Immunogenetics (2022) 74:129–147
https://doi.org/10.1007/s00251-021-01235-4

ORIGINAL ARTICLE

Recurrent expansions of B30.2-associated immune receptor families in fish

Jaanus Suurväli[1] · Colin J. Garroway[1] · Pierre Boudinot[2]

Received: 11 September 2021 / Accepted: 16 November 2021 / Published online: 1 December 2021

Abstract

B30.2 domains, also known as PRY/SPRY, are key components of specific subsets of two large families of proteins involved in innate immunity: the tripartite motif proteins (TRIMs) and the Nod-like receptors (NLRs). TRIM proteins are important, often inducible factors of antiviral innate immunity, targeting multiple steps of viral cycles through a variety of mechanisms. NLRs prime and regulate systemic innate defenses, especially against bacteria, and control inflammation. Large TRIM and NLR subsets characterized by the presence of a B30.2 domain have been reported from a few fish species including zebrafish and seem to be strongly prone to gene duplication/expansion. Here, we performed a large-scale survey of these receptors across about 150 fish genomes, focusing on ray-finned fishes. We assessed the number and genomic distribution of domains and domain combinations associated with TRIMs, NLRs, and other genes containing B30.2 domains and looked for gene expansion patterns across fish groups. We then used a model to test the impact of taxonomy, genome size, and environmental variables on the copy numbers of these genes. Our findings reveal novel domain structures, clade-specific gains and losses. They also assist with the timing of the gene expansions, reveal patterns associated with the MHC, and lay the groundwork for further studies delving deeper into the forces that drive the copy number variation of immune genes on a species level.

Keywords B30.2 · Copy number modeling · Fish immune duplications · NLR · NOD-like receptor · TRIM

Introduction

Domains frequently used in proteins of eukaryotic immune systems include the immunoglobulin domains, leucine-rich repeats, and DEATH-type interaction domains (CARD, PYD, etc.) (Buckley and Rast 2015). The B30.2 domain (also known as PRY/SPRY) (Henry et al. 1997) is another example, although perhaps less well known. Notable immune protein families associated with B30.2 domains in vertebrates include the tripartite motif proteins (TRIMs) (Nisole et al. 2005), NACHT and leucine-rich repeat-containing receptors (NLRs, also known as NOD-like receptors) (Laing et al. 2008; Stein et al. 2007), and butyrophilins (Afrache et al. 2012; Salim et al. 2017). B30.2 domains are also found in the venom of stonefish and snakes (Henry et al. 1997) and as single-domain proteins in many genomes. Another immune protein containing a B30.2 domain is the human pyrin (MEFV), which has a [PYD-BBox-B30.2] structure (Chae et al. 2000).

Structurally, the B30.2 domain is a β-barrel like the immunoglobulin domain. It forms a β-sandwich of two antiparallel β-sheets, made up of the PRY and SPRY subdomains (the latter of which can also be found independently in certain types of proteins) and connected by loops that create ligand-binding regions at the top of the domain (Biris et al. 2012; D'Cruz et al. 2013; Kelley et al. 2005; Munoz Sosa et al. 2021). These loops constitute hypervariable regions and define the specificity of the interactions with other proteins, thus specifying variations of the function of B30.2-containing receptors. The variability of the ligand-binding domain has been exploited for pathogen binding and immune functions of B30.2-containing proteins, as indicated by the evolution of loop sequences under strong diversifying selection in antiviral TRIMs such as primate TRIM5a (Newman et al. 2006; Sawyer et al. 2005) or zebrafish finTRIMs (van der Aa et al. 2009). Strikingly, subsets

✉ Jaanus Suurväli

✉ Pierre Boudinot
pierre.boudinot@inrae.fr

[1] Department of Biological Sciences, University of Manitoba, 50 Sifton Rd, Winnipeg, MB R3T 2N2, Canada

[2] Université Paris-Saclay, INRAE, UVSQ, Virologie et Immunologie Moléculaires, 78350 Jouy-en-Josas, France

of B30.2 domain-containing proteins are prone to repeated large expansions during evolution, a feature well-documented in, but not necessarily restricted to, teleost fish TRIMs and NLRs (Boudinot et al. 2011; Laing et al. 2008; Stein et al. 2007).

TRIM proteins are defined by the association of a C3HC4-type zinc finger (RING finger, short for Really Interesting New Gene) with one or two B-Box zinc fingers and coiled coil motifs, leading to their alternative abbreviation RBCC (RING-Bbox-Coiled Coil). They generally contain additional domain(s) at the C-terminus, which define the TRIM classes (Fig. 1A) (Short and Cox 2006). A C-terminal B30.2 domain is present in two classes of TRIM proteins: class I TRIMs in which it is associated with a fibronectin domain (RBBCC-Fn3-B30.2), and class IV TRIMs where the B30.2 domain is connected to the coiled coil motif (RBBCC-B30.2) (Fig. 1A). Both class I and class IV include important antiviral factors. The prototype member of class I is TRIM1; the mouse TRIM1 has antiretroviral activity against murine leukemia virus and activates NF-κB and AP-1 signaling (Uchil et al. 2013). Several class IV TRIMs play important roles in antiviral defense, and they can employ several different mechanisms to do so. For example, TRIM5a is a restriction factor of HIV1 in monkey cells, while TRIM25 is required for RIG-I sensing of viral RNA and TRIM27 and TRIM21 control the IRF3/IRF7 axis (reviewed in Ozato et al. (2008)).

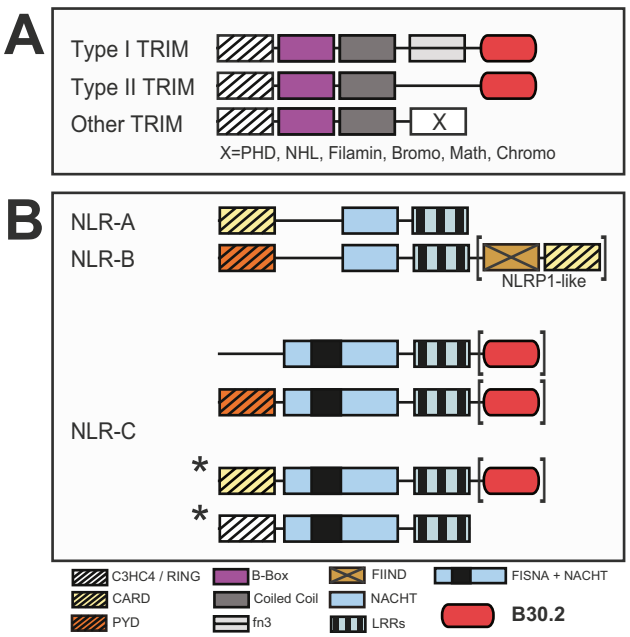


Fig. 1 Typical domain structures of TRIMs and NLRs in vertebrates. **A** All TRIMs share the structure of RING-BBox-CC, but class I and class IV TRIMs are the only ones with a C-terminal B30.2. In class I TRIMs, a fibronectin (fn3) domain is also present. **B** All NLRs share the central structure of NACHT-LRRs, but only NLR-C genes have the FISNA extension for NACHT and can have a C-terminal B30.2. Asterisks mark NLR-C domain structures not present in zebrafish, including the novel CARD-NLR-C-(B30.2) genes and the NLRs with an N-terminal C3HC4/RING. Domains that can optionally be either present or absent in different members of the family are surrounded with brackets

Class I TRIM genes usually have only one or a few copies per genome, are highly conserved across vertebrates, and are present in basal metazoans such as placozoans and cnidarians (Suurvali et al. 2014). In contrast, the repertoire of class IV TRIM genes shows large variations within vertebrates with frequent loss, duplication, and degeneration (Sardiello et al. 2008). Expansions of specific TRIM genes have been reported across vertebrates, such as the expansion of *trim64* in bovine species and large expansions of several class IV TRIMs in zebrafish, involving finTRIMs (*ftr*), bloodthirsty-like TRIMs (*btr*), and *trim35* (Boudinot et al. 2011). While these three subsets are also expanded, to variable extents, in other teleost species (such as pufferfish, rainbow trout, and medaka), other independent expansions of class IV TRIM genes have been identified in the coelacanth, a species much more closely related to terrestrial vertebrates (Boudinot et al. 2011, 2014). Furthermore, hypervariable loops at the top of the B30.2 domain of *ftr*, *btr*, and *trim35* expansions showed signatures of positive selection, suggesting that they have evolved to bind variable ligands (Boudinot et al. 2014; van der Aa et al. 2009). finTRIMs are induced by viral infection and type I IFN in rainbow trout and zebrafish (Aa et al. 2012; van der Aa et al. 2009), implicating their likely involvement in antiviral immunity similar to many mammalian class IV genes. Furthermore, a strong antiviral activity through type I IFN induction has been demonstrated for the zebrafish finTRIM (Langevin et al. 2017) and for other class IV fish TRIMs (Wang et al. 2017). Other members of this class have regulatory activities, such as FTRCA1 in crucian carp (Wu et al. 2019a).

Altogether, these examples illustrate the fast dynamics of the evolution of TRIM-B30.2 genes. Interestingly, there appears to be more than one mechanism involved in the TRIM gene expansions of fish: medaka *ftr* genes contain no introns, suggestive of the involvement of retrotransposon activity. On the other hand, in zebrafish and pufferfish, members of the *ftr*, *btr*, and *trim35* subsets tend to have a clear, conserved exon–intron structure (Boudinot et al. 2014). Overall, these observations suggest that genes encoding class IV TRIMs are highly prone to expansions. However, their diversity has been mainly characterized in zebrafish and a few other fish species, and the evolution of the repertoire across teleosts is poorly documented.

The other family of proteins that are often associated with B30.2 in fish is known as the NLRs. They are structurally very similar to the resistance genes of plants, proteins that are associated with both frequent expansions and segregating haplotypes with variable copy numbers (Jones et al. 2016; Van de Weyer et al. 2019). Similarly, in fish

and different invertebrates (sea urchin, amphioxus, sponges, etc.), there are notable expansions of NLR proteins (Huang et al. 2008; Yuen et al. 2014). The most conserved vertebrate NLRs carry out essential functions for the immune response and include NOD1 and NOD2 (cytoplasmic receptors for intracellular pathogens), NLRC5 and CIITA (transcription factors for Class I and II MHC, respectively), and most of the central genes for inflammasome assembly and regulation (NLRC3, NLRC4, NLRP if genes present in the genome) (Kim et al. 2016). In fish, some individual NLRs have been associated with inflammasome activity and pyroptosis as well (Chen et al. 2020; Kuri et al. 2017; Li et al. 2020b; Morimoto et al. 2021; Yang et al. 2018; Zhang et al. 2020).

At the core of nearly all vertebrate NLRs is the NACHT domain (Fig. 1B), which is a P-loop NTPase that is rarely found anywhere else (one exception is the telomerase component TEP1). The two key elements of NACHT are short protein motifs known as Walker A and Walker B, the former of which is so conserved that in some species (e.g., zebrafish) it contains enough phylogenetic signal to accurately distinguish between NLR subtypes (Howe et al. 2016). A typical Walker A motif of a fish NLR has the sequence G.AG.GK[TS]. The NACHT domain is often encoded as part of a large 1.8-kb exon that also codes for helical structures in the protein. This is then followed by a group of much smaller exons encoding leucine-rich repeats, resulting in the characteristic NACHT-LRR structure (Fig. 1B). Additional domains can be present in both C- and N-termini, most frequently DEATH-type protein interaction domains (more specifically, CARD and PYD), small accessory domains such as FISNA or FIIND (Fig. 1B), or short repeat sequences (Howe et al. 2016). It is generally thought that NLRs use their C-terminal domains for ligand binding and pathogen recognition, NACHT domain for oligomerization, and N-terminal domains for protein interactions and effector functions (Proell et al. 2008). Most mammalian NLRs are classified into NLRC (or NLR-A) and NLRP (or NLR-B) genes, depending on whether they are attached to an N-terminal CARD or PYD (Laing et al. 2008) (Fig. 1B).

Similar to TRIMs, early studies found a massive expansion of NLRs associated with B30.2 from the genome of zebrafish (> 400 copies, all of which also contain an N-terminal FISNA domain) (Howe et al. 2016; Laing et al. 2008; Stein et al. 2007). Based on sequence and structural similarities, this class of NLRs is also known as the NLR-C genes (Laing et al. 2008). However, not all NLR-C genes contain a C-terminal B30.2, as it is restricted to specific subsets and even then not always present (Adrian-Kalchhauser et al. 2020; Howe et al. 2016). In the N-terminus of NLR-C genes, additional effector domains can optionally be either present or not (Fig. 1B), although this also appears to be affected by the exact subset that the NLR belongs to (Howe et al. 2016). The exact role of B30.2 in NLR-C is unclear; however, it has been suggested that it is under positive selection and involved in pathogen recognition (Howe et al. 2016). In the Japanese flounder (*Paralichthys olivaceus*), an NLR-B30.2 with no other clear additional domains is a positive regulator of ATP-induced proinflammatory cytokine expression, suggestive of being involved in inflammasome activity even without possessing a clear N-terminal effector domain (Li et al. 2016b).

The expansion of NLR-C genes is thought to initially originate from the common ancestor of *Clupeocephala*, a superclass containing nearly all teleost species other than eels, tarpons, and bonytongues, but has probably been ongoing since. Later studies have led to estimates of NLR copy numbers in the genomes of additional species of fish, including round goby (353 genes), threespine stickleback (320 genes), Atlantic cod (178 genes), rainbow trout (157 genes), common carp (153 genes), channel catfish (estimates ranging from the initial 22 to more recent 160 genes), miiuyi croaker (48 genes), large yellow croaker (43 genes), fugu (estimates ranging from the initial 16 to more recent 76 genes), spotted gar (32 genes), turbot (29 genes), and tongue sole (29 genes) (Adrian-Kalchhauser et al. 2020; Ao et al. 2015; Biswas et al. 2016; Chen et al. 2021; Howe et al. 2016; Li et al. 2016a; Marancik et al. 2014; Rajendran et al. 2012; Schiffer et al. 2016; Torresen et al. 2018; Zhang et al. 2021). Additionally, some transcriptomic studies have provided the number of identified unigenes associated with NOD-like receptor signaling, for example 385 in the Siamese fighting fish (Amparyup et al. 2020) and 119 in Dabry's sturgeon (Chen et al. 2019). However, there is only one systemic survey of NLRs across fish genomes, and it included cod, haddock, and all of the 10 species with genomes available via Ensembl at the time (Torresen et al. 2018).

In this study, we made a large-scale survey of the repertoires of these receptors across fish species, using publicly available high-quality long-read-based full fish genomes, including those sequenced in the frame of the Vertebrate Genomes Project (Rhie et al. 2021; https://vertebrategenomesproject.org/). While a few species of *Agnatha* (jawless fishes) and *Chondrichthyes* (cartilaginous fishes) were included, we mainly focused on ray-finned fishes. We set up an approach to assess the number and genomic distribution of domains and domain combinations associated with TRIMs, NLRs, and other genes containing B30.2 domains independently of whether a full genome annotation is available or not, and used it to look for gene expansion patterns across taxonomic fish subsets. Finally, we created a model to test the impact of taxonomy, genome size, and environmental variables (geographic coordinates, and habitat in either freshwater or marine environments) on the copy numbers of these genes.

Materials and methods

Description of the dataset

We downloaded a total of 153 genome assemblies from NCBI Assembly, each selected from a different genus. For most genera, we selected either a reference genome or the latest available assembly, preferentially sequenced with long reads (PacBio or Oxford Nanopore), and with a contig N50 of at least 100 kb. For genomes with both RefSeq and GenBank versions available, we used the RefSeq assembly. The full list of all selected species with assembly accessions and references, genome parameters, and any additional data we collected is available in Supplementary Table 1. In the final dataset, there were only 6 assemblies with N50 < 100 kb (inshore hagfish, Pacific lamprey, coelacanth, spotted gar, turquoise killifish, and the tiny cyprinid *Paedocypris*), which were kept for their particular taxonomic position, and 102 of the 153 genome assemblies are chromosome-level. Of these 153 genomes, 145 are from species considered to be fish, i.e., defined here as non-tetrapod vertebrates. They belong to *Agnatha* (jawless fish), *Chondrichthyes* (cartilaginous fishes), *Actinopterygii* (ray-finned fishes, including teleosts), and *Sarcopterygii* (lobe-finned fishes, excluding tetrapods).

Spatial data describing the ranges of all fish species in the study were obtained from FishBase (Froese and Pauly 2021) and the database of the International Union for the Conservation of Nature's (IUCN) Red List of Threatened species (IUCN 2021). These databases contain spatial data on different (partially overlapping) subsets of the selected species. The two datasets were merged, and each species was associated with the central point of each range, determined from the minimum and maximum values available for longitude and latitude, which was then visualized by using the R packages ggplot2 v3.3.5, rnaturalearth v0.1.0, and sf v1.0–2 (Pebesma 2018; South 2017; Wickham 2016) (Fig. 2). One caveat of this approach is that species with a circumboreal, circum-Antarctic, or circum-global distribution all end up with the central point at a longitude of 0. For the handful of species with no spatial data available from either database, we obtained the min/max range values from the Global Biodiversity Information Facility (www.gbif.org), by limiting their available datapoints to countries that the species is known to naturally occur in.

For three of the genome assemblies, the exact species was not available; in those cases, we used spatial data from species of the same genus inhabiting the same geographic range instead:

Genome: *Coregonus* sp. "Balchen" – Data: *Coregonus alpinus.*

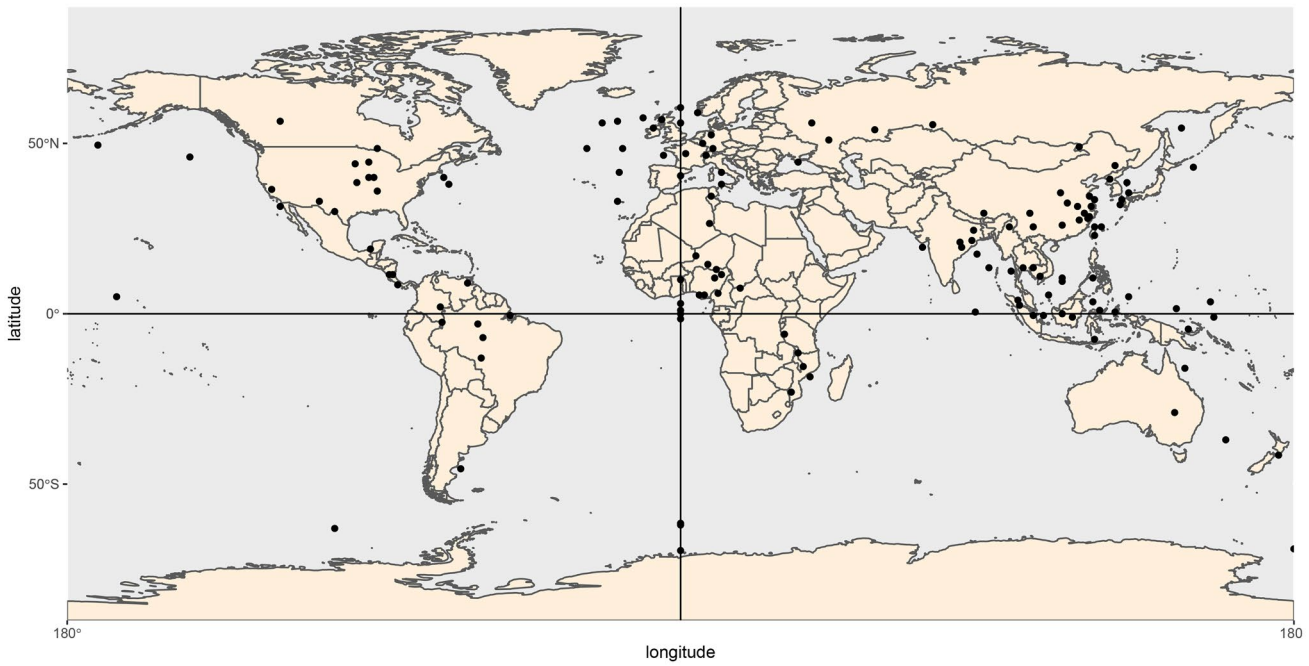


Fig. 2 Central points for the geographic ranges of each species of fish in the study. Each black dot represents one species. Note that the dots do not show sampling sites of sequenced individuals. Instead, they indicate that the area from which other fish of that species could be sampled from is centered on that point. In our dataset, almost all dots mapped to the vertical black line in the center where longitude = 0 correspond to species with a circum-global/-polar/-Antarctic range, with one exception: the common dragonet who inhabits the Atlantic basin at longitudes ranging from –32 to + 32

Genome: *Paedocypris* sp. Pulau Singkep – Data: *Paedocypris progenetica.*

Genome: *Pseudoliparis* sp. Yap Trench– Data: *Pseudoliparis amblystomopsis.*

Data on the preferred habitat (freshwater/marine) and the maximum length of the species were primarily obtained from FishBase (Froese and Pauly 2021). All species that can be found in both marine and freshwater environments, including migratory fish, were associated with both categories. Among the 145 fish species of the dataset, 62 and 57 were associated with either only freshwater or only marine environments, respectively. Twenty-six species were associated with both.

A phylogenetic tree of all selected species was obtained from the NCBI by using the R package *taxize* (version 0.9.99 (Chamberlain and Szocs 2013)).

Assessing gene copy numbers

Rather than look for gene copies from annotated proteomes, we chose to analyze genome sequences directly to standardize our approach. This also allowed us to analyze the many recently sequenced genomes for which the protein annotation does not exist yet. As a first step, *transeq* from the EMBOSS suite of bioinformatic tools (Rice et al. 2000) was used to translate the downloaded 153 genomes in all six reading frames (3 on the forward strand, 3 on the reverse strand), replacing stop codons with the character "X". For the genomes of human and zebrafish, contigs labeled as alternative haplotypes (e.g., "ALT_CONTIG") were removed.

We thus used *hmmsearch* from *HMMER3* v3.1 (Finn et al. 2010) on the translated genomes to retrieve the position of most domains that can be associated with TRIMs or NLRs. We did not search for leucine-rich repeats, as they are numerous in the genome, difficult to detect, and usually follow all vertebrate NACHT domains. Most of the hidden Markov models for protein domains were obtained from the Pfam-A database; however, we used only the selected domains and not the entire database for our searches. In addition, we included the "zf_B30.2" and "zf_FISNACHT" models that had been previously generated from zebrafish FISNA-NACHT and B30.2 sequences (Adrian-Kalchhauser et al. 2020). We disabled the composition bias filter and corrections of hmmsearch to reduce the chances of missing the zinc finger domains in TRIMs.

The final list of domain models used is as follows: zf_B30.2 (no PFAM id), zf_FISNACHT (no PFAM id), APAF1_C (PF17908.2), NACHT (PF05729.13), NB-ARC (PF00931.23), NLRC4_HD (PF17889.2), NLRC4_HD2 (PF17776.2), NOD2_WH (PF17779.2), Peptidase_C14 (PF00656.23), PRY (PF13765.7), SPRY (PF00622.29), PYRIN (PF02758.17), CARD (PF00619.22), CARD_2 (PF16739.6), zf-B_box (PF00643.25), zf-C3HC4 (PF00097.26), zf-C3HC4_2 (PF13923.7), zf-C3HC4_3 (PF13920.7), zf-C3HC4_4 (PF15227.7), zf-C3HC4_5 (PF17121.6), zf-RING_UBOX (PF13445.7), fn3 (PF00041.22), FISNA (PF14484.7), and FIIND (PF13553.7).

One of the main output files of *hmmsearch* was a table collecting domain hits, their associated *e*-values, and their coordinates in input sequences. We filtered this table with custom bash scripts by setting an *e*-value threshold of 1e − 5 and a minimum match length of 30 amino acids, which is shorter than any of the domains we were looking for. We also removed all matches from domains shorter than 100 amino acids for which the match was shorter than 90% of the model length and for which hmmsearch had reported a score value of "inf." We then converted the position in translated sequence back to genomic coordinates (for forward reading frames) or to reverse genomic coordinates (for reading frames on the opposite strand, starting from the last base pair of the chromosome as position "1").

The number of domains to analyze was then reduced by removing all cases of zf_FISNACHT, NLRC4_HD, NLRC4_HD2, NOD2_WH, Peptidase_C14, APAF1_C, and NB-ARC, as an initial analysis suggested that the NACHT domain from Pfam-A alone was indeed sufficient to detect the presence of the NLRs. At this point, the filters had produced lists of the following domain categories:

1. NACHT
2. B30.2 (actually, combining all matches from the models of PRY, SPRY, and zf_B30.2)
3. FISNA
4. FIIND
5. B-Box
6. C3HC4 (matches from the models of RING_UBOX and the C3HC4 subtypes)
7. fn3 (fibronectin 3)
8. CARD (matches from the models of CARD and CARD_2)
9. PYD (PYRIN)

We noticed many cases in which the data had partial domains of the same category right next to each other on the same strand, a pattern likely caused by a mixture of pseudogenization and sequencing errors introducing frameshifts. Because we were not able to distinguish sequencing errors from actual frameshifts based on genome sequence alone, we used *bedtools merge* to join all such cases of adjacent same-type-domain predictions, if separated by 30 base pairs or less.

In the final step, we used a custom R script and the R package *data.table* v1.12.8 (Dowle and Srinivasan 2019) to identify all cases in which the domains of interest were

positioned in an order consistent with TRIMs or NLRs, while being separated from each other by a distance of no more than 100 kb. An exception to this rule is the distance between FISNA and NACHT domains—we limited it to 1 kb or less, because these two domains are always in the same exon, and a FISNA without NACHT is not an NLR. From the per-contig domain counts, we also calculated (for each fish species) and plotted a measure of clustering: the minimum number of contigs containing at least 25% of the domains.

Modeling strategy

TRIMs and NLRs have key functions in the immune responses, and as such likely evolve under strong selective pressures elicited by a combination of pathogens and environmental parameters. However, copy numbers could also be driven by much more general mechanisms, e.g., the fluctuations in the size of the genome itself. In addition, simple correlations between two observable variables are probably not sufficient to explain the data since the phylogenetic context needs to be considered as well. Copy number variation will also be well predicted by shared evolutionary histories among species. We therefore chose to model the process of NLR and TRIM gene copy number variation with a linear mixed model approach that had the phylogenetic context included as a background effect. This allowed us to test the explanatory power of different variables accessible to us (maximum fish length, genome size, geographic coordinates, preferred habitat) while controlling for shared ancestry among species in our model by fitting a phylogeny.

In order to retain only the highest quality information for the modeling, all assemblies without full chromosomes were excluded from further analyses. Furthermore, we also removed all assemblies that had not involved long-read sequencing (Pacific BioSciences or Oxford NanoPore). Short-read-based assemblies often struggle with repetitive sequence, and gene families with hundreds of closely related paralogs can be considered repetitive elements in the genome.

All modeling was done with the tool *gls* (generalized least squares) from the R package *nlme* (Pinheiro et al. 2021). For phylogenetic background, we used the NCBI taxonomy obtained in the previous step as input and told the tool to estimate the effect of the tree on the data (Pagel's lambda/λ) before using it for a final model. λ is a parameter that varies from 0 to 1; in our case, 0 equals the tree having no effect on the data and 1 equals a strong effect of taxonomy that assumes a Brownian model of trait evolution. Different lambda values result in different model likelihoods and different estimates for the effect of predictor variables; the final λ is the one producing maximum likelihood for the model. Plots showing model likelihoods obtained with different λ values are presented in Supplementary Data 1.

For a few domain types, automatic λ estimation was not successful and produced errors or values falling outside the 0.0.1 range. These included the counts for B-Box domains and PYD-NLRs. In both cases, the λ value for final modeling was then selected by calculating model likelihoods for 1000 lambda values ranging from 0 to 1, then choosing the lambda values associated with maximum values and checking their sensibility from the plots mentioned above.

We were primarily interested in TRIM and NLR families as a whole (and we had the copy numbers for them), but in the previous step, we had also obtained the copy numbers for various different subsets of TRIMs and NLRs. We decided to model the drivers of copy number variation for all of them, with the reasoning that a truly robust effect should be present not only in the entire set of TRIMs/NLRs, but in the various subsets as well, providing us with an additional line of evidence similar to how biological replicates do.

We calculated the likelihoods of models including different combinations of distance from the equator (in degrees; the absolute value of latitude), longitude, maximum fish length, genome size, and habitat. The habitat information was encoded on a scale from − 1 to 1, with − 1 corresponding to freshwater and + 1 to marine fish. Stickleback, salmon, and other migratory/highly adaptable species that can be found in both were assigned an intermediate value of 0.

We tested different combinations of predictors on TRIM-B30.2 and NACHT-B30.2 copy numbers and found that not including the phylogeny lowers all model likelihoods greatly. The highest log likelihoods were produced by the following three models (showing the domain on the left and predictors on the right, with phylogenetic correction included into all three).

Genome_size + equator distance + longitude + habitat: − 114.4366 for NLR-B30.2, − 115.097 for TRIM-B30.2

Genome size + equator distance + habitat: − 114.4392 for NLR-B30.2, − 115.1087 for TRIM-B30.2

Genome size + longitude + habitat: − 114.4369 for NLR-B30.2, − 118.7815 for TRIM-B30.2

Anova test showed the first two options to be equivalent for TRIM-B30.2 and NACHT-B30.2, so we chose the one requiring less parameters (genome size + distance from the equator + habitat) and then applied the model to all domain combinations observed.

After running the models, an effect of the predictor on copy numbers was considered significant if the 95% confidence threshold of the parameter estimate did not overlap the value 0 (no effect), corresponding to a p value of 0.05 as reported by *gls* in model summaries.

Data visualization

Figure 1 was created with Adobe Illustrator. Unless stated otherwise, all our data was plotted with the R package *ggplot2* (Wickham 2016). The figures were finalized with Adobe Illustrator.

Results and discussion

A method to assess the number of genes belonging to TRIM and NLR families and their subclasses

Targeted analyses of multigene families in the genome of a given species typically involve the mining of existing annotations, complemented by blast analysis using related sequences, visual inspection of alignments, and domain searches with tools such as *HMMER3* and *InterProScan* (Buckley and Rast 2011). This is further complicated by complex exon–intron structure, leading to the genes spanning tens or even hundreds of kilobases. An important caveat of this approach is that it largely relies on the accuracy of the annotation; also, visual inspection is not possible for large numbers of species. Alternatively, one could work with transcriptomic data, preferably based on long reads and from multiple tissues/experiments, and hope that all genes of interest are expressed in it. However, for many newly sequenced species, only the genome sequence is available, requiring us to seek out alternative approaches.

Another complication is that for very large gene families, many of the duplicated genes can be recent duplicates, possibly even younger than the species itself. Such genes have nearly 100% identity to each other within the coding sequence (Howe et al. 2016), and without involving ultra-long sequencing (i.e., with the technologies spearheaded by Pacific Biosciences and Oxford Nanopore), their copy numbers are very likely to be underestimated due to assemblers collapsing the repeats to single genes. This becomes even more relevant for transcriptomic data, as introns acquire new mutations faster than the exons, making it easier to distinguish between the recent gene copies. On the transcriptome level, two or more genes with near-identical coding sequences would be often interpreted as a single gene with a higher expression.

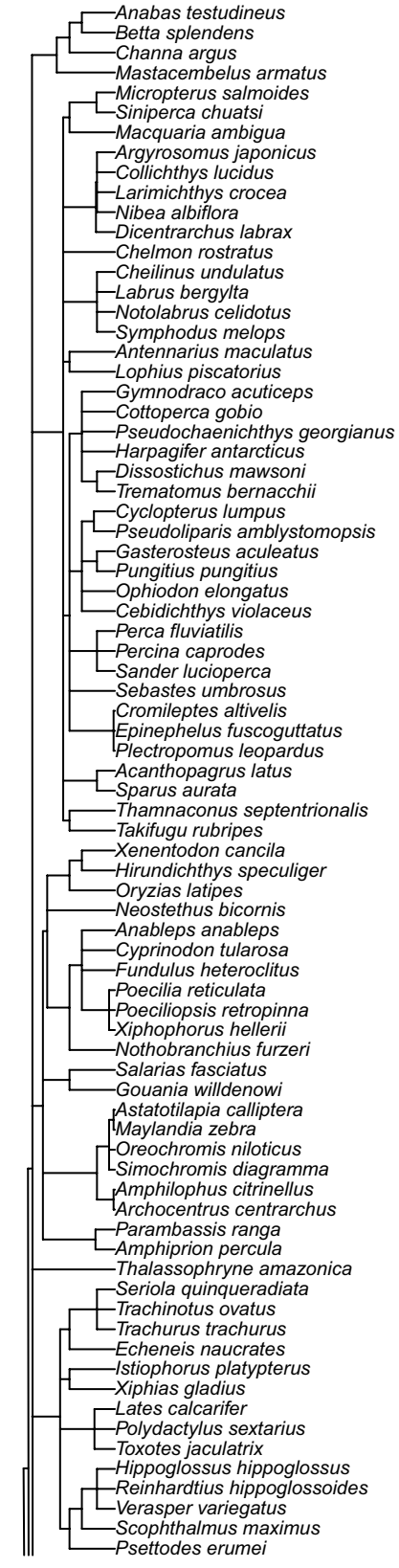
Analysis of the two large B30.2-associated immune receptor families in fish—NLRs and TRIMs—suffers from all of the complications above. While most of our knowledge about those gene families comes from just a few species (mostly zebrafish), they have both clearly had lineage-specific expansions, are associated with a very high degree of sequence similarity between paralogs, and have a complex exon–intron structure. It is possible to focus on species for which a genome assembly based on long-read sequencing is available, but many of those recent assemblies are not associated with any data beyond the genome sequence itself.

To compare TRIM and NLR repertoires across a large number of species based on available genome sequences alone, we therefore developed a method based on in silico translation and analysis of entire genome scaffolds and chromosomes, followed by domain searches with *hmmsearch*. A phylogenetic tree (taxonomy) of all analyzed species is presented in Fig. 3.

By defining a gene candidate as a stretch of sequence containing canonical combinations of TRIM- and NLR-associated domains in a defined order, we were able to assess both the total numbers and clustering of these gene families without having to rely on pre-existing protein annotations. Downstream filtering of the predicted domains improved our results further. To count the total number of NLRs, we followed previous studies which have shown that simply counting the number of NACHT domains is already sufficient to get a good estimate (Adrian-Kalchhauser et al. 2020; Torresen et al. 2018). When tested in human, this approach identified all 22 well-known NLR genes in human, but also the related highly conserved genes *NWD1* and *NWD2* (consisting of NACHT and WD repeats), as well as *NLRP2B* and two pseudogenes of *NAIP*.

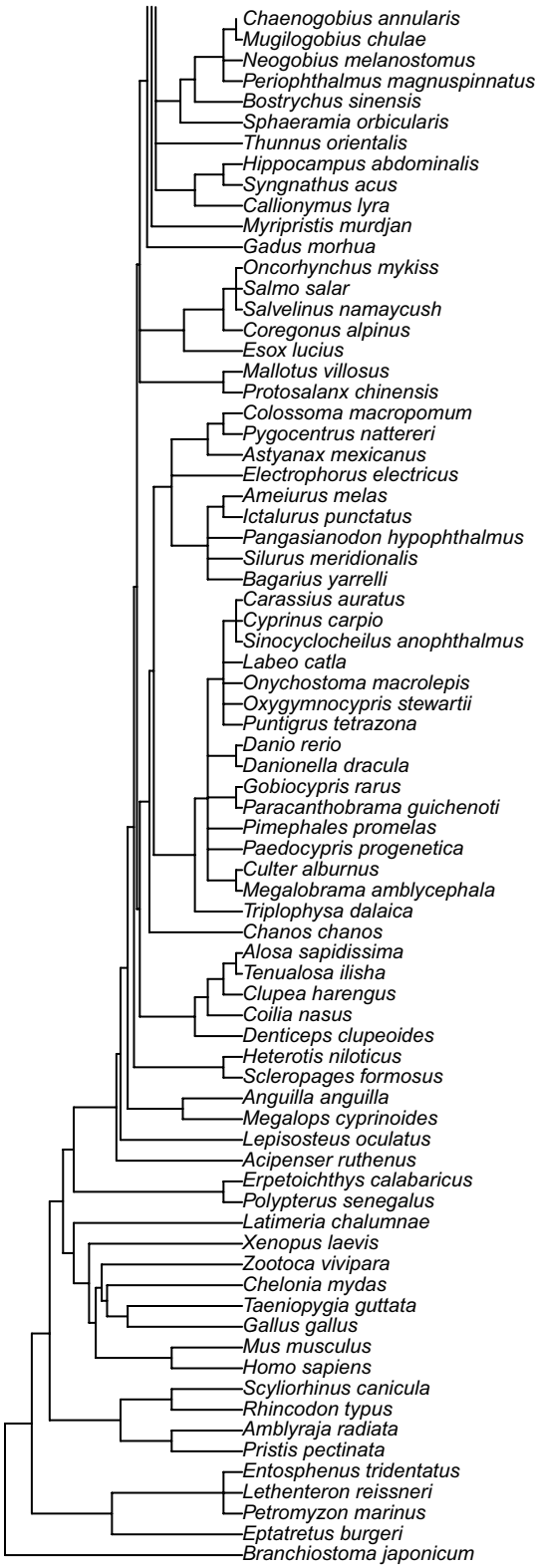
However, the short zinc finger motifs that form the core of TRIM proteins (B-Box and RING) can also be found elsewhere in the genome and are not always detected by all domain annotation tools. Our TRIM counts were thus obtained by counting domain combination occurrences, rather than individual domains. Our estimates for TRIM copy numbers without B30.2 are based on the number of occurrences of C3HC4 (RING finger) followed by B-Box without a B30.2. The estimates for TRIM numbers with B30.2 are based on at least one of these two zinc finger domains having been detected as preceding a B30.2 domain, with no NLR-related domains between them. We considered the sum of these counts (C3HC4-BBox, C3HC4-B30.2, BBox-B30.2) to be a good estimate of the total number of TRIMs in the genome. Nevertheless, we also counted the total number of B-Box domains per genome, and the summary value of this and the number of RING fingers followed by a B30.2 (without NLR-related domains in between). These values are presented in Supplementary Table 1 (columns "B-Box" and "TRIM_OR_B-Box"), along with counts for some other domain combinations such as FISNA-NACHT.

While our approach could be applied to any gene family, it does come with some limitations. First, we do not really annotate genes and their exons and instead only find the positions of specific domains in them. This allows us to estimate copy numbers and to use the coordinates to retrieve the sequence of these specific domains, but is not an actual full gene annotation. Second, our approach does not distinguish



--- Continued on the right ---

--- Continued from the left ---



◂**Fig. 3** Taxonomic relationships of the species used for the study. The top half of the tree is presented on the left side and contains only members of the clade *Euteleostei*, including stickleback, perch, cichlids, Antarctic icefish, croakers, anglerfish, pufferfish, flatfish, guppy, medaka, killifish, and ballan wrasse. The bottom half of the tree is on the right side and contains some more euteleosts (e.g., gobies, salmonids, pike, cod, capelin, seahorse, and its close relatives), but also catfish, cyprinids (zebrafish, carp, goldfish), herring, eels, bonytongues, bichirs, sturgeon, garfish, latimeria, tetrapods, sharks and rays, lampreys, hagfish, and, as an outgroup, the Japanese amphioxus. Based on the NCBI taxonomy database

between pseudogenes and functional genes, as evidenced by us finding the human *NAIP* pseudogenes. However, we took measures to not accidentally count pseudogenes more than once: we included an extra step in the analysis pipeline to merge adjacent partial gene models, usually originating from a frameshift that resulted in two different reading frames having a partial match for the domain. As the outcome of our approach was close to the known numbers for the repertoires of TRIMs and NLRs in humans, zebrafish, and other species (e.g., TRIM counts in fugu), we conclude that our results represent good approximations from which we can assess and compare gene repertoires across many fish genomes.

Striking variation of TRIM and NLR gene numbers across fish orders

As a result of the domain search strategy discussed above, we were able to determine copy number estimates for all the different domain combinations that we had associated with NLRs and/or TRIMs. The results were visualized as heatmaps (Table 1), from which we note that class I TRIMs ("TRIM-fn3") and NLRs with a C-terminal CARD domain (similar to human NLRP1) usually have very low numbers or are not found at all. These domains are known to have only a few copies in most vertebrate genomes. We do not know whether the number "0" means that the domain/domain combination was not detected or that it was not present in genome, but it is clear that most species have only a few copies at most.

Lineage-specific expansions of class I TRIMs

While most species have very low numbers for class I TRIMS, there seem to have been several independent expansions (Table 1). Salmonids and related species (e.g., salmon, trout, whitefish, and pike) all share an increased copy number of detectable TRIM gene attached to a fibronectin domain (6–15 paralogs per genome). The other group of fish exhibiting a similar pattern is *Clupeiformes* (herring and related species, 4–15 copies per genome). Our outgroup/control species *Branchiostoma japonicum* (Japanese amphioxus) has 6 copies as well. The highest numbers (21 copies) are found in the genome of the marine smelt capelin (*Mallotus villosus*), an important forage fish in the northern Atlantic ecosystem (Table 1).

Ancient origin of class IV TRIMs

The categories of "TRIM" and "TRIM-B30.2" have consistently close numbers (Table 1), indicating that the majority of TRIMs in nearly all analyzed species contain a C-terminal B30.2 domain. The only exception is the amphioxus, in whom only 9 TRIMs out of 183 contain a B30.2 (Table 1). Three of those are class IV TRIMs, as they do not possess an fn3 domain. Class IV TRIMs have never been described outside of vertebrates before to our knowledge; however, *Branchiostoma belcheri*, another species of amphioxus, has at least one official annotation of what does appear to be a class IV TRIM (NCBI gene ID: 109,472,165). Class IV TRIM-B30.2 genes were therefore most likely present at the common vertebrate ancestor. Furthermore, class IV TRIM-B30.2 gene expansion is not only seen in *Gnathostoma*, but also in multiple species of lamprey (e.g., 114 TRIM-B30.2 gene in sea lamprey and 203 in the Far Eastern brook lamprey, both of which appear to be class IV TRIMs without an fn3 domain) (Table 1). Class IV *trims* have never been described before in these species. Taking into account that lampreys most likely diverged from their common ancestor with jawed vertebrates between two whole genome duplications events (2R WGD) (Nakatani et al. 2021), it is possible that having high numbers of class IV TRIM genes had already become beneficial for the immunity by the end of the first round of genome duplications, or perhaps even before that. It is of particular interest that this immune expansion is shared by species with vastly different immune systems, since the LRR-based adaptive arm of jawless vertebrate immunity is quite unique and generates antigen receptor gene diversity with mechanisms that are different from jawed vertebrates (Kasahara and Sutoh 2014).

In any case, it seems fish have a lower limit on TRIM and in particular TRIM-B30.2 gene copy numbers, as the copy numbers of neither TRIMs in general nor TRIMs with B30.2 never drop to values close to 0 (even the big-belly seahorse has 41 detectable TRIM-B30.2 genes in its genome). For NLRs, if such a threshold does exist, it is much lower, as the seahorse genome has only 3 NLRs detectable by *HMMER3*, one of which is with a B30.2.

NLR-B30.2 genes represent a ray-finned fish expansion

While previously it has been proposed that NLR-C genes first acquired a B30.2 domain in the common ancestor of the *Neopterygii* subclass of ray-finned fish (Howe et al. 2016), here, we could date the event further back in time. We find

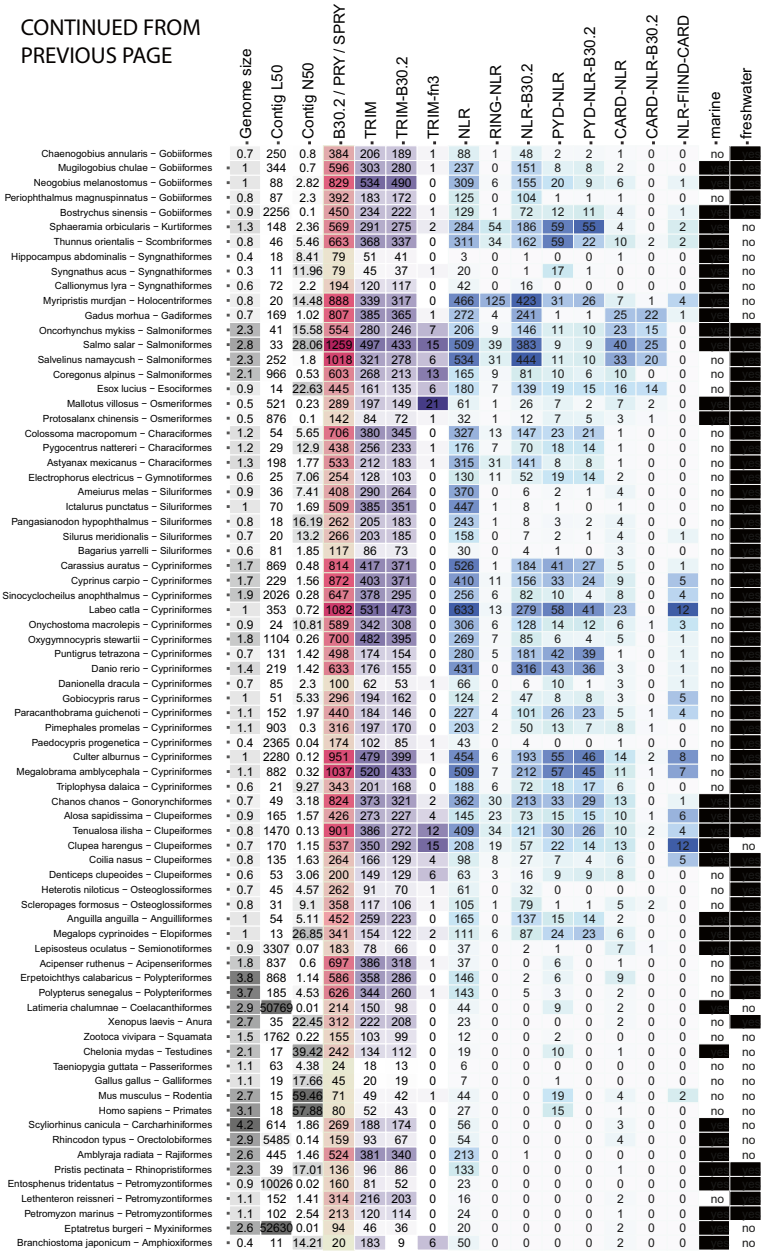
Table 1 Total counts of different protein domain combinations in 153 species, visualized as heatmaps. Each individual column contains a different heatmap with its own color scale. Color tones used for visualization represent the different types of data. The first column contains species and order names, sorted according to the taxonomic order of the tree in Fig. 3. Columns 2–4 (grayscale): general genome parameters. Column 5 (red): total count of all matches to the models of PRY, SPRY, and B30.2 in the genome. Columns 6–8 (purple): TRIM-associated counts. TRIM here is defined as a detectable RING domain (C3HC4) followed by either a B-box or a B30.2/PRY/SPRY (to include cases where detection fails for the B-box), or a B-box followed by B30.2/PRY/SPRY regardless of whether a preceding RING is found by the software or not. TRIM-B30.2 refers to the subset of TRIMs that contain a B30.2 domain. TRIM-fn3 refers to the class I TRIMs that include a fn3 and a B30.2 domain. Columns 9–16 (blue): NLR-associated counts. NLR here is defined as a detectable NACHT domain, regardless of whether FISNA is present or not. Data on FISNA-NACHT domains specifically (NLR-C counts) is available in Supplementary Table 1, and the numbers are generally close to the total NLR numbers. Each column contains what we defined as NLR in combination of some of the other associated domains in a specific order, given by name of the column. Columns 17–18 (grayscale): data on the habitat of the species. Most fish live either in freshwater or in marine environments, some are capable of living both, demonstrated by either having populations in both or by migrating from one to the other. This was color-coded into two columns that give a yes/no answer to the question of whether the fish lives there. Black = yes, white = no

	Genome size	Contig L50	Contig N50	B30.2 / PRY / SPRY	TRIM	TRIM-B30.2	TRIM-fn3	NLR	RING-NLR	NLR-B30.2	PYD-NLR	PYD-NLR-B30.2	CARD-NLR	CARD-NLR-B30.2	NLR-FIIND-CARD	marine	freshwater
Anabas testudineus – Anabantiformes	0.6	25	7.06	430	208	188	0	257	34	163	30	27	6	1	1	no	
Betta splendens – Anabantiformes	0.4	48	2.5	417	127	112	3	342	80	220	13	9	5	1	2	no	
Channa argus – Anabantiformes	0.6	18	11.79	470	216	206	1	245	43	171	11	10	4	2	1	no	
Mastacembelus armatus – Synbranchiformes	0.6	25	8.01	260	159	142	1	169	50	55	32	6	5	1	1	no	
Micropterus salmoides – Centrarchiformes	1	202	1.23	628	277	268	0	325	91	171	26	22	15	2	1	no	
Siniperca chuatsi – Centrarchiformes	0.8	21	12.19	494	266	257	0	209	59	117	20	19	3	1	0	no	
Macquaria ambigua – Centrarchiformes	0.7	757	0.24	273	149	137	0	96	18	48	16	8	3	1	0	no	
Argyrosomus japonicus – NA	0.8	19	13.11	600	348	333	1	347	136	117	49	42	9	2	6		no
Collichthys lucidus – NA	0.9	210	1.1	517	280	272	0	199	51	76	18	16	8	2	5		no
Larimichthys crocea – NA	0.7	69	2.83	489	305	297	0	233	76	94	32	24	11	3	8		no
Nibea albiflora – NA	0.6	39	4.42	378	206	197	0	181	39	84	21	19	5	1	3		no
Dicentrarchus labrax – NA	0.7	19	12.68	581	326	308	0	266	69	162	28	23	9	3	0		
Chelmon rostratus – Chaetodontiformes	0.6	15	16.96	207	112	103	0	74	17	53	4	4	3	1	1		no
Cheilinus undulatus – Labriformes	1.2	21	16.48	476	296	283	0	177	20	106	3	2	2	0	1		no
Labrus bergylta – Labriformes	0.8	282	0.7	1261	738	710	0	643	246	143	1	1	2	0	0		no
Notolabrus celidotus – Labriformes	0.8	64	3.75	388	242	238	0	159	14	100	0	0	2	0	0		no
Symphodus melops – Labriformes	0.6	343	0.46	545	397	378	0	224	63	60	1	0	8	1	0		no
Antennarius maculatus – Lophiiformes	0.5	17	10.11	87	48	43	1	17	0	4	2	1	2	0	0		no
Lophius piscatorius – Lophiiformes	0.7	1980	0.1	166	63	48	0	24	0	11	2	1	1	0	0		no
Gymnodraco acuticeps – Perciformes	1	503	0.53	477	334	313	0	147	59	45	12	8	5	0	1		no
Cottoperca gobio – Perciformes	0.6	27	6.33	282	203	190	0	75	13	20	5	5	1	0	0		no
Pseudochaenichthys georgianus – Perciformes	1	431	0.66	426	287	273	0	106	38	32	6	2	4	0	0		no
Harpagifer antarcticus – Perciformes	0.9	246	1.05	380	261	244	0	119	35	43	9	9	6	0	0		no
Dissostichus mawsoni – Perciformes	0.9	69	3.02	973	708	672	0	279	102	120	23	22	7	0	5		no
Trematomus bernacchii – Perciformes	0.9	160	1.4	756	526	502	0	320	162	100	42	32	14	1	1		no
Cyclopterus lumpus – Perciformes	0.6	35	4.95	231	167	158	0	38	0	6	1	1	1	0	0		no
Pseudoliparis amblystomopsis – Perciformes	0.8	298	0.52	758	610	579	0	193	2	4	0	0	0	0	0		no
Gasterosteus aculeatus – Perciformes	0.5	253	0.49	494	139	130	1	351	1	168	20	14	1	0	0		
Pungitius pungitius – Perciformes	0.5	50	2.78	243	112	107	2	81	0	67	13	13	1	0	0		
Ophiodon elongatus – Perciformes	0.6	104	1.54	400	279	265	0	97	3	44	13	11	2	0	0		no
Cebidichthys violaceus – Perciformes	0.6	25	5.51	355	232	219	0	99	6	50	16	10	7	2	0		no
Perca fluviatilis – Perciformes	1	57	4.2	964	521	497	0	323	21	206	51	41	8	1	2	no	
Percina caprodes – Perciformes	1	155	1.43	961	415	397	0	385	8	299	35	35	13	0	2	no	
Sander lucioperca – Perciformes	0.9	39	6.66	849	433	412	0	344	15	233	63	57	11	0	0	no	
Sebastes umbrosus – Perciformes	0.8	24	11.45	575	440	429	0	205	19	59	13	13	19	2	0		no
Cromileptes altivelis – Perciformes	1	18	18.27	371	241	227	0	96	26	46	10	9	6	1	0		no
Epinephelus fuscoguttatus – Perciformes	1	23	13.86	466	316	300	0	143	49	69	10	9	6	1	0		no
Plectropomus leopardus – Perciformes	0.8	200	1.14	252	162	150	0	70	7	41	5	5	4	1	0		no
Acanthopagrus latus – Spariformes	0.7	18	14.88	386	235	225	0	122	25	88	18	17	4	0	0		
Sparus aurata – Spariformes	0.8	75	2.86	1062	563	547	0	475	128	300	47	43	14	2	0		no
Thamnaconus septentrionalis – Tetraodontiformes	0.5	10	22.46	265	199	192	0	219	1	17	3	3	1	0	0		no
Takifugu rubripes – Tetraodontiformes	0.4	35	3.14	117	52	46	1	253	2	27	17	17	1	0	0		no
Xenentodon cancila – Beloniformes	0.7	83	2.2	427	107	102	0	254	2	208	5	5	22	14	0	no	
Hirundichthys speculiger – Beloniformes	1	225	1.09	324	139	127	0	173	33	96	12	11	13	6	0		no
Oryzias latipes – Beloniformes	0.8	76	3.26	188	132	125	0	190	10	16	1	1	1	0	0		
Neostethus bicornis – Atheriniformes	0.8	14	23.09	662	426	413	0	265	48	148	19	5	19	14	1	no	
Anableps anableps – Cyprinodontiformes	0.9	17	17.46	362	232	220	0	104	16	59	6	4	7	4	2	no	
Cyprinodon tularosa – Cyprinodontiformes	1.1	236	1.36	365	238	224	0	125	4	58	5	4	5	1	0	no	
Fundulus heteroclitus – Cyprinodontiformes	1.2	608	0.59	712	401	391	0	373	36	182	14	5	24	2	0		
Poecilia reticulata – Cyprinodontiformes	0.7	27	8.2	434	278	265	0	160	16	79	8	5	18	3	0	no	
Poeciliopsis retropinna – Cyprinodontiformes	0.6	15	15.06	342	213	202	0	106	8	63	8	4	10	3	0	no	
Xiphophorus hellerii – Cyprinodontiformes	0.7	28	7.08	521	291	275	0	311	13	154	8	5	10	1	0		
Nothobranchius furzeri – Cyprinodontiformes	1.1	10259	0.03	149	74	66	0	43	0	14	1	1	2	1	0	no	
Salarias fasciatus – Blenniiformes	0.8	86	2.6	481	223	208	0	368	6	169	65	59	6	4	4		no
Gouania willdenowi – Blenniiformes	0.9	132	1.84	432	237	229	0	142	0	102	0	0	1	0	0		no
Astatotilapia calliptera – Cichliformes	1.7	112	4.47	1180	600	568	1	562	54	290	75	33	50	30	0	no	
Maylandia zebra – Cichliformes	1	189	1.41	1037	407	393	1	540	68	325	64	40	36	15	0	no	
Oreochromis niloticus – Cichliformes	1	96	2.92	1068	462	453	1	552	67	364	60	42	25	16	0	no	
Simochromis diagramma – Cichliformes	0.8	108	2.23	744	386	369	0	339	47	165	47	30	36	14	0	no	
Amphilophus citrinellus – Cichliformes	1	67	3.84	866	391	352	0	488	49	286	26	23	43	23	1	no	
Archocentrus centrarchus – Cichliformes	0.9	129	2.15	724	284	271	0	356	38	262	23	20	20	11	0	no	
Parambassis ranga – NA	0.6	31	5.08	594	187	174	1	413	7	286	64	58	59	49	1	no	
Amphiprion percula – NA	0.9	84	3.12	253	138	129	0	126	9	61	39	25	1	0	0		no
Thalassophryne amazonica – Batrachoidiformes	2.4	310	2.33	229	103	93	1	87	0	57	5	1	2	0	0	no	
Seriola quinqueradiata – Carangiformes	0.6	239	0.87	334	181	169	1	145	34	71	12	7	4	0	0		no
Trachinotus ovatus – Carangiformes	0.6	107	1.85	327	174	157	0	116	17	56	12	9	7	0	2		no
Trachurus trachurus – Carangiformes	0.8	37	6.26	1022	652	619	0	520	122	283	53	42	11	0	2		no
Echeneis naucrates – Carangiformes	0.5	17	12.37	262	147	127	1	271	63	79	8	6	16	8	1		no
Istiophorus platypterus – Istiophoriformes	0.6	17	11.85	134	71	66	0	37	13	17	1	0	2	0	0		no
Xiphias gladius – Istiophoriformes	0.7	36	5.25	162	81	77	0	40	7	22	4	4	1	0	0		no
Lates calcarifer – NA	0.6	78	1.92	343	198	181	0	160	43	86	13	12	7	2	0		
Polydactylus sextarius – NA	0.6	13	18.42	406	225	213	2	203	28	89	12	9	5	0	1		no
Toxotes jaculatrix – NA	0.6	16	15.56	473	118	106	0	353	11	290	1	1	5	3	0		
Hippoglossus hippoglossus – Pleuronectiformes	0.6	23	7.02	293	240	225	1	104	73	10	3	3	1	0	0		no
Reinhardtius hippoglossoides – Pleuronectiformes	0.6	59	3.11	265	192	179	1	88	59	9	2	2	0	0	0		no
Verasper variegatus – Pleuronectiformes	0.5	15	14.45	162	109	99	1	45	17	8	3	3	1	0	0		no
Scophthalmus maximus – Pleuronectiformes	0.6	11	20.47	163	111	98	1	44	20	14	4	4	1	0	0		no
Psettodes erumei – Pleuronectiformes	0.5	56	2.45	118	84	69	2	24	4	9	4	4	1	0	1		no

CONTINUED ON NEXT PAGE

Table 1 (continued)

CONTINUED FROM PREVIOUS PAGE



	Genome size	Contig L50	Contig N50	B30.2 / PRY / SPRY	TRIM	TRIM-B30.2	TRIM-fn3	NLR	RING-NLR	NLR-B30.2	PYD-NLR	PYD-NLR-B30.2	CARD-NLR	CARD-NLR-B30.2	NLR-FIIND-CARD	marine	freshwater
Chaenogobius annularis – Gobiiformes	0.7	250	0.8	384	206	189	1	88	1	48	2	2	1	0	0	no	
Mugilogobius chulae – Gobiiformes	1	344	0.7	596	303	280	1	237	0	151	8	8	2	0	0		
Neogobius melanostomus – Gobiiformes	1	88	2.82	829	534	490	0	309	6	155	20	9	6	0	1		
Periophthalmus magnuspinnatus – Gobiiformes	0.8	87	2.3	392	183	172	0	125	0	104	1	1	1	0	0	no	
Bostrychus sinensis – Gobiiformes	0.9	2256	0.1	450	234	222	1	129	1	72	12	11	4	0	1		
Sphaeramia orbicularis – Kurtiformes	1.3	148	2.36	569	291	275	2	284	54	186	59	55	4	0	2		no
Thunnus orientalis – Scombriformes	0.8	46	5.46	663	368	337	0	311	34	162	59	22	10	2	2		no
Hippocampus abdominalis – Syngnathiformes	0.4	18	8.41	79	51	41	0	3	0	1	0	0	1	0	0		no
Syngnathus acus – Syngnathiformes	0.3	11	11.96	79	45	37	1	20	0	1	17	1	0	0	0		no
Callionymus lyra – Syngnathiformes	0.6	72	2.2	194	120	117	0	42	0	16	0	0	0	0	0		no
Myripristis murdjan – Holocentriformes	0.8	20	14.48	888	339	317	0	466	125	423	31	26	7	1	4		no
Gadus morhua – Gadiformes	0.7	169	1.02	807	385	365	1	272	4	241	1	1	25	22	1		no
Oncorhynchus mykiss – Salmoniformes	2.3	41	15.58	554	280	246	7	206	9	146	11	10	23	15	0		
Salmo salar – Salmoniformes	2.8	33	28.06	1259	497	433	15	509	39	383	9	9	40	25	0		
Salvelinus namaycush – Salmoniformes	2.3	252	1.8	1018	321	278	6	534	31	444	11	10	33	20	0	no	
Coregonus alpinus – Salmoniformes	2.1	966	0.53	603	268	213	13	165	9	81	10	6	10	0	0	no	
Esox lucius – Esociformes	0.9	14	22.63	445	161	135	6	180	7	139	19	15	16	14	0	no	
Mallotus villosus – Osmeriformes	0.5	521	0.23	289	197	149	21	61	1	26	7	2	7	2	0		
Protosalanx chinensis – Osmeriformes	0.5	876	0.1	142	84	72	1	32	1	12	7	5	3	1	0		
Colossoma macropomum – Characiformes	1.2	54	5.65	706	380	345	0	327	13	147	23	21	1	0	0	no	
Pygocentrus nattereri – Characiformes	1.2	29	12.9	438	256	233	1	176	7	70	18	14	1	0	0	no	
Astyanax mexicanus – Characiformes	1.3	198	1.77	533	212	183	1	315	31	141	8	8	1	0	0	no	
Electrophorus electricus – Gymnotiformes	0.6	25	7.06	254	128	103	0	130	11	52	19	14	2	0	0	no	
Ameiurus melas – Siluriformes	0.9	36	7.41	408	290	264	0	370	0	6	2	1	4	0	0	no	
Ictalurus punctatus – Siluriformes	1	70	1.69	509	385	351	0	447	1	8	1	0	1	0	0	no	
Pangasianodon hypophthalmus – Siluriformes	0.8	18	16.19	262	205	183	0	243	1	8	3	2	4	0	0	no	
Silurus meridionalis – Siluriformes	0.7	20	13.2	266	203	185	0	158	0	7	2	1	4	0	1	no	
Bagarius yarrelli – Siluriformes	0.6	81	1.85	117	86	73	0	30	0	4	1	0	3	0	0	no	
Carassius auratus – Cypriniformes	1.7	869	0.48	814	417	371	0	526	1	184	41	27	5	0	1	no	
Cyprinus carpio – Cypriniformes	1.7	229	1.56	872	403	371	0	410	11	156	33	24	9	0	5	no	
Sinocyclocheilus anophthalmus – Cypriniformes	1.9	2026	0.28	647	378	295	0	256	6	82	10	4	8	0	4	no	
Labeo catla – Cypriniformes	1	353	0.72	1082	531	473	0	633	13	279	58	41	23	0	12	no	
Onychostoma macrolepis – Cypriniformes	0.9	24	10.81	589	342	308	0	306	6	128	14	12	6	1	3	no	
Oxygymnocypris stewartii – Cypriniformes	1.8	1104	0.26	700	482	395	0	269	7	85	6	4	5	0	1	no	
Puntigrus tetrazona – Cypriniformes	0.7	131	1.42	498	174	154	0	280	5	181	42	39	1	0	1	no	
Danio rerio – Cypriniformes	1.4	219	1.42	633	176	155	0	431	0	316	43	36	3	0	1	no	
Danionella dracula – Cypriniformes	0.7	85	2.3	100	62	53	1	66	0	6	10	1	3	0	1	no	
Gobiocypris rarus – Cypriniformes	1	51	5.33	296	194	162	0	124	2	47	8	8	3	0	5	no	
Paracanthobrama guichenoti – Cypriniformes	1.1	152	1.97	440	184	146	0	227	4	101	26	23	5	1	4	no	
Pimephales promelas – Cypriniformes	1.1	903	0.3	316	197	170	0	203	2	50	13	7	8	1	0	no	
Paedocypris progenetica – Cypriniformes	0.4	2365	0.04	174	102	85	1	43	0	4	0	0	1	0	0	no	
Culter alburnus – Cypriniformes	1	2280	0.12	951	479	399	1	454	6	193	55	46	14	2	8	no	
Megalobrama amblycephala – Cypriniformes	1.1	882	0.32	1037	520	433	0	509	7	212	57	45	11	1	7	no	
Triplophysa dalaica – Cypriniformes	0.6	21	9.27	343	201	168	0	188	6	72	18	17	6	0	0	no	
Chanos chanos – Gonorynchiformes	0.7	49	3.18	824	373	321	2	362	30	213	33	29	13	0	1		
Alosa sapidissima – Clupeiformes	0.9	165	1.57	426	273	227	4	145	23	73	15	15	10	1	6		
Tenualosa ilisha – Clupeiformes	0.8	1470	0.13	901	386	272	12	409	34	121	30	26	10	2	4		
Clupea harengus – Clupeiformes	0.7	170	1.15	537	350	292	15	208	19	57	22	14	13	0	12		no
Coilia nasus – Clupeiformes	0.8	135	1.63	264	166	129	4	98	8	27	7	4	6	0	5		
Denticeps clupeoides – Clupeiformes	0.6	53	3.06	200	149	129	6	63	3	16	9	9	8	0	0	no	
Heterotis niloticus – Osteoglossiformes	0.7	45	4.57	262	91	70	1	61	0	32	0	0	0	0	0	no	
Scleropages formosus – Osteoglossiformes	0.8	31	9.1	358	117	106	1	105	1	79	1	1	5	2	0	no	
Anguilla anguilla – Anguilliformes	1	54	5.11	452	259	223	0	165	0	137	15	14	2	0	0		
Megalops cyprinoides – Elopiformes	1	13	26.85	341	154	122	2	111	6	87	24	23	6	0	0		
Lepisosteus oculatus – Semionotiformes	0.9	3307	0.07	183	78	66	0	37	0	2	1	0	7	1	0		
Acipenser ruthenus – Acipenseriformes	1.8	837	0.6	697	386	318	1	37	0	0	6	0	1	0	0	no	
Erpetoichthys calabaricus – Polypteriformes	3.8	868	1.14	586	358	286	0	146	0	2	6	0	9	0	0	no	
Polypterus senegalus – Polypteriformes	3.7	185	4.53	626	344	260	1	143	0	5	3	0	2	0	0	no	
Latimeria chalumnae – Coelacanthiformes	2.9	50769	0.01	214	150	98	0	44	0	0	9	0	2	0	0		no
Xenopus laevis – Anura	2.7	35	22.45	312	222	208	0	23	0	0	0	0	2	0	0	no	
Zootoca vivipara – Squamata	1.5	1762	0.22	155	103	99	0	12	0	0	2	0	0	0	0	no	no
Chelonia mydas – Testudines	2.1	17	39.42	242	134	112	0	19	0	0	10	0	1	0	0		no
Taeniopygia guttata – Passeriformes	1.1	63	4.38	24	18	13	0	6	0	0	0	0	0	0	0	no	no
Gallus gallus – Galliformes	1.1	19	17.66	45	20	19	0	7	0	0	1	0	0	0	0	no	no
Mus musculus – Rodentia	2.7	15	59.46	71	49	42	1	44	0	0	19	0	4	0	2	no	no
Homo sapiens – Primates	3.1	18	57.88	80	52	43	0	27	0	0	15	0	1	0	0	no	no
Scyliorhinus canicula – Carcharhiniformes	4.2	614	1.86	269	188	174	0	56	0	0	0	0	3	0	0		no
Rhincodon typus – Orectolobiformes	2.9	5485	0.14	159	93	67	0	54	0	0	0	0	4	0	0		no
Amblyraja radiata – Rajiformes	2.6	445	1.46	524	381	340	0	213	0	1	0	0	0	0	0		no
Pristis pectinata – Rhinopristiformes	2.3	39	17.01	136	96	86	0	133	0	0	0	0	1	0	0		
Entosphenus tridentatus – Petromyzontiformes	0.9	10026	0.02	160	81	52	0	23	0	0	0	0	0	0	0		
Lethenteron reissneri – Petromyzontiformes	1.1	152	1.41	314	216	203	0	16	0	0	0	0	2	0	0	no	
Petromyzon marinus – Petromyzontiformes	1.1	102	2.54	213	120	114	0	24	0	0	0	0	1	0	0		
Eptatretus burgeri – Myxiniformes	2.6	52630	0.01	94	46	36	0	20	0	0	0	0	2	0	0		no
Branchiostoma japonicum – Amphioxiformes	0.4	11	14.21	20	183	9	6	50	0	0	0	0	2	0	0		no

that while eels are the most distant lineage with a high copy number for this domain structure (137 genes in *Anguilla anguilla*), it is also present in gars (*Semionotiformes*) and reedfish/bichirs (*Polypteriformes*) (Table 1). The presence of NLRs in general has been reported previously for these species (He et al. 2019), but as can be seen here, a small fraction of those appear to contain B30.2 domains which were not detected by the earlier study. However, we did not find any NLR-B30.2 genes from the sturgeon *Acipenser ruthenus* (0 out of 37 genes). In order to confirm this observation, we blasted zebrafish NLR-B30.2 sequences against the available *Acipenseridae* sequences in EST (Expressed Sequence Tag) and TSA (Transcriptome Shotgun Assembly) databases and found all alignments to stop before the B30.2 domain. In *Sarcopterygii*, the sister group of *Actinopterygii* (contains *Latimeria* and all tetrapods—mammals, birds, reptiles, amphibians), there are also no NLRs with an N-terminal B30.2.

Lineage-specific reduction of the TRIM/NLR repertoire in species that lost MHC II

Both NLR-B30.2 and TRIM-B30.2 genes appear to share the property of going through many lineage-specific expansion

and contraction events (Table 1). However, there are a few cases in which both families are affected simultaneously in a clade. The two clades showing the most notable drop in the copy number of both families at the same time are *Syngnathiformes* (seahorses and pipefish) and *Lophiiformes* (anglerfish), groups of fish that have both independently lost Class II MHC from their genome. This outcome is likely the result of the need to reduce tissue compatibility reactions: in *Syngnathiformes*, males get pregnant and give birth (Roth et al. 2020). In anglerfish, males and females fuse their bodies permanently together (Dubin et al. 2019; Swann et al. 2020). However, the Atlantic cod (*Gadus morhua*), another species without Class II MHC (Jin et al. 2020; Star and Jentoft 2012; Star et al. 2011), has high numbers of NLRs and TRIMs, with and without B30.2 (Table 1). In the life history of cod, no particular tissue compatibility issues are known to exist, and it has been proposed that the high numbers of NLR genes (and it looks like TRIMs as well) might be what enables the species to thrive without Class II MHC (Jin et al. 2020). This brings up both more questions and more possibilities for future studies. Do NLRs and TRIMs contribute to tissue compatibility reactions in fish? Is the observed repertoire of immune receptors sufficient to substitute for the lack of a major part of adaptive immunity, or are there mechanisms to further enhance the immune repertoire somatically? What immune mechanisms do the anglerfish and seahorses have that replace the functionality of both Class II MHC and the multigene immune receptor families? These questions will hopefully be addressed in future studies.

In addition, fugu and related species (*Tetraodontiformes*) appear to have a reduction in both gene families as well, as in *Istiophoroformes* (sailfish and swordfish). The short-lived turquoise killifish *Nothobranchius furzeri*, as well as the tiny *Paedocypris progenetica* and *Danionella dracula*, have very low copy numbers (Table 1). For the first two, however, we note that their genomes are not as high quality as many of the others, as witnessed from the assemblies' L50 and N50 values (Table 1).

Rampant duplication in specific clades of fish

On the opposite end of the spectrum, several clades have nearly doubled the total numbers of both TRIM and NLR genes (Table 1). One such group is the cichlids, among whom the gene expansions are especially massive in Old World species from Africa (Nile tilapia, *Astatotilapia*, *Maylandia zebra*). *Astatotilapia calliptera* in particular has 600 TRIMs (568 with B30.2) and 562 NLRs (290 with B30.2). Other clades with clear expansions include the salmonids, perch, and closely related *Perciformes*, as well as several species of herring and shad. Within *Cypriniformes*, there are two subgroups with many paralogs of these immune receptors. The first one includes culter and bream, and the other one includes carp, goldfish, and related species (Table 1). The zebrafish, on the other hand, has less TRIMs than many other cyprinids, although its NLR repertoire is comparable to carp and the others (Table 1).

Protein subtype-specific gains and losses

It appears extremely uncommon for TRIM to have low copy numbers while a large number of NLRs are present. The opposite is not true, however, and there are many cases in which a repertoire of hundreds of TRIM-B30.2 genes is accompanied by only a handful of NLR-B30.2 genes. A notable example of this is the catfish (*Siluriformes*) which have many NLRs without B30.2 and many TRIM genes, but very few NLRs with a B30.2 or PYD (Table 1). In sharks, agnathans (lampreys and hagfish), and tetrapods, all NLR-related gene counts are below average, suggesting that the addition of B30.2 to NLRs in ray-finned fish may be related to the start of these large-scale gene expansions. On the other hand, rays (*Rhinipristiformes*, *Rajiformes*) seem to have their own expansions of NLRs without B30.2 or PYD, and flatfish (*Pleuronectiformes*), the group including flounders and turbot, also have fewer NLR copies than many others (Table 1).

Family-specific gains and losses

It appears surprisingly common for members of the same order to have greatly varying copy numbers, especially when it comes to NLRs with additional N-terminal domains (CARD and PYD). The most extreme example comes from Order *Labriformes*: ballan wrasse (*Labrus bergylta*) appears to have more NLRs in its genome than the other three analyzed members of this order combined (most of those without B30.2), even though their assembly sizes are similar (Table 1). Interestingly, this is due to a very large number of NLR genes without the B30.2 domain. It also has the highest number of TRIM-B30.2 genes of all species in our survey. Hence, repertoires of immune genes in the current genome assembly of this species look quite extraordinary. Other unique features of ballan wrasse's immune system are its extraordinarily high IgM expression in the gut and somatic hypermutation of T cell receptor genes (Bilal et al. 2019, 2018); however, these do not have an obvious direct link to an expansion of both NLRs and TRIMs.

In other cases, species even among the ray-finned fish have lost nearly all detectable NLRs with any of the N- and C-terminal extensions while keeping the TRIM diversity: examples include snailfish (*Pseudoliparis*) and lumpfish (*Cyclopterus lumpus*) (Table 1). In any case, reduced variability within an entire clade (as described above for cichlids,

salmonids, cyprinids, and flatfish) means that the signal for a change in a specific direction must be extremely strong.

Lineage-specific expansions of NLRP1-like proteins

A unique domain structure sometimes found in NLRs includes a C-terminal FIIND and CARD into the protein, similar to human NLRP1 (Fig. 1B). Proteins with this structure are generally key inflammasome components and are strongly associated with inflammasome activity, leading to caspase activation, proinflammatory cytokine production, and programmed cell death by pyroptosis. They are similar to class I TRIMs in that they can be hard to detect and usually have very low copy numbers. There are exceptions, however. Previously, a small expansion has been reported in the round goby (Adrian-Kalchhauser et al. 2020). Croakers and drums (*Larimichthus crocea*, *Collichthys lucidus*, etc.) also have expansions of this gene (up to the 8 copies seen in *Larimichthys crocea*), as do herrings, cyprinids closely related to carp and goldfish, and some others (Table 1). The highest detectable copy numbers (12 genes) were observed for two species of great economic importance, Atlantic herring and *Labeo catla*, a species that is frequently cultivated in commercial fisheries of Asian countries (Table 1).

RING-NLR genes

There is a poorly known subtype of NLRs that has nevertheless been reported at least twice independently over the years. Both turbot (a flatfish) and the Miiuyi croaker (a croaker) possess receptors that have a RING (C3HC4) domain commonly found in TRIMs, attached to the N-terminus of an NLR, in place of a CARD or PYD (Li et al. 2016a; Zhang et al. 2021) (Fig. 1B). These proteins have a FISNA domain and hence could be classified as NLR-C genes; however, we did not find any cases of them being attached to a B30.2 domain. While PYD and CARD are protein interaction domains, RING domains are associated with E3 ubiquitin ligase activity, so functionally these proteins are probably quite different from the other, "true" NLR-C genes. We found this domain structure to be present in most species, but it has been lost from zebrafish and some of the other cyprinids (Table 1), which explains why it has not received much attention thus far. This protein structure seems to have been repeatedly lost during evolution in general and it is also missing from the MHC-less anglerfish and seahorses (Table 1). However, other species have had RING-NLR expansions, and in the abovementioned ballan wrasse, about a half of the massive B30.2-less part of its NLR repertoire (246 out of 502 NLRs without B30.2) are RING-NLR genes (Table 1).

CARD-NLR-B30.2: a novel domain structure not found in zebrafish

There is another domain structure that seems to be quite rare, which is an NLR that contains both an N-terminal CARD, NACHT with FISNA, and a C-terminal B30.2 (Fig. 1B). Proteins with this structure appear to be either uncommon or missing from everywhere other than some very specific clades: *Beloniformes* other than medaka have up to 14 copies, cichlids have up to 30 copies, and salmonids/pike/cod have up to 25 copies (Table 1). To our knowledge, this is the first time that this domain combination has been described. It is remarkable to see it missing in some clades (like gobies, flatfish, and most cyprinids including zebrafish), while being expanded to 30 copies in others. In fact, salmonids have higher copy numbers for CARD-containing NLR-B30.2 genes than for PYD-containing NLR-B30.2 genes (Table 1). In African cichlids and the glassfish *Parambassis ranga*, the copy numbers for both PYD- and CARD-containing NLRs are higher than anywhere else in the dataset (for the glassfish: 64 PYD-NLRs, 58 of those with B30.2/59 CARD-NLRs, 49 of those with B30.2) (Table 1).

Taken together, it appears that the genomes of all fish contain genes resulting from large-scale duplications of TRIM and NLR genes, with or without B30.2. Why are so many genes needed? To address this question, we focus on the NLRs, as transcriptomic studies of immune activation in fish often report the observed effects on NLRs. In fact, in response to treatment, at least some NLRs typically show up as differentially expressed genes. When going through a large number of studies, some patterns become clear. First, the expression of different NLRs is often not affected to the same degree or is even downregulated for some genes and upregulated for others. Second, these patterns have been observed across a number of fish species, and in response to bacterial cell wall components (Alvarez et al. 2017; Biswas et al. 2016; Jin et al. 2020; Kim et al. 2019; Li et al. 2016a, 2018; Lv et al. 2017; Paria et al. 2016; Unajak et al. 2011; Xie and Belosevic 2018), Gram + bacteria (Biswas et al. 2016; Unajak et al. 2011; Wu et al. 2019b, c), Gram – bacteria (Biswas et al. 2016; Chen et al. 2019, 2021; Hou et al. 2017; Li et al. 2016a, 2017, 2019; Ling et al. 2019; Lv et al. 2017; Maekawa et al. 2017; Marancik et al. 2014; Pontigo et al. 2021; Qi et al. 2021; Rajendran et al. 2012; Unajak et al. 2011; Wang et al. 2021; Xie and Belosevic 2018; Xin et al. 2020; Zhang et al. 2021; Zhou et al. 2017), virus-mimicking agents (poly I:C and CpG) (Alvarez et al. 2017; Jin et al. 2020; Li et al. 2018; Paria et al. 2016), RNA viruses (Liu et al. 2020; Wang et al. 2020; Xiao et al. 2021), parasitic worms (Jin et al. 2020; Konczal et al. 2020), and protozoan parasites (Cheng et al. 2021; Jiang et al. 2019; Qiu et al. 2020; Syahputra et al. 2019). Combined with studies reporting tissue-specific expression

(Hou et al. 2017; Rajendran et al. 2012) and the fact that often the number of NLRs showing up in a transcriptome is less than 20, it becomes apparent that the gene copies likely undergo extensive subfunctionalization and have both tissue- and pathogen-specific expression patterns. If at least some of these hundreds of proteins interact with each other as well, then this creates an opportunity to create many diverse immune responses, although not on the same level as the diversity of antibodies and the T cell receptor.

We next decided to explore the results for NLR-B30.2 and TRIM-B30.2 domain combinations by plotting them in various ways. First, the distribution of counts per species indicated that TRIM-B30.2 genes generally have somewhat higher copy numbers than NLR-B30.2 genes (Fig. 4A), making the situation in zebrafish quite atypical for a species with such high copy numbers of TRIMs (431 NLRs, 176 TRIMs). This is largely caused by the fact that ~ 250 of the 431 zebrafish NLRs are part of a massive gene cluster on the long arm of a single chromosome (4q) that is heterochromatic (Howe et al. 2013, 2016). The zebrafish genome without 4q would have ~ 180 NLR-B30.2 genes and 174 TRIMs. In addition, NLR-B30.2 genes appear to have more variability in the proportion that is represented by tight clusters than TRIM do. However, in most cases, just a few chromosomes appear to contain at least 25% of the copies in both families (Fig. 4B). We see that there is a correlation between the total domain counts for these two gene families (Fig. 4C) and between the minimal numbers of chromosomes containing at least 25% of the two gene families in a given species (Fig. 4D). Individuals with higher NLR (but not necessarily TRIM) numbers tend to inhabit freshwater (Fig. 4E, F).

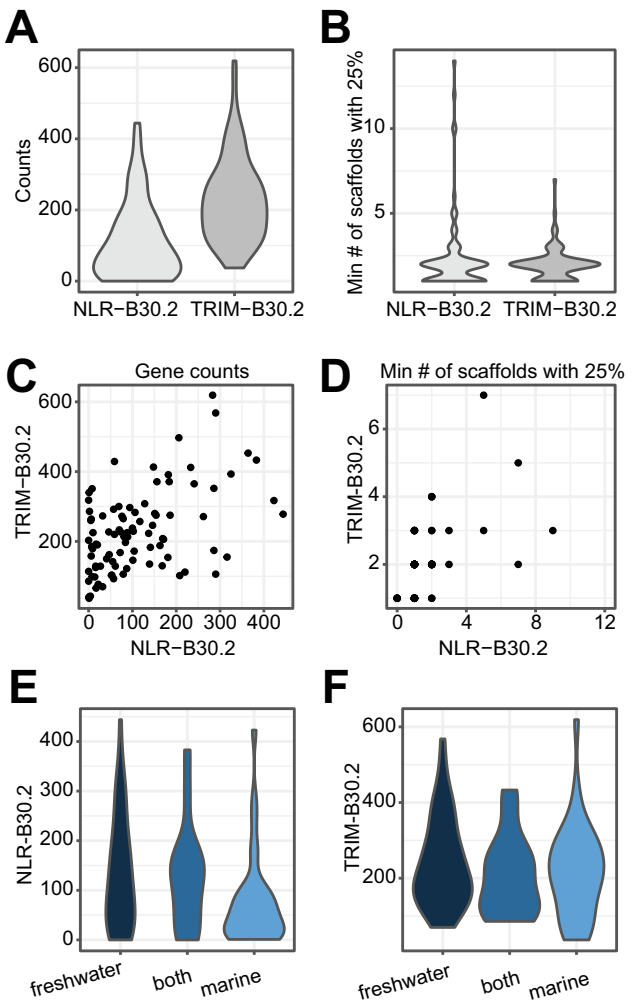


Fig. 4 Statistics for B30.2 containing immune receptors. **A** Violin plots showing the distribution of B30.2-containing gene copy numbers across species. **B** Violin plots showing the distribution of the minimal number of scaffolds that comprise 25% of all B30.2-containing genes of a specific type, again across species. **C** Scatterplot showing the relationship between numbers of the two different immune receptor types. **D** Scatterplot showing the relationship between the numbers of scaffolds that comprise 25% of genes of two different immune receptor types. **E**, **F** Violin plots showing the distribution of gene type copy numbers between species living in different types of water types (marine vs freshwater)

Modeling identifies unexpected correlations between copy numbers of (B30.2-) TRIM and NLR genes, taxonomic position, geographic range, and life in marine vs freshwater environments

Even though we are dealing with immune genes that have previously been shown to be under positive/diversifying selection, many of the observations still appear primarily affected by species relatedness on the phylogenetic scale. To test whether these observations listed above are independent effects or just a side effect of shared ancestry, we modeled the explanatory power of different variables in a phylogenetic context with a maximum likelihood-based linear mixed model (Fig. 5), after excluding tetrapods, amphioxus, and all species that did not have long-read-based full chromosome assemblies.

Selecting the likeliest λ value for each domain type independently led to an overall surprisingly clear pattern describing the effects of different variables. λ values appeared generally higher for TRIMs than for NLRs, reflecting a higher explanatory power of phylogeny for TRIM numbers. Selecting an optimal lambda value also improved the modeling results, as the observed effects appeared much more robust and widespread across multiple domain combinations.

The largest effect that we found was of genome size: we discovered that fish with larger genomes tend to harbor higher numbers of both TRIM and NLR genes (Fig. 5). This was statistically significant for nearly all TRIM-associated domain combinations, but also for PYD-NLRs and NLRs in general. No correlation was observed for rarer domain combinations such as class I TRIMs and the PYD- or CARD-containing NLR subtypes. In our data itself, salmonids,

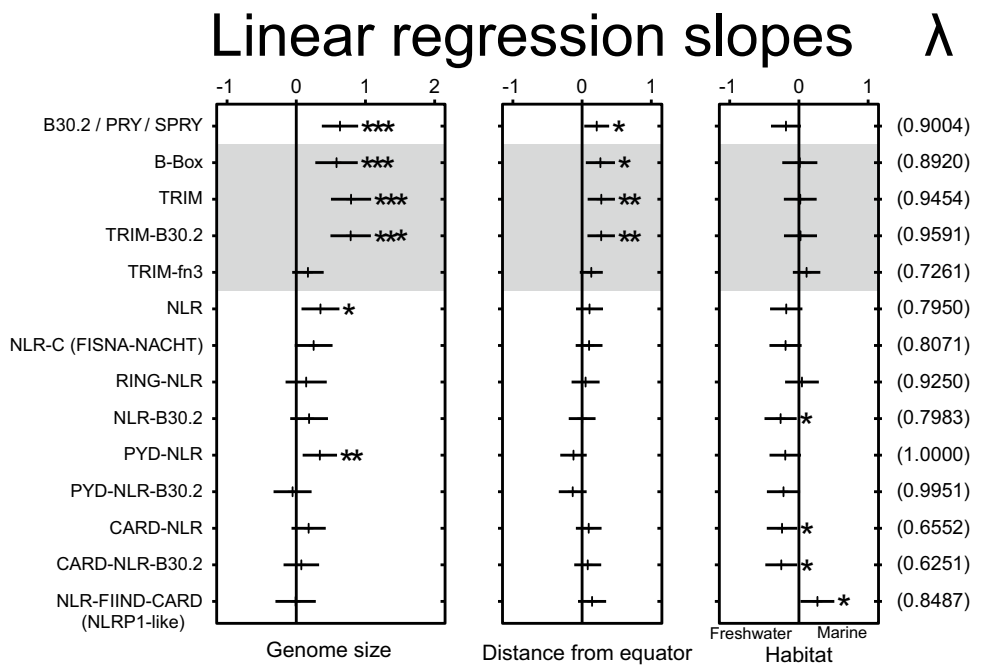


Fig. 5 Modeling results. The three plots present the values of regression slopes, along with 95% confidence intervals. On the right side, the amount of phylogenetic signal in the data (lambda) is shown and ranges from 0 to 1. In our models, the slope parameter is an indicator of the impact that the predictor variables tend to have on copy numbers and can be interpreted as the amount of increase that would be caused by a one-unit-increase in the predictor variable while statistically holding other variables in the model constant. Negative values mean a negative impact, but impact nonetheless. In the "Habitat" field, freshwater has been defined as −1 and marine as +1 so a positive value can be interpreted as marine fish having more copies in general, and negative as freshwater fish having more copies. In all cases, the value of 0 means that the copy numbers are not affected. The slope values can be assigned statistical significance if the 95% confidence interval does not cross zero. For the sake of clarity, asterisks were added to the plots based on *p*-values reported in model summaries: $*p<0.05$. $**p<0.01$. $***p<0.001$

whose ancestor went through a round of whole genome duplication (Berthelot et al. 2014; Macqueen and Johnston 2014), usually have both large genomes and high values for both NLR and TRIM genes (Table 1), which is consistent with the model.

We also observed that fish living further away from the equator, hence likely in colder environments, tend to have more TRIM genes in their genomes and that their copy numbers are significantly correlated with distance (Fig. 5). However, selection pressures leading to such correlations are difficult to infer based on the data that we have. If one were to speculate, it is clear that fish adaptive immunity—especially the T cell response—is slow and poorly efficient in cold conditions in several fish species (Avtalion 1969; Bly and Clem 1991); hence, it might be that cold water species may be subjected to selection pressures for strong innate antiviral immunity and large TRIM repertoires. This is supported by the observation that some of the highest TRIM copy numbers are found in *Pseudoliparis*, a snailfish caught from 7 km deep into the cold ocean depths (Mu et al. 2021), and in various notothenioid species from the Southern Ocean near the Antarctic. Furthermore, this observation remained significant even after correcting for phylogeny in the models.

Finally, the choice between a marine and freshwater habitat appears to have an impact on NLR copy numbers even in the model with phylogenetic correction: there is a clear trend for freshwater fish to have larger NLR complements, regardless of which domain structure we look for (Fig. 5). For NLRs attached to a CARD and/or B30.2 domain, the effect is also of statistical significance. The most obvious difference between freshwater and the sea is salinity, although its link to NLR duplication is unclear at this time. However, a recent study demonstrated that in at least one species, the silver pomfret (*Pampus argenteus*), NLR signaling pathway components are significantly enriched in the set of transcripts that are differentially affected by varying salt concentrations, suggesting that there is at least some link between fish NLRs and salinity (Li et al. 2020a).

Conclusion

Our data reveals a wide variation in TRIM and NLR gene numbers across ray-finned fish. The propensity for fish TRIM and NLR genes encoding a B30.2 domain (class IV TRIMs, NLR-C genes) to expand was suspected based on the well-characterized

repertoires of a few species, including zebrafish. Our approach shows that very large repertoires of TRIM-B30.2 and NLR-B30.2 are present in particular species (including members of *Salmoniformes*, *Cypriniformes*, *Gobiiformes*, *Cichliformes*, and *Perciformes*), but also that other species have only fairly modest sets of these genes; strikingly, the latter count several members of *Syngnathiformes* and *Lophiiformes*, known for their impoverished repertoires of immune genes and their particular adaptations to non-self tolerance. We also explored domain combinations for which little is known about, such as RING-NLR and CARD-NLR-B30.2, and we found classical class IV TRIM genes in unsuspected taxa (like the chordate *Amphioxus*). These findings illustrate the structural and evolutionary complexity of these groups of receptors and pave the way for future functional studies. TRIM-B30.2 and NLR-B30.2 are in fact more ancient than ray-finned fish and appear more than ever as fundamental components of the vertebrate/chordate defense system. Finally, our data allowed a tentative modeling to connect TRIM-B30.2 and NLR-B30.2 gene numbers with natural history traits of the species, as previously reported for the genetic diversity of the major histocompatibility complex (Yiming et al. 2021). Such approaches will certainly benefit from more complete phenotypic and geographic information in the future, but the links we found raise many questions about the impacts of life history traits on the genomic variation of immune gene repertoires.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1007/s00251-021-01235-4.

Acknowledgements We would like to thank the Vertebrate Genomes Project (https://vertebrategenomesproject.org/) for the early use of a number of high quality genome assemblies, and the authors of all original studies whose data contributed to this work.

Funding J.S. and C.J.G. were supported by the Great Lakes Fishery Commission and by a Natural Sciences and Engineering Research Council of Canada Discovery Grant to C.J.G. P.B. was supported by institutional grants from INRAE.

Availability of data and material A list of genome assemblies used in the study, as well as links to FishBase and NCBI, collected coordinates, and the calculated copy numbers and clustering are all available in Supplementary Table 1. Supplementary Data 1 contains plots of the likelihood distributions of the lambda parameter, which were of vital importance in determining the correct input values to use for the modeling.

Code availability The custom scripts used for the analyses are available from Github at https://github.com/jsuurvali/B302receptors

Declarations

Conflict of interest The authors declare no competing interests.

References

Aa LMVD, Jouneau L, Laplantine E, Bouchez O, Kemenade LV, Boudinot P, van der Aa LM, Van Kemenade L (2012) FinTRIMs, fish virus-inducible proteins with E3 ubiquitin ligase activity. Dev Comp Immunol 36:433–441

Adrian-Kalchhauser I, Blomberg A, Larsson T, Musilova Z, Peart CR, Pippel M, Solbakken MH, Suurväli J, Walser JC, Wilson JY, Alm Rosenblad M, Burguera D, Gutnik S, Michiels N, Topel M, Pankov K, Schloissnig S, Winkler S (2020) The round goby genome provides insights into mechanisms that may facilitate biological invasions. BMC Biol 18:11

Afrache H, Gouret P, Ainouche S, Pontarotti P, Olive D (2012) The butyrophilin (BTN) gene family: from milk fat to the regulation of the immune response. Immunogenetics 64:781–794

Alvarez CA, Ramirez-Cepeda F, Santana P, Torres E, Cortes J, Guzman F, Schmitt P, Mercado L (2017) Insights into the diversity of NOD-like receptors: Identification and expression analysis of NLRC3, NLRC5 and NLRX1 in rainbow trout. Mol Immunol 87:102–113

Amparyup P, Charoensapsri W, Samaluka N, Chumtong P, Yocawibun P, Imjongjirak C (2020) Transcriptome analysis identifies immune-related genes and antimicrobial peptides in Siamese fighting fish (Betta splendens). Fish Shellfish Immunol 99:403–413

Ao J, Mu Y, Xiang LX, Fan D, Feng M, Zhang S, Shi Q, Zhu LY, Li T, Ding Y, Nie L, Li Q, Dong WR, Jiang L, Sun B, Zhang X, Li M, Zhang HQ, Xie S, Zhu Y, Jiang X, Wang X, Mu P, Chen W, Yue Z, Wang Z, Wang J, Shao JZ, Chen X (2015) Genome sequencing of the perciform fish Larimichthys crocea provides insights into molecular and genetic mechanisms of stress adaptation. PLoS Genet 11:e1005118

Avtalion RR (1969) Temperature effect on antibody production and immunological memory, in carp (Cyprinus carpio) immunized against bovine serum albumin (BSA). Immunology 17:927–931

Berthelot C, Brunet F, Chalopin D, Juanchich A, Bernard M, Noel B, Bento P, Da Silva C, Labadie K, Alberti A, Aury JM, Louis A, Dehais P, Bardou P, Montfort J, Klopp C, Cabau C, Gaspin C, Thorgaard GH, Boussaha M, Quillet E, Guyomard R, Galiana D, Bobe J, Volff JN, Genet C, Wincker P, Jaillon O, Roest Crollius H, Guiguen Y (2014) The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. Nat Commun 5:3657

Bilal S, Lie KK, Dalum AS, Karlsen OA, Hordvik I (2019) Analysis of immunoglobulin and T cell receptor gene expression in ballan wrasse (Labrus bergylta) revealed an extraordinarily high IgM expression in the gut. Fish Shellfish Immunol 87:650–658

Bilal S, Lie KK, Saele O, Hordvik I (2018) T cell receptor alpha chain genes in the teleost ballan wrasse (Labrus bergylta) are subjected to somatic hypermutation. Front Immunol 9:1101

Biris N, Yang Y, Taylor AB, Tomashevski A, Guo M, Hart PJ, Diaz-Griffero F, Ivanov DN (2012) Structure of the rhesus monkey TRIM5alpha PRYSPRY domain, the HIV capsid recognition module. Proc Natl Acad Sci U S A 109:13278–13283

Biswas G, Bilen S, Kono T, Sakai M, Hikima J (2016) Inflammatory immune response by lipopolysaccharide-responsive nucleotide binding oligomerization domain (NOD)-like receptors in the Japanese pufferfish (Takifugu rubripes). Dev Comp Immunol 55:21–31

Bly JE, Clem LW (1991) Temperature-mediated processes in teleost immunity: in vitro immunosuppression induced by in vivo low temperature in channel catfish. Vet Immunol Immunopathol 28:365–377

Boudinot P, van der Aa LM, Jouneau L, Du Pasquier L, Pontarotti P, Briolat V, Benmansour A, Levraud JP (2011) Origin and evolu-

tion of TRIM proteins: new insights from the complete TRIM repertoire of zebrafish and pufferfish. PLoS One 6:e22022
Boudinot P, Zou J, Ota T, Buonocore F, Scapigliati G, Canapa A, Cannon J, Litman G, Hansen JD (2014) A tetrapod-like repertoire of innate immune receptors and effectors for coelacanths. J Exp Zool B Mol Dev Evol 322:415–437
Buckley KM, Rast JP (2011) Characterizing immune receptors from new genome sequences. Methods Mol Biol 748:273–298
Buckley KM, Rast JP (2015) Diversity of animal immune receptors and the origins of recognition complexity in the deuterostomes. Dev Comp Immunol 49:179–189
Chae JJ, Centola M, Aksentijevich I, Dutra A, Tran M, Wood G, Nagaraju K, Kingma DW, Liu PP, Kastner DL (2000) Isolation, genomic organization, and expression analysis of the mouse and rat homologs of MEFV, the gene for familial Mediterranean fever. Mamm Genome 11:428–435
Chamberlain SA, Szocs E (2013) taxize: taxonomic search and retrieval in R. F1000Res 2:191
Chen H, Ding S, Tan J, Yang D, Zhang Y, Liu Q (2020) Characterization of the Japanese flounder NLRP3 inflammasome in restricting Edwardsiella piscicida colonization in vivo. Fish Shellfish Immunol 103:169–180
Chen H, Wang B, Yu N, Qi J, Tang N, Wang S, Tian Z, Wang M, Xu S, Zhou B, Long Q, Chen D, Li Z (2019) Transcriptome analysis and the effects of polyunsaturated fatty acids on the immune responses of the critically endangered angtze sturgeon (Acipenser dabryanus). Fish Shellfish Immunol 94:199–210
Chen Z, Xu X, Wang J, Zhou Q, Chen S (2021) A genome-wide survey of NOD-like receptors in Chinese tongue sole (Cynoglossus semilaevis): identification, characterization, and expression analysis in response to bacterial infection. J Fish Biol
Cheng JX, Xia YQ, Liu YF, Liu PF, Liu Y (2021) Transcriptome analysis in Takifugu rubripes and Dicentrarchus labrax gills during Cryptocaryon irritans infection. J Fish Dis 44:249–262
D'Cruz AA, Kershaw NJ, Chiang JJ, Wang MK, Nicola NA, Babon JJ, Gack MU, Nicholson SE (2013) Crystal structure of the TRIM25 B30.2 (PRYSPRY) domain: a key component of antiviral signalling. Biochem J 456:231–240
Dowle M, Srinivasan A (2019) data.table: extension of 'data.frame'. R package version 1.12.8 edn
Dubin A, Jorgensen TE, Moum T, Johansen SD, Jakt LM (2019) Complete loss of the MHC II pathway in an anglerfish. Lophius Piscatorius Biol Lett 15:20190594
Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, Holm L, Sonnhammer EL, Eddy SR, Bateman A (2010) The Pfam protein families database. Nucleic Acids Res 38:D211–D222
Froese R, Pauly D (2021) FishBase. World Wide Web electronic publication
He Y, Pan H, Zhang G, He S (2019) Comparative study on pattern recognition receptors in non-teleost ray-finned fishes and their evolutionary significance in primitive vertebrates. Sci China Life Sci 62:566–578
Henry J, Ribouchon MT, Offer C, Pontarotti P (1997) B30.2-like domain proteins: a growing family. Biochem Biophys Res Commun 235:162–165
Hou Z, Ye Z, Zhang D, Gao C, Su B, Song L, Tan F, Song H, Wang Y, Li C (2017) Characterization and expression profiling of NOD-like receptor C3 (NLRC3) in mucosal tissues of turbot (Scophthalmus maximus L.) following bacterial challenge. Fish Shellfish Immunol 66:231–239
Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M, Collins JE, Humphray S, McLaren K, Matthews L, McLaren S, Sealy I, Caccamo M, Churcher C, Scott C, Barrett JC, Koch R, Rauch GJ, White S, Chow W, Kilian B, Quintais LT, Guerra-Assuncao JA, Zhou Y, Gu Y, Yen J, Vogel JH, Eyre T, Redmond S, Banerjee R, Chi J, Fu B, Langley E, Maguire SF, Laird GK, Lloyd D, Kenyon E, Donaldson S, Sehra H, Almeida-King J, Loveland J, Trevanion S, Jones M, Quail M, Willey D, Hunt A, Burton J, Sims S, McLay K, Plumb B, Davis J, Clee C, Oliver K, Clark R, Riddle C, Elliot D, Threadgold G, Harden G, Ware D, Begum S, Mortimore B, Kerry G, Heath P, Phillimore B, Tracey A, Corby N, Dunn M, Johnson C, Wood J, Clark S, Pelan S, Griffiths G, Smith M, Glithero R, Howden P, Barker N, Lloyd C, Stevens C, Harley J, Holt K, Panagiotidis G, Lovell J, Beasley H, Henderson C, Gordon D, Auger K, Wright D, Collins J, Raisen C, Dyer L, Leung K, Robertson L, Ambridge K, Leongamornlert D, McGuire S, Gilderthorp R, Griffiths C, Manthravadi D, Nichol S, Barker G et al (2013) The zebrafish reference genome sequence and its relationship to the human genome. Nature 496:498–503
Howe K, Schiffer PH, Zielinski J, Wiehe T, Laird GK, Marioni JC, Soylemez O, Kondrashov F, Leptin M (2016) Structure and evolutionary history of a large family of NLR proteins in the zebrafish. Open Biol 6:160009
Huang S, Yuan S, Guo L, Yu Y, Li J, Wu T, Liu T, Yang M, Wu K, Liu H, Ge J, Yu Y, Huang H, Dong M, Yu C, Chen S, Xu A (2008) Genomic analysis of the immune gene repertoire of amphioxus reveals extraordinary innate complexity and diversity. Genome Res 18:1112–1126
IUCN (2021) The IUCN Red List of Threatened Species. Version 2021–1. https://www.iucnredlist.org
Jiang B, Du JJ, Li YW, Ma P, Hu YZ, Li AX (2019) Transcriptome analysis provides insights into molecular immune mechanisms of rabbitfish, Siganus oramin against Cryptocaryon irritans infection. Fish Shellfish Immunol 88:111–116
Jin X, Morro B, Torresen OK, Moiche V, Solbakken MH, Jakobsen KS, Jentoft S, MacKenzie S (2020) Innovation in nucleotide-binding oligomerization-like receptor and toll-like receptor sensing drives the major histocompatibility complex-II free Atlantic cod immune system. Front Immunol 11:609456
Jones JD, Vance RE, Dangl JL (2016) Intracellular innate immune surveillance devices in plants and animals. Science 354
Kasahara M, Sutoh Y (2014) Two forms of adaptive immunity in vertebrates: similarities and differences. Adv Immunol 122:59–90
Kelley J, Walter L, Trowsdale J (2005) Comparative genomics of major histocompatibility complexes. Immunogenetics 56:683–695
Kim JH, Macqueen DJ, Winton JR, Hansen JD, Park H, Devlin RH (2019) Effect of growth rate on transcriptomic responses to immune stimulation in wild-type, domesticated, and GH-transgenic coho salmon. BMC Genomics 20:1024
Kim YK, Shin JS, Nahm MH (2016) NOD-like receptors in infection, immunity, and diseases. Yonsei Med J 57:5–14
Konczal M, Ellison AR, Phillips KP, Radwan J, Mohammed RS, Cable J, Chadzinska M (2020) RNA-Seq analysis of the guppy immune response against Gyrodactylus bullatarudis infection. Parasite Immunol 42:e12782
Kuri P, Schieber NL, Thumberger T, Wittbrodt J, Schwab Y, Leptin M (2017) Dynamics of in vivo ASC speck formation. J Cell Biol 216:2891–2909
Laing KJ, Purcell MK, Winton JR, Hansen JD (2008) A genomic view of the NOD-like receptor family in teleost fish: identification of a novel NLR subfamily in zebrafish. BMC Evol Biol 8:42
Langevin C, Aleksejeva E, Houel A, Briolat V, Torhy C, Lunazzi A, Levraud JP, Boudinot P (2017) FTR83, a member of the large fish-specific finTRIM family, triggers IFN pathway and counters viral infection. Front Immunol 8:617
Li J, Chu Q, Xu T (2016a) A genome-wide survey of expansive NLR-C subfamily in miiuy croaker and characterization of the NLR-B30.2 genes. Dev Comp Immunol 61:116–125
Li S, Chen X, Hao G, Geng X, Zhan W, Sun J (2016b) Identification and characterization of a novel NOD-like receptor family CARD domain containing 3 gene in response to extracellular ATP stimulation and its role in regulating LPS-induced innate immune response in Japanese

flounder (Paralichthys olivaceus) head kidney macrophages. Fish Shellfish Immunol 50:79–90
Li J, Xue L, Cao M, Zhang Y, Wang Y, Xu S, Zheng B, Lou Z (2020a) Gill transcriptomes reveal expression changes of genes related with immune and ion transport under salinity stress in silvery pomfret (Pampus argenteus). Fish Physiol Biochem 46:1255–1277
Li JY, Wang YY, Shao T, Fan DD, Lin AF, Xiang LX, Shao JZ (2020b) The zebrafish NLRP3 inflammasome has functional roles in ASC-dependent interleukin-1beta maturation and gasdermin E-mediated pyroptosis. J Biol Chem 295:1120–1141
Li S, Zhang Y, Cao Y, Wang D, Liu H, Lu T (2017) Trancriptome profiles of Amur sturgeon spleen in response to Yersinia ruckeri infection. Fish Shellfish Immunol 70:451–460
Li T, Shan S, Wang L, Yang G, Zhu J (2018) Identification of a fish-specific NOD-like receptor subfamily C (NLRC) gene from common carp (Cyprinus carpio L.): characterization, ontogeny and expression analysis in response to immune stimulation. Fish Shellfish Immunol 82:371–377
Li Z, Wang X, Chen C, Gao J, Lv A (2019) Transcriptome profiles in the spleen of African catfish (Clarias gariepinus) challenged with Aeromonas veronii. Fish Shellfish Immunol 86:858–867
Ling XD, Dong WT, Zhang Y, Qian X, Zhang WD, He WH, Zhao XX, Liu JX (2019) Comparative transcriptomics and histopathological analysis of crucian carp infection by atypical Aeromonas salmonicida. Fish Shellfish Immunol 94:294–307
Liu J, Yan Y, Yan J, Wang J, Wei J, Xiao J, Zeng Y, Feng H (2020) Multi-omics analysis revealed crucial genes and pathways associated with black carp antiviral innate immunity. Fish Shellfish Immunol 106:724–732
Lv Z, Wei Z, Zhang Z, Li C, Shao Y, Zhang W, Zhao X, Li Y, Duan X, Xiong J (2017) Characterization of NLRP3-like gene from Apostichopus japonicus provides new evidence on inflammation response in invertebrates. Fish Shellfish Immunol 68:114–123
Macqueen DJ, Johnston IA (2014) A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. Proc Biol Sci 281:20132881
Maekawa S, Byadgi O, Chen YC, Aoki T, Takeyama H, Yoshida T, Hikima JI, Sakai M, Wang PC, Chen SC (2017) Transcriptome analysis of immune response against Vibrio harveyi infection in orange-spotted grouper (Epinephelus coioides). Fish Shellfish Immunol 70:628–637
Marancik D, Gao G, Paneru B, Ma H, Hernandez AG, Salem M, Yao J, Palti Y, Wiens GD (2014) Whole-body transcriptome of selectively bred, resistant-, control-, and susceptible-line rainbow trout following experimental challenge with Flavobacterium psychrophilum. Front Genet 5:453
Morimoto N, Kono T, Sakai M, Hikima JI (2021) Inflammasomes in teleosts: structures and mechanisms that induce pyroptosis during bacterial infection. Int J Mol Sci 22
Mu Y, Bian C, Liu R, Wang Y, Shao G, Li J, Qiu Y, He T, Li W, Ao J, Shi Q, Chen X (2021) Whole genome sequencing of a snailfish from the Yap Trench (~7,000 m) clarifies the molecular mechanisms underlying adaptation to the deep sea. PLoS Genet 17:e1009530
Munoz Sosa CJ, Issoglio FM, Carrizo ME (2021) Crystal structure and mutational analysis of the human TRIM7 B30.2 domain provide insights into the molecular basis of its binding to glycogenin-1. J Biol Chem 296:100772
Nakatani Y, Shingate P, Ravi V, Pillai NE, Prasad A, McLysaght A, Venkatesh B (2021) Reconstruction of proto-vertebrate, proto-cyclostome and proto-gnathostome genomes provides new insights into early vertebrate evolution. Nat Commun 12:4489
Newman RM, Hall L, Connole M, Chen GL, Sato S, Yuste E, Diehl W, Hunter E, Kaur A, Miller GM, Johnson WE (2006) Balancing selection and the evolution of functional polymorphism in Old World monkey TRIM5alpha. Proc Natl Acad Sci U S A 103:19134–19139
Nisole S, Stoye JP, Saib A (2005) TRIM family proteins: retroviral restriction and antiviral defence. Nat Rev Microbiol 3:799–808
Ozato K, Shin DM, Chang TH, Morse HC 3rd (2008) TRIM family proteins and their emerging roles in innate immunity. Nat Rev Immunol 8:849–860
Paria A, Deepika A, Sreedharan K, Makesh M, Chaudhari A, Purushothaman CS, Thirunavukkarasu AR, Rajendran KV (2016) Identification of Nod like receptor C3 (NLRC3) in Asian seabass, Lates calcarifer: characterisation, ontogeny and expression analysis after experimental infection and ligand stimulation. Fish Shellfish Immunol 55:602–612
Pebesma E (2018) Simple features for R: standardized support for spatial vector data. The R Journal 10:439–446
Pinheiro J, Bates D, DebRoy S, Sarkar D, Team RC (2021) nlme: linear and nonlinear mixed effects models. R package version 3.1–152. https://CRAN.R-project.org/package=nlme
Pontigo JP, Yanez A, Sanchez P, Vargas-Chacoff L (2021) Characterization and expression analysis of Nod-like receptor 3 (NLRC3) against infection with Piscirickettsia salmonis in Atlantic salmon. Dev Comp Immunol 114:103865
Proell M, Riedl SJ, Fritz JH, Rojas AM, Schwarzenbacher R (2008) The Nod-like receptor (NLR) family: a tale of similarities and differences. PLoS One 3:e2119
Qi L, Chen Y, Shi K, Ma H, Wei S, Sha Z (2021) Combining of transcriptomic and proteomic data to mine immune-related genes and proteins in the liver of Cynoglossus semilaevis challenged with Vibrio anguillarum. Comp Biochem Physiol Part D Genomics Proteomics 39:100864
Qiu Y, Yin Y, Ruan Z, Gao Y, Bian C, Chen J, Wang X, Pan X, Yang J, Shi Q, Jiang W (2020) Comprehensive transcriptional changes in the liver of Kanglang white minnow (Anabarilius grahami) in response to the infection of parasite Ichthyophthirius multifiliis. Animals (Basel) 10
Rajendran KV, Zhang J, Liu S, Kucuktas H, Wang X, Liu H, Sha Z, Terhune J, Peatman E, Liu Z (2012) Pathogen recognition receptors in channel catfish: I. Identification, phylogeny and expression of NOD-like receptors. Dev Comp Immunol 37:77–86
Rhie A, McCarthy SA, Fedrigo O, Damas J, Formenti G, Koren S, Uliano-Silva M, Chow W, Fungtammasan A, Kim J, Lee C, Ko BJ, Chaisson M, Gedman GL, Cantin LJ, Thibaud-Nissen F, Haggerty L, Bista I, Smith M, Haase B, Mountcastle J, Winkler S, Paez S, Howard J, Vernes SC, Lama TM, Grutzner F, Warren WC, Balakrishnan CN, Burt D, George JM, Biegler MT, Iorns D, Digby A, Eason D, Robertson B, Edwards T, Wilkinson M, Turner G, Meyer A, Kautt AF, Franchini P, Detrich HW 3rd, Svardal H, Wagner M, Naylor GJP, Pippel M, Malinsky M, Mooney M, Simbirsky M, Hannigan BT, Pesout T, Houck M, Misuraca A, Kingan SB, Hall R, Kronenberg Z, Sovic I, Dunn C, Ning Z, Hastie A, Lee J, Selvaraj S, Green RE, Putnam NH, Gut I, Ghurye J, Garrison E, Sims Y, Collins J, Pelan S, Torrance J, Tracey A, Wood J, Dagnew RE, Guan D, London SE, Clayton DF, Mello CV, Friedrich SR, Lovell PV, Osipova E, Al-Ajli FO, Secomandi S, Kim H, Theofanopoulou C, Hiller M, Zhou Y, Harris RS, Makova KD, Medvedev P, Hoffman J, Masterson P, Clark K, Martin F, Howe K, Flicek P, Walenz BP, Kwak W, Clawson H et al (2021) Towards complete and error-free genome assemblies of all vertebrate species. Nature 592:737–746
Rice P, Longden I, Bleasby A (2000) EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet 16:276–277
Roth O, Solbakken MH, Torresen OK, Bayer T, Matschiner M, Baalsrud HT, Hoff SNK, Brieuc MSO, Haase D, Hanel R, Reusch TBH, Jentoft S (2020) Evolution of male pregnancy associated with remodeling of canonical vertebrate immunity in seahorses and pipefishes. Proc Natl Acad Sci U S A 117:9431–9439
Salim M, Knowles TJ, Baker AT, Davey MS, Jeeves M, Sridhar P, Wilkie J, Willcox CR, Kadri H, Taher TE, Vantourout P, Hayday

A, Mehellou Y, Mohammed F, Willcox BE (2017) BTN3A1 Discriminates gammadelta T cell phosphoantigens from nonantigenic small molecules via a conformational sensor in its B30.2 domain. ACS Chem Biol 12:2631–2643

Sardiello M, Cairo S, Fontanella B, Ballabio A, Meroni G (2008) Genomic analysis of the TRIM family reveals two groups of genes with distinct evolutionary properties. BMC Evol Biol 22:1–22

Sawyer SL, Wu LI, Emerman M, Malik HS (2005) Positive selection of primate TRIM5alpha identifies a critical species-specific retroviral restriction domain. Proc Natl Acad Sci U S A 102:2832–2837

Schiffer PH, Gravemeyer J, Rauscher M, Wiehe T (2016) Ultra large gene families: a matter of adaptation or genomic parasites? Life (Basel) 6

Short KM, Cox TC (2006) Subclassification of the RBCC/TRIM superfamily reveals a novel motif necessary for microtubule binding. J Biol Chem 281:8970–8980

South A (2017) rnaturalearth: World Map Data from Natural Earth

Star B, Jentoft S (2012) Why does the immune system of Atlantic cod lack MHC II? BioEssays: news and reviews in molecular, cellular and developmental biology 34:648–51

Star B, Nederbragt AJ, Jentoft S, Grimholt U, Malmstrøm M, Gregers TF, Rounge TB, Paulsen J, Solbakken MH, Sharma A, Wetten OF, Lanzén A, Winer R, Knight J, Vogel J-H, Aken B, Andersen O, Lagesen K, Tooming-Klunderud A, Edvardsen RB, Tina KG, Espelund M, Nepal C, Previti C, Karlsen BO, Moum T, Skage M, Berg PR, Gjøen T, Kuhl H, Thorsen J, Malde K, Reinhardt R, Du L, Johansen SD, Searle S, Lien S, Nilsen F, Jonassen I, Omholt SW, Stenseth NC, Jakobsen KS (2011) The genome sequence of Atlantic cod reveals a unique immune system. Nature 477:207–210

Stein C, Caccamo M, Laird G, Leptin M (2007) Conservation and divergence of gene families encoding components of innate immune response systems in zebrafish. Genome Biol 8:R251

Suurvali J, Jouneau L, Thepot D, Grusea S, Pontarotti P, Du Pasquier L, Ruutel Boudinot S, Boudinot P (2014) The proto-MHC of placozoans, a region specialized in cellular stress and ubiquitination/proteasome pathways. J Immunol 193:2891–2901

Swann JB, Holland SJ, Petersen M, Pietsch TW, Boehm T (2020) The immunogenetics of sexual parasitism. Science

Syahputra K, Kania PW, Al-Jubury A, Jafaar RM, Dirks RP, Buchmann K (2019) Transcriptomic analysis of immunity in rainbow trout (Oncorhynchus mykiss) gills infected by Ichthyophthirius multifiliis. Fish Shellfish Immunol 86:486–496

Torresen OK, Brieuc MSO, Solbakken MH, Sorhus E, Nederbragt AJ, Jakobsen KS, Meier S, Edvardsen RB, Jentoft S (2018) Genomic architecture of haddock (Melanogrammus aeglefinus) shows expansions of innate immune genes and short tandem repeats. BMC Genomics 19:240

Uchil PD, Hinz A, Siegel S, Coenen-Stass A, Pertel T, Luban J, Mothes W (2013) TRIM protein-mediated regulation of inflammatory and innate immune signaling and its association with antiretroviral activity. J Virol 87:257–272

Unajak S, Santos MD, Hikima J, Jung TS, Kondo H, Hirono I, Aoki T (2011) Molecular characterization, expression and functional analysis of a nuclear oligomerization domain proteins subfamily C (NLRC) in Japanese flounder (Paralichthys olivaceus). Fish Shellfish Immunol 31:202–211

Van de Weyer AL, Monteiro F, Furzer OJ, Nishimura MT, Cevik V, Witek K, Jones JDG, Dangl JL, Weigel D, Bemm F (2019) A species-wide inventory of NLR genes and alleles in Arabidopsis thaliana. Cell 178:1260–1272 e14

van der Aa LM, Levraud J-p, Yahmi M, Lauret E, Briolat V, Herbomel P, Benmansour A, Boudinot P (2009) A large new subset of TRIM genes highly diversified by duplication and positive selection in teleost fish. BMC Biol 23:7

Wang D, Sun S, Li S, Lu T, Shi D (2021) Transcriptome profiling of immune response to Yersinia ruckeri in spleen of rainbow trout (Oncorhynchus mykiss). BMC Genomics 22:292

Wang H, Tang X, Sheng X, Xing J, Chi H, Zhan W (2020) Transcriptome analysis reveals temperature-dependent early immune response in flounder (Paralichthys olivaceus) after Hirame novirhabdovirus (HIRRV) infection. Fish Shellfish Immunol 107:367–378

Wang Y, Kuang M, Lu Y, Lin L, Liu X (2017) Characterization and biological function analysis of the TRIM47 gene from common carp (Cyprinus carpio). Gene 627:188–193

Wickham H (2016) ggplot2: elegant graphics for data analysis. Springer-Verlag, New York

Wu M, Zhao X, Gong XY, Wang Y, Gui JF, Zhang YB (2019a) FTRCA1, a species-specific member of finTRIM family, negatively regulates fish IFN response through autophage-lysosomal degradation of TBK1. J Immunol 202:2407–2420

Wu W, Li L, Liu Y, Huang T, Liang W, Chen M (2019b) Multiomics analyses reveal that NOD-like signaling pathway plays an important role against Streptococcus agalactiae in the spleen of tilapia. Fish Shellfish Immunol 95:336–348

Wu XM, Cao L, Hu YW, Chang MX (2019c) Transcriptomic characterization of adult zebrafish infected with Streptococcus agalactiae. Fish Shellfish Immunol 94:355–372

Xiao F, Liao L, Xu Q, He Z, Xiao T, Wang J, Huang J, Yu Y, Wu B, Yan Q (2021) Host-microbiota interactions and responses to grass carp reovirus infection in Ctenopharyngodon idellus. Environ Microbiol 23:431–447

Xie J, Belosevic M (2018) Characterization and functional assessment of the NLRC3-like molecule of the goldfish (Carassius auratus L.). Dev Comp Immunol 79:1–10

Xin GY, Li WG, Suman TY, Jia PP, Ma YB, Pei DS (2020) Gut bacteria Vibrio sp. and Aeromonas sp. trigger the expression levels of pro-inflammatory cytokine: first evidence from the germ-free zebrafish. Fish Shellfish Immunol 106:518–525

Yang D, Zheng X, Chen S, Wang Z, Xu W, Tan J, Hu T, Hou M, Wang W, Gu Z, Wang Q, Zhang R, Zhang Y, Liu Q (2018) Sensing of cytosolic LPS through caspy2 pyrin domain mediates noncanonical inflammasome activation in zebrafish. Nat Commun 9:3052

Yiming L, Siqi W, Chaoyuan C, Jiaqi Z, Supen W, Xianglei H, Xuan L, Xuejiao Y, Xianping L (2021) Latitudinal gradients in genetic diversity and natural selection at a highly adaptive gene in terrestrial mammals. Ecography 44:206–218

Yuen B, Bayes JM, Degnan SM (2014) The characterization of sponge NLRs provides insight into the origin and evolution of this innate immune gene family in animals. Mol Biol Evol 31:106–120

Zhang L, Cao M, Li Q, Yan X, Xue T, Song L, Su B, Li C (2021) Genome-wide identification of NOD-like receptors and their expression profiling in mucosal tissues of turbot (Scophthalmus maximus L.) upon bacteria challenge. Mol Immunol 134:48–61

Zhang Y, Liu Q, Yin H, Li S (2020) Cadmium exposure induces pyroptosis of lymphocytes in carp pronephros and spleens by activating NLRP3. Ecotoxicol Environ Saf 202:110903

Zhou F, Zhan Q, Ding Z, Su L, Fan J, Cui L, Chen N, Wang W, Liu H (2017) A NLRC3-like gene from blunt snout bream (Megalobrama amblycephala): molecular characterization, expression and association with resistance to Aeromonas hydrophila infection. Fish Shellfish Immunol 63:213–219