Supporting Information

Nakamura et al. 10.1073/pnas.1302051110

SI Materials and Methods

Taxonomy and Choice of Specimen of the Tuna. ***Taxonomic classification.***

Kingdom: Animalia
Phylum: Chordata
Subphylum: Vertebrata
Superclass: Osteichthyes (bony fishes)
Class: Actinopterygii (ray-finned fishes)
Infraclass: Teleostei
Division: Euteleostei
Superorder: Acanthopterygii
Order: Perciformes
Suborder: Scombroidei
Family: Scombridae
Subfamily: Scombrinae Bonaparte 1831
Tribe: Thunnini Starks 1910
Genus: *Thunnus* South 1845
Species: *T. orientalis* (Temminck and Schlegel 1844) (Pacific bluefin tuna)

Specimen sequenced. A male, 155.5 cm in standard length (SL), 79 kg in weight, killed on July 25, 2007, at age 3+, reared in Amami Station of National Center for Stock Enhancement, caught in September 2004, Pacific, off Kochi, Japan.

Specimen for cDNA analysis. A female, 180 cm in SL, 171 kg in weight, killed on December 17, 2009, at age 5+, reared in Amami Station of National Center for Stock Enhancement, caught in September 2004, Pacific, off Kochi, Japan.

Genome sequencing and assembly. Genomic DNAs were extracted by the standard phenol–chloroform method. Preparation of sequence templates for 454 and Illumina followed the manufacturers' instructions. We first assembled 454 genomic shotgun and paired-end reads by Newbler assembler (Version 2.5; Roche Diagnostics) (Table S1). Scaffolds and contigs made up of 454 reads were then bridged by Illumina paired-end reads by mapping with Bowtie (Version 0.12.7) (1). Here we used only 1,426,325 (= 521,223 + 905,102) valid read pairs exactly matched and mapped near the end of sequences to be bridged (Table S2). Contigs and scaffolds <2,000 bp in length were not used for subsequent analysis. All Illumina reads worked for improvement of sequence accuracy mapping them onto the scaffolds by bwa (Version 0.5.9) (2). As a result, 7,259 nucleotide mismatches and 312,851 short indels were overridden by sequences called by Illumina.

Gene Analysis. ***cDNA sample preparation and sequencing.*** Total RNAs were isolated from the following 10 tuna tissues with TRIzol reagent (Invitrogen, Life Technologies) according to the manufacturer's instructions: fin, gill, heart, intestine, kidney, liver, red muscle, white muscle, pyloric caeca, and stomach. RNA integrity was assessed by using an Agilent Bioanalyzer 2100 with an RNA 6000 Nano Kit (Agilent Technologies). RNA integrity numbers were between 7.8 and 9.1. Equal amounts of total RNA were pooled from each tissue, and 1 μg of the pooled RNA was used to construct a normalized cDNA library. Full-length cDNAs for 454 sequencing were prepared as described (3). Briefly, first-strand cDNA was synthesized by using a SMART cDNA Library Construction Kit (Clontech Laboratories) with the following modifications: The primer used for first-strand cDNA synthesis was a modified oligo-dT primer (AAGCAGTGGTATCAACGCAGAGTCGCAGTCGGTACTTTTTTCTTTTTTV), and the cDNA was amplified by using the "CAP" primer (AAGCAGTGGTATCAACGCAGAGT). Amplified cDNA was purified with a Chromaspin-400 column (Clontech Laboratories) and normalized by using a Trimmer-Direct cDNA Normalization Kit (Evrogen) according to the manufacturer's instructions. The normalized cDNA was amplified and purified as described above. Approximately 5 μg of cDNA was sheared by nebulization and used for 454 sequencing.

cDNA contigs, isotigs, and isogroups. Reads by 454 were assembled by using Newbler (Version 2.5). We obtained 180,512 contigs, of which 5,741 have both 5′ and 3′ adaptor sequences and are regarded as full-length (Table S3).

Repeat masking. To obtain a tuna repeat sequence library, RepeatModeler (Version 1.0.4) (4) was performed to 16,801 scaffold sequences (a mitochondrial genome sequence is subtracted from 16,802 scaffolds), and 981 repeat families were predicted. By using the repeat library, the scaffold sequences were subsequently masked by RepeatMasker (Version 3.2.9) (5).

Gene inventory. The genome data of five teleost fishes (zebrafish, medaka, greenpuffer, fugu, and stickleback) were downloaded from the Ensembl database (Release 61; http://uswest.ensembl.org/index.html) (6). The cod genome data were downloaded from http:/codgenome.no (7).

In addition to 180,512 cDNA contigs from *T. orientalis* determined in this study, we obtained the sequences of cDNAs expressed in ovary, testis, and liver of the Atlantic bluefin tuna (8) from the National Center for Biotechnology Information Expressed Sequence Tags database (www.ncbi.nlm.nih.gov/dbEST/). The accession nos. are as follows: EC091633–EC93160, EG629962–EG631176, EC917676–EC919417, EG999340–EG999999, EH000001–EH000505, EH667253–EH668984, EL610526–EL611807, EC421414–EC422414, and EH379568–EH380065. We further removed rRNA and tRNA sequences by BLASTN and short sequences (<32 bp) and finally obtained 178,992 cDNA contigs from the Pacific bluefin tuna and 10,131 from the Atlantic bluefin tuna (Table S4). Among these, 174,351 (97.4%) contigs from the Pacific bluefin tuna and 8,322 (82.1%) from the Atlantic bluefin tuna were mapped to the scaffold by BLAT (9) (options: –mask = lower –qMask = lower) (Table S4).

We predicted protein-coding sequences using a gene-finding software AUGUSTUS (10) with a tuna-specific training model. At first, to collect the training set, four gene-finding software programs, GENSCAN (11), GlimmerHMM (12), GeneMark-ES (13), and AUGUSTUS, were performed to 16,801 scaffold sequences (Table S4). Next, all of the predicted amino acid sequences were queried against the protein sequence collection from the Ensembl five fish genomes by using BLASTP (14) with E–value $< 10^{-10}$, and the predicted genes were filtered following two criteria: Query and hit sequences were (*i*) <5% different in length and (ii) aligned over 95% of each length. We thus obtained 5,583 validated tuna genes and constructed the tuna-specific Hidden Markov Model by AUGUSTUS.

In addition, we selected cDNA sequences from the Atlantic/Pacific bluefin tunas and amino acid sequences from the five teleost genomes, both of which were mapped to the scaffold: (*i*) 156,265 of 174,351 cDNA sequences screened by a script filterPSL.pl included in the AUGUSTUS package, and (*ii*) 23,908 amino acid sequences by TBLASTN (E-value $< 10^{-5}$). These alignments to the scaffold were used as hints for UTR and protein-coding regions in AUGUSTUS to improve both the training model and gene prediction. Considering the functional domain size of proteins, we filtered the genes predicted by AUGUSTUS with the amino acid length ($\geq$50 aa). We manually revised those gene regions predicted when necessary.

Ribosomal RNA genes on the scaffold sequences were identified by BLASTN (15) searches (E-value $< 10^{-5}$) using the public sequences in the GenBank database (www.ncbi.nlm.nih.gov/): 28S from *Siniperca chuatsi* (Chinese perch; accession no. AY452491); 18S from *Oryzias latipes* (medaka; accession no. AB105163); 5.8S from *Auxis rochei* (bullet tuna; accession no. AB193734); and 5S from *Thunnus thynnus* (Atlantic bluefin tuna; accession no. HQ681117). Transfer RNA genes were predicted by tRNAscan-SE (16) with default options.

Expression analysis of tuna opsin genes. Total RNA was extracted from the eyes of 25-d-old juvenile Pacific bluefin tuna by using the RNeasy Lipid Tissue Mini Kit (QIAGEN). After DNase digestion, cDNAs were synthesized by the GeneRacer kit (Life Technologies). Primer sequences were designed based on the predicted coding sequences:

LWS-Fw/Rv, 5′-GGAGAGATGGGTCGTTGTGT-3′/5′-AACACCAGGGTCGTCACTTC-3′;

SWS1-Fw/Rv, 5′-TGGGATCCTTCATCACCTGT-3′/5′-TCCAAACACCATCTCCATGA-3′;

SWS2A-Fw/Rv, 5′-AAGCTTCGGTCTCACCTCAA-3′/5′-CAAAAGCAACCACAGCAAGA-3′;

SWS2B-Fw/Rv, 5′-GACCACTGGGATGCAAGATT-3′/5′-AAAGGGACAGCAAAGCAGAA-3′;

RH1-Fw/Rv, 5′-TTGGCTTTCCCATCAACTTC-3′/5′-AGCCAGAACAACCAGTGACC-3′;

RH2A1-Fw/Rv, 5′-TCTGTAAACCCATGGGAAGC-3′/5′-GATCAGGAAGACGGGAACAA-3′;

RH2A2-Fw/Rv, 5′-CGGTTTCACTTGGGTAATGG-3′/5′-GATCAGGAAGACGGGAATGA-3′;

RH2B-Fw/Rv, 5′-CCCAGGCTACAACAACGAAT-3′/5′-CCAACACCATCAGGACACAC-3′;

g6738-Fw/Rv, 5′-CAGGCTTCAACAACGAGTCA-3′/5′-TGTCACCTCCCTCTCAGCTT-3′;

g6740-Fw/Rv, 5′-GTCTGCTTTGGGTTCACCAT-3′/5′-ACTCCTGCGTGAGTTCCAGT-3′.

The cDNAs of tuna opsin were amplified by using the Phusion High-Fidelity PCR Master Mix with HF Buffer (Thermo Fisher Scientific) according to the manufacturer's instruction. Then, the reaction mixtures were electrophoresed on 2.0% agarose gels containing SYBR Safe DNA Gel Stain (Life Technologies).

Phylogenetic analysis on opsin genes. The multiple alignments were calculated by the MAFFT program (17), in which each alignment was first calculated by using amino acid sequences and then reversely translated to nucleotide sequences. The synonymous and nonsynonymous substitution rates (18) between each orthologous gene pair were calculated by the LPB93 method (19, 20) using the PAML program (21). For molecular phylogenetic analysis, distance matrices among nucleotide sequences were calculated by using Kimura's two-parameter method (22), and phylogenetic trees were constructed by using the neighbor-joining algorithm (23) in MEGA5 (24). The molecular clock based on synonymous substitution rate was calculated by using the divergence times among seven teleost genomes (Table S6). We conducted a Monte Carlo simulation by randomly changing divergence times in consideration of the deviations, in which we approximated the posterior distributions of estimated divergence time in the literature (25–27) by normal distributions. From the simulation, we obtained the 95% confidence intervals of duplication and conversion times among opsin genes.

Green and blue opsin genes for phylogenetic analysis. The nucleotide sequences were obtained from the GenBank and Ensembl (Release 61).

Green pigment genes. Atlantic cod (*Gadus morhua*), AF385824 and ENSGAUG00000017343; Atlantic salmon (*Salmo salar*), NM_001123707; barbell steed (*Hemibarbus labeo*), EU919546; bluefin killifish (*Lucania goodei*), AY296739; chicken (*Gallus gallus*), NM_205490; coelacanth (*Latimeria chalumnae*), AF131258–AF131262; coho salmon (*Oncorhynchus kisutch*), DQ309027 and AY214147; common carp (*Cyprinus carpio*), AB110602 and AB110603; fugu (*Takifugu rubripes*), NM_001033712; green anole (*Anolis carolinensis*), XM_003220346; greenpuffer (*Tetraodon nigroviridis*), AY598944; guppy (*Poecilia reticulata*), DQ234858 and DQ234859; halibut (*Hippoglossus hippoglossus*), AF156263; lamprey (*Geotria australis*), AY366494; lizard (*Uta stansburiana*), DQ100324; lungfish (*Neoceratodus forsteri*), EF526296; medaka (*Oryzias latipes*), AB223053, AB223054 and AB223055; Nile tilapia (*Oreochromis niloticus*), JF262086; rainbow trout (*Oncorhynchus mykiss*), NM_001124323; rock pigeon (*Columba livia*), AF149232–AF149233; stickleback (*Gasterosteus aculeatus*), ENSGACG00000001398 and ENSGACG00000001434; stone moroko (*Pseudorasbora parva*), EU919550; thickhead chub (*Opsariichthys pachycephalus*), EU410462, EU410463, EU410464, and EU410465; yellow-tail acei (*Pseudotropheus* sp. 'acei'), DQ088630, DQ088633 and DQ088645; zebrafish (*Danio rerio*), NM_131253, NM_182891, NM_182892, and NM_131254.

Blue pigment genes. Atlantic cod (*Gadus morhua*), AF385822 and ENSGAUG00000018054; Atlantic salmon (*Salmo salar*), NM_001123706; barbell steed (*Hemibarbus labeo*), EU919547; bluefin killifish (*Lucania goodei*), AY296736 and AY296737; chicken (*Gallus gallus*), NM_205517; coho salmon (*Oncorhynchus kisutch*), AY214149; common carp (*Cyprinus carpio*), AB113668; frog (*Xenopus laevis*), NM_001089142; fugu (*Takifugu rubripes*), AY598947; green anole (*Anolis carolinensis*), XM_003216903; greenpuffer (*Tetraodon nigroviridis*), CAAE01014528; guppy (*Poecilia reticulata*), JF303638 and DQ234860; halibut (*Hippoglossus hippoglossus*), AF316497; lamprey (*Geotria australis*), AY366492; lizard (*Uta stansburiana*), DQ100326; lungfish (*Neoceratodus forsteri*), EF526299; medaka (*Oryzias latipes*), AB223056 and AB223057; Nile tilapia (*Oreochromis niloticus*), JF262088; rainbow trout (*Oncorhynchus mykiss*), NM_001124322; rock pigeon (*Columba livia*), AF149238–AF149242; stickleback (*Gasterosteus aculeatus*), ENSGACG00000010229; stone moroko (*Pseudorasbora parva*), EU919552; thickhead chub (*Opsariichthys pachycephalus*), EU410467; yellow-tail acei (*Pseudotropheus* sp. 'acei'), DQ088636 and DQ088639; zebrafish (*Danio rerio*), NM_131192.

Accession nos. of rhodopsin genes examined in this study. DQ874804 (*Acanthocybium solandri*); DQ354577 (*Acanthopagrus butcheri*); EU919542 (*Acrossocheilus paradoxus*); EU637931 (*Aeoliscus strigatus*); EU637932 (*Agonopsis chiloensis*); HQ639184 (*Akihito futuna*); HQ639180 (*Akihito vanuatu*); JN544545 (*Alepocephalus agassizii*); EU637933 (*Alepocephalus antipodianus*); JN412584 (*Alepocephalus bairdii*); EU082664 (*Algansea lacustris*); AB084924 and AB458132 (*Altolamprologus calvus*); HM107824 (*Amphiprion melanopus*); FJ711156 (*Anableps anableps*); EU637936 (*Anarhichas lupus*); AJ249202 and AJ249203 (*Anguilla japonica*); XM_003224879 (*Anolis carolinensis*); JN412582 (*Anoplogaster cornuta*); EU637937 (*Antigonia capros*); EU637938 (*Aphanopus carbo*); EU637939 (*Aplodactylus punctatus*); EU637940 (*Apogon quadrifasciatus*); JN412571 (*Argyropelecus aculeatus*); JN412572 (*Argyropelecus gigas*); EU637942 (*Argyrosomus regius*); EU407248 (*Aristostomias scintillans*); EF517407 (*Aristostomias tittmanni*); AY846631 (*Aruma histrio*); EU637943 (*Aspasma minima*);

EU637944 (*Assessor flavissimus*); AB084925 (*Astatotilapia burtoni*); AB185213 and AB185216 (*Astatotilapia calliptera*); JN544530 (*Astronesthes macropogon*); AY775112 (*Aulonocara hueseri*); AB185214 (*Aulonocara* sp. 'copper'); AB084926 (*Aulonocara* sp. blueneon); AB457974 and AB458127 (*Aulonocranus dewindti*); EU637972 (*Aurigequula fasciata*); DQ874801 (*Auxis thazard*); HQ639147 and HQ639148 (*Awaous guamensis*); AB185217 (*Baileychromis centropomoides*); DQ874818 (*Balistes capriscus*); AF137212 (*Balistes* sp.); JN412581 (*Bassozetus compressus*); JN544531 (*Bathophilus pawneei*); JN544532 (*Bathophilus vaillanti*); AB185218 (*Bathybates fasciatus*); JN412585 (*Bathysaurus ferox*); JN412586 (*Bathysaurus mollis*); JN544540 (*Bathytroctes microlepis*); AB457850 and AB458117 (*Benthochromis horii*); AB457848 and AB458116 (*Benthochromis melanoides*); AB084927 (*Benthochromis tricoti*); JN412576 (*Benthosema suborbitale*); JN412574 (*Bolinichthys indicus*); JN544534 (*Bonapartia pedaliota*); JF745773 (*Bostockia porosa*); AB084928 (*Boulengerochromis microlepis*); JF966219 (*Brevibora dorsiocellata*); EU637945 (*Callanthias ruber*); EU637946 (*Callionymus schaapii*); AB457975 (*Callochromis macrops*); EF565167 (*Callorhinchus milii*); EF452898 (*Campostoma anomalum*); EU082626 (*Campostoma pullum*); EU919559 (*Candidia barbatus*); EU637947 and JN412580 (*Cataetyx laticeps*); EU637948 (*Cepola macrophthalma*); AB458018 and AB458133 (*Chalinochromis brichardi*); EU637949 (*Champsodon snyderi*); JN412564 (*Chauliodus danae*); EU407250 (*Chauliodus macouni*); JN412563 (*Chauliodus sloani*); EU637950 (*Cheilopogon heterurus*); HQ729682 and HQ729683 (*Chologaster cornuta*); AF149230, AF149231 (*Columba livia*); AB043817 and AB043818 (*Conger myriaster*); JN412577 (*Conocara salmonea*); JF745771 (*Coreoperca kawamebari*); JF745770 (*Coreoperca loona*); EU637951 (*Coryphaena equiselis*); DQ874824 (*Coryphaena hippurus*); JN544539 (*Coryphaenoides carapinus*); JN412578 (*Coryphaenoides guentheri*); JN544537 (*Coryphaenoides leptolepis*); JN544538 (*Coryphaenoides profundicolus*); HQ639165 and HQ639166 (*Cotylopus rubripinnis*); EU082603 (*Couesius plumbeus*); EU637952 (*Cubiceps gracilis*); AB457977 (*Cyathopharynx furcifer*); AB084929 (*Cyphotilapia frontosa*); AB457856, AB457858, AB457874, AB457892, AB457915 and AB588118 (*Cyprichromis coloratus*); AB084930, AB457921 and AB458118 (*Cyprichromis leptosoma*); AB457877 (*Cyprichromis microlepidotus*); AB457930, AB457933 and AB458119 (*Cyprichromis pavo*); AB457854, AB457937, AB457938, AB457939 and AB457940 (*Cyprichromis zonatus*); Z71999 (*Cyprinus carpio*); JQ614110 (*Danio aesculapii*); JQ614112 (*Danio* aff. choprai BOLD:AAU2157); JQ614115 (*Danio* aff. dangila BOLD:AAU2159); JQ614118 (*Danio* aff. kyathit BOLD:AAJ4326); JQ614121 (*Danio* albolineatus); JQ614124 (*Danio cf* kerri BOLD:AAU2162); JQ614127 (*Danio cf* rerio RCYY501-11); JQ614128 and JQ614129 (*Danio choprai*); JQ614136 (*Danio feegradei*); JQ614141 (*Danio margaritatus*); JQ614143 (*Danio meghalayensis*); JQ614145 and JQ614146 (*Danio nigrofasciatus*); AB087811, JQ614147 and NM_131084 (*Danio rerio*); EU637953 (*Dascyllus trimaculatus*); EU637954 (*Datnioides polota*); JQ614160 (*Devario cf* acuticephala BOLD:AAV9172); JQ614163 (*Devario cf* chrysotaeniatus BOLD:AAU3723); JQ614169 (*Devario pathirana*); JQ614182 (*Devario* sp. undet 2RAC); JN544536 (*Diaphus metopoclampus*); JN412587 (*Diaphus rafinesquii*); AB084931 and AY775059 (*Dimidiochromis compressiceps*); AB185220 (*Diplotaxodon macrops*); EU637955 (*Echiichthys vipera*); EU637956 (*Echiodon cryomargarites*); AB457978 (*Ectodus descampsii*); AY846565 (*Elacatinus digueti*); AY846623 (*Elacatinus gemmatus*); AY846568 (*Elacatinus inornatus*); AY846573 (*Elacatinus janssi*); AY846619 (*Elacatinus macrodon*); AY846575 and AY846576 (*Elacatinus multifasciatus*); AY846626 (*Elacatinus pallens*); JF745782 (*Elassoma alabamae*); JF745780 (*Elassoma boehlkei*); JF745776 (*Elassoma evergladei*); JF745784 (*Elassoma gilberti*); JF745774 (*Elassoma zonatum*); JF913270 (*Enoplosus armatus*); EU637959 (*Epigonus telescopus*); FJ426122 (*Epinephelus coioides*); DQ874825 (*Epinephelus guttatus*); JQ614184 (*Esomus metallicus*); JN544544 (*Eurypharynx pelecanoides*); JN544529 (*Eustomias acinosus*); EU637960 (*Favonigobius reichei*); JF745772 (*Gadopsis marmoratus*); AF137211 and AF385832 (*Gadus morhua*); NM_001030606 (*Gallus gallus*); DQ882021 (*Gasterochisma melampus*); DQ874812 (*Gempylus serpens*); AY366493 (*Geotria australis*); EU082602 (*Gila pandora*); AY846577 (*Ginsburgellus novemlineatus*); AB158262 (*Girella punctata*); EU637963 (*Gnathanodon speciosus*); AB084932 (*Gnathochromis permaxillaris*); AY846630 (*Gobiosoma chiquita*); JN412561 (*Gonostoma elongatum*); AB457989 (*Grammatotria lemairii*); EU637964 (*Grammicolepis brachiusculus*); AB185221 (*Greenwoodochromis bellcrossi*); EU606010 (*Gymnocypris eckloni*); HQ444181 (*Gymnothorax favagineus*); HQ444182 (*Gymnothorax reticularis*); EU606013 (*Hampala macrolepidota*); AF315354 (*Haplochromis burtoni*); AB084933 (*Haplochromis* sp. brownae); AB185390 and AB458142 (*Haplotaxodon microlepis*); AB458084 (*Haplotaxodon trifasciatus*); EU919548 (*Hemibarbus labeo*); AB185222 (*Hemibates stenosoma*); EF452899 (*Hemitremia flammea*); EU637965 (*Himantolophus groenlandicus*); AF156265 (*Hippoglossus hippoglossus*); JN544542 (*Histiobranchus bathybius*); NM_000539 (*Homo sapiens*); JN412583 (*Hoplostethus mediterraneus*); EU082653 (*Hybopsis boucardi*); JN412569 (*Ichthyococcus ovatus*); EU407249 (*Idiacanthus antrostomus*); JN412568 (*Idiacanthus fasciola*); EU637967 (*Indostomus paradoxus*); DQ874808 (*Istiophorus platypterus*); AB458021 (*Julidochromis dickfeldi*); GQ409787 (*Konia dikume*); JF764596 (*Kuhlia mugil*); AY775113 (*Labeotropheus fuelleborni*); AB185223 (*Labeotropheus trewavasae*); AB084934 (*Labidochromis caeruleus*); AY775120 (*Labidochromis chisumulae*); EU637969 (*Lamprogrammus shcherbachevi*); AB458029 and AB458030 (*Lamprologus callipterus*); AB458048 and AB458049 (*Lamprologus lemairii*); AB084935 (*Lamprologus ocellatus*); EU637970 (*Lates calcarifer*); EU637971 (*Lates niloticus*); AF131253, AF131254, AF131255, AF131256, AF131257 (*Latimeria chalumnae*); EF452915 and JQ614186 (*Laubuca dadiburjori*); JF966215 (*Laubuca laubuca*); HQ639177 and HQ639179 (*Lentipes armatus*); HQ639174 and HQ639176 (*Lentipes kaaea*); AB458031, AB458032 and AB458134 (*Lepidiolamprologus cunningtoni*); AB458037 (*Lepidiolamprologus elongatus*); AB458046 and AB458047 (*Lepidiolamprologus kendalli*); AB458059 (*Lepidiolamprologus nkambae*); EU407252 and EU407253 (*Lepidopus fitchi*); AB116382 (*Lethenteron camtschaticum*); AY775116 (*Lethrinops parvidens*); GU179273 (*Limia dominicensis*); AB185224 (*Limnochromis abeelei*); AB185225 (*Limnochromis staneri*); AY673743 (*Lipochromis melanopterus*); EU637973 (*Lopholatilus chamaeleonticeps*); AY296738 (*Lucania goodei*); EU637974 (*Lutjanus sebae*); EF452900 (*Luxilus chrysocephalus*); EU637976 (*Lycodapus antarcticus*); JF764597 (*Lycodes terraenovae*); DQ874810 (*Makaira nigricans*); AJ224691 (*Malacosteus niger*); AY775114 (*Maylandia zebra*); FJ940704 (*Melanotaenia australis*); EU637977 (*Menidia menidia*); EU637978 (*Microcanthus strigatus*); JQ614188 and JQ614189 (*Microdevario kubotai*); HM630114 (*Microlipophrys bauchotae*); HM630130 (*Microlipophrys caboverdensis*); HM630116 and HM630117 (*Microlipophrys dalmatinus*); HM630120 and HM630121 (*Microlipophrys velifer*); AY742587 (*Micropterus cataractae*); JQ614192 (*Microrasbora cf* rubescens BOLD:AAU1781); JQ614134 (*Microrasbora erythromicron*); JF966216 and JQ614195 (*Microrasbora rubescens*); AF137215 (*Mola mola*); EU637980 (*Monodactylus* sp. BL-2008); EU637981 (*Morone saxatilis*); EU637982 (*Mullus surmuletus*); NM_145383 (*Mus musculus*); AY775119 (*Mylochromis lateristriga*); EU637983 (*Myripristis* sp. BL-2008); EU637984 (*Naso lituratus*); EU637985 (*Nemadactylus monodactylus*); EF526295 (*Neoceratodus forsteri*); AB458135 (*Neolamprologus bifasciatus*); AY775110 (*Neolamprologus brichardi*); AB458065 (*Neolamprologus buescheri*); AB458066 (*Neolamprologus fasciatus*); AB084937 (*Neolamprologus leleupi*); AB458071 (*Neolamprologus obscurus*); AB458073 and AB458137 (*Neolamprologus sexfasciatus*); AB458075 (*Neolamprologus tetracanthus*); AB458000, AB458014 and AB458138 (*Neolamprologus ventralis*); DQ874819 (*Neomerinthe hemingwayi*); EU637986 (*Neopagetopsis ionah*); EU637987 (*Nerophis lumbriciformis*); EU637934 (*Niphon spinosus*); EF452903 (*Notropis atherinoides*); EU082658 (*Notropis calientis*); EU082651 (*Notropis imeldae*); AY214141 (*Oncorhynchus keta*); AY214146 (*Oncorhynchus kisutch*); NM_001124319 (*Oncorhynchus mykiss*); AY214156 (*Oncorhynchus nerka*); AY214136 (*Oncorhynchus tshawytscha*); EU637988 (*Ophiocara porocephala*); AB457981 (*Ophthalmotilapia nasuta*); AY775109 (*Ophthalmotilapia ventralis*); DQ874822 (*Opsanus pardus*); GU218596 (*Opsariichthys bidens*); EU919560 (*Opsariichthys pachycephalus*); XM_003439005 (*Oreochromis niloticus*); AB180742 and NM_001104695 (*Oryzias latipes*); AF137213 (*Ostracion* sp. IMCB-2002); AB185227 (*Otopharynx brooksi*); EF517408 (*Pachystomias microdon*); JQ614196 (*Paedocypris cf* carbunculus BOLD: AAV0810); EU637990 (*Pagetopsis macropterus*); AB185229 (*Pallidochromis tokolosh*); GU179284 (*Pamphorichthys araguaiensis*); HQ857439 (*Pamphorichthys hasemani*); HM630109 (*Parablennius sanguinolentus*); AB245433 (*Paracheirodon innesi*); AB457853, AB457880, AB458120 and AB458121 (*Paracyprichromis brieni*); AB457952 and AB458122 (*Paracyprichromis nigripinnis*); AY673747 (*Paralabidochromis cyaneus*); DQ874794 (*Paralichthys dentatus*); HQ413772 (*Paralichthys olivaceus*); EU637991 (*Pelates quadrilineatus*); EU637992 (*Periophthalmus barbarus*); AB458140 (*Perissodus eccentricus*); AB185230, AB458085 and AB458111 (*Perissodus elaviae*); AB458090, AB458115 and AB458141 (*Perissodus microlepis*); AB458114 (*Perissodus multidentatus*); AB458095, AB458113 and AB458139 (*Perissodus paradoxus*); AB458112 (*Perissodus straeleni*); EU637993 (*Photoblepharon palpebratum*); JN412566 (*Photostomias guernei*); JN412579 (*Phycis blennoides*); EU637994 (*Phycis phycis*); EU637989 (*Pinguipes chilensis*); AB185232 (*Placidochromis polli*); AB074484 (*Plecoglossus altivelis*); DQ912023, DQ912024 and GU179281 (*Poecilia reticulata*); JN544535 (*Pollichthys mauli*); JN544533 (*Polyipnus asteroides*); EU637995 (*Pomacanthus maculosus*); FN430608 (*Pomatoschistus marmoratus*); FN430606 and FN430607 (*Pomatoschistus microps*); FN430589, FN430590, FN430591, FN430592, FN430595, FN430596, FN430597, FN430602 and FN430603 (*Pomatoschistus minutus*); EU637996 (*Pontinus longispinis*); EU637997 (*Priacanthus arenatus*); HQ639145 and HQ639146 (*Protogobius attiti*); AY631036 (*Pseudopleuronectes americanus*); EU919553 (*Pseudorasbora parva*); GQ422475 (*Pseudotropheus* sp. 'acei'); AB084939 (*Pseudotropheus* sp. orange); EU637999 (*Ptereleotris zebra*); EU638000 (*Pterocaesio digramma*); EU638001 (*Pterycombus brama*); AY673740 (*Pundamilia azurea*); AY673738 (*Pundamilia nyererei*); AY673741 (*Pundamilia* sp. 'luanso'); AY673742 and AY673771 (*Pundamilia* sp. 'red head'); GQ409790 (*Pungu maclareni*); JQ614205 and JQ614206 (*Puntius assimilis*); JQ614210 (*Puntius cf* sarana RCYY056-11); JQ614220 (*Puntius denisonii*); JQ614240 (*Puntius jerdoni*); JQ614245 (*Puntius lateristriga*); JQ614246 (*Puntius manipurensis*); JQ614248 (*Puntius nigrofasciatus*); JQ614253 (*Puntius padamya*); JQ614261 (*Puntius semifasciolatus*); JQ614263 (*Puntius shalynius*); JQ614271 (*Puntius tetrazona*); AB185236 (*Rhamphochromis esox*); AB196147 (*Rhamphochromis longiceps*); HQ444180 (*Rhinomuraena quaesita*); HQ639143 and HQ639144 (*Rhyacichthys guilberti*); EF452906 (*Richardsonius balteatus*); AY846628 (*Risor ruber*); DQ874813 (*Ruvettus pretiosus*); NM_001123537 (*Salmo salar*); DQ874800 (*Sarda sarda*); AB084940 (*Sarotherodon melanotheron*); EU638003 (*Schedophilus medusophagus*); EU606012 (*Schizothorax waltoni*); EU637979 (*Sciaenidae* sp. BL-2008); DQ874797 (*Scomber scombrus*); DQ874799 (*Scomberomorus cavalla*); DQ874798 (*Scomberomorus maculatus*); DQ874805 (*Scomberomorus regalis*); EF517404 and EF517405 (*Scopelarchus analis*); EU638005 (*Scophthalmus rhombus*); DQ490124 (*Sebastolobus altivelis*); EU638006 (*Selene dorsalis*); EU638002 (*Selenotoca multifasciata*); HQ639164 (*Sicydiinae* sp. 'Keith et al., 2010'); HQ639173 (*Sicydium plumieri*); HQ639170 (*Sicydium punctatum*); HQ639163 (*Sicyopterus lagocephalus*); HQ639160 (*Sicyopterus pugnans*); HQ639159, HQ639193 and HQ639194 (*Sicyopus sasali*); HQ639196 (*Sicyopus* sp. 41B19B); EU638007 (*Siganus vulpinus*); JN412562 (*Sigmops bathyphilus*); EU638008 (*Sillago sihama*); EU605983 (*Sinocyclocheilus furcodorsalis*); EU605992 and EU605993 (*Sinocyclocheilus grahami*); EU605988 and EU605989 (*Sinocyclocheilus jii*); EU606009 (*Sinocyclocheilus jiuxuensis*); EU605986 and EU605987 (*Sinocyclocheilus macrolepis*); EU605990 and EU605991 (*Sinocyclocheilus macrophthalmus*); EU606003 and EU606004 (*Sinocyclocheilus maculatus*); EU605984 and EU605985 (*Sinocyclocheilus microphthalmus*); EU606001 and EU606002 (*Sinocyclocheilus purpureus*); EU606005 and EU606006 (*Sinocyclocheilus qiubeinsis*); EU605999 and EU606000 (*Sinocyclocheilus rhinocerous*); EU605994 and EU605995 (*Sinocyclocheilus tingi*); EU605982 (*Sinocyclocheilus xunlensis*); EU605996 (*Sinocyclocheilus yangzongensis*); EU605997 and EU605998 (*Sinocyclocheilus yimenensis*); EU638009 (*Solea solea*); AB084941 (*Spathodus erythrodon*); HQ729686 (*Speoplatyrhinus poulsoni*); EU638010 (*Sphaeramia nematoptera*); DQ874816 (*Sphyraena barracuda*); DQ874817 (*Sphyraena guachancho*); EU606011 (*Spinibarbus hollandi*); EU199442 (*Squalius pyrenaicus*); AB084942 (*Steatocranus casuarius*); EU407251 (*Stenobrachius leucopsarus*); DQ336173 (*Stereolepis gigas*); AY775122 (*Stigmatochromis modestus*); HQ639199 (*Stiphodon atratus*); HQ639154 (*Stiphodon elegans*); HQ639150 (*Stiphodon hydroreibatus*); HQ639155 and HQ639156 (*Stiphodon rutilaureus*); HQ639152 (*Stiphodon sapphirinus*); GQ409795 (*Stomatepia pindu*); JN412565 (*Stomias boa*); JN544528 (*Stomias brevibarbatus*); HQ444183 (*Strophidon sathete*); JF966222 (*Sundadanio axelrodi* 'green'); JQ614280 (*Sundadanio cf* axelrodi BOLD:AAU2681); JF913271 (*Synagrops bellus*); EU638011 (*Synanceia verrucosa*); DQ874823 (*Synodus foetens*); AF137214 and NM_001078631 (*Takifugu rubripes*); EU082608 (*Tampichthys ipni*); AB458080 and AB458081 (*Telmatochromis temporalis*); AB458083 (*Telmatochromis vittatus*); CAAE01015033 (*Tetraodon nigroviridis*); DQ874802 (*Thunnus atlanticus*); GQ409799 (*Tilapia bakossiorum*); AB084943 (*Tilapia buttikoferi*); GQ409803 (*Tilapia cf* guineensis 'Mampfe'); AB084944 (*Tilapia rendalli*); EU638012 (*Toxotes* sp. BL-2008); EU638013 (*Trachurus trachurus*); AB084945 (*Trematocara macrostoma*); AB185238 (*Trematocara unimaculatum*); EU638014 (*Trematomus bernacchii*); EU638015 (*Trinectes maculatus*); EU638016 (*Tripterygion delaisi*); HQ729688, HQ729689, HQ729690, HQ729692, HQ729693, HQ729697, HQ729701, HQ729703 and HQ729704 (*Typhlichthys subterraneus*); AY775117 (*Tyrannochromis maculatus*); DQ100323 (*Uta stansburiana*); EU638017 (*Valenciennea strigata*); AB458068 and AB458136 (*Variabilichromis moorii*); EU919558 (*Varicorhinus barbatulus*); JN412570 (*Vinciguerria nimbaria*); EU638018 (*Xeneretmus latifrons*); NM_001087048 (*Xenopus laevis*); AB458124 (*Xenotilapia* bathyphila); AB457985 (*Xenotilapia boulengeri*); AB185239 and AB458131 (*Xenotilapia caudafasciata*); AB185240 (*Xenotilapia flavipinnis*); AB457991 and AB458125 (*Xenotilapia leptura*); AB457992 (*Xenotilapia melanogenys*); AB185241 and AB458130 (*Xenotilapia nigrolabiata*); AB084947 (*Xenotilapia ochrogenys*); AB457995 (*Xenotilapia papilio*); AB457997 and AB458123 (*Xenotilapia rotundiventralis*); AB457998 (*Xenotilapia sima*); AB185242 and AB458128 (*Xenotilapia spiloptera*); EU638019 (*Xiphias gladius*); EU638020 (*Xyrichtys novacula*); EU638021 (*Zanclorhynchus spinifer*); EU638022 (*Zeugopterus punctatus*); EU638023 (*Zeus faber*). A rhodopsin gene sequence of *Gasterosteus aculeatus* was obtained from the Ensembl database.

1. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10(3):R25.
2. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754–1760.
3. Meyer E, et al. (2009) Sequencing and de novo analysis of a coral larval transcriptome using 454 GSFlx. *BMC Genomics* 10:219.
4. Smit A, Hubley R (2008–2010) RepeatModeler Open-1.0. Available at www.repeatmasker.org. Accessed December 16, 2010.
5. Smit A, Hubley R, Green P (1996–2010) RepeatMasker Open-3.0. Available at www.repeatmasker.org. Accessed December 9, 2010.
6. Flicek P, et al. (2011) Ensembl 2011. *Nucleic Acids Res* 39(Database issue):D800–D806.
7. Star B, et al. (2011) The genome sequence of Atlantic cod reveals a unique immune system. *Nature* 477(7363):207–210.
8. Chini V, et al. (2008) Genes expressed in Blue Fin Tuna (Thunnus thynnus) liver and gonads. *Gene* 410(1):207–213.
9. Kent WJ (2002) BLAT—the BLAST-like alignment tool. *Genome Res* 12(4):656–664.
10. Stanke M, Diekhans M, Baertsch R, Haussler D (2008) Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* 24(5): 637–644.
11. Burge C, Karlin S (1997) Prediction of complete gene structures in human genomic DNA. *J Mol Biol* 268(1):78–94.
12. Majoros WH, Pertea M, Salzberg SL (2004) TigrScan and GlimmerHMM: Two open source ab initio eukaryotic gene-finders. *Bioinformatics* 20(16):2878–2879.
13. Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M (2005) Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Res* 33(20):6494–6506.
14. Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25(17):3389–3402.
15. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403–410.
16. Lowe TM, Eddy SR (1997) tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25(5):955–964.
17. Katoh K, Toh H (2008) Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform* 9(4):286–298.
18. Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3(5):418–426.
19. Li WH (1993) Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *J Mol Evol* 36(1):96–99.
20. Pamilo P, Bianchi NO (1993) Evolution of the Zfx and Zfy genes: Rates and interdependence between the genes. *Mol Biol Evol* 10(2):271–281.
21. Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24(8):1586–1591.
22. Kimura M (1981) Estimation of evolutionary distances between homologous nucleotide sequences. *Proc Natl Acad Sci USA* 78(1):454–458.
23. Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4(4):406–425.
24. Tamura K, et al. (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28(10):2731–2739.
25. Peng Z, He S, Wang J, Wang W, Diogo R (2006) Mitochondrial molecular clocks and the origin of the major Otocephalan clades (Pisces: Teleostei): A new insight. *Gene* 370:113–124.
26. Setiamarga DH, et al. (2009) Divergence time of the two regional medaka populations in Japan as a new time scale for comparative genomics of vertebrates. *Biol Lett* 5(6):812–816.
27. Matschiner M, Hanel R, Salzburger W (2011) On the origin and trigger of the notothenioid adaptive radiation. *PLoS ONE* 6(4):e18911.

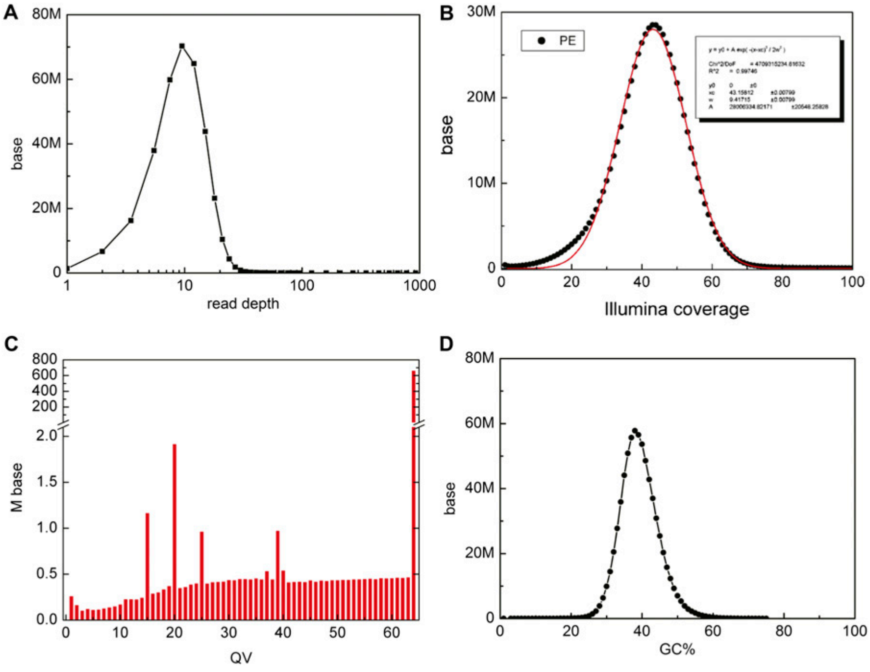
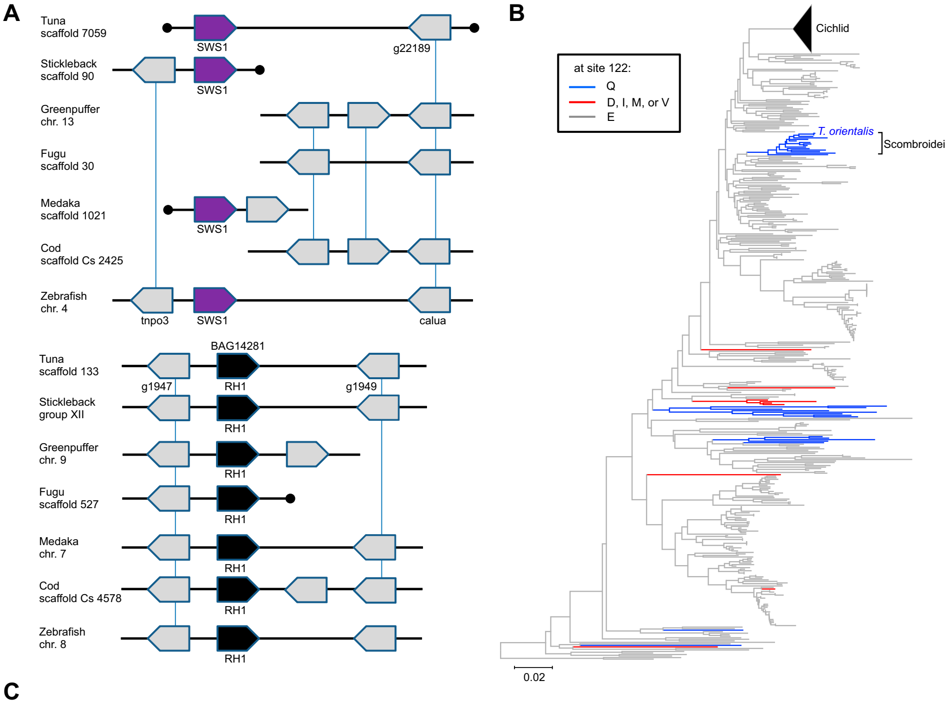


Fig. S1. Read depth, base quality and GC content. (*A* and *B*) Distribution of read depth by 454 (*A*) and Illumina (*B*) reads. (*C*) Site-by-site distribution of quality score estimated from assemblage of 454 reads. (*D*) GC content of 454 contigs with 500-bp window.



C

Red-sensitive (LWS)

Tuning site	zebrafish*		cod	medaka*		greenpuffer	fugu	stickleback	tuna
	LWS1	LWS2		LWS-A	LWS-B				
164	A	A	-	S	S	P	A	S	S
181	H	H	-	H	H	H	H	H	H
261	Y	F	-	Y	Y	Y	Y	Y	Y
269	T	T	-	T	T	T	T	T	T
292	A	A	-	A	A	A	A	A	A

*Expected λ_{max} values are summarized in Yokoyama S, et.al. (2008) *Genetics* 179(4):2037-2043.
zebrafish LWS1, 555nm; LWS2, 545nm; medaka LWS-A and B, 560nm.

UV-sensitive (SWS1)

Tuning site	zebrafish	cod	medaka	greenpuffer	fugu	stickleback	tuna
46	F	-	F	-	-	F	F
49	I	-	F	-	-	F	F
52	T	-	T	-	-	T	T
86	F	-	F	-	-	F	F
90	S	-	S	-	-	S	S
93	Q	-	Q	-	-	Q	Q
114	A	-	A	-	-	A	S
118	S	-	A	-	-	S	S

Blue-sensitive (SWS2)

Tuning site	zebrafish	cod		medaka		greenpuffer	fugu	stickleback	tuna	
		SWS2A	SWS2B	SWS2A	SWS2B	SWS2B	SWS2B	SWS2A	SWS2A	SWS2B
46	F	F	F	F	V	C	F	F	F	F
49	I	I	V	V	V	V	V	I	V	V
52	T	T	T	T	T	T	T	T	T	T
93	V	T	T	T	T	T	T	T	T	T
94	A	A	C	A	C	C	C	A	A	C
116	T	V	T	A	T	A	A	L	L	T
118	T	T	T	T	A	T	T	T	T	T
164	A	A	A	A	G	A	A	A	A	A
207	L	L	L	L	L	I	I	L	L	L
269	A	A	A	A	A	A	A	A	A	A
295	C	S	S	S	S	S	S	S	S	S

Rhodopsin (RH1)

Tuning site	zebrafish	cod	medaka	greenpuffer	fugu	stickleback	tuna
87	V	V	V	V	V	V	V
90	G	G	G	G	G	G	G
117	A	A	A	A	A	A	A
118	T	T	T	T	T	T	T
120	G	G	G	G	G	G	G
121	G	G	G	G	G	G	G
125	L	L	L	L	L	L	L
126	K	W	W	W	W	W	W
164	A	A	A	A	A	A	A
181	E	E	E	E	E	E	E
197	G	G	G	G	G	G	G
211	H	H	H	H	H	T	H
265	W	W	W	W	W	W	W
268	Y	Y	Y	Y	Y	Y	Y
269	A	A	A	A	A	A	A
295	A	A	A	A	A	A	A

Green-sensitive (RH2)

Tuning site	zebrafish				cod		medaka			greenpuffer	fugu	stickleback		tuna				
	RH2-1	RH2-2	RH2-3	RH2-4	RH2A1	RH2A2	RH2-A	RH2-B	RH2-C			RH2A1	RH2A2	RH2A1	g6738	RH2A2	g6740	RH2B
49	I	C	C	C	C	C	L	C	C	C	C	C	C	C	C	C	C	C
52	F	L	F	F	T	T	F	T	T	T	T	T	T	T	F	T	F	F
83	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
164	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
295	S	S	S	S	A	A	S	A	A	A	A	A	A	A	S	A	S	S

Fig. S2. Synteny of opsin gene loci and spectral tuning sites. (*A*) Synteny of UV pigment genes (*Upper*) and rhodopsin genes (*Lower*) compared with six teleosts whose genomes have been sequenced. Colors indicate different gene families (e.g., purple for UV pigment genes and black for rhodopsin genes). Gray-colored genes flank the pigment genes, and vertical bars indicate orthology. The gene names in zebrafish comply with previous studies, and others are named in this work (g22189, g1947, and g1949 of tuna are named in the order of genes predicted in scaffolds). Filled circles indicate the scaffold ends. Each gene is arranged in order and indicating transcriptional direction, and the physical distance is ignored. (*B*) A neighbor-joining tree of fish rhodopsin (RH1) sequences. A total of 557 nucleotide sequences including that for tuna (accession no. BAG14281) were collected from the GenBank database (*SI Materials and Methods*). The species that have glutamine (Q) at site 122 are colored in blue. For site 122, the species with aspartic acid (D), isoleucine (I), methionine (M), or valine (V) are colored in red, and those with glutamic acid (E) are colored in gray. (*C*) Representative amino acid sites involved in the light sensitivity of red, UV, blue, rhodopsin, and green opsin (1–8) in comparison with seven teleosts. For rhodopsin and green opsin, the amino acid sites that are not shown in Fig. 1 are summarized. The site numbers are standardized to those of bovine rhodopsin.

1. Yokoyama S (2000) Molecular evolution of vertebrate visual pigments. *Prog Retin Eye Res* 19(4):385–419.
2. Hunt DM, Dulai KS, Partridge JC, Cottrill P, Bowmaker JK (2001) The molecular basis for spectral tuning of rod visual pigments in deep-sea fish. *J Exp Biol* 204(Pt 19):3333–3344.
3. Shi Y, Radlwimmer FB, Yokoyama S (2001) Molecular genetics and the evolution of ultraviolet vision in vertebrates. *Proc Natl Acad Sci USA* 98(20):11731–11736.
4. Yokoyama S (2002) Molecular evolution of color vision in vertebrates. *Gene* 300(1-2):69–78.
5. Takahashi Y, Ebrey TG (2003) Molecular basis of spectral tuning in the newt short wavelength sensitive visual pigment. *Biochemistry* 42(20):6025–6034.
6. Yokoyama S, Tada T (2003) The spectral tuning in the short wavelength-sensitive type 2 pigments. *Gene* 306:91–98.
7. Chinen A, Matsumoto Y, Kawamura S (2005) Spectral differentiation of blue opsins between phylogenetically close but ecologically distant goldfish and zebrafish. *J Biol Chem* 280(10):9460–9466.
8. Yokoyama S, Yang H, Starmer WT (2008) Molecular basis of spectral tuning in the red- and green-sensitive (M/LWS) pigments in vertebrates. *Genetics* 179(4):2037–2043.

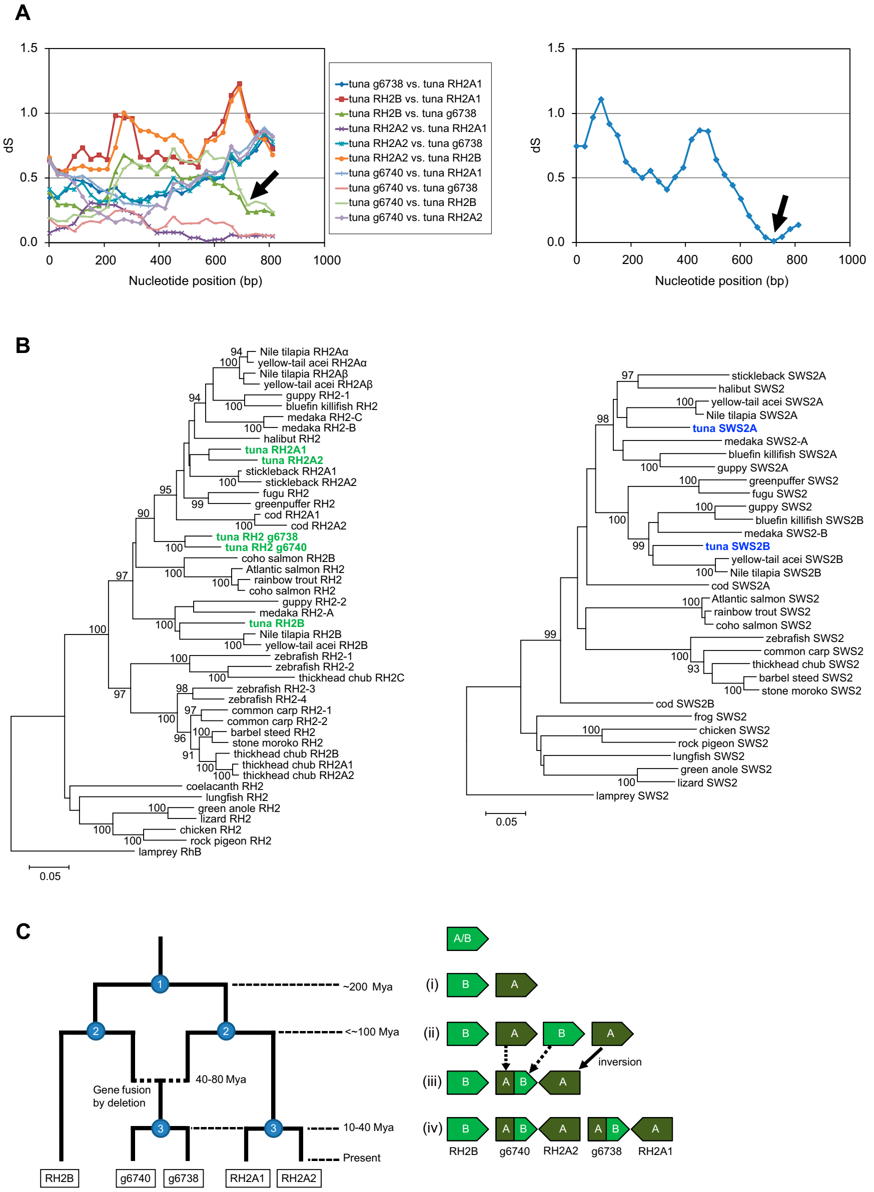


Fig. S3. Sequence comparison of green and blue opsin genes. (*A*) Local synonymous substitution rates of tuna opsin genes by window analysis. Green (*Left*) and blue (*Right*) opsin gene sequences in pairwise alignments were scanned with 240-bp window. Nucleotide position indicates the first site in window sequence. The positions around which *dS* values drop, probably due to gene conversion, are shown by arrows. Note that, in green-pigment genes, *dS* values of N-terminal regions are generally smaller than those of C-terminal regions. Therefore, it is uncertain whether gene conversion events have also occurred in N-terminal regions. (*B*) Phylogenetic trees of green (*Left*) and blue (*Right*) pigment genes. The trees were constructed by using N-terminal sequences in the multiple alignments. Numbers at branches indicate the bootstrap probabilities (≥90%) with 1,000 replicates. (*C*) Another possible evolutionary pathway of tuna RH2 genes. (*Left*) Each node with a number in the dendrogram indicates a duplication event. *Right* shows a diagram of the changes to the RH2 locus structure during the course of evolution. (*i*) An RH2 gene was duplicated into RH2A and RH2B (~200 Mya). (*ii*) A tandem duplication then doubled the two copies to four (<~100 Mya). (*iii*) A gene fusion by deletion reduced a pair of RH2A and RH2B to a single gene (40–80 Mya). (*iv*) A tandem duplication doubled two copies to four (10–40 Mya).

Table S1. Assembling and scaffolding of the Pacific bluefin tuna genome

	454 reads			Illumina reads	
Read type	Single	Paired-end		Paired-end	
		Insert = 3 kb	20 kb	Insert = 300 bp	600 bp
Number	24,865,968	4,949,157	2,228,697	281,702,212	177,817,566
Length (mean)	331 bp	190 bp	151 bp	75 bp	75 bp
		Contigs (>500 bp)*			
Number	192,169				
Length (total)		732.9 Mb			
(Longest)		79,054			
(Mean)		3,813			
	Scaffolds + single contigs (>500 bp)*			Scaffolds (>2,000 bp)[†]	
Number	98,673			16,802	
Length (total)	786.6 Mb			740.3 Mb	
(Longest)	589,244			1,021,118	
(Mean)	7,972			44,064	
(N50)	Not assessed			136,950	

*Assembly with Newbler v.2.5 (Roche Diagnostics).

[†] Manual bridging with Illumina paired-end reads mapped onto 454 scaffolds and contigs with Bowtie (Version 0.12.7; ref. 1).

1. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10(3):R25.

Table S2. Bridging 454 contigs and scaffolds with Illumina paired-end reads

Insert size	Total pairs (%)	Exact matched reads	Both mapped (%)	One mapped (%)	Unmapped (%)	Same contig (%)	Bridged* (%)	Valid[†]
300 bp	140,851,106 (100)	196,162,796	76,541,140 (54)	43,080,516 (31)	21,229,450 (15)	73,755,148 (96)	2,785,992 (4)	521,223
600 bp	88,908,783 (100)	117,206,899	42,976,076 (48)	31,254,747 (35)	14,677,960 (17)	40,736,369 (95)	2,239,707 (5)	905,102

*Only both exact matched reads were used for bridging.

[†] Only reads matched at less than the Illumina paired-end insert size from the end of 454 contig or scaffold were used for bridging.

Table S3. Overall metrics of the normalized cDNA assembly of 10 tissues from an adult *T. orientalis*

Assembly	Number	Full length*	N50, bp	Average	Longest	Total length, Mb
Reads	3,805,912					1,232.0
Contigs/isotigs[†]	180,512	5,741	836	783	13,541	141.4
Singletons	981,043	0	421	326	1,095	320.6

The 10 tissues were fin, gill, heart, intestine, kidney, liver, red muscle, white muscle, pyloric caeca, and stomach.

*Sequences at which both 5′ and 3′ adaptor sequences are recovered.

[†] Contigs containing reads which are shared by other contigs. Isotigs sharing reads may be splicing variants of a single gene region.

Table S4. Metrics of protein-coding regions

Mapping result	Input sequences	Mapped to scaffold	Predicted	Considered as true positives	Hinted
cDNA from *T. orientalis* (this study)	178,992	174,351*/151,887[†]			
cDNA from *T. thynnus*	10,131	8,322*/4,378[†]			
Total	189,123	156,265[†]			
ENSEMBL 5 fish genomes (tblastn)	161,585	23,908[‡]			
Ab initio gene prediction: program (model species)					
GenScan (human)			39,163	2,231	
GlimmerHMM (zebrafish)			103,734	2,493	
GeneMark-ES (none)			157,337	2,488	
Augustus (human)			25,866	2,623	
Nonredundant total				5,583	
Prediction result by tuna gene model: gene type					
Full-length			19,281		16,948
Only N-terminal			2,598		1,743
Only C-terminal			3,384		2,469
Internal			1,170		549
Total			26,433[§]		21,709

*Mapped by BLAT (1).

[†] After BLAT mapping, selected by a program filterPSL.pl included in the AUGUSTUS package.

[‡] Overlapped sequences are excluded.

[§] Five gene regions are subsequently and manually split into two genes, respectively.

1. Kent WJ (2002) BLAT—the BLAST-like alignment tool. *Genome Res* 12(4):656–664.

Table S5. Synonymous and nonsynonymous substitution rates between opsin genes

Gene pair	Aligned codons	*dS* ± SD	*dN* ± SD	*dN*/*dS*
Green pigment gene (RH2)				
Tuna RH2 g6738 vs. tuna RH2A1*	354	0.4400 ± 0.0534	0.0990 ± 0.0117	0.2250
Tuna RH2 g6740 vs. tuna RH2 g6738*	345	0.1561 ± 0.0279	0.0436 ± 0.0076	0.2793
Tuna RH2 g6740 vs. tuna RH2A1*	345	0.5038 ± 0.0595	0.1084 ± 0.0125	0.2152
Tuna RH2 g6740 vs. tuna RH2A2*	344	0.4505 ± 0.0551	0.0932 ± 0.0115	0.2069
Tuna RH2 g6740 vs. tuna RH2B*	345	0.3892 ± 0.0505	0.0858 ± 0.0109	0.2205
Tuna RH2A2 vs. tuna RH2 g6738*	351	0.4639 ± 0.0569	0.0936 ± 0.0115	0.2018
Tuna RH2A2 vs. tuna RH2A1*	351	0.1097 ± 0.0222	0.0610 ± 0.0090	0.5561
Tuna RH2A2 vs. tuna RH2B	344	0.7115 ± 0.0828	0.1430 ± 0.0145	0.2010
Tuna RH2B vs. tuna RH2 g6738*	345	0.4193 ± 0.0556	0.1033 ± 0.0121	0.2464
Tuna RH2B vs. tuna RH2A1	345	0.6941 ± 0.0793	0.1479 ± 0.0148	0.2131
Medaka RH2-B vs. medaka RH2-C*	348	0.0524 ± 0.0146	0.0259 ± 0.0058	0.4943
Medaka RH2-A vs. medaka RH2-B	345	1.1584 ± 0.1438	0.1540 ± 0.0151	0.1329
Medaka RH2-A vs. medaka RH2-C	345	1.0852 ± 0.1308	0.1433 ± 0.0145	0.1320
Nile tilapia RH2Aa vs. Nile tilapia RH2B	348	0.9014 ± 0.1038	0.1285 ± 0.0136	0.1426
Nile tilapia RH2Ab vs. Nile tilapia RH2B	342	0.8775 ± 0.1012	0.1184 ± 0.0130	0.1349
Yellow-tail acei RH2B vs. yellow-tail acei RH2Aa	349	0.9313 ± 0.1105	0.1353 ± 0.0140	0.1453
Yellow-tail acei RH2B vs. yellow-tail acei RH2Ab	350	1.0097 ± 0.1206	0.1301 ± 0.0136	0.1289
Average *dS* between RH2A and RH2B		0.921		
Blue pigment gene (SWS2)				
Tuna SWS2A vs. tuna SWS2 SWS2B*	351	0.4404 ± 0.0514	0.0865 ± 0.0109	0.1964
Medaka SWS2-A vs. medaka SWS2-B	349	1.0642 ± 0.1206	0.1430 ± 0.0146	0.1344
Bluefin killifish SWS2A vs. bluefin killifish SWS2B	351	1.1501 ± 0.1343	0.1797 ± 0.0166	0.1562
Nile tilapia SWS2B vs. Nile tilapia SWS2A	351	0.7695 ± 0.0848	0.1548 ± 0.0153	0.2012
Altantic cod SWS2A vs. Altantic cod SWS2B	351	0.6948 ± 0.0820	0.1600 ± 0.0156	0.2303
Yellow-tail acei SWS2A vs. yellow-tail acei SWS2B	348	0.8243 ± 0.0917	0.1633 ± 0.0158	0.1981
Guppy SWS2A vs. guppy SWS2	348	1.1678 ± 0.1575	0.1472 ± 0.0148	0.1260
Average *dS* between SWS2A and SWS2B		0.945		

*Not used for average calculation.

Table S6. Synonymous substitution rate and divergence time between teleosts

Teleost pair	Average *dS**	Divergence time, Mya	Reference
Zebrafish vs. tuna, cod, medaka, stickleback, fugu, and greenpuffer	1.4625	289–307	1, 2
Cod vs. tuna, medaka, stickleback, fugu, and greenpuffer	0.9672	158–208	2, 3
Tuna vs. medaka, stickleback, fugu, and greenpuffer	0.6074	131	3
Medaka vs. stickleback, fugu and greenpuffer	0.8321	125–180	2, 3
Stickleback vs. fugu and greenpuffer	0.6723	110–171	2, 3
Fugu vs. greenpuffer	0.3504	43–78	2, 3

*Of 6,170 orthologous pairs, 4,991 were used for calculation.

1. Peng Z, He S, Wang J, Wang W, Diogo R (2006) Mitochondrial molecular clocks and the origin of the major Otocephalan clades (Pisces: Teleostei): A new insight. *Gene* 370:113–124.
2. Setiamarga DH, et al. (2009) Divergence time of the two regional medaka populations in Japan as a new time scale for comparative genomics of vertebrates. *Biol Lett* 5(6):812–816.
3. Matschiner M, Hanel R, Salzburger W (2011) On the origin and trigger of the notothenioid adaptive radiation. *PLoS ONE* 6(4):e18911.