Evolutionary and Comparative Genomics to Drive Rational Drug Design, with Particular Focus on Neuropeptide Seven-Transmembrane Receptors

Michael Furlong and Jae Young Seong*

Graduate School of Biomedical Sciences, Korea University, Seoul 02841, Republic of Korea

Abstract

Seven transmembrane receptors (7TMRs), also known as G protein-coupled receptors, are popular targets of drug development, particularly 7TMR systems that are activated by peptide ligands. Although many pharmaceutical drugs have been discovered via conventional bulk analysis techniques the increasing availability of structural and evolutionary data are facilitating change to rational, targeted drug design. This article discusses the appeal of neuropeptide-7TMR systems as drug targets and provides an overview of concepts in the evolution of vertebrate genomes and gene families. Subsequently, methods that use evolutionary concepts and comparative analysis techniques to aid in gene discovery, gene function identification, and novel drug design are provided along with case study examples.

Key Words: Neuropeptide, 7TMR, G protein-coupled receptor, Coevolution, Gene duplication, Whole genome duplication, Evolutionary history

INTRODUCTION

Seven transmembrane receptors (7TMRs) represent the largest membrane bound receptor superfamily in humans, with over 840 members (Oh *et al*., 2006; Lagerstrom and Schioth, 2008). Genomic analysis of predictable pharmaceutical drug targets indicates that 7TMRs make up 19% of the drugable proteome, and that 36% of existing drugs target 7TMRs (Rask-Andersen *et al*., 2011). Worldwide, as of 2014, 7TMR targeting drugs have a market value of $100 billion, which is expected to grow to $115 billion by 2018 (Ufuk, 2014).

7TMRs are cylindrical receptor proteins usually found in the cellular membrane and are involved in signal transduction, by which chemical messengers found outside of the cell are able to alter intra-cellular protein activity and gene expression. 7TMRs possess seven hydrophobic transmembrane α-helices, which anchor the receptor into the membrane layer. The α-helices are connected by extra- and intracellular loops, in addition to an extracellular N-terminal strand and an intracellular C-terminal strand (Katritch *et al*., 2012). Ligands bind to pockets formed by the extracellular domains and/or transmembrane α-helices and induce conformational changes which modulate the intracellular domains' ability to interact with various intracellular messenger proteins such as G-proteins and β-arrestins (Conroy *et al*., 2015; M'Kadmi *et al*., 2015). Vertebrate 7TMRs can be categorised into the glutamate, rhodopsin, adhesion, frizzled/taste2, and secretin families based on the GRAFS system. The rhodopsin-like family can be further subdivided into α, β, γ, and δ subgroups (Fredriksson *et al*., 2003). Endogenous ligands for 7TMRs include peptides, amines, lipids, nucleotides, ions, and even photons.

7TMRs that are targeted by the neuropeptide category of peptide ligands are grouped in the secretin-like and β and γ groups of the rhodopsin-like 7TMRs (Fredriksson *et al*., 2003). The rhodopsin-like 7TMRs and cognate neuropeptides can then be further categorised into 5 clades (Yun *et al*., 2015) and the secretin-like 7TMRs and cognate neuropeptides into 5 families (Hwang *et al*., 2013), as displayed in Fig. 1. To date, over 30 7TMR-neuropeptide families have been identified in humans consisting of over 70 7TMR genes and over 60 neuropeptide genes, with a further 8 orphan 7TMRs in 6 families with no known ligands (Table 1).

Neuropeptide-interacting 7TMRs mediate a multitude of roles in the nervous system and peripheral organs and influ-
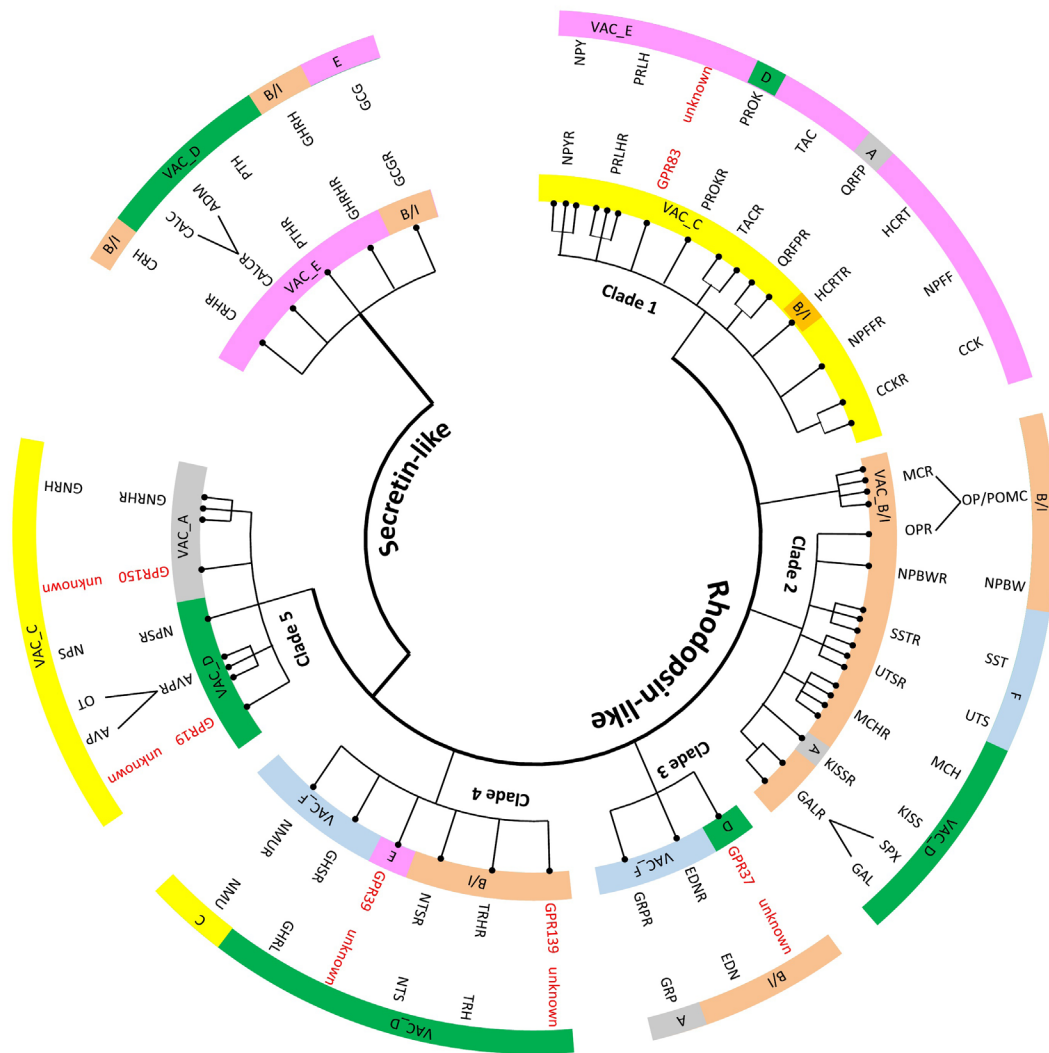
---

www.biomolther.org

Fig. 1. A simplified schematic of a phylogenetic tree of the pre-2R progenitor genes of the 5 clades of rhodopsin peptide-interacting 7TMRs, with the inclusion of the MCR family, and the single clade of secretin peptide-interacting 7TMRs. The inner ring consists of the phylogenetic schematic of the 7TMR progenitors and their VAC placement, and the outer ring consists of the neuropeptide ligand progenitors and their VAC placement. Lines between 7TMRs and their ligands indicate pre-2R interaction between multiple progenitor 7TMRs and ligands. 7TMRs annotated in red indicate the absence of known ligands for these receptors. The VAC bars are colour coded green, pink, tan, yellow, blue, and grey to represent VACs D, E, I/B, C, F, and A, respectively.

ence a number of physiological and psychological processes including reproduction, growth, homeostasis, metabolism, food intake, sleep, and social and sexual behaviours (Cho *et al*., 2007; Mirabeau and Joly, 2013; Vaudry and Seong, 2014). Mature neuropeptides are produced by cleavage of larger precursor proteins which share typical motifs in their amino acid sequences. These motifs include a signal peptide sequence at the N-terminus and an evolutionary conserved mature neuropeptide sequence, which is often flanked by cleavage sites (Steiner, 1998).

As larger quantities of biological data become available, rational, specifically tailored techniques are emerging. For example, novel peptide genes have been identified by tailoring genomic-analysis algorithms to search for specific peptide motifs (Mirabeau *et al*., 2007). In addition, comparative and evolutionary genomic analysis techniques are aiding in the

delineation of the mechanisms that drive the emergence of gene families, permitting the determination of novel peptide and 7TMR interaction (Hwang *et al*., 2013; Park *et al*., 2013; Kim *et al*., 2014a). Therefore, this article will discuss, using previously published data, how an understanding of gene family evolution, within the context of vertebrate genome evolution over the past 500 million years, can provide valuable insight for gene discovery, gene function identification, and drug design for peptide ligand-7TMR families.

GENE FAMILY EXPANSION VIA INDIVIDUAL GENE AND WHOLE GENOME DUPLICATION

Gene families are established through evolutionary processes such as gene/genome duplication followed by altera-

Table 1. Rhodopsin-like and Secretin-like 7TMR and their cognate neuropeptide gene families

Clade	Family	7TMR		Neuropeptide	
		Pre-2R progenitors	Post-2R members	Post-2R members	Pre-2R progenitors
1	Cholecystokinin	CCKR A/B	CCKAR	CCK	CCK/GAST
			CCKBR	GAST/2*	
		CCKR 3/4	CCK3R		
			CCK4R		
	Neuropeptide FF	NPFFR 1/2/3	NPFFR1	NPFF	NPFF/NPVF
			NPFFR2	NPVF	
			NPFFR3		
	Hypocretin	HCRTR 1/2	HCRTR1	HCRT	HCRT 1/2
			HCRTR2	HCRT2	
	Tachykinin	TACR 1/2/3	TACR1	TAC1	TAC 1/3/4
			TACR2	TAC3	
			TACR3	TAC4	
		TACR 4/5	TACR4/5*		
	Prokineticin	PROKR 1/2/3	PROKR1/2*	PROK1	PROK 1/2/3
			PROKR3	PROK2	
				PROK3	
	Orphan 83	GPR83 1/2/3	GPR83-1	Unknown	Unknown
			GPR83-2/3*		
	Prolactin releasing peptide	PRLHR 1	PRLHR1	PRLH1	PRLH 1/2
				PRLH2	
		PRLHR 2/3	PRLHR2		
			PRLHR3		
		PRLHR 4/5	PRLHR4		
			PRLHR5		
	Neuropeptide Y	NPYR 1/3/4/6	NPY1R	NPY/2*	NPY/PYY/PPY
			NPY3R	PYY/PYY*	
			NPY4R		
			NPY6R		
		NPYR 2/7	NPY2R		
			NPY7R		
		NPYR 5	NPY5R		
	Pyroglutamylated Rfamide peptide	QRFPR 1/2/3	QRFPR1/1ii*	QRFP	QRFP 1/2
			QRFPR2	QRFP2	
			QRFPR3		
		QRFPR 4	QRFPR4		
2	Galanin/Spexin	GALR1	GALR1A	GAL	GAL
			GALR1B	GALP	
		GALR2/3	GALR2A	SPX1	SPX
			GALR2B	SPX2/2b*	
			GALR3		
	Kisspeptin	KISSR	KISSR1	KISS1	KISS 1/2/3
			KISSR2	KISS2	
			KISSR3	KISS3	
			KISSR4		
	Urotensin-2	UTS2R-1/4	UTS2R-1	URP/2*	URP/1/2, UTS2
			UTS2R-4	UTS2	
				URP1	
		UTS2R-2	UTS2R-2		
		UTS2R-3/5	UTS2R-3		
			UTS2R-5		

Table 1. Continued

Clade	Family	7TMR		Neuropeptide	
		Pre-2R progenitors	Post-2R members	Post-2R members	Pre-2R progenitors
	Melanin-concentrating hormone	MCHR 1/3	MCHR1	PMCH1	PMCH1
			MCHR3		
		MCHR2	MCHR2		
		MCHR 4/5	MCHR4		
			MCHR5		
		MCHR 6/7	MCHR6		
			MCHR7		
	Somatostatin	SSTR 1/4/6	SSTR1	SST1/3/4*	SST 1-6
			SSTR4	SST2/6*	
			SSTR6	SST5	
		SSTR 2/3/5	SSTR2		
			SSTR3		
			SSTR5		
		SSTR 7	SSTR7		
	Neuropeptide-B/W	NPBWR 1/2	NPBWR1	NPW	NPB/NPW
			NPBWR2	NPB	
	Opioid	OPR D/K/L/M	OPRD	PENK	PDYN/PENK/
			OPRK	PDYN	PNOC/POMC
			OPRL	PNOC/POMC*	
			OPRM		
	Melanocortin	MCR 1	MC1R	POMC*	
		MCR 2	MC2R		
		MCR 3/5	MC3R		
			MC5R		
		MCR 4	MC4R		
3	Neuromedin-B/Bombesin subtype 3/ Gastrin-releasing peptide	NMBR/BRS3/GRPR	NMBR	GRP	GRP/NMB
			BRS3	NMB	
			GRPR		
	Endothelian	EDNR A/B/B2	EDNRA	EDN1	EDN
			EDNRB	EDN2	
			EDNRB2	EDN3	
				EDN4	
	Orphan 37	GPR37 /L1	GPR37	Unknown	Unknown
			GPR37L1		
4	Neuromedin-U	NMUR 1/2/3	NMUR1	NMU	NMU/NMS
			NMUR2	NMS	
			NMUR3		
	Growth hormone secretagogue/Motilin	GHSR 1/2/3 MLNR	GHSR	GHS	GHS/MLN
			GHSR2	MLN	
			GHSR3		
			MLNR		
	Orphan 39	GPR39-1/2	GPR39	Unknown	Unknown
			GPR39-2		
	Neurotensin	NTSR 1/2	NTSR1	NTS	NTS
			NTSR2	NTS2	
	Thyrotropin-releasing hormone	TRHR 1/2/3	TRHR1	TRH1/2*	TRH
			TRHR2		
			TRHR3		
	NMUR4	NMUR4	NMUR4	Unknown	Unknown
	Orphan 139/142	GPR 139/142	GPR139	Unknown	Unknown
			GPR142		
	NMOGPR**	Unclear	Unclear	Unknown	Unknown
	Orphan 139-like**	Unclear	Unclear	Unknown	Unknown

Table 1. Continued

Clade	Family	7TMR		Neuropeptide	
		Pre-2R progenitors	Post-2R members	Post-2R members	Pre-2R progenitors
5	Gonadotropin-releasing hormone	GnRHR1A	GnRHR1A	GnRH1 GnRH2 GnRH3	GnRH
		GnRHR1 B/C	GnRHR1B GnRHR1C		
		GnRHR2 A/B/C	GnRHR2A GnRHR2B GnRHR2C		
	Orphan 150	GPR150-1/2	GPR150 GPR150-2	Unknown	Unknown
	Neuropeptide-S	NSPR	NPSR	NPS	NPS
	Arginine vasopressin/Oxytocin	OTR/AVPR1 A/B	OTR AVPR1A AVPR1B	OXT/2*	OXT
				AVP	AVP
		AVPR 2 AVPR 3/4/5	AVPR2 AVPR3 AVPR4 AVPR5		
	Orphan 19	GPR19	GPR19/-2/-3*	Unknown	Unknown
Secretin-like	Corticotropin-releasing hormone/ Urocortin	CRHR 1/2	CRHR1 CRHR2	CRH UCN UCN2 UCN3	CRH/UCN
	Calcitonin/Islet amyloid polypeptide/ Adrenomedullin	CALCR /L	CALCR CALCRL	CALCA/B* IAPP ADM1 ADM2	CALC/IAPP ADM 1/2
	Parathyroid hormone	PTHR 1/2/3	PTH1R PTH2R PTH3R	PTH1 PTH2 PTHLH	PTH 1/2/LH
	Glucagon/Glucose-dependent insulinotropic polypeptide/ glucagon related peptide	GLP2R GLP1R GCGR/GIPR/ GCRPR**	GLP2R GLP1R GCGR GIPR GCRPR	GCG GIP GCRP	GCG/GIP/GCRP
	Growth hormone-releasing hormone/ Secretin/Vasoactive intestinal peptide/pituitary adenylate cyclase-activating polypeptide	GHRHR1 SCTR ADCYAP1R/ GHRHR2/3/ VIPR1/2**	GHRHR1 SCTR ADVCYAPR GHRHR2/3* VIPR1/2*	GHRH SCT PACAP VIP	GHRH/SCT/ PACAP/VIP

This table lists the known gene families of neuropeptide-interacting 7TMRs, divided into the 5 clades of the rhodopsin-like and single clade of the secretin-like. It lists the pre 2R progenitors of each family, as well as the post-2R products and subsequent duplications *Genes which are absent, or present as pseudogenes, in human are noted in grey. Some families possess unclear relationships. **Alongside each 7TMR family endogenous ligand genes for members of each 7TMR family are listed.

tions to gene function or gene loss (Abi-Rached *et al*., 2002; Larhammar *et al*., 2002; Holland, 2003; Santini *et al*., 2003; Vienne *et al*., 2003; Kim *et al*., 2011; Hwang *et al*., 2013). In particular, the quadruplication of genes by two rounds (2R) of whole genome duplication during early vertebrate evolution facilitated the rapid proliferation of genes within vertebrate gene families. These families emerged and were populated via continuous local gene duplications prior to 2R (Hwang *et al*., 2013). Subsequent to 2R, heavy gene loss and additional local gene duplications followed by differentiation and functionalization has created the current variety of vertebrate gene families (Lundin, 1993; Holland *et al*., 1994; Larhammar and

Salaneck, 2004; Hwang *et al*., 2013, 2014; Kim *et al*., 2014a; Sefideh *et al*., 2014).

Imperfections in the DNA replication process on an evolutionary time scale results in divergence of the amino acid sequences of duplicated genes. Advantageous changes, which enhance gene function and organism reproduction, are more likely to be retained in the species while detrimental changes are more likely to be lost. Neutral changes, which have no overall effect on organism survivability, accumulate at a slower rate (Lynch *et al*., 2001). Subsequent to duplication, daughter genes accumulate different mutations which result in a process of differentiation and functionalisation. There are numerous categories of functionalisation that are influenced by factors such as chromosomal location, the method of gene duplication, the location of the mutation, the gene type, and the replication rate of the species (Jensen and Bachtrog, 2011). However, the typical pattern is that one duplicate retains a larger proportion of the original functionality which leaves less function-conservation pressure on the other duplicate(s). This allows greater freedom for mutations to accumulate subsequent to duplication (Assis and Bachtrog, 2013). If the duplicate gene(s) is not rendered non-functional then a process of sub-functionalisation and specialisation often occurs; the expression patterns of the genes differentiate and each gene gains partial functionality of the pre-duplication gene. As the duplicates continue to diverge neofunctionalisation may occur and novel functions that did not exist in the pre-duplicate gene may be acquired (He and Zhang, 2005; Rastogi and Liberles, 2005; Gibson and Goldberg, 2009; Klingel *et al*., 2012).

PHYLOGENY AND SYNTENY FOR DELINEATING GENE FAMILY EMERGENCE

Phylogenetic analysis, using the amino acid sequence of gene products, is an important tool to delineate evolutionary relationships among genes from different taxa. Together with recent advances of bioinformatics tools to discover novel genes, a large amount of protostomian and deuterostomian data are rapidly being accumulated (Mirabeau *et al*., 2007). In addition, reverse pharmacological approaches in invertebrates (Hauser *et al*., 2006; Lindemans *et al*., 2009; Jiang *et al*., 2013) and vertebrates (Civelli *et al*., 2006) have facilitated discovery of a great number of peptide-7TMR families. It is of importance to note that phylogenetic analyses of protostomian and deuterostomian sequences of peptide 7TMRs show a large number of family subtrees containing both protostomian and deuterostomian 7TMRs, indicating that many vertebrate7TMR families originate prior to the divergence of deuterostomes and protostomes (Mirabeau and Joly, 2013). However, phylogenetic analysis without knowing the location of selected genes within the genomes of species often fails to correctly determine the evolutionary process for the establishment of a gene family (Abi-Rached *et al*., 2002; Larhammar *et al*., 2002).

Syntenic analysis involves the comparison of the locations of orthologous or paralogous genes between chromosomes, within or among species, which facilitates more accurate analyses of the origins and relationships of individual peptide-7TMR families (Cerda-Reverter *et al*., 2000; Lagerstrom *et al*., 2005; Lee *et al*., 2009; Kim *et al*., 2011, 2012; Dores, 2013; Osugi *et al*., 2014). However, small scale synteny analysis is only suitable for determining the evolutionary history of closely related families. Delineating the evolutionary mechanisms for larger gene families with less closely related members is particularly difficult due to the scattered distribution of the genes on many different chromosomes. Thus, to elucidate the evolutionary history of a superfamily, such as 7TMRs and their neuropeptide ligands, large scale synteny such as comparison of large segments of chromosomes among species that represent a wide selection of the vertebrate clade is also required.

Comparisons of entire genomes between evolutionarily distinct taxa have led to reconstructions of hypothetical ancestral chromosomes of early vertebrates or chordates (Nakatani *et al*., 2007; Putnam *et al*., 2008), which support the hypothesis that 2R occurred during early vertebrate emergence, approximately 500 million years ago. 2R produced, on average, four gnathastome ancestral chromosomes (GACs) that share related sets of genes, defined as ohnologs, from pre-2R progenitor vertebrate ancestral chromosomes (VACs) (Dehal and Boore, 2005; Meyer and Van de Peer, 2005; Nakatani *et al*., 2007; Putnam *et al*., 2008). Assigning GAC and VAC positions to members of a gene family provides a fast and relatively accurate tool to aid in tracing the origins of gene super families (Yegorov and Good, 2012; Hwang *et al*., 2013; Yun *et al*., 2015). For instance, the neuropeptide Y (*NPYR*), prolactin-releasing peptide (*PRLHR*), orphan G protein-coupled receptor 83 (*GPR83*), prokineticin (*PROKR*), tachykinin (*TACR*), neuropeptide FF (*NPFFR*), hypocretin (*HCRTR*), cholecystokinin (*CCKR*), and pyroglutamylated RFamide peptide (*QRFPR*) families are phylogenetically grouped in clade 1 of the rhosodpin-like neuropeptide-interacting 7TMRs and are located on VAC_C, except for the *HCRTR* family (Fig. 1). The receptor families in clade 5 consisting of the orphan GPR19, arginine vasopressin (*AVP*)/oxytocin (*OTR*), neuropeptide-S (*NPSR*), orphan GPR150, and gonadotropin-releasing hormone (*GnRHR*) are mainly located on VAC_D or VAC_A (Yun *et al*., 2015) (Fig. 1). The secretin-like 7TMR families comprising of corticotropin releasing hormone receptor (*CRHR1*), calcitonin receptor (*CALCR*), parathyroid hormone receptor (*PTHR*), growth hormone-releasing hormone receptor (*GHRHR*, which also includes the secretin, vasoactive intestinal peptide, and pituitary adenylate cyclase-activating polypeptide receptors), and glucagon receptor (*GCGR*, which also includes the glucagon-like peptide 1, glucagon-like peptide 2, and glucose-dependent insulinotropic polypeptide receptors) families are located on VAC_E, except for the GCGR family (Hwang *et al*., 2013) (Fig. 1). Thus, in general, it can be postulated that extensive tandem local duplication within ancestral chromosomes, which occurred prior to 2R, drove the emergence of these gene families. The presence of members of a single clade on two or more chromosomes is likely due to chromosome translocation before 2R. The Nakatani model only rebuilds putative VACs dated shortly before 2R, therefore this model does not account for prior translocation. Chromosomal translocations between the 1st and 2nd WGDs may also account for the spread of gene family members from a single VAC onto multiple distinct GACs.

When performing evolutionary comparative analysis, either syntenic or phylogenetic, it is important to ensure specific types of representative species are included. Within a lineage, some species will have particularly well-conserved genomes that have undergone lower rates of chromosomal change and retain a greater variety of genes produced by 2R (Yun *et al*.,

2015). Those with the most conserved genomes/gene families/gene families/morphology are referred to as 'basal' species. Therefore, a variety of the most basal species from across the desired spectrum of taxa should be selected as representative species. Furthermore, for comparative evolutionary analysis, the genomes of these species must be available and to fully exploit their genomic data their DNA must be allocated to chromosomes and their genes be well annotated. Therefore, the following species provide good representatives: the human genome, which possesses unrivalled annotation and is better conserved than many other available mammal genomes (Burt *et al.*, 1999); the chicken genome, which has some of the best preserved chromosomes of the tetrapods (Nishida *et al.*, 2008); and spotted gar, which has recently been considered the best vertebrate representative as, unlike most teleost fish, it has not undergone a 3rd whole genome duplication and has undergone fewer translocations than other vertebrates with mapped genomes (Amores *et al.*, 2011). Furthermore, the inclusion of experimentally important species, such as mouse, zebrafish, and Xenopus species help to increase the usefulness of data and ensure a diverse selection of vertebrate gene samples.

Unfortunately, the genomes of a number of important species such as coelacanth, a basal tetrapod (Amemiya *et al.*, 2010); elephant shark, the slowest known evolving vertebrate (Venkatesh *et al.*, 2014); and *Branchiostoma floridae*, a basal chordate and a useful outgroup for vertebrate analysis (Elphick and Mirabeau, 2014), have not yet been arranged into chromosomes and instead are available as short DNA sequences on scaffolds which diminishes their use for synteny analysis. Other species such as lamprey or hagfish, which could provide important perspectives on inter-2R genome evolution (Caputo Barucchi *et al.*, 2013; Mehta *et al.*, 2013), and *Asymmetron lucayanum*, which may be the most basal chordate discovered (Yue *et al.*, 2014), have only partial genomes available. The arrangement of genomic data onto chromosomes and annotation of the genes of these species would provide a large boon to vertebrate evolutionary research.

In addition to using evolutionarily divergent species simply to put human gene families into perspective, analysing species with particularly interesting physiological attributes could help to design novel therapeutic treatments. For example, the naked mole rat shows incredible longevity, resistance to mammalian age-related disease, and cancer (Lewis *et al.*, 2016), and the elephant shark possesses an adaptive immune system that lacks several constituent genes that are vital to the mammalian immune system, but are capable of mounting an immune response (Venkatesh *et al.*, 2014). Further exploring how gene families have evolved in these species could bring novel insights into treating human medical issues.

DISCOVERY OF NOVEL PEPTIDE GENES AND THEIR EVOLUTIONARY DEVELOPMENT

The use of evolutionary conserved sequences and motifs to bulk analyse genomes is an established approach to novel gene discovery and genome annotation (Mirabeau *et al.*, 2007). However, many species, including humans, have lost various ohnologs which may be retained in more basal species (Yun *et al.*, 2015) (Table 1). For instance, in human there is a single kisspeptin (*KISS*) gene and a single *KISS*
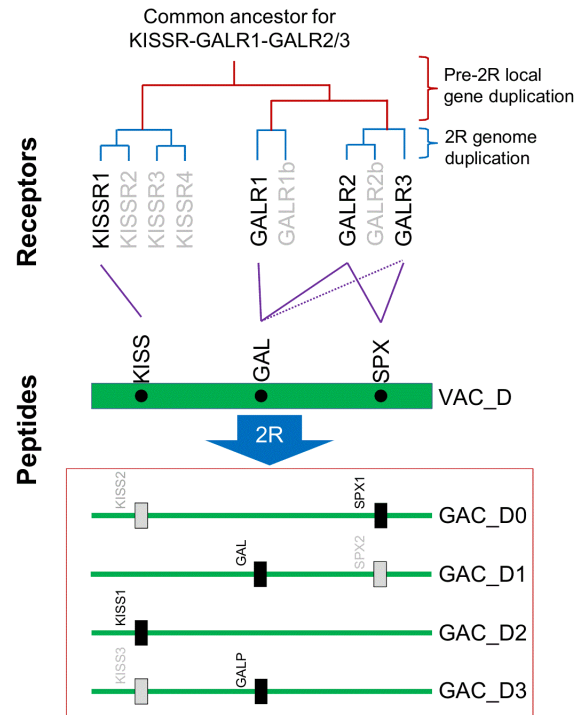

Fig. 2. A simple phylogenetic tree showing the development of the KISS and GAL branch of clade 2 via local duplication prior to 2R (red) and the ohnologs generated via 2R (blue). Genes present in humans are labelled in black, while those discovered in other vertebrate species are labelled in grey. Purple lines connect these 7TMRs to their ligand progenitors, with the dotted purple line between GAL and GALR3 indicating very low interaction. The ligand progenitors are placed on VAC_D and the red box below indicates how 2R expanded these peptide gene families from three progenitor genes into seven modern vertebrate genes, four of which are present in humans.

7TMR gene (*KISSR*), while spotted gar has three KISS genes and four *KISSR* genes (Lee *et al.*, 2009; Yun *et al.*, 2015). In human there are two *GnRH* genes and one *GnRHR* gene but coelacanth has three *GnRH* genes and four *GnRHR* genes. The loss of multiple members of a family may leave the resulting members too divergent for analyses using conserved motifs to identify each other without related genes to span the evolutionary divide. Therefore, performing blast searches using the full repertoire of protein sequences from a gene family, especially if selected from more basal species, gives a higher likelihood of success. However, basal species are not guaranteed to possess more genes in every family. For instance, humans have a single *NPS* gene and a single *NPSR* gene while neither spotted gar nor coelacanth appear to have genes from this family (Yun *et al.*, 2015).

In combination with phylogenetic analysis, small scale synteny (Yun *et al.*, 2015) and VAC/GAC models (Nakatani *et al.*, 2007), newly discovered genes can be correctly identified as orthologs (the same gene in different species), ohnologs (duplicates produced by WGD), or other paralogs (related genes created by other duplication events). For example, as displayed in Fig. 2, the single human *KISS* gene is located on a GAC_D2 linkage group while spotted gar possess an ad-

ditional two *KISS* genes on GAC_D0 and D3. Because these two genes are located on different GACs we can hypothesise they are separate ohnologs created by 2R from a single progenitor and not created by local duplication previously or subsequently. The use of synteny is particularly helpful when analysing peptide gene families because of the limited use of phylogeny to determine the exact relationships among peptide gene families. Therefore, it can be postulated that, because of the lower rates of change in species such as spotted gar, coelacanth and elephant shark, using gene identification algorithms in these basal species may return genes that have diverged too far to be detected in humans, and that by using VAC/GAC models, the regions of the genome that are most likely to harbour novel genes can be prioritised.

The evolutionary relationships among gene families can be examined by phylogenetic analysis. However, in the case of neuropeptide genes, signal peptide sequences are not conserved, and propeptide sequences, other than the mature peptide, are highly variable because these sequences are free from evolutionary conservational pressure (Lee *et al*., 2009). Sequence comparison of the short, conserved mature peptides is often not sufficient to extrapolate reliable relational information, particularly if they emerged prior to 2R (Cardoso *et al*., 2010; Hwang *et al*., 2013). Duplications that have occurred more recently, especially those subsequent to 2R, are more likely to be found in the same linkage block and share a high degree of amino acid sequence similarity (Yun *et al*., 2015). However, genes that emerged earlier, in pre-vertebrate evolution, or those that undergo particularly high rates of change accumulate mutations which lead to greater deviation in residue sequence and function. In contrast, 7TMR 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. Concerning the concept of co-evolution of peptides and their receptor genes, the evolutionary relationships among peptides can be extrapolated by matching them against their cognate 7TMR families. For instance, amino acid sequences of even the mature neuropeptides such as NPYR, PRLHR, PROKR, TACR, NPFFR, HCRTR, CCKR, and QRFPR families that are phylogenetically grouped in clade 1 (Fig. 1) are highly deviated such that phylogenetic analysis cannot be performed. However, when the genes for these neuropeptides are placed on VAC/GACs, 6 of the 8 neuropeptide gene families are found to be located on VAC_E (Yun *et al*., 2015). Similarly, the *AVP/OT*, *NPS*, and *GnRH* neuropeptide gene families of clade 5 are on VAC_C. These results indicate that neuropeptide families also multiplied through local duplications prior to 2R by the same pattern as their cognate 7TMRs.

COEVOLUTION OF NEUROPEPTIDES AND THEIR RECEPTOR GENES

Every known ligand gene for the related *GNRHR*, *NPSR*, and *AVPR* groups of 7TMRs found in clade 5 (Yun *et al*., 2015) can be found on a GAC_C linkage group (Fig. 1). Therefore, the ligands for the orphan GPR150 and GPR19 7TMRs, that are also present in clade five, can be postulated to also be present on a GAC_C linkage group. Furthermore, because the closest relative of *GPR150* is *NPSR*, then the ligand for GPR150 can be postulated to share similarity with the NPSR

ligand, NPS. By using these methods Kim *et al.* (2014a) were able to determine the receptor for the novel neuropeptide spexin (SPX) (Mirabeau *et al*., 2007). Syntenic analysis and relocating *SPX* genes and neighbouring genes on reconstructed VACs reveals that *SPXs* are located in the vicinity of *KISS* and galanin (*GAL*) family genes, suggesting that *SPX*, *GAL*, and *KISS* genes arose from a common ancestor through local duplications before 2R and that SPX may interact with receptors exhibiting similarity in amino acid sequence with those of GAL 7TMRs (GALRs) and KISSRs. KISS and GAL 7TMRs are phylogenetically closest among rhodopsin-like G protein-coupled receptors, and synteny revealed the presence of 3 distinct receptor progenitors *KISSR*, *GALR1*, and *GALR2/3* before 2R (Fig. 2). A ligand-receptor interaction study showed that SPX activates human, Xenopus, and zebrafish GALR2/3 but not GALR1, suggesting that SPX is a natural ligand for GALR2/3 (Kim *et al*., 2014a). Furthermore, linkage group analysis of the secretin neuropeptide and 7TMR family genes aided in the identification of a receptor for a novel GCRP neuropeptide (Hwang *et al*., 2013; Park *et al*., 2013).

Gradual duplication, differentiation, subfunctionalisation, and neofunctionalisation form the basis of the model of slow, consistent genome evolution where largely self-contained gene families are inherited from parents and passed down to offspring in a vertical pattern of inheritance. This allows genes to be placed into related families and their relationships be traced through evolutionary history, sometimes over a billion years (Nordstrom *et al*., 2011). However, occasionally sudden changes in gene interaction or lateral gene transfer defy this trend. For example, the pro-opiomelanocortin (*POMC*) gene possesses two evolutionary distinct subunits: the *ACTH* subunit and the opioid subunit, each of which is a ligand for two completely unrelated 7TMR families. It appears that the *ACTH* subunit existed as an individual gene in the vertebrate ancestor and, by chance, it's peptide product started to interact with the progenitor of the melanocortin family of 7TMRs (MCR) (Haitina *et al*., 2007). The *MCR* family progenitors emerged from the wider MECA group of receptors, none of which interact with peptide ligands, which means the *MCR* family is an evolutionary anomaly (Fredriksson *et al*., 2003). Subsequent to 2R, it appears that a duplication of an opioid gene, prepronociceptin (*PNOC*), placed an opioid coding region into the *ACTH* gene, resulting in a novel hybrid gene, *POMC*. Furthermore, intra-gene duplication of the *ACTH* subunit and proliferation of the *MCR* family has led to the development of an entirely novel system that has only been found in vertebrates (Harris *et al*., 2014). Lateral gene transfer can also produce sudden changes that do not fit into the standard evolutionary model. For example, it has been noted that the syncytin gene, which is critical for placental development in mammals (Dupressoir *et al*., 2009), in distantly related mammals sometimes have similar syncytin peptide sequences while closely related species sometimes have largely divergent syncetin peptide sequences (Redelsperger *et al*., 2014). It appears that mammalian syncytin genes originate from a viral gene and that periodically, in a pattern that does not follow standard evolutionary patterns, is updated via reinfection and gene adopted into evolving mammalian genomes (Cornelis *et al*., 2014).
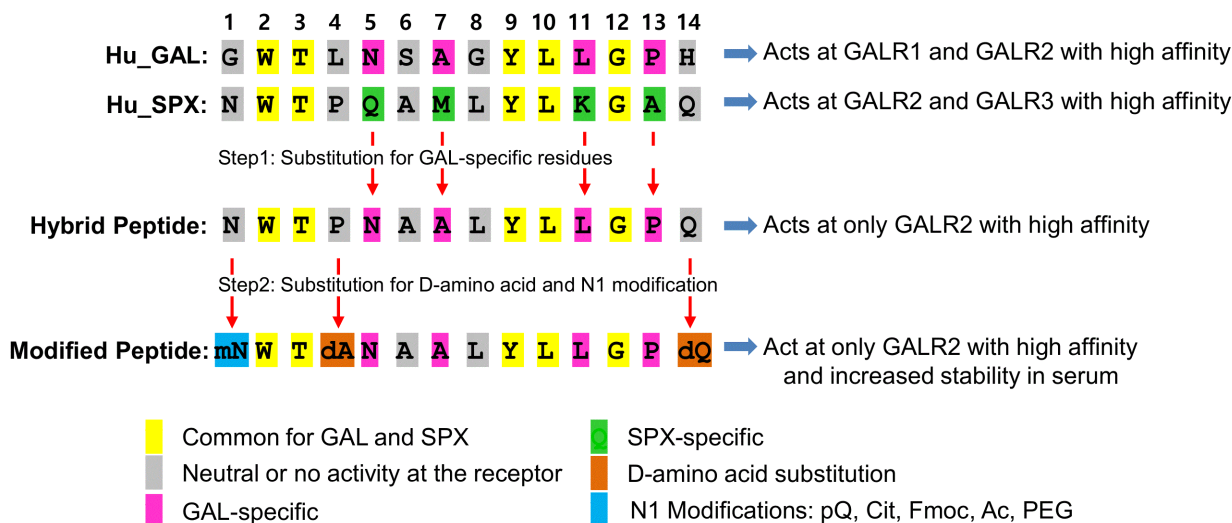
Fig. 3. Mature peptides of human GAL and SPX, below which is the hybrid peptide created via mutagenesis to specifically target GALR2, below which is the modified hybrid peptide with increased serum stability. Each residue is colour coded to indicate residues that are common to both GAL and SPX peptides (yellow), divergent residues that don't appear to alter receptor activity (grey), GAL specific residues (pink), and SPX specific residues (green). To provide serum stability, some residues were replaced with their D-amino isomer (orange) or N-terminal modifications were made (blue).

DRUG DESIGN USING EVOLUTIONARY COMPARATIVE ANALYSIS

Subsequent to gene duplication, subfunctionalisation and specialisation can result in related ligands with similar amino acid sequences which have greater or lesser ability to interact with 7TMR subtypes within a gene family (Kim *et al*., 2014a). The amino acid sequence of a peptide ligand defines the structural and chemical nature of that peptide, including the interactions that the peptide undergoes. Amino acids that are critical for protein function tend to be evolutionarily retained while those of decreasing importance will, on average, have increasing rates of variation. As the function of related peptides deviate further then variation even within the best conserved amino acids increases. As noted by Yun *et al*. (2015), humans often possess fewer, and different, paralogs within a gene family compared to other species. Therefore, by using a variety of paralogs from within a gene family from an array of species, different peptide sequences that are capable of binding human 7TMRs of interest at varying potencies and affinities can be analysed (Kim *et al*., 2014a) and the function of amino acids in receptor binding ascertained (Reyes-Alcaraz *et al*., 2016). Once the functions of individual amino acids within a peptide have been analysed then mutational experiments can be conducted to specifically alter the binding affinity of the peptide sequences to create novel ligands. Furthermore, alteration of the nature of the amino acids used in peptide synthesis can reduce susceptibility to proteases.

The *SPX* and *GAL* neuropeptide genes emerged through a local duplication from a common ancestor gene and both interact with members of the GAL 7TMR family (GALR1, 2, and 3). GAL can activate GALR1 and 2 to a high degree with a much lower ability to activate GALR3. SPX can activate GALR2 and 3 to a high degree. The mature neuropeptides of SPX and GAL share several conserved residues, includ-

ing Trp2, Thr3, Tyr9, Leu10, and Gly12 (Kim *et al*., 2014a) (Fig. 3). This indicates that these common residues may be required for activation of GALR2, the SPX-specific residues are likely involved in retaining the agonist activity toward GALR3 while GAL-specific residues may contribute to decreased affinity to GALR3. This observation can lead to a postulation that the replacement of SPX-specific residues with those of GAL can produce a novel agonist acting only on GALR2 with no cross reactivity with GALR3. Indeed, out of SPX-specific residues, Gln5, Met7, Lys11, and Ala13 were found to be critical for GALR3 activation. Replacement of these residues with Gal-specific residues (Gln5→Asn, Met7→Ala, Lys11→Phe, and Ala13→Pro) abolished the ability to activate GALR3 while retaining full activity to GALR2. This mutation study takes into account the evolutionary fates of duplicated neuropeptide ligand and receptor genes. The pre-2R local duplication followed by whole-genome duplication produced *GALR1*, *GALR2* and *GALR3*. Likewise, pre-2R local duplication produced *GAL* and *SPX* progenitors and following whole-genome duplication, generated the *GAL* family (*GAL* and *GALP*) and *SPX* family (*SPX1* and *SPX2*) (Fig. 2). During the divergence of the GAL/SPX and GALR1/2/3 system, GALR2 appears to have become an intermediate form as it responds to both SPX and GAL with high affinity, whereas GALR1 and GALR3 acquired significant preference to GAL and SPX, respectively (Reyes-Alcaraz *et al*., 2016).

Based on this concept, Reyes-Alcaraz *et al*. (2016) synthesised novel agonists that were capable of targeting and specifically activating GALR2. Furthermore, N-terminal modification and substitution of residues that were not found to alter GALR activity with D-isoforms of these residues greatly increased the stability of the peptide in serum. The endogenous ligands, SPX and GAL, and other synthetic ligands, such as M1145 and M1153, had cross-reactivity with multiple GALRs to some extent. The ability to activate GALR2, specifically, had inter-

esting therapeutic possibilities as each subtype of GALRs has been found to exhibit largely divergent physiological function. For instance, a recent observation demonstrates that GALR1 and GALR2 mediate opposite anxiety-like effects in rats: GALR1 and GALR2 agonists exerted anxiogenic and anxiolytic-like effects, respectively (Morais *et al*., 2016). GALR3 may also induce anxiogenic behaviour as GALR3-specific antagonists decrease anxiety and induce depression-like behavior in rats (Swanson *et al*., 2005). In addition, the actions of SPX and GAL in appetite behaviour appear to oppose each other as well: SPX is anorexic while GAL is orexigenic (Taylor *et al*., 2009; Shiba *et al*., 2010; Wong *et al*., 2013). Thus, the design of an agonist that discriminates GALR2 from GALR1 or GALR3 is of particular importance from therapeutic perspective.

Conversely, Moon *et al*. (2010) compared the amino acid sequences of two related secretin-like 7TMR interacting neuropeptides; glucagon like peptide (GLP-1), which binds GLP1R, and glucose-dependent insulinotropic polypeptide (GIP), which binds GIPR. These neuropeptides and 7TMRs share high similarity in amino acid sequence and both modulate insulin secretion from pancreatic B-cells, among other functions (Baggio and Drucker, 2007). However, despite their similarities they have no ability to activate each other's receptor. Moon *et al*. (2010) compared the two amino acid sequences and generated a hybrid that replaced four residues in GIP with those of GLP-1. This allowed the mutant peptide to activate both receptors with moderate potency, which allows two related but distinct messenger pathways to be activated with a single ligand. Subsequently, further research was conducted to increase the half-life and potency of GIP/GLP-1 hybrids (Moon *et al*., 2010; Kim *et al*., 2014b).

In addition, the use of comparative techniques where related proteins are compared to analyse the function of individual amino acids can be used to highlight amino acids in receptors that are highly conserved among vertebrate species. In 7TMRs, the intra- and extra-cellular loops as well as the N- and C-terminals of the receptors tend to deviate at a relatively high rate. Therefore, amino acid sequences in these regions that are well conserved among vertebrate species likely have a function in receptor activity or stability. By comparing GLP-1 receptor sequences and performing mutagenic studies, Moon *et al*. (2015) were able to find, by using point mutations of evolutionary conserved residues, that amino acid residue Arg[380] flanked by Leu[379] and Phe[381] in extra-cellular loop 3 may interact with Asp[9] and Gly[4] of the GLP-1 neuropeptide. This information helps to bring greater understanding to the mechanisms by which GLP-1 interacts with the GLP1R, which at the moment is not well known due to a lack of crystal structure data for the ligand-bound receptor complex.

CONCLUSIONS

Currently, a number of 'orphan' 7TMRs are predicted to bind peptidergic ligands, which may be expressed from currently undiscovered genes or known genes where functional relationship with orphan 7TMRs have not yet been identified. In these situations, the use of VAC/GAC maps and identification of peptide-receptor systems in more basal species, which have not diverged as greatly, may help to identify the ligand genes in humans. However, it has been postulated that some

7TMRs remain orphans, despite intensive efforts to identify endogenous ligands, because they simply do not have ligands (Davenport *et al*., 2013). Instead, some 7TMRs may influence signal transduction through dimerisation and modulation of other 7TMRs (Levoye *et al*., 2006). It is also possible that, as 7TMRs have constitutive activity rates irrespective of ligand binding, other mechanisms may be the primary mediator of some orphan 7TMR activity, such as pH, pressure, or temperature (Ahmad *et al*., 2015).

Prompted by the wide variety of important biological processes mediated by 7TMR signalling, there is a high demand for novel drugs that can target individual receptors and regulate these signalling pathways. However, high clinical standards with regards to new drugs being authorised for sale on the market combined with the high cost of drug development means that there is a strong interest in techniques that facilitate the design of drug candidates which can specifically target individual 7TMR mediated pathways. The discussed techniques and examples demonstrate how comparison of genomes, gene families, and individual protein sequences, using phylogeny and synteny, can aid in gene discovery, gene function identification, and the design of hybrid ligands.

ACKNOWLEDGMENTS

REFERENCES

Abi-Rached, L., Gilles, A., Shiina, T., Pontarotti, P. and Inoko, H. (2002) Evidence of en bloc duplication in vertebrate genomes. *Nat. Genet.* **31**, 100-105

Ahmad, R., Wojciech, S. and Jockers, R. (2015) Hunting for the function of orphan GPCRs - beyond the search for the endogenous ligand. *Br. J. Pharmacol.* **172**, 3212-3228.

Amemiya, C. T., Powers, T. P., Prohaska, S. J., Grimwood, J., Schmutz, J., Dickson, M., Miyake, T., Schoenborn, M. A., Myers, R. M., Ruddle, F. H. and Stadler, P. F. (2010) Complete HOX cluster characterization of the coelacanth provides further evidence for slow evolution of its genome. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 3622-3627.

Amores, A., Catchen, J., Ferrara, A., Fontenot, Q. and Postlethwait, J. H. (2011) Genome evolution and meiotic maps by massively parallel DNA sequencing: spotted gar, an outgroup for the teleost genome duplication. *Genetics* **188**, 799-808.

Assis, R. and Bachtrog, D. (2013) Neofunctionalization of young duplicate genes in Drosophila. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 17409-17414.

Baggio, L. L. and Drucker, D. J. (2007) Biology of incretins: GLP-1 and GIP. *Gastroenterology* **132**, 2131-2157.

Burt, D. W., Bruley, C., Dunn, I. C., Jones, C. T., Ramage, A., Law, A. S., Morrice, D. R., Paton, I. R., Smith, J., Windsor, D., Sazanov, A., Fries, R. and Waddington, D. (1999) The dynamics of chromosome evolution in birds and mammals. *Nature* **402**, 411-413.

Caputo Barucchi, V., Giovannotti, M., Nisi Cerioni, P. and Splendiani, A. (2013) Genome duplication in early vertebrates: insights from agnathan cytogenetics. *Cytogenet. Genome Res.* **141**, 80-90.

Cardoso, J. C., Vieira, F. A., Gomes, A. S. and Power, D. M. (2010) The serendipitous origin of chordate secretin peptide family members. *BMC Evol. Biol.* **10**, 135.

Cerda-Reverter, J. M., Martinez-Rodriguez, G., Zanuy, S., Carrillo, M. and Larhammar, D. (2000) Molecular evolution of the neuropeptide Y (NPY) family of peptides: cloning of three NPY-related peptides from the sea bass (Dicentrarchus labrax). *Regul. Pept.* **95**, 25-34.

Cho, H. J., Acharjee, S., Moon, M. J., Oh, D. Y., Vaudry, H., Kwon, H. B. and Seong, J. Y. (2007) Molecular evolution of neuropeptide receptors with regard to maintaining high affinity to their authentic ligands. *Gen. Comp. Endocrinol.* **153**, 98-107.

Civelli, O., Saito, Y., Wang, Z., Nothacker, H. P. and Reinscheid, R. K. (2006) Orphan GPCRs and their ligands. *Pharmacol. Ther.* **110**, 525-532.

Conroy, J. L., Free, R. B. and Sibley, D. R. (2015) Identification of G protein-biased agonists that fail to recruit β-arrestin or promote internalization of the D1 dopamine receptor. *ACS Chem. Neurosci.* **6**, 681-692.

Cornelis, G., Vernochet, C., Malicorne, S., Souquere, S., Tzika, A. C., Goodman, S. M., Catzeflis, F., Robinson, T. J., Milinkovitch, M. C., Pierron, G., Heidmann, O., Dupressoir, A. and Heidmann, T. (2014) Retroviral envelope syncytin capture in an ancestrally diverged mammalian clade for placentation in the primitive Afrotherian tenrecs. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E4332-E4341.

Davenport, A. P., Alexander, S. P., Sharman, J. L., Pawson, A. J., Benson, H. E., Monaghan, A. E., Liew, W. C., Mpamhanga, C. P., Bonner, T. I., Neubig, R. R., Pin, J. P., Spedding. M. and Harmar, A. J. (2013) International union of basic and clinical pharmacology. LXXXVIII. G protein-coupled receptor list: recommendations for new pairings with cognate ligands. *Pharmacol. Rev.* **65**, 967-986.

Dehal, P. and Boore, J. L. (2005) Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol.* **3**, e314.

Dores, R. M. (2013) Observations on the evolution of the melanocortin receptor gene family: distinctive features of the melanocortin-2 receptor. *Front. Neurosci.* **7**, 28.

Dupressoir, A., Vernochet, C., Bawa, O., Harper, F., Pierron, G., Opolon, P. and Heidmann, T. (2009) Syncytin-A knockout mice demonstrate the critical role in placentation of a fusogenic, endogenous retrovirus-derived, envelope gene. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 12127-12132.

Elphick, M. R. and Mirabeau, O. (2014) The evolution and variety of RFamide-type neuropeptides: insights from deuterostomian invertebrates. *Front. Endocrinol. (Lausanne)* **5**, 93.

Fredriksson, R., Lagerstrom, M. C., Lundin, L. G. and Schioth, H. B. (2003) The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol. Pharmacol.* **63**, 1256-1272.

Gibson, T. A. and Goldberg, D. S. (2009) Questioning the ubiquity of neofunctionalization. *PLoS Comput. Biol.* **5**, e1000252.

Haitina, T., Takahashi, A., Holmen, L., Enberg, J. and Schioth, H. B. (2007) Further evidence for ancient role of ACTH peptides at melanocortin (MC) receptors; pharmacology of dogfish and lamprey peptides at dogfish MC receptors. *Peptides* **28**, 798-805.

Harris, R. M., Dijkstra, P. D. and Hofmann, H. A. (2014) Complex structural and regulatory evolution of the pro-opiomelanocortin gene family. *Gen. Comp. Endocrinol.* **195**, 107-115.

Hauser, F., Cazzamali, G., Williamson, M., Blenau, W. and Grimmelikhuijzen, C. J. (2006) A review of neurohormone GPCRs present in the fruitfly Drosophila melanogaster and the honey bee Apis mellifera. *Prog. Neurobiol.* **80**, 1-19.

He, X. and Zhang, J. (2005) Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics* **169**, 1157-1164.

Holland, P. W., Garcia-Fernandez, J., Williams, N. A. and Sidow, A. (1994) Gene duplications and the origins of vertebrate development. *Dev. Suppl.* 125-133.

Holland, P. W. (2003) More genes in vertebrates? *J. Struct. Funct. Genomics* **3**, 75-84.

Hwang, J. I., Moon, M. J., Park, S., Kim, D. K., Cho, E. B., Ha, N., Son, G. H., Kim, K., Vaudry, H. and Seong, J. Y. (2013) Expansion of secretin-like G protein-coupled receptors and their peptide ligands via local duplications before and after two rounds of whole-genome duplication. *Mol. Biol. Evol.* **30**, 1119-1130.

Hwang, J. K., Yun, S., Moon M. J., Park C. R. and Seong, J. Y. (2014) Molecular evolution of GPCRs: GLP1/GLP1 receptors. *J. Mol. Endocrinol.* **52**, T15-T27.

Jensen, J. D. and Bachtrog, D. (2011) Characterizing the influence of effective population size on the rate of adaptation: Gillespie's Darwin domain. *Genome Biol. Evol.* **3**, 687-701.

Jiang, H., Lkhagva, A., Daubnerová, I., Chae, H. S., Šimo, L., Jung, S. H., Yoon, Y. K., Lee, N. R., Seong, J. Y., Žitňan, D., Park, Y. and Kim, Y. J. (2013) Natalisin, a tachykinin-like signaling system, regulates sexual activity and fecundity in insects. *Proc. Natl. Acad. Sci. U.S.A.* **110**, E3526-E3534.

Katritch, V., Cherezov, V. and Stevens, R. C. (2012) Diversity and modularity of G protein-coupled receptor structures. *Trends Pharmacol. Sci.* **33**, 17-27.

Kim, D. K., Cho, E. B., Moon, M. J., Park, S., Hwang, J. I., Kah, O., Sower, S. A., Vaudry, H. and Seong, J. Y. (2011) Revisiting the evolution of gonadotropin-releasing hormones and their receptors in vertebrates: secrets hidden in genomes. *Gen. Comp. Endocrinol.* **170**, 68-78.

Kim, D. K., Cho, E. B., Moon, M. J., Park, S., Hwang, J. I., Do Rego, J. L., Vaudry, H. and Seong, J. Y. (2012) Molecular coevolution of neuropeptides gonadotropin-releasing hormone and kisspeptin with their cognate G protein-coupled receptors. *Front. Neurosci.* **6**, 3.

Kim, D. K., Yun, S., Son, G. H., Hwang, J. I., Park, C. R., Kim, J. I., Kim, K., Vaudry, H. and Seong, J. Y. (2014a) Coevolution of the spexin/galanin/kisspeptin family: Spexin activates galanin receptor type II and III. *Endocrinology* **155**, 1864-1873.

Kim, H. Y., Hwang, J. I., Moon, M. J. and Seong, J. Y. (2014b) A novel long-acting glucagon-like peptide-1 agonist with improved efficacy in insulin secretion and β-cell growth. *Endocrinol. Metab. (Seoul)* **29**, 320-327.

Klingel, S., Morath, I., Strietz, J., Menzel, K., Holstein, T. W. and Gradl, D. (2012) Subfunctionalization and neofunctionalization of vertebrate Lef/Tcf transcription factors. *Dev. Biol.* **368**, 44-53.

Lagerstrom, M. C., Fredriksson, R., Bjarnadóttir, T. K., Fridmanis, D., Holmquist, T., Andersson, J., Yan, Y. L., Raudsepp, T., Zoorob, R., Kukkonen, J. P., Lundin, L. G., Klovins, J., Chowdhary, B. P., Postlethwait, J. H. and Schiöth, H. B. (2005) Origin of the prolactin-releasing hormone (PRLH) receptors: evidence of coevolution between PRLH and a redundant neuropeptide Y receptor during vertebrate evolution. *Genomics* **85**, 688-703.

Lagerstrom, M. C. and Schioth, H. B. (2008) Structural diversity of G protein-coupled receptors and significance for drug discovery. *Nat. Rev. Drug Discov.* **7**, 339-357.

Larhammar, D., Lundin, L. G. and Hallbook, F. (2002) The human Hox-bearing chromosome regions did arise by block or chromosome (or even genome) duplications. *Genome Res.* **12**, 1910-1920.

Larhammar, D. and Salaneck, E. (2004) Molecular evolution of NPY receptor subtypes. *Neuropeptides* **38**, 141-151.

Lee, Y. R., Tsunekawa, K., Moon, M. J., Um, H. N., Hwang, J. I., Osugi, T., Otaki, N., Sunakawa, Y., Kim, K., Vaudry, H., Kwon, H. B., Seong, J. Y. and Tsutsui, K. (2009) Molecular evolution of multiple forms of kisspeptins and GPR54 receptors in vertebrates. *Endocrinology* **150**, 2837-2846.

Levoye, A., Dam, J., Ayoub, M. A., Guillaume, J. L., Couturier, C., Delagrange, P. and Jockers, R. (2006) The orphan GPR50 receptor specifically inhibits MT1 melatonin receptor function through heterodimerization. *EMBO J.* **25**, 3012-3023.

Lewis, K. N., Soifer, I., Melamud, E., Roy, M., McIsaac, R. S., Hibbs, M. and Buffenstein, R. (2016) Unraveling the message: insights into comparative genomics of the naked mole-rat. *Mamm. Genome* **27**, 259-278.

Lindemans, M., Liu, F., Janssen, T., Husson, S. J., Mertens, I., Gade, G. and Schoofs, L. (2009) Adipokinetic hormone signaling through the gonadotropin-releasing hormone receptor modulates egg-laying in Caenorhabditis elegans. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 1642-1647.

Lundin, L. G. (1993) Evolution of the vertebrate genome as reflected in paralogous chromosomal regions in man and the house mouse. *Genomics* **16**, 1-19.

Lynch, M., O'Hely, M., Walsh, B. and Force, A. (2001) The probability of preservation of a newly arisen gene duplicate. *Genetics* **159**, 1789-1804.

Mehta, T. K., Ravi, V., Yamasaki, S., Lee, A. P., Lian, M. M., Tay, B. H., Tohari, S., Yanai, S., Tay, A., Brenner, S. and Venkatesh, B. (2013) Evidence for at least six hox clusters in the Japanese lamprey (Lethenteron japonicum). *Proc. Natl. Acad. Sci. U.S.A.* **110**, 16044-16049.

Meyer, A. and Van de Peer, Y. (2005) From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). *Bioessays* **27**, 937-945.

Mirabeau, O., Perlas, E., Severini, C., Audero, E., Gascuel, O., Possenti, R., Birney, E., Rosenthal, N. and Gross, C. (2007) Identification of novel peptide hormones in the human proteome by hidden Markov model screening. *Genome Res.* **17**, 320-327.

Mirabeau, O. and Joly, J. S. (2013) Molecular evolution of peptidergic signaling systems in bilaterians. *Proc. Natl. Acad. Sci. U.S.A.* **110**, E2028-E2037.

Moon, M. J., Kim, H. Y., Kim, S. G., Park, J., Choi, D. S., Hwang, J. I. and Seong, J. Y. (2010) Tyr1 and Ile7 of glucose-dependent insulinotropic polypeptide (GIP) confer differential ligand selectivity toward GIP and glucagon-like peptide-1 receptors. *Mol. Cells* **30**, 149-154.

Moon, M. J., Lee, Y. N., Park, S., Reyes-Alcaraz, A., Hwang, J. I., Millar, R. P., Choe, H. and Seong, J. Y. (2015) Ligand binding pocket formed by evolutionarily conserved residues in the glucagon-like peptide-1 (GLP-1) receptor core domain. *J. Biol. Chem.* **290**, 5696-5706.

Morais, J. S., Souza, M. M., Campanha, T. M., Muller, C. J., Bittencourt, A. S., Bortoli, V. C., Schenberg, L. C. and Beijamini, V. (2016) Galanin subtype 1 and subtype 2 receptors mediate opposite anxiety-like effects in the rat dorsal raphe nucleus. *Behav. Brain Res.* **314**, 125-133.

M'Kadmi, C., Leyris, J. P., Onfroy, L., Galés, C., Saulière, A., Gagne, D., Damian, M., Mary, S., Maingot, M., Denoyelle, S., Verdié, P., Fehrentz, J. A., Martinez, J., Banères, J. L. and Marie, J. (2015) Agonism, antagonism, and inverse agonism bias at the ghrelin receptor signaling. *J. Biol. Chem.* **290**, 27021-27039.

Nakatani, Y., Takeda, H., Kohara, Y. and Morishita, S. (2007) Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. *Genome Res.* **17**, 1254-1265.

Nishida, C., Ishijima, J., Kosaka, A., Tanabe, H., Habermann, F. A., Griffin, D. K. and Matsuda, Y. (2008) Characterization of chromosome structures of Falconinae (Falconidae, Falconiformes, Aves) by chromosome painting and delineation of chromosome rearrangements during their differentiation. *Chromosome Res.* **16**, 171-181.

Nordstrom, K. J., Sallman Almen, M., Edstam, M. M., Fredriksson, R. and Schioth, H. B. (2011) Independent HHsearch, Needleman--Wunsch-based, and motif analyses reveal the overall hierarchy for most of the G protein-coupled receptor families. *Mol. Biol. Evol.* **28**, 2471-2480.

Oh, D. Y., Kim, K., Kwon, H. B. and Seong, J. Y. (2006) Cellular and molecular biology of orphan G protein-coupled receptors. *Int. Rev. Cytol.* **252**, 163-218.

Osugi, T., Ubuka, T. and Tsutsui, K. (2014) Review: evolution of GnIH and related peptides structure and function in the chordates. *Front. Neurosci.* **8**, 255.

Park, C. R., Moon, M. J., Park, S., Kim, D. K., Cho, E. B., Millar, R. P., Hwang, J. I. and Seong, J. Y. (2013) A novel glucagon-related peptide (GCRP) and its receptor GCRPR account for coevolution of their family members in vertebrates. *PLoS ONE* **8**, e65420.

Putnam, N. H., Butts, T., Ferrier, D. E., Furlong, R. F., Hellsten, U., Kawashima, T., Robinson-Rechavi, M., Shoguchi, E., Terry, A., Yu, J. K., Benito-Gutiérrez, E. L., Dubchak, I., Garcia-Fernández, J., Gibson-Brown, J. J., Grigoriev, I. V., Horton, A. C., de Jong, P. J., Jurka, J., Kapitonov, V. V., Kohara, Y., Kuroki, Y., Lindquist, E., Lucas, S., Osoegawa, K., Pennacchio, L. A., Salamov, A. A., Satou, Y., Sauka-Spengler, T., Schmutz, J., Shin-I, T., Toyoda, A., Bronner-Fraser, M., Fujiyama, A., Holland, L. Z., Holland, P. W., Satoh, N. and Rokhsar, D. S. (2008) The amphioxus genome and the evolution of the chordate karyotype. *Nature* **453**, 1064-1071.

Rask-Andersen, M., Almen, M. S. and Schioth, H. B. (2011) Trends in the exploitation of novel drug targets. *Nat. Rev. Drug Discov.* **10**, 579-590.

Rastogi, S. and Liberles, D. A. (2005) Subfunctionalization of duplicated genes as a transition state to neofunctionalization. *BMC Evol. Biol.* **5**, 28.

Redelsperger, F., Cornelis, G., Vernochet, C., Tennant, B. C., Catzeflis, F., Mulot, B., Heidmann, O., Heidmann, T. and Dupressoir, A. (2014) Capture of syncytin-Mar1, a fusogenic endogenous retroviral envelope gene involved in placentation in the Rodentia squirrel-related clade. *J. Virol.* **88**, 7915-7928.

Reyes-Alcaraz, A., Lee, Y. N., Son, G. H., Kim, N. H., Kim, D. K., Yun, S., Kim, D. H., Hwang, J. I. and Seong, J. Y. (2016) Development of spexin-based human galanin receptor type II-specific agonists with increased stability in serum and anxiolytic effect in mice. *Sci. Rep.* **6**, 21453.

Santini, S., Boore, J. L. and Meyer, A. (2003) Evolutionary conservation of regulatory elements in vertebrate Hox gene clusters. *Genome Res.* **13**, 1111-1122.

Sefideh, F. A., Moon, M. J., Yun, S., Hong, S. I., Hwang, J. I. and Seong, J. Y. (2014) Local duplication of gonadotropin-releasing hormone (GnRH) receptor before two rounds of whole genome duplication and origin of the mammalian GnRH receptor. *PLoS ONE* **9**, e87901.

Shiba, K., Kageyama, H., Takenoya, F. and Shioda, S. (2010) Galanin-like peptide and the regulation of feeding behavior and energy metabolism. *FEBS J.* **277**, 5006-5013.

Steiner, D. F. (1998) The proprotein convertases. *Curr. Opin. Chem. Biol.* **2**, 31-39.

Swanson, C. J., Blackburn, T. P., Zhang, X., Zheng, K., Xu, Z. Q., Hökfelt, T., Wolinsky, T. D., Konkel, M. J., Chen, H., Zhong, H., Walker, M. W., Craig, D. A., Gerald, C. P. and Branchek, T. A. (2005) Anxiolytic- and antidepressant-like profiles of the galanin-3 receptor (Gal3) antagonists SNAP 37889 and SNAP 398299. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 17489-17494.

Taylor, A., Madison, F. N. and Fraley, G. S. (2009) Galanin-like peptide stimulates feeding and sexual behavior via dopaminergic fibers within the medial preoptic area of adult male rats. *J. Chem. Neuroanat.* **37**, 105-111.

Ufuk, E. (2014) G-protein coupled receptor (GPCR) targeting: technologies and global markets. *BCC Res.* BIO136A.

Vaudry, H. and Seong, J. Y. (2014) Neuropeptide GPCRs in Neuroendocrinology. *Front. Endocrinol. (Lausanne)* **5**, 41.

Venkatesh, B., Lee, A. P., Ravi, V., Maurya, A. K., Lian, M. M., Swann, J. B., Ohta, Y., Flajnik, M. F., Sutoh, Y., Kasahara, M., Hoon, S., Gangu, V., Roy, S. W., Irimia, M., Korzh, V., Kondrychyn, I., Lim, Z. W., Tay, B. H., Tohari, S., Kong, K. W., Ho, S., Lorente-Galdos, B., Quilez, J., Marques-Bonet, T., Raney, B. J., Ingham, P. W., Tay, A., Hillier, L. W., Minx, P., Boehm, T., Wilson, R. K., Brenner, S. and Warren, W. C. (2014) Elephant shark genome provides unique insights into gnathostome evolution. *Nature* **505**, 174-179.

Vienne, A., Rasmussen, J., Abi-Rached, L., Pontarotti, P. and Gilles, A. (2003) Systematic phylogenomic evidence of en bloc duplication of the ancestral 8p11.21-8p21.3-like region. *Mol. Biol. Evol.* **20**, 1290-1298.

Wong, M. K., Sze, K. H., Chen, T., Cho, C. K., Law, H. C., Chu, I. K. and Wong, A. O. (2013) Goldfish spexin: solution structure and novel function as a satiety factor in feeding control. *Am. J. Physiol. Endocrinol. Metab.* **305**, E348-E366.

Yegorov, S. and Good, S. (2012) Using paleogenomics to study the evolution of gene families: origin and duplication history of the relaxin family hormones and their receptors. *PLoS ONE* **7**, e32923.

Yue, J. X., Yu, J. K., Putnam, N. H. and Holland, L. Z. (2014) The transcriptome of an amphioxus, Asymmetron lucayanum, from the Bahamas: a window into chordate evolution. *Genome Biol. Evol.* **6**, 2681-2696.

Yun, S., Furlong, M., Sim, M., Cho, M., Park, S., Cho, E. B., Reyes-Alcaraz, A., Hwang, J. I., Kim, J. and Seong, J. Y. (2015) Prevertebrate local gene duplication facilitated expansion of the neuropeptide GPCR superfamily. *Mol. Biol. Evol.* **32**, 2803-2817.