

# Rapid Evolution of Beta-Keratin Genes Contribute to Phenotypic Differences That Distinguish Turtles and Birds from Other Reptiles

Yang I. Li\*, Lesheng Kong, Chris P. Ponting, and Wilfried Haerty\*

Department of Physiology, Anatomy and Genetics, MRC Functional Genomics Unit, University of Oxford, United Kingdom

\*Corresponding authors: E-mail: yang.li@dpag.ox.ac.uk; wilfried.haerty@dpag.ox.ac.uk.

Accepted: April 8, 2013

**Data deposition:** The accession numbers for the genome assemblies used in the analysis and some sequences from other species were retrieved from GenBank and are referenced in the Materials and Methods section.

## Abstract

Sequencing of vertebrate genomes permits changes in distinct protein families, including gene gains and losses, to be ascribed to lineage-specific phenotypes. A prominent example of this is the large-scale duplication of beta-keratin genes in the ancestors of birds, which was crucial to the subsequent evolution of their beaks, claws, and feathers. Evidence suggests that the shell of *Pseudomys nelsoni* contains at least 16 beta-keratins proteins, but it is unknown whether this is a complete set and whether their corresponding genes are orthologous to avian beak, claw, or feather beta-keratin genes. To address these issues and to better understand the evolution of the turtle shell at a molecular level, we surveyed the diversity of beta-keratin genes from the genome assemblies of three turtles, *Chrysemys picta*, *Pelodiscus sinensis*, and *Chelonia mydas*, which together represent over 160 Myr of chelonian evolution. For these three turtles, we found 200 beta-keratins, which indicate that, as for birds, a large expansion of beta-keratin genes in turtles occurred concomitantly with the evolution of a unique phenotype, namely, their plastron and carapace. Phylogenetic reconstruction of beta-keratin gene evolution suggests that separate waves of gene duplication within a single genomic location gave rise to scales, claws, and feathers in birds, and independently the scutes of the shell in turtles.

**Key words:** gene duplication, beta-keratins, Sauropsids, turtle shell.

## Introduction

Gene duplication is one of the principal evolutionary mechanisms that drive molecular diversity (Ohno 1970). Following duplication, genes may be subject to adaptive pressures that reflect species-specific biology, they may be more prone to random genetic drift, or they may be silenced and undergo nonfunctionalization. Among gene duplicates that appear to have been preferentially fixed owing to their novel, beneficial, function are those that encode visual pigments in insects and vertebrates (Yuan et al. 2010; Rennison et al. 2012), venoms in mollusks (Chang and Duda 2012), olfactory and gustatory receptors in insects (Robertson and Wanner 2006; McBride 2007), reptiles, and mammals (Meslin et al. 2011; Jiang et al. 2012), leukocyte receptor in human, mouse, and chicken, and the Hox gene family of transcription factors that regulate animal development (Barten et al. 2001; Nikolaidis et al. 2005; Pearson et al. 2005; Laun et al. 2006).

Many gene family expansions contribute to the emergence of novel, lineage-specific, morphological features. However, few are more striking than those of alpha-keratin genes that have led to the independent appearance of, for example, hair and nails in mammals, wool in sheep, and baleen in whales (Vandebergh and Bossuyt 2012). Beta-keratins, on the other hand, are specific to the Sauropsids (reptiles and birds) and add much more rigidity to the scales of reptiles than alpha-keratins. They contribute to the formation of the hard keratinous claws and scales of reptiles, as well as to the formation of the beaks and feathers in birds (Alibardi et al. 2009).

Previous genome-wide comparative analyses in chicken and zebra finch identified several clusters of beta-keratin genes, the largest two of which occur on chicken microchromosomes 25 and 27 (Greenwold and Sawyer 2010). When compared with expressed sequence tags (ESTs) derived from beak, claws, and feathers, the cluster on microchromosome

27 appears to be composed of beta-keratin genes whose proteins exclusively form the feathers. In contrast, the clusters on microchromosome 25 harbor beta-keratins involved in the formation of claws, feathers, and scales, which have been postulated to have originated from a single progenitor cluster of beta-keratin genes on the same microchromosome in ancestral birds (Greenwold and Sawyer 2010).

In turtles, beta-keratins are assembled into filaments in the outer corneous layers of the scutes (Alibardi 2002; Alibardi et al. 2009) that cover the ventral plastron (12–13 scutes) and the dorsal carapace (37–38 scutes), the two parts of the shell that are unique to the Chelonians. These scutes form hard structures that are thought to protect turtles from predators (Solomon et al. 1986). The evolution of the turtle shell has long fascinated biologists. However, the paucity of ancestral turtle fossils with intermediate shell forms has made its study difficult. Nevertheless, the formation of the shell has been extensively studied in the context of paleontology (Li et al. 2008; Joyce et al. 2009), comparative anatomy (Nagashima et al. 2009), and the development of the turtle bone plates (Nagashima et al. 2007; Lyson and Joyce 2012). In comparison, there have been very few investigations tackling the molecular evolution of the shell. Additionally, these studies have been limited to exploring genes involved in the musculoskeletal development of the shell (Kuraku et al. 2005). The major aim of this study is to gain a better understanding of the evolution of the turtle shell at a molecular level, by studying the evolution of beta-keratins that are components of the turtle scutes. In contrast to the alpha-keratins, for which studies revealed deep conservation across vertebrates (Eckhart et al. 2008), we still lack an understanding of the evolutionary relationships among the beta-keratins found in Reptiles.

Although a large number of beta-keratins have been annotated in bird genomes (International Chicken Genome Sequencing Consortium 2004; Warren et al. 2010), only 17 beta-keratin genes have thus far been identified as being expressed in the skin from the shell, limbs, neck, and tail of the turtle *Pseudemys nelsoni* (Dalla Valle, Nardi, et al. 2009). A phylogenetic analysis indicated that 16 of these 17 beta-keratin genes form a *P. nelsoni*-specific clade. The remaining beta-keratin gene, which is expressed specifically in the digits and claws, clustered with chicken sequences derived from scales and keratinocytes. Still, it remains unknown whether this set of beta-keratin genes is comprehensive, whether lineage-specific beta-keratin duplications can be identified among turtle species, and how these genes relate to avian or other reptilian beta-keratin genes.

With the newly sequenced genomes of three turtles (*Chrysemys picta*, *Chelonia mydas*, and *Pelodiscus sinensis*; Shaffer et al. 2013; Wang et al. 2013) that represent approximately 160 Myr of chelonian evolution, we have identified a total of 200 turtle beta-keratin genes. Further analysis of these genes reveals lineage-specific duplications of the beta-keratin gene family in turtles in the syntenic location to the expansions

of this family in birds. The timing of the emergence of the turtle-specific beta-keratin clade (173–273 Ma) is predicted to have coincided with the emergence of the turtle shell, 230–270 Ma (Li et al. 2008). Consequently, expansions of beta-keratin genes may have contributed to the evolution of two major morphological innovations, the turtle shell and avian feathers.

## Materials and Methods

### Genome Assemblies and Identification of an Initial Set of Turtle Beta-Keratin

The genome assemblies of *C. picta* (GenBank assembly ID: GCA\_000241765.1), *Che. mydas* (GenBank nucleotide ID: AJIM00000000.1), and *Pel. sinensis* (GenBank assembly ID: GCA\_000230535.1) were obtained as part of a collaborative effort between two turtle consortia (Shaffer et al. 2013; Wang et al. 2013). All remaining genomes and annotations were downloaded from Ensembl (release 66, <http://www.ensembl.org/>, last accessed April 26, 2013). Gene prediction by homology was used to predict genes in the *C. picta* genome with GPIPE (Heger and Ponting 2007) using human, anole lizard, chicken, and zebra fish. Orthology relationships were then assigned by OPTIC (Heger and Ponting 2007, 2008) between *C. picta*, human, mouse, opossum, anole lizard, chicken, zebra finch, zebrafish, and pufferfish. Beta-keratins from chicken and lizard were then identified and used to identify beta-keratins in *C. picta*.

### Identifying Additional Beta-Keratins in *C. picta*, *Che. mydas*, and *Pel. Sinensis*

To identify additional beta-keratins, we selected beta-keratins from OPTIC in *C. picta*, anole lizard, chicken, and zebra finch with a core box of 20 residues of a central filament region of beta-keratins, which is highly conserved throughout all reptiles and birds (Alibardi et al. 2009; Greenwold and Sawyer 2010). These beta-keratins were then mapped onto the *C. picta*, *Che. Mydas*, and *Pel. sinensis* genomes using both TBLASTN and BLASTN. Because beta-keratins are encoded within a single exon, the hits were then extended on both flanks to find the longest open reading frame. Finally, all proteins were visually inspected and discarded when containing a premature stop codon or undetermined bases.

### Identifying the Syntenic Location of *C. picta* and *Anolis carolinensis* Beta-Keratin Genes in Relation to the Birds

To identify the syntenic positions in chicken and zebra finch of the *C. picta* beta-keratin genes, one-to-one orthology predictions between *C. picta* and chicken or zebra finch for all nonkeratin genes flanking the beta-keratin gene clusters in *C. picta* were used. The same was done for *A. carolinensis*.

To expand our synteny analysis to *Pel. sinensis* and *Che. mydas*, we used a phylogenetic reconstruction of the

beta-keratin genes from the three turtles. Scaffolds containing two or more beta-keratin genes were considered for this reconstruction.

### Phylogenies of Beta-Keratins in Reptiles and Estimation of the Age of Beta-Keratin Clusters in Turtles

Maximum likelihood trees were built using phyML (Guindon et al. 2010) and the Jones, Taylor, and Thornton model based on the amino acid alignments of beta-keratins from *C. picta*, *Che. mydas*, *Pel. sinensis*, and *P. nelsoni* (GenBank ID: AM765814–AM765818, FM163386–FM163397) along with chicken, zebra finch, crocodile (*Crocodylus niloticus*, GenBank ID: AM765851, AM909650, AM765850, Dalla Valle et al. 2009), and anole lizard beta-keratins. Using the same alignments, we also reconstructed the phylogeny using MrBayes 3.2 (Ronquist et al. 2012, Jones model). Additionally, we used the nucleotide alignments of the beta-keratins in the three turtles and *P. nelsoni* to assess the phylogenetic relationships of these genes in Chelonians. The tree was built under the GTR + I +  $\Gamma$  model implemented in MrBayes (supplementary fig. S1, Supplementary Material online).

We also used the protein alignments to guide a nucleotide sequence alignment using tranalign from the EMBOSS package (Rice et al. 2000). The BEAST software (Drummond and Rambaut 2007) was then used to date beta-keratin duplication events. To find the appropriate model of evolution, we used jModelTest (Posada 2008) and found that the HKY +  $\Gamma$  model best fit the beta-keratin alignment in terms of Akaike Information criterion and number of parameters. The uncorrelated lognormal relaxed-clock mode was employed to allow evolutionary rates to vary along branches. We also used priors for the divergence time between birds and turtles (278.4 Myr) from Pereira and Baker (2006) and the Yule model as in Greenwold and Sawyer (2011). A Markov chain Monte Carlo run of 10 million episodes was used to determine the tree with the highest likelihood.

### Positive Selection Analysis

Using the guided nucleotide alignment and species tree, positive selection associated with gene duplication was tested using PAML (Yang 2007) after removal of gaps and ambiguous sites. Five models were tested to identify residues under positive selection: M0 (one ratio), M1a (neutral), M2a (selection), M3 (discrete), M7 (beta), and M8 (beta and  $\omega$ ). We applied a maximum likelihood ratio test to assess the significance of the difference between the models.

## Results

### Identification and Characteristics of Beta-Keratins in Turtles and Other Reptiles

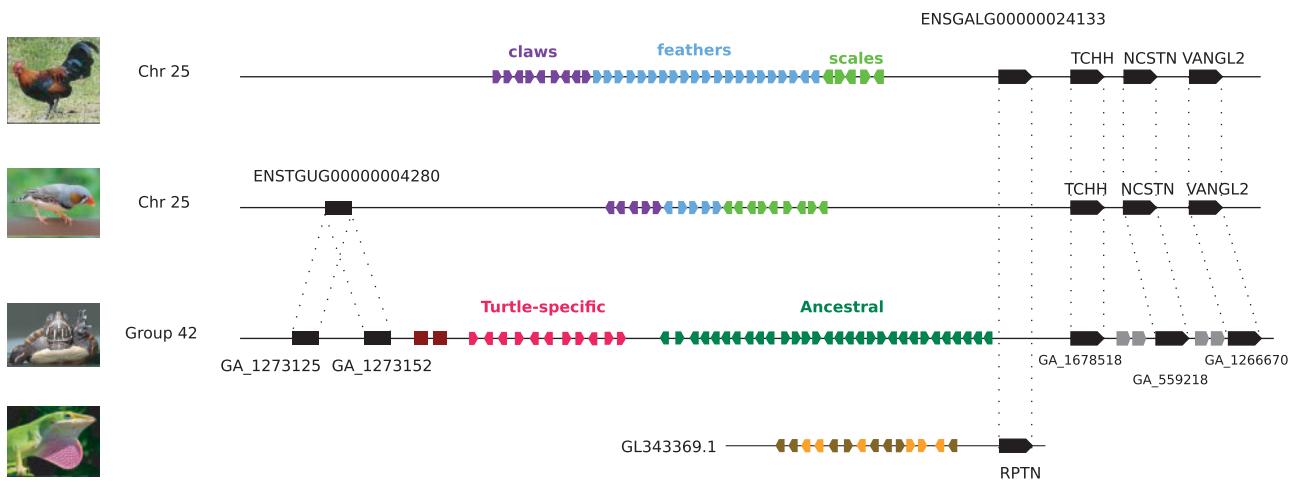
We identified, through homology searches and manual curation (see Materials and Methods), 89, 37, and 74 beta-keratin

genes in the genomes of three turtle species: *C. picta* (the painted turtle), *Che. mydas* (the green sea turtle), and *Pel. sinensis* (the Chinese softshell turtle), respectively (supplementary material S1, Supplementary Material online). Eleven, 31, and 43 additional loci containing beta-keratin-like sequences were either truncated or disrupted by a premature stop codon in *C. picta*, *Che. mydas*, and *Pel. Sinensis*, respectively, and thus were discarded from further analyses. To allow comparison, 106 and 133 beta-keratin genes were identified in the genomes of chicken and zebra finch, respectively, using OPTIC orthology predictions (Heger and Ponting 2007, 2008). However, because of likely differences in genome assembly qualities, it is too early at this stage to draw any conclusion based on differences in the numbers of beta-keratin genes found.

### Synteny of Ancestral Bird and Turtle Beta-Keratins

To identify bird genomic regions in synteny with the beta-keratin clusters of turtles, we performed analyses using 1:1 nonkeratin orthologous genes flanking beta-keratin genes. Of 89 beta-keratin genes in the *C. picta* genome, 43 were located on a single scaffold (Group42) that is syntenic to chicken microchromosome 25 (fig. 1). Although additional *C. picta* beta-keratin genes were located on smaller scaffolds, these have undetermined syntenic relationships with the chicken genome. This is because they consist of beta-keratin genes only or they lack one-to-one orthologs in either chicken or zebra finch, which unambiguously map to a single region. Each of the annotated beta-keratin sequences encodes a region of 20–30 amino acids, which represents the highly conserved core of beta-keratins from other reptiles (see supplementary fig. S2, Supplementary Material online). Similarly, we determined that a beta-keratin gene cluster on scaffold GL343369.1 of the green anole lizard (*A. carolinensis*, Alföldi et al. 2011) lies in conserved synteny with chicken microchromosome 25 (fig. 1). Furthermore, we were able to retrieve EST data for all beta-keratin genes on scaffold GL343369.1 but two and found that they are expressed in the skin of the green anole lizard.

To investigate synteny and orthology relationships among turtle beta-keratin genes, we annotated all beta-keratin genes located in scaffolds containing at least two beta-keratin genes and used phylogenetic reconstruction to determine their orthology relationships (supplementary fig. S3, Supplementary Material online). Although scaffolds harboring *Che. mydas* and *Pel. sinensis* beta-keratin genes are shorter due to their more fragmentary genome assemblies, beta-keratin genes with close orthology relationships to Group42 *C. picta* beta-keratin genes tend to be situated on the same scaffolds, suggesting that they are also located in regions with conserved synteny to microchromosome 25 of chicken (supplementary fig. S3, Supplementary Material online). Other *Che. mydas* and *Pel. sinensis* beta-keratin genes have



**Fig. 1.**—Synteny conservation for the beta-keratin cluster on the bird microchromosome 25 across reptiles. Data from Greenwold and Sawyer (2010) and Dalla Valle et al. (2009) were used to identify beta-keratins involved in the formation of the claws (violet), feathers (blue), and scales (green), and those that are turtle specific (red) and ancestral (dark green). Orthologous genes that are not beta-keratins are depicted in black. Protein coding genes without orthologous relationship are in gray. *Anolis carolinensis* genes displayed in brown are missing from figure 2 due to poor alignments.

undetermined syntenic relationships, suggesting either that they are in a region of the same microchromosome that is hard to assemble (often associated with repeat-rich regions) or on different chromosomes.

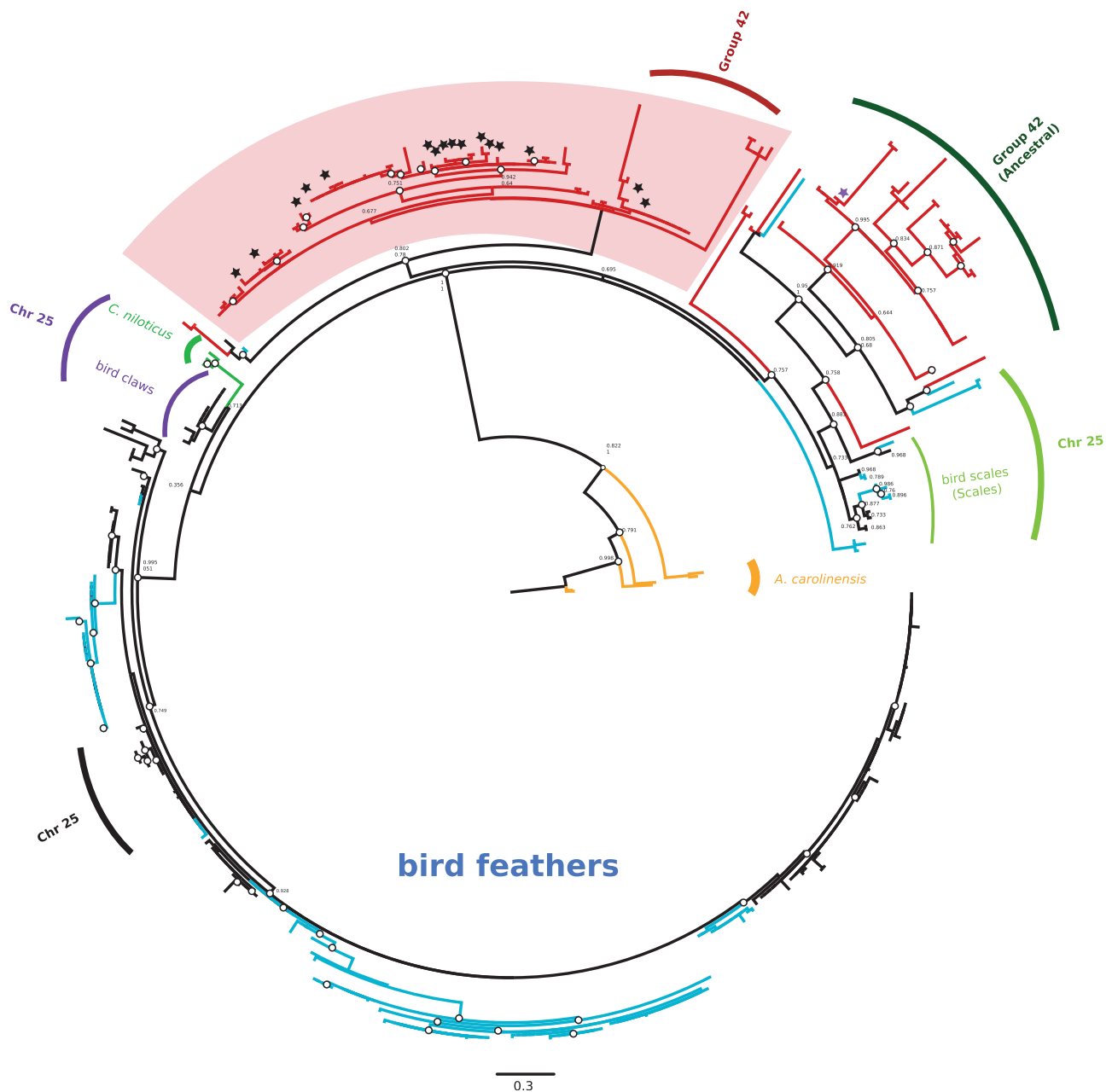
### Turtle-Specific and Avian-Specific Beta-Keratin Gene Family Expansions

Phylogenetic analyses of reptilian beta-keratin genes (figs. 2 and 3) revealed that turtle beta-keratin genes are represented within at least two major clades, one of which is turtle specific, whereas the other is shared with birds (figs. 2 and 3). Consistent with previous analyses (Dalla Valle, Nardi, et al. 2009), 16 of the 17 beta-keratins that were obtained in a study of the *P. nelsoni* shell skin and soft epidermis lie within the turtle-specific clade, whereas the one found to be expressed in the claws was located within the clade shared with birds (fig. 2). It is notable that the turtle-specific beta-keratins are less divergent from one another than are ancestral beta-keratins that clustered with avian beta-keratins (average amino acid identity  $81.1 \pm 1.0\%$  vs.  $56.2 \pm 1.7\%$ ). This would be consistent with turtle-specific beta-keratin genes tending to have arisen, through duplication, more recently than other such genes. The bird beta-keratin genes in this shared clade were previously associated with the formation of the scales based on comparison with ESTs (Greenwold and Sawyer 2010). The largest clade in this phylogeny is avian specific and contains both chicken and zebra finch beta-keratin genes, which are known to be expressed in feathers (Greenwold and Sawyer 2010). As expected, turtle beta-keratins lack a tail sequence specific to previously described feather beta-keratins (Sawyer et al. 2005, [supplementary fig. S2, Supplementary Material online](#)).

Several studies in Squamates and Sauropsids revealed that beta-keratins tend to differ in their amino acid composition according to the morphology and hardness of the tissue (Alibardi et al. 2007). Using the clustering of the turtle beta-keratins with previously identified beta-keratins from *P. nelsoni* (Dalla Valle, Nardi, et al. 2009), some of which were found to be highly expressed in the beta-layers of the scutes, we classified the beta-keratins into candidate (turtle specific) shell and ancestral keratins and compared the amino acid composition of the two groups. We found that turtle-specific beta-keratins tend to have a glycine- and tyrosine-rich tail, which is also observed in all beta-keratin genes of the Nile crocodile scales sequenced thus far (Dalla Valle et al. 2009) or nonfeather beta keratins in birds (Alibardi et al. 2009). In contrast, ancestral beta-keratins appear to have a cysteine and proline rich tail relative to turtle-specific beta-keratins (see [supplementary fig. S4, Supplementary Material online](#)).

### Evidence for Positive Selection in Turtle-Specific Beta-Keratin Genes

We investigated codons of duplicated beta-keratin paralogs for *C. picta*, *Che. Mydas*, and *Pel. sinensis*, independently, for evidence of positive selection acting on amino acid-changing nucleotide substitutions. We found no evidence for positive selection acting on ancestral beta keratins in *C. picta* (M1 vs. M2,  $P > 0.05$ , M7 vs. M8,  $P > 0.05$ ). For both *C. picta* and *Pel. sinensis* candidate shell beta-keratins, the likelihood ratio tests (M7 vs. M8) were significant ( $P = 1.8 \times 10^{-3}$  and  $P = 7.2 \times 10^{-4}$  respectively). However, the sites that were identified as being under positive selection differ (table 1,

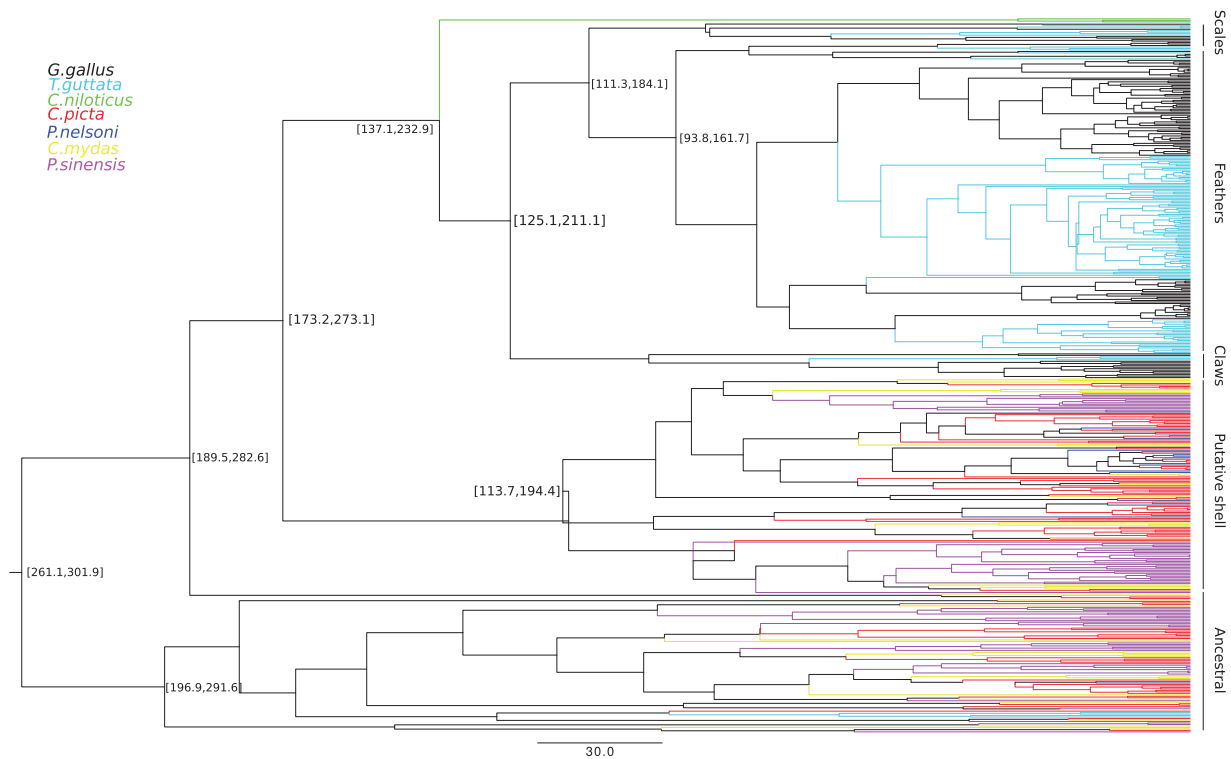


**FIG. 2.**—Identification of turtle beta-keratins potentially associated with the shell formation. Phylogenetic tree of the beta-keratins in *Chrysemys picta* (red), *Anolis carolinensis* (yellow), *Crocodylus niloticus* (green, three proteins), *Gallus gallus* (black), and *Taeniopygia guttata* (blue). The 16 translated cDNA sequences expressed in the skin from shell, soft skin, claws, and digit-tip of *Pseudemys nelsoni* are represented by black stars; the cDNA sequence with tissue specificity to claws and digit tip is labeled with a purple star. The red-shaded area highlights the putative “shell” beta-keratin clade in *C. picta*. The chromosomes and scaffolds represented in figure 1 are displayed above the tree. The association between beta-keratins and tissues was established according to phylogenetic affinity with ESTs (Dalla Valle et al. 2009; Greenwold and Sawyer 2010; bird scales: light green, nonshell: dark green, shell: dark red, bird claws: maroon, and feathers: light blue).

supplementary fig. S2, Supplementary Material online). No evidence for positive selection was found for *Che. mydas* beta-keratin genes.

We searched the scaffold containing the largest number of beta-keratin genes (43), Group42, in *C. picta* and more

specifically the beta-keratin cluster for an increased GC3 content that could be suggestive of biased gene conversion (Duret and Galtier 2009). Similar to microchromosome 25, its syntenic element, Group42 in *C. picta* displays an elevated G+C content relative to the rest of the genome



**Fig. 3.**—Phylogenetic trees of beta-keratins from all species and dating using BEAST. The datings (in Myr), which correspond to 95% confidence intervals, are in square brackets. The turtle-specific clade with 16 out of 17 beta-keratins from *Pseudemys nelsoni* is labeled as putative shell clade. The bird-specific clade has been labeled in the same way as in figure 2. The clade annotated as ancestral corresponds to the clade with turtle, zebra finch beta-keratin genes. A higher resolution and unedited version of this tree can be found as [supplementary figure S3, Supplementary Material](#) online.

**Table 1**

Test for Positive Selection in Beta-Keratins Genes from *Chrysemys picta*, *Pelodiscus sinensis*, and *Chelonia mydas*

Species	Tests	2ΔL	df	P
<i>C. picta</i>	M0 <sup>a</sup> vs. M3 <sup>b</sup>	10.08	3	0.0179
	M1 <sup>c</sup> vs. M2 <sup>d</sup>	4.74	2	0.0935
	M7 <sup>e</sup> vs. M8 <sup>f</sup>	12.68	2	0.0018
<i>P. sinensis</i>	M0 vs. M3	138.24	3	<2.2 × 10 <sup>-16</sup>
	M1 vs. M2	10.14	2	0.0063
	M7 vs. M8	14.48	2	0.0007
<i>C. mydas</i>	M0 vs. M3	22.428	3	5.31 × 10 <sup>-5</sup>
	M1 vs. M2	0.06	2	0.97
	M7 vs. M8	1.1	2	0.577

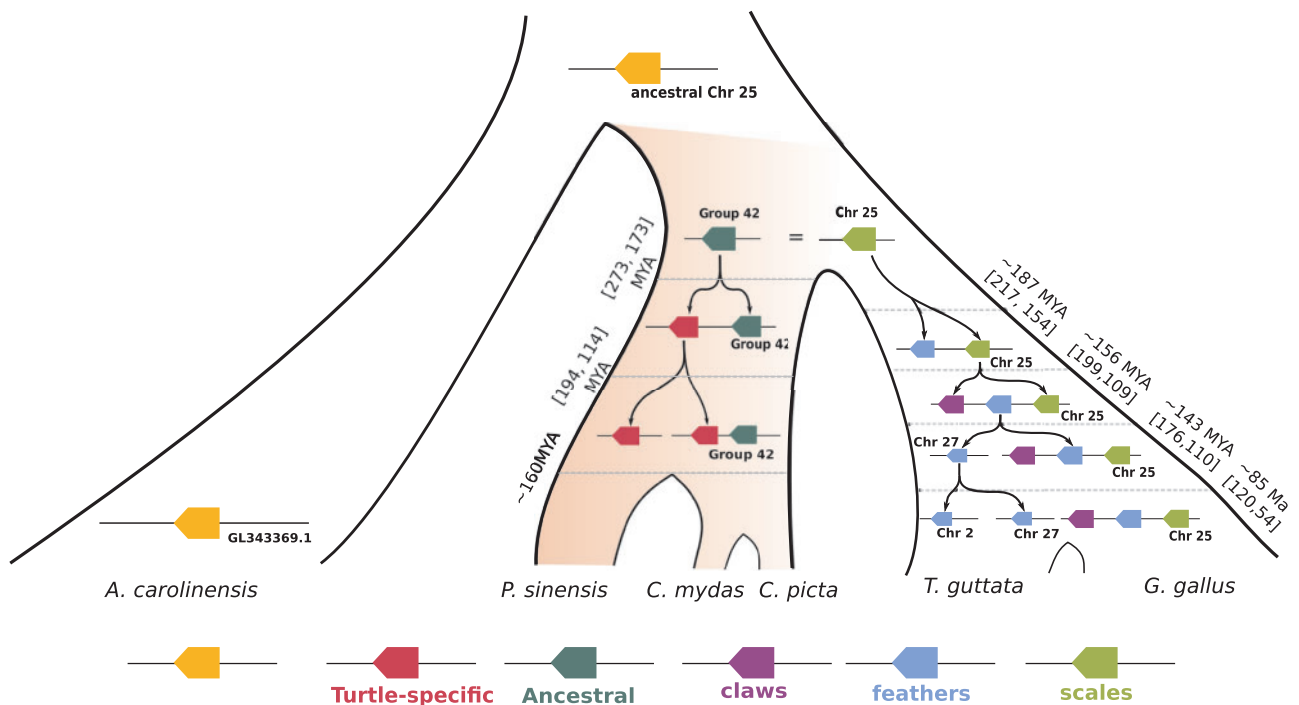
<sup>a</sup>M0: one ratio model.  
<sup>b</sup>M3: discrete.  
<sup>c</sup>M1: neutral.  
<sup>d</sup>M2: positive selection.  
<sup>e</sup>M7: beta.  
<sup>f</sup>M8: beta and ω.

([supplementary fig. S5, Supplementary Material](#) online). However, we found a weak but significant increased GC3 content within candidate shell relative to ancestral beta-keratin genes (Mann–Whitney test,  $P = 0.043$ , [supplementary](#)

[fig. S6, Supplementary Material](#) online). In addition, we ran GENECONV (Sawyer 1999) on the beta-keratin coding sequences located on Group42. However, no further evidence for gene conversion was reported.

### Turtle-Specific Beta-Keratins in the Softshell Turtle

Softshell turtles, such as *Pel. sinensis*, are characterized by the absence of scutes on the carapace and plastron leading to a soft and leathery shell. We therefore compared the beta-keratins of the Chinese softshell turtle *Pel. sinensis* to those of the hard shell turtles, *C. picta* and *Che. mydas*. Using PSIPRED (Jones 1999), we identified two beta-sheets within *Pel. sinensis* beta-keratins similar to those in *C. picta* (see [supplementary fig. S2, Supplementary Material](#) online). We also interrogated a preliminary assembly of the spiny softshell turtle *Apalone spinifera* and found 15 well-conserved beta-keratin genes using Basic Local Alignment Search Tool searches (see Materials and Methods). Contrary to previous speculations (Toni et al. 2007), we observed that beta-keratin genes in both softshell turtles possessed the conserved core box ([supplementary fig. S7, Supplementary Material](#) online).



**FIG. 4.**—Evolutionary model of the beta-keratin genes in the Sauropsids. Evolutionary scenario for the diversification of the beta-keratin clusters on the common ancestor of the chicken microchromosome 25 and *Chrysemys picta* Group42 scaffold. Dates on the turtle's lineages were estimated using BEAST (fig. 3), and dates for the bird lineage were estimated in Greenwold and Sawyer (2011). "Turtle-specific" and "ancestral" annotations are based on the phylogenetic affinity.

### Divergence Time of the Turtle-Specific Beta-Keratin Expansion

To estimate the divergence time of the turtle-specific beta-keratins from the other beta-keratin genes, we used BEAST (Drummond and Rambaut 2007), a Bayesian evolutionary analysis program (see Materials and Methods). The sole molecular dating prior used was the height of the tree, 278.4 Myr, which corresponds to the split between turtles and birds (Pereira and Baker 2006). Results from BEAST indicated that the turtle-specific beta-keratins diverged from the other beta-keratins between 173 and 273 Ma (95% confidence interval). A further estimate from BEAST indicated that additional beta-keratin duplications have taken place, between 114 and 194 Ma (95% confidence interval), after the divergence of the putative shell beta-keratins from other beta-keratins (fig. 4).

## Discussion

The feathers of birds and the shell of turtles are unique morphological traits that set apart the two groups from the other reptilian species. Several studies suggest that the feathers originated from modifications of the scales (Greenwold and Sawyer 2010). For turtles, research has been focused on the development of the shell owing to the extensive

transformation of the skeleton and muscles during turtle evolution (Gilbert et al. 2001), whereas the evolutionary origin of the scutes remains poorly studied. Our study reveals that, concomitant with the formation of the shell, the beta-keratin gene family underwent repeated duplications in the turtle lineage from beta-keratin genes involved in the formation of the claws and scales in the ancestor of the Sauropsids. Previous immunological studies of beta-keratins within the epidermis of various reptilian species suggested a correlation between the type and amount of beta-keratin expressed and epidermis hardness. Consequently, it is possible that duplicated beta-keratin genes were retained in Chelonians because they add rigidity to the shell by increasing the amount of beta-keratins (Alibardi et al. 2007).

### Beta-Keratins in the Scuteless Softshell Turtles

It is believed that the ancestors of softshell turtles possessed a hardshell, which subsequently became scaly and soft through the loss of their scutes (Reisz and Head 2008). According to our beta-keratin annotations, 43 beta-keratin-like genes in the genome of the Chinese soft-shelled turtle, *Pel. sinensis*, were either truncated or disrupted by a premature stop codon compared with only 11 in the painted turtle, *C. picta*. The higher number of disrupted beta-keratin genes in *Pel. sinensis*

compared with *C. picta* is not expected to be caused by the differences in genome assembly quality because the *Pel. sinensis* assembly was assembled using both long and short reads, whereas the assembly of *C. picta* used long reads (the N50 of the *Pel. sinensis* assembly was also higher: 22.2 kb vs. 12.2 kb). On the other hand, the *Che. mydas* genome assembly used short reads only, and comparison is questionable because short-read de novo assembly is known to be more problematic in tandemly duplicated regions such as the beta-keratin loci. The elevated number of disrupted beta-keratin genes in the *Pel. sinensis* genome compared with the *C. picta* genome suggests that the loss of scutes in the softshell turtles was caused by a relaxation of purifying selection on several beta-keratin genes. Retained beta-keratin genes lying in the turtle-specific clade could be more broadly expressed in the turtle epidermis.

An alternative hypothesis to the softshell turtles' loss of scutes was put forward by Toni et al. (2007). Their study of the spiny softshell turtle (*Apa. spinifera*) revealed that beta-keratin bundles generally present in turtles were absent in their soft epidermis (Toni et al. 2007). In situ analyses based on immunostaining showed the presence of beta-keratins without a core box in the epidermis of *Apa. spinifera*. They suggested that the absence—and possibly loss—of the core box, which is associated with two beta-sheets, in beta-keratin genes of *Apa. spinifera* could explain the soft epidermis of the softshell turtles. However, we identified numerous beta-keratins that possessed the core's two predicted beta-sheets within the genome of the Chinese softshell turtle *Pel. sinensis* (see [supplementary fig. S2, Supplementary Material](#) online). To further validate the presence of at least some beta-keratins with the core box in softshell turtles, we searched a preliminary assembly of the *Apa. spinifera* genome for beta-keratins. We identified 15 beta-keratins with conserved cores, which confirmed the existence of beta-keratins with a very well-conserved core box in *Apa. spinifera* (see [supplementary fig. S7, Supplementary Material](#) online) considering more than 160 Myr of divergence from *C. picta* and *Che. mydas*.

### Segmental Duplications Drive the Evolution of the Beta-Keratin Gene Family

Because of duplication processes such as nonallelic homologous recombination, members of a single gene family are likely to be clustered together (Lynch 2007). The gene family consisting of the scale, feather, beak, and claw beta-keratin genes is no exception and was found to be clustered on both microchromosomes 25 and 27 in birds (Greenwold and Sawyer 2010). We found that beta-keratin genes in the painted turtle genome and in anole lizard genome also clustered together and shared syntenic position with microchromosome 25 of the birds. Although we cannot exclude the possibility that beta-keratin genes were translocated early in reptile evolution, it is likely that the first beta-keratin genes originated on the ancestor of the chicken

microchromosome 25. Our phylogenetic analysis revealed that beta-keratin genes with affinity with bird beta-keratin genes are located within regions that are syntenic with Group42, which shares synteny with microchromosome 25 of birds. This further supports the hypothesis that ancestral beta-keratin genes were located on the ancestral microchromosome 25, which subsequently acted as a hotspot for tandem segmental duplication of turtle beta-keratin genes.

We have determined that, compared with the ancestral beta-keratins, the turtle-specific beta-keratins are more similar to one another. This can be explained by a larger number of more recent turtle-specific beta-keratin gene duplications. In addition, gene conversion between paralogous genes within a genomic region can also lead to reduced estimates of divergence and evolutionary rates (Innan and Kondrashov 2010). Because mismatch repair during gene conversion tends to increase G + C content in the affected genomic region (Meunier and Duret 2004), we searched and found a small but significant increase of GC3 within turtle-specific beta-keratin genes relative to ancestral beta-keratin genes. This suggests that biased gene conversion could have been stronger in the shell-related beta-keratin genes of *C. picta*.

Our analysis revealed in positive selection to have occurred in the beta-keratin genes of both *C. picta* and *Pel. sinensis*. Although the sites detected that have evolved under adaptive evolution differ between the two species, they are found in the C-terminal portion of the proteins, outside of the two beta sheets. This result was also reported in the analysis of the green anole genome, where three sites located after the beta sheets were found to have been under positive selection (Alföldi et al. 2011). Beta-keratins have been described to form filaments by interacting in an antiparallel manner at their beta sheets, and the N- and C- termini have been proposed to interact with other proteins, for example, through disulfide bonds (Alibardi et al. 2009). Such potential interactions might explain the observation of selected sites in the C-terminal part of the beta-keratins in multiple Sauropsid lineages.

We further found that beta-keratins within the turtle-specific clade had a glycine- and tyrosine-rich tail, whereas the beta-keratins in the clade shared with avians had a serine-rich tail (see [supplementary fig. S4, Supplementary Material](#) online). These results fit well with previous findings that turtle shell beta-keratins tended to be glycine–proline–tyrosine rich (Dalla Valle, Nardi, et al. 2009).

### Turtle-Specific Beta-Keratin Genes May Have Contributed to the Formation of the Modern Shell

It is generally agreed that beta-keratins are important in the evolution of hard skin appendages in reptiles (see Alibardi et al. [2009] for a review). Our analyses are the first to reveal that independent expansions occurred early (>160 Ma) in the turtle lineage. Additionally, we estimated the



duplication of the feather beta-keratins to be 111–184 Myr old, which agrees with previous estimates (Greenwold and Sawyer 2011). Although it is possible that ancestral beta-keratin-like genes were deleted on both turtle and bird lineages, the presence of only one small clade that contains both turtle and bird beta-keratin genes from our phylogenetic analyses suggests that the ancestor of turtles and birds likely had few beta-keratin genes (<30). In contrast, we found a large turtle-specific clade containing all 16 beta-keratin genes that are expressed in the precursor shell tissue in *P. nelsoni*, suggesting that this clade predominantly contains beta-keratin genes expressed in the shell tissue and whose functions are thus related to the shell (figs. 2 and 3). The only other beta-keratin gene found in *P. nelsoni* was clearly different in terms of sequence (see [supplementary fig. S7, Supplementary Material](#) online) and corresponds to a beta-keratin that shows expression in claws and digit-tips only (Dalla Valle, Nardi, et al. 2009).

#### Divergence Time of Turtle-Specific Beta-Keratin Clade Coincides with the Appearance of the First Turtles

We obtained an approximate estimate of 173–273 Myr for the divergence time between turtle-specific beta-keratins and the other beta-keratins. Gene conversion can lead to an underestimation of the age of duplication events (Teshima and Innan 2004; Innan and Kondrashov 2010). Nevertheless, our results are in line with previous estimates of over 160 Myr for the divergence time between *Pel. sinensis* and the common ancestor of *C. picta* and *Che. mydas* (Near et al. 2005) and estimates of the appearance of the first turtles, some 230–270 Ma (Li et al. 2008). Therefore, our evidence suggests that the beta-keratin expansion in turtles coincided with the emergence of turtles and the innovation of the turtle shell. Furthermore, we found that the turtle-specific beta-keratin clade was divided into two subclades, which diverged some time between 114 and 194 Ma. Again, gene conversion might have led to an underestimation of the divergence age, which means that these two subclades may have diverged earlier. Because *Pel. sinensis* beta-keratins were found in both subclades, we were able to further narrow the divergence time to at least 160 Ma, which suggests that this duplication happened early in turtle evolution.

Previously, our understanding of the turtle shell evolution has been limited by the scarcity of intermediate turtle forms. The lack of intermediate forms has reinforced the *de novo* model of shell evolution, and some have described the shell to have appeared within a short geological time frame through the differentiation of dermal bones (Gilbert et al. 2001). On the other hand, the composite model posits that the rigid armored body of turtles evolved gradually, in multiple steps (Lee 1996; Cebra-Thomas et al. 2005). The discovery of rapid beta-keratin gene evolution early in turtle history

supports the idea that intermediate forms existed for a short period (20–90 Myr) after the first Chelionans appeared and rapidly evolved thereafter into the modern turtles. In fact, the recent discovery of *Odontochelys* (Li et al. 2008), the oldest turtle fossils described thus far (220 Myr), which lack a carapace but possess a fully formed plastron, accompanied by comparative anatomy work between the Chinese soft-shelled turtle and other amniotes (Nagashima et al. 2009), supports a two-step scenario for the evolution of the shell. It is thus possible that the beta-keratin divergence early in turtle evolution corresponded to a subfunctionalization event in which duplicated plastron scute beta-keratins acted as a substrate for the origin of the beta-keratins in the scutes of the modern turtle carapace.

By determining the phylogeny of beta-keratin genes in *C. picta*, *Che. mydas*, and *Pel. sinensis*, together with beta-keratin sequences from chicken, zebra finch, anole lizard, and *P. nelsoni*, we identified monophyletic turtle-specific genes that show evidence of expression in turtle shell skin. Many of these genes lie in a region of these turtles' genomes that is syntenic to microchromosome 25 of the chicken. This primordial cluster of beta-keratin genes is thus the building block that conceivably allowed two independent phenotypic innovations—innovations that differentiate turtles and birds from other reptiles. We provide the sequence of 200 manually curated turtle beta-keratin genes, which can be used to study the evolution of the turtle shell. Our data along with previous analyses pertaining to the evolution of the beta-keratins in birds (Greenwold and Sawyer 2010, 2011) provide us with a better understanding of the evolutionary trajectory of the beta-keratins in the Sauropsids and how it relates to different morphological features (fig. 4).

This study is the first, to our knowledge, to report a large-scale expansion of beta-keratin genes in turtles and to propose an association between this expansion and the innovation of the turtle shell. Although further functional studies are needed to determine the role of turtle-specific beta-keratins in the formation of the turtle shell, this study is the first to suggest that large-scale independent expansions of a single gene family contributed to the evolution of two different synapomorphies. Additionally, we also envisage that our characterization of turtle beta-keratin genes will allow researchers to investigate the role of beta-keratin evolution on turtle scales and claws by comparing smaller turtle-specific clades with their sister clades, for example, within the ancestral beta-keratin clade.

#### Supplementary Material

Supplementary figures S1–S7 and material S1 are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

## Acknowledgments

The authors thank Prof. Bradley Shaffer, the painted turtle genome consortium, and the softshell/green shell turtle genome consortium for their data and support, as well as Matthew Fujita, Chris Rands, and Ron Nudel for reading through preliminary versions of this manuscript. This work was supported by the University of Oxford Nuffield Department of Medicine Prize scholarship to Y.I.L., by the European Research Council to L.K., and by the Medical Research Council UK to W.H. and C.P.P.

## Literature Cited

- Alföldi J, et al. 2011. The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature* 477(7366): 587–591.
- Alibardi L. 2002. Immunocytochemical observations on the cornification of soft and hard epidermis in the turtle *Chrysemys picta*. *Zoology (Jena)* 105(1):31–44.
- Alibardi L, Dalla Valle L, Nardi A, Toni M. 2009. Evolution of hard proteins in the sauropsid integument in relation to the cornification of skin derivatives in amniotes. *J Anat.* 214(4):560–586.
- Alibardi L, Toni M, Dalla Valle L. 2007. Hard cornification in reptilian epidermis in comparison to cornification in mammalian epidermis. *Exp Dermatol.* 16:961–976.
- Barten R, Torkar M, Haude A, Trowsdale J, Wilson MJ. 2001. Divergent and convergent evolution of nk-cell receptors. *Trends Immunol.* 22(1): 52–57.
- Cebra-Thomas J, et al. 2005. How the turtle forms its shell: a paracrine hypothesis of carapace formation. *J Exp Zool B Mol Dev Evol.* 304(6): 558–569.
- Chang D, Duda TF Jr. 2012. Extensive and continuous duplication facilitates rapid evolution and diversification of gene families. *Mol Biol Evol.* 29(8):2019–2029.
- Dalla Valle L, et al. 2009. Beta-keratins of the crocodylian epidermis: composition, structure, and phylogenetic relationships. *J Exp Zool B Mol Dev Evol.* 312(1):42–57.
- Dalla Valle L, Nardi A, Toni M, Emera D, Alibardi L. 2009. Beta-keratins of turtle shell are glycine-proline-tyrosine rich proteins similar to those of crocodylians and birds. *J Anat.* 214(2):284–300.
- Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol.* 7:214.
- Duret L, Galtier N. 2009. Biased gene conversion and the evolution of mammalian genomic landscapes. *Annu Rev Genomics Hum Genet.* 10:285–311.
- Eckhart L, et al. 2008. Identification of reptilian genes encoding hair keratin-like proteins suggests a new scenario for the evolutionary origin of hair. *Proc Natl Acad Sci U S A.* 105(47):18419–18423.
- Gilbert SF, Loredó GA, Brukman A, Burke AC. 2001. Morphogenesis of the turtle shell: the development of a novel structure in tetrapod evolution. *Evol Dev.* 3(2):47–58.
- Greenwold MJ, Sawyer RH. 2010. Genomic organization and molecular phylogenies of the beta (beta) keratin multigene family in the chicken (*Gallus gallus*) and zebra finch (*Taeniopygia guttata*): implications for feather evolution. *BMC Evol Biol.* 10:148.
- Greenwold MJ, Sawyer RH. 2011. Linking the molecular evolution of avian beta ( $\beta$ ) keratins to the evolution of feathers. *J Exp Zool B Mol Dev Evol.* 316(8):609–616.
- Guindon S, et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of phylml 3.0. *Syst Biol.* 59(3):307–321.
- Heger A, Ponting CP. 2007. Evolutionary rate analyses of orthologs and paralogs from 12 drosophila genomes. *Genome Res.* 17(12): 1837–1849.
- Heger A, Ponting CP. 2008. Optic: orthologous and paralogous transcripts in clades. *Nucleic Acids Res.* 36(Database issue):D267–D270.
- Innan H, Kondrashov F. 2010. The evolution of gene duplications: classifying and distinguishing between models. *Nat Rev Genet.* 11(2): 97–108.
- International Chicken Genome Sequencing Consortium. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432(7018):695–716.
- Jiang P, et al. 2012. Major taste loss in carnivorous mammals. *Proc Natl Acad Sci U S A.* 109(13):4956–4961.
- Jones DT. 1999. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol.* 292(2):195–202.
- Joyce WG, Lucas SG, Scheyer TM, Heckert AB, Hunt AP. 2009. A thin-shelled reptile from the Late Triassic of North America and the origin of the turtle shell. *Proc Biol Sci.* 276(1656):507–513.
- Kuraku S, Usuda R, Kuratani S. 2005. Comprehensive survey of carapacial ridge-specific genes in turtle implies co-option of some regulatory genes in carapace evolution. *Evol Dev.* 7(1):3–17.
- Laun K, et al. 2006. The leukocyte receptor complex in chicken is characterized by massive expansion and diversification of immunoglobulin-like loci. *PLoS Genet.* 2(5):e73.
- Lee MSY. 1996. Correlated progression and the origin of turtles. *Nature* 379:812–815.
- Li C, Wu X-C, Rieppel O, Wang L-T, Zhao L-J. 2008. An ancestral turtle from the late triassic of southwestern china. *Nature* 456(7221): 497–501.
- Lynch M. 2007. The origins of genome architecture. Sunderland (MA): Sinauer Associates.
- Lyson TR, Joyce WG. 2012. Evolution of the turtle bauplan: the topological relationship of the scapula relative to the ribcage. *Biol Lett.* 8: 1028–1031.
- McBride CS. 2007. Rapid evolution of smell and taste receptor genes during host specialization in drosophila sechellia. *Proc Natl Acad Sci U S A.* 104(12):4996–5001.
- Meslin C, et al. 2011. The evolutionary history of the *sal1* gene family in eutherian mammals. *BMC Evol Biol.* 11:148.
- Meunier J, Duret L. 2004. Recombination drives the evolution of GC-content in the human genome. *Mol Biol Evol.* 21(6):984–990.
- Nagashima H, et al. 2007. On the carapacial ridge in turtle embryos: its developmental origin, function and the chelonian body plan. *Development* 134(12):2219–2226.
- Nagashima H, et al. 2009. Evolution of the turtle body plan by the folding and creation of new muscle connections. *Science* 325(5937): 193–196.
- Near TJ, Meylan PA, Shaffer HB. 2005. Assessing concordance of fossil calibration points in molecular clock studies: an example using turtles. *Am Nat.* 165(2):137–146.
- Nikolaidis N, et al. 2005. Origin and evolution of the chicken leukocyte receptor complex. *Proc Natl Acad Sci U S A.* 102(11):4057–4062.
- Ohno S. 1970. The evolution by gene duplication. Berlin (Germany): Springer-Verlag.
- Pearson JC, Lemons D, McGinnis W. 2005. Modulating Hox gene functions during animal body patterning. *Nat Rev Genet.* 6(12): 893–904.
- Pereira SL, Baker AJ. 2006. A mitogenomic timescale for birds detects variable phylogenetic rates of molecular evolution and refutes the standard molecular clock. *Mol Biol Evol.* 23(9):1731–1740.
- Posada D. 2008. jmodeltest: phylogenetic model averaging. *Mol Biol Evol.* 25(7):1253–1256.
- Reisz RR, Head JJ. 2008. Palaeontology: turtle origins out to sea. *Nature* 456(7221):450–451.

- Rennison DJ, Owens GL, Taylor JS. 2012. Opsin gene duplication and divergence in ray-finned fish. *Mol Phylogenet Evol.* 62(3): 986–1008.
- Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* 16(6):276–277.
- Robertson HM, Wanner KW. 2006. The chemoreceptor superfamily in the honey bee, *Apis mellifera*: expansion of the odorant, but not gustatory, receptor family. *Genome Res.* 16(11):1395–1403.
- Ronquist F, et al. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol.* 61(3): 539–542.
- Sawyer RH, Rogers L, Washington L, Glenn TC, Knapp LW. 2005. Evolutionary origin of the feather epidermis. *Dev Dyn.* 232(2): 256–267.
- Sawyer S. 1999. GENECONV: a computer package for the statistical detection of gene conversion. Department of Mathematics, Washington University in St. Louis.
- Shaffer HB, et al. 2013. The western painted turtle genome, a model for the evolution of extreme physiological adaptations in a slowly evolving lineage. *Genome Biology*, doi:10.1186/gb-2013-14-3-r28.
- Solomon SE, Hendrickson JR, Hendrickson LP. 1986. The structure of the carapace and plastron of juvenile turtles, *Chelonia mydas* (the green turtle) and *Caretta caretta* (the loggerhead turtle). *J Anat.* 145: 123–131.
- Teshima KM, Innan H. 2004. The effect of gene conversion on the divergence between duplicated genes. *Genetics* 166(3):1553–1560.
- Toni M, Valle LD, Alibardi L. 2007. Hard (beta-)keratins in the epidermis of reptiles: composition, sequence, and molecular organization. *J Proteome Res.* 6(9):3377–3392.
- Vandebergh W, Bossuyt F. 2012. Radiation and functional diversification of alpha keratins during early vertebrate evolution. *Mol Biol Evol.* 29(3): 995–1004.
- Wang, et al. Forthcoming 2013. The draft genomes of the sift-shell turtle and green sea turtle yield insights into the development and the evolution of the turtle-specific body plan. *Nature Genetics*.
- Warren WC, et al. 2010. The genome of a songbird. *Nature* 464(7289): 757–762.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24(8):1586–1591.
- Yuan F, Bernard GD, Le J, Briscoe AD. 2010. Contrasting modes of evolution of the visual pigments in heliconius butterflies. *Mol Biol Evol.* 27(10):2392–2405.

**Associate editor:** Ross Hardison