The draft genomes of soft-shell turtle and green sea turtle yield insights into the development and evolution of the turtle-specific body plan

Wang, Zhuo; Pascual-Anaya, Juan; Zadissa, Amonida; Li, Wenqi; Niimura, Yoshihito; Huang, Zhiyong; Li, Chunyi; White, Simon; Xiong, Zhiqiang; Fang, Dongming; Wang, Bo; Ming, Yao; Chen, Yan; Zheng, Yuan; Kuraku, Shigehiro; Pignatelli, Miguel; Herrero, Javier; Beal, Kathryn; Nozawa, Masafumi; Li, Qiye; Wang, Juan; Zhang, Hongyan; Yu, Lili; Shigenobu, Shuji; Wang, Junyi; Liu, Jiannan; Flice, Paul; Searle, Steve; Wang, Jun; Kuratani, Shigeru; Yin, Ye; Aken, Bronwen; Zhang, Guojie; Irie, Naoki

Published in:
Nature Genetics

DOI:
10.1038/ng.2615

Publication date:
2013

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Wang, Z., Pascual-Anaya, J., Zadissa, A., Li, W., Niimura, Y., Huang, Z., ... Irie, N. (2013). The draft genomes of soft-shell turtle and green sea turtle yield insights into the development and evolution of the turtle-specific body plan. Nature Genetics, 45(6), 701-706. https://doi.org/10.1038/ng.2615
The draft genomes of soft-shell turtle and green sea turtle yield insights into the development and evolution of the turtle-specific body plan

Zhuo Wang1,12, Juan Pascual-Anaya2,12, Amonida Zadissa3,12, Wenqi Li4,12, Yoshihito Niumura5, Zhiyong Huang1, Chunyi Li4, Simon White2, Zhiqiang Xiong1, Dongming Fang1, Bo Wang1, Yao Ming1, Yan Chen1, Yuan Zheng1, Shigeru Kuratani2, Ye Yin4, Bronwen Aken3, Guojie Zhang1,10,11 & Naoki Irie2

The unique anatomical features of turtles have raised unanswered questions about the origin of their unique body plan. We generated and analyzed draft genomes of the soft-shell turtle (Pelodiscus sinensis) and the green sea turtle (Chelonia mydas); our results indicated the close relationship of the turtles to the bird-crocodilian lineage, from which they split ~267.9–248.3 million years ago (Upper Permian to Triassic). We also found extensive expansion of olfactory receptor genes in these turtles. Embryonic gene expression analysis identified a hourglass-like divergence of turtle and chicken embryogenesis, with maximal conservation around the vertebrate phylotypic period, rather than at later stages that show the amniote-common pattern. Wnt5a expression was found in the growth zone of the dorsal shell, supporting the possible co-option of limb-associated Wnt signaling in the acquisition of this turtle-specific novelty. Our results suggest that turtle evolution was accompanied by an unexpectedly conservative vertebrate phylotypic period, followed by turtle-specific repatterning of development to yield the novel structure of the shell.

The unique anatomy of turtles has raised questions about their evolution1. Their armor, even compared to other armored tetrapods (for example, the armadillo and Indian rhinoceros), is distinct in that the dorsal part of the shell (carapace) represents transformed vertebrae and ribs. In addition, their shoulder blades or scapulae display an inside-out topology against the rib cage (Supplementary Fig. 1 and Supplementary Note), and the lack of a temporal fenestra further complicates the reconstruction of their phylogenetic position1,2.

Three major hypotheses have been proposed for the evolutionary origin of turtles, including that they (i) constitute early-diverged reptiles, called anapsids3, (ii) are a sister group of the lizard-snake-tuatara (Lepidosauria) clade4 or (iii) are closely related to a lineage that includes crocodilians and birds (Archosauria)5–8. Even using molecular approaches, inconsistency still remains6–9. To clarify the evolution of the turtle-specific body plan, we first addressed the question of evolutionary origin of the turtle by performing the first genome-wide phylogenetic analysis with two turtle genomes sequenced in this project (the green sea turtle, C. mydas, and the Chinese soft-shell turtle, P. sinensis; Fig. 1a). In brief, the fragmented genomic DNA libraries of the two turtles were independently shotgun sequenced using the HiSeq 2000 sequencer and assembled using the SOAPdenovo assembler (Online Methods). The generated turtle genomes were both around 2.2 Gb in size, with the N50 lengths of scaffolds longer than 3.3 Mb (Table 1, Supplementary Figs. 2–5 and Supplementary Tables 1–9).

On the basis of the largest turtle data set so far, our phylogenetic analysis, with an orthologous set of 1,113 single-copy coding genes, robustly indicated that turtles are likely to be a sister group of crocodilians and birds (Fig. 1b, Supplementary Figs. 6 and 7 and Supplementary Tables 10–13), implying that the temporal fenestrae in the turtle skull were most likely secondarily lost in the turtle lineage1. A molecular evolutionary clock analysis with time constraints based on the fossil records estimated that turtles diverged from archosaursians approximately 257.4 million years ago, with a 95% credibility interval between 267.9 and 248.3 million years ago (Fig. 1b, Supplementary Fig. 7 and Supplementary Table 12). These results are consistent with the oldest turtle fossil (from 220 million years ago), named Odontocheilos10. The estimated time range corresponds...
to the Upper Permian to Triassic period (Fig. 1b), overlapping or following shortly after the Permian extinction event\(^1\); this raises the question of whether the emergence of the turtle group was related to this severe extinction event, which especially involved the extinction of marine species.

Taking into consideration the phylogenetic position of turtles, we next searched for genes that could potentially explain turtle-specific characteristics (Supplementary Fig. 8 and Supplementary Tables 14–23). Unexpectedly, we found that the olfactory receptor family was highly expanded in both turtle species (Fig. 2 and Supplementary Tables 14–20). In particular, the soft-shell turtle contained 1,137 intact, possibly functional olfactory receptor genes, a number comparable to or even greater than the number of olfactory receptor genes found in most mammals\(^15,16\). Olfactory receptor gene expansion was observed mainly in the \(\alpha\) subtype of the class I olfactory receptor genes, suggesting that turtles have superior olfaction ability against a wide variety of hydrophilic substances\(^15\) (Fig. 2a and Supplementary Tables 19 and 20). Detailed analyses with genomic sequences further clarified that the majority of the expansion occurred after the split of the two turtle species (Fig. 2b) and that the expansion was most likely facilitated by a gene duplication process, as inferred by the clustered distribution of the olfactory receptor genes in the genome (Fig. 2c,d). These results call into question the general proposition based on mammalian studies\(^13,14\) that vertebrates that expand their niche back into aquatic environments tend to reduce the number of olfactory receptor genes. Other than olfactory receptor gene expansion, we found that many genes involved in taste perception (Supplementary Tables 21–23) were lost in the two turtle species. Furthermore, we found that the gene for the hunger-stimulating and energy homeostasis–regulating hormone ghrelin was also lost specifically in the two turtle species (Supplementary Table 23), which could be related to their low-metabolic strategies. Further investigation of the lost genes in the two turtles identified the loss of many orthologs that are known to be important for normal development in different species, including the genes encoding UNC homeobox, FGF-binding protein 3, CXCL10 and Agouti signaling protein (Supplementary Table 23). These results, together with the identification of many other genes that show accelerated evolutionary rate in turtles (Supplementary Table 24; for example, Bmp receptor 1b, Kit, Juk1 and Eya4), suggest that turtle evolution has included many alterations of the signaling cascade that is presumably involved in morphogenesis. Finally, a possible connection to longevity in turtles was also found (Supplementary Table 24); the most accelerated gene in turtles, showing evidence of positive selection

<table>
<thead>
<tr>
<th>Table 1 Basic statistics of two turtle genomes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Estimated genome size</td>
</tr>
<tr>
<td>Sequencing depth</td>
</tr>
<tr>
<td>N50 scaffold</td>
</tr>
<tr>
<td>GC content</td>
</tr>
<tr>
<td>Number of coding genes</td>
</tr>
</tbody>
</table>

(with the rate of nonsynonymous nucleotide substitutions exceeding the rate of neutral mutations, \(dN/dS > 1\)), was microsomal glutathione S-transferase 3 (Mgst3; \(dN/dS = 5.68\)), which is reported to function in antioxidative stress, and disrupting the homolog Mgst3-like in Drosophila melanogaster reduces lifespan\(^15\).

In addition to changes in genomic sequences, we also investigated alterations in embryonic gene regulation that occurred after the split from the bird-crocodilian lineage. According to the recently supported developmental hourglass model\(^16–21\), the evolutionary changes underlying major adult morphological evolution occurred primarily in the developmental stages after the period of the vertebrate common plan or the period that serves as the source of the vertebrate basic body plan, namely, the vertebrate phylotypic period\(^22\). However, the hourglass model has not been tested in non-model organisms, particularly in those with the atypical anatomical features of turtles; therefore, we tested whether the model held true in turtle-chicken comparison. Taking advantage of RNA sequencing (RNA-seq) technology and our previously established method\(^21\) based on hierarchical Bayes statistics, our cross-species approach comparing whole-embryo gene expression profiles (GEPs) clearly demonstrated an hourglass-like GEP divergence in the embryogenesis of the soft-shell turtle and the chicken (Fig. 3a and Supplementary Figs. 9–12). However, the result was not robust enough to suggest that the most conserved developmental stage in turtles and birds corresponds to the vertebrate phylotype. The conserved stage could be one occurring later than the vertebrate phylotype, as a previous developmental study\(^23\) demonstrated that turtles have a typical amniote-common plan during embryogenesis and develop turtle-specific characteristics thereafter (for example, the scapula primordium first arises outside the rib cage and only later comes to lie inside the rib cage), as if the embryo is recapitulating its own evolutionary history\(^23\). If the most conserved developmental stage between the two species

![Figure 1](image-url)
Figure 2 Extensive expansion of olfactory receptor genes in turtles. (a) A neighbor-joining tree constructed with all the intact group α olfactory receptors from eight vertebrate species (soft-shell turtle, green sea turtle, chicken, zebra finch, anole lizard, human, dog and Western clawed frog), with the group β olfactory receptors as the outgroup. Bootstrap values (from 500 resamplings) are shown on the branches. The scale bar represents the number of amino-acid substitutions per site.

(b) Expansion of group α olfactory receptor genes in the evolution of tetrapods. Numbers in boxes indicate the current number of intact group α olfactory receptor genes in each species. The number of group α olfactory receptor genes in an ancestral species is shown in an ellipse at each node, and the numbers of gene gains and losses are shown on each branch with plus and minus signs, respectively. For divergence times, we used the median values obtained from TimeTree30. Note that the majority of the expansion of the group α olfactory receptor genes occurred independently in each turtle lineage. The same color code for species is used in a.b. (c,d) Genomic clusters of olfactory receptor genes in scaffolds 55 and 145 of the soft-shell turtle genome. Vertical red bars represent class I (c) and class II (d) olfactory receptor genes. Bars above and below the horizontal line indicate opposite directions of transcription. Long bars depict intact olfactory receptor genes, whereas short bars depict olfactory receptor pseudogenes or gene fragments.

Indeed corresponds to the stage of the amniote-common plan (approximately Tokita-Kuratani24 stage (TK) 13–14), this would indicate that the conserved stage may change depending on how distantly related the species are that are being compared, similar to the idea from the nested hourglasses model18 (Fig. 3b), justifying, in part, the hierarchical relationship between ontogeny and phylogeny once proposed by Karl von Baer25. Further investigation using a statistically robust cross-species comparative analysis indicated that the soft-shell turtle TK11 and the chicken HH16 developmental stages showed the most similar GXPs (Fig. 3c, Supplementary Figs. 13 and 14 and Supplementary Table 25).

Considering that the chicken stage corresponds to the previously identified phylotypic period21, turtle stage TK11 would be an attractive candidate for the vertebrate phylotypic period. In addition to the conservation between turtle and chicken at the level of gene regulation, the identified stages showed notable similarity in morphology (Fig. 3d and Supplementary Table 26), despite the large differences in their final

Figure 3 The molecular divergence of turtle and chicken embryos follows the hourglass model with a maximally conserved vertebrate phylotypic period. (a) Distances of whole-embryo GXPs (from depth-controlled, TMM (trimmed mean of M values)-normalized data) for 11,602 orthologs in selected developmental stages of the soft-shell turtle and chicken (Supplementary Fig. 16). Error bars, s.d. ANOVA P-value under heteroscedasticity = 7 × 10⁻⁷. (b) The hypothetical model (nested hourglass)18 in which both an hourglass-like divergence and a recapitulation-like relationship between ontogeny and phylogeny can be justified. The model infers that the most conserved developmental stage changes depending on how distantly related the species are that are being compared. Comparisons within vertebrate embryos gives a vertebrate phylotype (blue arrow), and comparisons within amniotes gives an amniote-type stage (red arrow) that emerges later than the vertebrate phylotype stage. (c) An all-to-all comparison of turtle and chicken GXP distances (total Manhattan) indicates that the highest similarity occurs between turtle stage TK11 and chicken stage HH16 embryos (see Supplementary Fig. 18 for statistical assessment). HH16 is the stage previously identified as the vertebrate phylotypic period21, which does not coincide with the model in b.

Error bars, s.d. (d) Morphological appearance of the soft-shell turtle (stage TK11) and chicken (stage HH16) embryos that showed the highest GXP similarity (Supplementary Table 25). Scale bars, 1 mm.
Figure 4 Molecular characteristics of turtle embryogenesis during and after the phylotypic period. (a) Shared expression of developmental genes (Supplementary Table 27) in the phylotypic stages of turtle and chicken embryos. The log_{10}-transformed relative expression levels of 11,602 orthologous genes from mapped-10M reads (data set based on randomly selected 10M tags mapped to the genome; Online Methods), with TMM-normalized data, were graphed on a scatterplot. Essentially the same results were obtained from other data sets (all-read-data, RPKM reads per kilobase per million mapped reads) and TMM normalizations). The results of a statistical test to determine the groups of genes that have more similar expression can be found in the Supplementary Note. (b) Genes that showed a statistically significant increase in their expression level after the phylotypic period (IAP; Online Methods). Each line represents the mean expression level of each IAP (increased expression after the phylotypic gene) calculated, with two biological replications, for each stage. The names of the genes with the top three highest expression levels in TK23 are shown. Consequently, 233 turtle IAP genes were found. See Supplementary Figure 18 for the expression pattern of the chicken orthologs of the turtle IAP genes. (c) Over-represented GO annotations for 233 turtle IAP genes with read depth–controlled, TMM-normalized data. Only the results corroborated by all of the data sets (mapped-10M reads (Online Methods), all reads, RPKM normalization and TMM normalization) are shown. Shown are P values calculated by Fisher’s exact test. (d) High numbers of tissue-specific miRNAs were identified (also in the carapacial ridge) in the embryo after the phylotypic period.

The above results suggest that turtle-specific global repatterning of gene regulation begins after TK11 or the phylotypic period. Although turtle and chicken express many shared developmental genes in the embryo during the putative phylotypic period (Fig. 4a and Supplementary Tables 27 and 28) and have the fewest expanded or contracted gene family members expressed (Supplementary Fig. 16) at this stage, later stages showed increasing differences in their molecular patterns. We found 233 genes that showed turtle-specific increasing expression patterns after the phylotypic type (Fig. 4b). Considering that the chicken orthologs did not show this type of increasing expression (Supplementary Figs. 17 and 18), these 233 genes represent attractive candidates for clarifying the genomic nature of turtle-specific morphological oddities. Furthermore, our Gene Ontology (GO)-based statistical analysis identified many genes that are potentially involved in ossification and extracellular matrix regulation (Fig. 4c), suggesting the involvement of morphological characteristics appearing in turtle embryogenesis, such as extensive ossification in the shell and folding of the body wall[23,24]. The morphological specifications of turtle embryogenesis after the identified phylotypic period include the formation of the novel turtle structure called the carapacial ridge[25,26], which is considered to be responsible for the flabellate expansion of the turtle ribs in late development[27]. Previous molecular studies[27,28] have identified many carapacial ridge–specific coding genes, whereas no study so far has investigated carapacial ridge–specific microRNA (miRNA)

Figure 5 Expression profiling of all 20 soft-shell turtle Wnt genes shows Wnt5a expression in the carapacial ridge. (a) Whole-mount in situ hybridization (ISH) was performed for all Wnt genes in the genome. Wnt5a (red outline) is specifically expressed in the carapacial ridge (red arrowheads), whereas most of the other genes show similar expression patterns to their known mouse and chicken counterparts. Scale bars, 0.5 mm. (b) Soft-shell turtle embryo at stage TK14. (c) Carapacial ridge expression of Wnt5a confirmed by ISH on a 6-µm paraffin transverse section (the sectioned level for ISH is indicated by the dashed line in b). The arrowhead indicates the carapacial ridge; the arrow indicates the body wall. NT, neural tube; NC, notochord. Scale bars in b and c, 0.5 mm.
expression, despite the increasing number of reports claiming the crucial roles of miRNA in various developmental processes. We therefore performed a small RNA-seq analysis of three tissues from soft-shell turtle embryos—limb, body wall and carapacial ridge (Fig. 4d and Supplementary Figs. 19–21)—and further predicted possible miRNAs by referring to the genome sequence (Supplementary Table 29). Unexpectedly, we found expression of a large number of specific miRNAs in all of the tissues (Fig. 4d and Supplementary Table 30), including the carapacial ridge (212 miRNAs). Although no definitive conclusion can be made regarding the functions of these miRNAs, our preliminary prediction-based analysis implied the possible involvement of Wnt signaling (Supplementary Fig. 21 and Supplementary Tables 31–33).

Ann Burke29 was the first to point out the similarity of the apical ectodermal ridge of limbs and the carapacial ridge of the turtle shell. Later, increasing molecular evidence supported this hypothesis. Previous studies27,28 have shown the carapacial ridge-specific shell. Later, increasing molecular evidence supported this hypothesis. Previous studies27,28 have shown the carapacial ridge–specific evolutionary novelties in morphological phenotypes.

In summary, our study both highlights the evolution of the turtle body plan and offers a model to explain, at the genomic level, how the vertebrate developmental program can change to produce major evolutionary novelties in morphological phenotypes.


METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. The Chinese soft-shell turtle and the green sea turtle draft genomes have been deposited in NCBI GenBank under accessions AGCU00000000 and AJIM00000000, respectively. The Chinese soft-shell turtle genome can also be accessed at the Ensembl database. Wnt gene sequences cloned for whole-mount ISH have been deposited in NCBI GenBank under accessions JQ968433–JQ968452. Soft-shell turtle chicken and RNA-seq data have been deposited in the DNA Data Bank of Japan (DDBJ) Sequence Read Archive under accession DRA000567. Soft-shell turtle RNA-seq data for small RNA are available under DDBJ Sequence Read Archive accession DRA000639. Additional information is provided in Supplementary Table 34.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

We would like to acknowledge the International Crocodilian Genomes Working Group, in particular, D.A. Ray of Mississippi State University, R.E. Green of the University of California at Santa Cruz and E.L. Braun of the University of Florida, for making their early drafts of the Crocodylus porosus (saltwater crocodile) and Alligator mississippiensis (American alligator) genomes available for the phylogenetic analysis. We acknowledge P. Martelli at Ocean Park Corporation Hong Kong for providing the green sea turtle samples and Daiyu-yoshoku Ltd. For providing the soft-shell turtle samples. We thank M. Noro, K. Yamanaka, K. Itomi and T. Kitazume for help collecting and sequencing the soft-shell turtle miRNA. The soft-shell turtle genome project was supported in part by Grants-in-Aid from the Ministry of Education, Culture, Sports, Science & Technology, Japan (22128003 and 22128001), the NIBB Cooperative Research Program (Next-generation DNA Sequencing Initiative, 11-730) and the European Molecular Biology Laboratory and the Wellcome Trust (grant numbers WT095908 and WT098951). The green sea turtle genome project was supported by the China National GenBank at BGI-Shenzhen.

AUTHOR CONTRIBUTIONS


COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

Note: this work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/3.0/.


This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/3.0/.


ONLINE METHODS
Source and sequencing of genomic DNA and error correction. The soft-shell turtle was purchased from a local farmer in Japan, and the green sea turtle was provided by the Genome 10K Project (originally collected in Ocean Park, Hong Kong). Genomic DNA was extracted from the whole blood of a female individual in each species, and we constructed a total of 18 (for the soft-shell turtle) and 17 (for the green sea turtle) libraries consisting of short-insert (170-bp, 500-bp and 800-bp) and long-insert (2-kb, 5-kb, 10-kb, 20-kb and 40-kb) libraries. Sequencing was performed using the Illumina HiSeq 2000 system, and read error correction was performed for the short-insert libraries (on the basis of the K-mer frequency distribution curve; Supplementary Note). Data accession numbers are given in Supplementary Table 34.

Genome assembly. Filtered and corrected data were assembled using SOAPdenovo11,32. We first generated contigs by constructing a de Bruijn graph with the reads from the K-mer–split short-insert library data. The graph was then simplified to generate the contigs by removing tips, merging bubbles and solving repeats. All sequenced reads were then realigned onto the contig sequences, and scaffolds were constructed by weighting the rates of consistent and conflicting paired-end relationships. Finally, we retrieved the read pairs with one end that uniquely mapped to the contig and the other end located in the gap region, and performed a local assembly for these collected reads to fill the gaps.

Repeat annotation and whole-genome alignment. Repeat detection was performed using the program RepeatMasker and the Genetic Information Research Institute (GIRI) repeat library. For homology-based prediction of repeats, we used the library of known repeats in the Repbase database (v2008-08-01, Repbase-16.02) with RepeatMasker (v3.2.6) and RepeatProteinMask to identify transposable elements at the DNA and protein levels, respectively. The de novo prediction of repeats involved building a de novo repeat library with RepeatModeler34 and subsequently employing RepeatMasker. Tandem repeats were searched with the Tandem Repeats Finder (TRF)35. Whole-genome pairwise alignments were generated by LASTZ36.

Gene prediction for the two turtles and crocodilians. Gene prediction for the two turtle genomes employed both the ab initio approach (GENSCAN37 [v2.5.5] and AUGUSTUS38 [v1.0]) and a homolog-based approach against the repeat-masked genome, and gene sets predicted by these two approaches were further consolidated with the GLEAN program. For the soft-shell turtle, an additional 146.7 Gb of RNA-seq data was used. The proteins of other vertebrate species were mapped to the genome using TBlastN (Legacy Blast40 [v2.2.3]) and aligned sequences were then filtered and passed to GeneWise41 (v2.2.0) along with the query sequences. The resulting data sets were integrated by GLEAN into a consensus gene set. The best BLASTP match to the SwissProt and TrEMBL databases was used to assign function. The motifs and domains of the gene products were annotated with InterProScan42 against the protein databases ProDom, PRINTS, Pfam, SMART, PANTHER and PROSITE. Gene Ontology IDs for each gene were obtained from the corresponding InterPro entries. The above prediction pipeline was applied to the saltwater crocodile and American alligator genomes (from the Crocodile Genome Consortium), except for the integration step in the latter case. Gene family identification was performed using TreeFam32.

Gene prediction for the soft-shell turtle by the Ensembl prediction pipeline. For gene expression comparison analyses between soft-shell turtle and chicken embryos, we generated and used another soft-shell turtle gene set that was created by the same Ensembl pipeline as the chicken gene set (see URLS).

GO analysis. Over-represented GO terms were investigated by testing (Fisher’s exact test) the bias in frequency toward other GO terms among certain gene sets, using the total set of defined GO terms as a control distribution. Developmental genes (5,659 in total) were defined as genes with developmental GO terms, and developmental GO terms were defined as those with GO:0032502 (developmental process) as an ancestor.

Animal care and use. Experimental procedures and animal care were conducted in strict accordance with guidelines approved by the RIKEN Animal Experiments Committee (Approval IDs H14-23 and H16-10).
RNA-seq for transcriptome identification. Three different types of sequencing were performed for transcriptome identification in the soft-shell turtle: (i) Titanium sequencing (about 2 Gb of clean sequence data), (ii) HiSeq strand-specific paired-end RNA-seq (two libraries were prepared by methods that retain strand-specific information, including a dUTP-based method) (19 Gb of clean data) that was modified to comply with the Illumina TruSeq RNA sample prep kit and an original method developed at BGI Sequencing that was performed with Illumina HiSeq 2000 (26 Gb of clean data) and (iii) HiSeq non-stranded RNA-seq (deep-sequencing data for gene expression analysis) was also used for transcriptome identification. Further details are given in the Supplementary Note.

RNA-seq for gene expression analysis and expression comparison. Biological replicates for each developmental stage were created from an independent sample pool. Extracted mRNA samples were then sequenced with an Illumina HiSeq 2000 instrument. We identified 11,602 one-to-one orthologous genes in the soft-shell turtle and chicken using RBBH information from BLAST+ (v2.2.25)61. Gene expression scores were obtained from RNA-seq data by mapping clean reads to the genome using Burrows-Wheeler Aligner (BWA)62 software (v0.5.9-r16). SAAtools63, BEDTools64 and the DEGseq package65 for R (v2.14.2) were used to calculate the tag count data that were mapped to the coding regions. Normalization of the orthologous gene expression scores was performed with all samples at once by either RPKM or TMM normalization66. Pearson’s correlation coefficients, Spearman correlation coefficients, total Euclidean distances (t-Euclidean) or total Manhattan distances (t-Manhattan) were used to estimate similarities in the gene expression profiles of the two samples being compared. Two independent random selections from all reads were performed to make the mapped-10M reads (sequencing depth–controlled dataset set based on randomly selected 10M tags mapped to the genome) data. The Welch two-sample t test or the Wilcoxon signed-rank test was used to detect the most conserved stages. The Holm-corrected α level was applied for these multiple comparisons. Only results reproduced by the data set from two different normalizations (RPKM67 and TMM66) were considered to be significant.

Genes with a significant increase in expression levels after the phylotypic period. Turtle IAP genes were selected using the following criteria: (i) the mean expression level after the phylotypic period (TK15–TK23) was more than five times higher (Wilcoxon test) than during earlier stages (gastula, neurula, TK7 and TK9) and (ii) the chicken orthologs of the turtle IAP genes (if any) did not show such increases in chicken (the average expression levels in HH28 and HH38 did not show more than five times higher expression than in the Prim–HH14 stages).

Wnt gene identification and cloning and whole-mount ISH. In addition to constructing the predicted gene sets, we manually searched for Wnt genes using TBLASTN. Cloning of the probes and whole-mount ISH were performed using standard methods28 (Supplementary Note).

miRNA extraction, prediction and expression analysis. Small RNA was extracted from dissected tissues using the mirVana microRNA Isolation kit (Life Technologies). Small RNA libraries were prepared and sequenced using an Illumina HiSeq 2000 (24 million reads per sample). These small RNA reads, together with the miRNA sequences from chicken, zebrafish and Anolis carolinensis from miBase (v.18), were used to predict miRNA sequences in the genome. The program mirDeep2 (v2.0.0.3)68 was used to predict miRNAs for this prediction. Only miRNA predictions that had P value lower than a significant Randfold α level (P<0.05 mononucleotide shuffling and 999 permutations; see ref. 68 for details) were taken into account for subsequent comparisons. miRNA target prediction was performed with mirRanda (v3.3) using the annotated 3’ UTRs of soft-shell turtle genes.

Statistical tests. To avoid an inflated type 1 error rate, an α level of 0.01 (further Bonferroni correction in case of multiple comparisons) was accepted for statistical significance throughout the analyses unless otherwise specified. Statistical methods were carefully chosen to properly reflect the population of interest. The Welch two-sample t test was used for two-sample comparisons when the data passed the Kolmogorov-Smirnov test for normal distribution; otherwise, the Wilcoxon signed-rank test was used.