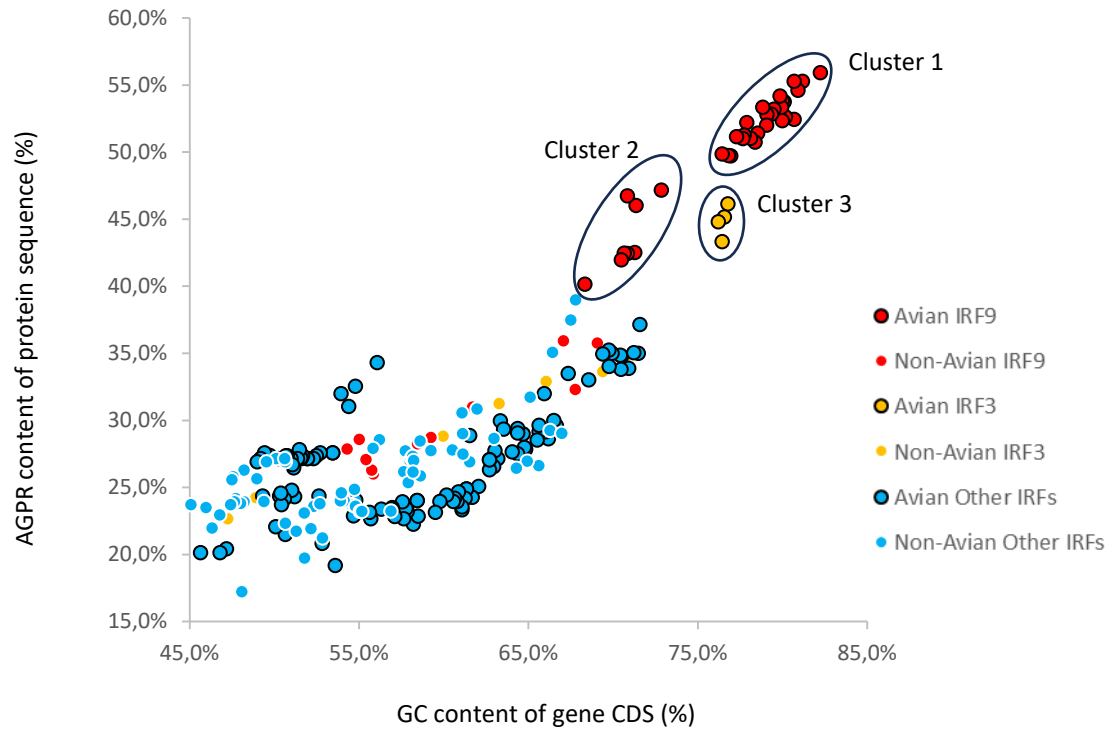




Cluster 1:

Avian IRF9 genes in original syntenic location on the dot chromosome equivalent of chicken chr 34 or avian IRF9 genes translocated to a different dot chromosome

Cluster 2:

Avian IRF9 genes translocated to either a macrochromosome or a non-dot microchromosome

Cluster 3:

Avian IRF3 genes in original syntenic position on the dot chromosome equivalent of chicken chr 31

Fig. S1 Percentages of amino acids with GC-rich codons (A, G, P, R) in IRF protein sequences correlate with GC contents of IRF coding sequences. The percentages were calculated using the sequences contained in and additional file (Additional file 1: Table S3). Data from all genes where calculation of GC and AGPR percentages are shown in the table were included. Duplicate values were eliminated. Percentages were not calculated for genes with inadequately complete sequences.

Caretta caretta, loggerhead sea turtle; asi – *Alligator sinensis*, Chinese alligator) and mammals (hsa – *Homo sapiens*, human; mmu – *Mus musculus*, house mouse; cfa – *Canis familiaris*, dog). Exonic structures of CDS are shown by alternating black and blue ink. Amino acids with codons split by introns are shown in red ink; in the case of phase 0 exon/intron boundaries, both bordering amino acids are bold. Five conserved tryptophans in DBD are indicated by a cyan highlight while two conserved serines in the C-terminus instrumental for activation of IRF3 by phosphorylation are indicated by a yellow highlight. Four-fold repetitive structure of the extension of the second exon of emu (dno) IRF3 is visualized using alternating solid and wavy underlines.

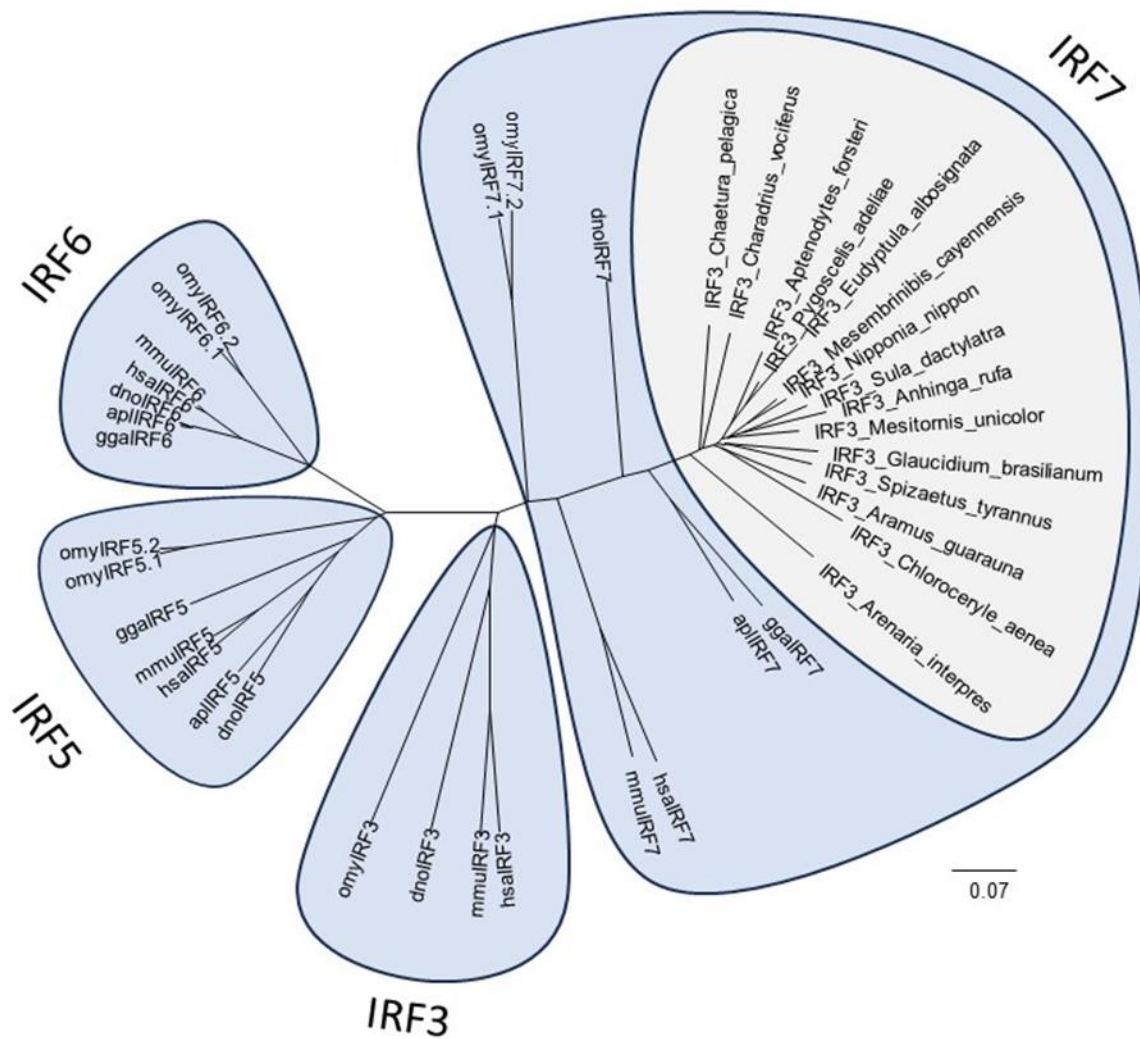


Fig. S3 IRF7 sequences of neognath birds that are erroneously annotated as IRF3 in GenBank (shown on gray background) cluster with vertebrate IRF7 sequences rather than with IRF3. NJ tree of Clustal X-aligned protein sequences. False IRF3 sequences are listed in an additional file (Additional file 1: Table S5). Control IRF sequences come from the following species: apl - *Anas platyrhynchos* (duck), dno - *Dromaius novaehollandiae* (emu), gga - *Gallus gallus* (chicken), hsa - *Homo sapiens* (human), mmu - *Mus musculus* (mouse), omy - *Oncorhynchus mykiss* (rainbow trout), and are available in an additional file (Additional file 1: Table S3).

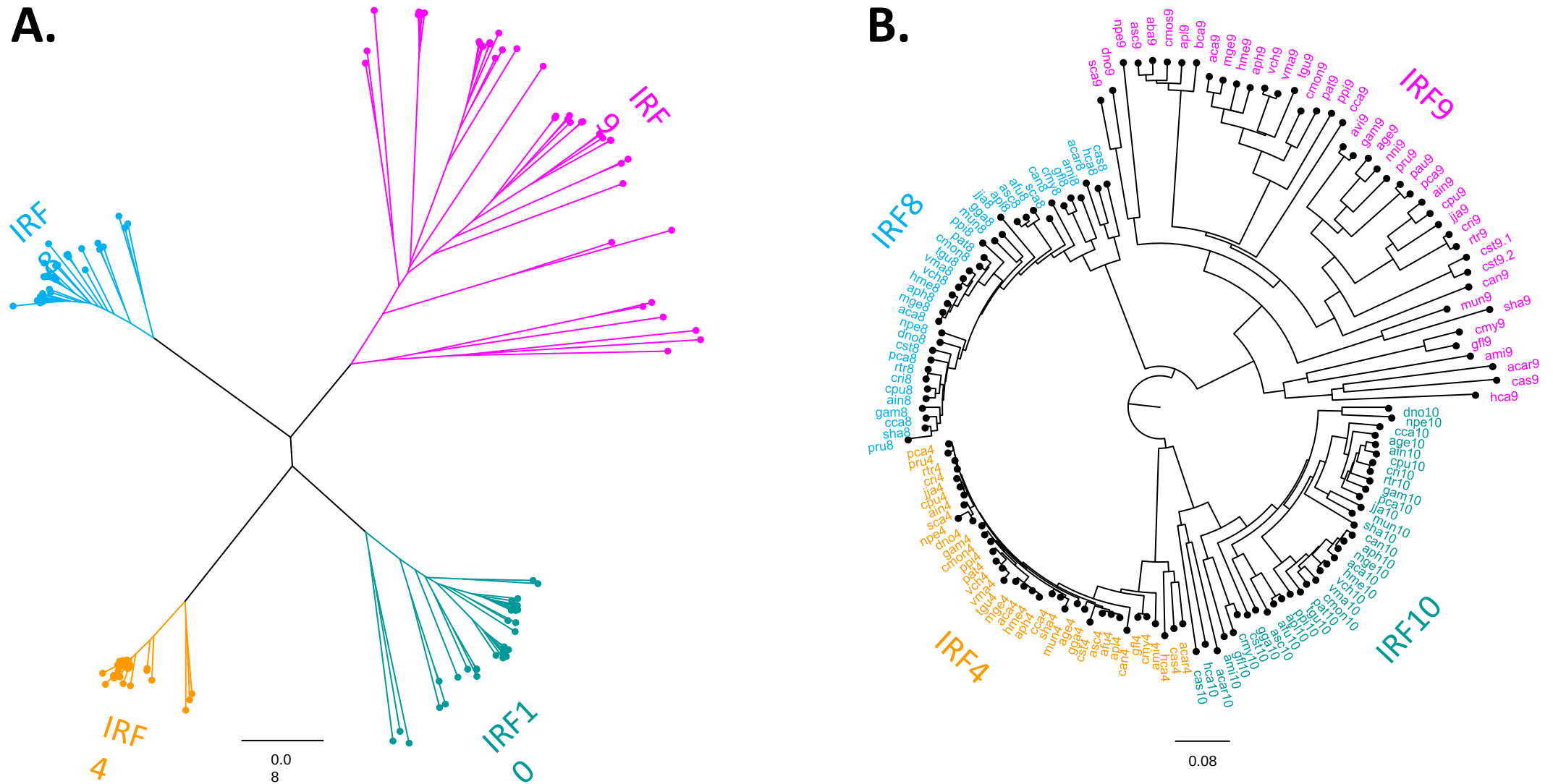


Fig. S4 Supplementary information for Fig. 5. Avian IRF9 protein sequences form a single clade in the phylogenetic tree. **A.** The tree shown in the same form of radial phylogram as is shown in Fig.5 in the main text. Each IRF family member is, however, distinctly color-coded. **B.** The same tree shown in form of circular cladogram with species labels included. For species labels look in the additional file (Additional file 1: Table S3).

duck IRF9 1 ATGACGT CAGGGCGGCGCGCCTCTGCCCTGGCTGCTGGCGCAGGCGCAGAGCGGCCCTTCCCGGGGCTGGAGTTTGACGACGCCCGCGCAGCGCCTCAGGG 106
synthesis 1 ATGACCAGCGGAAGAAGCGCCTGCTGCCCTGGCTGCTGGCTCAGGCACAGAGCGGAAGATTCCCGGGGCTGGAGTTTGATGACGCAGCTAGGTCGGCACTGCGCG 106

duck IRF9 107 TGGCCTGGGAGCGCGCGGAAGGGGCGGGGCGGGGGCGGGGCGCGCGCGCCTGTGCCAGGCGTGGGCGGAGTACAAAGGGGCCACGCCCGCGCGGCGCGCG 212
synthesis 107 TGCCCTGGGAAAGGGCAGGACGCGGCGGAGCGGGGGCGGAGCAGCAGCTGCAGTGTGCCAGGCATGGGCTGAGTACAAGGGAGCACTCCACCTCCAGGACCAGC 212

duck IRF9 213 CGTCTGCAAGACCAGGCTGCGCTGCGCCCTGCACAAAGAGCCCCGAGCTGCAGGAGGTGCCGAGCGCGCCGCTGGAAGGGCCCGCCCTACAAAGGTGTACCG 318
synthesis 213 TGTGTGCAAGACCAGGCTGAGATGCGCCCTGCACAAAGAGCCAGAACTGCAGGAAGTGCCGAAAGGGCTAGGCTGGATGGACCAAGGCCTACAAAGGTGTACAGA 318

duck IRF9 319 CTGCTGGGCGCACGCCACCAACGCAACGGCCCGCCCGCGCGGGCTCCAAGGGGCGCAAGGAGGAGCGGAGCCACGCGCGATGACGCGACCCGCCCCCA 424
synthesis 319 CTGCTGGGACCAAGGCCACCTAGAAACGGACAGCACCAGGAGAGGAAGCAAGGGCAGGAAGAGGAAGCAGAGCCTCGCGGAGATGACGCAACACCACTCA 424

duck IRF9 425 GCTCTGAGGAGGACGAGGAGGAGGCGGAGCCGAGCCCGCCCGCCCTCCGCTGCTGTGACCCCGCCCTTCCCGCGAGCTCGCCATCCTATTGGACAGCCC 530
synthesis 425 GCTCGAGGAAGACGAGGAGGAGGCTGAACCTTCCCGAGCACCACCTCCAGCAGCTGTGACTCCACCTCCAAGCAGACAGCTGGCTATCCTGCTGGATTCCCG 530

duck IRF9 531 GGAGCCCTCCCCAGCGAAGCGCAGCGGCAATTGGCGCTCTCCCTCTGGGTGGGCGGGGCTGGCTGGAGGGCGTGGCTCCCGCGGGCGAGTACCTCTA 636
synthesis 531 AGAGCCTCTGCCACCTGCAAGAAGGAGACGACGACTGGGAGTGAGCCTGTGGGTGGGCGGCGCACTGGCATGGAGGGCATGGCTCCAGCAGGAGAGTACCTGCTG 636

duck IRF9 637 TTGGCGCGGGCAGCCCGCCCGCGCGCCCTTGCCCGGCTCTGGTGCCGGCCCGCCCGCCCGCGCCCTGGAGGGCTGGGGGCGGGGCTGGGCGGG 742
synthesis 637 CTGGCTGCAGGATCCCCAGCTCCAGGACCTGCACCACTGCCAAGACTGCTGGTGCCCGCTCTCCACCTCCAGCACTGGAAGCACTGGAGCTGGACTGGGAAGG 742

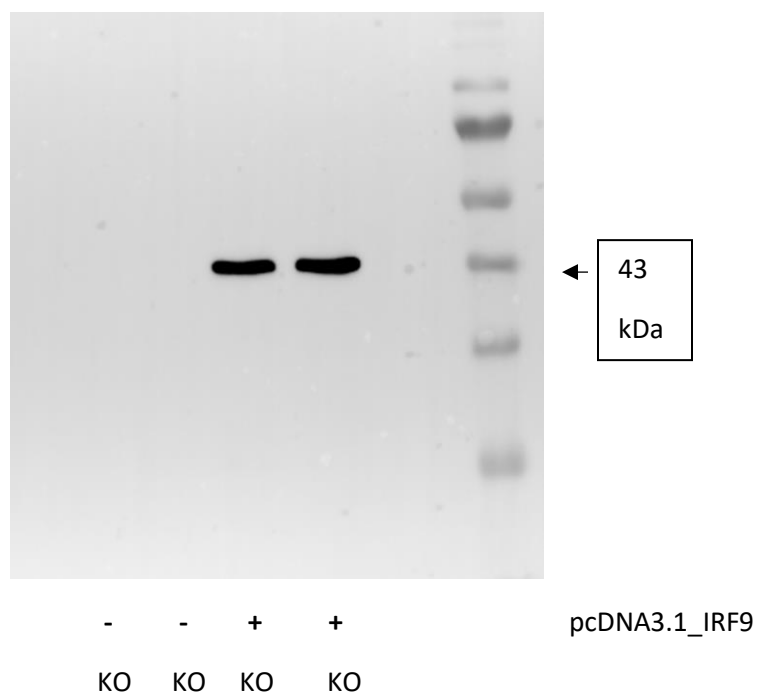
duck IRF9 743 GGTGCTGGTGCCAGCGGGGGCGGGGCTTTCTCTAGGACACGCCCCGGGGGCGGGGCTCGCCCTCAGCGCCCCACGCGGGGAGCCGGGGGCGGGG 848
synthesis 743 GACTGCTGGTGCAGCGGAGGAGAGGACTTTCTGAGGACAGAGCCAGGGGAGGGGACTGGCACTGTCCGCCCTCACGCTGGAGAGCCAGGCGGAGGGG 848

duck IRF9 849 CTGGCCAAGGGGAGGAGCCACTGGTGGCCGCTTCGACAGCGGGCGCTTCCGGGAGGAGCTGCAGCGGACCGCCAGGGGCTGGGCCCCCGCCGAGCACCG 954
synthesis 849 ATGGCCCAAGGGCGAAGAGCCTCTGGTGGCCGCTTCGATAGCAGGCGCTTAGGGAAAGCTGCAGAGGCACAGACAGGGAAGTGGGACCACCTCCAGAGCACAGG 954

duck IRF9 955 GTGCTGCTGGCGGCCACCCCGGGGAGCTCCGACGGGGGCGGGGCGAGCGCACAGGGGCGGTGCTACAGCTGGAGCAGGCCCTGGCCAGCGGCTCTGGGCC 1060
synthesis 955 GTGCTGCTGGCAGCAACCCAGGAGAACTGAGGACAGGCGGAGGGGCTTCGACAGGGAGCTGTGCTGACAGCTGGAGCAGGCACTGGCTCAGAGGCTGCTGGGAC 1060

duck IRF9 1061 CCGCGCGGATCCCGCAATTAA 1083
synthesis 1061 TGCACAGACCACTAATTGA 1083

Fig. S5 Pairwise alignment of IRF9 coding sequences of endogenous duck IRF9 with in vitro synthesized construct. The predicted protein coding sequence is identical for both. In the gene synthesis construct, the codons were optimized to break GC-rich stretches. Mismatches in the nucleotide sequence are highlighted in gray color.



(ad Fig. 7B)

Coomassie blue, 12% gel

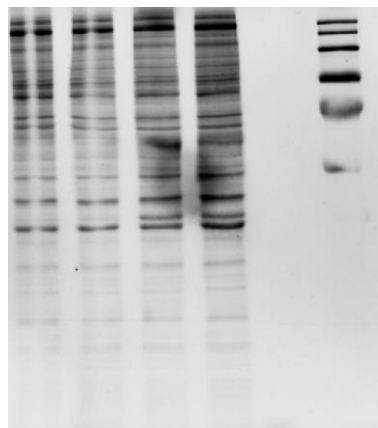


Fig. S6 Original gel and blot images. Western blot of lysate from duck KO cells transiently transfected by the dIRF9 expressing plasmid and corresponding Coomassie blue gel from figure 7B are shown.