Corticotropic-releasing hormone family evolution: five ancestral genes remain in some lineages

João C R Cardoso¹, Christina A Bergqvist², Rute C Félix¹ and Dan Larhammar²

¹Comparative Endocrinology and Integrative Biology, Centre of Marine Sciences, Universidade do Algarve, Campus de Gambelas, Faro, Portugal
²Department of Neuroscience, Science for Life Laboratory, Uppsala University, Uppsala, Sweden

Abstract

The evolution of the peptide family consisting of corticotropic-releasing hormone (CRH) and the three urocortins (UCN1-3) has been puzzling due to uneven evolutionary rates. Distinct gene duplication scenarios have been proposed in relation to the two basal rounds of vertebrate genome doubling (2R) and the teleost fish-specific genome doubling (3R). By analyses of sequences and chromosomal regions, including many neighboring gene families, we show here that the vertebrate progenitor had two peptide genes that served as the founders of separate subfamilies. Then, 2R resulted in a total of five members: one subfamily consists of CRH1, CRH2, and UCN1. The other subfamily contains UCN2 and UCN3. All five peptide genes are present in the slowly evolving genomes of the coelacanth Latimeria chalumnae (a lobe-finned fish), the spotted gar Lepisosteus oculatus (a basal ray-finned fish), and the elephant shark Callorhinchus milii (a cartilaginous fish). The CRH2 gene has been lost independently in placental mammals and in teleost fish, but is present in birds (except chicken), anole lizard, and the nonplacental mammals platypus and opossum. Teleost 3R resulted in an additional surviving duplicate only for crh1 in some teleosts including zebrafish (crh1a and crh1b). We have previously reported that the two vertebrate CRH/UCN receptors arose in 2R and that CRHR1 was duplicated in 3R. Thus, we can now conclude that this peptide–receptor system was quite complex in the ancestor of the jawed vertebrates with five CRH/UCN peptides and two receptors, and that crh and crhr1 were duplicated in the teleost fish tetraploidization.

Key Words

- CRH/UCN
- phylogeny
- gene duplication
- synteny
- chromosome duplication

Introduction

The corticotropic-releasing hormone (CRH) family members are a group of neuropeptides that are best known for their role in the regulation of the physiological response to stress, emotional behavior, and anxiety in vertebrates (Vale et al. 1981, Dunn & Berridge 1990, Koob & Heinrichs 1999, Lovejoy & Balment 1999, Gysling et al. 2004, Fox & Lowry 2013). Four independent lineages of structurally related peptides have been described in vertebrates, CRH, and three urocortins, abbreviated UCN1, 2, and 3. Vale and coworkers were pioneers in the discovery and functional characterization of the mammalian CRH family members. CRH was isolated...
from sheep hypothalamus and consists of 41 amino acids (aa) in mammals (Vale et al. 1981). Subsequently, UCN1 (40 aa in mammals) was obtained from rat midbrain and was concluded to be the ortholog of fish urotensin I and amphibian sauvagine (Vaughan et al. 1995). Both UCN2 (38 aa in mammals) (Reyes et al. 2001) and UCN3 (38 aa in mammals) (Lewis et al. 2001) were retrieved by searching mammalian nucleotide databases.

Studies in mammals revealed that CRH is the major stimulator of adrenocorticotropic hormone (ACTH) secretion from the pituitary and plays a central role in organizing the hypothalamic–pituitary–adrenal/interrenal (HPA/I) axis. The widespread distribution of CRH family peptides triggered further studies that have illuminated their functions in other physiological processes including appetite suppression (Spina et al. 1996, Contarino et al. 2000), cardiovascular adaptation to stress (Coste et al. 2000), regulation of the inflammatory response (Jain et al. 1991), and reproduction in vertebrates (Imperatore et al. 2006, Tao et al. 2007).

CRH family members have been described in many vertebrate clades and their identification in extant representatives of basal vertebrate lineages, such as the holocephalan elephant shark (Callorhinichus miliarii) (Nock et al. 2011) and the agnathan sea lamprey (Petromyzon marinus) (Roberts et al. 2014), shows that they are present throughout vertebrates. A CRH homolog has also been identified in invertebrate deuterostomes, namely the urochordates Ciona intestinalis and Ciona savignyi (Lovejoy & Barsyte-Lovejoy 2010), the cephalochordate Branchiostoma floridae, the echinoderm Strongylocentrotus purpuratus, and the hemichordate Saccoglossus kowalevskii (Mirabeau & Joly 2013). In protostomes, the arthropod diuretic hormones (DH31 and DH44) are the best-characterized homologs of the vertebrate CRH family peptides (Lovejoy & Jahan 2006, Lovejoy & de Lannoy 2013). Related peptides have been found also in other protostomes including nematodes (Lovejoy & Balment 1999) and lophotrochozoans (Mirabeau & Joly 2013), but their physiological roles and affinity for the receptors has not yet been studied. Despite the limited knowledge about the functions of the early vertebrate and invertebrate CRH-related peptides, the DH peptides are known to be key regulators of insect diuresis and food intake, and this has been taken to suggest that the ancestral CRH/DH peptide was probably involved in the regulation of ion homeostasis and energy-related metabolism. Their role in the regulation of the HPA/I axis was probably an innovation that arose in the vertebrate ancestor (Lovejoy & de Lannoy 2013).

The CRH/DH-like peptides bind and activate a group of receptors that belong to class B of the G-protein-coupled receptors (GPCRs) (Hwang et al. 2013, Cardoso et al. 2014, Lovejoy et al. 2014). Vertebrates usually have two highly conserved CRH receptors (CRHR1 and CRHR2) that arose by gene duplication in the two rounds of basal vertebrate genome doubling (Cardoso et al. 2014), the so-called 1R and 2R events (Nakatani et al. 2007, Putnam et al. 2008). The ancestor of teleost fish went through a third round of genome doubling called 3R (Jaillon et al. 2004) that generated a duplicate of the crhr1 gene, leading to Crhr1a and Crhr1b, which were only retained in a few species (Cardoso et al. 2014). In mammals, CRHR1 is strongly activated by CRH and UCN1 and this explains their similar pharmacological actions. CRHR2 is activated by all four peptides in mammals with UCN3 displaying the lowest affinity (Gysling et al. 2004).

Sequence comparisons of the CRH-related peptides suggested that two subfamilies exist in vertebrates, one consisting of CRH and UCN1 and the other harboring UCN2 and UNC3 (Lovejoy & Jahan 2006, Lovejoy & de Lannoy 2013). Two distinct evolutionary models have been proposed to explain the duplications leading from one to four CRH family members in vertebrates. The model by Lovejoy and coworkers proposed that a single ancestral gene was duplicated in the first tetraploidization event, resulting in the CRH/UCN1 ancestor and UCN2/UCN3 ancestor. Subsequently, the second tetraploidization generated the extant quartet (Lovejoy & de Lannoy 2013, Lovejoy et al. 2014). The other model put forth by Hwang and coworkers is based on the chromosomal locations of the four peptide genes in two separate paralogs (Hwang et al. 2013). A paralogon is a set of related chromosomes or chromosome regions, usually resulting from a tetraploidization. Vertebrate paralogons typically have four chromosome members in each paralogon as a consequence of the two tetraploidizations (1R and 2R) (Nakatani et al. 2007, Putnam et al. 2008), and as many as eight members in teleost fish due to their third tetraploidization (3R) (Jaillon et al. 2004). Hwang and coworkers suggested that the two ancestral CRH/UCN1 and UCN2/UCN3 precursor genes arose before the vertebrate tetraploidizations and ended up on separate chromosomes, whereupon each chromosome was quadrupled in the tetraploidizations, but only two peptide genes survived in each paralogon resulting in the four peptide genes that were included in their study, (Figure 4 in Hwang et al. 2013). Their analysis did not consider the CRH-like peptide previously
identified in cartilaginous fish, initially proposed to have arisen early in this lineage (Nock et al. 2011). This peptide was recently reported to be present in other vertebrate lineages and was named CRH2 (Grone & Maruska 2015b). The latter report suggested that CRH2 arose in the basal vertebrate tetraploidizations.

During the course of our previous analyses of the CRH-DH receptor family (Cardoso et al. 2014), we too found that CRH2 is present in representatives from different early-branching vertebrate lineages. We describe here our investigation of the CRH/UCN peptide family using the same approach as in our recent analysis of the CRH-DH receptor family (Cardoso et al. 2014), by combining in-depth analyses of sequences, conserved gene synteny, and paralogon comparisons. After comparison of large chromosomal regions, including phylogenetic analyses of several neighboring gene families, we conclude that two CRH family genes existed in the vertebrate ancestor already before the first tetraploidization (1R). Subsequently, the two tetraploidizations (1R and 2R) resulted in five CRH family peptides, all of which are still present in some of the major lineages. However, CRH2 was lost in placental mammals and in teleost genomes. We also confirm by synteny and paralogon analysis that the recently reported CRH-related peptide in teleost fish, named teleocortin and proposed to be an ancestral vertebrate duplicate (Hosono et al. 2015), is in fact a teleost-specific gene copy of CRH that arose in the teleost tetraploidization, as recently reported (Grone & Maruska 2015a). The evolutionary implications of this ancient ancestral complexity of the CRH system are discussed.

**Methods**

**Database mining and sequence retrieval**

Searches for CRH family members were performed in several vertebrate genomes with the human CRH (NP_000747), UCN1 (NP_003344), UCN2 (NP_149976), and UCN3 (NP_444277) peptide precursors and the deduced elephant shark (Callorhinchus milii) CRH2 (Nock et al. 2011) peptide precursor using the blast algorithm. Searches included the genomes of the early diverging ray-finned fish the spotted gar (Lepisosteus oculatus), the basal sarcopterygian representative the African coelacanth (Latimeria chalumnae), and also two mammalian key genomes, the duck-billed platypus (Ornithorhynchus anatinus) that is a member of the order Monotremata (an early divergent mammalian clade) and the marsupial gray short-tailed opossum (Monodelphis domestica).

Other vertebrate genomes were also analyzed including chicken (Gallus gallus), the anole lizard (Anolis carolinensis), the amphibian Xenopus tropicalis (Xenopus tropicalis) and the teleost fishes Japanese pufferfish (Takifugu rubripes), Nile tilapia (Oreochromis niloticus), zebrafish (Danio rerio), medaka (Oryzias latipes), three-spined stickleback (Gasterosteus aculeatus), and Atlantic cod (Gadus morhua) in the ENSEMBL database (http://www.ensembl.org/) and also the teleost Atlantic salmon (Salmon salar) that is available from the SalmonDB database (http://salmondb.cmm.uchile.cl). Searches were extended to two lampreys, the sea lamprey (Petromyzon marinus) (http://www.ensembl.org/) and the Arctic lamprey (Lethenteron camtschaticum) (http://jlampreygenome.imcb.a-star.edu.sg/). The cartilaginous fish the elephant shark (Callorhinchus milii) (http://esharkgenome.imcb.a-star.edu.sg) was also queried for any additional CRH family members.

The predicted peptide precursors were retrieved and identity was confirmed by searching against the human orthologs available at in the NCBI database using BLASTp and also in the Pfam database (http://pfam.xfam.org) to identify the protein family motifs. When short sequence motifs were found in nonannotated genome regions, sequences were retrieved from the genome assemblies extending 1 Mb upstream and downstream of the sequence hit and then translated into protein using the Expasy nucleotide translating tool (http://web.expasy.org/translate/) to obtain the most complete peptide precursor. Putative monobasic, dibasic, or tribasic peptide proteolytic consensus cleavage sites (RR, KR, KK) were identified or predicted using NeuroPred (http://stagbeetle.animal.uiuc.edu/cgi-bin/neuropred.py) (Southey et al. 2006) to retrieve the mature peptides for further analysis (Supplementary Data 1, see section on supplementary data given at the end of this article).

**Sequence alignments and phylogenetic analysis**

The deduced peptide precursors and mature peptides were compared across species and were aligned with ClustalW (http://www.genome.jp/tools/clustalw/) and edited in GeneDoc (http://iubio.bio.indiana.edu/). Phylogenetic trees were built using the maximum likelihood (ML) and neighbor-joining (NJ) methods and bootstrapped to assign accuracy to the clades (Felsenstein 1985). Trees were constructed based on the alignment of 78 mature peptide sequences produced in ClustalW that was submitted to ProtTest 2.4 to select the best model to study protein phylogeny according to the Akaike Information Criterion (AIC) statistical model (Abascal et al. 2005).
Table 1  Accession numbers and origin of the CRH/UCN sequences retrieved.

<table>
<thead>
<tr>
<th></th>
<th>CRH1</th>
<th>CRH2</th>
<th>UCN1</th>
<th>UCN2</th>
<th>UCN3</th>
<th>CRH/UCN1</th>
<th>UCN2/UCN3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TETRAPOD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human</td>
<td>NP_000747</td>
<td>ni</td>
<td>NP_00344</td>
<td>NP_149976</td>
<td>NP_444277</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Opossum</td>
<td>ENSMODP00000036360</td>
<td>na</td>
<td>ENSMODP00000028973</td>
<td>na</td>
<td>ENSMODP00000023062</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Platypus</td>
<td>ENSOANP0000003650</td>
<td>Ultra337_3</td>
<td>Ultra337_3</td>
<td>Ultra32_0.7</td>
<td>Ultra32_0.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chr 7.16</td>
<td></td>
<td>Chr 1.472</td>
<td>Chr 1.508.59</td>
<td>Chr 6.194</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>na</td>
<td>na</td>
<td>na</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CHICKEN</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>LIZARD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>XENOPUS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>LOBE-FINNED FISH</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coelacanth</td>
<td>ENSLACP0000002936</td>
<td>na</td>
<td>ENSLACP0000001213</td>
<td>na</td>
<td>na</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>JH128518.1_0.54557</td>
<td>JH127070.1_0.454</td>
<td>JH126901.1_1.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>RAY-FINNED FISH</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spotted gar</td>
<td>ENSLOCP0000002197</td>
<td>na</td>
<td>ENSLOCP00000022093</td>
<td>na</td>
<td>ENSLOCP00000022005</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LG 9_6</td>
<td>LG 9_6</td>
<td>LG 9_6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tilapia</td>
<td>a-ENSOPNP00000025716</td>
<td>GL31461.1_0.3789</td>
<td>ENSONIP00000025725</td>
<td>GL31520.1_0.242</td>
<td>ENSONIP00000025862</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GL831530.1_0.242</td>
<td>GL831157.1_0.8489</td>
<td>GL831167.1_2.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fugu</td>
<td>b-ENSTRUP00000022019</td>
<td>GL831167.1_0.8489</td>
<td>ENSSTRUP00000030463</td>
<td>ni</td>
<td>ni</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Scaffold 95_0.421</td>
<td>Scaffold 120_0.061</td>
<td>Scaffold 116_0.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zebrafish</td>
<td>a-ENSDARP0000118835</td>
<td>ni</td>
<td>NP_001025351</td>
<td>ni</td>
<td>ni</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Chr 2_42.21</td>
<td>Chr 2_42.21</td>
<td>Chr 2_42.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B-ENSGMOP0000001885</td>
<td>B-ENSGMOP0000001885</td>
<td>B-ENSGMOP0000001885</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>groupXXI_7306566</td>
<td>groupXXI_7306566</td>
<td>groupXXI_7306566</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ni</td>
<td>ni</td>
<td>ni</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ni</td>
<td>ni</td>
<td>ni</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stickleback</td>
<td>a-ENSGAC00000003885</td>
<td>ni</td>
<td>ENSGACP00000008094</td>
<td>groupXVIII_33095</td>
<td>groupXVIII_33095</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>groupXVIII_33095</td>
<td>groupXVIII_33095</td>
<td>groupXVIII_33095</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ni</td>
<td>ni</td>
<td>ni</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ni</td>
<td>ni</td>
<td>ni</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cod</td>
<td>a-ENSGMOG000000002152</td>
<td>ni</td>
<td>ENSGMOG00000017865</td>
<td>ni</td>
<td>ni</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GeneScaffold_378_17550</td>
<td>GeneScaffold_378_17550</td>
<td>GeneScaffold_378_17550</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ni</td>
<td>ni</td>
<td>ni</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medaka</td>
<td>a-ENPM000121990</td>
<td>ni</td>
<td>NP_001295911.1</td>
<td>NP_001295911.1</td>
<td>NP_001295911.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Chr 20_15301114</td>
<td>Chr 20_15301114</td>
<td>Chr 20_15301114</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B-ENSGMOP00000013071</td>
<td>B-ENSGMOP00000013071</td>
<td>B-ENSGMOP00000013071</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salmon</td>
<td>a-ENPM000121990</td>
<td>ni</td>
<td>NP_001295911.1</td>
<td>NP_001295911.1</td>
<td>NP_001295911.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Chr 20_15301114</td>
<td>Chr 20_15301114</td>
<td>Chr 20_15301114</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B-ENSGMOP00000013071</td>
<td>B-ENSGMOP00000013071</td>
<td>B-ENSGMOP00000013071</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ni</td>
<td>ni</td>
<td>ni</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ni</td>
<td>ni</td>
<td>ni</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The ML tree was built with PhyML 3.0 available from ATGC (http://www.atgc-montpellier.fr/phyml/) (Guindon et al. 2010) with the LG substitution model and with the following parameters: 4 gamma distributed rate categories and fixed gamma shape parameter (1.229). Reliability for internal branching was assessed using 100 bootstrap replicates. Sequence data were also analyzed with the NJ method (Saitou & Nei 1987) using the program MEGA version 6 (Tamura et al. 2013) with the JTT substitution model and fixed G (1.528) with pairwise deletion and 1000 bootstrap replicates. The consensus trees obtained with ML and NJ analysis shared similar topologies.

Gene synteny comparisons

The chromosomal environments of the five spotted gar CRH family genes were characterized and orthologous genes were identified in the human, chicken, zebrafish, and elephant shark genomes. Candidate genes were selected based on previous paralogon annotations of the opioid receptor and peptide genes (Sundstrom et al. 2010, Larhammar et al. 2015) and the visual opsin and oxytocin/vasopressin receptor genes (Lagman et al. 2013). Orthologous genes were sought by querying the genome assemblies for the different species, and gene identity (ortholog vs paralog) was investigated by phylogenetic analyses following a similar procedure as described above for the CRH family members.

Results

CRH gene family members in vertebrates

Searches for CRH family members in several vertebrate genome databases identified a large number of genes as listed in Table 1. A total of five distinct sequences were retrieved from three species representing lineages that diverged early in vertebrate evolution, namely coelacanth, spotted gar, and elephant shark. Extensive phylogenetic and chromosomal analyses described below identified four of these as orthologs of mammalian CRH, UCN1, UCN2, and UCN3 peptides. The CRH2 peptide (Nock et al. 2011, Grone & Maruska 2015a) was found to be orthologous between coelacanth, gar, and elephant shark, and is present also in the anole lizard and probably also in platypus and opossum (see below), but is missing in other vertebrate lineages including placental mammals and teleosts. For clarity, we will use the name CRH1 for the ‘standard’ CRH gene to distinguish it from the novel CRH2 gene.
For each of the CRH family members, we also identified and analyzed gene families with members in the same chromosomal regions. This showed that the CRH family genes are located in two paralogons (sets of related chromosome regions), which have been previously analyzed in detail by us, namely the paralogon that contains the opioid peptide and receptor genes (Dreborg et al. 2008, Sundstrom et al. 2010, Larhammar et al. 2015) and the one that contains the visual opsin genes and the oxytocin and vasopressin receptor genes (Lagman et al. 2013). This allowed us to distinguish orthologs and paralogs for the CRH family genes and to deduce that the CRH family duplications took place as part of the basal vertebrate tetraploidizations (with one additional copy arising in the teleost tetraploidization, see below).

Mammals

In human and other placental mammals, four CRH family members exist. The genomes of two nonplacental mammals, the duck-billed platypus and the gray short-tailed opossum, were found to have CRH1, UCN2, and UCN3 genes. Opossum also had UCN1, but this gene could not be found in platypus. Both of these nonplacental mammals have a putative CRH2 gene located in platypus on chromosome Ultra337_3Mb and in opossum on Chr1_472Mb (Table 1). This gene is absent from the genomes of placental mammals.

Nonmammalian tetrapods

The genomes of chicken, anole lizard, and X. tropicalis contain the family members CRH1 and UCN3. The lizard also has the CRH2 gene and the UCN1 gene, but lacks UCN2. Chicken has UCN2, but lacks CRH2 and UCN1. The frog, finally, lacks CRH2, UCN1, and UCN2. It remains possible that the missing genes are indeed present in the genomes but have not yet been sequenced and included in the assembly. This possibility appears especially likely for chicken because putative CRH2 genes were identified in other bird genomes such as duck (Anas platyrhynchos, KB744325_0.354Mb), flycatcher (Ficedula albicollis, JH603360_1.093Mb), and zebra finch (Taeniopygia guttata, chr_20_8.814Mb). The microchromosomes of chicken are known to be underrepresented in the genome assemblies (Lovell et al. 2014, 2015, Hton et al. 2015). This might apply also to the lizard as it has microchromosomes too. In X. tropicalis, two CRH1 gene copies are predicted to be localized in tandem on scaffold GL173273 that are 100% identical at nucleotide level, suggesting that they may be a consequence of an error in the genome assembly rather than the result of a species-specific recent duplication.

Ray-finned fish

Two crh1 genes were found in most of the teleosts investigated including zebrafish, stickleback, medaka, and cod. Phylogenetic analyses and synteny comparisons (see below) showed that these two copies derive from the teleost 3R event; hence, we have named them crha and crhb, as recently done also by Grone and Maruska (2015a). Only crhb was found in the Takifugu genome. Single copies of ucn1 and ucn3 genes were found in all teleosts investigated, implying that the 3R tetraploidization duplicates have been lost. Likewise, the ucn2 gene was found in a single copy in the teleosts with the exception of zebrafish that has none and Atlantic salmon that received a second copy, presumably in the salmonid-specific fourth tetraploidization (4R). As a consequence of the extra genome doubling, a second copy of both crha and crhb were gained in the Atlantic salmon (crha1 and crha2; crhb1 and crhb2). These presumed 4R duplicates encode identical mature Crha and Crhb peptides (Grone & Maruska 2015a). All of the teleost genomes analyzed lack an ortholog of the novel crh2 gene. All five CRH family members, including Crh2, were identified both in the spotted gar that diverged before the teleost radiation and the coelacanth, a lobe-finned fish (Table 1).

Elephant shark and lampreys

Genes for four CRH family members have been previously identified in the elephant shark, namely CRH1, UCN1, UCN3, and the novel CRH2. Our investigation revealed that also the UCN2 gene exists. Thus, this cartilaginous fish displays the same gene repertoire as spotted gar and coelacanth (Table 1). The presence of the full set of CRH family peptides in the elephant shark makes it even more intriguing that so far only one receptor has been identified, CRHR1 (Cardoso et al. 2014).

Homologs of the bony vertebrate CRH/UCN members were found in the genomes of two lampreys, the sea lamprey Petromyzon marinus and the Arctic lamprey Lethenteron camtschaticum. Two CRH family genes were found in the former and four in the latter. Although it is possible that P. marinus may have lost two of the peptide genes, the recent divergence of these two species from each other only 10–30Mya (Kuraku & Kuratani 2006) makes it more likely that two genes are missing in the
Some of the agnatha sequences tend to cluster with the gnathostome peptide clades, and putative CRH1, CRH2, and UCN2 seem to exist in the Arctic lamprey and UCN2 in the sea lamprey genomes suggesting that the lamprey genomes possess a similar gene content to the other vertebrates; however, more detailed analysis of gene synteny is required to confirm orthology.

Sequence comparison of the vertebrate CRH/UCN peptides

Specific amino acid motifs were identified that are common to all family members or characteristic for the CRH1/CRH2/UCN1 and UCN2/UCN3 clades (Fig. 1). The S-L/I-D motif at the N-terminus is present in all five peptides and conserved between human, coelacanth, spotted gar, and elephant shark. The N near the carboxy-terminus and the A located three positions before are also conserved across the four species. Within the CRH1/CRH2/UCN1 subfamily, the P before the S-L/I-D motif and the L-T-F-H-L/I-L-R motif after it are also conserved. Five other amino acids were found to be totally conserved between CRH1 and CRH2. In the UCN2/UCN3 subfamily, several conserved motifs were also identified (Fig. 1).

Phylogeny of the CRH family members

The mature peptide sequences of the CRH family members (Supplementary Fig. 1) were used to construct evolutionary trees with both ML (Fig. 2 and Supplementary Fig. 2A) and NJ (Supplementary Fig. 2B) methods that produced similar branch topologies. Both phylogenetic trees are composed of two major and clearly separate peptide clusters, one that contains the CRH1/CRH2/UCN1 and the other comprising the UCN2/UCN3 subfamilies, respectively. Positions shown with ‘*’ are common to CRH1 and CRH2. The consensus peptide sequence for each subfamily is represented below and annotated partial conservation, small case letters indicate different evolutionary pressures. The peptides CRH1 and UCN3 evolve more slowly and presumably have been subjected to stronger conservative selection pressure (shorter tree branches) than the other
three members. Moreover, within each peptide cluster, the sequences do not radiate according to the established species trajectories as a result of lineage-specific events and rate differences. The lamprey peptides group with either the CRH1/CRH2/UCN1 clade or the UCN2/UCN3 clade, showing that the two subfamilies arose before the divergence of cyclostomes and gnathostomes.

**Analyses of neighboring gene families**

The chromosomal locations of the CRH family genes show that they are located in two distinct paralogs: CRH1, CRH2, and UCN1 reside in the same paralogon as the previously described opioid peptide and receptor genes (Dreborg *et al.* 2008, Sundstrom *et al.* 2010, Larhammar *et al.* 2015), and the UCN2 and UCN3 genes are located in the paralogon harboring the visual opsin and oxytocin/vasopressin receptor genes (Lagman *et al.* 2013) (Fig. 3).

Each of these two paralogs has numerous other gene families represented on 2–4 of the chromosomes within each paralogon (Dreborg *et al.* 2008, Sundstrom *et al.* 2010, Lagman *et al.* 2013, Larhammar *et al.* 2015). Extensive conservation of gene synteny was found for several neighboring gene families in the genomes of human, chicken, and spotted gar. Eight of these gene families that are syntenic with the CRH family members in either of the paralogs are shown in Fig. 3, namely STMN, NCOA, XKR, HCK/LCK/BLK/LYN, GATA (two subfamilies), CAMK1, PLXNA, and CACNA2D (two subfamilies). The gene families were investigated in detail
Gene synteny analysis of the CRH1/CRH2/UCN1 and UCN2/UCN3 paralogons. Chromosomal locations of the CRH family members and neighboring genes in five different vertebrate clades are shown. The numbers to the left and right of the chromosome lines refer to chromosome or scaffold number. Gene positions (Mb) are given below each gene. The order of the genes on the chromosomes has been reshuffled and members of the same gene family were aligned to highlight the similarities both between species and between the chromosomes within each species, i.e. paralogon. The gene abbreviations stand for: CACNA2D, voltage-dependent calcium channel subunit alpha2delta family; CAMK/IPCK, Ca2+/calmodulin-dependent protein kinase family; GATA, GATA transcription factor family; HCK/ILCK/BLK/KLYN, tyrosine protein kinases; NCOA, nuclear receptor coactivator family; PLXNA, plexin A family; STMN, stathmin; XKR, X Kill blood group precursor-related family. A full colour version of this figure is available at http://dx.doi.org/10.1530/JME-16-0051.
with alignments and phylogenetic analyses as described for the CRH family. Phylogenetic trees are shown in Supplementary Fig. 3. All of these gene families display tree topologies and species distribution of the family members that comply with duplications in the 2R and 3R events. Four of the gene families are complete quartets from 2R: STMN, HCK/LCK/BLK/LYN, CAMK1, and PLXNA. Two additional neighboring gene families, TRIM and DNAJC5, are included in Supplementary Fig. 4. These families have less well-resolved phylogenies, but their close proximity to CRH1/CRH2/UCN1 provides further support that these four chromosome regions resulted from a common ancestral chromosome.

Also, the zebrafish displays extensive conserved synteny, but as usual has a more complex situation due to teleost 3R, gene losses and chromosomal rearrangements. Combining information from a few teleost species, we found multiple syntenies reflecting the common ancestry with the spotted gar and the tetrapod ancestor. None of the four gene families that were complete quartets after 2R has become a complete octet in zebrafish after 3R, but the camk1 family is close with seven members. As mentioned above, note that for the CRH family only CRH1 gave rise to a surviving duplicate after the 3R tetraploidization. CRH2 seems to have been lost in the teleost ancestor. Ucn2 could not be identified in the zebrafish genome but is present in other teleost fish genomes.

The elephant shark genome assembly has not yet allowed assignment of chromosomes, but all five CRH family genes are located on scaffolds that harbor at least one of the neighboring genes found in the other vertebrates. For instance, the new CRH2 gene maps to elephant shark scaffold 6 together with GATA5, HCK, and STMN3, thereby displaying conserved synteny with CRH2 in spotted gar.

Interestingly, both of the CRH subfamilies are linked to GATA family members as noted previously (Hwang et al. 2013). Our phylogenetic analysis shows that GATA too forms two distinct subfamilies, one consisting of GATA1, 2, and 3 and the other containing GATA 4, 5, and 6 (Supplementary Fig. 3). This suggests that the ancestor of the vertebrates underwent a duplication of a chromosomal region containing both a CRH family gene and a GATA gene, and that the duplicated regions were translocated to a different chromosome. Subsequently, both of the CRH-GATA pairs were duplicated in 2R, and CRH1, GATA1, and GATA2 were also duplicated in teleost 3R.

Taken together, the duplications of the two CRH subfamilies and the eight neighboring gene families (two of which consist of two subfamilies), as well as several additional gene families, located in these chromosomal regions (Dreborg et al. 2008, Sundstrom et al. 2010, Lagman et al. 2013, Larhammar et al. 2015) can most parsimoniously be explained by chromosome duplications (Fig. 3).

**Discussion**

The evolution of the CRH family has been exceedingly difficult to resolve due to uneven evolutionary rates among its members and gene losses in different evolutionary lineages. This has led to conflicting models of the evolutionary history of the peptide family. We, therefore, applied our combined approach of sequence-based phylogenetic analyses and comparison of chromosomal locations across species. This approach allowed us to deduce the evolutionary history of the family including the recently discovered CRH2 gene, the fifth ancestral vertebrate member of this family (Nock et al. 2011, Grone & Maruska 2015b).

Members of the CRH peptide family were identified in the genomes of several vertebrates. Phylogenetic analysis of the predicted mature peptides revealed that the gnathostome CRH family members form five distinct clades (Fig. 2): the previously described CRH1, UCN1, UCN2, and UCN3 and an additional peptide most closely related to CRH1 and UCN1. This peptide was previously identified in a shark (Nock et al. 2011) and was suggested to have arisen in this lineage. Our identification of this gene also in coelacanth and spotted gar (Fig. 1), along with its chromosomal location (Fig. 3), shows that it arose in the basal vertebrate tetraploidizations before the cartilaginous fishes diverged from the lineage leading to tetrapods, coelacanth, and ray-finned fishes (Fig. 4), in agreement with a recently published independent study (Grone & Maruska 2015b).

In order to date more accurately the CRH family gene duplications, we investigated the chromosomal regions of their genes in several species, primarily human, African coelacanth, spotted gar, and zebrafish. This approach allowed us to identify extensive conservation of synteny for the CRH family members between these species. Incidentally, these chromosomal regions have been investigated in great detail previously in our laboratory in connection with our studies of the evolution of the opioid peptide and receptor gene families (Dreborg et al. 2008, Sundstrom et al. 2010, Larhammar et al. 2015) and the visual opsin gene family (Lagman et al. 2013).
Hence, the two paralogons harboring these gene families have been called by us the opioid paralogon and the opsin paralogon, respectively. The opioid and opsin gene families are not included in Fig. 3 because they involve some additional rearrangements in some of the species, see (Dreborg et al. 2008, Sundstrom et al. 2010, Lagman et al. 2013, Larhammar et al. 2015). The CRH1/CRH2/UCN1 subfamily is located in the opioid paralogon and the UCN2/UCN3 subfamily is present in the opsin paralogon (Fig. 3). A recently published report on CRH2 proposed that the two subfamilies arose in 2R, but did not provide synteny and paralogon information to show this (Grone & Maruska 2015b). The previous report proposing origin in 2R had not identified CRH2 and did not perform phylogenetic analyses of the neighboring gene families to show that their gene duplications were simultaneous (Hwang et al. 2013).

The chromosomal regions for the two CRH gene subfamilies are shown in Fig. 3 and Supplementary Fig. 4. The order of the genes on the chromosomes has been adjusted to highlight the similarities both between the chromosomes within each species, thus displaying the paralogon relationship, and between species, thus exposing the conservation of synteny. It is well known that intrachromosomal rearrangements happen frequently (Ocampo Daza et al. 2012, Schmid et al. 2015), so it is not surprising that some differences in gene order have arisen between the four chromosomes in each quartet, as well as between species for the corresponding chromosomal regions. The positions of the genes are shown below each box, and it can be seen that many of the genes are still in reasonably close proximity to one another.

It should be emphasized that the gene family topologies are highly similar for all of these gene families.
(Supplementary Fig. 3), thereby supporting shared evolutionary history by duplication in the 2R events. It should also be noted that the paralogon relationships between the chromosomes shown in Fig. 3 are just a small subset of the neighboring gene families, with four complete quartets, four triplets, and two pairs, after the 2R tetraploidizations. Several additional gene families are present in these paralogons and their sequence-based phylogenies have been analyzed in detail by us and found to add further support to concomitant duplications by chromosome doubling; see our previous articles for further details (Dreborg et al. 2008, Lagman et al. 2013, Larhammar et al. 2015, Sundstrom et al. 2010).

**CRH2, the fifth ancient member of the peptide family**

The presence of the newly discovered CRH2 gene in the three vertebrate lineages represented by the elephant shark, the spotted gar, and the coelacanth is probably no coincidence, because it has been found that these genomes evolve slowly in terms of both rate of amino acid changes and chromosomal organization (Amores et al. 2011, Venkatesh et al. 2014). Interestingly, we found that the CRH2 gene is also present in platypus and opossum as well as in the anole lizard and also in birds with the exception of chicken. This suggests that CRH2 has been lost independently at least twice, namely in ray-finned fishes and in placental mammals. The persistence of CRH2 in at least four vertebrate lineages will hopefully make it possible to address its functional role, and this may in turn lead to clues how it could be lost in some lineages.

**Lineage-specific differences in CRH family evolution**

While the CRH1 and UCN3 genes have persisted in all of the genomes analyzed, the evolution of UCN1, UCN2, and CRH2 has been affected by lineage-specific events. The UCN1 gene is absent from the genomes of *X. tropicalis*, anole lizard, chicken, and platypus; UCN2 is absent from *X. tropicalis* and the anole lizard and also from zebrafish. The CRH2 gene is absent from placental mammals, chicken, *X. tropicalis*, and teleost fishes. The physiological consequences of these gene losses have not yet been explored, and it must also be kept in mind that gene absence may be due to incomplete genome assemblies.

In this study, orthologs of the UCN2 were identified in several vertebrate classes. This peptide was initially described in mammals, and only recently a putative gene was predicted in chicken and in several teleost genomes (Hwang et al. 2013, Lovejoy & de Lannoy 2013). In the elephant shark, a putative nonfunctional UCN2 gene was described as a result of the presence of a premature stop codon in the coding sequence (Nock et al. 2011, Lovejoy & de Lannoy 2013). However, we could identify the UCN2 gene as being potentially functional in the most recent elephant shark genome assembly.

After completion of this study, a member of the CRH peptide family was reported in a few species of teleost fish and was named teleocortin (Tcn) (Hosono et al. 2015). The gene was reported to be missing in zebrafish. The phylogenetic analyses led the authors to propose that Tcn arose before the teleost tetraploidization, but was not duplicated in this event. We had identified this gene in several additional teleosts including zebrafish and concluded from our synteny and paralogon analyses that this gene is indeed a duplicate of CRH1. The same conclusion was drawn independently by Grone and Maruska (2015a), and we use the same name as them for this gene: *crha*.

**Conclusions**

Analysis of vertebrate genomes suggests that five ancestral CRH family genes existed at the origin of the vertebrates as a result of the two ancient genome doublings of two distinct peptide genes that were present in the vertebrate progenitor. Thus, together with the two CRH receptors that resulted from the same events, the CRH/UCN system displayed impressive complexity already approximately 500 million years ago (Fig. 4). The fifth CRH2 gene still exists in four separate vertebrate lineages, namely cartilaginous fish (elephant shark), coelacanth, ray-finned fish (spotted gar), and several amniotes (birds, lizard, nonplacental mammals), although it has been lost in placental mammals, chicken, and teleost fish. Therefore, its functional roles will be interesting to explore. Additional gene gains in teleost fishes through their third tetraploidization, as well as lineage-specific gene losses, have resulted in distinct gene repertoires and variable evolutionary rates between the peptide family members and evolutionary lineages. The vertebrate genome doublings have once again provided great explanatory power for the ancient vertebrate gene complexity.

**Supplementary data**

This is linked to the online version of the paper at http://dx.doi.org/10.1530/JME-16-0051.
Declaration of interest
The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Funding
This study received Portuguese national funds from FCT – Foundation for Science and Technology through project UID/Multi/04326/2013. J C R C has received grant support from an auxiliary research contract (CCMAR/UID/Multi/04326/2013) and R C F by FCT SFRH/BPD/89811/2012. D L has received grant support from The Swedish Research Council.

Author contribution statement
J C R C, C A B, and R C F collected the data. J C R C, C A B, R C F, and D L performed the analysis and discussed the results. J C R C and D L wrote the manuscript.

References
Grone BP & Maruska KP 2015b A second corticotropin-releasing hormone gene (CRH2) is conserved across vertebrate classes and expressed in the hindbrain of a basal neopterygian fish, the spotted gar (Lepisosteus oculatus). Journal of Comparative Neurology 523 1125–1143. (doi:10.1002/cne.23729)
Lagman D, Ocampo Daza D, Widmark J, Abalo XM, Sundstrom G & Larhammar D 2013 The vertebrate ancestral repertoire of visual opsins, transducin alpha subunits and oxytocin/vasopressin receptors was established by duplication of their shared genomic region in the two rounds of early vertebrate genome duplications. BMC Evolutionary Biology 13 238. (doi:10.1186/1471-2148-13-238)


Received in final form 18 May 2016
Accepted 24 May 2016
Accepted Preprint published online 24 May 2016