 2a
Precerebellin 2b
Precerebellin 4
Caprin 2
C1qL4
C1qL3b
EMILIN 2b
EMILIN 1
multimerin 2a

SUPPLEMENTAL FIGURE S6. Phylogenetic trees constructed by neighbor-joining method using MEGA4.0 program and schematic diagrams representing domain structures of C1qDC proteins. Selected C1qDC-encoding sequences are from zebrafish (Z), fugu (F), and tetraodon (T). Proteins with similar domain organization are grouped together. The EMI domain, collagen-like region, and C1q domain are indicated. The sequence accession number in Genbank or Ensemble are as follows:
Z_adiponectin, NP_001038890.1; Z_collagen X a1, AAX54479.1; Z_ otolin 1a, NP_001093211.1; Z_C1qtnf4, NP_001017702; Z_caprin2, NP_001013291.1; Z_cbln2a, NP_001122160.1; Z_cbln2b, NP_001092208.1; Z_cbln4, NP_001103861.1; Z_C1qtnf9, NP_001107934.1; Z_C1qtnf2, XP_700791.1; Z_COL8A2, XP_690914.3; Z_COL8A1, NP_001135846.1; Z_C1qL4, XP_695928.3; Z_C1qL3b, NP_991264.1; Z_C1qtnf6, XP_699453.2; Z_C1qtnf5, NP_001025124.1; Z_C1qtnf8, XP_693502.3; Z_C1qtnf7, XP_693031.2; Z_emilin2a, XP_001341669.2; Z_emilin2b, XP_700066.3; Z_multimerin2a, XP_001923804.1; Z_C1qtnf1, NP_001017875.1; Z_emilin1b, ENSDARP00000033239; Z_emilin1a, NP_001025378.1; Z_C1qL1, XP_699016.3; T_C1qtnf2, ENSTNIP00000011863; T_C1qtnf4, emb|CAG03423.1|; ;T_C1qtnf5, ENSTNIP00000010609; T_C1qtnf6a, emb|CAG12528.1|; T_C1qtnf6b, emb|CAF95529.1|; T_C1qtnf7, emb|CAG10999.1|; T_C1qtnf9, emb|CAG04206.1|; T_COL8A1, emb|CAG00884.1|; T_COL8A2, emb|CAF99434.1|; T_C1qA, emb|CAF95737.1|; T_C1qC, emb|CAF95736.1|; T_C1qB, mRNACR720865partial; T_cbln1, GSTENP00008752001; T_cbln2, GSTENP00015542001; T_cbln4, GSTENP00023416001; T_C1qL2, GSTENP00035278001; T_C1qL1, GSTENP00022029001; T_C1qL3, GSTENP00027650001; T_C1qL4, GSTENP00028104001; T_emilin1a, STENP00035501001; T_emilin1b, GSTENP00019367001; T_emilin2a, GSTENP00034516001; T_emilin2b, GSTENP00031781001; F_C1qL1, ENSTRUP00000038558; F_C1qL4b, NSTRUP00000012152; F_C1qL1b, ENSTRUP00000033292; F_C1qL2a, ENSTRUP00000003096; F_C1qL2b, ENSTRUP00000032791; F_C1qL4a, ENSTRUP00000034386; F_Caprin2, ENSTRUP00000032753; F_C1qtnf2, ENSTRUP00000002973; F_C1qtnf7, ENSTRUP00000035837; F_C1qtnf9, ENSTRUP00000018781; F_C1qtnf4, ENSTRUP00000018012; F_C1qtnf5, ENSTRUP00000038969; F_emilin1, ENSTRUP00000002880; F_emilin2, ENSTRUP00000025497.



SUPPLEMENTAL FIGURE S7. Emu family phylogenetic tree constructed using EMI domain sequences by neighbor-joining method using MEGA4.0 program. Sequence accession number in Genbank or Ensemble are as follows: H_emilin, NP_008977; H_emilin3, NP_443078.1; H_emilin2, NP_114437.2; H_emiDC2, NP_597714.2; H_emiDC1, NP_597712.2; H_multimerin 1, NP_031377.2; H_multimerin 2, NP_079032.1; Z_emilin 1a, NP_001025378.1; Z_emilin 1b, NP_001038807.1; Z_Semilin2, XP_687089.3; Z_SemiDC2, XP_001920406.1; Z_Semilin3, XP_683693.3; Z_emilin2b, XP_700066.3; Z_SemiDC1, XP_001337708.2; Z_Semilin3, XP_694616.2; Z_multimerin2a, XP_001923804.1; M_emilin1, NP_598679.1; M_emilin3, NP_878260.1; M_emilin2, NP_660140.1; M_emiDC2, NP_077794.2; M_emiDC1, NP_542162.1; M_emilin3, NP_694767.3; M_multimerin1, XP_284198.1. Abbreviations: H, Human; Z, zebrafish; M, mouse; C1qdc, C1q-domain-containing sequences; col, collagen-like sequence; C1qcoldc, C1q-collagen-domain-containing sequences.