In Silico Biology 7 (2007) 355–367
IOS Press

Positive Darwinian Selection on Crustacean Hyperglycemic Hormone (CHH) of the Green Shore Crab, *Carcinus maenas*

Abinash Padhi[a,*], Bindhu Verghese[a], Subhendu K. Otta[b], Binu Varghese[c] and Karri Ramu[d]

[a]*Department of Biological Science, University of Tulsa, 600 S. College Ave., Tulsa, OK-74104, USA*
[b]*Baylor College of Medicine, Department of Pediatrics-Hematology/Oncology, 6621 Fannin Street, MC 3-3320, Houston, TX-77030, USA*
[c]*Central Marine Fisheries Research Institute, Vizhinjam Research Center, Trivandrum, Kerala, India*
[d]*Center for Marine Environmental Studies, Ehime University, Bunkyo Cho 2-5, Matsuyama 790-8577, Japan*

Edited by E. Wingender; received 28 December 2006; revised 12 March 2007; accepted 3 April 2007; published 16 June 2007

ABSTRACT: The tissue-specific expression and differential function of the crustacean hyperglycemic hormone (CHH) in *Carcinus maenas* indicate an interesting evolutionary history. Previous studies have shown that CHH from the sinus gland X-organ (XO-type) has hyperglycemic activity, whereas the CHH from the pericardial organ (PO-type) neither shows hyperglycemic activity nor it inhibits Y-organ ecdysteroid synthesis. Here we examined the types of selective pressures operating on the variants of CHH in *Carcinus maenas*. Maximum likelihood-based codon substitution analyses revealed that the variants of this neuropeptide in *C. maenas* have been subjected to positive Darwinian selection indicating adaptive evolution and functional divergence among the CHH variants leading to two unique groups (PO and XO-type). Although the average ratio of nonsynonymous to synonymous substitution (ω) for the entire coding region is 0.5096, few codon sites showed significantly higher ω (10.95). Comparison of models that incorporate positive selection ($\omega > 1$) with models not incorporating positive selection ($\omega < 1$) at certain codon sites failed to reject ($p = 0$) evidence of positive Darwinian selection.

KEYWORDS: Molecular evolution, crustacean hyperglycemic hormone, positive Darwinian selection, synonymous and nonsynonymous substitution, *Carcinus maenas*

INTRODUCTION

Crustacean hyperglycemic hormone (CHH), a polypeptide neurohormone, plays a crucial role in regulating carbohydrate metabolism [Santos and Keller, 1993] and in other physiological processes including reproduction [Khayat *et al.*, 1998], osmoregulation [Serrano *et al.*, 2003], lipid metabolism [Santos *et al.*, 1997] and molting [Chung *et al.*, 1999]. Phylogenetic and structural analyses of CHH, molt-inhibiting hormone (MIH), vitellogenesis-inhibiting hormone (VIH) and mandibular organ-inhibiting hormone (MOIH) revealed high structural similarities among these neuropeptides [Keller, 1992; Lacombe *et al.*, 1999; Chan *et al.*, 2003; Chen *et al.*, 2005], thus suggesting that these neuropeptides diverged from a common ancestor. The CHH/MIH/GIH gene family in crustaceans has an interesting evolutionary

*Corresponding author. Tel.: +1 918 631 3062; Fax: +1 918 631 2762; E-mail: abinash-padhi@utulsa.edu.

Electronic publication can be found in *In Silico Biol.* **7**, 0024 <http://www.bioinfo.de/isb/2007/07/0024/>, 16 June 2007.

history, as it contains several isoforms that are expressed in different tissues (e.g. reviews in Chan *et al*., 2003; Fanjul-Moles, 2006). Chan *et al.*, 2003, proposed that like prolactin and growth hormone gene family in vertebrates [e.g. Ohta, 1993; Wallis *et al.*, 2005; Wallis *et al.*, 2001; Wallis and Wallis, 2001], the crustacean CHH/MIH/GIH gene family might have also diverged from a common ancestral gene through multiple gene duplication events. However, the pattern of nucleotide substitution among the duplicated genes is unclear. Gene duplication followed by functional divergence of duplicated genes is one of the most important underlying mechanisms for the evolution of novel gene function [Ohno, 1970; Zhang *et al.*, 1998]. Many genes in vertebrates [e.g. Ohta, 1993; Zhang *et al.*, 1998; Goldstone and Stegeman, 2006; Kitano *et al.*, 2006; Shiu *et al.*, 2006] are shown to have accelerated amino acid substitutions following gene duplication and positive Darwinian selection is the likely cause of such accelerated amino acid substitutions among the duplicated genes.

In decapods, although the sinus gland X-organ (XO) is considered as the main locus for neuropeptide production [Fanjul-Moles, 2006], CHH/MIH has also been detected in other sites of the crustacean organs. For example, CHH has been detected in paraneurons of the fore and hind gut [Chung *et al.*, 1999; Webster *et al.*, 2000], the abdominal pericardial peripheral neurons [Chung and Webster, 2004] and in the pericardial organ (PO) of green shore crab *Carcinus maenas* [Dircksen and Heyn, 1998; Dircksen *et al.*, 2000]. It has been proposed that the non-eyestalk CHH could correspond to a different form of CHH than that produced in the X-organ of the sinus gland [Fanjul-Moles, 2006]. Dircksen *et al.*, 2001, reported production of two types of peptides from the neurosecretory cells of PO, which lack hyperglycemic activity. We hypothesized that following gene duplication positive Darwinian selection could be the likely cause of the rapid functional diversification resulting in different variants of CHH in *C. maenas*.

Neutral theory predicts that most variations in the genes/genome are the outcome of the stochastic nature of the mutational process (purifying selection) [Kimura, 1983]. However, many of the functional proteins in all living forms are influenced by natural selection (positive Darwinian selection) [e.g., Ford, 2002], therefore proven to be highly adaptive. Thus, detecting positive Darwinian selection in functionally important genes is important towards understanding the molecular evolution. While purifying selection (functionally constrained) favors the preservation of existing phenotypes, positive selection causes functional diversification of protein coding genes. A number of statistical approaches have been developed to detect positive selection of protein-coding genes [Yang and Bielawski, 2000; Nielsen, 2001; Ford, 2002]. One of the most widely used approaches to detect positive selection is by comparing the number of nonsynonymous substitution per nonsynonymous site (d_N) with the number of synonymous substitution per synonymous site (d_S) [Kimura, 1983; Hughes and Nei, 1989]. Mutation at a codon site that results in an amino acid change is known as nonsynonymous substitution, whereas in synonymous substitution amino acid remains unchanged. If the protein-coding gene is functionally constrained, then the rate of nonsynonymous changes will be lower than neutral rate resulting in $\omega < 1$, and the gene is considered to be subjected to strong purifying selection [Kimura, 1983]. Alternatively, if nonsynonymous mutations are beneficial, then the average rate of nonsynonymous changes is expected to be higher than the neutral rate resulting in $\omega > 1$, which indicates functional diversification of the gene and the possibility that it has been subjected to positive selection [Hughes and Nei, 1989; Ford, 2002].

Like many functional genes [e.g. Ford, 2002], positive selection might also have caused functional differences in XO-type and PO-type CHH in *C. maenas*. Although all amino acid differences between these two forms of CHH might not be adaptive, the functional differences between these two forms could be due to adaptive evolution of certain amino acids. Therefore, identifying the amino acid site on which positive selection has been operating would gain insight into understanding the evolution of XO

and PO-type CHH in *C. maenas*. Maximum likelihood (ML)-based codon substitution models [Yang *et al.*, 2000] that account for variable ω among codon sites have been commonly used [e.g. Swanson *et al.*, 2001; Swanson and Aquadro, 2002] to detect positive selection. Based on the phylogenetic perspective of evolution, here we report the types of selective pressures operating on the variants of *C. maenas*. We used ML-based codon substitution models of Yang *et al.*, 2000, and random effects likelihood (REL) method of Kosakovsky Pond and Frost, 2005, to detect positive Darwinian selection across the codon sites of CHH variants of *C. maenas*.

MATERIALS AND METHODS

Phylogenetic Analyses

To infer evolutionary relationship among 23 CHH variants/isoforms representing six crab species, nucleotide coding sequences were retrieved from GenBank (Table 1). Sequences were aligned using Mesquite ver. 1.11 [Maddison and Maddison, 2006] and DAMBE ver. 4.5 [Xia, 2000; Xia and Xie, 2001]. Aligned amino acid sequences were mapped to corresponding codons using DAMBE. Phylogenetic relationships among these six species representing 23 CHH variants/isoforms were inferred from maximum likelihood (ML), Bayesian inference (BI), maximum parsimony (MP) and neighbor-joining (NJ) methods. As described by Chen *et al.*, 2005, CHH-A of lobster *H. americanus* was chosen as out-group. Kimura-2-parameter (K80) with the proportion of invariable site ($I = 0.161$) gamma distribution shape parameter ($G = 2.8284$) was the best-fit model selected by hierarchical likelihood ratio tests (hLRTs) implemented in Modeltest ver. 3.5 [Posada and Crandall, 1998]. MP analysis was conducted using heuristic search option, implementing stepwise addition with 100 random addition replicates and TBR branch swapping using PAUP* 4.0b10 [Swofford, 2002]. PHYML ver. 2.4.4 [Guindon and Gascuel, 2003] was used for ML analyses and MrBayes 3.04 [Huelsenbeck and Ronquist, 2001] was used for BI. The resulting trees were drawn using TreeView [Page, 1996]. Nodal support for MP and ML trees were estimated using 1000 non-parametric bootstrap replicates. MrBayes was used to conduct a Bayesian approach to phylogenetic inference by running 20 million generations (10 000 burn-in) with four Metropolis-coupled Markov chain Monte Carlo to optimize efforts to find peaks in tree-space. Parameters were set as: *nst* $= 2$ and *rates* $=$ *invgamma*, and one tree was sampled in every 100. Convergence of tree was checked using Tracer ver.1.3.1 [Rambout and Drummond, 2003] and resulting trees were used to generate a majority consensus tree with posterior probability values. NJ trees based on Kimura-2-parameter with gamma corrected model and based on uncorrected pairwise distance methods were also reconstructed using MEGA ver. 3.1 [Kumar *et al.*, 2004]. Using the same program nodal supports were estimated with 10 000 bootstrap replicates.

We reconstructed unrooted ML, BI, MP and NJ phylogenies among the variants of *C. maenas* and these inferred trees were independently used to tests for selection using PAML program [Yang, 1997]. To account for uncertainty regarding the true topology, we repeated the tests for positive selection using trees from ML, BI, MP and NJ analyses.

Tests for Selection

To account for among-site variations and to test for positive selection on different codon sites, we performed tests for positive selection using two approaches; (1) ML-based codon substitution analyses [Yang *et al.*, 2000] and (2) random effects likelihood method of Kosakovsky Pond and Frost, 2005.

Table 1
GenBank accession number, species, product, code and amino acid sequence length of the taxon used in the study

Accession number	Species	Product	Code	aa*	Source
AF286081	*Carcinus maenas*	CHH (PO-type) variant 1	CM-PO1	139	Dircksen *et al.*, 2001
AF286082	,,	CHH (PO-type) variant 2	CM-PO-2	139	Dircksen *et al.*, 2001
AF286083	,,	CHH (PO-type) variant 3	CM-PO-3	139	Dircksen *et al.*, 2001
AF286084	,,	CHH (PO-type) variant 4	CM-PO-4	137	Dircksen *et al.*, 2001
AF286085	,,	CHH (PO-type) variant 5	CM-PO-5	137	Dircksen *et al.*, 2001
AF286086	,,	CHH (PO-type) variant 6	CM-PO-6	139	Dircksen *et al.*, 2001
AF286092	,,	CHH (PO-type) variant 11	CM-PO-11	128	Dircksen *et al.*, 2001
AF286093	,,	CHH (PO-type) variant 12	CM-PO-12	125	Dircksen *et al.*, 2001
AF286094	,,	CHH (PO-type) variant 13	CM-PO-13	137	Dircksen *et al.*, 2001
AF286078	,,	CHH (XO-type) variant 1	CM-XO-1	112	Dircksen *et al.*, 2001
AF286079	,,	CHH (XO-type) variant 2	CM-XO-2	112	Dircksen *et al.*, 2001
AF286080	,,	CHH (XO-type) variant 3	CM-XO-3	112	Dircksen *et al.*, 2001
X17596	,,	CHH	CM-XO-5	142	Weidemann *et al.*, 1989
AY372181	*Scylla olivacea*	CHH	SO	141	GenBank
AY180334	*Pachygrapsus marmoratus*	CHH-B (XO-type)	PM-XO-B	150	Toullec *et al.*, 2006
AY094983	,,	CHH	PM	150	GenBank
AY180333	,,	CHH-A (XO-type)	PM-XO-A	150	Toullec *et al.*, 2006
AY180335	,,	CHH-B (PO-type)	PM-PO-B	145	Toullec *et al.*, 2006
DQ176431	*Potamon ibericum*	CHH (XO-type)	PI-XO	147	Toullec *et al.*, 2006
DQ176432	,,	CHH (PO-type)	PI-PO	142	Toullec *et al.*, 2006
AY536012	*Callinectes sapidus*	CHH-1	CS-1	139	Choi *et al.*, 2006
	,,	CHH-2	CS-2	135	GenBank
AF241264	*Bythograea thermydron*	CPRP/cHH	BT	147	Toullec *et al.*, 2002
DQ181791	*Homarus gammarus*	CHH-A	HG-A	134	GenBank

*Amino acid sequence length.

ML-Based Codon Substitution Analyses

We estimated parameters under seven different codon substitution models [Yang, *et al.,* 2000] and their performances were evaluated using likelihood ratio tests (LRTs). LRTs were used to compare models that assume no positive selection ($\omega < 1$) with those that assume positive selection ($\omega > 1$). The seven codon substitution models are: M0 (one-ratio), M1a (nearly neutral), M2a (positive selection), M3 (discrete), M7 (β- distribution; $0 \leqslant \omega \leqslant 1$), M8 ($\beta + \omega > 1$: continuous) [Yang *et al.*, 2000] and M8a ($\beta + \omega = 1$) [Swanson *et al.*, 2003]. The M1a model estimates a single parameter p_0, the frequency of conserved sites with $\omega_0 = 0$ and the remaining sites with frequency p_1 ($p_1 = 1 - p_0$) assuming $\omega_1 = 1$. The M2a model adds a class of positively selected sites with frequency p_2 (where $p_2 = 1 - p_1 - p_0$), with ratio ω_2 estimated from the data. Thus, whereas M1a estimates a single parameter (p_0), M2a estimates three parameters (p_0, p_1, and ω_2). In the M7 model, ω follows a beta distribution such that $0 \leqslant$ u $\omega \leqslant 1$ and the two parameters (p and q) of the beta distribution are estimated from the data. In the M8 model, a proportion p_0 of sites have ω drawn from the beta distribution. The remaining sites with proportion p_1 are positively selected and have $\omega_1 > 1$. Thus, M7 estimates two parameters (p and q), while M8 estimates four parameters (p, q, p_0, and ω_1).

The LRTs between nested models were conducted by comparing twice the difference in log-likelihood values ($2ln\Delta l$) against a χ^2- distribution, with degrees of freedom equal to the difference in the number of parameters between models [Yang *et al.*, 2000]. Five LRTs were conducted. The first comparison was made between M0, a model that fits a single ω for all sites with M2a, which allows three site classes ($0 < \omega < 1$, $\omega = 1$ or $\omega > 1$). The second comparison was between M0 and M3. The third comparison was between M1a, which allows for two site classes ($0 < \omega < 1$, $\omega = 1$) with M2a. The fourth comparison

was between M7, a model of beta-distributed selective pressures that allows for 10 site classes, each with $\omega < 1$ and M8, which has 11 site classes, one of them allowed for $\omega > 1$. The last comparison was between M8 and M8a, in which an additional parameter was constrained to have $\omega = 1$ [Swanson *et al.*, 2003]. In all LRTs good evidence for positive selection is found if the LRT indicates that models that allow for positive selection (i.e. M2a and M8) are significantly better than their respective null models (M1a, M7 and M8a) without positive selection ($\omega < 1$) [Yang *et al.*, 2000]. M1a-M2a comparison is a simplified version with fewer parameters than M7-M8, and M8-M8a comparisons.

Random Effects Likelihood (*REL*)

The REL method of Kosakovsky Pond and Frost, 2005, implemented in HyPhy package (http://www.datamonkey.org), was used to detect codon sites that are under positive and negative selection. We applied a hierarchical model selection procedure [Kosakovsky Pond and Frost, 2005] to choose a model of nucleotide substitution. F81 (matrix: 111111) was selected as the optimal nucleotide substitution model using the implementation in the HyPhy package. Independent rate distribution for d_N and d_S were obtained by crossing MG94 [Muse and Gaut, 1994] model and the nucleotide substitution model fitted to the data [Kosakovsky Pond and Muse, 2005]. Using these parameters with an initial NJ tree, the program computed two Bayes factors, one for the event that $d_N < d_S$ at that site (negative selection), and another for the event that $d_N > d_S$ (positive selection). When these Bayes factors are sufficiently large (100 or more), the sites are said to be selected [Kosakovsky Pond and Frost, 2005]. The REL method is least conservative than the other methods implemented in HyPhy package [Kosakovsky Pond and Frost, 2005]. However, since our sample size is less than 25, we performed REL analyses to detect positively and negatively selected sites in CHH variants of *C. maenas*.

Protein Structure of C. maenas CHH

We used Geno3D version 2 (http://geno3d-pbil.ibcp.fr), an automated web server for protein molecular modeling [Combet *et al.*, 2002] to predict the structure of XO-type CHH of *C. maenas*. The MIH crystal structure of kuruma prawn (Protein Data Bank number = 1J0T; Katayama *et al.*, 2003; http://www.rcsb.org/pdb/) was used as reference point to predict the CHH structure of *C. maenas*. The structure was displayed and positively selected sites were mapped using Chimera software (available at http://www.cgl.ucsf.edu/chimera, Pettersen *et al.*, 2004). Amino acid variable sites among 12 CHH variants of *C. maenas* were detected using MEGA 3.1 [Kumar *et al.*, 2004].

RESULTS

Phylogenetic Analyses

Phylogenetic analyses revealed that both XO-type and PO-type variants of *C. maenas* CHH are monophyletic (Fig. 1) with strong nodal support, thus allowing us to perform subsequent codon substitution analyses to detect positively selected sites. The CHH gene tree also revealed that *Callinectes sapidus*, *Scylla olivacea* and *Carcinus maenas* shared a common ancestor. However, the divergence of *Scylla olivacea* and *C. sapidus* occurred after the divergence of *C. maenas* from *Scylla olivacea* and *C. sapidus* (Fig. 1). It appeared that both types of CHH in *C. maenas* diverged from their common ancestor relatively more recently. Although all the four variants of CHH in *Pachygrapsus marmoratus* formed a

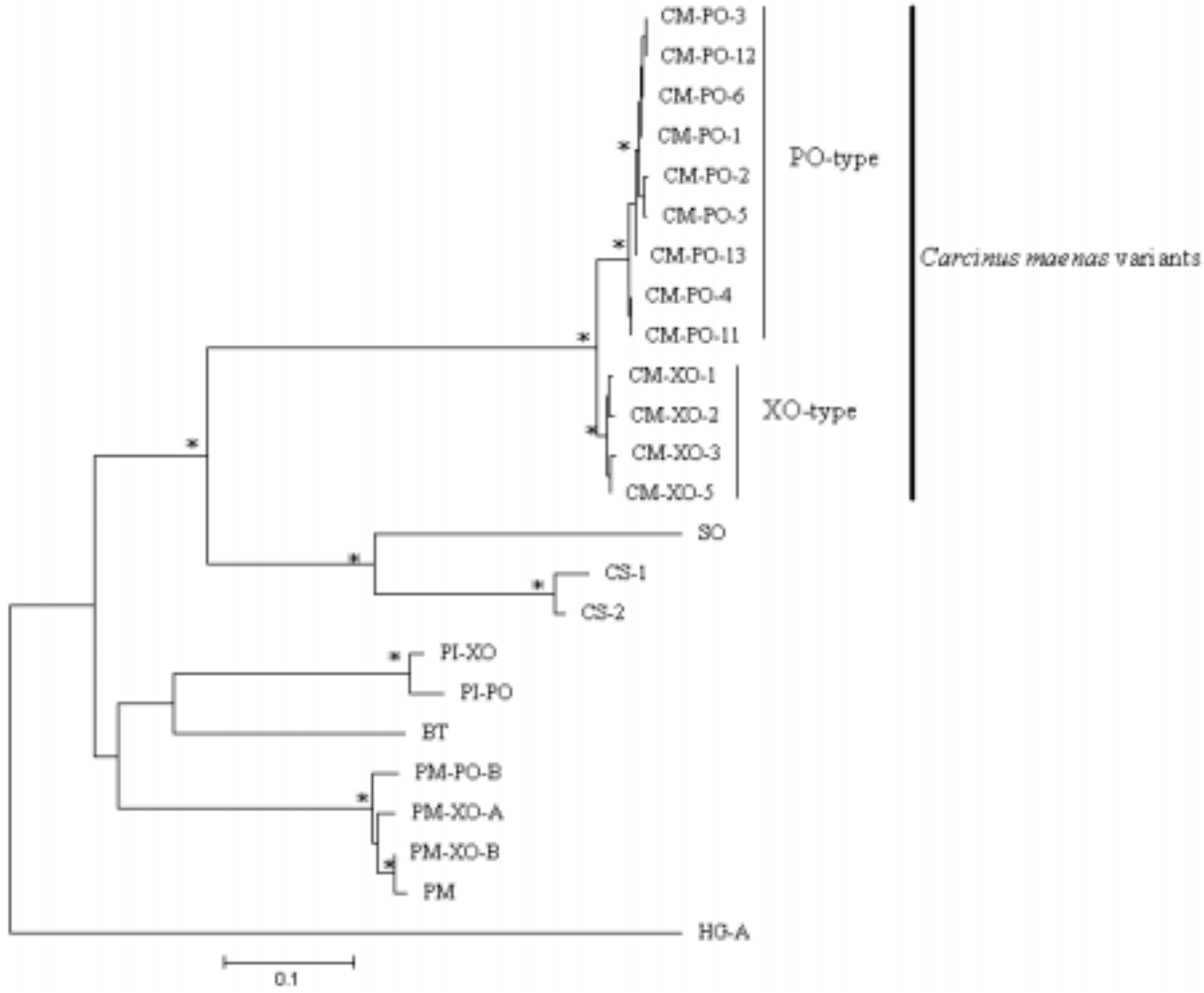

Fig. 1. Neighbor joining tree inferred from CHH variants of six crab species. The tree is rooted with CHH-A of lobster *Homarus gammarus* (GenBank No. **DQ181791**). Nodal support (ML/BI/MP/NJ) $\geq$ 70 is indicated by asterisk. CM: *Carcinus maenas* (GenBank No. **AF286078** – **AF286086**, **AF286092** – **AF286094**, **X17596**), CS: *Callinectes sapidus* (**AY536012**, **DQ343306**), PM: *Pachygrapsus marmoratus* (**AY180333** – **AY180335**, **AY094983**), SO: *Scylla olivacea* (**AY372181**), PI: *Potamon ibericum* (**DQ176431**, **DQ176432**), BT: *Bythograea thermydron* (**AF241264**).

monophyletic group, a relatively very small sample size (< 8), did not allow us to perform codon substitution analyses. Simulation studies showed that a minimum of 8 sequences are required for maximum likelihood codon substitution analyses [Anisimova *et al.*, 2001; Anisimova *et al.*, 2002; Wong *et al.*, 2004].

Tests for Selection

ML-Based Codon Substitution Analyses

Parameter estimates and log likelihood values under models of variable ω among codon sites and their LRTs are shown in Table 2a and 2b, respectively. Under the simplest model, which allows for only a single ω across all codon sites (M0), the ML estimate of ω was 0.5096. This estimate is statistically

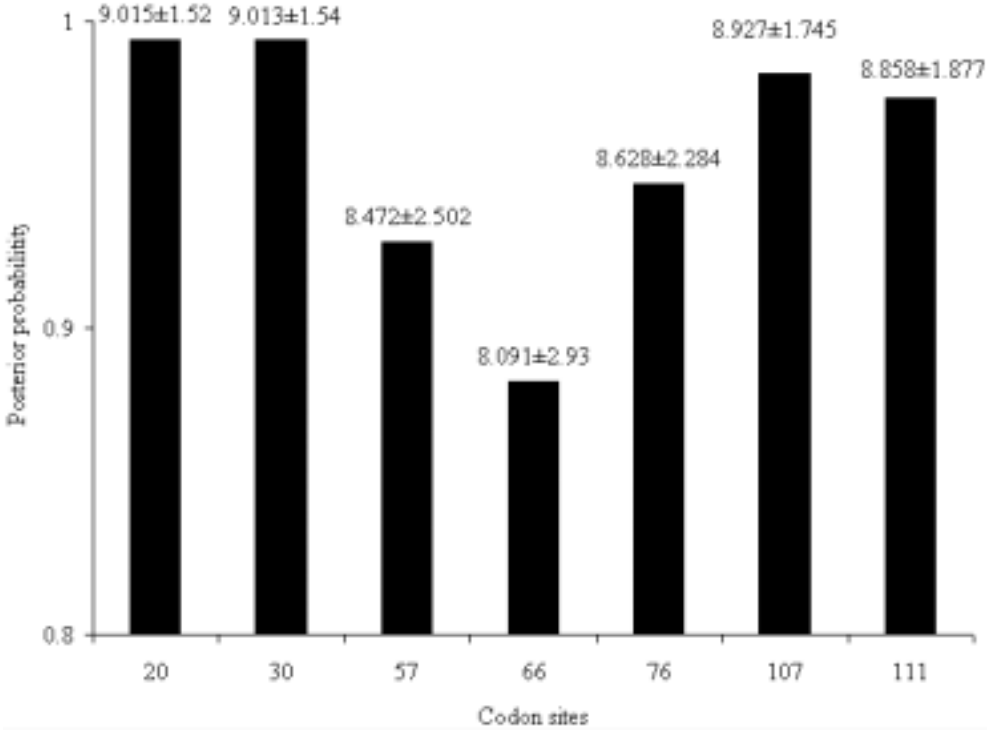

Fig. 2. Posterior probabilities for each codon of CHH (M2a) of *Carcinus maenas*. The ω values with standard errors are shown for each site.

indistinguishable from $\omega = 1$, the expected value under a completely neutral model of sequence evolution. However, a model allowing for variation among sites (M3) provides a significantly better fit to the data showing that there is variation among sites in the strength of selection. However, M0 is a highly unrealistic model and the M0/M3 comparison thus provides a test for variation in ω among sites rather than variation in strength of selection among sites [Yang *et al.*, 2000]. Comparison of models M7 and M8 is a more stringent test to detect positive selection [Anisimova *et al.*, 2001]. The LRT comparing M7 and M8 provides strong evidence for positive selection (Table 2b; $p = 0$). Similar results were obtained using models M1a and M2a. There is a large degree of overlap in the positively selected sites identified from models M2a and M8. [Yang, 1997], (in PAML ver. 3.15) reported that the M1a/M2a comparison seems more robust than the M7/M8 pair. If the true null model assumes several classes of conserved sites with $\omega < 1$ as well as neutral sites with $\omega = 1$, the M7/M8 comparison may often be significant, and among half of such cases, the ω estimate under M8 will be >1, and will produce false positives [Yang, 1997] (PAML ver. 3.15). In such cases, the M8a/M8 comparison or M1a/M2a comparison may be more robust than M7/M8 comparison. Nevertheless, our analyses showed that all the models revealed consistent results ($p = 0$; Table 2b), indicating evidence of positive Darwinian selection across the codons of CHH variants in *C. maenas*. The posterior probability along with ω of each positively selected codon sites identified by M2a model is shown in Fig. 2. All the positively selected codon sites have ω ranged from 8.091 to 9.015 with posterior probabilities greater than 0.89. The sites (identified by M2a/M8 model) predicted to be subjected to positive Darwinian selection are plotted on the PDB structure (Fig. 3).

Table 2
Maximum likelihood-based codon substitution analyses. (a) Parameter estimates and log-likelihood values under models of variable ω-ratios among sites. (b) Likelihood ratio statistics among different models

Table 2a

Model	Free Parameters	Parameter estimates	Likelihood scores	Positively selected sites[a]
M0: One-ratio	1	$\omega = 0.5096$	−642.239453	None
M1a: Nearly neutral	1	$\omega_0 = 0$, $\omega_1 = 1$, ($p_0 = 0.84$, $p_1 = 0.16$)	−628.765617	Not allowed
M2a: Positive selection	3	$\omega_0 = 0$, $\omega_1 = 1$, $\omega_2 =$ **10.95**; ($p_0 = 0.92$, $p_1 = 0$, $p_2 = 0.08$)	−616.425383	**20**, **30**, 57, 66, **76**, **107**, **111**
[b]M3: discrete	5	$\omega_0 = 0, \omega_1 = 0, \omega_2 = 10.95$, ($p_0$ = 0.38, $p_1 = 0.54$, $p_2 =$ 0.08)	-616.425367	**20**, **30**, **57**, **66**, **76**, **107**,**110**, **111**
M7: β	2	$p = 0.005$, $q = 0.047$	−629.399197	Not allowed
M8: $\beta + \omega_s > 1$	4	$p_0 = 0.92$, $p_1 = 0.08$, $p =$ 0.005, $q = 99$, $\omega = 10.95$	−616.425352	**20**, **30**, 57, 66, **76**, **107**, **111**
M8a: $\beta + \omega_s = 1$	3	$p_0 = 0.84$, $p_1 = 0.16$, $p =$ 0.005, $q = 2.88$, $\omega = 1$	−628.765603	Not allowed

Positively selected sites with posterior probability $\geqslant$95 are in bold, 0.9–0.95 are underlined, 0.8–0.9 in italics, and 0.5–0.8 in plain text. M0: one-ratio ω value is the average for all codon sites, whereas M2a, M3 and M8 ω values are the estimated values for the positively selected codon sites under respective models.
[a]Sites with a posterior probability $> 50\%$ of having $\omega > 1$.
[b]Model detected posterior probabilities based on Naive Empirical Bayes (NEB) analysis.

Table 2b

Comparison	2 Δ l	df[a]	p
M0 vs M2a	51,628	2	0.0000
M0 vs M3	51,628	4	0.0000
M1a vs M2a	24,680	2	0.0000
M7 vs M8	25,948	2	0.0000
M8 vs M8a	24,681	1	0.0000

[a]Degrees of freedom.

REL

The positively and negatively selected codon sites detected by REL method are shown in Table 3a and 3b, respectively. Based on this method there are 30 amino acid sites in CHH variants of *C. maenas* are positively selected with posterior probability = 100 (Table 3a), whereas only 4 sites are negatively selected (Table 3b). All the codon sites that were positively selected by ML-based codon substitution models were also positively selected by REL methods.

Positively selected amino acid sites

The positively selected amino acid sites by M2a/M8a model are mapped in the predicted 3D-protein structure of XO-type (Fig. 3). While serine, a less polar amino acid, is found at both sites (107 and 111) in XO-type, the polar amino acids asparagine and glutamic acid are unique to positions 107 and 111 in PO-type, respectively (Figs 2 and 3). We speculate that these changes at amino acid sites 107 and 111 in PO-type might have caused the loss of CHH activity. However, further study is needed to validate this hypothesis.

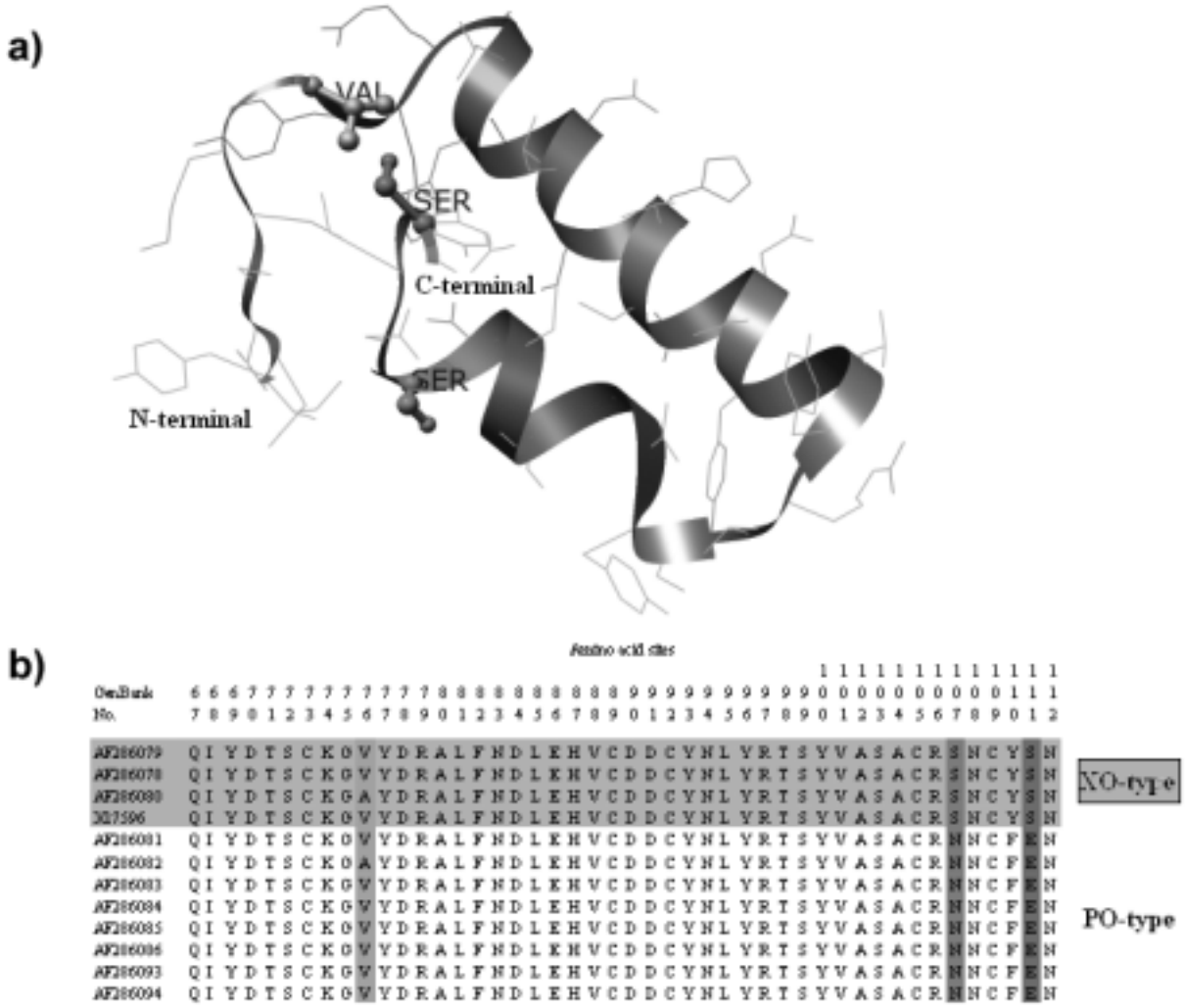

Fig. 3. The three-dimensional structure of *Carcinus maenas* XO-type CHH predicted based on molting inhibiting hormone PDB structure (Protein Data Bank number: 1J0T, http://www.rcsb.org/pdb/). Sites shown are those sites predicted to be under positive selection with posterior probability >95%. Ribbon colors in Fig. 3a are corresponding to the amino acid sites positively selected sites highlighted with same color in Fig. 3b.

DISCUSSION

Phylogenetic analyses revealed that PO-type and XO-type are monophyletic for *C. maenas*. It could be possible that these two groups in *C. maenas* might have diverged from a common ancestor through gene duplication. Chan *et al.*, 2003, also postulated that the neuropeptides of CHH/MIH gene families diverged as a result of mutations and gene duplication events. Dircksen *et al.*, 2001, reported XO-type and PO-type CHH in *C. maenas* to have differential function. In contrast to XO-type, PO-type neither exhibits hyperglycemic activity, nor does it inhibit Y-organ ecdysteroid synthesis [Dircksen *et al.*, 2001]. Thus, the question of what factors are responsible to promote functional diversification of these neuropeptides remains unclear. We suggest that the positive Darwinian selection (adaptive evolution) is a likely cause for the accelerated rate of amino acid substitutions in the CHH variants of

Table 3
Positively and negatively selected codon sites identified by REL method implemented in HyPhy package. (a) Positively selected sites. (b) Negatively selected sites

Codon	$E[d_N - d_S]$	Posterior probability $[d_N > d_S]$	Bayes factor $[d_N > d_S]$
(a)			
20	14.863	1	1.00E + 26
30	14.0969	1	1.00E + 26
57	14.0476	1	1.00E + 26
66	13.8381	1	1.00E + 26
76	14.5913	1	1.00E + 26
107	12.4321	1	1.00E + 26
110	8.40389	1	3.83E + 10
111	13.2026	1	1.00E + 26
113	13.8242	1	1.00E + 26
116	12.2199	1	1.00E + 26
117	9.65145	1	1.00E + 26
119	14.0375	1	2.05E + 11
120	13.9851	1	3.40E + 11
121	13.3325	1	1.00E + 26
123	13.2951	1	1.00E + 26
124	13.1607	1	1.00E + 26
125	12.1609	1	1.00E + 26
126	13.672	1	1.69E + 11
127	13.249	1	1.00E + 26
128	12.812	1	1.00E + 26
129	13.3127	1	1.15E + 11
130	13.3931	1	1.00E + 26
131	12.1786	1	1.00E + 26
132	14.1751	1	1.00E + 26
133	13.4525	1	9.59E + 10
134	13.8724	1	1.74E + 11
135	12.7321	1	3.61E + 11
136	9.03004	1	1.00E + 26
137	12.476	1	3.75E + 11
138	12.6408	1	1.00E + 26
(b)			
10	−10.0919	0.0293342	50.7165
24	−10.1194	0.0275895	54.0206
109	−10.0893	0.0243621	61.3802
112	−10.0416	0.0242983	61.5453

C. maenas. Irrespective of variant types (XO or PO), the majority of the amino acid sequences in *C. maenas* variants are longer (112–139 aa) than those of the other isoforms in the CHH/MIH/GIH family (72–87aa; 8 to 11kDa) [Chen *et al.*, 2005]. Dircksen *et al*., 2001, proposed that these discrepancies in the length of amino acid sequences among all the variants of *C. maenas* are possibly the outcome of alternative splicing. Nevertheless, both PO and XO types are reported to have differential function [Dircksen *et al*., 2001; Fanjul-Moles, 2006]. Thus, it is unequivocal that positive Darwinian selection is the likely explanation for the accelerated amino acid substitution rate. Although our analyses showed discrepancies in the positively selected sites identified by REL and ML-based codon substitution models, both methods are consistent with the fact that CHH variants in *C. maenas* are under the influence of positive Darwinian selection, therefore indicating rapid functional diversification (adaptive evolution). However, codons that are identified as positively selected sites by ML-codon substitution analyses [Yang *et al.*, 2000] are also identified by REL method. Nevertheless, our results showed evidence of adaptive

Table 4
Amino acid variable sites in CHH variants of *Carcinus maenas*

Amino acid variable sites																													
						1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	2	3	5	6	7	0	1	1	1	1	1	1	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3
Code	0	0	7	6	6	7	0	1	3	6	7	9	0	1	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8
CM-PO1	A	P	H	R	V	N	F	E	E	D	V	V	Y	E	Y	F	P	D	H	E	E	Y	L	R	S	R	D	G	L	K
CM-PO-2	.	Q	.	K	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CM-PO-3	P	.	.	K	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CM-PO-4	T	Q	.	.	.	.	.	.	.	.	.	.	.	Q	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	-
CM-PO-5	.	Q	.	K	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-
CM-PO-6	.	.	.	K	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CM-PO-13	.	Q	N	.	.	.	.	.	.	.	.	.	.	Q	H	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-
CM-PO-12	P	.	.	K	.	.	.	.	.	.	.	.	.	.	.	.	.	-	-	-	-	-	-	-	-	-	-	-	-	-
CM-XO-1	.	Q	.	K	.	S	Y	S	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CM-XO-2	P	.	.	K	.	S	Y	S	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CM-XO-3	.	Q	.	.	A	S	Y	S	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CM-XO-5	.	Q	N	.	.	S	Y	S	L	R	Q	M	D	D	L	M	M	.	E	F	D	Q	Y	A	R	K	V	Q	M	V

Positively selected sites (M2a; Tab. 2a) are shaded. Amino acid sites 1 - 26: signal peptide, 27- 64: CPRP (corazonin-precursor-related peptide), 65- 66 cleavage sites, and 67 - 138: mature peptide.

evolution. This line of evidence is consistent with the evolutionary patterns that are seen in prolactin and growth hormone gene families in vertebrates. The duplicated genes in prolactin [Ohta, 1993; Wallis *et al*., 2005; Kitano *et al.*, 2006] and growth hormone [Wallis and Wallis, 2001; Wallis *et al.*, 2001] gene family are also under the influence of positive Darwinian selection.

Thus, our results present the convincing evidence that multiple copies of CHH in *C. maenas* have evolved by gene duplication, and positive Darwinian selection (adaptive evolution) is the likely cause of accelerated rate of amino acid substitutions among these duplicated genes.

ACKNOWLEDGEMENT

We thank the University of Tulsa for providing facilities to carry out this work.

REFERENCES

- Anisimova, M., Bielawski, J. P. and Yang, Z. (2001). Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. Mol. Biol. Evol. **18**, 1585-1592.
- Anisimova, M., Bielawski, J. P. and Yang, Z. (2002). Accuracy and power of Bayes prediction of amino acid sites under positive selection. Mol. Biol. Evol. **19**, 950-958.
- Chan, S.-M., Gu, P.-L., Chu, K.-H. and Tobe, S. S. (2003). Crustacean neuropeptide genes of the CHH/MIH/GIH family: implications from molecular studies. Gen. Comp. Endocrinol. **134**, 214-219.
- Chen, S. H., Lin, C. Y. and Kuo, C. M. (2005). *In silico* analysis of crustacean hyperglycemic hormone family. Mar. Biotechnol. **7**, 193-206.

- Choi, C. Y., Zheng, J. and Watson, R. D. (2006). Molecular cloning of a cDNA encoding a crustacean hyperglycemic hormone from eyestalk ganglia of the blue crab, *Callinectes sapidus*. Gen. Comp. Endocrinol. **148**, 383-387.
- Chung, J. S. and Webster, S. G. (2004). Expression and release patterns of neuropeptides during embryonic development and hatching of the green shore crab, *Carcinus maenas*. Development **131**, 4751-4761.
- Chung, J. S., Dircksen, H. and Webster, S. G. (1999). A remarkable, precisely timed release of hyperglycemic hormone from endocrine cells in the gut is associated with ecdysis in the crab *Carcinus maenas*. Proc. Natl. Acad. Sci. USA **96**, 13103-13107.
- Combet, C., Jambon, M., Deléage, G. and Geourjon, C. (2002). Geno3D: automatic comparative molecular modelling of protein. Bioinformatics **18**, 213-214.
- Dircksen, H. and Heyn, U. (1998). Crustacean hyperglycemic hormone-like peptides in crab and locust peripheral intrinsic neurosecretory cells. Ann. N. Y. Acad. Sci. **839**, 392-394.
- Dircksen, H., Böcking, D., Heyn, U., Daufeldt, S., Baggerman, G., Verhaert, P., Chung, J. S., Soyez, D., Plösch, T., Jaros, P. P. and Webster, S. G. (2000). Crustacean hyperglycaemic hormone-like peptides in similar peripheral and central neurosecretory cells of crustaceans and insects as spliced products of multiple genes. Zoology **103** (suppl. III), 81.
- Dircksen, H., Böcking, D., Heyn, U., Mandel, C., Chung, J. S., Baggerman, G., Verhaert, P., Daufeldt, S., Plösch, T., Jaros, P. P., Waelkens, G., Keller, R. and Webster, S. G. (2001). Crustacean hyperglycemic hormone (CHH)-like peptides and CHH-precursor-related peptides from pericardial organ neurosecretory cells in the shore crab, *Carcinus maenas*, are putatively spliced and modified products of multiple genes. Biochem. J. **356**, 159-170.
- Fanjul-Moles, M. L. (2006). Biochemical and functional aspects of crustacean hyperglycemic hormone in decapod crustaceans: Review and update. Comp. Biochem. Physiol. C. Toxicol. Pharmacol. **142**, 390-400.
- Ford, M. J. (2002). Applications of selective neutrality tests to molecular ecology. Mol. Ecol. **11**, 1245-1262.
- Goldstone, H. M. and Stegeman, J. J. (2006). A revised evolutionary history of the CYP1A subfamily: gene duplication, gene conversion, and positive selection. J. Mol. Evol. **62**, 708-717.
- Guindon, S. and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. **52**, 696-704.
- Huelsenbeck, J. P. and Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics **17**, 754-755.
- Hughes, A. L. and Nei, M. (1989). Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. Proc. Natl. Acad. Sci. USA **86**, 958-962.
- Katayama, H., Nagata, K., Ohira, T., Yumoto, F., Tanokura, M. and Nagasawa, H. (2003). The solution structure of molt-inhibiting hormone from the kuruma prawn Marsupenaeus japonicus. J. Biol. Chem. **278**, 9620-9623.
- Keller, R. (1992). Crustacean neuropeptides: structures, functions and comparative aspects. Experientia **48**, 439-448.
- Khayat, M., Yang, W., Aida, K., Nagasawa, H., Tietz, A., Funkenstein, B. and Lubzens, E. (1998). Hyperglycemic hormones inhibit protein and mRNA synthesis in in vitro-incubated ovarian fragments of the marine shrimp Penaeus semisulcatus. Gen. Comp. Endocrinol. **110**, 307-318.
- Kimura, M. (1983). The neutral theory of molecular evolution. New York: Cambridge University Press.
- Kitano, T., Tian, W., Umetsu, K., Yuasa, I., Yamazaki, K., Saitou, N. and Osawa, M. (2006). Origin and evolution of gene for prolactin-induced protein. Gene **383**, 64-70.
- Kosakovsky Pond, S. L. and Frost, S. D. W. (2005). Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics **21**, 2531-2533.
- Kosakovsky Pond, S. and Muse, S. V. (2005). Site-to-site variation of synonymous substitution rates. Mol. Biol. Evol. **22**, 2375-2385.
- Kumar, S., Tamura, K. and Nei, M. (2004). MEGA3.1: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Bioinformatics **5**, 150-163.
- Lacombe, C., Greve, P. and Martin, G. (1999). Overview on the sub-grouping of the crustacean hyperglycemic hormone family. Neuropeptides **33**, 71-80.
- Maddison, W. P. and Maddison, D. R. (2006). Mesquite: a modular system for evolutionary analysis. http:// mesquiteproject.org.
- Muse, S. V. and Gaut, B. S. (1994). A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates with application to the chloroplast genome. Mol. Biol. Evol. **11**, 715-724.
- Nielsen, R. (2001). Statistical tests of neutrality in the age of genomics. Heredity **86**, 641-647.
- Ohno, S. (1970). Evolution by gene duplication. Springer, New York.
- Ohta, T. (1993). Pattern of nucleotide substitutions in growth hormone-prolactin gene family: A paradigm for evolution by gene duplication. Genetics **134**, 1271-1276.
- Page, R. D. M. (1996). TREEVIEW: An application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. **12**, 357-358.
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. and Ferrin, T. E. (2004). UCSF Chimera - A Visualization System for Exploratory Research and Analysis. J. Comput. Chem. **25**, 1605-1612.

- Posada, D. and Crandall, K. A. (1998). Modeltest: testing the model of DNA substitution, Bioinformatics **14**, 817-818.
- Rambout, A. and Drummond, A. J. (2003). Tracer. Version 1.3.1. Available from http://evolve.zoo.ox.ac.uk/.
- Santos, E. A. and Keller, R. (1993). Crustacean hyperglycemic hormone (CHH) and the regulation of carbohydrate metabolism: current perspectives. Comp. Biochem. Physiol. A **106**, 405-411.
- Santos, E. A., Nery, L. E. M., Keller, R. and Gonçalves, A. A. (1997). Evidence for the involvement of the crustacean hyperglycemic hormone in the regulation of lipid metabolism. Physiol. Zool. **70**, 415-420.
- Serrano, L., Blanvillain, G., Soyez, D., Charmantier, G., Grousset, E., Aujoulat, F. and Spanings-Pierrot, C. (2003). Putative involvement of crustacean hyperglycemic hormone isoforms in the neuroendocrine mediation of osmoregulation in the crayfish *Astacus leptodactylus*. J. Exp. Biol. **206**, 979-988.
- Shiu, S.-H., Byrnes, J. K., Pan, R., Zhang, P. and Li, W.-H. (2006). Role of positive selection in the retention of duplicate genes in mammalian genomes. Proc. Natl. Acad. Sci. USA **103**, 2232-2236.
- Swanson, W. J. and Aquadro, C. F. (2002). Positive darwinian selection promotes heterozygosity among members of the antifreeze protein multigene family. J. Mol. Evol. **54**, 403-410.
- Swanson, W. J., Yang, Z., Wolfner, M. F. and Aquadro, C. F. (2001). Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. Proc. Natl. Acad. Sci. USA **98**, 2509-2514.
- Swanson, W. J., Nielsen, R. and Yang, Q. (2003). Pervasive adaptive evolution in mammalian fertilization proteins. Mol. Biol. Evol. **20**, 18-20.
- Swofford, D. L. (2002). PAUP*. Phylogenetic Analysis Using Parsimony (and other methods), version 4. Sinauer Associates, Sunderland, MA.
- Toullec, J.-Y., Vinh, J., Le Caer, J.-P., Shillito, B. and Soyez, D. (2002). Structure and phylogeny of the crustacean hyperglycemic hormone and its precursor from a hydrothermal vent crustacean: the crab *Bythograea thermydron*. Peptides **23**, 31-42.
- Toullec, J.-Y., Serrano, L., Lopez, P., Soyez, D. and Spanings-Pierrot, C. (2006). The crustacean hyperglycemic hormones from an euryhaline crab P*achygrapsus marmoratus* and a fresh water crab *Potamon ibericum*: Eyestalk and pericardial isoforms. Peptides **27**, 1269-1280.
- Wallis, O. C. and Wallis, M. (2001). Molecular evolution of growth hormone (GH) in Cetartiodactyla: cloning and characterization of the gene encoding GH from a primitive ruminant, the chevrotain (*Tragulus javanicus*). Gen. Comp. Endocrinol. **123**, 62-72.
- Wallis, O. C., Zhang, Y.-P. and Wallis, M. (2001). Molecular evolution of GH in primates: characterisation of the GH genes from slow loris and marmoset defines an episode of rapid evolutionary change. Mol. Endocrinol. **26**, 249-258.
- Wallis, O. C., Mac-Kwashie, A. O., Makri, G. and Wallis, M. (2005). Molecular evolution of prolactin in primates. J. Mol. Evol. **60**, 606-614.
- Webster, S. G., Dircksen, H. and Chung, J. S. (2000). Endocrine cells in the gut of the shore crab *Carcinus maenas* immunoreactive to crustacean hyperglycaemic hormone and its precursor-related peptide. Cell Tissue Res. **300**, 193-205.
- Weidemann, W., Gromoll, J. and Keller, R. (1989). Cloning and sequence analysis of cDNA for precursor of a crustacean hyperglycemic hormone. FEBS Lett. **257**, 31-34.
- Wong, W. S. W., Yang, Z., Goldman, N. and Nielsen, R. (2004). Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. Genetics **168**, 1041-1051.
- Xia, X. (2000). Data analysis in molecular biology and evolution. pp276.
- Xia, X. and Xie, Z. (2001). DAMBE: Data analysis in molecular biology and evolution. J. Hered. **92**, 371-73.
- Yang, Z. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. **13**, 555-556.
- Yang, Z. and Bielawski, J. P. (2000). Statistical methods for detecting molecular adaptation. Trends Ecol. Evol. **15**, 496-503.
- Yang, Z., Nielsen, R., Goldman, N. and Pederson, A. M. (2000). Codon-substitution models for heterogeneous selection pressures at amino acid sites. Genetics **155**, 431- 449.
- Zhang, J., Rosenberg, H. F. and Nei, M. (1998). Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proc. Natl. Acad. Sci. USA **95**, 3708-3713.