Gene duplication plays a major role in gene co-option: Studies into the evolution of the motilin/ghrelin family and their receptors

HE Jing¹, Irwin M. DAVID^{1,2} & ZHANG YaPing^{1,3*}

¹ State Key Laboratory of Genetic Resource and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China;

² Department of Laboratory Medicine and Pathobiology and Banting and Best Diabetes Centre, University of Toronto, Toronto, M5S1A1, Canada;

³ Laboratory for Conservation and Utilization of Bioresource, Yunnan University, Kunming 650091, China

Received December 20, 2010; accepted May 16, 2011

Extant genes can be modified, or ‘tinkered with’, to provide new roles or new characteristics of these genes. At the genetic level, this often involves gene duplication and specialization of the resulting genes into particular functions. We investigate how ligand-receptor partnerships evolve after gene duplication. While significant work has been conducted in this area, the examination of additional models should help us better understand the proposed models and potentially reveal novel evolutionary patterns and dynamics. We use bioinformatics, comparative genomics and phylogenetic analyses to show that preproghrelin and prepromotilin descended from a common ancestor and that a gene duplication generated these two genes shortly after the divergence of amphibians and amniotes. The evolutionary history of the receptor family differs from that of their cognate ligands. GPR39 diverges first, and an ancestral receptor gives rise to receptors classified as fish-specific clade A, *GHSR* and *MLNR* by successive gene duplications occurring before the divergence of tetrapods and ray-finned fish. The ghrelin/*GHSR* system is maintained and functionally conserved from fish to mammals. Motilin-*MLNR* specificity must have arisen by ligand-receptor coevolution after the *MLN* hormone gene diverged from the *GHRL* gene in the amniote lineage. Conserved molecular machinery can give rise to new neuroendocrine response mechanisms by the co-option of duplicated genes. Gene duplication is both parsimonious and creative in producing elements for evolutionary tinkering and plays a major role in gene co-option, thus aiding the evolution of greater biological complexity.

ghrelin, motilin, gene family, ligand, receptor, coevolution, gene interaction

Citation: He J, David I M, Zhang Y P. Gene duplication plays a major role in gene co-option: Studies into the evolution of the motilin/ghrelin family and their receptors. Chinese Sci Bull, 2011, 56: 2690–2697, doi: 10.1007/s11434-011-4614-9

Although the general concept of evolutionary tinkering, that natural selection did what it could with the materials at its disposal, was inherent in Darwin’s thinking [1], it has become prominent in contemporary evolutionary biology following the publication of Jacob’s landmark paper [2]. Evolutionary tinkering might be more apparent at the molecular level because multicellular organisms use large sets of similar gene products while exhibiting significant biodiversity. Existing genes can be co-opted to generate new functions by changing their regulatory control and/or the functions of the proteins they encode, and this often involves gene du-

plication followed by specialization of the resulting genes [3]. Although the importance of gene co-option has been a topic of much recent discussion, and the promiscuity of proteins has drawn considerable attention [4,5], further studies are needed to understand the evolutionary dynamics by which gene duplication enables co-option of novel gene functions, and how specificity and promiscuity of the proteins coincide.

Motilin and ghrelin are members of a gastrointestinal hormone family, the motilin/ghrelin family, that regulate gastrointestinal function [6]. Ghrelin is a circulating peptide hormone derived by posttranslational cleavage from preproghrelin (*GHRL*). It is mainly secreted by the stomach

*Corresponding author (email: zhangyp1@263.net.cn)

and acts as an afferent signal on the hypothalamus and hindbrain [6]. Ghrelin acts through the growth hormone secretagogue receptor (*GHSR*), a G protein-coupled receptor, to stimulate the release of growth hormone (*GH*) from the pituitary [6]. Accumulating evidence in mammals suggests that, in addition to its ability to stimulate *GH* secretion, ghrelin stimulates gastric motility, increases appetite and food intake, and induces a positive energy balance leading to body weight gain, as well as a variety of other functions [7]. Posttranslational maturation of prepromotilin (*MLN*) yields a secreted peptide that is further cleaved at a paired basic amino acid site to give rise to motilin [8]. Motilin increases gastrointestinal motility by activating neural pathways or by directly stimulating smooth muscles. Furthermore, the physiological role of motilin appears to be the regulation of a motor pattern associated with the fasted state in humans and dogs, and possibly other species [8]. It is also interesting to note that motilin has a weak *GH*-releasing effect [9]. The orphan G protein-coupled receptor GPR38 was identified as the motilin receptor, *MLNR*, in a remarkable process of reverse pharmacology [10]. Not only do the ligands motilin and ghrelin show structural similarity, but their receptors also have marked sequence similarity with an overall identity of 44%, which rises to 87% in the transmembrane regions. Despite the sequence similarity, there is no evidence for cross-reactivity of the ligands, which is consistent with the different activities of the peptides [6].

Ghrelin and motilin each have a unique but related receptor, which regulate distinct but related physiological functions. The precise evolutionary history of these hormones has not yet been established, and the history of *GHSR*-like receptor diversification remains largely unknown. To gain a better understanding of the evolutionary dynamics of gene duplication on gene products involved in molecular interactions, to determine whether the origins of specific receptors parallel those of their ligands since there is an intimate association between the ligand and receptor, and to gain further insight into the ligand–receptor interactions and changes in their physiological functions, we reconstructed the evolutionary histories of these 2 hormones and their specific receptors. Our analysis has important implications on our understanding of the early steps in the evolution of new protein functions.

1 Materials and methods

1.1 Sequence identification

The sequences of *GHRL*, *MLN*, *GHSR*, *MLNR* and *GPR39* were retrieved from GenBank. *GHSR*, *MLNR* and *GPR39* are all members of the β -group of rhodopsin-like receptor family, and the sequences of these receptors are more similar to each other than they are to any other characterized receptor [11]. TBlastN [12] was used to search for

GHRL/MLN-like and *GHSR*-like sequences in genome sequences for chimpanzee (*Pan troglodytes*) panTro2, orangutan (*Pongo pygmaeus abelii*) ponAbe2, rhesus (*Macaca mulatta*) rheMac2, marmoset (*Callithrix jacchus*) calJac1, mouse (*Mus musculus*) mm9, rat (*Rattus norvegicus*) rn4, guinea pig (*Cavia porcellus*) cavPor3, cat (*Felis catus*) felCat3, dog (*Canis lupus familiaris*) canFam2, horse (*Equus caballus*) equCab2, cow (*Bos taurus*) bosTau4, opossum (*Monodelphis domestica*) monDom4, platypus (*Ornithorhynchus anatinus*) ornAna1, chicken (*Gallus gallus*) galGal3, lizard (*Anolis carolinensis*) anoCar1, *X. tropicalis* (*Xenopus tropicalis*) xenTro2, zebrafish (*Danio rerio*) danRer5, tetraodon (*Tetraodon nigroviridis*) tetNig1, fugu (*Takifugu rubripes*) fr2, stickleback (*Gasterosteus aculeatus*) gasAcu1 and medaka (*Oryzias latipes*) oryLat2, with the previously known genes used as queries. To complement the genomic data we also searched the NCBI nonredundant nucleotide sequences database for sequences with potential similarity.

For accurate prediction of genes, we performed a series of computational procedures. The identified homologous genomic DNA sequences were downloaded and characterized using Wise2 (version 2.1.20 stable), a DNA and protein sequence analysis program [13] that provides highly reliable exon/intron structures and predicted full-length cDNA sequences. Reciprocal blast searches (i.e., searches of well-annotated mammalian genomes with newly identified genomic sequences) were used to eliminate genomic sequences most similar to other rhodopsin-like receptors. Finally, putative *GHSR*-like protein sequences were examined using HMMTOP software (an automatic server that predicts transmembrane helices and topology of proteins) [14] to determine the presence of a seven-transmembrane domain. Intron–exon boundary consensus rules (i.e., AG/GT) were observed and the intron phase of homologous introns was maintained in both the prehormone and receptor genes. The sequences used in this study are presented in the Appendix material.

1.2 Phylogeny and divergence times

We used the generally accepted vertebrate phylogeny [15,16] to calculate evolutionary rates for *GHSRs* and *MLNRs*. We used the assumed divergence date of 450 million years ago (MYA) for tetrapods and ray-finned fishes, 360 MYA for amniotes and amphibians, and 310 MYA for mammals and birds [17]. The divergence time for other lineages are: Therian and Monotreme, 225 MYA; Marsupialia and Eutheria, 190 MYA; Laurasiatheria and Euarchontoglires, 94 MYA; Primates and Rodentia, 87 MYA; Primates and Lagomorpha, 87 MYA; pig and ruminant artiodactyls, 65 MYA; cow and sheep/goat, 20 MYA; *Mus* and *Rattus*, 23 MYA; Old world monkeys and hominoids, 36 MYA; *Macaca fascicularis* and *Macaca mulatta*, 1.92 MYA; humans and chimpanzees, 5.5 MYA [16–19].

1.3 Evolutionary analyses

DNA sequence alignments were generated with ClustalW [20] guided by alignments of the amino acid sequences, after manual adjustments. Phylogenies of the receptors and prehormones were inferred using PHYLIP version 3.65 software [21]. Gap sites in the alignments were excluded in the phylogenetic reconstructions (complete-deletion option). The alignments were bootstrapped using SEQBOOT. The SEQBOOT files were then analyzed with three different methods: maximum parsimony, neighbor-joining, and maximum likelihood, as implemented in PHYLIP. Consensus trees were obtained for each method using CONSENSE. Amino acid p-distances for each branch of *GHSR* and *MLNR*, based on the vertebrate phylogeny, were estimated by Codeml using PAML software [22]. PAML [22] was applied to detect sites potentially under positive selection.

1.4 Molecular modeling

The three-dimensional (3D) structure of *MLNR* was predicted by homology modeling with 2VT4A [23], 2J4YA [24] and 2Z73A [25], which had the best matching protein fold (E -value is 3×10^{-15} , 9×10^{-14} and 7×10^{-7} , respectively), as the templates. The two-dimensional (2D) structure of *MLNR*, 2VT4, 2J4Y and 2Z73 were retrieved from SWISS-PROT (<http://www.expasy.org/sprot/>). Global sequence alignment guided by the 2D structure was used to construct the target-template alignments, and MODELLER (<http://www.salilab.org/modeller/>) was used to create a 3D model of *MLNR* under its automodel class. Loop regions were refined using Accelrys discovery studio 2.1 (<http://accelrys.com/products/discovery-studio/>). Rasmol [26] was used to visualize the *MLNR* structure.

2 Results

2.1 Vertebrate *GHRL*, *MLN* and *GHSR*-like receptor genes

Unlike many other hormone genes, such as the growth hormone, insulin and proglucagon genes [27], only a single *GHRL/MLN*-like prehormone sequence was found in each fish species studied (Tables S1 and S2), suggesting that only one of the duplicates from the fish lineage-specific genome duplication was retained [28]. The sequences of *GHSR*-like receptors are mostly derived from mammals and are few in number; with *GHSR*-like receptors from only two species of fish having been previously identified: *Acanthopagrus schlegelii* and *Sphoeroides nephelus* [29,30]. Our bioinformatic searches identified a large number of *GHSR*-like receptors from diverse vertebrates (Table S3). Our predicted *GHRL*, *MLN*, *GHSR*, *MLNR* and *GPR39* genes (Tables S1–S3) exist in conserved gene neighborhoods; that is, the genes flanking the predicted genes in most genomes were

homologous, strongly supporting orthology. When we compared the genomic neighborhoods of the *GHRL* and *MLN* genes, we found no similarities. Interestingly, when we compared the human *GHRL* and *MLN* genomic regions with those of other vertebrates (UCSC, <http://genome.ucsc.edu/>), we found that the homologous genes flanking the human *GHRL* and *MLN* genes, are located on different chromosomes in amniotes but reside on the same chromosome as the *GHRL* locus in medaka, tetraodon, zebrafish and *X. tropicalis* (Figure S1). There was no overlap in the genomic neighborhoods of *GHRL* and *MLN* gene locus of amniotes. This seems to suggest that there was only one copy of the *GHRL/MLN* gene in the ancestor of fish and tetrapods, and that a genomic region was duplicated on the amniote lineage with the random loss of one of the duplicates from each region, except for *GHRL/MLN*, from the duplicated region. This process yielded genomic neighborhoods that do not share similarity. There was also no evidence for homology between the gene neighborhoods of the receptor genes. Therefore, the genomic context did not support the hypothesis that genomic duplications generated the receptor genes, although it does not exclude this possibility.

2.2 Evolutionary relationship of the motilin/ghrelin gene family

Although *GHRL* and *MLN* have diverged considerably, they still retain some features that help trace their evolutionary history. We were able to recover *GHRL* sequences from all vertebrate classes studied, including fish, amphibians, reptiles, birds and mammals. In contrast, we could only identify *MLN* in reptiles, birds and mammals. The amino acid sequences of ghrelin are well conserved between species, particularly in the N terminal region, and the same principle holds for motilin. The sequence and overall structure of *GHRL* show similarities with *MLN*; both are encoded by four exons with a similar exon/intron organization and intron phase. In both genes, the bioactive peptide coding sequence spans two exons with the intron between these exons interrupting the 14th residue of the mature peptide, and has an identical phase. They also share similar predicted endoprotease cleavage sites, which suggests common ancestry (Figure 1). Further inspection of the *GHRL* sequences from fish and amphibians revealed that they possess one putative endoprotease recognition site, which is located on the C-terminal side of the signal peptidase cleavage site. This should produce a single posttranslational-processed peptide, i.e., ghrelin. In contrast, *GHRL* in amniotes (i.e., reptiles, birds and mammals) possess three putative endoprotease recognition sites, potentially giving rise to a second posttranslational-processed peptide, a 24-residue ghrelin-associated peptide (Figure 1), which was recently identified as obestatin [31]. All *MLNs*, which are only found in reptiles, birds and mammals, possess three putative

Figure 1 Schematic representation of ghrelin and motilin preproteins. The preproteins of ghrelin and motilin are represented by boxes divided into protein domains, proportional to their length. The open and filled triangles indicate the locations of cleavage sites used by signal peptidase and proprotein convertase, respectively. The sequences of the putative endoprotease cleavage sites of various vertebrate classes are shown below. R in parentheses for avian and bony fish represents other alternatives. “↑” denotes an intron position with an intron phase shown beside the arrow. SP, signal peptide. “/”, no putative endoprotease recognition sites. The amino acid sequence alignments of the ghrelin and motilin presequences are shown in the supplementary materials.

endoprotease recognition sites, thus potentially giving rise to two posttranslational-processed peptides, namely motilin and a 17-residue peptide in a position homologous to obestatin (Figure 1).

Phylogenetic analysis of the aligned *GHRL* and *MLN* sequences was conducted to assess their relationships. The relationships among these sequences are consistent with the accepted phylogeny of vertebrates, except for the placement of bullfrog *GHRL* (Figures 2, S2 and S3). Bullfrog *GHRL* groups with amniote *MLN* rather than amniote *GHRL*, although the bootstrap support is weak. A previous study suggested that bullfrog *GHRL* has all of the characteristics of *GHRL*, including the threonine residue at position 3, which is modified by *n*-octanoic acid [32]. We failed to recover a complete *Xenopus GHRL* coding sequence. We only identified exons 1 and 3, which are very similar to the bullfrog sequence, but no additional *GHRL/MLN*-like sequences were identified. Therefore, the *Xenopus* sequence could not be used in the tree construction. Based on the distribution of *GHRL* and *MLN* in the species studied, comparative genomics analysis between human and other vertebrates, and the distribution of endoprotease cleavage sites in the *GHRL/MLN* genes, we surmise that *MLN* was generated by gene duplication on the early amniote lineage, and the tree (Figure 2) is rooted accordingly, offering the most parsimonious scenario. Other potential phylogenetic relationships (e.g., duplication before the fish–tetrapod divergence) require larger numbers of gene deletion events and a parallel gain or loss of endoprotease cleavage sites.

Figure 2 Phylogenetic relationships within the motilin/ghrelin gene family. A bootstrapped tree for diverse vertebrate preproghrelin and pre-promotilin sequences, reconstructed by the maximum likelihood method with Jones-Taylor-Thornton distances and time to jumble set to 3. Bootstrap percentages are shown for the interior branches. *GHRL*, preproghrelin. *MLN*, prepromotilin. *MLNs* are shown in bold.

2.3 Evolutionary relationships of *GHSR*-like receptors

We reconstructed a phylogenetic tree of *GHSR*-like receptor sequences from diverse vertebrates using both the maximum likelihood and neighbor-joining methods, yielding essentially the same results. Receptors could be classified into four major clades, with three being represented by mammalian receptors (Figure 3). These results strongly support the monophyly of each receptor type and they were used to name each clade (Figure 3). The fourth clade was a fish-specific clade, and was named clade A. No more than one copy of each type of receptor was identified in any species studied.

Our tree shows that *GPR39* diverged first, and the remaining *GHSR*-like receptors can be classified into 3 clades; A, *GHSR* and *MLNR*. Palyha et al. [30] searched the pufferfish (*Spherooides nephelus*) genome and identified three distinct *GHSR*-like genes (SnAF082209, SnAF082210 and SnAF082211), which fall into the *GHSR*, *MLNR* and A clades, respectively. Clade A consists of genes found only within fish, and no orthologs were identified in any non-fish vertebrates. The functions of these genes remain unknown. Clade A genes have some peculiarities, including long extracellular loops 2 and 3, generally about 100 residues longer than the analogous loops in *GHSR* and *MLNR*. Loop 2 also displays variations within *MLNR*, with mammalian *MLNR* having a second extracellular loop that is about 40 amino acid residues longer than the analogous loop in

Figure 3 Phylogenetic relationship of 64 *GHSR*-like receptors from diverse vertebrates. The phylogenetic tree for 64 vertebrate *GHSR*-like receptor sequences was generated by the maximum likelihood method. We also used the neighbor-joining method, which yielded essentially the same results. Bootstrap percentages are shown for the interior branches. Human NMUR1 and NMUR2 were used as the outgroups. Hs, *Homo sapiens*; Pt, *Pan troglodytes*; Rm, *Macaca mulatta*; Mf, *Macaca fascicularis*; Mm, *Mus musculus*; Rn, *Rattus norvegicus*; Oc, *Oryctolagus cuniculus*; Cf, *Canis familiaris*; Bt, *Bos taurus*; Oa, *Ovis aries*; Ss, *Sus scrofa*; Et, *Echinops telfairi*; Md, *Monodelphis domestica*; Ora, *Ornithorhynchus anatinus*; Gg, *Gallus gallus*; Ac, *Anolis carolinensis*; Xt, *Xenopus tropicalis*; Dr, *Danio rerio*; Tr, *Takifugu rubripes*; Tn, *Tetraodon nigroviridis*; Sn, *Spheroides nephelus*; As, *Acanthopagrus schlegelii*; Ga, *Gasterosteus aculeatus*; Ol, *Oryzias latipes*.

non-mammalian *MLNR* and *GHSR*. Residues at both ends of these loops are functionally important in hormone binding and action; however, the identities of the functionally significant residues in this loop are not yet clear [33].

2.4 Evolutionary rate of *GHSR*-like receptors in different lineages

The evolutionary rates of *GHSRs* and *MLNRs* were calculated using the generally accepted vertebrate phylogeny [15,16]. The amino acid *p*-distances of *GHSR* and *MLNR* were estimated by Codeml in PAML [22], using fish receptors as an outgroup, and the amino acid substitution rates were calculated. *GHSR* evolved at a rather low but constant rate of 0.139×10^{-9} – 0.71×10^{-9} substitutions/site/year through

vertebrates (Figure S4(a)). In contrast, for *MLNR*, the branch leading to amniote *MLNR* (labeled a–b in Figure S4(b)) showed a burst of rapid evolution where the rate increased to 2.907×10^{-9} amino acid substitutions/site/year. Accelerated evolution of a protein could be the result of either positive Darwinian selection or relaxation of functional constraints. PAML [22] was used to identify the signature of positive selection (Table 1), and the null hypothesis of no positive selection was marginally rejected. Branch-site model testing suggested that about 7.5% of sites were evolving under positive selection along the lineage ancestral to amniote *MLNR*. To understand their biological significance, the positively selected sites along the lineage leading to the amniote *MLNR* were mapped onto a homology modeled 3D-structure of *MLNR*. Mutational mapping of

the ligand-binding pocket in many peptide receptors has revealed the critical role of the extracellular domains [34]. That study also revealed that some of the peptides have additional points of interactions in the transmembrane binding pocket, and that all of the identified residues appear to be located on the surface of the predicted binding pocket [34]. The positively selected sites cluster together in our model, with most facing the putative binding pocket (Figure 4), suggesting that positive selection was driven by their cognate ligand. Since the rapid evolution occurred over a short period of time, after the divergence of amphibians but before the bird-mammal split, because the sequences then resumed a normal rate of evolution, and that this time period coincides with the *GHRL/MLN* gene duplication event, we speculate that this rapid evolutionary burst in the *MLNR* gene was a consequence of coevolution with its new ligand, motilin.

3 Discussion

Given our evolutionary analysis of the motilin/ghrelin gene family, the most parsimonious scenario is that the ancestral *GHRL/MLN* prepropeptide possessed only a single putative endoprotease cleavage site, posterior to the signal peptide. Furthermore, the other two C-terminally located putative endoprotease recognition sites originated *de novo* on the amniote lineage after the divergence of amphibians but before the gene duplication that give birth to the *GHRL* and *MLN* genes. In contrast, the receptor phylogeny suggests that successive duplications of an ancestral *GHSR*-like receptor happened more than 450 MYA, before the divergence of ray-finned fish and tetrapods. The duplications allowed the GPR39 receptor to diverge earliest, followed by the clade A receptors, and finally the *GHSR* and *MLNR* genes (Figure 3). The genes for clade A receptors have been lost in tetrapods, either because of pseudogenization and degeneration, or deletion, whereas the remaining receptors have been retained in both actinopterygians and tetrapods.

Because there is an intimate association between the lig-

and and receptor, one might expect that the genes for receptors and hormones should co-evolve, and that new ligand-receptor pairs would evolve from parallel duplications of the hormone and receptor genes [35,36]. However, the evolutionary history of *MLNR* does not match that of its cognate ligands. Our receptor phylogeny indicates that the duplications that generated distinct genes for the ghrelin and motilin receptors occurred more than 450 MYA, before the divergence of ray-finned fish and tetrapods. Ghrelin—the physiological function of which is to stimulate *GH* release from the pituitary—is conserved from ray-finned fish to mammals [37–39]. Consistent with the conservation of the ghrelin-*GHSR* system from ray-finned fish to mammals, *GHSR* has evolved at a rather low and homogeneous rate (Figure S4(a)). Motilin emerged much later than its receptors, only after the amphibian–amniote divergence. Nevertheless, since their origin, *MLNR* has evolved as if it was under sustained functional constraints (Figure 3), and this occurred before the emergence of its cognate ligands. How the motilin-*MLNR* partnership evolved, which generated a specific neuroendocrine response distinct from that of ghrelin-*GHSR*, is of great interest.

Existing genes can be modified, or tinkered with, to generate new protein structures and functions that are related to those of their ancestors, and which depends on the availability of an evolutionary starting point [2,4,5]. There has been much work in this area, with other hormone–receptor pairs, where the acquisition of specificity for a new effector exploits the promiscuity of existing receptors [40]. Ohno's model [41] suggests that the generation of a redundant gene copy and its release from selective constraints is the first step in the evolution of a new function. This assumption is driven by the notion that negative trade-offs dominate evolutionary processes at all levels and it is impossible to evolve a new function without compromising the original one first [42]. Some studies have suggested that improvements in promiscuous functions are not subject to a trade-off with parallel decreases in the original function [4,42]. Our phylogeny strongly supports the monophyly of each receptor type and that they have diverged considerably

Figure 4 Structural model of MLNR. The three-dimensional structure of MLNR was generated by homology modeling. The MLNR is displayed in the ribbon mode with group coloring. Sites under positive selection along the lineage ancestral to amniote MLNR, are shown in red and in the ball & stick models. The first four images are side views rotated nearly 90° from each other. The last image is the top view from the extracellular domains.

(Figure 3). Studies have shown that *GHSR* orthologs in fish can be activated by growth hormone secretagogues (GHSs), while *MLNR* orthologs in fish are not activated by GHSs or mammalian motilin [30]. It is reasonable to speculate that *GHSR* and *MLNR* experienced functional diversification shortly after their duplication; for example, *MLNR* experienced sub- and/or neo-functionalization and had a different spatial and/or temporal regulation of expression than did *GHSR*. Although identifying the ancient binding state for a receptor is challenging, and the source of the functional constraint on the *MLNR* before the emergence of the motilin hormone is unknown, we can conclude that the ancestral binding properties of the *MLNR* were unlikely to be a promiscuous substrate binding property, which allowed motilin binding, since neither *GHSR* and fish *MLNR* can bind motilin [30]. A specific interaction between motilin and *MLNR* evolved after the origin of the *MLN* hormone gene, which occurred after the amphibian–amniote divergence, and this recruitment corresponds to the period when positive selection was observed in the ligand binding pocket of *MLNR*. Although ghrelin and motilin share substantial sequence similarities, there is no evidence for cross-reactivity of the ligands, and there are fundamental differences between their modes of binding with specific receptors [6]. The activities of the mature peptides are quite different, and a unique and important feature of ghrelin is the octanoylation of serine 3 (which is threonine and modified by *n*-octanoic acid in bullfrog); studies have shown that removing the octanoyl group significantly decreases the potency of *GHRL* [6,32]. This suggests that the acquisition of highly proficient binding occurs at the expense of the old function.

Our study extends our understanding of the early steps in the evolution of new protein functions at the molecular interaction level following gene duplication. The functions of each part of a network do not remain static, and parts of the system can be co-opted for novel functions, with promiscuity serving as the starting point. A major role for gene co-option in gene duplication and divergence should be recognized and that this process produces elements upon which selection can act within the biological network and evolve new functions [3]. The growth of gene families allows for more flexible gene expression profiles and/or the evolution of new biochemical specificities, and has facilitated the evolution of greater biological complexity.

This work was supported by the National Basic Research Program of China (2007CB411600), the National Natural Science Foundation of China (30621092, 30623007), and the Bureau of Science and Technology of Yunnan Province (0803481101).

- 1 Darwin C. On the Various Contrivances by Which British and Foreign Orchids Are Fertilized by Insects and on the Good Effects of Intercrossing. London: Murray, 1886
- 2 Jacob F. Evolution and tinkering. *Science*, 1977, 196: 1161–1166
- 3 True J R, Carroll S B. Gene co-option in physiological and morphological evolution. *Annu Rev Cell Dev Biol*, 2002, 18: 53–80
- 4 Khersonsky O, Roodveldt C, Tawfik D S. Enzyme promiscuity:

- Evolutionary and mechanistic aspects. *Curr Opin Chem Biol*, 2006, 10: 498–508
- 5 Tokuriki N, Tawfik D S. Protein dynamism and evolvability. *Science*, 2009, 324: 203–207
- 6 Peeters T L. Ghrelin: A new player in the control of gastrointestinal functions. *Gut*, 2005, 54: 1638–1649
- 7 Murray C D, Kamm M A, Bloom S R, et al. Ghrelin for the gastroenterologist: History and potential. *Gastroenterology*, 2003, 125: 1492–1502
- 8 Walsh J H, Dockaray G J. *Gut Peptides: Biochemistry and Physiology*. New York: Raven Press, 1993. 261–303
- 9 Samson W K, Lumpkin M D, Nilaver G, et al. Motilin: A novel growth hormone releasing agent. *Brain Res Bull*, 1984, 12: 57–62
- 10 Feighner S D, Tan C P, McKee K K, et al. Receptor for motilin identified in the human gastrointestinal system. *Science*, 1999, 284: 2184–2188
- 11 Holst B, Holliday N D, Bach A, et al. Common structural basis for constitutive activity of the ghrelin receptor family. *J Biol Chem*, 2004, 279: 53806–53817
- 12 Altschul S F, Madden T L, Schaffer A A, et al. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res*, 1997, 25: 3389–3402
- 13 Birney E, Clamp M, Durbin R. GeneWise and genomewise. *Genome Res*, 2004, 14: 988–995
- 14 Tusnady G E, Simon I. The HMMTOP transmembrane topology prediction server. *Bioinformatics*, 2001, 17: 849–850
- 15 Murphy W J, Eizirik E, O'Brien S J, et al. Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science*, 2001, 294: 2348–2351
- 16 van Rheede T, Bastiaans T, Boone D N, et al. The platypus is in its place: Nuclear genes and indels confirm the sister group relation of monotremes and Therians. *Mol Biol Evol*, 2006, 23: 587–597
- 17 Hedges S B, Kumar S. Genomic clocks and evolutionary timescales. *Trends Genet*, 2003, 19: 200–206
- 18 Hayasaka K, Fujii K, Horai S. Molecular phylogeny of macaques: Implications of nucleotide sequences from an 896-base pair region of mitochondrial DNA. *Mol Biol Evol*, 1996, 13: 1044–1053
- 19 Springer M S, Murphy W J, Eizirik E, et al. Placental mammal diversification and the Cretaceous-Tertiary boundary. *Proc Natl Acad Sci USA*, 2003, 100: 1056–1061
- 20 Thompson J D, Higgins D G, Gibson T J. ClustalW: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res*, 1994, 22: 4673–4680
- 21 Felsenstein J. PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle, USA 1993
- 22 Yang Z. PAML: A program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci*, 1997, 13: 555–556
- 23 Warne T, Serrano-Vega M J, Baker J G, et al. Structure of a beta1-adrenergic G-protein-coupled receptor. *Nature*, 2008, 454: 486–491
- 24 Standfuss J, Xie G, Edwards P C, et al. Crystal structure of a thermally stable rhodopsin mutant. *J Mol Biol*, 2007, 372: 1179–1188
- 25 Murakami M, Kouyama T. Crystal structure of squid rhodopsin. *Nature*, 2008, 453: 363–367
- 26 Sayle R A, Milner-White E J. Rasmol: Biomolecular graphics for all. *Trends Biochem Sci*, 1995, 20: 374
- 27 Roch G J, Wu S, Sherwood N M. Hormones and receptors in fish: Do duplicates matter? *Gen Comp Endocrinol*, 2009, 161: 3–12
- 28 Taylor J S, Braasch I, Frickey T, et al. Genome duplication, a trait shared by 22000 species of ray-finned fish. *Genome Res*, 2003, 13: 382–390
- 29 Chan C B, Cheng C H. Identification and functional characterization of two alternatively spliced growth hormone secretagogue receptor transcripts from the pituitary of black seabream *Acanthopagrus schlegelii*. *Mol Cell Endocrinol*, 2004, 214: 81–95
- 30 Palyha O C, Feighner S D, Tan C P, et al. Ligand activation domain of human orphan growth hormone (GH) secretagogue receptor (GHS-R) conserved from Pufferfish to humans. *Mol Endocrinol*,

- 2000, 14: 160–169
- 31 Zhang J V, Ren P G, Avsian-Kretchmer O, et al. Obestatin, a peptide encoded by the ghrelin gene, opposes ghrelin's effects on food intake. *Science*, 2005, 310: 996–999
 - 32 Kaiya H, Kojima M, Hosoda H, et al. Bullfrog ghrelin is modified by n-octanoic acid at its third threonine residue. *J Biol Chem*, 2001, 276: 40441–40448
 - 33 Matsuura B, Dong M, Miller L J. Differential determinants for peptide and non-peptidyl ligand binding to the motilin receptor. Critical role of second extracellular loop for peptide binding and action. *J Biol Chem*, 2002, 277: 9834–9839
 - 34 Gether U. Uncovering molecular mechanisms involved in activation of G protein-coupled receptors. *Endocr Rev*, 2000, 21: 90–113
 - 35 Fryxell K J. The coevolution of gene family trees. *Trends Genet*, 1996, 12: 364–369
 - 36 Irwin D M. Evolution of hormone function: Proglucagon-derived peptides and their receptors. *BioScience*, 2005, 55: 583–591
 - 37 Kaiya H, Kojima M, Hosoda H, et al. Amidated fish ghrelin: Purification, cDNA cloning in the Japanese eel and its biological activity. *J Endocrinol*, 2003, 176: 415–423
 - 38 Kaiya H, Kojima M, Hosoda H, et al. Identification of tilapia ghrelin and its effects on growth hormone and prolactin release in the tilapia, *Oreochromis mossambicus*. *Comp Biochem Physiol B Biochem Mol Biol*, 2003, 135: 421–429
 - 39 Unniappan S, Lin X, Cervini L, et al. Goldfish ghrelin: Molecular characterization of the complementary deoxyribonucleic acid, partial gene structure and evidence for its stimulatory role in food intake. *Endocrinology*, 2002, 143: 4143–4146
 - 40 Bridgham J T, Carroll S M, Thornton J W. Evolution of hormone-receptor complexity by molecular exploitation. *Science*, 2006, 312: 97–101
 - 41 Ohno S. *Evolution by Gene Duplication*. New York: Springer, 1970
 - 42 Kondrashov F A. In search of the limits of evolution. *Nat Genet*, 2005, 37: 73–76

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Supporting Information

Figure S1 Comparative genomics between human and other vertebrates (chicken, *X.tropicalis*, medaka, tetraodon, zebrafish) of GHRL and MLN neighborhood region. Homologous region identified in other vertebrates are show below human schematic GHRL/MLN embedding region, with genomic location indicated on left column.

Figure S2 Complete maximum parsimony phylogeny of motilin/ghrelin family.

Figure S3 Complete NJ phylogeny of motilin/ghrelin family.

Figure S4 Variation in evolutionary rate of GHSR and MLNR within vertebrates.

Table S1 Preproghrelin sequences from diverse vertebrata.

Table S2 Prepromotilin sequences from diverse vertebrata.

Table S3 GHSR, MLNR and GPR39 sequences from diverse vertebrata.

The supporting information is available online at csb.scichina.com and www.springerlink.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.