Recurrent convergent evolution at amino acid residue 261 in fish rhodopsin

Hill, Jason; Enbody, Erik D; Pettersson, Mats E; Sprehn, C Grace; Bekkevold, Dorte; Folkvord, Arild; Laikre, Linda; Kleinau, Gunnar; Scheerer, Patrick; Andersson, Leif

Published in:
Proceedings of the National Academy of Sciences of the United States of America

Link to article, DOI:
10.1073/pnas.1908332116

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Recurrent convergent evolution at amino acid residue 261 in fish rhodopsin

Jason Hill, Erik D. Enbody, Mats E. Pettersson, C. Grace Sprehn, Dorte Bekkevold, Arild Folkvord, Linda Laikre, Gunnar Kleinau, Patrick Scheerer, and Leif Andersson

Department of Medical Biochemistry and Microbiology, Uppsala University, SE-751 23 Uppsala, Sweden; National Institute of Aquatic Resources, Technical University of Denmark, 8600 Silkeborg, Denmark; Department of Biological Sciences, University of Bergen, 5020 Bergen, Norway; Institute of Marine Research, 5018 Bergen, Norway; Department of Zoology, Division of Population Genetics, Stockholm University, SE-106 91 Stockholm, Sweden; Department of Analytical Breeding and Genetics, Swedish University of Agricultural Sciences, SE-750 07 Uppsala, Sweden; and Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, TX 77843

Contributed by Leif Andersson, July 26, 2019 (sent for review May 14, 2019; reviewed by Michel Georges and Philip W. Hedrick)

The evolutionary process that occurs when a species colonizes a new environment provides an opportunity to explore the mechanisms underlying genetic adaptation, which is essential knowledge for understanding evolution and the maintenance of biodiversity. Atlantic herring has an estimated total breeding stock of about 1 trillion (10¹²) and has colonized the brackish Baltic Sea within the last 10,000 y. Minute genetic differentiation between Atlantic and Baltic herring populations at selectively neutral loci combined with this rapid adaptation to a new environment facilitated the identification of hundreds of loci underlying ecological adaptation. A major question in the field of evolutionary biology is to what extent such an adaptive process involves selection of novel mutations with large effects or genetic changes at many loci, each with a small effect on phenotype (i.e., selection on standing genetic variation). Here we show that a missense mutation in rhodopsin (Phe261Tyr) is an adaptation to the red-shifted Baltic Sea light environment. The transition from phenylalanine to tyrosine differs only by the presence of a hydroxyl moiety in the former, but this results in a red-shifted light absorbance of the receptor. Remarkably, an examination of the rhodopsin sequences from 2,056 species of fish revealed that the same missense mutation has occurred independently and been selected for during at least 20 transitions between light environments across all fish. Our results provide a spectacular example of convergent evolution and how a single amino acid change can have a major effect on ecological adaptation.

Significance

Fish occupy a wide range of light environments that vary depending on the wavelengths absorbed by the local environment, including brackish and freshwater environments in which dissolved organic matter produces a red-shifted visual environment compared with marine waters. Atlantic herring colonized the brackish Baltic Sea after its formation approximately 10,000 y ago, which required adaptation to a red-shifted light environment. We show that visual adaptation to the new light environment was achieved through a recent and rapid selective sweep on a mutation in the rhodopsin gene. Furthermore, this exact same amino acid change has occurred at least 20 separate times in fish species transitioning from marine to brackish or freshwater environments. This is a remarkable example of convergent evolution.


Reviewers: M.G., University of Liege; and P.W.H., Arizona State University.

The authors declare no conflict of interest.

Data deposition: The sequence reported in this paper has been deposited in the European Nucleotide Archive (ENA) database (accession no. ENA PRJEB32358).

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

Published online August 26, 2019.

www.pnas.org/cgi/doi/10.1073/pnas.1908332116

PNAS | September 10, 2019 | vol. 116 | no. 37 | 18473–18478
Ile213Thr occurs in the same exon as Phe261Tyr, only 144 bp apart. These 2 missense mutations are absent in herring from the Atlantic Ocean, but occur in populations from the Baltic Sea, which supports their derived status, and Phe261Tyr is fixed in some populations from the inner Baltic (Fig. 1 C and D).

We further explored the hard sweep associated with the rhodopsin locus, using a dataset comprising 1,170 fish from 24 localities both from the Swedish west coast (high salinity) and the brackish Baltic Sea (SI Appendix, Supplementary Text: Estimating the speed and age of the selective sweep at the rhodopsin locus). These fish have been individually genotyped for a previously developed 45k SNP chip (1), which fortunately included both missense mutations, Ile213Thr and Phe261Tyr. This analysis confirmed strong differentiation between Atlantic and Baltic populations for both SNPs and revealed complete linkage disequilibrium (LD) between the 2 sites, showing that the Ile213Thr substitution must have arisen on the Tyr261 haplotype (SI Appendix, Table S1 and Fig. S1). The younger age of Thr213 was further supported by the observation that its associated sweep region is larger than that associated with Tyr261 (SI Appendix, Fig. S2). Furthermore, an analysis of the LD data indicated that the time for Tyr261 to sweep through the population was in the range of 23 to 65 generations, corresponding to about 140 to 390 y, given an estimated generation interval of about 6 y (SI Appendix, Supplementary Text: Speed of the sweep and Fig. S3). A very high selection coefficient (s ≥ 0.2) per generation is required to generate such a rapid sweep. Similarly, we used the LD decay between the missense mutations and flanking markers to estimate their age, which resulted in estimates of 6,400 and 42,000 y for Ile213Thr and Phe261Tyr, respectively (SI Appendix, Supplementary Text: Age of the sweep and Fig. S4). The latter is older than the Baltic Sea (formed about 10,000 y ago), and the estimate may be inflated, since it is highly dependent on the estimated recombination rate for the region, but it is possible that Phe261Tyr mutation occurred before the colonization of the Baltic Sea and swept to high frequency during the colonization.

Structure–Function Relationships. To explore whether Phe261Tyr and Ile213Thr could be causal mutations underlying genetic adaptation in the Baltic Sea, we built a structural model of herring rhodopsin based on the crystal structure of bovine rhodopsin to estimate the potential effect of amino acid substitutions Phe261Tyr and Ile213Thr on variations in light sensitivity. Residue 261 is located in transmembrane helix 6 (TMH6) close to the retinal binding site (Fig. 2A). The latter is older than the Baltic Sea (formed about 10,000 y ago), and the estimate may be inflated, since it is highly dependent on the estimated recombination rate for the region, but it is possible that Phe261Tyr mutation occurred before the colonization of the Baltic Sea and swept to high frequency during the colonization.
and Tyr/Phe261, as observed for the corresponding position in active or inactive bovine rhodopsin structures (6–9). Of note, exactly at this region between TMH5 and TMH6, a channel-like opening for potential retinal release or entry was previously described (10, 11). The exact structural function of amino acids at residue 213 is not yet established (SI Appendix, Supplementary Text: Analysis of structural herring rhodopsin model), but substitutions at this site are associated with divergence between riverine and lake dwelling cichlid species, suggesting it may affect vision (5).

Residue 261 has been identified as 1 of 27 known key spectral tuning sites in opsins, where amino acid changes lead to shifts in light absorption (12). Blue, green, and red cone opsins are paralogs to rhodopsin, with differences in absorption maxima, and are responsible for color vision in humans and other primates. Residue 261 in rhodopsin corresponds to positions 258 and 277 in blue and green opsins, respectively, which have phenylalanine at this position, similar to the Atlantic herring rhodopsin, whereas red opsin (position 277), which has a red-shifted light absorption, has tyrosine, similar to the Baltic herring rhodopsin (SI Appendix, Fig. S5). This cone opsin Phe277Tyr substitution, among other mutations, was previously identified as causal for the spectral red shift from 547 to 556 nm in primate visual pigments for squirrel (Phe) versus tamarin (Tyr) monkeys, respectively (13). Furthermore, introduction of a Phe261Tyr substitution in bovine rhodopsin by in vitro mutagenesis causes a significant 10-nm red shift from 500 to 510 nm (14). The Mexican cave fish (Astyanax mexicanus), similar to the Baltic herring, has tyrosine at residue 261 in rhodopsin, and in vitro mutagenesis found that a Tyr261Phe substitution caused an 8-nm blue shift in this molecule (3). It has previously been shown that the rod cells in the retina of the Baltic herring have a 10-nm red-shifted light absorption compared with the Atlantic herring (15), in perfect agreement with the effect of the Phe261Tyr substitution (14). Thus, the retina of a Baltic herring with the rhodopsin Tyr261 variant can catch more photons than a retina with the Phe261 variant in the red-shifted light environment in the Baltic Sea (Fig. 1D), and likely improves the ability of the fish to detect prey and avoid predators in this environment. We also tested the hypothesis that regulatory and structural changes could occur in parallel in a locus under selection by testing for differences in gene expression between the Phe261 and Tyr261 allele. However, no evidence for differential gene expression between the 2 alleles (i.e., allelic imbalance) was found in the retina from Atlantic × Baltic herring hybrids (SI Appendix, Fig. S6).

Convergent Evolution of Rhodopsin at Residue 261 among Fish Adapted to Brackish and Freshwater. The coding sequence of rhodopsin has been widely implemented as a phylogenetic marker, and as a consequence, sequences from the single rhodopsin exon in 2,056 species of fish are publicly available (16). We leveraged this dataset to explore variation at residues 213 and 261, the codons carrying missense mutations in the Baltic herring, across the fish tree of life. The data show that only phenylalanine and tyrosine are tolerated at residue 261 (SI Appendix, Table S2). In fact, an aromatic residue is highly conserved at this position among all G protein-coupled receptors (Ballesteros & Weinstein numbering 6.44 [17]; sequence conservation ~75% Phe, ~7% Tyr; SI Appendix, Table S3). The aromatic character of this residue is essential for ligand-induced receptor activation by a complex interplay with Trp6.48 (Fig. 2B) and other hydrophobic residues in the TMH5–TMH6 interface (positions 5.47 and 5.51). Residue 213 is less constrained and not as structurally conspicuous as residue 261, and the 2 variants present in herring (Ile and Thr) are not the most common amino acids at this residue across fish species (SI Appendix, Table S2).

The evidence that Phe261Tyr is an adaptive mutation in the Baltic herring encouraged us to explore whether there is an association between the presence of Tyr261 and habitat, as brackish waters in general, as well as freshwater environments, have a red-shifted light environment compared with oceanic waters (15). Habitat data from 2,056 fish species (18) revealed that there is a highly significant association ($P = 1.2 \times 10^{-85}$, Fisher’s Exact test) between the presence of Tyr261 and freshwater or brackish environment (Fig. 3 and Table 1). As many as 356 species living in freshwater and/or brackish water carry this variant. In contrast, only 10 of 885 marine fish species carry Tyr261: 6 of

![Crystal structure of bovine rhodopsin and structural homology model of herring rhodopsin.](image-url)
them are bathymetric (deep-sea) species (Fig. 3, clades 8, 11, and 13), and 4 are nocturnal, shallow reef- and lagoon-dwelling Indo-Pacific squirrelfish (Fig. 3, clade 15); it has recently been shown that vision in several deep-sea fishes, including those in the Myctophidae family (clade 13 in Fig. 3), has evolved by rhodopsin gene duplications and diversification (19). Furthermore, it is clear from examining the fish phylogeny that the adaptive change Phe261Tyr must have occurred independently more than 20 times during the evolution of bony fishes (Fig. 3), providing the most striking example of recurrent convergent evolution at a single amino acid residue in a vertebrate.

Residue 213 has not been identified as a key spectral tuning site in opsins (12). It is more difficult to assess the biological significance of an association to habitat for the polymorphism at residue 213 due to the high variability at this position, and the highly significant association with habitat may primarily reflect a phylogenetic signal (SI Appendix, Table S4). In herring, the fact that the haplotype carrying the Ile213Thr substitution reaches a frequency of 0.50 and above in populations from the inner Baltic Sea (SI Appendix, Fig. S1), combined with our LD analysis showing that the Ile213Thr mutation is considerably younger than Phe261-Tyr (SI Appendix, Supplementary Text: Estimating the speed and age of the selective sweep at the rhodopsin locus), provide evidence that this haplotype is under selection because genetic drift is minute in the herring (1, 20); the great majority of SNPs show no allele frequency difference between the inner Baltic Sea and the Atlantic.
Ocean. However, since there is not yet conclusive evidence for a functional significance of Ile213Thr, as there is for Phe261Tyr, we cannot exclude the possibility that the former mutation is hitchhiking to high frequency on another second causal mutation from this genomic region. Nevertheless, the results illustrate the evolution of adaptive haplotypes by accumulation of multiple closely linked causal variants in a natural population, as previously documented during phenotypic evolution in domestic animals (21).

Natural Selection at Rhodopsin and Life History Parameters. Functional evidence and the signal of selection presents a very convincing association between the presence of rhodopsin Tyr261 and the light conditions in the Baltic Sea, but there are some discrepancies worth considering. One discrepancy concerns populations spawning in the southwest Baltic (numbers 2 and 18 to 23) that all have a low frequency of Tyr261 (Fig. 1D). The light conditions in this area are less red-shifted than in other parts of the Baltic (Fig. 1D), and many populations that spawn in this area forage in the marine waters of Kattegat and Skagerrak (22). The other discrepancy concerns differences between autumn- and spring-spawning herring, which is highlighted by populations from the Bay of Gävle and the Gulf of Riga, all collected at spawning (Fig. 1D). The samples from Gävle (numbers 24, 29, and 32) were collected in May (spring), July (summer), and September (autumn), and their frequencies of Tyr261 were 77%, 66%, and 26%, respectively. In the Gulf of Riga, the 2 samples of spring-spawning fish (numbers 33 and 37) were fixed or close to fixation for Tyr261, whereas the 2 autumn-spawning collections (numbers 25 and 26) had allele frequencies as low as 24%. One possible explanation for these differences is that Tyr261 first occurred in spring-spawning herring and has not yet reached high frequencies in autumn-spawning fish due to limited gene flow (20). An alternative explanation is that spring- and autumn-spawning herring are foraging for plankton in different areas or at different depths with different light conditions, thus experiencing different selective pressures. Interestingly, it has been documented that spring-spawning Baltic herring have larger eyes than autumn-spawning herring, supporting a difference in visual ability between these ecotypes of Baltic herring (23). A third possibility is that seasonal fluctuations in phytoplankton population dynamics and/or seasonal river runoff dynamics would result in spring-spawning herring larvae having to cope with a more red-shifted light environment than autumn-spawning juveniles at the same age (SI Appendix, Fig. S7). Under this hypothesis, climate-driven fluctuations in Baltic turbidity may differentially affect spring and autumn-spawning populations because larval survival is critical for herring recruitment and is influenced by temperature (24).

We addressed the question of whether selection at the rhodopsin locus is most important at the larval/juvenile stage or at the adult stage by determining a partial rhodopsin sequence from 29 individuals representing 3 populations each of Atlantic salmon (Salmo salar) and brown trout (Salmo trutta). Both species have anadromous populations in which breeding, egg laying, and the juvenile stage take place entirely in freshwater, but adults spend the majority of their lives in open marine waters to feed. The 3 populations sampled from each species were chosen such that in 1 population, adults remained in freshwater, in 1 they moved to feed in brackish water, and in 1 they moved to feed in marine waters. All individuals from both species, irrespective of the habitat of adult fish, carried only Tyr261, the expected residue for a freshwater fish (SI Appendix, Table S5). In contrast, the European eel (Anguilla anguilla) and the Japanese eel (Anguilla japonica), which have their juvenile stage in the ocean but live most of their life in freshwater, both have Phe261. The data are consistent with the hypothesis that selection on rhodopsin codon 261 is most important during the larval and juvenile stages, when mortality rates are the highest (25) in species in which visual acuity is critical for predator avoidance and feeding.

Conclusions

The genes encoding visual pigments (blue, red, and green opsins, as well as rhodopsin) have been under strong positive selection as species have adapted to different light environments (12). This is illustrated by a recent finding that vision in deep-sea fish has evolved by copy number expansion of rhodopsin genes followed by positive selection for gene diversification, resulting in an expansion of the retinal light absorption spectra (19). Our findings demonstrate that selection has co-opted the exact same amino acid substitution in rhodopsin that occurs in Baltic herring more than 1 million times independent times. This has created a broad pool of alleles adapted to similar red-shifted light environments, and it is remarkable that a single amino acid site can have such an important role in ecological adaptation. Other examples in which the exact same mutation is associated with the same or very similar phenotypic effects are rare enough to be notable; for instance, a missense mutation His207Arg in the melanocortin 1 receptor is associated with light plumage color in red-footed boobies and in the ruff (26). However, we believe that the number of described instances of evolution converging on a phenotype by way of identical recurrent mutations will increase as high-quality, well-annotated genome assemblies become available for an increasing number of species.

Methods

Whole-Genome Sequencing. Tissues from 47 to 100 fish were collected from different localities in the Baltic Sea, Skagerrak, Kattegat, the Atlantic Ocean, and the Pacific Ocean (Dataset S1). Pooled samples were sequenced as previously described (1). Illumina 2 × 100 bp short read sequences for the new populations were analyzed together with read sets that were generated for populations used previously (1, 20, 27). Reads from each population pool were mapped to a chromosome-scale reference genome of Clupea harengus (ACCC GCA_90007045), using Bowtie2 v.2.2.1 (28), followed by duplicate removal and read group assignment with picard v2.10.3 (http://broadinstitute.github.io/picard). SNP calling was performed using a standard GATK pipeline. Quality filtering of raw variant calls was performed using GATK v3.8.0 (29), with the following cutoffs: DP ≥ 10, DP ≤ 400, and GQ ≥ 10. To reduce sampling error due to low coverage and inaccurate variation estimates due to incorrect mapping to repetitive regions the median read coverage was calculated to be 40× coverage across the 38 populations, and filtered variants were processed with a custom R script to filter populations at a variant site with less than 10, or more than 100, reads mapped. Sites for which more than 12 of 38 populations (>30%) of populations were missing data or filtered were also removed to prevent inflated significance values caused by unbalanced comparison groups. Inflated significance values could also result from higher-than-average read depth in some samples, so allele counts for each population at each site were downsampled to 40 when the read depth exceeded the median value. For the genome-wide scan for selection (Fig. 1A), populations were grouped by salinity as marine (Atlantic and Pacific Ocean) or brackish (Baltic Sea) and a contingency $\chi^2$ test was performed for the allele frequency estimates at each position. For the delta allele frequency estimates of Baltic Sea alleles (Fig. 1B), we used all SNPs from the interval 10 to 15 Mb on chromosome 4 that were fixed for 1 allele in Atlantic populations but were variable in Baltic populations.

Allelic Imbalance. We tested for the presence of allelic imbalance as a possible contributor to the phenotype by sequencing rhodopsin RNA from 5 hybrids of Baltic and Atlantic herring. We extracted RNA from the eye using a Qiagen
RNAeasy minikit (Qiagen) and DNA from tissue using custom salt preparations. We generated cDNA using applied biosystems High-Capacity cDNA Reagents Kit (Thermo Fisher Scientific) and amplified a 315-bp region of the single rhodopsin exon using primers designed to span both the Phe261Tyr and Ile213Thr mutations (F: AGGGTGTTGAATGAGCAATG, R: CGCGTCTACAACGAGCTG). PCR conditions were 10 min at 95 °C followed by 35 cycles of 15 s at 95 °C, 30 s at 58 °C, and 30 s at 72 °C and a final extension of 10 min at 72 °C. Both genomic DNA and cDNA of the eye was sequenced to detect any sequencing bias. Segments were sequenced by the Eurofins sequencing platform and allelic imbalance checked by eye. No detectable level of allelic imbalance was detected in the cDNA of rhodopsin. Sanger sequencing traces of the region near position 261 are shown in SI Appendix, Fig. S5.

Environmental Variables. To visualize marine absorbance (Fig. 1D), we extracted average absorbance between 2002 and 2019 at 412-nm wavelength from the NASA Moderate Resolution Imaging Spectroradiometer (MODIS-Aqua), accessed on February 4, 2019, from OceanColor Web (30). Marine absorbance is derived from a provisional Generalized Inherent Optical Properties algorithm, which returns spectral marine absorbance for water column constituents based on MODIS-Aqua satellite data. This study has been conducted using E.U. Copernicus Marine Service Information to extract salinity data.

Trout and Salmon Rhodopsin Sequencing. Atlantic salmon and brown trout samples were collected from locations around Sweden and Norway (SI Appendix, Table S5) and were kept in a frozen tissue bank collection at Stockholm University (maintained by L.L. and Prof. Nils Ryman). Genomic DNA was extracted from muscle tissue using Qiagen’s DNeasy Blood and Tissue Kit according to the manufacturer’s protocol (Qiagen, Hilden, Germany), including RNase A treatment. Sample DNA quality was assessed by visual inspection of DNA fragmentation on agarose gels and absorbance at 260/280. We designed primers spanning the rhodopsin exon positions 213 and 216 using the salmon rhodopsin sequence (accession: EU492188.1; F: GGCTCGGATATCCCGA, R: TTGTTACAGGCGGCAACTCT). We amplified a 380-bp fragment for 29 salmon and trout (SI Appendix, Table S5). PCR conditions were 10 min at 95 °C, followed by 35 cycles of 15 s at 95 °C, 30 s at 59 °C, and 30 s at 72 °C and a final extension of 10 min at 72 °C. Samples were sequenced across the Eurofins sequencing platform, and sequences were edited manually.

1. A. Martínez Barrio et al., The genetic basis for ecological adaptation of the Atlantic herring revealed by genome sequencing. elife 6, 1–32 (2016).
33. D. A. Case et al., AMBER 7 (University of California, San Francisco, 2002).