Title:

Pangolin scales as adaptations for innate immunity against pathogens

Authors:

Xuechen Tian[1,2,3#], Li Chen[4#], Jinfeng Zhou[5], Enbo Wang[1], Mu Wang[6], Nicholas Jakubovics[7], Jing Li[6], Kunping Song[1], King Tong Lau[6], Klaus-Peter Koepfli[8,9], Siyuan Zhang[5], Geok Yuan Annie Tan[4], Yixin Yang[1,2,3,10] and Siew Woh Choo[1,2,3,10*]

Affiliations:

[1]College of Science, Mathematics and Technology, Wenzhou-Kean University, Ouhai, Wenzhou, Zhejiang Province, 325060 China.

[2]Zhejiang Bioinformatics International Science and Technology Cooperation Center, Wenzhou-Kean University, Ouhai, Wenzhou, Zhejiang Province, 325060 China.

[3]Wenzhou Municipal Key Laboratory for Applied Biomedical and Biopharmaceutical Informatics, Wenzhou-Kean University, Ouhai, Wenzhou, Zhejiang Province, 325060 China.

[4]Institute of Biological Sciences, Faculty of Science, Universiti Malaya, 50603 Kuala Lumpur, Malaysia.

[5]China Biodiversity Conservation and Green Development Foundation, Empark International Apartment, No. 69, Banding Road, Haidian District, Beijing, China.

[6]Department of Biological Sciences, Xi'an Jiaotong-Liverpool University, Suzhou, China.

[7]School of Dental Sciences, Faculty of Medical Sciences, Newcastle University, Framlington Place, Newcastle upon Tyne, NE2 4BW, UK.

[8]Smithsonian-Mason School of Conservation, George Mason University, Front Royal, VA 22630, USA.

[9]Center for Species Survival, Smithsonian's National Zoo and Conservation Biology Institute, Washington, D.C., 20008, USA.

[10]Dorothy and George Hennings College of Science, Mathematics and Technology, Kean University, 1000 Morris Ave, Union, NJ 07083, USA.

 # The authors have contributed equally to this work

*** Corresponding author:**

Siew Woh Choo, College of Science, Mathematics and Technology, Wenzhou-Kean University,

32 Wenzhou, Zhejiang Province, China 325060. Email: cwoh@wku.edu.cn
33

Supplementary Methods

1.0 Animal ethics approval

This work was approved (reference number: GF (2019) BASE08) by the Biology and Science Ethics Committee under the China Biodiversity Conservation and Green Development Foundation (CBCGDF).

2.0 Samples

Manis javanica was the sole species analyzed in this study. All scales used in this study were provided by the China Biodiversity Conservation and Green Development Foundation (CBCGDF). Crude scales were sourced from different pangolins and collected in a sterile sample bag until utilized. The samples collected from the skin surface were obtained during the rescue operation of pangolins by CBCGDF in Guangzhou, China. The skin surface samples were collected using the Icleanhcy Genetic Testing Sample Collection Kit (Huachenyang Technology, China) according to the instructions of the manufacturer. The mouse hair was collected from a domestic mouse.

3.0 Scale preprocessing

Crude scales were pre-processed based on the method previously reported with slight modifications [47]. In particular, all scale samples were cleaned by rinsing and wiping their surfaces vigorously with ethanol, followed by washing with sterile water to remove contaminants from the scales further. The scales were submerged with or without pure ethanol for 12 hours and then washed with sterile water several times. Scales were pulverized using a sterile masher after being dried in a 60°C incubator, and powders were collected for subsequent experiments.

4.0 Exosome detection through TEM and NanoFCM technology

Exosomes were extracted using super-centrifugation according to the method previously reported [48-49]. After removing contamination from the surface of crude pangolin scales with ethanol and sterile water, scales were pulverized with nitrogen using a sterile masher. Scale powders were collected into a 50 ml centrifuge tube with an equivalent of 0.2% type I collagenase solution and digested for 1 hour at a 37°C incubator with gently shaking. Dulbecco's Modified Eagle Medium (DMEM) medium consisting of 10% FBS (exosome free) was used to terminate the digestion when the suspension was emulsion texture. The sample treated was centrifuged at 2,000xg, 4°C for 30 min, and the supernatant was collected and further centrifuged at 10,000xg, 4°C for 45 min to

remove more giant vesicles. Then, the supernatant was filtered by a 0.45μm filter and centrifuged at 100,000xg, 4°C for 70 min. After discarding the supernatant, 10 ml of precooling 1 X PBS (Sangon Biotech, Shanghai) was used to suspend the pellet and again centrifuged at 100,000xg, 4°C for 70 min. After removing the supernatant, exosomes were obtained and suspended with 100 μ L of precooling PBS. Exosomes were visualized using a transmission electron microscope (HT-700, Hitachi) and were further quantified by a Flow NanoAnalyzer (NanoFCM INC, China).

5.0 Procedure for SEM data collection

The pangolin scale samples were used without any (washing or chemical) pretreatment. Since the original crude pangolin scales were too big for viewing directly, they were sliced into smaller size samples with a sterile scalpel. The samples were typically of ca. 0.2 cm wide x 0.5 cm length with sufficient top area and cross-sectional area exposed for viewing. The cut samples were then put into a gold sputter machine (JEOL model JEC-3000FC) for 2 minutes for each side so that the whole sample was covered with a gold layer. The gold-sputtered samples were then ready for viewing. When viewing the top (or ulterior) of the scale, the sample was placed on the sample holder with the bottom of the scale placed on carbon film, hence exposing the top for observation. The same sample was then tilted with the cut side placed upwards, exposing the cross-section of the scale for observation. Different magnifications were used to visualize the detailed structure of the pangolin scale under the SEM microscope (SEM- JEOL model JSM-6510).

6.0 Scale metagenomics

6.1 DNA extraction

In this analysis, three crude scales (CS1, CS2, and CS3) from different pangolins were used. For each scale, the microbial community DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen, Valencia, USA) following the manufacturer's instructions. DNA was quantified with a Qubit Fluorometer by using the Qubit dsDNA BR Assay kit (Invitrogen, USA) and the quality was checked by running an aliquot on a 1% agarose gel.

6.2 Library construction and high-throughput sequencing

1μg genomic DNA was randomly fragmented using a Covaris ultrasonicator (Covaris, Woburn, USA). The fragmented DNA was selected by magnetic beads to an average size of 200-400bp. The

selected fragments were end-repaired, 3' adenylated, ligated with adapters, and then PCR amplified. PCR products were purified using magnetic beads. The double-stranded PCR products were heat-denatured and circularized using the splint oligo sequence. The single-strand circle DNA (ssCir DNA) was formatted as the final library and qualified by QC. The qualified libraries were sequenced on the MGISEQ-2000 platform (BGI-Shenzhen, China).

6.3 Metagenomic data analysis

The raw read data were trimmed using SOAPnuke v.1.5.2 [50]. The trimmed reads were mapped to the reference genome sequence of the host (Malayan pangolin; *Manis javanica*; Accession number: GCA_001685135.1) using the SOAP2 [51] software to identify and remove host-derived reads (Additional file B: Table S1). High-quality reads were *de novo* assembled using IDBA-UD software [52]. Assembled contigs with lengths less than 300 bp were discarded in the following analysis. Genes were predicted from contigs using MetaGeneMarker (2.10) [53]. Redundant genes were removed using CD-HIT [54] with an identity cutoff of 95% and sequence completeness of at least 90%. To generate the taxonomic information, the protein sequences of genes were aligned against the NR database using DIAMOND [55] with an e-value cutoff of $1e^{-5}$. Based on the MEGAN [56] LCA algorithm, the taxonomic annotation was assigned. To reconstruct the microbial genomes, for each sample we mapped reads onto the reference genome sequences of the two most abundant microbial species using Bowtie2 [57] and calculated genome completeness based on the mapped reads. All reference genome sequences were downloaded from NCBI (Additional file B: Table S1).

7.0 High-throughput mass spectrometric (MS) analysis

7.1 Samples

Three individual crude scales (CS4, CS5, and CS6) of Malayan pangolins were used for this analysis. For each scale, we had two replicates, one was from the proximal region of the scale (labeled as PS), and another was from its distal region (labeled as DS). In total, we had six experimental sets (CS4: PS1 and DS1; CS5: PS2 and DS2; CS6: PS3 and CS3) for mass spectrometric analysis. For skin samples, due to the difficulty in getting living healthy Malayan pangolins, which are *Critically Endangered* species, we only managed to collect skin surface samples from one living Malayan pangolin which was rescued by CBCGDF in China. Samples

were collected from three different sites of this individual: head skin, anal skin, and dorsal skin. In total, we had three experimental sets.

7.2 Trypsin digestion

All pangolin scale samples were ground into powders and dissolved in 8M urea. Samples were homogenized by sonication (QSONICA) on ice, and then centrifuged at 11,400 rpm at 4 °C for 20 min. Concentrations were determined using a Nanodrop 2000 spectrophotometer (Thermo Scientific). Samples were diluted to 1mg/ml with 50 mM ammonium bicarbonate, reduced with 10 mM dithiothreitol (DTT), and alkylated with 10 mM iodoacetamide (IAA) at room temperature for 30 minutes in dark. The samples were then digested by modified trypsin (Promega) at 37°C overnight and desalted using Monospin C18 columns (GL Sciences).

7.3 Nanoflow LC–MS/MS

An LTQ Orbitrap Elite mass spectrometer (Thermo-Fisher Scientific) equipped with a nanoelectrospray source (heated capillary temperature = 300°C) was used to acquire the MS and MS/MS data. Peptides were separated on a 15 cm analytical RSLC column (Acclaim™ PepMap™ 100 C18 HPLC Columns, 2 µm, 150 mm length, 0.05 mm I.D.). Each sample (10 µL) was auto-injected and separated using a 60-min gradient at a flow of 300 nL/min ranging from 100% buffer A (0.1% formic acid, 99.9% double-deionized water) and 0 % buffer B (99.9% acetonitrile and 0.1% formic acid) to 5% buffer B in 1 min, followed by a linear gradient from 5% buffer B to 25% buffer B in 45 min, ramped to 40 % buffer B in 5 min, then increased and stayed in 95% buffer B for 9 min. The LTQ Orbitrap Elite mass spectrometer was operated in a data dependent acquisition (DDA) mode, with 10 most intense ions from full scan for tandem mass spectrometry. The normalized collision energy was 35V and the default charge state was 2 in HCD mode. Scans were collected in the positive polarity mode.

7.4 Database searching

Tandem mass spectra were extracted using Mascot Distiller 2.7 (Matrix Science, London, UK; version 2.4.1) and searched against the in-house built pangolin protein database with a MS/MS mass tolerance of 0.05 Da and a precursor ion tolerance of 10 PPM. Scaffold version 4.9.0, (Proteome Software Inc., Portland, OR, USA) was used to validate MS/MS based peptide and

protein identifications. Peptide Probabilities from Mascot for all samples were assigned by the Peptide Prophet algorithm [58] with Scaffold delta-mass correction. The peptide identification False Discovery Rate (FDR) was set at 1%. For each sample, protein identifications were accepted if they contained at least two unique identified peptides and an FDR<1%. Protein probabilities were assigned using the Protein Prophet algorithm [59].

7.5 Identification of prominent proteins in scales and the skin surface of pangolins

In this study, we used stringent criteria to identify prominent proteins for downstream analyses. For scale proteins, we reported a prominent protein if a protein presented in at least five out of six experimental sets with a protein identification probability of at least 96%. For skin surface proteins, we defined a prominent protein as a protein that presented in at least two out of three experimental sets with a protein identification probability of at least 96%.

7.6 STRING network analysis

Network analysis was conducted to create scale protein-protein interaction (PPI) networks by using the STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) database version 11.0 with default settings [20]. Networks were clustered using the MCL clustering method provided under the STRING Clusters tab with an inflation parameter of 3. Biological process enrichment analysis was performed and results with multiple testing corrections were used for further analysis. An FDR threshold <0.05 was used to define significant processes, which were color-coded using the STRING analysis tool tab.

8.0 Scale metabolomics

8.1 Samples

The samples used for metabolomics analyses were the same samples that we used for proteomic analyses. Therefore, we had six experimental sets in total (CS4: PS1 and DS1; CS5: PS2 and DS2; CS6: PS3 and CS3).

8.2 Preparation of metabolic extracts and sample analysis

The extraction protocol for all samples was referenced to previous methods [60]. Briefly, the metabolic extracts from all samples were prepared using the extraction protocol as previously

described [60]. Chromatographic separation was performed on reversed-phase ACQUITY UPLC HSS T3 1.8 μm, 2.1×100 mm columns (Waters, Dublin, Ireland) using an ultra-performance LC system (Agilent 1290 Infinity II; Agilent Technologies). MS was performed using high-resolution mass spectrometry (5600 Triple TOF Plus, Applied Biosystems/MDS Sciex, Concord, ON, Canada) equipped with an ESI source. Data was acquired in positive and negative ion modes, respectively.

8.3 Identification of prominent metabolites in pangolin scales and its skin surface

Analyses were performed as previously described [61]. Briefly, MarkerView 1.3 (AB SCIEX) was used to extract peak areas of all the detected ions. The resulting two-dimensional matrices, including the observations (sample names) in columns, the variables (m/z-retention time pairs) in rows, and the peak areas, were imported into an R language program (self-compiling) for multivariate analysis. The following Isotopic internal standards were used to correct their corresponding metabolite MS data, including phenylalanine-D8, tryptophan-D8, isoleucine-D10, asparagine-13C4, methionine-D3, valine-D8, proline-D7, alanine-D4, glycine-D2, serine-D3, glutamate-D5, aspartate-D3, arginine-D7, glutamine-D5, lysine-D9, histidine-D5, taurine-D2. Other metabolite MS data were corrected using retention-time adjacent standards of the above isotopic references. For each sample, metabolite identification was compared with standard references and the open-source databases HMDB [62] and METLIN [63]. For differential metabolomics analysis, the peak area of each metabolite was compared using a two-tailed unpaired t-test and $P<0.05$ was considered significant. All statistical analyses were performed in R 4.0.2.

To identify prominent metabolites in pangolin scales and on the skin surface, we used stringent criteria. For scale metabolites, a prominent metabolite was defined as a metabolite that presented in all six experimental sets. For the metabolites on the skin surface, a metabolite was prominent and reported if it was presented in all three experimental sets.

9.0 Bacterial 16S gene PCR amplification

To investigate the presence of bacteria in pangolin scales, we used universal bacterial 16S gene primers, generating a product size of approximately 1,500 bps. All samples were subjected to a series of pre-treatments before DNA extraction. Since we studied the potential bacteria found inside pangolin scales, we used rigorous methods to remove any potential contamination on the top layer of scales. We first scraped off any strains on the top layer of the scale using a knife,

followed by washing several times with sterile water to further remove any contamination from the scales. All samples were then submerged in 70% ethanol for 12h during which tubes were inverted several times. After removing the ethanol by careful pipetting, all samples were immersed in 7% (v/v) bleach solution containing moderate Tween 20 for 4h and inverted for further sterilization. After that, all samples were washed at least 7 times with sterile water until no bubbles were observed. The pangolin scale samples were pulverized using a sterile masher after drying in a 60°C incubator, whereas the mouse hairs were also sterilized using the same method and were cut into small pieces with a sterile scissor for subsequent experiments. 25mg of scale powder/hair pieces were used to extract DNA with the DNeasy Blood & Tissue Kit (Qiagen, Valencia, USA) based on the manufacturer's specifications. All extracted DNAs were amplified using primers 27F (5'-AGAGTTTGATCMTGGCTCAG-3) 'and 1492R (5'-TACGGYTACCTTGTTACGACTT-3'). The primers were used for PCR using the following protocol.

PCR reaction:

ddH_2O	up to 30µL
Platinum Green PCR Master Mix (2X)	15µL
Forward primer(10µM)	3µL
Reverse primer(10µM)	3µL
Platinum GC Enhancer	6µL
DNA template	1µL

The PCR was performed as follows: 95ºC for 5 minutes for pre-denaturation; 40 cycles of 95ºC 30s for denaturation, 60ºC 30s for annealing; 72 ºC 60s for extension; 1 cycle of 72ºC 10min for re-extension; and 4ºC for storage. The PCR products were examined using agarose gel electrophoresis, which was set at 80 V voltage to run the 1% of agarose gel for 100min. All amplified DNA bands were visualized with the Bio-Rad imaging system (Bio-Rad Laboratories, Hercules, CA, USA).

10.0 Species identification of pangolin scales

To identify the pangolin species from which the scales used in this study were sourced, scales CS1, CS2, CS3, CS4, CS5, and CS6 were used. DNA extraction using the same extraction method as metagenomics analysis, we then amplified and sequenced the mitochondrial Cytochrome *b* (Cytb)

gene for phylogenetic analysis. Briefly, the DNA of pangolin scales was extracted using the DNeasy Blood & Tissue Kit (Qiagen, Valencia, USA) following the manufacturer's instructions. The extracted DNAs were amplified using primers: Cytb1 (5'-CCATCCAACATCTCAGCATGATGAAA-3') and Cytb2 (5'-GCCCCTCAGAATGATATTTGTCCTCA-3') [64]. The primers were used for PCR using the following protocol.

PCR reaction:

ddH_2O	up to 20μL
2 × Phanta Max Master Mix	10μL
Forward primer(10μM)	1μL
Reverse primer(10μM)	1μL
DNA template(10-50ng/μL)	1-2μL

The PCR was performed as follows: 95ºC for 2 minutes for pre-denaturation; 40 cycles of 95ºC 30s for denaturation, 49ºC 30s for annealing; 72 ºC 45s for extension; 1 cycle of 72ºC 3min for re-extension. The PCR products were purified by standard methods and directly sequenced with the same primers using an ABI 3730 DNA Sequencer (Thermofisher Scientific).

To perform phylogenetic inference, all Cytb sequences were initially aligned using MUSCLE [65]. The alignments were manually curated to ensure accuracy. Phylogenetic trees were constructed using MEGA-X [66]. The maximum-likelihood trees were inferred using the Tamura-Nei DNA substitution model and nodal support was estimated using 1,000 bootstrap replicates.

11.0 Antibacterial assays

All metabolites used to test antibacterial effects were bought from commercial companies. Bacterial cell lines used in this study were *Escherichia coli* ATCC 25922, *Staphylococcus aureus* ATCC 25923, *Pseudomonas aeruginosa* ATCC27853, and *Serratia marcescens* BNCC107931.

11.1 Preparation of bacterial suspensions

All bacterial cell lines used in this experiment were revived. Individual colonies were picked out on plates of the purified bacteria to be tested using an inoculating loop, inoculated in 2 mL of

lysozyme broth (LB, Qingdao Hope Bio-Technology Co., Ltd, China) and incubated in an incubator at 37°C with 200 rpm. The logarithmic growth phase of the bacterial broth was corrected to a concentration of 0.5 Mackay's turbidity standard (i.e., absorbance of 0.08-0.10 at 600 nm), containing approximately 1×10^8-2×10^8 CFU/mL.

11.2 Pre-screening of antibacterial metabolites

Antibacterial metabolites preparation and experimental design were based on the previously reported methods with minor modification [67]. The concentrations of the metabolites (Psaitong, Beijing, China) were set to C in the same proportions as in the pangolin scales, all in a total volume of 2 ml. The antibacterial effect of metabolites was determined by adding 200 ul of metabolites at concentration C to 1800 ul of LB broth containing an inoculum of 10^5 CFU/mL. The control contained only LB broth containing 10^5 CFU/mL of bacteria and was incubated at 37°C for 12h. The antibacterial effect of metabolites was measured by detecting the OD value at 600 nm by Varioskan Flash (Thermo Scientific, China).

11.3 Preparation of antibacterial metabolites

Metabolite (s) with concentrations of 4C, 2C, C, 0.5C, 0.25C, 0.125C, 0.0625C, 0.03125C were prepared by the two-fold dilution method. Eight sterile test tubes were taken and numbered sequentially, and 1ml of sterile water was added to 2-8 test tubes. 2 ml of 4C antibacterial metabolite(s) was prepared in the first test tube, 1 ml of the solution from the first test tube was added to the second test tube and mixed well, and then 1 ml of the solution from the second test tube was added to the third test tube and diluted in turn.

11.4 Antibacterial assays

Microdilution assays were used to investigate the MIC, MBC and bacterial growth curves. In a 96-well plate, 200 μL of LB broth medium was added to column 1 as a negative control, and 180 μL of prepared bacterial suspension was added to a final concentration of 1×10^5 CFU/mL in columns 2-10. 20 μL of sterile water was added to column 2 as a positive control, and 20 μL of drug concentrations were added to columns 3-10 in the order of 0.4, 0.2, 0.1, 0.05, 0.025, 0.0125, 0.00625 and 0.003125C. The plates were incubated in a 37°C incubator for 24 h. The MIC value was the lowest concentration in which no growth of the organism was observed. The growth curve

of the bacteria was measured by detecting the OD value at 600 nm. After determination of the MIC of the metabolite, 50 µl aliquots of all tubes with no visible bacterial growth were inoculated onto LB agar plates (Qingdao Hope Bio-Technology Co., Ltd, China) and incubated at 37°C for 24 h. When the agar plate was below 5 colonies at the lowest concentration of an antimicrobial agent, it was termed as an MBC endpoint. All experiments were repeated three times.

11.5 Bacterial morphology examination by Scanning Electron Microscopy (SEM)

Briefly, logarithmic growth stage *E. coli* and *S. aureus* cultures (~10^5 CFU/mL) were untreated (control) and treated with the 4-metabolite combination at its MIC value and incubated at 37°C for 24 h. Bacteria cells were collected after removing the medium and were washed twice with PBS (0.1 M). Bacterial cells were fixed in 2.5% glutaraldehyde and subsequently filtered and dehydrated in an ethanol solution. Dehydration was performed in a series of ethanol solutions (30, 50, 70, 80, 90 and 100%; each twice, for 15 min). Finally, all samples were air-dried at room temperature. SEM micrographs were acquired in a Field emission scanning electron microscope (Hitachi, Ltd., SU8010, Japan) with magnifications at 10,000X (left panel), 30,000X (middle panel), and 60,000X (right panel). All experiments were performed using three biological replicates per group.

References could be available in the main manuscript.