**Supplementary File 3:** Detailed explanation of Table 1

Main text Table 1 shows the lowest percentages of amino acid sequence identities between membrane-distal domains (1+2 for MHC I, 1 for MHC IIA, 1 for MHC IIB) of same category MHC molecules that we were able to find for reported sequences within the same species. Whether sequences are allelic or from different loci was not considered. In some species no genes for particular categories were found (black boxes), while in other instances only one seemingly intact gene sequence was found (1 sequence) or only pseudogenes were found (pseudogene). Only the membrane-distal domains were considered because they are important for antigen-fragment binding and for interaction with T cell receptors. For a division of teleost MHC class I sequences into lineages U, Z, L, P and S, see reference1. The category "U classical" only considers U lineage molecules in which at least six of the eight characteristic "key" amino acids for binding of the peptide ligand termini have been conserved2. Bony fishes do not have a conserved non-classical MHC class II lineage, and the emerging picture is that most teleost fish have classical MHC class II genes belonging to the DA lineage, plus often a number of non-classical genes which were derived from duplicates of the classical genes at some point in evolution and which are not consistently inherited3. The DA lineage separated from DB category sequences probably early in the teleost line, with the relationship between the various extant DB genes and DA genes not fully resolved3. The MHC class II lineage DE already separated from the DA/DB lineage before the separation between the ancestors of teleost fish and gar, and this very ancient lineage appears to have been lost in most teleost fish3. For finding the lowest percentages of amino acid sequence identities in the respective species, we started with the alignments of teleost MHC class I and II sequences that we showed previously1,3 in combination with sequence analysis software that determined the sequence identity percentages (GENETYX Version 12.0.3, Tokyo, function "Create %Identity Matrix" with setting "Gaps are NOT taken into account"). Then, by various blastp searches, we tried to find even more divergent sequences in the nonredundant sequence section of the NCBI database, and incorporated them into the same analysis for calculation of the amino acid identity percentages. Finally, for all seven fish species we retrieved all GenBank protein sequence database accessions which were found with the search word "MHC" (>300 for Atlantic cod, also hundreds for most of the other fishes), and separated them into the respective MHC categories based on their designations and our phylogenetic analysis (Muscle alignment followed by neighbor joining tree analysis including reference sequences). We then aligned those sequences with seemingly intact membrane-distal domains together with our other set of sequences (using GENETYX Version 12.0.3, Tokyo, function "Muscle Alignment"), and then determined the identity percentages as described above. For finding cod Z lineage sequences we analysed the sequence scaffolds published by Malmstrøm *et al.*4 as we did elsewhere1, and then determined the lowest percentage of amino acids between their 1+2 sequences as done for the other sequences. Below follow the names plus GenBank accessions (between brackets) and/or relevant article references for the sequences whose comparisons gave the sequence identity percentages shown in the table. Although not necessarily canonical full-length MHC molecules, all the compared sequences might form intact molecules.

In reference1, estimations of the gene+pseudogene numbers in the haploid genomes of the several MHC categories in the seven here listed fish species were given. For the U lineage sequence numbers in zebrafish, salmon, medaka, fugu, stickleback and tilapia these estimates were 4, 7, 13, 8, 29, and 45, respectively (reference1). Although these estimations rely on the quality of the available information, there is little doubt that these fish species have considerably fewer U lineage sequences than the estimated 80-100 in cod (although that was analysed by a different method4). Despite the limited number of U lineage genes in zebrafish and salmon, in these species impressive allelic/haplotype variation can be found among the expressed classical U lineage sequences1, reflected in the here presented 40% and 47% amino acid sequence identities between their most divergent alleles.

Note: Our analyses were performed as part of an intended correspondence article submitted in its last form to Nature Genetics in February 2017. It then took 15 months before we received the first response, which was final and negative. We did not re-perform the analyses, although new fish MHC sequences keep being published. The reason for not updating our analyses is that it is a real lot of work, while we believe that essentially the data did not change.

**The sequence pairs whose distal domain identity percentages are shown in Table 1:**

 **Zebrafish**

**MHC class I** U classical UBA (NM\_131471) and UJA (NM\_200406); see reference5

 U all UIA (KC626502) and UMA (KR086347); see references5,6

 Z mhc1ze (NM\_194425) and ZJA (KC607869); see reference7

 L LFA (NM\_001327913) and (XM\_001340377); see references1,8

**MHC class II** DA IIA D8.37\_A1 (XM\_005167233) and D8.37\_A4 (NM\_001005975); (two loci . on Chr. 8) see reference3

DA IIB D8.35\_B2 (NM\_001080633) and MHCII (NM\_001005943) ; see references3,9

DB IIA D8.46\_A (XM\_694857) and D8.45\_A2 (XM\_009304226); (two loci on Chr.8) see reference3

DB IIB D18\_B (XM\_005159194) and D8.45\_B2 (XM\_009304199); (a locus on Chr.18 and a locus on Chr.8) see reference3

 **Salmon**

**MHC class I** U classical UBA\*1001 (AF504024) and UBA\*4001 (JN897012); see reference1 and . site https://www.ebi.ac.uk/ipd/mhc/fish/index.html

U all UDA\*0101 (FJ969490) and UHA1\*0101 (FJ969489); see reference1

 Z ZBA\*0101 (GQ505860) and ZE\*0101 (DQ099914); see reference1

 L SAA\*0101 (FJ969488) and (XM\_014188252); see reference1

**MHC class II** DA IIA DAA\*0801 (AY780910) and DAA\*1201 (AY780914); see references3,10 and site <https://www.ebi.ac.uk/ipd/mhc/fish/index.html>

DA IIB DAB\*0101 (X70166) and DAB\*0601 (X70165); see references3,11 and site <https://www.ebi.ac.uk/ipd/mhc/fish/index.html>

 DB IIA DCA (XM\_014150051) and DDA (XM\_014130418); see reference3

 DB IIB DCB\*0103 (KC316031) and DDB (XM\_014130415); see reference3

 DE IIA DEA\*0101 and DEA\*0102 (KC316033); see reference3

 DE IIB DEB\*0101 and DEB\*0102 (KC316036); see reference3

 **Medaka**

**MHC class I** U classical UBA\*0206 (AB451005) and UGA\*2001 (AB604103); see references12,13

*Note: Classical nature of medaka UGA has been questioned, but there is also only 53% identity between the recognized classical medaka sequences UBA\*0206 (BAJ07301) and UBA\*0201 (BAB83850) (references11,12)*

U all UEA\*0201 (BA000027) and UHA\*2201 (AB604115); see references1,13

 Z OL17 (XM\_011493562) and (XM\_011480916); see reference1

**MHC class II** DA IIA DCA\*21 (JQ743250 ) and DFA\*21 (JQ743256); see references3,14

 DA IIB DAB\*206 (JQ743290) and DCB (AB033214); see references3,14

 DB IIA DDA\*21 (JQ743252) and DEA\*11 (JQ743255); see references3,14

 DB IIB DDB1\*21 (JQ743261) and DEB\*21 (JQ743264); see references3,14

 **Fugu**

**MHC class I** U classical TR3 (XM\_011609132) and I103 (CAC13116); see references1,15

U all TR7 and TR13; see reference1 (sequences can be retrieved from Text S2 in that article)

P TR26 and TR29; see reference1 (sequences can be retrieved from Text S2 in that article)

**MHC class II** DA IIA (XM\_011619962) and (XM\_011621535); see reference3

 DA IIB (XM\_011620545) and (XM\_003976765); see reference3

 **Stickleback**

**MHC class I** U classical GA1 and Gaac-UAA\*01 (ABN14358); see reference1 (sequence . GA1 can be retrieved from Text S2 in that article) and reference1

U all GA12 and GA16; (two loci on linkage group X) see reference1 (sequences can be retrieved from Text S2 in that article)

**MHC class II** DA IIA GVII\_A and DAA\*01 (AY713945); see reference3 (sequence GVII\_A can be . retrieved from Text S1 in that article) and reference17

DA IIB GVII\_B and DBB (AY713945); see reference3 (sequence GVII\_B can be retrieved from Text S1 in that article) and reference17

 **Tilapia**

**MHC class I** U classical ON1 (XM\_005457584) and ON20 (XM\_013265954); see reference1

U all ON32 and ON45 (XM\_013267852); see reference1 (sequences can be retrieved from Text S2 in that article)

 Z ON22 (XM\_005463504) and ON24 (XM\_005464317); see reference1

**MHC class II** DA IIA O22\_A1 (XM\_005460587) and O745\_A (XM\_005465903); see . reference3

 DA IIB O31\_B1 (XM\_013274958) and MHC II beta (JN983489); see reference3

DB IIA O29\_A3 (XM\_003451009) and O57\_A; see reference3 (sequences can be retrieved from Text S1 in that article)

DB IIB O9\_B (XM\_005471522) and O57\_B; see reference3 (sequences can be retrieved from Text S1 in that article)

 **Cod**

**MHC class I** U classical clone 0122 (KF033069) and clone 0101 (JX567634); see references4,18

 U all clone 0122 (KF033069) and clone 0101 (JX567634); see references4,18

Z Scaffold\_CAEA020024072\_Z and Scaffold\_CAEA020024073\_Z; see reference4

**References used in this supplementary file**

1. Grimholt, U., Tsukamoto, K., Azuma, T., Leong, J., Koop, B.F. & Dijkstra, J.M. A comprehensive analysis of teleost MHC class I sequences. *BMC Evol. Biol.* **15**, 32 (2015).

2. Hashimoto, K., Okamura, K., Yamaguchi, H., Ototake, M., Nakanishi, T. & Kurosawa, Y. Conservation and diversification of MHC class I and its related molecules in vertebrates. *Immunol. Rev.* **167**, 81-100 (1999).

3. Dijkstra, J.M., Grimholt, U., Leong, J., Koop, B.F., Hashimoto, K. Comprehensive analysis of MHC class II genes in teleost fish genomes reveals dispensability of the peptide-loading DM system in a large part of vertebrates. *BMC Evol. Biol.* **13**, 260 (2013).

4. Malmstrøm, M., Matschiner, M., Tørresen, O.K., Star, B., Snipen, L.G., Hansen, T.F., Baalsrud, H.T., Nederbragt, A.J., Hanel, R., Salzburger, W., Stenseth, N.C., Jakobsen, K.S. & Jentoft, S. Evolution of the immune system influences speciation rates in teleost fishes. *Nat. Genet.* **48**, 1204-1210 (2016).

5. McConnell, S.C., Restaino, A.C. & de Jong, J.L. Multiple divergent haplotypes express completely distinct sets of class I MHC genes in zebrafish. *Immunogenetics* **66**, 199-213 (2014).

6. Dirscherl, H. & Yoder, J.A. A nonclassical MHC class I U lineage locus in zebrafish with a null haplotypic variant. *Immunogenetics* **67**, 501-513 (2015).

7. Dirscherl, H. & Yoder, J.A. Characterization of the Z lineage Major histocompatability complex class I genes in zebrafish. *Immunogenetics* **66**, 185-98 (2014).

8. Dirscherl, H., McConnell, S.C., Yoder, J.A. & de Jong, J.L. The MHC class I genes of zebrafish. *Dev. Comp. Immunol.* **46**, 11-23 (2014).

9. Ono, H., Klein, D., Vincek, V., Figueroa, F., O'hUigin, C., Tichy, H. & Klein, J. Major histocompatibility complex class II genes of zebrafish. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 11886-11890 (1992).

10. Consuegra, S., Megens, H.J., Leon, K., Stet, R.J. & Jordan, W.C. Patterns of variability at the major histocompatibility class II alpha locus in Atlantic salmon contrast with those at the class I locus. *Immunogenetics* **57**, 16-24 (2005).

11. Stet, R.J., de Vries, B., Mudde, K., Hermsen, T., van Heerwaarden, J., Shum, B.P. & Grimholt, U. Unique haplotypes of co-segregating major histocompatibility class II A and class II B alleles in Atlantic salmon (Salmo salar) give rise to diverse class II genotypes. *Immunogenetics* **54**, 320-331 (2002).

12. Nonaka, M.I. & Nonaka, M. Evolutionary analysis of two classical MHC class I loci of the medaka fish, Oryzias latipes: haplotype-specific genomic diversity, locus-specific polymorphisms, and interlocus homogenization. *Immunogenetics* **62**, 319-32 (2010).

13. Nonaka, M.I., Aizawa, K., Mitani, H., Bannai, H.P. & Nonaka, M. Retained orthologous relationships of the MHC Class I genes during euteleost evolution. *Mol. Biol. Evol.* **28**, 3099-3112 (2011).

14. Bannai, H.P. & Nonaka, M. Comprehensive analysis of medaka major histocompatibility complex (MHC) class II genes: implications for evolution in teleosts. *Immunogenetics* **65**, 883-895 (2013).

15. Clark, M.S., Pontarotti, P., Gilles, A., Kelly, A. & Elgar, G. Identification and characterization of a beta proteasome subunit cluster in the Japanese pufferfish (Fugu rubripes). *J. Immunol.* **165**, 4446-4452 (2000).

16. Schaschl, H. & Wegner, K.M. Contrasting mode of evolution between the MHC class I genomic region and class II region in the three-spined stickleback (Gasterosteus aculeatus L.; Gasterosteidae: Teleostei). *Immunogenetics* **59**, 295-304 (2007).

17. Reusch, T.B., Schaschl, H. & Wegner, K.M. Recent duplication and inter-locus gene conversion in major histocompatibility class II genes in a teleost, the three-spined stickleback. *Immunogenetics* **56**, 427-37 (2004).

18. Star, B. *et al.* The genome sequence of Atlantic cod reveals a unique immune system. *Nature* **477**, 207-210 (2011).