Defense Threat Reduction Agency
8725 John J. Kingman Road, MS
6201 Fort Belvoir, VA 22060-6201

DTRA-TR-16-18

TECHNICAL REPORT

A Metagenomic Approach to Unbiased Identification of Pathogens Endemic to Pakistan

October 2018

HDTRA1-10-1-0081

Eric T. Harvill et al.

Prepared by:
Pennsylvania State University
University Park, PA 16802

DESTRUCTION NOTICE:

Destroy this report when it is no longer needed.
Do not return to sender.

PLEASE NOTIFY THE DEFENSE THREAT REDUCTION AGENCY, ATTN: DTRIAC/ RD-NTF, 8725 JOHN J. KINGMAN ROAD, MS-6201, FT BELVOIR, VA 22060-6201, IF YOUR ADDRESS IS INCORRECT, IF YOU WISH IT DELETED FROM THE DISTRIBUTION LIST, OR IF THE ADDRESSEE IS NO LONGER EMPLOYED BY YOUR ORGANIZATION.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
00-10-2018	Technical	June 2010 - July 2015

4. TITLE AND SUBTITLE

A Metagenomic Approach to Unbiased Identification of Pathogens Endemic to Pakistan

5a. CONTRACT NUMBER

5b. GRANT NUMBER
HDTRA 1-10-1-0081

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Eric T. Harvill
Bodo Linz
Walter R. McWey, Jr.
Yury V. Ivanov
Bhushan M. Jayarao
Zubair Shabbir

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Pennsylvania State University, University Park, PA 16802, USA University of Veterinary and Animal Sciences, Syed Abdul Qadir Jillani (Out Fall) Road, Lahore, Pakistan

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Defense Threat Reduction Agency
8725 John J. Kingman Road, STOP 6201
Fort Belvoir, VA 22060-6201

10. SPONSOR/MONITOR'S ACRONYM(S)
DTRA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
DTRA-TR-16-18

12. DISTRIBUTION / AVAILABILITY STATEMENT

Distribution Statement A. Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Identification of new pathogens, which nearly always begin as zoonotic infections that spill over from animals, is a critical and rate-limiting step in the study of disease. 'Historically, discovery of new pathogens was itself limited by the slow advance of techniques for the culture of organisms, most of which remain unculturable. New bioinformatic tools, developed at Penn State, allowed to identify and study organisms associated with disease. This report focuses on identification and characterization of the diversity of zoonotic organisms, endemic to livestock, in Pakistan.

15. SUBJECT TERMS

microbiome, bacterial pathogens, infectious diseases, livestock

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	SAR	33	Heather Meeks
Unclassified	Unclassified	Unclassified			**19b. TELEPHONE NUMBER** *(include area code)* 571-616-3059

UNIT CONVERSION TABLE

U.S. customary units to and from international units of measurement*

U.S. Customary Units	Multiply by → / ← Divide by†		International Units
Length/Area/Volume			
inch (in)	2.54	$\times 10^{-2}$	meter (m)
foot (ft)	3.048	$\times 10^{-1}$	meter (m)
yard (yd)	9.144	$\times 10^{-1}$	meter (m)
mile (mi, international)	1.609 344	$\times 10^{3}$	meter (m)
mile (nmi, nautical, U.S.)	1.852	$\times 10^{3}$	meter (m)
barn (b)	1	$\times 10^{-28}$	square meter (m^2)
gallon (gal, U.S. liquid)	3.785 412	$\times 10^{-3}$	cubic meter (m^3)
cubic foot (ft^3)	2.831 685	$\times 10^{-2}$	cubic meter (m^3)
Mass/Density			
pound (lb)	4.535 924	$\times 10^{-1}$	kilogram (kg)
unified atomic mass unit (amu)	1.660 539	$\times 10^{-27}$	kilogram (kg)
pound-mass per cubic foot (lb ft^{-3})	1.601 846	$\times 10^{1}$	kilogram per cubic meter (kg m^{-3})
pound-force (lbf avoirdupois)	4.448 222		newton (N)
Energy/Work/Power			
electron volt (eV)	1.602 177	$\times 10^{-19}$	joule (J)
erg	1	$\times 10^{-7}$	joule (J)
kiloton (kt) (TNT equivalent)	4.184	$\times 10^{12}$	joule (J)
British thermal unit (Btu) (thermochemical)	1.054 350	$\times 10^{3}$	joule (J)
foot-pound-force (ft lbf)	1.355 818		joule (J)
calorie (cal) (thermochemical)	4.184		joule (J)
Pressure			
atmosphere (atm)	1.013 250	$\times 10^{5}$	pascal (Pa)
pound force per square inch (psi)	6.984 757	$\times 10^{3}$	pascal (Pa)
Temperature			
degree Fahrenheit (^oF)	$[T(^oF) - 32]/1.8$		degree Celsius (^oC)
degree Fahrenheit (^oF)	$[T(^oF) + 459.67]/1.8$		kelvin (K)
Radiation			
curie (Ci) [activity of radionuclides]	3.7	$\times 10^{10}$	per second (s^{-1}) [becquerel (Bq)]
roentgen (R) [air exposure]	2.579 760	$\times 10^{-4}$	coulomb per kilogram (C kg^{-1})
rad [absorbed dose]	1	$\times 10^{-2}$	joule per kilogram (J kg^{-1}) [gray (Gy)]
rem [equivalent and effective dose]	1	$\times 10^{-2}$	joule per kilogram (J kg^{-1}) [sievert (Sv)]

*Specific details regarding the implementation of SI units may be viewed at http://www.bipm.org/en/si/.

†Multiply the U.S. customary unit by the factor to get the international unit. Divide the international unit by the factor to get the U.S. customary unit.

Abstract

Precise and definitive microorganism identification and pathogen detection is essential for disease diagnosis and treatment of infection. However, bacterial identification methods primarily rely on the prior knowledge of the organism and have been using traditional phenotypic analyses that include morphological characterizations, such as the Gram staining, culturing techniques, chemical sensitivities, cell motility assays, growth on selective agar, and some other conventional tests. Only after advancements of Next-Generation Sequencing (NGS) technology, it became possible to do an unbiased identification of all bacteria present in a sample. This is particularly important because the majority of microorganisms inside our body, and that of other animals, cannot be cultured and therefore remain unknown or uncharacterized.

Pakistan is considered to be one of the hot-spot regions for the emergence of new zoonotic diseases. Within these hotspot regions, identification of new pathogens—which nearly always begin as zoonotic infections that spill over from animals—is a critical and rate-limiting step in the study of disease. We aimed to determine new zoonotic pathogens in livestock of animals, within Pakistan, using Next-Generation Sequencing technology. To increase efficiency of sample preparation and quality of samples, this project utilized and further strengthened a certified University Diagnsotic Laboratory (UDLL) at the University of Veterinary and Animal Sciences (UVAS), under the guidance of Penn State University team, suitable for Biosafety Level-2 pathogen handling. This laboratory of Pakistani scientists, most of which are veterinarians, was fully trained and employed to process animal samples for further sequencing.

Overall, this project was proven to be successful. The UDL and allied labs at UVAS has processed 1,013 samples (including both study inclusive and exclusive) and trained 246 young scientists and prepared them for future science-related jobs. This project assessed microbial communities of both healthy and diseased/deceased farm animals and, by performing comparisons between the two, identified etiological agents that caused the most severe disease outbreaks in different farms within

Pakistan. By confirming Koch's postulates, this project revealed for the first time the animal reservoir (sheep) for *Acinetobacter baumanii* infections.

Contents

Abstract 2

Acknowledgements 5

Introduction 7

Methods, Assumptions, and Procedures 9
- **Ethics Statement** 9
- **Respiratory lavage of birds** 9
- **Sample collection (nasal swab and broncho-alveolar lavage) from sheep** 10
- **Prokaryotic DNA isolation and purification** 10
- **Genome Sequencing** 11
- **Preparation of PCR Amplicons for 454-pyrosequencing** 11
- **Data processing and analysis** 12
  - MOTHUR 13
  - MEGAN 13
  - METAGENASSIST 13

Results and Discussion 14
- **A new, hypervirulent *Acinetobacter baumannii* strain causing high mortality in sheep** 14
- **Microbial communities present in the lower respiratory tract of clinically healthy birds in Pakistan** 17
- **Molecular Identification of Respiratory Microbiota of a Young Ostrich (*Struthio camelus*) Using Tracheo Broncho Alveolar Lavage** 18
- **Culture Independent Analysis of Respiratory Microbiome of Houbara Bustard (*Chlamydotis undulata*) Revealed Organisms of Public Health Significance** 19

Conclusions 20

References 21

Appendices 22
- **Appendix A: Sample code to run MOTHUR** 22
- **Apendix B: Broader impact** 27
- **Human resource development** 27
- **Strengthening of laboratory infrastructure** 28
- **Outreach** 29

List of Symbols, Abbreviations, and Acronyms 30

Acknowledgements

We thank Drs. Khushi Muhammad (Dean/Professor Faculty of Veterinary Sciences at UVAS), Tahir Yaqub (Professor, Department of Microbiology at UVAS), Abdul Rehman (Director, Poultry Resaerch Institute, Rawalpindi), Arfan Ahmad (Assistant Professor at UVAS), Muhammad Abbas (Senior Veterinary Officer, Okara district), Atta Subhani (District Livestock Officer, Faisalabad district), Muhammad Shafee (Assistant Professor, Baluchistan University, Quetta) for their necessary support and facilitation in sample collection from birds and animals; Drs. Muhammad Abu Bakr Shabbir (Lab. Manager K&Ns Poultry, Abbottabad) and Muhammad Husnain (Lab. Manager Big Bird Poultry, Lahore) for their support in data management, on-site sample collection and genome extraction procedures; Laboratory supervisors and technicians Mr. Muhammad Arshad, Muhammad Naeem Akhtar, Imran Shafique, Babar Nawab, Muhammad Asif and Muhammad Tariq for their help in sample processing in the laboratory. In particular we would like to thank Livetsock and Dairy Development Department of the Punjab province, field veterinarians and farm owners for their unconditional support and facilitation in sample collection from farms. Thanks to administration at UVAS for their necessary support in the smooth running of the project. We thank the Defense Threat Reduction Agency for the financial support.

Summary

Rapid and unbiased identification of new zoonotic pathogens that circulate in a livestock and spill over from those animals is critical in the study of human disease. Among several countries in South Asia, Pakistan is one of the "hotspots" of animal-transmitted diseases. Most conventional techniques and practices of microbial identification still rely on culturing bacteria in a laboratory setting. With advancements of Next-Generation Sequencing (NGS) techniques, unbiased culture-independent methods became available for identification of microbial and viral communities. Using NGS approaches, this study sought to identify the diversity of organisms endemic in both healthy and diseased animals, in Pakistan. During the 5-year period of this project, we founded a disease surveillance and diagnostic laboratories in Pakistan, trained dozens of young scientists—both Americans and Pakistani—in team-based infectious disease study that includes surveillance and sample collection, optimization of pipelines of nucleic acid purification and sequencing, and application of "Big Data" (high-throughput) Bioinformatics analyses for pathogen identification and characterization. This highly collaborative project established partner relations among the University of Veterinary & Animal Sciences (UVAS, Pakistan), Pennsylvania State University (USA), and various Disease Surveillance Departments within Pakistan. While DNA samples, collected from sick and healthy animals, were successfully sequenced and analyzed, it is not possible to know whether identified organisms are alive or dead, at a time of the sample collection. To make sure that only alive bacteria are sequenced, one should collect and process RNA samples instead. These samples, however, are labile and their integrity highly depends on proper handling. This is particularly challenging when bacterial (prokaryotic) RNA samples are isolated directly from an animal body fluid, which is *ab initio* enriched with lots of eukaryotic RNAs. Furthermore, these RNA samples are also problematic to ship to a different country and, therefore, should be prepared, purified, quantified for integrity, and sequenced, preferably all three steps together, on site.

Introduction

The evolution among micro-organisms is a constant and everlasting process since the emergence of life on earth. The process is continuing with the emergence and/or re-emergence of new species and strains, including organisms of zoonotic potential, and will remain till end of the day. A number of very advanced approaches are available to identify the potential etiological agent resulting in significant economic losses in terms of sufferings to animals and subsequent devastation of economics of dependent individuals; however all of these rely on extensive prior knowledge of the organism in terms of its isolation and identification from clinical sample. Such approaches are adequate to monitor KNOWN pathogens only within populations of animals in Pakistan, but do not have the ability to identify NOVEL pathogens. Even in some of the cases, the labs in Pakistan are neither well equipped nor the human resources are available to isolate and identify such pathogens. Furthermore, the great majority of organisms within the gut of the animals that are the source of zoonotic diseases, have not been cultured, remain unknown, and cannot be studied or monitored with the techniques used on known pathogens.

Isolation and identification of new pathogens, which nearly always begin as zoonotic infections that spill over from animals, is a critical and rate-limiting step in the study of disease. Historically, identification of new/novel pathogens was itself limited by the slow advance of techniques for the culture of organisms, most of which remain unculturable. However, with the current advancement in high throughput DNA sequencing machines, identification of novel pathogen has been made possible. Notably, identifying the unknown sequence read in a given clinical sample but closest to its known ancestors provide substantial information on many aspect of that particular organism with unknown sequence read including isolation and its characterization.

Here in this DTRA sponsored project, the main focus of collaborative research was to identify and determine the novel and diverse microbiome of present in the diseased and healthy animals together with development and further strengthening of laboratory infrastructure, human resource development and public awareness about

animal diseases and its prevention. With this project, the institute in Pakistan (UVAS) has been able to further strengthen the collaborative research between institutes of both countries and within research institutes of Pakistan.

Methods, Assumptions, and Procedures

Ethics Statement

The International Animal Care and Use Committee (IACUC), USA, and the Ethical Research Committee of University of Veterinary and Animal Sciences, Lahore, Pakistan, approved all the protocols, required for this study. The sample collection procedure was performed as aseptically as possible using all necessary personal protective equipment. Sample collections were performed by either veterinarians or by trained veterinarian technicians. Additional owner's consents have been received for sampling farm animals. The animals used for the study did not include endangered and/or protected species; no additional permissions were neither required nor needed.

Respiratory lavage of birds

Birds were euthanatized by intravenous injection of potassium chloride (Merck, Germany), KCl at 1-2 mEq per kg of body weight. Following procedures described in Fulton et al. (1990), avian respiratory lavage was taken from morbid and/or recently deceased as well as apparently healthy birds aseptically in a sterile falcon tubes, in a Biosafety Level-2 cabinet (BH-EN 2004-5 CE, Italy) at University Diagnostic Laboratory. Briefly, the birds were placed in dorsal recumbency and a ventral midline incision was made from the angle of the mandible to the thoracic inlet. The trachea was dissected freed of surrounding fascia and cut transversely at the mid-cervical region. A catheter (32 French, Tom Cat Catheter; Monoject, Division of Sherwood Medical, St. Louis, Missouri) was placed into the trachea to form an airtight seal. A four-way stopcock (Medex Inc., Hillard, Ohio) was attached to the catheter and a 35-mL plastic disposable syringe, containing 30 mL of phosphate-buffered saline (PBS; Oxoid, UK), and an empty 12-mL disposable syringe. A vacuum was formed by withdrawing air from the respiratory tract until there was a partial collapse of the extra-thoracic portion of the trachea. After this, the stopcock valve was turned to maintain the vacuum, and 5 to 10 mL of the PBS was instilled into the trachea. The body of the bird was rocked from side to side to allow the fluid to reach all parts of the respiratory tract. The fluid was then withdrawn slowly until there was a partial

collapse of the trachea. The fluid was measured and then filtered through sterile two-ply gauze to remove mucus clumps before collection into chilled glass tubes. Ten mL of PBS were instilled into and withdrawn from the trachea for two more lavages. This technique allows the lavage fluid to reach all parts of the respiratory tract, as was evident by wet lungs and air sacs during necropsy. Preliminary studies with toluidine blue added to PBS and lavage with a radiopaque dye followed by radiography also revealed that all parts of the respiratory tract came in contact with the lavage medium.

Sample collection (nasal swab and broncho-alveolar lavage) from sheep

Sampling was performed from 2011 to 2012 at a Government Livestock research farm in Khizarabad in the Sargodha district in Pakistan. Sampling was performed with all necessary personal protective equipment, following approved protocols. Samples were collected from 30 sheep, 15 from healthy and 15 from diseased animals. Clinically sick sheep showing the respiratory signs were selected and sampled. Nasal swabs and bronchioalveolar lavage were collected using sterile nasal swabs and locally modified aspirators. Similarly, nasal swabs and bronchioalveolar lavage from healthy animals were also collected aseptically. The sheep was anesthetized using Xylaz intravenously (2 mL). Using sterile Stomach Catheter, about 5-10 mL of sterile Phosphate Buffered Saline (PBS) was added, and the sheep was moved on each side several times to allow fluid to contact all parts of the lower respiratory system. Bronco-alveolar liquid (BAL, 3 - 5mL) was collected aseptically passing Stomach Tube (Curity(R), Tyco Healthcare, Thailand) through trachea to the lungs and aspirating with a 60 mL catheter tip disposable syringe (STAR, Jiangsu Kanghua, Medical Equipment Co. Ltd., China). The samples were properly labeled and were transported in cold chain to the laboratory, along with the history of treatment and disease outbreak at the farm.

Prokaryotic DNA isolation and purification

All described procedures were carried out in a Biohazard Safety Cabinet BSL-2 (BH-EN 2004-5 CE, Italy) at University Diagnostic Laboratory (UVAS, Lahore, Pakistan). Lavage was collected aseptically from trachea and lung samples. Bacterial genomic DNA was isolated using BiOstic® FFPE Tissue DNA Isolation Kit (Mobio,

USA), following the manufacturer's instructions. The extracted genome was quantified with NanoDrop (Thermo Scientific, USA) and genomic material (concentration of DNA ≥ 1 ng per μl) was sent to the Pennsylvania State University (University Park, PA, USA) for sequencing.

Genome Sequencing

During the 5-year period, sequencing methods rapidly evolved. Initial sequencing pipeline relied on 454-pyrosequening, while towards the end, the protocols were amended to adopt new, improved, sequencing platforms, such as Illumina MiSeq and HiSeq. Multiple sequencing runs were performed either at Genomics Core Facility (Penn State University, University Park, PA) or at Singapore Centre for Environmental Life Sciences Engineering (Singapore).

Preparation of PCR Amplicons for 454-pyrosequencing

One-way read amplicons (Lib-L) were prepared using bar-coded fusion primers with the universal 16S rRNA gene primers (27F, 5'-AGA GTT TGA TCM TGG CTC AG-3' and 907R, 5'-TAC GGG AGG CAG CAG-3') and adaptor sequences (forward: 5'-CCA TCT CAT CCC TGC GTG TCT CCG ACT CAG -MID- AGT TTG ATC MTG GCT CAG-3' and reverse: 5'-CCT ATC CCC TGT GTG CCT TGG CAG -TCTCAG- TAC GGG AGG CAG CAG-3'), as illustrated on **Figure 1**. The samples were denatured at 94°C for 3 minutes, followed by 35 cycles of 94°C for 15 sec, 55°C for 45 seconds, and 72°C for 60 seconds and a final extension at 72°C for 8min (Gene AMP PCR System 9700; Applied Biosystems, Foster City, CA). The PCR products (approximately 900 bp) were separated on agarose gels and were extracted, purified (Agencourt AMPure technology; Beckman Coulter, Brea, CA), and quantified (Qubit fluorometer; Lifetech, Carlsbad, CA and Biosystems library quantification kit; Kapa Biosystems, Woburn, MA). Pyrosequencing on a 454/Roche GS FLX+ instrument using titanium chemistry (Roche Diagnostics, Indianapolis, IN) was performed in accordance with the manufacturer's instructions at Penn State Genomic Core facility (State College, PA).

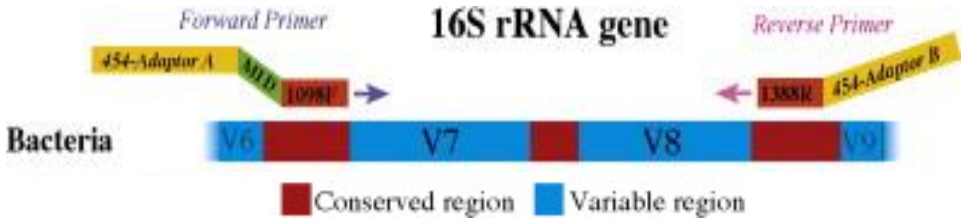


Figure 1. A schematic of a DNA library construction via PCR amplification of hypervariable regions of 16S rRNA gene. Universal sequences, attached to the MID and adapter, for forward and reverse primers, are shown in red.

Data processing and analysis

Next-Generation Sequencing data has been steadily increasing, in size, over the years. Data represents "short reads", or sequences, that are read by DNA sequencer. Each output file has to be processed and pass a quality check, to ensure that only good-quality amplicons are analyzed, and not random contaminations (if any). The schematic diagram of these steps is represented on **Figure 2**.

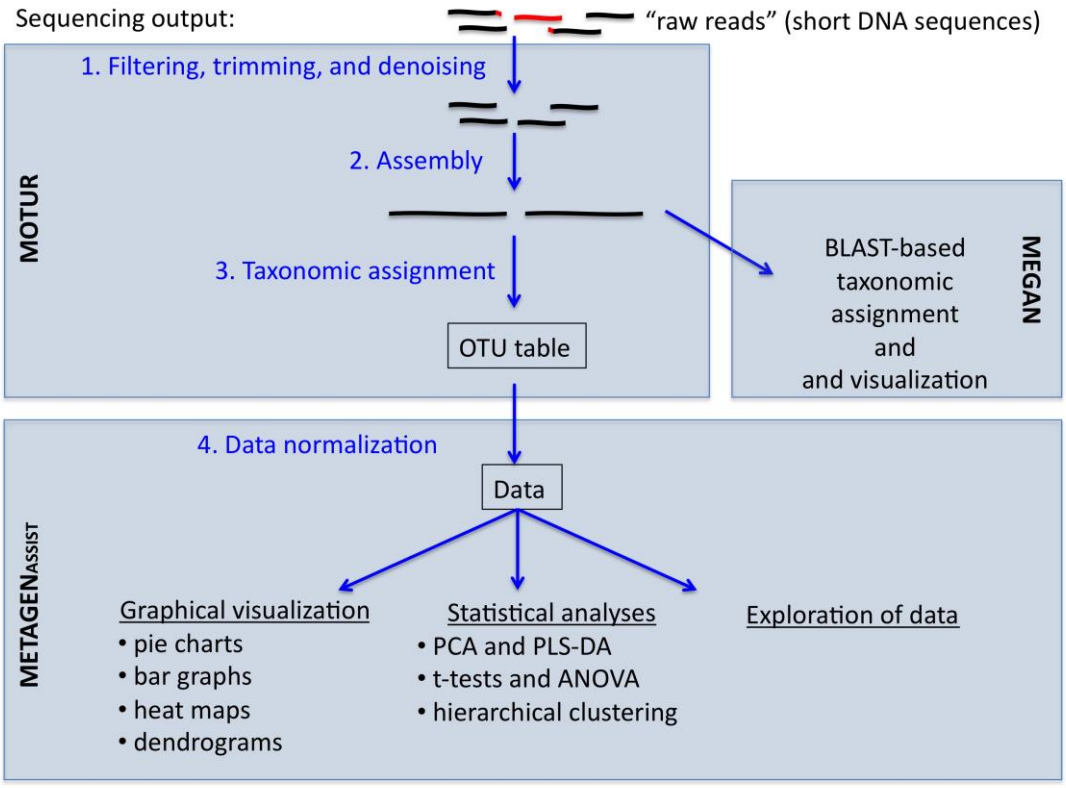


Figure 2. NGS data processing and analyses. This pipeline uses 3 different programs. MOTHUR processes sequencing reads, does QC, assembles reads into contigs and assigns the contigs to taxonomic units, i.e., produces an OTU table.

MEGAN is a taxonomic browser that takes an output file of the Basic Local Alignment Search Tool (BLAST) and visualizes results in the form of a phylogenetic tree. METAGENASSIST performs statistical analyses and visualizes the data.

MOTHUR

MOTHUR (Schloss et al., 2009) is publicly available at www.mothur.org. It is a command-line-based software that is capable of processing "Big NGS Data" on either a personal computer or a Linux server. We used MOTHUR to process output (.sff) files from 454/Roche GS FLX+ instrument or (.fastq) files from Illumina MiSeq or HiSeq sequencers. The code to process (.sff) file, via MOTHUR, is included in Appendix A. Several additional examples can be also found on MOTHUR wiki page at http://www.mothur.org/wiki/Analysis_examples.

MEGAN

MEGAN, or MEtaGenome ANalyzer (Huson, Auch, Qi, & Schuster, 2007), is publicly available at http://www.megan-db.org. In order to identify the most closely related type strain, MEGAN relates each read to sequences deposited to the GenBank, thereby defining the closest known relative. When matches are imperfect, the read is not identified as the same species, but as some higher taxonomic group (*e.g.*, genus or family). In this way, every read can identify the organism from which it came, *albeit* with variable precision, dependent on how closely related it is to a known, type strain present in GenBank.

METAGENASSIST

The METAGENASSIST (Arndt et al., 2012) is a publicly available at www.metagenassist.ca. METAGENASSIST takes OTU tables—produced by MOTHUR, for example—and performs normalization of the data, statistical analyses, and visualizations. The program requires a customly-made meta-file, which lists all the samples and all the information or groupings.

Results and Discussion

A new, hypervirulent *Acinetobacter baumannii* strain causing high mortality in sheep

An outbreak - symptoms and mortality. In 2011-2012, a severe outbreak of respiratory disease occurred in a flock of Kajli sheep in the Sargodha district of the Punjab province in Pakistan. During this outbreak, 415 out of 1,500 animals (28%) rapidly died, often within 3 days after the onset of symptoms. The sheep experienced nasal discharge, coughing and pyrexia (108°F, or 42.2°C); they were weak, emaciated with an increased frequency of breathing. Many infected animals showed signs of abdominal breathing and some animals had yellow-whitish diarrhea.

Post-mortem examination of the animals revealed anemic, swollen and congested lungs covered with white patches and with lesions on the anterior part of the lobes. In addition, the pericardial sacs were filled with a yellowish fluid. In some animals, parts of both the small and large intestine were congested, the kidneys were soft and fragile and the liver was hard and congested, sometimes blackish. Based on the observed symptoms, contagious caprine pleuropneumonia (CCPP) was diagnosed, a respiratory disease that is common across Africa, the Middle East and Western Asia. However, CCPP is usually caused by *Mycoplasma capricolum* and sometimes by *Mycoplasma mycoides*, bacteria that are usually susceptible to macrolides, tetracyclines and quinolones. Prescribed antibiotics, including tylosine (macrolide), rasomycin (oxytetracycline) and cereflox (quinol`one), failed to cure diseased animals. Other, subsequently administered antibiotics also proofed ineffective, including penicillin, the broad-spectrum antibiotic tribersin and the aminoglycosides streptomycin and gentamycin. This suggests a broad-spectrum antibiotic resistance of the causative agent.

Metagenomics revealed putative etiological agent of the disease. Our analysis of 16S rRNA gene sequences revealed the apparent difference in the bacterial composition of samples obtained from apparently healthy versus that of samples from diseased sheep. Samples from healthy animals contained 69 bacterial genera (Fig. 3A), including *Planococcus* (55% of the sequencing reads, class *Bacilli*), *Bacteroides* (13%, *Bacteroidetes*), *Pseudomonas* (6%, *Gammaproteobacteria*) and *Streptococcus*

(5%, *Bacilli*). In contrast, the three analyzed samples obtained from diseased animals were dominated by *Acinetobacter* (*Gammaproteobacteria*) with 42.6%, 48.8% (Fig. 3B) and 97.7% (Fig. 3C) of the reads. Both absence of *Acinetobacter* in healthy sheep (Fig. 3A) and high loads of *Acinetobacter* in diseased sheep suggest it to be the causing agent responsible for the disease outbreak.

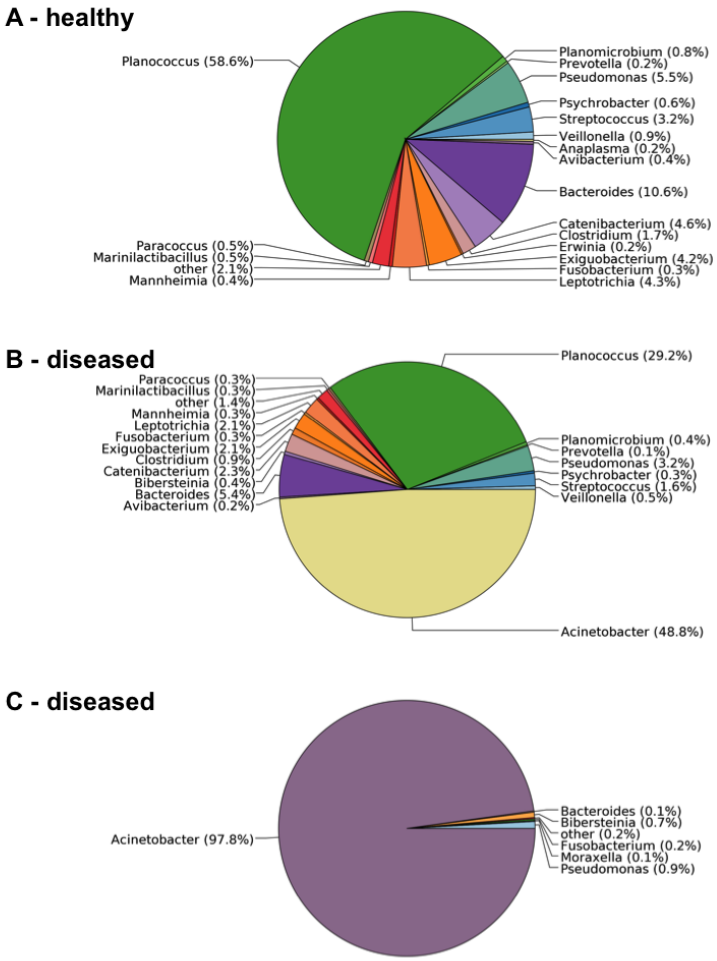
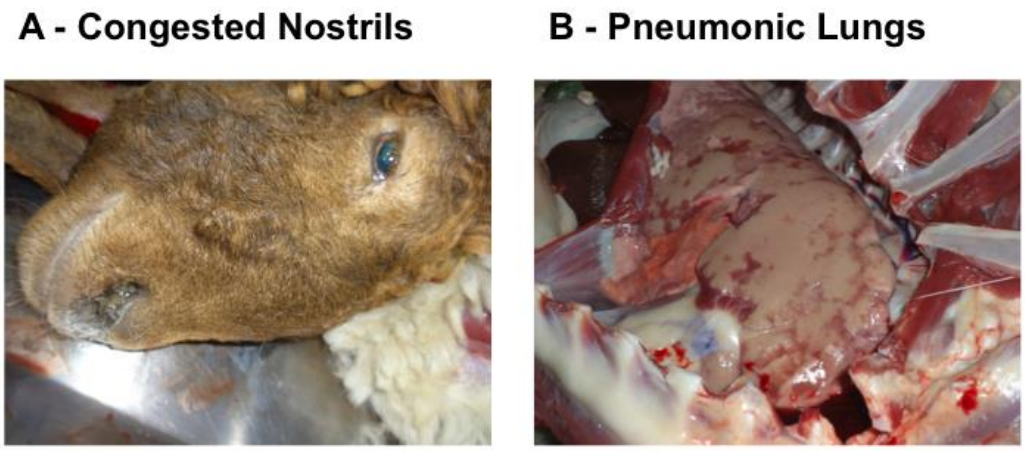


Figure 3. Distribution of bacterial 16S rRNA gene sequences in T-BAL samples obtained from diseased and healthy sheep. A) healthy sheep; B) diseased sheep; C) another diseased sheep.

Following up on the prediction made by the metagenomic test, bacteria was grown on solid media—using standard isolation procedures for *Acinetobacter spp.*—and overnight cultures derived from single colonies were preserved long term. The Gram-negative, non-motile isolate was oxidase negative and catalase positive and did

not ferment lactose, when grown on MacConkey agar, all of which are matching characteristics for the genus *Acinetobacter*. Subsequent PCR amplification, sequencing, and analysis of the 16S rRNA gene confirmed that this bacterium belongs to the genus *Acinetobacter*. More specifically, the sequence was 99% identical to that of *A. baumannii* species. The 99% identity of 16S rRNA gene sequences is considered reliable for the species-level identification.

Animal infection study – confirmation of the etiological agent. In compliance with the Ethical Research Committee of University of Veterinary and Animal Sciences (Lahore, Pakistan) protocol, we performed a controlled animal infection experiment, in Pakistan, to fulfill Koch's postulates and to confirm the etiological agent of the disease. Out of 4 female animals at 12-13 months of age each, 2 were kept as uninfected controls and the other two were inoculated intra-tracheally with 5×10^9 CFU of a liquid culture containing previously isolated *A. baumannii*. While the control sheep did not develop any symptoms, both of the inoculated sheep showed clinical symptoms within 24-48 hours post inoculation, including sneezing, coughing and abdominal breathing indicating severe respiratory problems. The infected animals were weak, showed abnormal behavior and had urination problems. The sheep were sacrificed on day-7 post inoculation. Post-mortem examination revealed congested nostrils, pneumonic lungs (Fig. 4), a fibrinous covering on the heart and a fragile liver.



Bacteria—re-isolated from lungs, liver, and kidneys of each animal—formed small, rounded colonies on Trypticase Soy Agar. This isolate was catalase positive, oxidase negative and had morphology of Gram-negative coccobacilli that exhibited a non-fermentor phenotype on MacConkey agar, all of which are matching characteristics for the genus *Acinetobacter*. PCR amplification, sequencing and analysis of the 16S rRNA gene showed a 100% identity of this organism to the that used for the inoculum, confirming that *A. baumannii* was, in fact, the etiological agent of the disease outbreak in sheep.

***A. baumanii* genome sequence and MLST**. We have sequenced a genome of the *A. baumanii* isolate that we used to inoculate healthy sheep. This genome was sequenced as 300×300-overlapping DNA libraries on Illumina MiSeq. We are currently analyzing and improving this draft genome assembly. Meanwhile, Multi Locus Sequence Typing (MLST) revealed a novel type, related to the sequence type 2 (or ST2), with the following allelic profile: *fusA*-2, *pyrG*-2, *rplB*-2, *rpoB*-2, *cpn60*-2, *gltA*-2 and a new *recA* allele that differed from *recA*-2 by a single nucleotide. This MLST suggests that this *A. baumanii* isolate is new and indeed corresponds to this species.

Microbial communities present in the lower respiratory tract of clinically healthy birds in Pakistan

This study was completed as a part of graduate research of one of faculty members at University of Veterinary and Animal Sciences, Lahore Pakistan. With the understanding that commercial poultry is an important agricultural industry worldwide particularly in Pakistan and that, besides increase in meat and egg production, dense living conditions and large flocks also increase the risk of disease outbreaks and zoonoses. Here in this study, we did collect tracheobronchoalveolar lavage of 14 birds raised at 3 different farms representing three different management system of rearing in the Punjab province of Pakistan. To characterize the lower respiratory microbiome of these birds, we sequenced hyper-variable regions of the 16S ribosomal subunit gene. Although dominated by bacteria belonging to a small number of taxonomic classifications, the lower respiratory microbiome from each farm was far more diverse and novel than previously known. The differences in

microbiome among farms suggest that inter-farm differences affect the microbiome of birds more than breed, geographic location, or management system. The presence of potential and known pathogens in genetically similar specialty breeds of chickens kept at unnaturally high densities and under variable conditions presents an extraordinary opportunity for the selection of highly pathogenic bacteria. In some instances, opportunistic respiratory pathogens were observed in apparently healthy birds. Understanding and monitoring the respiratory microbiome of such populations may allow the early detection of future disease threats.

The said research work has been published in Poultry Science journal in 2015 and can be accessed through the given link (http://www.ncbi.nlm.nih.gov/pubmed/25667427, Shabbir et al., 2015).

Molecular Identification of Respiratory Microbiota of a Young Ostrich (*Struthio camelus*) Using Tracheo Broncho Alveolar Lavage

This sub-project assessed microbial communities present in respiratory airway of ostriches from a farm located in Punjab province of Pakistan (30°48’29”N 73°36’00”E). It this first assessment study, we determined microbiomes of 37 out of 90 total ostriches housed in that farm, using 16S-rRNA-gene–based approach with 454-pyrosequencing. The microbiome of the deceased ostrich revealed abundance of Gram-negative bacterial species. Comparative taxonomic searches revealed that several identified sequences were similar (but not exactly the same) to known human pathogens from different genera: *Sphingobacterium*, *Shewanella*, *Myroides*, *Flavobacterium*, *Chrysobacterium*, *Aeromonas*, and even *Acinetobacter* species. While 16S rRNA gene sequences provide reliable assessment of microbial communities down to genus-level taxonomic assignment, it is almost impossible, with a rare exception, to identify the exact species. Nonetheless, this approach can be used as a first step towards pathogen identification and should be followed up with microbiome-guided culturing of particular organisms of interest. This study was published in the *Journal of Animal & Plant Sciences* (Shabbir Z. et al., 2013).

Culture Independent Analysis of Respiratory Microbiome of Houbara Bustard (*Chlamydotis undulata*) Revealed Organisms of Public Health Significance

In a similar study, we assessed respiratory microbiome of a houbara bustard, a migratory bird capable of a potential spread of pathogens over wide geographic areas. This endangered bird—which suffered from an unidentified disease—was brought dead into the Quality Operations Laboratory for necropsy examination. The necropsy revealed hemorrhages on tracheal mucosa but no apparent gross lesions on the lungs, suggesting an upper respiratory disease. Collected tracheoalveolar lavage was processed for sequencing. Total microbial community present was dominated by *Myroides* spp. MY15, suggesting that these species might have been involved in the respiratory disease and/or the cause of death. Other microbes present included some pathogens of interest: *Anaplasma phygocytophilum*, *Borrelia burgdorferi*, *Pasturella multocida*, *Campylobacteri jejuni*, *Clostridium botulinum*, *Pseudomonas aeroginosa*, and *Mycobacterium avium*. This study was published in the *International Journal of Agriculture & Biology* (Shabbir M., et al., 2014).

Conclusions

This first step toward understanding the uncultivable component of the lower respiratory microbiome of clinically diseased and healthy animals demonstrates the feasibility and value of a broader effort. Although we observed the animal microbiome to be dominated by bacteria belonging to a small number of taxonomic classifications, a far more diverse consortium of microbiome was identified than previously known, including bacteria of significance to both animal and human health. Many known bacterial taxa in our samples are not previously associated with the lower respiratory tracts of animals, including several known pathogens. We also identify multiple novel organisms and observe differences in taxa between sampled farms and animals. Based upon general conclusions drawn that are interested and compelling, the study thus, opens the door to advances in culture-independent identification of microbiome.

The said, DTRA sponsored research work has provided a baseline laboratory infrastructure, human resource development, encouragement to do collaborative research within and between the institute at national and international level, opened-up a diverse research avenues for indigenous researchers in Pakistan that were otherwise limited to conventional and routine research methodologies with no advancement to existing knowledge of disease and pathogens in Pakistan.

References

Arndt, D., Xia, J., Liu, Y., Zhou, Y., Guo, A. C., Cruz, J. A., et al. (2012). METAGENassist: A comprehensive web server for comparative metagenomics. *Nucleic Acids Research, 40*(Web Server issue), W88-95. doi:10.1093/nar/gks497

Huson, D. H., Auch, A. F., Qi, J., & Schuster, S. C. (2007). MEGAN analysis of metagenomic data. *Genome Research, 17*(3), 377-386. doi:gr.5969107 [pii]

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology, 75*(23), 7537-7541. doi:10.1128/AEM.01541-09

Appendices

Appendix A: Sample code to run MOTHUR

This code is written for Mac (UNIX) shell. To run Windows code, refer to MOTHUR user manual or MOTHUR wiki page.

1. Extract .fasta, .qual, and .flow files from H2NRBCL01.sff input file.

```
mothur > sffinfo(sff=./H2NRBCL01.sff, flow=T)
```

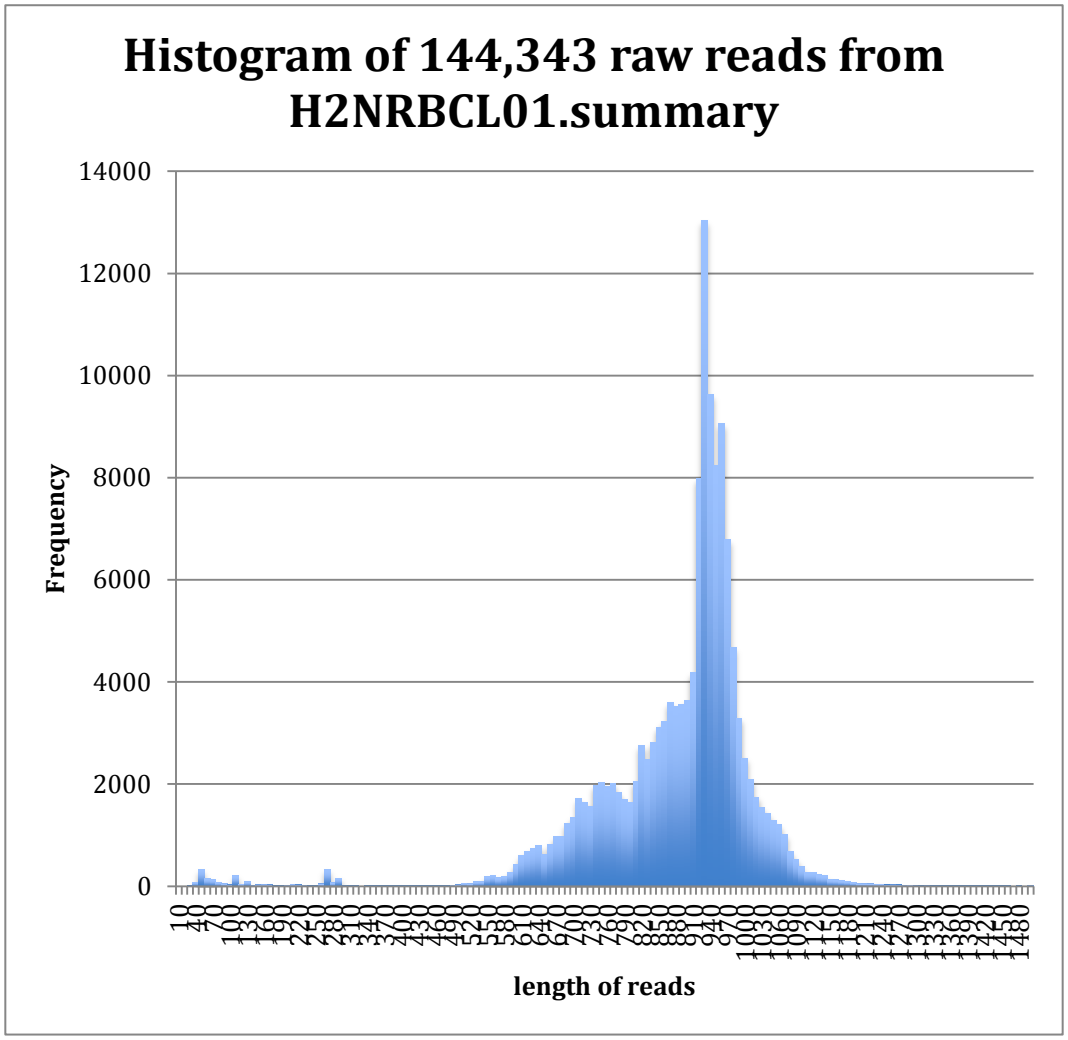
2. Generate a summary.

```
mothur > summary.seqs(fasta=./H2NRBCL01.fasta)
```

3. Save the summary as a text file.

```
mothur > system(cp ./H2NRBCL01.summary ./H2NRBCL01.txt)
```

4. Open .txt in Excel and create a histogram showing distribution of read lengths:



5. Reduce sequencing errors with trim.flows command. Supply a .oligos file. For faster parallel processing, indicate the number of processors that can be used (processors=X, where X is the number of CPU available).

mothur > trim.flows(flow=./H2NRBCL01.flow, oligos=./H2NRBCL01.oligos, bdiffs=1, minflows=450, processors=X)

6. Denoise with PyroNoise algorithm.

mothur > shhh.flows(file=./H2NRBCL01.flow.files, processors=X)

7. Remove barcodes and primer sequences (adaptors) from sequencing reads. Filter out sequences with length < 200 bp and with > 8-bp homopolymers.

mothur > trim.seqs(fasta=./H2NRBCL01.shhh.fasta, name=./H2NRBCL01.shhh.names, oligos=./H2NRBCL01.oligos, bdiffs=1, maxhomop=8, minlength=200, processors=X)

8. Generate a summary.

mothur > summary.seqs(fasta=./H2NRBCL01.shhh.trim.fasta, name=./H2NRBCL01.shhh.trim.names)

9. Simplify the data to reduce CPU and RAM use, by working with unique sequences.

mothur > unique.seqs(fasta=./H2NRBCL01.shhh.trim.fasta, name=./H2NRBCL01.shhh.trim.names)

10. Generate a summary.

mothur > summary.seqs(fasta=./H2NRBCL01.shhh.trim.unique.fasta, name=./H2NRBCL01.shhh.trim.unique.names)

11. Align the data against the SILVA-compatible reference alignment database of 16S rRNA sequences. Alternatively, use greengeens reference alignment.

mothur > align.seqs(reference=./silva.bacteria/silva.bacteria.fasta, processors=X)

12. Generate a summary of the alignment.

mothur > summary.seqs(fasta=./H2NRBCL01.shhh.trim.unique.align, name=./H2NRBCL01.shhh.trim.unique.names)

13. Adjust for overlap in the alignment.

mothur > screen.seqs(fasta./H2NRBCL01.shhh.trim.unique.align, name=./H2NRBCL01.shhh.trim.unique.names, group=./H2NRBCL01.shhh.groups, start=1044, optimize=end, criteria=95, processors=X)

14. Generate a summary of the overlapping.

mothur > summary.seqs(fasta=./H2NRBCL01.shhh.trim.unique.good.align, name=./H2NRBCL01.shhh.trim.unique.good.names)

15. Filter this alignment file.

mothur > filter.seqs(fasta=./H2NRBCL01.shhh.trim.unique.good.align, vertical=T, trump=., processors=X)

16. Simplify the dataset.

mothur > unique.seqs(fasta./H2NRBCL01.shhh.trim.unique.good.filter.fasta, name=./H2NRBCL01.shhh.trim.unique.good.names)

17. Merge sequence counts of samples that are within 2 bp of a more abundant sequence.

mothur > pre.cluster(fasta=./H2NRBCL01.shhh.trim.unique.good.filter.unique.fasta, name=./H2NRBCL01.shhh.trim.unique.good.filter.names, group=./H2NRBCL01.shhh.good.groups, diffs=2)

18. Generate a summary of pre-clustering.

mothur > summary.seqs(fasta=./H2NRBCL01.shhh.trim.unique.good.filter.unique.precluster.fasta, name=./H2NRBCL01.shhh.trim.unique.good.filter.unique.precluster.names)

19. Perform *de novo* removal of chimeras using chimera.uchime command.

mothur > chimera.uchime(fasta=./H2NRBCL01.shhh.trim.unique.good.filter.unique.precluster.fasta, name=./H2NRBCL01.shhh.trim.unique.good.filter.unique.precluster.names, group=./H2NRBCL01.shhh.good.groups, processors=X)

mothur > remove.seqs(accnos=./H2NRBCL01.shhh.trim.unique.good.filter.unique.precluster.uchime.accnos, fasta=./H2NRBCL01.shhh.trim.unique.good.filter.unique.precluster.fasta, name=./H2NRBCL01.shhh.trim.unique.good.filter.unique.precluster.names, group=./H2NRBCL01.shhh.good.groups, dups=T)

20. Generate a summary.

mothur > summary.seqs(name=current)

21. Remove additional eukaryotic contaminants (if any). Supply RDP training set (download from www.mothur.org)

mothur >
classify.seqs(fasta=./H2NRBCL01.shhh.trim.unique.good.filter.unique.precluster.pick.fasta,
name=./H2NRBCL01.shhh.trim.unique.good.filter.unique.precluster.pick.names, group=./H2NRBCL01.shhh.good.pick.groups,
template=./rdp_v9.pds/trainset9_032012.pds.fasta,
taxonomy=./rdp_v9.pds/trainset9_032012.pds.tax, cutoff=80, processors=X)

22. Remove sequences matching chloroplastsm mitochondria, or unknowns (if any).

mothur >
remove.lineage(fasta=./H2NRBCL01.shhh.trim.unique.good.filter.unique.precluster.pick.fasta,
name=./H2NRBCL01.shhh.trim.unique.good.filter.unique.precluster.pick.names, group=./H2NRBCL01.shhh.good.pick.groups,
taxonomy=./H2NRBCL01.shhh.trim.unique.good.filter.unique.precluster.pick.pds.wang.taxonomy, taxon=Mitochondria-Chloroplast-Archaea-Eukaryota-unknown)

23. Generate a summary.

mothur > summary.seqs(name=current)

24. Rename some files.

mothur > system(cp
./H2NRBCL01.shhh.trim.unique.good.filter.unique.precluster.pick.pick.fasta
./H2NRBCL01.final.fasta)

mothur > system(cp ./H2NRBCL01.shhh.good.pick.pick.groups
./H2NRBCL01.final.groups)

mothur > system(cp
./H2NRBCL01.shhh.trim.unique.good.filter.unique.precluster.pick.pick.names
./H2NRBCL01.final.names)

mothur > system(cp
./H2NRBCL01.shhh.trim.unique.good.filter.unique.precluster.pick.pds.wang.pick.taxonomy ./H2NRBCL01.final.taxonomy)

25. Cluster the data and prepare inputs for analysis.

mothur > dist.seqs(fasta=./H2NRBCL01.final.fasta, cutoff=0.15, processors=X)

mothur > cluster(column=./H2NRBCL01.final.dist, name=./H2NRBCL01.final.names)

mothur > make.shared(list=./H2NRBCL01.final.an.list, group=./H2NRBCL01.final.groups, label=0.03)

mothur > count.groups()

26. Subsample all the samples by the one with a fewest number of sequences. X is number of sequences to subsample to.

mothur > sub.sample(shared=./H2NRBCL01.final.an.shared, size=X)

27. Get the taxonomy information for each of the OTU.

mothur > classify.otu(list=./H2NRBCL01.final.an.list, name=./H2NRBCL01.final.names, taxonomy=./H2NRBCL01.final.taxonomy, label=0.03)

28. Assign sequences to phylotypes. Subsample to size=X, where X is a number.

mothur > phylotype(taxonomy=./H2NRBCL01.final.taxonomy, name=./H2NRBCL01.final.names, label=1)

mothur > make.shared(list=./H2NRBCL01.final.tx.list, group=./H2NRBCL01.final.groups, label=1)

mothur > sub.sample(shared=./H2NRBCL01.final.tx.shared, size=X)

29. Get the taxonomy of each phylotype.

mothur > classify.otu(list=./H2NRBCL01.final.tx.list, name=./H2NRBCL01.final.names, taxonomy=./H2NRBCL01.final.taxonomy, label=1)

mothur > dist.seqs(fasta=./H2NRBCL01.final.fasta, output=phylip, processors=X)

30. Make phylip-formatted distance matrics and make Neighbor-Joining phylogenetic tree.

mothur > clearcut(phylip=./H2NRBCL01.final.phylip.dist)

31. Estimate Chao1 richness with inverse Simpson diversity index

mothur > collect.single(shared=./H2NRBCL01.final.an.shared, calc=chao-invsimpson, freq=100)

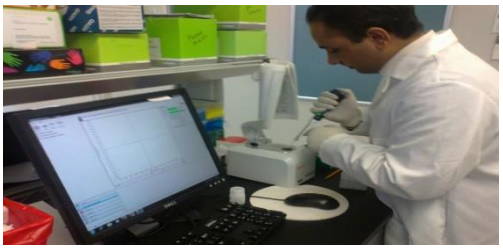
32. Generate rarefraction curve

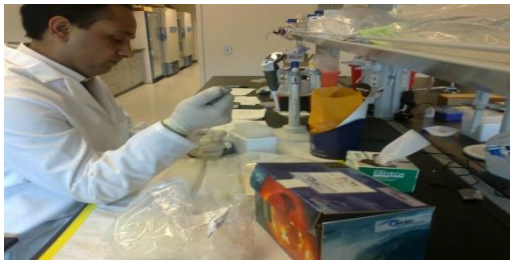
mothur > rarefaction.single(shared=./H2NRBCL01.final.an.shared, calc=sobs, freq=100)

Apendix B: Broader impact

Human resource development

As this area of advanced bioinformatics was relatively novel for indigenous researchers in Pakistan, a comprehensive training plan was executed where well-motivated faculty, post-graduate researchers were selected to participate on a short training courses at The Pennsylvania State University, USA. Two trainings were organized at Penn State University, one in July-August 2011 and the other one in January 2014. During 2011, five select candidates were given hand-on training on sample processing procedures coupled with bioinformatics analysis. Of these five, three were from University of Veterinary and Animal Sciences (Drs. Masood Rabbani, Tahir Yaqub, Muhammad Zubair Shabbir) and one each were from Veterinary Research Institute, Lahore (Dr. Mushtaq Ahmed) and Dr. Tanvir Ahsan (Aga Khan University, Karachi). For training in January 2014, two PhD students (Ms Nadia Mukhtar and Hassan bin Aslam) from Department of Microbiology at UVAS visited Pennstate and learned latest bioinformatics tools while processing genome-sequencing data recovered from the project derived samples. Added to this, one of faculty member (Muhammad Zubair Shabbir) and postgraduate student (Muhammad Husnain) did complete his PhD and M. Phil thesis (2013) respectively. The PhD thesis was entitled “A metagenomic analysis of respiratory microbiota of birds” while M. Phil thesis was entitled as “Culture-independent analysis of respiratory microbiome of clinically healthy and diseased broiler





breeders". Both of thesis involved sample processing related things in Pakistan while data analysis part was completed at Penn State University, USA.

Upon the return to Pakistan, trained individuals mentored faculty, research personnel, and students in number of training courses/workshops organized at UVAS. Participant represented renowned institutes and various areas in Pakistan. Throughout the project period (2010 – 2015), a total of 13 training workshops were organized where 246 participants were trained on sample collection process/procedures, genome extraction procedure and data analysis through various bioinformatics tools.

Strengthening of laboratory infrastructure

The existing labs in three different departments of UVAS were upgraded with state-of the art equipment that could be used for genome extraction, its measurement, storage and molecular research related work. Namely, University Diagnostics Laboratory, Quality Operations Laboratory, and Department of Microbiology. These facilities were not only used by the individuals involved in the project but is also being used by the students enrolled in the department.

Outreach

Beside project activities, the research group at UVAS organized extension activities. These were conducted to make farmers well aware of animal diseases, pathogens and their potential of being zoonotic. Such activities were conducted each a year at places where farmers are illiterate and are not well known to animal pathogens. They were guided to the best how they can handle their animal in the vent of disease outbreak and what should they do for disease diagnosis, its treatment and subsequent control.

As women play an important role in routine socio-economic set up in rural areas in Pakistan, their participation was encouraged in particular.

List of Symbols, Abbreviations, and Acronyms

T-BAL – tracheo-bronchoalveolar lavage.

OTU – operational taxonomic unit.

NGS – next-generation sequencing.

PCR – polymerase chain reaction.

primer – oligonucleotide DNA (for PCR).

CCPP – contagious caprine pleuropneumonia.

16S rRNA – The subunit 16S of a ribosomal RNA.

CFU – colony-forming units (bacterial numbers).

MLST – Multi Locus Sequence Typing.

UVAS – University of Veterinary and Animal Sciences, Lahore, Pakistan.

QOL – Quality Operations Laboratory, Pakistan.