Prevertebrate Local Gene Duplication Facilitated Expansion of the Neuropeptide GPCR Superfamily

Seongsik Yun,1,4 Michael Furlong,4,1 Mikang Sim,2 Minah Cho,2 Sumi Park,1 Eun Bee Cho,1 Arfaxad Reyes-Alcaraz,1 Jong-Ik Hwang,1 Jaebum Kim,2 and Jae Young Seong*,1

1Graduate School of Medicine, Korea University, Seoul, Republic of Korea
2Department of Animal Biotechnology, Konkuk University, Seoul, Republic of Korea
†These authors contributed equally to this work.

*Corresponding author: E-mail: jyseong@korea.ac.kr.
Associate editor: David Irwin

Abstract

In humans, numerous genes encode neuropeptides that comprise a superfamily of more than 70 genes in approximately 30 families and act mainly through rhodopsin-like G protein-coupled receptors (GPCRs). Two rounds of whole-genome duplication (2R WGD) during early vertebrate evolution greatly contributed to proliferation within gene families; however, the mechanisms underlying the initial emergence and diversification of these gene families before 2R WGD are largely unknown. In this study, we analyzed 25 vertebrate rhodopsin-like neuropeptide GPCR families and their cognate peptides using phylogeny, synteny, and localization of these genes on reconstructed vertebrate ancestral chromosomes (VACs). Based on phylogeny, these GPCR families can be divided into five distinct clades, and members of each clade tend to be located on the same VACs. Similarly, their neuropeptide gene families also tend to reside on distinct VACs. Comparison of these GPCR genes with those of invertebrates including Drosophila melanogaster, Caenorhabditis elegans, Branchiostoma floridae, and Ciona intestinalis indicates that these GPCR families emerged through tandem local duplication during metazoan evolution prior to 2R WGD. Our study describes a presumptive evolutionary mechanism and development pathway of the vertebrate rhodopsin-like GPCR and cognate neuropeptide families from the urbilaterian ancestor to modern vertebrates.

Key words: neuropeptide, GPCR, coevolution, gene duplication, whole-genome duplication, evolutionary history.

Introduction

Neuropeptides play a plethora of roles in the nervous system and peripheral organs, mediating a number of physiological processes including reproduction, growth, homeostasis, metabolism, food intake, sleep, and social and sexual behaviors (Cho et al. 2007; Mirabeau and Joly 2013; Vaudry and Seong 2014). Mature neuropeptides are derived from large precursor proteins that share typical structures in their primary amino acid sequences. These structures include a signal peptide sequence at the N-terminus; an evolutionary conserved mature neuropeptide sequence with canonical dibasic processing sites, which are cleaved by prohormone convertases (Steiner 1998); and often a glycine residue, which can be subjected to amidation, at the C-terminus of the mature peptide (Eipper et al. 1992). To date, more than 100 neuropeptides have been identified in humans, and the vast majority of these act through G protein-coupled receptors (GPCRs) (Fredriksson et al. 2003; Oh et al. 2006), which are characterized by their possession of seven transmembrane helices. Phylogenetic analysis indicates that the majority of neuropeptide GPCRs are clustered in the β and γ groups of the rhodopsin-like (class A) or in the secretin-like (class B) GPCR families (Fredriksson et al. 2003).

Rapid accumulation of protostomian and deuterostomian data, recent advances of bioinformatics tools for neuropeptide and GPCR identification (Mirabeau et al. 2007), and reverse pharmacological approaches in insects (Hauser et al. 2006; Jiang et al. 2013), nematodes (Lindemans et al. 2009), and vertebrates (Civelli et al. 2006) have facilitated discovery of a substantial number of neuropeptide-GPCR systems. Recently, phylogenetic analysis of neuropeptide GPCRs revealed subtrees containing mixed sets of protostomian and deuterostomian sequences, which indicates that many vertebrate and arthropod GPCR families that were previously thought to be phyla specific actually have a bilaterian origin (Mirabeau and Joly 2013). Recent studies indicate that the continuous emergence of new genes through duplication in ancestral species followed by neo-/subfunctionalization or gene loss (Krishnan et al. 2012; Assis and Bachtrog 2013; Hwang et al. 2013) during metazoan evolution established the current diversified range of neuropeptide-GPCR families.

Syntenic analysis of vertebrate genome fragments and comparison of whole chromosomes of evolutionarily distant taxa support the concept of two rounds of whole-genome duplication (2R WGD) during early vertebrate emergence (Ohno 1970). These events produced, on average, four paralogous chromosomes that share related sets of genes, defined as ohnologs (Dehal and Boore 2005; Meyer and Van de Peer 2005; Nakatani et al. 2007; Putnam et al. 2008). The quadruplication of genes by 2R WGD facilitated the rapid proliferation of genes within vertebrate gene families. Additionally, local gene duplications shortly before, during, and after 2R WGDs contributed to the diversification of GPCR families.
WGD further contributed to the proliferation within gene families (Lundin 1993; Holland et al. 1994; Larhammar and Salanek 2004; Hwang et al. 2013, 2014; Kim et al. 2014; Sefideh et al. 2014). More recent local duplications within gene families are easier to identify, as these duplications tend to share higher sequence similarity and be located close to each other. However, genes that emerged earlier in metazoan and vertebrate evolution have accumulated mutations, leading to greater deviation in function and residue sequence. Furthermore, chromosomal shuffling has translocated these genes to new inter- or intrachromosomal regions (Nakatani et al. 2007; Hwang et al. 2013). Thus, phylogenetically related genes created by local duplication are often located on different chromosomes, despite originating from the same ancestral location. Large-scale gene loss and occasional single gene translocation further complicate syntenic analysis (Sefideh et al. 2014).

The evolutionary relationships of gene families can be examined by phylogenetic analysis; however, in the case of neuropeptide genes phylogenetic analysis often does not provide sufficiently reliable results to ascertain the order of emergence of, or relationships within, neuropeptide gene families (Cardoso et al. 2010; Hwang et al. 2013; Kim et al. 2014). In general, signal peptide sequences are not conserved, and pro-peptide sequences, other than the mature peptide, are highly variable, as these sequences are free from evolutionary selective pressure (Lee et al. 2009). Sequence comparison of the short, conserved mature peptides is not sufficient to extrapolate reliable relational information. Furthermore, paralogue peptide genes that arose by local gene duplication prior to vertebrate emergence display considerable variation, even in the mature peptide sequences (Cardoso et al. 2010; Hwang et al. 2013). In contrast to peptide genes, evolutionary relationships among related GPCR families can be more accurately assessed by phylogenetic analysis. GPCR transmembrane domains are reasonably well conserved across vertebrate and invertebrate species, and the amino acid sequences are long enough to generate relatively reliable phylogenetic trees. Thus, exploration of relationships among related GPCRs can help us to trace the evolution of the cognate neuropeptides (Kim et al. 2014; Roch et al. 2014).

In addition to phylogenetic analysis, mapping related genes onto reconstructed pre-2R vertebrate ancestral chromosomes (VACs) can be used to further explore the relationships among related gene families (Yegorov and Good 2012; Hwang et al. 2013). This method is based on the fact that paralogous genes that emerged through local duplication of an ancestral gene prior to and during vertebrate emergence tend to be located in the same vicinity on VAC linkage groups (Kim et al. 2014). Pre-2R VACs and post-2R gnathostome chromosomes (GACs) are mapped by building linkage groups through whole-genome comparison of paralogous regions in multiple vertebrate taxa (Nakatani et al. 2007; Putnam et al. 2008). The Nakatani model shows that linkage groups representing a single GAC can be mapped to multiple regions of multiple chromosomes in extant vertebrates, as chromosomal shuffling has fractured ancestral GACs into linkage groups and dispersed them around the genomes of evolving vertebrates (Nakatani et al. 2007). Therefore, contiguous reconstructed chromosomes show that phylogenetically related genes are indeed present on the same GACs/VACs despite being located on different chromosomes (Hwang et al. 2013; Kim et al. 2014; Sefideh et al. 2014).

Previous detailed analyses of the evolution of individual neuropeptide-GPCR families in vertebrates have predicted pre-2R progenitor genes (Cerdà-Reverter and Larhammar 2000; Lagerstrom et al. 2005; Dores 2013; Osugi et al. 2014). However, few articles have analyzed the evolution of neuropeptide-GPCR families from a more ancient and holistic perspective than the putative vertebrate ancestor (Hwang et al. 2013; Mirabeau and Joly 2013; Yun et al. 2014). In this study, the development of vertebrate rhodopsin-like neuropeptide GPCRs was analyzed through comparison with nonvertebrate chordate and protostomian GPCRs. Phylogenetic analysis combined with extensive synteny comparison revealed that rhodopsin-like neuropeptide GPCRs can be divided into five clades and that members in each clade tend to be located on the same one or two putative VACs, with their cognate neuropeptides following a parallel pattern. This study suggests that local duplication prior to the emergence of vertebrates produced the current diversity of neuropeptide and GPCR superfamilies.

**Results**

*Data Retrieval and Phylogeny of the Rhodopsin-Like Neuropeptide GPCR Superfamily*

The amino acid sequences of rhodopsin-like neuropeptide GPCRs and cognate neuropeptides from representative taxa were obtained using the ENSEMBL and NCBI genome browsers. Representative taxa consisted of tetrapods (human, mouse, chicken, anole lizard, and Xenopus), a lobe-finned fish (sarcopertygian coelacanth), ray-finned fish (basal actinopterygian spotted gar, zebrafish, medaka, stickleback, and tetraodon), a cartilaginous fish (elephant shark *Callorhinchus milii*), a vertebrate of the Agnatha clade (sea lamprey *Petromyzon marinus*), early chordates (lancelet *Branchiostoma floridae*, and sea squirt *Ciona intestinalis*), and protostomes (*Caenorhabditis elegans* and *Drosophila melanogaster*). Using phylogenetic and syntenic analyses, orthologous and paralogous relationships among the GPCR genes (supplementary tables S1–S5 and S11, Supplementary Material online) were identified. Table 1 displays the size of extant gene families in humans, coelacanth, spotted gar, and elephant shark. It reveals that duplications prior to 2R WGD have been followed by widespread gene loss, and occasional additional 2R duplications, throughout subsequent vertebrate evolution.

To further elucidate the evolutionary relationships within this GPCR superfamily, representative vertebrate GPCRs from human (if the human genes were absent, representative orthologs from coelacanth, spotted gar, or elephant shark were used), representative GPCRs from the nonvertebrate chordates *C. intestinalis* and *B. floridae*, and representative GPCRs from the protostomes *C. elegans* and *D. melanogaster*
were subjected to phylogenetic analysis. Rhodopsin-like neuropeptide GPCRs can be classified into five clades containing mixed protostomian and deuterostomian (vertebrate and nonvertebrate chordate) sequences (fig. 1), indicating that these clades, and many families within these clades, originated before the divergence of protostomes and deuterostomes. It should be noted that the bootstrap values of the more ancient branches in the phylogenetic trees derived from figure 1 are low, likely because the time of divergence of these ancient GPCR families far predate the divergence of protostomes and deuterostomes, potentially over a billion years ago (Nordstrom et al. 2011). The bootstrap values of branches that represent relatively more recent divergence tend to be much higher as these gene families have had less time to differentiate. Despite the low bootstrap values of ancient branches, phylogenetic trees that show similar clades were consistently obtained, regardless of the use of complete GPCR sequences or partial sequences consisting of only the transmembrane regions or the use of a variety of alignment options. The phylogenetic tree was further confirmed by Bayesian phylogenetic analysis. General topology was similar between the maximum likelihood and Bayesian trees except for the GPR139/142 family, which is located out of the clades in the Bayesian tree (supplementary fig. S1, Supplementary Material online).

### Syntenic Analysis of Clade 1 GPCR and Neuropeptide Genes

Clade 1 GPCRs include the neuropeptide Y (NPYR), prolactin-releasing peptide (PRLHR), orphan G-protein-coupled receptor 83 (GPR83), prokineticin (PROKR), tachykinin (TACR),

### Table 1. Paralog Numbers of Neuropeptide GPCRs and Their Cognate Neuropeptides in Vertebrates.

<table>
<thead>
<tr>
<th>Clade</th>
<th>GPCR Family</th>
<th>Number of Paralogs</th>
<th>Peptide Family</th>
<th>Number of Paralogs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Pre-2R</td>
<td>ES</td>
<td>SG</td>
</tr>
<tr>
<td>1</td>
<td>NPYR</td>
<td>3</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>PRLHR</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>PROKR</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>TACR</td>
<td>2</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>NPFVR</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>HCRTR</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>CCKR</td>
<td>2</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>QRFP</td>
<td>2</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>GPR83</td>
<td>1 or 2*</td>
<td>1(1)</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>OPR</td>
<td>1</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>MCR</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>NPBWR</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>SSTR</td>
<td>3 or 2*</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>UTS2R</td>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>MCHR</td>
<td>4</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>GALR</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>KISSR</td>
<td>1 or 2*</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>EDNR</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>NMBR/GRPR/BRIS3</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>GPR37</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>TRHR</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>NTSR</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>GHSR/MLNR</td>
<td>1</td>
<td>3(1)</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>NMUR</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>NMUR4</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>GPR39</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>GPR139/142</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>OXT/AVPR</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>NPSR</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>GnRHR</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>GPR19</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>GPR150</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

**Note.**—Gene families are grouped into their respective clades. The number of genes in each family is displayed for vertebrate representatives including the pre-2R ancestor; ES, elephant shark; SG, spotted gar; Co, coelacanth; Hu, humans. The total number of extant genes in these species is shown in the “Total” column, in which all known 2R-generated GPCRs can be found. The total is often higher than any individual species as different GPCR orthologs have been conserved in different lineages. Additional, post-2R, duplications in the representative species are indicated in parentheses. The absence of elephant shark, and to a lesser extent coelacanth, genes may be due to incomplete genome mapping.

*Lack of clarity in the quantity of pre-2R progenitors.*
neuropeptide FF (NPFF), hypocretin (HCRTR), cholecystokinin (CCKR), and pyroglutamylated RFamide peptide (QRFPR) families. In the previous phylogenetic tree by Mirabeau and Joly (2013), CCKRs and OxR (HCRTRs) are on independent branches from clade 1 whereas they are within the branch of clade 1 in the present tree. The tree by Fredriksson et al. (2003) also shows the clustering of these GPCRs in the clade 1 branch. Individual GPCR and neighboring genes in this clade were subjected to syntenic analysis to compare their locations in human, chicken, and spotted gar genomes (fig. 2). To explore the evolutionary relationships among the vertebrate clade 1 GPCR genes, we examined the location of each gene on reconstructed GACs, as proposed by Nakatani et al. (2007). The Nakatani model suggests the presence of ten pre-2R VACs, defined as A–J. These VACs underwent 2R WGD to generate approximately 40 post-2R GACs A0–J1. Most clade 1 GPCRs are located on chromosomal regions with reliable synteny that correspond to GAC_Cs. For instance, NPY4R, TACR2, NPFFR1, QRFPR3, and PRLHR1 are located in the 44–120 Mb region of human chromosome 10 that corresponds to GAC_C0 (fig. 2A). CCKAR, NPFFR2, TACR3, QRFPR1, NPY2R, NPY1R, and NPY5R are located in the 72–169 Mb region of human chromosome 4 that is part of GAC_C1 (fig. 2B). NPYSR and NPY7R are in human chromosome 5 on the GAC_C2 block (fig. 2C). The GAC_C3 block, containing PROKR1, QRFPR4, TACR1, QRFPR2, PRLHR3, NPFFR3 and NPY3R, is split between two human chromosomal regions (68–75 Mb region of chromosome 2 and 20–41 Mb region of chromosome 8); however in chicken, this synteny block is located on a single short region (0.2–2 Mb) of chromosome 22 (fig. 2D), which indicates that this GAC_C3 linkage group is more conserved in chicken.

Although most clade 1 GPCR genes are located on GAC_Cs, there are a few exceptions. The Nakatani model places HCRTR1 and HCRTR2 along with the neighboring COL genes on GAC_Cs (fig. 2E). The fact that paralogous COL genes are found near clade 1 GPCR genes on GAC_C and GAC_B suggests a translocation of this region from VAC_C to VAC_B prior to 2R WGD. PRLHR4 and PRLHRS, which appear to be ohnologs, are located on GAC_Cs and GAC_G0, respectively (fig. 2E). The PRLHR4/S progenitor gene was also likely translocated to VAC_G prior to 2R WGD, as other PRLHR paralogs are located on GAC_Cs. CCKBR and PROKR2 and 3 are located on GAC_D1 and GAC_G1,

![Fig. 1. Maximum-likelihood phylogenetic tree of the rhodopsin-like neuropeptide GPCR superfamily. Representative vertebrate GPCRs from human, representative nonvertebrate chordate GPCRs from *Ciona intestinalis* and *Branchiostoma floridae*, and representative protostome GPCRs from *Caenorhabditis elegans* and *Drosophila melanogaster* were subjected to phylogenetic analysis. Coelacanth, spotted gar, or elephant shark GPCRs, respectively, were used if the genes are absent in human. Colored regions highlight distinct GPCR groups according to taxon grouping, which is also indicated as an abbreviation in front of the GPCR names (v, vertebrate; c, chordate; p, protostome). Yellow represents vertebrate species, green represents nonvertebrate chordate species, orange represents protostome species, and purple represents orthologs only found in elephant shark. The symbols (●) represent the branches of the five individual clades. The number of paralogs in each given family is indicated in parentheses.](https://academic.oup.com/mbe/article-abstract/32/11/2803/982638)
respectively, but their ohnologs and the paralogs of their neighboring genes are located on GAC_Cs. Thus, CCKBR and PROKRs 2 and 3 were translocated subsequent to 2R WGD. Another translocation is found in the region containing GPR83. GPR83 1, 2, and 3 are located on GAC_F3 (the 94–100 Mb region of human chromosome 11) and GAC_F4 (the 66–94 Mb region of human chromosome X), but ohnologs of their neighboring genes (MAML, ARHGAP, and NR3C) are located on GAC_C1, and GAC_C2. This result indicates that between the first and second rounds of WGD a translocation from one of the two GAC_Cs to a GAC_F occurred. The second round of WGD then produced ohnologous regions on GAC_C1 and GAC_C2, and two translocated GAC_F3 and GAC_F4 regions; however, the possibility of misannotation of the regions in the Nakatani model cannot be excluded.

To further corroborate the results shown in figure 2E, which uses only a small number of GPCR and neighboring genes, all protein-encoding genes found on chromosome 4 were compared against the protein-coding genes in the rest of the human genome to identify genes with the highest sequence similarity (cut off < 1e-10; fig. 3). Chromosome 4 was chosen as the reference because the Nakatani model describes this chromosome as being entirely GAC_C1, and most clade 1 GPCR families are represented on this chromosome. The results show that chromosomes 2, 5, 8, 10, 11, and X contain regions in which genes have the highest similarity to the genes on chromosome 4. These regions are postulated to contain GAC_C linkage groups, and in which the GPCR and neighboring genes presented in figure 2E can be found. In chromosome 5 these sequences are concentrated in the 140–160 Mb region which indicates preservation of the linkage group, but in other chromosomes the regions with high sequence similarity are less condensed which indicates intrachromosomal translocation and possible hybridization. This whole-genome analysis supports the results of the small-scale comparison shown in figure 2E.

Synteny analysis of the cognate neuropeptide genes for clade 1 GPCR families was also performed to investigate
intrafamily relationships and for assignment to GAC linkage groups (fig. 4). The HCRT, NPY, NPFF, TAC, and CCK family genes and PRLH2 were allocated to GAC_Es. PRLH1, PROKs, and QRFPs (fig. 4E) are located on GAC_Fs, GAC_Ds, and GAC_As, respectively. The fact that PRLH1’s ohnolog and the paralogs of neighboring genes are located on GAC_Es indicates a misannotation in the Nakatani model and also indicates that PRLH1 was present on GAC_E0 post-2R WGD. The PROK and QRFP genes, however, were likely translocated from VAC_E to VAC_D and VAC_A, respectively, prior to 2R WGD.

**Relationship between Syntenic and Phylogenetic Analyses of Clade 1 GPCR and Neuropeptide Genes**

Clade 1 GPCR progenitors were aligned on VACs (fig. 5) based on their position in the phylogenetic tree derived from figure 1. Peptide progenitors were placed on VACs in parallel to their cognate GPCRs (fig. 5), as phylogenetic trees for neuropeptides are inherently unreliable. Two progenitors were predicted for each of the QRFP, TACR, and CCKR families, and three pre-2R progenitors were predicted for the NPYR and PRLHR families. In contrast to the GPCR families, multiple pre-2R progenitors could not be identified in the neuropeptide families.

Paralogous and ohnologous relationships among family members can be deduced from figure 5. For instance, NPYRs 1, 3, 4, and 6 are grouped in a single branch of the phylogenetic tree, which splits and then splits again, and are distributed on different GAC Cs, that is, C1, C3, C0, and C2, respectively. This pattern is iconic of ohnologs, which are produced from a single progenitor gene through 2R WGD. NPY5R from the neighboring branch is also located on GAC_C1 but because it is phylogenetically more distant it can be deduced that separate progenitor genes produced the NPYR1/3/4/6 and NPY5R groups. These progenitor genes can be postulated to have been generated through local duplication of a common ancestral gene indicated by the node at which these two branches separate. Further back, the vertebrate (black) and protostome (orange) lineages join indicating the point at which the deuterostome and protostome lineages diverged. NPY2R and NPY7R are grouped more distantly in the phylogenetic tree and distributed on GAC_C1 and GAC_C2, respectively. Thus, NPY2R and NPY7R

**Fig. 3.** Whole-genome analysis for chromosomal regions which frequently encode proteins with high sequence similarity to proteins encoded on chromosome 4, in which GAC_C1 is found in its entirety. Regions with high sequence similarity to chromosome 4 on chromosomes 2, 5, 8, 10, and X are connected by colored lines which match their destination chromosomes (chromosomal representation based on the UCSC chromosome browser). Protein matches that correspond to genes presented in figure 2E are color coded to match.
are orthologous to each other and originate from an NPY2/7R progenitor. Therefore, it can be postulated that the vertebrate NPYR family emerged from three separate pre-2R progenitor genes, NPY1/3/4/6R, NPY5R, and NPY2/7R, which all emerged through prior local duplication. The number of NPYR family genes then expanded through 2R WGD, although many of these genes have been subsequently lost. The pattern of gene loss is specific to species. For instance, coelacanth retains seven members whereas humans retain only four members (table 1).

Phylogenetic analysis demonstrates that protostomes contain NPYR-like, TACR-like, and CCKR-like GPCRs (fig. 5). Thus, these families are likely to constitute the most ancient GPCR groups in this clade. Nonvertebrate chordates possess QRFPR-like, PRLHR-like, PROKR-like, NPFFR-like, and HCRTR-like GPCR families, whereas GPR83-like genes were only identified in FIG. 4. Synteny for chromosomal regions containing the clade 1 peptide genes. This figure was generated as in figure 2 with some alterations and additions, as described below. Conserved genomic synteny of clade 1 peptide genes is presented on GAC_E0 (A), GAC_E1 (B), GAC_E2 (C), and GAC_E3 (D) linkage groups. In some cases genes are located on scaffolds and, thus, the last four digits of the respective scaffold numbers are shown (e.g., JH375467 is indicated as 5467) above the gene instead of the chromosome number. Additionally, question marks indicate that synteny analysis could not place absent genes in putative locations. (E) Arrangement of human chromosomal regions onto post-2R GAC linkage groups follows the description provided in figure 2.
vertebrates. It could be postulated that these families were generated in deuterostome/early chordate and early vertebrate development; however, as demonstrated in the NPYR family, in which vertebrate and protostomian GPCRs but no other chordate GPCRs could be found, gene families may become extinct in various lineages. Further detailed information on the evolutionary history of each GPCR family is described in the supplementary information, Supplementary Material online.

Syntenic and Phylogenetic Analysis of Clade 2 GPCR and Neuropeptide Genes

Clade 2 GPCRs include the opioid (OPR), neuropeptide-B/W (NPBWBR), somatostatin (SSTR), melanin-concentrating hormone (MCHR), urotensin-2 (UTS2R), galanin/spexin (GALR), kisspeptin (KISSR), and, nominally, melanocortin (MCR) families. Fredriksson et al. (2003) placed the clade 2 GPCRs in a distinct branch of the gamma-group of rhodopsin-like GPCRs because they did not find close structural similarity between the clade 2 GPCRs and beta-group GPCRs. However, the present phylogenetic tree (fig. 1) is comprised both beta-group and clade 2 GPCRs because the clade 2 family, like the betagroup, interacts with neuropeptide ligands that are produced through a maturation process by prohormone convertases.

The MCR family is phylogenetically distant from the clade 2 GPCR families and is classified as part of the alpha-group of rhodopsin-like GPCRs (Fredriksson et al. 2003). The MCR family is only included in this clade because the cognate MSH/ACTH peptides are derived from the POMC gene that also produces a peptide ligand for the clade 2 OPRs. Syntenic and phylogenetic analysis to place clade 2 GPCR and cognate neuropeptide genes onto GACs and their progenitors onto VACs was conducted with the same methods and rationale as presented for clade 1. Syntenic analysis and GAC placement of clade 2 GPCR and cognate peptide genes are found in supplementary figures S2 and S3, Supplementary Material online. VAC placement of the clade 2 GPCR progenitors parallel to their positions on the clade 2 phylogenetic tree is presented in figure 6.

Clade 2 GPCRs are mostly split between VAC_B and VAC_I, which contain five and eight pre-2R progenitor genes, respectively. Coincidentally, the MCR family also has four pre-2R progenitors on VAC_B (fig. 6). The translocation that separated the MCHR and SSTR families between VAC_I and VAC_B likely occurred subsequent to the divergence of the vertebrate and nonvertebrate chordate lineages but prior to 2R WGD, similarly to the MCHR6/7 translocation to VAC_C. Because the KISSR family emerged prior to chordate and vertebrate divergence, estimating the time of translocation from VAC_I to VAC_A is more difficult. Further syntenic
and phylogenetic analysis of the clade 2 GPCRs focused on translocations that occurred between the first and second rounds of WGD and ancestral gene family emergence is discussed in the supplementary information, Supplementary Material online.

The neuropeptide genes in this clade were allocated to VAC_B, VAC_I, VAC_F, and VAC_D (fig. 6). The presence of the PDYN/ENK/NOC/OMC on VAC_B, while NPB/W is located on VAC_I, raises the possibility that these progenitors emerged prior to the translocation from VAC_I to VAC_B, possibly being translocated with their cognate GPCR progenitors. Therefore, the cognate NPBW1/2 and OPRL/K/M/D progenitors, which appear to be vertebrate specific, may have emerged and diversified prior to the VAC_I to VAC_B translocation. The closely related URP and SST progenitors likely emerged from a single ancestral gene on VAC_F. This is consistent with a recent study demonstrating that somatostatin and urotensin peptides and receptors are derived from a single ancestral ligand-receptor pair before 2R WGD (Tostivint et al. 2014). Likewise, the SSTRs and UTS2Rs are mainly located on VAC_I, suggesting that these GPCRs likely belong to the same family. The remaining ligands PMCH1, SPX, GAL, and KISS are all located on VAC_D. The presence of half of the clade 2 neuropeptide progenitors on VAC_D, the split of their neighboring genes VAMP1, 2, and 3 over both GAC_F and GAC_D, and the split of AURKA/B/C over GAC_D, GAC_B, and GAC_F implicates VAC_D as the original location of the clade 2 peptides, as all related gene groups can be found on VAC_D (supplementary fig. S3, Supplementary Material online).

Syntenic and Phylogenetic Analysis of Clade 3 GPCR and Neuropeptide Genes

Clade 3 GPCRs include the orphan GPR37, endothelin (EDNR), and neuromedin-B (NMBR)/bombesin receptor subtype (BRS)/gastrin-releasing peptide (GRPR) receptor families. Syntenic analysis and GAC placement of clade 3 GPCRs and neuropeptides are found in supplementary figures S4 and S5, Supplementary Material online, respectively. VAC placement against the clade 3 phylogenetic tree is presented in figure 7A. The NMBR/GRPR/BRS3 gene family and the paralogs of neighboring gene families are divided over GAC_F3, GAC_F4, GAC_B2, and GAC_B3, indicating an intra-2R translocation from GAC_F to GAC_B. The EDNR and neighboring gene families are split over GAC_F3, GAC_F4, and GAC_C1, indicative of mislabeling of the GAC_C1 region in the Nakatani model, and the GPR37 progenitor was likely translocated from VAC_F to VAC_D prior to 2R WGD. The GPR37 and EDNR progenitors appear to have arisen from a local duplication of an ancestral gene prior to vertebrate and nonvertebrate chordate divergence. This GPR37/EDNR ancestral gene originated from a duplication of an even older ancestral gene which also produced the GRP/BRS/NMBR progenitor which, barring gene loss in the protostome lineage, occurred after protostome/deuterostome divergence but shares ancestry with the protostomian drCCHA1R and ceNPR14 GPCRs (fig. 7A).

Syntenic analysis indicates that a translocation prior to 2R WGD also separated the clade 2 peptide progenitors over two VACs with EDN1/2/3/4 and GRP/NMB being located on VAC_B and VAC_A, respectively (supplementary fig. S5,
Clade 4 GPCRs include the orphan GPR139/142, thyrotropin-releasing hormone (TRHR), neurotensin (NTSR), orphan GPR39, growth hormone secretagogue (GHSR)/motilin (MLNR), and neuromedin-U (NMUR) families. Additionally, the NMOGPR, coNMUR4 (which, despite the name, is not part of the NMUR family), and elephant shark-specific CAPAR/GPR139-like families are also included. Syntenic analysis and GAC placement of clade 4 GPCRs and neuropeptides are described in supplementary figures S6 and S7, Supplementary Material online. VAC placement of clade 4 genes against the clade 4 phylogenetic tree is presented in figure 7B.

Mixed vertebrate-protostome branches indicate that the NMUR, TRHR, and coNMUR4 families emerged prior to the urbilatarian ancestor (fig. 7B). The absence of protostome representatives for the GHSR/MLNR/GPR39/NTSR and GPR139/142/NMOGPR families indicates gene loss in the protostome lineage, in addition to wide-spread nonvertebrate chordate gene loss.

Clade 4 GPCRs are located on VAC_B, VAC_E, VAC_F, and VAC_I. Due to the isolation of GPR139/142, the timing of the translocation to VAC_I that separated it from the rest of clade 4 cannot be estimated. The GHSR/MLNR and NTSR/GPR39 lineages both contain nonvertebrate chordate representatives. Therefore, the time of translocation that separated the two lineages over VAC_F (GHSR/MLNR), VAC_B (NTSR) and VAC_E (GPR39) cannot be defined except that it likely occurred after protostome and deuterostome divergence. Conversely, the duplication that created the GPR39 and NTSR progenitors from an ancestral gene appears to be vertebrate specific, thus, the translocation that placed the GPR39 progenitor onto VAC_E likely occurred subsequent to the divergence of vertebrate and nonvertebrate chordate lineages but prior to 2R WGD (fig. 7B). The CAPAR/GPR139-like, NMOGPR, and coNMUR4 families do not appear to have amniote representatives, which indicates loss of these entire gene families in both nonvertebrate and amniote lineages.

In contrast to the GPCR family, the majority of the clade 4 cognate neuropeptides are located on VAC_D, and the NMU/S progenitor is located on VAC_C (fig. 7B). The presence of the related GRM genes on GAC_Ds and GAC_C2 and the MGAT4 family genes, on GAC_Cs and GAC_D2, shows close evolutionary history between these VAC_D and VAC_C linkage groups (supplementary fig. S7, Supplementary Material online). Therefore, prior to 2R WGD, the NMU/S progenitor gene, along with neighboring genes, was likely translocated from VAC_D to VAC_C. As the cognate neuropeptides for the GPCRs phylogenetically closest to GPR39 are located on VAC_D, the cognate neuropeptide for GPR39 is also likely located in a VAC_D linkage group, possibly sharing sequence similarity with NTS1/2. As obestatin, a product of the GHR1 gene, was claimed not to be a ligand for GPR39 (Zhang et al. 2005; Chartrel et al. 2007), we did not make a connection between GHRI and GPR39.

Supplementary Material online. The unknown progenitor peptide for GPR37 has been tentatively placed on VAC_B as the ligand of GPR37’s closest relative, EDN1/2/3/4, is found on VAC_B.

Syntenic and Phylogenetic Analysis of Clade 4 GPCR and Neuropeptide Genes

Clade 5 GPCRs include the orphan GPR19, arginine vasopressin (AVP)/oxytocin (OTR), neuropeptide-S (NPSR), orphan GPR150, and gonadotropin-releasing hormone (GnRHR)
receptor families. Syntenic analysis and GAC placement of clade 5 GPCR and peptide genes are provided in supplementary figures S8 and S9, Supplementary Material online. VAC placement against the clade 5 phylogenetic tree is presented in figure 8.

VAC placement showed that the genes in this clade are split between VAC_D and VAC_A (fig. 8). The GnRHR, AVPR, and NPSR families possess closely related protostomian GPCRs, indicating that each of these lineages emerged in a preurbilaterian ancestor. The lack of a protostomian GPR150 suggests either loss of this branch in protostomes or emergence from a common ancestor of the NPSR family during deuterostome development, followed by rapid differentiation. The translocation that separated the related GPR150 and NPSR progenitors between VAC_D and VAC_A occurred subsequent to the generation of these two families. The GPR19 family, which is also present on VAC_D, is somewhat distant from the rest of clade 5, indicating loss of this distinct branch in the protostome lineage.

Synteny analysis revealed that all the discovered cognate neuropeptides for clade 5 GPCRs are located on VAC_C (fig. 8). The Nakatani model labels AVP, OXT, and GnRH2 with unknown synteny, but our analysis confirmed that this linkage group is located on GAC_C1. The two orphan GPCR progenitors in this clade are GPR19 and GPR150, and our analysis supports the notion that the cognate neuropeptides for each of these may also be found on VAC_C (supplementary fig. S9, Supplementary Material online).

**Discussion**

Although neuropeptides and their GPCRs are well documented in vertebrates, the evolutionary mechanisms for the development and interfamily relationships of neuropeptides and GPCRs remain unclear. In this study, we retrieved rhodopsin-like GPCR and cognate neuropeptide gene sequences from representative vertebrate and invertebrate species and analyzed their phylogenetic relationships and the locations of progenitor genes on reconstructed VACs. The most important finding of this study is that the progenitors of phylogenetically related families tend to be located on the same VACs, which strongly supports the notion that these gene families emerged through sequential local duplications during early metazoan evolution. Previous studies have shown proliferation within gene families by local duplication, such as the duplications within the NPYR and SSTR families prior to 2R WGD (Lagerstrom et al. 2005; Liu et al. 2010; Larhammar and Bergqvist 2013; Tostivint et al. 2014). The current analysis reveals a more ancient pattern of local duplications that gave rise to new families. This pattern is most evident in clade 1, in which 13 of the 16 GPCR family progenitors are located on VAC_C and 7 of the 9 cognate neuropeptide progenitors are located on VAC_E. Additionally, the majority of clade 4 and 5 peptide progenitors are also located on single VACs. This observation reveals that the cognate neuropeptide progenitor genes of phylogenetically related GPCR families tend to be located on parallel VACs, strongly suggesting that neuropeptide families also multiplied through local duplications by the same pattern as their cognate GPCRs.

In contrast to clade 1, the majority of GPCR progenitors in clades 2, 3, and 5 are split between two VACs. The split of a single clade between two, or more, chromosomes prior to vertebrate emergence can occur by the same mechanisms by which translocations subsequent to 2R WGD lead to the splitting of single GACs over multiple locations in extant
genomes (Nakatani et al. 2007). The Nakatani model only rebuilds putative VACs dated shortly before 2R WGD; therefore, the model does not account for prior translocations. Phylogenetic and syntactic analysis can be used to predict the time periods of some translocations such as the clade 2 \textit{VAC}_I-VAC_B and clade 4 \textit{VAC}_B-VAC_E translocations that occurred subsequent to the divergence of the vertebrate and nonvertebrate chordate lineages.

In addition, chromosomal translocations between the first and second WGDs can account for the spread of ohnologous genes, originally located on a single VAC, across multiple distinct GACs. Furthermore, distinct linkage groups that have been translocated next to each other may undergo secondary translocations which hybridize the two linkage groups. These complications can lead to misannotation in the Nakatani model. Confirmation of chromosomal translocation through in-depth synteny of invertebrate genomes and detailed analysis of the veracity of the Nakatani model are both beyond the scope of this article but may be addressed in the future.

This study also reveals a highly dynamic process of emergence and loss of GPCR gene families through metazoan evolution, demonstrated in the clade 4 \textit{conNMUR4}, \textit{GPR139/142}, and \textit{NMOGP/GPR139-like} families (independent branch in Bayesian phylogenetic analysis) which proliferated, separately, in the urbilatarian, protostome, and vertebrate ancestors, but have been subsequently lost in nonvertebrate chordates and suffered extensive gene loss in amniotes. Ancient genes, as a result of local duplication followed by subsequent diversification during metazoan evolution gave rise to a variety of gene families. Emerging gene families undergo intrafamily and interfamily competition which can result in the loss of a family member, gene family, or entire class of gene (Krishnan et al. 2012; Assis and Bachtrog 2013). Therefore, in addition to lineage-specific duplications, gene loss in protostomes may have generated what appear to be deuterostome-specific gene families, and vice versa. This pattern of duplication, differentiation, and lineage-specific gene loss produced the progenitors for the current diversity of vertebrate gene families (Nordstrom et al. 2009). More recently, 2R of WGD in the vertebrate ancestor led to, on average, a 4-fold increase in gene family members. Subsequent duplications, gene loss, and neo-/subfunctionalization have given rise to the current number and diversity of vertebrate gene family members.

The conventional process of GPCR family expansion is based on gene duplication followed by gradual differentiation and functionalization in which interaction is preserved within established GPCR-neuropeptide gene family systems. However, there are other ways of differentiation; one example is the \textit{MCR-ACTH} system found on \textit{VAC}_B in clade 2. This system does not share strong phylogenetic ancestry with other neuropeptide GPCRs, despite its cognate neuropeptides being located in the same gene, \textit{POMC}, as a neuropeptide for the opioid GPCRs. The MCRs appear to be more closely related to the \(\alpha\)-group of rhodopsin-like GPCRs (Fredriksson et al. 2003). By chance, the \textit{MCR} progenitor likely arose by local duplication of a member of the \textit{MECA} gene group and came under interaction with the ACTH peptide prior to 2R WGD (Haitina et al. 2007). A local duplication of the \textit{PNOC} gene, likely subsequent to 2R WGD, placed the opioid neuropeptide into the same region as the \textit{ACTH} gene, creating a hybrid gene. Subsequently, intragenic duplications of the \textit{MSH} subunit of \textit{ACTH}, along with proliferation of the \textit{MCR} family, led to the creation of an entirely new system that has only been found in vertebrates (Harris et al. 2014).

The rates of gene loss through various stages of vertebrate evolution can be estimated based on the number of pre-2R progenitors and extant GPCR ohnologs in various vertebrate species (table 1). These data demonstrate the loss of many ohnologous genes subsequent to 2R WGD. For instance, one may postulate that the presence of three pre-2R \textit{NPYR} paralogs may give rise to 12 ohnologs post-2R; however, humans have only four members, coelacanth and spotted gar have seven, and elephant shark has six. In general, loss of ohnologous genes in the human lineage throughout GPCR and cognate neuropeptide families is greater than that of lineages which have undergone lower rates of genomic change, such as coelacanth (Amemiya et al. 2010), spotted gar (Amores et al. 2011), and elephant shark (Venkatesh et al. 2014). When tracing the evolutionary history of a gene family the more retained paralogs that we find the better understanding we acquire. Therefore, choosing specific representative species is very important. Of mammals, although humans are not the most conserved genome the manually annotated and the sequences are more conserved than many other mammals (Burt et al. 1999). Of the tetrapods, avians, chicken in particular, have some of the best preserved chromosomes (Nishida et al. 2008). Teleost fish, that underwent a third WGD, are exceedingly difficult to analyze due to the doubling of linkage groups and large-scale gene loss. However, spotted gar has recently been considered the best vertebrate representative as this species represents an ancestral pre-3R teleost and has undergone fewer translocations than other vertebrates with mapped genomes (Amores et al. 2011). In the future a fully mapped coelacanth genome could replace chicken and human as the tetrapod representative, and may be even more genomically conserved than gar. Additionally, the elephant shark, which is both a cartilaginous fish and the slowest evolving vertebrate known (Venkatesh et al. 2014), is ideal for vertebrate genomic evolution research. Additionally, lamprey could provide an interesting perspective on the early vertebrate genome because it may have diverged from the jawed vertebrate lineage shortly after 2R WGD (Caputo Barucchi et al. 2013), although there is uncertainty about the number of WGDs that it has undergone (Mehta et al. 2013). Finally, although \textit{B. florisdae} is a useful basal chordate representative (Elphick and Mirabeau 2014) an even slower evolving basal chordate, \textit{Asymmetron lucayanum}, exists (Yue et al. 2014). Unfortunately, the genomes of coelacanth, elephant shark, lamprey, and \textit{A. lucayanum} remain unmapped, their chromosomes broken into many separate scaffolds. Arrangement of these scaffolds into full chromosomes would provide a great benefit to many areas of genomic research.

An advantage of this study is that it provides clues to identify novel neuropeptide genes for orphan GPCRs. In our analysis, we included orphan GPCRs that are phylogenetically
close to established neuropeptide GPCRs. It was found that many of these orphan GPCRs reside on VACs with their related neuropeptide GPCRs. By aligning the known cognate neuropeptide genes against GPCR genes on VACs, the relative positions of cognate neuropeptides for orphan GPCRs can be postulated. While speculative, this determination can help to target gene discovery both by identifying the likely linkage group of the unidentified peptide genes and by showing the sequences of the closest known relatives of the peptides. For example, the ligand for GPR150 is currently unknown; however, based on our analysis, the ligand is most likely on a VAC_C linkage group and likely has sequence similarity to NPS. Using this method, spexin was correctly predicted to be a natural ligand for galanin GPCRs (Kim et al. 2014), and linkage group analysis of secretin peptide and GPCR family genes (Hwang et al. 2013) was helpful for the identification of a receptor for a novel GCPR peptide (Park et al. 2013). In addition, once a new gene is identified, comparison of the gene to its closest evolutionary relatives allows for faster analysis of its putative functions.

Taken together, this study suggests a provisional model for the evolutionary history of rhodopsin-like neuropeptide GPCRs. Multiple duplications of an ancestral rhodopsin-like GPCR, which likely arose in early metazoan-fungal evolution from a duplication of a cyclic-AMP receptor (Krishnan et al. 2012), during early metazoan evolution gave rise to at least five progeny each of which, through further duplications, acted as the progenitors of the five clades of rhodopsin-like GPCRs found in extant vertebrates. Concurrently, chromosomal shuffling spread these initial five GPCR genes, and their subsequent duplications, around the genomes of successive vertebrate ancestors. Correspondingly, cognate neuropeptide gene families have coevolved in a manner similar to that of GPCRs. In the future, complete assembled genomes of invertebrate and early vertebrate species as well as reconstructed vertebrate linkage groups with higher resolution will provide better means for further delineation of vertebrate genome evolution.

**Materials and Methods**

**Data Retrieval**

The amino acid sequences of the rhodopsin-like neuropeptide GPCRs and cognate peptides were downloaded from the Ensembl Genome Browser (http://www.ensembl.org), the NCBI GenBank database (http://www.ncbi.nlm.nih.gov/Entrez), or the B. floridæ genome database (http://genomejgi-psf.org/Braf11/Braf11.home.html). Search tools provided by the Ensembl Genome Browser were used for the initial identification of orthologous and paralogous genes from vertebrate and invertebrate species. This data set was used to identify the vertebrate and invertebrate members of rhodopsin-like neuropeptide GPCRs that were not already registered in these databases. The genome databases of human, mouse, chicken, anole lizard, Xenopus tropicalis, coelacanth, zebrafish, medaka, fugu, stickleback, tetraodon, spotted gar, elephant shark, lamprey, Ci. intestinalis, B. floridæ, C. elegans, and Drosophila were manually searched with the TBLASTN algorithm. Database GPCR prediction was reasonably reliable, requiring only occasional manual editing; however, neuropeptide prediction was less reliable due to low sequence conservation. Thus, many recorded sequences were manually revised for incorrect splice sites and stop/start codons. To determine full-length open-reading frame sequences, exons and splice junctions were defined manually using both the Ensembl protein report service and an HMMgene/ (http://www.cbs.dtu.dk/services/HMMgene/) that was provided by the CBS Prediction Service. Putative signal peptides were predicted using SignalP 3.0 (http://www.cbs.dtu.dk/services/SignalP/). Duplicates were removed from this data set using phylogenetic analysis. Orthology and paralogy of the genes were determined based on phylogeny and synteny analyses. Accession numbers of the sequences are listed in supplementary tables S1–S11, Supplementary Material online.

**Phylogenetic Analysis of Rhodopsin-Like Neuropeptide GPCRs**

Complete GPCR amino acid sequences were aligned using MUSCLE (multiple sequence comparison by log-expectation) as implemented in MEGA 6.06 (Tamura et al. 2013). Maximum-likelihood trees were constructed using the Jones–Thornton–Taylor model with 100 bootstrap replicates and partial deletion with 95% site coverage cut-off selected in the gaps/missing data treatment. Numbers on the nodes of the trees indicate the bootstrap replicates corresponding to each branch. Bayesian analysis was conducted with the same maximum-likelihood database with Bayesian Evolutionary Analysis Sampling Tree (BEAST) (Drummond and Rambaut 2007). BEAUti v1.8.2 was used to create BEAST input file using Blosum62 substitution, the Gamma site heterogeneity model, a strict clock, and a Markov Chain Monte Carlo length of 10⁶.

**Genome Synteny Analysis and Tracing the Duplication History of Rhodopsin-Like Neuropeptide and GPCR Families**

Multispecies genome synteny analysis was performed by comparing genome regions, which contain the rhodopsin family neuropeptide and GPCR genes. The chromosomal locations of the orthologs and paralogs of neighboring genes were obtained from the Ensembl Genome Browser to create synteny blocks (Cunningham et al. 2015). According to the method of Yegorov and Good (2012), chromosome fragments with reliable synteny were matched to reconstructed, putative pre-2R VAC models created by Nakatani et al. (2007). The Nakatani model provides putative GAC linkage groups that are displayed alongside their location in individual human chromosomes and also indicates the location of human orthologs and ohnologs in medaka, chicken, and mouse (Nakatani et al. 2007). Thus, we were able to compare the chromosome fragments containing our genes of interest against VAC linkage groups to resolve the positions of the genes of interest at consecutive stages of vertebrate genome evolution (pre-2R VACs and post-2R GACs). Other species were analyzed by first matching the synteny blocks with...
their human orthologs and then comparing against the Nakatani model.

Whole-Genome Analysis for Chromosomal Regions with High Sequence Similarity to Chromosome 4
The amino acid sequences of known proteins in human chromosome 4 were downloaded from the Ensembl biomart database (http://www.ensembl.org). Sequence similarity analysis against all other proteins encoded in the human genome was then performed using a custom BLASTP pipeline implemented in the Perl programming language. For each query protein a target protein, which had the lowest matching e value, was identified. Chromosomal regions with a match density of less than 3 per 10 Mb were discarded. The syntenic relationships between chromosomes were plotted using the Circos software package (Krzywinski et al. 2009).

Supplementary Material
Supplementary information, figures S1–S9, and tables S1–S11 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments
This work was supported by grants from the Brain Research Program (2011-0019205) and Basic Science Research Program (2013R1A1A2010481) of the National Research Foundation of Korea funded by the Ministry of Education, Science, and Technology.

References
Larhammar D, Bergqvist CA. 2013. Ancient grandeur of the vertebrate neuropeptide Y system shown by the coelacanth Latimeria chalumnae. Front Neurosci. 7:27.
multiple forms of kisspeptins and GPR54 receptors in vertebrates.
Endocrinology 150:2837-2846.


