APJAI
Asian Pacific Journal of
Allergy and Immunology

ORIGINAL ARTICLE

Comparative analysis of asian carps parvalbumin reveals the divergence pattern of major fish allergen

Judy Kin Wing Ng,[1] Qing Xiong,[1,2] Ling Shi,[1] Christine Yee Yan Wai,[3] Soo Kyung Shin,[1] Fu Kiu Ao,[1] Agnes Sze Yin Leung,[3,5] Nicki Yat Hin Leung,[3] Ting Fan Leung,[3,5] Stephen Kwok Wing Tsui[1,2,4]

Abstract

Background: Asian carps, a popular freshwater fish globally, are valued for their flavor and serve as a crucial protein source, especially for infants. However, grass carp parvalbumin is highly allergenic, surpassing the allergenicity of fish like salmon and cod. The allergenic potential of parvalbumin in other Asian carps remains unknown, underscoring the need for allergen identification to improve the precision of fish allergy diagnosis and treatment.

Objectives: To identify all parvalbumin homologs in Asian carps and investigate the role of gene divergence in allergenic homolog formation.

Method: Three annotated genomes of Asian carp, including grass carp, black carp and bighead carp, were constructed using a hybrid assembly approach. Through sequence homology at the genomic level, all the homologs of major fish allergens were identified. Bioinformatics tools were then employed to reveal the gene structures, expression levels, and protein conformations of parvalbumin.

Results: Grass carp genome analysis showed nine parvalbumin homologs, with Cid_PV2 most similar to Cten i 1. Bighead and black carp genomes had ten homologs, including potentially allergenic Mpi_PV7 and Hno_PV7. Tissue-specific expression patterns revealed alternative usage of parvalbumin homologs. Gene duplication events expanded parvalbumin copies in bony fish, with two gene clusters identified in Asian carp genomes.

Conclusion: All the homologs of Asian carps' parvalbumin were accurately identified and gene divergence contributed to the formation of allergenic homologs. Together with a comprehensive gene sequence profile of carps' parvalbumin, those could be applied to achieve a more precise clinical diagnostic test.

Keywords: Fish allergy, Asian carp, allergen, genome, parvalbumin

Citation:
Ng, J. K. W., Xiong, Q., Shi, L., Wai, C. Y. Y., Shin, S. K., Ao, F. K., Leung, A. S. Y., Leung, N. Y. H., Leung, T. F., Tsui, S. K. W. (0000). Comparative analysis of asian carps parvalbumin reveals the divergence pattern of major fish allergen.
Asian Pac J Allergy Immunol, 00(0), 000-000.
https://doi.org/10.12932/ap-200823-1673

Affiliations:
[1] School of Biomedical Sciences,
The Chinese University of Hong Kong, Shatin, Hong Kong
[2] Hong Kong Bioinformatics Centre,
The Chinese University of Hong Kong, Shatin, Hong Kong
[3] Department of Paediatrics, The Chinese University of Hong Kong, Shatin, Hong Kong
[4] Centre for Microbial Genomics and Proteomics,
The Chinese University of Hong Kong, Shatin, Hong Kong
[5] Hong Kong Hub of Paediatric Excellence,
The Chinese University of Hong Kong, Shatin, Hong Kong

Corresponding author:
Stephen Kwok-Wing Tsui
E-mail: kwtsui@cuhk.edu.hk

Abbreviation:

BLAST	Basic Local Alignment Search Tool
BUSCO	Benchmarking Universal Single-Copy Orthologs
CPV3	Parvalbumin 3 from chicken
MYA	Million Years Ago
NCBI	National Center for Biotechnology Information
PV	Parvalbumin
SPT	Skin Prick Test
WGD	Whole Genome Duplication
WHO/IUIS	World Health Organization/International Union of Immunological Societies

Abbreviations of species name were listed in Table S2

Introduction

Fish allergy is one of the eight most common human food allergies, and the highest prevalence was 2.3% in Asian countries.[1,2] Our territory-wide survey revealed that fish was a main cause of adverse food reactions in Hong Kong pre-school children.[3] Collectively, fish was also a major food triggers of pediatric anaphylaxis, which affected a wide range of children from infants to school-age children in Hong Kong.[4] Understanding the spectrum of fish allergens is crucial for developing optimal diagnostic approaches and designing effective immunotherapeutic strategies. According to the WHO/IUIS Allergen Nomenclature database http://www.allergen.org/,[5] up to thirteen groups of fish allergens have been reported in fifteen species, of which the major allergen, parvalbumin, was first identified in Atlantic cod *Gadus (G.) morhua*.[6] Parvalbumin from cod can cross-react with those from other fishes such as salmon *Salmon (S.) salar* and tuna *Thunnus (T.) albacares*[7] and the allergenic property of parvalbumin cannot be destroyed by heat or enzymatic digestion.[8,9] Previous study investigated the content of parvalbumin in different muscle types and revealed that white muscle responsible for short burst swimming in fish contained higher level of parvalbumin.[10,11]

Freshwater fishes constitute a major proportion of fish consumption in Asian countries.[12] In Hong Kong, daily consumption of freshwater fish is over 140 tons[13] with the most popular species being Asian carps, typically referring to four freshwater fish species under family Xenocyprididae including grass carp *Ctenopharyngodon (C.) idella*, black carp *Mylopharyngodon (M.) piceus*, bighead carp *Hypophthalmichthys (H.) nobilis* and sliver carp *Hypophthalmichthys (H.) molitrix*.[14] Asian carps are widely consumed and valued for their affordability, nutritional value, and easy-to-ingest texture. Grass carp, recommended for infant nutrition, is both a healthy and affordable option. However, it is essential to recognize that grass carp can also cause allergies with symptoms ranging from mild skin irritation to severe anaphylaxis.

While a recent publication has highlighted the high allergenicity of *C. idella* compared to cod and salmon,[15] there is still a lack of studies on the allergen profiles of other Asian carp species. Additionally, Asian carps are not included in the 26 commercially available skin prick tests (SPTs) and 28 fish extracts for ImmunoCAP.[16,17] To improve fish allergy diagnosis, we studied the gene family evolution of parvalbumin in three Asian carp species commonly found in Hong Kong's wet markets (*C. idella*, *M. piceus*, and *H. nobilis*). Our findings contribute insights for developing tailored diagnostic tests for fish allergy, addressing the limitations of current diagnostic approaches.

Methods

Please refer to the supplementary materials.

Ethics approval

This study was approved by Clinical Research Ethics Committee (Reference no. 2019.612) and written informed consent was obtained from all individual participants and/or their parents.

Results

Genome assembly and phylogenomic analysis

To ensure that the fish used in our study were consistently sourced as those commonly consumed by local citizens, we obtained them from a local fish farm that supplies the local wet markets in Hong Kong. We generated three high-quality Asian carp' genomes (**Table S1**), and the size was around 860 to 880 Mbp. Based on the phylogenetic tree constructed by COX 1 gene sequence with 74 bony fish under the same family, the result indicated the de novo assembled genome aligned with the reported mitochondrial sequence (**Figure S3**). The completeness of those genomes was over 95% and the annotation completeness ranged from 87.0 to 90.5% assessed by BUSCO analysis (vertebrata_odb9). Other genomes were obtained from NCBI GenBank database (**Table S2**), and the genome and annotation completeness were ranging from 70.4 to 96.8% and 83.5 to 99.0% respectively. The quality of genomes were satisfactory and high-quality genomes are essential for downstream analysis.

To provide a deeper understanding of the origin and evolution of fish parvalbumins, phylogenetic tree was constructed to illustrate the phylogenetic relationship of selected Asian carps with other fish species (**Figure 1**). We selected four species (*C. carpio*, *G. morhua*, *S. salar* and *P. hypophthalmus*) in Teleost with complete genome on the NCBI GenBank database. Order Cypriniformes consisted of most of the freshwater fish (including zebrafish as model organism) and *C. idella*, *M. piceus* and *H. nobilis* were closely related species under family Xenocyprididae. The phylogenetic tree provides a valuable reference for studying the evolutionary relationships between species and the divergence times of various taxonomic groups. Based on the phylogenetic tree of Teleost fishes, it is estimated that their divergence occurred over 300 million years ago (MYA), during the Carboniferous period. Moreover, the Cypriniformes order, which includes the carp species analyzed in our study, emerged with a most recent common ancestor at 70 MYA. The family Xenocyprididae, to which the grass carp (*C. idella*) belongs, evolved more recently, at approximately 15 MYA. The phylogenetic tree and the estimated divergence times of fish species were used as a reference for downstream analysis of parvalbumin homologs. This analysis aimed to identify parvalbumin genes and their homologs across different fish species, which can aid in understanding the evolution and diversification of these proteins.

Genome-wide identification of fish allergens in Asian carps

We first identified potential allergenic proteins in Asian carps using a genome sequence homology approach. To achieve this, the references allergen protein sequences were retrieved from all the reported fish allergen group in WHO/IUIS database. Next, those reference sequences from closest species were used to search for putative allergens in Asian carp genomes. We identified 11 putative allergen groups (group 1-4, 6-11, 13) in Asian carps (**Table S3**). For each allergen group we also determined the number of homologs and identified the homolog with the highest percentage matching as the putative allergen.

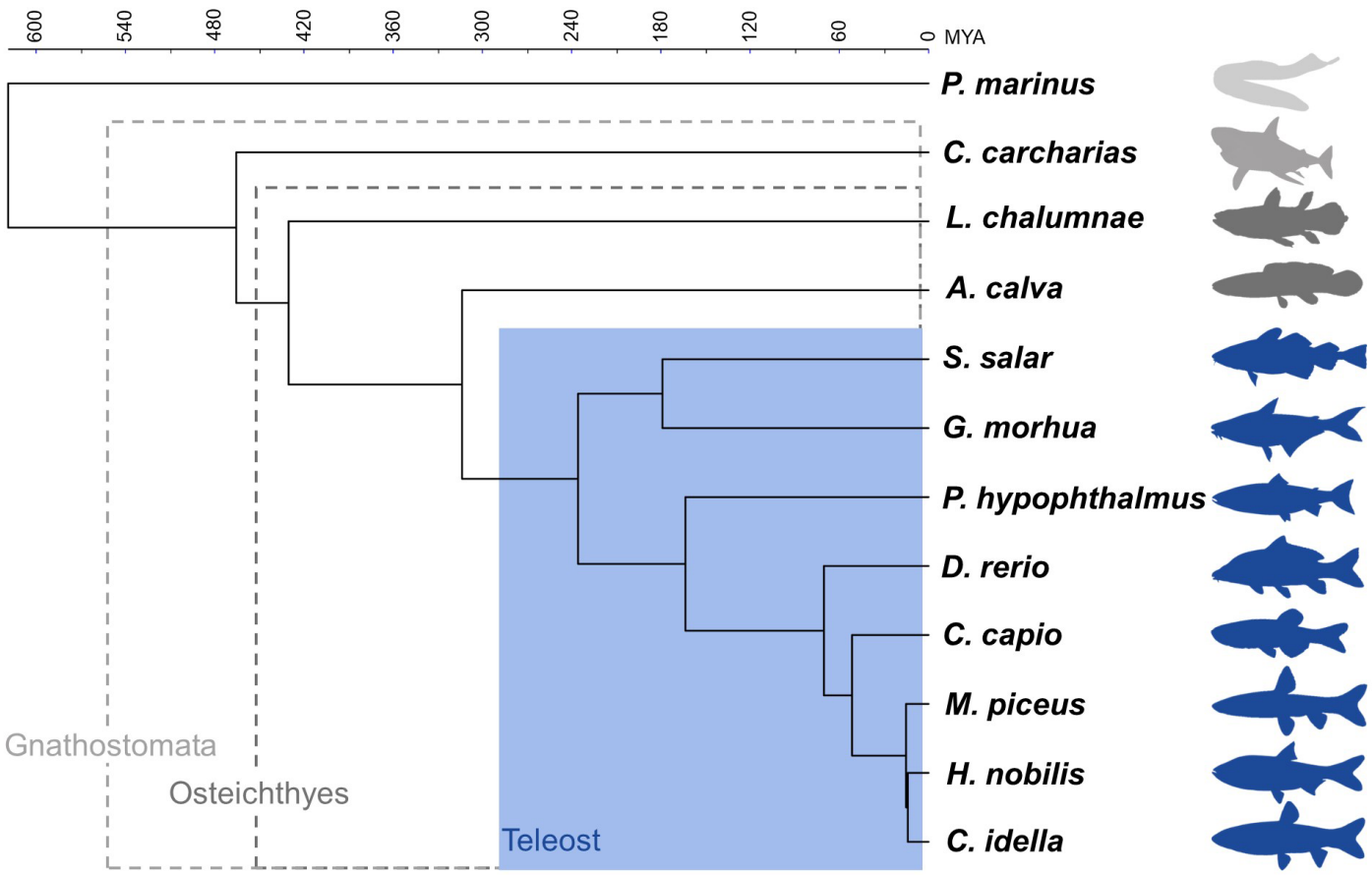


Figure 1. Phylogenetic tree of Gnathostomata. The tree was constructed based on 315 BUSCO gene sequences of 12 species and the time scale was in million years ago (MYA). *D. rerio*, *C. carpio*, *C. idella*, *M. piceus* and *H. nobilis* were under the order of Cypriniformes. Five species under Cypriniformes together with *G. morhua*, *S. salar*, *P. hypophthalmus* and *A. calva* were Actinopterygii (ray-finned fish) and *L. chalumnae* was Sarcopterygii (lobe-finned fish), both groups were under Osteichthyes (bony fish). All species excepted *P. marinus* were Gnathostomata (jawed vertebrates). The silhouettes of organisms were not drawn in scale.

Interestingly, we found that the putative allergens in *C. idella* shared the highest sequence homology (over 85%) with the reference sequences, indicating their potential allergenicity. Similarly, the putative allergens in *M. piceus* and *H. nobilis* also demonstrated high sequence identity with reported sequences, suggesting their potential as allergens.

We retrieved protein sequences from the allergen groups, including group 1, putative group 2, and 3 allergens in *C. idella* with the highest sequence identity to the previously reported allergens in the WHO/IUIS database, including Cten i 1 (GenBank accession: QCY53440.1), Cyp c 2 (AWS00995.1), and Sal s 3 (ACH70901). To test the allergenicity of these proteins, we conducted an indirect ELISA using the sera of patients who had reported allergic reactions to grass carp ingestion (**Table S3**). The patients had been diagnosed with grass carp allergy by a physician and experienced a range of symptoms, including anaphylaxis, angioedema, contact urticaria, erythema, gastrointestinal, neurologic, oral, respiratory, and urticaria, occurring within two hours of grass carp intake. Our results indicated that parvalbumin (group 1) was the major allergen in *C. idella*, and we did not detect the allergenicity of aldolase (**Figure S1**). The allergenicity of recombinant parvalbumin and beta enolase was found to be 45% and 15% respectively.

Based on the results of indirect ELISA among group 1 to 3 allergen in *C. idella*, parvalbumin had the highest IgE reactivity and it was also reported as the major allergen salmon and catfish previously.[35] Thus, we aimed to investigate the evolutionary divergence and the emergence of allergenic parvalbumin.To achieve this goal, we focused on the identification of parvalbumin homologs in Asian carps. Through genome analysis, we identified nine homologs of parvalbumin in the *C. idella* genome. Among them, Cid_PV2 showed the highest sequence identity (99.083%) to the previously identified allergenic homolog Cten i 1 and was regarded as the putative allergenic homolog in *C. idella* (**Table 1**). Similarly, ten parvalbumin homologs were identified in the genomes of *M. piceus* and *H. nobilis*, respectively. Based on their sequence identities to Cten i 1, we identified Mpi_PV7 and Hno_PV7 as the putative allergenic homologs in these species.

Table 1. Parvalbumin homologs identified in Asian carps.
The homologs were identified based on sequence homolog to Cten i 1 (GenBank (GenBank accession: QCY53440.1) by TBLASTN algorithm.

Species	Homologs	Identity (%)	Expression[c]
C. idella	Cid_PV2[a]	99.1	3.83
	Cid_PV6	85.3	2.39×10^{-2}
	Cid_PV5	83.5	8.79×10^{-4}
	Cid_PV3	86.2	3.31×10^{-4}
	Cid_PV8	63.3	1.62×10^{-5}
	Cid_PV9	64.8	4.97×10^{-5}
	Cid_PV1	58.9	0
	Cid_PV4	59.3	0
	Cid_PV7	53.3	0
M. piceus	**Mpi_PV7[b]**	**97.2**	5.37×10^{-1}
	Mpi_PV2	87.2	4.93×10^{-2}
	Mpi_PV3	82.6	8.72×10^{-5}
	Mpi_PV6	88.5	5.40×10^{-4}
	Mpi_PV4	62.4	1.13×10^{-5}
	Mpi_PV10	64.8	0
	Mpi_PV8	58.9	0
	Mpi_PV9	59.8	7.89×10^{-2}
	Mpi_PV5	59.3	7.94×10^{-6}
	Mpi_PV1	53.3	0
H. nobilis	**Hno_PV7[b]**	**97.3**	1.85
	Hno_PV3	86.2	8.91×10^{-2}
	Hno_PV2	82.5	6.28×10^{-5}
	Hno_PV6	88.5	5.55×10^{-4}
	Hno_PV1	63.3	1.20×10^{-4}
	Hno_PV9	64.8	3.72×10^{-4}
	Hno_PV8	57.9	0
	Hno_PV10	60.7	3.41
	Hno_PV5	59.2	3.86×10^{-4}
	Hno_PV4	53.2	3.92×10^{-6}

[a]Highest sequence similarity to Cten i 1 in *C. idella*
[b]Putative allergenic parvalbumin in *M. piceus* and *H. nobilis*
[c]Normalized Transcripts Per Million (TPM) by GAPDH

Parvalbumin is a calcium-binding protein that belongs to the calmodulin family and contains two EF-hand motifs. It is commonly found in the muscle of vertebrates as well as in GABAergic neurons.[18,19] In our study, we constructed a 3D model of parvalbumin based on the sequence of Cid_PV2 and Cyp c 1 (CAC83659.1) (**Figure 2A**). Previous studies have determined the epitope sequences of parvalbumin in Gad m 1, Sal s 1, Sco j 1, Cyp c 1 and Lat c 1.[20-23,39] To study the conserved and variable regions of parvalbumin homologs in different species, we aligned the homologs sequences with reference to Cyp c 1. The result revealed that the epitopes were located at the non-functional EF-hand motif and the EF-hand motif near the C-terminal in the sequence alignment (**Figure 2B-C**). Among different homologs, the dN/dS value indicated positive selection at the non-functional part of the protein, while the conserved sequences encode the functional EF-hand motif (**Figure 2B**). Conversely, the IgE epitopes at the non-functional part were more conserved compared to the rest of the sequences in this region. These results suggested that the IgE epitope sequences were more conserved among different homologs. In particular, our analysis revealed that Cid_PV2 and Cid_PV6 have epitope sequences that are nearly identical, with only one base difference observed at Ser36. These similarities with Cyp c 1 suggest that Cid_PV6 may also possess allergenic properties in *C. idella*. Moreover, we observed that Mpi_PV7 and Hno_PV7 demonstrated high epitope sequence identity with Cyp c 1 among the identified homologs. These findings indicate that these parvalbumin homologs may also have potential allergenicity, and further studies are needed to confirm their allergenicity and report them as isoallergens of parvalbumin.

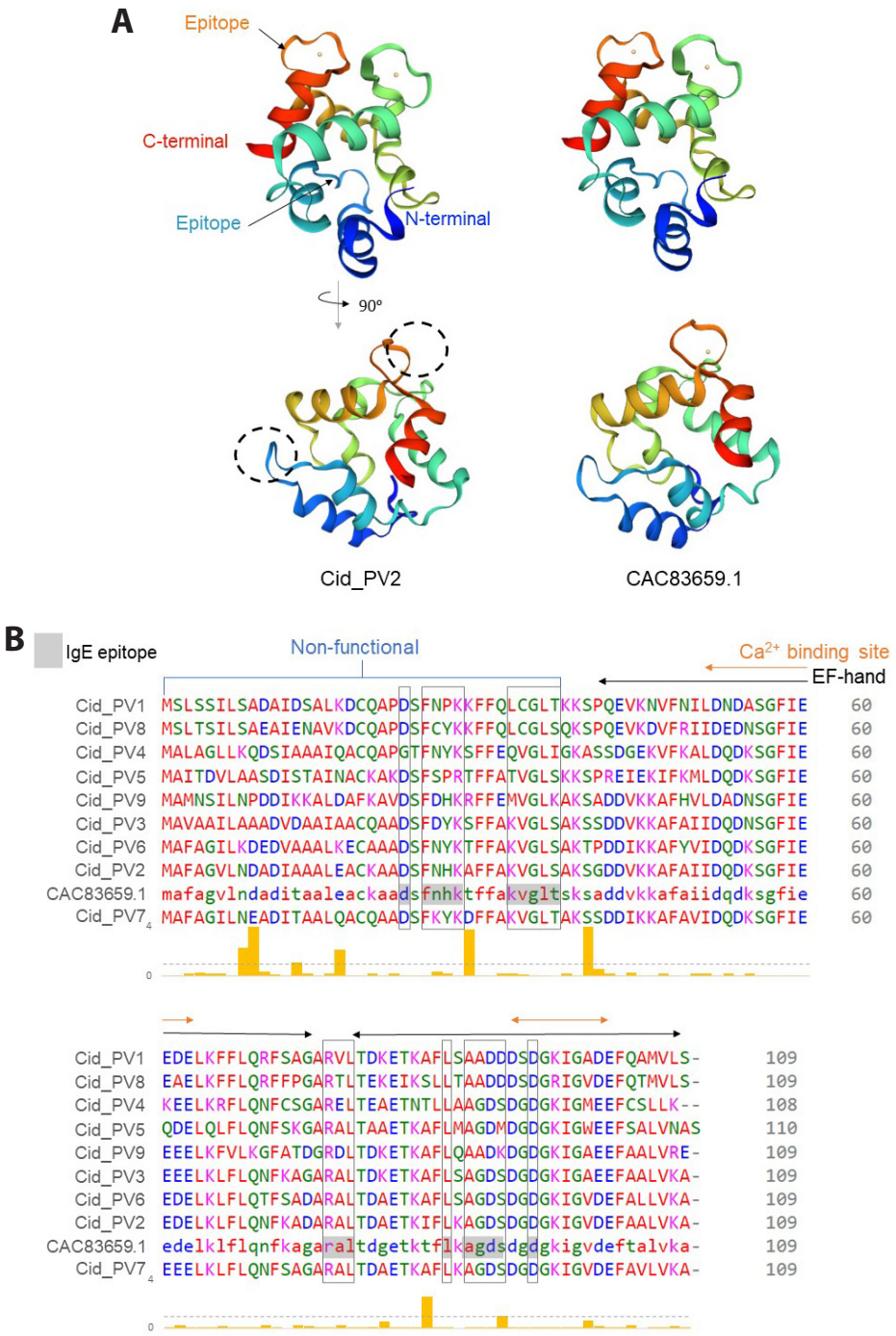


Figure 2. Alignment of parvalbumin homologs with common carp parvalbumin. (A) Protein structures of parvalbumins were constructed using SWISS-MODEL. Two IgE epitopes were located at the non-functional part near the N-terminal and next to the Ca2+ binding site near the C-terminal (circled by dotted line). Nine parvalbumin homologs from (B) *C. idella* were aligned with reported allergen Cyp c 1 (GenBank accession number: CAC83659.1). Shaded areas referred to the IgE epitopes of Cyp c 1 and the corresponding regions of IgE epitopes were marked in square. Based on the alignment, sequence variations among the homologs occurred at non-function part of the protein, while the sequences encode functional EF-hand motif were conserved. The bar plot indicated the dN/dS value of each amino acid and positive selection was represented by dN/dS > 1 (marked with dotted line). (C) Ten parvalbumin homologs from *H. nobilis* (left) and *M. piceus* (right) were aligned with reported allergen Cyp c 1.

C

```
Mpi_PV1     MSLTSILSAEAIENAVKDCQAPDSFSYKKFFQLCGLSQKSPQEVKDVFRIIDEDNSGFIE  60
Mpi_PV8     MSLSSILSADAIDSALKDCQAPDSFNPKKFFQLCGLTKKSPQEVKNVFNILDNDASGFIE  60
Mpi_PV10    MAMNSILNPDDIKKALDAFKAVDSFDHKRFFEMVGLKAKSADDVKKAFHVLDADNSGFIE  60
Mpi_PV9     MAMKNLLKDDDIKKAIDQFKAVDSFDHKKFFDVVGLKALSADNVKLVFKALDVDASGFIE  60
Mpi_PV5     MALAGLLKQDSIAAAIQACQAPGTFNYKSFFEQVGLIGKASSDGEKVFKALDQDKSGFIE  60
Mpi_PV4     MAITDVLAASDISTAINACKAKDSFSPRTFFATIGLSKKSPREIEKIFKMLDQDKSGFIE  60
Mpi_PV6     MAVAAMLAAADVDAAIAACQAADSFDYKSFFAKVGLSAKSSDDVKKAFAIIDQDNSGFIE  60
Mpi_PV3     MAFAGILKDEDVAAALKECAAADSFNYKTFFAKVGLSAKTPDDIKKAFYVIDQDKSGFIE  60
Mpi_PV2     MAFAGILNEADITAALQACQAADSFKYKDFFAKVGLSAKSPDDIKKAFAVIDQDKSGFIE  60
Mpi_PV7     MAFAGVLNDADIAAALEACKAADSFNHKAFFAKVGLSAKSGDDVKKAFAIIDQDKSGFIE  60
CAC83659.1  mafagvlndaditaaleackaadsfnhktffakvgltsksaddvkkafaiidqdksgfie  60

Mpi_PV1     EAELKFFLQRFFPGARTLTEKEIKSLLTAADDDSDGRIGVDEFQTMVLS-      109
Mpi_PV8     EDELKFFLQRFSAGARVLTDKETKAFLSAADDDSDGKIGADEFQAMVLS-      109
Mpi_PV10    EEELKFVLKGFATDGRDLTDKETKAFLQAADKDGDGKIGAEEFAALVRE-      109
Mpi_PV9     EEELKFVLKGFSADGRDLTDKETKAFLAAADKDGDGKIGIDEFEALVHE-      109
Mpi_PV5     KEELKRFLQNFCSGARELTEAETNTLLAAGDSDGDGKIGMEEFCSLLK--      108
Mpi_PV4     QDELQLFLQNFSKGARALTAAETKAFLMAGDMDGDGKIGWEEFSALVNAS      110
Mpi_PV6     EEELKLFLQNFKAGARALTDKETKAFLAAGDSDGDGKIGVDEFAALVKA-      109
Mpi_PV3     EDELKLFLQTFSASARALTDAETKAFLSAGDSDGDGKIGVDEFALLVKA-      109
Mpi_PV2     EDELKLFLQNFSAGARALTDAETKAFLKAGDSDGDGKIGVDEFAVLVKA-      109
Mpi_PV7     EDELKLFLQNFKAGARALTDAETKTFLKAGDSDGDGKIGVDEFAALVKA-      109
CAC83659.1  edelklflqnfkagaraltdgetktflkagdsdgdgkigvdeftalvka-      109

Hno_PV4     MSLTSILSAEAIENAVKDCQAPDSFCYKKFFQLCGLSQKSPQEVKDVFRIIDEDNSGFIE  60
Hno_PV8     MSLSSILSADAIDSALKDCQAPESFNPKKFFQLCGLTKKSPQEVKNVFNILDNDASGFIE  60
Hno_PV9     MAMNSILNPDDIKKALDAFKAVDSFDHKRFFEMVGLKAKSADDVKKAFHVLDADNSGFIE  60
Hno_PV10    MAMKNLLKDDDIKKAIDQFKAADSFDHKKFFDVVGLKALSADNVKLVFKALDVDASGFIE  60
Hno_PV5     MALAGLLKQDSIAAAIQACQAPGTFNYKSFFEQVGLIGKASSDGEKVFKALDQDKSGFIE  60
Hno_PV1     MAITDVLAASDISTAINACKAKDSFSPRTFFATVGLSKKSPREIEKIFKMLDQDKSGFIE  60
Hno_PV6     MAVAAMLAAADVDAAIAACQAADSFDYKSFFAKVGLSAKSSDDVKKAFAIIDQDNSGFIE  60
Hno_PV7     MAFAGVLNDADIAAALEACKDADSFNHKAFFAKVGLSAKSGDDVKKAFAIIDQDKSGFIE  60
CAC83659.1  mafagvlndaditaaleackaadsfnhktffakvgltsksaddvkkafaiidqdksgfie  60
Hno_PV2     MAFAGILKDDDVAAALKECSAADSFNYKSFFAKVGLTAKTPDDIKKAFYVIDQDKSGFIE  60
Hno_PV3     MAFAGILNEADVTAALQACQAADSFKYKDFFAKVGLSAKSPDDIKKAFAVIDQDKSGFIE  60

Hno_PV4     EAELKFFLQRFFPGARTLTEKEIKSLLTAADDDSDGRIGVDEFQTMVLS-      109
Hno_PV8     EDELKFFLQRFSAGARVLTDKETKAFLSAADDDSDGKIGADEFQAMVLS-      109
Hno_PV9     EEELKFVLKGFATDGRDLTDKETKAFLQAADKDGDGKIGAEEFAALVRE-      109
Hno_PV10    EEELKFVLKGFSADGRDLTDKETKAFLAAADKDGDGKIGIDEFEALVHE-      109
Hno_PV5     KEELKRFLQNFCSGARELTEAETNTLLAAGDSDGDGKIGMEEFCSLLK--      108
Hno_PV1     QDELQLFLQNFSKGARALTAAETKAFLMAGDMDGDGKIGWEEFSALVNAS      110
Hno_PV6     EEELKLFLQNFKAGARALTDKETKAFLSAGDSDGDGKIGVDEFAALVKA-      109
Hno_PV7     EDELKLFLQNFKAGARALTDAETKIFLKAGDSDGDGKIGVDEFAALVKA-      109
CAC83659.1  edelklflqnfkagaraltdgetktflkagdsdgdgkigvdeftalvka-      109
Hno_PV2     EDELKLFLQNFSAGARALTDAETKAFLSAGDSDGDGKIGVDEFALLVKA-      109
Hno_PV3     EDELKLFLQNFSAGARALTDAETKAFLKAGDSDGDGKIGVDEFAVLVKA-      109
```

Figure 2. (Continued)

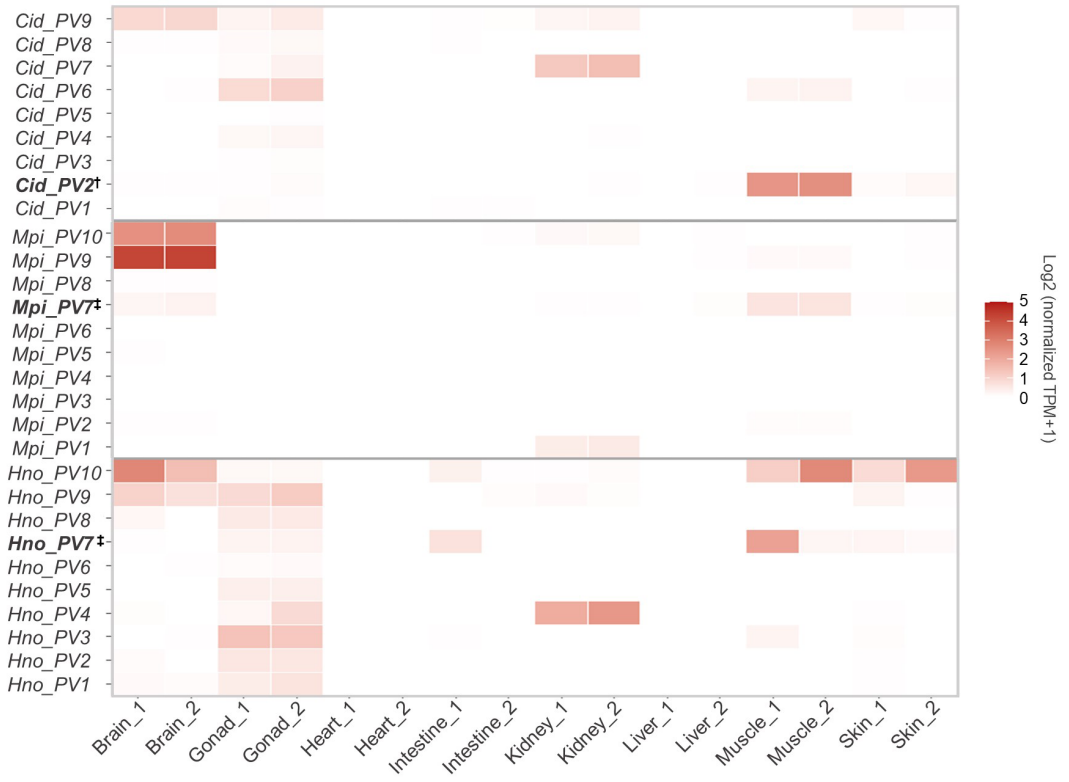


Figure 3. Gene expression levels of parvalbumin homologs in different organs among carps. The upper, middle and lower panels referred to *C. idella*, *M. piceus* and *H. nobilis* respectively. The expression level was calculated by Log2 (normalized TPM + 1) and TPM was normalized by the TPM level of GAPDH. *Cid_PV2* (protein sequence exhibited highest sequence similarity of 99.083% to the reported allergen, Cten i 1) was the dominant form in *C. idella* and was highly expressed in muscle. However, the expression of homologs (*Mpi_PV7* and *Hno_PV8*) with protein sequences similar to Cten i 1 were not highly expressed in the muscle of *M. piceus* and *H. nobilis*.

We also investigated the expression profile of parvalbumin homologs in different tissues of the selected fish species (**Figure 3**). The results revealed alternative usage of homologs in different tissues, suggesting tissue-specific expression patterns of parvalbumin genes. Of particular interest, the allergic response triggered by consumption of grass carp could be attributed to high Cid_PV2 expression in its muscle. However, the putative allergenic parvalbumins *Mpi_PV7* and *Hno_PV8* were not highly expressed in the muscle tissue of *M. piceus* and *H. nobilis*, respectively. This suggests that these species could potentially serve as substitutes for individuals who are allergic to Cid_PV2 in the muscle tissue of *C. idella*. However, further comparative studies on the allergenicity of these species are needed. Notably, there were three homologs expressed in the muscle tissue of *M. piceus* (*Mpi_PV2*, *Mpi_PV7*, and *Mpi_PV10*) and *H. nobilis* (*Hno_PV3*, *Hno_PV7*, and *Hno_PV10*), despite some of the expressions being low. However, there were only two homologs (*Cid_PV2* and *Cid_PV6*) expressed in the muscle tissue of *C. idella*. This suggested that *C. idella* may be lacking an α-subtype parvalbumin in muscle, as indicated by the clustering in **Figure 4**. Furthermore, we observed tissue-specific expression patterns of parvalbumin homologs in other organs, such as the brain and kidney. Specifically, the brain tissue of *C. idella* only utilized *Cid_PV9* (α-subtype) rather than other homologs of parvalbumin, and *Cid_PV7* was relatively highly expressed in the kidney of *C. idella*. These findings highlight the subfunctionalization of parvalbumin homologs in Asian carps and the tissue specificity in terms of the expression level.

Evolution and divergence of parvalbumin

We identified all parvalbumin homologs in the selected fish species and determined which homolog corresponded to the reported allergenic form (**Table S4**). To achieve this, we aligned a total of 109 parvalbumin protein sequences from Teleost, as well as bowfin *Amia (A.) calva* and coelacanth *Latimeria (L.) chalumnae* species under Osteichthyes. The amino acid sequences were extracted from the proteome of each species, and they were divided into three clades, as shown in the phylogenetic tree (**Figure 4a**). It is important to note that both the α- and β-subtypes of parvalbumin are found in Teleost fish. However, the β-subtype is the predominant form in fish, whereas the α-subtype is more commonly observed in cartilaginous fish and other vertebrates.[10,24,25,37] Interestingly, nearly all reported allergenic parvalbumins were clustered in the most diverse clade (blue), composed of β-parvalbumins, except for Phy_PV9. The blue clade was further divided into three gradients (light blue, blue, and deep blue), with the allergens concentrated in the light blue clade.

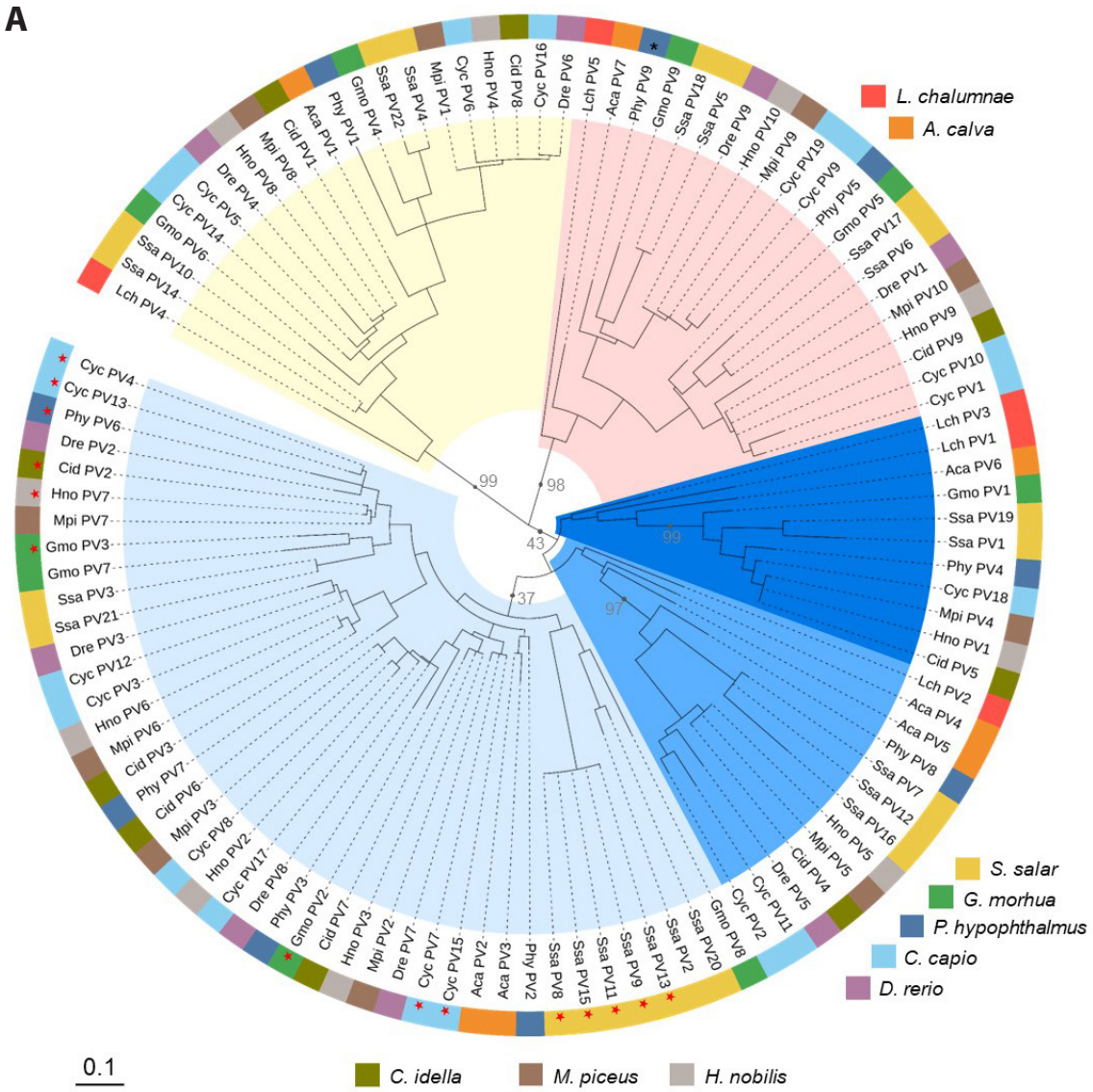


Figure 4. Phylogenetic tree and gene synteny of parvalbumins in Osteichthyes. (A) A total of 109 protein sequences of parvalbumin from ten species in Osteichthyes were aligned with MUSCLE, and the tree was generated by maximum likelihood algorithm (bootstrap = 500) based on the alignment result. All parvalbumins were divided into three clades included α-like (red), thymic CPV3-like (yellow) and β-like (blue) subtypes. According to the BLAST result in **Table S1**, reported and putative allergenic parvalbumins were marked with red asterisk. The clustering analysis revealed the allergenic were clustered in the most diverse clade (blue) except Phy_PV9. (B) The position of each gene was illustrated in coloured arrows, and the color was corresponded to the clade color shown in the phylogenetic tree of parvalbumins. Two gene clusters of parvalbumin were found in the genome of *C. idella*, *M. piceus* and *H. nobilis*. The distances between genes were less than 5 kb within each cluster and the genes of reported allergenic parvalbumin were marked in red.

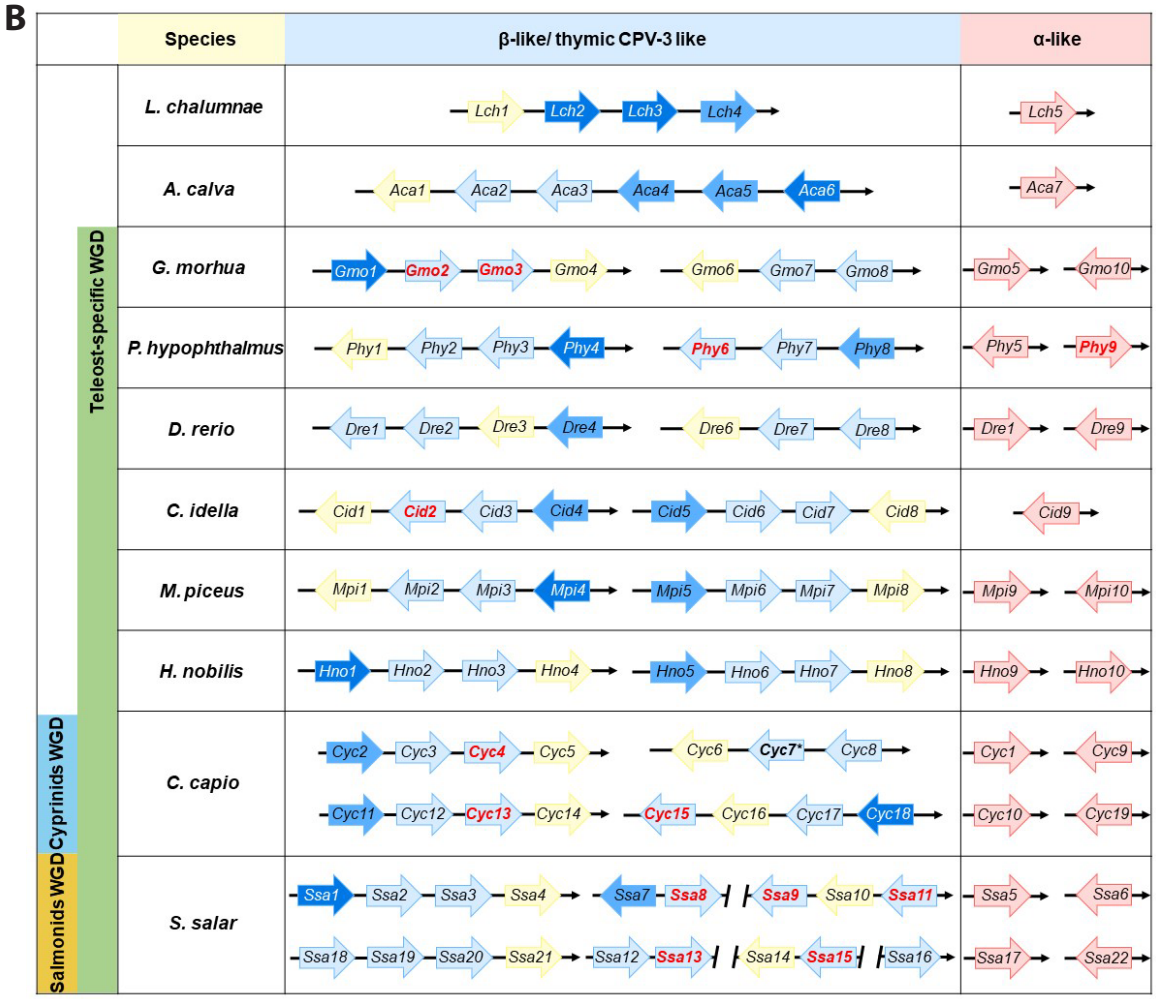


Figure 4. (Continued)

Most importantly, the clade in light blue marked an expansion of genes that mainly occurred in Teleost. This suggests that the expansion of parvalbumin genes in Teleost may have contributed to the emergence of allergenic parvalbumin homologs in these species.

In addition, we investigated the gene synteny of parvalbumin homologs across different fish species. In particular, we examined the early gene arrangement in *L. chalumnae* and *A. calva*, as well as the changes that occurred after the Teleost-specific whole-genome duplication (WGD). We found that in *L. chalumnae*, the parvalbumin genes were arranged in a cluster with one thymic CPV3-like subtype (yellow) and three β-subtype (blue) parvalbumins, as well as a single copy of the α-subtype (red) parvalbumin (**Figure 3B**). This pattern was preserved in *A. calva*, with the addition of two extra copies of the β-like subtype (blue) parvalbumins. After the Teleost-specific WGD, the gene copies of parvalbumin nearly doubled in Teleosts, and more copies of β-parvalbumins existed in the light blue clade and tandem arrayed gene clusters. These results suggested that the emergence of allergenic parvalbumins coincided with the gene duplication event. Moreover, we found that the allergenic parvalbumins were more closely related among species and recently evolved by gene duplication. These findings provide insights into the significant role of gene duplication in the molecular mechanisms underlying the allergenicity of parvalbumin homologs.

The investigation of parvalbumin evolution was further extended to Gnathostomata, including both Osteichthyes and Chondrichthyes. The phylogenetic tree revealed that three out of six homologs in the great white shark *Carcharodon (C.) carcharias* were excluded from other homologs in other species, and those homologs were phylogenetically distinct from the α- and β-parvalbumin homologs. Interestingly, only one homolog in *C. carcharias* belonged to the β-subtype (**Figure S2A**). This further confirmed that the expansion of β-subtype was preferentially occurred in Osteichthyes, and subsequently expanded in Teleost. Moreover, our analysis identified a single copy of parvalbumin in the sea lamprey *Petromyzon (P.) marinus*, which is a jawless vertebrate. This parvalbumin gene contained three shared intron sites, providing evidence that parvalbumin is a conserved protein in vertebrates (**Figure S2B**). Even though parvalbumin was thought to be a protein in vertebrates, two potential sequences of parvalbumin were found in amphioxus *Branchiostoma (B.) floridae*, which is a jawless vertebrate. The sequences were aligned with parvalbumins in *P. marinus* and *C. carcharias*,

and Bfl_PV2 was identified as homolog of parvalbumin with two shared intron sites (**Figure S2B**). Also, the homologs were further confirmed by reciprocal BLAST using the protein sequences from *P. marinus* and *B. floridae* to search the closest match in the genome of *C. carcharias*. Overall, our analysis suggests that the existence of a single parvalbumin gene can be traced back to chordate and the allergenic homologs arose later in teleost by gene duplication and diversification.

Discussion

Four domestic fish species are widely farmed and consumed in China. For our study, we focused on three Asian carp species, *C. idella*, *M. piceus*, and *H. nobilis*, selected based on their prevalence in local markets, potential allergenicity, and dietary significance in the region. Genomic approaches provide a comprehensive allergen profile by retrieving allergen transcripts and amino acid sequences, enabling the identification of potential allergens. Despite thorough cooking, parvalbumin in Asian carps remains allergenic. Our study identified putative allergenic parvalbumins in *M. piceus* and *H. nobilis*, expressed in muscle tissue with conserved IgE epitopes. However, further research is needed to confirm allergenicity and investigate potential cross-reactivity, advancing diagnostics and therapies for individuals with multiple fish allergies.

Previously, Leung et al discovered two isoforms of parvalbumin in *C. idella* and the major IgE-binding parvalbumin was 9 kDa, while another isoform with 11 kDa was found to be non-allergenic.[15] In fact, nine homologs of parvalbumin existed in *C. idella* instead of two isoforms in the genome. Experiments done by Leung et al was based on the protein extracted from muscle and we believed that the allergenic isoforms were corresponded to Cid_PV2 in our result. The presence of isoforms in parvalbumin among different fish species had been reported and antibody reactivity to parvalbumin would be affected by isoforms.[26-28] But we did not observe alternative splicing in gene sequences and different "isoforms" or forms of parvalbumin could be attributed to the existence of homologs in the genome. Coupled with the transcriptome data, the expression profile also revealed that the allergenic parvalbumin Cid_PV2 was highly expressed in muscle. Hence, our data supported their findings with more accurate details of homologs and allergen expression in various tissues. Indifferent with Asian carps, common carp undergone cyprinids whole genome duplication (WGD) attributed to a double of parvalbumin genes in its genome and there were more homologs (two α, two β1 and eight β2) expressed in muscle.[29] In contrast, only two parvalbumin β were expressed in the muscle of grass carp, and one extra parvalbumin α was found in both bighead carp and black carp. While parvalbumin α was predominantly expressed in brain and gonad in all Asian carps mentioned, α-subtype could be found in most common carp's tissues as more α gene copies were available. Taken together, the effect of gene duplication in homolog usage was revealed in accordance with previous results.

A collection of sequences of parvalbumins from each species can be retrieved along with their corresponding genomes, which enabled us to investigate the evolution and origin of allergens. The selection of allergens cannot be explained solely by their biological functions, and the emergence of allergenic parvalbumins is believed to be the result of gene duplication and divergence.[30] Of particular interest is the case of *S. salar* (Atlantic salmon), which has the highest number of parvalbumin gene copies due to salmon specific WGD. Out of the 22 parvalbumin sequences identified in *S. salar*, five homologs corresponded to the reported allergenic parvalbumin (Sal s 1) in the WHO/IUIS allergen database. However, despite having a greater number of gene copies of allergenic parvalbumin, the allergenicity of Sal s 1 was found to be lower than that of Cten i 1 in terms of IgE reactivity. These findings suggest that the allergenicity of parvalbumin may not be directly related to the quantity of gene copies, and the structure and sequence of IgE epitopes may also play a crucial role. Further investigation is necessary to gain a deeper understanding of the structural and immunological aspects of these allergens. Specifically, future studies could focus on characterizing the structural IgE epitopes to elucidate the mechanisms underlying their allergenicity.

Furthermore, this study also provided insight on the evolution of parvalbumin in vertebrates. Similar to the previous finding, we discovered three groups of parvalbumins in teleost and the gene number of β-subtype was expanded after WGD.[29,31] Based on the sequence homology with BLAST algorithm, we described one group as thymic CPV3-like subtype which reported to be found in lower vertebrates and related to oncomodulin (a type of β-parvalbumin) in mammals.[32] The ancestral parvalbumin can be found in *B. floridae* with one copy in the genome. The number of genes increased to six in *C. carcharias*, but only one homolog was closely related to β-parvalbumin. The gene expansion of β-parvalbumin emerged in Osteichthyes with three copies of tandem arrayed β-parvalbumin genes in *L. chalumnae*. Besides teleost, the number of β-parvalbumin also increased in the genome of *A. calva*. This reflects the importance of duplication and divergence of β-parvalbumin for survival not only in teleost but also in other species under Osteichthyes, and incorporation of β-parvalbumin as well as usage of calcium ions in fast-moving muscle is essential for local movement of fishes in aquatic environments.

Currently, the clinical tests for fish allergy are primarily based on commonly consumed European and seawater fish species, such as Atlantic cod (*G. morhua*) and salmon (*S. salar*), as well as freshwater species, including common carp (*C. carpio*). In contrast, Asian carps are very common fish species in Asia, but they are seldomly served as food in western countries. Due to the cultural difference, a more suitable reference species could be used in Asian countries for fish allergy testing. *C. idella* could be a potential candidate for allergy testing, which its parvalbumin is reported to be more allergenic than other fishes,

and a more precise test could be provided based on individual proteins rather than the whole fish extract. For instance, recombinant *C. idella* parvalbumin could be used for testing the level of serum specific IgE level[33] and to enhance the accuracy of serum specific IgE test. With the clear spectrum of fish parvalbumins we described, more accurate gene sequences can be retrieved from the genome. A high-resolution and species-specific allergen test based on a single protein (i.e., parvalbumin) can be achieved. The identification of additional parvalbumin homologs can have significant implications for the improvement of diagnostics and care for patients with fish allergies. For instance, other parvalbumin homologs such as Cid_PV6 were also identified in muscle tissue and were found to share the same epitope sequences as Cid_PV2. These homologs have the potential to serve as candidates for diagnostics and immunotherapy. Moreover, a comprehensive parvalbumin expression profile was provided for patients to select fish species with low allergic parvalbumin expression. Ultimately, it is important to subject patients to the gold standard of double-blind, placebo-controlled fish challenge[34] to ascertain their allergy status so as to accurately define the diagnostic values of different carp allergens.

Author Contributions

- Judy Kin Wing NG performed the experimental works, data curation, interpretation of the results and wrote the manuscript.
- Stephen Kwok Wing Tsui designed the study.
- Qing XIONG reviewed and edited the manuscript.
- Ling SHI, Christine Yee Yan WAI, Soo Kyung SHIN, Fu Kiu AO contributed to data collection and experimental works.
- Agnes Sze Yin LEUNG, Nicki Yat Hin LEUNG, Ting Fan LEUNG supported conceptualization of the study.
- All authors read and approved the final manuscript.

Acknowledgment and Funding

This project was supported by the General Research Funds (Ref. No.: 14119420 and 14121222) of the Hong Kong Research Grants Council and Innovation and Technology Fund (Ref. No.: ITS/082/17) of Hong Kong SAR Government.

Availability of data and Ethics approval

The genomes have been deposited under NCBI BioProject PRJNA890423, PRJNA892279 and PRJNA891927. This study was approved by Joint Chinese University of Hong Kong-New Territories East Cluster Clinical Research Ethics Committee (Reference no. 2019.612) and written informed consent was obtained from all individual participants and/or their parents.

Authors' consent for publication

All the authors approved the manuscript and gave their consent for submission and publication.

Competing Interests

Christine Yee Yan WAI, Agnes Sze Yin LEUNG, Nicki Yat Hin LEUNG and Ting Fan LEUNG published a Non-Provisional Patent under publication no. US2020/0191797 A1 on 18 Jun 2020. Other authors have no conflict of interest to declare.

References

1. Kalic T, Radauer C, Lopata AL, Breiteneder H, Hafner C. Fish allergy around the world-precise diagnosis to facilitate patient management. Front Allergy. 2021;2:732178.
2. Kourani E, Corazza F, Michel O, Doyen V. What do we know about fish allergy at the end of the decade? J Investig Allergol Clin Immunol. 2019;29:414-21.
3. Leung TF, Yung E, Wong YS, Lam CW, Wong GW. Parent-reported adverse food reactions in Hong Kong Chinese pre-schoolers: epidemiology, clinical spectrum and risk factors. Pediatr Allergy Immunol. 2009;20:339-46.
4. Leung ASY, Li RMY, Au AWS, Rosa Duque JS, Ho PK, Chua GT, et al. Changing pattern of pediatric anaphylaxis in Hong Kong, 2010–2019. Pediatr Allergy Immunol. 2022;33:e13685.
5. Goodman RE, Breiteneder H. The WHO/IUIS Allergen Nomenclature. Allergy. 2019;74:429-31.
6. Aas K, Elsayed SM. Characterization of a major allergen (cod). Effect of enzymic hydrolysis on the allergenic activity. J Allergy. 1969;44:333-43.
7. Van Do T, Elsayed S, Florvaag E, Hordvik I, Endresen C. Allergy to fish parvalbumins: Studies on the cross-reactivity of allergens from 9 commonly consumed fish. J Allergy and Clini Immunol. 2005;116:1314-20.
8. Liang J, Taylor SL, Baumert J, Lopata AL, Lee NA. Effects of thermal treatment on the immunoreactivity and quantification of parvalbumin from Southern hemisphere fish species with two anti-parvalbumin antibodies. Food Control. 2021;121:107675.
9. Freidl R, Gstöttner A, Baranyi U, Swoboda I, Stolz F, Focke-Tejkl M, et al. Resistance of parvalbumin to gastrointestinal digestion is required for profound and long-lasting prophylactic oral tolerance. Allergy. 2020;75:326-35.
10. Celio M, Heizmann C. Calcium-binding protein parvalbumin is associated with fast contracting muscle fibres. Nature. 1982;297:504-6.
11. Kobayashi A, Tanaka H, Hamada Y, Ishizaki S, Nagashima Y, Shiomi K. Comparison of allergenicity and allergens between fish white and dark muscles. Allergy. 2006;61:357-63.
12. Mohan Dey M, Rab MA, Paraguas FJ, Piumsombun S, Bhatta R, Ferdous Alam M, et al. Fish consumption and food security: a disaggregated analysis by types of fish and classes of consumers IN SELECTED ASIAN COUNTRIES. Aquac Econ Manag. 2005;9:89-111.
13. Average daily consumption of fresh fish in Hong Kong from 2015 to 2021, by type (in metric tons) [database on the Internet]. AFCD. 2022.
14. Kočovský PM, Chapman DC, Qian S. "Asian carp" is societally and scientifically problematic. Let's replace it. Fisheries. 2018;43:311-6.
15. Leung NYH, Leung ASY, Xu KJY, Wai CYY, Lam CY, Wong GWK, et al. Molecular and immunological characterization of grass carp (Ctenopharyngodon idella) parvalbumin Cten i 1: A major fish allergen in Hong Kong. Pediatr Allergy Immunol.. 2020;31:792-804.
16. Ruethers T, Taki AC, Nugraha R, Cao TT, Koeberl M, Kamath SD, et al. Variability of allergens in commercial fish extracts for skin prick testing. Allergy. 2019;74:1352-63.
17. Tong WS, Yuen AW, Wai CY, Leung NY, Chu KH, Leung PS. Diagnosis of fish and shellfish allergies. J Asthma Allergy. 2018;11:247-60.
18. Cates MS, Teodoro ML, Phillips GN, Jr. Molecular mechanisms of calcium and magnesium binding to parvalbumin. Biophys J. 2002;82:1133-46.
19. Erickson JR, Moerland TS. Functional characterization of parvalbumin from the Arctic cod (Boreogadus saida): similarity in calcium affinity among parvalbumins from polar teleosts. Comp Biochem Physiol A-Mol Integr Physiol. 2006;143:228-33.
20. Perez-Gordo M, Pastor-Vargas C, Lin J, Bardina L, Cases B, Ibáñez MD, et al. Epitope mapping of the major allergen from Atlantic cod in Spanish population reveals different IgE-binding patterns. Mol Nutr Food Res. 2013;57:1283-90.

21. Perez-Gordo M, Lin J, Bardina L, Pastor-Vargas C, Cases B, Vivanco F, et al. Epitope Mapping of Atlantic Salmon Major Allergen by Peptide Microarray Immunoassay. Int Arch Allergy Immunol. 2012;157:31-40.
22. Kumeta H, Nakayama H, Ogura K. Solution structure of the major fish allergen parvalbumin Sco j 1 derived from the Pacific mackerel. Sci Rep. 2017;7:17160.
23. Untersmayr E, Szalai K, Riemer AB, Hemmer W, Swoboda I, Hantusch B, et al. Mimotopes identify conformational epitopes on parvalbumin, the major fish allergen. Mol Immunol. 2006;43:1454-61.
24. Cowan RL, Wilson CJ, Emson PC, Heizmann CW. Parvalbumin-containing GABAergic interneurons in the rat neostriatum. J Comp Neurol. 1990;302:197-205.
25. Kuehn A, Swoboda I, Arumugam K, Hilger C, Hentges F. Fish allergens at a glance: variable allergenicity of parvalbumins, the major fish allergens. Front immunol. 2014;5:179-.
26. Saptarshi SR, Sharp MF, Kamath SD, Lopata AL. Antibody reactivity to the major fish allergen parvalbumin is determined by isoforms and impact of thermal processing. Food chem. 2014;148:321-8.
27. Van Do T, Hordvik I, Endresen C, Elsayed S. The major allergen (parvalbumin) of codfish is encoded by at least two isotypic genes: cDNA cloning, expression and antibody binding of the recombinant allergens. Mol immunol. 2003;39:595-602.
28. Friedberg F. Parvalbumin isoforms in zebrafish. Mol biol rep. 2005;32: 167-75.
29. Mukherjee S, Bartoš O, Zdeňková K, Hanák P, Horká P, Musilova Z. Evolution of the parvalbumin genes in teleost fishes after the whole-genome duplication. Fishes. 2021;6:70.
30. Taylor JS, Raes J. Duplication and divergence: the evolution of new genes and old ideas. Annu Rev Genet. 2004;38:615-43.
31. Dijkstra JM, Kondo Y. Comprehensive sequence analysis of parvalbumins in fish and their comparison with parvalbumins in tetrapod species. Biology. 2022;11:1713.
32. Climer LK, Cox AM, Reynolds TJ, Simmons DD. Oncomodulin: The Enigmatic Parvalbumin Protein. Front Mol Neurosci. 2019;12:
33. Swoboda I, Bugajska-Schretter A, Verdino P, Keller W, Sperr WR, Valent P, et al. Recombinant carp parvalbumin, the major cross-reactive fish allergen: A tool for diagnosis and therapy of fish allergy. Journal Immunol. 2002;168:4576-84.
34. Leung ASY, Leung NYH, Wai CYY, Xu KJY, Lam MCY, Shum YY, et al. Characteristics of Chinese fish-allergic patients: Findings from double-blind placebo-controlled food challenges. J Allergy Clin Immunol Pract. 2020;8:2098-100.e8.

Supplementary materials

Materials and methods

Genomic DNA extraction and sequencing

Three Asian carps species included *C. idella*, *M. piceus*, and *H. nobilis* were obtained from Hong Kong local wet market. The fish was euthanized by 100 mg/L tricaine methanesulfonate (TMS) and sacrificed by cervical dislocation. Genomic DNA (gDNA) was extracted from 5 g of muscle tissue. Homogenized muscle tissue was subjected to DNA extraction using MagAttract HMW DNA Kit (Qiagen, Germany). DNA was extracted by the protocol provided by the manufacturer. DNA concentration and absorbance were measured by Qubit and Nanodrop (Thermo Fisher Scientific, USA) respectively. Size of extracted genomic DNA was determined by gel electrophoresis. For the extracted DNA, single-molecule real-time (SMRT) genomic libraries (10-20 kbp in length) were constructed with 1 μg of gDNA as input and SMRT sequence data were generated using the Oxford Nanopore Sequencing platform (GridION Mk1) until a coverage of over 30X was achieved. The gDNA was also subjected to short read sequencing in Groken Bioscience (Hong Kong) to generate paired-end 150-bp reads by DNBseq platform.

De novo assembly and annotation

Sequencing data generated by Oxford Nanopore GridION platform was trimmed out by Porechop v0.2.4.[1] Low quality reads were filtered out by NanoFilt v2.7.1.[2] The minimum read length of 500 and minimum read quality score of 7 were used in the read filtering step. Filter reads were assembled by Flye v2.8.2[3] based on repeat graph method to obtain a draft genome. The genome sequence was further polished by Pilon v1.23[4] with Illumina DNBseq data. Scaffolding was performed after the genome was polished. SSPACE-longread v 1.1[5] was designed for long reads data which joined contigs into larger scaffolds. The completeness and contiguity of genome was assessed by Benchmarking Universal Single-Copy Orthologs (BUSCO) v3.1.0[6] and QUAST v5.0.2.[7] The repetitive sequences were masked by RepeatModeler v2.0.1[8] and RepeatMasker v4.0.8.[9] Prediction of repeat family was done by RECON v1.05[10] and RepeatScout v1.0.6[11] that build in the RepeatModeler. Based on the constructed genome sequence and aligned transcriptome sequence, annotation of genome was done by Maker v2.3.2[12] pipeline. Gene prediction was carried out by SNAP v 18.04.6 LTS,[13] Augustus v3.3.1[14] and GeneMark v4.38,[15] and the prediction results were used as input for Maker pipeline together with aligned transcriptome sequence and protein sequences from related species. The gene annotation completeness was accessed by BUSCO v3.1.0. The genomes have been deposited under NCBI BioProject PRJNA890423, PRJNA892279 and PRJNA891927.

Phylogenomic analysis of bony fish

The genome, transcriptome, and protein sequences of 8 species including sea lamprey *Petromyzon (P.) marinus*, great white shark *Carcharodon (C.) carcharias*, coelacanth *Latimeria (L.) chalumnae*, bowfin *Amia (A.) calva*, Atlantic cod *Gadus (G.) morhua*, Atlantic salmon *Salmo (S.) salar*, striped catfish *Pangasianodon (P.) hypophthalmus*, zebrafish *Danio (D.) rerio* and common carp *Cyprinus (C.) capio* were downloaded from NCBI database (**Table S2**). To understand the evolutionary relationship of jawed vertebrates as well as bony fish, gene sequences of 315 overlapped BUSCO genes from 11 species based on vertebrata_odb9 database were extracted. The codon sequences were aligned by MAFFT v7.310 and transformed to PHYLIP format by online tool MABL http://phylogeny.lirmm.fr/phylo_cgi/data_converter.cgi. The divergence time of species was calculated by MCMCTree in PAML package v4.9.

In silico identification of allergens

For the parvalbumins in selected Asian carps, it was identified by NCBI BLAST Toolkit with TBLASTN algorithm by matching the reported *C. idella* parvalbumin (GenBank accession number: QCY53440.1) to the genome sequence. While the putative allergens were identified by TBLASTN algorithm by matching the reported protein sequences from related species to the genome sequence. The Evalue cutoff was 1e-6 in applying the BLAST algorithm. All the reported allergens can be found on the World Health Organization (WHO) allergen database released on the webpage of WHO/IUIS Allergen Nomenclature http://allergen.org/. The protein sequence of reported allergens was downloaded from NCBI with the accession number provided in the WHO/IUIS allergen database.

RNA extraction and sequencing

Eight types of organs were selected (brain, gonad, heart, intestine, kidney, liver, muscle and skin) for RNA extraction. Tissue samples with TRIzol (Invitrogen, USA) were homogenized, 0.2 mL of chloroform was added to 1 mL of lysate and the mixture was centrifuged at 12,000 × g for 15 minutes at 4°C. Approximately 400 µL of upper phase (containing RNA) was transferred to a new 1.5 mL tube and an equal volume of 70% ethanol was added. RNA extracted in the upper phase was purified by PureLink RNA Mini kit (Thermo Fisher, USA). RNA concentration and absorbance were measured by Qubit and Nanodrop respectively. The quality of extracted RNA was assessed by Agilent 2100 Bioanalyzer (Agilent Technologies, USA) RNA sequencing was performed using Illumina HiSeq 2500 by Groken Bioscience (Hong Kong) to generate paired-end 150-bp reads.

Transcriptome analysis

Both reference-mapping and de novo assembly of transcriptome data were performed by Hisat2 v2.0.4[16] and Trinity v2.8.4[17] respectively. The mapping output from Hisat2 was converted to FASTA format by StringTie v1.3.0[18] and gffread v0.12.1.[19] The transcriptome data were combined and ready for gene annotation. For allergen expressions in various organs, the pair-end clean reads from each organ were processed by salmon v0.13.0[20] and the expression level of each gene was calculated in terms of transcripts per million (TPM).

Sequence alignment and phylogenetic analysis of gene family

The protein homologs of parvalbumin were identified by BLASTP algorithm to the protein sequence files from each species. The position of each homolog inside the genome was extracted from the gene feature format (gff) files and further confirmed by visualizing the genome in IGV v 2.9.2.[21] Faulty annotated genes were corrected by manual curation if mis-annotation by software occurred. The target sequences were put in a FASTA file, and the sequence alignment was done by MUSCLE algorithm in MEGA-X.[22] The alignment result was applied for maximum-likelihood tree construction using partial deletion mode with set coverage cutoff of > 90% and bootstrap = 500. The tree was visualized and edited in online tool iTOL https://itol.embl.de/.

Cloning and indirect ELISA of recombinant protein

The coding sequence (CDS) of *C. idella* parvalbumin (GenBank accession: OP787027), β-enolase (OP787028) and aldolase (OP787029) were retrieved from the genome. Synthesized CDS sequences were inserted into pET-30a+ vector and expressed in TOP10 *E. coil*. The expression of recombinant proteins was performed by Sangon Biotech (Shanghai, China). The allergenicity of recombinant proteins was tested by indirect enzyme-linked immunosorbent assay (ELISA) and serum samples from 20 fish ImmunoCAP positive patients and 14 healthy controls were used in the experiment (**Table S3**). Purified proteins (5 µg/ml) were coated onto 96-well microtiter plate in sodium bicarbonate coating buffer (100 mM, pH 9.6) and incubated at 37°C for 3 hours. The plate was washed by 0.05% Tween-20/PBS (PBST) for 3 times and blocked with 8% fetal bovine serum (FBS, Gibco) in PBS at room temperature for 2 hours. Serum samples were diluted 1:20 with blocking buffer and 50 µL of diluted serum was added to each well overnight at 4°C. After washing, horseradish peroxidase (HRP) conjugated anti-human IgE antibodies (Thermo Fisher, USA) at 1:1000 dilution was added and incubated at room temperature for 1 hour. The plate was than wash with PBST for five times and incubated with TMB-ELISA substrate (Abcam, UK) for color development. The reaction was stopped by adding 0.1 M sulphuric acid and the absorbance was recorded at 450 nm by microplate reader (Bio-Rad, USA). This study was approved by Clinical Research Ethics Committee (Reference no. 2019.612) and written informed consent was obtained from all individual participants and/or their parents.

Mitochondrial phylogeny

To confirm the identity of the target species, a mitochondrial phylogeny of 74 fish species under Cyprinidae was constructed. All the complete mitochondrial genomes of 74 species were downloaded from NCBI database. While the complete mitochondrial genome was extracted from the draft genome constructed and the annotation of genes was done by MITOS web server.[23] The gene sequence of Cox 1 was extracted and aligned with the sequences 74 mitochondrial genomes by MEGA-X[22] using MUSCLE[24] algorithm. Only the sequences aligned with Cox 1 were extracted for tree construction. The tree was generated by maximum likelihood algorithm (bootstrap = 100) based on the alignment result.

References

1. Wick RR, Judd LM, Gorrie CL, Holt KE. Completing bacterial genome assemblies with multiplex MinION sequencing. Microb Genom. 2017;3:e000132.
2. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics. 2018;34:2666-9.
3. Freire B, Ladra S, Parama JR. Memory-efficient assembly using Flye. IEEE/ACM Trans Comput Biol Bioinform. 2021;Pp:
4. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLOS ONE. 2014;9:e112963.
5. Boetzer M, Pirovano W. SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. BMC Bioinformatics. 2014;15:211.
6. Seppey M, Manni M, Zdobnov EM. BUSCO: Assessing genome assembly and annotation completeness. Methods Mol Biol. 2019;1962:227-45.
7. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013;29:1072-5.
8. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci. 2020;117:9451-7.
9. Tarailo-Graovac M, Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinformatics. 2009; Chapter 4:Unit 4.10.
10. Levitsky VG. RECON: a program for prediction of nucleosome formation potential. Nucleic Acids Res. 2004;32:W346-9.
11. Price AL, Jones NC, Pevzner PA. De novo identification of repeat families in large genomes. Bioinformatics. 2005;21 Suppl 1:i351-8.
12. Holt C, Yandell M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics. 2011;12:491.
13. Korf I. Gene finding in novel genomes. BMC Bioinformatics. 2004;5:59.
14. Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics. 2008;24:637-44.
15. Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M. Gene identification in novel eukaryotic genomes by self-training algorithm. Nucleic Acids Res. 2005;33:6494-506.
16. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015;12:357-60.
17. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29:644-52.
18. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015;33:290-5.
19. Pertea G, Pertea M. GFF Utilities: GffRead and GffCompare. F1000Res. 2020;9:
20. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. 2017;14:417-9.
21. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29:24-6.
22. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol. 2018;35:1547-9.
23. Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, et al. MITOS: Improved de novo metazoan mitochondrial genome annotation. Mol Phylogenet Evol. 2013;69:313-9.
24. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792-7.

Supplementary table 1. Genome assembly of Asian carps.

	C. idella	*M.piecus*	*H. nobilis*
Genome size (bp)	880,848,479	877,217,469	861,003,699
Scaffold number	1,949	3,348	2,001
Scaffold N50 (bp)	7,888,681	3,165,062	5,775,568
GC (%)	37.46	37.51	37.35
Largest scaffold	30,034,836	14,523,745	24,842,118
Genome completeness (%)	97.7	95.8	95.8
Annotation completeness (%)	87.9	90.5	88.9

Supplementary table 2. Genome data from NCBI GenBank.

Abbr.	Species	Common name	Genome size	Completeness[a]	Accession
Pma	*P. marinus*	Sea lamprey	1,089,050,413	70.4%/83.5%	GCF_010993605.1
Cca	*C. carcharias*	Great white shark	4,286,311,191	89.1%/98.3%	GCF_017639515.1
Lch	*L. chalumnae*	Coelacanth	2,860,591,921	87.4%/95.5%	GCF_000225785.1
Aca	*A. calva*	Bowfin	897,731,172	95.7%/87.1%	GCA_016984155.1
Ssa	*S. salar*	Atlantic salmon	3,412,113,583	92.6%/98.2%	GCF_000233375.1
Gmo	*G. morhua*	Atlantic cod	646,392,667	94.2%/97.9%	GCF_902167405.1
Phy	*P. hypophthalmus*	Striped catfish	758,973,678	95.6%/98.3%	GCF_009078355.1
Dre	*D. rerio*	Zebrafish	1,373,454,788	95.5%/99.0%	GCF_000002035.6
Cyc	*C. capio*	Common carp	1,680,134,903	96.8%/99.0%	GCF_018340385.1

[a]Genome completeness/ annotation completeness

Supplementary Table 3. Putative allergens identified in *C. idella* genome. The result was based on the protein sequence BLAST to the *C. idella* genome, and the identity of homolog with highest percentage matching was shown.

Allergen ID	Ref. GenBank ID	Biochemical name	Ref. species	Identity (%)	No. of homologs
Cten i 1	QCY53440	Beta-parvalbumin	*C. idella*	99.1	9
Cyp c 2	AWS00995.1	Beta-enolase	*C. carpio*	99.1	4
Pan h 3	XP_026771637.1	Aldolase A	*P. hypophthalumus*	93.1	5
Sal s 4	NP_001117128.1	Tropomyosin	*S. salar*	93.3	7
Sal s 6	XP_014059932.1	Collagen alpha	*S. salar*	89.2	5
Sal s 7	ACH70914.1	Creatine kinase	*S. salar*	92.7	8
Sal s 8	ACM09737.1	Triosephosphate isomerase	*S. salar*	86.5	2
Pan h 9	XP_026775867.1	Pyruvate kinase PKM-like	*P. hypophthalumus*	87.2	3
Pan h 10	XP_026774991.1	L-lactate dehydrogenase	*P. hypophthalumus*	93.1	5
Pan h 11	XP_026782721.1	Glucose 6-phosphate isomerase	*P. hypophthalumus*	92.6	2
Pan h 13	XP_026782131.1	Glyceraldehyde-3-phosphate dehydrogenase	*P. hypophthalumus*	95.2	2

Supplementary Table 4. Putative allergens identified in *H. nobilis* genome. The result was based on the protein sequence BLAST to the *H. nobilis* genome, and the identity of homolog with highest percentage matching was shown.

Allergen ID	Ref. GenBank ID	Biochemical name	Ref. species	Identity (%)	No. of homologs
Cten i 1	QCY53440	Beta-parvalbumin	*C. idella*	97.3	10
Cyp c 2	AWS00995.1	Beta-enolase	*C. carpio*	99.1	4
Pan h 3	XP_026771637.1	Aldolase A	*P. hypophthalumus*	93.1	5
Sal s 4	NP_001117128.1	Tropomyosin	*S. salar*	82.6	8
Sal s 6	XP_014059932.1	Collagen alpha	*S. salar*	88.3	5
Sal s 7	ACH70914.1	Creatine kinase	*S. salar*	91.3	8
Sal s 8	ACM09737.1	Triosephosphate isomerase	*S. salar*	86.5	2
Pan h 9	XP_026775867.1	Pyruvate kinase PKM-like	*P. hypophthalumus*	88.9	3
Pan h 10	XP_026774991.1	L-lactate dehydrogenase	*P. hypophthalumus*	81.9	5
Pan h 11	XP_026782721.1	Glucose 6-phosphate isomerase	*P. hypophthalumus*	92.7	2
Pan h 13	XP_026782131.1	Glyceraldehyde-3-phosphate dehydrogenase	*P. hypophthalumus*	96.1	2

Supplementary Table 5. Putative allergens identified in *H. molitrix* genome. The result was based on the protein sequence BLAST to the *H. molitrix* genome, and the identity of homolog with highest percentage matching was shown.

Allergen ID	Ref. GenBank ID	Biochemical name	Ref. species	Identity (%)	No. of homologs
Cten i 1	QCY53440	Beta-parvalbumin	*C. idella*	97.3	10
Cyp c 2	AWS00995.1	Beta-enolase	*C. carpio*	98.8	4
Pan h 3	XP_026771637.1	Aldolase A	*P. hypophthalumus*	93.1	5
Sal s 4	NP_001117128.1	Tropomyosin	*S. salar*	74.8	7
Sal s 6	XP_014059932.1	Collagen alpha	*S. salar*	87.7	5
Sal s 7	ACH70914.1	Creatine kinase	*S. salar*	91.1	8
Sal s 8	ACM09737.1	Triosephosphate isomerase	*S. salar*	85.7	2
Pan h 9	XP_026775867.1	Pyruvate kinase PKM-like	*P. hypophthalumus*	88.2	3
Pan h 10	XP_026774991.1	L-lactate dehydrogenase	*P. hypophthalumus*	83.3	3
Pan h 11	XP_026782721.1	Glucose 6-phosphate isomerase	*P. hypophthalumus*	92.6	2
Pan h 13	XP_026782131.1	Glyceraldehyde-3-phosphate dehydrogenase	*P. hypophthalumus*	92.0	2

Supplementary Table 6. Serum sample information.

Patient no.	Age (at sample collection) (y)	Sex	Fish-related symptoms[a]	Grass Carp sIgE (kUA/L; ImmunoCAP)	Cod sIgE (kUA/L; ImmunoCAP)
1	0.9	M	*Ae, R*	4.41	2.05
2	3.7	F	*Ae, Ery, O, U*	43.2	12.4
3	1.2	M	*Ae, Ery, O, U*	9.13	2.11
4	2.8	M	*Ae, Neu, U*	0.47	0.28
5	5.1	M	*NA*	11.9	4.91
6	0.6	M	*Ae, Ery*	2.15	0.56
7	4.6	M	*NA*	5.06	1.15
8	3.9	F	*NA*	1.27	0.51
9	5.6	M	*Ae, Ery, O, U*	1.49	0.76
10	2.4	M	*Ery, O*	0.53	0.17
11	3.2	M	*Ery, O*	1.35	0.28
12	6.6	F	*Ae, Ery, GI, U*	13.4	4.56
13	7	M	*Ae, GI, O*	19.8	7.79
14	10.4	M	*Ery, U*	1.71	0.56
15	10.3	F	*Ae, R, I, N, V, Ez*	23.2	12.3
16	33	M	*Ae, I*	1.71	0.67
17	2.5	F	*R, I*	3.12	1.09
18	8.8	M	*NA*	1.36	0.47
19	21.1	F	*NA*	1.52	0.48
20	4	M	*NA*	0.62	0.34

[a]Abbreviations for symptoms: Ae = angioedema; Ery = erythema; Ez = eczema flair; GI = gastrointestinal; I = itchy mouth/throat; Neu = neurologic; O = oral; Res = respiratory; R = rash; U = urticaria; V = vomiting; NA = not available

Supplementary Table 7. Parvalbumin homologs identified among fish species.

Allergen ID	Isoallergen	GenBank Protein	Homologs	Identity (%)	Protein ID
Pan h 1	Pan h 1.0101	XP_026772003	**Phy_PV6[a]**	**100**	**XP_026772003.1**
			Phy_PV7	77.982	XP_026772004.1
			Phy_PV2	78.899	XP_026770387.1
			Phy_PV8	65.421	XP_026772007.1
			Phy_PV3	73.349	XP_026771441.1
			Phy_PV5	63.889	XP_026771856.2
			Phy_PV4	57.944	XP_026771440.1
			Phy_PV9	57.944	XP_026803769.1
			Phy_PV1	48.598	XP_026771248.1
	Pan h 1.0201	XP_026803769	**Phy_PV9[a]**	**100**	**XP_026803769.1**
			Phy_PV5	72.477	XP_026771856.2
			Phy_PV7	53.271	XP_026772004.1
			Phy_PV4	51.852	XP_026771440.1
			Phy_PV1	48.598	XP_026771248.1
			Phy_PV2	57.009	XP_026770387.1
			Phy_PV8	52.778	XP_026772007.1
			Phy_PV6	57.944	XP_026772003.1
			Phy_PV3	54.206	XP_026771441.1

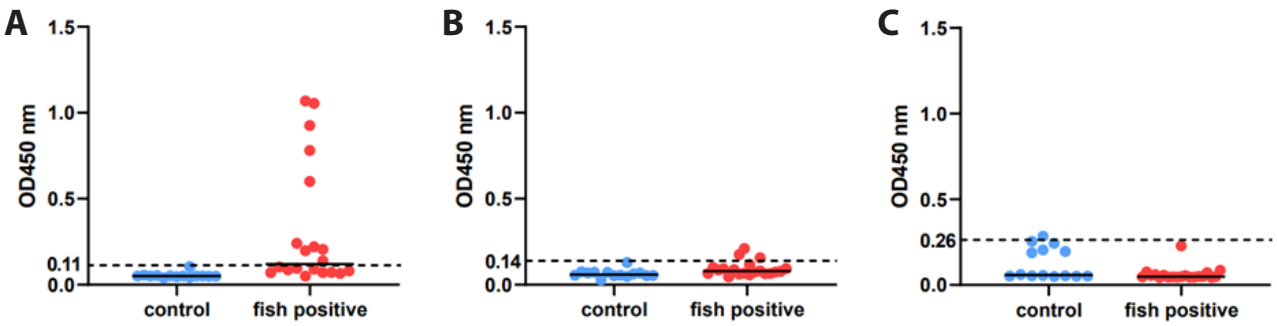
Supplementary Table 7. (Continued)

Allergen ID	Isoallergen	GenBank Protein	Homologs	Identity (%)	Protein ID
Gad m 1	Gad m 1.0101	AAK63086	**Gmo_PV2[a]**	**100**	**XP_030196388.1**
			Gmo_PV7	72.477	XP_030205636.1
			Gmo_PV3	71.560	XP_030196389.1
			Gmo_PV8	73.394	XP_030204978.1
			Gmo_PV1	58.716	XP_030196982.1
			Gmo_PV4	48.624	XP_030196391.1
			Gmo_PV6	55.046	XP_030203868.1
			Gmo_PV5	55.630	XP_030198131.1
			Gmo_PV9	46.296	XP_030208759.1
	Gad m 1.0102	CAM56785	**Gmo_PV2[a]**	**99.083**	**XP_030196388.1**
			Gmo_PV7	71.560	XP_030205636.1
			Gmo_PV3	70.642	XP_030196389.1
			Gmo_PV8	72.477	XP_030204978.1
			Gmo_PV1	59.633	XP_030196982.1
			Gmo_PV4	49.541	XP_030196391.1
			Gmo_PV5	55.556	XP_030198131.1
			Gmo_PV6	54.128	XP_030203868.1
			Gmo_PV9	45.370	XP_030208759.1
	Gad m 1.0201	AAK63087	**Gmo_PV3[a]**	**100**	**XP_030196389.1**
			Gmo_PV7	77.064	XP_030205636.1
			Gmo_PV2	71.560	XP_030196388.1
			Gmo_PV1	60.550	XP_030196982.1
			Gmo_PV8	65.138	XP_030204978.1
			Gmo_PV4	47.706	XP_030196391.1
			Gmo_PV6	53.211	XP_030203868.1
			Gmo_PV5	54.630	XP_030198131.1
			Gmo_PV9	49.533	XP_030208759.1
	Gad m 1.0202	CAM56786	**Gmo_PV3[a]**	**99.083**	**XP_030196389.1**
			Gmo_PV7	77.982	XP_030205636.1
			Gmo_PV2	71.560	XP_030196388.1
			Gmo_PV1	60.550	XP_030196982.1
			Gmo_PV8	65.138	XP_030204978.1
			Gmo_PV5	54.630	XP_030198131.1
			Gmo_PV4	46.789	XP_030196391.1
			Gmo_PV6	52.294	XP_030203868.1
			Gmo_PV9	49.533	XP_030208759.1
Sal s 1	Sal s 1.0101	CAA66403	**Ssa_PV13[a]**	**100**	**NP_001117190.1**
			Ssa_PV15[a]	**99.08**	**XP_014058479.1**
			Ssa_PV8[a]	**99.083**	**XP_014049624.1**
			Ssa_PV9[a]	**99.083**	**XP_014049696.1**
			Ssa_PV11[a]	**99.083**	**XP_014049698.1**
			Ssa_PV2	61.111	XP_014064366.1
			Ssa_PV7	55.556	XP_014049628.1
			Ssa_PV12	55.556	XP_014058477.1
			Ssa_PV21	72.093	NP_001117189.1
			Ssa_PV3	72.093	XP_014064375.1
			Ssa_PV6	59.633	XP_014047899.1
			Ssa_PV17	59.633	XP_014059854.1
			Ssa_PV20	62.693	NP_001133167.1
			Ssa_PV19	54.206	NP_001134150.1
			Ssa_PV18	54.630	XP_013985935.1
			Ssa_PV5	54.630	XP_014038140.1
			Ssa_PV4	49.074	XP_014064684.1
			Ssa_PV1	53.125	XP_014064352.1
			Ssa_PV14	51.402	XP_014058478.1
			Ssa_PV10	51.402	XP_014049697.1
			Ssa_PV22	47.222	XP_014035202.1
			Ssa_PV16	47.692	XP_014058485.1

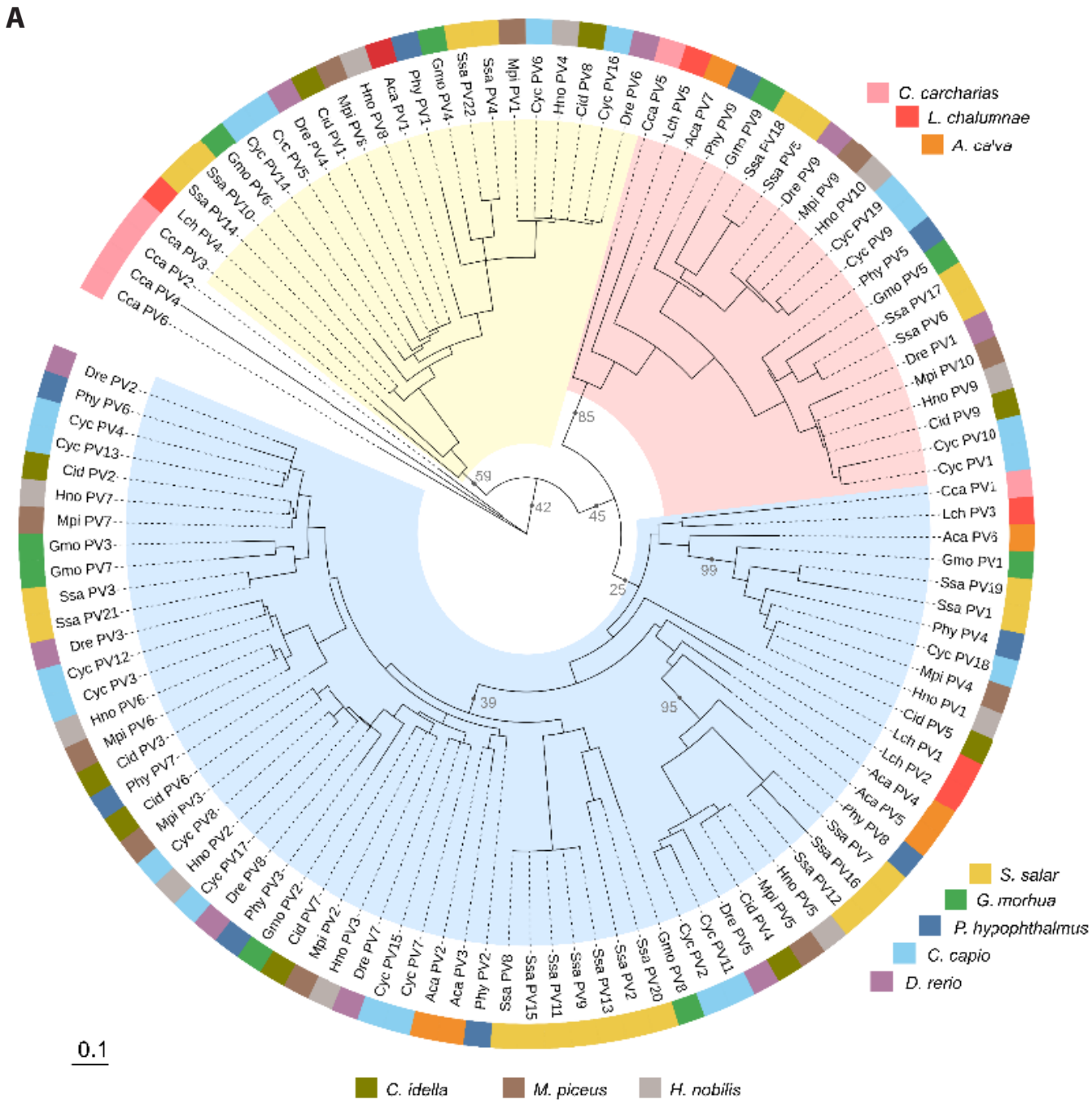
Supplementary Table 7. (Continued)

Allergen ID	Isoallergen	GenBank Protein	Homologs	Identity (%)	Protein ID
Cyp c 1	Cyp c 1.0101	CAC83658	**Cyc_PV7[a]**	**100**	**XP_018969314.1**
			Cyc_PV15[a]	**98.165**	**XP_018922924.1**
			Cyc_PV13	84.404	XP_018966236.1
			Cyc_PV8	88.073	XP_018969313.1
			Cyc_PV17	88.073	XP_018922925.1
			Cyc_PV4	82.569	XP_018966256.1
			Cyc_PV3	85.057	XP_018966257.1
			Cyc_PV18	65.138	XP_042591230.1
			Cyc_PV12	83.908	XP_018966237.1
			Cyc_PV5	57.944	XP_042586368.1
			Cyc_PV1	60.185	XP_018964600.1
			Cyc_PV14	55.140	XP_018966352.2
			Cyc_PV10	59.813	XP_018932906.1
			Cyc_PV16	54.206	XP_042591229.1
			Cyc_PV6	54.206	XP_018969360.1
			Cyc_PV9	57.009	XP_042567957.1
			Cyc_PV19	57.009	XP_042604735.1
			Cyc_PV2	59.259	XP_042586360.1
			Cyc_PV11	56.481	XP_042577342.1
	Cyp c 1.0201	CAC83659	**Cyc_PV13[a]**	**100**	**XP_018966236.1**
			Cyc_PV4[a]	**97.248**	**XP_018966256.1**
			Cyc_PV15	85.321	XP_018922924.1
			Cyc_PV7	84.404	XP_018969314.1
			Cyc_PV8	79.817	XP_018969313.1
			Cyc_PV17	78.899	XP_018922925.1
			Cyc_PV3	85.057	XP_018966257.1
			Cyc_PV18	64.220	XP_042591230.1
			Cyc_PV12	82.759	XP_018966237.1
			Cyc_PV1	62.037	XP_018964600.1
			Cyc_PV5	59.813	XP_042586368.1
			Cyc_PV14	57.944	XP_018966352.2
			Cyc_PV10	58.879	XP_018932906.1
			Cyc_PV9	57.944	XP_042567957.1
			Cyc_PV16	52.336	XP_042591229.1
			Cyc_PV19	57.009	XP_042604735.1
			Cyc_PV6	52.336	XP_018969360.1
			Cyc_PV2	58.333	XP_042586360.1
			Cyc_PV11	57.407	XP_042577342.1

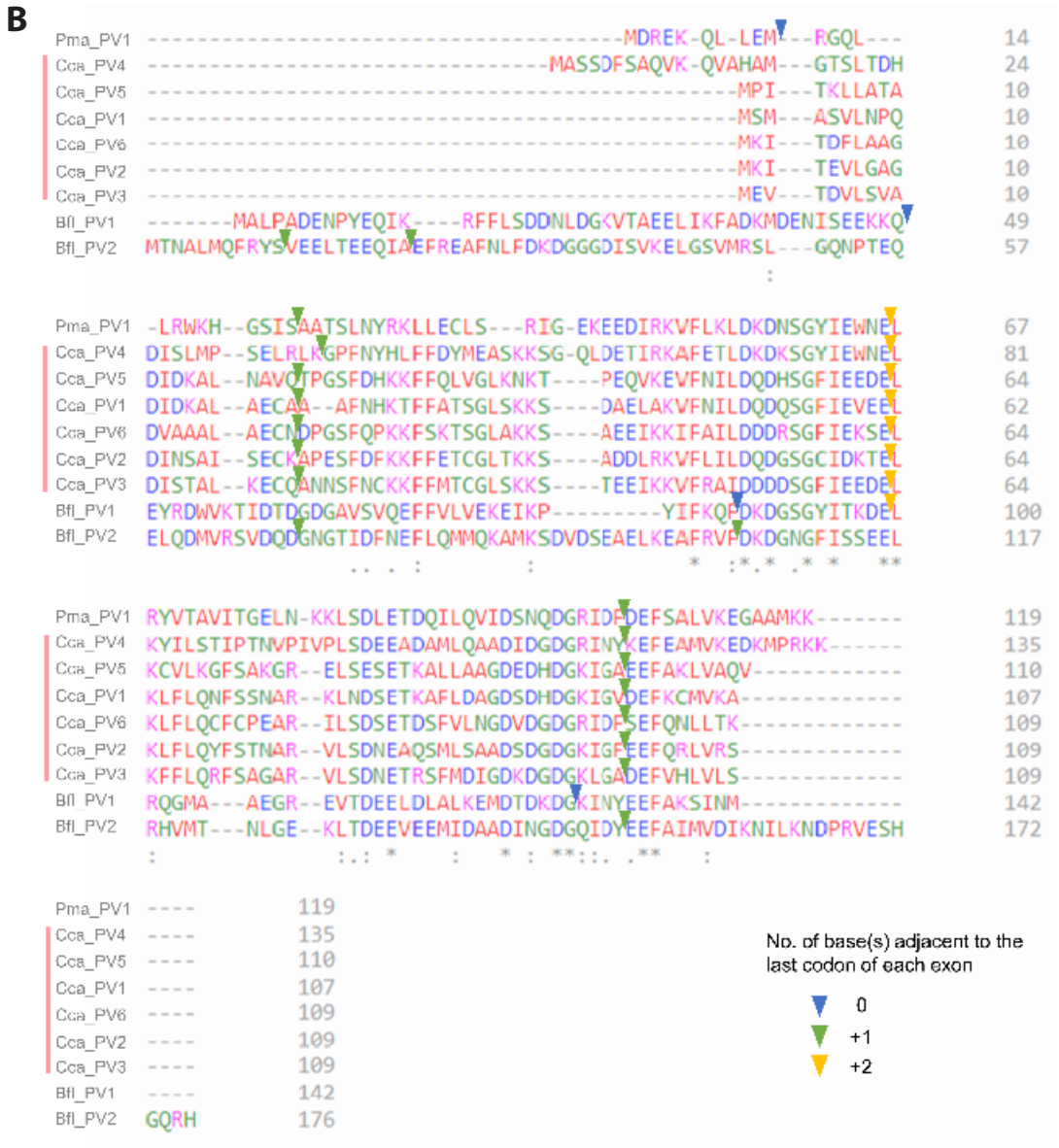
[a]Highest sequence similarity to the reported allergen sequences in WHO/IUIS allergen database



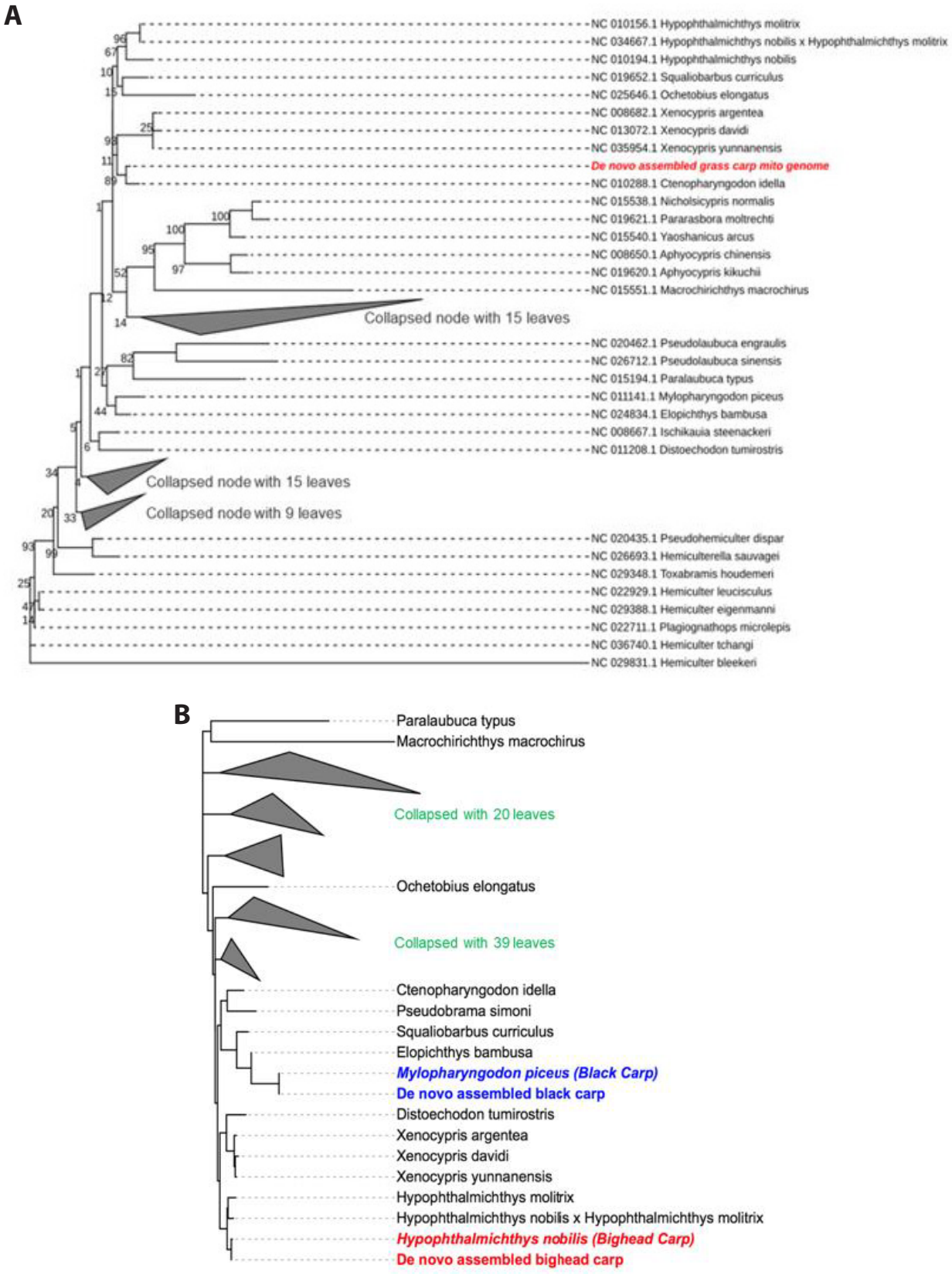
Supplementary Figure 1. Specific IgE against recombinant *C. idella* proteins. The recombinant proteins were synthesized based on the gene sequences from the homolog with highest sequence identity to the reported allergen Cten i 1, Cyp c 2 and Sal s 3. The allergenicity of the recombinant proteins (A: Parvalbumin, B: Beta Enolase, C: Aldoase A) were tested with 20 sera samples from fish allergy patients. The dotted line indicated the positive threshold value, 2-fold of the average value obtained from the negative controls. The allergenicity of each recombinant protein were PV: 45%, BE: 15% and AA: 0% respectively.



Supplementary Figure 2. Phylogenetic tree of parvalbumins in Gnathostomata and alignment of *C. carcharias* parvalbumins with closest protein sequences in *P. marinus* and *B. floridae*. (A) A total of 114 protein sequences of parvalbumin from 8 bony fishes and *C. carcharias* were aligned with MEGA, and the tree was generated by maximum likelihood algorithm (bootstrap = 500) based on the alignment result. Parvalbumins were divided into three clades included α-like (red), thymic CPV3-like (yellow) and β-like (blue) subtypes. (B) The arrows indicated the intron sites, and the color represented the number of additional base(s) next to the exon. The result indicated one of the closest parvalbumin homologs in invertebrate to vertebrates was Bfl_PV2 with two shared intron sites.



Supplementary Figure 2. (Continued)



Supplementary Figure 3. (A) A complete mitochondrion genome 16,609 bp was assembled and the tree was constructed based on 74 bony fishes. Based on the DNA sequence of Cox 1 gene, result indicated the genome assembled belongs *C. idella*. (B) Mitochondrial phylogenetic relationship of black carp and bighead carp together with Xenocyprididae subfamily species.