

Table S1. The third complementarity region (CDR3) of IgM heavy chain variable domain (IGHV) identified in this study

Variants	Catfish IGHV ¹	Human IGHV ²	CDR3 nucleotide sequence ³ (V region end/V-D junction/D region/D-J junction/J start)	CDR3 Amino Acid
1C17	1	3-66*02	GC/ACGAG/AACTCC/TGGGGCC/CTTTGACTAC	ARELLGP FDY
4D9	1	3-74*03	CA/CGCCGG/GGTGG/CACGAAA/TACTTTGACTAC	ARRGGTKY FDY
1E12	2	3-11*04	GA/x/x/TACCGCGGGCCCG/ACTACTTTGACTAC	VRVRGPDY FDY
(2B1) ⁴	2	3-11*04	AG/GCGAGGGG/GTGGCTA/TGGGCT/TTTGATTAC	ARGEGVAMG FxL
3D12	2	3-11*04	AG/TACGAG/GGGGGA/TCAG/CTACTTTGACTAC	ARVRGGISY FDY
2E5	2	3-11*04	GC/AAAAAT/AGCAGCTGG/AAGGTGGGCT/TTTGATTAC	AKIAAGRWA FDY
1E5	2	3-11*06	GC/ACTTAGG/GGCTGG/ACT/TTTGATTAC	ALRGWT FDY
1E8	2	3-11*06	GA/AAGGAA/TATAAC/GAAGGTGCT/TTTGATTAC	ARKEYNEGA FDY
1E3	2	3-13*01	GAG/GGCCCTTCCTC/CTACG/A/CTACTTTGACTAC	ARGPSSYDY FDY
1-35	2	3-38*03	GA/CGG/CCAGCTG/GGAGAGGGGGG/GCTTTTGAT	ARRPAGRGA FDY
1E2	2	3-38*03	GA/ACTTC/CTACG/A/CTACTTTGACTAC	ARTSYDY FDY
1E6	2	3-38*03	AT/CGGGTAGGGTGG/ATTATAA/TGCT/TTTGGCTAC	ARSGRVDYNA FGY
(1E9)	2	3-38*03	GA/GGTT/TGGCTACG/GCTTTAAGCT/TTTGGCTAC	LEVWLRLxA FGY
2E9	2	3-38*03	TA/x/x/GCGGG/TACTTTGACTAC	ARYSGY FDY
3D8	2	3-38*03	AA/GA/ACTAC/CTACG/ACTACTTTGACTAC	ARTTYDY FDY
1C8	2	3-48*01	GC/AGGATTC/GATGGC/GCCGATCTCGCTTTTGG/CTAC	AGFDGADLA FGY
3B46	2	3-48*03	GA/x/CTACTAC/G/ACTACTTTGACTAC	ARDYYDY FDY
3B44	2	3-53*03	GG/GTGGCAGTA/ATGCT/T/TTGACTAC	ARGGSNA FDY
3D13	2	3-64*04	AG/GGCCCTTCCTC/CTACG/A/CTACTTTGACTAC	ARGPSSYDY FDY
3D4	2	3-64*04	AG/GGGG/TACTAC/x/GATGCTTTTGAT	ARGYYDA FDY
1B15	2	3-66*01	GC/AGAAAGTATCGGCAC/CTGGA/GG/TTTGACTAC	AESIGTWR FDY
1E4	2	3-66*01	GC/AGAAAGTATCG/GCAGCTGG/AGG/TTTGACTAC	AESIGSWR FDY
1E7	2	3-66*01	GC/AGAAAGTATCG/GCAGCTGG/AGG/TTTGACTAC	AESIGSWR FDY
2E2	2	3-66*01	GC/AAAT/ACTGGGG/TGGCACTCTATAATGCT/TTTGGCTA	ANTGVALYNA FGY
3B42	2	3-66*01	AG/GGGTA/TTGGGG/CTCTTGAT/TTTGATTAC	ARGVLGALD FDY
3B43	2	3-66*01	GA/TCCCCGAATGGGA/ATGCT/T/TTGACTAC	ARSPNGNA FDY
3D14	2	3-66*01	GA/TCCCCGAATGGGA/ATGCT/T/TTGACTAC	ARSPNGNA FDY
3D6	2	3-66*01	GA/x/TCGGGG/GTGGCACTA/TACTTTGACTAC	ARSGVALY FDY
3D9	2	3-66*01	GA/T/TGGATAT/CGCTT/ACTACTTTGAGTAC	ARLDIAYY FEY
1E10	5	2-5*02	GC/TCGTTTCCC/CAGCAGCTGG/AGTAGCGGGT/ACTACTT	ARFPSSWSSGYFDY
2E3	3	3-21*02	GA/C/GTAGC/GGGA/ACTACTTTGACTAC	TRRSGNY FDY
1A32	3	3-30*03	GA/CGG/ACTGGGG/TG/TACTTTGACTAC	TRRTGVY FDY
1A53	3	3-30*03	GA/AGA/GCGACTA/AC/TACTTTGACTAC	TRRATNY FDY
2E1	3	3-64*04	GC/CCGTGACCCCACTC/ACTACGG/ACG/CTTTGACTAC	ARDPTHYGR FDY
4-7	4	3-33*03	GC/ACCCGCCTTTATAGC/TGGGGTAGTTAT/AATGCT/TTT	TRLYSWGSYNA FDY
(1C14)	5	2-5*02	GC/TCGTTTCCGGG/TCGGGG/CGGCGA/GACTAC	ARFRVGAARL x
2E13	5	2-70*02	TG/CGCTCGTGAAGCCA/ACTACGG/CGCCAGGACGGTGCT/TTT	ARGSQLRRQDGA FDY
4B42	5	2-70*10	TG/GTCGTGGGGGGGGCCTGGGGTGGCAC/TATGCT/T/TTG	GRGGGLGWHYA FDY
4B47	5	2-70*10	TG/GTCGTGGGGGGGGCCTGGGGTGGCACTATGCT/T/TTG	GRGGGLGWHYA FDY
3C8	6	1-2*02	GC/CCGGCG/GTACCA/CGCT/TTTGACTAC	ARRYHA FDY
3D7	6	1-2*02	GC/TCGGCAGGC/GGCTACGA/CGG/ACTTTGACTAC	ARQAATTD FDY
1E11	6	1-18*04	GC/TCGTGGG/GCATAT/AGCGC/CTTTGACTAC	ARGAYSA FDY
2-12	6	1-46*03	GA/AGAGGGG/GGTGG/CACG/TACTTTGACTAC	AREEGGGTY FDY
2E8	6	1-46*03	GA/A/CAGCAGCTGG/AA/TACTTTGACTAC	AREQQLEY FDY
3D5	6	1-46*03	CT/CGTGCGAGTC/ACTACGG/x/CTTTGACTAC	ARASHYG FDY
1B32	6	2-70*02	TG/CGCTCGTGAACGAG/ACGGTGG/CGATGCT/TTTGGCTA	ARERDGGDA FGY
4B41	6	3-66*01	GC/TCGTGAACC/CTGGGG/TGTCCGTGCT/TTTGATTAC	AREPWGVRA FDY
4D4	6	3-66*01	GC/ (GC) /TCGTC/AGGGGGTCGCT/TTTGACTAC	ARQGVA FDY
2E-1	6	3-66*04	CG/ (AC) /TGGGG/GGGTCCG/ACTACTTTGACTAC	ARLGGSY FDY
1B31	7	3-69-1*02	GC/CCGACGGTATG/GCAGCTGG/AG/CTACTTTGACTAC	ARRYGSWSY FDY

3D3	9	1-46*03	GA/ACTG/GCAGCGG/GTATT/TTTGACTAC	ARELAAGI	FDY
1E1	12	4-59*03	TG/CACACGAGT/TAAGTGGG/CT/TTTGATTAC	TRVNW	FDY

¹ The family grouping was based on Bengtén, E., Clem, L.M., Miller, N.W., Warr, G.W., Wilson, M. 2006. Channel catfish immunoglobulins: Repertoire and expression. Developmental & comparative Immunology. 30:77-92.

² The grouping and identification of variable (V), diversity (D), and joint (J) gene segments in CDR3 was based on Ye, J., Ma, N., Madden, T.L., Ostell, J.M. 2013. IgBLAST: an immunoglobulin variable domain sequence analysis tool. Nucleic Acid Research 41:w34-40.

³ The missing site in CDR3 sequence (or having a stop codon) was marked by x.

⁴ Variants with parentheses were pseudogenes; they either have a stop codon or are out of translational frame.