Available online at www.sciencedirect.com

SCIENCE DIRECT

Gene 350 (2005) 161–171

GENE
SECTION
EVOLUTIONARY GENOMICS

www.elsevier.com/locate/gene

Convergent loss of an anciently duplicated, functionally divergent RH2 opsin gene in the fugu and *Tetraodon* pufferfish lineages

Daniel E. Neafsey*, Daniel L. Hartl

Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138, USA

Received 10 December 2004; received in revised form 21 January 2005; accepted 22 February 2005
Available online 7 April 2005

Abstract

We describe the complete opsin gene families from the sequenced fugu and *Tetraodon* pufferfish genomes. We report the convergent loss of function of an anciently duplicated, functionally divergent RH2 or "green-sensitive" opsin gene in both pufferfish lineages, designated *RH2-2*. In fugu, *RH2-2* apparently ceased to function very recently following a transposon-induced deletion that truncated the N-terminal 115 amino acids from the translated protein. Although a lack of frameshift or nonsense mutations in the fugu *RH2-2* pseudogene suggests that the gene was lost very recently in this lineage, we were unable to detect any evidence of a selective sweep associated with the fixation of the truncated allele from population data. Interspecific comparison of the remaining fugu *RH2-2* coding sequence paradoxically indicates that the gene was under strong purifying selection until the truncation occurred.

Keywords: Pseudogenes; Gene duplication; Transposable elements; *Takifugu rubripes*; Pufferfish

1. Introduction

Gene family evolution is a dynamic process, in which the birth of new member genes from tandem or whole-genome duplication is balanced by mutational decay of some duplicates into pseudogenes. Theory and empirical observation suggest that most gene duplicates that will be lost begin to decay soon after they originate, unless one copy assumes a new function or mutation subdivides the functions of the original copy among the pair (Ohno, 1970; Li, 1980; Force et al., 1999). The loss of anciently duplicated genes exhibiting evidence of functional divergence or subdivision is expected to be a rare occurrence under these models.

Here we report the recent convergent loss of an opsin gene duplicate exhibiting strong evidence of functional divergence that we estimate to be ~80 million years old. The opsins compose a small family of retinal transmembrane proteins that facilitate vertebrate color vision via sensitivity to different wavelengths of light. Opsins are members of the G-protein-coupled receptor (GPCR) family that are expressed in the membranes of retinal cone cells. Like rhodopsin, which is expressed in retinal rod cells, they bind with a retinylidene chromophore via a protonated Schiff base at Lys-296. Illumination isomerizes the chromophore and initiates the G-protein signal transduction cascade. Interactions between the opsin protein and the chromophore determine the maximal wavelength sensitivity (λ_{max}) of the chromophore. Amino acid substitutions in opsin proteins affecting these interactions shift the λ_{max}, yielding visual pigments maximally sensitive to wavelengths of light in the ultra-violet (<380 nm) or "visual" (380–740 nm) range of the electromagnetic spectrum.

While characterizing the opsin gene families in the sequenced *Tetraodon nigroviridis* and *Takifugu rubripes* (fugu) pufferfish genomes, we discovered the convergent

Abbreviations: LTR, long terminal repeat; GPCR, G-protein-coupled receptor.

* Corresponding author. The Broad Institute of MIT and Harvard, 320 Charles St., Cambridge, MA 02141, USA. Tel.: +1 617 324 5259; fax: +1 617 452 4588.
E-mail address: neafsey@broad.mit.edu (D.E. Neafsey).

0378-1119/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2005.02.011

loss of a medium-wavelength-sensitive ("green") opsin paralog, designated *RH2-2*, from both the *Tetraodon* and fugu pufferfish genomes. Evidence from extensive structural, biochemical, and comparative evolutionary studies of opsins and the related rhodopsin protein (for review see Sakmar et al., 2002) strongly suggests that the spectral sensitivity of *RH2-2* is distinctly blue-shifted relative to its sister paralog *RH2-1*. Comparison to intact orthologs in medaka indicates that pufferfish *RH2-1* and *RH2-2* were, until recently, subject to similarly strong levels of purifying selection.

In this manuscript we describe our efforts to understand the evolutionary circumstances surrounding the loss of *RH2-2* in these two pufferfish lineages. The *Tetraodon* copy of *RH2*-2 is highly degenerate, and may have become a pseudogene soon after the *Tetraodon* and fugu lineages diverged 18–30 million years ago (Crnogorac-Jurcevic et al., 1997). We focus our efforts on fugu, where *RH2-2* was lost much more recently as the result of a nearby non-Long Terminal Repeat (LTR) retroelement insertion. A deletion that likely accompanied the retroelement insertion caused the truncation of the N-terminal 115 codons of fugu *RH2-2* (Fig. 1), removing the first two of seven transmembrane domains. We checked for expression of the truncated gene in a fugu cDNA library to substantiate its nonfunctional status, because the truncation abuts a methionine codon (which could act as a novel start codon) and the remaining coding sequence exhibits no fixed or segregating indels or nonsense mutations. The presence of non-truncated, apparently functional copies of *RH2-2* in five other members of the genus *Takifugu* indicates that the truncation fixed in *T. rubripes* in the recent past, but we were unable to detect strong evidence of selection associated with the fixation event from polymorphism data. We conclude that the fixation event may have occurred neutrally by genetic drift, or alternatively, that if truncation was driven to fixation by selection, the fixation may not have occurred sufficiently recently for a detectable signal of natural selection to persist.

2. Materials and methods

2.1. Identification of pufferfish opsin genes

Versions 3 and 6 of the fugu and *Tetraodon* genome assemblies, respectively, were queried with T-BLASTX using a zebrafish rhodopsin (RH1) coding sequence (accession no. NM131084). Highly similar nucleotide sequences were aligned to each other and to a selection of publicly available teleost opsin and rhodopsin sequences with Clustal X using the default parameter settings (see Appendix for accession numbers). Medaka *RH2-1* was assembled from two expressed sequence tags (ests; ID=BJ491781, BJ495952) downloaded from the MBase medaka est library Olestall0309 (http://mbase.bioweb.ne.jp). A neighbor-joining tree was constructed from the aligned sequences in Clustal X with a Kimura correction for multiple substitutions to identify fugu and *Tetraodon* orthologs belonging to the four major clades of the vertebrate opsin gene family: RH2 ("green-sensitive"), SWS1 ("UV-sensitive"), SWS2 ("blue-sensitive"), and LWS ("red-sensitive"). The tree was rooted with rhodopsin (RH1) sequences and nodal support was determined with 1000 bootstrap replicates. All teleost opsin sequences were tested for statistical evidence of gene conversion with Sawyer's Runs Tests, v1.4.1 (Sawyer, 1989) using both polymorphic informative sites and silent informative sites.

Sequences belonging to the RH2 clade were further analyzed phylogenetically using a maximum likelihood approach to substantiate the timing of the duplication event leading to pufferfish *RH2-1* and *RH2-2*. Fugu rhodopsin (accession no.AF137214) was used as an outgroup. Hier-

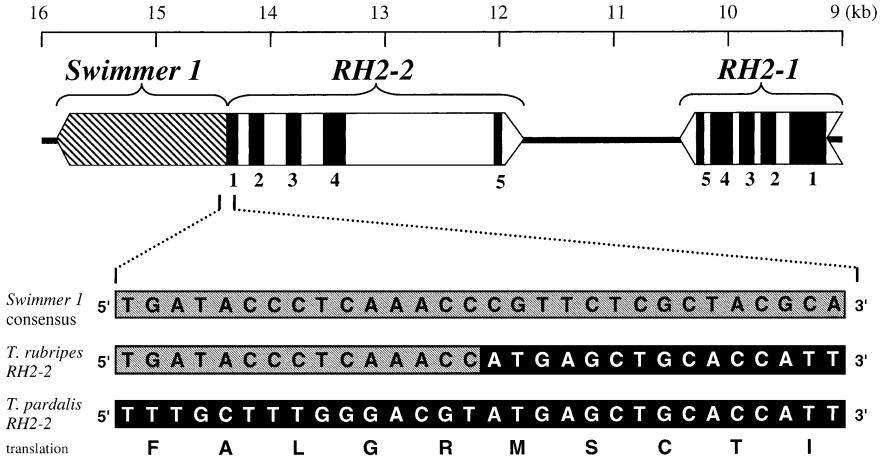

Fig. 1. Physical map of fugu *RH2-1* and *RH2-2* genes, showing truncation of *RH2-2* induced by *Swimmer 1* non-LTR element. Kilobase scale at top shows location of genes on scaffold 2650 from the JGI Fugu Genome Assembly v. 3.0. Open reading frames are left-to-right for *RH2-2* and right-to-left for *Swimmer 1* and *RH2-1*. Numbers below dark bars indicate exons. An expanded view of the junction between the *Swimmer 1* insertion and *RH2-2* is presented below the physical map, together with non-truncated *Swimmer 1* and *RH2-2* sequences.