Received: 21 November 2021 | Revised: 21 July 2022 | Accepted: 31 July 2022

DOI: 10.1111/jeb.14097

JOURNAL OF Evolutionary Biology eseb

SHORT COMMUNICATION

The eggshell-matrix protein gene *OC-17* is functionally lost in the viviparous Chinese crocodile lizard

Hong-Xin Xie[1,2] | Xi-Xi Liang[1] | Wei-Ming Li[1,2] | Zhi-Qiang Chen[3] | Xi-Feng Wang[1] | Zi-Han Ding[1,2] | Xu-Ming Zhou[1] | Wei-Guo Du[1]

[1]Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing, China

[2]University of Chinese Academy of Sciences, Beijing, China

[3]Novogene Bioinformatics Institute, Beijing, China

Correspondence
Wei-Guo Du or Xu-Ming Zhou, Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China.
Emails: duweiguo@ioz.ac.cn; zhouxuming@ioz.ac.cn

Funding information
National Natural Science Foundation of China, Grant/Award Number: 31821001, 31901223 and 32170528

Abstract

Thickness reduction or loss of the calcareous eggshell is one of major phenotypic changes in the transition from oviparity to viviparity. Whether the reduction of eggshells in viviparous squamates is associated with specific gene losses is unknown. Taking advantage of a newly generated high-quality genome of the viviparous Chinese crocodile lizard (*Shinisaurus crocodilurus*), we found that ovocleidin-17 gene (*OC-17*), which encodes an eggshell matrix protein that is essential for calcium deposition in eggshells, is not intact in the crocodile lizard genome. Only *OC-17* transcript fragments were found in the oviduct transcriptome, and no OC-17 peptides were identified in the eggshell proteome of crocodile lizards. In contrast, OC-17 was present in the eggshells of the oviparous Mongolia racerunner (*Eremias argus*). Although the loss of *OC-17* is not common in viviparous species, viviparous squamates show fewer intact eggshell-specific proteins than oviparous squamates. Our study implies that functional loss of eggshell-matrix protein genes may be involved in the reduction of eggshells during the transition from oviparity to viviparity in the crocodile lizard.

KEYWORDS
eggshell reduction, eggshell-matrix proteins, ovocleidin-17, squamates, viviparity

1 | INTRODUCTION

The evolution of viviparity was estimated to have occurred independently more than 100 times in squamates (Blackburn, 2006). During the transition from oviparity to viviparity, thickness reduction of eggshells is one of the key phenotypic changes, which allows for oxygen diffusion between the maternal uterus and developing embryos (Murphy & Thompson, 2011). There are emerging studies to investigate the genetic basis of eggshell reduction in squamates. For example, previous genomic and transcriptomic analyses suggested that gene-expression changes could account for the degeneration of eggshell glands in the viviparous lizard (*Phrynocephalus vlangalii*), therefore leading to eggshell reduction (Gao et al., 2018). A recent genome-wide association study (GWAS) on the common lizard (*Zootoca vivipara*) also revealed that the lizard's eggshell thickness was genetically determined and identified 17 SNPs associated with eggshell traits (Recknagel et al., 2021). However, whether gene loss plays a role in eggshell reduction of squamates is unknown. Gene loss is important in the evolution of new traits (Blumer et al., 2022; Hecker et al., 2019; Huelsmann et al., 2019). Loss of genes prohibits reversion to the ancestral state and could be an adaptive force in evolution (Albalat & Cañestro, 2016; Olson, 1999; Smith & Rausher, 2011). The loss of eggshells in viviparous squamates could be associated with key genes that play eggshell-specific roles.

Eggshells are formed in the maternal oviduct after ovulation and mainly composed of deposited calcium carbonate and proteins (termed eggshell-matrix proteins; Rose & Hincke, 2009; Stewart et al., 2010). Eggshell-matrix protein components consist of three protein groups:

Hong-Xin Xie and Xi-Xi Liang contributed equally.

egg white proteins expressed in the egg white and the eggshell, ubiquitous proteins present in many tissues, and eggshell-specific proteins unique to the eggshell (Mann, 2015; Mann & Mann, 2015; Rose & Hincke, 2009). Two possible functions of eggshell-specific proteins have been proposed: regulation of calcium deposition in eggshells and antimicrobial defence (Rose & Hincke, 2009). A total of six eggshell-specific proteins have been reported in the eggshells of birds: ovocleidin-17 (OC-17), ovocleidin-116 (OC-116), ovocalyxin-32 (OCX-32), ovocalyxin-36 (OCX-36), ovocalyxin-25 (OCX-25), and ovocalyxin-21 (OCX-21; Mann & Mann, 2015). OC-17 is the first reported eggshell-specific protein, and it is one of the major proteins in chicken eggshells (Hincke et al., 1995). It has biomineralization activity and is suggested to function in the initiation of calcite deposition in bird eggshells (Freeman et al., 2010, 2011; Lakshminarayanan et al., 2002; Reyes-Grajeda et al., 2004; Yu et al., 2013). OC-17 also has antimicrobial activity (Mine et al., 2003; Wellman-Labadie et al., 2008). OC-116 is also a major component of chicken eggshell and is suggested to function in maintaining calcium deposition in eggshells (Hincke et al., 1999). OCX-32 is mainly present in uterine fluid during the late stage of eggshell calcification and is therefore surmised to be involved in the termination of calcification (Gautron et al., 2001; Hincke et al., 2003). A correlation between haplotypes of *OCX-32* and eggshell strength was also reported (Dunn et al., 2009; Takahashi et al., 2010). Ocx-36 is homologous to proteins that perform innate immune functions, so it is supposed to function in antimicrobial defence (Gautron et al., 2007). Reports about OCX-25 and OCX-21 are scarce, and their functions in eggshells are unknown. OCX-25 is a protease inhibitor with WAP and Kunitz domains (Mann & Mann, 2015). OCX-21 is identical to chicken gastrokie-2 which belongs to the gastric mucosal secretome (Marie et al., 2015).

The Chinese crocodile lizard (*Shinisaurus crocodilurus*) is a viviparous squamate endemic to evergreen mountain forests in southern China and northern Vietnam and is currently endangered (Huang, 2009; van Schingen et al., 2015). It is the only living species of the family Shinisauridae. The divergence time of the crocodile lizard and the closest living relatives (monitor lizards) is about 100 Mya (Xie et al., 2022; Zheng & Wiens, 2016). The crocodile lizard was estimated to have an ancient origin of viviparity during the Cretaceous period, which is older than most Cenozoic origins of viviparity in squamates (Wright et al., 2015). In captivity, female crocodile lizards give birth to fully developed young after a gestation time of up to 10 months (Li et al., 2019), which is rare even in viviparous lizards (Olsson & Shine, 1998). Neonates are surrounded with a transparent, water-rich membrane (Figure 1a), which is often broken during parturition. The loss of calcareous eggshells in the crocodile lizard enables us to explore relevant genetic changes of viviparity. Due to its special characteristics and endangered status, a high-quality reference genome of the crocodile lizard has been published recently, along with genome and transcriptome sequencing data (Xie et al., 2022). Here, we make use of currently available squamate genomes to explore the possible gene loss associated with the reduction of eggshells in viviparous squamates, focusing on eggshell-specific proteins and the crocodile lizard. By manually annotating eggshell-specific genes in the crocodile lizard genome and subsequent transcriptomic and proteomic explorations, we present here evidence that gene loss is potentially relevant to the loss of eggshells in the crocodile lizard and provide a preliminary view of eggshell-matrix proteins of viviparous and oviparous squamates.

2 | MATERIALS AND METHODS

2.1 | Annotation of eggshell-specific genes

The six avian eggshell-specific proteins were selected for manual annotation in genomes of the crocodile lizard and other 17 representative squamate species (Tables S1 and S2). The 18 species we analysed include all viviparous lizards with the genome available to date (three species), four viviparous snakes, six oviparous lizards, and five oviparous snakes to facilitate comparison between modes of reproduction. Annotation was performed using a homology-based method, following manual check. Eggshell-specific proteins were first identified from chicken, so the protein sequences of chicken were used as references, which were obtained from the UniProtKB or NCBI's NR database. Then, Tblastn (Altschul et al., 1990) was done using protein queries against the genome, and the hit region was extracted with 2000-bp extensions. In this step, the orthologue sequences from *Anolis carolinensis* were used as reference sequences for squamates because it is the first squamate genome published and is well-annotated in NCBI's NR database (Alföldi et al., 2011). In the case that the orthologue sequence of *A. carolinensis* is absent, the sequence from *Podarcis muralis*'s chromosome-level genome was used as the reference query. A further annotation was performed against the extracted region using GeneWise v2.4.1 (Birney et al., 2004). The same annotation process was performed independently on every squamate genome. Phylogenetic trees were then built using IQtree v1.6.12 (Nguyen et al., 2015) to check the orthology of annotated genes. Finally, the integrity of the genes was checked manually using the online tool "CD-search" at the NCBI website (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). Incomplete genes were reannotated from the second step with 10 000-bp extensions to exclude possible annotation errors. After annotation, orthologue genes missing conserved protein domains or with premature stop codons were assigned as fragmented genes. For fragmented genes predicted in the crocodile lizard genome, we further checked possible genome assembly errors by (1) mapping the sequencing reads used for genome assembly (Xie et al., 2022) back to the genome; (2) checking the synteny information of neighbouring genes; and (3) conducting Sanger sequencing on the gene region.

2.2 | Transcriptome data to facilitate annotation

We first collected transcriptome data to validate expression of genes annotated in the crocodile lizard genome and search for possible exons missing in the genome. The oviduct of a female crocodile lizard, which was used for genome sequencing, was sampled for RNA extraction.

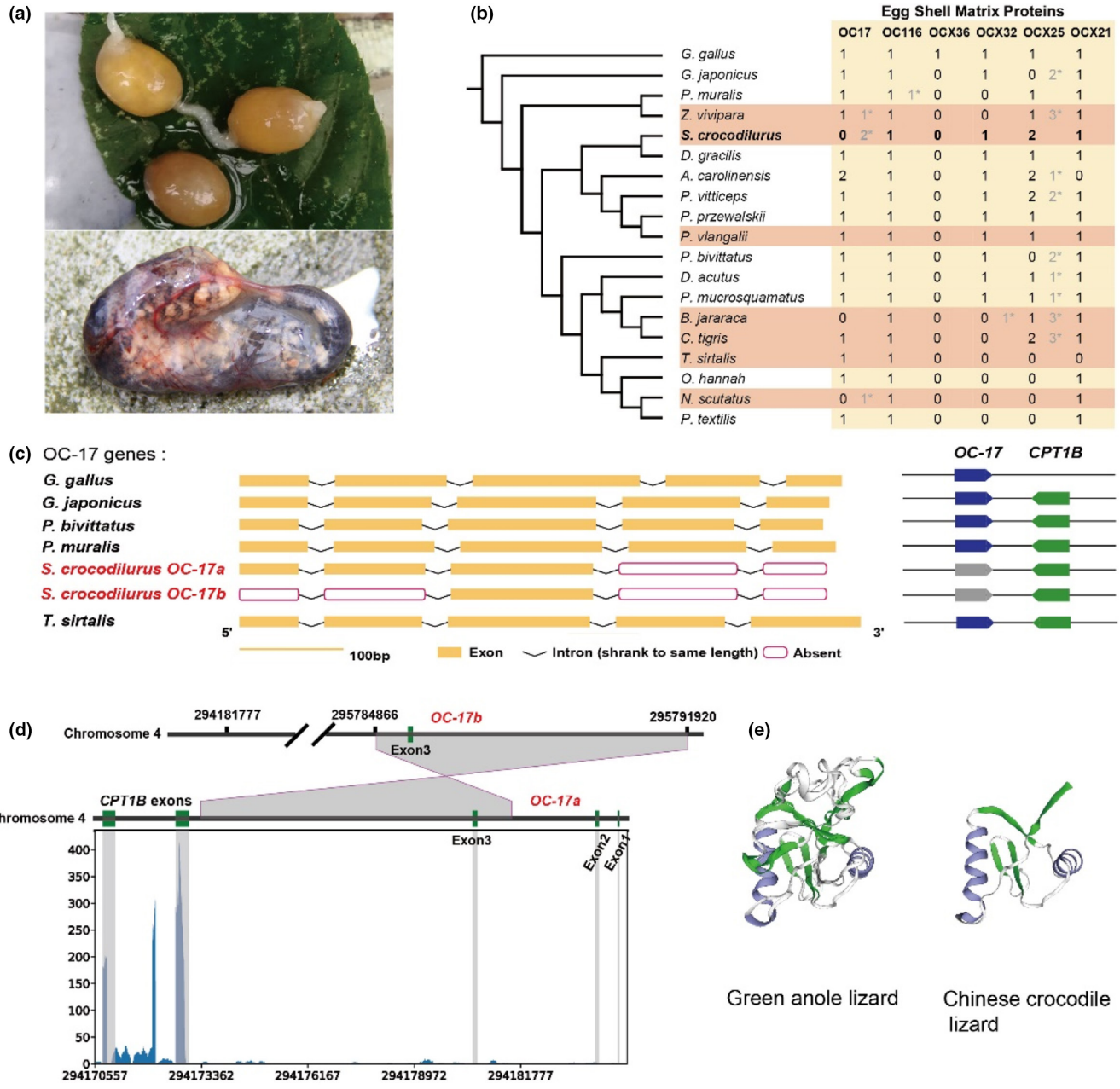

FIGURE 1 Functional loss of OC-17 in the crocodile lizard. (a) Pictures of unfertilized eggs (upper) of crocodile lizards and a newborn lizard with unbroken membrane (lower) showing the absence of calcareous eggshells. Note that the membrane surrounding neonates is usually broken during parturition. (b) Annotation of six eggshell-specific proteins in 18 squamates. Names of reference proteins of chicken are presented. Species tree was adopted from TimeTree (Kumar et al., 2017). Chicken was added as an outgroup. Viviparous species are highlighted in orange. Black numbers indicate intact genes. Grey numbers following an asterisk represent fragmented genes. (c) Schematic view of OC-17 gene structure and syntenic relationship with the downstream gene *CPT1B* in squamates (*CPT1B* is absent in chicken). Fragmented genes of the crocodile lizard are shown. Species are shown in alphabetical order. See Figure S1 for the OC-17 gene structure and synteny of all the species we looked at. (d) Gene loci of *OC-17a* and *OC-17b* and transcriptome read mapping of *OC-17a* in the crocodile lizard genome. (e) Comparison of modelled 3D structures of the OC-17 of the green anole lizard and the crocodile lizard. Helices and strands are shown in blue and green, respectively. Loss of two exons in the crocodile lizard led to tremendous alteration of the protein

The lizard died of an infection in August (during reproductive season), 2017, at the breeding center of Daguishan Nature Reserve, Guangxi, China, and was sampled immediately with the approval of the local committee. Unfertilized eggs were present in its oviduct. Libraries for sequencing were constructed using the NEBNext UltraTM RNA Library Prep Kit for Illumina (NEB, USA) following the manufacturer's recommendations. A paired-end library was then sequenced on Illumina NovaSeq platform. Transcriptome data of other six tissues (Gonad, heart, kidney, liver, lung, and skin) of this individual have been published before and were also collected (Xie et al., 2022). Sequencing reads were filtered by removing adapters and low-quality reads (reads with more than 10% unidentified nucleotides or reads with more than

50% low-quality bases [Phred Q-score < 10]). Both reference-genome-based mapping analysis and *de novo* assembling of transcripts were then conducted. Clean reads were mapped to the genome assembly using HISAT2 v2.0.4b (Pertea et al., 2016). Only one mismatch was allowed for a properly mapped read. *De novo* assembling of transcripts was conducted using Trinity (version r2012-10-05) (Grabherr et al., 2011) with default parameters. We then collected transcriptome evidence for other species from public resources. Most species we analysed have genome annotation information on the NCBI. So for fragmented genes in the genome, gene integrity was further manually checked by aligning the gene to NCBI's NR database to find possible intact transcripts using Tblastn. Genome information for *P. vlangalii* was not available on the NCBI. So, we collected *P. vlangalii*'s transcriptome data from Gao et al. (2018). We used the transcriptome data of the oviduct from the stage before ovulation because this is the stage when the eggshell gland is active. Data were analysed using the same methods implemented on the transcriptome of the crocodile lizard.

2.3 | Protein structure prediction

Secondary structure was predicted using the PSIPRED web server (http://bioinf.cs.ucl.ac.uk/psipred/; Buchan & Jones, 2019). The signal peptide was predicted on the SignalP-5.0 web server (http://www.cbs.dtu.dk/services/SignalP/; Almagro Armenteros et al., 2019). Homology modelling of protein 3D structure was conducted by SWISS-MODEL (https://swissmodel.expasy.org/; Waterhouse et al., 2018).

2.4 | Lizard eggshell proteome

Unfertilized eggs of crocodile lizards were collected from the breeding center of Daguishan Nature Reserve (24°09′N, 111°81′E), Guangxi, China. To serve as a positive control, eggshells of an oviparous lizard, the Mongolia racerunner (*Eremias argus*), were collected from wild populations in Shierliancheng Field Station of the Institute of Grassland Research of the Chinese Academy of Agricultural Sciences (40°12′N, 111°07′E), Inner Mongolia, China for comparison with viviparous crocodile lizards. The racerunner was chosen because fresh eggshells could be collected from semi-natural enclosures set up for academic research (Hao et al., 2021), and more importantly, the whole-genome sequence resources and gene annotation of this lizard, which are needed for proteome identification, are also available (Li et al., 2022). Eggshells were stored in −20°C before processing. Four eggshells were combined to form a proteome sample, and three samples were analysed for each species. An extra sample of the Mongolia racerunner was analysed to confirm the existence of ovocleidin-17 (OC-17) in lizard eggshells. Whole eggshell matrix proteins were extracted for SDS-PAGE and LC–MS analyses. Raw data were processed using MaxQuant v1.6.1.0 (Tyanova et al., 2016) to search against predicted protein sequences of the crocodile lizard and the Mongolia racerunner. A sample from Crocodile lizards showed extremely low protein concentration, so after initial protein identification, the sample with smallest protein number was excluded for both species to facilitate fair comparison. The final accepted proteins for each species were obtained from two samples. See supplementary note for details of protein extraction, LC–MS analysis, and parameters for protein identification.

2.5 | Statistical test of the number of intact eggshell-specific proteins

We implemented the Phylogenetic Generalized-Least Squares (PGLS) model using the PHYTOOLS package in R v4.0.5 to test the correlation between modes of reproduction and the number of intact proteins. In our univariate PGLS model, we scored the oviparous mode of reproduction "0" and the viviparous mode of reproduction "1" to quantify the relationship between the modes of reproduction and the protein number of the squamate species we annotated.

3 | RESULTS AND DISCUSSION

3.1 | Functional loss of *OC-17* in the crocodile lizard

Of the six eggshell-specific genes known in birds, we found there is one intact *OC-116* in all squamates examined, while *OCX-36* was absent in squamates. For the other four genes, the number of intact genes vary in different species (Figure 1b). Specifically, we found two fragmented *OC-17*, which share 98.55% sequence identity, in the crocodile lizard genome (Figure 1c,d). Multi-copies of *OC-17* are also observed in other species, such as in *A. carolinensis* (two intact, 90.16% identity), in *Z. vivipara* (one intact and one fragmented, 98.45% identity), and in birds like the ostrich (Struthiocalcin-1 and Struthiocalcin-2, 43.18% identity; Mann & Siedler, 2004). The two copies of OC-17 in the crocodile lizard and other squamates may originate from recent lineage-specific gene duplication as they share high sequence identity and are within the same gene order with the downstream *CPT1B* (Figure S1), while the homologues in birds may have more ancient origin. Further explorations of the OC-17 genes in the crocodile lizard revealed that the exon 4 and exon 5 are missing in the longer copy of *OC-17* (*OC-17a*), and the other copy (*OC-17b*) consists of a single exon 3 (Figure 1c). (1) The last two exons of *OC-17* are missing in the crocodile lizard genome. Both the Tblastn analyses using the OC-17 protein sequence of *A. carolinensis* and the last two exons of *Dopasia gracilis* against the chromosome-level genome (Xie et al., 2022) and another published scaffold-level genome (Gao et al., 2017) of the crocodile lizard received no hit. (2) No assembly error was found in the genomic region of *OC-17*. The location of *OC-17* and the intergenic region between *OC-17* and the downstream *CPT1B* are fully covered by sequencing reads, and no assembly anomalies exist (Figure S2). Overlapping PCR experiments (about 1000 bp amplicon) succeeded in amplifying over 4000 bp from the third exon of *OC-17* towards *CPT1B* (Figure S3, Table S3), which confirmed no assembly anomalies in the gene region. (3) *OC-17*

transcript is also fragmented. The crocodile lizard's *OC-17* expressed at very low levels in the oviduct (Figure 1d, Table S4). Direct mapping of the crocodile lizard's RNA-seq to intact genes of other squamate species received no hits. The *de novo* assembled transcripts of the crocodile lizard's oviduct only contained a fragmented transcript of *OC-17*, which consistently does not contain the last two exons.

Comparison of the structure of OC-17 proteins of squamates and birds showed that they have similar secondary structure containing two α-helices and eight β-sheets (Figure S4a), which is a common characteristic of the C-type lectin family (Ruiz-Arellano et al., 2015). The homology-based 3D-structure model of squamate OC-17 (XP_008121228.2) showed high similarity with Struthiocalcin-1 (the OC-17 of ostrich; GMQE = 0.78, QMEAN = −1.80; Benkert et al., 2010; Waterhouse et al., 2018; Studer et al., 2019; Figures S4b,c). A signal peptide was also predicted in the N-terminal of the protein, a feature of extracellular proteins (Figure S4d). These results indicate that OC-17 in squamates may serve a homologous function to the one observed in birds. The loss of last two exons in crocodile lizard *OC-17* results in the loss of four conserved β-sheets (β5, β6, β7, and β8), of which β5, β6, and β7 make up the upper side of the protein and β8 stabilizes the lower subdomain through Cys31-Cys128 binding with helix α1 (Ruiz-Arellano et al., 2015). Therefore, we surmised that the encoded protein structure of *OC-17* in crocodile lizards is dramatically altered (Figure 1d).

The *OC-116* (Figure S5), *OCX-32*, *OCX-25*, and *OCX-21* genes are intact in crocodile lizards. However, except for *OCX-32* that was highly expressed in several tissues, other genes showed limited expression or no expression (Table S4). Interestingly, OCX-32 was the only eggshell-specific protein identified in crocodile lizards' eggshells (Table S5).

3.2 | Identification of OC-17 in squamate eggshells

We analysed eggshell proteomes of the viviparous crocodile lizard and the oviparous Mongolia racerunner. PAGE separation of total protein showed different band patterns in these two species, with the eggshell matrix of Mongolia racerunner more prominent in bands below 20 kDa than that of the crocodile lizard (Figure 2a). OC-17 in

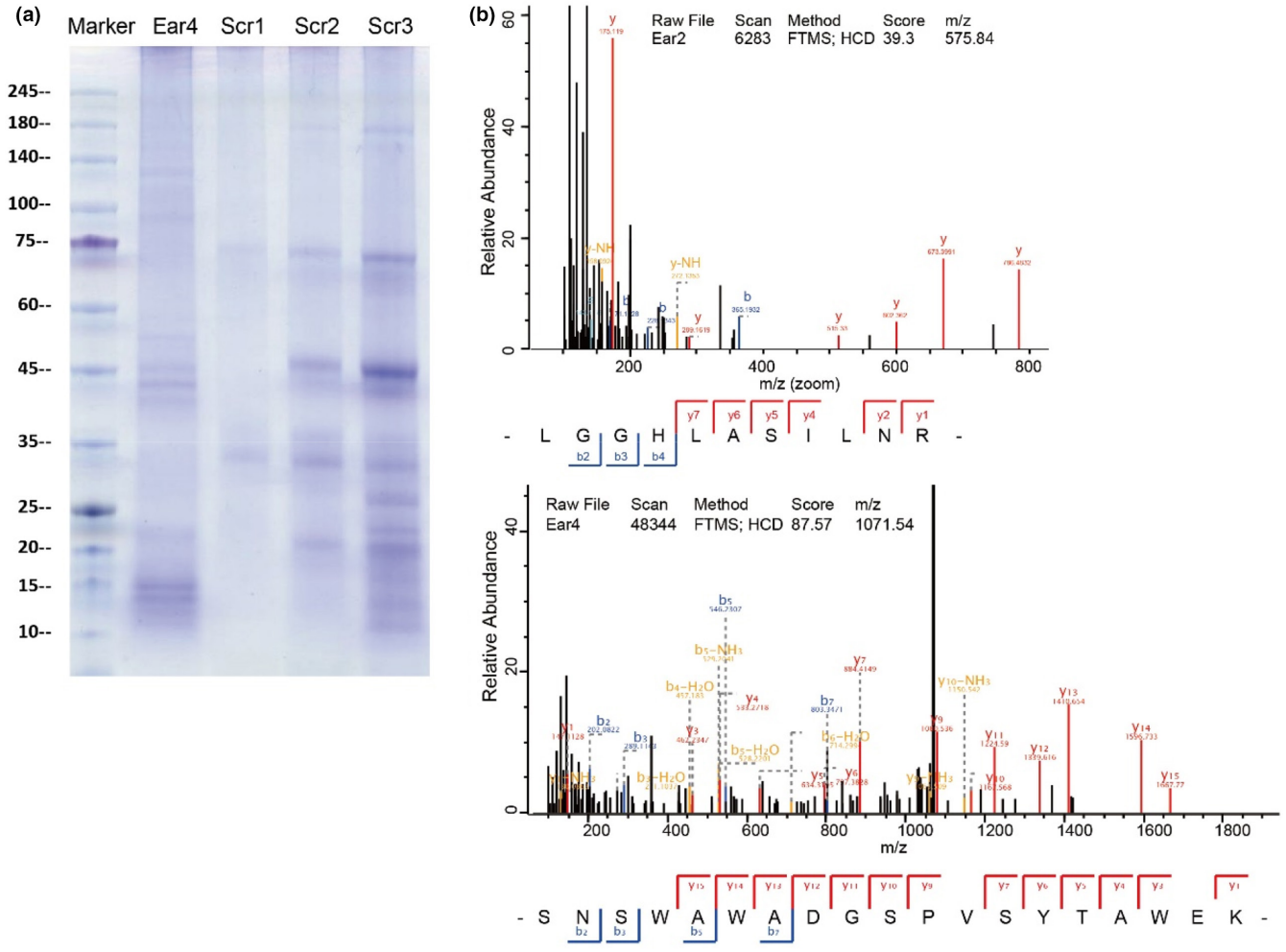

FIGURE 2 Identification of OC-17 in eggshell proteome. (a) Comparison of SDS-PAGE separation of total eggshell matrices. Lane ear, the eggshell matrix of the oviparous Mongolia racerunner. Lanes Scr1-Scr3, eggshell matrices of the viviparous Chinese crocodile lizard. Molecular weight of markers is shown in the unit of kDa. (b) Spectra of the two OC-17 peptides identified in the eggshell matrix of Mongolia racerunner. Y-ions and b-ions are shown in red and blue, respectively. Ions with water or ammonia losses are in orange

birds is located in this mobility region (Mann & Mann, 2015). We identified the protein sequence of Mongolia racerunner's OC-17 in its eggshell with two unique peptides (Figure 2b). On the contrary, no related OC-17 peptides were found in crocodile lizard eggshells. Apart from OC-17, OCX-32 and OCX-25 were identified in the crocodile lizard and the Mongolia racerunner, respectively (Tables S5 and S6). OCX-25 is also one of the most abundant proteins in eggshells of Mongolia racerunner (Table S6).

3.3 | Viviparous squamates show fewer intact eggshell-specific proteins

The loss of *OC-17* is not commonly shared in viviparous squamates. For the other six viviparous species we annotated, intact *OC-17* was found in *Z. vivipara*, *P. vlangalii* (fragmented in the genome, but the complete coding sequence was obtained with transcriptome data), *Crotalus tigris*, and *Thamnophis sirtalis*. *Notechis scutatus* and *Bothrops jararaca* presumably lost the functional *OC-17* as they do not contain an intact gene in the genome and transcriptome data. However, other genes, including *OCX-32*, *OCX-25*, and *OCX-21*, are lost or fragmented in some squamates, indicating that the correlation of integrity of eggshell-specific protein genes and the reduction of eggshells are more complex. We then implemented a PGLS model to detect the relationship between modes of reproduction and the number of intact eggshell-specific proteins, and we found that modes of reproduction negatively correlate with the number of proteins ($\beta = -0.426 \pm 0.154$, $t_{[18]} = -2.767$, $p = 0.014$), suggesting that viviparous species have a reduced intact eggshell-specific protein number when compared to oviparous species in our data, which is especially the case in snakes (Figure 1b). For some species with recent origins of viviparity, such as *Z. vivipara* and *P. vlangalii*, they might have achieved eggshell reduction through genetic changes other than gene loss, which have been explored based on transcriptome data and GWAS (Gao et al., 2018; Recknagel et al., 2021). We provide evidence that *OC-17* is functionally lost in the genome of crocodile lizards. However, given the long independent evolutionary history of the crocodile lizard and lack of fossil records, it is of challenge to infer the timeline of the loss of *OC-17* in the genome and the loss of calcareous eggshells, so a causal link between the gene loss and trait evolution is not possible. Based on the observation that *OC-17* is intact in some viviparous species and other eggshell-specific genes is lost in different species, the loss of *OC-17* in the crocodile lizard might more likely have happened after the evolution of viviparity due to the relaxed selection. However, the loss of *OC-17* prevents reversal to oviparity and therefore may serve as an evolutionary constraint.

In conclusion, our analyses on eggshell-specific genes in squamates identified a species-specific functional loss of OC-17, a protein essential for the formation of calcareous eggshells, in the viviparous crocodile lizard. This, to the best of our knowledge, is the first report of the functional loss of a gene that is associated with the reduction of eggshell thickness in a squamate reptile (Blackburn, 1995; Cabej, 2019; Foster et al., 2020; Gao et al., 2018; Griffith et al., 2016; Murphy & Thompson, 2011). Future studies making use of whole-genome sequence of closely related oviparous and viviparous species could help understand the molecular mechanisms involved in multi-origins of viviparity in squamates.

AUTHOR CONTRIBUTIONS

W.-G.D., X.-M.Z., H.-X.X., and X.-X.L. conceived the study. H.X.X., X.-X.L., Z.-Q.C., and W.-M.L. performed data analyses. X.-F.W. and Z.-H.D. conducted lab work. H.-X.X., X.-X.L., X.-M.Z., and W.-G.D. wrote and revised the manuscript. W.-G.D. and X.-M.Z. jointly supervised this work.

ACKNOWLEDGEMENTS

We thank Shanghai Bioprofile Technology Co. Ltd (Shanghai, China) for help in proteome analyses. We thank Gideon Deme Gywa for help in PGLS model analyses. We thank Shu-Yi Luo for providing pictures of crocodile lizards' eggs and the newborn lizard. We thank Xin Hao and Yu-Jie Yang for help in collecting lizard eggshells. This project was supported by National Natural Science Foundation of China (31821001, 31901223, and 32170528).

CONFLICT OF INTEREST

The authors declare no competing interests.

DATA AVAILABILITY STATEMENT

The oviduct RNA-seq data have been deposited in NCBI's BioProject under the accession number PRJNA756104. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD029080. The raw sequencing data generated and other datasets supporting this work, including the annotation results of eggshell matrix genes, *de novo* assembly of transcripts, and the structure modelling results of OC-17 proteins, have also been deposited in Science Data Bank (http://doi.org/10.57760/sciencedb.02480).

ORCID

Hong-Xin Xie https://orcid.org/0000-0001-8012-6575
Wei-Guo Du https://orcid.org/0000-0002-2275-2591

REFERENCES

Albalat, R., & Cañestro, C. (2016). Evolution by gene loss. *Nature Reviews Genetics*, *17*, 379–391.

Alföldi, J., Di Palma, F., Grabherr, M., Williams, C., Kong, L., Mauceli, E., Russell, P., Lowe, C. B., Glor, R. E., Jaffe, J. D., Ray, D. A., Boissinot, S., Shedlock, A. M., Botka, C., Castoe, T. A., Colbourne, J. K., Fujita, M. K., Moreno, R. G., ten Hallers, B. F., … Lindblad-Toh, K. (2011). The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature*, *477*, 587–591.

Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., von Heijne, G., & Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology*, *37*, 420–423.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*, 403–410.

Benkert, P., Biasini, M., & Schwede, T. (2010). Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*, *27*, 343–350.

Birney, E., Clamp, M., & Durbin, R. (2004). Genewise and Genomewise. *Genome Research*, *14*, 988–995.

Blackburn, D. G. (1995). Saltationist and punctuated equilibrium models for the evolution of viviparity and placentation. *Journal of Theoretical Biology*, *174*, 199–216.

Blackburn, D. G. (2006). Squamate reptiles as model organisms for the evolution of viviparity. *Herpetological Monographs*, *20*, 131–146.

Blumer, M., Brown, T., Freitas, M. B., Destro, A. L., Oliveira, J. A., Morales, A. E., Schell, T., Greve, C., Pippel, M., Jebb, D., Hecker, N., Ahmed, A.-W., Kirilenko, B. M., Foote, M., Janke, A., Lim, B. K., & Hiller, M. (2022). Gene losses in the common vampire bat illuminate molecular adaptations to blood feeding. *Science Advances*, *6494*, eabm6494.

Buchan, D. W. A., & Jones, D. T. (2019). The PSIPRED protein analysis Workbench: 20 years on. *Nucleic Acids Research*, *47*, W402–W407.

Cabej, N. R. (2019). Origins of evolutionary novelty. In N. R. Cabej (Ed.), *Epigenetic principles of evolution* (pp. 379–492). Elsevier.

Dunn, I. C., Joseph, N. T., Bain, M., Edmond, A., Wilson, P. W., Milona, P., Nys, Y., Gautron, J., Schmutz, M., Preisinger, R., & Waddington, D. (2009). Polymorphisms in eggshell organic matrix genes are associated with eggshell quality measurements in pedigree Rhode Island red hens. *Animal Genetics*, *40*, 110–114.

Foster, C. S. P., Thompson, M. B., Van Dyke, J. U., Brandley, M. C., & Whittington, C. M. (2020). Emergence of an evolutionary innovation: Gene expression differences associated with the transition between oviparity and viviparity. *Molecular Ecology*, *29*, 1315–1327.

Freeman, C. L., Harding, J. H., Quigley, D., & Rodger, P. M. (2010). Structural control of crystal nuclei by an eggshell protein. *Angewandte Chemie*, *49*, 5135–5137.

Freeman, C. L., Harding, J. H., Quigley, D., & Rodger, P. M. (2011). Simulations of ovocleidin-17 binding to calcite surfaces and its implications for eggshell formation. *Journal of Physical Chemistry C*, *115*, 8175–8183.

Gao, J., Li, Q., Wang, Z., Zhou, Y., Martelli, P., Li, F., Xiong, Z., Wang, J., Yang, H., & Zhang, G. (2017). Sequencing, *de novo* assembling, and annotating the genome of the endangered Chinese crocodile lizard *Shinisaurus crocodilurus*. *GigaScience*, *6*, 1–6.

Gao, W., Sun, Y., Zhou, W., Xiong, Z., Chen, L., Li, H., Fu, T.-T., Xu, K., Xu, W., Ma, L., Chen, Y.-J., Xiang, X.-Y., Zhou, L., Zeng, T., Zhang, S., Jin, J.-Q., Chen, H.-M., Zhang, G., Hillis, D. M., … Che, J. (2018). Genomic and transcriptomic investigations of the evolutionary transition from oviparity to viviparity. *Proceedings of the National Academy of Sciences of the United States of America*, *116*, 3646–3655.

Gautron, J., Hincke, M. T., Mann, K., Panheleux, M., Bain, M., McKee, M. D., Solomon, S. E., & Nys, Y. (2001). Ovocalyxin-32, a novel chicken eggshell matrix protein. Isolation, amino acid sequencing, cloning, and immunocytochemical localization. *The Journal of Biological Chemistry*, *276*, 39243–39252.

Gautron, J., Murayama, E., Vignal, A., Morisson, M., McKee, M. D., Réhault, S., Labas, V., Belghazi, M., Vidal, M.-L., Nys, Y., & Hincke, M. T. (2007). Cloning of ovocalyxin-36, a novel chicken eggshell protein related to lipopolysaccharide-binding proteins, bactericidal permeability-increasing proteins, and plunc family proteins. *The Journal of Biological Chemistry*, *282*, 5273–5286.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., … Regev, A. (2011). Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nature Biotechnology*, *29*, 644–652.

Griffith, O. W., Brandley, M. C., Belov, K., & Thompson, M. B. (2016). Reptile pregnancy is underpinned by complex changes in uterine gene expression: A comparative analysis of the uterine transcriptome in viviparous and oviparous lizards. *Genome Biology and Evolution*, *8*, 3226–3239.

Hao, X., Zou, T.-T., Han, X.-Z., Zhang, F.-S., & Du, W.-G. (2021). Grow fast but don't die young: Maternal effects mediate life-history tradeoffs of lizards under climate warming. *The Journal of Animal Ecology*, *90*, 1550–1559.

Hecker, N., Sharma, V., & Hiller, M. (2019). Convergent gene losses illuminate metabolic and physiological changes in herbivores and carnivores. *Proceedings of the National Academy of Sciences of the United States of America*, *116*, 3036–3041.

Hincke, M. T., Gautron, J., Mann, K., Panhéleux, M., McKee, M. D., Bain, M., Solomon, S. E., & Nys, Y. (2003). Purification of ovocalyxin-32, a novel chicken eggshell matrix protein. *Connective Tissue Research*, *44*(Suppl 1), 16–19.

Hincke, M. T., Gautron, J., Tsang, C. P., McKee, M. D., & Nys, Y. (1999). Molecular cloning and ultrastructural localization of the core protein of an eggshell matrix proteoglycan, ovocleidin-116. *The Journal of Biological Chemistry*, *274*, 32915–32923.

Hincke, M. T., Tsang, C. P. W., Courtney, M., Hill, V., & Narbaitz, R. (1995). Purification and immunochemistry of a soluble matrix protein of the chicken eggshell (ovocleidin 17). *Calcified Tissue International*, *56*, 578–583.

Huang, W. (2009). Present status of Chinese xenosaurs and its protection. *Chinese Journal of Wildlife*, *30*, 287–289.

Huelsmann, M., Hecker, N., Springer, M. S., Gatesy, J., Sharma, V., & Hiller, M. (2019). Genes lost during the transition from land to water in cetaceans highlight genomic changes associated with aquatic adaptations. *Science Advances*, *5*, eaaw6671.

Kumar, S., Stecher, G., Suleski, M., & Hedges, S. B. (2017). TimeTree: A resource for timelines, timetrees, and divergence times. *Molecular Biology and Evolution*, *34*, 1812–1819.

Lakshminarayanan, R., Kini, R. M., & Valiyaveettil, S. (2002). Investigation of the role of ansocalcin in the biomineralization in goose eggshell matrix. *Proceedings of the National Academy of Sciences of the United States of America*, *99*, 5155–5159.

Li, Q., Luo, S., Yang, C., Li, S., Guo, J., He, J., Chen, Y., Huang, C., Wu, Z., & Du, W. (2019). Impacts of maternal characteristics and temperature on juvenile survival in the crocodile lizard: Implications for conservation. *Zoo Biology*, *38*, 272–280.

Li, W.-M., Du, J., Yang, L.-Y., Liang, Q.-Q., Yang, M.-Y., Zhou, X.-M., & Du, W.-G. (2022). Chromosome-level genome assembly and population genomics of Mongolian racerunner (*Eremias argus*) provide insights into high-altitude adaptation in lizards. Under review.

Mann, K. (2015). The calcified eggshell matrix proteome of a songbird, the zebra finch (*Taeniopygia guttata*). *Proteome Science*, *13*, 1–20.

Mann, K., & Mann, M. (2015). Proteomic analysis of quail calcified eggshell matrix: A comparison to chicken and Turkey eggshell proteomes. *Proteome Science*, *13*, 1–19.

Mann, K., & Siedler, F. (2004). Ostrich (*Struthio camelus*) eggshell matrix contains two different C-type lectin-like proteins. Isolation, amino acid sequence, and posttranslational modifications. *Biochimica et Biophysica Acta*, *1696*, 41–50.

Marie, P., Labas, V., Brionne, A., Harichaux, G., Hennequet-Antier, C., Nys, Y., & Gautron, J. (2015). Quantitative proteomics and bioinformatic analysis provide new insight into protein function during avian eggshell biomineralization. *Journal of Proteomics*, *113*, 178–193.

Mine, Y., Oberle, C., & Kassaify, Z. (2003). Eggshell matrix proteins as defense mechanism of avian eggs. *Journal of Agricultural and Food Chemistry*, *51*, 249–253.

Murphy, B. F., & Thompson, M. B. (2011). A review of the evolution of viviparity in squamate reptiles: The past, present and future role of molecular biology and genomics. *Journal of Comparative Physiology B*, *181*, 575–594.

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating

maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*, 268–274.
Olson, M. V. (1999). When less is more: Gene loss as an engine of evolutionary change. *American Journal of Human Genetics*, *64*, 18–23.
Olsson, M., & Shine, R. (1998). Timing of parturition as a maternal care tactic in an alpine lizard species. *Evolution*, *52*, 1861–1864.
Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., & Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols*, *11*, 1650–1667.
Recknagel, H., Carruthers, M., Yurchenko, A. A., Nokhbatolfoghahai, M., Kamenos, N. A., Bain, M. M., & Elmer, K. R. (2021). The functional genetic architecture of egg-laying and live-bearing reproduction in common lizards. *Nature Ecology & Evolution*, *5*, 1546–1556.
Reyes-Grajeda, J. P., Moreno, A., & Romero, A. (2004). Crystal structure of ovocleidin-17, a major protein of the calcified *Gallus gallus* eggshell. *Journal of Biological Chemistry*, *279*, 40876–40881.
Rose, M. L. H., & Hincke, M. T. (2009). Protein constituents of the eggshell: Eggshell-specific matrix proteins. *Cellular and Molecular Life Sciences*, *66*, 2707–2719.
Ruiz-Arellano, R. R., Medrano, F. J., Moreno, A., & Romero, A. (2015). Structure of struthiocalcin-1, an intramineral protein from *Struthio camelus* eggshell, in two crystal forms. *Acta Crystallographica Section D: Biological Crystallography*, *71*, 809–818.
Smith, S. D., & Rausher, M. D. (2011). Gene loss and parallel evolution contribute to species difference in flower color. *Molecular Bology and Evolution*, *28*, 2799–2810.
Stewart, J. R., Mathieson, A. N., Ecay, T. W., Herbert, J. F., Parker, S. L., & Thompson, M. B. (2010). Uterine and eggshell structure and histochemistry in a lizard with prolonged uterine egg retention (Lacertilia, Scincidae, Saiphos). *Journal of Morphology*, *271*, 1342–1351.
Studer, G., Rempfer, C., Waterhouse, A. M., Gumienny, R., Haas, J., & Schwede, T. (2019). QMEANDisCo – Distance constraints applied on model quality estimation. *Bioinformatics*, *36*, 1765–1771.
Takahashi, H., Sasaki, O., Nirasawa, K., & Furukawa, T. (2010). Association between ovocalyxin-32 gene haplotypes and eggshell quality traits in an F2 intercross between two chicken lines divergently selected for eggshell strength. *Animal Genetics*, *41*, 541–544.
Tyanova, S., Temu, T., & Cox, J. (2016). The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nature Protocols*, *11*, 2301–2319.
van Schingen, M., Schepp, U., Pham, C. T., Nguyen, T. Q., & Ziegler, T. (2015). Last chance to see? A review of the threats to and use of the crocodile lizard. *Traffic Bulletin*, *27*, 19–26.
Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. (2018). SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research*, *46*, W296–W303.
Wellman-Labadie, O., Lakshminarayanan, R., & Hincke, M. T. (2008). Antimicrobial properties of avian eggshell-specific C-type lectin-like proteins. *FEBS Letters*, *582*, 699–704.
Wright, A. M., Lyons, K. M., Brandley, M. C., & Hillis, D. M. (2015). Which came first: The lizard or the egg? Robustness in phylogenetic reconstruction of ancestral states. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, *324*, 504–516.
Xie, H., Liang, X., Chen, Z., Li, W.-M., Mi, C., Li, M., Wu, Z.-J., Zhou, X.-M., & Du, W.-G. (2022). Ancient demographics determine the effectiveness of genetic purging in endangered lizards. *Molecular Biology and Evolution*, *39*, msab359.
Yu, D., Zhou, C., & Li, B. (2013). Research advances in ovocleidin-17, the eggshell-specific matrix protein from avians. *Chinese Journal of Animal Nutrition*, *025*, 1164–1168.
Zheng, Y., & Wiens, J. J. (2016). Combining phylogenomic and supermatrix approaches, and a time-calibrated phylogeny for squamate reptiles (lizards and snakes) based on 52 genes and 4162 species. *Molecular Phylogenetics and Evolution*, *94*, 537–547.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Xie, H.-X., Liang, X.-X., Li, W.-M., Chen, Z.-Q., Wang, X.-F., Ding, Z.-H., Zhou, X.-M., & Du, W.-G. (2022). The eggshell-matrix protein gene *OC-17* is functionally lost in the viviparous Chinese crocodile lizard. *Journal of Evolutionary Biology*, *00*, 1–8. https://doi.org/10.1111/jeb.14097