Evolution of IFN subgroups in bony fish - 2. Analysis of subgroup appearance and expansion in teleost fish with a focus on salmonids.

Fuguo Liu[a], Tiehui Wang[a], Jules Petit[b], Maria Forlenza[c], Xinhua Chen[d,e], Liangbiao Chen[f], Jun Zou[a,e,f] and Christopher J. Secombes[a,*]

[a]*Scottish Fish Immunology Research Centre, School of Biological Sciences, University of Aberdeen, Aberdeen, AB24 2TZ, Scotland, UK*
[b]*Wageningen University & Research, Aquaculture and Fisheries Group, Department of Animal Science, 6708WD Wageningen, The Netherlands*
[c]*Wageningen University & Research, Cell Biology & Immunology Group, Department of Animal Science, 6708WD Wageningen, The Netherlands*
[d]*Key Laboratory of Marine Biotechnology of Fujian Province, Institute of Oceanology, Fujian Agriculture and Forestry University, Fuzhou 350002, China*
[e]*Laboratory for Marine Biology and Biotechnology, Qingdao National Laboratory for Marine Science and Technology, Qingdao, China*
[f]*Key Laboratory of Exploration and Utilization of Aquatic Genetic Resources, Ministry of Education, Shanghai Ocean University, Shanghai, 201306, China*

* Corresponding author
E-mail address: c.secombes@abdn.ac.uk

Key words: Type I interferon, evolution, teleosts, salmonids.

Abstract

A relatively large repertoire of type I interferon (IFN) genes is apparent in rainbow trout/Atlantic salmon, that includes six different IFN subgroups (IFNa-IFNf) belonging to the three known type I IFN groups (1-3) in bony fish. Whether this is true for other salmonids, and how the various type I subgroups evolved in teleost fish was studied using the extensive genomic resources available for fish. This confirmed that salmonids, at least the Salmoninae, indeed have a complex (in terms of IFN subgroups present) and large (number of genes) IFN repertoire relative to other teleost fish. This is in part a consequence of the salmonid 4R WGD that duplicated the growth hormone (GH) locus in which type I IFNs are generally located. Divergence of the IFN genes at the two GH loci was apparent but was not seen in common carp, a species that also underwent an independent 4R WGD. However, expansion of IFN gene number can be found at the CD79b locus of some perciform fish (both freshwater and marine), with expansion of the IFNd gene repertoire. Curiously the primordial gene order of GH-IFNc-IFNb-IFNa-IFNe is largely retained in many teleost lineages and likely reflects the tandem duplications that are taking place to increase IFN gene number. With respect to the evolution of the IFN subgroups, a complex acquisition and/or loss has occurred in different teleost lineages, with complete loss of IFN genes at the GH or CD79b locus in some species, and reduction to a single IFN subgroup in others. It becomes clear that there are many variations to be discovered regarding the mechanisms by which fish elicit protective (antiviral) immune responses.

1. Introduction

Interferons (IFN) exist in all extant Gnathostome vertebrates, and function as a key component of the antiviral defences. Three types (I-III) of IFN are broadly recognized, with type III apparently lost in bony fish [1]. Type II IFN have remained present in the genomes of all jawed vertebrates but in teleost fish have been expanded, likely as a result of tandem gene duplication at the IFN-γ locus, to include a related gene called IFN-γ-rel [2]. In contrast, type I IFNs are highly diverse in terms of the groups/subgroups and copy number present in different vertebrate groups and species. All of these IFNs have relatedness to the IL-10 family of cytokines (i.e. class II cytokines), and appear to have evolved from a primordial class II cytokine gene that gave rise to the IL-10 cytokines and an IFN type I/III precursor, with the latter subsequently diverging into the type I and III IFNs [3].

Some IFN genes may have separated early from the ancestral type I IFN, giving rise to distinct lineages that have been expanded or lost during vertebrate evolution. For example, three groups (1-3) of type I IFN genes are known in the ray finned fish, but group 3 genes (also called IFNf) appear to have evolved quite early and are also found in cartilaginous fish and amphibians [1]. In the ray finned fish a putative group 1/2 IFN ancestor evolved that had diverged into distinct group 1 and group 2 genes by the appearance of the Chondrostean fish (eg sturgeon). Hence these fish possess 3 groups of type I IFNs; group 1 represented by IFNe, group 2 by IFNb and group 3 by IFNf [4,5]. Diversification of the group 2 IFNs into two subgroups (ie IFNb and IFNc) is apparent in Holosteans (eg gar) [5], whilst further expansion of the group 1 IFNs into additional subgroups (IFNa, IFNd, IFNh) has occurred in teleost fish [6,7]. This further expansion of group 1 genes in the teleost fish lineage could potentially be linked to the teleost specific whole-genome duplication (3R/TS-WGD) event, which generated two IFN loci [8], that are referred to below as linked to growth hormone (GH) or CD79b. However, subsequent expansion or even loss of these subgroups seems to have happened in a lineage-specific fashion within teleosts. In this second of two papers looking at IFN evolution in ray-finned fish, we examine these issues.

Past studies of the IFN groups/subgroups in teleost fish suggest that salmonids (rainbow trout, Atlantic salmon) have the largest IFN repertoire; not only in terms of the groups/subgroups that they possess but also in the number of genes present [6,9]. However, this statement has been based on BAC clone analysis and to date the salmonid genomic loci have not been defined/described. With an increasing number of teleost genomes available to interrogate, in this study we revisit this finding to verify if this is true for other Protacanthopterygian species. We have analysed a variety of salmonid species (i.e.- rainbow trout, Atlantic salmon, chinook salmon, coho salmon and Arctic charr) that have undergone a 4R WGD event, as well as Northern pike that have not, to see if the mechanism(s) by which IFN gene expansion has occurred is influenced by WGD. In addition, we have analysed the type I IFN genes, subgroups and loci present in a variety of other teleost fish groups (Elopomorpha, Osteoglossomorpha,

Ostariophysi, Paracanthopterygii, Acanthopterygii), to give a broader view of subgroup expansion in teleosts, especially of the group 1 IFNs since only a single subgroup (IFNe) appears to have been present prior to the emergence of this infraclass [4,5]. This included a species (common carp) that has undergone an independent 4R WGD event. Our findings show that salmonids, at least the Salmoninae (one of three salmonid subfamilies), indeed have a complex IFN repertoire relative to other teleost fish. This is in part a consequence of the salmonid WGD that duplicated the growth hormone (GH) locus. However, divergence of the IFN genes at the two GH loci was apparent and was not seen in common carp. Interestingly, expansion of IFN gene number was found at the CD79b locus of some perciform fish, where the IFNd gene repertoire has increased.

2. Materials and Methods

2.1 Teleost fish genomes

Currently, the genomes or whole genome contigs of many fish species are available at the National Centre for Biotechnology Information (NCBI: https://www.ncbi.nlm.nih.gov/) or Ensembl (https://www.ensembl.org/index.html) databases. They include a good coverage of different teleost superorders, such as the Elopomorphs, Osteoglossomorpha, Ostariophysi, Protacanthopterygii, Paracanthopterygii and the Acanthopterygii. These available genome sequences can facilitate the identification and evolutionary analysis of fish type I IFN. In this study we focused initially on salmonid species, including rainbow trout (*Oncorhynchus mykiss*), chinook salmon (*Oncorhynchus tshawytscha*), coho salmon (*Oncorhynchus kisutch*), Atlantic salmon (*Salmo salar*) and Arctic charr (*Salvelinus alpinus*). We then analysed other species within the above mentioned superorders, including species to allow a comparison of the impact of a 4R WGD in relation to 3R relatives within the Ostariophysi and Protacanthopterygii. The species analysed included the Japanese eel (*Anguilla japonica*), Asian bonytongue (*Scleropages formosus*), Northern pike (*Esox lucius*), nile tilapia (*Oreochromis niloticus*), common carp (*Cyprinus carpio*), cod (*Gadus morhua*), haddock (*Melanogrammus aeglefinus*), olive flounder (*Paralichthys olivaceus*), turbot (*Scophthalmus maximus*), large yellow croaker (*Larimichthys crocea*), tetraodon (*Tetraodon nigroviridis*), medaka (*Oryzias latipes*), seabass (*Dicentrarchus labrax*), the white-blooded icefish (*Chaenocephalus aceratus*) that lacks hemoglobin in its blood and the cold-adapted Antarctic toothfish (*Dissosticus mawsoni*). Whilst some of the IFN genes present in these species have been published previously (Lutfalla et al. [10] (tetraodon); Casani et al. [11] (sea bass); Kitao et al. [12] (carp); Pereira et al. [13] (turbot); Maekawa et al. [14] (medaka); Hu et al. [15] (flounder); Ding et al. [16] (croaker); Huang et al. [17] (Japanese eel)), the exact number of each subgroup present and their genomic location were not typically available. Data for the IFN loci/genes in zebrafish (*Danio rerio*) and stickleback (*Gasterosteus aculeatus*) were already available and included without further analysis [8].

2.2 In silico identification of fish IFN genes

The fish IFN genes were obtained by tBLASTn against the fish genome database using

previously published IFN sequences (e.g. Zou et al. [6]). The identified IFN sequences were then recorded according to their positions in the genome. The genomic DNA sequences that partially matched the IFN sequences were also recorded and analysed by the GenScan program [18] or by Splign (https://www.ncbi.nlm.nih.gov/sutils/splign/splign.cgi). ExPASy-translate (https://web.expasy.org/translate/) was used to determine whether the predicted sequences could be correctly translated. The predicted transcripts were also confirmed by BLASTp search using default parameters on the NCBI website (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins&). The accession numbers of identified IFN genes are listed in **Tables S1-S15**, but when no accession number was available we have provided the predicted sequences in supplementary **Figures S8-S56**. Subsequently, alignment of protein sequences using Clustal Omega was performed to sort out any wrongly annotated IFN sequences, which were then re-predicted by GenScan program or by Splign. Due to the low identities of IFN genes between different IFN subgroups and among fish species, the queries used in BLAST search varied a lot, e.g. IFNh of large yellow croaker was used to search the IFNh genes in other fish, and 4 published Japanese eel IFN genes [17] were used to predict the additional IFN genes in the genome of Japanese eel. The synteny between the type I IFN loci was predicted using the Genomicus program (database version 96.01) or information extracted from recently released genomes or whole genome contigs at NCBI or Ensembl databases, with a focus on identifying linkage to GH and CD79b, to confirm the evolutionary changes occurring at particular loci.

2.3 Phylogenetic tree analysis of fish IFN genes

A series of phylogenetic trees were generated to verify the IFN subgroups present in different fish species and to understand the evolution of fish IFN genes. These included a salmonids IFN phylogenetic tree, salmonid and pike IFN phylogenetic tree, and a teleost fish IFN phylogenetic tree. Phylogenetic trees were constructed by the Neighbour-joining method using the MEGA7.0 program on full-length amino acid (aa) alignments and bootstrapped 1,000 times. The evolutionary distances were computed using the JTT matrix-based method with all ambiguous positions removed for each sequence pair.

2.4 Terminology

Having identified the IFN gene repertoires, it was clear that a large number of IFN genes are present in some lineages/loci. So we have introduced a terminology to name the genes by IFN subgroup, followed by locus (with the GH locus/loci numbered first) and then gene number for the locus being described (ie IFNa1.1, a1.2, b1.1, etc). In addition, where a gene was fully identified but there was a premature stop codon, it was termed a pseudogene (pIFN), and the subgroup designation was given. If only part of a gene was found (ie several exons), usually due to incomplete sequencing (ie multiple N's), then it was reported in our synteny analysis but the subgroup designation was not always possible to ascribe.

2.5 Sequence analysis

Protein translation was performed using Virtual Ribosome-version 2.0. Identity and similarity analysis were performed using the matrix BLOSUM62 within the MatGAT program [19], with a gap open penalty of 10 and gap extension penalty of 1. Multiple aa alignment was performed using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) and the conserved aa were shaded using

the BoxShade program (https://embnet.vital-it.ch/software/BOX_form.html).

3. Results/Discussion

Seven type I IFN subgroups are known in teleost fish, that were named as discovered; IFNa-f and IFNh [7,8], with IFNg avoided to prevent confusion with IFN-γ, a type II IFN. Whilst IFNa-f are present in the salmonids (e.g. rainbow trout), only IFNa, c and d have been found in cyprinids [8,12,20], although in black carp IFNc (see **Fig. 7**) has been described as IFNb [21]. In percomorphs initially IFNd was discovered [11,22,23], followed by the new subgroup IFNh [7], but most recently it has become apparent that three subgroups are present in some perciform species, namely IFNc, IFNd and IFNh [16,24,25]. This may be true in other percomorph orders since olive flounder (Pleuronectiformes) also possess these three subgroups [15] and turbot have an IFNc and IFNh gene [13], so most likely will have IFNd in common with all other Acanthopterygian species studied to date. Whilst medaka are reported to have an IFNa and IFNd gene [14], we found that the IFNa is in fact IFNh (and there are multiple IFNd genes – see **Fig. 7**), and hence is in line with the above. The discovery of the IFNe and IFNf subgroups in salmonids initially led to the hypothesis that these could be salmonid-specific IFNs. However, this was quickly dispelled with the realization that IFNf is in fact an ancient IFN also present in cartilaginous fish [1], and that IFNe genes were present in Chondrostean and Holostean fish [4,5], and therefore these subgroups were likely lost in particular teleost lineages. Nevertheless, a relatively large repertoire of IFN genes is apparent in rainbow trout/Atlantic salmon. Whether this is true for other salmonids, potentially influenced by the 4R WGD event in this lineage, and more generally how the IFN subgroups evolved in teleost fish warrants further analysis. This was undertaken here using the extensive genomic resources available for fish.

3.1 What happened post-genome duplication in salmonid species?

To understand the impact of the 4R WGD in salmonids on IFN diversity, we have analysed the IFN loci in five salmonid species (rainbow trout, chinook salmon, Atlantic salmon, coho salmon, Arctic charr) and in Northern pike. In pike, as with other 3R teleost species, there are two IFN loci, one linked to GH and one linked to CD79b (**Fig. 1**). A single IFNd gene is present at the CD79b locus, whilst at the GH locus 12 IFN genes were found, with subgroups verified by phylogenetic tree analysis as 3x IFNa, 1x IFNb, 5x IFNc, 2x IFNe and 1x IFNf (**Figs. 1** and **2**). Therefore it is apparent that salmonids are not the only teleost species to possess 6 IFN subgroups, and that the GH locus expanded prior to the 4R WGD. In salmonids two GH-linked IFN loci were found in all species (**Figs. 3** and **4**). The first GH locus (locus 1) looked quite similar to the pike locus, in that 4 IFN subgroups are present, with multiple IFNc and a single IFNf (**Fig. 3**). However, only a single IFNa and IFNe exist at this locus in salmonids, where 3 or 2 genes are present, respectively, in pike. IFNb is also present at this locus but as one (or two) pseudogene(s), with the exception of charr where no IFNb could be

identified. This probably reflects the fact that the genome assembly is not as good in charr. Indeed two different scaffolds were included in the analysis; one linked to GH and a second where the IFN subgroups and gene number (1x IFNa, 1x IFNe and 1x IFNf) suggested it was part of locus 1, especially as the IFNa and IFNe genes clustered with the respective Atlantic salmon genes from this locus. At the second GH locus (locus 2) all five subgroups were found (**Fig. 4**), except for charr which again apparently lacked IFNb, but now with multiple IFNa, IFNb, IFNe and IFNf as well as multiple IFNc present. In comparison to the 12 genes present in pike at the GH locus, the number of IFN genes at this second salmonid GH locus ranged from 15-17 genes in chinook salmon, Atlantic salmon and coho salmon, to 28 genes in rainbow trout. The number of IFNe in particular was greatly expanded in rainbow trout at this locus. Whilst the Arctic charr had relatively few IFN genes (7) at this locus the genome assembly was probably not sufficiently robust to allow detection of all genes present. Three charr scaffolds were included in the analysis, one linked to GH and two where the subgroups/gene number present suggested the IFN genes detected are from locus 2 by comparison to the other salmonids studied (see **Fig. 4** legend). Lastly, only a single CD79b locus was identified, that was linked with a single IFNd gene in each species (**Fig. 5**).

The total number of IFN genes present in chinook, Atlantic and coho salmon is close to double the number present in pike, which might be predicted due to the duplication caused by the salmonid 4R WGD. However, the loci are not identical to pike and in general there is a small reduction of IFN genes at locus 1 and a small expansion at locus 2. The number of IFN genes identified at the second GH locus in rainbow trout seems exceptional, but perhaps also reflects a better quality genome being analysed. Only resequencing through this region for the other species will confirm if more IFN genes are present at GH locus 2. Indeed, it should be noted that a large number of IFN pseudogenes and IFN partial sequences were detected at the GH loci in salmonids (**Figs 3** and **4**). This might be as expected for sites of high gene birth and death [1,6] but perhaps some will prove to be transcribed genes in future analysis.

One of the most interesting findings was that the two GH loci do seem to be diverging. This is evidenced by 1) the loss of IFNb genes at locus 1, where only a pseudogene is now present, 2) the major expansion of IFNe genes at locus 2, and 3) the divergence of IFNc genes between locus 1 and locus 2, as seen in the phylogenetic tree analysis (**Fig. 2**) and aa alignments (**Suppl Fig. 3**). The latter can be seen in trout IFNc1.1, as a representative molecule of the IFNc at locus 1, where aa 19 (F), 34 (T), 93 (T), 99 (M), 107 (Y), 171 (E), 175 (K) and 184 (S) are different to the IFNc equivalent aa at locus 2. With the other IFN subgroups a high degree of sequence conservation was apparent (**Suppl Figs1, 2, 4-6**), despite the large increase in gene number in some cases. However, a divergence from the pike sequences was seen, with the N-terminal sequence of IFNc showing the greatest difference.

In relation to the previously published BAC sequence analysis in rainbow trout [6], it was difficult to find exact congruence of the data. However, with the multiple IFNe

present in Clones RT282J16 and RT303F02 it is clear they are from GH locus 2, with regions in the above analysis containing 3x IFNe and an IFNa or 3x IFNe and an IFNf in agreement with these BAC clones. BAC clone RT292E06 was more difficult to place but again appeared to be from GH locus 2, since there are more IFNb and IFNc genes (3 and 4 respectively) than found by genomic analysis of GH locus 1. The region immediately downstream of GH at this locus also contains IFNa and IFNf genes, as observed in clone RT292E06.

Altogether, it is clear from the salmonid IFN loci analysis that the 4R WGD generated two GH loci, although only a single CD79b locus appears to have been retained. So the total number of IFN genes present is approximately double, or has been expanded further in the case of rainbow trout (at GH locus 2). Divergence between the number of genes per subgroup is also apparent, as seen with the two GH loci. The salmonids examined are all members of the Subfamily Salmoninae, and therefore it is not impossible that a different scenario will be found in species belonging to other subfamilies (Coregoninae and Thymalinae). Indeed, future analysis of other species within the Salmoninae, such as the Danube salmon, and Thymalinae (e.g. Grayling) may also help confirm whether the large IFN repertoire is associated with anadromy (all the salmonid species examined here are anadromous), or whether it is a subgroup or salmonid wide phenomenon.

3.2 Does genome duplication per se result in IFN gene expansion?

From the above findings in salmonids, the question remains as to whether WGD has contributed to IFN gene number and loci divergence at other times during teleost evolution. One comparison that can be made to answer this question is to look at the IFN genes in gar [5], a Holostean ray-finned fish, compared to a basal teleost such as Japanese eel [26] that has undergone the teleost wide 3R WGD [27]. In gar we have previously identified an IFN locus linked to both GH and CD79b that contains 7 IFN genes (1x IFNb, 4x IFNc, 2x IFNe), and a separate scaffold (that cannot be linked currently) that contains an IFNf gene [5]. The Japanese eel was studied recently by Huang et al. [17], where five putative IFN genes were found at a single locus linked to GH, with four verified by cDNA sequencing. These genes included 1x IFNa, 1x IFNb, 2x IFNc and 1x IFNe. Our analysis of the eel genome discovered an additional IFNb, IFNc and a partial sequence for an IFNa gene at the GH locus, as well as an IFNc and IFNf at the CD79b locus (**Figs. 6** and **7**). Finally we discovered 2x IFNf on a separate scaffold that is likely linked to one of these loci, but it was not clear which (**Fig. 6**). This helps confirm that following the 3R WGD two loci were generated in early teleosts, as postulated from studies of zebrafish and stickleback [8], with one linked to GH (with CD79b lost) and one linked to CD79b (with GH lost). It is possible that IFNc and IFNf are present at both, depending on where scaffold 364684 eventually links (**Fig. 6**). However, IFNa, IFNb and IFNe are present at only the GH locus. Thus the eel GH locus looks quite similar to the single gar IFN locus, in having 7-9 IFN genes (depending on where the IFNf will be located) vs 7-8 genes in gar, with IFNb, IFNc and IFNe genes

present in both species. The CD79b locus has a reduced IFN/subgroup number, with only a single IFNc and 1-3 IFNf. Hence, whilst the 3R WGD resulted in two IFN loci, the number of genes and subgroups has only expanded marginally in the Elopomorphs. However, as will be outlined below, this is actually a unique situation in terms of the eel CD79b locus, where in all other studied teleosts IFNd genes are exclusively located at this site. To see if any other basal teleosts may have similar IFN loci, we also examined the genome of the Asian bonytongue, as a representative of the Osteoglossomorpha [26]. Again two IFN loci were found linked to GH or CD79b (Fig. 6), but with only a single IFNa and IFNb at the GH locus and a single IFNc at the CD79b locus. This suggests that IFNe and IFNf has been lost in these fish as Elopomorphs are considered more ancient, and again shows retention of an IFNc at the CD79b locus in basal teleosts.

Another comparison that can be made is between 3R cyprinids such as zebrafish, with 4R cyprinids such as the common carp (**Fig. 8**). It is known that zebrafish have two loci, with 1x IFNa and 2x IFNc at the GH locus and 1x IFNd at the CD79b locus (Boudinot et al. [8] – see **Fig. 8** for reference to phi terminology for these genes). Our analysis of the carp genome has confirmed that these loci are duplicated exactly in carp, giving two GH loci each with 1x IFNa and 2x IFNc, and two CD79b loci with a single IFNd gene (**Fig. 8**). There has been no gene loss or gain at the loci, but clearly the number of IFN loci and gene number has doubled. However, it should be noted that the 4R WGD in carp was more recent than the salmonid 4R WGD, and was an allotetraploidization event vs the autotetraploidization that occurred in salmonids, and these differences may have impacted the above findings.

Thus it is apparent that genome duplication has indeed increased the number of IFN loci in teleosts. However, gene loss, gene gain or no change can occur at the duplicated loci, with loss of entire loci also possible (as seems to have occurred with one of the salmonid CD79b loci).

3.3 When did the IFN group 1 subgroups appear?

In sturgeon (Chondrostean) and gar (Holostean) only a single type of group 1 IFNs is present, IFNe [4,5]. However, already in eel representing an early teleost group (Elopomorphs) a second group 1 subgroup is apparent, IFNa (**Fig. 6**), and this is also the case in bonytongues (Osteoglossomorpha). It is found at the GH locus and hence is likely derived from IFNe. IFNa is also found in the cyprinids and salmonids but appears to be lost in the neoteleosts, as is not present in gadoids and percomorphs (see below). IFNe is also lost in these groups, and is even absent in the cyprinids analysed to date, and so could have been lost independently on several occasions. Once more teleost genomes are available to interrogate the timing of these events should become clearer. Similarly, IFNf has been lost alongside IFNe, and from both loci, since IFNf is present at the CD79b locus in Japanese eel (see **Fig. 6**). However, further group 1 subgroups have appeared in these fish. In all Euteleosts and Otocephala examined to date, IFNd is

present at the CD79b locus. It is not clear how it has arisen, since no other group 1 genes are present at the CD79b locus in eels and bonytongues, that have only group 2/IFNc (in both) and group 3/IFNf (eels) genes. However, IFNe could have been present at both loci following the 3R WGD, and so perhaps IFNd was derived from IFNe later in teleost evolution, but that loss of IFNe occurred at the CD79b locus in Elopomorpha and Osteoglossomorpha. Indeed, in the phylogenetic tree of the salmonid and pike IFN molecules (that include the vast majority of the IFNe genes known), it does suggest that IFNe is basal to both IFNa and IFNd, in support of this hypothesis (**Fig. 2**).

Another group 1 subgroup that has emerged is IFNh, initially discovered in the percomorphs [7]. In our examination of several percomorph species (turbot, tetraodon, large yellow croaker, tilapia, sea bass, stickleback) it is apparent that IFNh is present, or as a partial sequence, at the GH locus (**Figs. 9** and **10**, **Suppl Fig. 7**). This linkage was not able to be verified in medaka or flounder (**Suppl Fig. 7**), but it seems likely that the scaffolds/genes shown will eventually be found to be linked. Whether IFNa or IFNe gave rise to IFNh is less clear but this would be the most likely origin. Curiously, we have also found IFNh in gadoids (cod, haddock), confirmed to be at the GH locus in cod alongside IFNb (**Figs. 7** and **10**, **Suppl. Fig. 7**). This shows that this subgroup emerged earlier, and was present in neoteleosts before the divergence of the Paracanthopterygii and Acanthopterygii. In the case of haddock, two scaffolds were found with 1x IFNh and 1x IFNb respectively, and so in comparison to cod we predict the haddock genes will be linked to GH.

A model of the appearance (and loss) of IFN subgroups during teleost evolution is presented below.

3.4 Can expansion of the CD79b locus occur?

The CD79b locus seems to have reduced to a single gene quite early in teleost evolution, as a single IFNc in Osteoglossomorpha or a single IFNd in the Otocephala and Euteleosts, as seen in the Ostariophysii (eg cyprinids) and Protacanthopterygii (eg esociformes and salmoniformes). However, there is evidence that the IFN genes at this locus have also been expanded later in teleost evolution, as seen in the Percomorphs. In some species, such as turbot, flounder, stickleback, tetraodon, medaka, ice fish, toothfish and large yellow croaker 2-4 IFNd genes are present (**Fig. 9**, **Suppl Fig. 7**), and in some cases (tetraodon, icefish/toothfish) this is the only IFN subgroup present (with no functional IFN genes at the GH locus). However, in species such as tilapia and seabass major expansion of the CD79b locus has occurred with 12-18 IFNd genes present (**Fig. 10**). In terms of the mode of gene duplication occurring, en bloc duplication seems to be a common theme. For example, three linked blocks are identifiable in tilapia that form a single clade (IFNd2.1-2.6) in the phylogenetic tree, and six continuous blocks (IFN2.7-2.19) are present downstream, such that each block has a gene/genes that belong to two independent clades (**Figs. 7** and **10**). Similarly, en bloc duplication may have occurred at the salmonid IFN locus 2 (**Fig. 5**). In contrast to

these perciform fish, in cod (that was also examined in this study) the CD79b locus had no detectable IFN genes present. Similarly, no IFNd genes could be found in haddock, suggesting IFNd has been lost in gadoids/ Paracanthopterygii (**Fig. 10**).

So precedents exist that show expansion of IFN genes at the CD79b locus, as seen in some perciform species.

4. Conclusion

This analysis has confirmed that salmonids, at least the Salmoninae, indeed have a complex (in terms of IFN subgroups present) and large (number of genes) type I IFN repertoire relative to other teleost fish. Whilst 6 IFN subgroups were already present in pike, the salmonid WGD gave rise to a second GH locus substantially increasing the number of IFN genes. The IFN genes at these two GH loci are clearly diverging, with expansion of several group 1 genes (IFNa, IFNe) particularly apparent in rainbow trout. In contrast the WGD event in cyprinids has not driven (as yet) a comparable gene loss or gain, although the loci are duplicated, thus effectively increasing IFN gene number. The salmonids have also been shown to have a large number of (IFN induced) Mx genes [28,29], and hence the antiviral defences in these fish is likely augmented at several levels, perhaps reflecting their anadromous life cycle. However, expansion of IFN gene number can be found at the CD79b locus in some perciform fish (both freshwater and marine), with expansion of IFNd genes, which is most intriguing. That these loci are sites of high gene gain and loss is also apparent from the large number of pseudogenes present, independently of whether this occurs at the GH loci in salmonids or the CD79b locus in perciformes. Curiously the primordial gene order of GH-IFNc-IFNb-IFNa-IFNe is largely retained in many teleost lineages and likely reflects the tandem duplications that are taking place to increase IFN gene number.

With respect to the evolution of the type I IFN subgroups, a complex acquisition and/or loss has occurred in different teleost lineages, as illustrated in **Figure 11**, with complete loss of IFN genes at the GH or CD79b locus seen in some species, and even reduction to a single IFN subgroup. The evolutionary pressures leading to IFN reduction or expansion will be important to establish, to understand more fully how antiviral defences adapt to different life history traits. For example, gadoids possess a single IFNb and IFNh, but have lost their Mx genes [30] as well as other immune molecules [31,32] yet are able to produce a clear antiviral response following viral infection [33] or stimulation with poly I:C [34]. Some evidence for IFN subgroup functional diversification exists, mainly in the relatively well studied salmonid IFN genes. In rainbow trout, IFNa transcripts can undergo alternative splicing to generate intracellular IFNs that may have a selective advantage [35]. Furthermore, analysis of subgroup induction following viral infection has shown some subgroups are induced rapidly but not substantially, whereas others (especially group 2 genes) can be highly upregulated later in the response [6]. These group 2 IFN genes are apparently highly (co)expressed by a discrete cell population in salmon [36], rather similar to the situation in mammals

with IFN production by plasmacytoid dendritic cells [37,38]. There may also be functional divergence between the group 1 and group 2 IFN molecules in terms of receptor signalling, as seen in zebrafish where these two IFN groups have been shown to signal via different receptors [39]. It is interesting to see that group 2 genes (unlike group 3 IFNf) have been retained through to the perciforms, although loss of IFNb or IFNc has happened in different lineages. Nevertheless some perciform species have lost the group 2 genes, and so it is certainly possible to survive without them! As exemplified by the unusual immune system present in gadoids, it is clear there are many variations to be discovered regarding the mechanisms by which fish elicit protective (antiviral) immune responses.

Acknowledgements

FL was supported by a Newton International Fellowship funded by the Academy of Medical Sciences, UK (AMS, NIF004\1036). Thanks go to Mingli Liu (Shanghai Ocean University) for help with the bioinformatics analysis of the Icefish/Toothfish, and to Drs Dan Macqueen and Manu Gundappa (Roslin Institute, University of Edinburgh) for helpful discussions and advice on the analysis.

References

[1] A.K. Redmond, J. Zou, C.J. Secombes, D. Macqueen, H. Dooley, Discovery of all three types in cartilaginous fishes enables phylogenetic resolution of the origins and evolution of interferons. Frontiers in Immunology 10 (2019) 1558.

[2] C.J. Secombes, J. Zou, Evolution of interferons and interferon receptors. Frontiers in Immunology 8 (2017) 209.

[3] P. Siupka, O.J. Hamming, M. Fretaud, G. Luftalla, J.P. Levraud, R. Hartmann, The crystal structure of zebrafish IL-22 reveals an evolutionary, conserved structure highly similar to that of human IL-22. Genes and Immunity 15 (2014) 292-302.

[4] Q. Xu, K. Luo, S. Zhang, W. Gao, W. Zhang, Q. Wei, Sequence analysis and characterization of type I interferon and type II interferon from the critically endangered sturgeon species, *A. dabryanus* and *A. sinensis*. Fish & Shellfish Immunol. 84 (2019) 390-403.

[5] F. Liu, N.C. Bols, P.H. Pham, C.J. Secombes, J. Zou, Evolution of IFN subgroups in bony fish - 1 : Group I-III IFN exist in early ray-finned fish, with group II IFN subgroups present in the Holostean spotted gar, *Lepisosteus oculatus*. Fish Shellfish Immunol. 95 (2019) 163-170.

[6] J. Zou, B. Gorgoglione, N.G.H. Taylor, T. Summathed, P.-T. Lee, A. Panigrahi, C. Genet, Y.-M. Chen, T.-Y. Chen, M. Ul Hassan, S.M. Mughal, P. Boudinot, C.J. Secombes, Salmonids have an extraordinary complex type I interferon system: Characterisation of the IFN locus in rainbow trout *Oncorhynchus mykiss* reveals two novel IFN subgroups. J. Immunol. 193 (2014) 2273-2286.

[7] Y. Ding, J. Ao, X. Huang, X. Chen, Identification of two subgroups of type I IFNs in Perciforme fish large yellow croaker *Larimichthys crocea* provides novel insights into function and regulation of fish type I IFNs. Front. Immunol. 7 (2016) 343.

[8] P. Boudinot, C. Langevin, C.J. Secombes, J.-P. Levraud, The peculiar characteristics of fish type I interferons. Viruses 8 (2016) 298.

[9] B. Sun, B. Robertsen, Z. Wang, B. Liu, Identification of an Atlantic salmon IFN multigene cluster encoding three IFN subtypes with very different expression properties. Dev. Comp. Immunol. 33 (2009) 547–558.

[10] G. Lutfalla, H. Roest Crollius, N. Stange-thomann, O. Jaillon, K. Mogensen, D. Monneron, Comparative genomic analysis reveals independent expansion of a lineage-specific gene family in vertebrates: The class II cytokine receptors and their ligands in mammals and fish. BMC Genomics 4 (2003) 29.

[12] D. Casani, E. Randelli, S. Costantini, A.M. Facchiano, J. Zou, J., S. Martin, C.J. Secombes, G. Scapigliati, F. Buonocore, Molecular characterisation and structural analysis of an interferon homologue in sea bass (*Dicentrarchus labrax* L.). Molec. Immunol. 46 (2009) 943-952.

[13] Y. Kitao, T. Kono, H. Korenaga, T. Lizasa, K. Nakamura, R. Savan, M. Sakai, Characterization and expression analysis of type I interferon in common carp *Cyprinus carpio* L. Molec. Immunol. 46 (2009) 2548-2556.

[14] P. Pereiro, M.M. Costa, P. Diaz-Rosales, S. Dios, A. Figueras, B. Novoa, The first characterization of two type I interferons in turbot (*Scophthalmus maximus*) reveals their differential role, expression pattern and gene induction. Dev. Comp. Immunol. 45 (2014) 233-244.

[15] S. Maekawa, Y.A. Chiang, J. Hikima, M. Sakai, C.F. Lo, H.C. Wang, T. Aoki, Expression and biological activity of two types of interferon genes in medaka (*Oryzias latipes*). Fish & Shellfish Immunol. 48 (2016) 20-29.

[15] Y.W. Hu, T. Yoshikawa, S. Chung, I. Hirono, H. Kondo, Identification of 2 novel type I IFN genes in Japanese flounder, *Paralichthys olivaceus*. Fish & Shellfish Immunol. 67 (2017) 7-10.

[16] Y. Ding, Y.Y. Guan, X.H. Huang, J.Q. Ao, X.H. Chen, Characterization and function of a group II type I interferon in the perciform fish, large yellow croaker (*Larimichthys crocea*). Fish & Shellfish Immunol. 86 (2019) 152-159.

[17] B. Huang, Z.X. Wang, Y. Liang, S.W. Zhai, W.S. Huang, P. Nie, Identification of four type I IFNs from Japanese eel with differential expression profiles and Mx promoter inducibility. Dev. Comp. Immunol. 91 (2019) 62-71.

[18] C. Burge, S. Karlin, Prediction of complete gene structures in human genomic DNA. J. Molec. Biol 268 (1997) 78-94.

[19] J.J. Campanella, L. Bitincka, J. Smalley, MatGAT: an application that generates similarity/identity matrices using protein or DNA sequences. BMC Bioinformatics 4 (2003) 29.

[20] D.M. Li, W.L. Tan, M.S. Ma, X.J. Yu, Q.N. Lai, Z.Q. Wu, G. Lin, C.Y. Hu, Molecular characterization and transcription regulation analysis of type I IFN gene in grass carp (*Ctenopharyngodon idella*). Gene 504 (2012) 31-40.

[21] H. Wu, L.Q. Liu, S.Z. Wu, C.Y. Wang, C.L. Feng, J. Xiao, H. Feng, IFNb of black carp functions importantly in host innate immune response as an antiviral cytokine. Fish & Shellfish Immunol. 74 (2018) 1-9.

[22] Q. Wan, W.D.N. Wicramaarachchi, I. Whang, B.S. Lim, M.J. Oh, S.J. Jung, H.C. Kim, S.Y. Yeo, J. Lee, Molecular cloning and functional characterization of two duplicated two-cysteine containing type I interferon genes in rock bream *Oplegnathus fasciatus*. Fish & Shellfish Immunol. 33 (2012) 886-898.

[23] Y.-M. Chen, C.-E. Kuo, G.-R. Chen, Y.-K. Kao, J. Zou, C.J. Secombes, T.-Y. Chen,, Functional analysis of an orange-spotted grouper (*Epinephelus coiodes*) interferon gene and characterization of its expression in response to nodavirus infection. Dev. Comp. Immunol. 46 (2014) 117-128.

[24] D.J. Milne, C. Campoverde, K.B. Andre, X. Chen, J. Zou, C.J. Secombes, The discovery and comparative expression analysis of three distinct type 1 interferons in the perciform fish, meagre (*Argyrsomus regius*). Dev. Comp. Immunol. 84 (2018) 123-132.

[25] Z.A. Laghari, S.N. Chen, L. Li, B. Huang, Z. Gan, Y. Zhou, H.J. Huo, J. Hou, P. Nie, Functional, signalling and transcriptional differences of three distinct type I IFNs in a perciform fish, the mandarin fish *Siniperca chuatsi*. Dev. Comp. Immunol. 84 (2018) 94-108.

[26] R. Betancur-R, E.O. Wiley, G. Arratia, A. Acero, N. Bailly, M. Miya, G. Lecointre, G. Orti, Phylogenetic classification of bony fishes. BMC Evolutionary Biology 17

(2017) 162.

[27] Y. Nakatani, H. Takeda, Y. Kohara, S. Morishita, Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. Genome Research 17 (2007) 1254-1265.

[28] B. Robertsen, L. Greiner-Tollersrud, L.G. Jorgensen, Analysis of the Atlantic salmon genome reveals a cluster of Mx genes that respond more strongly to IFN gamma than to type I IFN. Dev. Comp. Immunol. 90 (2019) 80-89.

[29] T.-Y. Wang, F. Liu, G. Tian, C.J. Secombes, T. Wang, Lineage/species-specific expansion of the Mx gene family in teleosts: Differential expression and modulation of nine Mx genes in rainbow trout *Oncorhynchus mykiss*. Fish & Shellfish Immunology 90 (2019) 413-430.

[30] M.H. Solbakken, M.L. Rise, K.S. Jakobsen, S. Jentoft, Successive losses of central immune genes characterize the Gadiformes' alternate immunity. Genome Biol. Evol. 8 (2016) 3508-3515.

[31] B. Star, A.J. Nederbragt, S. Jentoft, U. Grimholt, M. Malmstrom, T.F. Gregers, T.B. Rounge, J. Paulsen, M.H. Solbakken, A. Sharma, O.F. Wetten, A. Lanzen, R. Winer, J. Knight, J.H. Vogel, B. Aken, O. Andersen, K. Lagesen, A. Tooming-Klunderud, R.B. Edvardsen, K.G. Tina, M. Espelund, C. Nepal, C. Previti, B.O. Karlsen, T. Moum, M. Skage, P.R. Berg, T. Gjoen, H. Kuhl, J. Thorsen, K. Malde, R. Reinhardt, L. Du, S.D. Johansen, S. Searle, S. Lien, F. Nilsen, I. Jonassen, S.W. Omholt, N.C. Stenseth, K.S. Jakobsen, The genome sequence of Atlantic cod reveals a unique immune system. Nature 477 (2011) 207-210.

[32] M. Malmstrom, M. Matschiner, O.K. Torresen, B. Star, L.G. Snipen, T.F. Hansen, H.T. Baalsrud, A.J. Nederbragt, R. Hanel, W. Salzburger, N.C. Stenseth, K.S. Jakobsen, S. Jentoft, Evolution of the immune system influences speciation rates in teleost fishes. Nature Genetics 48 (2016) 1204-1210.

[33] A. Krasnov, O. Kileng, S. Skugor, S.M. Jorgensen, S. Afanasyev, G. Timmerhaus, A.I. Sommer, I. Jensen, Genomic analysis of the host response to nervous necrosis virus in Atlantic cod (*Gadus morhua*) brain. Molec. Immunol. 54 (2013) 443-452.

[34] K. Eslamloo, A. Ghorbani, X. Xue, S.M. Inkpen, M. Larijani, M.L. Rise, Characterization and transcript expression analyses of Atlantic cod viperin. Front. Immunol. 10 (2019) 311.

[35] M.-X. Chang, J. Zou, P. Nie, B. Huang, Z. Yu, B. Collet, C.J. Secombes, Intracellular interferons in fish: A unique means to combat viral infection. PLoS Pathogens 9 (11) (2013) e1003736.

[36] T. Svingerud, T. Solstad, B. Sun, M.L.J. Nyrud, Ø. Kileng, L. Greiner-Tollersrud, B. Robertsen, Atlantic salmon type I IFN subtypes show differences in antiviral activity and cell-dependent expression: Evidence for high IFNb/IFNc-producing cells in fish lymphoid tissues. J. Immunol. 189 (2012) 5912–5923.

[37] Y.J. Liu, IPC: Professional type I interferon-producing cells and plasmacytoid dendritic cell precursors. Ann. Rev. Immunol. 23 (2005) 275-306.

[38] P. Björck, H.X. Leong, E.G. Englemann, Plasmacytoid dendritic cell dichotomy: Identification of IFN-α producing cells as a phenotypically and functionally distinct subset. J. Immunol. 186 (2011) 1477-1485.

[39] D. Aggad, M. Mazel, P. Boudinot, K.E. Mogensen, O.J. Hamming, R. Hartmann, S. Kotenko, P. Herbomel, G. Lutfalla, J.P. Levraud, The two groups of zebrafish virus-induced interferons signal via distinct receptors with specific and shared chains. J. Immunol. 183 (2009) 3924–3931.

[40] J. Zou, C. Tafalla, J. Truckle, C.J. Secombes, Identification of a second group of type I IFNs in fish sheds light on IFN evolution in vertebrates. J. Immunol. 179 (2007) 3859-3871.

[41] M. Chang, P. Nie, B. Collet, C.J. Secombes, J. Zou, Identification of an additional two-cysteine containing type I interferon in rainbow trout *Oncorhynchus mykiss* provides evidence of a major gene duplication event within this gene family in teleosts. Immunogenetics 61 (2009) 315-325.

Figure Legends

Figure 1. Figure showing the GH and CD79b loci in pike, with the associated IFN genes, with different colours representing the different IFN subgroups present.

Figure 2. Phylogenetic tree of all salmonid and pike IFN molecules known to date. The phylogenetic tree was constructed using amino acid multiple alignments of IFN molecules from salmonids and pike, and the neighbour-joining method within the MEGA7.0 program. The evolutionary distances were computed using the JTT matrix-based method with all ambiguous positions removed for each sequence pair. Node values represent percent bootstrap confidence derived from 1,000 replications. Note the subdivision of the IFNc subgroup into two clades that represent molecules at the two GH loci in the salmonid species. The pike molecules are highlighted with a red dot.

Figure 3. Figure showing the GH locus 1 in salmonids, with the associated IFN genes,

with different colours representing the different IFN subgroups present. Note the IFNb pseudogenes shown with a solid line and the additional partial IFN genes shown with broken lines. In the case of charr two scaffolds are presented that were considered to be from locus 1. Scaffold 1253 contains GH whilst scaffold 807 has IFN subgroups and gene number (1x IFNa, 1x IFNe and 1x IFNf) that suggest it is part of locus 1, especially as the IFNa and IFNe genes cluster with the respective Atlantic salmon genes from this locus (see **Fig. 2**). Note that the previously published trout IFN2 [40] is IFNa1.1.

Figure 4. Figure showing the GH locus 2 in salmonids, with the associated IFN genes, with different colours representing the different IFN subgroups present. Note the pseudogenes shown with a solid line and the additional partial IFN genes shown with broken lines. In the case of charr three scaffolds are presented that were considered to be from locus 2. Scaffold 4096 had an IFNc gene that grouped with other locus 2 IFNc molecules, whilst scaffold 3499 had two IFNe where only a single gene is present at locus 1. In addition, the charr IFNa gene clustered with the respective Atlantic salmon IFN genes from locus 2, and the charr IFNe genes showed similar associations (see **Fig. 2**). Note that the previously published trout IFN1, IFN3 and IFN4 [40, 41] are IFNa2.6, IFNb2.1 and IFNb2.2 respectively.

Figure 5. Figure showing the CD79b loci in salmonids, with the associated IFNd genes. Note that the previously published trout IFN5 [41] is IFNd3.1.

Figure 6. Figure showing the IFN locus of A) gar (associated with GH and CD79b) in comparison to the two loci in B) Japanese eel and C) bonytongue. Different colours represent the different IFN subgroups present. Note the two IFNf genes could not be linked to GH or CD79b. A partial IFNa gene was also found at locus 1.

Figure 7. Phylogenetic tree of all teleost IFN molecules reported in this study. A) The salmonid and pike IFN subgroup clades are condensed (shown as black triangles), as well as the percomorph IFNd genes (pink triangle). B) the percomorph IFNd genes alone. The phylogenetic tree was constructed using amino acid multiple alignments of the IFN molecules, and the neighbour-joining method within the MEGA7.0 program. The evolutionary distances were computed using the JTT matrix-based method with all ambiguous positions removed for each sequence pair. Node values represent percent bootstrap confidence derived from 1,000 replications.

Figure 8. Figure showing the GH and CD79b loci and associated IFN genes found in A) zebrafish and B) common carp. Different colours represent the different IFN subgroups present. Locus 1 was derived from contigs 26878, 18220 and 2101, locus 2 from contigs 56270 and 4163, locus 3 from contig 13361 and locus 4 from contig 56953. Note that as the cyprinid type I IFN nomenclature is different from other teleost groups, a translation has been provided. All genes indicated with IFNa correspond to IFNphi1 in cyprinids, genes indicated with IFNd correspond to IFNphi4. Genes indicated with

IFNcx.1 correspond to IFNphi3, and genes indicated with IFNcx.2 correspond to IFNphi2.

Figure 9. Figure showing the GH and CD79b loci and associated IFN genes found in A) turbot, B) tetraodon, C) icefish and D) large yellow croaker. Different colours represent the different IFN subgroups present. Note the partial IFNh sequence in tetraodon shown with a broken line. Also, note that the previously published turbot IFN1 = IFNc1.1 and IFN2 = IFNh1.1 [12].

Figure 10. Figure showing the GH and CD79b loci and associated IFN genes found in A) tilapia, B) seabass and C) Atlantic cod. Different colours represent the different IFN subgroups present. Note partial IFNd sequences in tilapia shown with a broken line, and seabass scaffolds 3867 and 1156 (locus 2) were combined following our analysis. Homologous blocks of tilapia IFN genes are underlined with red and green lines, respectively.

Figure 11. Possible model of type I IFN evolution in teleosts.

Supplementary Figure Legends

SFig. 1. Multiple amino acid alignment of all salmonid IFNa molecules.

SFig 2. Multiple amino acid alignment of all salmonid IFNb molecules.

SFig 3. Multiple amino acid alignment of all salmonid IFNc molecules.

SFig 4. Multiple amino acid alignment of all salmonid IFNd molecules.

SFig 5. Multiple amino acid alignment of all salmonid IFNe molecules.

SFig 6. Multiple amino acid alignment of all salmonid IFNf molecules.

SFig 7. Figure showing the GH and CD79b loci and associated IFN genes found in A) toothfish, B) medaka, C) flounder, D) stickleback and E) haddock. Different colours represent the different IFN subgroups present. Note that the medaka IFNh was not proven to be linked to GH and was based on the sequence provided in Maekawa et al. [15]. Similarly the two haddock genes have not been shown to be linked.

Figure 1.

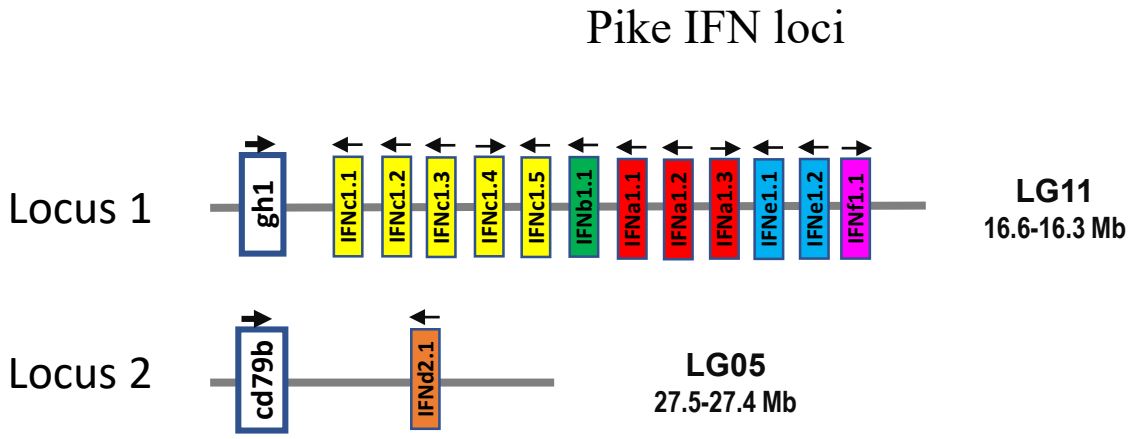

Pike IFN loci
Locus 1
gh1
IFNc1.1
IFNc1.2
IFNc1.3
IFNc1.4
IFNc1.5
IFNb1.1
IFNa1.1
IFNa1.2
IFNa1.3
IFNe1.1
IFNe1.2
IFNf1.1
LG11
16.6-16.3 Mb
Locus 2
cd79b
IFNd2.1
LG05
27.5-27.4 Mb

Figure 2.

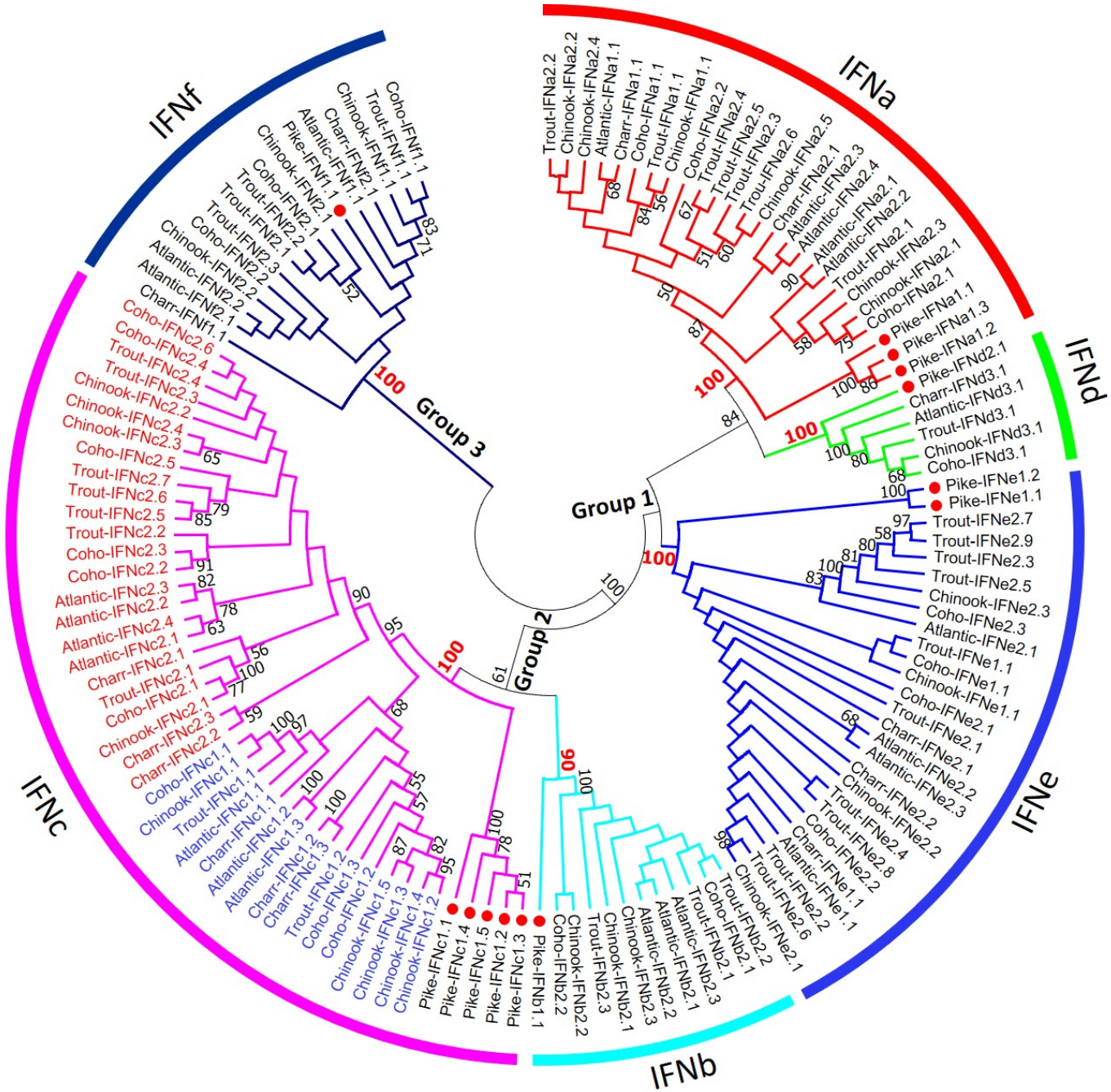

IFNa
IFNd
IFNe
IFNb
IFNc
IFNf
Group 1
Group 2
Group 3
Trout-IFNa2.2
Chinook-IFNa2.2
Chinook-IFNa2.4
Atlantic-IFNa1.1
Charr-IFNa1.1
Coho-IFNa1.1
Trout-IFNa1.1
Chinook-IFNa1.1
Coho-IFNa2.2
Trout-IFNa2.4
Trout-IFNa2.5
Trout-IFNa2.3
Trou-IFNa2.6
Chinook-IFNa2.5
Charr-IFNa2.1
Atlantic-IFNa2.3
Atlantic-IFNa2.4
Atlantic-IFNa2.1
Atlantic-IFNa2.2
Trout-IFNa2.1
Chinook-IFNa2.3
Chinook-IFNa2.1
Coho-IFNa2.1
Pike-IFNa1.1
Pike-IFNa1.3
Pike-IFNa1.2
Pike-IFNd2.1
Charr-IFNd3.1
Atlantic-IFNd3.1
Trout-IFNd3.1
Chinook-IFNd3.1
Coho-IFNd3.1
Pike-IFNe1.2
Pike-IFNe1.1
Trout-IFNe2.7
Trout-IFNe2.9
Trout-IFNe2.3
Trout-IFNe2.5
Chinook-IFNe2.3
Coho-IFNe2.3
Atlantic-IFNe2.1
Trout-IFNe1.1
Coho-IFNe1.1
Chinook-IFNe1.1
Coho-IFNe2.1
Trout-IFNe2.1
Charr-IFNe2.1
Atlantic-IFNe2.2
Atlantic-IFNe2.3
Charr-IFNe2.2
Chinook-IFNe2.2
Trout-IFNe2.2
Coho-IFNe2.2
Trout-IFNe2.4
Trout-IFNe2.8
Charr-IFNe1.1
Atlantic-IFNe1.1
Trout-IFNe2.6
Chinook-IFNe2.1
Trout-IFNb2.2
Coho-IFNb2.1
Atlantic-IFNb2.1
Trout-IFNb2.3
Chinook-IFNb2.1
Trout-IFNb2.3
Atlantic-IFNb2.2
Atlantic-IFNb2.3
Chinook-IFNb2.2
Coho-IFNb2.2
Pike-IFNb1.1
Pike-IFNc1.3
Pike-IFNc1.2
Pike-IFNc1.5
Pike-IFNc1.4
Pike-IFNc1.1
Chinook-IFNc1.4
Chinook-IFNc1.3
Chinook-IFNc1.5
Coho-IFNc1.2
Trout-IFNc1.3
Charr-IFNc1.3
Charr-IFNc1.2
Atlantic-IFNc1.3
Atlantic-IFNc1.2
Trout-IFNc1.1
Charr-IFNc1.1
Coho-IFNc1.1
Chinook-IFNc1.1
Trout-IFNc1.1
Atlantic-IFNc1.1
Coho-IFNc1.1
Charr-IFNc2.2
Charr-IFNc2.3
Chinook-IFNc2.1
Coho-IFNc2.1
Trout-IFNc2.1
Charr-IFNc2.1
Atlantic-IFNc2.1
Atlantic-IFNc2.4
Atlantic-IFNc2.2
Atlantic-IFNc2.3
Coho-IFNc2.2
Coho-IFNc2.3
Trout-IFNc2.2
Trout-IFNc2.5
Trout-IFNc2.6
Trout-IFNc2.7
Coho-IFNc2.5
Chinook-IFNc2.3
Chinook-IFNc2.4
Chinook-IFNc2.2
Trout-IFNc2.3
Trout-IFNc2.4
Coho-IFNc2.4
Coho-IFNc2.6
Charr-IFNf1.1
Atlantic-IFNf2.1
Atlantic-IFNf2.2
Chinook-IFNf2.2
Coho-IFNf2.3
Trout-IFNf2.2
Trout-IFNf2.1
Coho-IFNf2.2
Chinook-IFNf2.1
Pike-IFNf1.1
Atlantic-IFNf1.1
Charr-IFNf2.1
Chinook-IFNf1.1
Trout-IFNf1.1
Coho-IFNf1.1

Figure 3.

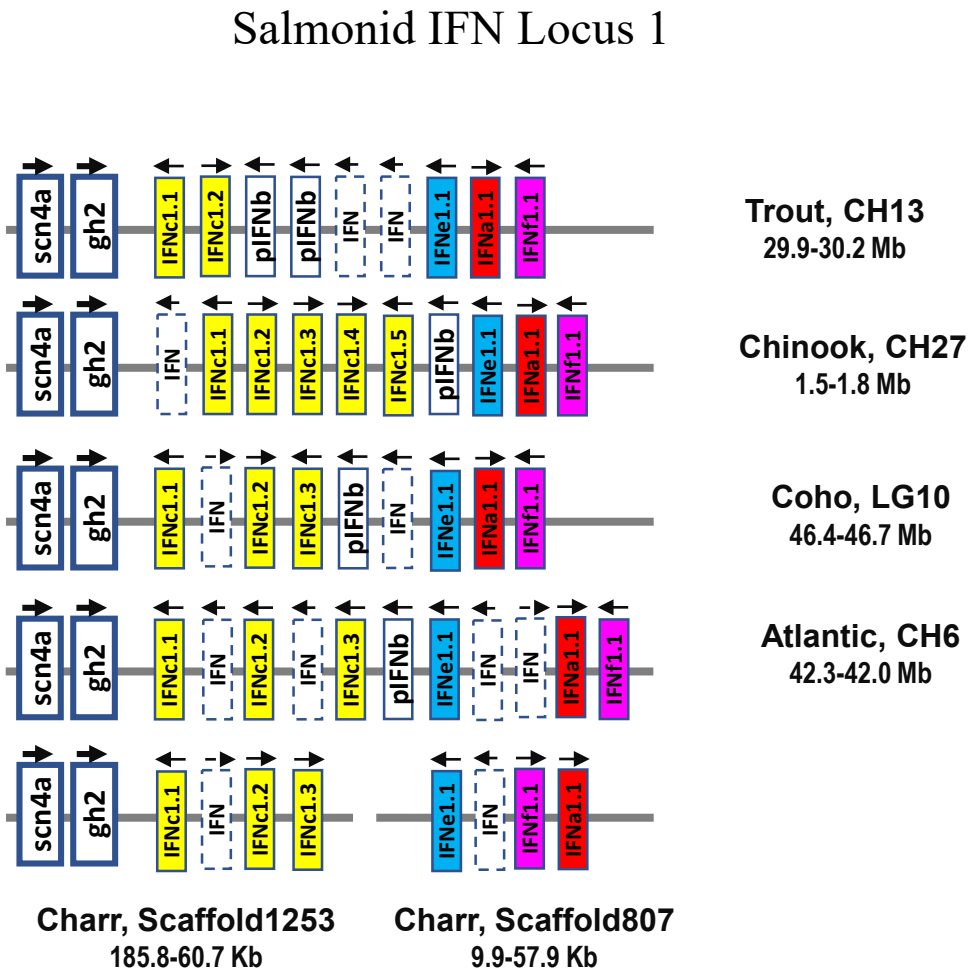

Salmonid IFN Locus 1
scn4a
gh2
IFNc1.1
IFNc1.2
pIFNb
pIFNb
IFN
IFN
IFNe1.1
IFNa1.1
IFNf1.1
Trout, CH13
29.9-30.2 Mb
IFN
IFNc1.1
IFNc1.2
IFNc1.3
IFNc1.4
IFNc1.5
pIFNb
IFNe1.1
IFNa1.1
IFNf1.1
Chinook, CH27
1.5-1.8 Mb
IFNc1.1
IFN
IFNc1.2
IFNc1.3
pIFNb
IFN
IFNe1.1
IFNa1.1
IFNf1.1
Coho, LG10
46.4-46.7 Mb
IFNc1.1
IFN
IFNc1.2
IFN
IFNc1.3
pIFNb
IFNe1.1
IFN
IFN
IFNa1.1
IFNf1.1
Atlantic, CH6
42.3-42.0 Mb
IFNc1.1
IFN
IFNc1.2
IFNc1.3
IFNe1.1
IFN
IFNf1.1
IFNa1.1
Charr, Scaffold1253
185.8-60.7 Kb
Charr, Scaffold807
9.9-57.9 Kb

Figure 4.

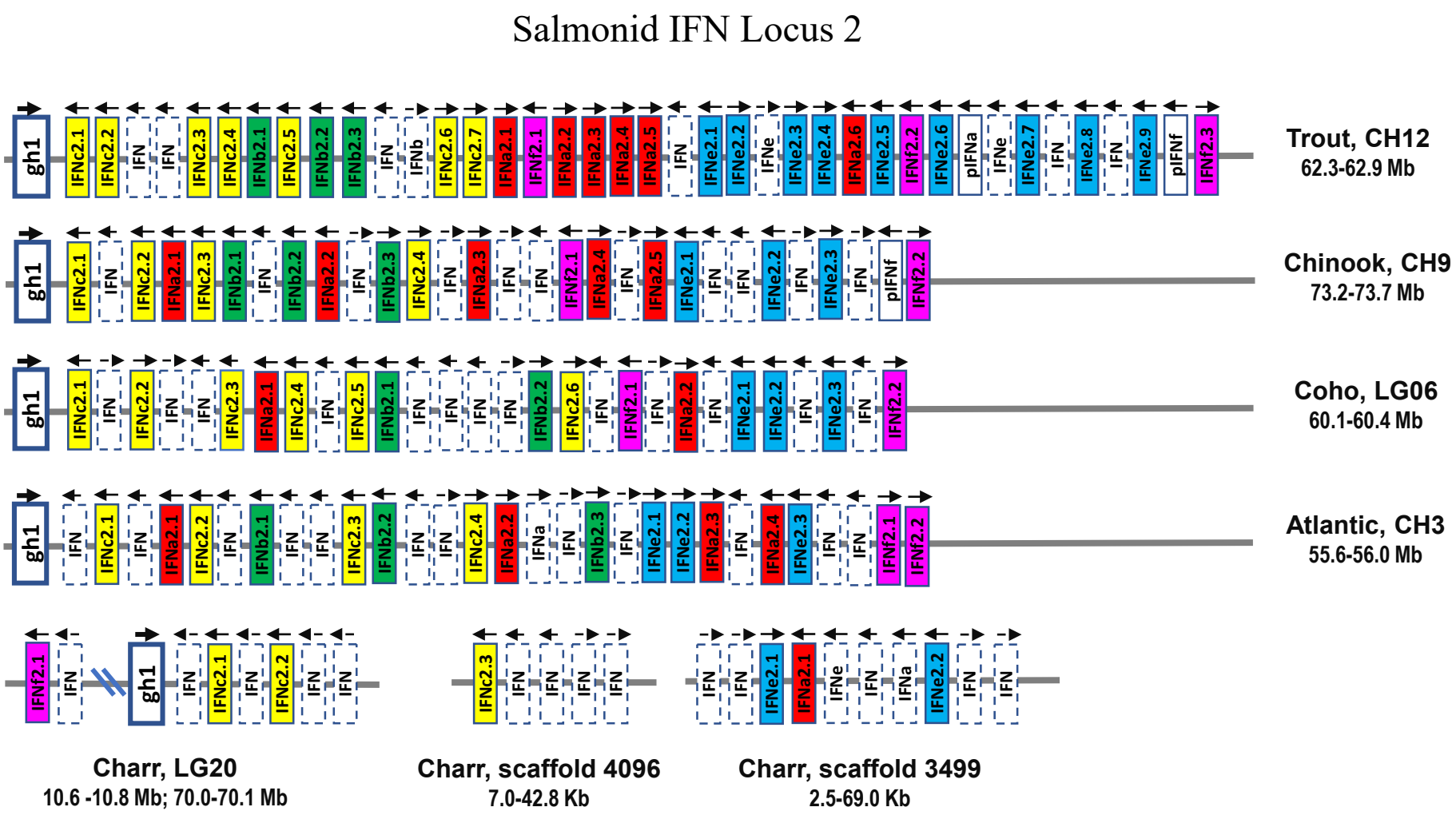

Salmonid IFN Locus 2
gh1
IFNc2.1
IFNc2.2
IFN
IFNc2.3
IFNc2.4
IFNb2.1
IFNc2.5
IFNb2.2
IFNb2.3
IFNb
IFNc2.6
IFNc2.7
IFNa2.1
IFNf2.1
IFNa2.2
IFNa2.3
IFNa2.4
IFNa2.5
IFNe2.1
IFNe2.2
IFNe
IFNe2.3
IFNe2.4
IFNa2.6
IFNe2.5
IFNf2.2
IFNe2.6
pIFNa
IFNe2.7
IFNe2.8
IFNe2.9
pIFNf
IFNf2.3
Trout, CH12
62.3-62.9 Mb
IFNa2.1
IFNc2.3
IFNb2.1
IFNb2.2
IFNa2.2
IFNb2.3
IFNc2.4
IFNa2.3
IFNa2.4
IFNa2.5
IFNe2.1
IFNe2.2
IFNe2.3
pIFNf
IFNf2.2
Chinook, CH9
73.2-73.7 Mb
IFNa2.1
IFNc2.4
IFNc2.5
IFNb2.1
IFNb2.2
IFNc2.6
IFNf2.1
IFNa2.2
IFNe2.1
IFNe2.2
IFNe2.3
IFNf2.2
Coho, LG06
60.1-60.4 Mb
IFNa2.1
IFNc2.2
IFNb2.1
IFNc2.3
IFNb2.2
IFNc2.4
IFNa2.2
IFNa
IFNb2.3
IFNe2.1
IFNe2.2
IFNa2.3
IFNa2.4
IFNe2.3
IFNf2.1
IFNf2.2
Atlantic, CH3
55.6-56.0 Mb
IFNf2.1
IFNc2.1
IFNc2.2
IFNc2.3
Charr, LG20
10.6 -10.8 Mb; 70.0-70.1 Mb
IFNc2.3
Charr, scaffold 4096
7.0-42.8 Kb
IFNe2.1
IFNa2.1
IFNe
IFNa
IFNe2.2
Charr, scaffold 3499
2.5-69.0 Kb

Figure 5.

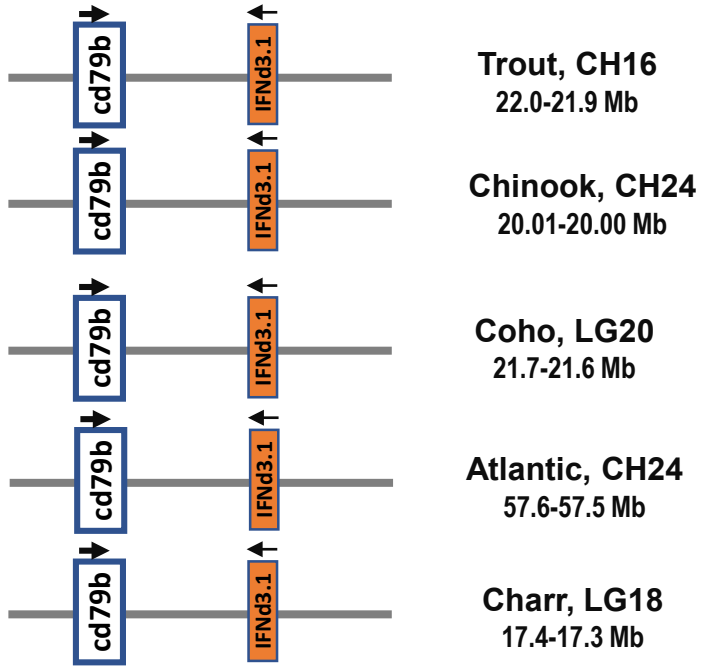
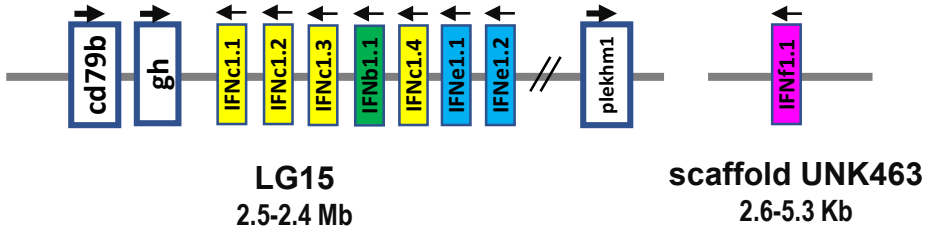
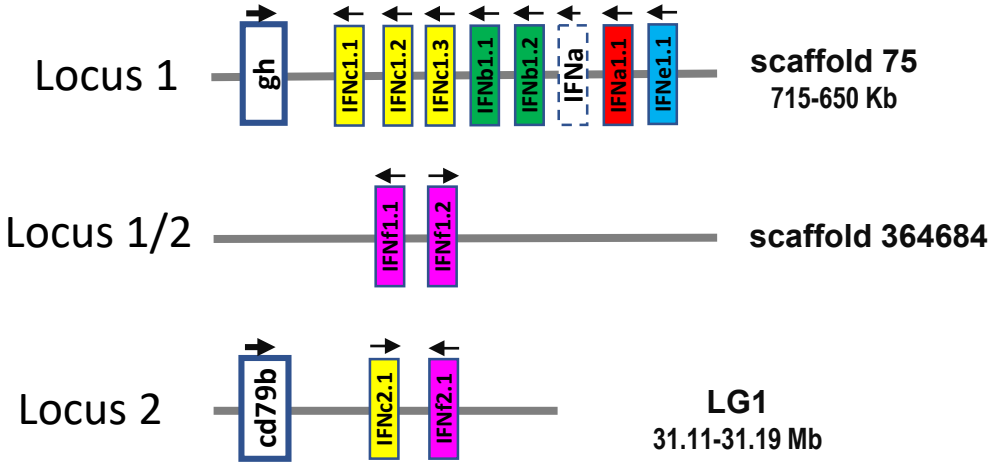

Salmonid IFN Locus 3
cd79b
IFNd3.1
Trout, CH16
22.0-21.9 Mb
Chinook, CH24
20.01-20.00 Mb
Coho, LG20
21.7-21.6 Mb
Atlantic, CH24
57.6-57.5 Mb
Charr, LG18
17.4-17.3 Mb

Figure 6.

A. Gar IFN locus



B. Japanese eel IFN loci



C. Bonytongue IFN loci

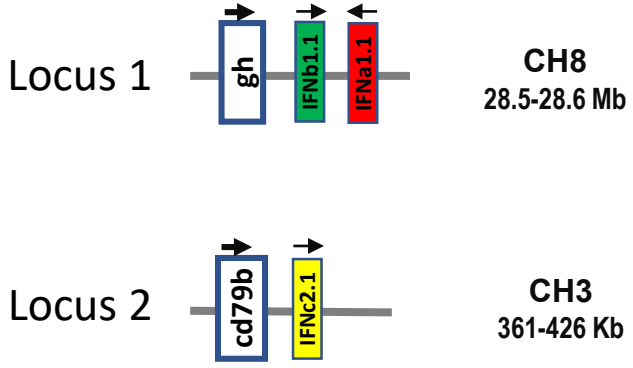


Figure 7.

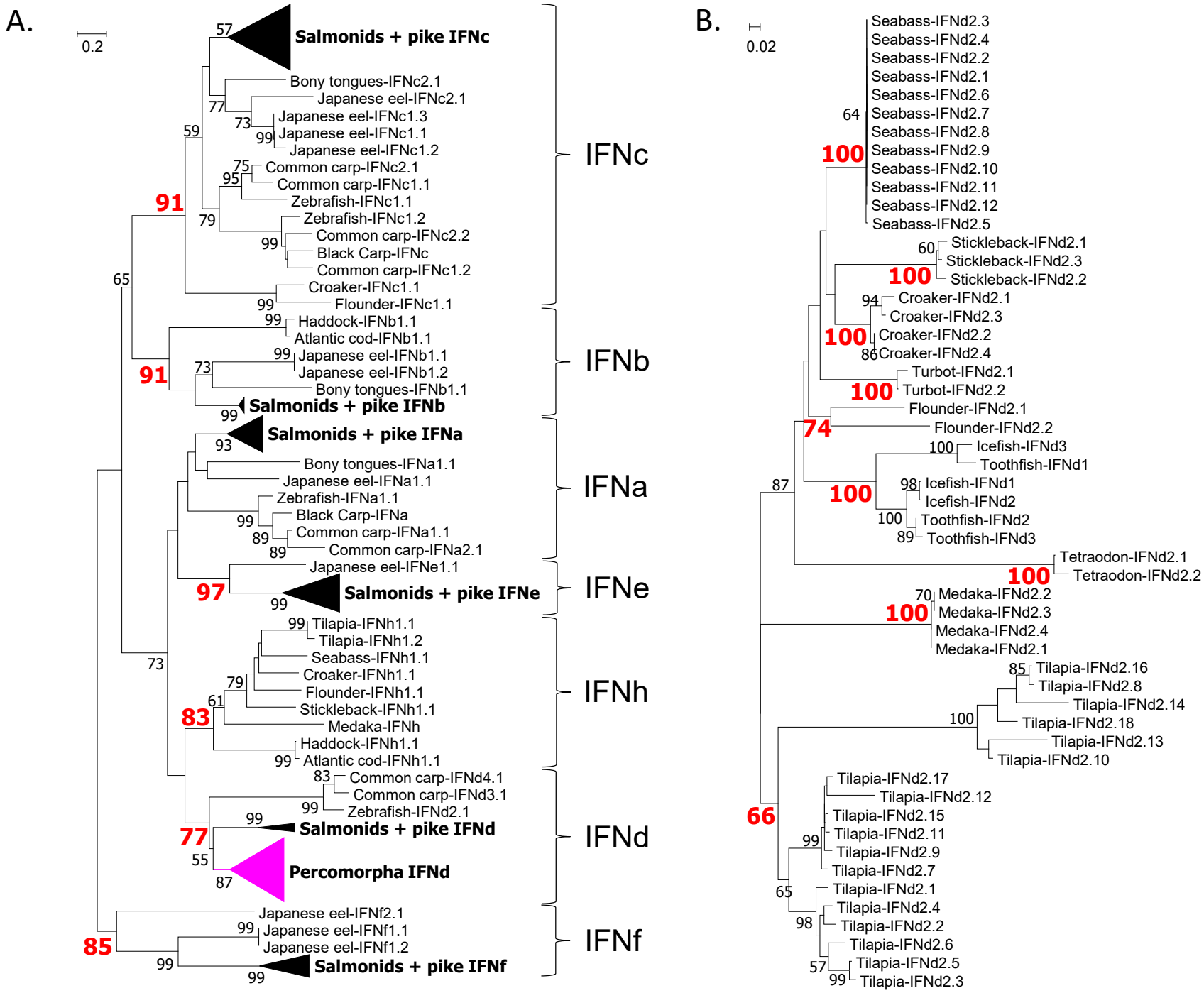

A.
0.2
Salmonids + pike IFNc
Bony tongues-IFNc2.1
Japanese eel-IFNc2.1
Japanese eel-IFNc1.3
Japanese eel-IFNc1.1
Japanese eel-IFNc1.2
Common carp-IFNc2.1
Common carp-IFNc1.1
Zebrafish-IFNc1.1
Zebrafish-IFNc1.2
Common carp-IFNc2.2
Black Carp-IFNc
Common carp-IFNc1.2
Croaker-IFNc1.1
Flounder-IFNc1.1
IFNc
Haddock-IFNb1.1
Atlantic cod-IFNb1.1
Japanese eel-IFNb1.1
Japanese eel-IFNb1.2
Bony tongues-IFNb1.1
Salmonids + pike IFNb
IFNb
Salmonids + pike IFNa
Bony tongues-IFNa1.1
Japanese eel-IFNa1.1
Zebrafish-IFNa1.1
Black Carp-IFNa
Common carp-IFNa1.1
Common carp-IFNa2.1
IFNa
Japanese eel-IFNe1.1
Salmonids + pike IFNe
IFNe
Tilapia-IFNh1.1
Tilapia-IFNh1.2
Seabass-IFNh1.1
Croaker-IFNh1.1
Flounder-IFNh1.1
Stickleback-IFNh1.1
Medaka-IFNh
Haddock-IFNh1.1
Atlantic cod-IFNh1.1
IFNh
Common carp-IFNd4.1
Common carp-IFNd3.1
Zebrafish-IFNd2.1
Salmonids + pike IFNd
Percomorpha IFNd
IFNd
Japanese eel-IFNf2.1
Japanese eel-IFNf1.1
Japanese eel-IFNf1.2
Salmonids + pike IFNf
IFNf
B.
0.02
Seabass-IFNd2.3
Seabass-IFNd2.4
Seabass-IFNd2.2
Seabass-IFNd2.1
Seabass-IFNd2.6
Seabass-IFNd2.7
Seabass-IFNd2.8
Seabass-IFNd2.9
Seabass-IFNd2.10
Seabass-IFNd2.11
Seabass-IFNd2.12
Seabass-IFNd2.5
Stickleback-IFNd2.1
Stickleback-IFNd2.3
Stickleback-IFNd2.2
Croaker-IFNd2.1
Croaker-IFNd2.3
Croaker-IFNd2.2
Croaker-IFNd2.4
Turbot-IFNd2.1
Turbot-IFNd2.2
Flounder-IFNd2.1
Flounder-IFNd2.2
Icefish-IFNd3
Toothfish-IFNd1
Icefish-IFNd1
Icefish-IFNd2
Toothfish-IFNd2
Toothfish-IFNd3
Tetraodon-IFNd2.1
Tetraodon-IFNd2.2
Medaka-IFNd2.2
Medaka-IFNd2.3
Medaka-IFNd2.4
Medaka-IFNd2.1
Tilapia-IFNd2.16
Tilapia-IFNd2.8
Tilapia-IFNd2.14
Tilapia-IFNd2.18
Tilapia-IFNd2.13
Tilapia-IFNd2.10
Tilapia-IFNd2.17
Tilapia-IFNd2.12
Tilapia-IFNd2.15
Tilapia-IFNd2.11
Tilapia-IFNd2.9
Tilapia-IFNd2.7
Tilapia-IFNd2.1
Tilapia-IFNd2.4
Tilapia-IFNd2.2
Tilapia-IFNd2.6
Tilapia-IFNd2.5
Tilapia-IFNd2.3

Figure 8.

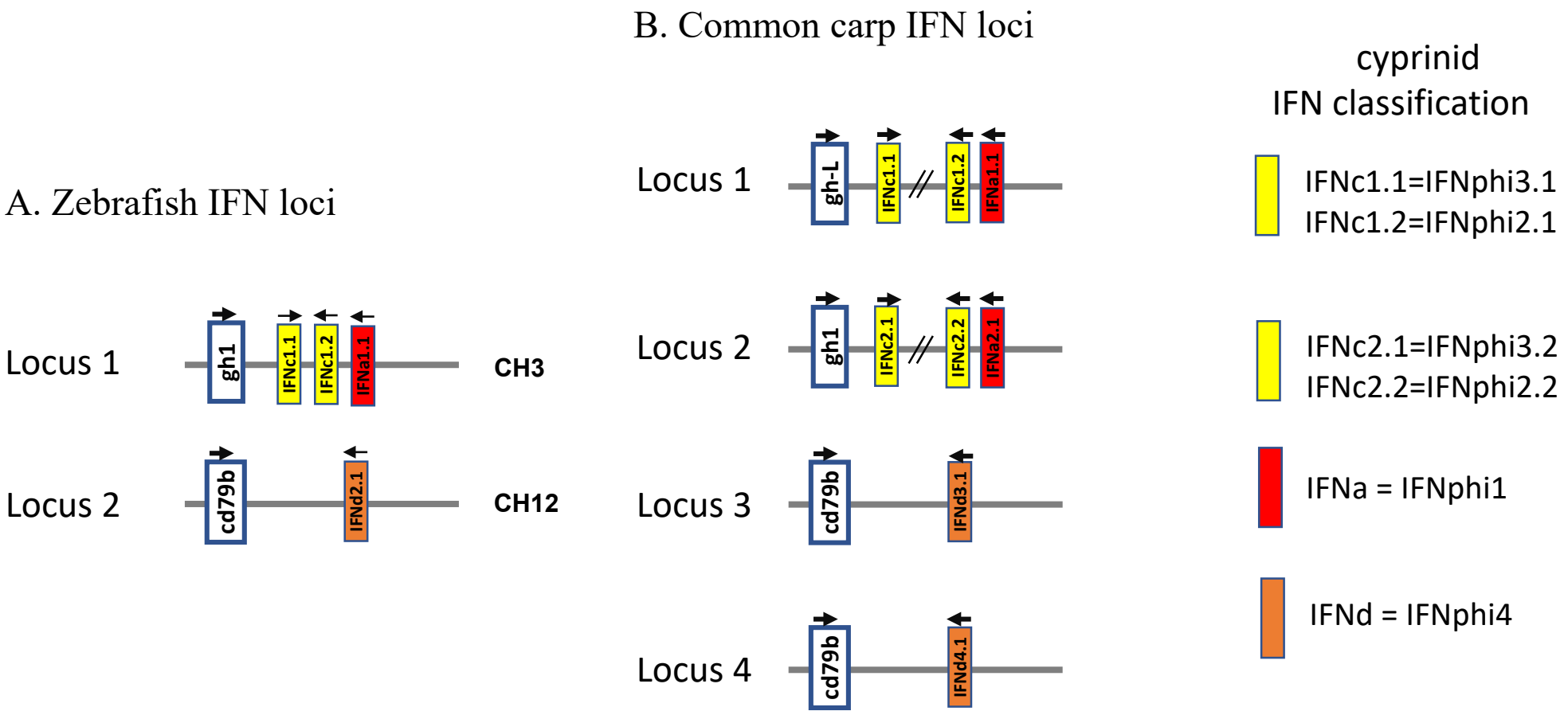

A. Zebrafish IFN loci
Locus 1
gh1
IFNc1.1
IFNc1.2
IFNa1.1
CH3
Locus 2
cd79b
IFNd2.1
CH12
B. Common carp IFN loci
Locus 1
gh-L
IFNc1.1
IFNc1.2
IFNa1.1
Locus 2
gh1
IFNc2.1
IFNc2.2
IFNa2.1
Locus 3
cd79b
IFNd3.1
Locus 4
cd79b
IFNd4.1
cyprinid
IFN classification
IFNc1.1=IFNphi3.1
IFNc1.2=IFNphi2.1
IFNc2.1=IFNphi3.2
IFNc2.2=IFNphi2.2
IFNa = IFNphi1
IFNd = IFNphi4

Figure 9.

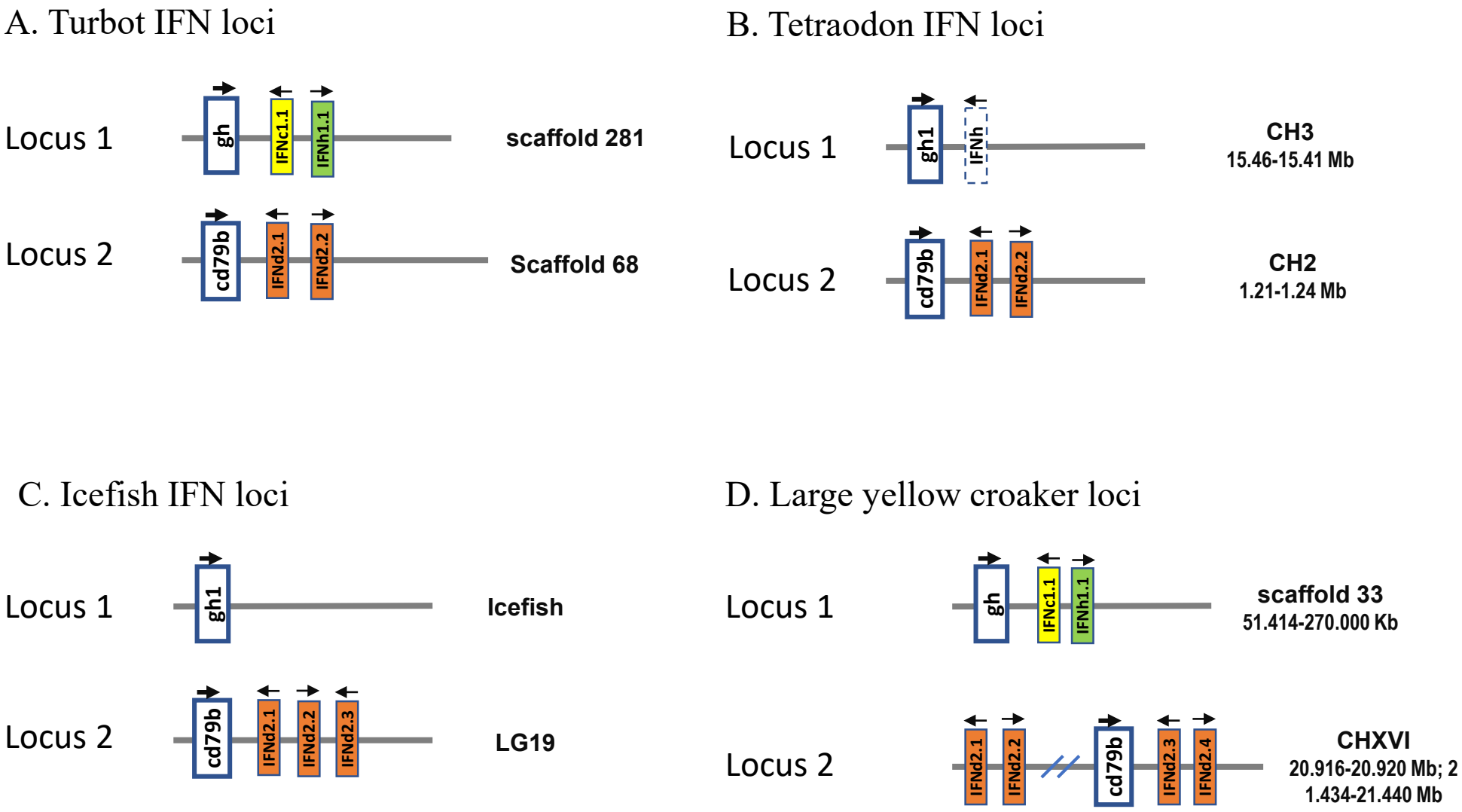

A. Turbot IFN loci
Locus 1
gh
IFNc1.1
IFNh1.1
scaffold 281
Locus 2
cd79b
IFNd2.1
IFNd2.2
Scaffold 68
B. Tetraodon IFN loci
Locus 1
gh1
IFNh
CH3
15.46-15.41 Mb
Locus 2
cd79b
IFNd2.1
IFNd2.2
CH2
1.21-1.24 Mb
C. Icefish IFN loci
Locus 1
gh1
Icefish
Locus 2
cd79b
IFNd2.1
IFNd2.2
IFNd2.3
LG19
D. Large yellow croaker loci
Locus 1
gh
IFNc1.1
IFNh1.1
scaffold 33
51.414-270.000 Kb
Locus 2
IFNd2.1
IFNd2.2
cd79b
IFNd2.3
IFNd2.4
CHXVI
20.916-20.920 Mb; 2
1.434-21.440 Mb

Figure 10.

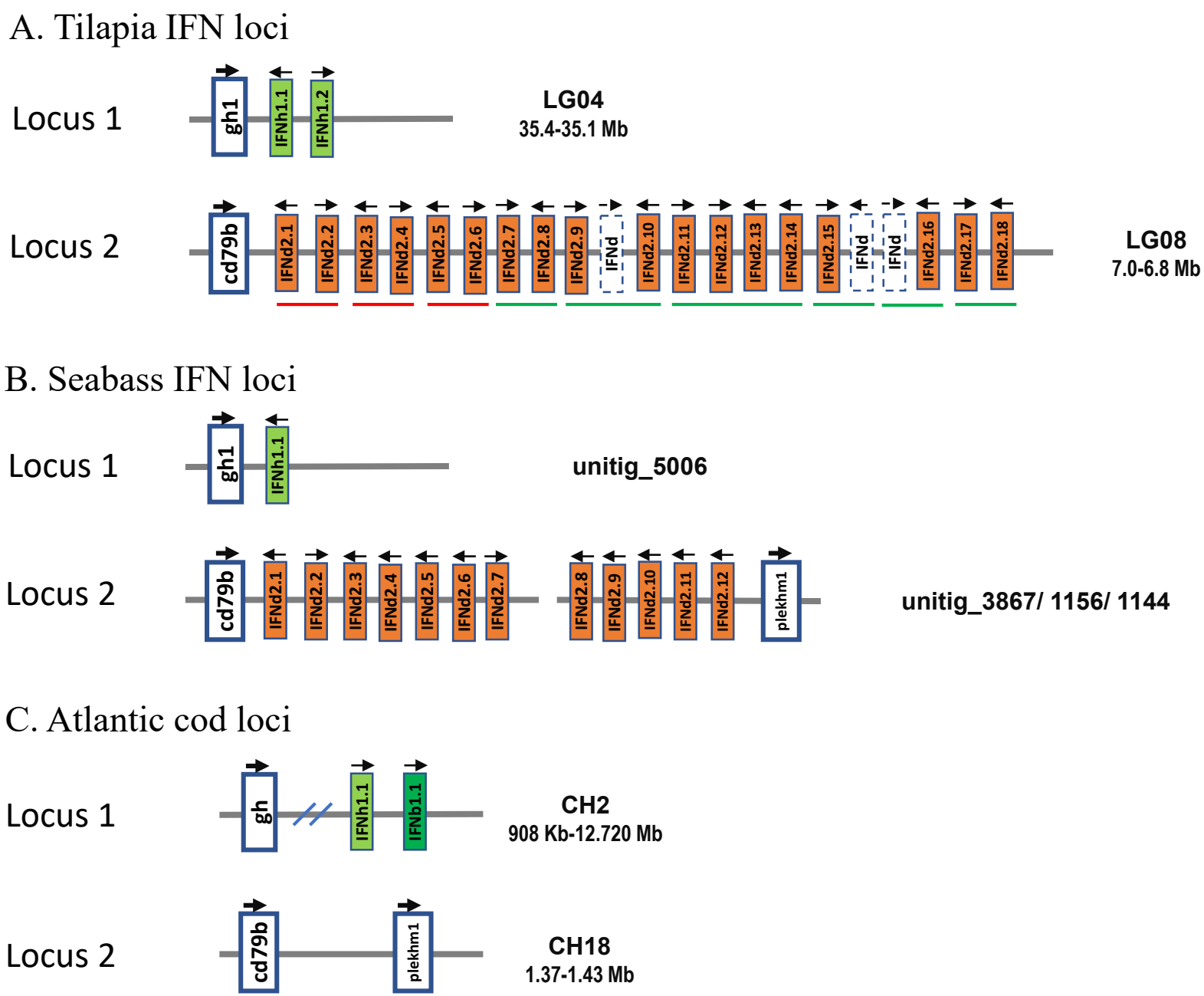
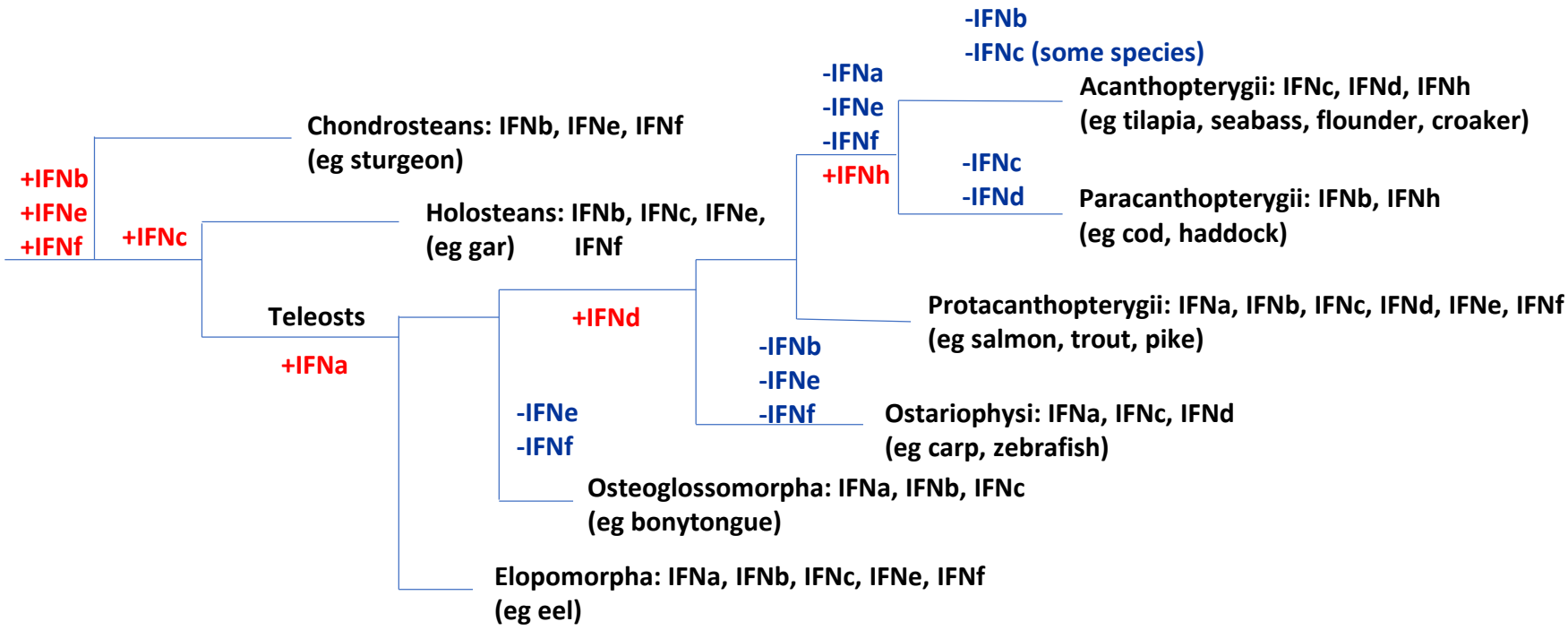

A. Tilapia IFN loci
Locus 1
gh1
IFNh1.1
IFNh1.2
LG04
35.4-35.1 Mb
Locus 2
cd79b
IFNd2.1
IFNd2.2
IFNd2.3
IFNd2.4
IFNd2.5
IFNd2.6
IFNd2.7
IFNd2.8
IFNd2.9
IFNd
IFNd2.10
IFNd2.11
IFNd2.12
IFNd2.13
IFNd2.14
IFNd2.15
IFNd
IFNd
IFNd2.16
IFNd2.17
IFNd2.18
LG08
7.0-6.8 Mb
B. Seabass IFN loci
Locus 1
gh1
IFNh1.1
unitig_5006
Locus 2
cd79b
IFNd2.1
IFNd2.2
IFNd2.3
IFNd2.4
IFNd2.5
IFNd2.6
IFNd2.7
IFNd2.8
IFNd2.9
IFNd2.10
IFNd2.11
IFNd2.12
plekhm1
unitig_3867/ 1156/ 1144
C. Atlantic cod loci
Locus 1
gh
IFNh1.1
IFNb1.1
CH2
908 Kb-12.720 Mb
Locus 2
cd79b
plekhm1
CH18
1.37-1.43 Mb

Figure. 11


+IFNb
+IFNe
+IFNf
Chondrosteans: IFNb, IFNe, IFNf
(eg sturgeon)
+IFNc
Holosteans: IFNb, IFNc, IFNe,
(eg gar) IFNf
Teleosts
+IFNa
+IFNd
-IFNe
-IFNf
Osteoglossomorpha: IFNa, IFNb, IFNc
(eg bonytongue)
Elopomorpha: IFNa, IFNb, IFNc, IFNe, IFNf
(eg eel)
-IFNb
-IFNe
-IFNf
Ostariophysi: IFNa, IFNc, IFNd
(eg carp, zebrafish)
-IFNa
-IFNe
-IFNf
+IFNh
Protacanthopterygii: IFNa, IFNb, IFNc, IFNd, IFNe, IFNf
(eg salmon, trout, pike)
-IFNb
-IFNc (some species)
Acanthopterygii: IFNc, IFNd, IFNh
(eg tilapia, seabass, flounder, croaker)
-IFNc
-IFNd
Paracanthopterygii: IFNb, IFNh
(eg cod, haddock)