BMC Evolutionary Biology

BioMed Central

Research article

Ultraconserved coding regions outside the homeobox of mammalian Hox genes

Zhenguo Lin*[1,2], Hong Ma[1,2] and Masatoshi Nei[1,2]

Address: [1]Department of Biology and Institute of Molecular Evolutionary Genetics, Pennsylvania State University, University Park, PA 16802, USA and [2]Huck Institutes of Life Sciences, Pennsylvania State University, University Park, PA 16802, USA

Email: Zhenguo Lin* - zul102@psu.edu; Hong Ma - hxm16@psu.edu; Masatoshi Nei - nxm2@psu.edu

* Corresponding author

Published: 24 September 2008

BMC Evolutionary Biology 2008, **8**:260 doi:10.1186/1471-2148-8-260

Received: 27 May 2008

Accepted: 24 September 2008

This article is available from: http://www.biomedcentral.com/1471-2148/8/260

Abstract

Background: All bilaterian animals share a general genetic framework that controls the formation of their body structures, although their forms are highly diversified. The Hox genes that encode transcription factors play a central role in this framework. All Hox proteins contain a highly conserved homeodomain encoded by the homeobox motif, but the other regions are generally assumed to be less conserved. In this study, we used comparative genomic methods to infer possible functional elements in the coding regions of mammalian Hox genes.

Results: We identified a set of ultraconserved coding regions (UCRs) outside the homeobox of mammalian Hox genes. Here a UCR is defined as a region of at least 120 nucleotides without synonymous and nonsynonymous nucleotide substitutions among different orders of mammals. Further analysis has indicated that these UCRs occur only in placental mammals and they evolved apparently after the split of placental mammals from marsupials. Analysis of human SNP data suggests that these UCRs are maintained by strong purifying selection.

Conclusion: Although mammalian genomes are known to contain ultraconserved non-coding elements (UNEs), this paper seems to be the first to report the UCRs in protein coding genes. The extremely high degree of sequence conservation in non-homeobox regions suggests that they might have important roles for the functions of Hox genes. We speculate that UCRs have some gene regulatory functions possibly in relation to the development of the intra-uterus child-bearing system.

Background

An unexpected feature of mammalian genomes is that they contain a large number of ultraconserved DNA elements [1]. These elements have been shown to be under strong purifying selection, and therefore they are believed to have some important biological functions [2]. The specific functions of these elements and the mechanism that led to formation of these regions remain unclear. Some studies have suggested that these regions may play a role in the regulation of their neighboring developmental genes [3,4]. These ultraconserved elements have been identified almost exclusively from noncoding regions of the genome.

During the course of studying DNA sequence divergence of Hox genes among different mammalian orders, we noticed that many ultraconserved regions exist in the protein coding regions outside the homeobox. Hox genes encode a group of transcription factors that control the segmentation identities of developing animal embryos along the head-to-tail axis. These proteins contain a domain called the homeodomain encoded by the homeobox motif. The amino acid sequences of the homeodomain have been conserved even between mammals and insects [5]. At the nucleotide level, however, synonymous nucleotide substitutions occur with reasonably high frequencies [6]. Therefore, the homeobox motifs are not ultraconserved regions.

Hox genes tend to be organized into gene clusters, and there is a striking correlation between the order of Hox genes in the cluster and the spatial patterns of their expression in the developing embryo. The Hox genes at the 3' end of the cluster are expressed in the anterior regions of the embryo and those genes at the 5'end are expressed in the posterior regions. This cluster organization and the expression pattern of Hox genes are highly conserved from arthropods to mammals [7]. Multiple duplication events of these gene clusters have led to a significant expansion of the Hox gene family in vertebrates. As a result, four Hox clusters (denoted as HoxA, HoxB, HoxC and HoxD) are present on separate chromosomes in mammalian species. According to the position in the clusters, the Hox genes in the four clusters can be classified into thirteen cognate (orthologous gene) groups. However, some members of these cognate groups have been lost, and only 39 Hox genes are currently present in the human and other mammalian genomes (Fig. 1A).

Although the homeodomain is highly conserved, the sequences outside the homeodomain in Hox proteins are generally quite divergent and do not contain conserved domains except some small motifs such as the MXSXFE motif at the N-terminus and the YPWM motif near the homeodomain. Some studies have been conducted on these non-homeodomain regions of *Drosophila* Hox genes. For example, the C-terminal region of the *Drosophila Ubx* protein was shown to serve as a repressor domain, and sequence changes in this region are correlated with limb pattern differences between insects and crustaceans [8,9]. Another study on the *Drosophila Abd-B* gene indicates that protein domains outside the homeodomain influence the activation or repression of target gene expression [10]. These studies suggest that protein regions outside the homeodomain of other Hox proteins might also have some effects on embryonic development and morphological evolution. In this study, we examined the rates of nucleotide substitutions in different coding regions of Hox genes to identify potentially important functional elements. Interestingly, we found that many ultraconserved regions are present between orthologous mammalian Hox genes, and most of them are located outside the homeobox motif, indicating that they are probably important for the functions of Hox genes.

Results and discussion

Ultraconserved Coding Regions (UCRs) present in many mammalian Hox genes

A preliminary sliding window analysis was performed to study the synonymous and nonsynonymous nucleotide substitutions between orthologous Hox genes from human, dog and mouse to identify conserved coding regions. We then found that some coding regions are ultraconserved between different species. Neither nonsynonymous nor synonymous substitutions were observed in these regions, which may include as many as hundreds of nucleotide sites (Fig. 2 and 3). Because these species have diverged over 90 million years ago [11,12], it is unlikely to observe these conserved regions if the synonymous substitutions are free from natural selection [13]. Considering that the nucleotide identity between the human and mouse genomes is about 40% [14], the probability of observing two identical sequences with more than 100 consecutive nucleotide sites by chance will be = $(0.4)^{-100}$. Therefore, it appears that the synonymous sites as well as the nonsynonymous sites in these regions have been highly conserved.

In the present study, we defined an Ultraconserved Coding Region (UCR) as a region with at least 120 nucleotides (40 codons) that do not contain any synonymous substitutions between two distantly related species (divergence time ~90 million years or more). Using 120 nucleotides as the threshold is not based on considerations of biological functions but merely statistical convenience. It can be shown that the probability of occurrence of a UCR by chance is extremely small if this criterion is applied.

To detect the UCRs in the mammalian lineages, we performed a sliding window analysis of synonymous and nonsynonymous nucleotide substitutions for each pair of Hox orthologous genes from six species (human, dog, cow, mouse, opossum and platypus) representing six different mammalian orders (Fig 1B). We found that the UCRs in Hox genes are frequently detected in placental mammals (Table. 1 and Additional file 1). For example, 33 UCRs were detected in 21 Hox genes in the dog/cow comparison (Table 1). We also found 26 UCRs in 19 Hox orthologous genes between human and dog, 23 UCRs in 19 Hox genes in the human/cow comparison, and 14 UCRs in 13 Hox genes in the human/mouse comparison. In summary, at least 12 UCRs were found in the Hox genes in each pairwise comparison of placental mammals. Furthermore, many of these UCRs were found repeatedly in different pairwise comparisons (see Additional file 1), indicating that these regions have been conserved throughout the evolution of placental mammals. Consid-

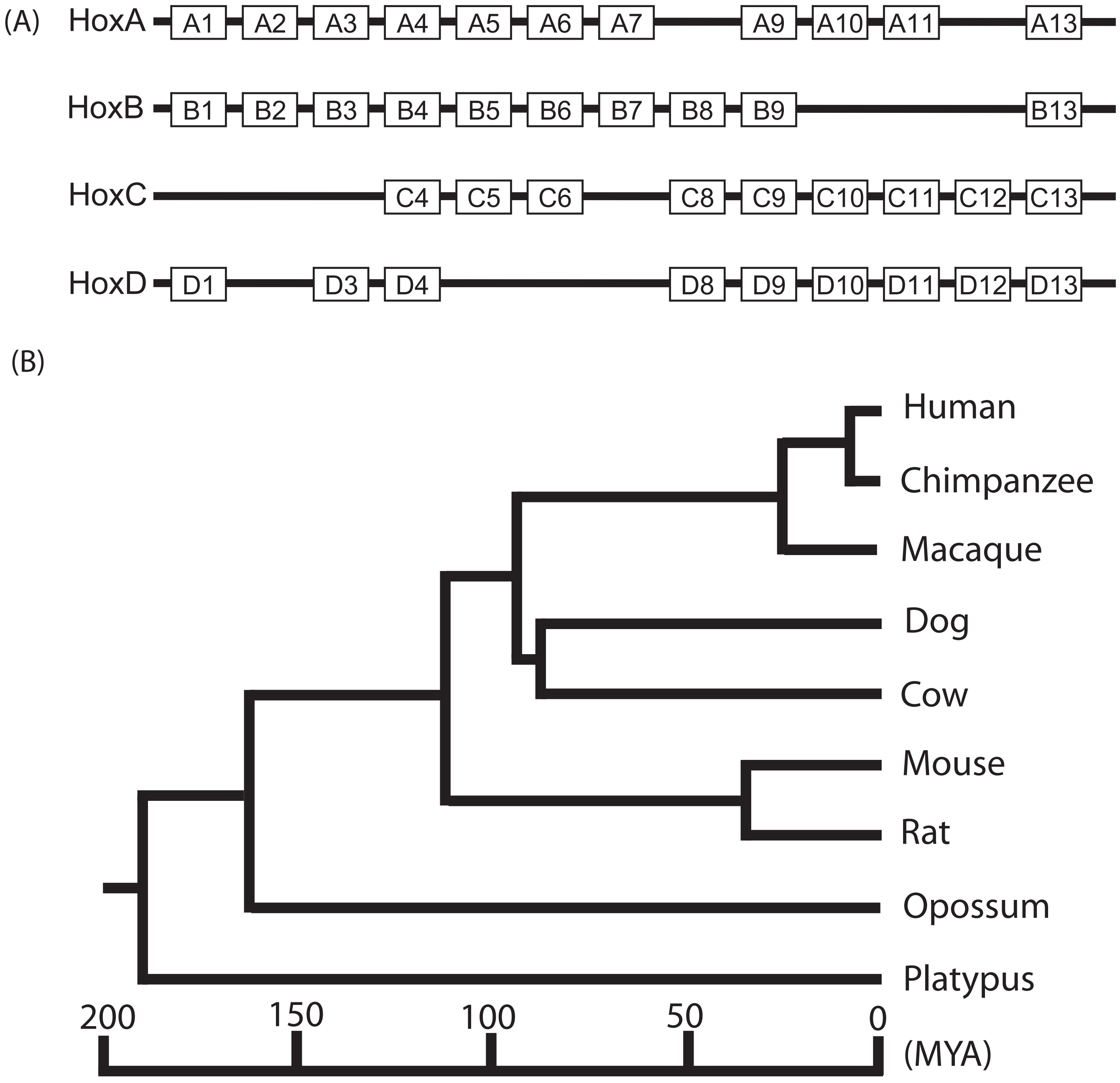
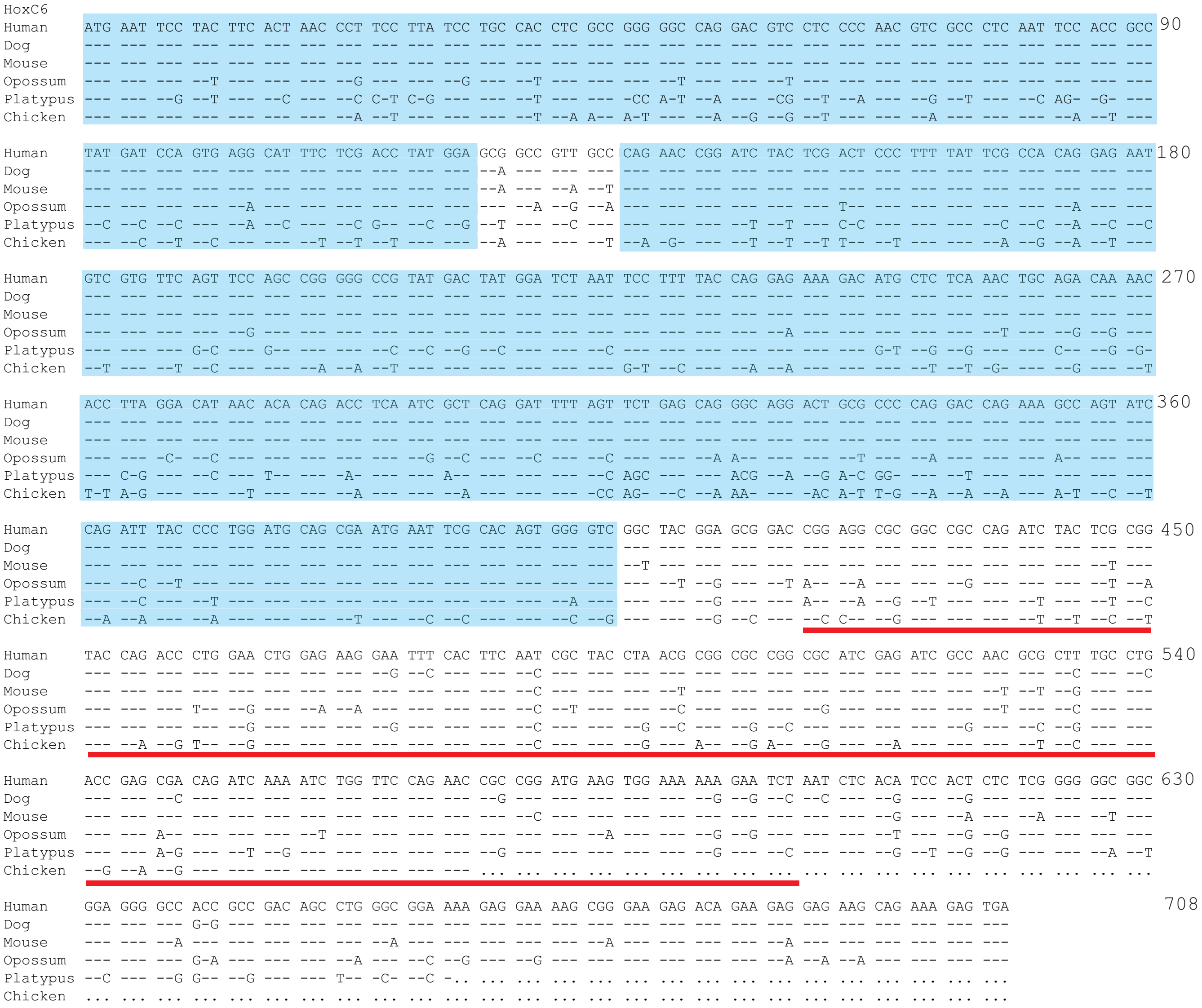


Figure 1

A. Diagram showing the chromosomal organization of Hox genes in human. Each horizontal thick line represents a gene cluster, with cluster name shown at the left side. Clusters are shown from 3' end to 5' end. The 13 cognate gene groups are defined vertically across the clusters and consisting of up to four genes, each on a different chromosome. Some genes in each cognate group have been lost during evolution. B. A phylogenetic tree illustrating the evolutionary relationships and approximate divergence times of representative mammalian species. The branch lengths represent the estimated divergence times. The divergence times were obtained from the TimeTree database[50]. MYA: million years ago.

ering that the divergence time between dog and cow is less than 90 million years and their Hox sequences are highly similar, we decided not to include the UCRs that are only found in the two species for further analysis.

In contrast, the number of UCRs is significantly reduced in the comparison between a placental mammal and the marsupial opossum (*Monodelphis domestica*) or duck-billed platypus (*Ornithorhynchus anatinus*). This paucity of



```
HoxC6
Human    ATG AAT TCC TAC TTC ACT AAC CCT TCC TTA TCC TGC CAC CTC GCC GGG GGC CAG GAC GTC CTC CCC AAC GTC GCC CTC AAT TCC ACC GCC 90
Dog      --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Mouse    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Opossum  --- --- --- --T --- --- --- --G --- --- --G --- --T --- --- --- --T --- --- --T --- --- --- --- --- --- --- --- --- ---
Platypus --- --- --G --T --- --C --- --C C-T C-G --- --- --T --- --- -CC A-T --A --- -CG --T --A --- --G --T --- --C AG- -G- ---
Chicken  --- --- --- --- --- --- --- --A --T --- --- --- --T --A A-- A-T --- --A --G --G --T --- --- --A --- --- --- --A --T ---

Human    TAT GAT CCA GTG AGG CAT TTC TCG ACC TAT GGA GCG GCC GTT GCC CAG AAC CGG ATC TAC TCG ACT CCC TTT TAT TCG CCA CAG GAG AAT 180
Dog      --- --- --- --- --- --- --- --- --- --- --- --A --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Mouse    --- --- --- --- --- --- --- --- --- --- --- --A --- --A --T --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Opossum  --- --- --- --- --A --- --- --- --- --- --- --- --A --G --A --- --- --- --- --- --- T-- --- --- --- --- --- --A --- ---
Platypus --C --C --C --- --A --C --- --C G-- --C --G --T --- --C --- --- --- --- --T --T --- C-C --- --- --- --C --C --A --C --C
Chicken  --- --C --T --C --- --- --T --T --T --- --- --A --- --- --T --A -G- --- --T --T --T T-- --T --- --- --A --G --A --T ---

Human    GTC GTG TTC AGT TCC AGC CGG GGG CCG TAT GAC TAT GGA TCT AAT TCC TTT TAC CAG GAG AAA GAC ATG CTC TCA AAC TGC AGA CAA AAC 270
Dog      --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Mouse    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Opossum  --- --- --- --- --G --- --- --- --- --- --- --- --- --- --- --- --- --- --- --A --- --- --- --- --- --T --- --G --G ---
Platypus --- --- --- G-C --- G-- --- --- --C --C --G --C --- --- --C --- --- --- --- --- --- --- G-T --G --G --- --- C-- --G -G-
Chicken  --T --- --T --C --- --- --A --A --T --- --- --- --- --- --- G-T --C --- --A --A --- --- --- --T --T -G- --- --G --- --T

Human    ACC TTA GGA CAT AAC ACA CAG ACC TCA ATC GCT CAG GAT TTT AGT TCT GAG CAG GGC AGG ACT GCG CCC CAG GAC CAG AAA GCC AGT ATC 360
Dog      --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Mouse    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Opossum  --- --- -C- --C --- --- --- --- --- --G --C --- --C --- --C --- --- --A A-- --- --- --T --- --A --- --- --- A-- --- ---
Platypus --- C-G --- --C --- T-- --- -A- --- --- A-- --- --- --- --C AGC --- --- ACG --A --G A-C GG- --- --T --- --- --- --- ---
Chicken  T-T A-G --- --- --T --- --- --A --- --- --A --- --- --- -CC AG- --C --A AA- --- -AC A-T T-G --A --A --A --- A-T --C --T

Human    CAG ATT TAC CCC TGG ATG CAG CGA ATG AAT TCG CAC AGT GGG GTC GGC TAC GGA GCG GAC CGG AGG CGC GGC CGC CAG ATC TAC TCG CGG 450
Dog      --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Mouse    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --T --- --- --- --- --- --- --- --- --- --- --- --- --T ---
Opossum  --- --C --T --- --- --- --- --- --- --- --- --- --- --- --- --- --T --G --- --T A-- --A --- --- --G --- --- --- --T --A
Platypus --- --C --- --T --- --- --- --- --- --- --- --- --- --A --- --- --- --G --- --- A-- --A --G --T --- --- --T --- --T --C
Chicken  --A --A --- --A --- --- --- --T --- --C --C --- --- --C --G --- --- --G --C --- --C C-- --G --- --- --- --T --T --C --T

Human    TAC CAG ACC CTG GAA CTG GAG AAG GAA TTT CAC TTC AAT CGC TAC CTA ACG CGG CGC CGG CGC ATC GAG ATC GCC AAC GCG CTT TGC CTG 540
Dog      --- --- --- --- --- --- --- --- --G --C --- --- --C --- --- --- --- --- --- --- --- --- --- --- --- --- --- --C --- --C
Mouse    --- --- --- --- --- --- --- --- --- --- --- --- --C --- --- --- --T --- --- --- --- --- --- --- --- --T --T --G --- ---
Opossum  --- --- --- T-- --G --- --A --A --- --- --- --- --C --T --- --- --C --- --- --- --G --- --- --- --- --T --- --C --- ---
Platypus --- --- --- --- --G --- --- --- --G --- --- --- --C --- --- --G --C --- --G --C --- --- --- --- --G --- --C --G --- ---
Chicken  --- --A --G T-- --G --- --- --- --- --- --- --- --C --- --- --G --- A-- --G A-- --G --- --A --- --- --- --T --C --- ---

Human    ACC GAG CGA CAG ATC AAA ATC TGG TTC CAG AAC CGC CGG ATG AAG TGG AAA AAA GAA TCT AAT CTC ACA TCC ACT CTC TCG GGG GGC GGC 630
Dog      --- --- --C --- --- --- --- --- --- --- --- --G --- --- --- --- --- --G --G --C --C --- --G --- --G --- --- --- --- ---
Mouse    --- --- --- --- --- --- --- --- --- --- --- --- --C --- --- --- --- --- --- --- --- --- --G --- --A --- --A --- --T ---
Opossum  --- --- A-- --- --- --- --T --- --- --- --- --- --- --- --A --- --- --G --G --- --- --- --T --- --G --G --- --- --- ---
Platypus --- --- A-G --- --T --G --- --- --- --- --- --G --- --- --- --- --- --G --- --C --- --- --G --T --G --G --- --- --A --T
Chicken  --G --A --G --- --- --- --- --- --- --- --- ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...

Human    GGA GGG GCC ACC GCC GAC AGC CTG GGC GGA AAA GAG GAA AAG CGG GAA GAG ACA GAA GAG GAG AAG CAG AAA GAG TGA                 708
Dog      --- --- --- G-G --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Mouse    --- --- --A --- --- --- --- --- --A --- --- --- --- --- --A --- --- --- --- --A --- --- --- --- --- ---
Opossum  --- --- --- G-A --- --- --- --A --- --C --G --- --G --- --- --- --- --- --- --A --A --A --- --- --- ---
Platypus --C --- --G G-- --G --- --- T-- -C- --C -.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
Chicken  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
```

Figure 2

Nucleotide sequences of orthologous HoxC6 genes from six mammalian species. The ultraconserved coding regions (UCRs) are shaded by blue and the homeobox motif is underlined in red.

UCRs could possibly be due to longer divergence times between these species, and therefore more synonymous substitutions have accumulated. Although another possible reason is that the coverage of genomic sequences of opossum and platypus are incomplete and several Hox genes were not retrieved, this is unlikely to affect our results because we did uncover most of the Hox genes from these species and only about half of the missing ones have UCRs in comparison of genes from placental mammals. Similarly, UCRs were not detected in Hox orthologs between two closely related puffer fishes, *Takifugu rubripes* and *Tetraodon nigroviridis*, even though the time of divergence between these two species has been estimated to be only 18–30 million years [15]. Furthermore, we did not identify any UCR for 8 Hox gene pairs between two *Drosophila* species, *D. melanogaster* and *D. pseudoobscura*, whose divergence time is less than 30 million years [16]. Therefore, the presence of UCRs in the Hox genes is likely to be unique to placental mammals.

Although we used a threshold of 120 nucleotides, the length of a UCR was often much longer than this. For example, a UCR of 348 nucleotides is present in HoxC6 genes between the human and dog, and this region is also highly conserved in all other placental mammals (Fig. 2). Note that this region is located outside the highly con-

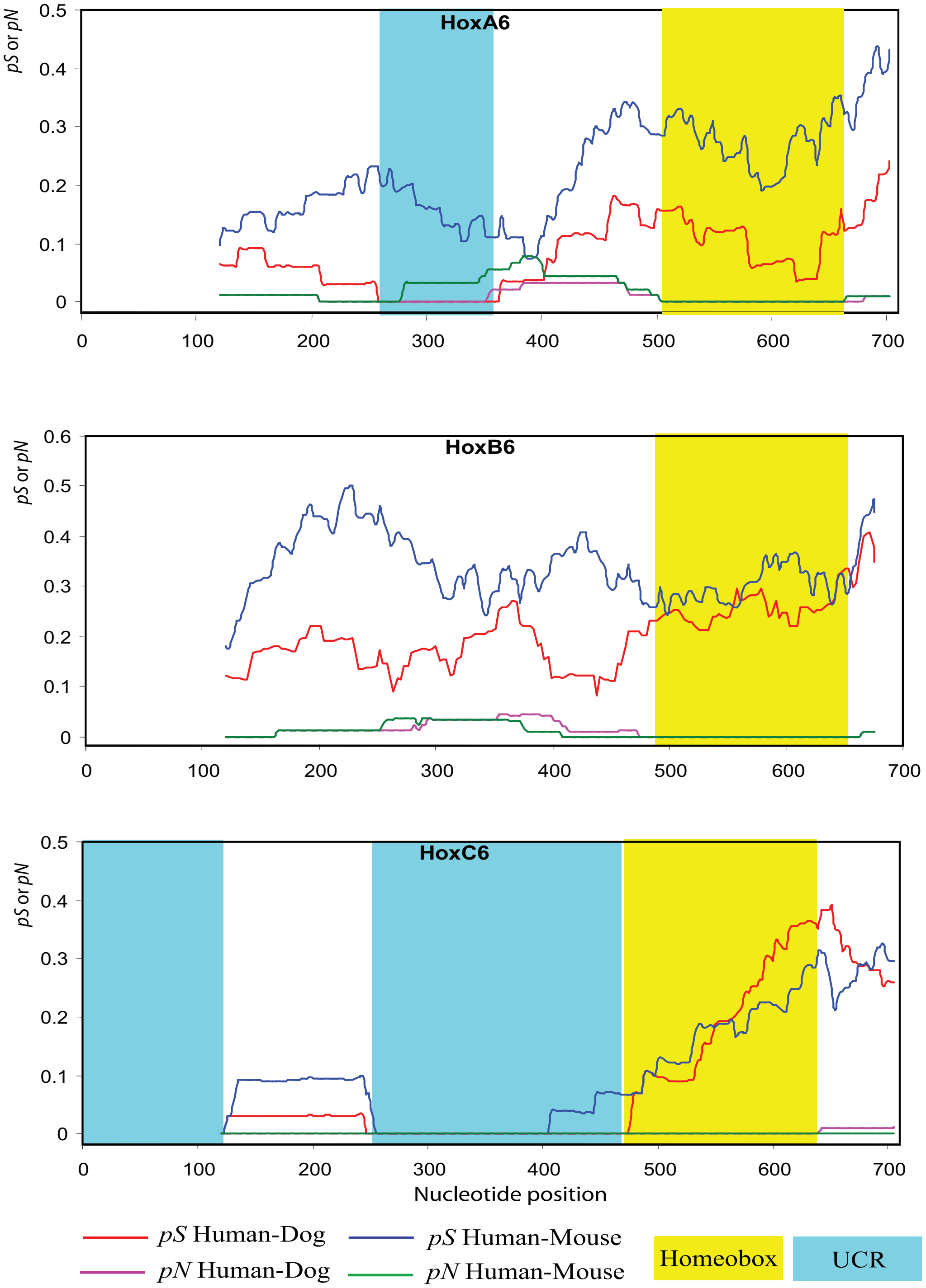


Figure 3

Sliding window analysis of synonymous and nonsynonymous nucleotide substitutions between human and dog or human and mouse for the orthologous gene group 6. The window size is 120 nucleotides and moves every single codon. The X axis indicates the position of the last nucleotide of each sliding window comparison. The regions of homeobox motif and UCRs are highlighted in yellow and light blue, respectively. The synonymous and nonsynonymous substitutions are measured by the proportion of synonymous differences per synonymous site (*pS*), and the proportion of nonsynonymous differences per nonsynonymous site (*pN*), respectively. We used *pS* and *pN* here because these measures are model free and range from 0 to 1 (see Nei and Kumar 2000).

Table 1: List of numbers of UCRs identified in the mammalian Hox genes

	Dog	Cow	Mouse	Opossum	Platypus
Human	19/26	19/23	13/14	2/2	0
Dog	-	21/33	10/12	2/2	0
Cow		-	14/16	2/2	0
Mouse			-	2/2	0
Opossum				-	0

Notes: The first number represents the number of Hox genes that contain UCRs. The second number refers to the total number of UCRs identified in the pairwise comparison between two species.

served homeobox motif. In the homeobox motif, a number of synonymous substitutions were observed although the nonsynonymous sites are identical (Fig. 3). The difference of the conservation level of synonymous sites between the homeobox and UCR regions indicates that the occurrence of UCRs is more likely to be due to purifying selection at both synonymous and nonsynonymous sites. It also indicates that the appearance of UCRs is unlikely to be due to conservation of protein sequence, although it was previously found that highly conserved nonsynonymous sites are usually associated with conserved synonymous sites [17]. The mammalian Hox genes usually contain two exons and one intron. The homeobox motif is located in the second exon in all mammalian Hox genes. However, most UCRs are found in the first exon, and many of them appear at the 5'end of the coding sequences (CDS) (see Additional file 1), indicating that they are located in non-homeobox regions.

Although a long UCR is present in HoxC6 gene, the other Hox6 group genes do not contain such a long UCR. For example, many synonymous substitutions are detected in the corresponding region of HoxB6 (Fig. 3). Such differences were also observed between members of other cognate groups (see Additional file 1). In addition, even if UCRs are present in all genes of a cognate group (e.g., group 5 and 8), UCRs are observed at somewhat different locations (see Additional file 1).

Comparison of the frequency of UCRs among different gene clusters and different regions of clusters showed that the frequency varies significantly with gene pair. First, UCRs are more frequent in the HoxC cluster than in any other clusters. Ten UCRs are detected for 8 out of 9 HoxC cluster genes, and HoxC genes usually contain long UCRs. In contrast, only 20 UCRs are found in the other three gene clusters (30 genes). Second, more UCRs are present in the central cognate groups (Hox4-8) than those genes of anterior (Hox1-3) and posterior groups (Hox9-13). For example, 11 out of 15 central Hox genes contain at least one UCR (total = 15), while only 11 UCRs were detected among 24 Hox genes of the other two groups. Furthermore, there is a significant difference in the fraction of DNA sequence involved in UCRs between central and non-central Hox genes. The UCRs in the central group of genes contain 3684 nucleotides, accounting for 32% of coding sequences, whereas only 10% of nucleotides are involved in UCR sequences in the other groups of Hox genes.

Origin and evolution of UCRs

To study the origin and evolution of the UCRs, we concatenated the nucleotide sequences of 32 UCRs which were shared by 22 mammalian Hox sequences (see Additional file 2) for the seven eutherian species. We also concatenated the corresponding UCR region of the Hox genes from opossum, platypus and chicken, and these sequences were aligned with the eutherian sequences (The multiple sequence alignments are available in Additional file 3). We then computed the numbers of synonymous (d_S) and nonsynonymous (d_N) substitutions for all pairs of ten species. We then related the d_S and d_N values for a pair species to their divergence time. The results of the study are presented in Fig. 4A. This figure shows that both d_S and d_N do not increase appreciably for the first 100 million years during which placental mammals (or eutherians) evolved. As the divergence time increased beyond 100 million years, however, d_S rapidly increased up to 340 million years ago (MYA) when eutherians and chicken diverged. The rate of increase of d_N was slow beyond the first 100 million years but steadily increased up to 340 MYA. These results clearly showed that the eutherian UCRs evolved around the time of origin of placental mammals. This conclusion is consistent with previous observation that UCRs exist primarily between eutherian genomes.

To clarify this situation graphically, we constructed the Neighbor-Joining tree of the species involved using the d_S and d_N values with the chicken sequence as the outgroup (Fig. 4B, C). Fig. 4B shows that in the eutherian lineage d_S increased relatively rapidly before the eutherian radiation but virtually stopped increasing thereafter. This again supports the evolution of UCRs in the early stage of eutherian evolution. A similar result was observed by d_N, though the reliability of estimates of branch lengths is low in this case because of the small d_N values.

Why did UCRs evolved in placental mammals and how are they conserved in the genome? It is very difficult to answer these questions at the present. However, we noticed that the origin of placental mammals depended on the evolution of placenta, which grows inside the mother's uterus and functions as a way to exchange gas and nutrients between the mother and fetus during gestation [18]. It has been suggested that the placenta of eutherian mammals evolved from a much simpler tissue

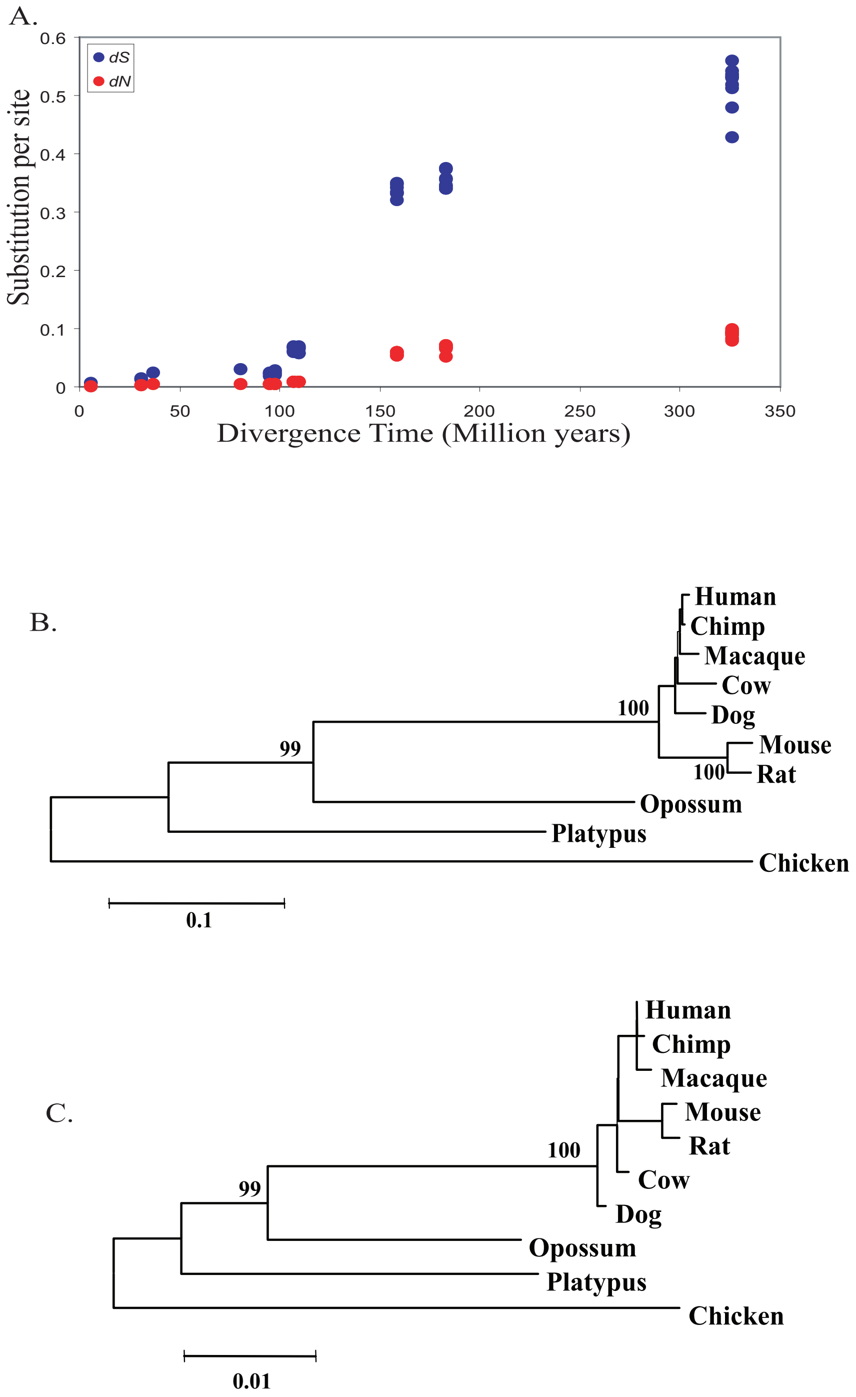


Figure 4

Non-linear accumulation of synonymous substitutions in the Hox UCRs. **A**. The numbers of synonymous and non-synonymous substitutions between two species are plotted against their divergence times, which were obtained from the TimeTree database[50]. **B**. Neighbor-Joining trees based on synonymous substitutions (d_S) which were computed by the pair-wise deletion option (Nei and Kumar 2000). The chicken sequence was used as the outgroup. This tree shows short branches for placental mammals which diverged during the last 100 million years. Opossum and placental mammals have diverged about 145 MYA, but the branch length for opossum is shorter than that for placental mammals. This result indicates that the d_S for placental mammals increased rather rapidly until their emergence. C. Neighbor-Joining tree for d_N. The evolutionary pattern of this tree is similar to that of the tree for d_S.

attached to the eggshell of birds and reptiles [18]. Many Hox genes, including HoxA4, HoxA7, HoxA11, HoxC4, HoxC5, HoxC6, and HoxC8, have been shown to be expressed during placental development [19-21]. It is interesting to note that most of these Hox genes are located in the central regions of Hox gene clusters and contain at least one UCR in the coding regions. Although the specific functions of these Hox genes in placental development are still unclear, these observations suggest that they are involved in the growth and differentiation of trophoblasts. Considering that the Hox genes play important roles in the development of animal embryo development, we can postulate that the eutherian UCRs might have developed in association with the evolution of placentae. According to this hypothesis, we could further speculate that new mutations in the UCRs could have been deleterious after the formation of placentae and these mutations have been eliminated by purifying selection. In summary, the UCR sequence might be important for the function of Hox genes in formation of the placenta and have contributed to the evolution of placental mammals.

Mutation rate and the maintenance of UCRs

One of the possible explanations of the maintenance of UCRs for a long evolutionary time in the eutherian genome could be a low mutation rate that might be observed in this specific genomic region. If this hypothesis is correct, one would expect that the rate of nucleotide substitution is lower in the intronic and intergenic regions as well as in the UCRs in these Hox gene regions than in other regions. Actually, there is some evidence that the mutation rate is not constant throughout the genome but varies with genomic region in mammals [14,22]. We therefore tested this hypothesis by computing the rate of nucleotide substitution in the intronic and intergenic regions of the Hox gene complex as well as the rate of synonymous substitution. We studied this problem in relation to the locations of Hox genes because the genes located in the central region of the Hox gene complex contain UCRs more often than the genes in the noncentral regions.

The rate of nucleotide substitution between a pair of species was measured by the number of nucleotide substitution (d) between them by using the Jukes-Cantor formation [23]. Similarly, the rate of synonymous substitution was measured by the number of synonymous nucleotide substitutions (d_S) using the modified Nei and Gojobori method [23]. The values were computed for all pairs of human, dog, and mouse and opossum. The results obtained are presented in Fig 5. Fig. 5A shows that the d_S values are significantly correlated with the positions of the genes in the Hox gene cluster. The relationship is more or less U-shaped, and the Hox genes at both ends of the cluster (e.g., cognate groups 1 and 13) have higher d_S than the genes in the central region (cognate groups 4–8). The d_S values of central cognate genes are significantly lower than those of other Hox genes ($P < 3.47 \times 10^{-7}$). This pattern of d_S values was not observed in the comparison of Hox genes between the two closely related puffer fishes and between *D. melanogaster* and *D. pseudoobscura*. There-

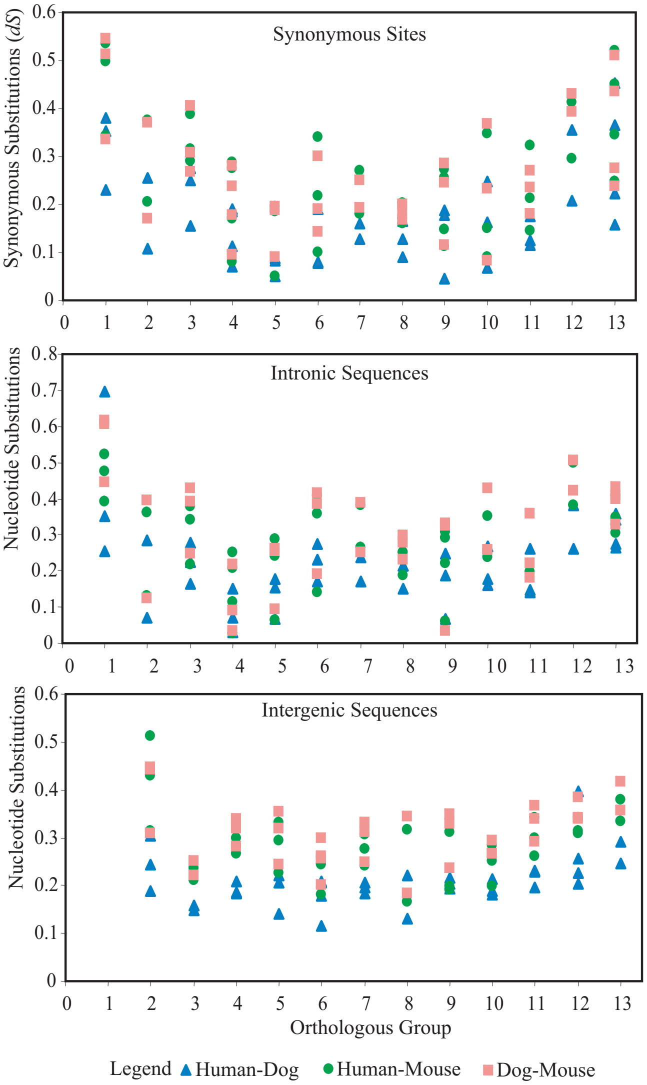


Figure 5
Reduced substitution rates in the central region of Hox gene clusters. **A**. Correlation between the synonymous mutation rates (d_S) of coding regions of Hox genes and their locations in the Hox cluster. The d_S values are plotted against the corresponding number of cognate groups. **B**. Correlation between nucleotide substitution rates of intronic sequences of each Hox genes and their positions on the cluster. **C**. Correlation between nucleotide substitution rates of intergenic sequences and their positions on the Hox cluster. The numbers on the x axis represent the numbers of cognate group at 5' flanking end of each intergenic sequence.

fore, it is likely that only mammalian Hox gene clusters show such a position effect, which is consistent with the occurrence of UCRs only in eutherians.

In addition, the numbers of nucleotide substitutions (*d*) of intronic sequences in Hox genes shows a similar U-shaped curve (Fig. 5B), and these values showed a significant correlation with the d_S values ($R = 0.6465$, $P < 1.488 \times 10^{-5}$). Similarly, the *d* values for the intergenic sequences also showed a U-shape pattern (Fig. 5C). Therefore, all sites in the central regions of Hox gene clusters (including coding, intronic and intergenic regions) are more conserved than those of the anterior or posterior genes. This suggests that low mutation rate in the central region of the Hox gene cluster is one factor for generating the higher frequency of UCRs in this region. However, the d_S values in the central region of Hox gene complex was significantly lower than the *d* values for the intronic and intergenic region ($P < 1.29 \times 10^{-4}$). This higher degree of conservation of the synonymous sites suggests that purifying selection is a more important factor for generating UCRs than the lower mutation rate.

UCRs have reduced density of synonymous SNPs and nucleotide diversity

The highly conserved synonymous sites of the UCRs indicate that most mutations at these sites should have some deleterious effects and purifying selection eliminate the mutations. Therefore, if the conservation of UCRs is due to purifying selection, we would expect a decreased frequency of synonymous Single Nucleotide Polymorphisms (SNPs) and reduced gene diversity (heterozygosity) in the UCRs in the human population. To test this hypothesis, we used human SNP data to estimate the frequency of SNPs inside and outside the UCRs. As expected, the frequency of synonymous SNPs in UCRs (2.88/kb) was significantly lower than that in the non-UCRs regions (9.04) of human Hox genes ($P < 0.01$). The frequency of nonsynonymous SNPs within UCRs was 1.44/kb, which is also lower than that of non-UCRs (2.73/kb), although the difference was not significant ($P = 0.34$). The significantly reduced frequency of synonymous SNPs in the UCRs further supports the hypothesis that the synonymous sites in the UCR regions have been constrained by purifying selection.

We also compared the nucleotide diversity (π) between UCR and non-UCR regions based on SNP data and observed a significant difference between them. For the synonymous sites, $\pi = 6.01 \times 10^{-4}$ in the UCR regions, compared to $\pi = 0.017$ in the non-UCR regions. The $\pi = 1.9 \times 10^{-4}$ for nonsynonymous sites in the UCRs, but 4.17×10^{-4} in the non-UCR sequences. Therefore, the UCR regions have much lower nucleotide diversity at both synonymous and nonsynonymous sites than those of non-UCRs. The reduction of nucleotide diversity in the UCRs further supports the idea that population frequencies of deleterious SNP alleles has been reduced by purifying selection. The evolutionary conservation of UCRs, in turn, suggests that the UCR sequences may be important for the proper function of Hox genes.

UCRs might overlap with novel transcripts or function as regulatory elements

Recent studies have shown that synonymous sites could be affected by natural selection [24-28]. For example, the synonymous codons of many highly expressed genes are not randomly used in many organisms such as bacteria, plants, fungi and invertebrates [29-31]. Moreover, the synonymous sites in the Exonic Splicing Enhancers (ESEs), which affect splicing of pre-mRNA, are also under purifying selection [27,32-34]. However, we did not detect a significant correlation between codon usage and occurrence of UCRs, and there is no statistically significant difference in the density of ESEs between UCRs and other regions (data not shown).

In this study, we found that many UCRs are located either at the 5'end or 3'end of coding regions of genes and their flanking noncoding regions are also highly conserved in some cases. As a consequence, large conserved blocks covering both coding and noncoding regions are present in some Hox genes. For example, the noncoding sequences are highly conserved near the UCRs in HoxC4, HoxC5 and HoxC6 (Additional file 4A), forming long conserved blocks (over 500 nucleotides). Further analysis indicated that these regions are only conserved in placental mammals, similar to the UCRs (Additional file 4B). Currently, information about the functions and evolutionary origins of these conserved blocks is not available. At present time, we can not exclude the possibility that these blocks might be the overlapping regions of Hox genes and other unknown RNA transcript genes.

Such potential overlapping genes can be transcribed in the same or reverse direction to the Hox genes. Antisense RNA transcription has been implicated in various forms of gene regulation, including RNAi-like degradation of corresponding sense RNA transcripts and competition with sense transcription. For example, a 177-nt antisense RNA is transcribed from the central coding region of the photosynthesis gene *IsiA* in cyanobacteria and suppresses the expression of gene *IsiA* under some conditions [35]. A recent study has shown that the HoxA clusters are enriched in antisense transcripts and many of these transcripts overlap with coding regions [36]. Among them, the *HoxA11* antisense RNA transcript has been well characterized and shown to have a function in regulating the expression of the *HoxA11* gene [37,38]. We noticed that a *HoxA11* antisense transcript overlaps with the coding region of the first HoxA11 exon [37], where a UCR is identified in this study. Although we did not find any information in the literature for other antisense transcripts that overlap with UCRs studying Hox genes, the possibility of antisense transcription still remains. A recent detailed analysis on 1% of the human genome indicates that substantial portions of the unannotated sequences are transcribed and many transcripts overlap one another extensively [39]. Therefore, we should consider the possibility that the occurrence of UCRs in the mammalian Hox genes is related to the presence of overlapping RNA transcripts, including antisense RNAs.

Some ultraconserved elements in the noncoding regions of mammalian genomes have been shown to function as regulatory elements of neighboring developmental genes [3,4]. The Hox genes are the master control genes for animal embryonic development and are organized into conserved clusters. There is evidence that this organization is critical for the control of the proper spatial and temporal expression of Hox genes [40]. A number of conserved noncoding sequences of 60 to hundreds of nucleotides have been identified in the intergenic regions of Hox clusters by comparative genomic studies, and they were considered to be putative regulatory elements of Hox genes [41]. Although the UCRs are found in the coding regions, it is possible that these regions contain regulatory elements for their own genes or downstream genes. A similar scenario has been observed about the expression regulation of β-globin gene cluster, which has been the subject of extensive studies [42]. Different β-globin genes are expressed at various development stages. The elements close to the globin genes and the arrangement of the globin genes control their expression switching in different developmental stages[42]. The Hox genes have a similar cluster organization and they are also expressed differently at various developmental stages. Therefore, the UCRs might also serve as regulatory elements to control their neighboring Hox genes. If this is the case, selection on both regulatory elements and coding sequences in the same region has led to the formation and/or maintenance of these UCRs. Because the UCRs are much longer than usual protein binding motifs, it is possible that multiple binding motifs are present in a single UCR. The sequence, number and order of these motifs could be important for their regulatory functions, and the combination of these factors might explain the formation of the UCRs [42]. Alternatively, these extraordinarily long UCRs might interact with RNAs or DNAs in other genomic regions and such interactions might require more highly conserved sequences. Therefore, it is reasonable to postulate that these UCRs are involved in the regulation of expression of Hox genes.

Conclusion

It appears that ultraconserved elements are frequently present in the mammalian genomes, especially near or in regulatory genes [1,28]. The appearance of UCRs in mammalian Hox genes suggests that they might play important roles during embryo development. Placental mammals are fundamentally different from other mammals in regard to the early developmental environment and fetal nourishment. Previous studies have shown that a number of Hox genes that contain UCRs are expressed in the placenta, although their functions are still unclear [19-21]. Taking into account the important roles of the Hox genes in animal embryonic development and morphological evolution, we cannot ignore the potential connection between the appearance of UCRs and advent of long gestation periods in placental mammals. Therefore, it would be particularly important to study the functions of the UCRs during fetal development. Point mutations at the synonymous sites in the UCRs could be introduced to test if these sites are required for the proper function of mammalian Hox genes. The experimental test could also provide further understanding about the functions of sequences outside the homeobox motif and their roles in the evolution of placental mammals.

Experimental procedures

Data mining

The protein coding sequences (CDS) and genomic sequences of Hox genes of human (*Homo sapiens*), mouse (*Mus musculus*), dog (*Canis familiaris*) and the fruitfly *Drosophila melanogaster* were obtained from the NCBI database. These Hox protein sequences were used as queries to blast the NCBI *nr* database for Hox genes from other representative organisms. TBLASTN was performed against the genomic sequences from NCBI genomic databases to obtain as many Hox sequences as possible from the following species: chimpanzee (*Pan troglodytes*), rhesus macaque (*Macaca mulatta*), cow (*Bos Taurus*), dog (*Canis familiaris*), rat (*Rattus norvegicus*), gray short-tailed opossum (*Monodelphis domestica*), duck-billed platypus (*Ornithorhynchus anatinus*), chicken (*Gallus gallus*), zebrafish (*Danio rerio*), Japanese pufferfish (*Takifugu rubripes*), and green pufferfish (*Tetraodon nigroviridis*). The putative coding regions and their protein products were predicted from their genomic sequences based on sequence homology.

Sequence analysis

The protein sequences of each Hox orthologous group from human, dog, cow, mouse, opossum and platypus were used to generate a multiple sequence alignment by using ClustalX 1.83 [43]. Corresponding nucleotide sequence alignments were then reexamined to improve the alignments based on these protein sequence alignments to avoid frame-shift errors in GeneDoc [44]. The alignments of all available Hox genes (43 pairs) from Japanese pufferfish and green pufferfish were generated separately using the same procedure. Similarly, the alignments of eight Hox genes from two *Drosophila* species, *D. melanogaster* and *D. pseudoobscura*, were also produced. In order to detect a coding region without any synonymous substitution in a region of 120 nucleotides or longer, a sliding window analysis was performed. The size of each sliding window was 120 nucleotides (40 codons), and it moved by every single codon. In each window, the proportion of synonymous differences per synonymous site (p_S) and proportion of nonsynonymous differences per nonsynonymous site (p_N) were calculated by using the Nei and Gojobori method to minimize the estimation error [45,46]. The ultraconserved coding region (UCR) was detected if $p_S = 0$ and $p_N = 0$ in these sliding windows. The *perl* script for sliding window analysis was provided by Masafumi Nozawa (Personal communication). The multiple sequence alignment of each UCR was then further inspected to detect the presence of insertions and deletions (indels). Only UCRs without any indel were used in this study.

Multiple nucleotide sequence alignments of the UCRs were constructed using sequences from human, chimpanzee, macaque, dog, mouse, rat, opossum, platypus and chicken. A concatenated alignment was generated by joining all the UCRs. The number of synonymous substitutions per synonymous site (d_S) and the number of nonsynonymous substitutions per nonsynonymous site (d_N) for each pair of concatenated UCR sequences were estimated by using the modified Nei-Gojobori method in MEGA4 [6,45,47]. Neighbor Joining (NJ) trees of the concatenated UCR sequences were reconstructed separately based on their synonymous and nonsynonymous substitutions in MEGA4.

Pairwise d_S values of whole coding regions of Hox genes between human, dog and mouse were calculated by the modified Nei-Gojobori method in MEGA4. The intron and intergenic sequences of Hox clusters were retrieved from human (build 36.2), mouse (Build 37.1) and dog genomic sequences and aligned by ClustalX. Pairwise rates of nucleotide substitutions of these intronic and intergenic sequences were estimated by the same method between the three species.

SNP, ESE, and Codon usage bias analysis

Single Nucleotide Polymorphisms (SNPs) data for each human Hox gene were obtained from the NCBI dbSNP database. The numbers of polymorphic synonymous and nonsynonymous per site were determined for UCR and non-UCR regions. The nucleotide diversity (average number of nucleotide differences per site between two sequences) was estimated by $\hat{\pi} = \frac{n}{n-1}\sum_{ij}\hat{x}_i\hat{x}_j\pi_{ij}$ [23], where n is the number of DNA sequences examined, and $\hat{x}_i$ is the population frequency of the i-th allele, and π_{ij} is the proportion of different nucleotides between the i-th and j-th type of DNA sequences. The numbers of detected Exonic Splicing Enhancers (ESEs) in UCRs and non-UCRs of each human Hox gene were obtained by examining the human RESCUE-ESE WebServer [48]. The values of Effective Number of Codons of the Hox genes from different lineages were estimated using the web server of CodonW [49].

Authors' contributions

ZL and MN designed and conducted data analysis. ZL, MN and HM wrote the manuscript.

Additional material

Additional file 1

A complete list of identified UCRs based on pairwise comparisons among mammalian Hox genes.

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2148-8-260-S1.pdf]

Additional file 2

***A list of UCRs of Hox genes that are used for concatenated multiple sequence alignment and phylogenetic tree construction**. The nucleotide positions of each UCR are listed in the right column. The name of each UCR is denoted in parentheses.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2148-8-260-S2.pdf]

Additional file 3

Multiple alignments of the nucleotide sequences of UCRs

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2148-8-260-S3.pdf]

Additional file 4

***Conserved noncoding sequences flanking the UCRs**. **A**. The conservation of noncoding sequences flanking the first exons of HoxC4, HoxC5 and HoxC6 from UCSC Human Genome Brower. The position of the coding regions is highlighted by red bar. Transcription direction is indicated by arrow. **B**. The accumulations of nucleotide mutations in the conserved regions with divergence times of the three Hox genes. The number of substitution per site was estimated using the Jukes-Cantor's method in MEGA4 on the coding region and flanking conserved noncoding sequences of HoxC4, HoxC5 and HoxC6.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2148-8-260-S4.pdf]

Acknowledgements

We thank Masafumi Nozawa, Dimitra Chalkia, Saby Das, John Diller, Zhengjia Wang and three anonymous reviewers for critical reading and valuable comments on the manuscript. This work was supported by National Institutes of Health Grant GM020293 (to MN).

References

1. Bejerano G, Pheasant M, Makunin I, Stephen S, Kent WJ, Mattick JS, Haussler D: **Ultraconserved elements in the human genome.** *Science* 2004, **304(5675):**1321-1325.
2. Katzman S, Kern AD, Bejerano G, Fewell G, Fulton L, Wilson RK, Salama SR, Haussler D: **Human genome ultraconserved elements are ultraselected.** *Science* 2007, **317(5840):**915.
3. Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, Shoukry M, Minovitsky S, Dubchak I, Holt A, Lewis KD, *et al.*: **In vivo enhancer analysis of human conserved non-coding sequences.** *Nature* 2006, **444(7118):**499-502.
4. Woolfe A, Goodson M, Goode DK, Snell P, McEwen GK, Vavouri T, Smith SF, North P, Callaway H, Kelly K, *et al.*: **Highly conserved non-coding sequences are associated with vertebrate development.** *PLoS biology* 2005, **3(1):**e7.
5. Gehring WJ, Qian YQ, Billeter M, Furukubo-Tokunaga K, Schier AF, Resendez-Perez D, Affolter M, Otting G, Wuthrich K: **Homeodomain-DNA recognition.** *Cell* 1994, **78(2):**211-223.
6. Zhang J, Nei M: **Evolution of Antennapedia-class homeobox genes.** *Genetics* 1996, **142(1):**295-303.
7. Krumlauf R: **Hox genes in vertebrate development.** *Cell* 1994, **78(2):**191-201.
8. Galant R, Carroll SB: **Evolution of a transcriptional repression domain in an insect Hox protein.** *Nature* 2002, **415(6874):**910-913.
9. Ronshaugen M, McGinnis N, McGinnis W: **Hox protein mutation and macroevolution of the insect body plan.** *Nature* 2002, **415(6874):**914-917.
10. Kuziora MA: **Abdominal-B protein isoforms exhibit distinct cuticular transformations and regulatory activities when ectopically expressed in Drosophila embryos.** *Mechanisms of development* 1993, **42(3):**125-137.
11. Kumar S, Hedges SB: **A molecular timescale for vertebrate evolution.** *Nature* 1998, **392(6679):**917-920.
12. Nei M, Xu P, Glazko G: **Estimation of divergence times from multiprotein sequences for a few mammalian species and several distantly related organisms.** *Proceedings of the National Academy of Sciences of the United States of America* 2001, **98(5):**2497-2502.
13. King JL, Jukes TH: **Non-Darwinian evolution.** *Science (New York, NY)* 1969, **164(881):**788-798.
14. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, *et al.*: **Initial sequencing and comparative analysis of the mouse genome.** *Nature* 2002, **420(6915):**520-562.
15. Crnogorac-Jurcevic T, Brown JR, Lehrach H, Schalkwyk LC: **Tetraodon fluviatilis, a new puffer fish model for genome studies.** *Genomics* 1997, **41(2):**177-184.
16. Russo CA, Takezaki N, Nei M: **Molecular phylogeny and divergence times of drosophilid species.** *Molecular biology and evolution* 1995, **12(3):**391-404.
17. Wyckoff GJ, Malcom CM, Vallender EJ, Lahn BT: **A highly unexpected strong correlation between fixation probability of nonsynonymous mutations and mutation rate.** *Trends Genet* 2005, **21(7):**381-385.
18. Knox K, Baker JC: **Genomic evolution of the placenta using co-option and duplication and divergence.** *Genome research* 2008, **18(5):**695-705.
19. Amesse LS, Moulton R, Zhang YM, Pfaff-Amesse T: **Expression of HOX gene products in normal and abnormal trophoblastic tissue.** *Gynecologic oncology* 2003, **90(3):**512-518.
20. Ishii M, Hayakawa S, Satoh K: **Roles of HOX Genes in the Growth, Differentiation and Malignant Transformation of Human Trophoblasts.** *Nihon Univ J Med* 1999, **41(6):**339-350.
21. Wolgemuth DJ, Viviano CM, Gizang-Ginsberg E, Frohman MA, Joyner AL, Martin GR: **Differential expression of the mouse homeobox-containing gene Hox-1.4 during male germ cell differentiation and embryonic development.** *Proceedings of the National Academy of Sciences of the United States of America* 1987, **84(16):**5813-5817.
22. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, *et al.*: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409(6822):**860-921.
23. Nei M, Kumar S: **Molecular Evolution and Phylogenetics.** USA: Oxford University Press; 2000.
24. Shields DC, Sharp PM, Higgins DG, Wright F: **"Silent" sites in Drosophila genes are not neutral: evidence of selection among synonymous codons.** *Molecular biology and evolution* 1988, **5(6):**704-716.
25. Chamary JV, Hurst LD: **Evidence for selection on synonymous mutations affecting stability of mRNA secondary structure in mammals.** *Genome biology* 2005, **6(9):**R75.
26. Chamary JV, Parmley JL, Hurst LD: **Hearing silence: non-neutral evolution at synonymous sites in mammals.** *Nature reviews* 2006, **7(2):**98-108.
27. Parmley JL, Chamary JV, Hurst LD: **Evidence for purifying selection against synonymous mutations in mammalian exonic splicing enhancers.** *Molecular biology and evolution* 2006, **23(2):**301-309.
28. Schattner P, Diekhans M: **Regions of extreme synonymous codon selection in mammalian genes.** *Nucleic acids research* 2006, **34(6):**1700-1710.
29. Akashi H: **Translational selection and yeast proteome evolution.** *Genetics* 2003, **164(4):**1291-1303.
30. Ermolaeva MD: **Synonymous codon usage in bacteria.** *Current issues in molecular biology* 2001, **3(4):**91-97.
31. Akashi H: **Synonymous codon usage in Drosophila melanogaster: natural selection and translational accuracy.** *Genetics* 1994, **136(3):**927-935.
32. Cartegni L, Chew SL, Krainer AR: **Listening to silence and understanding nonsense: exonic mutations that affect splicing.** *Nature reviews* 2002, **3(4):**285-298.
33. Chamary JV, Hurst LD: **Biased codon usage near intron-exon junctions: selection on splicing enhancers, splice-site recognition or something else?** *Trends Genet* 2005, **21(5):**256-259.
34. Carlini DB, Genut JE: **Synonymous SNPs provide evidence for selective constraint on human exonic splicing enhancers.** *Journal of molecular evolution* 2006, **62(1):**89-98.
35. Duhring U, Axmann IM, Hess WR, Wilde A: **An internal antisense RNA regulates expression of the photosynthesis gene isiA.** *Proceedings of the National Academy of Sciences of the United States of America* 2006, **103(18):**7054-7058.
36. Mainguy G, Koster J, Woltering J, Jansen H, Durston A: **Extensive polycistronism and antisense transcription in the Mammalian hox clusters.** *PLoS ONE* 2007, **2:**e356.
37. Hsieh-Li HM, Witte DP, Weinstein M, Branford W, Li H, Small K, Potter SS: **Hoxa 11 structure, extensive antisense transcription, and function in male and female fertility.** *Development (Cambridge, England)* 1995, **121(5):**1373-1385.
38. Potter SS, Branford WW: **Evolutionary conservation and tissue-specific processing of Hoxa 11 antisense transcripts.** *Mamm Genome* 1998, **9(10):**799-806.
39. Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE, *et al.*: **Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project.** *Nature* 2007, **447(7146):**799-816.
40. Crawford M: **Hox genes as synchronized temporal regulators: implications for morphological innovation.** *Journal of experimental zoology Part B* 2003, **295(1):**1-11.
41. Santini S, Boore JL, Meyer A: **Evolutionary conservation of regulatory elements in vertebrate Hox gene clusters.** *Genome research* 2003, **13(6A):**1111-1122.
42. Martin DI, Fiering S, Groudine M: **Regulation of beta-globin gene expression: straightening out the locus.** *Current opinion in genetics & development* 1996, **6(4):**488-495.
43. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG: **The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools.** *Nucleic Acids Res* 1997, **25(24):**4876-4882.
44. Nicholas KB, Nicholas HB Jr, Deerfield DW II: **GeneDoc: Analysis and Visualization of Genetic Variation.** *EMBNET News* 1997, **4(14):**1-4.

45. Nei M, Gojobori T: **Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions.** *Mol Biol Evol* 1986, **3(5):**418-426.
46. Zhang J, Rosenberg HF, Nei M: **Positive Darwinian selection after gene duplication in primate ribonuclease genes.** *Proceedings of the National Academy of Sciences of the United States of America* 1998, **95(7):**3708-3713.
47. Tamura K, Dudley J, Nei M, Kumar S: **MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0.** *Molecular biology and evolution* 2007:2007.
48. Fairbrother WG, Yeh RF, Sharp PA, Burge CB: **Predictive identification of exonic splicing enhancers in human genes.** *Science* 2002, **297(5583):**1007-1013.
49. Peden J: **Analysis of Codon Usage.** Nottingham: University of Nottingham; 1999.
50. Hedges SB, Dudley J, Kumar S: **TimeTree: a public knowledgebase of divergence times among organisms.** *Bioinformatics (Oxford, England)* 2006, **22(23):**2971-2972.