Food Research International 160 (2022) 111670

Contents lists available at ScienceDirect

Food Research International

journal homepage: www.elsevier.com/locate/foodres

Revealing the species-specific genotype of the edible bird's nest-producing swiftlet, *Aerodramus fuciphagus* and the proteome of edible bird's nest

Hang-kin Kong [a,b], Zoe Chan [a], Sau-woon Yan [a], Pak-yeung Lo [a], Wing-tak Wong [a], Ka-hing Wong [a,b,*], Chun-lap Lo [a,b,*]

[a] Department of Applied Biology and Chemical Technology, Faculty of Applied Science and Textiles, The Hong Kong Polytechnic University, Hung Hom, Hong Kong SAR, China

[b] Research Institute for Future Food, The Hong Kong Polytechnic University, Hung Hom, Hong Kong SAR, China

ARTICLE INFO

Keywords:
Aerodramus fuciphagus
Edible bird's nest
Genome
CCDC63
Proteome
Epidermal growth factor

ABSTRACT

Only a few species of swiftlets in the *Aerodramus* and *Collocalia* genera can produce edible bird's nests (EBN). These saliva-cemented nests have been consumed as delicacies for centuries in Asia. Many researches have reported the aqueous extract of EBN has epidermal growth factor-like (EGF-like) activity. However, no standalone EGF has been identified in EBN. Moreover, proteome of EBN remained unclear due to lack of genomic data base of an EBN-producing swiftlet to support proteomic analysis of EBN. To address this, the first genome of the EBN-producing swiftlet, *Aerodramus fuciphagus*, was constructed. Orthology comparison of *A. fuciphagus* with 10 other avian species were conducted. The results revealed that the number of predicted paralogous coiled-coil domain-containing protein 63 (CCDC63) coding sequences (CDSs) in *A. fuciphagus* was found to be significantly expanded in comparison to *Gallus gallus*. There were 3 paralogous CCDC63 genes in the genome of *A.fuciphagus*. The CDSs predicted from the genome of *A. fuciphagus* were used to construct a database for proteomic analysis of EBN. In total, 398 proteins have been identified in EBN. The proteome of EBN was significant enriched with extracellular proteins as well as proteins related to extracellular matrix (ECM) organization and immune response. A few proteins with Ca^{2+}-binding EGF-like domains were found in the proteome of EBN, like fibrillin-1, protocadherin fat 4 and coagulation factor X. No standalone EGF protein was identified. This indicated that the proteins with EGF-like domains might be responsible for the EGF-like activity of EBN. In addition, acidic mammalian chitinase and lysyl oxidase in EBN were found to be active when extracting with distilled water at room temperature. The current study has not just revealed the species-specific genotype of the EBN-producing swiftlet, *A. fuciphagus*, but also revealed the proteome of EBN. This established an important foundation for subsequently studies on efficacies of EBN.

1. Introduction

Edible bird's nests (EBNs) are saliva-cemented nests only built by a few swiftlet species found in Southeast Asia (Chua & Zukefli, 2016). The EBNs traded worldwide today are mainly from the white-nest swiftlet (*Aerodramus fuciphagus*), the black-nest swiftlet (*Aerodramus maximus*) and some species of the *Collocalia* genera (Ma & Liu, 2012). Among these species, EBNs from *A. fuciphagus* are generally much preferred in the market as they are relatively clean, and seldom contaminated by feathers or bird droppings (Chua & Zukefli, 2016). According to the Chinese Compendium of Materia Medica, EBNs are associated with a wide range of health benefits, including improving asthmatic condition, boosting immune system, preventing stomach illness, as well as restoring stamina (Wong, 2013). As the consumption of EBNs is thought to be beneficial to human health, they have been consumed as a traditional Chinese tonic food and delicacy for centuries.

In the past few decades, many research have analyzed compositions of EBN and investigated efficacies associated with the consumption of EBN. EBN was found to be rich in polyunsaturated fatty acids (PUFA), sialic acids, and essential amino acids (Haghani et al., 2016, Hun Lee et al., 2020, Shangguan et al., 2018). Some the chemical compositions of EBN can account for the efficacies of EBN. The sialic acid in EBN exerted

* Corresponding authors at: Department of Applied Biology and Chemical Technology, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong SAR, China.

E-mail addresses: hang-kin.kong@polyu.edu.hk (H.-k. Kong), kahing.wong@polyu.edu.hk (K.-h. Wong), samuel.chun-lap.lo@polyu.edu.hk (C.-l. Lo).

https://doi.org/10.1016/j.foodres.2022.111670
Received 29 March 2022; Received in revised form 23 May 2022; Accepted 7 July 2022
Available online 12 July 2022

anti-viral activities (Haghani et al., 2016). The PUFA content in EBN might improve cardiovascular conditions (Hun Lee et al., 2020). Apart from these efficacies, extracts of EBN were found to have anti-oxidant, anti-inflammatory, anti-ageing, and anti-cancer effects (Lee et al., 2021). Besides those, the consumption of EBN was found to improve bone strength, immunity, intelligence and memory (Yeo et al., 2021). However, the active ingredients in EBN responsible for those efficacies remained unknown.

Among various efficacies, epidermal growth factor-like (EGF-like) activity of EBN was one of the most famous and well-studied. The aqueous extracts of EBN were found to stimulate growths of human peripheral blood monocytes, adipose-derived stem cells and rabbit corneal keratocytes (Kong et al., 1987, Matsukawa et al., 2011, Ng et al., 1986, Zainal Abidin et al., 2011). Although the chemical compositions of EBN were well characterized, it still failed to fully account for the EGF-like activities of EBN. As around 65% of the dry weight of EBN was composed of protein, so it was suspect that some of the proteins in EBN were responsible to the efficacies reported in the previous studies (Quek et al., 2018). However, the proteome of EBN remained largely unknown. In the literature, a few studies have attempted to identify proteins that composed of EBN using different techniques. EBN was found to contain some glycoproteins by examining Raman shift using Raman micro-spectroscopy (Shim et al., 2016). Another study, using 2D-gel electrophoresis and MALDI-TOF-TOF mass spectrometry, had identified acidic mammalian chitinase in EBN (Liu et al., 2012). In our previous study, with *de-novo* sequencing using MALDI-TOF-TOF mass spectrometry, a dozen of proteins have been identified in EBN, including mucin 2 and 5 (Kong et al., 2016). However, no standalone EGF has ever been identified in EBN. Although more proteins in EBN have been identified with the advancement of *de-novo* sequencing, the number of proteins that had been identified were still limited due to the lack of an appropriate genomic or transcriptomic database to support proteomic analysis and bioinformatic searches.

Looi et al., (2017) have conducted a comparative transcriptomic analysis of salivary glands between EBN-producing swiftlets and non EBN-producing swiftlets (Looi et al., 2017). The transcriptomes of the salivary glands from the EBN-producing swiftlets could be used as the database for identification of proteins composing EBN. However, gene expression varied due to different intrinsic and extrinsic factors, like gender, age and growth conditions. Thus, the transcriptome of the salivary glands may not contain all the coding sequences (CDSs) of the proteins composing EBN. To identify a more complete proteome of EBN, the first genome of the EBN-producing swiftlet *A. fuciphagus* has been constructed in order to establish a complete CDS database for proteomic analysis of EBN. In addition, an orthology comparison between the genome of *A. fuciphagus* and 10 others non EBN-producing avian species was conducted to reveal species-specific genotypes of *A. fuciphagus*.

2. Material and methods

2.1. Chemicals and materials

Chemicals and solvents were purchased from Sigma Aldrich (USA) and Merck Millipore (Germany). They were either in molecular grade, RNase and DNase-free or compatible to mass spectrometry analysis. Plastic consumables, which were purchased from Eppendorf (USA), were either compatible to mass spectrometry or molecular analysis.

2.2. Collecting live specimens from swiftlet house

To construct the genome of the EBN-producing swiftlet, live specimens were required for extracting genomic DNA. An EBN supplier in Cirebon, Indonesia, kindly donated 4 of their EBN-producing swiftlets from their swiftlet houses. The supplier was responsible to produce and supply EBNs to a local vendor, Hing Kee Java Edible Bird's Nest (HKJEBN) Co. Ltd. in Hong Kong. Nine intact EBNs used for subsequent proteomic analysis were also donated by HKJEBN Co. Ltd. The live specimens were transported to Prof. Kris Herawan Timotius's research laboratory in Faculty of Medicine, Krida Wacana Christian University Cirebon, Indonesia. Within 24 h after capture, the live specimens were euthanized before being dissected to collect tight skeletal muscles. Before the dissection, the working space was cleaned with RNase decontamination solution (Invitrogen, USA). All the dissection tools were cleaned and autoclaved. All the plastic consumables were DNase- and RNase-free (Neptune Scientific, US). All the tissues were cut into around 2 mm × 2 mm pieces and persevered in RNAlater solution (Thermo Fisher Scientific) before DNA extractions.

2.3. Genomic DNA extraction

Genomic DNA (gDNA) of the specimens was extracted using Genomic-tip 500/G column according to protocols from the manufacturer (Qiagen, USA). In brief, 100 mg of tight muscle was ground into powder with liquid nitrogen. The powder was resuspended with 20 ml of lysis buffer with RNase A and Protease K. After incubation at 50℃ for 2 h, the lyse was loaded to a 500/G column. The column was washed, subsequently eluted. The gDNA in the elute was precipitated by adding isopropanol (0.7-fold of volume) at room temperature, and then centrifuged at 5000g for 15 min at 4℃. The pellet was washed with 4 ml of ice cold 70% ethanol twice before air dried. The pellet was re-dissolved in 0.1 ml of 10 mM Tris-Cl, pH 8.5. The length of the gDNA was visualized by gel electrophoresis. The length of the gDNA was < 40 kb.

2.4. Genomic DNA library preparation and sequencing

A library of gDNA was prepared by using Chromium Genome HT Library Kit & Gel Bead Kit v2 (10xGenomic, USA). Around 1 ng of was used for preparing linked reads (barcoded reads) for sequencing following the protocols from the manufacturer. The resultant linked reads were comprised with a standard Illumina P5 adapter, followed by a 16 bp 10X Genomics barcode at the start of read 1, the gDNA insert, an 8 bp sample index and standard Illumina P7 adapter at the end. Subsequently, the linked reads were amplified for 10 thermal cycles (denaturing at 98℃ for 20 s, annealing at 54℃ for 30 s, elongating at 72℃ for 20 s). The library was then sequenced with an Illumina NovaSeq 5000. The sequencing data was then cleaned up and assembled using Supernova v2.1.1 (10xGenomic, USA). The details of the library preparation and sequencing strategy were provided in the supplementary material.

2.5. Gene prediction and annotation

GeneWise was used to predict genes by homology searching with coding sequences (CDS) and transcriptomic data from other avian species in the NCBI database (Birney et al., 2004). The species used were *Chaetura pelagica, Gallus gallus, Meleagris gallopavo, Anas platyrhychos, Columba livia,* and *Calypte anna*. The predictions were then aligned and assembled by Program to Assemble Spliced Alignments (PASA) (Haas et al., 2003). EvidenceModeler was used to compute weighted consensus gene structure annotations based on the above data and the CDSs were exported in GFF3 format. The CDSs were extracted from the genome based on the information of GFF3 file using getfasta in BEDTools v2.29.0 (Quinlan, 2014). The CDSs were annotated using Blast2GO with NCBI non-reductant database (RefSeq, release 93, 13th March 2019, 52,033,004,779 amino acid residues) as well as Swiss-Prot (UniProtKB/Swiss-Prot UniProt release 2019_02, 13th Feb 2019, 559,228 entries). The details of gene prediction and annotation were provided in the supplementary material.

2.6. Assessing quality of genome assembly

The quality of the new genome assembly was assessed by using BUSCO in genome mode with tBLASTn v2.10, Augustus v3.3.3, and HMMER v3.3 (Seppey et al., 2019). The DNA sequence of the genome in fasta format was submitted for the assessment. The BUSCO datasets used was Aves Odb10 which was the most taxonomically related dataset for the assessment.

2.7. Orthology analysis

Orthology analysis was conducted between *A. fuciphagus* and 10 other avian species using OrthoFinder v.2.2.7 (Emms & Kelly, 2019). The species used were *C. pelagica, G. gallus, M. gallopavo, A. platyrhychos, C. livia, C. anna, Antrostomus carolinensis, Taeniopygia guttata, Hirundo rustica* and *Zonotrichia albicollis.* The coding sequences of the above avian species were downloaded from the NCBI genome database (https://www.ncbi.nlm.nih.gov/genome/). The information related to the version of the genomes was summarized in Table 1. The coding sequences from all the genomes were submitted to the OrthoFinder coupled with Blast + v.2.8.1, FastME v.2.1.6.1_1, and Markov cluster algorithm (MCL). After homology search, the orthologous genes among the genomes were clustered into orthogroups. A phylogenetic tree of the avian species was generated based on the sequence homology. The gene counts of each avian species in each orthogroup and the phylogenetic tree were imported to CAFExp for further statistical analysis (Han et al., 2013). CAPExp used birth and death statistical model to estimate expansions and contractions of gene families (Hahn et al., 2005). Before importing the data, the orthogroups were filtered according to the guidelines provided by the developer of CAFExp. The orthogroups with 1 or more avian species that had a CDS count of more than 100 were discarded. Since there were only 11 avian species in the orthology analysis, it was assumed that the results from CAFExp would not be reliable if orthogroups with 7 or more species which had a CDS count equal to 0. Thus, those orthogroups were also discarded. As it was suspected that avian species in different orders might have different rates of gain and loss, 4 λs (birth–death parameters) were estimated according to their clades/ orders. Genome assembly errors were also estimated and applied to control inflations of λ due to the errors in the genome assembly.

2.8. Phylogenetic analysis

To reveal the taxonomy of the specimens from Indonesia, mitochondrial cytochrome *b* sequences (Cytb, 840 nucleotides in length) from the sequencing data were aligned with other avian Cytb sequences in the NCBI nucleotide database using NCBI Blastn (Forni et al., 2019, Sanita Lima & Smith, 2017). To construct a phylogenetic tree of the specimen, the Cytb sequences from *Aerodramus*, *Apus*, *Chaetura*, and *Collocalia* (genus under *Apodinae*) with at least 800 nucleotides in length (full length of Cytb should be around 1100 bp) were exported in txt file format and then imported to NGPhylogeny.fr for the phylogenetic analysis (Lemoine et al., 2019). NGPhylogency.fr was a pipeline that included tools for aligning sequences, cleaning up the alignment, and building the tree. A multiple alignment program for amino acid or nucleotide sequences (MAFFT) v.7.407_1 was used to align the cytb sequences (Katoh & Standley, 2013). Block Mapping and Gathering with Entropy (BMGE) v.1.12_1 was used to select highly conserved regions of the alignment for constructing the tree (Criscuolo & Gribaldo, 2010). The tree was built using FastME v.2.1.6.1_1 (distance algorithms) (Lefort et al., 2015). Finally, the tree was visualized using Newick display v.1.6 (Junier & Zdobnov, 2010).

2.9. Functional enrichment analysis

PANTHER v14 (Protein Analysis Through Evolutionary Relationships) was also used for the functional enrichment analysis (Mi et al., 2019). As the genome of *A. fuciphagus* was annotated with NCBI non-reductant database as well as Swiss-Prot, the UniProt entries and gene symbols of the annotated genes (16081 annotated genes) were first mapped to databases in PANTHER. The model organism was *G. gallus.* After mapping, the lists of mapped UniProt entries or gene symbols of the genes and proteins of interest were submitted to an over-representation test. *G. gallus* was used as the model organism. Fisher's exact test with false discovery rate correction was used to compute significances of the analysis. As recommended by the developer, all annotated genes in the genome of *A. fuciphagus* that could be mapped to the databases of PANTHER (13747 genes mapped to PANTHER) were used as reference set for the analysis (Mi et al., 2019). The Gene ontology terms in molecular function, biological process and cellular component with false discovery rate < 0.05 were considered significantly enriched. The details of the function enrichment analysis ware provided in the supplementary material.

2.10. Mapping of scaffolds to reference genome

Minimap2 was used to map the scaffolds in the genome of *A. fuciphagus* to the chromosomes from the reference genome of *G. gallus*

Table 1
Summary of the assembly versions of 10 avian species used for the orthological analysis.

Specie	Common name	Clade/ Order	Assembly version	link
Gallus gallus	chicken	Galloanserae	Genome Reference Consortium Chicken Build 6a (GRCg6a)	https://www.ncbi.nlm.nih.gov/assembly/GCF_000002315.6
Meleagris gallopavo	turkey	Galloanserae	Turkey_5.1	https://www.ncbi.nlm.nih.gov/assembly/GCF_000146605.3
Anas platyrhynchos	mallard	Galloanserae	IASCAAS_PekingDuck_PBH1.5	https://www.ncbi.nlm.nih.gov/assembly/GCF_003850225.1
Taeniopygia guttata	Zebra finch	Passeriformes	bTaeGut2.pat.W.v2	https://www.ncbi.nlm.nih.gov/assembly/GCF_008822105.2
Hirundo rustica	Barn swallow	Passeriformes	Chelidonia	https://www.ncbi.nlm.nih.gov/assembly/GCA_003692655.1
Zonotrichia albicollis	white-throated sparrow	Passeriformes	Zonotrichia_albicollis-1.0.1	https://www.ncbi.nlm.nih.gov/assembly/GCF_000385455.1
Columba livia	rock pigeon	Columbiformes	Cliv_2.1	https://www.ncbi.nlm.nih.gov/assembly/GCA_000337935.2
Antrostomus carolinensis	chuck-will's-widow	Caprimulgiformes	ASM70074v1	https://www.ncbi.nlm.nih.gov/assembly/GCF_000700745.1/
Calypte anna	Anna's hummingbird	Apodiformes	bCalAnn1_v1.p	https://www.ncbi.nlm.nih.gov/assembly/GCF_003957555.1
Chaetura pelagica	Chimney swift	Apodiformes	ChaPel_1.0	https://www.ncbi.nlm.nih.gov/assembly/GCF_000747805.1/

(version: GRCg7b) using default settings (Li, 2018). The alignments were exported in PAF format, and subsequently visualized using D-Genise (Cabanettes & Klopp, 2018).

2.11. Prediction of TATA box

Online OProf from Signal Search Analysis Server was used to predict the TATA boxes of the CDSs of interest (Ambrosini et al., 2003). The upstream 1000 base pair nucleotide sequences from the start codons in fasta format were submitted to OProf (https://ccg.epfl.ch/ssa/oprof.php). Search mode was set to "Forward". Window size and window shift were 20 and 1 respectively. Position weight matrix used for the prediction was TATA-box (length = 15) in a promoter motif library. The cut-off value was −8.16.

2.12. Prediction of transcription start site

The transcription start sites were predicted using Promoter2.0 in the online server of DTU Health Tech (https://services.healthtech.dtu.dk/service.php?Promoter-2.0) (Knudsen, 1999). The upstream 1000 base pair nucleotide sequences from the start codons in fasta format were submitted to Promoter2.0 for the prediction.

2.13. Protein extraction from edible bird nest

The EBNs were originated from Cirebon, Indonesia. Proteins were extracted according to our previous study with modifications (Kong et al., 2016). In brief, dried EBN was grinded into around 1 mm^3 small cubes. Distilled water was added to the grinded EBN in 30 to 1 vol to weight ratio. The grinded EBN was soaked for an hour at room temperature and the cold aqueous extract (CAE) was harvested by centrifuging the EBN suspension at 4000 rpm for 10 min at 4℃. The grinded EBN was soaked for twice before stewing. The soaked EBN was stewed with 60 ml of distilled water at 80℃ for 45 min. The hot aqueous extract (the distilled water for stewing, HAE) was harvested by centrifugation at 4000 rpm for 10 min at 4℃. The CAE and HAE were concentrated by Amicon ultrafiltration with a 3 K Da molecular weight cutoff membrane (Merck Millipore, Germany). The extracts were dried by vacuum centrifugal evaporator and stored at −20℃ before further analysis.

2.14. Simulated gastric digestion of stewed edible bird nest

After stewing, the EBN was partially digested by simulated gastric fluid (SGF: 32 mg of pepsin in 10 ml of 35 mM NaCl and 0.7% HCl, pH 1.2). The stewed EBN was incubated with 30 ml of SGF at 37℃ for 4 h. After digestion, the SGF was harvested by centrifugation at 4000 rpm for 10 min at 4℃. The insoluble and undigested residual of EBN was stored at −80℃ for further analysis. The SGF was concentrated by Amicon ultrafiltration with a 3 K Da molecular weight cutoff membrane (Merck Millipore, Germany). The SGF was dried by vacuum centrifugal evaporator and stored at −20℃ before further analysis.

2.15. Sample preparations for proteomic analysis

The extracted proteins (100ug) from CAE, HAE and SGF were dissolved in 50ul of 25 mM NH_4HCO_3, and then reduced with 10 mM dithiotheritol at 56℃ for 45 min, followed by alkylation with 55 mM iodoacetamide at room temperature for 30 min in dark. The samples were precipitated with 4-fold volume of ice-cold acetone at −20℃ overnight. The precipitated proteins were harvested by centrifugation at 13,000 g for 10 min at 4℃. The pellets were washed with 80% ice cold acetone twice before air dried. The pellets were dissolved in 25 mM NH_4HCO_3 and mass spectrometry graded trypsin (Promega, USA) was added to digest the proteins at 37 ℃ overnight. The tryptic peptides were desalted with C_{18} spin cartridge according to protocol provided by the manufacturer (Thermo Fisher Scientific, USA). The elute from the C_{18} column was dried with vacuum centrifugal evaporator.

2.16. Acquisition of mass spectrometry analysis

Before mass spectrometry analysis, the desalted peptides were dissolved in 10 μl of 0.1% FA in ultrapure water. The sample was analysis with Orbitrap Fusion Lumus mass spectrometer (Orbitrap, Thermo Fisher Scientific, USA) coupled with nano-flow ultra-performance liquid chromatography (UPLC) in data dependent acquisition (DDA) mode. Around 1 μg of peptide was injected into UPLC with a nano-flow C_{18} column (25 cm in length, 75 μm in diameter, practical size: 2 μm, pore size: 100 Å, ThermoFisher Scientific, USA) equilibrated with 0.1% FA. The peptides were eluted with a liner gradient of ACN, from 6 to 30% in 0.1% FA for an hour at 25℃. The Orbitrap was operated in positive mode with voltage of electrospray set at 2300 V. The temperature of the ion source was kept in 150 ℃. The range of mass-per-charge ratio for precursor ions scan was 350 to 1500 *m/z* with a resolution of 60000. The precursor ions with intensity higher than threshold (5 × 104, max. injection time: 20 ms) were selected for fragmentations (MS2) with isolation window of 1.6 *m/z*. High energy collisional dissociation was employed to fragmentate the precursor ions. The fragments were detected by Orbitrap with a resolution of 15000.

2.17. MS data searching

The MS data was processed by Mascot Distiller v2.5 (Matrix Science, USA) with a Mascot Server v2.5. The database used for searching was the CDSs predicted from our newly constructed genome of *A. fuciphagus*. The database contained around 16,800 sequences. The mass tolerances of MS and MS2 were 10 ppm. The charge states were set be 2+, 3 + and 4 +. Oxidation on methionine and carbamidomethylation on cysteine were set as variable and fixed modification respectively. Cleavage sites were set as C-terminal of arginine and lysine and one missed cleave was permitted. A random sequence database was used for decoy search. False discovery rate (FDR) was set to 1%. Results of searching were exported in Excel file format for reporting.

2.18. Protein homology search

The coding sequences of the annotated proteins in the genome of *A. fuciphagus* were searched against the coding sequences of human using BLAST all-vs-all in NCBI blast + v.2.10. The coding sequences of human were from Genome Reference Consortium Human Build 38 patch release 13 (GRCh38.p13).

2.19. Protein-protein interaction

STRING v11 (Search Tool for the Retrieval of Interacting proteins database) was used for protein–protein interaction as well as functional enrichment analysis (Szklarczyk et al., 2019). The UniProt entries or gene symbols of the proteins identified in EBN were submitted to STRING multiple proteins search against *G. gallus* database. Aggregate fold change was used to test the significances of the functional enrichment analysis conducted by SRTING (Szklarczyk et al., 2019). Pathways, gene ontology terms or other entries in STRING database with false discovery rate < 0.05 were considered as significantly enriched. The details of performing the protein–protein interaction were provided in the supplementary material.

2.20. SDS-PAGE and western blot

Ten micrograms of proteins from CAE, HAE and SGF were mixed with 2 × SDS-PAGE loading buffer [0.25 M Tris HCl, pH 6.8, 10% (w/v) SDS, 50% (v/v) glycerol, 1% (v/v) β-mercaptoethanol, 0.01% (v/v) bromophenol blue] and heated in a boiling water bath for 5 min. Subsequently, the samples were resolved on a 10% acrylamide gel using a

5% stacking gel. SDS-PAGE was electrophoresed at constant 60 V until the dye front reached the resolving gel. Then, SDS-PAGE was electrophoresed at constant 120 V until the dye front reached the bottom of the resolving gel. After SDS-PAGE separation, the samples were transferred onto 0.45 μm nitrocellulose membranes (Millipore, USA) by electroblotting. Subsequently, the membranes were blocked with 5% bovine serum albumin (BSA) in a TTBS buffer (20 mM Tris-HCl, 137 mM NaCl, 0.05% Tween 20, pH 7.6) for 1 h under room temperature prior to incubation with specific primary antibodies. To detect acidic mammalian chitinase (AC), rabbit polyclonal antibody against AC (ARP42601, Aviva Systems Biology) was used with dilution at 1:2000. To detect lysyl oxidase (LOX), rabbit polyclonal antibody against LOX (L4669, Sigma-Aldrich) was used with dilution at 1:500. The primary antibodies were incubated at 4 ℃ overnight. After sequential washing 6 times with TTBS (5 min each), the membranes were further incubated with HRP conjugated goat anti-rabbit IgG antibody (1:3000, Santa Cruz, USA) for 3 h at room temperature. Finally, the membranes were further washed with TTBS, while chemiluminescence for detection was developed using SuperSignal West Pico Chemiluminescence Kit (Pierce, USA) according to the manufacturer's instructions. Image acquisition was performed using the ChemiDoc XRS + system (Bio-Rad, USA).

2.21. Acidic mammalian chitinase activity assay

Chitinase assay kit (CS0980, Sigma Aldrich) was used to detect the AC activities in CAE, HAE and SGF according to protocol provided by the manufacturer. Around 0.5 μg of protein from CAE, HAE and SGF (in 10 μl of PBS, pH7.4) were mixed with 90 μl of assay buffer with 0.2 mg of 4-Nitrophenyl β-D-N,N',N"-triacetylchitotriose (N8638, Sigma Aldrich). Around 0.1 μg of chitinase from *Trichoderma viride* (C6242, Sigma Aldrich, in 10 μl of PBS, pH7.4) was used as positive control. After incubating at 37 ℃ for 30 min, the reactions were stopped by adding 100 μl of 0.5 M of Na_2CO_3. The abundances at 405 nm of the samples were measured using CLARIOstar from BMG Labtech. All the groups were assayed in triplicates and ANOVA followed by Bonferroni pos hoc test was used to test for the significances among groups using SPSS v.28.

2.22. Lysyl oxidase activity assay

Lysyl oxidase activity assay kit (ab112139, Abcam) was used to detect the LOX activities in CAE, HAE and SGF according to protocol provided by the manufacturer. Around 0.5 μg of protein from CAE, HAE and SGF (in 50 μl of PBS, pH7.4) were mixed with 50 μl of assay buffer with 0.01 unit of Horseradish Peroxidase and its substrate. Around 0.1 μg of recombinant human lysyl oxidase like 2 (SRP0375, Sigma Aldrich, in 50 μl of PBS, pH7.4) was used as positive control. After incubating at 37 ℃ for 30 min in dark, the fluorescence with excitation wavelength at 540 nm and emission wavelength at 590 nm were measured using CLARIOstar from BMG Labtech. All the groups were assayed in triplicates and ANOVA followed by Bonferroni pos hoc test was used to test for the significances among groups using SPSS v.28.

2.23. Data availability

The sequencing data for constructing the genome of *A. fuciphagus* was deposited in the NCBI Sequence Read Archives (SRA) under Bioproject ID: PRJNA607340.

3. Results

3.1. Construction of the genome dataset of A. Fuciphagus

Four live *A. fuciphagus* swiftlets were collected from a heavily guarded and gated swiftlet house in Cirebon, Indonesia (supplementary fig. S1, Supplementary Material online) (Kalbfleisch et al., 2018). After extraction of genomic DNA, linked-read sequencing was carried out with an Illumina NovaSeq 5000 with 150 bp pair-end reads and assembled with Supernova v.2.1.1 (Weisenfeld et al., 2017). The genome assembly was of 1,126.08 megabases (Mb) in length with GC content at around 41.95%. The genome was assembled up to scaffold level. N50 and L50 of scaffold were 4.92 and 55 Mb, respectively. There were 21,285 predicted CDSs and 16,081 of them were annotated with the Swiss-Prot database (release 2019_02, 559,228 entries). To assess the quality and completeness of the newly assembled genome, it was blasted against Aves Odb10 dataset using BUSCO v.4.0.2 (Seppey et al., 2019). The percentages of complete and single as well as complete and duplicated genes were 94.3% and 0.6%, respectively while those of fragmented and missing genes were 1.8% and 3.3%, respectively. The mitochondrial cytochrome *b* (Cytb) sequence of our specimens was aligned with other avian Cytb sequences in the NCBI nucleotide database using NCBI Blastn and the alignment was analyzed by NGPhylogeny.fr (Lemoine et al., 2019). The alignment results confirmed that our specimens obtained from Cirebon were indeed of *A. fuciphagus* (Fig. 1).

3.2. Orthology analysis of the genome of an EBN-producing swiftlet, A. fuciphagus

To identify any species-specific genetic variations in *A. fuciphagus*, orthology analysis was performed using OrthoFinder v.2.2.7 (Emms & Kelly, 2019) between *A. fuciphagus* and 10 other model avian species, namely *Chaetura pelagica, Gallus gallus, Meleagris gallopavo, Anas platyrhychos, Antrostomus carolinensis, Columba livia, Calypte anna, Taeniopygia guttata, Hirundo rustica*, and *Zonotrichia albicollis*. Avian species in the same clades and orders were successfully clustered in the phylogenetic tree by OrthoFinder (Fig. 2A).

After clustering orthologous CDSs into orthogroups and data filtering, the number of CDSs of each species in 13,517 orthogroups and the corresponding phylogenetic tree were imported to CAFExp to identify the orthogroups in *A. fuciphagus* that showed significant changes, either expansions or contractions, in the number of orthologous CDSs (Han et al., 2013). The estimated error of genome assembly (ε) across the 11 avian species was 0.095 and the estimated gain and loss rates (λ) were corrected with this estimated error. Since it was believed that different clades or orders might have different rates in gain and loss, so 4 λs were estimated accordingly. The estimated λs of *Galloanserae*, *Apodiformes*, *Caprimulgiformes* and *Columbiformes* as well as *Passeriformes* were hence calculated as 13.889, 2.378, 0.433, and 0.234 respectively (Fig. 2A). Twenty orthogroups were missing in both *A. fuciphagus* and *C. pelagica* (common chimney swift of the same clad) ($\rho < 0.05$) while another 50 orthogroups were missing in *A. funiphagus* only ($\rho < 0.05$). Moreover, the number of the orthologous CDSs in 196 orthogroups was found to be significantly changed, either expanded or contracted in the same manner in both *A. fuciphagus* and *C. pelagica* (Fig. 2A and Supplementary table 1). Furthermore, 332 orthogroups were found to show significant expansions or contractions ($\rho < 0.05$) that were uniquely in *A. fuciphagus* (Fig. 2A). The orthogroups with significant expansions or contractions (OSEC) were submitted to functional enrichment analysis using PANTHER v. 14 (Mi et al., 2019). To establish a reference set for the functional enrichment analysis, annotated genes from the *A. fuciphagus* genome were mapped to the database of *G. gallus* (a model organism) in PANTHER. In total, 13,747 annotated genes were mapped and used as the reference set for subsequent analysis. No enrichment was obtained with the 196 OSEC found in both *A. fuciphagus* and *C. pelagica* as well as those 70 missing orthogroups in *A. fuciphagus*. The 332 OSEC found only in *A. fuciphagus* were significantly enriched (FDR < 0.05) in various generations, organizations and regulations of cellular components and organelles (GO:0032989, GO:0044087, GO:0051239, GO:0033043, & GO:0120039) (Fig. 2B). Several gene ontology (GO) terms related to the nervous system and neurons (GO:0030900, GO:0021954, GO:0031175, GO:0050877 & GO:0007611) were also found to be significantly enriched (Fig. 2B). When all 528 OSEC (332 + 196) in *A. fuciphagus* were analyzed for

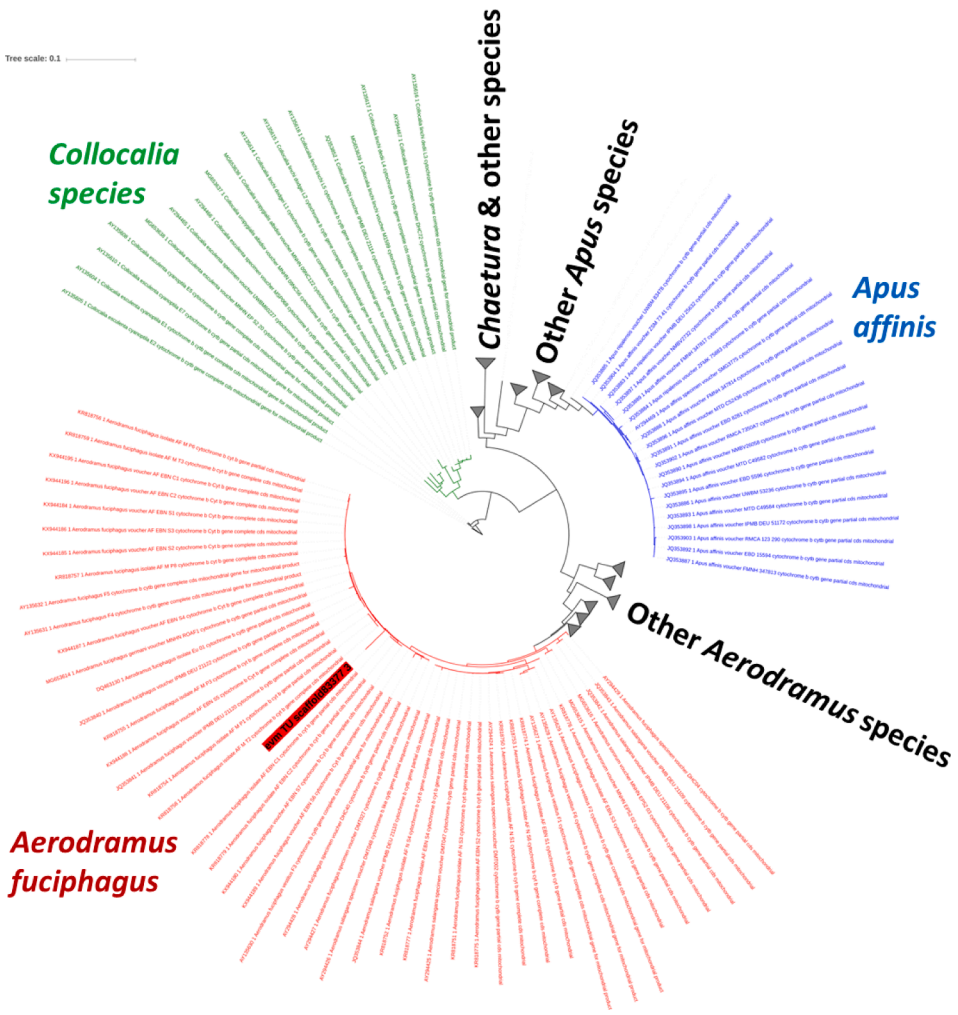


Fig. 1. Speciation of EBN-producing swiftlets from swiftlet's house in Cirebon, Indonesia. The phylogenetic trees were constructed using mitochondrial cytochrome *b* (CytB) sequence. The CytB sequence from the specimen (highlighted with a red rectangle) was aligned with other avian CytB sequences in the NCBI database using NCBI Blastn. The alignment was then imported to NGPhylogeny.fr to construct the phylogenetic tree. The CytB sequence from the specimen were clustered with other CytB sequences from other *A. fuciphagus* in the database. This result confirmed that our specimen is *A. fuciphagus*. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

functional enrichment, it was found that various regulations of epigenetic, transcription and histone methylation were enriched (GO:0040029, GO:0006355, & GO:0031060) (Fig. 2C). Among the OSEC, most of them were found to be contracted, like forkhead box protein P1 and P2 (FOXP1 and FOXP2), while only a few were found to be expanded, like FOXJ1 and coiled-coil domain-containing protein 63 (CCDC63) (Supplementary table 1).

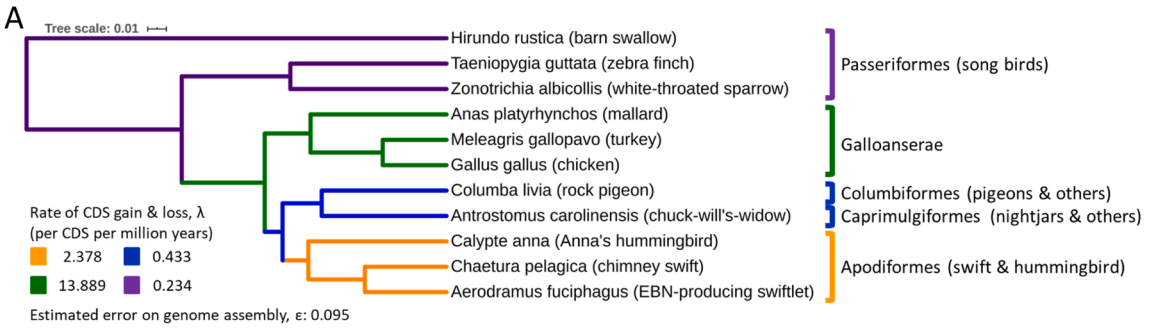
3.3. Multiple paralogous CCDC63 CDSs found in the genome of *A. fuciphagus*

There were 14 paralogous CDSs of CCDC63 found in the genome of *A. fuciphagus* that were significantly expanded only in *A. fuciphagus* (CAFExp $\rho < 0.00001$, Supplementary table 1). These 14 paralogous CCDC63 CDSs were located in different scaffolds in the genome. They were denoted as "evm.model.", followed by the number of the scaffold where the paralogous CDS was located. However, the high number of paralogous CDSs in a genome could be a result of errors during sequencing, assembly and gene prediction. To validate the number of paralogous genes of CCDC63 in the genome of *A. fuciphagus*, the scaffolds from the genomes were mapped to the genome of *G. gallus* (version: GRCg7b) using Minimap2 and visualized with D-GENIES (Cabanettes & Klopp, 2018, Li, 2018). The scaffolds that were annotated with at least 3 CDSs from different genes were considered and included in the analysis. The use of scaffolds that had at least 3 annotated genes would allow a more reliable mapping. Scaffolds 85, 100, and 138 were found to contain 32, 57, and 138 CDSs, respectively (Supplementary table 2). Each of them had a paralogous CCDC63 CDS (evm.model.scaffold85.1, evm.model.scaffold100.19, & evm.model.scaffold138.70). The rest of the scaffolds with other paralogous CCDC63 CDSs in the genome of *A. fuciphagus* were found to have only 1 to 2 CDSs. It would be difficult to correctly map such a short scaffold to the chromosomes of *G. gallus* without referencing chromosomal coordinates of other genes. Moreover, multiple paralogous CDSs located on short scaffolds would be likely to be a result of errors of sequencing and assembly. Thus, they were discarded from the mapping. Scaffolds 85, 100, and 138 were mapped to chromosomes 2 (NC_052533.1), 3 (NC_052534.1) and 15 (NC_052546.1) in the genome of *G. gallus* respectively (Fig. 3). Evm.model.scaffold138.70 was mapped to the only CCDC63 gene on chromosome 15 of *G. gallus* (Supplementary table 2). Transcription start sites and TATA box of those 3 paralogous CCDC63 CDSs (evm.model.scaffold85.1, evm.model.scaffold100.19, & evm.model.scaffold138.70) were predicted using Promoter2.0 (Knudsen, 1999) and OProf (Ambrosini et al., 2003) respectively. The transcription start sites and TATA boxes were predicted within 750 base pairs upstream from the start codon of each paralogous CCDC63 CDS (Fig. 4).

3.4. Proteins identified in edible bird nest

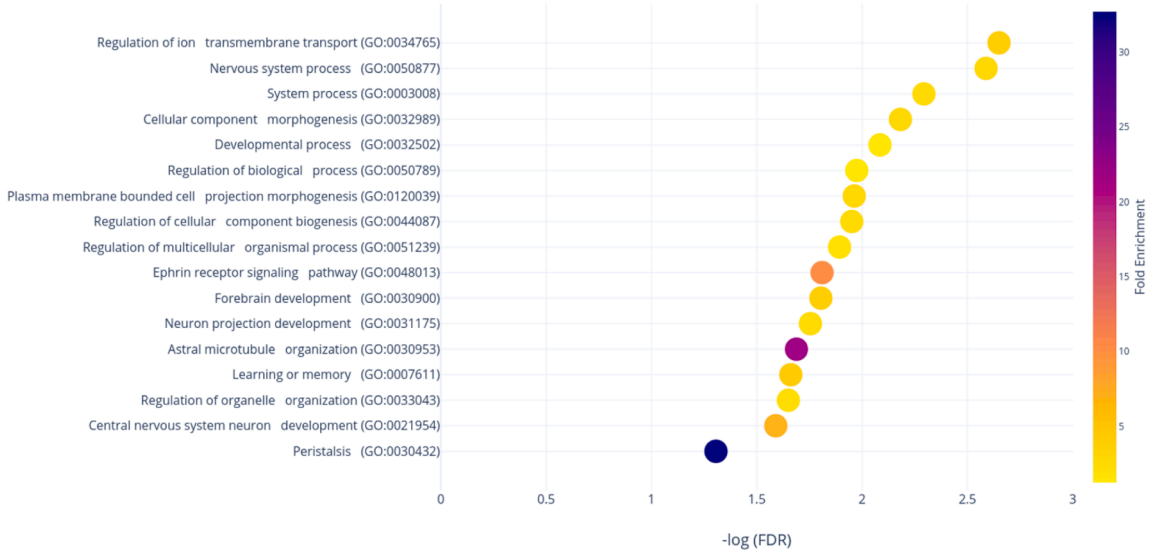
With the genome of *A. fuciphagus*, the CDSs were predicted to construct the database for proteomic analysis of EBN. Three hundred and ninety-eight proteins have been identified in EBN using nano-flow C_{18} UPLC coupled Orbitrap mass spectrometer with an estimated false discovery rate of<1% (Supplementary table 3). One hundred and forty-

A

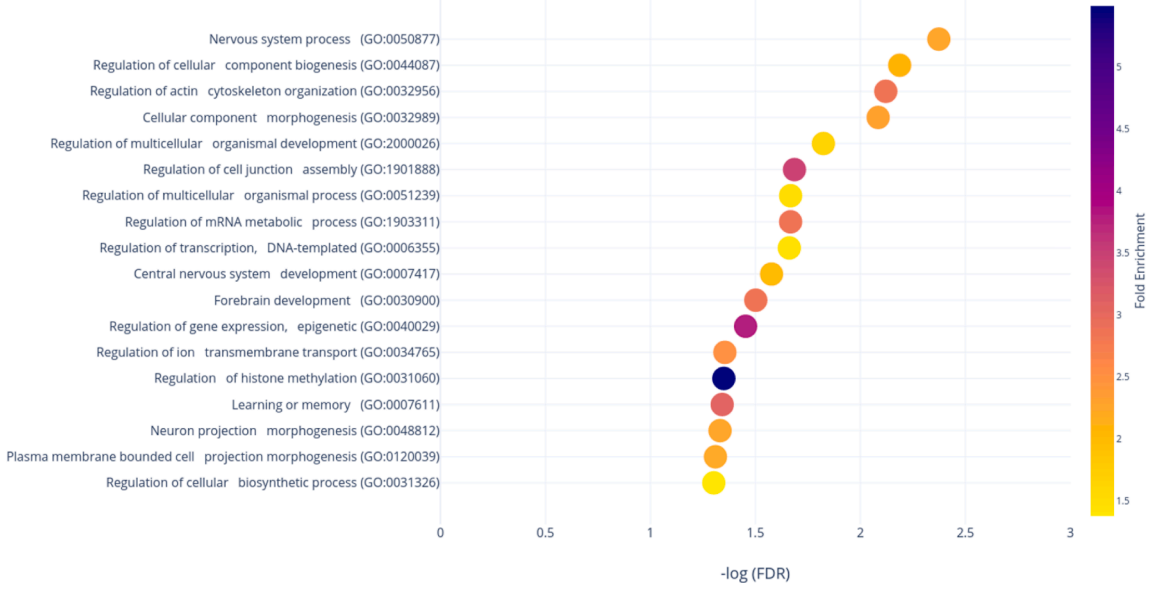


Orthogroups	*C. Pelagica* & *A. fuciphagus*	Uniquely in *A. fuciphagus*
Significantly changed in numbers of orthologues CDSs	196	332
Missing	20	50

B



C



(caption on next page)

Fig. 2. Expansions and contractions of paralogous CDSs in *A. fuciphagus*. CDSs from the genome of *A. fuciphagus* as well as those from *C. pelagica, G. gallus, M. gallopavo, A. platyrhychos, A. carolinensis, C. livia, C. anna, T. guttata, H. rustica* and *Z. albicollis* were analyzed by Orthofinder v.2.2.7 to cluster orthologous genes among the avian species. The results from Orthofinder were imported into CAFExp to identify the orthogroups/ genes that had significant changes in the number of paralogous CDSs in *A. fuciphagus*. The estimated error of genome assembly (ε) across 11 avian species was 0.095 (a). The estimated gain and loss rates (λ) were corrected with the estimated error. The estimated λs of *Galloanserae*, *Apodiformes*, *Caprimulgiformes* and *Columbiformes* as well as *Passeriformes* were 13.889, 2.378, 0.433, and 0.234 respectively (A). There were 196 orthogroups found to be significantly changed in the number of paralogous CDSs, either expanded or contracted ($\rho < 0.05$) in both *A. fuciphagus* and *C. pelagica* in the same manner (A). CDSs from 20 genes were missing in both species and CDSs from another 50 genes were missing only in *A. fuciphagus*. There were 332 orthogroups found to be significantly changed ($\rho < 0.05$) only in *A. fuciphagus* (A). The orthogroups with significant expansions or contractions (OSEC) were submitted to functional enrichment analysis using PANTHER v. 14. No enrichment was obtained with those 196 OSEC commonly found in both *A. fuciphagus* and *C. pelagica* as well as those 70 genes that were missing in *A. fuciphagus*. The remaining 332 OSEC found only in *A. fuciphagus* were significantly enriched (FDR < 0.05) in various regulations and developmental processes (B). All 528 OSEC in *A. fuciphagus* were also enriched in similar regulatory and developmental processes, except for regulations of cell junction assembly (GO:1901888), transcription (GO:0006355) and epigenetic (GO:0040029) as well as histone methylation (GO:0031060) (C).

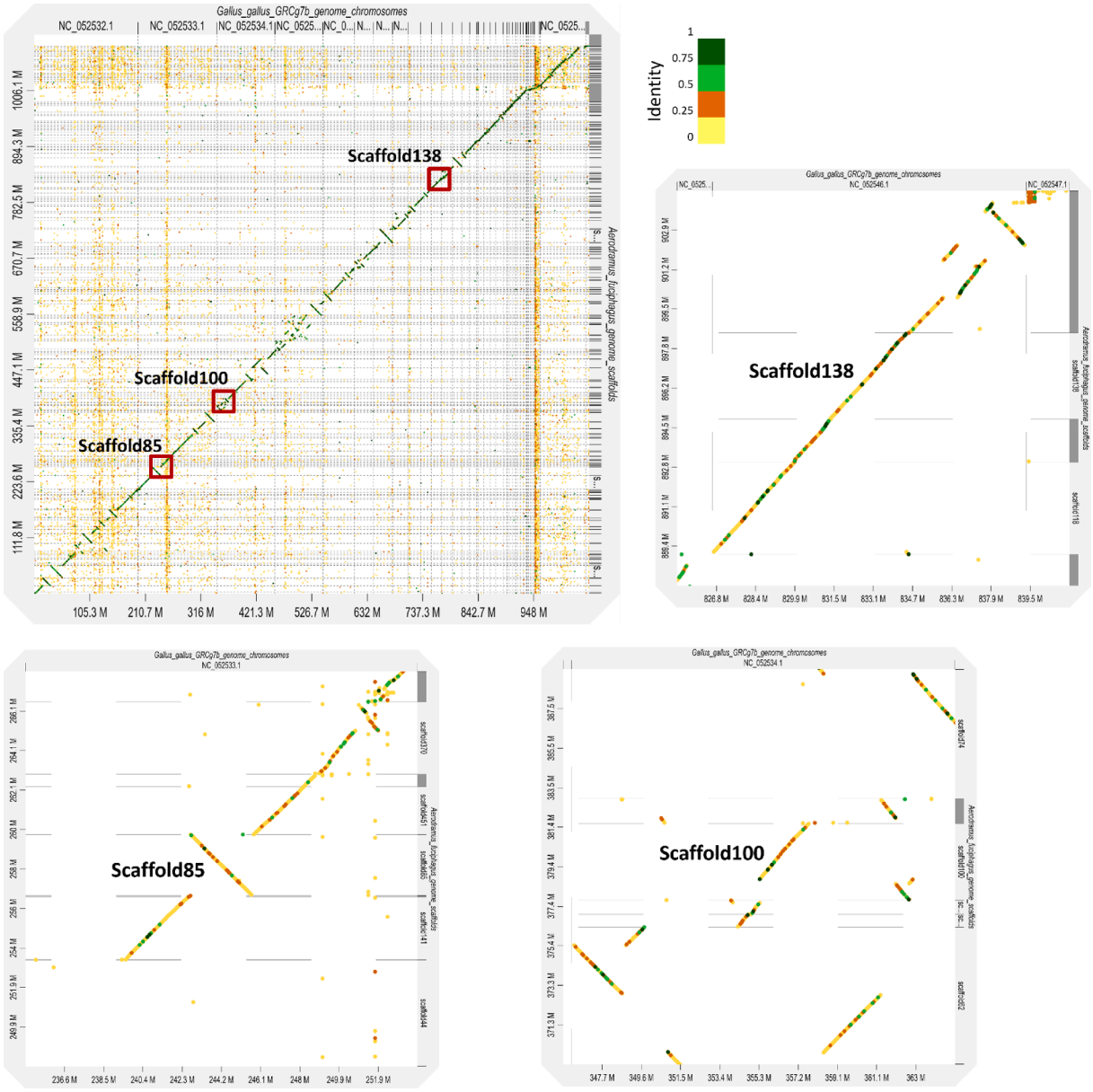


Fig. 3. Mapping between the scaffolds from *A. fuciphagus* and the chromosomes from *G. gallus*. The scaffolds from the genome of *A. fuciphagus* were mapped to the chromosomes from *G. gallus* (version: GRCg7b) using Minimap2 and visualized using D-Genies. The top left panel showed the results of the mappings. The mappings of scaffolds 85, 100, and 138 were highlighted in red squares. Enlarged alignments of these 3 scaffolds were provided separately. Scaffold 85 was mapped to chromosome 2 (NC_052533.1), starting from around 46.04 to 49.03 million base pair. Scaffold 100 was mapped to chromosome 3 (NC_052534.1), starting from around 9.17 to 16.62 million base pair. Scaffold 138 was mapped to chromosome 15 (NC_052546.1), starting from around 4.75 to 9.18 million base pair. The detailed mappings between CDSs from these 3 scaffolds and genes from the chromosomes of *G. gallus* were reported in Supplementary table S5 in Supplementary Material online. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

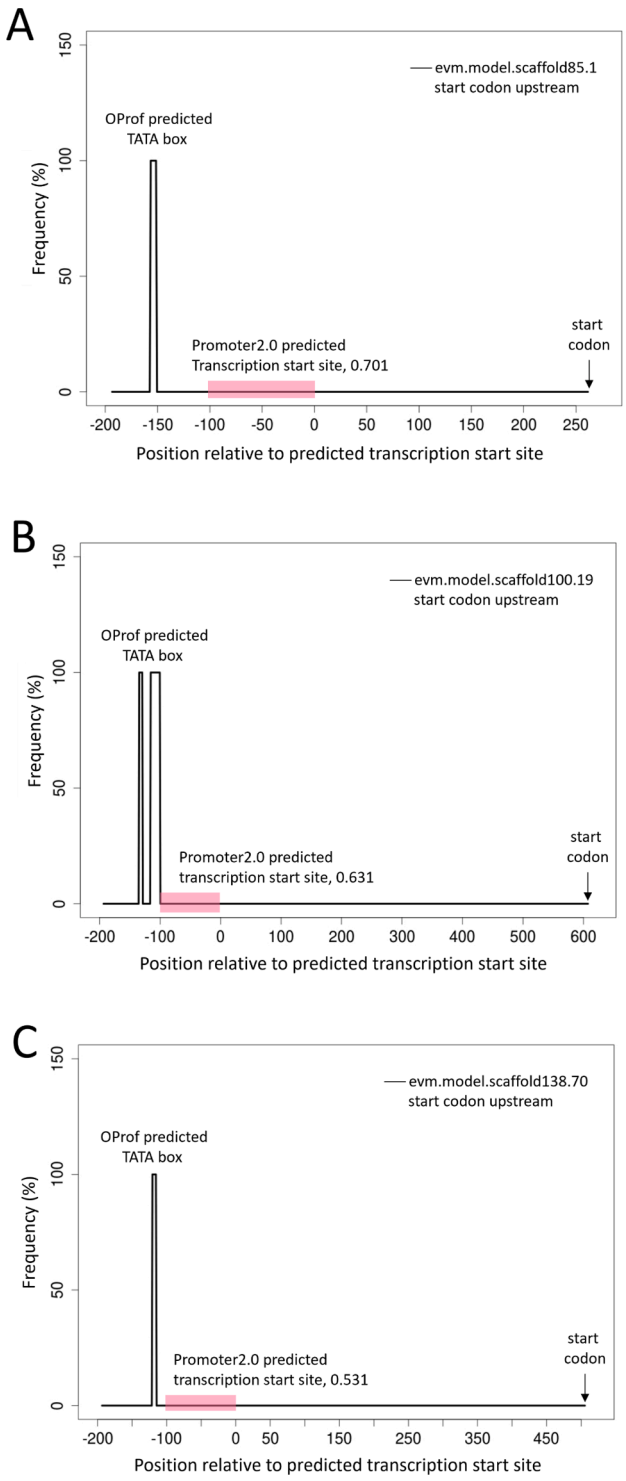
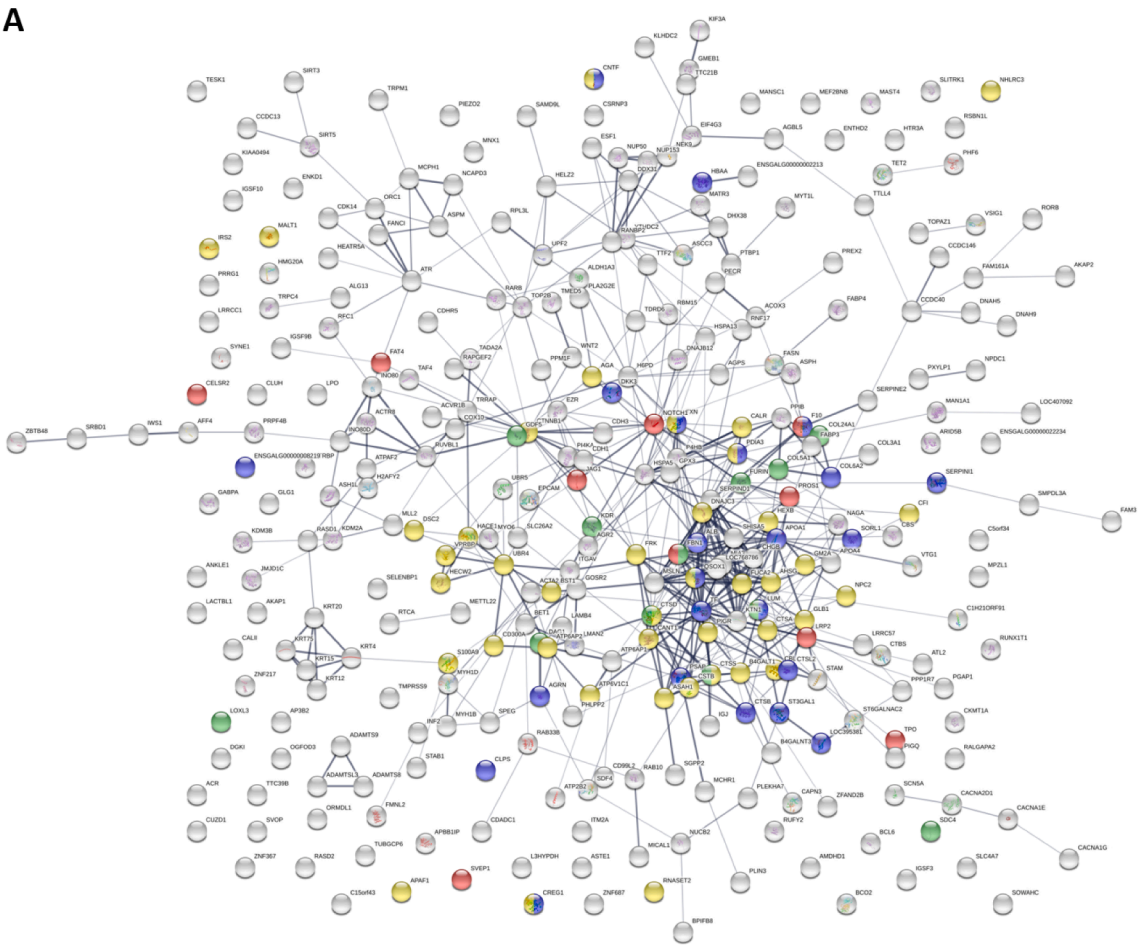


Fig. 4. Predictions of TATA boxes and transcription start sites of 3 paralogous CCDC63 scaffolds. The upstream nucleotide sequence (1000 base pairs) from the start codon of each paralogous CCDC63 CDS of interest (evm.model.scaffold85.1, evm.model.scaffold100.19 & evm.model.scaffold138.70) was analyzed individually by OProf and Promoter2.0 to predict the TATA box and transcription start site respectively. The transcription start site of each CDS of interest was predicted. Position scores of the predictions ranged from 0.531 to 0.701. The positions scoring ranged from 0.5 to 0.8 indicated that there was 65% of chance of having a true transcription start site within 100 base pairs upstream (highlighted in red rectangle). The TATA box of each CDS of interest was predicted. The TATA boxes were found to be in the upstream of the corresponding transcription start sites. The TATA box sequence of evm.model.scaffold85.1 was 5′ TTTAAA 3′ (A). There were 2 predicted TATA box sequences in the upstream of evm.model.scaffold100.19: 5′ TTTAAA 3′ and 5′ TACATATAAGTATTA 3′ (B). The TATA box sequence of evm.model.scaffold138.70 was 5′ TTAAAA 3′ (C). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



Color	Database	GO-term/ Domain	Description	Count in gene set	False discovery rate
Red	SMART Protein Domains	SM00179	Calcium-binding EGF-like domain	10 of 79	0.0024
Blue	Cellular Component (GO)	GO:0005576	Extracellular region	23 of 299	3.61e-05
Green	Reactome Pathways	GGA-1474244	Extracellular matrix organization	12 of 145	0.0052
Yellow	Reactome Pathways	GGA-168256	Immune system	41 of 936	0.00086

Fig. 5. Functional enrichment analysis of proteins in the EBN proteome. Protein-protein interaction network was generated by STRING v11. Spheres in the network represented proteins that were identified in EBN (A). Edges between the spheres represented interactions among the proteins. Thicker edges indicated higher confidences that the proteins jointly contribute to a shared function. Interactions in the STRING v11 included direct (physical) and indirect (functional) associations, which were derived from genomic context predictions, high-throughput lab experiments, (conserved) co-expression, automated text mining and knowledge in other databases. Thickness of the edges between nodes (proteins) indicated strength (confidence) of interactions between the proteins supported by the information in the STRING database. Proteins in the proteome of EBN was significantly enriched (FRD < 0.05) in proteins with Ca2 + -binding EGF-like domain (SM00179, red). The proteome of EBN was also enriched with extracellular proteins (GO:0005576, blue) as well as proteins that involved extracellular matrix organization (GGA-1474244, green) and immune system (GGA-168256, yellow). Further functional enrichment analysis on biological process (B), cellular component (C) and molecular function (D) were performed using PANTHER v14. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

eight of them were homologs to proteins found in human saliva (Supplementary table 3) (Sivadasan et al., 2015), supporting the notion that EBNs are mainly composed of swiftlet saliva. The average sequence similarity was 68.4% with a standard deviation of 17.2%. Functional enrichment performed with STRING v.11 revealed that the proteome of EBN was significant enriched with extracellular proteins (GO:0005576) and proteins related to extracellular matrix (ECM) organization (GGA-1474244) and immune response (GGA-168256) (Fig. 5A). No standalone biochemical EGF molecule could be identified in the EBN, but proteins which contained Ca^{2+}-binding EGF-like domains (smart00179) were found in EBN (Fig. 5A). As shown in the supplementary table 3, LRP2, PROS1 and SVEP1 were proteins with EGF-like domains that were found in both human saliva and EBN. Furthermore, Fibrillin-1, Protocadherin Fat 4 and Coagulation factor X were proteins with EGF-like domains that were only found in EBN (Supplementary table 3). Functional enrichment analysis using PANTHER v.14.1 further

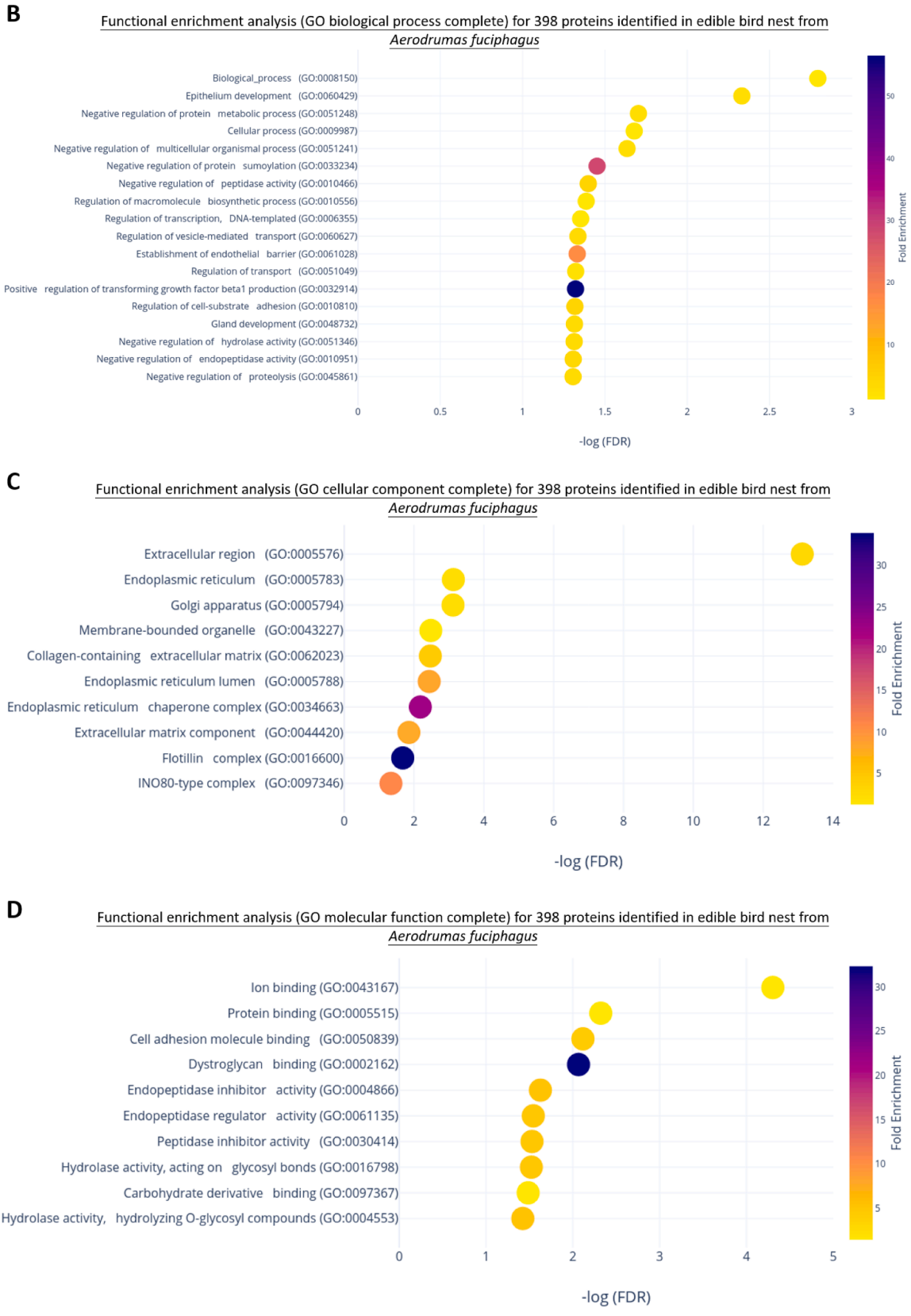


Fig. 5. (*continued*).

revealed that the proteins composed of EBN was significantly enriched in regulation of transcription (GO:0006355), TGFβ production (GO:0032914) and also involve in epithelium development (GO:0060429), just like those 234 differentially expressed genes that had significant changes in gene count (Fig. 3B). Moreover, gland development (GO:0048732), regulation of vesicle-mediated transport (GO:0060627), proteolysis (GO:0045861) and cell-substrate adhesion (GO:0010810) were also found to be enriched (Fig. 5B).

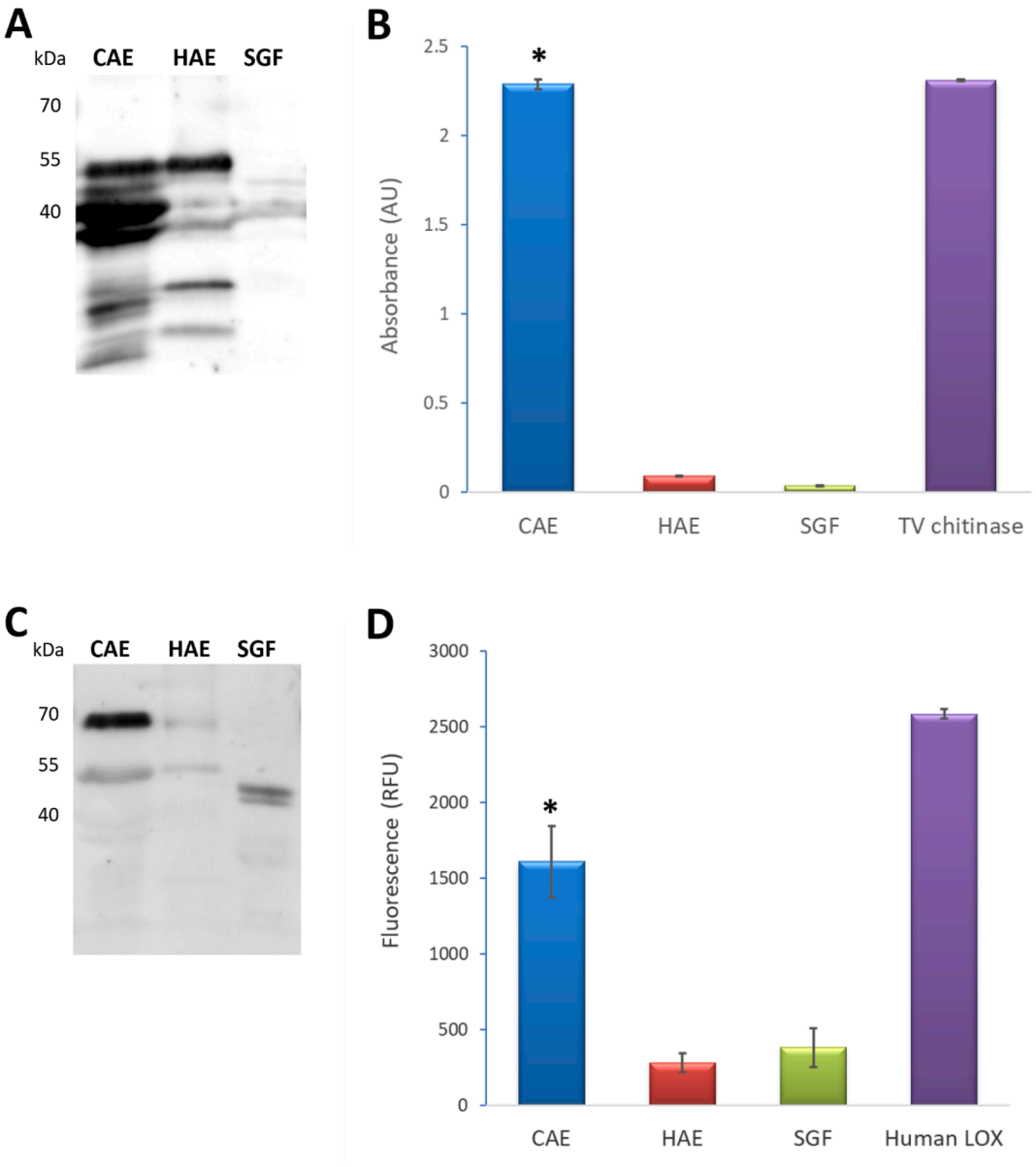


Fig. 6. Western blots and functional assays of acidic mammalian chitinase (AC) and lysyl oxidase (LOX). A few proteins in the cold aqueous extract of EBN (CAE) were detected by using anti-AC antibody (A). In brief, the signal of CAF was much intense than those of hot aqueous extract (HAE) and simulated gastric fluid after digesting EBN (SGF). For the chitinase activity assay, the absorbance of CAE was significantly higher than those of HAE and SGF (B, n = 3 for each group, *, $\rho < 0.001$). The absorbance of the chitinase assay was directly proportional the chitinase activity in the sample. Chitinase from *Trichoderma viride* was used as a positive control for the chitinase activity assay. An intense signal was detected in CAE using anti-LOX antibody, but not in HAE and SGF (C). For the LOX activity assay, the fluorescence emitted from CAE was significantly higher than those from HAE and SGF (D, n = 3 for each group, *, $\rho < 0.001$). The fluorescence emitted in the LOX activity assay was directly proportional to the LOX activity in the sample. Human recombinant LOX was used as a positive control in the LOX activity assay. The original western blot results have been provided in the supplementary material.

3.5. Western blots and activity assays of acidic mammalian chitinase and lysyl oxidase

To validate the proteomic results, western blots were conducted to detect acidic mammalian chitinase (AC) and lysyl oxidase (LOX) in the cold aqueous extract (CAE), and the hot aqueous extract (HAE) of EBN as well as the simulated gastric fluid (SGF) after digesting EBN. A few proteins with apparent molecular weight (aMW) at around 50 to 35 kDa were detected in CAE using the anti-AC antibody (ARP42601, Aviva Systems Biology) and the signal of these proteins in CAE were much intense than those in HAE and SGF (Fig. 6A). A protein with aMW at around 70 kDa in CAE was detected by using anti-LOX antibody (L4669, Sigma-Aldrich) and the signal of the protein in CAE was much intense than other proteins detected in HAE and SGF (Fig. 6C). To further validate the presences of AC and LOX in EBN, the functional assays on AC and LOX enzymatic activities were also conducted. A strong enzymatic activity of AC was observed in CAE, which were significantly higher than those in HAE and SGF ($\rho < 0.001$, Fig. 6B). A high enzymatic activity of LOX was also observed in CAE, which were significantly higher than those in HAE and SGF ($\rho < 0.001$, Fig. 6D).

4. Discussion

The first genome of the EBN-producing swiftlet, *A. fuciphagus* have been constructed. The results from BUSCO indicated the completeness of the newly constructed genome, in terms of expected gene contents, was

high. In addition, the low percentages of the duplicated and fragmented genes meant that the genome should have little wrong phasing haplotypes. The overall quality and completeness of the *A. fuciphagus* genome were similar to other newly constructed avian genomes (Prost et al., 2019, Pujolar et al., 2017). The results from orthology analysis clearly indicated that *A. fuciphagus* at least has 3 CCDC63 genes in its genome (Figs. 3 and 4). *C. pelagica* was the closet relative to *A. fuciphagus* which included in the orthology analysis (Fig. 2A). It has only a single CCDC63 gene in its genome (NCBI accession no: XM_009998451.1). This implied that the expansion of CCDC63 gene in *A. fuciphagus* should occur after the divergence between *A. fuciphagus* and *C. pelagica* from their common ancestors. According to literature, CCDC63 was an important protein composing axonemes in motile cilia and flagella in sperm (Young et al., 2015). As cilia can be found in many different cell types, including ciliated stratified epithelial cells in salivary ducts of lingual apparatus of avian species (Jerber et al., 2014, Karunakaran et al., 2020, Onoufriadis et al., 2013, Triantafyllou et al., 2018, Zur Lage et al., 2019), so it was suspected that the expansion of CCDC63 may affect the morphologies and functions in these cell types. Beside CCDC63, there were other 331 OSCE found only in *A. fuciphagus* which were involved in various biological functions, like ephrin receptor signaling pathway (GO:0048013, Fig. 2A). Ephrin receptor signaling pathway was found to play an important role in tumorigenesis of salivary glands in human (Pergaris et al., 2021). The genetic changes in ephrin receptor signaling pathway in *A. fuciphagus* implied that the development of the salivary glands in *A. fuciphagus* might be different from other avian species. Other OSCE were involved in central nervous system development (GO:0021954), as well as learning or memory (GO:0007611, Fig. 2A). This suggested that the accumulation of OSCE in the nervous system along the evolution may alter neuronal functions or behaviors of *A. fuciphagus.* However, further investigations were required to study and validate the effects of these unique genotypes on the EBN-producing phenotype and the nesting behavior of *A. fuciphagus*.

Using the CDSs predicted from the first genome of *A. fuciphagusas* the database for proteomic analysis of EBN, around 400 proteins have been identified, no standalone EGF was found in the proteome of EBN, but a dozen of proteins which contained Ca^{2+}-binding EGF-like domains (smart00179) were found in EBN (Fig. 5A). This implied that the EGF-like domains in those proteins in EBN should be responsible to the EGF-like activities of EBN reported in the previous studies. To validate the hypothesis, the EGF-like domains in EBN should be either purified or cloned for subsequent bioactivity assays. The functional enrichment analysis and the protein–protein interaction analysis revealed that EBN was a saliva-like ECM containing the proteins with EGF-like domains and other proteins that are known to regulate various biological processes, like immune responses (Fig. 5A). Cathepsin B, D and S were identified in the proteome of EBN (Supplementary table 3). Cathepsin protein family was known to regulate both innate and adaptive immune responses against bacterial infections and cancers (Conus & Simon, 2010, Yan et al., 2020). Thus, cathepsins and other immune related proteins in EBN may be responsible for the immunoregulating effects of the extracts of EBN (Zhao et al., 2016).

Interestingly, enzymes like AC and LOX in EBN were found to be still active after the extraction with distilled water at room temperature (Fig. 6B & 6D). Their activities would significantly reduce when extracting EBN with hot distilled water or after simulated gastric digestion. The findings implied that extraction methods have a significant impact on the biological activity of the extract of EBN. Thus, extra caution has to be paid on choosing suitable extraction methods to preserve the biological activities of the extract of EBN for subsequent functional assays.

5. Conclusion

The current study reported the first genome of EBN-producing *A. fuciphagus* and the species-specific genotype of *A. fuciphagus*, which was the expansion of CCDC63 paralogous genes, has been revealed. With the genome, the current study has identified 398 proteins in EBN. The function enrichment analysis revealed that EBN contained various extracellular proteins and other proteins related to various biological functions, including ECM organization and immune functions. The proteins with Ca2 + -binding EGF-like domains in EBN might be responsible for the EGF-like activity in the aqueous extract of EBN. In addition, the enzymatic activities of AC and LOX in EBN were found to susceptible to the methods of extractions. The current study has not just revealed the species-specific genotype of the EBN-producing swiftlet, *A. fuciphagus*, but also revealed the proteome of EBN. This established an important foundation for subsequently studies on efficacies of EBN.

CRediT authorship contribution statement

Hang-kin Kong: Conceptualization, Methodology, Formal analysis, Writing – original draft. **Zoe Chan:** Software, Formal analysis. **Clare Sau-woon Yan:** Investigation. **Pak-yeung Lo:** Investigation. **Wing-tak Wong:** Funding acquisition. **Ka-hing Wong:** Funding acquisition, Writing – review & editing, Supervision. **Samuel Chun-lap Lo:** Funding acquisition, Writing – review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by research grants from the Hong Kong Polytechnic University (A/C:1-99X3 and 1-ZVKU). We gratefully acknowledge the support of the Hong Kong Polytechnic University Research Facilities in Chemical and Environmental Analysis (UCEA), Life Sciences (ULS), Big Data Analysis (UBDA) as well as the Centralized Animal Facilities (CAF). We also like to acknowledge the great help from Dr. Kevin HK Kwok of the HK PolyU and Prof. Kris Herawan Timotius of the Krida Wacana Christian University in Cirebon, Indonesia, for helping the research team to obtain the live EBN-producing swiftlet specimens.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodres.2022.111670.

References

Ambrosini, G., Praz, V., Jagannathan, V., & Bucher, P. (2003). Signal search analysis server. *Nucleic Acids Research, 31*(13), 3618–3620. https://doi.org/10.1093/nar/gkg611

Birney, E., Clamp, M., & Durbin, R. (2004). GeneWise and Genomewise. *Genome Research, 14*(5), 988–995. https://doi.org/10.1101/gr.1865504

Cabanettes, F., & Klopp, C. (2018). D-GENIES: Dot plot large genomes in an interactive, efficient and simple way. *PeerJ, 6*, Article e4958. https://doi.org/10.7717/peerj.4958

Chua, L. S., & Zukefli, S. N. (2016). A comprehensive review on edible bird nests and swiftlet farming. *Journal of Integrative Medicine, 14*(6), 415–428. https://doi.org/10.1016/S2095-4964(16)60282-0

Conus, S., & Simon, H. U. (2010). Cathepsins and their involvement in immune responses. *Swiss Medical Weekly, 140*, 4–11. https://doi.org/10.4414/smw.2010.13042

Criscuolo, A., & Gribaldo, S. (2010). BMGE (Block Mapping and Gathering with Entropy): A new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evolutionary Biology, 10*, 210. https://doi.org/10.1186/1471-2148-10-210

Emms, D. M., & Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology, 20*(1), 238. https://doi.org/10.1186/s13059-019-1832-y

Forni, G., Puccio, G., Bourguignon, T., Evans, T., Mantovani, B., Rota-Stabelli, O., & Luchetti, A. (2019). Complete mitochondrial genomes from transcriptomes: Assessing pros and cons of data mining for assembling new mitogenomes. *Scientific Reports, 9*(1), 14806. https://doi.org/10.1038/s41598-019-51313-7

Haas, B. J., Delcher, A. L., Mount, S. M., Wortman, J. R., Smith, R. K., Jr., Hannick, L. I., … White, O. (2003). Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Research, 31*(19), 5654–5666. https://doi.org/10.1093/nar/gkg770

Haghani, A., Mehrbod, P., Safi, N., Aminuddin, N. A., Bahadoran, A., Omar, A. R., & Ideris, A. (2016). In vitro and in vivo mechanism of immunomodulatory and antiviral activity of Edible Bird's Nest (EBN) against influenza A virus (IAV) infection. *J Ethnopharmacol, 185*, 327–340.

Hahn, M. W., De Bie, T., Stajich, J. E., Nguyen, C., & Cristianini, N. (2005). Estimating the tempo and mode of gene family evolution from comparative genomic data. *Genome Research, 15*(8), 1153–1160. https://doi.org/10.1101/gr.3567505

Han, M. V., Thomas, G. W., Lugo-Martinez, J., & Hahn, M. W. (2013). Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Molecular Biology and Evolution, 30*(8), 1987–1997. https://doi.org/10.1093/molbev/mst100

Hun Lee, T., Hau Lee, C., Alia Azmi, N., Kavita, S., Wong, S., Znati, M., & Ben Jannet, H. (2020). Characterization of Polar and Non-Polar Compounds of House Edible Bird's Nest (EBN) from Johor, Malaysia. *Chem Biodivers, 17*(1), Article e1900419. https://doi.org/10.1002/cbdv.201900419

Jerber, J., Baas, D., Soulavie, F., Chhin, B., Cortier, E., Vesque, C., … Durand, B. (2014). The coiled-coil domain containing protein CCDC151 is required for the function of IFT-dependent motile cilia in animals. *Human Molecular Genetics, 23*(3), 563–577. https://doi.org/10.1093/hmg/ddt445

Junier, T., & Zdobnov, E. M. (2010). The Newick utilities: High-throughput phylogenetic tree processing in the Unix shell. *Bioinformatics, 26*(13), 1669–1670. https://doi.org/10.1093/bioinformatics/btq243

Kalbfleisch, T. S., Rice, E. S., DePriest, M. S., Jr., Walenz, B. P., Hestand, M. S., Vermeesch, J. R., …, & MacLeod, J. N. (2018). Improved reference genome for the domestic horse increases assembly contiguity and composition. Communications Biology, 1, 197. https://doi.org/10.1038/s42003-018-0199-z.

Karunakaran, K. B., Chaparala, S., Lo, C. W., & Ganapathiraju, M. K. (2020). Cilia interactome with predicted protein-protein interactions reveals connections to Alzheimer's disease, aging and other neuropsychiatric processes. *Scientific Reports, 10*(1), 15629. https://doi.org/10.1038/s41598-020-72024-4

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution, 30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Knudsen, S. (1999). Promoter2.0: For the recognition of PolII promoter sequences. *Bioinformatics, 15*(5), 356–361. https://doi.org/10.1093/bioinformatics/15.5.356

Kong, H. K., Wong, K. H., & Lo, S. C. (2016). Identification of peptides released from hot water insoluble fraction of edible bird's nest under simulated gastro-intestinal conditions. *Food Research International, 85*, 19–25. https://doi.org/10.1016/j.foodres.2016.04.002

Kong, Y. C., Keung, W. M., Yip, T. T., Ko, K. M., Tsao, S. W., & Ng, M. H. (1987). Evidence that epidermal growth factor is present in swiftlet's (Collocalia) nest. *Comparative Biochemistry and Physiology. B, Comparative biochemistry, 87*(2), 221–226. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=3497769.

Lee, T. H., Wani, W. A., Lee, C. H., Cheng, K. K., Shreaz, S., Wong, S., … Azmi, N. A. (2021). Edible Bird's Nest: The Functional Values of the Prized Animal-Based Bioproduct From Southeast Asia-A Review. *Front Pharmacol, 12*, Article 626233. https://doi.org/10.3389/fphar.2021.626233

Lefort, V., Desper, R., & Gascuel, O. (2015). FastME 2.0: A Comprehensive, Accurate, and Fast Distance-Based Phylogeny Inference Program. *Molecular Biology and Evolution, 32*(10), 2798–2800. https://doi.org/10.1093/molbev/msv150

Lemoine, F., Correia, D., Lefort, V., Doppelt-Azeroual, O., Mareuil, F., Cohen-Boulakia, S., & Gascuel, O. (2019). NGPhylogeny.fr: New generation phylogenetic services for non-specialists. *Nucleic Acids Research, 47*(W1), W260–W265. https://doi.org/10.1093/nar/gkz303

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics, 34*(18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Liu, X., Lai, X., Zhang, S., Huang, X., Lan, Q., Li, Y., … Yang, G. (2012). Proteomic profile of edible bird's nest proteins. *Journal of Agricultural and Food Chemistry, 60*(51), 12477–12481. https://doi.org/10.1021/jf303533p

Looi, Q. H., Amin, H., Aini, I., Zuki, M., & Omar, A. R. (2017). De novo transcriptome analysis shows differential expression of genes in salivary glands of edible bird's nest producing swiftlets. *BMC Genomics, 18*(1), 504. https://doi.org/10.1186/s12864-017-3861-9

Ma, F., & Liu, D. (2012). Sketch of the edible bird's nest and its important bioactivities. *Food Research International, 48*(2), 559–567.

Matsukawa, N., Matsumoto, M., Bukawa, W., Chiji, H., Nakayama, K., Hara, H., & Tsukahara, T. (2011). Improvement of bone strength and dermal thickness due to dietary edible bird's nest extract in ovariectomized rats. *Bioscience, Biotechnology, and Biochemistry, 75*(3), 590–592. https://doi.org/10.1271/bbb.100705

Mi, H., Muruganujan, A., Huang, X., Ebert, D., Mills, C., Guo, X., & Thomas, P. D. (2019). Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). *Nature Protocols, 14*(3), 703–721. https://doi.org/10.1038/s41596-019-0128-8

Ng, M. H., Chan, K. H., & Kong, Y. C. (1986). Potentiation of mitogenic response by extracts of the swiftlet's (Collocalia) nest. *Biochemistry International, 13*(3), 521–531. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=3790144.

Onoufriadis, A., Paff, T., Antony, D., Shoemark, A., Micha, D., Kuyt, B., …, Mitchison, H. M. (2013). Splice-site mutations in the axonemal outer dynein arm docking complex gene CCDC114 cause primary ciliary dyskinesia. American Journal of Human Genetics, 92(1), 88-98. https://doi.org/10.1016/j.ajhg.2012.11.002.

Pergaris, A., Danas, E., Goutas, D., Sykaras, A. G., Soranidis, A., & Theocharis, S. (2021). The Clinical Impact of the EPH/Ephrin System in Cancer: Unwinding the Thread. *Int J Mol Sci, 22*(16). https://doi.org/10.3390/ijms22168412

Prost, S., Armstrong, E. E., Nylander, J., Thomas, G. W. C., Suh, A., Petersen, B., … Irestedt, M. (2019). Comparative analyses identify genomic features potentially involved in the evolution of birds-of-paradise. *Gigascience, 8*(5). https://doi.org/10.1093/gigascience/giz003

Pujolar, J. M., Dalen, L., Olsen, R. A., Hansen, M. M., & Madsen, J. (2017). First de novo whole genome sequencing and assembly of the pink-footed goose. *Genomics, 110*(2), 75–79. https://doi.org/10.1016/j.ygeno.2017.08.008

Quek, M. C., Chin, N. L., Yusof, Y. A., Law, C. L., & Tan, S. W. (2018). Characterization of edible bird's nest of different production, species and geographical origins using nutritional composition, physicochemical properties and antioxidant activities. *Food Research International, 109*, 35–43. https://doi.org/10.1016/j.foodres.2018.03.078

Quinlan, A. R. (2014). BEDTools: The Swiss-Army Tool for Genome Feature Analysis. Current Protocols in Bioinformatics, 47, 11 12 11-34. https://doi.org/10.1002/0471250953.bi1112s47.

Sanita Lima, M., & Smith, D. R. (2017). Pervasive Transcription of Mitochondrial, Plastid, and Nucleomorph Genomes across Diverse Plastid-Bearing Species. *Genome Biology and Evolution, 9*(10), 2650–2657. https://doi.org/10.1093/gbe/evx207

Seppey, M., Manni, M., & Zdobnov, E. M. (2019). BUSCO: Assessing Genome Assembly and Annotation Completeness. *Methods in Molecular Biology, 1962*, 227–245. https://doi.org/10.1007/978-1-4939-9173-0_14

Shangguan, G., Liang, X., Huang, G., Li, X., & Zeng, Q. (2018). Determination of amino acids in different cubilose by method of online pre-column OPA-FMOC derivatization HPLC. *Sci Technol Food Industry, 39*, 250–254.

Shim, E. K., Chandra, G. F., Pedireddy, S., & Lee, S. Y. (2016). Characterization of swiftlet edible bird nest, a mucin glycoprotein, and its adulterants by Raman microspectroscopy. *Journal of Food Science and Technology, 53*(9), 3602–3608. https://doi.org/10.1007/s13197-016-2344-3

Sivadasan, P., Gupta, M. K., Sathe, G. J., Balakrishnan, L., Palit, P., Gowda, H., … Sirdeshmukh, R. (2015). Human salivary proteome–a resource of potential biomarkers for oral cancer. *Journal of Proteomics, 127*(Pt A), 89–95. https://doi.org/10.1016/j.jprot.2015.05.039

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., … Mering, C. V. (2019). STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research, 47*(D1), D607–D613. https://doi.org/10.1093/nar/gky1131

Triantafyllou, A., Mikkelsen, L. H., Gnepp, D. R., Andreasen, S., Hunt, J. L., Devaney, K. O., … Ferlito, A. (2018). Salivary myoepithelial cells: An addendum. *Ultrastructural Pathology, 42*(6), 465–476. https://doi.org/10.1080/01913123.2018.1551259

Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M., & Jaffe, D. B. (2017). Direct determination of diploid genome sequences. *Genome Research, 27*(5), 757–767. https://doi.org/10.1101/gr.214874.116

Wong, R. S. (2013). Edible bird's nest: Food or medicine? *Chinese Journal of Integrative Medicine, 19*(9), 643–649. https://doi.org/10.1007/s11655-013-1563-y

Yan, X., Wu, Z., Wang, B., Yu, T., Hu, Y., Wang, S., … Zhang, X. (2020). Involvement of Cathepsins in Innate and Adaptive Immune Responses in Periodontitis. *Evid Based Complement Alternat Med, 2020*, 4517587. https://doi.org/10.1155/2020/4517587

Yeo, B. H., Tang, T. K., Wong, S. F., Tan, C. P., Wang, Y., Cheong, L. Z., & Lai, O. M. (2021). Potential Residual Contaminants in Edible Bird's Nest. *Front Pharmacol, 12*, Article 631136. https://doi.org/10.3389/fphar.2021.631136

Young, S. A., Miyata, H., Satouh, Y., Kato, H., Nozawa, K., Isotani, A., … Ikawa, M. (2015). CRISPR/Cas9-Mediated Rapid Generation of Multiple Mouse Lines Identified Ccdc63 as Essential for Spermiogenesis. *Int J Mol Sci, 16*(10), 24732–24750. https://doi.org/10.3390/ijms161024732

Zainal Abidin, F., Hui, C. K., Luan, N. S., Mohd Ramli, E. S., Hun, L. T., & Abd Ghafar, N. (2011). Effects of edible bird's nest (EBN) on cultured rabbit corneal keratocytes. *BMC Complementary and Alternative Medicine, 11*, 94. https://doi.org/10.1186/1472-6882-11-94

Zhao, R., Li, G., Kong, X. J., Huang, X. Y., Li, W., Zeng, Y. Y., & Lai, X. P. (2016). The improvement effects of edible bird's nest on proliferation and activation of B lymphocyte and its antagonistic effects on immunosuppression induced by cyclophosphamide. *Drug Des Devel Ther, 10*, 371–381. https://doi.org/10.2147/DDDT.S88193

Zur Lage, P., Newton, F. G., & Jarman, A. P. (2019). Survey of the Ciliary Motility Machinery of Drosophila Sperm and Ciliated Mechanosensory Neurons Reveals Unexpected Cell-Type Specific Variations: A Model for Motile Ciliopathies. *Frontiers in Genetics, 10*, 24. https://doi.org/10.3389/fgene.2019.00024