Integrating *De Novo* Transcriptome Assembly and Cloning to Obtain Chicken Ovocleidin-17 Full-Length cDNA

Quan Zhang1*, Long Liu1,*, Feng Zhu2, ZhongHua Ning1, Maxwell T. Hincke3, Ning Yang1, ZhuoCheng Hou1*

1 National Engineering Laboratory for Animal Breeding and MOA Key Laboratory of Animal Genetics and Breeding, Department of Animal Genetics and Breeding, China Agricultural University, Beijing, China, 2 Institute of Animal Genetics and Breeding, College of Animal Science and Technology, Sichuan Agricultural University, Ya’an, China, 3 Department of Cellular and Molecular Medicine, University of Ottawa, Ottawa, Ontario, Canada

Abstract

Efficiently obtaining full-length cDNA for a target gene is the key step for functional studies and probing genetic variations. However, almost all sequenced domestic animal genomes are not ‘finished’. Many functionally important genes are located in these gapped regions. It can be difficult to obtain full-length cDNA for which only partial amino acid/EST sequences exist. In this study we report a general pipeline to obtain full-length cDNA, and illustrate this approach for one important gene (Ovocleidin-17, OC-17) that is associated with chicken eggshell biomineralization. Chicken OC-17 is one of the best candidates to control and regulate the deposition of calcium carbonate in the calcified eggshell layer. OC-17 protein has been purified, sequenced, and has had its three-dimensional structure solved. However, researchers still cannot conduct OC-17 mRNA related studies because the mRNA sequence is unknown and the gene is absent from the current chicken genome. We used RNA-Seq to obtain the entire transcriptome of the adult hen uterus, and then conducted *de novo* transcriptome assembling with bioinformatics analysis to obtain candidate OC-17 transcripts. Based on this sequence, we used RACE and PCR cloning methods to successfully obtain the full-length OC-17 cDNA. Temporal and spatial OC-17 mRNA expression analyses were also performed to demonstrate that OC-17 is predominantly expressed in the adult hen uterus during the laying cycle and barely at immature developmental stages. Differential uterine expression of OC-17 was observed in hens laying eggs with weak versus strong eggshell, confirming its important role in the regulation of eggshell mineralization and providing a new tool for genetic selection for eggshell quality parameters. This study is the first one to report the full-length OC-17 cDNA sequence, and builds a foundation for OC-17 mRNA related studies. We provide a general method for biologists experiencing difficulty in obtaining candidate gene full-length cDNA sequences.

Introduction

Messenger RNA sequence information is necessary for a variety of studies including mRNA expression, designing mRNA microarrays and genome annotation. Many genome projects provided new solutions for identifying mRNA sequences after the release of the human genome. However, only 32% of the genomes in the GOLD database are ‘complete’ or ‘closed’, meaning that they contain no gaps [1]. An even smaller number have been “finished” by manually correcting errors and adding annotations. Repetitive elements, sequencing biases and other complicating factors all come together to make some regions difficult or impossible to assemble [2]. It is estimated that bases in the gapped regions can account for about 1.3% (chicken) to 7.2% (rhesus macaque) to 12.8% (Purple Sea Urchin) of the genome for 12 representative high quality draft assemblies of highly studied species [2]. Almost all the major domestic animal reference genomes contained many gapped regions. Obviously, some functionally important genes are located in these gapped/mis-assembled regions.

How do we efficiently obtain those full-length cDNAs for genes which are not yet covered by the reference genome? RNA-Seq provides a new efficient method to obtain the entire transcribed mRNA in such examples. The *de novo* transcriptome assembly method has been extensively used in non-model organisms to obtain transcript sequences [3,4]. In contrast, researchers generally apply mapping-based methods to detect mRNA structure for model organisms. However, this method will miss important genes that are located in gapped regions of the reference genome. The chicken genome was the first well sequenced domestic animal genome [5] and also contains the least gapped regions [2]. One important gene, *Ovocleidin-17* (OC-17), is of intense interest with
Figure 1. The de novo transcriptome assembly and cloning pipeline. This pipeline has four key steps: (1) Sampling the suitable tissues. Target genes must be relatively highly expressed in the sampled tissues; (2) Transcriptome assembling. The de novo assembly is necessary to obtain the target gene whether we have the reference genome or not; (3) Alignment of the partial target gene protein/EST to the assembled transcripts. We used the reference OC-17 protein sequence (GenBank No.: Q9PR58) to conduct tBLASTn against the assembled transcripts. (4) RACE. The last step is to use RACE to confirm the assembled transcripts and obtain the full-length cDNA. doi:10.1371/journal.pone.0093452.g001

respect to the chicken eggshell biomineralization process. In this study we use OC-17 as example to illustrate a general framework for cloning and obtaining full-length cDNA sequences to pave the way for OC-17 genetic variation and functional studies.

Biomineralization is an important process in a wide range of scientific disciplines including chemistry, biology, medicine, and materials science [6]. Organic macromolecules are important in the regulation of mineral growth, particularly crystal morphology and particle aggregation. Discovering the mechanisms that regulate biomineral formation will provide important insights for advances in human health, material sciences and biology [7]. All birds and most reptiles lay calcareous eggs, while the avian egg is considered to represent the most advanced amniotic egg in oviparous vertebrates [8]. As a typical biogenic mineral, the avian eggshell has attracted considerable attention due to its very rapid rate of formation, well defined ultrastructure and superior mechanical properties that are important features of eggshell quality [8]. The eggshell is a very good model for understanding the mechanisms underlying the phenomena of CaCO₃ biomineralization [9]. OC-17 is a major eggshell-specific protein and was the first eggshell-specific matrix protein to be isolated and characterized [10]. Chicken OC-17 is one of the best candidates to control and regulate the deposition of calcium carbonate in the calcified eggshell layer [9]. For these reasons, the crystal structure of OC-17 protein has been determined [9]; computer simulations based on this structure suggest that OC-17 protein may act as a catalyst in the transformation of amorphous calcium carbonate to calcite crystals [11]. However, in spite of significant efforts, it has not yet been cloned, and nothing is known about the regulation of OC-17 mRNA expression [8]. Only 20 amino acids of OC-17 are found on the ChrUn in the chicken reference genome. This is a relatively frequent case in biology where researchers know the amino acid sequence, have conducted many protein-based studies, but lack mRNA sequences for expression studies.

In this study we illustrate a novel approach to this problem by combining de novo transcriptome assembly and RACE, using the successful example of obtaining the OC-17 full-length cDNA. In doing so, we provide a general method for biologists experiencing difficulty in obtaining full-length cDNAs.

Materials and Methods

Ethics Statement

Animal experiments were approved by the Animal Care and Use Committee of China Agricultural University. Euthanasia was performed by cervical dislocation in order to quickly obtain the tissue samples to minimize any effect on gene expression changes. All experiments were performed according to regulations and guidelines established by this committee.

Animal Samples

All birds were maintained in the China Agricultural University poultry resources station, with free access to standard feed and water. Uterus, isthmus, hypothalamus, hypophysis, heart, liver, spleen, kidney, pectoral muscles, pancreas, magnum, ovary, jejunum, cerebrum and cerebellum were harvested from four normal White Leghorn hens at 49 weeks of age. Uterus samples were also collected at different physiological stages (four birds at 13 weeks, 16 weeks, 20 weeks, 27 weeks, respectively), from birds which were reared under the same environmental conditions. Birds at 13, 16 and 20 weeks of age are not in the reproduction cycle and the birds at the same weeks of age were slaughtered at the same time of day. Birds at 27 weeks of age were slaughtered at 2 h following ovulation. In order to measure correlations between OG-17 mRNA expression level and eggshell quality, we collected eggs from a Rhode Island White Layer line which exhibits divergent eggshell strength. At 40 weeks of age, 3 eggs from each hen were collected for measuring eggshell quality traits. Eggshell strength and eggshell thickness were measured within 12 h after collecting eggs. The average value of 3 eggs per hen for each eggshell quality trait was used in the following analysis. The eggshell strength and eggshell thickness were measured as published before [12]. Hens were rank ordered according to eggshell strength. We sampled the actively calcifying uterus (22 h after ovulation, egg with eggshell) of the top 8 hens (highest eggshell strength) and bottom 8 hens (lowest eggshell strength) for mRNA expression analysis. The entire uterus was taken out and rapidly wiped to remove adhering fluid. The middle part of the uterus was excised with scissors and transferred to a tube. The tube was immediately plunged into liquid nitrogen and then stored at −80°C until RNA extraction.

RNA-Seq and Bioinformatics Analysis

Uterus tissue (active uterus, 22 h after ovulation, egg with eggshell) was obtained from one White Leghorn hen at 49 weeks,
and used for RNA-Seq. Approximately 10 μg of sheared cDNA were prepared for Illumina sequencing according to the manufacturer’s protocols. Paired-End libraries were prepared from a 200–230 bp size-selected fraction following adapter ligation and agarose gel separation. The library was sequenced using a multiplexed paired ends protocol with 101 bp of data collected per run on the Illumina Hiseq 2000 (the data are archived at the NIH Short Read Archive under accession number SRX180570 http://ncbi.nlm.nih.gov/sra). Base calling was performed by the Illumina instrument software.

The reads quality value (Q-value) distribution showed that 5–15 base quality in the 5′-end was lower than Q20. All reads were trimmed by 20 base pairs at the 5′-end to keep the same length using FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). Trimmed reads were used in the following bioinformatics analysis.

Trinity [13] software (version: r20131110) was used to conduct de novo transcriptome assembly with default parameters. We only kept contigs with length longer than 300 bp for further BLAST analysis [14]. The putative signal peptide was predicted using SignalP 4.0 [15].
We used bowtie2 (version 2.2.0) [16] to map short-reads on the candidate OC-17 transcript with the default parameters. SAMTools (version 0.1.19) [17] was used to process the alignment file. SeqMonk (version: 0.27.0) package (http://www.bioinformatics.bbsrc.ac.uk/) was used to quantify and display the coverage of the paired-end reads. Tools (version 0.1.19) [17] was used to process the alignment file. The quality of the mapped reads was assessed using Qualimap (version 1.0.1) [17]. The alignment file was used to calculate the expression of OC-17 using the RSEM tool (version 1.2.4) [17]. The expression was normalized using TPM (transcripts per million) values.

To compare OC-17 expression between different tissues and at different developmental stages, we used the one way ANOVA analysis to analyze the means differences among samples. The mean values were compared using a t-test with a significance level of 0.05.

Results and Discussion

Transcriptome Assembly: Successfully Assembling Target Transcripts

The tBLASTn results confirmed that the OC-17 gene is located in the gap region of the chicken reference genome. This means that we could not use mapping-based methods to align short-reads onto the reference genome. As OC-17 protein is predominantly present in the adult hen uterus during eggshell formation, we collected a mature hen uterus sample to process for RNA-Seq. The resulting RNA-Seq data was used to perform de novo assembly for uterus RNA-Seq. This resulted in 37,390,288 paired-end reads (GenBank SRA No.: SRX180570). We used the Trinity package to conduct de novo assembly for uterus RNA-Seq data to obtain the transcripts. This resulted in 76,068 transcripts, ranging in length from 201–441 bp, with 423 bp median length. The N50 of the assembled transcript was 1,172 bp. We only kept contigs with a length longer than 300 bp. The final assembled contig dataset which was used for BLAST analysis had 51,512 contigs with 649 bp median length.

We used the published OC-17 amino acid sequence (GenBank No: Q9PRS8) to perform BLASTn against the assembled uterine expressed contigs. The analysis pipeline that we developed is shown in Figure 1. The BLASTn results revealed that OC-17 amino acid sequences were fully covered by two contigs. The

| Primer category | Primer name | Nucleotide sequences | Size (bp) |
|-----------------|-------------|----------------------|-----------|
| RACE primers    | 5’-GRP      | 5'-CGACTGAGAGGACGAGGACACTGA-3' | 212 |
|                 | 5’-GAGTCCGCTGACAGAGGACACTGA-3' | 235 |
|                 | 3’-GRP      | 5'-GCTGACGAGGACGAGGACACTGA-3' | 44 |
|                 | 3’-GSP      | 5'-GGCTGACGAGGACGAGGACACTGA-3' | 60 |
|                 | SP-F        | 5'-GGCTGACGAGGACGAGGACACTGA-3' | 96 |
|                 | SP-R        | 5'-GGCTGACGAGGACGAGGACACTGA-3' | 110 |
|                 | 3’-Oligo    | 5'-GGCTGACGAGGACGAGGACACTGA-3' | 96 |
| qRT-PCR primers | P-OC-17-F   | 5'-CGTCTGACCCCTGTTGAGG-3' | 96 |
|                 | P-OC-17-R   | 5'-CGTCTGACCCCTGTTGAGG-3' | 110 |
|                 | P-β-Actin-F | 5'-TATGCAAGCAGCGGTTC-3' | 96 |
|                 | P-β-Actin-R | 5'-TATGCAAGCAGCGGTTC-3' | 110 |

Table 1. Primers for amplifying cDNA and quantifying mRNA of OC17.

GRP means Gene RACE Primer which corresponded to the adaptor sequence and GSP means Gene Special Primer which was complementary to the OC17 cDNA. Primers for each sequence end are indicated either 5’ or 3’. Forward primers and reverse primers for qRT-PCR have prefixes ‘F’ and ‘R’, respectively. 5’GRP and 5’GSP were used for the cloning of 5’ end of the OC17 cDNA while 3’GRP and 3’GSP were used for the 3’ end. SP-F and SP-R were used for validation of the other parts of the OC17 cDNA. 3’Oligo and 3’Oligo were provided by the Gene Racer Kit.

doi:10.1371/journal.pone.0093452.t001
longer contig almost fully covered the shorter contig except for a 10-bases difference. We used the long contig as the assembled OC-17 transcript for the following analysis and primer design for molecular cloning. We also plotted the reads distribution on the assembled OC-17 transcript and found the reads were evenly distributed across the entire OC-17 transcript (Figure S1). In total, there 120,023 reads were mapped on the assembled OC-17 transcript which represented very high coverage. Alignments showed that the assembled transcripts not only included all the reference OC-17 amino acid sequence, but also included an extra deduced 19 amino acids at the N-terminus of the mature protein (Figure 2, brown box). Analysis of the new sequence with the SignalP 4.1 server [16] (http://www.cbs.dtu.dk/services/SignalP/) revealed that a signal peptide cleavage site is predicted between residues 19 and 20, suggesting that the extra amino acids correspond to a putative signal peptide commencing with the methionine start codon. This is consistent with OC-17’s function as a secreted extracellular matrix protein [10]. Sequence analysis showed that these proteins all belong to a family of C-type lectin-like proteins [18]. However, OC-17 only showed 32–47% sequence identity with the other identified matrix proteins [18]. Furthermore, none of the cDNA sequences for these eggshell matrix proteins are yet available. It is difficult to design degenerate primers for cDNA cloning based on poor multiple protein sequence alignments. All of these reasons have hindered previous attempts to obtain the OC-17 mRNA sequence.

Development of the next-generation sequencing technology makes it easier than before to obtain tissue transcriptomes for model and non-model organisms. Transcriptome de novo assembly has been widely used in the analysis of non-model organisms [21,22], but less so for model organisms. If we tried to map RNA-Seq short-reads to the chicken reference genome, we would never be successful in assembling transcripts located in the gapped

![Graphs showing eggshell biomechanical measurements and expression analysis of Ovocleidin-17.](https://example.com/graph.png)

Figure 3. Eggshell biomechanical measurements and expression analysis of Ovocleidin-17. (A) Eggshell phenotypes showed significant differences between the two tails of the eggshell strength groups; (B) the OC-17 mRNA expression levels were significantly higher in hen uterus and isthmus in the low eggshell strength groups compared with the high eggshell strength groups; (C) OC-17 mRNA can only be detected in hen uterus and isthmus among the 15 sampled adult hen tissues. One way ANOVA analysis suggested there were significant expression differences among different tissues. OC-17 showed a much higher expression level in uterus than in other tissues (pairwise t-test, P value < 0.05); (D) One way ANOVA analysis suggested that significant expression differences existed among different stages. OC-17 showed an increasing pattern of expression between the immature and mature laying stages (pairwise t-test, P value < 0.05).

doi:10.1371/journal.pone.0093452.g003
regions. Therefore, we also need to do de novo transcriptome assembly for species for which a reference genome exists, for this purpose. There are a number of situations that require this approach for integrating transcriptome assembly and RACE to obtain target full-length mRNA. Firstly, targeted genes may be located in the gapped/UniAssembled reference genome region. Secondly, there may be species-specific target genes which do not have high similarity to orthologs in other species. Thirdly, target genes may correspond to purified and sequenced protein products that fail to be cloned using RACE based on the partial amino acid sequences. In this specific example, once we obtained the target gene cDNA, it was necessary to design primers based on our assembled OC-17 transcript sequence and apply RACE to confirm the accuracy of the transcriptome assembly.

Full-Length cDNA confirmed the Accuracy of Transcriptome Assembly

Based on the assembled transcripts, gene-specific primers (Table 1, Figure 2) were designed to amplify the full-length cDNA of chicken OC-17 from chicken uterus. We used the 3'-RACE and 5'-RACE kit to clone the 3' and 5' sequences, respectively. The full-length OC-17 cDNA is 626 bp in length, with an open reading frame of 483 bp (GenBank Accession No.:KF833610; Figure 2). Sequence analyses confirmed that full-length OC-17 cDNA included the extra 19 amino acids corresponding to the putative signal peptide, compared to the previously published OC-17 amino acid sequence (Figure 2, with brown box). The RACE cDNA was the same as the assembled contigs, and showed the power of de novo assembly in discovering transcripts for species with/without a reference genome. The newly obtained OC-17 cDNA sequence can now be exploited for mRNA-related studies. OC-17 may be critically important in CaCO3 transformation and crystal growth [9,11], and also in the regulation of eggshell quality [23]. However, the amino acids that directly interact with the mineral are not yet known. The OC-17 cDNA sequence can now be used to introduce mutated sites to produce recombinant proteins to verify the functional consequences for CaCO3 crystal growth in vitro.

We explored the relationship between eggshell mineralization and OC-17 by investigating OC-17 expression at different hen uterine developmental stages, and determining the specificity of OC-17 expression.

Expression Analyses of OC-17 Supports its association with Biomineralization

Although thickness is the main factor contributing to mechanical strength of the eggshell, the structural organization of the eggshell at different levels has a significant effect. Ultrastructure (the organization of major structural units) and texture (the size of crystals, their shape and crystallographic orientation) are especially important. The soluble matrix proteins of calcitic biomaterials can modify crystal growth, and thus regulate the macroscopic properties of the resulting bioceramic [8,24].

We quantified uterus and isthmus OC-17 expression level in two hen groups laying eggs with different eggshell strength (Figure 3A). The OC-17 mRNA expression level in the high eggshell strength group is significantly lower than in the low eggshell strength group (Figure 3B). This is consistent with the OC-17 protein concentrations in eggshell of eggs laid by young and old hens, two groups where eggshell strength is significantly different [25]. This suggested that OC-17 protein concentration might be negatively correlated with eggshell strength, as originally reported [25].

As tissue-specific gene expression can indicate gene function in certain tissues [26], we also wanted to determine the OC-17 mRNA expression status in a variety of tissues. Among all the 15 investigated tissues, OC-17 was only detected in uterus and isthmus (Figure 3C). OC-17 has been considered as a uterine-specific protein based on Western blotting [10], although sensitive proteomic studies have detected low levels of this protein in other egg compartments, such as egg white [27,28], vitelline membrane [29] and egg yolk plasma and granules [30]. Indeed, we also detected expression of the OC-17 in the isthmus, the oviduct segment where the eggshell membranes are synthesized. Does the OC-17 expression correlate with uterine developmental stage? To address this question, uterine tissue from four white leghorn hens at each stage (developing, immature uterus: 13 weeks, 16 weeks and 20 weeks; actively laying uterus: 27-weeks) were harvested and analyzed by qRT-PCR. The OC-17 gene was highly expressed in mature uterus of 27-week hens and much less expressed in the immature uterus of 13-week, 16-week and 20-week hens (Figure 3D). OC-17 showed an increasing trend of expression during the progressive developmental and laying stages.

OC-17 begins to be expressed during the uterine developmental stages (13-week, 16-week), albeit at low levels. Does this mean that OC-17 has functions during development of the immature uterus? Further studies are needed to explore OC-17 potential functions in hen uterine development. In summary, this study provides new evidence that OC-17 is expressed in a hen oviduct-specific fashion, and may have potential new roles during uterine development. Moreover, its expression levels in the mature, laying uterus are inversely correlated with eggshell quality.

Conclusions

We have demonstrated that de novo transcriptome assembly can be critically important for certain purposes in model organisms. This study is the first one to report the full-length OC-17 cDNA sequence, and provides preliminary data describing physiological regulation of its expression, as a foundation for OC-17 mRNA related studies. Furthermore, we provide a general method for biologists experiencing difficulty in obtaining candidate gene full-length cDNA sequences.

Supporting Information

Figure S1 Coverage plot of the OC-17 transcript. Red color represents the reads from the 5'-end while the blue color represents reads from the 3'-end. Last row represents the coverage plot of the mapped short-reads on the OC-17 transcript. (TIF)

Acknowledgments

The authors thank the China Agricultural University poultry resources station for maintaining and providing birds.

Author Contributions

Conceived and designed the experiments: ZHN MTH NY ZCH. Performed the experiments: QZ LL Analyzed the data: EZ ZCH. Wrote the paper: QZ LL EZ ZCH. Performed bioinformatics and statistical analysis with the RNA-seq data and comparative genomics: ZCH FZ. Performed the laboratory experiments, sequence analysis and sample collecting: QZ LL. Paper revisions: NY ZHN MTH ZCH. All authors read and approved the final manuscript: QZ LL EZ ZHN MTH NY ZCH.
References

1. Koren S, Harhay GP, Smith TP, Bono JL, Harhay DM, et al. (2013) Reducing assembly complexity of microbial genomes with single-molecule sequencing. Genome biology 14: R101.

2. English AC, Richards S, Han Y, Wang M, Vee V, et al. (2012) Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. PLoS one 7: e47768.

3. Robertson G, Schen J, Chin R, Corbett R, Field M, et al. (2010) De novo assembly and analysis of RNA-seq data. Nature methods 7: 909–912.

4. Li C, Wang Y, Huang X, Li J, Wang H (2013) De novo assembly and characterization of fruit transcriptome in Litchi chinensis Sonn and analysis of differentially regulated genes in fruit in response to shading. BMC genomics 14: 532.

5. International Chicken Genome Sequencing C (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432: 695–716.

6. Mann S, Archibald DD, Didymus JM, Douglas T, Heywood BR, et al. (1993) Crystallization at Inorganic-organic Interfaces: Biominerals and Biomimetic Synthesis. Science 261: 1296–1292.

7. Colfen H (2010) Biominalization: A crystal-clear view. Nature materials 9: 960–961.

8. Hincke MT, Nys Y, Gastron J, Mann K, Rodriguez-Navarro AB, et al. (2012) The eggshell: structure, composition and mineralization. Frontiers in bioscience 17: 1266–1280.

9. Reyes-Grajeda JP, Moreno A, Romero A (2004) Crystal structure of ovocleidin-17, a major protein of the calcified Gallus gallus eggshell: implications in the calcite mineral growth pattern. The Journal of biological chemistry 279: 40876–40881.

10. Hincke MT, Tsang CP, Courtney M, Hill V, Nairnitz R (1995) Purification and immunochromistry of a soluble matrix protein of the chicken eggshell (ovocleidin 17). Calcified tissue international 56: 578–583.

11. Freeman CL, Harding JH, Quigley D, Rodger PM (2012) Protein binding on stepped calcite surfaces: simulations of ovocleidin-17 on calcite [31.16] and [31.8]. Physical chemistry chemical physics : PCCP 14: 7287–7295.

12. Zhang Y, Hou ZC, Chen ZX, Zheng JX, Chen SR, et al. (2011) Low-density lipoprotein receptor-related protein 2 gene is associated with egg-quality traits in layers. Poultry science 90: 2718–2722.

13. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, et al. (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature protocols 8: 1494–1512.

14. Ashual SF, Gish W, Miller W, Myers EW, Lifman DJ (1990) Basic local alignment search tool. Journal of molecular biology 215: 403–410.

15. Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature methods 8: 785–786.

16. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9: 357–359.

17. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078–2079.

18. Mann K, Siedler F (2006) Amino acid sequences and phosphorylation sites of eggshell matrix proteins (ovalbumin, ovotransferrin) in shell from young and old hens. Poultry science 79: 580–588.

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.

20. Garg R, Jain M (2013) RNA-Seq for transcriptome analysis in non-model plants. Methods in molecular biology 1069: 43–51.

21. Hoeijmakers WA, Bartfai R, Stunnenberg HG (2013) Transcriptome analysis using RNA-Seq. Methods in molecular biology 923: 221–239.

22. Panthleux M, Nys Y, Williams J, Gastron J, Boldtke I, et al. (2000) Extraction and quantification by ELISA of eggshell organic matrix proteins (ovocleidin-17, ovalbumin, ovotransferrin) in shell from young and old hens. Poultry science 79: 580–588.

23. Nys Y, Garcia-Ruiz JM, Hincke MT (2004) Avian eggshell mineralization: biochemical and functional characterization of matrix proteins. Comptes Rendus Palevol: 549–562.

24. Panthleux M, Nys Y, Williams J, Gastron J, Boldtke I, et al. (2000) Extraction and quantification by ELISA of eggshell organic matrix proteins (ovocleidin-17, ovalbumin, ovotransferrin) in shell from young and old hens. Poultry science 79: 580–588.

25. Rose ML, Hincke MT (2009) Protein constituents of the eggshell: eggshell-specific matrix proteins. Cellular and molecular life sciences : CMLS 66: 2707–2719.

26. Hou Z, Romero R, Uddin M, Than NG, Wildman DE (2009) Adaptive history of single copy genes highly expressed in the term human placenta. Genomics 93: 33–41.

27. D’Ambrosio C, Arena S, Scaloni A, Guerrier L, Bouchetti E, et al. (2008) Exploring the chicken egg white proteome with combinatorial peptide ligand libraries. J Proteome Res 7: 3461–3474.

28. Mann K, Mann M (2011) In-depth analysis of the chicken egg white proteome using an LTQ Orbitrap Velos. Proteome Sci 9: 7.

29. Mann K (2008) Proteomic analysis of the chicken egg vitelline membrane. Proteomics 8: 2322–2332.

30. Mann K, Mann M (2006) The chicken egg yolk plasma and granule proteomes. Proteomics 6: 178–191.