Heterotopy is now recognized as a generative force in the formation of new proteins through modification of existing proteins. We report that heterotopy in the N-terminal region of the mature growth/differentiation factor 5 (GDF5) protein occurred during evolution of teleosts. N-terminal length variation of GDF5 was found among teleost families and interorders but not within teleost families or among tetrapods. We further show that increase of proline and glutamine to the N-terminal region of mature GDF5 occurred in Eurypterygii, the higher lineage of teleosts. Because the basic amino acids, believed to control diffusion, are conserved in this region across all species examined, we suggest that the N-terminal elongation of the mature GDF5 protein during evolution has altered the protein diffusion in Eurypterygii, leading to high concentrations of the protein in the joint of the pharyngeal skeleton, the location of cartilage formation during development.

Morphological changes are the most visible features of vertebrate evolutionary development. From the viewpoint of evolutionary developmental biology, it has been proposed that mutations in the cis-regulatory regions are the predominant source of morphological changes (Carroll et al. 2001). Such mutations can cause gene expression changes in 3 ways (Gilbert 2006): heterotopy, a change in location; heterochrony, a change in temporal expression; heterometry, a change in amount. These changes can occur even if the amino acid sequence remains unaltered. However, heterotopy has recently been recognized as an important force in the generation of a new protein via modification of an existing protein (Gilbert 2006). Several reports suggest that new proteins that are generated by length variations in tandemly repeated amino acid sequences are a major source of morphological variation (Galant and Carroll 2002; Fondon and Garner 2004; Caburet et al. 2005). Tandem repeats are thought to be generated mainly by polymerase slippage during DNA replication and are abundant in coding sequences (Fondon and Garner 2004; Caburet et al. 2005). Proteins containing repeats of alanine, glutamine, glycine, proline, and serine, in particular, are frequently involved in development (Fondon and Garner 2004; Caburet et al. 2005).

Here, we report that the heterotopy in the N-terminal region of growth/differentiation factor 5 (GDF5) mature protein occurred during evolution of teleosts. Gdf5 is a member of the transforming growth factor β (TGFβ) superfamily, which is a large group of genes encoding secreted signaling molecules that play critical roles in several important processes, such as cell proliferation, embryonic patterning, and cell-type specification (Hogan 1996; Moses and Serra 1996; Massague et al. 2000). Mutations at the Gdf5 locus have been shown to reduce the length of limb long bones (Storm et al. 1994; Thomas et al. 1996). GDF5 plays an important role in the differentiation of cartilage and in joint formation (Francis-West et al. 1999; Storm and Kingsley 1999). The zebrafish ortholog of Gdf5 is expressed during development not only in pectoral fin buds but also in the pharyngeal skeleton and in the endoskeletal supports of the dorsal, anal, and caudal fins (Bruneau et al. 1997; Crotwell et al. 2001). The members of the TGFβ superfamily are synthesized as large precursor molecules and processed at a group-specific cleavage site (Cui et al. 1998; Thomas et al. 2006; fig. 1A). Therefore, the precursor molecule is composed of 2 major domains: an N-terminal “prodomain” containing the signal peptide and the cleavage site and a C-terminal “mature domain” (fig. 1A). Members of the TGFβ superfamily can be subdivided into related groups on the basis of amino acid sequence similarities within the mature domain (Rissi et al. 1995; Hogan 1996; Bruneau and Rosa 1997; Bruneau et al. 1997; Davidson et al. 1999; supplementary fig. 1A, Supplementary Material online). Within each group, the mature domain is highly conserved and contains 7 invariant cysteine residues (McDonald and Hendrickson 1993; Schreuder et al. 2005; fig. 1A). The mature domain itself is composed of 2 regions: an N-terminal region (black shading in fig. 1A) that forms a tail-like structure within GDF5 dimers and a C-terminal region (gray shading in fig. 1A) that is involved in formation of homo- and heterodimers and is comprised of the remaining portion of the mature domain beginning with the first conserved cysteine residue. Terai et al. (2002) reported that evolutionary changes of bone morphogenetic protein 4 (BMP4), other member of the TGFβ superfamily, among East African cichlids have been restricted to the prodomain, suggesting that the regulation of BMP4 protein may underlie some morphological aspect of the cichlid diversity.

To further investigate the evolution of other genes of the TGFβ superfamily among the cichlids, we first isolated the cichlid homolog of Gdf5. As previously described (Terai et al. 2002), partial fragments of cichlid Gdf5 were amplified by PCR using the following degenerate primers: BmpF (5′-GTRGGSTGGGAATGACAGGAT-3′) and BmpR (5′-CTCCACWACCATGTCCTGRTA-3′). Genomic DNA from the Lake Malawi cichlid, Labidochromis caeruleus, was used as a template and the amplified DNA products were cloned and sequenced. We confirmed that the cichlid homolog of Gdf5 is a member of the TGFβ superfamily and is highly conserved among cichlids. To further investigate the evolution of Gdf5 in cichlids, we isolated the Gdf5 gene from 2 different species of cichlids, Cynotilapia afra and Pseudocrenicichla sharpeyi, and determined the amino acid sequence of the mature GDF5 protein from each species. We found that the amino acid sequence of the mature GDF5 protein is conserved among cichlids, suggesting that the regulation of BMP4 protein may underlie some morphological aspect of the cichlid diversity.

Key words: growth/differentiation factor 5 (GDF5), teleosts, heterotopy, Eurypterygii.

E-mail: nokada@bio.titech.ac.jp.

doi:10.1093/molbev/msn041
Advance Access publication February 23, 2008

© The Author 2008. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved. For permissions, please e-mail: journals.permissions@oxfordjournals.org
fragments were sequenced. Subsequently, cDNA from the whole body of an adult specimen of *L. caeruleus* was used in combination with sequence information from the partial fragments to obtain sequences of longer cichlid *Gdf5* fragments via 3’ or 5’ rapid amplification of cDNA ends (RACE). RACE experiments were carried out with the following nested primers for cichlid *Gdf5*: SF1 (5’-AGGTATCTTTTCAACATCAGCTCTCTG-3’), SF2

![Diagram](image-url)
(5'-CGAGCTACATCTGGAGAAG-3'), SR1 (5'-CTATGCTCTGATTGCTTGA-3'), and SR2 (5'-CGAGTCGATGAGGATGCTG-3'). Using this technique, we determined the partial sequence of cichlid Gdf5 (1.056 bp), which contained part of the prodomain and the full length of the mature domain. The predicted amino acid sequence of the translated GDF5 suggested that the cleavage site of L. caeruleus GDF5 was RRMRR, the same as that of zebrafish, and that the C-terminal region of the mature domain was also highly conserved and contained seven invariant cysteine residues.

Interestingly, the N-terminal region of the L. caeruleus GDF5 mature domain was 12 and 14 residues longer than those of tetrapod and zebrafish GDF5 mature domains, respectively. The N-terminal regions of all 15 tetrapod GDF5 mature domains obtained from the database were highly conserved and were 18 residues in length (supplementary fig. 1B and table 1, Supplementary Material online). Therefore, we examined the N-terminal regions of mature GDF5 from 23 East African cichlid species, for which morphological adaptive diversity has been well established. The N-terminal regions among the cichlid species were highly conserved and were 30 residues in length (supplementary fig. 1C and table 1, Supplementary Material online). Hence, variation of length was not observed in our comparison among species belonging to the same family of teleosts or among tetrapsods.

The N-terminal region of cichlid GDF5 mature domain is 14 residues longer than that of zebrafish; next, we were interested in determining when the elongation of GDF5 occurred during evolution. Therefore, the N-terminal regions of 23 teleosts GDF5 mature domains were amplified using the following nested primers: ui1 (5'-GCCACACAAACAGTC-3'), ui2 (5'-CTACGAAGTCCTGTTACACCA-3'), di1 (5'-GTCCGAGTGATAGGCGTC-3'), and di2 (5'-CATCTCTTGAGTCTGA CG-3') and the sequences were determined (fig. 1B; supplementary table 1, Supplementary Material online). Regions of zebrafish and medaka were obtained from the database (fig. 1B; supplementary table 1, Supplementary Material online). The N-terminal regions of mature GDF5 in the basal lineage of teleosts, which is composed of Ostariophysi, Clupeomorpha, Protacanthopterygii, and Stomiiformes, were 16–22 residues in length (fig. 1C). In contrast, the higher lineage of teleosts, which is systematically called “Eurypterygii,” had highly diversified N-terminal regions in lengths (fig. 1C): The N-terminus of Myctophiformes (Myctophum affine) was 49 residues, whereas those of other Eurypterygii ranged from 29 to 37 residues. Thus, length variation was found among the families or among the orders of teleosts, and the elongation of GDF5 occurred in the higher lineage of teleosts.

Finally, we examined the identities of the residues within elongated sequences of the N-terminal regions of GDF5 homologs. As shown in blue in fig. 1B, among the 5 repeated residues (alanine, glycine, proline, serine, and glutamine), proline and glutamine were more abundant in fishes than in tetrapods (supplementary table 2, Supplementary Material online). Furthermore, an increase in the number of glutamine residues mainly contributed to the elongation of the N-terminal region during evolution of teleosts (fig. 1B; supplementary table 2, Supplementary Material online). Fondon and Garner (2004) reported a correlation between the variation of repeated residues and morphological modification in dog breeds and proposed that repeats in coding regions contribute to rapid generation of new shapes. However, because length variation of GDF5 was not found among the East African cichlids, GDF5 may not be involved in rapid morphological variation within teleost families. However, their hypothesis could explain the rapid morphological evolution occurred in Eurypterygii.

In contrast to the repeated residues in the N-terminal regions, the characteristic basic residues in the N-terminal regions of mature GDF5 were relatively conserved with respect to number and position despite the length diversification (red in fig. 1B; supplementary table 3, Supplementary Material online). Ohkawara et al. (2002) noticed that the basic residues were also observed in this region among members of the BMP family (BMP2, BMP4, GDF6, and GDF7, colored red in supplementary fig. 1D, Supplementary Material online) but not among members of Activin or TGFβ3 families. The diffusion of BMP2 and BMP4 is tightly restricted within cells, whereas Activin and TGFβ3 are assumed to be highly diffusible. The authors demonstrated that the basic residues in the N-terminal region of the mature BMP4 are required for binding to heparin sulfate proteoglycans, leading to restricted diffusion. From the analogy of BMP4, the elongation of the N-terminal region of mature GDF5 containing the conserved basic residues might play a role in modulating the diffusion rate in the morphogenetic field that leads to morphological change.

Teleosts, comprising about 28,000 species estimated, are the most diversified group of all vertebrates (Nelson 2006). The presence of the interhyal between the ceratohyal and the hyomandibular, forming the dorsal element of hyoid bar, is considered to be a typical feature among teleosts (de Beer 1937; Schultze 1993). McAllister (1968) extensively studied the evolution of hyoid bars among teleosts and found that elements of the hyoid are incomplete in the lower lineage of teleosts, whereas all of them are present among Eurypterygii surveyed by McAllister (Acantopterygii). More articulated hyoid bars play a central role in the feeding efficiency and respiration of teleosts (Stiassny 2000). Zebrafish Gdf5 is a downstream target in the endothelin 1 pathway and Gdf5 expression is required for pharyngeal skeleton joint formation (Miller et al. 2003). Therefore, heterotypy in the N-terminal region of mature GDF5 may cause the high concentration of GDF5 at the joint. Furthermore, Kimmel et al. (1998) observed that the joint acts as an organizing center for cartilage formation. Therefore, the high concentration of GDF5 at the joint may have contributed to the modification of the hyoid bars during teleost evolution. Experiments will be required to reveal the role of amino acid variation in the N-terminal region of mature GDF5 in teleosts.

**Supplementary Material**

Supplementary figure 1 and tables 1–3 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).
Acknowledgments

This work was supported by a grant-in-aid for Scientific Research on Priority Areas “Molecular Mechanism of Speciation” and “Genome Science” to N.O. from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

Literature Cited


Naoko Takezaki, Associate Editor

Accepted February 4, 2008