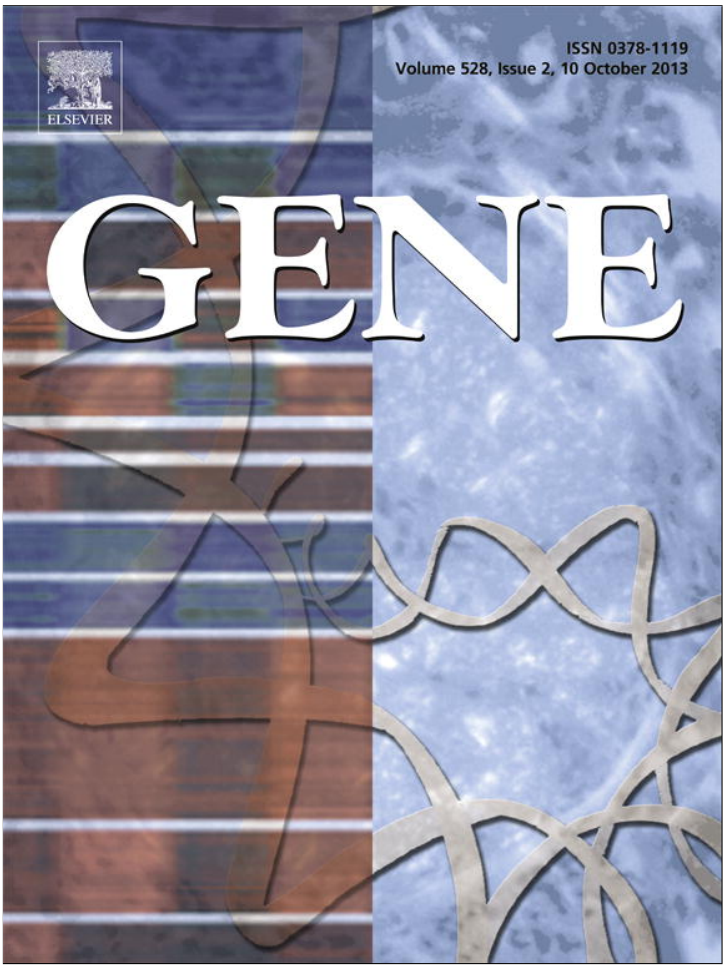

Gene 528 (2013) 201–205

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Computational analysis and functional expression of ancestral copepod luciferase

Yasuhiro Takenaka [a,*], Akiko Noda-Ogura [b], Tadashi Imanishi [b], Atsushi Yamaguchi [c], Takashi Gojobori [d], Yasushi Shigeri [a,*]

[a] Health Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), 1-8-31 Midorigaoka, Ikeda, Osaka 563-8577, Japan
[b] Molecular Profiling Research Center for Drug Discover, National Institute of Advanced Industrial Science and Technology (AIST), 2-42 Aomi, Koto-ku, Tokyo 135-0064, Japan
[c] Faculty of Fisheries Science, Hokkaido University, 3-1-1 Minato-cho, Hakodate 041-0821, Japan
[d] Laboratory for DNA Data Analysis, Center for Information Biology, National Institute of Genetics, Yata 1111, Mishima, Shizuoka 411-8540, Japan

ARTICLE INFO

Article history:
Accepted 14 July 2013
Available online 23 July 2013

Keywords:
Ancestral gene
Bioluminescence
Coelenterazine
Copepod
Luciferase
Luciferin

ABSTRACT

We recently reported the cDNA sequences of 11 copepod luciferases from the superfamily Augaptiloidea in the order Calanoida. They were classified into two groups, Metridinidae and Heterorhabdidae/Lucicutiidae families, by phylogenetic analyses. To elucidate the evolutionary processes, we have now further isolated 12 copepod luciferases from Augaptiloidea species (*Metridia asymmetrica*, *Metridia curticauda*, *Pleuromamma scutullata*, *Pleuromamma xiphias*, *Lucicutia ovaliformis* and *Heterorhabdus tanneri*). Codon-based synonymous/nonsynonymous tests of positive selection for 25 identified copepod luciferases suggested that positive Darwinian selection operated in the evolution of Heterorhabdidae luciferases, whereas two types of Metridinidae luciferases had diversified *via* neutral mechanism. By *in silico* analysis of the decoded amino acid sequences of 25 copepod luciferases, we inferred two protein sequences as ancestral copepod luciferases. They were expressed in HEK293 cells where they exhibited notable luciferase activity both in intracellular lysates and cultured media, indicating that the luciferase activity was established before evolutionary diversification of these copepod species.

1. Introduction

Copepods, dominant among zooplankton, are multicellular organisms with a wide distribution and variety of species throughout the world. Some species of copepods are thought to produce unknown biological materials and have exhibited bioluminescence, but their mechanism and physiological role of this ability have remained a mystery (Haddock et al., 2010; Herring, 1988).

Bioluminescence of marine organisms is mainly produced by an enzymatic reaction of a luciferase and luciferin with oxygen (Shimomura, 2006). In copepods, coelenterazine with the imidazopyrazinone structure was identified as a representative luciferin (Oba et al., 2009; Shimomura, 2006). The order Calanoida contains almost all of the bioluminescent species. *Gaussia* luciferase (GLuc) from *Gaussia princeps* (Calanoida: Metridinidae) was first cloned in 2002 (Verhaegen and Christopoulos, 2002). GLuc possessed an N-terminal signal sequence for extracellular secretion, and showed robust stability and remarkable luminescent activity. Subsequently, the cDNAs of *Metridia longa* luciferase (MLuc) were isolated (Markova et al., 2004) and we reported the two forms of luciferase genes from *Metridia pacifica* (MpLuc1 and 2) (Takenaka et al., 2008). These copepod luciferases were all secreted luciferases and reacted with coelenterazine to produce light. We recently isolated 11 copepod luciferase cDNAs from the five luminous species in the order Calanoida and evaluated their relative luciferase activities using zooplankton homogenates and recombinant proteins (Takenaka et al., 2012). We had also examined the presence of luciferases in more primitive Calanoida, but failed to isolate luciferases or luciferase-like genes. To date, all copepod luciferases have been isolated from species in the Augaptiloidea superfamily from Calanoida, but not from other superfamilies, such as Pseudocyclopoidea or Centropagoidea, demonstrating the difficulties in elucidation of the origin and ancient ancestry of copepod luciferases.

In this study, we isolated 12 additional luciferase genes from Augaptiloidea species, conducted phylogenetic analyses, and deduced two amino acid sequences as ancestral copepod luciferases by *in silico* analyses using Molecular Evolutionary Genetics Analysis Software (MEGA5) (Tamura et al., 2011). Furthermore, we confirmed their functional activities using a mammalian expression system and estimated the evolutionary mechanism of copepod luciferases.

2. Materials and methods

Procedures for zooplankton sampling were described previously (Takenaka et al., 2012). Total RNA was prepared from one specimen using RNeasy Plus Micro kit (Qiagen). cDNA was synthesized from total RNA using the SMART RACE cDNA Synthesis Kit (Takara Bio).

Abbreviations: ASR, ancestral sequence reconstruction; Coe, coelenterazine; GLuc, *Gaussia* luciferase; ML, maximum likelihood; MLuc, *Metridia longa* luciferase.

* Corresponding authors. Tel.: +81 72 751 8324; fax: +81 72 751 9517.
E-mail addresses: pc_tert0617@yahoo.co.jp (Y. Takenaka), yasushi.shigeri@aist.go.jp (Y. Shigeri).

0378-1119/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.gene.2013.07.011

Internal sequences of copepod luciferase cDNAs were amplified using White_luc UP1, 5′-GGCTGCACYAGGGGATGYCTKATMTG-3′, and White_luc LP1, 5′-ACATTGGCAAGACCYTTVAGRCA-3′, which were designed based on the conserved region of previously isolated copepod luciferases (Takenaka et al., 2012). PCR products (200–300 bp) were gel-purified, cloned into a pCR2.1 vector (Life Technologies) and sequenced using an ABI PRISM 3100 Genetic Analyzer (Life Technologies). The 5′ and 3′-ends of copepod luciferase cDNAs were amplified using a gene-specific primer and Universal Primer Mix (UPM) following manufacturer's instructions. After sequencing the 5′ and 3′-RACE products, the full-length cDNA was amplified using gene-specific primers. The cDNA sequences of novel luciferases were deposited in the DDBJ/EMBL/GenBank databases with accession numbers as follows; MaLuc1 (*Metridia asymmetrica*, AB519700), MaLuc2, (*M. asymmetrica*, AB519701), McLuc1 (*Metridia curticauda*, AB519702), McLuc2, (*M. curticauda*, AB519703), PsLuc1 (*Pleuromamma scutullata*, AB519706), PsLuc2, (*P. scutullata*, AB519707), PxLuc1-7 (*Pleuromamma xiphias*, AB519708), PxLuc1-8 (*P. xiphias*, AB716975), PxLuc2 (*P. xiphias*, AB519709), LoLuc1-3 (*Lucicutia ovaliformis*, AB716356), HtLuc1-1-3 (*Heterorhabdus tanneri*, AB519712) and HtLuc1-2-2 (*H. tanneri*, AB519713).

For phylogenetic analyses, the full-length amino acid sequences of copepod luciferases were aligned using the web-based ClustalW program with default parameters (http://www.genome.jp/tools/clustalw/). Phylogenetic trees were generated with the maximum likelihood (ML) method of the WAG model using the MEGA5 program (Tamura et al., 2011). The reliability of the trees was computed by 1000 replications of bootstrap analyses. On the basis of the ML tree, an estimation of the ancestral sequences of copepod luciferases was performed. We chose the most likely amino acid at every branching point of the tree and estimated two ancestral sequences of copepod luciferases (aCopLuc43 and aCopLuc48) by MEGA5 program. For a test of positive selection, the number of synonymous and nonsynonymous nucleotide substitutions per site was estimated from the pairs of sequences (408 bp) that encode two core domains and the interstitial fragment. A Z-test of selection was conducted using the modified Nei–Gojobori (assumed transition/transversion bias = 1) method in MEGA5. The variance of the difference was computed using the bootstrap method (1000 replicates). All ambiguous positions were removed for each sequence pair.

The coding sequences of the putative ancestral copepod luciferases (aCopLuc43 and aCopLuc48) were synthesized *de novo* (Operon). The coding regions of copepod luciferases were amplified by Pyrobest DNA Polymerase (Takara Bio) to subclone into pcDNA3.1/V5-His-TOPO (Life technologies) for protein expression in HEK293 cells (Health Science Research Resources Bank). Cells were transfected with the expression construct using FuGENE HD (Roche Applied Science). Culture medium and cellular lysate were harvested for 72 h after transfection. 5 μl of culture medium or cellular lysate was mixed with 10 μl of 1 ng/μl coelenterazine, 20 mM Tris-HCl (pH 8.0), 50 mM $MgCl_2$, and measured for 10 s using the luminometer MiniLumat LB 9506 (Berthold).

3. Results and discussions

Firstly, we isolated copepod luciferase genes from *M. asymmetrica* (MaLuc1 and MaLuc2), *M. curticauda* (McLuc1 and McLuc2), and *P. scutullata* (PsLuc1 and PsLuc2) (Fig. 1). From these species, two different isotypes of luciferase genes were also identified as previously reported (Takenaka et al., 2008, 2012). We further cloned three distinct luciferases (PxLuc1–7, PxLuc1–8 and PxLuc2) from *P. xiphias* cDNAs. The sequences of the C-terminal half of PxLuc1–7 and PxLuc1–8 were completely identical, but those of the N-terminus were highly diversified (Fig. 1). PxLuc2 was cloned as a partial cDNA because it encoded no secretion signal at its N-terminus. Finally, we isolated an additional luciferase cDNA from *L. ovaliformis* (LoLuc1–3) and two luciferases from *H. tanneri* (HtLuc1-1-3 and HtLuc1-2-2). LoLuc1–3 contained the same number of amino acid residues (223 amino acids) as that of LoLuc, which was previously reported (Takenaka et al., 2012), and identity between them was 98.2%. On the other hand, both HtLuc1-1-3 and HtLuc1-2-2 were also highly similar to previously identified HtLuc1 (90 and 87%, respectively) (Takenaka et al., 2012).

The newly identified full-length luciferases were relatively small proteins, ranging from 19.8 to 24.3 kDa. They were secreted into the culture media when they were expressed in HEK293 cells (data not shown). The identity of the deduced amino acid sequences of these novel and previously-identified copepod luciferases was significantly greater at the internal and C-terminal regions compared to the N-terminus (Fig. 1). Alignment of amino acid sequences of all known copepod luciferases, including the luciferases reported in this paper, further confirmed the presence of two short, repeat sequences that contain consensus amino acid residues of C-x(3)-C-L-x(2)-L-x(4)-C-x(8)-P-x-R-C (x, amino acid residue) in the primary structures of copepod luciferases. The substrate specificity and bioluminescent spectra of the novel luciferases were similar to those of previously identified copepod luciferases (Table 1).

As shown in Fig. 2, we obtained an expected ML phylogenetic tree by analyzing the amino acid sequences of 12 additional copepod luciferases and 13 other, previously identified luciferases. Phylogenetic analysis of these protein sequences suggested that the type I and II luciferase genes from *Metridia* and *Pleuromamma* were generated by gene duplication at node **c** after separation of Metridinidae and Lucicutiidae superfamilies at node **a** (Fig. 2). Gene duplication plays an important role in evolution by giving opportunities for acquisition of novel gene functions. Therefore, we suspected that positive Darwinian selection operated after gene duplication of two types of *Metridia* and *Pleuromamma* luciferase genes. To examine this hypothesis, we used a simple method of calculating the numbers of synonymous (d_S) and nonsynonymous (d_N) nucleotide substitutions per site between sequences. The number of synonymous and nonsynonymous nucleotide substitutions per site was estimated from the pairs of sequences of two core domains and the interstitial fragment (408 bp) by using MEGA5 software. Testing the probability of the null hypothesis of neutral evolution ($d_N = d_S$) in favor of the alternative hypothesis of positive selection ($d_N > d_S$) is, unexpectedly, only rejected for Heterorhabdidae luciferase sequences, not for those of Metridinidae (Table 2). Therefore, these results suggest that divergence between type I and type II luciferase genes from *Metridia* and *Pleuromamma* occurred by neutral evolution at node **c** in Fig. 2, whereas ancient Heterorhabdidae luciferases did not have gene duplication, but exhibited positive Darwinian selection and the creation of adaptive luciferases. These findings were unexpected because activities of *Heterorhabdus* and *Heterostylites* luciferases were much weaker than those of *Metridia* and *Pleuromamma* luciferases, illustrating a loss of gene function. One hypothesis is that bioluminescence was disadvantageous for ancient *Heterorhabdus* and *Heterostylites* species, since their feeding mechanism changed from suspension feeding to predation (Nishida and Ohtsuka, 1996). Alterations of feeding behavior of *Heterorhabdus* might consequently lead to the loss of luciferase functions, driven by positive Darwinian selection.

We tried to isolate luciferase-like genes from Pseudocyclopoidea, an evolutionary older superfamily than Augaptiloidea, using several degenerate PCR primers, but this attempt failed (data not shown). Therefore, we inferred the amino acid sequences of ancestral copepod luciferases by computational analysis using MEGA5. The aCopLuc43 (position at node **a**) is a common ancestor for Metridinidae and Lucicutiidae superfamilies, while aCopLuc48 (node **b**) is a common ancestor for the Heterorhabdidae superfamily (Figs. 1 and 2). Both ancestral luciferases contained highly conserved amino acid residues of C-x(3)-C-L-x(2)-L-x(4)-C-x(8)-P-x-R-C in the tandem repeat sequences (Fig. 1). Coding sequences of ancestral copepod luciferases were synthesized *de novo* and subcloned into a mammalian expression vector to demonstrate whether they were enzymatically active or not.

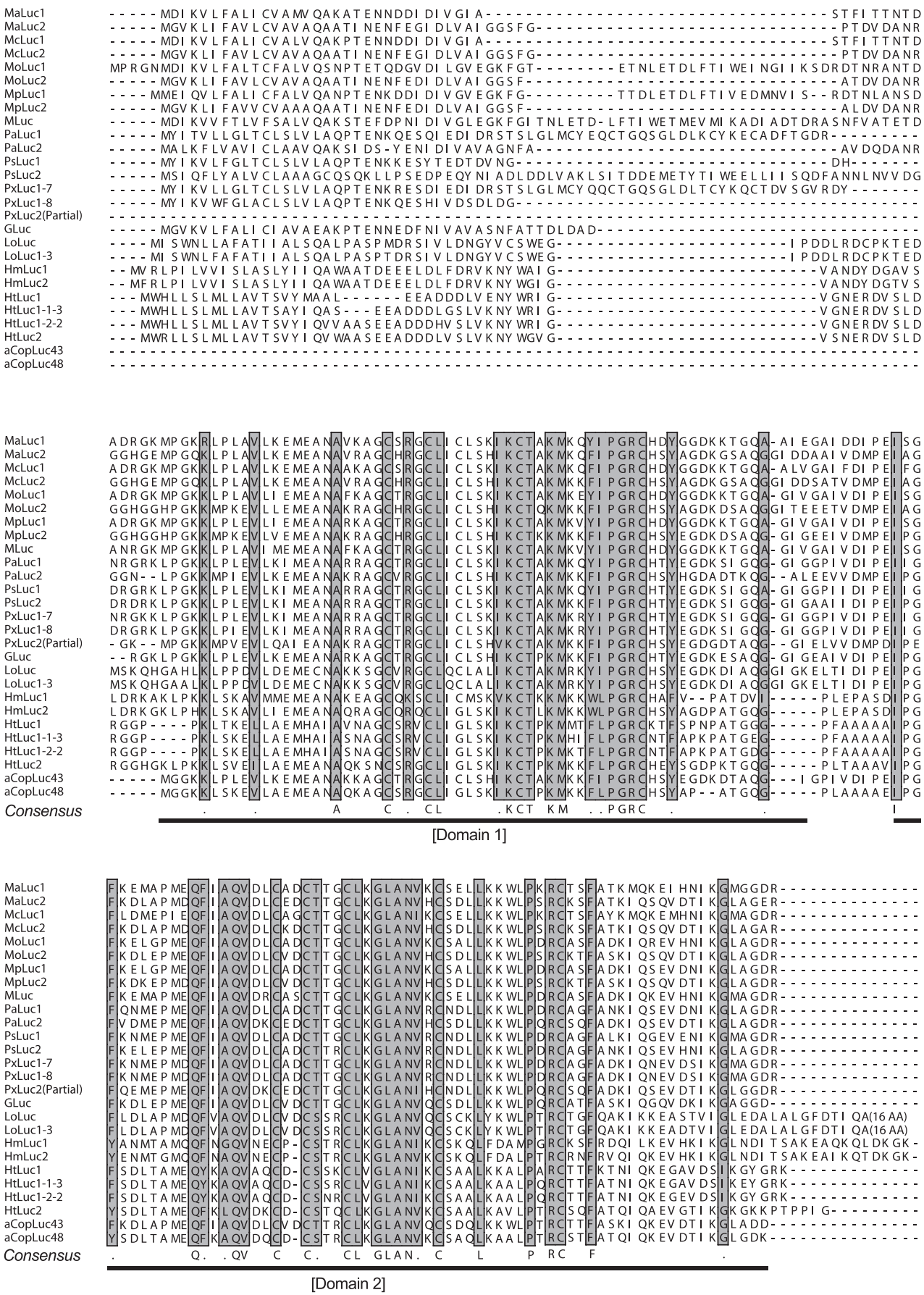

Fig. 1. Alignment of the amino acid sequences of 25 copepod luciferases and two inferred ancestral copepod luciferases (aCopLuc43 and aCopLuc48). Dark shade represents highly conserved amino acid residues. Positions of two similar, repeat domains are underlined.

Table 1
Substrate specificity and bioluminescence spectra of copepod luciferases.

Luciferase	MpLuc1	MpLuc2	MoLuc1	MoLuc2	McLuc1	McLuc2	MaLuc1	MaLuc2
Coe	100%	100%	100%	100%	100%	100%	100%	100%
Cyp luciferin	0.01	0.12	0.16	0.06	0.53	0.30	0.12	0.47
Coe-f	4.23	2.92	1.00	1.38	1.83	0.98	4.04	1.53
Coe-n	0.44	0.12	0.42	0.06	0.31	0.18	0.21	0.20
Coe-cp	0.00	0.01	0.02	0.01	0.07	0.04	0.02	0.09
Coe-hcp	0.01	0.09	0.08	0.04	0.37	0.22	0.10	0.30
λ_{max} (nm)	485	485	492	493	482	493	493	492
Luciferase	PsLuc1	PsLuc2	PaLuc1	PaLuc2	PxLuc1-7	PxLuc1-8	LoLuc1-1	LoLuc1-3
Coe	100%	100%	100%	100%	100%	100%	100%	100%
Cyp luciferin	0.05	0.60	0.07	0.18	0.14	0.21	0.27	0.44
Coe-f	1.85	1.38	2.66	0.42	3.83	4.44	0.11	0.22
Coe-n	0.02	0.28	0.03	0.04	0.11	0.10	0.10	0.16
Coe-cp	0.01	0.20	0.02	0.04	0.04	0.07	0.06	0.11
Coe-hcp	0.04	0.59	0.05	0.14	0.13	0.21	0.18	0.30
λ_{max} (nm)	492	491	493	490	492	ND	ND	493

Luciferase activity of culture medium of HEK293 cells expressing each copepod luciferase was measured by adding synthetic coelenterazine (Coe), Cyp luciferin or coelenterazine analogs (Coe-f, Coe-n, Coe-cp and Coe-hcp). Values were given as a relative to reaction with Coe (defined as 100%). Cyp luciferin is a synthetic *Cypridina* luciferin. Bioluminescence spectra (λ_{max}) were measured for 1 min after addition of Coe into HEK293 culture medium containing recombinant copepod luciferases by ATTO AB-1850 lumiFL-Spectrocapture (ATTO, Japan). ND, not determined.

Both ancestral copepod luciferases exhibited remarkable luciferase activities in the intracellular lysates and culture media when they were mixed with coelenterazine (Fig. 3). As shown in Fig. 3, luciferase activities in cellular lysates of both ancestral luciferases were significantly higher than those in culture media, presumably due to the lack of N-terminal secretion signals. Both ancestral luciferases did not react with *Cypridina* luciferin at all (data not shown). These results suggested that the ancient Augaptiloidea luciferases might have been active before evolutionary diversification of the species. Ancestral sequence reconstruction (ASR) is a recently developed strategy for

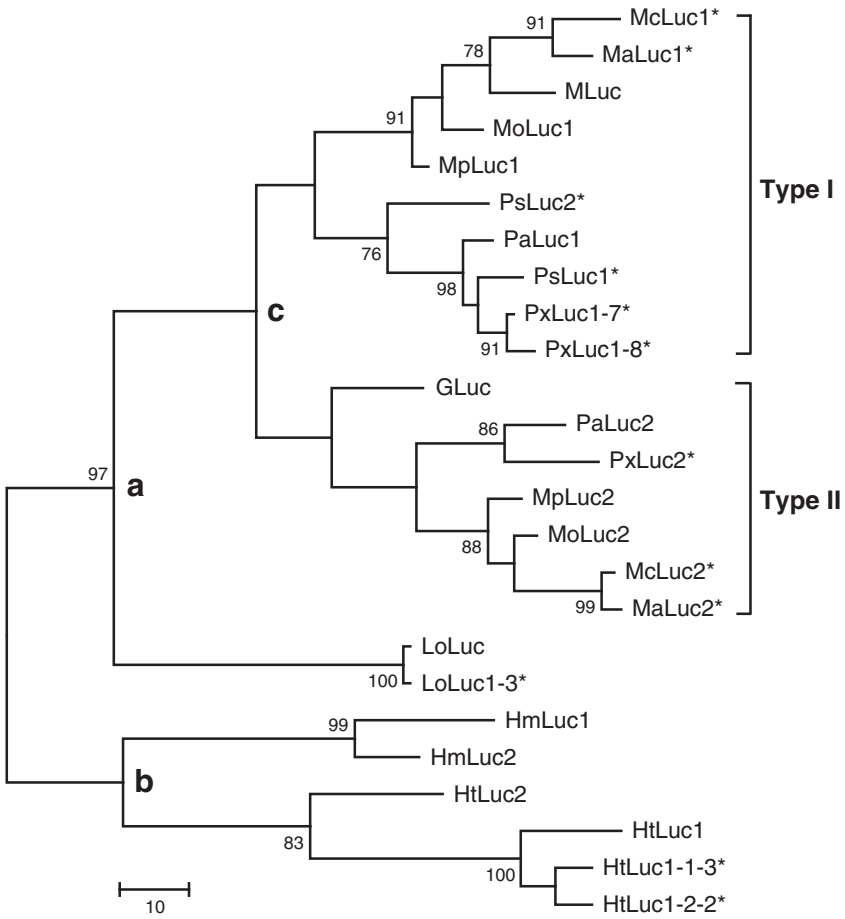

Fig. 2. Molecular phylogenetic analysis of 25 copepod luciferases by the ML method conducted in MEGA5. Branches corresponding to partitions reproduced in less than 50% of bootstrap replicates are collapsed. The percentages of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown on only highly or moderately supported nodes (>74%) next to the branches. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. aCopLuc43 and aCopLuc48, inferred ancestral copepod luciferases, are located at nodes **a** and **b**, respectively. Newly identified luciferases in this study are marked with an asterisk.

Table 2
Codon-based test of positive selection for analysis between tandem repeat and interstitial sequences of six Heterorhabditidae luciferases.

	1	2	3	4	5	6
1. HtLuc1		1.8525	0.9130	3.2708	−2.6861	−2.1383
2. HtLuc1-1-3	0.0332*		−1.1752	2.8788	−2.6665	−2.2678
3. HtLuc1-2-2	0.1815	1.0000		2.4993	−3.0389	−2.7822
4. HtLuc2	0.0007*	0.0024*	0.0069*		−3.5768	−3.7044
5. HmLuc1	1.0000	1.0000	1.0000	1.0000		3.0416
6. HmLuc2	1.0000	1.0000	1.0000	1.0000	0.0014*	

The probability of rejecting the null hypothesis of strict-neutrality ($d_N = d_S$) in favor of the alternative hypothesis ($d_N > d_S$) (below diagonal) is shown in lower-left matrix. Values of P, less than 0.05, are considered as significant at the 5% level and are marked with an asterisk. The test statistic ($d_N - d_S$) is shown above the diagonal. d_S and d_N are the numbers of synonymous and nonsynonymous substitutions per site, respectively.

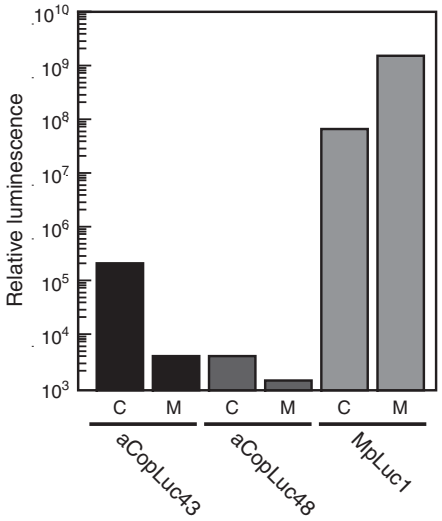

Fig. 3. Activities of ancestral copepod luciferases expressed in HEK293 cells. C and M indicate cellular lysate and culture medium, respectively.

molecular evolution (Thornton, 2004). Using phylogenetic analysis, synthesizing ancestral sequences and testing functional expression using molecular techniques, ASR can help reconstruct ancestral proteins, such as alcohol dehydrogenase (Thomson et al., 2005), chymase protease (Chandrasekharan et al., 1996), EF-Tu (Gaucher et al., 2003), GFP-like proteins (Field and Matz, 2010), RNase (Zhang and Rosenberg, 2002), steroid receptors (Harms and Thornton, 2010) and Opsins (Yokoyama et al., 2008). To our knowledge, this is the first report of an inferred ancestral gene related to bioluminescence. Compared to other resurrected genes, luciferase has considerable advantages in the analyses of ancestral gene reconstruction. In terms of dynamic range, copepod luciferases, such as MpLucs and GLuc show greater linearity in their activity (10^3 ~ 10^5 RLU). This provides for an accurate quantitative assessment of reconstructed ancestral genes. Furthermore, the measurement of luciferase activity is very simple and fast, necessitating only the mixing of a protein and its substrate, followed by photon counting for several seconds in a luminometer. Additional advantages of using copepod luciferases is that they are relatively small molecules (~20 kDa) and have a robust stability even at high temperature (>60 °C) (Takenaka et al., 2008). Smaller sequences facilitate computational analysis and artificial synthesis of ancestral genes. These advantages of luciferases make them useful models for future ASR research.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgments

This study was supported by an AIST research grant, Japan Foundation for Applied Enzymology (Y. Shigeri) and the NIG Cooperative Research Program (Y. Shigeri and T. Gojobori, 2011-A64, 2012-A54, 2013-A67). We thank Prof. Leslie Sargent Jones (Appalachian State University) for her careful reading of the manuscript.

References

Chandrasekharan, U., Sanker, S., Glynias, M., Karnik, S., Husain, A., 1996. AngiotensinII-forming activity in a reconstructed ancestral chymase. Science 271, 502–505.
Field, S., Matz, M., 2010. Retracing evolution of red fluorescence in GFP-like proteins from Faviina corals. Mol. Biol. Evol. 27, 225–233.
Gaucher, E., Thomson, M., Burgan, M., Benner, S., 2003. Inferring the palaeoenvironment of ancient bacteria on the basis of resurrected proteins. Nature 425, 285–288.
Haddock, S.H.D., Moline, M.A., Case, J.F., 2010. Bioluminescence in the sea. Annu. Rev. Mar. Sci. 2, 443–493.
Harms, M., Thornton, J., 2010. Analyzing protein structure and function using ancestral gene reconstruction. Curr. Opin. Struct. Biol. 20, 360–366.
Herring, P.J., 1988. Copepod luminescence. Hydrobiologia 167, 183–195.
Markova, S.V., Golz, S., Frank, L.A., Kalthof, B., Vysotski, E.S., 2004. Cloning and expression of cDNA for a luciferase from the marine copepod *Metridia longa*. A novel secreted bioluminescent reporter enzyme. J. Biol. Chem. 279, 3212–3217.
Nishida, S., Ohtsuka, S., 1996. Specialized feeding mechanism in the pelagic copepod genus *Heterorhabdus* (Calanoida: Heterohabdidae), with special reference to the mandibular tooth and labral glands. Mar. Biol. 126, 619–632.
Oba, Y., Kato, S., Ojika, M., Inouye, S., 2009. Biosynthesis of coelenterazine in the deep-sea copepod. *Metridia pacifica*. Biochem. Biophys. Res. Commun. 390, 684–688.
Shimomura, O., 2006. Bioluminescence. Chemical principles and methods.World Scientific Publishing Co. Ltd., Singapore.
Takenaka, Y., et al., 2008. Two forms of secreted and thermostable luciferases from the marine copepod crustacean, *Metridia pacifica*. Gene 425, 28–35.
Takenaka, Y., Yamaguchi, A., Tsuruoka, A., Torimura, M., Gojobori, T., Shigeri, Y., 2012. Evolution of bioluminescence in marine planktonic copepods. Mol. Biol. Evol. 29, 1669–1681.
Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731–2739.
Thomson, J., et al., 2005. Resurrecting ancestral alcohol dehydrogenases from yeast. Nat. Genet. 6, 630–635.
Thornton, J., 2004. Resurrecting ancient genes: experimental analysis of extinct molecules. Nat. Rev. Genet. 5, 366–375.
Verhaegen, M., Christopoulos, T.K., 2002. Recombinant *Gaussia* luciferase. Overexpression, purification, and analytical application of a bioluminescent reporter for DNA hybridization. Anal. Chem. 74, 4378–4385.
Yokoyama, S., Tada, T., Zhang, H., Britt, L., 2008. Elucidation of phenotypic adaptations: molecular analyses of dim-light vision proteins in vertebrates. Proc. Natl. Acad. Sci. U. S. A. 105, 13480–13485.
Zhang, J., Rosenberg, H., 2002. Complementary advantageous substitutions in the evolution of an antiviral RNase of higher primates. Proc. Natl. Acad. Sci. U. S. A. 99, 5486–5491.