Diverse Immunoglobulin Light Chain Organizations in Fish Retain Potential to Revise B Cell Receptor Specificities[1]

Ellen Hsu[2]* and Michael F. Criscitiello[†]

We have characterized the genomic organization of the three zebrafish L chain isotypes and found they all differed from those reported in other teleost fishes. Two of the zebrafish L chain isotypes are encoded by two loci, each carrying multiple V gene segments. To understand the derivation of these L chain genes and their organizations, we performed phylogenetic analyses and show that IgL organization can diverge considerably among closely related species. Except in zebrafish, the teleost fish IgL each contain only two to four recombinogenic components (one to three V, one J) and exist in multiple copies. BCR heterogeneity can be generated, but this arrangement apparently provides neither combinatorial diversification nor an opportunity for the secondary rearrangements that, in mammals, take place during receptor editing, a process crucial to the promotion of tolerance in developing lymphocytes. Examination of the zebrafish IgL recombination possibilities gave insight into how the suppression of self-reactivity by receptor editing might be managed, including in miniloci. We suggest that, despite the diverse IgL organizations in early and higher vertebrates, two elements essential to generating the Ab repertoire are retained: the numerous genes/loci for ligand-binding diversification and the potential for correcting unwanted specificities that arise. *The Journal of Immunology,* 2006, 177: 2452–2462.

The adaptive immune system of vertebrates is based on a vast repertoire of Ag receptors expressed on their lymphocytes. Each Ig-producing B cell expresses a receptor of single specificity, composed of two H chains and two L chains whose N termini, the V regions, form the Ag-combining site. During B cell development in mouse, the H chain V region is generated by a recombination process consisting of the combinatorial assembly of one of >100 tandemly arrayed V gene segments, to one of at least nine D genes and one of four J gene segments (1). dsDNA breaks are introduced by the recombinase RAG, and during the repair and rejoining process, considerably even greater diversity is achieved through flank sequence deletion as well as addition of extra nucleotides at the junctions of the V/D and D/J joins (for a review, see Ref. 2). The V region of L chain, a VJ recombination, is similarly derived from a choice of ∼100 V and four J gene segments (3). This process, V(D)J rearrangement, generates as diverse a repertoire of ligand-binding sites as possible, in anticipation of a lifetime of pathogenic onslaught. Equally important, the product—a BCR of randomly generated specificity—must not be self-reactive. In the case of an autoreactive receptor, the specificity is altered, primarily by replacement of the L chain component through secondary rearrangements in a process called receptor editing (4–6). This is the first step leading to self-tolerance (7).

Receptor editing is crucial in the generation of the Ab repertoire, because it occurs at a frequency involving up to half of B cells (8, 9). Thus, the nature of the germline Ig gene organization must not

only enable the generation of V region diversity but provide an opportunity for replacing a nonviable H and L chain combination. For the κ L chain in mouse and humans, replacement mechanisms require the presence of multiple V followed by multiple J gene segments. If the first VJ rearrangement is unwanted because it is not in frame or incompatible with H chain or with the H chain combination forms a self-reactive receptor, secondary rearrangements take place in nested fashion between V genes upstream of the VJ and any available downstream J genes. With the exhaustion of replacement possibilities, the κ locus can be inactivated by deletion of its C exon; rearrangement continues at the allele, followed by attempts at the second L chain locus, λ (10, 11).

Although ensuring self-tolerance is fundamental to the adaptive immune system, this subject has been hardly considered in nonmammalian models. The organization of germline L chain genes in the earliest vertebrates, fishes, is unlike the tetrapod κ-style arrangement. In both cartilaginous and teleost fishes, the L chain genes are arranged in multiple "clusters," defined as independently rearranging miniloci consisting of few gene segments (one to three V gene segments, one J) and one C region exon (12). The nurse shark carries >70 such L chain clusters (13); the catfish, 85 (14, 15). This alternative gene arrangement still provides for the diversity aspect of V(D)J rearrangement, but, with so few gene segments per cluster, there appears no obvious opportunity for a secondary rearrangement to replace the first.

In this study, we have scanned the zebrafish genome and mapped some of its L chain genes. To our surprise, all three L chain isotypes display germline gene organizations differing from the conventional cluster configuration described above, one that is accepted as the IgL organization found in cartilaginous fishes and teleost fishes. To understand the relationship of the zebrafish L chain genes and the derivation of their organization from all of the other teleost fishes so far described, phylogenetic comparisons were performed. Based on taxonomy and available and deduced genomic data, we propose that all present-day teleost fish L chains derived from two isotypes that were present at Teleostei radiation, 240 million years ago (16). Based on the extensive sequence information available in the zebrafish database, which provided a

---

*Department of Physiology and Pharmacology, State University of New York Health Science Center at Brooklyn, Brooklyn, NY 11203; and [†]Department of Microbiology and Immunology, University of Maryland, Baltimore, MD 21201

[2] Address correspondence and reprint requests to Dr. Ellen Hsu, Department of Physiology and Pharmacology, State University of New York Health Science Center at Brooklyn, Brooklyn, NY 11203. E-mail address: hsue@hscbklyn.edu

more comprehensive overview of the IgL in its genomic context, we suggest how receptor editing might be managed in all fishes.

Materials and Methods

Animals

Five- to 8-mo-old zebrafish (*Danio rerio*) of the AB line were purchased from Zebrafish International Resource Center (Eugene, OR). Genomic DNA was obtained from individual carcasses using standard methods. C and V sequences of types 1–3 (17) were obtained by PCR from genomic DNA (18) and cloned (accession nos. DQ343247–DQ323252). Genomic Southern blotting was performed as previously described (19), using hybridization conditions of routine stringency.

Database searches

The NCBI database (⟨www.ncbi.nlm. nih.gov/BLAST/⟩) was used to search for V and C regions of zebrafish L chains (17) (type 1, LC1–8, AF246185; type 2, NCL106, AF246180 and AF246181; type 3, LC1–7, AF246193). Sequences that contained matches for both V and C were downloaded and screened for the presence of additional recombination signal sequences (RSS).[3] The tblastn program was efficacious in finding variant C sequences of the same isotype.

Phylogenetic analysis

Predicted amino acid sequences of V and C regions from cDNA and genomic sequences were aligned in ClustalX, version 1.81 (20), using blosum weight matrix, gap opening penalty of 10, gap extension penalty of 0.05, and 40% divergence cutoff. The V alignment was trimmed to the framework regions. Distance matrices and neighbor joining trees were generated with PhylipW (21) using the Protdist and Neighbor programs. PhylipW, Seqboot, and Consense were used for bootstrap analysis (1000 times) to test the reliability of the bifurcations. Dendrograms with similar topography were also drawn from these data using other software and algorithms in Bioedit, including ProML (maximum likelihood), Fitch (Fitch-Margoliash), and Protpars (parsimony). Treeview, version 1.66 (22), was used to draw the trees rooted with IgHV (for IgLV) and IgHC4 (for IgLC) from mouse as outgroups. Pairwise identity matrices were created from Clustal alignments using Bioedit.

Results

The zebrafish L chain isotypes were classified by Haire et al. (17) by C region identity (~30%); this degree of shared sequence is what distinguishes mammalian κ and λ C regions, 35–37%. Contig sequences from the database were selected for analysis based on the presence of matches with both L chain V and C regions. The organizations of loci encoding L chain types 1–3 are shown in Fig. 1. The diversity of the V and C genes has already been discussed (17); some considerations as to their derivation are in the Fig. 1 legend. The organization and rearrangement potentials of the three L chain isotypes will be presented in the order: type 2, followed by type 3 and type 1. The type 2 locus is the most complex, containing multiple V, J, and C genes. The type 3 locus has a different organization, similar to one of the type 1 loci; the remaining type 1 loci are in the well-known cluster configuration.

Type 2 L chain organization

Type 2 genes are located over an area of 36 kb on contig NW_644395 (1759 kb). There are 12 V gene segments, 4 J gene segments and 2 C exons (Fig. 1). There are no C sequences in the 4-kb interval between *J2b* and *V2i*, or after *J2d*. The *C2a* is identical with the published type 2 C region (17). The *C2b* sequence is 59% identical with *C2a* at the nucleotide level and 45% in the derived amino acid sequence. That is, it is well diverged from *C2a* but nonetheless distinguishable from type 1/type 3 C sequences (22–31% identity in all inter-isotype pairwise comparisons).

All of the genes are in the same transcriptional orientation and are so closely packed that they might operate as one locus. The distance between *C2a* and *V2g* is <2.5 kb; if they are not differentially activated, conventional recombination between V and J would occur by deletion of intervening DNA to form a VJ (Fig. 2*A*) or deletion of C (Fig. 2*B*). In any case, genes *V2a* through *V2f* can rearrange to *J2c* and be spliced to *C2b*. It is not clear whether any rearrangement to *J2a* or *J2b* would be efficiently spliced to *C2b*.

The multiplicity of J segments scattered throughout the locus provide the potential for recombination events that could inactivate the locus. Recombination between *J2a* and *V2g-V2k* would delete the intervening DNA carrying *C2a* (Figs. 1 and 2*B*), and assuming *J2d* is part of the activated locus, any rearrangement with *J2d* excises the intervening DNA carrying one or both C exons. The biological significance of these possibilities will be considered in *Discussion*.

Type 3 L chain organization

A type 3 locus is located on chromosome 5. There are eight V gene sequences, one J gene segment, and one C exon over 18 kb (Fig. 1, *bottom*); no other Ig sequences were detected in over 1600 kb on contig NW_634729. Six V gene segments are located upstream of the J and C; two more are 3′ of the C. Of the eight V genes, seven are in opposite transcriptional polarity to *V3b*, *J3a*, and *C3a*.

V genes on both sides of the J/C can rearrange functionally to the J. The V genes downstream can recombine by inversion of the J and C (Fig. 2*D*), whereas those upstream will themselves invert to join J (Fig. 2*C*). The finding that both downstream V gene segments *V3g* and *V3h* appear to be functional argue for their active usage. Thus, this type 3 locus consists of eight V gene segments all potentially able to recombine with the J.

Type 1 L chain organization

A search with the zebrafish type 1 C sequence showed good matches from contigs NW_646408 and NW_633979 (Fig. 1, *top*); the former carries two C regions that are 99 and 100% identical with the LC1–8 cDNA sequence and the latter, one C region sharing 87% identity. Both contigs contain multiple V region sequences that were distinguished after searches with the type 1 V sequence and with the type 1 RSS. Thus, type 1 L chains are encoded by several loci, as originally deduced from the finding of different C sequences in a cDNA library (17).

NW_646408 carries four V gene segments interspersed among two J and C genes, as shown in Fig. 1, *top left*; its 5′ region is unknown, being at the very end of the contig. NW_633979 is assigned to chromosome 1 and contains seven V genes, one J gene segment, and one C exon over 18 kb (Fig. 1, *top right*). The organization and rearrangement potential is like the type 3 locus; it too apparently exists as an isolated Ig locus.

Additional matches were found using tblastn program with the *C1a* amino acid sequence. On NW_644842 (34 kb long), an isolated V-V-J-C cluster was detected; this C sequence (*C1k*; see supplemental data)[4] shared 58–62% identity with *C1a-c* at the amino acid level. On NW_633913, assigned to chromosome 19, seven partial and complete C sequences were found accompanied by J gene segments; the functional ones share 43–62% identity with *C1a,b,c,k* and 90–100% with each other, showing that they duplicated among themselves and diverged long ago from the others. Ten V sequences were detected, placed in the opposite transcriptional orientation to the nearest J and C, about one or two per cluster. Over a span of 516 kb, these genes form four clusters, separated by intervals of 3.7–418 kb. Thus, in contrast to the IgL on chromosome 1, these type 1 loci resemble the kind of small IgL clusters reported in other bony fishes.
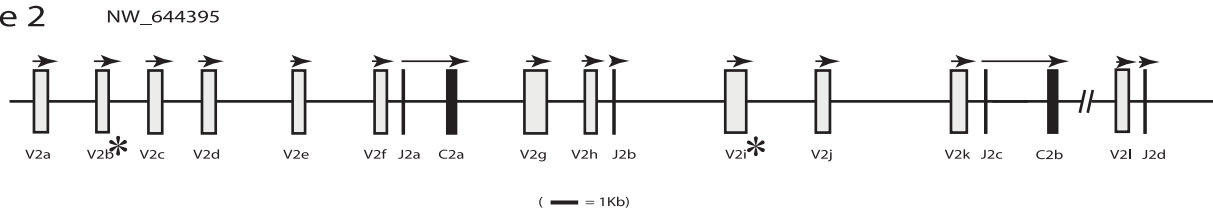
---

[3] Abbreviation used in this paper: RSS, recombination signal sequence.

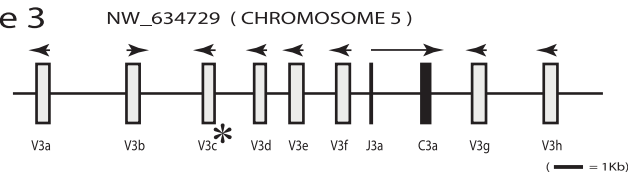[4] The online version of this article contains supplemental material.
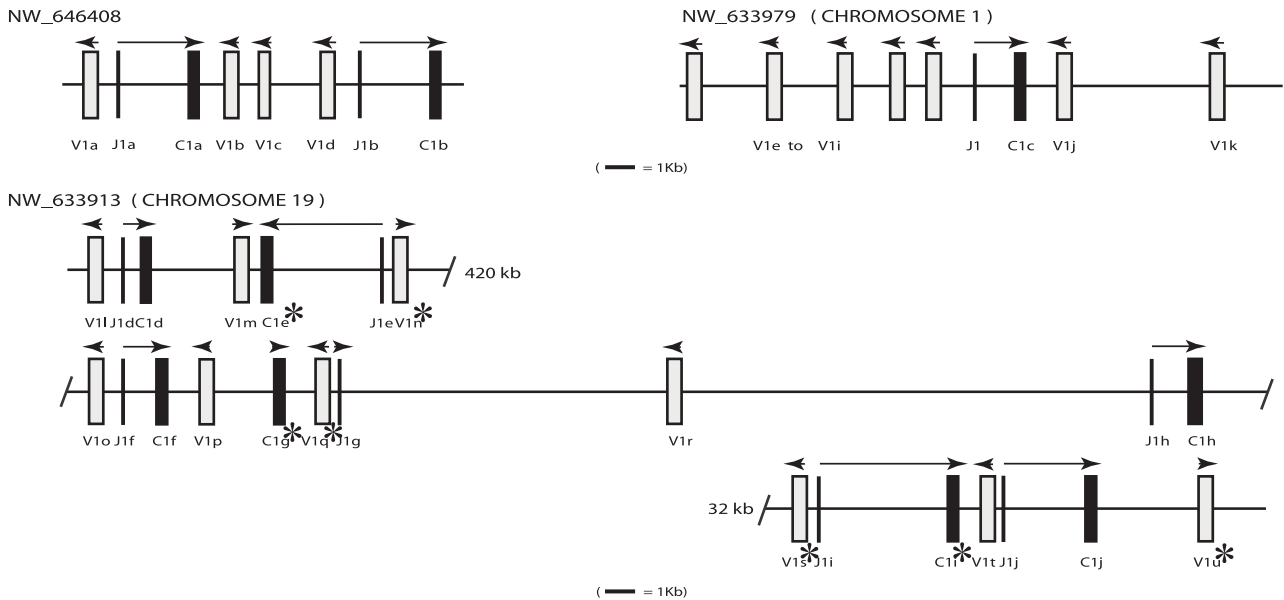
FIGURE 1. Organization of representative genes encoding zebrafish L chains. Five contigs from the zebrafish database were analyzed for the arrangement of L chain V (gray boxes) and J gene segments and C exons (black boxes). The transcriptional polarity is indicated by overhead arrows. Each gene is labeled, and an asterisk indicates presence of stop codon or discontinuous coding sequence. The sequences of functional genes and their positions are available in supplemental Fig. 1*b*. Type 1, Three contigs (NW_633979, NW_646408, NW_633913). NW_633979, The coding sequences of *V1f/V1h* 85% share identity and their leader introns 99% identity, suggesting that they are duplications. The complete sequence of *V1j* is not available from the current database. Most of these sequences closely resemble the four V genes on contig NW_64608: *V1a* is overall 85% identical with *V1i*, *V1b* 89% with *V1k*, and *V1c* is 97% identical with *V1e*. *V1d* is 92% identical with *V1a*. The end of the contig sequence is nearby; there probably exist additional upstream V genes, because a J gene segment (not shown) is present 10 kb upstream of *V1a*. NW_633913 contains a series of clusters extending over 516 kb. There are four functional C exons; the overall V sequences are well diverged from each other (31–100% identity) and from the nonlinked ones described above (31–67% identity). The V gene segments appear to be functional except for *V1n*, *V1q*, *V1s*, and *V1u*, which have stops; all RSS are intact. There is no J gene segment between *V1p* and *C1g*, nor is there a C exon between *J1g* and *V1r*. Type 2 (contig NW_644395), The functional V fall into two groups: *V2b*, *V2e*, *V2f*, *V2g*, *V2h*, *V2k* and *V2a*, *V2c*, *V2d*, *V2i*, *V2j*, *V2l*. Within a group, they are 89–95% at the nucleotide level in the V gene segment; between the two groups, they are 50–55% identical. Thus, these V genes are the least diverse among the three isotypes. We tried to ascertain how the four J gene segments are related by including the nearest upstream V gene in the sequence comparisons. A search using the sequence encompassing *V2h-J2b* (1144 bp) revealed that *V2f-J2a* share >90% identity over 920 bp, from the leader extending into the intervening DNA 450 bp beyond the RSS of the V gene. This suggests that they are relatively recent duplications. *V2k* is 90% identical with them, but homology drops off after the RSS. *J2c* also differs from the identical *J2a/J2b/J2d*, but the overall impression is that this type 2 locus arose from a series of tandem duplications. *V2l-J2d* are located 22 kb from *C2b*. All of the V gene segments appear to be functional except for *V2b*, which may have a nonfunctional RSS and *V2i*, which has an insertion in CDR3. Type 3 (contig NW_634729), The seven V3 gene segments overall share 46–94% nucleotide identity in the coding regions. The location of the V genes that are similar (*V3a/V3h*, 91%, and *V3e/V3g*, 95% identity) give little clue as to how the locus evolved; extensive changes have occurred that are difficult to reconstruct. *V3a* through *V3f* could have been generated by tandem duplication, but it is also possible that two clusters could have become merged, followed by deletion of a J and C formerly located after *V3h*. The *V3b* segment is in a different transcriptional orientation, but because the coding sequence is not outstanding, its position probably resulted from a meiotic inversion event. *V3c* is a pseudogene.

The type 1 L chain in zebrafish are encoded both by small V-J-C clusters and an "expanded" cluster like those in types 2 and 3. This is a range of variant gene organizations not previously observed in other fishes, perhaps due to the limitations of cloning with bacteriophage vectors.

Genomic Southern blotting results from our laboratory (not shown) for C sequences were identical with those obtained by Haire et al. (17), who also screened a zebrafish BAC library. The many (7–9) C1-hybridizing bands found in *Eco*RI- or *Hin*dIII- or *Pst*I-cut genomic DNA corroborate the maps in Fig. 1, showing
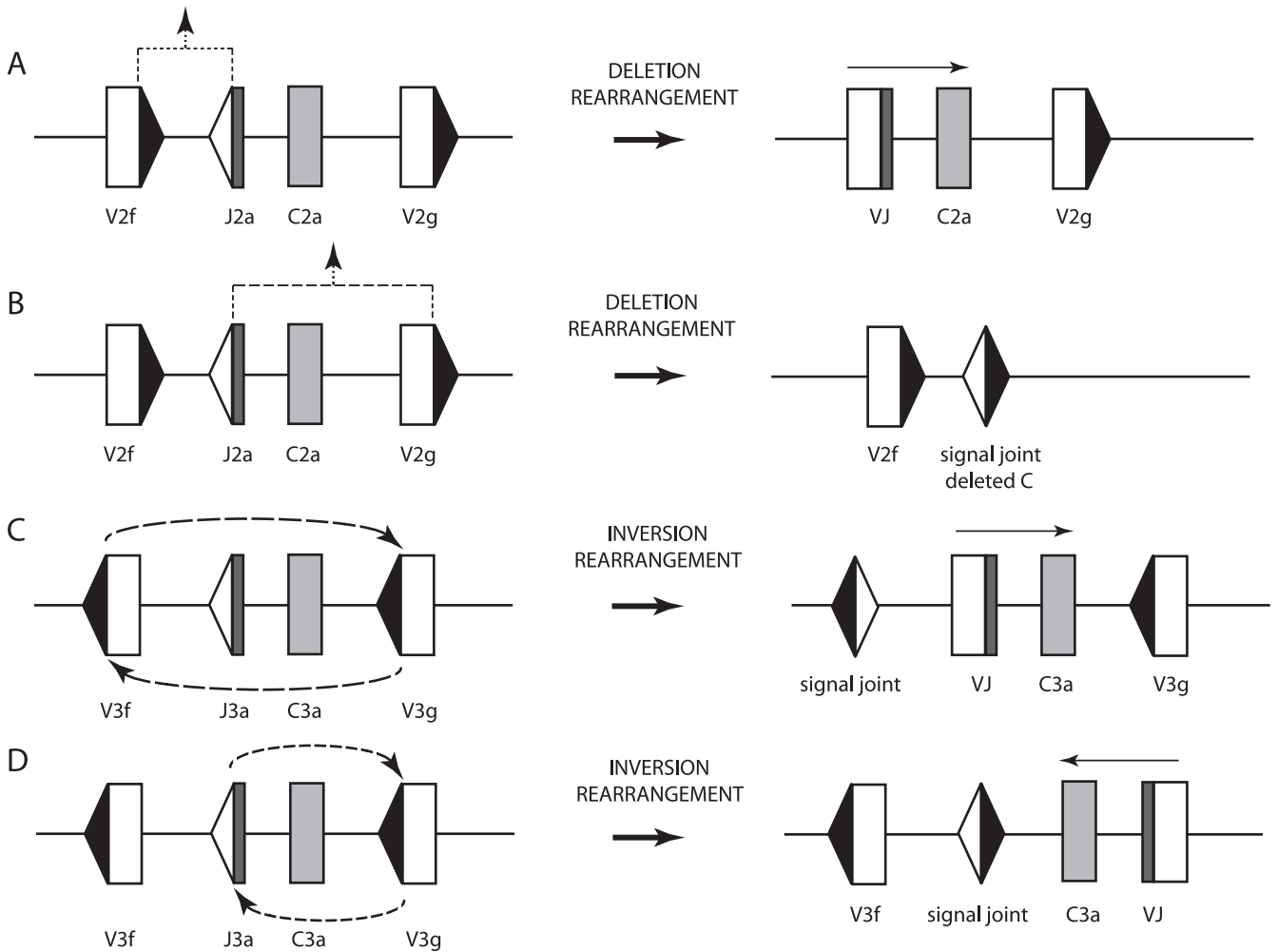
FIGURE 2. Recombination at the type 2 and type 3 loci. *A*, Deletion rearrangement at type 2 locus. *B*, Possible deletion rearrangement excising *C2a* exon. The excised region is indicated by brackets. *C*, Inversion rearrangement at type 3 (or type 1) locus. *D*, Inversion rearrangement to 3′ V gene segments in type 3 (or type 1) locus. Arrows show the sites of DNA breakage and inversion. The transcription polarity of the rearranged VJ, at the *right*, is indicated for *A*, *C*, and *D*. The RSS with 12-bp spacer is indicated as a black triangle, the RSS with 23-bp spacer is indicated as a white triangle.

that the type 1 L chain is encoded by multiple clusters, whereas the few (2, 3) C-hybridizing bands for types 2 and 3 suggest about two loci each.

Type 1 and 3 share a common ancestor

Together with representative zebrafish sequences, the V and C regions of various teleost L chains were examined for their phylogenetic relationship (Fig. 3). These particular species were selected based on the existence of organizational information for one or all of their L chain types (Table I). The deduced organization of salmon type 2 is based on the finding of a germline transcript containing an unrearranged V gene, the V-J intergenic sequence, and a J spliced to the C sequence. Such transcripts are only possible when the V, J, and C are in close proximity and in the same transcriptional orientation, like *V2f-J2a-C2a* (Fig. 1). The two reported in zebrafish and in salmon are both type 2 L chains (17, 23).

In Fig. 3, *top*, it can be seen that the type 2 V regions grouped strongly together (>99% of bootstrap resamplings) and are distinct from the type 1/3, which appear to be intermixed. The close relationship of type 1 to type 3 is further supported by the similar configuration of the germline genes, in contrast to all the type 2, where the V are in the same transcriptional orientation as J and C (Table I). The statistically robust phylogenetic groupings of sequences, together with the shared organizational characteristics,

are the strongest evidence for the common derivation of type 1 and type 3 L chains.

The classification of a teleost fish L chain is most reliably established through its C region. The phylogenetic tree of the C regions (Fig. 3, *bottom*) shows strongly supported branches where sequences of the same isotype are clustered (>98% of bootstrap resamplings), suggesting common descent of all type 2 and of all type 1 C region genes among these fishes. The one exception among type 3 C regions is the salmon sequence named "type 3," which is found amid the type 1 genes. Catfish, zebrafish, and carp, whose type 3 sequences group together and independently of the other C regions, all belong to Ostariophysi; in contrast, salmon belongs to Protacanthopterygii (see Fig. 4). The type 3 L chain genes in Ostariophysi fishes thus descend from a common ancestral duplication from type 1 L chain genes. We suggest that, in contrast, the third salmon L chain characterized by Solem and Jørgensen (23) derives from an event specific to Protacanthopterygii or Salmoniformes (hence the asterisked "L3" is distinguished in Fig. 4). A more recent divergence could explain why the salmon "type 3" and type 1 C regions share 61.5–63.3% amino acid identity, a relationship that is more comparable to zebrafish *C1a-c* with *C1k*, which share 58–62% identity. In contrast, between isotypes, the zebrafish type 1 and 3 share 23–27% and the catfish G (type 1)
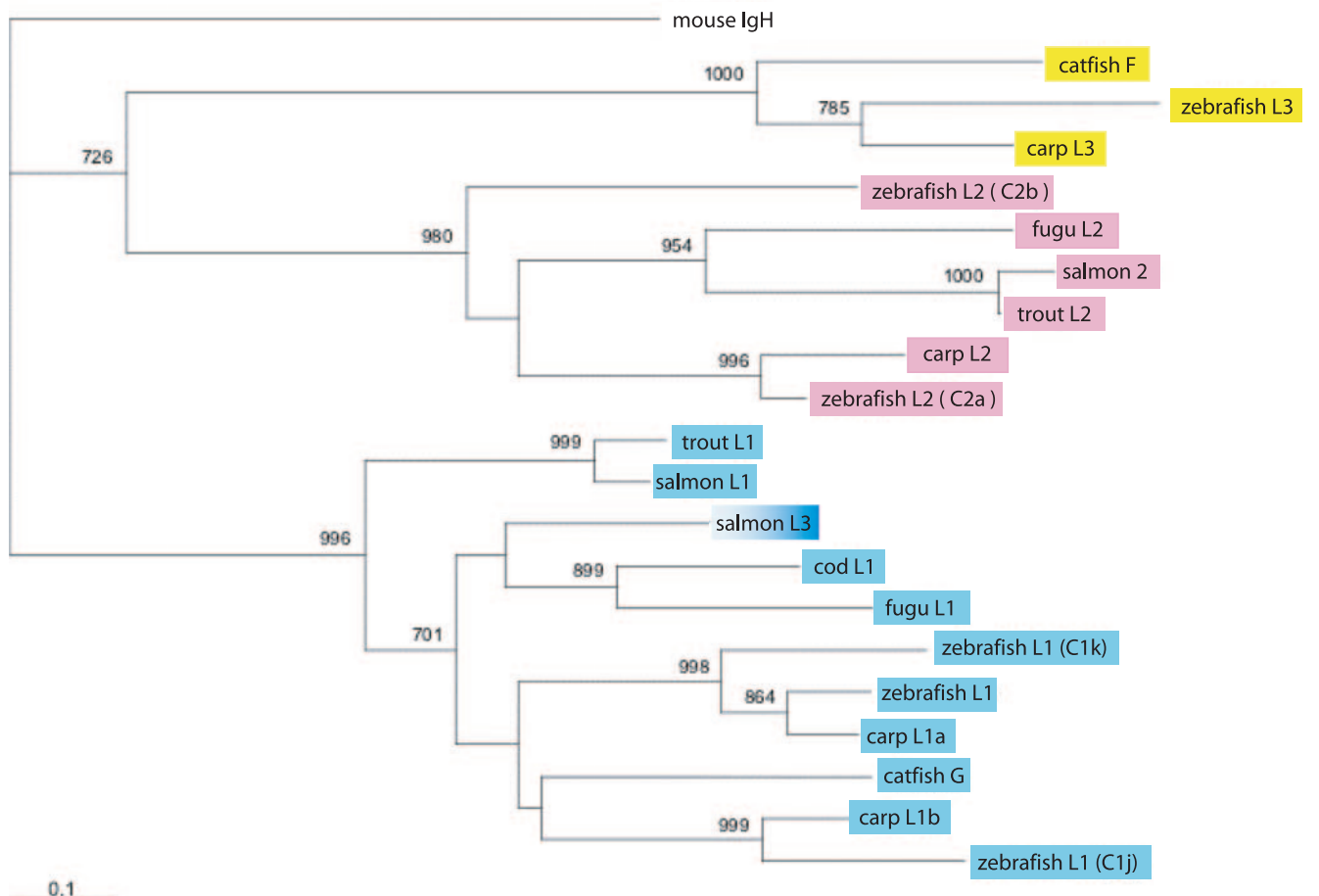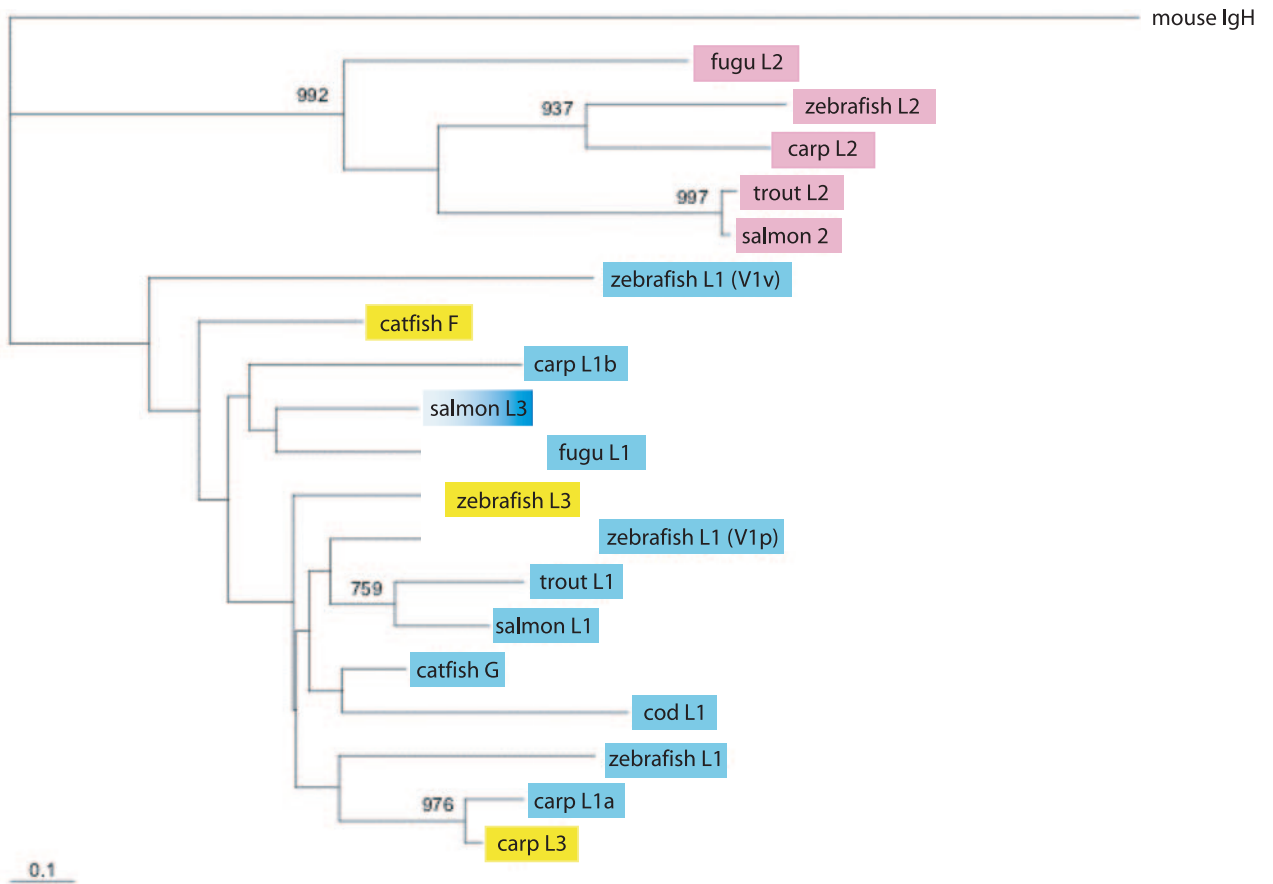
FIGURE 3. (*Legend continues*)

Table I. *Teleost L chains grouped according to isotype and organization*

L Chain Type[a]	Gene Segment Configuration	Organization Type
Zebrafish type 1	V̂1–V̂2–V̂3–J–C	Cluster (17); extended cluster (this study)
Trout L1	V̂ –J–C	Cluster (28)
Codfish	V̂1–J–C	Cluster (29)
Catfish G	V̂1–V̂2–J–C	Cluster (14)
Fugu type 1	V̂1–V̂2–J–C	Cluster (47)
Salmon type 1	Unknown	Probably cluster, many V and C (23)
Zebrafish type 2	V-JC germline transcript	(17; cDNA clone ZFNLC 3)
	V̂1–V̂2–...V̂6–J–C	Extended cluster (this study)
Trout L2	V̂1–V̂2–V̂3–J–C	Cluster (28)
Salmon type 2	V-JC germline transcript	Unknown (23; cDNA clone 01–14, AF297518)
	V̂n–J–C (deduced)	
Fugu type 2	V̂ –J–C	Cluster (this study)
Zebrafish type 3	V̂1–V̂2–V̂3–...V̂6–J–C	Extended cluster (this study)
Catfish F	V̂1–J–C	Cluster (15)
Salmon type 3	V̂n–J–C	Unknown (23)

[a] Except for catfish, all sequences are classified according to investigators (type 1–3 or L1–3).

and F (type 3) share <35% (15); this level of identity distinguishes mammalian κ and λ C regions.

Because type 2 is present in Ostariophysi, Protacanthopterygii, as well as Acanthopterygii (see below), its apparent absence in catfish and others might be attributed to species-specific deletion or to a L chain subpopulation that has yet to be identified. In fact, there exists in the catfish pronephros EST library a L chain sequence (accession no. CK403931) that may be a candidate for the catfish type 2 L chain. A preliminary analysis shows greater similarity to bony fish type 2 sequences than to the catfish F or G L chains; however, the establishment of its identity requires further investigation (M. Criscitiello, unpublished observation).

We suggest that type 1 and type 2 L chains have been present since the emergence of Euteleostei. Type 3 could have evolved early after Ostariophysi divergence because no type 3 homolog has so far been reported in other fish orders; if indeed that is the case, type 3 is by definition not a teleost isotype but a subtype.

Type 2 L chain sequence in puffer fish

We used the zebrafish type 2 V sequence to search the fugu database (⟨http://fugu.biology.qmul.ac.uk/⟩). Two sequences (clones M007644, accession no. DQ471453; M006921, DQ471454) are >2 kb in length and carry one V and one J gene segment and one C exon. All are in the same transcriptional orientation; the analysis of M007644 is shown in supplemental Fig. 1a. Searches using the fugu C sequence resulted in five more matches at 99–100% identity. Because M006921 has a V with a stop codon (supplemental Fig. 1a), M007644 a functional V, and another clone contains two J segments and one C (M007311), the heterogeneity suggests an organization of multiple clusters (Table I). The fugu V and C sequences group with the other type 2 (Fig. 3) and provide additional proof that type 1 and type 2 loci existed early in teleost radiation.

Searches were also performed with zebrafish C3 and C1 sequences; among fugu sequences matching to both, all showed greater similarity to C1. One sequence with similarity to only C3 encodes an Ig-like domain; however, there were no V, J, or RSS sequences within a 10-kb vicinity, nor were there any matches to transcribed fugu or tetraodon sequences.

Discussion

We have characterized the genomic organization of the three zebrafish L chain isotypes and found that all three displayed features that have not been reported in other bony fishes. This provides further evidence of the plasticity Ig organization (for a review, see Ref. 24). In an effort to understand the phylogenetic position of these L chains and their organizations, we compared them with sequences from other fishes and propose a hypothesis on their relationships and origins. We also suggest that despite these diverse IgL organizations, two elements crucial to generating the Ab repertoire are retained: sequence diversity and receptor editing, or recognition of nonself and avoidance of autoimmunity, respectively.

Origin of teleost L chain types

The subclass Actinopterygii, ray-finned fishes, consists mostly of teleost fish species, and all the L chain sequence information available is limited to this group, with the exception of sturgeon (*Acipenser baerii*) as a representative of older extant fishes (see taxonomic relationship depicted in Fig. 4). Phylogenetic trees have been generated in several studies, but there has not been a recent attempt to unify the information and sort out the relationships among the many L chains reported. In this study, we limited the

FIGURE 3. Phylogenetic analyses of L chain V and C from various teleost fishes. *Top*, V domain. *Bottom*, C domain. All type 1/L1 sequences are indicated in blue, type 2/L2 in pink, and all type 3/L3 in yellow except salmon, which was named type 3, but which we designate as a type 1 (see *Results*). The following sequences were used in alignment. *IGH*, Mouse *Mus musculus* V MUSIGHVP, C MUSIGCD10. *IGL*, Channel catfish *Ictalurus punctatus* F (U25705, G L25533); zebrafish *D. rerio* L1(AF246185); L1(V1v) and L1(C1k) from NW_644842; L1(V1p) and L1(C1j) from NW_633913; L2 V (AF246183); L2(C2a) from NW_644395; L2(C2b) from NW_644395; L3 (AF246193); fugu pufferfish *Takifugu rubripes* (L1 AB126061, L2 M007644); rainbow trout *Oncorhynchus mykiss* (L1 X65260, L2 V OMU69988, L2 AJ251648); carp *Carpio cyprinus* L1a V (AB073328); L1a C (AB015905); L1b V (AB073332); L1b C (AB035728); L2 V (AB091112); L2 C (AB091120); L3 V (AB073335); L3 C (AB035730); Atlantic salmon *Salmo salar* L1 (AF273012); L2 V (AF406963 and AF406964); L2 C (AF297518); L3 (AF406956); Atlantic cod *Gadus morhua* L1 (AF104898). Bootstrap values >70% are shown at nodes.
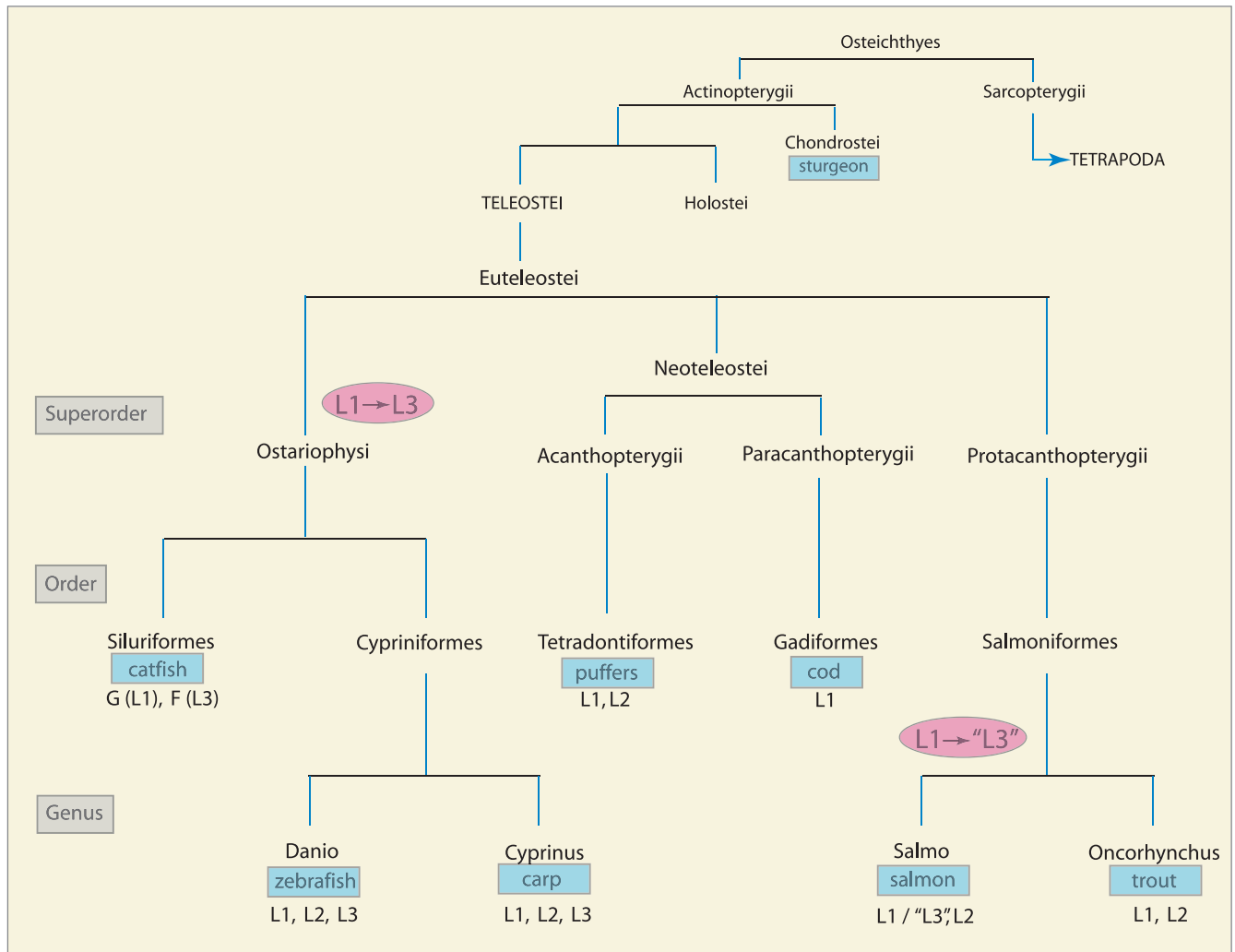
FIGURE 4. Taxonomic relationships among several teleost species. The taxonomic classifications of the teleost fishes listed in Table I are displayed according to superorder, order, and genus (48). The L chains that have been reported here and in the references in Table I are listed below the common name of the fish model (blue box). We propose that the type 3 (L3) is derived from type 1 (L1) in separate events in Ostariophysi and in Protacanthopterygii (pink ovals) (see *Results*). The carp L1A and L1B, considered by some investigators to be two isotypes (49), carry C regions that are highly related to zebrafish C1k and C1j, respectively (see *Results* and Fig. 3, *bottom*); we consider all these to be variants of type 1 (L1).

comparisons to seven species where genomic information was available or could be deduced. A correlation between the genomic configuration of the genes, in terms of transcriptional polarity, and position in the phylogenetic trees distinguishes two groups of V genes.

The type 2 V gene sequences from zebrafish, trout, salmon, fugu, and carp not only clustered together (Fig. 3, *top*) but are all organized in the same transcriptional orientation as the J and C genes (Table I). This suggests a common ancestry and, given their presence in different orders of fish, the existence of a type 2 gene early in Teleostei.

The V genes of type 1 and type 3 cluster together on the phylogenetic tree, without isotypic distinction among the different species (Fig. 3, *top*). Moreover, in all eight known instances, the V genes of both types are inverted with respect to the J and C genes (Table I). This, together with the clustering of the type 3 C sequences from catfish, carp, and zebrafish (Fig. 3, *bottom*), suggests that type 3 was derived from type 1 at least during Ostariophysi evolution (Fig. 4). The relationship between type 1 and 3 has been suggested previously (for instance, see Ref. 25), but this study is the first to demonstrate the point in several species and by phylo-

genetic and genomic comparisons as well as taxonomic relationships. At this time, we failed to find a type 3 C region homolog in fugu (Acanthopterygii), but if in the future such sequences are found in non-Ostariophysi species, this would place the L1/L3 divergence at or before emergence of Euteleosts.

It is not known when type 1 and type 2 L chain genes diverged, but at some time the type 1 V genes acquired their inverted orientation. This aspect is particular to teleost clusters, because the Ig genes in cartilaginous fishes are all in the same transcriptional orientation. Perhaps the inversion was the event that served to isolate the primordial type 1 locus. The situation in nonteleost bony fishes has not been much investigated, the only example being sturgeon, where the relative V-J orientation is unknown (M. Lundqvist, personal communication). On phylogenetic analysis (23), the sturgeon C region appears unique and the V region clusters with teleost type 1/3, but the relationship of this L chain to the teleost L chains is not clear, in the absence of more information. The interest of the sturgeon L chain lay in its κ-type organization (26), with multiple tandem V, multiple tandem J, and single C exon, which is discussed below.

Evolution of IgL gene organization

The cluster organization of L chain in fishes was initially discovered in trout and cod (27), and subsequent studies, the most comprehensive ones in cartilaginous fishes (horn shark, nurse shark, sandbar shark, little skate), catfish, cod, and trout reinforced an idea of its universality among fishes. In the various teleost species, the L chain isotypes are encoded by multiple loci, each consisting of one C exon, one J, and one to three V gene segments (14, 15, 28, 29), and in elasmobranchs each IgL locus contains one V gene segment. In this context, the finding of a noncluster organization in sturgeon was the first indication that early IgL evolution involved extensive diversification of gene configurations. Lundqvist et al. (26) suggested that a primordial V-J-C cluster in the ancestral vertebrate expanded by duplicating the entire locus, which led to the multiple clusters in Chondrichthyes, or by duplicating V and J gene segments, which resulted in the κ-type organization found in sturgeon and in Sarcopterygians (essentially, all tetrapods). The teleosts, which diverged later, were proposed to have recreated the cluster type organization independently, by trimming down the number of V genes (and inverting them) and duplicating the downsized locus.

Up until now, the multiple clusters appeared to be an either-or alternative to the κ-type arrangement: the three nurse shark L chain isotypes encoded by 70 IgL loci are all V-J-C clusters, whereas the three *Xenopus* IgL isotypes are encoded by three loci all in κ-type organization (references in Refs. 11 and 12). Therefore, we were surprised to find that, in zebrafish, the IgL organization of two of the isotypes are organized as loci with multiple V and one to two C and another can consist of clusters but also acquire multiple V gene segments per locus. This characterization of the zebrafish database information is supported by genomic Southern blotting results.

Although the ancestral type 1 must have had the cluster organization, because that is its configuration in all orders shown in Table I, at least one zebrafish type 1 locus (at NW_633979) is different, suggesting extensive genomic restructuring events during zebrafish evolution. If an L chain isotype is encoded by a series of closely linked loci, they can be pared down or expanded by a large unequal crossing-over event; if the loci are scattered throughout the genome, they are less prone to wholesale alteration. We would argue that the original type 2 loci in zebrafish were linked, as were the type 3, whereas the type 1 loci were, and are, at different locations, as suggested by polymorphism between individuals of the AB line on the Southern blots (not shown) and the multiple matches in the database. Just as segmental duplication may have expanded the type 1 clusters (as in NW_633913), segmental deletions may have whittled down the number of type 2/3 clusters.

The zebrafish type 2 and 3 have an increase in the number of V gene segments in the locus, together with a reduction in the total number of loci, which is present at about two per genome. They appear to be an intermediate between cluster and κ-type arrangements, and might be described as extended clusters. How quickly did these changes take place? We have shown that the zebrafish type 3 is the homolog of the catfish F L chains. The F L chain genes are well established as existing in typical V-J-C cluster formation: this suggests that the original type 3 gene in Ostariophysi had the capability of evolving to either kind of configuration within the time after catfish and zebrafish divergence, 65–127 million years ago (30). Moreover, if a κ-type organization in the sturgeon-teleost ancestor can be converted to cluster-type at the time of teleost divergence (26), this process was reversed within a species like zebrafish. We suggest that the L chain gene organization can be more volatile than heretofore thought.

Then how unique is zebrafish? At this time, there are only six teleost L chains, two to three loci representing each, whose genomic configurations have been characterized (Table I), and possibly further investigation would reveal greater heterogeneity of IgL organization than is apparent in the limited stretch of DNA obtained in bacteriophage vectors. For instance, the close taxonomical relationship of zebrafish and carp (Fig. 4), together with the phylogenetic grouping of zebrafish L1 and carp L1a (Fig. 3, *bottom*) suggests that those carp genes could be candidates organized like the zebrafish homolog on chromosome 1 (Fig. 1). The more promising situations to investigate would be those where only few C region and multiple V region bands can be visualized in genomic Southern blotting.

One curious feature in the extended clusters is the presence of V gene segments 3′ of the C exon (*C1a, C1c, C2a, C3a*). In the case of type 1 and type 3, these genes are available for conventional V-J joining. We suggest in the last section that possibly there may exist another reason for maintaining this arrangement.

Relationship of teleost L chains to mammalian κ and λ

The evolutionary ancestry of the mammalian κ and λ has not been fully investigated with respect to all vertebrate classes. In particular, incomplete sampling of animals in crucial phylogenetic positions, such as sturgeon, reptiles, and nonplacental mammals, leaves significant gaps in the natural history of the L chain. Through phylogenetic analyses, κ orthologs have been proposed in the amphibian *Xenopus* (rho L chain; Refs. 31 and 32), in bony fishes (type 1 and type 3; Refs. 17 and 32), and in the shark (NS4, type III; Refs. 32 and 33). λ orthologs in contrast have been found in chicken and suggested in *Xenopus* (λ; Ref. 32), and shark (NS3, type II; Ref. 32), but none so far in bony fishes. Instead, the bony fish type 2 is most closely related to the *Xenopus* σ (sigma; Ref. 32) and a fourth shark L chain (M. Criscitiello, unpublished observations) but has no known orthologs among higher vertebrates.

Thus, the overall picture of L chain descent is incomplete, sometimes due to the deletion of the L chain genes, like κ in chickens, or like *Xenopus* σ presumably in early reptiles antecedent to birds and mammals. The absence of λ in bony fishes could be attributed to its deletion after Teleostei divergence or to as-yet-undiscovered isotypes; this can be investigated once the entire zebrafish genome is available for such analyses. In summary, of four L chain isotypes found in the representatives of the earliest vertebrates, sharks, one (NS5, type I) is restricted to cartilaginous fishes (32) and three varyingly have orthologs in bony fishes and tetrapods (M. Criscitiello, unpublished observations).

Why more than one L chain locus?

Should a self-reactive Ig receptor be created or the H-L pairing is so poor that inadequate levels of receptor reach the surface, rearrangement continues in the mammalian pre-B cell. A new VJ combination can be produced by nested rearrangement, if possible, or the locus is inactivated by deletion of the C exon (Fig. 5A). The latter process entails recombination between the RSS ("RS element") downstream of the C exon and either the isolated heptamers ("IRS") in the J-C intron or, more frequently, V genes upstream of the unwanted VJ rearrangement (Fig. 5A) (34–36). We will refer to the RS element and the IRS together as "RS/kde." With successful κ rearrangement and expression, feedback mechanisms leave the λ gene in germline configuration; in λ-expressing B cells, the κ locus has usually rearranged and is nonfunctional (37). The λ locus can conceivably be viewed as a reserve in the event of failure at the κ locus to produce an appropriate BCR (38),
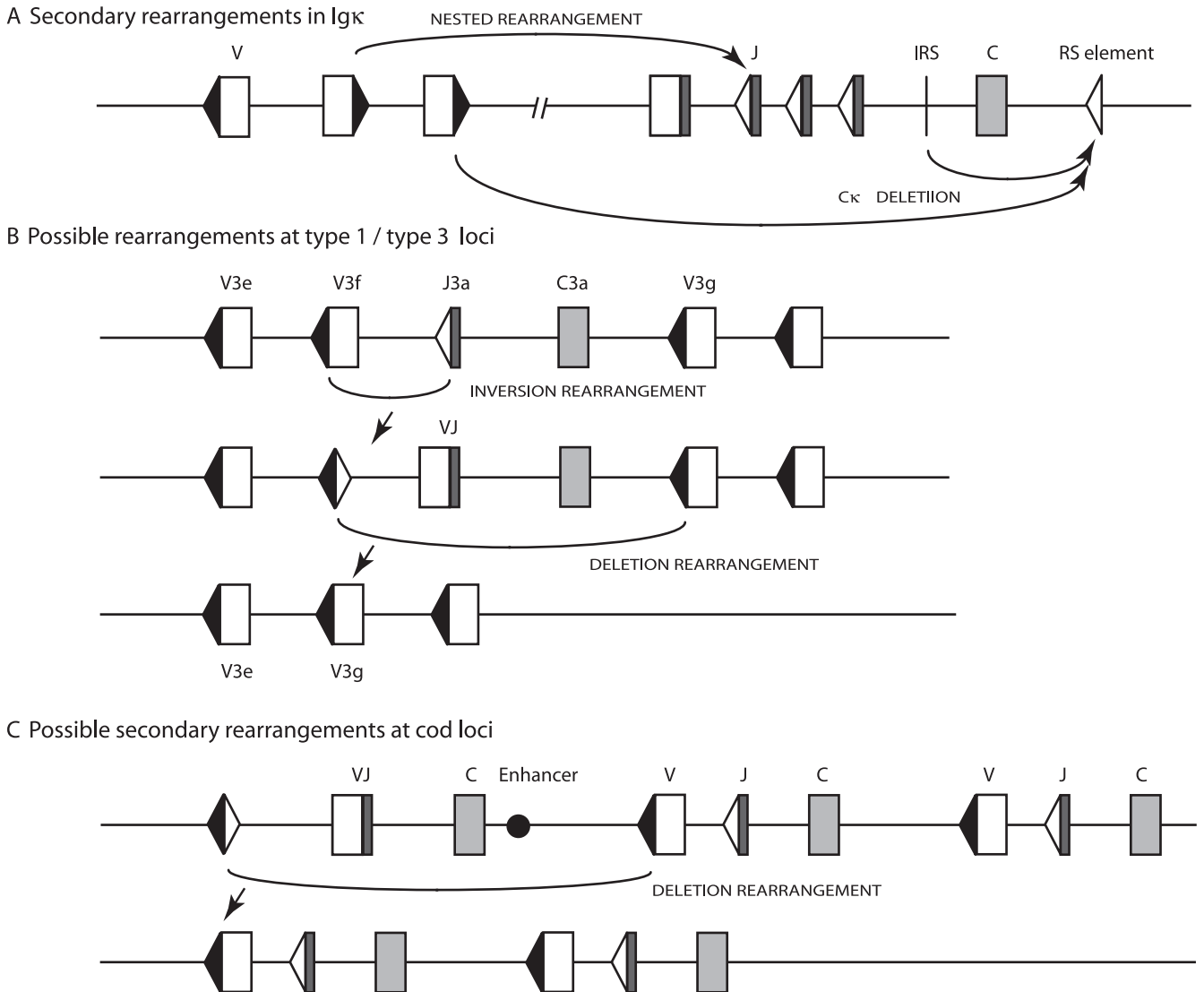
FIGURE 5. Receptor editing possibilities at fish IgL. *A*, Secondary rearrangements at the mouse κ locus. Recombination involving nested rearrangement of upstream V gene segment and downstream J gene segment can replace and delete unwanted VJ. Recombination between upstream V gene segment and RS element deletes C exon and inactivate locus, as does recombination between a heptamer (IRS) in the J-C intron and the RS element. After Moore et al. (35). *B*, Possible secondary rearrangement at a type 1 (or type 3) locus. Downstream V gene segments can rearrange with RSS in the signal joint to delete unwanted VJ. *C*, Possible secondary rearrangement among cod L chain clusters. Modeled after genomic clone CgL10, where enhancer activity was found 3′ of only the first cluster; it was proposed that one enhancer can regulate several clusters (29). With an unwanted VJ, rearrangements continue at the activated region. Another VJ recombination may occur, or the first VJ may be deleted, as shown, by rearrangement between the RSS in the signal joint and an inverted, downstream V gene segment. If the latter occurs first, the enhancer is deleted and the area may no longer be active. The RSS with 12-bp spacer is indicated as a black triangle, and the RSS with 23-bp spacer is indicated as a white triangle.

and its importance cannot be underrated because, in humans, the κ-λ expression ratio is 60:40, although the κ locus is activated first (11, 39).

Does receptor editing occur in nonmammals? In chickens, there is one L chain locus, with one functional V and one J segment; in such a situation, the two germline sequences and their very few CDR3 possibilities (in the absence of terminal transferase additions and before gene conversion) could have been selected, at the species level, to not generate autoimmune specificities. In *Xenopus*, all three L chain loci carry multiple V and J gene segments that can potentially participate in secondary nested rearrangements. In sturgeon, there are multiple V and at least seven J gene segments (26). In teleosts and in cartilaginous fishes, however, there are multiple clusters, but all those described contain only one functional J gene segment next to the C exon. This means that the

replacement of any unwanted fish VJ combination entails at least two secondary rearrangement events, one that brings about its elimination and one at another available cluster.

As can be seen from the maps in Fig. 1, most type 1/3 C exons are followed downstream by an inverted V gene segment. Although these 3′ V gene segments can participate in generating the repertoire, as described in *Results*, their frequent presence may indicate an additional and important function—the provision of downstream RSS for receptor editing. As shown in Fig. 5*B*, the inversion rearrangement with any upstream V gene generates a signal joint (RSS-12/RSS-23). If the initial VJ is unwanted, secondary recombination in this already activated region could involve the RSS-23 of the signal joint and the RSS-12 of a downstream V gene segment (Fig. 5*B*). The ability of the RSS components of the signal joint to participate in recombination

events has been documented (40, 41). The secondary rearrangement illustrated in Fig. 5B would cause deletion of the VJ and C exon in the intervening DNA at the type 1 and type 3 loci.

If the downstream elements have evolved to be used for receptor editing, why have they not become V pseudogenes or isolated RS/kde? Recombination to the downstream RSS is theoretically as likely as to the upstream ones, all of which involve inversion rearrangement (Fig. 2, C and D). Then, there is a probability of one in three that the type 3 J will recombine to a downstream RSS, uselessly, if the RSS is not flanked by a functional V sequence. Thus, if the first (and only potentially functional) rearrangement occurs to a downstream RSS, then this event can still generate a productive VJ. This kind of rearrangement (Fig. 2D) cannot be readily eliminated in the same manner as the others (Fig. 5B), but a casual search of the type 3 locus, for instance, reveals heptamer sequences located downstream of V3h, and one of these heptamers recombining with non-rearranged upstream V gene segments could enable C exon excision, in the manner of RS/kde and V in the mouse (35).

In the zebrafish type 2 locus, all rearrangements involve deletion, and secondary recombination of available upstream V genes with the J2c and J2d gene segments can effect deletion of VJ rearrangements transcribed with C2a and C2b, respectively (Fig. 1).

Is there potential for secondary rearrangements and receptor editing in species other than zebrafish? The various reported teleost L chain clusters are tightly linked, and the distance between the C exon of one cluster and the downstream inverted V of the next cluster is <5 kb in trout L1, <5 kb in cod L1, <6 kb in catfish G, and <10 kb in catfish F. Although it is usually remarked in reviews that rearrangement events are limited to within a cluster, this really only applies to observations at elasmobranch Ig clusters, which are not close together. Intracluster IgM H chain recombination was demonstrated in three horn shark rearrangements of the λ1113 IgH locus (42) and 16 nurse shark rearrangements of the G2-V1 and G2-V2 IgH loci (43). Further characterization of 56 nurse shark IgM H chain cDNA revealed that all were rearrangements of V gene segments from groups 1, 2, 4, and 5 (44) to their respective group-specific J gene segments (V. Lee and E. Hsu, unpublished results). The rearrangement patterns for the closely packed IgL in teleost fishes are currently unknown.

The teleost V-J-C clusters might not be individually distinct—several might be controlled by one set of cis-regulatory elements, as suggested in cod. Bengtén et al. (29) discovered that the enhancers in cod IgL are located downstream of the C exon, and that of six IgL clusters tested for enhancer activity, two 3′ regions contained strong activity and one weak when transfected with a reporter into fish B cell lines. Although the presence of multiple enhancers was established, it also appeared that not every cluster possessed its own enhancer. Fig. 5C is modeled after cod IgL clone CgL10 (29), and we suggest as an example that the unwanted VJ at one cluster can be deleted by secondary rearrangement between the RSS in the signal joint and a downstream V gene segment, part of another accessible cluster under the control of the shared enhancer. In this case, if the enhancer is deleted in the process, rearrangement would proceed elsewhere, at a group of clusters regulated by a different enhancer.

Is receptor editing possible in cartilaginous fishes? In the nurse shark, the L chain clusters can be rather distantly spaced (45); unlike in teleost fishes, we have not found more than one L chain cluster per bacteriophage clone in our nurse shark genomic DNA libraries (our unpublished observations). In the course of characterizing one nurse shark L chain cluster, NS5-2 (13), which is composed of one V, J, and C gene each, we found an isolated RSS sequence (nonamer, 23-bp spacer, heptamer) in the 4-kb J-C intron and a heptamer located immediately 3′ of the C exon (E. Hsu, unpublished observations). This is the same combination of elements that enables C exon deletion in the mouse κ L chain system, and we suggest that inactivation of an unwanted shark L chain VJ could occur this way.

In mammals, receptor editing occurs at high levels in normal B cells, but in cold-blooded vertebrates, it is not yet clear whether promotion of tolerance occurs this way or by clonal deletion/anergy. There are many IgL loci in fishes, and we propose that they serve not only to provide BCR diversity but also a reserve of genes/RSS sequences to be used for the purpose of replacing rearrangements encoding nonviable or self-reactive specificities. The maintenance of self-tolerance is an innate part and corollary of a process that generates a vast repertoire, including anti-self specificities that must be removed, preferably with less cell wastage, which is more costly in cold-blooded vertebrates (46). Diversification and receptor editing are both achieved by rearrangement, and as such, the potential to enact both processes will be selected for in the organizational structure of Ag receptor genes during vertebrate evolution.

Disclosures

The authors have no financial conflict of interest.

References

1. Riblet, R. 2004. Immunoglobulin heavy chain genes of mouse. In *Molecular Biology of B Cells*. T. Honjo, F.W. Alt, and M. S. Neuberger, eds. Elsevier Academic Press, Amsterdam, pp. 19–26.
2. Fugmann, S. D., A. I. Lee, P. E. Shockett, I. J. Villey, and D. G. Schatz. 2000. The RAG proteins and V(D)J recombination: complexes, ends, and transposition. *Annu. Rev. Immunol.* 18: 495–527.
3. Zachau, H. G. 2004. Immunoglobulin κ genes in human and mouse. In *Molecular Biology of B Cells*. T. Honjo, F.W. Alt, and M. S. Neuberger, eds. Elsevier Academic Press, Amsterdam, pp. 27–36.
4. Gay, D., T. Saunders, S. Camper, and M. Weigert. 1993. Receptor editing: an approach by autoreactive B cells to escape tolerance. *J. Exp. Med.* 177: 999–1008.
5. Radic, M. Z., J. Erikson, S. Litwin, and M. Weigert. 1993. B lymphocytes may escape tolerance by revising their antigen receptors. *J. Exp. Med.* 177: 1165–1173.
6. Tiegs, S. L., D. M. Russell, and D. Nemazee. 1993. Receptor editing in self-reactive bone marrow B cells. *J. Exp. Med.* 177: 1009–1020.
7. Nemazee, D., and M. Weigert. 2000. Revising B cell receptors. *J. Exp. Med.* 191: 1813–1817.
8. Retter, M. W., and D. Nemazee. 1998. Receptor editing occurs frequently during normal B cell development. *J. Exp. Med.* 188: 1231–1238.
9. Casellas, R., T. Y. Shih, M. Kleinewietfeld, J. Rakonjac, D. Nemazee, K. Rajewsky, and M. C. Nussenzweig. 2001. Contribution of receptor editing to the antibody repertoire. *Science* 291: 1541–1544.
10. Gorman, J. R., and F. W. Alt. 1998. Regulation of immunoglobulin light chain isotype expression. *Adv. Immunol.* 69: 113–181.
11. Klein, F., N. Feldhahn, J. L. Mooster, M. Sprangers, W. K. Hofmann, P. Wernet, M. Wartenberg, and M. Muschen. 2005. Tracing the pre-B to immature B cell transition in human leukemia cells reveals a coordinated sequence of primary and secondary IGK gene rearrangement, IGK deletion, and IGL gene rearrangement. *J. Immunol.* 174: 367–375.
12. Litman, G. W., M. K. Anderson, and J. P. Rast. 1999. Evolution of antigen binding receptors. *Annu. Rev. Immunol.* 17: 109–147.
13. Fleurant, M., L. Changchien, C.-T. Chen, M. F. Flajnik, and E. Hsu. 2004. Shark immunoglobulin light chain junctions are as diverse as in heavy chains. *J. Immunol.* 173: 5574–5582.
14. Ghaffari, S. H., and C. J. Lobb. 1993. Structure and genomic organization of immunoglobulin light chain in the channel catfish. *J. Immunol.* 151: 6900–6912.
15. Ghaffari, S. H., and C. J. Lobb. 1997. Structure and genomic organization of a second class of immunoglobulin light chain in the channel catfish. *J. Immunol.* 159: 250–258.
16. Plough, F. H., C. M. Janis, and J. B. Heiser. 2002. *Vertebrate Life*, 6th Ed. Prentice Hall, Upper Saddle River, NJ.
17. Haire, R. N., J. P. Rast, R. T. Litman, and G. W. Litman. 2000. Characterization of three isotypes of immunoglobulin light chains and T-cell antigen receptor α in zebrafish. *Immunogentics* 51: 915–923.

18. Wilson, M., E. Hsu, A. Marcuz, M. Courtet, L. Du Pasquier, and C. Steinberg. 1992. What limits affinity maturation of antibodies in *Xenopus*—the rate of somatic mutation or the ability to select mutants? *EMBO J.* 11: 4337–4347.

19. Lee, A., S. Desravines, and E. Hsu. 1993. IgH diversity in an individual with only one million B lymphocytes. *Dev. Immunol.* 3: 211–222.

20. Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 24: 4876–4882.

21. Felsenstein, J. 1989. PHYLIP—Phylogeny Inference Package (version 3.2). *Cladistics* 5: 164–166.

22. Page, R. D. 1996. Treeview: an application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* 12: 357–358.

23. Solem, S. T., and T. Ø. Jørgensen. 2002. Characterization of immunoglobulin light chain cDNAs of the Atlantic salmon, *Salmo salar* L.; evidence for three IgL isotypes. *Dev. Comp. Immunol.* 26: 635–647.

24. Hsu, E., N. Pulham, L. Rumfelt, and M. F. Flajnik. 2006. The plasticity of immunoglobulin gene systems in evolution. *Immunol. Rev.* 210: 8–26.

25. Pilström, L., G. W. Warr, and S. Strömberg. 2005. Why is the antibody response of Atlantic cod so poor? The search for a genetic explanation. *Fisheries Science* 71: 961–971.

26. Lundqvist, M., E. Bengtén, S. Strömberg, and L. Pilström. 1996. Ig light chain gene in the Siberian sturgeon. *J. Immunol.* 157: 2031–2038.

27. Daggfeldt, A., E. Bengtén, and L. Pilström. 1993. A cluster type organization of the loci of the immunoglobulin light chain in Atlantic cod (*Gadus morhua* L.) and rainbow trout (*Oncorhynchus mykiss* Walbaum) indicated by nucleotide sequences of cDNA and hybridization analysis. *Immunogenetics* 38: 199–209.

28. Timmusk, S., S. Partula, and L. Pilstrom. 2000. Different genomic organization and expression of immunoglobulin light-chain isotypes in the rainbow trout. *Immunogenetics* 51: 905–914.

29. Bengtén, E., S. Strömberg, A. Daggfeldt, B. G. Magor, and L. Pilström. 2000. Transcriptional enhancers of immunoglobulin light chain genes in Atlantic cod (*Gadus morhua*). *Immunogenetics* 51: 647–658.

30. Fink, S. V., and W. L. Fink. 1997. Interrelationships of the Ostariophysi. In: *Interrelationships of Fishes.* M. Stiassney, L. Parenti, and D. Johnson, eds. Academic Press, New York, pp. 209–249.

31. Hsu, E., and L. A. Steiner. 1992. Primary structure of immunoglobulins through evolution. *Curr. Biol.* 2: 422–431.

32. Ota, T., T. Sitnikova, and M. Nei. 2000. Evolution of vertebrate immunoglobulin variable gene segments. *Curr. Top. Microbiol. Immunol.* 248: 221–245.

33. Greenberg, A. S., L. Steiner, M. Kasahara, and M. F. Flajnik. 1993. Isolation of a shark immunoglobulin light chain cDNA clone encoding a protein resembling mammalian κ light chains: implication for the evolution of light chains. *Proc. Natl. Acad. Sci. USA* 90: 10603–10607.

34. Durdik, J., M. W. Moore, and E. Selsing. 1984. Novel κ light-chain gene rearrangements in mouse λ light chain producing B lymphocytes. *Nature* 307: 749–752.

35. Moore, M. W., J. Dirdik, D. M. Persiani, and E. Selsing. 1985. Deletions of κ chain constant region genes in mouse λ-producing B cells involve intrachromosomal DNA recombination similar to V-J joining. *Proc. Natl. Acad. Sci. USA* 82: 6211–6215.

36. Shimiziu, T., T. Iwasato, and H. Yamagishi. 1991. Deletions of immunoglobulin Cκ region characterized by the circular excision products in mouse splenocytes. *J. Exp. Med.* 173: 1065–1072.

37. Nadel, B., P. A. Cazenave, and P. Sanchez. 1990. Murine λ gene rearrangements: the stochastic model prevails over the ordered model. *EMBO J.* 9: 435–440.

38. Wardemann, H., J. Hammersen, and M. C. Nussenzweig. 2004. Human autoantibody silencing by immunoglobulin light chains. *J. Exp. Med.* 200: 191–199.

39. Bernier, G. M., and J. J. Cebra. 1965. Frequency distribution of α, γ, κ and λ polypeptide chains in human lymphoid tissues. *J. Immunol.* 95: 246–253.

40. Lewis, S., A. Gifford, and D. Baltimore. 1985. DNA elements are asymmetrically joined during the site-specific recombination of κ immunoglobulin genes. *Science* 228: 677–685.

41. Meier, J. T., and S. M. Lewis. 1993. P nucleotides in V(D)J recombination: a fine-structure analysis. *Mol. Cell. Biol.* 13: 1078–1092.

42. Hinds-Frey, K. R., H. Nishikata, R. T. Litman, and G. W. Litman. 1993. Somatic variation precedes extensive diversification of germline sequences and combinatorial joining in the evolution of immunoglobulin heavy chain diversity. *J. Exp. Med.* 178: 815–824.

43. Malacek, K., J. Brandman, J. E. Brodsky, Y. Ohta, M. F. Flajnik, and E. Hsu. 2005. Somatic hypermutation and junctional diversification at immunoglobulin heavy chain loci in the nurse shark. *J. Immunol.* 175: 8105–8115.

44. Rumfelt, L. L., R. L. Lohr, H. Dooley, and M. F. Faljnik. 2004. Diversity and repertoire of IgW and IgM VH families in the newborn nurse shark. *BMC Immunol.* 5: 8.

45. Lee, S. S., D. S. Tranchina, Y. Ohta, M. F. Flajnik, and E. Hsu. 2002. Hypermutation in shark immunoglobulin light chain genes results in contiguous substitutions. *Immunity* 16: 571–582.

46. Du Pasquier, L. 1982. Antibody diversity in lower vertebrates—why is it so restricted? *Nature* 296: 311–313.

47. Saha, N. R., H. Suetake, and Y. Suzuki. 2004. Characterization and expression of the immunoglobulin light chain in the fugu: evidence of the solitary type. *Immunogenetics* 56: 47–55.

48. Carroll, R. L. 1998. *Vertebrate Paleontology and Evolution.* W. H. Freeman, New York.

49. Ishikawa, J., E. Imai, T. Moritomo, M. Nakao, T. Yano, and M. Tomana. 2003. Characterization of a fourth immunoglobulin light chain isotype in the common carp. *Fish Shellfish Immunol.* 16: 369–379.

Figure 1a. Type 2 L chain cluster from fugu. M007644 and M006921 (pseudogene). Sequence retrieved from fugu database is analyzed as to coding regions (underlined with derived amino acid sequence) for L chain: leader, V gene segment, J gene segment, C exon are all in the same transcriptional direction. Splice signals, RSS, and polyadenylation signals are bolded. Stop is indicated by asterisk.

```
M007644

CAGGTTGATAGTCAGCACACAGTGGCAACATGCTGGAGAGCCTCTGCACTCTCATCACTGCTCTAACATGTGAGGAGGCTGACTTACTGCAGCTCTGAGCTC
                                 M  L  E  S  L  C  T  L  I  T  A  L  T  C
ATGCTGGATTCATCAGTCTGTGTCTTTCTCTTGACTTCATGTCCTCTTGTTGTTTCCAGGTGTGAGTGCTGTGACGGTGCTGACACAGAAGCCCACTGTTGT
                                                                V  S  A  V  T  V  L  T  Q  K  P  T  V  V
GTCTCTGAGCAGAGGAGAGTCAGTCACCATGGACTGTAACCTGGGAACTAGTACTAACTATGCTGCCTGGTATAAACAGGTTCCAGGAGGAGTTCCTCAGTA
  S  L  S  R  G  E  S  V  T  M  D  C  N  L  G  T  S  T  N  Y  A  A  W  Y  K  Q  V  P  G  G  V  P  Q  Y
TGTGCTGAGCTGGCGTAGCAGCTGGAGTTCTGTCAAATATGGTTCTGGTTTCTCATCTCCAAGATTCACTTCTACTCATCAGTCAACATCAGATTATCGTTT
  V  L  S  W  R  S  S  W  S  S  V  K  Y  G  S  G  F  S  S  P  R  F  T  S  T  H  Q  S  T  S  D  Y  R  L
GATGATCAATAATCTGGAGGAGGGAGACTCAGCAGTCTATTACTGTGGAACATGGGACAGCTCTGCTAAAGAGAACGTATCACAGTGATTCAGACTGTGACA
  M  I  N  N  L  E  E  G  D  S  A  V  Y  Y  C  G  T  W  D  S  S  A  K  E  N  V
AAAACCTCCTCACTCAATACTTCTGCTTTTTAGTGTCTGATACATCAGCTCATGTTCAGTCTCACTAAATTACAAATCAACTCATGTTCACACTCAGTTTCC
AGAGAAACTTTCAGATTGTTGGAAACTTTCTGCCGTTTCATCACTGAAAATAACAGNAACCAAAACCAACTCCCTCACTCATCCTTCATCCTCCATCATCAC
CATTAAAATACACCTTTAGATCATCAAAACGACTTTATGGGGAATATGATGAAGAGTTTGTTGATTAAAGTTCTAACAGAGAATATGTGGTGATGATGATGA
TGATGATGATAATGATGATAGTGGTGGTGGTGGTGATGATGATGATGATGATAATGATGTGATGTTGATGATGATGATGATGATGGTGATGATGATGGTGAT
GATGATGGTGATGGTGATGATGATGATGGTGATGATGATGGTGGTGATTGTGATGATGATGATGATGATGTGGTGATGATGATGGTGGTGATGATGATGAT
GATGATGGTGGTGTCACCAGTGAAGGCTGTTCTCATGTCTCAGTGTCTCCACTCATCAGTGACTCTCAGCAGGTTTTTGTACAGCTGTGTGTACAAACTGAA
TCACTGTGGTATTCCGACCAGGAACCAAGCTGATTGTCACCAGTAAGTACTTTTACACTCAGCAGAAACAACAAAGATTTGATGCTTTTTCTGTCAGCTAAA
         V  F  R  P  G  T  K  L  I  V  T  S
GGAAACATCTCCAAAGATTACATCATTTCAATCAGTTTCTCACCAAATATTGTTGAATTTTCTTCGTTTTATTTTAATAACTGCAGAACAGATTTTCAAACT
AAATGTATTTTAAATTGACTTTGGTTTCTTGGCTCGTGTTTGATGGAAGTGAGTGAAAAAGGTTTTGGAAAATTCATCCAACTTCATTCAGCATAAATTTTA
AATATAAACTTTATATTTTTGTATCATTTGCTGGTCATAATTTTACTTTGGTCAAATATTTAATTGATGCTGTTTATGTTCAAAATGACTCTGAAATGGTG
TAAAATATCTGGTTTAGTGAGATTTTCCTCATATTGTTGCTGAAGATTGAAGGTTTTTATTTGCTCAGTGTGTTTGTGATGTTTTCAACCAAACTCCTAAATT
CATTAGTTCCTGTTGAACTTTGTGTGAGTGTGTGAATTTCTTCTTCATGTATTTTCAGGCTCCAGCCTCCCTCCTCCTGTCCTGACAGTCTTCCCTCCGTCC
                                                        S  S  L  P  P  P  V  L  T  V  F  P  P  S
AGAGCTGAGCTCCAGTCCAACAAAGCCACTCTGGTCTGTCTGTCCAGACTCTCTGCACCTTTTGCAGAGGTGAGCTGGTTGCTTGGTGACACTTCAGTGAGC
  R  A  E  L  Q  S  N  K  A  T  L  V  C  L  S  R  L  S  A  P  F  A  E  V  S  W  L  L  G  D  T  S  V  S
AGCGGGATCTCCACCAGCACTCCTGTCCAACAAGCAGACCAGACTTTCCAAATCAGCAGCCATCTGTCCATCCTGACGTCAGACTGGGACGCGCAGAAGGTT
  S  G  I  S  T  S  T  P  V  Q  Q  A  D  Q  T  F  Q  I  S  S  H  L  S  I  L  T  S  D  W  D  A  Q  K  V
TACACCTGTAAAGTGTCTGTGGGCTCCCAGACTGCAGAGAAAAGCATCAAGAAGTCCGAGTGTGAAGAATAGTGGAGGAGCAGCAGCACCATTTCACATCTG
  Y  T  C  K  V  S  V  G  S  Q  T  A  E  K  S  I  K  K  S  E  C  E  E  *
CTTCTTTAGTCTTTGTGCTGCATCAGCAGCTTCACATGTCAGAGGAGACCAGTCTGCATGCTTGGAATCAATGTTGTCACGCTGAGCTGCTGCTCTTGGAGC
AGCTTTGCAGCTGCTCTGTTTTTACAATCTTTCAATAAAAACTCATCAAGTTGAAGAAAAAGAACTCTTTCTGTTTATTTGGACTCTTCCAAATGTTGTTTT
CAACAATAATCAATAAAAGTTCAGTGATTTTAGTGTCTTGTCTCAGGACAACAGAAAAATCAGATCTATAATAAAGTGAAACTGAGTTTTCATGCTTCATCT
GTGAAGCGTGGCCTCTGGACGCCTCAGTACTTTCCCAGGATGCACCTGCAGGCCTCCTCACTCATTTATGAGCTGCATCTAAATGGAGCTCAAATGTCAGCTG
TCTAAAAAGAGAGGGAACCATTTTATCATTTTAAAAAAATATTATATTTGAAGAACAATGATTTTGTTGTTTAATATGTGGAGGATATTTATGTGTTTAAATG
TCACAACAGGATCATCAAGGGTGAAAGACTCGATGTTCCAGATAATTCATGATGTGATCCACGATGATGATACG
```

M007644

TTCA*ATGTAAA*TGAAACGTTGGAGCTCCAATCTGCTCTTTTCCACATCACACACAGGAGAAAAGCTGCAGATTGAAGAGAAGCTCCTCCTCCTGTCAGTC
CCTCTCAGCTCCAGTGGGCTATTAAATTGCTCCTCCAGCACCTCCTCTCCTCATCCTCCAAAGCCTGAATCTGTCAGCAGTAAAGTCACTTTCCAGGTTG
ATAGTCAGCACACAGTGGCAAC<u>ATGCTGGAGAGCCTCTGCACTCTCATCACTGCTCTAACAT</u>**GT**GAGGAGGCTGACTTACTGCAGCTCTGAGCTCATGCT
     M L E S L C T L I T A L T C
GGATTCATCAGTCTGTGTCTTTCTCTTGACTTCATGTCCTCTTGTTGTTTCC**AG**<u>GTGTGAGTGCTGTGACGGTGCTGACACAGAAGCCCACTGTTGTGTC
                                 V S A V T V L T Q K P T V V S
TCTGAGCAGAGGAGAGTCAGTCACCATGGACTGTAACCTGGGAACTAGTACTAACGATGCTGGCGCTGGTATAAACAGGTTCCAGGAGGAGTTCCTCAGT
 L S R G E S V T M D C N L G T S T N D A G A G I N R F Q E E F L S
ATGTGCTGAGCTGGTATCAGCGGCTGGAGTTCTGTTGATTATGGTTCTGGTTTCTCATCTCCAAGATTCACTTCTACTCATCAGTCAACATCAGATTATC
 M C * A G I S G W S S S V D Y G S G F S S P R F T S T H Q S T S D Y R
GTTTGATGATCAATAATCTGGAGGAGGGAGACTCAGCAGTCTATTACTGTCAAACATGGGACGGCTCTGCTAAAGAGGTCGTAT</u>**CACAGTG**ATTCAGACT
  L M I N N L E E G D S A V Y Y C Q T W D G S A K E V V
GTG**ACAAAAACC**TCCTCACTCAATACTTCTGCTTTTTAGTGTCTGATACATCAGCTCATGTTCAGTCTCACTAAATTACAAATCAACTCATGTTCACACT
CAGTTTCTCAGAGAAACTTTCAGATTGTTGGAAACTTTCTGCTGTTTCATCACTGAAAATAACAAGAACCAAAACCAACTCCCTCACTCATCCTTCATCC
TCCATCATCACCATTAAAATACACCTTTAAGTCATCAAAACGACTTTATGGGGAATATGAAGAGTTTGTTGATTAAAGTTCTAACAGAGAGTCTGTTGTG
ATGATGATGATGATGATGATAATGATGATGATGGTGTTGATAATATTGATGATGATGATGGTGATGATGATGATGATGGTGTTGATAATATTGATGATGA
TGGTGATGATGATGATGGTGATGATGATGATAACAATGATGATGGTGATGGTGATGGTGATGATGATGATGATGATGATGGTGATGGTGATGGTGAT
GATGATGGTGATGGTGATGGTGATGATGGTGTCACCAGTGAAGGCTGTTCTCATGTCTCAGTGTCTCCACTCATCAGTGACTCTCAGCA**GGTTTTTGT**AC
AGCTGTGTGTACAAACTGAAT**CACTGTG**GTATTCGGACCAGGAACCAAGCTGATTGTCACCA**GT**AAGTACTTTTACACTCAGCAGAAACAACAAAGATTT
                        V F G P G T K L I V T S
GATGCTTTTGATACAAACTAGAAGCTTTTTCATAGAAATGTGGAGTGTGGAAATGATTTGAAATATTCTTCTGTTAGTCTTTGACCTGCTTTGATACTTT
ATCTGTCAGCTAAAGGAAACATCTCCAAAGATTACATCATTTCAATCAGTTTCTCACCAAATATTGTTGAGTTTTCTTCATTTTCACCTGTTTTTAATAA
CTGCAGAACAGATTTTCTACCTAAACGTATTTTAAATTGACTTTGGTTTCTTGGCTCGTGTTTGATGGAAGTGAGTGAAAAAGGTTTTGTAAAATTCACC
CAAATTAATTCAGTTGCTGAAATTCTCATTTTAATTATGTAACTTTATATTTTTGTATCATTTGCTGGTTATAATTTTACTTTGGTCAAATATTTAATTG
ATGCTGTTTATATTCAAAATGACTCTGAAATGGTGTAAAATATCTGGTTTAGTGAGATTTTCCTCATATTGTTGCTGAAGATTGAAGGTTTTATTTGCTC
AGTGTGTTTGTGATGTTTTCAACCAAACTCCTAAATTCATTAGTTCCTGTTGAACTTTGTGTGAGTGTGTGAATTTCTTCTTCATGTATTTTC**AG**<u>GCTCC
                                                                                      S
AGCCTCCCTCCTCCTGTCCTGACAGTCTTCCCTCCGTCCAGAGCTGAGCTCCAGTCCAACAAAGCCACTCTGGTCTGTCTGTCCAGACTCTCTGCACCTT
 S L P P P V L T V F P P S R A E L Q S N K A T L V C L S R L S A P F
TTGCAGAGGTGAGCTGGTTGCTTGGTGACACTTCAGTGAGCAGCGGGATCTCCACCAGCACTCCTGTCCAACAAGCAGACCAGACTTTCCAAATCAGCAG
  A E V S W L L G D T S V S S G I S T S T P V Q Q A D Q T F Q I S S
CCATCTGTCCATCCTGACGTCAGACTGGGACGCGCAGAAGGTTTACACCTGTAAAGTGTCTGTGGGCTCCCAGACTGCAGAGAAAAGCATCAAGAAGTCC
  H L S I L T S D W D A Q K V Y T C K V S V G S Q T A E K S I K S
GAGTGTGAAGAA</u>TAGTGGAGGAGCAGCAGCACCATTTCACATCTGCTTCTTAGTCTTTGTGCTGCATCAGCAGCTTCACATGTCAGAGGAGACCAGTCT
 E C E E *
GCATGCTTGGAATCAATGTTGTCACGCTGAGCTGCTGCTCTTGGAGCAGCTTTGCAGCTGCTCTGTTTTTACAATCTTTC**AATAAA**AACTCATCAAGTTG
AAGTAAAAGAACACTTTCTGTTTATTTGGACTCTTCCAAATGTTGTTTTCAACAATAATC**AATAAA**AGTTCAGTGATTTTAGTGTCTTGTCTCAGGACAA
CAGAAAAATCAGATCTAT**AATAAA**GTGAAACTGAGTTTTCATGCTTCATCTGTGAAGCGTGGCCTCTGGACGCCTCAGTACTTTCCCAGGATGCACCTGC
AGGCCTCCTCACTCATTTATAGCTGCATCTAAATGGAGCTCAAATGTCAG

Figure 1b. Sequences of functional V, J, and C zebrafish light chain genes. The sequences of functional genes depicted in Figure 1 of text are shown below along with their positions on the contigs. The order is as follows: Type 1 V, J, C; Type 2 V, J, C; Type 3 V, J, C. For V and J the coding sequences are underlined; RS sequences are bolded. The derived amino acid sequences are only shown for C. C1k and J1k are from contig NW_644842 and are not depicted in Figure 1.

TYPE 1

>V1a (155,306–155,819)
ATGACCTTCATCATCATCTTCATCGGGACTCTCACAGTGTTTGCTCGAGGTTTGTCCAGCGCTGCTTTCATATTCATTCATTTCTTCAAGTGGATTGTCGATATATTTAGAGCTGATTTTCTTCTTGTTGTGTT
TCAGAGTGTGGAGGACAGATCAGCGTCACTCAGAGTCCCTCAATGACAGCAGAACCAGACAAACCTGTCACCATCACCTGCAGAACCAGTAGAGATGTGCAGAGATGGAGTGATGGAGATGAAGGTTTAGCCTG
GTATTTGCAGAAACCTGGAGAAGCTCCTAAACTCCTGATTCGTTATGTGAAAAAGCTTCTGTCAGGAACTGCATCTAGATTCAGTGGCAGTGGATCTAACAGTGATTTCACTCTGAGCATCAGTGGAGTCCAGA
CTGAAGATGCTGGACATTATTACTGTCAGAGTTTACACTGCAGAACAGATTCGAACAATAACTGTGTGGGTGATGTTGTGTTCA**CACAGTG**ATGAAAAGTGGA**ACAAAAACC**

>V1b (150,597–151,092)
ATGACCTTCAGCAGCATATTGATCTGGACACTGGCAGCATTATGTGGAGGTGTGTGTCCCTCAAGATTATGTGTGTTATTGACTTGTGTTTCTGTGTTGTGATCTTATATATGATTATACAAATGATTTCTAAT
GTTTCTCCTCTGGCAGAATGTATTGGTCAGGTGACAGTGACTCAAACTCCCTCAGTGCTGCTCTCTGAACCTGGACGATCTCTCACTGTCCACTGCAGATTTAGCAGTAATCCAGACTGCTGTTATCACGTTAG
CTGGTACTTACAGAAACCTGGAGAAGCTCCTAAACTCCTGATTTATTATACAGATGTCCTCCAGCCAGGAACTGCATCTAGATTCAGTGGCAGTGGATCTAACAGTGATTTCACTCTGAGCATCAGTGGAGTCC
AGACTGAAGATGCTGGACATTATTACTGTCAGAGCTACCACTACATCAACAGTAAATATGTGTTCA**CACAGTG**ATAAAGAGTGAT**ACAAAAACC**

>V1c (149,385–149,941)
ATGAAGATCATCAGCGTCTGTTTCTGGACACTCGCTCTTTCTGTTCAAGGTTCATTTACACTGTTATTGACTCTAAATGAGGTTTAAATGCAGCTGTTTGACAGTAAAGGTGTTTTTGTTACGGACACTTCAAC
TATATTAAGTGTTGATGAATATGAGCTGAAATAATGTTCTTCTCAAATGTCCTTCACAGAAATGTCAGGACAGCTTTCAGTGACTCAAACTCTTTCAGCAGCAGCTCCAGCAGGAGAAACTGTGACTATAAACT
GTCAGACTAGCAGAGGAATTGGAGATCGAACACAGGGATGTGGCTCTTGTCTCTCCTGGTATCTGCAGAAACCTGGAGAAGCTCCTAAACTCCTCATTTCCTATATAAGATCACTGCAGTCAGGAACTCCAGCC
AGATTCAGTGGCAGTGGAGAATATTATGGCACTGATTTCACTTTGACCATCAGTGGAGTCCAGACTGAAGATTCTGGACATTATTACTGTATGGGTGTACACTTGATCAACAGTAAATATGTGTTGA**CACAGTG**
ATAAAGAGTGAT**ACAAAAACC**

>V1d (147,360–147,872)
ATGACCTTCATCATGATCTTCATCTGGACTCTCACAGTGTTTGCTCGAGGTTTGTCCAGCGCTGCTTTCATATTCATTCATTCTTCAAGTGGATTGTCGATATATTTAGAGCTGATTTTCTTCTTGTTGTGTTT
CAGAGTGTGGAGGACAGATCAGCGTCACTCAGAGTCCCTCAATGACAGCAGAACCAGACCAAACCGTCACAATCACCTGCAGAACCAGTAGAGATGTGTACAGATATGATGATGGAGATGATGCTTTTTTTCTGG
TATTTACAGAAACTTGGAGAAGCTCCTAAACTTCTGGTTCGTTATGCAAAAGTCTGGAATCAGAAACTCCGTCTAGATTCAGTGGCAGAGGATCTAACAGTGATTTCACTCTGAGCATCAATGGAGTCCAGAC
TGAAGATGCTGGACATTATTACTGTCAGAGTTTCCACTGCAGAACAGATTCGAACAATAACTGTGTGGGTGATGTTGTGTTGA**CACAGTG**ATAAAGAGTGAT**ACAAAAACC**

>V1e (891,778–892,334)
ATGAAGATCATCAGCGTCTGTTTCTGGACACTCGCTCTTTCTGTTCAAGGTTCATTTACTCTGTTATTGACTCTAAATGAGGTTTAAATGCAGATGTTTGACAGTAAAGGTGTTTTTGTTACGGACACTTCAAC
TATATTAAGTGTTGATGAATATGAGCTGAAATAATGTTCTTCTCAAATGTCCTTCACAGAAATGTCAGGACAGCTTTCAGTGACTCAAACTCCTTCAGCAGCAGCTCCAGCAGGAGAAACTGTGACTATAAACT
GTCAGACTAGCAGAGGAATTGGAGATCGGTCACAGGGATGTGGCTCTTGTCTCTCCTGGTATCTGCAGAAACCTGGAGAAGCTCCTAAACTCCTCATTTATTATATAAGATCACTGCAGTCAGGAACTCCAGCC
AGATTCAGTGGCAGTGGAGCAGATTATGGCACAGATTTCACTTTGACCATCAGTGGAGTCCAGACTGAAGATGCTGGACATTATTACTGTATGAGTATGCACTGGATCAACAGTAACTGGGTGTTCA**CACAGTG**
ATAAAGAGTGAT**ACAAAAACC**

>V1f (889,130–889,651)
ATGACCTTCATCATCGTCTTCATCTGGACTCTTTCATACAGCTTCATCCTGCGAGGTGACCAGTTTTATATCTGAATATATTATAGCTGTAGTGTCATCTGAACAAGCTTAAAGATGGACTGATAAATGATGATC
TTTGTAGACTAATAGTGTGAATGATTTGTTTCTCAGGATGTGCAGCTCAGATCACAGTGACTCAAAGTCCAGTGGAACCTGTAAAACCTGGAGAAAATGTGCAGATGAGCTGCAAACTTAGCGAACCTTCAAAC
TGCAGCCCAAACCCCTGTGTGTCCTGGTACTTACAGAAACCTGGAGAAGCTCCTAAACTCCTGATTTATAAAACAGACATCCGTGAGTCAGGAACTGCATCTAGATTCAGTGGCAGTGGATATGATACAGATTT
CACTCTGAGCATCAGTGGAGTCCAGACTGAAGATGCTGAACACTATTACTGTCAGAGTGCACACTTCATCAGCAATATTGGGTGGGTGTTGA**CACAGTG**ATAAAGAGTGGT**ACAAAAACC**

>V1g (886,794-887,286)
ATGACCTTCATCATCATCTTCATCTGGACTCTCACAGCGTTTGCTCGAGGTTTGTCCAGCTCTGCTTTCATGTTCACTCATTTCTTCAAGTGGATTGTCGATATATTTACAGCTGATTTTCTTCTTGTTGTGTT
TCAGAGTGTGGAGGACAGATCAGCGTCACTCAGAGTCCCTCAATCACAGCAGAACCAGACAAACCCGTCACCATCACCTGCAGAACCAGCACAAATGTGTACAGTGACAGTACATATGGGCCACGTTTAGCCTG
GTACTTACAGAAACCTGGAGAAGCTCCTAAACTCCTGATTCATTTAACAAACCGTCTCCAGTCAGGAACTGCATCTAGATTCAGTGGCAGTGGATCTAACAGAGATTTCACTCTGAGCATCAGTGGAGTCCAGA
CTGAAGATGCTGGACATTATTACTGTCAGAGTTACCACTACATCAACAGTAAAGATGTGTTCA**CACAGTG**ATAAAGAGTGAT**ACAAAAACC**

>V1h (885,014-885,532)
ATGACCTTCATCATCATCTTCATCTGGACTCTTTCATTCAGCTTCATCCTGCGAGGTGACCAGTTTATATCTGAATATATTATAGCTGTAGAGTCATCTGAACAAGCTTAAAGATGGACTGATAAATGATGATC
TTTGTAGACTAATAGTGTGAATGATTTGTTTCTCAGGATGTGCAGCTCAGATCACAGTGACTCAAAGTCCAGTGGAACCTGTTAAACCTGGAGAAAATGTGCAGATGAACTGCAAACTCAGTCAACCTGCACCC
TGCAGCCCAGACTCATGTGTGTCCTGGTACTTACAGAAACCCGGAGAAGCTCCTAAACTCTTGATTTACTATCAAGCAGCCGTTATTCAGGAACTGCATCTAGATTCAGTGGCAGTGGATCTAATGTTGATTT
CACTCTGAGCATCAGTGGAGTCCAGACTGAAGATGCTGGACATTATTACTGTCAGACTTACCACTGGATCAACAGTAAAGATGTGTTGA**CACAGTG**ATAAAGAGTGGT**ACAAAAACC**

>V1i (883,708-884,320)
ATGCTGATCTTGCTGATCTGTGTTCAGTGTGAGCTGGAGGTTTTGAATGTGTGTCCTGCCCCTGACACTGGAGCTCAATATCAGTCTGACTGTGAATCAGCATCTCTGTCCTTCAGTTCTCTGAGCTTCAGTGA
AGATGAAGACGCTCGCTCTGATCATCTGCTCTCTGCTGCTGATCCTCCACGGTCAGACTCACTTTCACTGCATCTCACAGCTCAGACATTCAGGACTAATGCTGACTTTCTTCTTGTGTTTCAGAGTGTGGAGG
ACAGATCAGCGTCACTCAGAGTCCCTCAATCACAGCAGAACCAGGCAAACCCGTCACCATCACCTGCAGAACCAGCACAAATGTGTACAGGGATAGTGATGGAGATGACTACTTATTCTGGTATTTGCAGAAAC
CTGGAGAAGCTCCTAAACTCCTGATTCGTTATGTGAAAAAGCTTCAGTCAGGAACTGCATCTAGATTCAGTGGCAGTGGATCTAACAGAGATTTCACTCTGAGCATCAGTGGAGTCCAGACTGAAGATGCTGGA
CAATATTACTGTCAGAGTTACCACTACATCAACAGTAATAATGTGTTGA**CACAGTG**ATAAAGAGTGGT**ACAAAAACC**

>V1k (871,305-874,806)
ATGAGCTTCAGCAGCATATTGATCTGGACACTGGCAGCATTATGTGGAGGTGTGTGTCCCTCAACATCATGTGTGCTATTGACTTGTGTTTCGGTGTTGTGATCTTATATATGATTATACAAATGATTTCTAAT
GTTTCTCCTCTGGCAGAATGTATTGGTCAGGTGACAGTGACTCAAACTCCCTCAGTGCTGCTTTCTGAACCTGGACAATCCCTCACTGTCCACTGCAGATTCAGCAGAGATCCACCTTGCTGTTATGGCAGTGG
AACTTATCTTGTTAGCTGGTACTTACAGAGACCTGGAGAAGCTCCTAAACTCCTGATTTATTACACAACTGTTCTCCAGTCAGGAACTGCATCTAGATTCAGTGGCAGTGGATCTAAGACTGATTTCACTCTGA
GCATCAGTGGAGTCCAGACTGAAGATGCTGGACATTATTACTGTCAGAGTGAACACAGTGGTCCAGTGTTCA**CACAGTG**ATAAAGAGTGAT**ACAAAAACC**

>V1l (593,433-593,931)
ATGATTATCTTCATCTGGGCATTTTACTGCATCCAAGGTAATTCAAGAAAGGTTTATTTATTTGTCATGCACTAATATTGGTTTGACTTCAGCAGAGTTTTCTTTCTCTCACAGGCTGTTTGGGGCAGATTACT
ATTACTCAACCTAATGTTAAATTAGTTCAGCTCAGTCAAACAGTCTCTATTGACTGTAAGGTCAACACTGTAGTTTACAGGTATCCAACTGGTTCAATCTCTAAAGATTATTACATCTCCTGGTATCTTCAGAA
GCCAGAGGAGCCTTCTAAACTCCTTATGTATTTCACAACAATGAAACTTCCAAGTACTTCATCTAGAATGAGTGGAGGAGGTAGTTTTCACTCTGGAGGAAACGGTCTTGACTTCACTTTGACCATCAGTGAAG
CTCAGCTTGAAGATGCTGGAGTTTATTACTGTCAGAGTTACCATGAGATAAGTGGTAAACCATTGTTCA**CACAGTA**AAAAAGAGCTTA**ACAAAAACC**

>V1m (588,833-588,330)
ATGATTTATTTTATTCTACTGTTGGAGACACTTTTTCTGCTCACTCAAGGTTTGGATCATTTTTATTGAGATGACACATTTTTTAAAATATACATTTTTGCAGAAATGTTAAGCTTTTAGTTATATTTTTATGG
ATTCATTTCAATTTCATTTCTTGTTATTAATTTTTTGTTCTCTCTCAGTCTCTCATCAAGAGGTTCAGTCTCCTGAGTCTCTGTCTGTTAAAGCAGGGGATTCTGCCTCTATCAGCTGTACTGGGACCAGTGGA
GTTGGTTATGACATGAGCTGGTATTTGCAGAAACCTGGAAAAGCTCCTAAACTCCTGATCTATGGAGTAAGTTACCTTGAATCTGGAGTTTCAGATCGATTCAGTGGAAATGGATCTGAACCTGCTTTTACACT
GAACATACGTGGAGTCCAGCCTGAAGATGCAGGAGTTTATTACTGTATGGGGTACTATGGTGGTGGAAGGTGCA**CACAGTG**AAACAACATCAA**ACAAAAACC**

>V1o (164,429-164,927)
ATGATTATCTTCATCTGGGCATTTTACTGCATCCAAGGTAATTCAAGAAAGGTTTATTTATTTGTCATGCACTAATATTGGTTTGACTTCAGCAGAGTTTTCTTTCTCTCACAGGCTGTTTGGGGCAGATTACT
ATTACTCAACCTAATGTTAAATTAGTTCAGCTCAGTCAAACAGTCTCTATTGACTGTAAGGTCAACACTGTAGTTTACAGGTATCCAACTGGTTCAATCTCTAAAGATTATTACATCTCCTGGTATCTTCAGAA
GCCAGAGGAGCCTTCTAAACTCCTTATGTATTTCACAACAATGAAACTTCCAAGTACTTCATCTAGAATGAGTGGAGGAGGTAGTTTTCACTCTGGAGGAAACGGTCTTGACTTCACTTTGACCATCAGTGAAG
CTCAGCTTGAAGATGCTGGAGTTTATTACTGTCAGAGTTACCATGAGATAAGTGGTAAACCATTGTTCA**CACAGTA**AAAAAGAGCTTA**ACAAAAACC**

>V1p (160,112-160,613)
ATGTCCTTTATCATCTGGATGCTCTACAGCCTACAAGGTAAAATAAAAACATTATTACAAATAACTTTAAAATATATTTCAGTATTAATGTAATTATTCAGTGCTGTTATTTCTTGTTCTACACAGGTTGCTTT
GGACAAATTGTTGTGACTCAGTATGAGGTTAAATCAGTCCAGCCTGGTCAAACAGTCTCTATAAACTGTAAAGTCAATACTGTGGTTTTTAAAAATGATCAAGGCACCACCTCCAAAGGCCGTTATTATATACA
CTGGTACAGCCAAAAGCCTGAAGAGGCTCCCAGACTCTTGATTTATTACACAAGCGAAAAAAACTCAGGAATTCCATCTCGATTCAGTGGCAGTGGACAGGGAAACGGTCTTGACTTCACTTTGAGCATCAGTG
AAGCTCAGCATGAAGATGCAGGAGATTATTACTGTCAGAGTGTCCATAAGATCAGTGATAACTGGGTGTTCA**CACAGTG**AAAATGTGCCAT**ACAAAAACC**

>V1r (142,547-143,388)
ATGACTGTAATAGCTGTCATTATTTGGACACTCTCAGTCTGCACATGGAGTAAGTATATAAAAATTCACATCATGAAATTACAGAAGCTGAAATCAAGCATTTACTATGTTTCAGTGAGCATTTACATGCATTA
TGTTGTAAACATGAGCATGTGAAGTCAACAAAACATTAGCATTAAATGTCTTAGTATTTGTAGTGTGAAATAAATATTGTGTGCTCATTCTGTGGAGTTTAATTTTATCAGGTTCTGTAGAGTATACATGATTT
ACTAATACAAATCATATATTGTTTATGATCTGTGCATATATGTTTATTTGATTCAAGATCTTTATTGTCATCAATTAAGTTGTACATTATATGCAAATAAATAAACTTAGTATATAGTGCAAGAGTAAGATAT
ATTGAATGTTCAGTCTGTGGAGTTATTTTACAGTGTTTTCTTTCTTTTTTTTCAGGTTCTGTGGGGCAGATTCTGACTCAATCTGAAAATCAAACTGTGCTGCCTGATCAGACTGTCACTATTCACTGTAATCA
TAAACCAGCATTGAATTGTTGGAGTTTAAAAAAACACCTCCCAGTACTGTATGTCCTGGTATCATCATTTTCCTGGAAAAACACATAAACTCTTAATCTATTACACTTCAAACAGAGCTTCAGGTGTTTCTTCTA
GATACAGTGGCAGTGAAAATGCAGACCACACAGTTTTTTACTTTGACCATCAGTGGAGTCCAGCCTAAAGATTCAGGAGTTTATTACTGCCAGAGTCAATACACTAAGGGAGTCTCTGAGTCTGGGAGACATGAC
TGGGTGTTCA**CACAGTG**AAAAAACCTTGT**ACAAAAACA**

>V1t (86,458-86,961)
ATGATTTATTTTATTCTACTGTTGGAGACACTTTTTCTGCTCACTCAAGGTTTGGATCATTTTTATTGAGATGACACATTTTTTAAAATATACATTTTTGCAGAAATGTTAAGCTTTTAGTTATATTTTTATGG
ATTCATTTCAATTTCATTTCTTGTTATTAATTTTTTTGTTCTCTCTCAGTCTCTCATCAAGAGGTTCAGTCTCCTGAGTCTCTGTCTGTTAAAGCAGGGGATTCTGCCTCTATCAGCTGTACTGGGACCAGTGGA
GTTGGTTATGACATGAGCTGGTATTTGCAGAAACCTGGAAAAGCTCCTAAACTCCTGATCTATGGAGTAAGTTACCTTGAATCTGGAGTTTCAGATCGATTCAGTGGAAATGGATCTGAACCTGCTTTTTACACT
GAACATACGTGGAGTCCAGCCTGAAGATGCAGGAGTTTATTACTGTATGGGGTACTATGGTGGTGGAAGGTGCA**CACAGTG**AAACAACATCAA**ACAAAAACC**

>J1a (154,684-154,608)
GGTTTTTGTTCAGCCGCTCTATTAGTTTGTAT**CACTGTG**GTGGACTTTCGGCAGCGGCACCAAACTGGATGTTGGCA
>J1b (146,588-146,512)
GGTTTTTGTTCAGCCGCTCTATTAGTTTGTAT**CACTGTG**GTACACTTTCGGCAGCGGCACCAAACTGGATGTTGGCA
>J1c (882,654-882,578)
GGTTTTTGTTCAGCCGCTCTATTAGTTTGTAT**CACTGTG**GTACACTTTCGGCGGCGGCACCAAACTGGATGTTGGCA
>J1d (592,746-592,670)
GGTTTTTGTACAACTGATGGCTATGTTTGTAT**CACTGTG**GTACACGTTTGGTGGAGGCACTCGACTCTCAGTTGGAA
>J1e (583,809-583,885)
GGTTTCTGTACAGCTGATGATTGTGTTTGTGT**TACAGTG**GCGCACGTTCGGTGGAGGCACTTGACTCTCAATTGGAA
>J1f (163,742-163,666)
GGTTTTTGTACAACTGATGGCTATGTTTGTAT**CACTGTG**GTACACGTTTGGTGGAGGCACTCGACTCTCAGTTGGAA
>J1g (154,648-154,572)
GCATGAATTAGTGCAGATGATTGGGTTTATAT**CACTGTG**GTGAACTTTCGGTAGAGGCACTCGACTGTCAGTTGGAA
>J1h (127,189-127,113)
GGTTTTTGTACAGCTGATGATTGGGTTTGTAT**CACTGTG**GTGGACTTTCGGAGGAGGCACTCGACTCTCTGTTGGAA
>J1i (92,367-92,291)
GGTTTCTGTACAGCTGATGATTGTGTTTGTGT**TACAGTG**GCGCACGTTCGGTGGAGGCACTTGACTCTCAATTGGAA
>J1j (86,169-86,093)
GGTTTTTGTACAGCTGATGATTGGCTTTGTAT**CACTGTG**GTACACATTTGGTGGAGGCACTCGACTCTCAGTTGGAA
>J1k (NW_644842/18,009-17,933)(not shown)
GGTTTTTGTTCAGCCGCTCCATTAGTTTGTAT**CACTGTG**ATACACTTTCGGCGCCGGCACCAGACTGGATGTTGGCA

>C1a (152,269-151,932)
CCGTCTCTCCTCCTCAAGTGAGCGTCCTGCAGCCCTCCAGTGCAGAGATCTCCTCCAAGCGGACGGCTACACTGCTGTGTGTGGCCAACAAGGGCTTCCCGTCAGACTGGAGGCTGGTCTGGAAGGTCCTCAAG
CCTGACGGGAGCAGCAGCAGCGGGGGTCAGGAGAGCAGCAGTGCTGGGCTCCTGGAGAAGGACGGTCTGTACAGCTGGAGCAGCAGCCTGACTCTCACTGAGCAGCAGTGGATGGAGAGCGCCTCAGTGAGCTG
TGAGGCCACACGGAGCGGCCAGCCTGCGCTTACTGCAGGAGACACTGTGAGGAGACAGCAGTGTTCAGAG

>C1a (amino acid)
VSPPQVSVLQPSSAEISSKRTATLLCVANKGFPSDWRLVWKVLKPDGSSSSGGQESSSAGLLEKDGLYSWSSSLTLTEQQWMESASVSCEATRSGQPALTAGDTVRRQQCSE

>C1b (144,171-143,834)
CCGTCTCTCCTCCTCAAGTGAGCGTCCTGCCGCCCTCCAGTGCAGAGATCTCCTCCAAGCGGACGGCTACACTGCTGTGTGTGGCCAACAAGGGCTTCCCGTCAGACTGGAGGCTGGTCTGGAAGGTCCTCAAG
CCTGACGGGAGCAGCAGCAGCGGGGGTCAGGAGAGCAGCAGTGCTGGGCTCCTGGAGAAGGACGGTCTGTACAGCTGGAGCAGCAGCCTGACTCTCTCTGAGCAGCAGTGGATGGAGAGCGCCTCAGTGAGCTG
TGAGGCCACACGGAGCGGCCAGCCTGCGCTCACTGCAGGAGACACTGTGAGGAGACAGCAGTGTTCAGAG

>C1b (amino acid)
VSPPQVSVLPPSSAEISSKRTATLLCVANKGFPSDWRLVWKVLKPDGSSSSGGQESSSAGLLEKDGLYSWSSSLTLSEQQWMESASVSCEATRSGQPALTAGDTVRRQQCSE

>C1c (881,354-881,032)
CCGTCTCTCCTCCTCAAGTGAGCGTCCTGTCGCCCTCCAGTGCAGAGATCTCCTCCAAGCGGGCGGCTACACTGCTGTGTGTGGCCAACAAGGGCTTCCCGTCAGACTGGAGGCTGGTCTGGAAGGTGTCGGGC
AGCAGCGGGGGTCAGGAGAGCAGCAGTGCTGGGCTCCTGGAGAAGGACGGTCTGTACAGCTGGAGCAGCAGCCTGACTCTCTCTGAGCAGCAGTGGATGGAGAGCGTCTCAGTGAGCTGTGAGGCCACACGGAG
CGGCCAGCCTGCGCACACTGCAGAATACACTGTGACGAGACAGCAGTGTTCAGAG

>C1c (amino acid)
VSPPQVSVLSPSSAEISSKRAATLLCVANKGFPSDWRLVWKVSGSSGGQESSSAGLLEKDGLYSWSSSLTLSEQQWMESVSVSCEATRSGQPAHTAEYTVTRQQCSE

>C1d (591,774-592,090)
CTGCGGCTCGTCCCACTGTGGCGGTGCTTTGGCCCTCCAGAGATGAGCTGCAGAAGGGTCAGGCTACAGTGATGTGTGTGGCCAGTAAGGGATTTCCCTCTGACTGGACGCTGAGCTGGAAGGTGGGCGACAGC
AGCAGGAGCTCTGGTGTGAATATGAGTCCCAGTGAGCTGCAGGAGGATGGACTGTACAGCTGGAGCAGCAGCCTGAGCATCACTGAGAGTGAGTGGAGCAGAGCAACAAGCATCAGCTGTGACGCCACACACCC
GTCCCAGCACACAGTGAGCCAGAGGATACACACACAGCAGTGCCATGAT

>C1d (amino acid)
AARPTVAVLWPSRDELQKGQATVMCVASKGFPSDWTLSWKVGDSSRSSGVNMSPSELQEDGLYSWSSSLSITESEWSRATSISCDATHPSQHTVSQRIHTQQCHD

>C1f (162,444-162,137)
CTCGTCCCACTGTGATGGTGCTTTGGCCCTCCAGAGATGAGCTGCAGAAGGGTCAGGTTACAGTGATGTGTGTGGCCAGTAAAGGATTCCCTTCTGACTGGACGCTAAGCTGGAAAGTGGGCGACAGCAGCAGG
AGCTCTGGGGTAAATGTGAGTCCCAGTGAGCTGCAGGAAGATGGACTGTACAGCTGGAGCAGCAGCCTGAGCCTCACTGAGAGTGAGTGGAGCAGAGCAACAAGCATCAGCTGTGACGCCACACACCCGTCACA
GCACACAGTGAGCCACAAGACAAACACACAGCAGTGCCAT

>C1f (amino acid)
RPTVMVLWPSRDELQKGQVTVMCVASKGFPSDWTLSWKVGDSSRSSGVNVSPSELQEDGLYSWSSSLSLTESEWSRATSISCDATHPSQHTVSHKTNTQQCH

>C1h (125,858-125,542)
CTGCGGCTCGTCCCACTGTGGCAGTACTTTGGCCCTCCAGAAATGAACTGCAGAAGGGTCAGGCTACAGTACTGTGTGTGGCCAGTAAGGGATTCCCATCTGACTGGACCCTGAGCTGGAAGGTAGATGACAGC
AGCAGGAGCTCTGGTGTGAATGTGAGTCCCAGTGAGCTGCAGGAGGATGGACTGTACAGCTGGAGCAGCAGCCTGAGCCTCACTGAGAGTGAGTGGAGCAGAGCAACAAGCATCAGCTGTGACGCCACACACCC
GTCCCAGCACACAGTGAGCCACAAGATACGTACACAGCAGTGCCATGAT

>C1h (amino acid)
AARPTVAVLWPSRNELQKGQATVLCVASKGFPSDWTLSWKVDDSSRSSGVNVSPSELQEDGLYSWSSSLSLTESEWSRATSISCDATHPSQHTVSHKIRTQQCHD

>C1j (83,338-83,022)
CTGCGGCTCGTCCCACTGTGGCGGTGCTTTGGCCCTCCAGAGATGAGCTGCAGAAGGGTCAGGCTACAGTGATGTGTGTGGCCAGTAAGGGATTTCCCTCTGACTGGACGCTGAGCTGGAAGGTGGGCGACAGC
AGCAGGAGCTCTGGTGTGAATATGAGTCCCAGTGAGCTGCAGGAGGATGGACTGTACAGCTGGAGCAGCAGCCTGAGCATCACTGAGAGTGAGTGGAGCAGAGCAACAAGCATCAGCTGTGACGCCACACACCC
GTCCCAGCACACAGTGAGCCAGAGGATACACACACAGCAGTGCCATGAT

>C1j (amino acid)
AARPTVAVLWPSRDELQKGQATVMCVASKGFPSDWTLSWKVGDSSRSSGVNMSPSELQEDGLYSWSSSLSITESEWSRATSISCDATHPSQHTVSQRIHTQQCHD

>C1k (NW_644842/16,697-16,399)(not shown)
TGAGCGTCCTGCCGCCGTCCAGTGCAGAAATCTCCTTGAGCAAGAGGGCGACACTGGTTTGTGTGGCCAGCGAGGGCTTTCCGTCAGACTGGAAGCTGAGCTGGAAGGTGAACGGCCACATCAGGTCTCAGGAG
AGCAGTGCTGGGCTGATGCAGGAGAACGCTCTGTACAGCTGGAGCAGCAGCCTGACTCTCACTGAGGACCAGTGGATGACGGAGAGAGGCTCTGTGAGCTGTGAGGCCACACGCAGCGGCCAGCCTACAGTCAG
TGCACAGCTGACCAGAGATCAGTGCTCACAG

>C1k (amino acid)
SVLPPSSAEISLSKRATLVCVASEGFPSDWKLSWKVNGHIRSQESSAGLMQENALYSWSSSLTLTEDQWMTERGSVSCEATRSGQPTVSAQLTRDQCSQ

TYPE 2

>V2a (593,799-593,304)
<u>ATGCTGATCATCTTCTACTGCTCTCTCCTCACTCTTCTGTCATGTAAGTCACACATTTCACTGCAGACATGCCGATCTCTCAGCAGAGACGACAGATTCATCACATTCATATTATTGATTATTATTTTACAGGT</u>
<u>GTCAGATGTGCGAAAGTGCTCACTCAGAAACCCTCCATAATGACTGAAGACAAGGGCAGATCTGTGACTCTGGACTGTAATATTGAAAGAGATGATGGTTTTTATTATGTCAGCTGGTATAAACAGACTCCAAA</u>
<u>CGCAGCTCCTCAGTTTGTGTTGAGATATTATCATTCTCTCAGCTCACCTGATCTCTATGGGGATGGTTTTACCTCCAGCCGCTTCACATCTAAAGCTCAGTCAAAGATTGATTATAGTCTGATCATCAGTAATG</u>
<u>TGGATGTGAGTGATTCTGCAGTGTATTACTGTCAGACATGGGACGACACTGTT</u>AGTGAGGGCGTAT**CACAGTG**ATTCACATCATG**ACAAAAACC**

>V2c (589.648-589,156)
<u>ATGCTGATCATCTTCTACTGCTCTCTCCTCACTCTTCTGTCATGTAAGTCACACATTTCACTGCAGACATGCCGATCTCTCAGCAGAGACAACAGATTCATCACATTCATATTATTGATTATTATTTTACAGGT</u>
<u>GTCAGATGTGCGAAAGTCTTTACTCAGAAACCCTCCATAATGACTGAAGACAAGGGCAGATCTGCGACTCTGGACTGTAATATTGAGAAGGATGAAAGTCAATATGTCAGCTGGTATAAACAGACTCCAAACGC</u>
<u>AGCTCCTCAGTTTGTGTTGAGATATTATCATTCTCACAGCTCACCTGATAGATATGGGGATGGTTTTACCTCCAGCCGCTTCACATCTAAAGCTCAGTCAAGTATTGACTATAGTCTGATCATCAGTAATGTGG</u>
<u>ATGTGAGTGATTCTGCAGTGTATTACTGTGCAACATGGGACGACACTGCT</u>AAAGAGTGGGTAT**CACAGTG**ATTCATATCATG**ACAAAAACC**

>V2d (587,705-587,216)
<u>ATGCTGATCATCTTCTACTGCTCTCTCCTCACTCTTCTGTCATGTAAGTCACACATTTCACTGCAGACATGCCGATCTCTCAGCAGAGACAACAGATTCATCACATTCATATTATTGATTATTATTTTACAGGT</u>
<u>GTCAGATGTGCGAAAGTGCTCACTCAGAAACCCTCCATAATGACTGAAGACAAGGGCAGATCTGTGACTCTGGACTGTAATGTTGAGAAGGATGAGAATTATGTCAGCTGGCATAAACAGACTCCAAACGCAGC</u>
<u>TCCTCAGTTTGTGTTGAGATATTATCATTCTCTCAGCTCACCTGATAGATATGGGGATGGTTTTACCTCCAGCCGCTTCACATCTAAAGCTCAGTCAAGTATTGATTATAGTCTGATCATCAGTAATGTGGCTG</u>
<u>TGAGTGATTCTGCAGTGTATTACTGTCAGACATGGGACAGCTCTGTT</u>AGTGAGGGCGTAT**CACAGTG**ATTCACATCATG**ACAAAAACC**

>V2e (584,456-583,987)
<u>ATGATCTCAATATTCTGCTCTGTCTGCACTCTTCTGACCTGTGAGTTTCCTCATGTCTGATGGACAGTTTAAAGTGTTGGACACATGATCTCATGTAAAGCTTGTGTTTGTTCTGCAGGTGTCAGTGGAGTGAC</u>
<u>TGTAGTGACTCAGAGTCCAGTAATCACAGTCAGTAAAGGACAAACCGCTCAACTGGACTGTAATCTGGGCACTGTGACGAATTATCCTGCTGTATGGTACAAACAGACTCCAGCAGGAGTTCCTCAGTATGTGT</u>
<u>TATACGATTATCATAGCAGTTCTCCTATCTATGGATCTGGTTTCTCTGCTCCTAAATTCACATCAACACACTCATCTACATCAGATTATCGTCTGATTATCAGTAATGTGGATGTGAGTGATTCTGCAGTGTAT</u>
<u>TACTGTGCAACATGGGACGACACTGCT</u>AGTGAGTGGGTAT**CACAGTG**ATTCACATCATG**ACAAAAACC**

>V2f (581,483-581,011)
<u>ATGATCTCAATATTCTGCTCTGTCTGCACTCTTCTGACCTGTGAGTTTCCTCATGTCTGATGGACAGTTTAAAGTGTTGGACACACGATCTCATGTAAAGCTTGTGTTTGTTCTGCAGGTGTCAGTGGAGTGAC</u>
<u>TGTAGTGACTCAGAGTCCAGTAATCACAGTCAGTAAAGGACAAACCGCTCAACTGGACTGTAATCTGGGCACTGTGACTGGTTCTGCTGCTAGATGGTACAAACAGACTCCAGCAGGAGTTCCTCAGTATGTGT</u>
<u>TAAGATTTTATCATGACTGGAGTTCACCTTACTATGGATCTGGTTTCTCTGCTCCTAAATTCACATCAACACATTCATCTACATCAGATTATAGTCTGATTATCAGTAATGTGGATGTGAGTGATTCTGCAGTG</u>
<u>TATTACTGTCAGACATGGGACAGCTCTGTT</u>AGAGAGTATGTAT**CACAGTG**ATTCACATCATG**ACAAAAACC**

>V2g (575,795-575,237)
<u>ATGCTGACGCTCTTCATCCTCCTCATCATCACTCTATCATGTGAGACGACACTAGACTGCAGATCCAGAGGCACATACGACTGTCGAGTGTGAGAGTTGTGTGGCATTAATGGTGTGTGTGTGTGTGTGTGTGT</u>
<u>GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTTTGTGTGTGTGTTTGTGTTCTGCAGGTGTCAGTGGAGTGACTGTAGTGACTCAGAGTCCAGTAATCACAGTCAGTAAAGGACAAAC</u>
<u>CGCTCAACTGGACTGTAATCTGGGGACGGTGACTCATGTTGCTGGATGGTACAAACAGACTCCAGCAGGAGCTCCTCAGTATGTGACAGTGTTTTATCACAGCTGGAGTTCACCTCGCTATGGATCTGGTTTCT</u>
<u>CTGCTCCTAAATTCACATCAACATCTTCATCTTATTCGGATTATAGTCTGATCATCAGTAATGTGGATGTGAGTGATTCTGCAGTGTATTACTGTCAGACATATGACAGCTCTGTTAGAGAGTATGTAT</u>**CACAG**
TGATTCACATCATG**ACAAAAACC**

>V2h (573,874-573,403)
ATGATCTCAATATTCTGCTCTGTCTGCACTCTTCTGACCTGGAGTTTCCTCATGTCTGATGGACAGTTTAAAGTGTTGGACACACGATCTCATGTAAAGCTTGTGTTTGTTCTGCAGGTGTCAGTGGAGTGACT
GTAGTGACTCAGAGTCCAGTAATCACAGTCAGTAAAGGACAAACCGCTCAACTGGACTGTAATCTGGGGACGGTGACTGATTATCCTGCTGTATGGTACAAACAGACTCCAGCAGGAGTTCCTCAGTATGTGCT
ACAATTTCGTCATGACTGGAGATCACCTAGCTATGGATCTGGTTTCTCTGCTCCTAAATTCACATCAACATCTTCATCTAGATCAGATTATAGTCTGATCATCAGTAATGTGGATGTGAGTGATTCTGCAGTGT
ATTACTGTCAGACATGGGACAGCTCTGTTAGTGAGTACGTAT**CACAGTG**ATTCACATCATG**ACAAAACC**

>V2j (565,592-565,100)
ATGCTGATCATCTTCTACTGCTCTCTCCTCACTCTTCTGTCATGTAAGTCACACATTTCACTGCAGACATGCCGATCTCTCAGCAGAGACGACAGATTCATCACATTCATATTATTGATTATTATTTTACAGGT
GTCAGATGTGCGAAAGTGCTCACTCAGAAACCCTCCATAATGACTGAAGACAAGGGCAGATCTGTGACTCTGGACTGTAATATTGAAAAGGAGGAGAATTATTATGTCAGCTGGTATAAACAGACTCCAAACAC
AGCTCCTCAGTTTGTGTTGAGATATTATCATGCTCACAGCTCACCTGGTGCATATGGGGATGGTTTTACCTCCAGCCGCTTCACATCTAAAGCTCAATCAAGTATTGACTATAGTCTGATCATCAGTAATGTGG
CTGTGAGTGATTCTGCAGTGTATTACTGTCAGACATGGGACAACACTGCTAAAGAGTATGTAT**CACAGTG**ATTCACATCATG**ACAAAACC**

>V2k (560,520-560,029)
ATGCTGACACTCTTCATCCTCCTCATCATCACTCTATCATGTGAGACGCCACTAGACTGCAGATCCAGAGGCACATCAGACTGTCATGTGGGTGTGTTGTGTGGCATTAATGCTGTGTGTGTGTGTTTGTGTTCTG
CAGGTGTCAGTGGAGTGACTGTAGTGACTCAGAGTCCAGTAATCACAGTCAGTAAAGGACAAACTGCTCAACTGGACTGTAATCTGGGGACTGTGACAAGTTCTGAAGCTAGATGGTACAAACAGACTCCAGCA
GGAGTTCCTCAGTATGTGTTAAGATTTTATCATACCATGAGATCACCTGACTATGGATCTGGTTTCTCTGCTCCTAAATTCACATCAACACATTCATCTAAATCAGATTATAGTCTGATCATCAGTAATGTGGC
TGTGAGTGATTCTGCAGTGTATTGCTGTAAGACATGGGACCACTCTGCTAATGAATATGTAT**CACAGTG**ATTCACATCATG**ACAAAACC**

>V2l (534,354-533,866)
ATGCTGATCATCTTCTACTGCTCTCTCCTCACTCTTCTGTCATGTAAGTCACACATTTCACTGCAGACATGCCGATCTCTCAGCAGAGACGACAGATCATCACATTCATATTATTGATTATTATTTTACAGGTG
TCAGATGTGCGAAAGTGCTCACTCAGAAACCCTCCATAATGACTGAAGACAAGGGCAGATCTGTAACTCTGGACTGTAATGTTGAGAAGGATGAGAGTTATATCTACTGGTATAAACAGACTCCAAACGCAGCT
CCTCAGTTTGTGTTGAGATATTATCATTCTCACAGCTCACCTGATCTCTATGGGAATGGTTTTACCTCCAGCCGCTTCACATCTAAAGCTCAGTCAAAGATTGATTATAGTCTGATCATCAGTAATGTGGTTGT
GAGTGATTCTGCAGTGTATTACTGTCAGACATGGGACAGCTCTGTTAGTGAGTACATAT**CACAGTG**ATTCACATCATG**ACAAAACC**

>J2a (580,421-580,349)
GGTTTTTGTACGGCTGAACATATCAAATCTTT**CACTGTG**GTATTCGGACAAGGAACTAAACTCATTGTGACGG
>J2b (572,802-572,730)
GGTTTTTGTACGGCTGAACATATCAAATCTTT**CACTGTG**GTATTCGGACAAGGAACTAAACTCATTGTGACGG
>J2c (559,441-559,369)
GGTTTTTGTATGAGTGAAAGTGAGAAATGAAT**CACTGTG**GTCTTTGGCCAAGGAACAAAACTGAACGTCATCG
>J2d (533,275-533,203)
GGTTTTTGTACGGCTGAACATATCAAATCTTT**CACTGTG**GTATTCGGACAAGGAACTAAACTCATTGTGACGG

>C2a (578,855-578,533)
ACGCTGCTGCTGCTCCTCCTGTGCTCACCATCCTCCGTCCATCATCCAGAGAAGAGCTGAGCTCCAGTAAAGTCACTCTGCTGTGTCTGATCAATCACATGTCTGTGGCTTTCGCTGATGTGCGTTGGCTTGTG
AACGGGAACTCGGTTACCGAAGGCGTCTTCACCGGCTCTGCTGAACAGCAGCCTGATCACAAATTCAAGAGCAGCAGTTCTTTAACCATCCAGAGGTCCGAGTGGGACAAAGACAGAGAGCTGACATGTGAAGC
CACTGTGGCCTCCAAAACCAGCAGAGCAAGCATCAGGAAATCTGAATGCAGCGAC

>C2a (amino acid)
AAAAPPVLTILRPSSREELSSSKVTLLCLINHMSVAFADVRWLVNGNSVTEGVFTGSAEQQPDHKFKSSSSLTIQRSEWDKDRELTCEATVASKTSRASIRKSECSD

>C2b (557,168-556,852)
ACTTTCCTCTGACTCCTCCTGTGGTGACTCTCTTGCCTCCGTCCGCTGAGGCTCTGAAGGACAGTAATGCTACACTGGTGTGTCTGGCCGAGGAGATTTCTGTGGCTCTGCTTGACGTCAGCTGGACAGTCAAT
GGACACTCCGTCACTGATGGCGTTTGGAGCAGCTCTACTGAGAAACAGCAAGACCAAACCTTCAAAATGAGCAGCTTTCTGACCATCCAATCCTCTGACTGGACCAGCAATGCGCAATTCTCCTGTCGAGTGTC
GCTGGGCTCAAAAATGACGGAGAAAAGCATCTCCAGTTCACAGTGTAAC

>C2b (amino acid)
DFPLTPPVVTLLPPSAEALKDSNATLVCLAEEISVALLDVSWTVNGHSVTDGVWSSSTEKQQDQTFKMSSFLTIQSSDWTSNAQFSCRVSLGSKMTEKSISSSQCN

TYPE 3

>V3a (1,027,144-1,026,655)
ATGATCAACATCTGCATCTGCATCTGGACTCTGACTCTCTGCATTCAGGGTAATTCTCCTCTTATTAATTAGCACTGATCAATAATCAGCTTCTTCTCTTCTACACTTCTAGATCCACTAATGAGGAAAATCTC
TTTATTTCAGAGAGTTGGGGACAAGTGACAGTGACTCAGAGTCCAGCAGTCAAATCCATCACACAAGGAGAAACTCTGACCATCAGCTGTAAAACCAGTCAGCCCATATCTGGCTGTAGTAACTGTGTGTCCTG
GTATCAGCAGAAACCTGGAGAAGCTCCTAAACTCCTGATTAAATATATAAACAGCCGTTACTCAAACACTCCATCTAGATTCAGTGGTGGTGGATCTGACTATGGCAGTGATTTTACTCTGAGCATCAGTGGAG
TCCAGACTGAAGATGCTGGAGATTATTACTGTCTGAGTTACCACAGTGGTGGAGTGTTCA**CACAGTG**ATAAAGCGCCGT**ACAGAAACC**

>V3b (1,029,792-1,030,271)
ATGATCAACATCTGCATCTGCATCTGGACTCTGACAAAGGTCGCTCGAGGTTTGTGTAACTCACGTTCATCCCGTGTGTGTGTAATTGATCATATTCATACTGTATATGTTGCTGTAATAGTTGTGTTGTGTTT
CAGAGTGTCGAGGACAGATCACTGTTACCCAGAGTCCATCAGTAACAGCTCAAGCTGGACAAGCCGTCTCTATAACCTGCAAAACCAGCACTGCAGTGTGGAATGGTAATCATTTAGCCTGGTATCAGCAGAAA
CCTGGAGAAGCTCCTAAACTCCTGATTTATTATGCAACAACCCGTTACTCAAACACTCCATCTAGATTCAGTGGCGGTGGATCCAGAACTGATTTTACTCTGAGCATCAGTGGAGTCCAGACTGAAGATGCTGG
AGATTATTACTGTCAGAGTTTCCACTATCCGGGCAGCACTAGGCTGTTCA**CACAGTG**ATAAAGCGCCGT**ACAAAAACC**

>V3d (1,034,848-1,034,341)
ATGAAGCTCGTTCTCTTGTTCTTGACCTCACAGCTGTATTTTCAGGGTGGGAAATCATCACTAGATGTGTAAAGTTAAACTGAATCAGGGTCATTTGAGTGTGTCGTGTAGTCGTTCCTCATTTTCTGTGTTCT
TCAGCTTCCACAGGTCAGGTGACTGTAATCCAGACCCCCGCAGAGCATCATGTTCAGCACACACACACAGTCACTCTGAGCTGTAAAACCAGCAGCTCCATTGGACGTTGGTCAGATGGTGACTGTAAGTCTTG
TGTCGCCTGGTACCTGCAGAAACCCGGAGAAACTCCTAAACCTTTAGTTTACAACATCAGAACCCTTCGTGCTGGAACTCCATCTAGATTCAGTGGAAATGGAGCTGATCATGGGACTGATTTTACTCTGAGCA
TCAGTGGAGTCCAGACTGAAGATGCTGGAGATTATTACTGTCAGAGTTACCACTGGATCAACAGTAAAGCTGTGTTCA**CACCGTG**ATAAAGCGCCGT**ACAAAAACC**

>V3e (1,035,869-1,035,356)
ATGATCAACATCTGCATCTGCATCTGGACCATTGCTGTCTTTATTCATGGTGAGTCTTTCTGTTTAATATGAATGAATTTGATTTTTAGCATTTTAATATGGAAAAGTGATGTATTAAATTGAAATATTAAATA
AACCATTTTATTTTCTCTTCTTCAGGATCATCAGGAGTCACTTTGACTCAACCTGAAGTTAAAACTGTCCAACAAGACGAAACAGCGACAATAGAGTGTCACATAGATGTGGGAATATACAGCAACTATTTAGC
CTGGTATCAGCAGAAACCTGGAGAAACTCCTAAACTCCTGATTAAATATATAAACAGCCGTTACTCAAACACTCCATCTAGATTCAGTGGCAGTGGAACAGTGAATACTGGAAAAGATTTTACTCTGAGCATCA
GTGGAGTCCAGACTGAAGATGCTGGAGATTATTACTGTCAGAGTTTACACTGTCCTAGTACTAGCTGTGTGTTCA**CACAGTG**ATAAAGCGCCGT**ACAAAAACC**

>V3f (1,037,464-1,036,969)
ATGATCAACATCTGCATCTGGACGCTCTATATGTGTGTTCAAGGTATTGAAATGGTCTTTTATTATAATATTAACACAGTTAAATATAAGACACATGTAGAGTGAATTCTGCATTTCTAACTGTATTTCTAACA
GGTTTGAGTGGACAGGTGACTGTAACTCAGAGTCCTGCTGGAAAAATGGTTCAAGTTGGTGAATCAGTGACGATAACCTGCAAATCAAACCCAGAAGTGGGAGATTACTGTGTAGGAACACACACCTGTATGGC
CTGGTATCAGCAGAAACCTGGAGAAGCTCCTAAACTCCTGATTCATTCTGCAAATCGGAGAGATTCAAACACTCCATCTAGATTCAGTGGCAGTGGATCTTATAGTGATTTTACTCTGAGCATCAGTGGAGTCC
AGACTGAAGATGCTGGAGATTATTACTGTCAGAGTTTCCACTATCCGGGCAGCACTAGGCTGTTCA**CACAGTG**ATAAAGTGCCGT**ACAGAAACC**

>V3g (1,042,142-1,041,636)
ATGATCAACATCTGCATCTGCATCTGGACCGTTGCTGTCTTCATTCATGGTGAGTCTTTTCTGTTTAATATGAATGAATTTGATTTTTAGCATTTTAATATGGAAATGTGATGTATGAAATTGAAATATTAAAA
AAAAACATTTTATTTTCTCTTCTTCAGGATCATCAGGAGTCACTTTGACTCAACCTGAAGTTAAAACTGTCCAACAAGACGAAACAGCCACAATAGAGTGTCACATAGATGTGGGAATATACAGCAACTATTTA
GCCTGGTATCAGCAGAAACCTGGAGAAACTCCTAAACTCCTGATCTATCGGATAAACAACCGTTACTCAAACACTCCATCTAGATTCAGTGGTGGTGGAACAGTGAATACTGGAAAAGATTTTACTCTGAGCAT
CAGTGGAGTCCAGACTGAAGATGCTGCAGATTATTACTGTCAGAATGTAGACTGTTCTGGTAGTAGCTGTGTGTTCA**CACAGTG**ATAAAGCGCCGT**ACAAAAACC**

>V3h (1,044,595-1,044,097)
ATGATCAACATCTGCATCTGCATCTGGACTCTGACTCTCTGCATTCAGGGTAATTCTCCTCTTATTAATTAGCACTGATCAATAAACAGCTTCTTCTCTTCTACACTTCAAGATCCACTAATGAGGAAAATCTC
TTTATTTCAGAGAGTTGGGGACAAGTGACAGTGACTCAGAGTCCAGCAGTCAGATCCATCACACAAGGAGAAACTCTGACCATCAGCTGCAAAACTACTCCGCTGGTGTCTGGCTGTAGTGACTGTGTGTCCTG
GTATCAGCAGAAACCTGGAGAAACCCCTAAACTCCTCATCTATGATATAAGCAGCCCTTACTCAAACACTCCATCTAGATTCAGTGGTGTAGGATCTAACTTTGGCAGTGATTTTACTCTGATCATCAGTGGAG
TCCAGACTGAAGATGCTGGAGATTATTACTGTCTGAGTGTAAACAAGATCAGTGATAGTTGGGTGTTCA**CACAGTG**ATAAAGCGCCGT**ACAAAAACC**

>J3a (1,038,140-1,038,216)
GGTTTTTGTACGAGTGTTTCAGCACATTCAAG**CACTGTG**GTACACATTCGGTGGAGGAACCAAACTGCTGATCTCAA

>C3a (1,039,915-1,040,219)
CTGGTTCCCCAGTGAAGCCCTCAGTGTCTCTGCTCGGCTCCTCTTCTCTGCAGAGCTCTGGAGACTCGGCCGCTCTGCTCTGCCTGCTCAGCTCTTACTCTCCACCGGGGGCGAAGGTGAGCTGGACACGGGAC
GGCTCTGAGCTCAGCGAGGGGTTCCTGACCAGTGCAGAGAGCCAGCAGGACGGCCGCTACAGCTGCACCAGCGTCCTGAAGCTCAAGAGAGAAGAATGGGAGAAGCGAGAGCGCTACGCCTGCAGAGTCACTCA
TGCTGGAGGTGATCAGGAGATCCCGTTCCCCAAGTGC

>C3a (amino acid)
GSPVKPSVSLLGSSSLQSSGDSAALLCLLSSYSPPGAKVSWTRDGSELSEGFLTSAESQQDGRYSCTSVLKLKREEWEKRERYACRVTHAGGDQEIPFPKC