

**THE ANIMAL-HUMAN INTERFACE AND INFECTIOUS
VIRAL DISEASES: GENOMICS AND MOLECULAR
EPIDEMIOLOGICAL APPROACHES IN “ONE HEALTH”
CONTEXT**

Ph.D. Thesis

By
BASANTA PRAVAS SAHU



**DEPARTMENT OF BIOSCIENCE AND BIOMEDICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY INDORE
FEBRUARY, 2021**

**THE ANIMAL-HUMAN INTERFACE AND INFECTIOUS
VIRAL DISEASES: GENOMICS AND MOLECULAR
EPIDEMIOLOGICAL APPROACHES IN “ONE HEALTH”
CONTEXT**

A THESIS

*Submitted in partial fulfilment of the
requirements for the award of the degree
of*
DOCTOR OF PHILOSOPHY

by
BASANTA PRAVAS SAHU



**DEPARTMENT OF BIOSCIENCE AND BIOMEDICAL
ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY INDORE
FEBRUARY, 2021**



INDIAN INSTITUTE OF TECHNOLOGY INDORE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled **THE ANIMAL-HUMAN INTERFACE AND INFECTIOUS VIRAL DISEASES: GENOMICS AND MOLECULAR EPIDEMIOLOGICAL APPROACHES IN "ONE HEALTH" CONTEXT** in the partial fulfilment of the requirements for the award of the degree of **DOCTOR OF PHILOSOPHY** and submitted in the **DEPARTMENT OF BIOSCIENCES AND BIOMEDICAL ENGINEERING, Indian Institute of Technology Indore**, is an authentic record of my own work carried out during the time period from December 2015 to December 2020 under the supervision of Dr. Debasis Nayak, Associate Professor, Department of Biosciences and Biomedical Engineering.

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other institute.

Basanta Pravas Sahu
4.02.2021
(BASANTA PRAVAS SAHU)

This is to certify that the above statement made by the candidate is correct to the best of my/our knowledge.

Debasis Nayak
04.02.2021

(DR. DEBASIS NAYAK)
Signature of Thesis Supervisor

Mr. BASANTA PRAVAS SAHU has successfully given his Ph.D. oral Examination held on **12.07.2021**.

Debasis Nayak
12.07.2021

(Dr. DEBASIS NAYAK)
Signature of Thesis Supervisor
Date: 12.07.2021

ACKNOWLEDGEMENTS

This work carries out for the fulfillment of the Doctoral degree is a result of teamwork. This is an opportunity for me to convey my sincere gratitude to all of them for being with me through this journey and help me to complete the Ph.D. research successfully.

Firstly, I appreciate *God* to provide this life and strength to overcome all the rough conditions during this tenure.

This piece of work was not possible without the guidance of my thesis supervisor. I am lucky enough to get *Dr. Debasis Nayak* as my thesis guide. I convey my deep gratitude to him for bracing me in my professional as well as personal life. His worthwhile suggestions and recommendation in the field of molecular and classical virology, helped me to carry out my thesis work.

I am privileged to *Dr. Debasis Nayak* (Associate Professor, Department of Biosciences and Biomedical Engineering, IIT Indore) for giving me a chance to work under his supervision for the Ph.D. degree. His expertise, enthusiasm, and constant encouragement were always there for us. Under his guidance, I got an opportunity to work with viruses, which was always my aspiration being a microbiologist. The rigorous scientific discussion, opening our heart-out, the parties at his home, the delicious dishes cooked by Sir, everything will be etched as lifetime memories. The greatest learning from my association with him was simplicity, dedication, and happiness that will be the key mantra for the rest of my life.

I want to extend my appreciation to my PSPC members, *Dr. Sharad Gupta* and *Dr. Anjan Chakraborty*, for their valuable suggestions and support during the Ph.D. period. Their evaluation of my thesis work helped me rectify the mistakes and improve for a better scientific approach towards the problems.

My sincere thanks also go to *Dr. Prashant Kodgire* (DPGC, BSBE), for his valuable suggestions and constant encouragement. I am grateful to all the faculties of BSBE for maintaining a nice atmosphere in the department and always encouraging us to do better. I would like to express my sincere thanks to *Prof. Neelesh Kumar Jain*, the Director, IIT Indore, and the entire IIT Indore for providing me the platform to complete my degree.

The Ph.D. work could have never been complete without the support of staff members. I want to express special thanks to *Dr. Ravinder Kumar* from SIC facility, *Mr. Tapesh Parihar* from the Academics section, and *Mr. Amit Mishra*, *Mr. Arif Patel*, *Mr. Murphy Bhaskar Ganveer*, *Mr. Gaurav Singh* from the BSBE department. I am obliged to all the library staff, medical team, hostel staff, and academic section, all other technical and non-technical staff for their unhesitating help whenever required. Without them, life at IIT Indore would not be so simple.

I would especially thank the security personnel and the housekeeping staff who helped to keep us safe and sound during our late-night stays in labs.

I would like to express my token of love and appreciation to all my lab-mates. Thank you, guys, for tolerating me through

these years, *Ritudhwaj Tiwari, Anurag Mishra, Suman Bishnoi, Prativa Majee, Shruti Pyasi, Soumya Jaiswal, Ramesh Kumar, Dr. Venkata Narayan* Are from the DN lab. A special thanks to *Ravi Raj Singh, Sanchit Neema, Dr. Surjendu Dutta, Sheeba Rehman, Arpita Baa and Priyanka* for their support.

I express my heartiest gratitude to all my friends, batchmates and juniors, *Uma Shankar, Dr. Eshan Khan, Dr. Subodh Kumar Mishra, Rinki Singh, Prashant Kharey, Anjali Roy, Dr. Anshu Kumari, Yogajivan Raut, Rajesh Barik, Debasais Panda, Gaurav Pandey, Monika Jain*, for their constant appreciation and motivation. I would also like to acknowledge our collaborators, *Prof. Niranjana Sahoo* and *Dr Anjan Sahoo*, Department of Epidemiology and Preventive Medicine, and College of Veterinary Science and Animal Husbandry, Odisha University of Agriculture and Technology, Bhubaneswar, India.

At last, I would like convey my immense gratitude towards my family. The five long years of my Ph.D. have been a challenging time for me and my family, having a lot of obstacles and unsuceed dreams, I began the journey, and today I am ready here to submit my thesis. All credit goes to my uncle *Braja Kishore Sahoo* and parents and my wife *Rezismita Sahoo*. I am happy to have my wife Rezi, who has provide me space, freedom and support in every aspects during this journey. Thank you, Rezi.

I would like to thank the University Grants Commission (UGC), New Delhi, for my JRF as well as SRF fellowship, which fortified me to conduct my thesis work comfortably.

My Ph.D. tenure at IIT Indore was lucrative, exciting, and rewarding. I am taking a bagful of experiences and memories from here and would like to acknowledge each and everyone who was a part of this beautiful journey.

Thank You...

Basanta Pravas Sahu

IIT Indore

SYNOPSIS

A virus is a small biological entity that resides within the cells of hosts. During the infection, the virus replicates within the host's cell and produces thousands of identical copies of the original. Since the discovery of plant virus in tobacco plants by Dmitri Ivanovsky's 1892 and animal virus in cattle (aphthovirus) by Martinus Beijerinck in 1898, the number is continuously growing with the identification of more than 6,000 viruses. They exhibit all life forms, such as animals, plants, bacteria, and fungi. Viruses are prone to recombination through which they can exchange genes between or within to increase genetic diversity and maintains evolution. Host range differed from virus to virus, which classified them as narrow host range having species-specific host such as smallpox virus, infect only humans or broad host range, such as rabies, a virus that can infect different species of mammals.

The broad host range includes zoonotic diseases, which are defined as infections that are naturally transmitted from animals to humans. Zoonotic pathogens have diverse taxonomic groups like viruses, bacteria, fungi, and protozoa, which cause 75% of emerging human infectious diseases. Some examples include the highly pathogenic avian influenza (HPAI) virus, severe acute respiratory syndrome coronavirus (SARS-CoV), and the Ebola virus. Animals are generally used as a common source for food, dairy, clothing, companions, guards, research, and learning. That increases the chance of contact by which pathogens from carrier host transferred to susceptible animal and cause infection. There are growing calls for integrated 'One Health' approaches to controlling zoonotic diseases. This approach brought together veterinary,

ecological, conservation, and human medical perspectives to address several problems like emerging diseases, food security, and preservation of endangered animals. We followed this approach to decipher viral infectious agents related to zoonosis, food security, and companion animals during our study.

2. Objective

The present study aims to achieve three broad "One Health" objectives by studying incidence of emergence of pathogens associated with following scenarios.

1. Zoonotic viral diseases
2. Food security
3. Companion animal

1. Zoonotic viral diseases

1.1 The prevalence and epidemiological analysis of Orf virus from Eastern India.

1.2 Development of microsatellite markers within the Orf virus to decipher viral strain.

1.3 Complete genome characterization of first Indian Orf virus through next-generation sequencing (NGS).

2. Food security

2.1 Emergence of the fowlpox virus and pigeonpox virus were identified through epidemiology.

2.2 Molecular marker development for these viruses.

2.3 Complete genome analysis of Indian fowlpox virus utilizing NGS platform.

3. Companion animal

3.1 Complete genome analysis of Indian pigeonpox virus.

3.2 Comprehensive analysis of Simple Sequence Repeats in Picornaviruses.

4. Summary of thesis work

Chapter 1: Review of literature regarding infectious viral diseases, zoonosis, and one health approach. This chapter describes a detailed review of the literature regarding the zoonotic potential of several viruses. This portion also elaborates on one health's niches and its importance to control infectious disease, food security, companion animal, and maintaining a healthy environment. Additionally, we have insight into the epidemiological surveillance of virus outbreak location, implementation of molecular markers for viral strain demarcation, and utilization of next-generation sequencing strategies to decipher the complete genome of viruses.

Chapter 2: Material and method. This chapter highlights the materials, methodology, and principles involved in conducting the research studies. Additionally, this chapter briefs about the basic principle of various techniques used for virus identification, complete genome characterization using various genomics tools. It describes the techniques such as Polymerase Chain Reaction (PCR), Cycle sequencing reaction, Sanger sequencing, Next-generation sequencing (NGS) technology along with their fundamentals to decipher

virus biology. Furthermore, the bioinformatics and computational tools to identify molecular markers such as microsatellite followed by polymorphism analysis in the viral genome and several population genetics and assembly software like Bio edit, MAFFT, MEGA, BWA, DNasp were described in this chapter.

Chapter 3: Molecular characterization, comparative and evolutionary analysis of the recent Orf outbreaks among goats in the Eastern part of India (Odisha). This chapter elucidated the epidemiological characteristics of a zoonotic virus, Orf (ORFV), the causative agent of contagious ecthyma disease that primarily occurs within small ruminants and animal handlers. We surveyed the fifteen outbreak locations of eastern India during this study and collected clinical samples from infected goats. Total DNA was extracted and screened with the universal PCR primers to confirm the respective virus. Further amplification using gene-specific primers targeting fragments of ORFV011, ORFV020, ORFV059, ORFV108 genes, followed by sequencing and GenBank submission to get accession numbers. The comparative genomics approach was deployed to construct the gene-specific phylogenetic tree and evolutionary parameters of this virus, suggesting that the viral strain was closely related to Chinese isolates. The evolutionary study revealed the purifying selection is maintaining the heterogeneity within the viral strain.

Chapter 4: Development of microsatellite markers within the Orf virus to decipher viral strain. Genome-wide *in-silico* investigation of microsatellites markers within the Orf virus (ORFV) genome, the etiological agent of contagious

ecthyma disease has been accomplished to decipher the type, distribution, and potential role in the genome evolution, during this study. We have scrutinized eleven ORFV strains during this study, which showed the existence of 1,036–1,181 microsatellites or Simple Sequence Repeats (SSRs) per strain. Further analysis deciphered residence of 83–107 compound SSRs (cSSRs) per genome.

Classified microsatellites indicate that the dinucleotide (76.9%) SSRs to be predominant in comparison with trinucleotide (17.7%), mononucleotide (4.9%), tetranucleotide (0.4%), and hexanucleotide (0.2%) repeat. Other parameters like Relative Abundance (RA) and Relative Density (RD) of those microsatellites differed between 7.6–8.4 and 53.0–59.5 bp/kb. In compound microsatellites, the RA and RD varied from 0.6–0.8 and 12.1–17.0 bp/kb. Regression analysis revealed all parameters such as the incident of microsatellites, RA, and RD remarkably correlated with the GC content. But in the case of genome size, excluding the incidence of microsatellites, all other parameters were non-significantly correlated. Almost all compound microsatellites were made up of two SSRs, which showed unbiased to a particular motif. Motif duplication pattern, like, (C)-x-(C), (TG)-x-(TG), (AT)-x-(AT), (TC)- x-(TC) and self-complementary cSSRs, such as (GC)-x-(CG), (TC)-x-(AG), (GT)-x-(CA) and (TC)-x-(AG) were noticed during the study. Finally, *in-silico* polymorphism was evaluated, followed by *in-vitro* validation utilizing PCR and sequencing analysis. The thirteen polymorphic microsatellite markers developed during this investigation were characterized by mapping with the GenBank database. The outcome of this study suggests that these microsatellites

could be a tool for strain demarcation, genetic diversity estimation, and evolutionary analysis of the virus.

Chapter 5: Recombination may drive the emergence of Orf virus diversity: Evidence from the first complete genome of Indian Orf virus and comparative genomic analysis. We have explored the 2017 ORFV outbreak within goats in Madhya Pradesh, a state of central India, during this study. The outbreak was distinguished by a moderate morbidity rate (up to 20%) with a lack of mortality. Phylogenetic analysis deciphered the transboundary perspective of the virus by indicating its correlation with a distinct geographical location. We elucidated the draft genome of this viral strain, named Ind/MP, and carry out a comparative genomic analysis. The Ind/MP complete genome was having 139,807 bp with GC content 63.7%. The genome was made up of 132 open reading frames (ORFs) flanked by inverted terminal repeats (ITRs) of 3,910 bp at both ends with terminal BamHI sites and conserved telomere resolution sequences. Population genetic indices like nucleotide diversity (π), selection pressure analysis ($\theta=dN/dS$), etc., indicate that the ORFV inhabit under purifying selection. Like Simple Sequence Repeats (SSRs) and compound SSRs (cSSRs), the versatile molecular markers were prevalent within the functional protein region, having the value 70% and 67%. A sum of forty recombination incidents was recognized. Ind/MP strain actively contributed to 21 events indicating the potentiality of this strain for recombination to generate new variants.

Chapter 6: The emergence of subclades A1 and A3 Avipoxviruses in India. In this chapter, we described the

surveillance report of outbreaks of Avipoxvirus (APV) infected domestic chickens and pigeons belong to the Eastern Indian state of Odisha during the years 2010–2018. Analyzing typical lesions over the body, followed by molecular techniques, the overall morbidity was observed 18%–19.23% and 16.92%–23% in chickens and pigeons, respectively. PCR amplicons were intending the viral P4b core protein-coding gene and the DNA polymerase gene confirmed APV strains within 10 birds. Phylogenetic analysis of two genes evident the circulating strains were members of APV clade A. The subclade analysis deciphered the introduction of A1 and A3 subclades in Indian chickens and pigeons, respectively. This study reveals APV's presence in eastern Indian birds (Odisha) and utilization of the polymerase gene for the first time to elucidate the spreading clades of Indian APVs.

Chapter 7: Genome-wide identification and characterization of microsatellite markers within Avipox virus.

In this chapter, we conducted a genome-wide analysis to decipher the type; distribution pattern of eight complete genomes derived from the Avipox virus genus, the causative agent of pox like lesions above 300 avian species and one of the major diseases for the extinction of endangered avian species as well a severe economic threat to livelihood. The *in-silico* screening deciphered the existence of 1531-2473 SSRs per strain. In the case of compound SSRs (cSSRs), the value opted 83-107 per genome. Our analysis indicates that the dinucleotide (53.85%) repeats are the most abundant, followed by mononucleotide (34.20%), trinucleotide (11.11%), tetranucleotide (0.61%), pentanucleotide (0.11%) and hexanucleotide (0.11%) repeats. The specific parameters like Relative Abundance (RA) and Relative Density (RD) of

microsatellites ranged within 6.87-8.12 and 45.8-53.58 bp/kb. The analysis of RA and RD value of compound microsatellites resulted between 0.5-0.71 and 9.39-12.92 bp/kb, respectively. The analysis of motif composition of cSSR revealed that most of the compound microsatellite was made up of two microsatellites, with some unique duplicated pattern of motif like, (TA)-x-(TA), (TCA)-x-(TCA), etc. and self-complementary motifs, as (TA)-x-(AT). Finally, we have validated forty sets of compound microsatellite markers through the in-vitro approach utilizing clinical specimens, followed by mapping the sequencing product with the database through comparative genomics approaches.

Chapter 8: Complete genome analysis of Indian fowlpox virus utilizing NGS platform. For the first time, we have deciphered the complete genome sequence of a fowlpox virus of Indian origin through a next-generation sequencing platform and utilize a comparative genomics approach to decipher more about this virus biology. The NextSeq 500 NGS generated a total of ~2.16 GB of quality data with 7,303,963 numbers of quality reads. The previously published complete genome isolate fowlpox virus (Acc. No-AF198100) was used as the reference genome, to which the sequence generated through NGS mapped and assembled. The assembled genome of this novel virus exhibited 260,066 bp in length. A lower percentage of GC content (28.5) was observed in this newly assembled genome, in comparison to fowlpox virus (30.83), shearwater pox virus (30.23), canarypox virus (30.37), and turkeypox virus (29.78) but similar to other APV such as pigeonpox virus, flamingo pox virus, magpiepox virus and penguinpox virus. The extreme left nucleotide was considered as base 1, and the beginning region of the Inverted

Terminal Repeats (ITRs), which consists of a total length 9350 bp.

Chapter 9: Complete genome analysis of Indian pigeonpox

virus. The emergence of poxviruses has a severe health hazard to both domestic and free-ranging wild birds regardless of their age and sex due to its exorbitant contagious property, which influences both the economic and conservation aspects. The NextSeq 500 NGS generated a total of ~2.72 GB of quality data with 9,541,974 numbers of quality reads. A lower percentage of GC content (29.5) was observed in this newly assembled genome, in comparison to other Avipox viruses. The extreme left nucleotide was considered as base 1, and the beginning region of the Inverted Terminal Repeats (ITRs), which consists of total length 4688 bp. The ITR spanned throughout PPV-001 to PPV-004a and PPV-256 to PPV-259. A number of recombination events were overlapped within intergenic (events 5, 7, 8, 9, 13, 16, and 20), hypothetical protein (events 1, 2, 3 17, and 18), transcription factor VLTF (events 6 and 12), ankyrin repeat protein (events 15 and 21). Out of 21, potential recombination events, 15 events include PPV/Ind as recombinant, minor, and significant parental sequences. The recombinant PPV/Ind (event 7, 8) were observed within the intergenic region, using KX857215 (SWPV-2), (MK903864) (crusty tissue/MGP) as a minor parent and MF766431 (16069_trachea_170323), KJ859677 (PSan92) as a major parent.

Chapter 10: Comprehensive analysis of Simple Sequence

Repeats in Picornaviruses. Genome-wide identification of Simple Sequence Repeats (SSRs) of Picorna viruses was carried out to investigate type, distribution, and potential role

in genome evolution. Investigation on 88 Picornavirus species revealed the presence of 2,488 SSRs and 100 compound SSRs. The relative abundance and relative density of SSR varied between 1.953 bp/kb-5.763 bp/kb and 13.39 bp/kb-45.02 bp/kb, while that of cSSR ranged from 0.108bp/kb-0.636bp/kb and 1.36bp/kb-26.84bp/kb. Regression analysis revealed a significant correlation of genome size and GC content with the incidence of SSRs. Duplication pattern of motifs, like (C)-x-(C), (TG)-x-(TG), etc., and self-complementary motifs, such as (GC)-x-(CG), (TC)-x-(AG), etc. were observed in cSSR. Polymorphism analysis revealed that most of the cSSR were prone to instability, followed by consensus motifs. Finally, recombination analysis revealed that the breakpoints were rich in dinucleotide repeat, especially GT. However, further experimental validation is needed to elucidate the correlation between recombination hotspots and microsatellites.

Chapter 11: Scope of the thesis work and the future perspective. We investigated a handful of disease outbreaks through a molecular surveillance approach during the study. One of the important aspects of the investigation was to attend to emerging neglected pathogens, which gave us valuable information regarding many endemic animal viruses prevalent across India. While working with Orf, fowlpox, and pegionpox viruses, we deciphered the presence of molecular markers such as microsatellite and validated those using clinical samples. These molecular markers can be used to construct a multiplex panel, followed by genotyping to evaluate virus evolution. However, further molecular, evolutionary and genetic analysis of larger sample size and the

wider geographical area would provide a clear understanding of emerging patterns of viral pathogens. Finally, we revealed the complete genome of several poxviruses that would help develop effective vaccine candidates and several other future research applications.

LIST OF PUBLICATIONS

(A) Publications from the Thesis

(a) Published

1. **Sahu, B.P.**, Majee, P., Sahoo, A. and Nayak, D. (2019), Molecular characterization, comparative and evolutionary analysis of the recent Orf outbreaks among goats in the Eastern part of India (Odisha). *Agri Gene*, 12: p. 100088 (DOI: 10.1016/j.aggene.2019.100088). (IF = 0.580)
2. **Sahu, B.P.**, Majee, P., Mishra, C., Dash, M., Biswal, S., Sahoo, N. and Nayak, D. (2020), The emergence of subclades A1 and A3 avipoxviruses in India. *Transboundary and Emerging Diseases*, 67(2): p. 510-517 (DOI: 10.1111/tbed.13413). (IF = 4.188)
3. **Sahu, B.P.**, Majee, P., Singh, R.R., Sahoo, A. and Nayak, D. (2020), Comparative analysis, distribution, and characterization of microsatellites in Orf virus genome. *Scientific reports*, 10(1): p. 1-13 (DOI: 10.1038/s41598-020-70634-6). (IF=4.12)

(b) Under peer review

4. **Sahu, B.P.**, Majee, P., Singh, R., & Nayak, D. (2020). Recombination may drive the emergence of Orf virus diversity: Evidence from the first complete genome of Indian Orf virus and comparative genomic analysis. (Scientific report) (Manuscript ID: 34c9727b-a610-43d1-a18e-009e128ba703).

5. **Sahu B.P.**, George B, Majee P., Singh., R.R., Mishra A., Tiwari R, Nayak D. comprehensive analysis of Simple Sequence Repeats in Picornaviruses. (Computational biology and chemistry) (Manuscript ID: CBAC_2020_1340).

6. **Sahu B.P.**, Majee P., Singh. R.R., Sahoo N., Nayak D Genome-wide identification and characterization of microsatellite markers within Avipox virus (Infection Genetics Evolution) (Manuscript ID: with editor).

(c) Conference proceedings

1. **Sahu B.P.**, Sahoo A., Nayak D; Molecular detection and first complete genome characterization of Indian Orf virus. Presented during the conference of Emerging Areas in Biosciences and Biomedical Technologies (eBBT2) organized by IIT Indore.

2. **Sahu B.P.**, Sahoo A., Nayak D; Molecular characterization of the Orf virus isolates from Eastern India. Presented during the conference of 5th Molecular Virology meeting organized by THSTI.

3. **Sahu B.P.**, Nayak D; Orf virus: A potential agent for solid tumor oncolytic activity. Presented during ISMPO 2019 Conference in the category of “Basic & Applied Oncology” organized by RRCAT.

4. **Sahu B.P.**, Sahoo N., Nayak D., Emergence of subclade A1 and A3 in Indian Avipox virus. Presented during the conference of 6th Molecular Virology meeting organized by IIT Kharagpur.

(B) Publication apart from the Thesis

(a) Published

1. Panigrahi, A., **Sahu, B.P.**, Mandani, S., Nayak, D., Giri, S. and Sarma, T.K. (2019), AIE active fluorescent organic nanoaggregates for selective detection of phenolic-nitroaromatic explosives and cell imaging. *Journal of Photochemistry and Photobiology A: Chemistry*, 374: p.194-205 (DOI: 10.1016/j.jphotochem.2019.01.029). (IF=3.30)

2. Pyasi, S., **Sahu, B.P.**, Sahoo, P., Dubey, P.K., Sahoo, N., Byrareddy, S.N. and Nayak, D. (2020), Identification and phylogenetic characterization of bovine ephemeral fever virus (BEFV) of Middle Eastern lineage associated with 2018–2019 outbreaks in India. *Transboundary and Emerging Diseases*, 2020: p (1–7) (DOI: 10.1111/tbed.13531). (IF = 4.188)

(b) Manuscript submitted

1. Pyasi S., Singh R.R., **Sahu B.P.**, Hegde N.R., and Nayak D (2020) Genome-wide identification and characterization of polymorphic microsatellite markers in Ephemerovirus. (*Infection Genetics Evolution*) (Manuscript ID: MEEGID-D-20-01192)

TABLE OF CONTENTS

LIST OF FIGURES	xxxiii-xliv
LIST OF TABLES	xliv-xlviii
NOMENCLATURE	xlix-l
ACRONYMS	li
Chapter 1: Introduction	1-54
1.1 Viruses	1-6
1.1.1 Baltimore classification of viruses	2-4
1.1.2 Virus evolution	5
1.1.3 Viral diseases	5-6
1.1.4 The One Health concept	6-14
1.1.4.1 Zoonosis	8-12
1.1.4.2 Food safety	13-14
1.1.4.3 Companion animals	14
1.2 Microsatellite marker	15-23
1.2.1 Microsatellite markers in animals	15-18
1.2.2 Microsatellite markers in plants	19-20
1.2.3 Microsatellite markers in bacteria	21
1.2.4 Microsatellite markers in viruses	21-22
1.3 Evolutions of sequencing strategies	23-27
1.3.1 First-generation sequencers	23-24
1.3.2 Second-generation sequencers	24-26
1.3.3 Third-generation sequencers	26-27
1.4 Aim of the thesis	28
1.5 Organization of the thesis	29-34
1.6 Scope of the thesis	34
1.7 References	35-54

Chapter 2: Material, methods, and instrumentation	55-74
2.1 Material procurement details	55
2.2 Instrument Specifications	56
2.2.1 Thermal cycler	56
2.2.2 Gel Imaging	56
2.2.3 Sanger sequencer	56
2.2.4 NextSeq500 NGS platform	56
2.2.5 D1000 ScreenTape	57
2.3 Sample preparation and Methodology	57-59
2.3.1 DNA isolation	57
2.3.2 PCR amplification	57
2.3.3 Cycle Sequencing PCR	58
2.3.4 Preparation of NextSeq 500 Shotgun library	58
2.3.5 Quality check (QC), cluster generation, and sequencing	59
2.4 Principle behind the techniques/ methods used	59-66
2.4.1 Polymerase chain reaction (PCR)	59-61
2.4.2 Sanger sequencing	62-64
2.4.3 NextSeq500 next generation sequencing	64-66
2.5 Bio-informatic tools and server used in the Studies	67-70
2.5.1 Molecular Evolutionary Genetics Analysis (MEGA)	67

2.5.2 BioEdit	67
2.5.3 DnaSP	68
2.5.4 Trimmomatic	68-69
2.5.6 Burrows-Wheeler Alignment (BWA)	69
2.5.7 Sequin	69
2.5.8 Recombination Detection Program (RDP)	70
2.6 Server used in the studies	70-72
2.6.1 Imperfect Microsatellite Extractor (IMEx)	70
2.6.2 Expert Protein Analysis System (ExPASy)	71
2.6.3 MicroSATellite identification tool (MISA)	71
2.6.4 The Genome Annotation Transfer Utility (GATU)	71-72
2.7 References	72-74

Chapter 3: Molecular characterization,	75-104
comparative and evolutionary analysis of	
the recent Orf outbreaks among goats in the	
Eastern part of India (Odisha)	
3.1 Introduction	75-78
3.2 Results	78-90
3.2.1 Outbreaks description, preliminary	78
screening, and viral isolation	
3.2.2 PCR amplification and sequencine	79
3.2.3 Sequence analysis and Phylogenetic tree	79-90
construction	
3.2.3.1 ORFV011 gene	79-82
3.2.3.2 ORFV020 gene	83-85
3.2.3.3 ORFV059 gene	86-88
3.2.3.2 ORFV108 gene	89-90
3.3 Discussion	91-94
3.4 Materials and methods	94-97
3.4.1 Sample and outbreak data collection	94
3.4.2 DNA isolation and PCR	95
3.4.3 Primer design and preliminary screening	95
by PCR	
3.4.4 Amplification and sequencing of fragments	96
containing the ORFV011, ORFV020, ORFV059, and	
ORFV108 genes	

3.4.5 Estimation of population genetics parameters	96-97
3.5 Conclusion	97
3.6 References	97-104
Chapter 4: Comparative analyses, distribution, and characterization of microsatellites in Orf virus genome	105-146
4.1 Introduction	105-108
4.2 Results	108-121
4.2.1 Distribution of SSRs and cSSRs in ORFV genome	108-112
4.2.2 Genomic parameters influencing SSR and cSSR distribution	112
4.2.3 The frequency of classified repeat types	112-113
4.2.4 Motif complexity of compound microsatellites	113-114
4.2.5 Identification of polymorphic microsatellite through <i>in-silico</i> approach	115-116
4.2.6 Development and characterization of SSRmarkers	117-121
4.3 Discussion	122-126
4.4 Materials and methods	127-131
4.4.1 Genome sequences	127
4.4.2 Microsatellite identification, investigation, and statistical analysis	127
4.4.3 Multiple sequence alignment and	128

identification of polymorphic SSRs	
4.4.4 Sample and outbreak data	128-129
collection	
4.4.5 Development of polymorphic SSRs	130
4.4.6 Sequencing data analysis and	131
Phylogenetic tree construction	
4.5 Conclusion	131
4.6 References	132-146

Chapter 5: Recombination drives emergence 147-185

of Orf virus diversity: Evidence from the first complete genome of Indian Orf virus and comparative genomic analysis

5.1 Introduction	147-149
5.2 Results	149-167
5.2.1 Clinical gross pathological changes	149
5.2.2 Virus conformation through PCR and sequencing	149
5.2.3 NGS output and structural analysis of ORFV	153
5.2.4 Mutations, DNA polymorphism, and evolutionary analysis	155-160
5.2.5 Microsatellite and tandem repeat detection analysis	161-162
5.2.6 Phylogenetic and recombination analysis	162-167
5.3 Discussion	167-170
5.4 Materials and Methods	170-175

5.4.1 Goat herds and tissue collection	170
5.4.2 DNA extraction and virus conformation	171
5.4.3 Library construction and Illumina	172
NextSeq500 sequencing	
5.4.4 Raw sequence data processing, mapping, assembly, and genome annotations	173
5.4.5 Mutations, DNA polymorphism, and evolutionary analysis	174
5.4.6 Detection of Simple Sequence Repeats (SSRs) or Microsatellites	174
5.4.7 Phylogenetic and recombination analysis	175
5.5 Conclusion	175
5.6 References	176-185
Chapter 6: The emergence of subclade	187-210
A1 and A3 Avipoxviruses in India	
6.1 Introduction	187-189
6.2 Results	189-198
6.2.1 Clinical pathology	189-191
6.2.2 Phylogenetic analysis of the P4b (FPV 167) and DNA polymerase (FPV094) gene	192-198
6.3 Discussion	199-202
6.4 Materials and methods	202-204
6.4.1 Sample and outbreak data collection	202
6.4.2 DNA isolation	203
6.4.3 PCR amplification	203
6.4.4 DNA sequence analysis of the P4b and	204

DNA polymerase genes	
6.5 Conclusion	204
6.6 References	205-210
Chapter 7: Genome-wide identification	211-246
and characterization of microsatellite markers	
within the Avipox virus	
7.1 Introduction	211-213
7.2 Results	213
7.2.1 Distribution of SSRs and cSSR in APV	213-215
Genome	
7.2.2 Frequency of classified repeat types	215-218
7.2.3 Motif complexity of compound	218-220
microsatellites	
7.2.4 Development and characterization of	220-227
cSSR markers	
7.3 Discussion	228-231
7.4 Materials and methods	231-233
7.4.1 Genome sequences	231
7.4.2 Microsatellite identification, investigation,	231-232
and statistical analysis	
7.4.3 Microsatellite PCR analysis and sequencing	232
7.4.4 Sequencing data analysis and construction	233

Of the Phylogenetic tree	
7.5 Conclusion	233
7.6 References	234-246
Chapter 8: Complete genome analysis of Indian fowlpox virus	247-258
8.1 Introduction	247-248
8.2 Results	248-251
8.2.1 NGS output and general features of Indian FPV	248-249
8.2.2 Phylogenetic and recombination analysis	249-251
8.3 Discussion	251-252
8.4 Materials and methods	252-255
8.4.1 Nucleic acid extraction	252
8.4.2 Library construction and Illumina NextSeq500 sequencing	253
8.4.3 Raw sequence data processing, mapping, assembly, and genome annotations	254
8.4.4 Phylogenetic and recombination analysis	254-255
8.5 Conclusion	255
8.6 References	255-258

Chapter 9: Complete genome analysis of	259-272
Indian pigeonpox virus	
9.1 Introduction	259-261
9.2 Results	261-264
9.2.1 NGS output and general features of	261-262
Indian PPV	
9.2.2 Phylogenetic and recombination analysis	262-264
9.3 Discussion	265
9.4 Materials and methods	266-268
9.4.1 Nucleic acid extraction	266
9.4.2 Library construction and Illumina	266
NextSeq500 sequencing	
9.4.3 Raw sequence data processing, mapping,	267
assembly, and genome annotations	
9.4.4 Phylogenetic and recombination analysis	268
9.5 Conclusion	268
9.6 References	268-272
Chapter 10: A comprehensive analysis of	273-313
Simple Sequence Repeats in Picornaviruses	
10.1 Introduction	273-276
10.2 Results	276-298
10.2.1 Number, relative abundance and relative	276-282
density of SSR in picornaviruses	
10.2.2 Number, relative abundance and relative	283-285
density of cSSR in picornaviruses	
10.2.3 Effect of dMAX on cSSR incidence	286
10.2.4 Diversity of derived motifs	287-289

10.2.5 Over and under-representation of Mononucleotide repeat	289-292
10.2.6 Motif complexity and polymorphism of Picornavirus genome	292-295
10.2.7 Recombination analysis	295-296
10.2.8 Genomic parameters influencing SSR and cSSR distribution	296-298
10.3 Discussion	299-301
10.4 Materials and methods	302-304
10.4.1 Genome sequences	302
10.4.2 Microsatellite identification and investigation	302
10.4.3 Estimation of the number of expected mononucleotide repeats	303
10.4.4 Multiple sequence alignment and identification of polymorphic cSSR	303
10.4.5 Recombination analysis	304
10.4.6 Statistical analysis	304
10.5 Conclusion	304
10.6 References	305-313
Chapter 11: Conclusion and scope for future work	314-321
11.1 Conclusion drawn from the present work	314-319
11.2 Future scope of the present work	319-322
Annexure	
Annexure A: Published isolates of ORFV used for phylogenetic and population genetic analysis.	323-335
Annexure B: Published isolates of APV used for phylogenetic analysis.	335-345

LIST OF FIGURES

Chapter 1: Introduction

- Figure 1.1:** Schematic diagram representing the 4
Baltimore classification of viruses depending on their genome organization and mode of replication. It consists of three groups of viruses. These groups are further classified as ss: single-stranded, ds: double-stranded, +: positive sense, -: negative sense, RT: reverse transcription depending upon the intermediate steps of mRNA synthesis.
- Figure 1.2:** One Health umbrella: The umbrella 8
illustrating various niches of One Health and their interaction to solve critical problems.
- Figure 1.3:** Goat infected with ORFV: figure 10
illustrating ORFV infection in Black Bengal goat showing lesions around the lip recorded during the outbreak investigation.
- Figure 1.4:** Human Orf: Orf virus infection on the 10
hand of a person working in a butcher shop in Indore district, Madhya Pradesh, India with an ulcerative lesion or nodule over the right hand.
- Figure 1.5:** Linkage map of chicken. Diagram 18
illustrating a chromosomal linkage map of chicken, where the left side represents the distance of SSR markers in centimorgan and the right side representing the name of microsatellite markers.
- Figure 1.6:** Linkage map of rice. Diagram illustrating 20
a chromosomal linkage map of rice, where the left side

represents the distance of SSR markers in centimorgan and the right side representing the name of microsatellite markers.

Figure 1.7: Microsatellite in a virus. Diagram 23 illustrating electropherogram of a compound microsatellite having motif composition (AAT)₃-x9-(T)₆ within Orf virus.

Figure 1.8: Polymorphism within microsatellite. The 23 figure depicts (GGC)₃ microsatellite polymorphism within the ankyrin protein gene of the Orf virus.

Chapter 2: Material, methods, and instrumentation

Figure 2.1: Principle of PCR. The figure represents 61 PCR steps such as Denaturation, Annealing, and Elongation of specific DNA fragments during an amplification reaction.

Figure 2.2: Working principle of Sanger sequencing. 64 The sequencing reaction starts with primer annealing and chain extension followed by chain termination, capillary electrophoresis, and sequence analysis.

Figure 2.3: Principle behind NextSeq500 NGS. This 66 NGS platform generally utilizes the sequencing by synthesis (SBS) chemistry. The sequencing reaction begins with library preparation followed by hybridization and bridge amplification, sequencing, and assembly.

Chapter 3: Molecular characterization, comparative and evolutionary analysis of the recent Orf outbreaks among goats in the Eastern part of India (Odisha)

Figure 3.1: Goat infected with ORFV: figure 78
illustrating clinical cases of ORFV infection in Black
Bengal goat showing lesions around the lip recorded
during the outbreak investigation.

Figure 3.2: Phylogenetic trees constructed for 82
ORFV011 based on partial nucleotide sequences
retrieved from GenBank. Bootstrapped maximum
likelihood (ML) phylogenetic trees were constructed
with MEGA6 based on the Tamura-Nei model for 1000
replications using aligned sequences. Black circles
represent the new isolates from Odisha, India,
mentioned in our study, while the black triangles
represent the other Indian isolates.

Figure 3.3: Phylogenetic trees constructed for the 84
ORFV020 based on partial nucleotide sequences
retrieved from GenBank. Bootstrapped maximum
likelihood (ML) phylogenetic trees were constructed
with MEGA6 based on the Tamura-Nei model for 1000
replications using aligned sequences. Black circles
represent the new isolates from Odisha, India,
mentioned in our study, while the black triangles
represent the other Indian isolates.

Figure 3.4: Phylogenetic trees constructed for the 88
ORFV058 based on partial nucleotide sequences
retrieved from GenBank. Bootstrapped maximum
likelihood (ML) phylogenetic trees were constructed
with MEGA6 based on the Tamura-Nei model for 1000
replications using aligned sequences. Black circles
represent the new isolates from Odisha, India
mentioned in our study while the black triangles
represent the other Indian isolates.

Figure 3.5: Phylogenetic trees constructed for the ORFV108 based on partial nucleotide sequences retrieved from GenBank. Bootstrapped maximum likelihood (ML) phylogenetic trees were constructed with MEGA6 based on the Tamura-Nei model for 1000 replications using aligned sequences. Black circles represent the new isolates from Odisha, India, mentioned in our study, while the black triangles represent the other Indian isolates. 90

Chapter 4: Comparative analyses, distribution, and characterization of microsatellites in Orf virus genome

Figure 4.1: Pie-chart illustrating the percentage of differential distribution of SSR within coding (UTR and intergenic region) and non-coding (functional and hypothetical proteins) regions of ORFV genome. 109

Figure 4.2: Pie-chart illustrating the percentage of differential distribution of cSSR within coding (UTR and intergenic region) non-coding (functional and hypothetical proteins) regions of ORFV genome. 110

Figure 4.3: Frequency of cSSR in relation to varying dMAX (10–100) across eleven ORFV complete genomes represented on the right side of the graph. A higher cSSR incidence was observed with increasing dMAX in the genomes. 111

Figure 4.4 A-C: Distribution of different motifs of mono, di, and tri microsatellites in the different genomes of ORFV. 113

Figure 4.5: Circos Plot showing the Genome size, CDS, Distribution of SSR, selected SSR markers, 116

cSSR, and GC content in ORFV (OV-SA00) genome. From outer track to inner track: Genome size, CDS, SSR, selected SSR markers (Black lines within the SSR), cSSR, and GC content.

Figure 4.6: *Clinical samples evaluation by universal ORFV primers. Electrophoresis gel showing the PCR amplicon of four suspected ORFV clinical samples were collected from Black Bengal goats. M: 100 bp DNA ladder; -C: Negative controls (PCR using nuclease-free water as DNA template); 1-4: Clinical samples.* 119

Figure 4.7: *Clinical samples validation using SSR markers. Electrophoresis gel showed the PCR amplicon of the developed SSR markers in ORFV. SSR markers from SSR1 to SSR13; M: 50 bp DNA ladder; -C: Negative controls (PCR using nuclease-free water as DNA template).* 121

Figure 4.8: *The concatenated phylogenetic tree was constructed using bootstrap consensus tree building method of neighbor joining with bootstrap value 500 using MEGA 5. Black triangle represents the ORFV isolates of present investigation showing its relationship with eleven global strains.* 121

Figure 4.9: *ORFV infection in goat. Representative figure clinical cases of ORFV infection in Black Bengal goats having proliferative lesions around the lip recorded in the study area.* 129

Chapter 5: Recombination drives emergence of Orf virus diversity: Evidence from the first complete genome of Indian Orf virus and comparative genomic analysis

Figure 5.1 A-D: Phylogenetic trees constructed based on the four ORFV genes, namely ORFV011, ORFV020, ORFV059, and ORFV108. The phylogenetic relationship was constructed by the GTR model of the maximum likelihood method using MEGA 6.0 software. Numbers at the branching points indicate the bootstrap support calculated for 1,000 replicates (A) ORFV011, (B) ORFV020, (C) ORFV059, and (D) ORFV108 based on partial nucleotide sequences. The black triangle represents the new isolate mentioned in our study.

Figure 5.2: Construction of concatenated phylogenetic tree to determine reference genome. Concatenated DNA sequences comprising ORFV011, ORFV020, ORFV059, and ORFV108 genes of ORFV were used to construct a phylogenetic tree with bootstrap value 1000 using the GTR model of maximum likelihood method. The black triangle represents the ORFV isolate of present investigation showing its close relationship with the strain KP010354, which is considered as the reference.

Figure 5.3: Analysis of ITR-ends of ORFV. (A) The left end (5') sequence alignment of 14 complete ORFV genomes consists of a terminal BamHI site (green box) along with telomere resolution motifs (red box) (ATTTTTT-N(8)-TAAAT). (B) The right end (3') sequence alignment of 14 complete ORFV genomes consists of a terminal BamHI site (green box) along with telomere resolution motifs (red box) (ATTTTTT-N(8)-TAAAT).

Figure 5.4: *Circos plot illustrating the complete genome of the Ind/MP ORFV strain. Schematic presentation of the complete genome: from outer to inner track: ORFV genome size with 1Kb division, genes on a negative- strand (black), genes on a positive-strand (red), SSRs, cSSRs, Synonymous mutation, Nonsynonymous mutation. These mutations, along with SSRs and cSSR, were ubiquitously present throughout the genome irrespective of gene orientation.* 155

Figure 5.5: *Classification of SSR distribution in ORFV genome. Distribution of different motifs of SSRs within ORFV genomes. Dinucleotide repeats were found to be the most abundant, followed by trinucleotide and mononucleotide repeats in the reported genome. Among the di and trinucleotide repeats CG/GC and CGC/GCG having the highest number among all classified repeats.* 162

Figure 5.6: *Comparison of PPVs by constructing Phylogenetic tree. Nineteen complete genome sequences, including the terminal repetitions, were aligned to construct a phylogenetic tree with bootstrap value 1000 using the GTR model of maximum likelihood method. The black triangle represents the ORFV isolates of the present investigation showing its relationship with fourteen global strains.* 163

Chapter 6: The emergence of subclade A1 and A3 Avipoxviruses in India

Figure 6.1: *Chickens infected with APV: figure*

195

illustrating clinical cases of APV infection in chickens showing lesions around the beak, eyelid and featherless parts recorded during the outbreak investigation.

Figure 6.2 A-B: *PCR confirmation of Avipoxvirus genes: (a) The left panel shows the PCR amplification of the P4b gene from the ten APV isolates in this study. The last lane represents the marker (100 bp), while the second to last lane represents the negative control, i.e., no template. (b) The right panel shows the PCR amplification of the DNA polymerase gene from the ten APV isolates in this study. The last lane represents the marker (100 bp), while the second to last lane represents the negative control, i.e., no template.*

Figure 6.3: *Phylogenetic tree construction for Avipoxvirus isolates. Concatenated DNA sequences comprising P4b and DNA polymerase genes Avipoxviruses was used to construct a phylogenetic tree with posterior probability values of the Bayesian trees (1,000 replicates), neighbor-joining, and maximum likelihood bootstrap values (1,000 replicates). The left side of the branch points having some numeric, which represents the Bayesian posterior probability values/ neighbor-joining bootstrap values/maximum likelihood bootstrap values and subclades, labeled according to the nomenclature of Mapaco et al. 2018 and Offerman et al., 2013. The number of substitutions per site was represented in the scale.*

Figure 6.4: The phylogenetic tree of nucleotide sequences for the P4b core protein gene of the APV isolates. Black circles represent the APVs previously isolated in India; black triangles represent the isolates of the current study. 197

Figure 6.5: The phylogenetic tree of nucleotide sequences for the P4b core protein gene of the APV isolates. Black circles represent the APVs previously isolated in India; black triangles represent the isolates of the current study. 198

Chapter 7: Genome-wide identification and characterization of microsatellite markers within the Avipox virus

Figure 7.1 A-B: Distribution of SSR and cSSR within AVP genome. Graphical representation showing the distribution of SSRs and cSSR within the different genomic locations, including the noncoding and coding regions. 215

Figure 7.2: Effect of dMAX. Percentage of compound microsatellite differed with various dMAX values (10–100). The accession number of each species of the APV genus is mentioned on the right side of the graph. An increased incidence of compound microsatellite was observed with a higher value of dMAX. 216

Figure 7.3 A-C: Distribution of classified microsatellites among APV genomes. (A) Distribution pattern of mononucleotide SSRs within APV genomes, (B) Distribution pattern of dinucleotide SSRs within APV genomes, and (C) Distribution pattern of 218

trinucleotide SSRs within APV genomes.

Figure 7.4: *Circos map construction. Circos plot 220 demonstrating the genome-wide study of SSRs and cSSRs across eight different species of Avipoxvirus. From outer to inner ring: Species, Genome size, Genes, SSRs, SSRs frequency/1kbs, cSSRs, developed cSSR markers, and GC content.*

Figure 7.5: *Microsatellite markers development using 225 clinical samples of APV. Validation of compound microsatellites primers using a clinical sample. M:100 bp DNA ladder; -C: Negative controls (PCR using nuclease-free water as DNA template). (a) Fowlpox virus (FPV SSR1-FPV SSR10) (b) Fowlpox virus (FPV SSR11-FPV SSR20) (c) Pigeonpox virus (PPV SSR1-PPV SSR10) (d) Pigeonpox virus (PPV SSR11-PPV SSR20).*

Figure 7.6: *BLAST analysis depicts that the tested 226 PPV SSR 15 (marked yellow) shows polymorphism with the other strain like pigeonpox virus isolate FeP2 (KJ801920) having an additional T.*

Figure 7.7: *Construction of a concatenated 227 phylogenetic tree. A consensus phylogenetic tree constructed using a concatenated sequence of microsatellites by the neighbor-joining method with bootstrap value 1000 of MEGA7. During the study, the clinical isolate represented a black triangle and black circle showing its relationship and Feral pigeon strain.*

Chapter 8: Complete genome analysis of Indian fowlpox virus

Figure 8.1: Phylogenetic relationship between Indian 250
FPV and other AVPs. The General Time Reversible
(GTR) model of MEGA6 was used to create a
phylogenetic tree based on the Maximum likelihood
(ML) phylogeny with 500 bootstraps. The
abbreviations and GenBank accession details for
seven core protein sequences of the selected AVPs
have used pigeonpox virus (FeP2; KJ801920),
fowlpox virus (FWPV; AF198100, MF766430-32,
MH709124, MH719203, and MH734528),
shearwaterpox virus 1 (SWPV-1; KX857216),
shearwaterpox virus 2 (SWPV-2; KX857215).

Chapter 9: Complete genome analysis of Indian pigeonpox virus

Figure 9.1: Phylogenetic relationship between Indian 263
PPV and other AVPs. The General Time Reversible
(GTR) model of MEGA6 was used to create a
phylogenetic tree based on the Maximum likelihood
(ML) phylogeny with 500 bootstraps. The
abbreviations and GenBank accession details for
seven core protein sequences of the selected AVPs
have used pigeonpox virus (FeP2; KJ801920),
fowlpox virus (FWPV; AF198100, MF766430-32,
MH709124, MH719203, and MH734528),
shearwaterpox virus 1 (SWPV-1; KX857216),
shearwaterpox virus 2 (SWPV-2; KX857215).

Chapter 10: A comprehensive analysis of Simple Sequence Repeats in Picornaviruses

Figure 10.1: <i>The Circos map is representing</i>	88	282
<i>Picornavirus genomes. From the outer ring to inner:</i>		
<i>Selected Picornavirus genomes, genome size,</i>		
<i>polyprotein gene, mature proteins, SSRs, cSSRs, and</i>		
<i>GC content.</i>		
Figure 10.2: <i>Effect of cSSR% on varying dMAX in five</i>	286	286
<i>different species of Picornavirus: For this analysis,</i>		
<i>IMEx software was used, which allowed varying the</i>		
<i>dMAX value from 10 to 50.</i>		
Figure 10.3: <i>Differential distribution of SSR in the</i>	288	288
<i>coding/noncoding region of Picornavirus.</i>		
Figure 10.4: <i>Distribution of mono-, di- and tri-</i>	288	288
<i>nucleotide SSR motifs (%) across coding/noncoding</i>		
<i>regions of Picornavirus.</i>		
Figure 10.5: <i>Differential distribution of cSSR in the</i>	294	294
<i>coding/noncoding region of Picornavirus.</i>		
Figure 10.6 A-B: <i>Regression analysis of genomic</i>	297	297
<i>features, namely, the genome size (A) and GC content</i>		
<i>(B) with Incidence, RA, and RD of SSRs.</i>		
Figure 10.7 A-B: <i>Regression analysis of genomic</i>	298	298
<i>features, namely, the genome size (A) and GC content</i>		
<i>(B) with Incidence, RA, and RD of cSSRs.</i>		

LIST OF TABLES

Chapter 3: Molecular characterization, comparative and evolutionary analysis of the recent Orf outbreaks

among goats in the Eastern part of India (Odisha)

Table 3.1: Table showing the characteristics of the animals included in the study. The characteristics of animals used for sample collection in the investigation and their corresponding morbidity percentage. M = Male, F=Female, Y = years, Mb = Morbidity. 80

Table 3.2: The table consists of the primers used in our study. The first set of primers, Orf1, and Orf2 are used for screening and confirming the ORFV isolates. The rest of the primers are used for amplification and sequencing of the genes, namely, ORFV011, ORFV020, ORFV059, and ORFV108 of the ORFV. 85

Table 3.3: The list of ORFV field isolates of Odisha (n=15) characterized in this study. The table includes the isolate name, area of collection, animal or host from where the samples were collected and the GenBank accession number for the four ORFV genes, ORFV011, ORFV020, ORFV059, and ORFV108. 86

Table 3.4: Table enlists the different population genetic parameters estimated for ORFV globally. 89

Chapter 4: Comparative analyses, distribution, and characterization of microsatellites in Orf virus genome

Table 4.1: Overview of microsatellites in ORFV complete genome sequences. 111

Table 4.2: Characteristics of the 13 microsatellite markers developed for the ORFV. 117

Table 4.3: Alignment of the 13 sequenced microsatellite markers (partial) against the complete genome present in the NCBI database. 120

Chapter 5: Recombination drives emergence of Orf virus diversity: Evidence from the first complete genome of Indian Orf virus and comparative genomic analysis

Table 5.1: Mutational analysis ORF isolates. The number of unique synonymous and unique nonsynonymous amino acid substitutions across all isolates sequenced for this study is presented in the table. Individual nonsynonymous substitutions are also shown in the column on the right of the table. 156

Table 5.2: Predicted potential recombination events between ORFV. Details recombination events detected between ORFV isolates were analyzed in this study using the RDP4 program. Detection method coding R, G, B, M, C, S, P, L, and T represents methods RDP, GENECONV, Bootscan, MaxChi, Chimaera, SiScan, PhylPro, LARD, and 3Seq, respectively. The P-value for the detection method is shown in bold. To emphasize the recombination potential of Indian isolate, (Ind/MP) has been bold. Ind/MP strain consists of 21 potential events in which it acts as recombinant, minor, and major parental sequences. 164

Chapter 6: The emergence of subclade A1 and A3 Avipoxviruses in India

Table 6.1: Details of Avipoxvirus-infected domestic birds and the location of the samples collected in this study. 190

Table 6.2: Table showing the characteristics of the birds included in the study with their corresponding morbidity and mortality percentage (W=Weeks). 191

Chapter 7: Genome-wide identification and characterization of microsatellite markers within the Avipox virus

Table 7.1: Overview of SSRs within various species of APV complete genome sequences and their parameters obtained during this study. 214

Table 7.2: The characteristics of forty compound SSR developed using a clinical sample. 222

Table 7.3: Characterization of forty cSSR markers against the NCBI database. 226

Chapter 8: Complete genome analysis of Indian fowlpox virus

Table 8.1: Predicted potential recombination events between FPV/Ind. Details recombination events detected between FPV isolates were analyzed in this study using the RDP4 program. Detection method coding R, G, B, M, C, S, P, L, and T represents methods RDP, GENECONV, Bootscan, MaxChi, Chimaera, SiScan, PhylPro, LARD, and 3Seq, respectively. The P-value for the 250

detection method is shown in bold.

Chapter 9: Complete genome analysis of Indian pigeonpox virus

Table 9.1: Predicted potential recombination events between PPV/Ind. Details recombination events detected between PPV isolates were analyzed in this study using the RDP4 program. Detection method coding R, G, B, M, C, S, P, L, and T represents methods RDP, GENECONV, Bootscan, MaxChi, Chimaera, SiScan, PhylPro, LARD, and 3Seq, respectively. The P-value for the detection method is shown in bold. 264

Chapter 10: A comprehensive analysis of Simple Sequence Repeats in Picornaviruses

Table 10.1: Overview of selected 88 species of Picornaviridae family. The dataset includes specification of the viral strain, specific host, geo-location, isolation source, and disease caused. 200

Table 10.2: Survey of various microsatellites (SSRs and cSSRs) with their relative abundance, relative density, and cSSR % in selected Picornaviridae family. 200

Table 10.3: Mononucleotide microsatellite representation in Picornaviridae family. 200

Table 10.4: The existence of the consensus motif in diverse species of Picornaviridae family. 200

NOMENCLATURE

ds	Double-stranded
ss	Single-stranded
ICTV	International Committee on Taxonomy
RV	Rabies virus
VSV	Vesicular stomatitis virus
TMV	Tobacco mosaic virus
CaMV	Cauliflower mosaic virus
TYLCV	Tomato yellow leaf curl virus
ACMV	African cassava mosaic virus
PVX	Potato virus X
CDC	Centers for Disease Control
WHO	World Health Organization
FAO	Food and Agricultural Organization
FMD	Foot and mouth disease virus
APV	Avipoxvirus
PPR	Peste des petits ruminants
SSR	Simple sequence repeat
cSSR	Compound simple sequence repeat
PCR	Polymerase chain reaction
QTL	Quantitative trait loci
MAS	Marker Assisted Selection
VNTR	Variable number of tandem repeat
STR	Short tandem repeats

RA	Relative abundance
RD	Relative density
hCMV	Human cytomegalovirus
WSSV	White spot syndrome virus
ORFV	Orf virus
NGS	Next Generation Sequencing
EDTA	Ethylene diaminetetraacetic acid
DMEM	Dulbecco's Modified Eagle's Medium
FBS	Fetal bovine serum
PBS	Phosphate buffer saline
NCCS	National Centre for Cell Sciences
ddNTPs	Dideoxynucleotides
dNTPs	Deoxynucleotides
SBS	Sequencing by synthesis
DnaSP	DNA Sequence Polymorphism
SNP	Single nucleotide polymorphism
Indel	Insertion-deletion
MEGA	Molecular Evolutionary Genetics
BWA	Burrows-Wheeler Alignment
NCBI	National Center for Biotechnology
RDP	Recombination Detection Program
IMEx	Imperfect Microsatellite Extractor
ExPASy	Expert Protein Analysis System
SIB	Swiss Institute of Bioinformatics
MISA	MicroSAtellite identification tool

GATU	Genome Annotation Transfer Utility
ORFs	Open reading frames

ACRONYMS

+ve	Positive-sense
-ve	Negative-sense
cM	Centimorgan
MB	Megabases
GB	Gigabases
mV	Millivolt
pA	Picoampere
μ l	Microliter
pmol	Picomole
°	Degree
C	Centigrade
bp	Basepair
s	Second

Chapter 1

1. Introduction

1.1 Viruses

Viruses, considered one of the most diverse microorganisms, usually act as infectious agents with genetic material inside a coat protein. The virus always requires a living system, such as humans, animals, plants, or even bacteria, for its growth and proliferation. This character is considered an 'obligate intracellular parasite' [1]. After the initial discovery of tobacco mosaic virus (TMV) in plant by Dmitri Ivanovsky in 1892 and foot and mouth disease virus (FMDV), in cattle by Martinus Beijerinck in 1898, the number of discovery of new viruses is increasing continuously with the identification of more than 6,000 viruses [2]. Viruses are prone to recombination through which they can exchange genes between or within to increase genetic diversity and maintains evolution [3]. Host range differs from virus to virus, which classified them as a narrow host range with a species-specific host such as smallpox virus, infecting only humans [4] or broad host range such as rabies virus (RV) that can infect different mammal species [5].

The size of viruses varied from nanometers in nodamura virus to micrometers in mimivirus. The shapes vary from icosahedral in rhinovirus, filamentous in potyvirus, head-tail conformation in T7 bacteriophages etc. While morphologically, it differs from having an envelope; the capsid structure may vary, contain spikes in the outer membrane, etc. Their replication procedure and replication site vary. For example most RNA viruses prefer the host cytoplasm to replicate and assemble. In contrast, most of the

DNA viruses make use of the host nucleus for the same. Viruses are classified based on different criteria, but the most conventional and widely accepted classification system is the Baltimore classification discussed below [6].

1.1.1 Baltimore classification of viruses

Unlike other organisms, the viruses lack a common ancestor, which leads to increased difficulties in categorizing them [7]. In 1971, considering the genomic composition (DNA or RNA), strandedness (single-stranded or double-stranded), sense or polarity (positive or negative), replication method, David Baltimore suggested the classification of the known viruses into specific viral families [8] (Figure 1.1). Concerning the Baltimore classification, virus families classified into seven broad categories mentioned below:

1. Category I: Double-stranded (ds) DNA virus (e.g., Vaccinia virus)
2. Category II: Single-stranded (ss) DNA virus (e.g., human bocavirus)
3. Category III: Double-stranded (ds) RNA virus (e.g., Bluetongue virus)
4. Category IV: Single-stranded (ss) positive-sense (+ve) RNA virus (e.g., West Nile virus)
5. Category V: Single-stranded (ss) negative-sense (-ve) RNA virus (e.g., Rift Valley fever virus)
6. Category VI: Single-stranded (ss) RNA having DNA intermediate with the help of reverse transcriptase enzyme (e.g., Feline immunodeficiency virus)

7. Category VII: Double-stranded DNA (ds) having RNA intermediate with the help of reverse transcriptase enzyme (e.g., Hepatitis B virus).

The replication process differs from one another in which they utilize their genomes to synthesis the viral mRNA that translates to proteins. For example, the ds DNA virus, ds RNA virus, or ss -ve RNA virus directly transcribe their genome into mRNA. In contrast, the other four groups of viruses' genomic material go through certain intermediate stages to make the mRNA.

The International Committee on Taxonomy of Virus (ICTV) classifies and categorizes recorded viral species into different hierarchical ranks [2]. The present update from ICTV classified the known viruses into 55 orders, eight suborders, 168 families, 1421 genera, 6590 species (Virus Taxonomy: 2019).

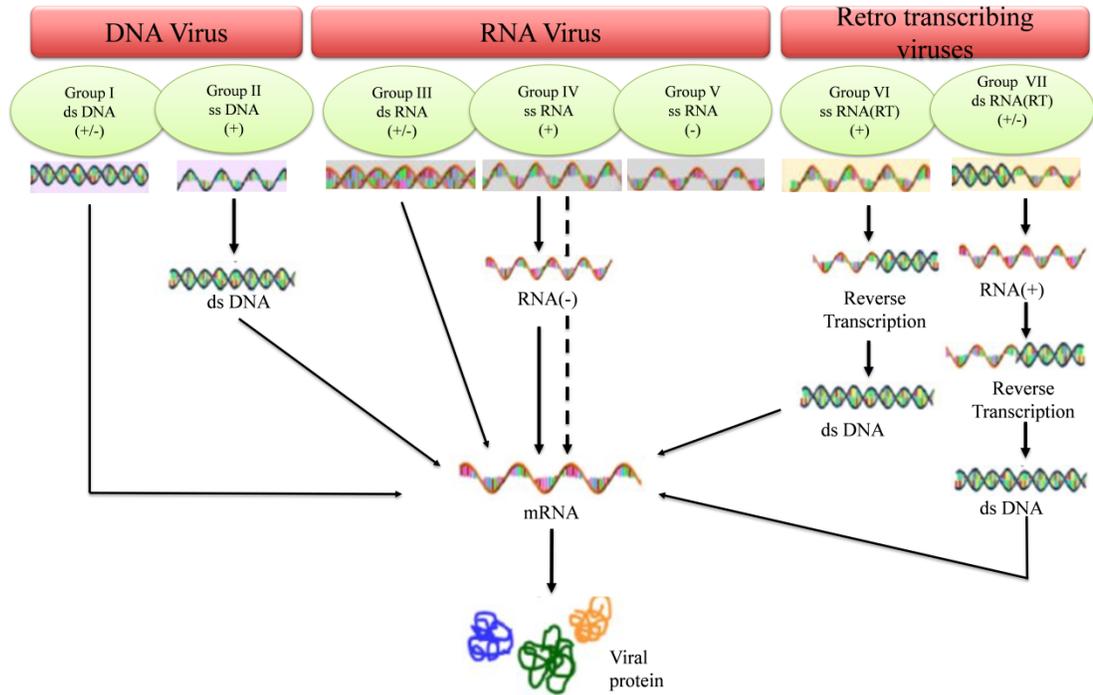


Figure 1.1: Schematic diagram representing the Baltimore classification of viruses depending on their genome organization and mode of replication. It consists of three groups of viruses. These groups are further classified as ds: double-stranded, ss: single-stranded, -: negative sense, +: positive sense, RT: reverse transcription depending upon the intermediate steps of mRNA synthesis.

1.1.2 Virus evolution

Virus evolution usually deals with how virus responds to host immune responses, antiviral stresses, etc in due course of time. Molecular mechanisms, such as mutations, recombination, and reassortment, leading to insertion, deletion, synonymous or nonsynonymous mutations viral genome, gain or loss of genes, change in reading frames are part of the evolutionary process. Usually, the RNA viruses are more inclined towards mutation than the DNA viruses due to differences in proofreading activity of DNA/RNA polymerase enzymes. The average rates of misincorporation of the nucleotide bases during RNA synthesis may range from 10^{-4} to 10^{-5} in RNA viruses, which is almost one mutation per genome in each replication cycle [9]. In contrast, the DNA viruses' mutation rate is lower as they have robust proofreading and mismatch repair mechanisms. Therefore, it is quite apparent that newly emerging and re-emerging viruses belong to RNA viruses. The change in genome sequence can be beneficial to the virus by enhancing the adaptability in the changing environment but may sometimes result in harmful consequences, leading to virus extinction [10].

1.1.3 Viral diseases

Viruses infect almost all living organisms. The severity of viral infection usually depends upon the virus type and host immune system. In most cases, the viral infection activates the host immune response and kills the infected cells to stop viral proliferation. For example, in vesicular stomatitis virus (VSV), a cytopathic virus that damages the cell by expressing pro-apoptotic proteins [11]. Few pathogenic virus-like chicken-pox viruses lie in dormant (latent) condition, in which

virus replication ceases and waits for favorable conditions to increase [12]. Interestingly, some viruses like Muvirus encode the gene integrase through which they insert the viral genome into the host genetic makeup [13].

Viral diseases are of great concern to all living organisms. Plant viral diseases caused by Tobacco mosaic virus (TMV), African cassava mosaic virus (ACMV), Potato virus X (PVX), etc., mostly affect crop production and subsequent economic losses [14]. Certain animal viruses, such as the Hantavirus, Hendra virus, Menangle virus, Equine morbillivirus, etc., affect animal health and often result in a heavy toll on the animal husbandry sector [15, 16]. The situation aggravates as some of these animal viruses have a zoonotic potential and can jump off the host barrier to infect humans.

1.1.4 The One health concept

One health is described as a combined effort of multiple disciplines like pharmaceutical, veterinary medicine, scientific, and non-scientific people working locally, nationally, and internationally to improve human and animal health as well as the ecosystem [17]. Within this approach, collaboration occurs among multidisciplinary professionals, such as veterinarians, physicians, wildlife experts, environmental researchers, public health personnel, and many others [18]. The One Health concept began by introducing the One Medicine theme by Rudolf Virchow, a German physician during the 19th century. He introduced the term "zoonosis" was also introduced by him.

Latter, Sir William Osler, James Steele, Canadian physician and pathologist laid the foundation of veterinary public health

at the Centers for Disease Control (CDC) in the USA. One health facilitates health through multidisciplinary study and regulation among all animals to address complex public health problems. Examples of such complex issues include emerging infectious diseases, food safety, and climate change. An 'umbrella' symbol recently created by Swedish One Health cooperation with the One Health Initiative, which encircles all relevant aspects. Several scientific domains are represented under one health's umbrella, such as public health, environmental chemistry, veterinary science, human medication, and the environment. One health approach forms a scientific surrounding where lab facilities are shared by physicians, microbiologists, veterinarians, and access to samples among themselves. The examples may include tissue, blood, saliva, or any information, which could be tedious, costly, and ethically complex to collect. Therefore, we need to emphasise a few terms used to elucidate one health (Figure 1.2).

These are:

1.1.4.1 Zoonosis

1.1.4.2 Food safety

1.1.4.3 Companion animals

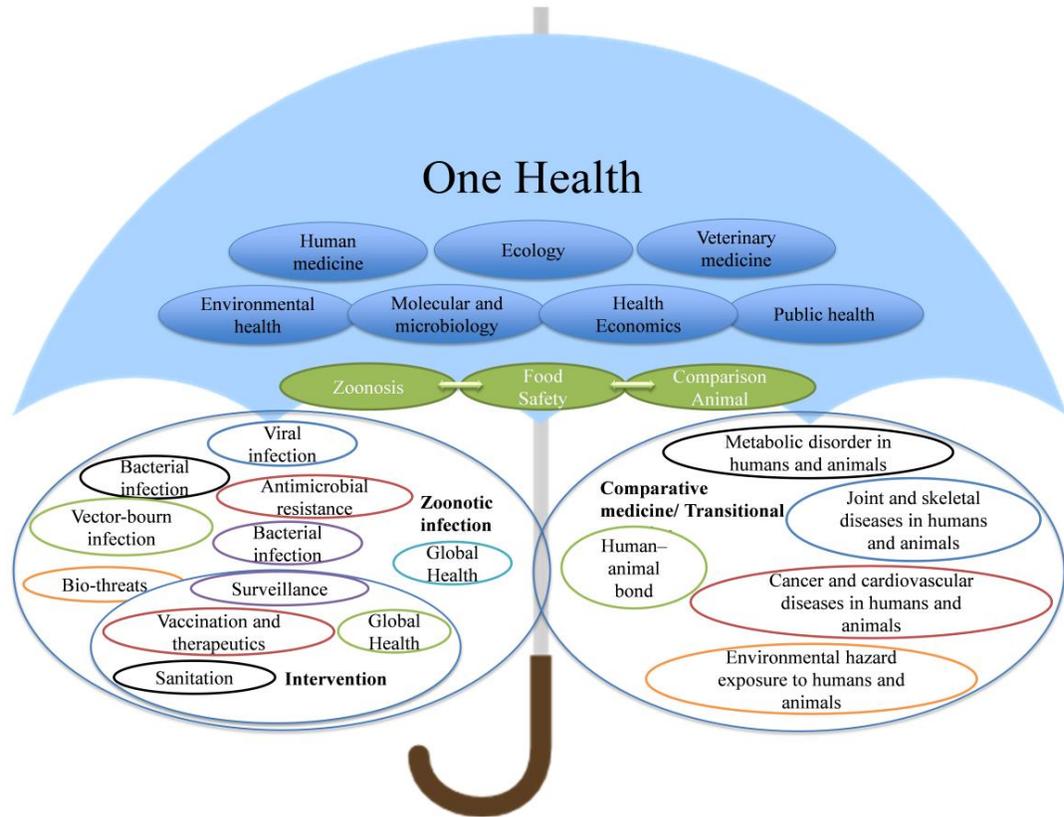


Figure 1.2: One health umbrella: The umbrella illustrating various niches of One health and their interaction to solve critical problems.

1.1.4.1 Zoonosis

According to the World Health Organization (WHO) and the Food and Agricultural Organization (FAO), zoonosis is defined as naturally transmitted disease from animals to humans. The information regarding all zoonotic pathogens is still unknown. Some reports recommend that 75% of emerging human diseases occurred through zoonosis [19]. The first zoonotic disease, rabies recorded during the eighteenth century BC Babylonian era [20]. The emergence of plague among the Philippians and the 'plague of Athens' considered

classic examples of zoonotic infections. Thucydides, the plague's survivor, narrated the symptoms like vomiting, fever, diarrhea, headache, bleeding, rashes, and erythema, and that lead to ulcers [21]. With regards to this issue, epidemiological scholars tried to find out the cause of outbreaks of typhus, smallpox, Rift Valley fever, anthrax, or dengue, those having fever as a common symptom [21].

To evaluate the emergence of a pandemic from the starting, one should decipher the correlation between humans and animals. Humans contact animals primarily for food sources by practicing hunting from the pre-historic times. The domestication of animals corroborated a closer relationship between humans and animals and increased the animal-source food product. Over the past several years, these relations have evolved to more intimacy having huge economic benefit as well as risks to humans. Animals are predominantly used for food, clothing, pets, guards, and research. The pathogens cross the barrier through direct and indirect contact and spread through ingestion, aerosol transmission, bites, and are vector-borne [22] (Figure 1.3 & 1.4). So, understanding the reservoirs, infection sources, and risk factors for disease transfer can often be challenging.



Figure 1.3: Goat infected with ORFV: figure illustrating ORFV infection in Black Bengal goat showing lesions around the lip recorded during the outbreak investigation.



Figure 1.4: Human Orf: Orf virus infection on the hand of a person working in a butcher shop in Indore district, Madhya Pradesh, India with an ulcerative lesion or nodule over the right hand.

➤ 1.1.5.1.1 Ingestion

Ingestion or oral communication for zoonotic bacterias like *Campylobacter* genus and *Salmonella* genus occurs when infected food, water, and unpasteurized milk, are ingested. For example, unpasteurized milk consumption causes the transmission of brucellosis [23, 24] and bovine tuberculosis to humans [25]. In Tanzania, improperly cooked meat consumption caused Q fever [26]. Food contaminated with fecal materials is regarded as the main route of transmission for *Salmonella* bacterium [27]. In developing countries, several zoonotic agents, like helminthes, are transmitted to humans after consumption of undercooked pork and beef [28]. The recent outbreak of the Ebola virus in West Africa is directly connected to wildlife as its epicenter. Even though the pathogen's natural reservoirs remain unknown, bats are strongly linked as potential reservoirs [29]. Wild animals are considered as a source of zoonotic pathogens and play an important role during zoonoses. Humans can bear the infection when close contact with the infected animal during bushmeat consumption [29].

Wild birds are considered as the main source of highly pathogenic avian influenza virus from which the domestic birds get an infection as they have efficient migratory abilities that help them to introduce the new epidemic within poultry [30]. After the establishment of human-to-human transmission, it rapidly across broad geographical regions and spread the emerging zoonotic diseases [31].

➤ **1.1.5.1.2 Aerosol**

Zoonotic diseases can be transmitted through air droplets, sneezing, and coughing. Anthrax and Avian influenza-like pathogens can be transmitted through the air. Transmission could be possible when fluid and feces of infected animals contaminate soil and, dried, and inhaled as dust particles. Wool, hide, and skin of infected animals act as a source of anthrax spores. Inhalation directly from aerosolized air brings a common occupational hazard, where routine anthrax livestock vaccination is not practiced [32].

➤ **1.1.5.1.2 Vectors**

Vector-borne diseases are responsible for more than 17% of all infectious diseases, followed by 1 million deaths annually [33]. Tsetse fly acts as a vector for the transmission of zoonotic pathogens such as Trypanosoma species, which causes African sleeping sickness [34]. Another example includes the tabanid fly that transmits bacterium Francisella tularensis, which cause lethal pneumonia in human [35]. Rodents are primarily responsible for the Hantavirus transmission to humans [36]. Dengue, Japanese encephalitis virus, chikungunya that causes acute encephalitis in humans are mosquito-borne viruses, which are transmitted to humans through mosquito bites from infected animals [37].

➤ **Bites and scratches**

Rabies virus is considered one of the oldest viruses with zoonotic potential. This virus usually transmits to humans

through bites or scratches from several hosts like dogs, raccoons, bats, cats, and monkeys [38]. However, pet dogs are regarded as important source of rabies transmission to humans in Africa [39].

1.1.4.2 Food Safety

The One health approach includes multiple industry sectors such as Agriculture and Animal Husbandry and their economic importance. Food-borne illnesses' growing incidences should require public awareness of food safety and security, and sustainability in food production. Implementation of a large amount of pesticide havoc the residue of antibiotics, toxins, and hormones from animal-sourced foods followed by an increase of unaddressed health issues. In India, the agricultural sector (Livestock, Forestry & Logging, Fishing, and related farming) still employs 47% of its workforce and contributes to 16.8% of the national GDP. Foot and mouth disease virus (FMD), one of the devastating viruses, usually affects all cloven-footed animals, including cattle, goats, sheep, and pigs having high cross-species transmission potential, thus impacting food production. An outbreak of FMD in India had a tremendous impact on economic repercussions and brought a loss of Rs.20, 000 Crores annually (FAO). According to the 19th Livestock census (2019), the total poultry birds in farms/hatcheries in India's rural and urban areas were around 214 million layers, 282 million broilers, 5 million ducks, and 10 million other birds (Turkey, Emu, Ostrich, etc.). Avipoxvirus (APV) infection usually infect the skin and gastrointestinal tract of avian species and act as a major disease to the poultry industry which can cause severe economic losses in backyard poultry

farming by causing the reduction of egg-laying abilities in layers and stunted growth in young birds. Morbidity and mortality rates in fowlpox may be above 50% [40, 41]. Predominantly found in small ruminants such as sheep and goats, one of the significant sources of meat production in India, susceptible to infectious viral disease, such as Peste des petits ruminants (PPR) causes gastroenteritis, oculo-nasal discharge, and sometimes, pneumonia. The condition is characterized by up to 100% mortality. Altogether, these viral diseases required concrete measures to control these infections.

1.1.4.3 Companion animal

Companion animal is considered as pet those lives in human surroundings with household care. Pet animals are usually treated as family members in several developed countries like United Kingdom [42]. In the USA, dogs have gained the highest popularity among all pet animals, whereas in Europe, the cat contributes significant portions of all pets [43]. A country like China, where animals were banned as pets till 1992, has recently increased pet ownership spontaneously in few cities. During the interaction between animals live with humans, they can exchange their emotions and benefit one another. For the past few years, dogs and cats spend their lives indoors with their owners having close physical contact. Several zoonotic infectious diseases and resistant bacteria can be transferred through direct or indirect contact with these species [44, 45]. Pet animals shared a familiar surrounding with their owners. They were usually responsible for food contamination, the transmission of infectious disease, and environmental contamination [46].

1.2 Microsatellite marker

1.2.1 Microsatellite markers in animals

Litt and Luty first coined the microsatellite term during the study of (TG) repeat within cardiac actin [47]. These repeats discovered during this study were responsible for neurological diseases in humans, and later, their utilization in the different molecular fields made them unique molecular markers. Microsatellites, otherwise known as simple sequence repeats (SSRs), are defined as tandemly repeated genomic sequences, predominantly dominated within the genomes having increased allele polymorphism levels. These are considered as codominant molecular markers with limited size and are amplifiable easily by the polymerase chain reaction (PCR). These unique features make them versatile tags followed by their utility in various applied and fundamental biology and medicine, comprising epidemiology, forensics science, population and conservation genetics, molecular mapping, and complex traits genetic analysis. Within the genome, SSRs are attributed to maintaining DNA structure, chromatin modification, gene expression, recombination, and cell cycle regulation.

The occurrence of microsatellite markers in eukaryotic genomes has been illustrated from the 1970s [48]. Since then, its presence and distribution within yeast to vertebrate's genome demonstrated [49]. Using the hybridization technique, different microsatellite sequences within various organisms evaluated utilizing the genomic DNA [50]. The classified SSRs such as mono to hexanucleotide well-investigated throughout the eukaryotes suggest that poly mononucleotide SSR tracts in primates most frequent classes of SSRs [51].

The comparative analysis of other repeats illustrated that dinucleotides 1.5-fold higher than tri-, tetra-, penta- and hexanucleotides vertebrates [52]. Within the human genome, SSR was exhibited once every 6 kb, and one CA repeat type every 30 kb [51]. However, in Japanese pufferfish, one SSR was observed in every 1.87 kilobases (kb) of DNA, and a CA repeat within every 6.56 kb of DNA.

SSRs ubiquitously distributed in protein-encoding and noncoding DNA, with a higher percentage in noncoding regions than other regions [53]. The lower frequency of microsatellites within coding regions could elaborate as the adverse selection act against the frameshift mutations within the coding region [54]. In invertebrates, the frequency of di and tetra-nucleotide motifs are 42- and 30-fold less within the exons region compared with intronic and intergenic regions, respectively [52]. The G+C rich motifs, such as CCG and CAG, are the most common between trinucleotides repeat motifs. In humans, the expansion of trinucleotides repeats coding for polyglutamine (CAG)_n, polyarginine (CGG)_n and polyproline (CCG)_n, and polyalanine (GCC)_n tracts within exons of the human genome form several neurodegenerative and neuromuscular diseases, like myotonic dystrophy, spinocerebellar ataxia, Huntington's disease, and fragile X syndrome [55]. The critical feature of microsatellites as a molecular marker is their hypermutability. The estimated rate of SSR mutation is at 10^{-2} – 10^{-6} per locus per generation [56], greater than the mutation rate of regular nonrepetitive DNA [57], which leads to hypervariability in species and populations. The slippage of the DNA polymerase enzyme and unequal recombination act as a major source of microsatellite generation and its evolution within the genome

[58, 59]. Dinucleotide motifs are prone to recombination due to their high affinity towards recombinase enzymes [60]. Like GT, CA, CT, GA, and others, few SSR sequences may directly influence recombination by modifying the genome organization [60]. Microsatellite control replication in rat cells, where DNA amplification is prohibited by a $d(GA)_{27} \times d(TC)_{27}$ tract. This sequence is represented at the end of an amplicon by forming a loop, which acts as a stop signal for DNA polymerase [61]. Microsatellites are considered a codominant molecular marker as a progeny inherits one allele from each parent through recombination. This pattern of inheritance can be utilized for paternity testing, construction of linkage map and Marker Assisted Selection (MAS) in agriculturally important animals, like cattle [62], pig [63], sheep [64], chicken [65], and fish [66] (Figure 1.5) Due to the small size and relative stability in degraded DNA, a panel of multiple SSR loci widely used in forensic science [67]. Microsatellite markers are highly sensitive compare with allozymes to evaluate the dynamics of populations, demographic bottlenecks [68, 69, 70], effective population sizes [71, 72, 73], and conservation genetics [74]. Polymorphic SSR marker determines the genetic diversity by fluctuating heterozygosity and allelic diversity of a particular population [75].

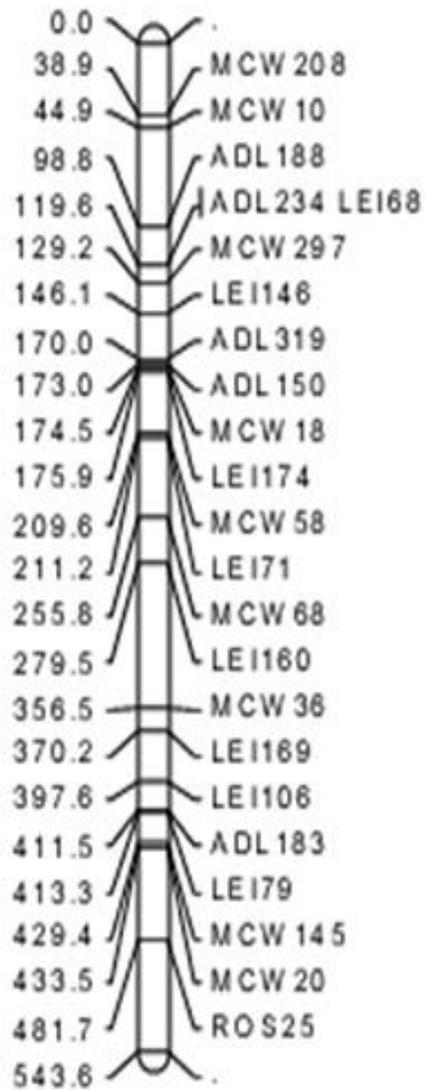


Figure 1.5: Linkage map of chicken. Diagram illustrating a chromosomal linkage map of chicken, where the left side represents the distance of SSR markers in centimorgan and the right side representing the name of microsatellite markers.

1.2. Microsatellite markers in plants

Microsatellites represented within the plant genome with high polymorphism ubiquitous distribution across the genome makes them one of the versatile molecular markers for plant genetic improvement programs [76, 77]. However, several other characters like low cost, high-throughput, reproducibility, and cross-species transfer made them a marker of choice [78]. (AT)_n motif is the most abundant SSR type in plants, whereas (AC)_n motif is the most abundant in the human genome. The human genome size (3.0×10^9 bp) is ~6.6 times larger than rice's genome size (0.45×10^9 bp) [79]. The presence of one (AC)_n motif nearly every 360–450 kb in rice genome, compared with one every 40–80 kb in the human genome. Similarly, there is one (GA)_n motif in every 225–330 kb of rice genome [80,81]. The trinucleotide motif (CGG)_n has been widespread in rice and is interspersed throughout the genome [82]. A quantitative trait is a character that influences several measurable phenotypic variations related to genetic and environmental influences. Generally, quantitative traits are pleiotropic effects governed by a number of polymorphic genes and environmental factors. One of many quantitative trait loci (QTLs) that contributes to a trait or a phenotype. For this purpose, a high-density genetic map saturated by microsatellite tags was constructed to detect QTL. The distance between two markers usually calculated as centimorgan (Figure 1.6).

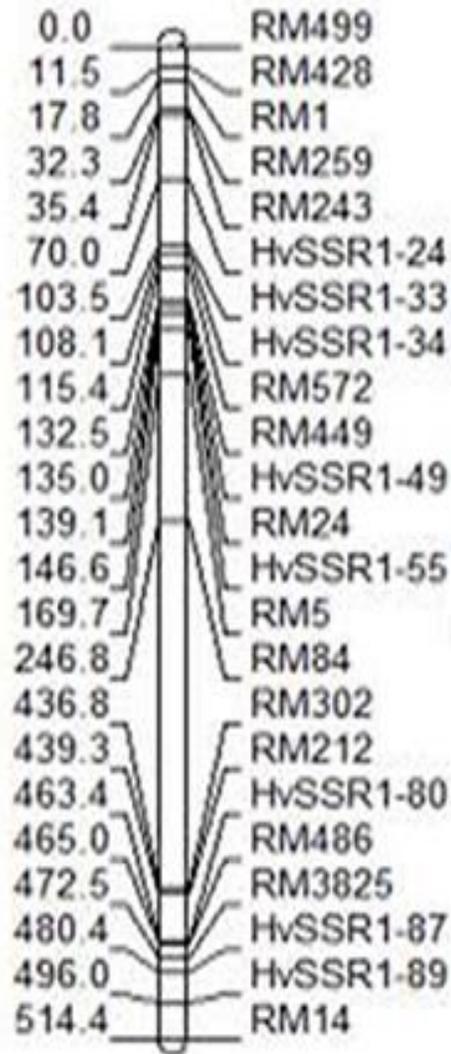


Figure 1.6: Linkage map of rice. Diagram illustrating a chromosomal linkage map of rice, where the left side represents the distance of SSR markers in centimorgan and the right side representing the name of microsatellite markers.

1.2.3 Microsatellite markers in bacteria

Microsatellites are of 1–6 bp units tandemly repeated DNA sequences and acquired up to 5% of the total genome [83, 84]. These are classified into various types according to the presence of repeated sequence, like (i) perfect microsatellite, with no interruption, e.g., (AG)₁₈ (ii) Imperfect microsatellite, with interruption by different nucleotides, e.g.,(TC)₁₂GC(TC)₈, and (iii) composite microsatellite, having two or more different motifs in tandem, e.g., (GA)₈(TA)₅. Many bacteria like the *Lactobacillus* genome riched with imperfect compound SSRs within the coding regions. Microsatellite markers were observed predominantly in prokaryotic genomes and used to investigate population genetics due to their polymorphism, reproductivity, and codominance nature [85, 86]. A long stretch of tetranucleotide abundantly present in the genome of Hi strain of *Haemophilus influenza* (RdKW20), which regulates the genes responsible for commensal and virulence behavior [87]. In *Neisseria* site-specific recombination, homologous recombination, genome duplications increase the rate of polymorphism and control the gene expression and regulate phase variations, promoter activity, and gene transcription in different bacterial genomes [88, 89]. The microsatellite markers present within *Helicobacter pylori*'s genome change its length through replication slippage followed by alternating the gene's specific function [90]. Unit mutations stimulate the formation of Fimbriae in *Haemophilus influenza* in SSR sequences present within the promoter spacing [91].

1.2.4 Microsatellite markers in viruses

Microsatellites acted as important genomic resources in viruses and used to discriminate several viruses, like cytomegalovirus (CMV) [92, 93], white spot syndrome virus (WSSV) [94, 95, 96], Herpes Simplex virus [97], Adenovirus [98], Ostreid herpesvirus [99, 100], Marek's disease virus [101], and Spodopteralittoralis multiple nucleopolyhedrovirus (SpliMNPV) [102] due to its polymorphic nature. The functional evaluation of microsatellite repeats has been highlighted in menovirus [103], vesicular stomatitis virus [104], hepatitis C virus [105], and human respiratory syncytial virus (RSV) [106]. In human papillomavirus (HPV) [107] and Herpesviruses [108], higher relative abundance (RA) and relative density (RD) variation was observed. Regression analysis suggested that the number of both SSRs and cSSRs, RA of cSSRs, was dependent on genome size but independent of GC concentration, which was similar to that of HPV [107] but the opposite of human immunodeficiency viruses (HIV) [109], Potexvirus [110] Carlavirus [112] and Tobamovirus [111]. The cSSRs percentages of Orf virus ranged from 7.9-9.0%, which is lower compared with HIV-1, 0-24.2% (Geminivirus, 0-27.2%, Herpesvirus, 8.1-33.3%) (Figure 1.7 & 1.8). Generally, the incidence of cSSRs decreases with an increase in complexity [112]. Microsatellites present within the genic region provide adaptive variation and virulence to avian influenza virus and encephalo-myocarditis virus [113, 114]. Recently, a microsatellite present in HSV-1 glycoprotein coding region US4 was utilized for strain demarcation [115]. A single mononucleotide repeat was able to decipher the dynamics of transmission of human adenovirus [98].

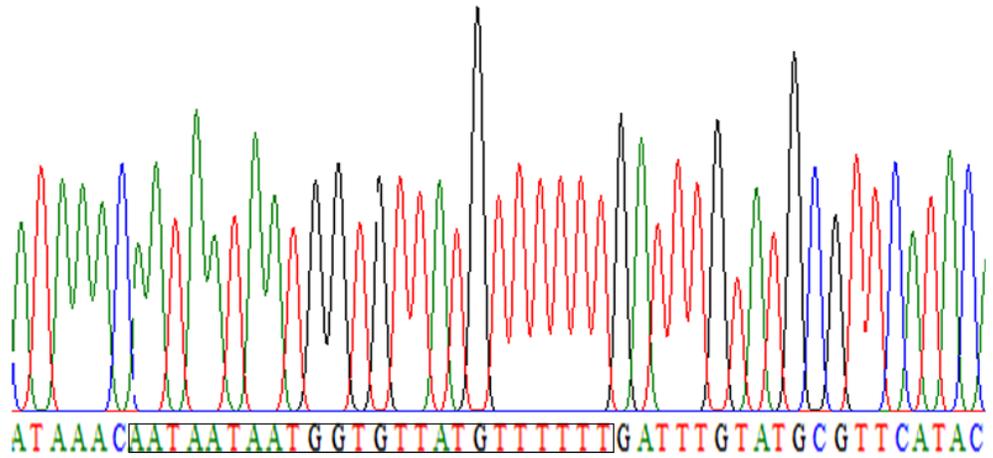


Figure 1.7: Microsatellite in a virus. Diagram illustrating electropherogram of a compound microsatellite with motif composition $(AAT)_{3-x9}-(T)_6$ within Orf virus.



Figure 1.8: Polymorphism within microsatellite. The figure depicts $(GGC)_3$ microsatellite polymorphism within the ankyrin protein gene of the Orf virus.

1.3 Evolutions of sequencing strategies

1.3.1 First-generation sequencers

Sanger's and Maxam-Gilbert's DNA sequencing strategy were regarded as the first-generation sequencing methods from the

beginning of the sequencing era [116,117]. Sanger's method sequencing technique utilized chain-terminating inhibitors, while Maxam-Gilbert's strategy utilized the chemical modification of nitrogenous bases and cleavage of the nucleotide backbone [118]. Sanger's DNA sequencing method was considered immensely popular for its comfort and automation procedure. To date, Sanger's DNA sequencing method acts as the gold standard for specific applications [119]. The sequencing strategy procedure usually progresses by amplifying templates of DNA with fluorescently labeled chain-terminating nucleotides and capillary electrophoresis. The camera present within the sequencer reads the signals and provides the sequence information up to 800 bp. During larger genome sequencing, this sequencing strategy was not suitable due to its low throughput, higher cost, time, and labor intensity. So, it would not be preferable for metagenomic applications for mixed clinical samples. Thus, next-generation sequencing technology became comparatively much faster than classical techniques and led to discovering several viruses and their genotype demarcation [120, 121].

1.3.2 Second-generation sequencers

The second-generation sequencing, otherwise known as Next Generation Sequencing (NGS) overcame the Sanger sequencers' technological constraints. These NGS technologies were developed with many innovative technologies advanced sequencing chemistry, microfluidics, Bioinformatics, and imaging technologies [122]. There were three novel second-generation sequencers named Roche, Illumina, and Life Technologies used in metagenomic studies. These NGS platforms can generate only short sequence reads

in a massively parallel manner with millions to billions of reads during a single run. The data usually developed a gigabase (GB) of nucleotide sequence. These technologies are otherwise known as massive parallel sequencing strategies, but depending upon their sequencing chemistry and difference in data output, every platform is suitable for certain specific applications. Thus, the NGS platform is carefully chosen as per requirements. During virus discovery, NGS platforms producing more extended sequence read preferred over the others. Longer sequences are useful for denovo assembly and produce supercontigs to increase the possibility of finding out related sequences in GenBank. Conversely, to determine mutation and quasi-species of the virus, analysis of variants, a platform providing quality reads with less error, and higher depth became the choice over longer read lengths. Recently, Illumina sequencing technology has got popularity worldwide. It utilized sequencing by synthesis (SBS) chemistry with a higher sequence data yield with a low cost per base [123, 124, 125, 126]. Another sequencing platform named Roche 454 amplifies DNA by emulsion PCR, generating DNA clones using a single template and sequencing through sequencing-by-synthesis chemistry. This technology can produce long reads with a high error rate in homopolymers (127, 128). The NGS reads have higher error rates than generally Sanger sequencers, requiring sophisticated bioinformatics tools and statistical measures prior to data processing and assembly [129]. To increase ease of application, several benchtops forms different NGS platforms were recently developed and named 454 GS Junior (Roche) and MiSeq (Illumina) [130]. Apart from these two, the third type of NGS platform from Life Technologies known as the SOLiD is commercially

available, having a complexity of data processing and assembly compared to others [131]. The sequencing chemistry utilized sequencing by ligation. However, SOLiD is considered a reliable NGS strategy for the de-novo assembly of mammalian genomes [131].

1.3.3 Third-generation sequencers

Recently NGS technology evolves as "third-generation" sequencing, which produces longer reads measured in kilobases than other sequencing platforms, which have relatively short reads. The read length was produced during the third generation. It includes the Ion Torrent (Life Technologies), Single-Molecule Real-Time technology SMRT (Pacific Biosciences), and the Nanopore sequencing technology (Oxford Nanopore Technologies). These sequencers are differed from the rest by two main features: (1) template amplification is not required before sequencing and cost (2) the signal is recorded in real-time, i.e., at the time of enzymatic action. The Ion Torrent sequencing chemistry is based on pH-dependent semiconductors and does not need fluorescence or chemiluminescence image scanning, resulting development of a cost and time effective sequencing platform. During nucleotide incorporation into the DNA, a proton is released, which causes a change in voltage and subsequently is detected by the microfluidic chip [130, 132].

Except for Ion Torrent, the rest of the third-generation sequencing platforms are relatively recent and still need rigorous evaluation. The SMRT technology single can be sequenced the DNA fragments up to 7 kb, having of 3-4 kb

average read length [133, 134]. The sequencing chemistry of SMRT cells includes Zero-Mode waveguides incorporated with a single set of enzymes and templates. The enzyme inserts a nucleotide into the complementary strand during their action, cleaving off fluorescent dye linked with the nucleotide, and this fluorescent signal is captured.

A nanopore is a tiny biopore with a diameter in nanoscale and involves a heptameric transmembrane channel α - (α HL) derived from *Staphylococcus aureus*. This protein can bear high voltage and current (up to 100 mV, 100 pA). When a DNA fragment is crossed through the channel of haemolysin, the current is varied depending upon the size difference between deoxyribonucleoside monophosphate (SNMP). The contrast of current is observed by electrophysiological techniques, followed by identification dNMP. The Nanopore sequencing chemistry is devoid of fluorescence/chemiluminescence and less sensitive to temperature and other conditions.

Among the various NGS technology available today, selecting the most appropriate depends upon the project goal followed by their correct application. During a metagenomic project, the unavailability of a reference genome leads to a de-novo assembly for virus discovery. Such an assembly requires a high throughput computational tool with longer reads with high coverage. When the reference genome is available, the short reads could be implemented for metagenome analysis [129]. The Illumina technology is the most widely used platform for virus discovery through de novo assembly or mapping with reference genome due to its considerable read

length and lower error rate among the entire NGS platform available to date.

1.4 Aim of the thesis

The present study aims to achieve three broad "One Health" objectives by studying the emergence of pathogens associated with the following scenarios.

1. Zoonotic viral diseases

2. Food security

3. Companion animal

1. Zoonotic viral diseases

1.1 Molecular characterization, comparative and evolutionary analysis of the recent Orf outbreaks among goats in the Eastern part of India (Odisha).

1.2 Development of microsatellite markers within the Orf virus to decipher viral strain.

1.3 Recombination may drive the emergence of Orf virus diversity: Evidence from the first complete genome of Indian Orf virus and comparative genomic analysis.

2. Food security

2.1. The emergence of subclades A1 and A3 Avipoxviruses in India.

2.2. Genome-wide identification and characterization of microsatellite markers within the Avipox virus.

2.3. Complete genome analysis of Indian fowlpox virus utilizing NGS platform.

3. Companion animal

3.1. Complete genome analysis of Indian pigeonpox virus.

3.2. Comprehensive analysis of simple sequence repeats in Picornaviruses.

1.5 Organization of the thesis

The entire thesis is broken down into eleven different chapters. The organization of the chapters is discussed below

Chapter 1: Review of literature regarding infectious viral diseases, zoonosis, and one health approach.

This chapter describes a detailed review of the literature regarding the zoonotic potential of several viruses. This portion also elaborates on one health's niches and its importance to control infectious disease, food security, companion animal, and maintaining a healthy environment. Additionally, we give insight into the epidemiological surveillance of virus outbreak location, implementation of molecular markers for viral strain demarcation, and utilization of next-generation sequencing strategies to decipher the complete genome of viruses.

Chapter 2: This includes the details of the materials and methods used during the present study. It includes a brief description of the major techniques involved in the findings.

Chapter 3: Molecular characterization, comparative and evolutionary analysis of the recent Orf outbreaks among goats in the Eastern part of India (Odisha). This chapter elucidated the epidemiological characteristics of a zoonotic virus, Orf (ORFV), the causative agent of Contagious ecthyma

disease that primarily occurs within small ruminants and animal handlers. We surveyed the fifteen outbreak locations of eastern India during this study and collected clinical samples from infected goats. Total DNA was extracted and screened with the universal PCR primers to confirm the respective virus. Further amplification using gene-specific primers targeting fragments of ORFV011, ORFV020, ORFV059, ORFV108 genes, followed by sequencing and GenBank submission to get accession numbers. The comparative genomics approach was deployed to construct the gene-specific phylogenetic tree and evolutionary parameters of this virus, suggesting that the viral strain is closely related to Chinese isolates. The evolutionary study revealed the purifying selection is maintaining the heterogeneity within the viral strain.

Chapter 4: Development of microsatellite markers within the Orf virus to decipher viral strain. Genome-wide/in-silico investigation of microsatellites markers within ORFV genome, the etiological agent of contagious ecthyma disease, has been accomplished to decipher the type, distribution, and its potential role in the genome evolution, during this study. We have scrutinized eleven ORFV strains during this study, which showed the existence of 1,036–1,181 microsatellites per strain. Further analysis deciphered residence of 83–107 compound SSRs (cSSRs) per genome. Finally, in-silico polymorphism was evaluated, followed by in-vitro validation utilizing PCR and sequencing analysis. The thirteen polymorphic microsatellite markers developed during this investigation were characterized by mapping with the GenBank database. The outcome of this study suggests that these microsatellites could be a tool for strain demarcation,

genetic diversity estimation, and evolutionary analysis of the virus.

Chapter 5: Recombination may drive the emergence of Orf virus diversity: Evidence from the first complete genome of Indian Orf virus and comparative genomic analysis. We have explored the 2017 ORFV outbreak within goats in Madhya Pradesh, a central India state, during this study. Phylogenetic analysis deciphered the transboundary perspective of the virus by indicating its correlation with a distinct geographical location. We elucidated this viral strain's complete genome, named Ind/MP, and carry out a comparative genomic analysis. The Ind/MP complete genome was having 139,807 bp with GC content 63.7%. The genome was made up of 132 open reading frames (ORFs) flanked by inverted terminal repeats (ITRs) of 3,910 bp at both ends with terminal BamHI sites and conserved telomere resolution sequences. A sum of forty recombination incidents was recognized. Ind/MP strain actively contributed to twenty-one events indicating this strain's potential for recombination to generate new variants.

Chapter 6: The emergence of subclades A1 and A3 avipoxviruses in India. In this chapter, we described the surveillance report of outbreaks of avipoxvirus (APV) infected domestic chickens and pigeons belongs to the eastern Indian state of Odisha during the years 2010–2018. PCR amplicons and phylogenetic analysis using the viral P4b core protein and the DNA polymerase gene confirmed the APV strains and suggested the introduction of A1 and A3 subclades in Indian chickens and pigeons, respectively. This study reveals APV's presence in eastern Indian birds (Odisha) and the polymerase

gene's utilization for the first time to elucidate the spreading clades of Indian APVs.

Chapter 7: Genome-wide identification and characterization of microsatellite markers within the Avipox virus. In this chapter, we conducted a genome-wide analysis to decipher the type; distribution pattern of eight complete genomes derived from the Avipox virus genus, the causative agent of pox like lesions above 300 avian species and one of the major diseases for the extinction of endangered avian species as well a severe economic threat to livelihood. The in-silico screening deciphered the existence of 1531-2473 SSRs per strain. In the case of compound SSRs (cSSRs), the value opted 83-107 per genome. Finally, we have validated forty sets of compound microsatellite markers through the in-vitro approach utilizing clinical specimens, followed by mapping the sequencing product with the database through comparative genomics approaches.

Chapter 8: Complete genome analysis of Indian fowlpox virus utilizing NGS platform. For the first time, we have deciphered the complete genome sequence of a fowlpox virus of Indian origin through a next-generation sequencing platform and utilize a comparative genomics approach to solve more about this virus biology. The NextSeq 500 NGS generated a total of ~2.16 GB of quality data with 7,303,963 numbers of quality reads. The previously published complete genome isolate fowlpox virus (Acc. No- AF198100) was used as the reference genome, to which the sequence generated through NGS mapped and assembled. The assembled genome of this novel virus exhibited 260,066 bp in length with a lower percentage of GC content (28.5) in comparison to fowlpox

virus (30.83), shearwater pox virus (30.23), canarypox virus (30.37), and turkeypox virus (29.78)

Chapter 9: Complete genome analysis of Indian pigeonpox

virus. The emergence of poxviruses has a severe health hazard to domestic and free-ranging wild birds regardless of their age and sex due to its exorbitant contagious property, which influences both the economic and conservation aspects. The NextSeq 500 NGS generated a total of ~2.72 GB of quality data with 9,541,974 numbers of quality reads. The extreme left nucleotide was considered base 1, and the beginning region of the Inverted Terminal Repeats (ITRs), which consists of a total length of 4688 bp. The ITR spanned throughout PPV-001 to PPV-004a and PPV-256 to PPV-259. A number of recombination events were overlapped within intergenic (events 5, 7, 8, 9, 13, 16, and 20), hypothetical protein (events 1, 2, 3 17 and 18), transcription factor VLTF (events 6 and 12), ankyrin repeat protein (events 15 and 21).

Chapter 10: Comprehensive analysis of simple sequence

repeats in Picornaviruses. Genome-wide identification of simple sequence repeats (SSRs) of picornaviruses was carried out to investigate type, distribution, and potential role in genome evolution. Investigation on 88 picornavirus species revealed the presence of 2,488 SSRs and 100 compound SSRs. The relative abundance and relative density of SSR varied between 1.953 bp/kb-5.763 bp/kb and 13.39 bp/kb-45.02 bp/kb, while that of cSSR ranged from 0.108 bp/kb-0.636 bp/kb and 1.36 bp/kb-26.84 bp/kb. Regression analysis revealed a significant correlation of genome size and GC content with the incidence of SSRs. Duplication pattern of motifs, like (C)-x-(C), (TG)-x-(TG), etc., and self-

complementary motifs, such as (GC)-x-(CG), (TC)-x-(AG), etc. were observed in cSSR. Polymorphism analysis revealed that most of the cSSR were prone to instability, followed by consensus motifs. Finally, recombination analysis revealed that the breakpoints were rich in dinucleotide repeat, especially GT.

Chapter 11: Concluding remark and the future perspective. This chapter reflects the concluding remark and the future aspects of the thesis work.

1.6 Scope of the thesis

We investigated a handful of disease outbreaks through a molecular surveillance approach during the study. One of the essential aspects of the investigation was to attend to emerging neglected pathogens, which gave us valuable information regarding many endemic animal viruses prevalent across India. While working with Orf, fowlpox, and pegiopox virus, we deciphered the presence of molecular markers such as microsatellite and validated those using clinical samples. These molecular markers can be used to construct a multiplex panel, followed by genotyping to evaluate virus evolution. However, further molecular, evolutionary, and genetic analyses of larger sample size and the wider geographical area would clearly understand emerging viral pathogens patterns. Finally, we revealed the complete genome of several poxviruses that would help develop effective vaccine candidates and several other future research applications.

1.7 References

1. Ryu, W.S. (2016), Molecular virology of human pathogenic viruses. Academic Press.
2. Lefkowitz, E.J., Dempsey, D.M., Hendrickson, R.C., Orton, R.J., Siddell, S.G. and Smith, D.B. (2018), Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic acids research*. 46(D1): p. D708-D717 (DOI: 10.1093/nar/gkx932).
3. Martin, D.P., Murrell, B., Golden, M., Khoosal, A. and Muhire, B. (2015), RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus evolution*. p. 1(1) (DOI: 10.1093/ve/vev003).
4. Breman J. G., Henderson D. A. (2002), Diagnosis and management of smallpox. *New England Journal of Medicine*. 346(17): p. 1300-1308 (DOI: 10.1056/NEJMra020025).
5. Wunner W. H. (2007), Rabies virus. In *Rabies* (pp. 23-68). Academic Press
6. Simmonds P. and Aiewsakun P. (2018), Virus classification – where do you draw the line? *Archives of Virology*. 163(8): p. 2037-2046 (DOI: 10.1007/s00705-018-3938-z).
7. Brüssow H. (2009), The not so universal tree of life or the place of viruses in the living world. *Philos Trans R Soc Lond B Biol Sci*. 364(1527): p. 2263-74 (DOI: 10.1098/rstb.2009.0036).
8. Baltimore D. (1971), Expression of animal virus genomes. *Bacteriological reviews*. 35(3): p. 235.

9. Holland J. and Domingo E. (1998), Origin and Evolution of Viruses. *Virus Genes*. 16(1): p. 13-21 (DOI: 10.1023/A:1007989407305).
10. Luring A.S., Frydman J., and Andino R. (2013), The role of mutational robustness in RNA virus evolution. *Nature Reviews Microbiology*. 11(5): p. 327-336 (DOI: 10.1038/nrmicro3003).
11. Boya P., Pauleau A. L., Poncet D., Gonzalez-Polo R. A., Zamzami N., Kroemer G. (2004), viral proteins targeting mitochondria: controlling cell death. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*. 1659(2-3): p. 178-189 (DOI: 10.1016/j.bbabi.2004.08.007).
12. Payne S. (2017), Introduction to Animal Viruses. *Viruses*: p. 1-11 (DOI: 10.1016/B978-0-12-803109-4.00001-5).
13. Desfarges S. and Ciuffi A. (2012), Viral Integration and Consequences on Host Gene Expression. In *Viruses: Essential Agents of Life*, G. Witzany, Editor. Springer Netherlands. Dordrecht: p. 147-175
14. Scholthof K. B.G., Adkins S., Czosnek H., *et al.* (2011), Top 10 plant viruses in molecular plant pathology. *Molecular Plant Pathology*. 12(9): p. 938-954 (DOI: 10.1111/j.1364-3703.2011.00752.x).
15. Ahmad K. (2000), Malaysia culls pigs as Nipah virus strikes again. *Lancet*. 356 (9225): p. 230 (DOI: 10.1016/S0140-6736(05)74483-4).
16. Glennon E.E., Restif O., Sbarbaro S.R., *et al.* (2018), Domesticated animals as hosts of henipaviruses and filoviruses: A systematic review. *Vet J*. 233: p. 25-34 (DOI: 10.1016/j.tvjl.2017.12.024).

17. King L., Anderson L., Blackmore C. (2008), Executive summary of the AVMA one health initiative task force. *J Am Vet Assoc* 233: p. 259–260 (DOI: 10.2460/javma.233.2.259).
18. Atlas R.M. (2013), One Health: its origins and future. *Curr Top Microbiol Immunol.* 365: p. 1-13 (DOI: 10.1007/82_2012_223. PMID: 22527177).
19. Taylor L.H., Latham S.M., Woolhouse M. (2001), Risk factors for human disease emergence. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 356: p. 983–9 (DOI: 10.1098/rstb.2001.0888).
20. Kruse H., Kirkemo A.M., Handeland K. (2004), Wildlife as source of zoonotic infections. *Emerg. Infect. Dis.* 10: p. 2067–2072 (DOI: 10.3201/eid1012.040707).
21. Baum S. G. (2004), Zoonoses. *Current infectious disease reports.* 6(5): 343-344 (DOI: 10.1007/s11908-004-0029-y).
22. Woolhouse M.E., Taylor L.H., Haydon D.T. (2001), Population biology of multihost pathogens. *Science.* 292: 1109–1112 (DOI: 10.1126/science.1059026).
23. Corbel M.J. (2006), Brucellosis in humans and animals. *WHO-FAO-OIE:* p. 1-102 (DOI: 10.2105/AJPH.30.3.299).
24. Díaz-Aparicio E. (2013), Epidemiology of brucellosis in domestic animals caused by *Brucellamelitensis*, *Brucellasuis* and *Brucellaabortus*. *Rev. Sci. Tech.* 32, 43–51: p. 53–60 (DOI: 10.20506/RST.32.1.2187).
25. Kazwala R., Daborn C., Kusiluka L., Jiwa S., Sharp J., Kambarage D. (1998), Isolation of *Mycobacterium* species from raw milk of pastoral cattle of the Southern Highlands of Tanzania. *Trop Anim Heal. Prod.* 30: p. 233–9 (DOI: 10.1023/a:1005075112393).

26. Prabhu M., Nicholson W.L., Roche A.J., Kersh G.J., Fitzpatrick K.A., Oliver L.D., Massung R.F., Morrissey A.B., Bartlett J.A., Onyango J.J., Maro V.P., Kinabo G.D., Saganda W., Crump J.A. (2011), Q fever, spotted fever group, and typhus group rickettsioses among hospitalized febrile patients in northern Tanzania. *Clin. Infect. Dis.* 53: p. 8–15 (DOI: 10.1093/cid/cir411).
27. Thorns C.J. (2000), Bacterial food-borne zoonoses changing perspectives :20th and 21 st centuries *Revi. Sci. Tech. Off. Int. Epiz.* 19: p. 226–239 (DOI: 10.20506/rst.19.1.1219).
28. Gilman R.H., Gonzalez A.E., Llanos-Zavalaga F., Tsang V.C.W., Garcia H.H. (2012), Prevention and control of *Taeniasoliumtaeniasis/cysticercosis* in Peru. *athog. Glob. Health.* 106: p. 312–8 (DOI: 10.1179/2047773212Y.0000000045).
29. Calisher C.H., Childs J.E., Field H.E., Holmes K. V., Schountz T. (2006), Bats: important reservoir hosts of emerging viruses. *Clin. Microbiol. Rev.* 19: p. 531–545 (DOI: 10.1128/CMR.00017-06).
30. Parrish C.R., Murcia P.R., Holmes E.C. (2015), Influenza virus reservoirs and intermediate hosts: dogs, horses, and new possibilities for influenza virus exposure of humans. *J Virol.* 89: p. 2990–2994 (DOI: 10.1128/JVI.03146-14).
31. Liu Q., Cao L., Zhu X.Q. (2014), Major emerging and re-emerging zoonoses in China: a matter of global health and socioeconomic development for 1.3 billion. *Int. J. Infect. Dis.* 25: p. 65–72 (DOI: 10.1016/j.ijid.2014.04.003).

32. WHO (2008), Anthrax in humans and animals. Geneva: World Health Organization. 4th Edition. Available from: <http://www.ncbi.nlm.nih.gov/books/>.
33. WHO (2016a), Vector-borne diseases. <http://www.who.int/mediacentre/factsheets/fs387/en/>.
34. MacGregor P., Savill N.J., Hall D., Matthews K.R. (2011), Transmission stages dominate trypanosome within-host dynamics during chronic infections. *Cell Host Microbe*. 9: p. 310–318 (DOI: 10.1016/j.chom.2011.03.013).
35. Petersen J.M., Mead P.S., Schriefer M.E. (2009), *Francisella tularensis*: an arthropod-borne pathogen. *Vet. Res.* 40: p. 7 (DOI: 10.1051/vetres:2008045).
36. Chrispal A., Boorugu H., Gopinath K.G., Chandy S., Prakash J.A.J., Thomas E.M., Abraham A.M., Abraham O.C., Thomas K. (2010), Acute undifferentiated febrile illness in adult hospitalized patients: the disease spectrum and diagnostic predictors - an experience from a tertiary care hospital in south India. *Trop. Doct.* 40: p. 230–234. (DOI: 10.1258/td.2010.100132)
37. Fever H., Hospital T.G., Papakonstantinou I., Pagdatoglou K. (2014), Phylogenetic analysis of West Nile Virus genome. *Iran J. Microbiol.* 20: p. 1419–1421 (DOI: 10.3201/eid2008.131321).
38. Susilawathi N.M., Darwinata A.E., Dwija I.B., Budayanti N.S., Wirasandhi G.A., Subrata K., Susilarini N.K., Sudewi R.A., Wignall F.S., Mahardika G.N. (2012), Epidemiological and clinical features of human rabies cases in Bali 2008-2010. *BMC Infect. Dis.* 12: p. 81 (DOI: 10.1186/1471-2334-12-81).

39. Cleaveland S. (1998), Royal Society of Tropical Medicine and Hygiene meeting at Manson House, London, 20 March 1997. Epidemiology and control of rabies. The growing problem of rabies in Africa. *Trans R Soc Trop Med Hyg.* 92(2): p. 131 (DOI: 10.1016/s0035-9203(98)90718-0.).
40. Tripathy, D.N. (1993), Avipox viruses. *Virus infections of birds.* p. 1-15.
41. Skinner, M.A., 2008. Poxviridae. In *Poultry Diseases.* p. 333-339.
42. McNicholas J., Gilbey A., Rennie A., Ahmedzai S., Dono J.A., Ormerod E. (2005), Pet ownership and human health: A brief review of evidence and issues. *Br. Med. J.* 331: p. 1252–1254 (DOI: 10.1136/bmj.331.7527.1252).
43. AVMA (2018), *Pet Ownership and Demographics Source book 2017–2018.* Available online: www.avma.org/news/press-release/avma-releases-latest-stats-pet-ownership-and-veterinary-care.
44. Day M.J. (2010) One Health: The small animal dimension. *Vet. Rec.* 167: p. 847–849 (DOI: 10.1136/vr.c6492).
45. Baede V.O., Broens E.M., Spaninks M.P., Timmerman A.J., Graveland H., Wagenaar J.A., Duim, B., Hordijk J. (2017), Raw pet food as a risk factor for shedding of extended-spectrum beta-lactamase-producing Enterobacteriaceae in household cats. *PLoS ONE.* 12: p. e0187239 (DOI: 10.1371/journal.pone.0187239. eCollection 2017).

46. Schmidt P.L. (2009), Companion animals as sentinels for public health. *Vet. Clin.N. Am. Small Anim. Pract.* 39: p. 241–250 (DOI: 10.1016/j.cvsm.2008.10.010).
47. Litt M., Luty J.A. (1989), A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am J Hum Genet. Mar.* 44(3): p. (397-401).
48. Bruford M.W., Cheesman D.J., Coote T., Green H.A.A., Haines S. A., Oryan C. (1996), Microsatellites and their application to conservation genetics. In: Smith, T.B., Wayne, R.K. (Eds.), *Molecular Genetic Approaches in Conservation*. Oxford University Press, Oxford: p. (278–297).
49. Hamada H., Petrino M., Kakunaga T. (1982), A novel repeated element with Z-DNA-forming potential is widely found in evolutionary diverse eukaryotic genomes. *Proc. Natl. Acad. Sci. U. S. A.* 79: p. 6465–6469 (DOI: 10.1073/pnas.79.21.6465).
50. Tautz, D., Renz, M. (1984), Simple sequences are ubiquitous repetitive components of eukaryote genomes. *Nucleic Acids Res.* 17: p. 4127–4138 (DOI: 10.1093/nar/12.10.4127).
51. Beckmann J.S., Weber J.L. (1992), Survey of human and rat microsatellites. *Genomics* 12: p. 627–631 (DOI: 10.1016/0888-7543(92)90285-z).
52. Toth G., Gaspari Z., Jurka J. (2000), Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res.* 10: p. 967–981 (DOI: 10.1101/gr.10.7.967).
53. Metzgar D., Wills C. (2000), Evolutionary changes in mutation rates and spectra and their influence on the

- adaptation of pathogens. *Microbes Infect.* 2: p. 1513–1522 (DOI: 10.1016/s1286-4579(00)01306-x).
54. Li Y.C., Korol A.B., Fahima T., Nevo E. (2004), Microsatellites within genes: structure, function, and evolution. *Mol. Biol. Evol.* 21: p. 991–1007 (DOI: 10.1093/molbev/msh073).
55. Brown L.Y., Brown S.A. (2004), Alanine tracts: the expanding story of human illness and trinucleotide repeats. *Trends Genet.* 20: p. 51–58 (DOI: 10.1016/j.tig.2003.11.002).
56. Ellegren H. (2000), Microsatellite mutations in the germline: implications for evolutionary inference. *Trends Genet.* 16: p. 551–558 (DOI: 10.1016/s0168-9525(00)02139-9).
57. Li W.H. (1997), *Molecular Evolution*. Sinauer Associates, Sunderland, MA: p. 177–213.
58. Schlötterer C., Tautz D. (1992), Slippage synthesis of simple sequence DNA. *Nucleic Acids Res.* 20: p. 211–215 (DOI: 10.1093/nar/20.2.211).
59. Richards R.I., Sutherland G.R. (1994), Simple repeat DNA is not replicated simply. *Nat. Genet.* 6: p. 114–116 (DOI: 10.1038/ng0294-114).
60. Biet E., Sun J., Dutreix M. (1999), Conserved sequence preference in DNA binding among recombinant proteins: abnormal effect of ssDNA secondary structure. *Nucleic Acids Res.* 27: 596–600 (DOI:10.1093/nar/27.2.596).
61. Li Y.C., Korol A.B., Fahima T., Beiles A., Nevo E. (2002), Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Mol.*

- Ecol. 11: p. 2453–2465 (DOI: 10.1046/j.1365-294x.2002.01643.x).
62. Maillard J.C., Berthier D., Chantal I., Thevenon S., Sidibe I., Stachurski F., Belemsaga D., Razafindraibe H., Elsen J.M. (2003), Selection assisted by a BoLA-DR/DQ haplotype against susceptibility to bovine dermatophilosis. *Genet. Sel. Evol.* 35: p. S193–S200 (DOI: 10.1186/1297-9686-35-S1-S193).
63. Rothschild M.F. (2003), From a sow's ear to a silk purse: real progress in porcine genomics. *Cytogenet. Genome Res.* 102. p. 95–99 (DOI: 10.1159/000075732).
64. Notter D.R, Cockett N.E. (2005), Opportunities for detection and use of QTL influencing seasonal reproduction in sheep: a review. *Genet Sel. Evol.* 37: p. 39–53 (DOI: 10.1186/1297-9686-37-S1-S39).
65. Malek M., Lamont S.J. (2003), Association of INOS, TRAIL, TGFbeta2, TGF-beta3, and IgL genes with response to *Salmonella enteritidis* in poultry. *Genet. Sel. Evol.* 35: p. S99–S111 (DOI: 10.1186/1297-9686-35-s1-s99).
66. Sahoo L., Patel A., Sahu B.P., Mitra S., Meher P.K., Mahapatra K.D., Dash S.K., Jayasankar P., Das P. (2015), Preliminary genetic linkage map of Indian major carp, *Labeorohita* (Hamilton 1822) based on microsatellite markers. *Journal of genetics.* 94(2): p. 271-277 (DOI: 10.1007/s12041-015-0528-7).
67. Schneider P.M., Bender K., Mayr W.R., Parson W., Hoste B., Decorte R. (2004), STR analysis of artificially degraded DNA results of a collaborative

- European exercise. *Forensic Sci. Int.* 139: 123–134 (DOI: 10.1016/j.forsciint.2003.10.002).
68. Spencer C.C., Neigel J.E., Leberg P.L. (2000), Experimental evaluation of the usefulness of microsatellite DNA for detecting demographic bottlenecks. *Mol. Ecol.* 9: p. 1517–1528 (DOI: 10.1046/j.1365-294x.2000.01031.x).
69. Guinand B., Scribner K.T. (2003), Evaluation of methodology for detection of genetic bottlenecks: inferences from temporally replicated lake trout populations. *C.R. Biol.* 326: p. S61–S67 (DOI: 10.1016/s1631-0691(03)00039-8).
70. Ramstad K.M., Woody C.A., Sage G.K., Allendorf F.W. (2004), Founding events influence genetic population structure of sockeye salmon (*Oncorhynchus nerka*) in Lake Clark, Alaska. *Mol. Ecol.* 13: 277–290 (DOI:10.1046/j.1365-294x.2003.2062.x).
71. Gold J.R., Burrige C.P., Turner T.F. (2001), A modified steppingstone model of population structure in red drum, *Sciaenops ocellatus* (Sciaenidae), from the northern Gulf of Mexico. *Genetica.* 111: p. 305–317 (doi: 10.1023/a:1013705230346).
72. Bérubé M., Urban J., Dizon A.E., Brownell R.L., Palsboll P.J. (2002), Genetic identification of a small and highly isolated population of a fin whales (*Balaenoptera physalus*) in the Sea of Cortez, Mexico. *Conserv. Genet.* 3: p. 183–190 (DOI:10.1023/a:1015224730394).
73. Waples R.S. (2002), Effective size of fluctuating salmon populations. *Genetics.* 161: p. 783–791.

74. Zierdt H., Hummel S., Herrmann B. (1996), Amplification of human short tandem repeats from medieval teeth and bone samples. *Hum. Biol.* 68: p. 185–199.
75. Pujolar J.M., Maes G.E., Vancoillie C., Volckaert F.A.M. (2005), Growth rate correlates to individual heterozygosity in European eel, *Anguilla anguilla* L. *Evolution*. 59: p. 189–199 (DOI:10.1111/j.0014-3820.2005.tb00905.x).
76. Morgante M., Hanafey H., Powell W. (2002), Microsatellites are preferentially associated with nonrepetitive DNA in plant genome. *Nat. Genet.* 30: p. 194–200 (DOI: 10.1038/ng822. Epub 2002 Jan 22).
77. Wright J.M., Bentzen P. (1994), Microsatellites: Genetic markers for the future. *Rev. Fish Biol. Fish.* 4: p. 384–388.
78. Gupta P.K., Varshney R.K., Sharma P.C., Ramesh B. (1999), Molecular markers and their applications in wheat breeding. *Plant Breed.* 118: p. 369–390 (DOI: 10.1046/j.1439-0523.1999.00401.x).
79. Arumunagathan K., Earle E.D. (1991), Nuclear DNA content of some important plant species. *Plant Mol. Biol.* 9: p. 208–218 (DOI: 10.1007/BF02672069).
80. Panaud O., Chen X., McCouch S.R. (1995), Frequency of microsatellite sequences in rice (*Oryza sativa* L.) *Genome*. 38: p. 1170–1176 (DOI: 10.1139/g95-155).
81. Wu K.S., Tanksley S.D. (1993), Abundance, polymorphism and genetic mapping of microsatellites in rice. *Mol. Gen. Genet.* 241: p. 225–235 (DOI: 10.1007/BF00280220).

82. Zhao X., Kochert, G. (1992), Characterization and genetic mapping of a short, highly repeated, interspersed DNA sequence from rice (*Oryza sativa* L.), *Mol. Gen. Genet.* 231: p. 353–359 (DOI: 10.1007/BF00292702).
83. Ussery D.W., Binnewies T.T., Gouveia-Oliveira R., Jarmer H. and Hallin P.F. (2004), Genome update: DNA repeats in bacterial genomes. *Microbiology.* 150(11): p. 3519-3521 (DOI: 10.1099/mic.0.27628-0).
84. Wheeler G.L., Dorman H.E., Buchanan A., Challagundla L., Wallace L.E. (2014), A review of the prevalence, utility, and caveats of using chloroplast simple sequence repeats for studies of plant biology. *Applications in plant sciences.* 2(12): p. 1400059 (DOI: 10.3732/apps.1400059).
85. Field D., Wills C. (1998), Abundant microsatellite polymorphism in *Saccharomyces cerevisiae*, and the different distributions of microsatellites in eight prokaryotes and *S. cerevisiae*, result from strong mutation pressures and a variety of selective forces. *Proceedings of the National Academy of Sciences.* 95(4): p. 1647-1652 (DOI: 10.1073/pnas.95.4.1647)
86. Schlötterer C. (2000), Evolutionary dynamics of microsatellite DNA. *Chromosoma.* 109(6): p. 365-371 (DOI: 10.1007/s004120000089).
87. Hood D.W., Deadman M.E., Jennings M.P., Bisercic M., Fleischmann R.D., Venter J.C. Moxon E.R. (1996), DNA repeats identify novel virulence genes in *Haemophilus influenzae*. *Proceedings of the National Academy of Sciences.* 93(20): p. 11121-11125 (DOI: 10.1073/pnas.93.20.11121).

88. Dawid S., Barenkamp S.J., Geme J.W.S. (1999), Variation in expression of the *Haemophilus influenzae* HMW adhesins: a prokaryotic system reminiscent of eukaryotes. *Proceedings of the National Academy of Sciences*. 96(3): p. 1077-1082 (DOI: 10.1073/pnas.96.3.1077).
89. Martin P., Makepeace K., Hill S.A., Hood D.W., Moxon E.R. (2005), Microsatellite instability regulates transcription factor binding and gene expression. *Proceedings of the National Academy of Sciences*. 102(10): p. 3800-3804 (DOI: 10.1073/pnas.040680510).
90. Tomb J.F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G., Fleischmann R.D., Ketchum K.A., Klenk H.P., Gill S., Dougherty B.A., Nelson K. (1997), The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature*. 388(6642): p. 539-547 (DOI: 10.1038/41483).
91. Moxon E.R., Rainey P.B., Nowak M.A., Lenski R.E. (1994), Adaptive evolution of highly mutable loci in pathogenic bacteria. *Current biology*. 4(1): p. 24-33 (DOI: 10.1016/s0960-9822(00)00005-1).
92. Davis C.L., Field D., Metzgar D., Saiz R., Morin P.A., Smith I.L., Spector S.A., Wills C. (1999), Numerous length polymorphisms at short tandem repeats in human cytomegalovirus. *Journal of virology*. 73(8); p. 6265-6270 (DOI: 10.1128/JVI.73.8.6265-6270.1999).
93. Walker A., Petheram S.J., Ballard L., Murph J.R., Demmler G.J., Bale J.F. (2001), Characterization of human cytomegalovirus strains by analysis of short tandem repeat polymorphisms. *Journal of clinical*

- microbiology. 39(6): p. 2219-2226 (DOI: 10.1128/JVI.73.8.6265-6270.1999).
94. Hoa T.T., Hodgson R.A., Oanh D.T., Phuong N.T., Preston N.J., Walker, P.J. (2005), Genotypic variations in tandem repeat DNA segments between ribonucleotidoreductase subunit genes of white spot syndrome virus (WSSV) isolates from Vietnam. *Diseases in Asian Aquaculture V.* p. 339-351.
95. Kiatpathomchai W., Taweetungtragoon A., Jittivadhana K., Wongteerasupaya C., Boonsaeng V., Flegel T.W. (2005), Target for standard Thai PCR assay identical in 12 white spot syndrome virus (WSSV) types that differ in DNA multiple repeat length. *Journal of virological methods.* 130(1-2): p. 79-82 (DOI: 10.1016/j.jviromet.2005.06.006).
96. Pradeep B., Shekar M., Karunasagar I., Karunasagar I. (2008), Characterization of variable genomic regions of Indian white spot syndrome virus. *Virology.* 376(1): p. 24-30 (DOI: 10.1016/j.virol.2008.02.037).
97. Szpara M.L., Gatherer D., Ochoa A., Greenbaum B., Dolan A., Bowden R.J., Enquist L.W., Legendre M., Davison A.J. (2014), Evolution and diversity in human herpes simplex virus genomes. *Journal of virology.* 88(2): p. 1209-1227 (DOI: 10.1128/JVI.01987-13).
98. Hounq H.S.H., Lott L., Gong H., Kuschner R.A., Lynch J.A. Metzgar D. (2009), Adenovirus microsatellite reveals dynamics of transmission during a recent epidemic of human adenovirus serotype 14 infection. *Journal of clinical microbiology.* 47(7): p. 2243-2248 (DOI: 10.1128/JCM.01659-08).

99. Segarra A., Pépin J.F., Arzul I., Morga B., Faury N., Renault T. (2010), Detection and description of a particular *Ostreidherpes virus 1* genotype associated with massive mortality outbreaks of Pacific oysters, *Crassostreagigas*, in France in 2008. *Virus research*. 153(1): p. 92-99 (DOI: 10.1016/j.virusres.2010.07.011).
100. Renault T., Tchaleu G., Faury N., Moreau P., Segarra A., Barbosa-Solomieu V., Lapègue S. (2014), Genotyping of a microsatellite locus to differentiate clinical *Ostreidherpesvirus 1* specimens. *Veterinary research*. 45(1): p. 3 (DOI: 10.1186/1297-9716-45-3).
101. Spatz S.J., Silva R.F. (2007), Polymorphisms in the repeat long regions of oncogenic and attenuated pathotypes of Marek's disease virus 1. *Virus genes*. 35(1): p. 41-53 (DOI: 10.1007/s11262-006-0024-5).
102. Atia M.A., Osman G.H., Elmenofy W.H. (2016), Genome-wide in silico analysis, characterization and identification of microsatellites in *Spodopteralittoralis* multiple nucleopolyhedrovirus (SpliMNPV). *Scientific Reports*. 6(1): p. 1-9 (DOI: 10.1038/srep33741).
103. Duke G.M., Osorio J.E., Palmenberg A.C. (1990), Attenuation of Mengo virus through genetic engineering of the 5' noncoding poly (C) tract. *Nature*. 343(6257): p. 474-476 (DOI: 10.1038/343474a0).
104. Barr J.N., Whelan S.P., Wertz G.W. (1997), cis-Acting signals involved in termination of vesicular stomatitis virus mRNA synthesis include the conserved AUAC and the U7 signal for polyadenylation. *Journal of virology*. 71(11): p. 8718-8725 (DOI: 10.1128/JVI.71.11.8718-8725.1997).

105. Yamada N., Tanihara K., Takada A., Yorihuri T., Tsutsumi M., Shimomura H., Tsuji T., Date T. (1996), Genetic organization and diversity of the 3' noncoding region of the hepatitis C virus genome. *Virology*. 223(1): p. 255-261 (DOI: 10.1006/viro.1996.0476).
106. Garcia-Barreno B., Delgado T., Melero J.A. (1994), Oligo (A) sequences of human respiratory syncytial virus G protein gene: assessment of their genetic stability in frameshift mutants. *Journal of virology*. 68(9): p. 5460-5468 (DOI: 10.1128/JVI.68.9.5460-5468.1994).
107. Singh A.K., Alam C.M., Sharfuddin C., Ali S. (2014), Frequency and distribution of simple and compound microsatellites in forty-eight Human papillomavirus (HPV) genomes. *Infection, Genetics and Evolution*. 24: p. 92-98 (DOI: 10.1016/j.meegid.2014.03.010).
108. Ouyang Q., Zhao X., Feng H., Tian Y., Li D., Li M., Tan Z. (2012), High GC content of simple sequence repeats in Herpes simplex virus type 1 genome. *Gene*. 499(1): p. 37-40 (DOI: 10.1016/j.gene.2012.02.049).
109. Chen M., Tan Z., Zeng G., Zeng Z. (2012), Differential distribution of compound microsatellites in various Human Immunodeficiency Virus Type 1 complete genomes. *Infection, Genetics and Evolution*. 12(7): p. 1452-1457 (DOI: 10.1177/1747493018778713).
110. Alam C.M., Singh A.K., Sharfuddin C., Ali S. (2014), Incidence, complexity and diversity of simple sequence repeats across potexvirus genomes. *Gene*. 537(2): p.189-196 (DOI: 10.1016/j.gene.2014.01.007).

111. Alam C.M., Singh A.K., Sharfuddin C. Ali S. (2013), In-silico analysis of simple and imperfect microsatellites in diverse tobamovirus genomes. *Gene*. 530(2): p. 193-200 (DOI: 10.1016/j.gene.2013.08.046).
112. Kruglyak S., Durrett R., Schug M.D, Aquadro C.F. (2000), Distribution and abundance of microsatellites in the yeast genome can be explained by a balance between slippage events and point mutations. *Molecular Biology and Evolution*. 17(8): p.1210-1219 (DOI: 10.1093/oxfordjournals.molbev.a026404).
113. Hahn H., Palmenberg A.C. (1995), Encephalomyocarditis viruses with short poly (C) tracts are more virulent than their mengovirus counterparts. *Journal of Virology*. 69(4): p. 2697-2699 (DOI: 10.1128/JVI.69.4.2697-2699.1995).
114. Perdue M.L., García M., Senne D., Fraire M. (1997), Virulence-associated sequence duplication at the hemagglutinin cleavage site of avian influenza viruses. *Virus research*. 49(2): p.173-186 (DOI: 10.1016/s0168-1702(97)01468-8).
115. Deback C., Boutolleau D., Depienne C., Luyt C.E., Bonnafous P., Gautheret-Dejean A., Garrigue I., Agut H. (2009), Utilization of microsatellite polymorphism for differentiating herpes simplex virus type 1 strains. *Journal of clinical microbiology*. 47(3): p. 533-540 (DOI: 10.1128/JCM.01565-08).
116. Sanger F., Nicklen S., Coulson A.R. (1977), DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci. USA*. 74: p. 5463-5467 (DOI: 10.1073/pnas.74.12.5463).

117. Maxam A.M., Gilbert W. (1992), A new method for sequencing DNA. *Biotechnology*. 24: p. 99-103 (DOI: 10.1073/pnas.74.2.560).
118. Friedmann T. (1979), Rapid nucleotide sequencing of DNA. *Am J Hum Genet*. 31: p. 19-28
119. Grada A., Weinbrecht K. (2013), Next-generation sequencing: methodology and application. *J Invest Dermatol*. 133: p. e11 (DOI: 10.1038/jid.2013.248).
120. Haagmans B.L., Andeweg A.C., Osterhaus A.D. (2009), The application of genomics to emerging zoonotic viral diseases. *PLoS Pathog*. 5: p. e1000557 (DOI: 10.1371/journal.ppat.1000557)
121. Ansorge W.J. (2009), Next-generation DNA sequencing techniques. *N Biotechnol*. 25: p. 195-203 (DOI: 10.1016/j.nbt.2008.12.009)
122. Buermans H. P., den Dunnen J.T. (2014), Next generation sequencing technology: Advances and applications. *Biochim Biophys Acta*. 1842: p. 1932-1941 (DOI: 10.1016/j.bbadis.2014.06.015).
123. Erlich Y., Mitra P.P., delaBastide M., McCombie W.R., Hannon G.J. (2008), Alta-Cyclic: a self-optimizing base caller for next-generation sequencing. *Nat Methods*. 5: p. 679-682 (DOI: 10.1038/nmeth.1230).
124. Dolan P.C., Denver D.R. (2008), TileQC: a system for tile-based quality control of Solexa data. *BMC Bioinformatics*. 9: p. 250 (DOI: 10.1186/1471-2105-9-250).
125. Schröder J., Bailey J., Conway T., Zobel J. (2010), Reference-free validation of short read data. *PLoS*

- One. 5: p. e12681 (DOI: 10.1371/journal.pone.0012681)
126. Nakamura K., Oshima T., Morimoto T., Ikeda S., Yoshikawa H., Shiwa Y., Ishikawa S., Linak M.C., Hirai A., Takahashi H., Altaf-Ul- Amin M., Ogasawara N., Kanaya S. (2011), Sequence-specific error profile Of Illumina sequencers. *Nucleic Acids Res.* 39: p. e90 (DOI: 10.1093/nar/gkr344)
127. Quince C., Lanzén A., Curtis T.P., Davenport R.J., Hall N., Head I.M., Read L.F., Sloan W.T., (2009), Accurate determination of microbial diversity from 454 pyrosequencing data. *Nat Methods.* 6: p. 639-641 (DOI: 10.1038/nmeth.1361)
128. Gomez-Alvarez V., Teal T.K., Schmidt T.M. (2009), Systematic artifacts in meta genomes from complex microbial communities. *ISME J.* 3: p. 1314-1317 (DOI: 10.1038/ismej.2009.72).
129. Escalante A.E., Lev Jardón Barbolla., Santiago Ramírez-Barahona., Luis E., Eguiarte. (2014). The study of biodiversity in the era of massive sequencing. *Rev Mex de Biodivers.* 85: p. 1249-1264 (DOI: 10.7550/rmb.43498).
130. Loman N.J., Misra R.V., Dallman T.J., Constantinidou C., Gharbia S.E., Wain J., Pallen M.J. (2012), Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol.* 30: p. 434-439 (DOI: 10.1038/nbt.2198).
131. Flicek P., Birney E. (2009), Sense from sequence reads: methods for alignment and assembly. *Nat. Methods.* 6: p. S6-S12 (DOI: 10.1038/nmeth.1376).

132. Mardis E.R. (2008), Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet.* 9: p. 387-402 (DOI: 10.1146/annurev.genom.9.081307.164359).
133. Barzon L., Lavezzo E., Militello V., Toppo S., Palù G. (2011), Applications of next-generation sequencing technologies to diagnostic virology. *Int J MolSci.* 12: p. 7861-7884 (DOI: 10.3390/ijms12117861).

Chapter 2

2.1 Material, methods, and instrumentation

All the clinical samples were brought from outbreak areas. DNA oligonucleotides used in the studies were purchased from Sigma-Aldrich Chemicals Ltd. (St. Louis, MO, USA) and Integrated DNA Technologies (Iowa, United States), respectively.

The chemical reagents used in various experiments such as MgCl₂, D₂O, Tris base, Boric acid, Ethylene diaminetetraacetic acid (EDTA), DMSO, ethanol, isopropanol, and urea, etc. were procured from Sigma-Aldrich Chemicals Ltd. (St. Louis, MO, USA). The dNTPs used for PCR, ladder, agarose, and Taq polymerase enzyme were also purchased from Sigma-Aldrich Chemicals Ltd. (St. Louis, MO, USA). The DNeasy blood and tissue purification Kits, MiniElute gel extraction kit, and PCR purification kit were brought from Qiagen (QIAGEN, Germany). The IlluminaTruSeq Nano DNA Library Prep Kit and AMPure XP beads were brought from (Invitrogen™, USA). The Vero (Cercopithecusaethiops, CCL-81) cell lines used in the cell culture studies were purchased from the National Centre for Cell Sciences (NCCS). All other cell culture media and reagents included Dulbecco's Modified Eagle's Medium (DMEM), Opti-MEM, fetal bovine serum (FBS), phosphate buffer saline (PBS), Trypsin enzyme, and Pen-Strep solution were procured from Gibco Life Technologies, USA. While all the typical plastic wares used were from Parsons Products Pvt. Ltd. (Kolkata, India), but the plastic ware used in cell culture were specifically from Corning (Kennebunk, USA), Molecular

biology grade water used in various experiments to prepare buffer and media was obtained with the help of the Sartorius (Sartorius Corporate, Germany) water purification system.

2.2 Instrument Specifications

2.2.1 Thermal cycler

To carry out the PCR amplification, the ProFlex™ thermal cycler was utilized (Applied Biosystem, USA). This instrument has three blocks to carry out the PCR reaction with different programs.

2.2.2 Gel Imaging

The gel images were visualized using the ImageQuant LAS4000 (GE Healthcare, Biosciences Ltd, Sweden), and the radiolabelled images were acquired using the Fuji FLA-5000 PhosphorImager (Fujifilm, Tokyo).

2.2.3 Sanger sequencer

Applied Biosystems 3730xl has been utilized to sequence the purified PCR product. This instrument is well versed with advances in automation and optics with proprietary Applied Biosystems reagents and bioinformatics tools to carry out a wide range of genetic analysis.

2.2.4 NextSeq500 NGS platform

The Illumina® NextSeq™ 500 NGS platform enables the sequencing of exomes, whole genomes, and transcriptomes and supports TruSeq® and Nextera® libraries. The control software guides you through the steps to set up a sequencing run.

2.2.5 D1000 ScreenTape

The D1000 ScreenTape enables the separation and analysis of fragmented DNA and NGS libraries from 35 - 1,000 bp. Depending on the project, we have to select D1000 and the High Sensitivity D1000 ScreenTape assay. Both can measure accurate and reliable sizing and quantitation. These instruments require only 2 μL sample as input material and have a sensitivity of 5 $\text{pg}/\mu\text{L}$ per fragment. The consumables, buffer, as well as ladder are available with the kit provided by the manufacture.

2.3 Sample preparation and methodology

2.3.1 DNA isolation

All clinical samples were crushed in 1 ml of sterile 0.1 M PBS and then suspended with tissue lysis buffer having proteinase K. The mixture was kept at 56°C overnight for cell lysis. Finally, the cell lysates were filtered by a charged column (Qiagen). DNA was extracted from the column standard phenol-chloroform and stored at -20°C [1].

2.3.2 PCR amplification

We needed the template, a forward primer, and a reverse primer to carry out a PCR reaction. The PCR mix (25 μl) having 5 pmol of each primer, 2mM dNTPs in $1\times$ buffer with 2.5U Taq polymerase. PCR amplified products (5 μl) were loaded onto a 1.5% agarose gel in 1X TAE buffer. The image was captured using the ImageQuant LAS 4000.

2.3.3 Cycle Sequencing PCR

The cycle sequencing reaction requires a DNA template, a sequencing primer, a thermal stable DNA polymerase, deoxynucleotides (dNTPs), dideoxynucleotides (ddNTPs), and buffer. The mixture is subjected to annealing, extension by using fluorescent nucleotide, and denaturation in a thermal cycler. The reactions are followed by amplification, extension, and termination by one of the four dideoxynucleotides.

2.3.4 Preparation of NextSeq 500 Shotgun library

The paired-end sequencing library was constructed from the QC-qualified viral DNA sample utilizing IlluminaTruSeq Nano DNA Library Prep Kit. Covaris M220 fragmented approximately 200 ng of QC-passed DNA to produce a mean fragment of 350 bp. Covaris shearing produces dsDNA fragments having 3' or 5' overhangs. The fragments further subjected to end-repair using End Repair Mix. The 3' to 5' exonuclease activity of this mix releases the 3' overhangs, and the 5' to 3' polymerase activity fills in the 5' overhangs, followed by adapter ligation. This step forms a low rate of the concatenated template formation. The ligated products were size-selected using AMPure XP beads (Invitrogen™, USA). The size-selected products have PCR amplified following the program: 72 °C initial denaturations for 3 min; 95 °C denaturation for 30 s; 12 cycles of 95 °C for 10 s, 55 °C for 30 s, and 72 °C for 30 s; and 72 °C final extension step for 5 min. The adapters were ligated to the DNA fragments' ends, preparing them for hybridization onto a flow cell.

2.3.5 Quality check (QC), cluster generation, and sequencing

The enriched libraries were analyzed on 4200 TapeStation systems using high sensitivity D1000 Screentape as per manufacturers' instruction. After getting the optimum concentration, the Illumina library loaded onto NextSeq500 for cluster generation and sequencing. Paired-end sequencing allows the template fragments to be sequenced in both the direction. The components provided by the kit were utilized to bind samples to complementary adaptor oligos on the paired-end flow cell. The adaptors were designed to allow selective cleavage of the forward strand after re-synthesis of the reverse strand during sequencing. The copied reverse strand was further used to sequence from the opposite end of the fragment.

2.4 Principle behind the techniques/ methods used

2.4.1 Polymerase chain reaction (PCR)

PCR can generate a number of copies of a particular DNA fragment. The reaction is accomplished in a reaction containing template DNA, Taq polymerase, the primers, and the four deoxyribonucleoside triphosphates (dNTPs) in a suitable buffer. The reaction mixture having tubes are subjected to a thermal cycler's heating block followed by repetitive temperature. The reaction process is subdivided into three steps as follows (Figure 2.1):

➤ The denaturation

Here, the separation of the double-strand template DNA separate from one another by raising the temperature. A

temperature of 94°C provided to the mixture, called denaturation temperature, using a thermal cycler. The hydrogen bonds present within the DNA cannot withstand this high temperature and denatured to single-stranded DNA.

➤ **Hybridization**

This step is usually carried out between 40 to 70°C and is called hybridization temperature. During low temperature, the hydrogen bonds reform; this leads to hybridization of complementary strands. The primers' size lies between 10 and 30 oligo-nucleotides and is synthesized chemically. Furthermore, the primers must complement both ends of the sequence of interest to amplify the fragments 5'–3' strand.

➤ **Elongation**

The average temperature for elongation is 72°C when the synthesis of the complementary strand occurs. At 72°C, Taq polymerase, a derivative DNA polymerase enzyme derived from an extremophilic bacterium, *Thermus aquaticus*, and resists temperatures above 100°C. Taq polymerase binds to the single-stranded DNAs of primers and amplifies utilizing the dNTPs present in the reaction mixture. In the next cycle, the amplification product synthesized in the previous process used as a template. It takes around 20–40 cycles to synthesize 0.1 µg of DNA during a PCR reaction. PCR can amplify fragment size less than 6 kilobases within 2–3 hours for a PCR of 30 cycles.

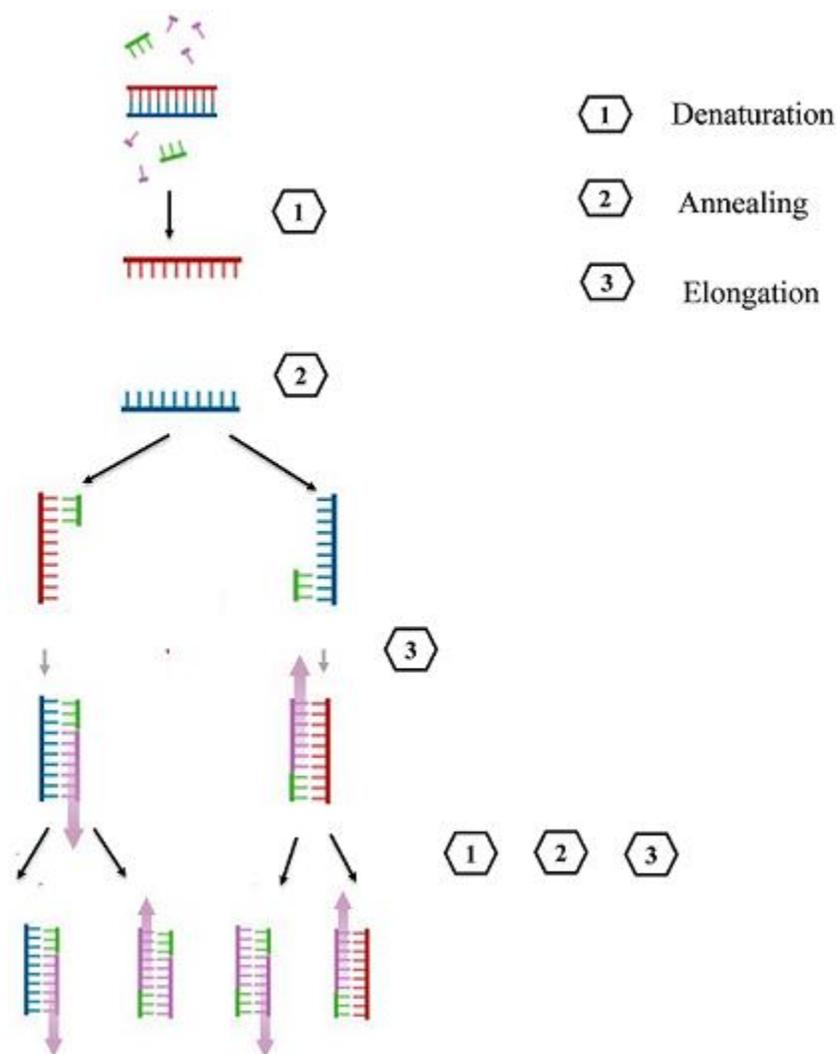


Figure 2.1: Principle of PCR. The figure is representing steps of PCR such as Denaturation, Annealing, and Elongation of specific DNA fragment during an amplification reaction.

2.4.2 Sanger sequencing

The Sanger sequencing principles having the following steps (Figure 2.2):

➤ Template preparation

Genomic DNA can be derived from animals, cell cultures, and plants using the manual phenol-chloroform method or QIAamp Blood Kit (QIAGEN, GmbH) kit. The PCR amplification of the specific target sequence occurs using genomic DNA to get sufficient copies for fluorescent DNA sequencing before proceeding with cycle sequencing.

➤ Cycle sequencing

Cycle sequencing requires a DNA template, a sequencing primer, a thermal stable DNA polymerase, nucleotides (dNTPs), dideoxynucleotides (ddNTPs), and buffer. This reaction uses fluorescent dyes to label the extension products. All required components are combined in a reaction followed by annealing, extension, and denaturation in a thermal cycler. The sequencing reactions amplify extension products and are terminated by one of the four dideoxynucleotides.

➤ Purification after cycle sequencing

The presence of unlabeled and dye-labeled nucleotide components during cycle sequencing reaction interfere with electrokinetic injection, electrophoresis, and data analysis. Purification of amplification products can minimize or discard unwanted substances. Ethanol precipitation methods are usually followed due to their cost effectiveness. Several commercially available products such as the BigDye®XTerminator Purification kit, spin columns, size-

exclusion membranes, and magnetic beads are also used for this purpose.

➤ **Capillary electrophoresis**

The capillary has an opaque, very fragile, polyamide external coating electrophoresis except in the detection window area. At the time of electrophoresis, the laser and detector read samples through an uncoated window area. So the window area remains free of dust. The polyacrylamide gel (POP6) is usually used within the capillary during electrophoresis. The polymer delivery system checked the presence of sufficient polymer and the absence of bubbles or particles before the machine's start. The presence of dust within the polymer can make noisy data. The capillary ends were stored in the buffer or deionized water to avoid the capillary ends to dry. The used buffer and water and discarded daily after each set of runs. Each capillary has the potential to run around 300 runs for 3730/3730xl instruments.

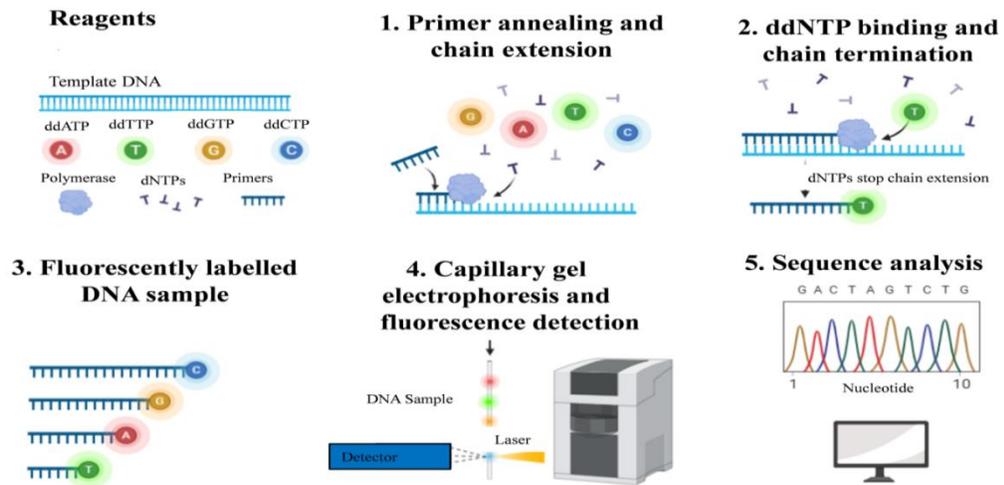


Figure 2.2: Working principle of Sanger sequencing. The sequencing reaction starts with primer annealing and chain extension followed by chain termination, capillary electrophoresis, and sequence analysis.

2.4.3 NextSeq500 next-generation sequencing

The principle of NextSeq500 NGS technology is coincidence with capillary electrophoresis sequencing chemistry. DNA polymerase incorporates fluorescently labeled dNTPs into a DNA template during sequential cycles of DNA synthesis. In each cycle, at the time of dNTP incorporation, recognized by fluorophore excitation. The main difference is that, rather than sequencing a single DNA fragment, NGS processed millions of fragments in a massively parallel fashion. Most genome projects across the world are obtained through Illumina sequencing by synthesis (SBS) chemistry. It provides high accuracy, a greater yield of error-free reads, and an elevated

percentage of base calls. Illumina NGS workflows include four basic steps (Figure 2.3):

- **Library Preparation**—The sequencing library is processed through random fragmentation of the DNA or cDNA sample, followed by 5' and 3' adapter ligation. Alternatively, "tagmentation" combines the fragmented sequence to adaptor through ligation reactions, which greatly increases the efficiency of the library preparation. Then the adaptor-ligated fragments are amplified through PCR followed by gel purification.
- **Cluster Generation**—During cluster generation, the library is loaded in a flow cell containing surface-bound oligos complementary to the library adapters. The fragments are captured over the lawn of the flow cell. The amplification of each fragment is then amplified into distinct, clonal clusters through bridge amplification. After completion of cluster generation, the templates are ready for sequencing.
- **Sequencing**—Illumina SBS technology utilizes a reversible terminator-based method that detects single bases at the time of incorporation into the DNA template. This sequencing chemistry can identify accurate base-by-base nucleotide sequences within repeat sequence and homopolymers, thus reduce error rates.
- **Data Analysis**— The novel sequences are usually aligned to a reference genome. During alignment, many variations of analysis are possible, like single nucleotide polymorphism (SNP) or insertion-deletion (indel) identification, phylogenetic or metagenomic analysis, and more.

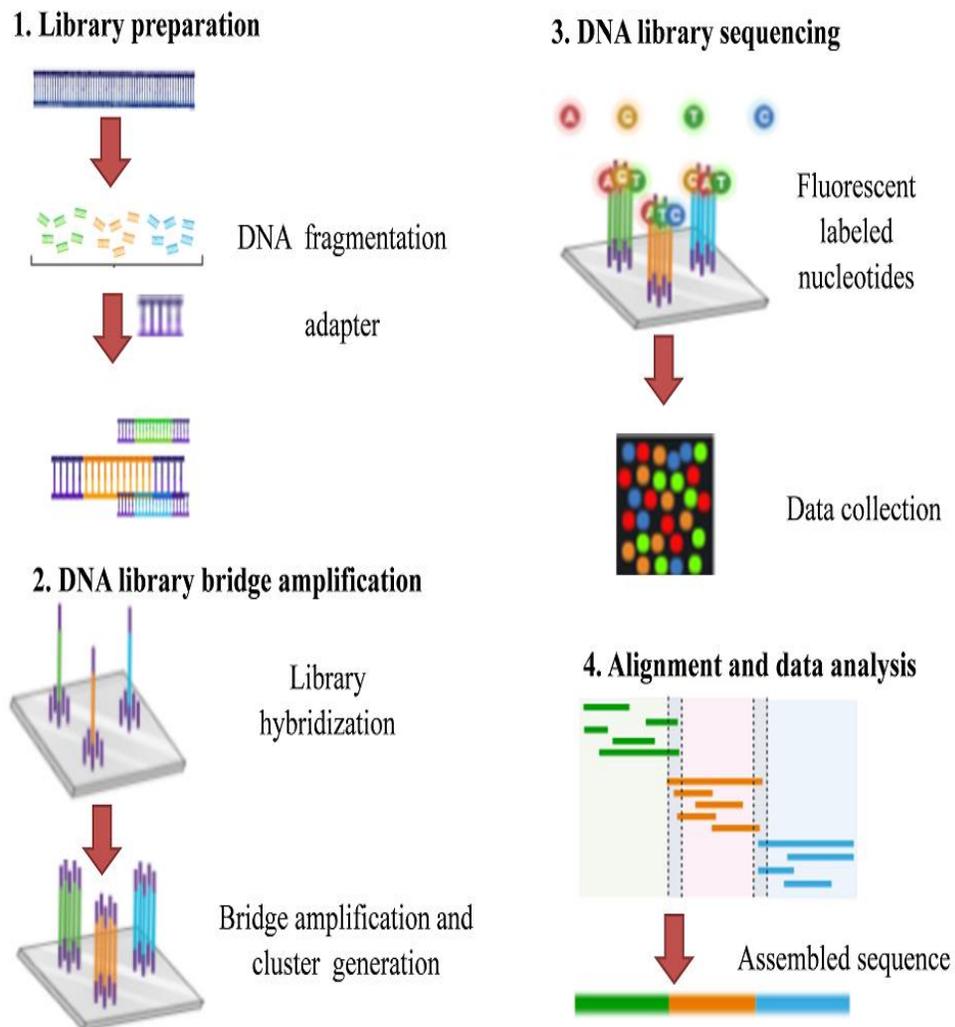


Figure 2.3: Principle behind NextSeq500 NGS. This NGS platform generally utilizes the sequencing by synthesis (SBS) chemistry. The sequencing reaction begins with library preparation followed by hybridization and bridge amplification, sequencing, and assembly.

2.5 Bio-informatic tools and server used in the studies

2.5.1 Molecular Evolutionary Genetics Analysis (MEGA)

MEGA is online free software, which provides tools for analyzing DNA and protein to decipher evolutionary perspectives [2]. Due to its vast application, many version of this software updated from time to time to provide a better experience. MEGA 7 is the latest version, including many applications such as sequence alignment, evolutionary distances estimation, phylogenetic tree construction, selection determination, construction of likelihood trees, etc. The input data file usually contained DNA or protein sequence with FASTA format and .TXT extension. The output of MEGA software comes with .MEG file.

2.5.2 BioEdit

BioEdit is a user-friendly software usually applied to check the DNA and protein sequences [3]. The software usually provides a well-illustrated graphical view for various projects. The FASTA format of DNA and protein sequences act as an input file for this software. Its application includes automatically and manually annotation of DNA and protein sequences, group sequences into various color, phylogenetic tree viewer, edit ABI trace files from ABI auto sequencer model 377, 373, and 3700, image copy/paste and bitmap, edit both amino acid and nucleic acid sequences, reads and writes GenBank, Fasta, NBRF/PIR, Philip 3.2 and Philip 4.

2.5.3 DnaSP

Population genetics is a part of evolutionary biology that evaluates the level of genetic polymorphism in natural populations and determines the evolutionary forces (mutation, migration, selection, and drift) that shapes the genetic pattern in wild populations. Comparison of DNA sequences is the ideal way to measure genetic variation in natural populations [4].

The DnaSP (DNA Sequence Polymorphism) is a bioinformatics tool used to address population genetics and can estimate several measures of DNA sequence differentiation within and between populations in non-coding, is synonymous or in nonsynonymous sites; gene flow, gene conversion, recombination and linkage disequilibrium parameters [5]. Nevertheless, DnaSP used the input formats generated by MEGA [1], NBRF/PIR [6], NEXUS [7], PHYLIP [8], HapMap3 Phased Haplotypes. The analysis usually deciphers many population genetics parameters such as polymorphic sites, InDel (Insertion-Deletion) polymorphism, DNA divergence between populations, synonymous and nonsynonymous substitutions, codon usage bias, gene flow, and genetic differentiation, linkage disequilibrium etc.

2.5.4 Trimmomatic

Trimmomatic is a bioinformatics tool that can be used to trim and crop Illumina (FASTQ) data and remove adapters. The adapter sequence increases the redundancy and creates problems during sequence annotation. Trimmomatic accepts the FASTQ files. Files having ".gz" or ".bz2" extensions

format are also supported by Trimmomatic. The command line, such as ILLUMINACLIP, LEADING, TRAILING used during this. ILLUMINACLIP command used to cut adapter whereas LEADING and TRAILING command cut bases off the start and end of a read, respectively.

2.5.6 Burrows-Wheeler Alignment (BWA)

BWA tool is utilized to map small reads generated by NGS platforms like Illumina/Solexa [9]. The Burrows-Wheeler Transform (BWT) is used [10] to get the alignment of these short reads along the reference genomes. The mapping process requires a small memory computer rather than a high computational tool used during de novo assembly. BWA produces a SAM/BAM file depending on the user's parameters, followed by mapping with the reference genome.

2.5.7 Sequin

Sequin tool generally used to submit or edit a novel biological sequence(s) to National Center for Biotechnology (NCBI) to get GenBank accession number. Here the scientist is prompted to include his/her contact information, authors' information, and host information from which sequence derived. The definition line describes the sequence's name and other information. Sequin also includes nucleotides, the name of the sources, strain, and proteins. After putting all this information, sequin can automatically assemble a record suitable GenBank format. The annotated record was submitted to NCBI to get the accession number.

2.5.8 Recombination Detection Program (RDP)

RDP is a computer program that was developed for the identification of recombination events within aligned nucleotide sequences. It also provides recombination breakpoints and sequence information such as recombinant, major and minor parents. RDP4 is the latest version of this software which include many recombination detection programs such as, BOOTSCAN [11], MaxChi [12], CHIMAERA [13], 3SEQ [14], GENECONV [15], LARD [16], and SCAN [17]. Along with the recombination hot and cold spots identification, this software can construct the phylogenetic tree, tests of purifying selection acting against recombination.

2.6 Servers used in the studies

2.6.1 Imperfect Microsatellite Extractor (IMEx)

IMEx server is used for both perfect and imperfect microsatellite identification [18]. The IMEx tool is having two modes: as a stand-alone program and also in the form of a web server. IMEx scans the nucleotide sequence from hexanucleotide to mononucleotide. The nucleotide sequence having the following parameters: (a) a number of motifs; (b) percentage imperfection within the sequences ; (c) minimum repeat number (n); (d) coding information. The web version provides three different modes: basic, intermediate, and advanced. Using IMEx, the user can evaluate the pattern of the microsatellite, whether it contains a perfect repeat or a combination of perfect and imperfect repeats. Server available at <http://43.227.129.132:8008/IMEX/>

2.6.2 Expert Protein Analysis System (ExPASy)

The ExPASy server was developed by the Swiss Institute of Bioinformatics (SIB), which provides various databases and bioinformatic tools to analyze proteins and proteomics related projects. These databases are named SWISS-PROT, SWISS-2DPAGE, PROSITE, ENZYME, and SWISS-MODEL. However, several proteomics tools like sequence alignment, post-translational modification, topology prediction, and primary, secondary, tertiary structure analysis are available on this server at <https://www.expasy.org/>

2.6.3 MicroSATellite identification tool (MISA)

The MISA microsatellite finder is a web-based tool for the identification of microsatellites in nucleotide sequences. It can identify perfect as well as compound microsatellites [19]. MISA can take two types of the input file, MISA.ini file, FASTA file having the nucleotide sequence. MISA-web server output format in GFF3 file. MISA can take a genome length of 2 Mb and analysis in an efficient, error-free manner. Server available at <https://www.ipk-gatersleben.de/en/bioinformatics-tools/marker-data/misa/>

2.6.4 The Genome Annotation Transfer Utility (GATU)

Several viral genomic sequences have been deposited in GenBank without annotations, generally limiting their value to researchers. However, sequin can annotate some of the small viral genome sequences, such as HIV genomes having identical gene content [20]. But, a sequence having a larger and variable genome like pox, sequin could not annotate. GATU is an ideal annotation tool for those viruses, which compare the orthologues in closely related genomes.

Therefore it reduces the time, tidiness, and complexity of the genome annotation procedure. It can also determine the open reading frames (ORFs) present in the genome and save the final annotated as a GenBank, EMBL, or XML-format. Server available at <https://4virology.net/virology-ca-tools/gatu/>

2.7 References

1. Sambrook J., Fritsch E.F., Maniatis T. (1989), Molecular cloning: a laboratory manual (No. Ed. 2). Cold spring harbor laboratory press
2. Kumar S., Tamura K., Nei. M. (1994), MEGA: molecular evolutionary genetics analysis software for microcomputers. *Bioinformatics*. 10(2): p. 189-191 (DOI: 10.1093/bioinformatics/10.2.189)
3. Hall T., Biosciences I., Carlsbad C. (2011), BioEdit: an important software for molecular biology. *GERF Bull Biosci*. 2(1): p. 60-61
4. Kreitman M. (1983), Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. *Nature*. 304(5925): p. 412-417 (DOI: 10.1038/304412a0).
5. Rozas J. Rozas R. (1999), DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* (Oxford, England). 15(2): p. 174-175 (DOI: 10.1093/bioinformatics/15.2.174).

6. Sidman K.E., George D.G., Barker W.C. Hunt L.T. (1988), The protein identification resource (PIR). *Nucleic acids research*. 16(5): p. 1869-1871 (DOI: 10.1093/nar/16.5.1869).
7. Maddison D.R., Swofford D.L., Maddison W.P. (1997), NEXUS: an extensible file format for systematic information. *Systematic biology*. 46(4): p. 590-621 (DOI: 10.1093/sysbio/46.4.590).
8. Felsenstein J. (1993), PHYLIP (phylogeny inference package), version 3.5 c. Joseph Felsenstein
9. Li H., Durbin R. (2009), Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 25(14): p.1754-1760 (DOI: 10.1093/bioinformatics/btp324).
10. Burrows M., Wheeler D.J. (1994), A block-sorting lossless data compression algorithm
11. Salminen M.O., Carr J.K., Burke D.S., McCutchan F.E. (1995), Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. *AIDS research and human retroviruses*. 11(11): p. 1423-1425 (DOI: 10.1089/aid.1995.11.1423).
12. Maynard Smith J., Smith N.H. (1998), Detecting recombination from gene trees. *Molecular biology and evolution*. 15(5): p. 590-599 (DOI: 10.1093/oxfordjournals.molbev.a025960).
13. Posada D., Crandall K.A. (2001), Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proceedings of the National Academy of Sciences*. 98(24): p.13757-13762 (DOI: 10.1073/pnas.241370698).

14. Boni M.F., Zhou Y., Taubenberger J.K., Holmes E.C. (2008), Homologous recombination is very rare or absent in human influenza A virus. *Journal of virology*. 82(10): p. 4807-4811 (DOI: 10.1128/JVI.02683-07. Epub 2008 Mar 19).
15. Padidam M., Sawyer S., Fauquet C.M. (1999), Possible emergence of new geminiviruses by frequent recombination. *Virology*. 265(2): p. 218-225 (DOI: 10.1006/viro.1999.0056).
16. Holmes E.C., Worobey M., Rambaut A. (1999), Phylogenetic evidence for recombination in dengue virus. *Molecular biology and evolution*. 16(3): p. 405-409 (DOI: 10.1093/oxfordjournals.molbev.a026121).
17. Gibbs M.J., Armstrong J.S., Gibbs A.J. (2000), Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics*. 16(7): p. 573-582 (DOI: 10.1093/bioinformatics/16.7.573).
18. Mudunuri S.B., Nagarajaram H.A. (2007), IMEx: imperfect microsatellite extractor. *Bioinformatics*. 23(10): p.1181-1187 (DOI: 10.1093/bioinformatics/btm097).
19. Beier S., Thiel T., Münch T., Scholz U., Mascher M. (2017), MISA-web: a web server for microsatellite prediction. *Bioinformatics*. 33(16): p. 2583-2585 (DOI: 10.1093/bioinformatics/btx198).
20. Tcherepanov V., Ehlers A., Upton C. (2006), Genome Annotation Transfer Utility (GATU): rapid annotation of viral genomes using a closely related reference genome. *BMC genomics*. 7(1): p. 150 (DOI: 10.1186/1471-2164-7-150).

Chapter 3

Molecular characterization, comparative and evolutionary analysis of the recent Orf outbreaks among goats in the Eastern part of India (Odisha)

3.1 Introduction

Contagious ecthyma or ORF is a zoonotic viral disease of sheep, goats, and other small ruminants, which is characterized by proliferative skin lesions in and around the oral cavity in the form of erythematous macule, papule, vesicle, pustule, and scab. The causative agent is the Orf virus (ORFV), a member of the genus Parapoxvirus of the *Poxviridae* family. The virus is highly contagious and is quite stable in the environment and even remains in the infectious form in wools, hides, and animal excreta for months to years [1, 2]. The disease is manifested by proliferative lesions on the mouth and muzzle that usually get resolved in 1–2 months [3]. A mammalian viral vascular endothelial growth factor (VEGF), having similarity with viral VEGF, was the first virulence factor, played a critical role in ORFV pathogenesis by transcribing early during infection [4]. VEGF-E enhance epidermal and endothelial cell proliferation by interacting with VEGF receptor-2, which helps for vascular permeability and promote epidermal regeneration by supplying cellular substrates for viral replication dermal angiogenesis. Those facial scabs and oral lesions in lambs may interfere with suckling. Lesions on the udder may interfere in feeding neonates and often result in the abandonment of offspring. Similarly, foot lesions can cause transient lameness in infected animals.

Infection progress by forming smaller, not proliferative lesions and resolve more rapidly, usually within 2–3 weeks, due to the absence of protective immunity for that particular virus [1]. Orf is not normally fatal but occurs with a high morbidity rate that can be fatal to the neonates as it deprives them of suckling milk from the infected udder or predisposes them to secondary bacterial or fungal infections [4, 5]. The morbidity of the disease may reach up to 100%, while the mortality rate may reach up to 15%, primarily due to complications of secondary infections, as mentioned earlier [6]. There is increasing evidence of the ability of ORF to cross-infect other species of animals other than sheep and goats, such as camels, gazelles, reindeers, musk ox, and Japanese serows. The virus can infect human being; particularly those are closely associated with animal handling. Zoonoses occur most frequently during lambing, shearing, docking, drenching, or slaughtering of affected animals [1, 3].

Most infections in human appear in hand [7] but occasionally seen in the face [8], nose [9], axilla [10], scalp [11], genitals [12, 13], urethral [8], pericanthal eyelid skin, conjunctiva, and the wound heals spontaneously [3]. However, in immune suppressive individuals, large-sized, poorly healing lesions could remain for an extended period of up to a couple of months [14]. This poses a significant health risk to animal handlers and veterinarians who often get infected by direct contact and develop painful pustular lesions in the skins. Complications of Orf with secondary bacterial infections are potentially life-threatening and need urgent medical attention.

The ORFV is a classic epitheliotropic virus, having a double-stranded DNA genome with a higher (64%) GC content [15]. The genome of this virus consists of central conserved and terminal variable domains. Its genome length is varied 134 to 139 kb having ~130 putative genes, 88 of which are conserved to Chordopoxviruses [16, 17].

Most published phylogenetic and molecular analyses are based on highly conserved genes such as ORFV011 and ORFV059, both of them located at the conserved region, which can induce strong immune responses [18, 19, 20, 21, 22]. Virus interferon resistance gene ORFV020, ATPase gene ORFV108, encodes a dsRNA-binding protein that inhibits the antiviral activity of interferon [23, 24] and viral DNA packaging protein [17, 25], respectively. These genes were used for viral strain differentiation due to their high heterogeneity. Although the presence of the virus has been reported throughout the world [15, 16] as well as in India [26, 27], still a comprehensive analysis of the molecular epidemiology ORFV isolates lack in the prevalent strains of virus isolates from the eastern part of India.

Although several reports from India regarding the ORFV epidemiology suggested its ubiquitous distribution, the eastern part (Odisha) has got less attention. To address the virus circulation in Odisha, we have investigated two Orf outbreaks in the year 2016. Along with three conventional primers sets, for the first time, molecular epidemiology of any Indian ORFV isolates was evaluated on the basis of the ORFV020 gene, which interferes with the interferon (IFN) response of the host and shed insight into the population genetic dynamics of that virus worldwide. The phylogenetic

trees were constructed by using a comparative genomic approach, revealed that most of the isolates belong to Indian origin with a lower nucleotide as well as haplotype diversity.

3.2 Results

3.2.1 Outbreaks description, preliminary screening, and viral isolation

After the late monsoon, the virus infections were noticed in different villages of Khordha, followed by the Puri district. The morbidity rate was recorded to be 8% in each heard, lacking no mortality or human infections at that time of collection. Geographical information about the sample collection site is mentioned in Table-3. 1. Typical clinical symptoms, such as lack of energy, lethargy, weight loss, anorexia, and cutaneous infectious pustules, or scabs on lips, tongue or around the mouth were observed, which were similar to other previously reported [1] (Figure 3.1). Fifteen out of sixteen samples collected during these outbreaks were found to be positive with the Orf 1 and Orf 2 primer [28].



Figure 3.1: ORFV infection in Black Bengal goat. Representative figure depicting clinical cases of ORFV infection in Black Bengal goat having proliferative cutaneous lesions around the lip recorded during outbreak occurred at Khordha district (20.1301° N, 85.4788° E) in the year 2016.

3.2.2 PCR amplification and sequencing

The respective genes, such as ORFV011, ORFV020, ORFV059, and ORFV108, were amplified by using both forward and reverse sequencing primers. The amplicon produced the band at the respective size (Table 3.2). The amplicons were then purified, sequenced, manual edited, and finally uploaded into Genbank. The accession numbers were generated which ranges MG365656-MG365670, MH172306-MH172320, MG365641-MG365655, and MG365671-MG365685 for ORFV011, ORFV020, ORFV059, and ORFV108, respectively (Table 3.3).

3.2.3 Sequence analysis and Phylogenetic tree construction

3.2.3.1 ORFV011 gene

A total of 181 ORFV strains for the ORFV011 gene which includes 15 isolates from this study with another 166 from Genbank having identity level, 92.9–100% and 80.7-100%, for nucleic acid and amino acid, respectively, were used as input to check the population genetic parameters. Out of the 781 sites (aligned nucleotide), 579 conserved, 200 variables, 137 parsimony informative, 63 singleton, and 128 haplotypes. The Hd, π and $\theta = dN/dS$ estimated as 0.992, 0.02541 and 0.05443, respectively (Table 3.4).

The phylogenetic analysis using MEGA 4.0 included a neighbour-joining tree for all ORFV011 gene present in the database as well as the strain obtained from the present study

(Annexure A). The reliability of the tree topology was assessed by 1000 bootstrap replications to infer the relationship between strains. The sequences were broadly grouped into four; group I and group II comprised of the ORFV strain from India and China, group III comprises of the ORFV strains from different continents such as the USA, Uruguay, Brazil, Korea and Taiwan. Group IV consists of only Iraq and one Indian ORFV strain. Interestingly, only one strain of India, which was found rather than its normal host, such as camel, present in group-IV.

Here, all Odisha isolates represent themselves in group II clustered with Assam isolates (JN846834). Therefore, we assumed that the Odisha ORFV was phylogenetically closer to the Assam strains. However, geographically, these two states are nearby each other. So, we can expect that, during the export of their breeds, there may be an exchange of the strains (Figure 3.2).

Name of village/district of sampling	Sex/age of clinical sample	Total number of Infected animals having infected lessons						Total number animals investigated				
		Age Group						Total	Age Group			Total
		<1 years		1-3 years		>3 years			<1 Y	1-3 Y	>3 Y	
		No. of Infected animals	Mb %	No. of Infected animals	Mb %	No. of Infected animals	Mb %					
Balianta /Khordha	F/1.5 years	3	100	2	10	0	0	15	3	10	0	30
		5	100	5	71.4	0	0		5	12	0	
Achyutpur /Khordha	F/4 months	8	100	2	50	0	0	25	8	4	0	30
		6	100	9	75	0	0		6	12	0	
Bidharpur /Khordha	M/1.3 years	6	100	1	33.3	0	0	15	6	3	0	24
		5	83.	3	30	0	0		6	9	0	

			3									
Bhabanipur /Khordha	M/ 5 months	14	93.3	3	60	0	0	30	15	5	0	42
		7	53.8	6	53.8	0	0		13	9	0	
Dhalapat har/Khordha	M/1.2 years	9	100	2	66.7	0	0	12	9	3	0	24
		1	16.7	0	0	0	0		6	6	0	
Hare-krushnapur /Khordha	F/1.5 years	10	100	2	15.4	0	0	20	10	13	0	35
		7	100	1	50	0	0		7	2	1	
Gangapada /Khordha	M/1.3 years	5	71.4	2	15.4	0	0	15	7	13	0	28
		7	100	7	100	0	0		7	1	0	
Matiapada /Khordha	F/ 3 months	8	80	1	50	0	0	23	10	2	0	30
		11	91.7	3	50	0	0		12	6	0	
Biswanathpur /Puri	M/5 months	7	100	8	66.7	0	0	23	7	12	0	27
		7	100	1	100	0	0		7	1	0	
Chandanpur/Puri	F/1.5 years	15	88.2	5	50	0	0	25	17	10	0	40
		4	80	1	14.3	0	0		5	7	1	
Jagan-nathpur /Puri	F/1.8 years	5	100	5	33.3	0	0	17	5	15	0	30
		3	100	4	80	0	0		3	5	2	
Balanga /Puri	M/2 years	9	100	5	50	0	0	27	9	10	0	50
		7	100	4	20	0	0		7	20	4	
Berhampur /Puri	F/1.2 years	4	44.4	1	20	0	0	16	9	5	0	26
		7	100	4	80	0	0		7	5	0	
Bijipur /Puri	M/2 years	7	46.6	3	60	0	0	22	15	5	0	40
		11	73.3	1	33.3	0	0		15	3	2	
Kantapada /Puri	F/5 months	12	92.3	1	20	0	0	25	13	5	0	30
		10	100	2	100	0	0		10	2	0	

Table 3. 1: Table showing the characteristics of the animals included in the study. The characteristics of animals used for sample collection in the investigation and their corresponding morbidity percentage. M=Male, F=Female, Y=years, Mb= Morbidity.

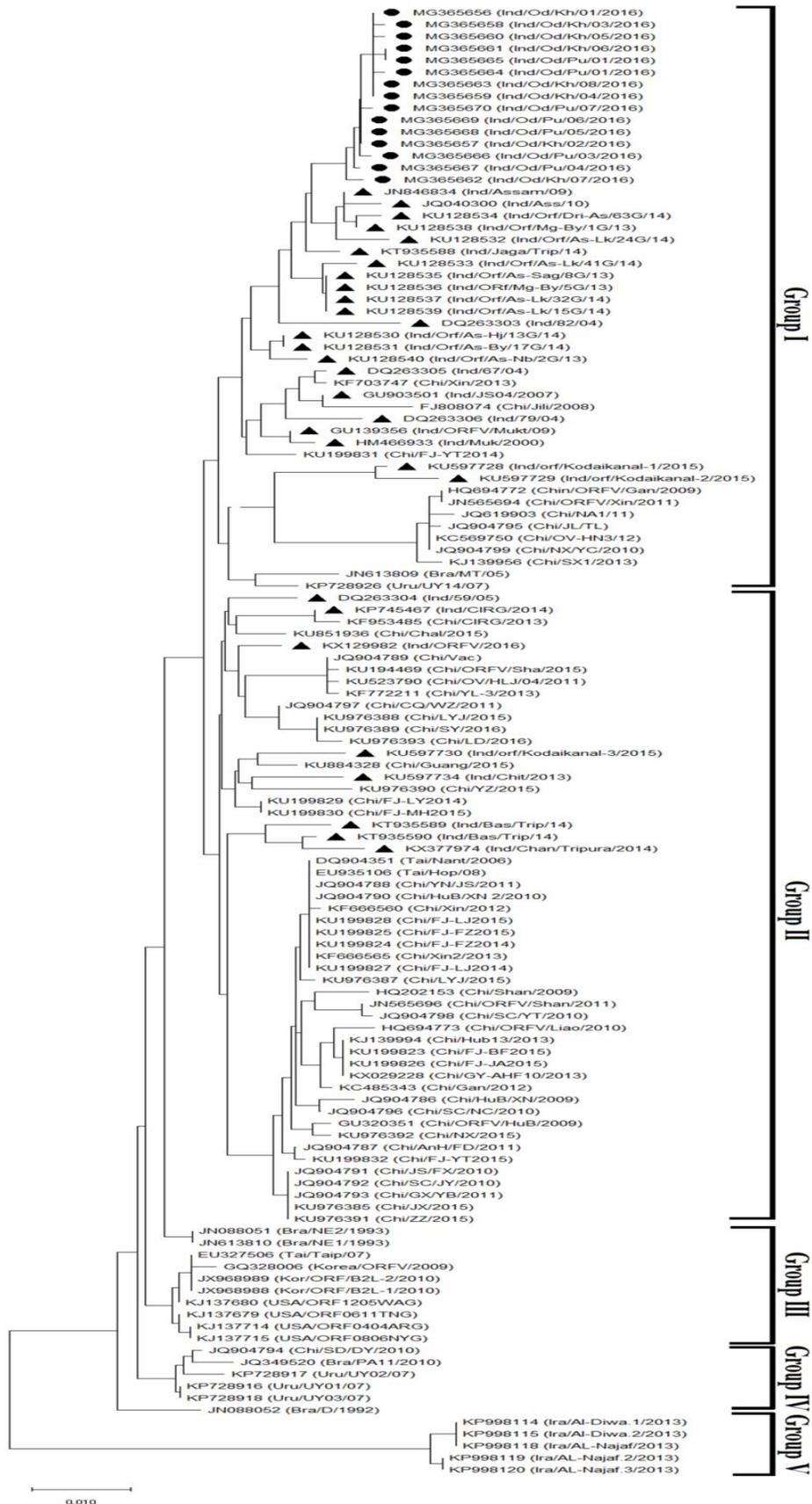


Figure 3.2: Phylogenetic trees constructed for ORFV011 based on partial nucleotide sequences retrieved from GenBank. Bootstrapped maximum likelihood (ML) phylogenetic trees were constructed with MEGA6 based on the Tamura-Nei model for 1000 replications using aligned sequences. Black circles represent the new isolates from Odisha, India, mentioned in our study, while the black triangles represent the other Indian isolates.

3.2.3.2 ORFV020 gene

Analysis of 86 ORFV strains for ORFV020 gene having nucleotide identity 93.9-100% and amino acid similarity 84.9–100%, resulted, 482 sites, 369 were conserved, 113 variable, 77 parsimony informative sites, 40 singletons and 86 haplotype. The Hd, π and $\theta = dN/dS$ estimated as 0.988, 0.03679 and 0.05295, respectively (Table 3.3).

The phylogenetic tree of the ORFV020 gene revealed that there are three groups that exist throughout the world. The strains of India, China, Brazil, USA, UK, Greece, Argentina, Italy, Japan represent themselves in groups. Group-II consists of India and China isolates, whereas group-III consists of the strains of India, China, Korea, the USA, and Zambia. All fifteen strains of the present study exist in group-III and form clade with China (KU199846) (Figure 3.3).

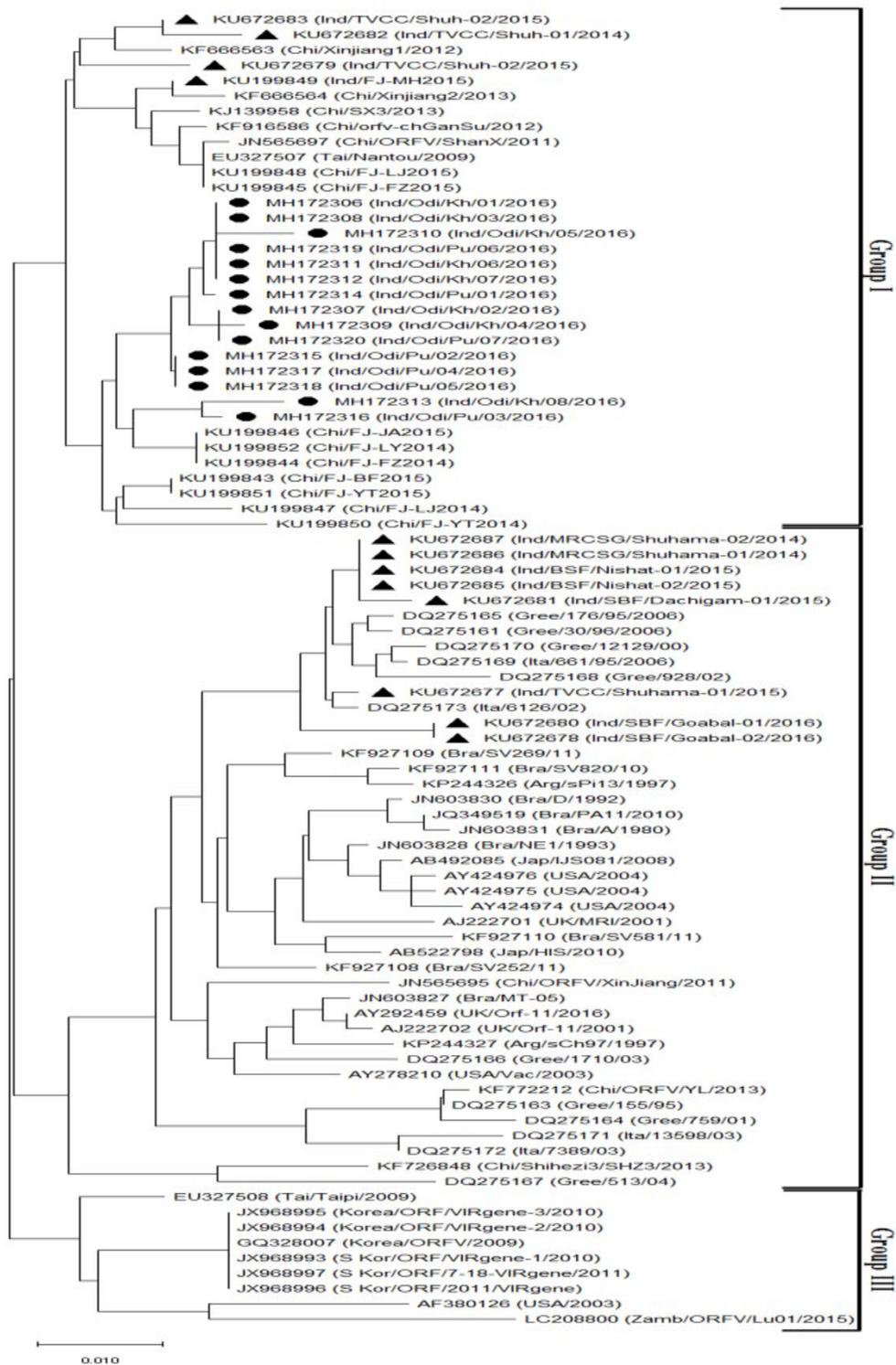


Figure 3.3: Phylogenetic trees constructed for the ORFV020 based on partial nucleotide sequences retrieved from GenBank. Bootstrapped maximum likelihood (ML) phylogenetic trees were constructed with MEGA6 based on the Tamura-Nei model for 1000 replications using aligned sequences. Black circles represent the new isolates from Odisha, India, mentioned in our study, while the black triangles represent the other Indian isolates.

Primer name	Gene function	Sequences (5'–3')	Application	PCR product size
Orf1	Preliminary screening	CGCAGACGTGGCTGAGTACGT	PCR	140bp
Orf2		TGAGCTGGTTGGCGCTGTCCT		
ORFV011-F	Major envelope protein	TCCCTGAAGCCCTATTATTTTTGTG	PCR/Sequencing	1210bp
ORFV011-R		GCTTGCGGGCGTT CGGACCTTC		
ORV020-F	Immuno regulatory gene	ATGGCCTGCGAGTGCGCGTC	PCR/Sequencing	552bp
ORV020-R		AGCGTAATCTGGAACATCGTATGGGTA		
ORFV059-F	Immunodominant envelope protein	ACGTCATCACATGCGGGTCAGAG	PCR/Sequencing	1023bp
ORFV059-R		CTTCCTGTTCTGGCGGGCAT		
ORFV108-F	ATPase	CTCCATTTAGAGGCCGTGAG	PCR/Sequencing	1137bp
ORFV108-R		CGTGTTATGTGCCATCTTGC		

Table 3.2: The table consists of the primers used in our study. The first set of primers, *Orf1*, and *Orf2*, are used for screening and confirming the ORFV isolates. The rest of the primers are used for amplification and sequencing of the genes, namely, *ORFV011*, *ORFV020*, *ORFV059*, and *ORFV108* of the ORFV.

3.2.3.3 ORFV059 gene

A total of 56 ORFV strains for the ORFV059 gene, which includes 15 isolates from this study with another 41 from Genbank having nucleotide and amino acid identity, 79.2-100% and 78.6–100%, respectively. Out of the 677 sites, 472 were conserved, 87 variables, 47 parsimony-informative, 40 singletons, and 43 haplotypes. The Hd , π , $\theta = dN/dS$, estimated as 0.977, 0.01982, 0.03700, respectively (Table 3.3).

Sr. No.	Isolate name	Area of collection	ORFV011 (Accession No.)	ORFV020 (Accession No.)	ORFV059 (Accession No.)	ORFV108 (Accession No.)
1	Od/Kh/01/2016	Balianta	MG365656	MH172306	MG365641	MG365671
2	Od/Kh/02/2016	Achyutpur	MG365657	MH172307	MG365642	MG365672
3	Od/Kh/03/2016	Bidharpur	MG365658	MH172308	MG365643	MG365673
4	Od/Kh/04/2016	Bhabanipur	MG365659	MH172309	MG365644	MG365674
5	Od/Kh/05/2016	Dhalapathar	MG365660	MH172310	MG365645	MG365675
6	Od/Kh/06/2016	Harekrushnapur	MG365661	MH172311	MG365646	MG365676
7	Od/Kh/07/2016	Gangapada	MG365662	MH172312	MG365647	MG365677
8	Od/Kh/08/2016	Matiapada	MG365663	MH172313	MG365648	MG365678
9	Od/Pu/01/2016	Biswanathpur	MG365664	MH172314	MG365649	MG365679
10	Od/Pu/02/2016	Chandanpur	MG365665	MH172315	MG365650	MG365680
11	Od/Pu/03/2016	Jagannathpur	MG365666	MH172316	MG365651	MG365681
12	Od/Pu/04/2016	Balanga	MG365667	MH172317	MG365652	MG365682
13	Od/Pu/05/2016	Berhampur	MG365668	MH172318	MG365653	MG365683
14	Od/Pu/06/2016	Bijipur	MG365669	MH172319	MG365654	MG365684
15	Od/Pu/07/2016	Kantapada	MG365670	MH172320	MG365655	86MG365685

Table 3.3: *The list of ORFV field isolates of Odisha (n=15) characterized in this study. The table includes the isolate name, area of collection, animal or host from where the samples were collected, and the GenBank accession number for the four ORFV genes, ORFV011, ORFV020, ORFV059, and ORFV108.*

The ORFV059 sequences were formed into three groups. Group-I and group-III consist of strains from India and China, whereas the group only represents only Indian strains. The lack of enough sequences with regards to geographical diversification concise the knowledge of strain relationship here. All fifteen strains examined in the present study were placed in group-1 and group-2. Fourteen isolates lay in one clade and close to Bangalore (KY412872), and the single isolate (MG365649) form the clade with strains of Mukteswar (KY412864) (Figure 3.4).

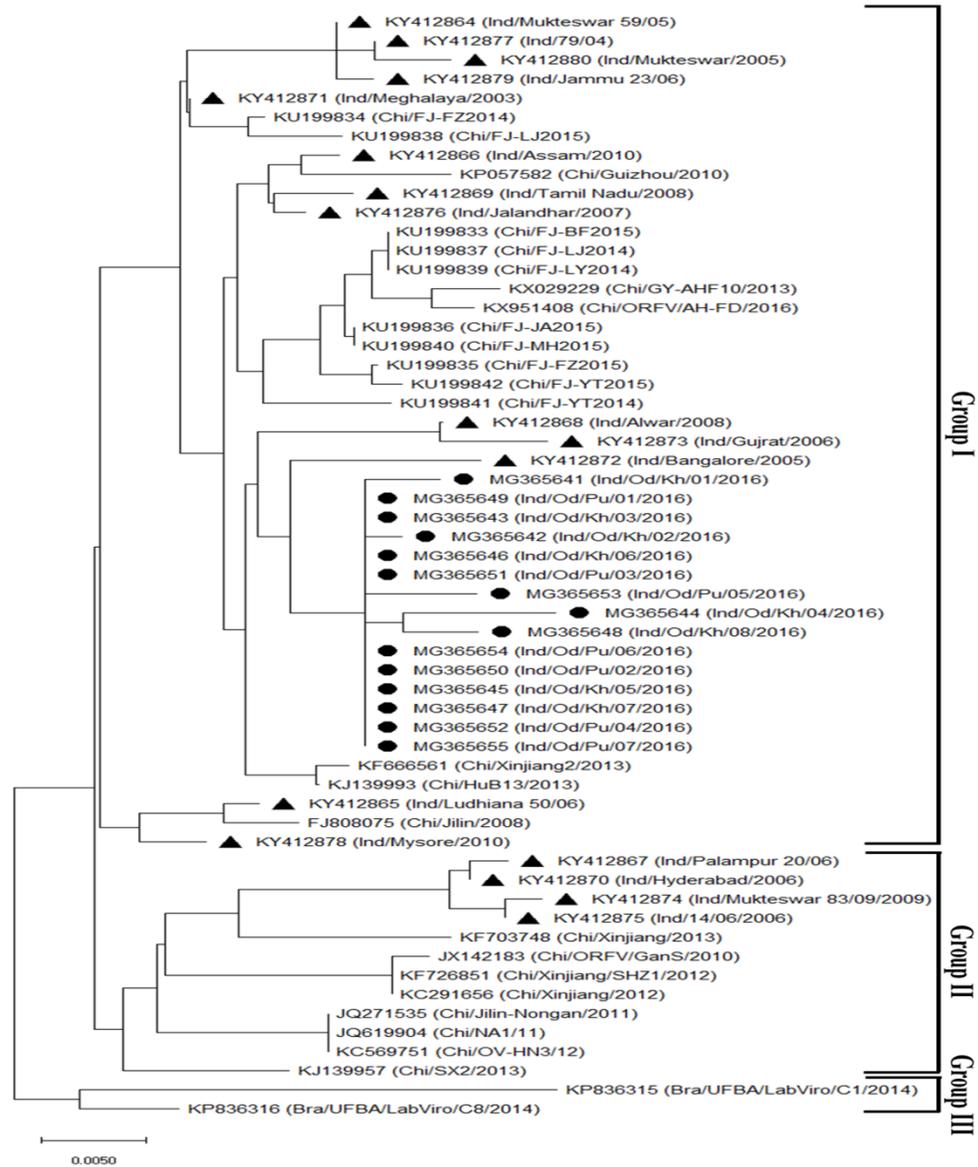


Figure 3.4: Phylogenetic trees constructed for the ORFV058 based on partial nucleotide sequences retrieved from GenBank. Bootstrapped maximum likelihood (ML) phylogenetic trees were constructed with MEGA6 based on the Tamura-Nei model for 1000 replications using aligned sequences. Black circles represent the new isolates from

Odisha, India mentioned in our study while the black triangles represent the other Indian isolates.

Parameters	ORFV011	ORFV020	ORFV059	ORFV108
Full length of the gene in bp	1137	552	1050	1098
Number of sequence analysed	179	86	58	38
Total aligned nucleotide sites	781	482	677	770
Number of Haplotypes	128	86	43	28
Conserved nucleotides	579	379	472	714
Variable nucleotides	202	113	87	56
Partimony informative nucleotides	137	77	47	27
Haplotype diversity	0.992	0.988	0.977	0.966
Nucleotide diversity	0.02541	0.03679	0.01982	0.01129
Θ (per site)	0.05443	0.05295	0.037	0.01793
Nutrality tests D	-1.694	-1.0281	-1.6247	-1.3378
Significance	NS	NS	NS	NS

Table 3.4: Table enlists the different population genetic parameters estimated for ORFV globally.

3.2.3.2 ORFV108 gene

By analyzing 38 ORFV strains for the ORFV108 gene, including 23 from Genbank having nucleotide and amino acid identity, 97.9–100% and 93.3-100%, respectively, resulted in 770 sites (nucleotides). Among them, 714 conserved 56 variables, 27 parsimony-informative, 29 singletons, and 28 haplotypes. The Hd and π and $\theta = dN/dS$ estimated as 0.966, 0.01129 and 0.01793, respectively (Table-3.4).

Due to the lack the of enough reference sequence of ORFV108 gene in GenBank, all analyzed sequences represent themselves in one group. The phylogeny revealed that 14 isolates of the present study exist in one clade and were closely related to Meghalaya (JN183073), which was part of Assam. Only a single isolate (MG365639) of the present study was grouped with strains of Ludhiana (JN183072). As most of

the strains are related to Assam, it makes a clear sense like ORFV011 gene fragment analysis (Figure 3.5).

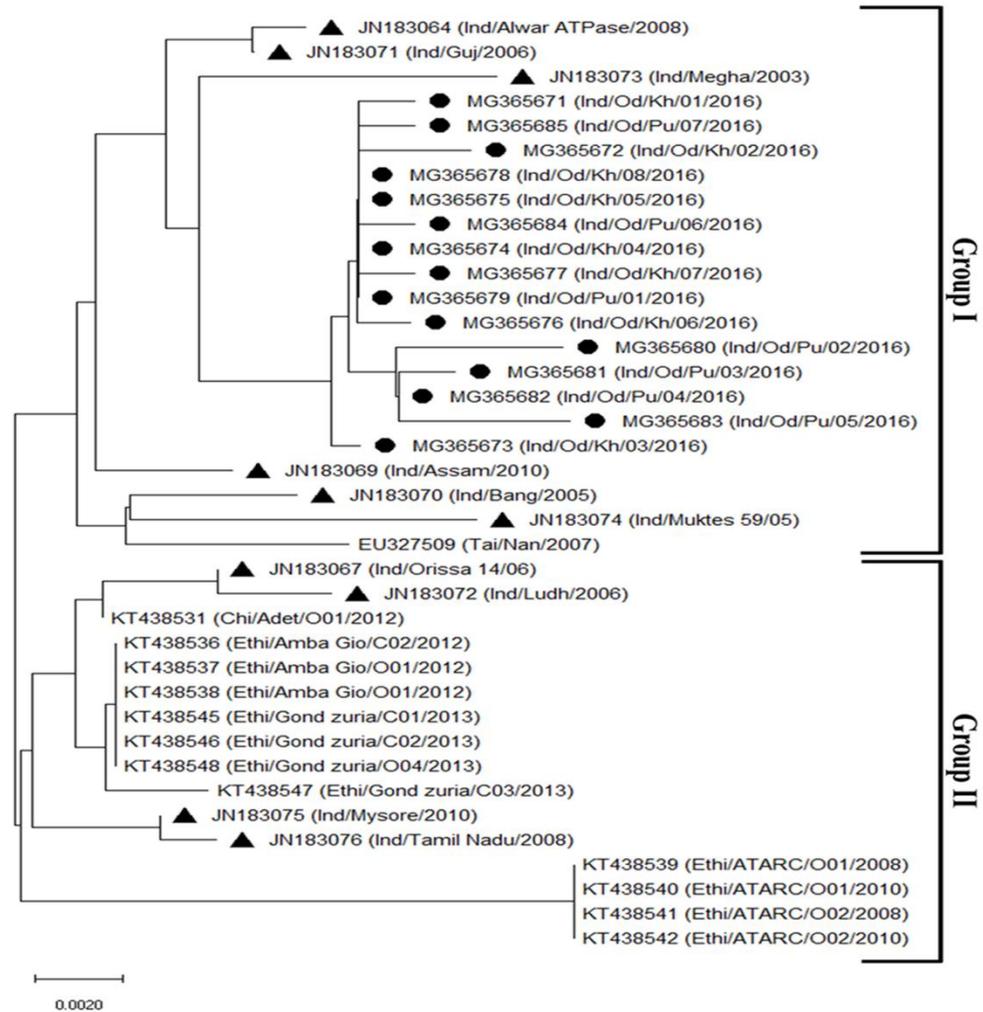


Figure 3.5: Phylogenetic trees constructed for the ORFV108 based on partial nucleotide sequences retrieved from GenBank. Bootstrapped maximum likelihood (ML) phylogenetic trees were constructed with MEGA6 based on the Tamura-Nei model for 1000 replications using aligned sequences. Black circles represent the new isolates from Odisha, India, mentioned in our study, while the black triangles represent the other Indian isolates.

3.3 Discussion

ORFV is an environmentally stable, highly contagious, and transboundary pathogen that affects other small ruminants. The virus has the ability to spread into disease-free areas by sharing pasture lands, migratory patterns, improper husbandry practices, and the intimate contact between the affected and the susceptible animals (sheep and goat). In India, contagious ecthyma is endemic, and it has been reported in both sheep and goats now and then from different agro-climatic regions [27]. Given this, continued epidemiological surveillance and monitoring can reduce ORFV outbreaks and arrest their further spread. This study provides identification, phylogenetic information of ORFV strains and their molecular evolution globally.

In recent years, a few studies using DnaSP have reported high haplotype diversity values in plant RNA virus populations [29], but these studies were mostly focused on Tajima's D similarly. In the present study, haplotype number and frequency analysis using DnaSP showed that there was considerable haplotype diversity in the individual genes.

Among the four genes, Hd and π ranged from 0.966 to 0.992 and 0.01129 to 0.03679, respectively (Table 3.4). The ORFV011 population had the highest Hd (0.992), followed by ORFV020 (0.988), ORFV059 (0.977), and ORFV108 (0.966), respectively. In the case of nucleotide diversity, the gene ORFV20 (0.03679) has the highest value, followed by ORFV011 (0.02541), ORFV059 (0.01982), and ORFV108 (0.01129). These values are a sensitive index of the genetic diversity of a population [29]. Interestingly, having a larger sample size variation between the analyzed genes (Table-3.4),

still, a mere difference is observed between haplotype diversity. In the case of nucleotide diversity, having less than half of the sequences of the ORFV011 gene, ORFV020 resulted in the highest value of nucleotide diversity. According to García-Arenal et al., 2001, many RNA and DNA plant viruses exhibiting nucleotide diversity lower than 0.070 in diverse genomic regions have a low genetic variation [30]. In our study, all genes having nucleotide diversity lower than 0.070, suggesting that there is little variation exists within the isolates. However, our assumption is very much similar to the Esposito et al., 2006, while studying the evolution of the Variola (Smallpox) Virus [31]. The low diversity should assist the development of targeted, efficacious antiviral drug development in eradicating the disease.

Selection pressure analyses showed that ORFV011, ORFV020, ORFV059, ORFV108 genes were under purifying selection with (dN/dS) values 0.05443, 0.05295, 0.037, 0.01793, and, respectively. No positively selected site was found in any of the four genes studied. This result coincides with the evolution study of avipoxvirus [31]. The synonymous and nonsynonymous mutations drive the selection pressure to fluctuate the evolution rate. The positive selection for genes involved in virulence and played a major role in for host-pathogen interaction [33]. Indeed, a positive selection of viruses could also contribute to an increase of substitution rate, which can be observed either for viruses inducing a strong host immune response or for those which have been the object of vaccination [34]. ORFV vaccination is not systematic worldwide, thus making selection pressure assessment difficult. Except for vaccination, several other circumstances like ecological and epidemiological factors

should be further investigated for a proper understanding of the observed selection phenomenon [32].

Tajima's D test is based on the differences between the numbers of segregating sites and the average number of nucleotide differences. Tajima's D test of neutrality was performed on the four genes to evaluate whether the sequence variant distribution followed the model of neutral evolution. The significantly negative value obtained (Table-3.4) (statistical significance $P < 0.001$) confirms that purifying selection is acting on the part of the ORFV genome studied. Statistically significant negative values were obtained for Tajima's suggesting that this viral group may be undergoing a period of evolutionary expansion.

The molecular phylogenetic tree illustrates here to prove that most of the Indian strain is closely related to China as they exist in the same group. This result is very much similar to the previous result [27]. However, it is difficult to determine the precise route through which those Chinese strains were introduced in India. The report regarding the outbreaks in Odisha is very rare. Most of the analyzed genes (ORFV011 and ORFV0108) of this outbreak samples represent that they are similar to Assam isolates. As Assam geographically nearer to Odisha, one can assume that during the transportation of breeds, that virus may be transferred. However, there is some heterogeneity observed in some genes (ORFV020 and ORFV059) that suggest a close relation of that virus with Chinese and Bangalore strains. How these mixed populations exist in a single pool is an important question to address in our future research. The absence of uniform sampling may result in that ambiguity, which needs further validation by doing extensive sampling and sequencing from

the different geographical regions of India. ORFV undergo genetic recombination during which many nonessential genes may be deleted, or new host genes may be acquired, leading to the generation of new variants in the population [35, 36]. To investigate whether recombination plays a crucial role in virus evolution, it needs, increases the genomic resources of that particular virus in the NCBI database by doing several complete genome sequences of that particular virus, as, still a complete genome of ORFV is not reported in Indian point of view. Due to heterogenicity, it is difficult to develop an effective vaccine is lacking in that virus. Our findings provide evidence of genetic variation in circulating ORFV strains in India, which may contribute to an improved understanding of ORFV epidemiology to develop a proper vaccine. Our search for clues into the evolution of ORFV is provocative, but it remains limited because the repository represents a somewhat unsystematic collection.

3.4 Materials and methods

3.4.1 Sample and outbreak data collection

In October and Nov 2016, two outbreaks were noticed at two districts of Odisha in the geographical locations of Khordha (Phulnakhara) (20.1301° N, 85.4788° E) and Puri (Nimapara) (19.8510° N, 85.7256° E) districts. Scab tissue samples were collected from the animals (goats) presented to veterinary clinics at both infective and recovery/convalescent phases. Upon collection of the scabs, simultaneous treatment of infected wounds was done with 2% boro glycerine and parenteral application of Enrofloxacin@ 5mg/kg IM as prescribed by veterinarians. The samples were collected randomly from eight animals spread in the eight different

villages of each district mentioned earlier. Briefly, about 5gm of tissue samples were collected from each animal and placed in a labeled sterile universal tube containing phosphate-buffered saline (PBS), pH 7.2 supplemented with antibiotics and antifungal reagents. Samples were immediately transferred into a cold box and then stored at -20°C until further processing and laboratory analysis.

3.4.2 DNA isolation and PCR

Briefly, samples collected from the vesicular lesions around the lower lip from the affected animals were triturated in sterile 0.1 M PBS. The homogenized sample was then treated with tissue lysis buffer containing proteinase K, and the mixture was incubated at 56°C overnight for further cell lysis. Finally, the mixture was passed through a charged column (QUAGEN, USA), and DNA was purified from the column by using the standard phenol-chloroform method as described by Sambrook et al., 1989 and stored at -20°C until further use [37].

3.4.3 Primer design and preliminary screening by PCR

To confirm the presence of ORFV, all samples were screened by PCR using ORFV-specific primers as mentioned by Toferson et al., 2002 to get the band at 140bp. After preliminary screening, the positive samples were amplified using ORFV011, ORFV20, ORFV059, and ORFV108 genes coding for the viral major envelope protein, interferon resistance gene, immune dominant envelope protein, and ATPase, respectively, followed by Sanger sequencing. All sets of primers used are listed in Table-3.2.

3.4.4 Amplification and sequencing of fragments containing the ORFV011, ORFV020, ORFV059, and ORFV108 genes

Molecular characterization was conducted by using a partial fragment of ORFV011, ORFV20, ORFV059, and ORFV108 genes. The PCR reaction was conducted in a reaction volume of 25 μ L containing 500 nM forward primer and reverse primer, using Taq polymerase as per the supplier's instruction (NEB). The thermal cycling conditions for all genes were as follows: initial denaturation at 95°C for 5 min, followed by 35 cycles at 95 °C for 50 s, annealing at 55 °C for 60 s and extension at 72 °C for 90 s with a final extension at 72 °C for 7 min. The same protocol was used for all, except the ORFV011 gene, where the annealing temperature was at set 52°C. For each PCR run, a negative control consisting of DI water instead of the template DNA was included. Aliquots of the PCR products were checked using gel-electrophoresis on a 1.5 % agarose gel stained with ethidium bromide. The DNA from the PCR products was extracted using a gel extraction kit (Invitrogen, USA) according to the manufacturer's instructions. Automated nucleotide sequencing was performed using a 3730XL DNA analyzer (Applied Biosystems, USA).

3.4.5 Estimation of population genetics parameters

Multiple nucleotide sequence alignments were performed using CLUSTALW in Bioedit 7.4 to analyze the presence of heterogeneity in terms of nucleotide as well as amino acid [38]. To elucidate The ORFV evolution, we retrieved all sequences of respective genes present in GenBank to measure several population genetics parameters such as haplotype diversity (Hd), nucleotide diversity (π), selection pressure (θ),

and neutrality test (D). Aligned sequences were assessed using DnaSP software (version 5.10.01) to evaluate all the above-said parameters [39]. To assess selection pressure imposed upon ORFV protein-coding regions, non-synonymous (dN) and synonymous (dS) substitution rates and their associated ratios ($\theta = dN/dS$) were estimated for each segment by using the bootstrap method with 100 replicates under the Kumar method in DnaSP. Bootstrapped maximum likelihood (ML) phylogenetic trees were constructed with MEGA6 based on the Tamura-Nei model for 1000 replications using aligned sequences, excluding primer sequences to establish phylogenetic relations among ORFV sequences [40]. To infer the neutrality test, Tajima's D, based on the differences between the numbers of segregating sites and the average number of nucleotide differences, was evaluated in those genes.

3.5 Conclusion

The conclusions in this present study may be informative with regards to genotype classification of that less studied etiological agent, which will pave the path to improve the strategy for disease prevention and possible eradication.

3.6 References

1. Haig D.M., Mercer A.A. (1998), Ovine diseases Orf. *Vet Res.* p. 29:311-326
2. Gokce H.I., Genc O., and Gokce G. (2005), Sero-prevalence of Contagious Ecthyma in Lambs and Humans in Kars, Turkey. *Turk. J. Vet. Anim. Sci.* 29: p. 95-101

3. Hosamani M., Scagliarini A., Bhanuprakash V., McInnes C.J., Singh R.K. (2009), Orf: an update on current research and future perspectives. *Expert Rev Anti Infect Ther.* 7:879-893 (DOI: 10.1586/eri.09.64).
4. Haig, D. M., & Mercer, A. A. (1998), *Ovine diseases.* Orf. *Veterinary research.* 29(3-4): p. 311-326.
5. Pugh D.G., Baird N.N. (2012), *Sheep and goat medicine.* Elsevier Health Sci.
6. Bora D.P., Barman N.N., Kumar Das., S.K., Bhanuprakash V., Yogisharadhya R., Venkatesan G., Kumar A., Rajbongshi G., Khatoon E., Chakraborty A., and Bujarbaruah K.M. (2012), Identification and phylogenetic analysis of Orf virus isolated from outbreaks in goats of Assam, a North Eastern state of India. *Virus Genes.* 45: p. 98-104 (DOI: 10.1007/s11262-012-0740-y).
7. Uzel M., Sasmaz S., Bakaris S., Cetinus E., Bilgic E., Karaoguz A., Ozkul A. and Arican O., (2005), A viral infection of the hand commonly seen after the feast of sacrifice: human orf (orf of the hand). *Epidemiology & Infection.* 133(4): p. 653-657 (DOI: 10.1017/s0950268805003778).
8. Duchateau N.C., Aerts O., Lambert J. (2014), Autoinoculation with orf virus (ecthymacontagiosum). *Int J Dermatol.* 53: p. e60-e62 (DOI: 10.1111/j.1365-4632.2012.05622.x).
9. Maki A. Jr., Hinsberg A., Percheson P., Marshall D.G. (1988), Orf: contagious pustular dermatitis. *CMAJ.*;139: p. 971-972
10. Weide B., Metzler G., Eigentler T.K., Fehrenbacher B., Sonnichsen K., Garbe C. (2009), Inflammatory

- nodules around the axilla: an uncommon localization of orf virus infection. *ClinExpDermatol.* 34: p. 240-242 (DOI: 10.1111/j.1365-2230.2007.02567.x).
11. Glass J., Ghali F.E., Sinkre P., Ricotti C.A., Cockerell C.J. (2009), Acute onset of erythematous scalp nodules in a child—quiz case. *Arch Dermatol.* 145: p. 1053-1058 (DOI: 10.1001/archdermatol.2009.184-a).
 12. Kandemir H., Ciftcioglu M.A., (2008), Yilmaz E. Genital orf. *Eur J Dermatol.* 18: p. 460-461 (DOI: 10.1684/ejd.2008.0444).
 13. Vogel T.A., Schuttelaar M.L. (2013), Generalized orfsuperinfection in a child with atopic dermatitis. *Eur J Dermatol.* 23: p. 538-539 (DOI: 10.1684/ejd.2013.2076).
 14. Caravaglio J.V., and Khachemoune A. (2017), Orf Virus Infection in Humans: A Review With a Focus on Advances in Diagnosis and Treatment. *Journal of drugs in dermatology: JDD.* 16(7): p. 684-689.
 15. Cottone R., Büttner M., Bauer B., Henkel M., Hettich E., and Rziha H. J. (1998). Analysis of genomic rearrangement and subsequent gene deletion of the attenuated Orf virus strain D1701. *Virus research.* 56(1): p. 53-67 (DOI: 10.1016/s0168-1702(98)00056-2).
 16. Delhon G., Tulman E.R., Afonso C.L., Lu Z., De la Concha-Bermejillo A., Lehmkuhl H.D., Piccone M.E., Kutish G.F., Rock D.L. (2004), Genomes of the parapoxviruses ORF virus and bovine papular stomatitis virus. *Journal of virology.* 78(1): p. 168-177 (DOI: 10.1128/jvi.78.1.168-177.2004).

17. Mercer A.A., Ueda N., Friederichs S.M., Hofmann K., Fraser K.M., Bateman T., Fleming S.B. (2006), Comparative analysis of genome sequences of three isolates of Orf virus reveals unexpected sequence variation. *Virus research*. 116(1): p. 146-158 (DOI:10.1016/j.virusres.2005.09.011).
18. Billinis C., Mavrogianni V.S., Spyrou V., Fthenakis G.C. (2012), Phylogenetic analysis of strains of Orf virus isolated from two outbreaks of the disease in sheep in Greece. *Virology Journal*. 9(1): p. 24 (DOI: 10.1186/1743-422X-9-24).
19. Li W., Ning Z., Hao W., Song D., Gao F., Zhao K., Liao X., Li M., Rock D.L., Luo S. (2012), Isolation and phylogenetic analysis of orf virus from the sheep herd outbreak in northeast China. *BMC veterinary research*. 8(1): p. 229 (DOI: 10.1186/1746-6148-8-229).
20. Yang, H., Meng, Q., Qiao, J., Peng, Y., Xie, K., Liu, Y., Zhao, H., Cai, X. and Chen, C., 2014. Detection of genetic variations in Orf virus isolates epidemic in Xinjiang China. *Journal of basic microbiology*, 54(11), pp.1273-1278. (DOI: 10.1002/jobm.201300911)
21. Zhang K., Liu Y., Kong H., Shang Y., Liu X., (2014a). Comparison and phylogenetic analysis based on the B2L gene of orf virus from goats and sheep in China during 2009–2011. *Arch. Virol*. 159, 1475–1479 (DOI: 10.1007/s00705-013-1946-6)
22. Zhang K., Liu Y., Kong H., Shang Y., Liu X. (2014b), Human infection with Orf virus from goats in China, 2012. *Vector Borne Zoonotic Dis*. 14: p. 365–367 (DOI: 10.1089/vbz.2013.1445).

23. McInnes C.J., Wood A.R., Mercer A.A. (1998), Orf virus encodes a homolog of the vaccinia virus interferon-resistance gene E3L. *Virus genes*. 17(2): p. 107-115 (DOI: 10.1023/a:1026431704679).
24. Kottaridi C., Nomikou K., Teodori L., Savini G., Lelli R., Markoulatos P., Mangana, O. (2006), Phylogenetic correlation of Greek and Italian orf virus isolates based on VIR gene. *Veterinary microbiology*. 116(4): p. 310-316 (DOI: 10.1016/j.vetmic.2006.04.020).
25. Yogisharadhya R., Bhanuprakash V., Venkatesan G., Balamurugan V., Pandey A. B., Shivachandra S.B.(2012), Comparative sequence analysis of poxvirus A32 gene encoded ATPase protein and carboxyl terminal heterogeneity of Indian orf viruses. *Veterinary microbiology*. 156(1-2): p. 72-80 (DOI: 10.1016/j.vetmic.2011.10.021).
26. Mondal B., Bera A.K., Hosamani M., Tembhurne P.A., Bandyopadhyay S.K.(2006), Detection of orf virus from an outbreak in goats and its genetic relation with other parapoxviruses. *Veterinary research communications*. 30(5): p. 531-539 (DOI: 10.1007/s11259-006-3270-z).
27. Venkatesan G., Balamurugan V., Bora D.P., Yogisharadhya R., Prabhu M., Bhanuprakash V. (2011), Sequence and phylogenetic analyses of an Indian isolate of orf virus from sheep. *Vet. Ital.* 47: p. 323-332
28. Torfason E.G., Gunadóttir S. (2002), Polymerase chain reaction for laboratory diagnosis of orf virus infections. *Journal of Clinical Virology*. 24(1-2): p. 79–84 (DOI: 10.1016/s1386-6532(01)00232-3).

29. Nguyen H.D., Tran H.T.N., Ohshima K. (2013), Genetic variation of the Turnip mosaic virus population of Vietnam: a case study of founder, regional and local influences. *Virus Research*. 171(1): p. 138-149 (DOI: 10.1016/j.virusres.2012.11.008).
30. García-Arenal F., Fraile A., Malpica J.M. (2001), Variability and genetic structure of plant virus populations. *Ann Rev Phytopathol*. 39: p. 157–186 (DOI: 10.1146/annurev.phyto.39.1.157).
31. Esposito J.J., Sammons S.A., Frace A.M., Osborne J.D., Olsen-Rasmussen M., Zhang M., Li Y. (2006), genome sequence diversity and clues to the evolution of variola (smallpox) virus. *Science*. 313(5788): p. 807-812 (DOI: 10.1126/science.1125134)
32. Le Loc’h G., Bertagnoli S., Ducatez M.F. (2015), Time scale evolution of avipoxviruses. *Infection, Genetics and Evolution*. 35: p. 75-81 (DOI: 10.1016/j.meegid.2015.07.031).
33. McLysaght A., Baldi P.F., Gaut B.S. (2003). Extensive gene gain associated with adaptive evolution of poxviruses. *Proc. Natl. Acad. Sci*. 100: p. 15655–15660 (DOI: 10.1073/pnas.2136653100).
34. Firth C., Kitchen A., Shapiro B., Suchard M.A., Holmes E.C., Rambaut A. (2010), Using time-structured data to estimate evolutionary rates of double-stranded DNA viruses. *Mol. Biol. Evol*. 27: p. 2038–2051 (DOI: 10.1093/molbev/msq088).
35. Fleming S.B., Lyttle D. J., Sullivan J. T., Mercer A. A., Robinson A. J. (1995), Genomic analysis of a transposition-deletion variant of orf virus reveals a 3.3 kbp region of non-essential DNA. *Journal of General*

- Virology. 76(12): p. 2969-2978 (DOI: 10.1099/0022-1317-76-12-2969).
36. McInnes C. J., Wood A. R., Nettleton P. F., Gilray J. A. (2001), Genomic comparison of an avirulent strain of Orf virus with that of a virulent wild type isolate reveals that the Orf virus G2L gene is non-essential for replication. *Virus Genes*. 22(2): p. 141-15 (DOI: 10.1023/a:1008117127729).
 37. Sambrook J., Fritsch E.F., Maniatis T. (1989), *Molecular cloning: a laboratory manual* (No. Ed. 2). Cold spring harbor laboratory press
 38. Hall T., Biosciences I., Carlsbad C. (2011), BioEdit: an important software for molecular biology. *GERF Bull Biosci*, 2(1), p. 60-61
 39. Rozas J. Rozas R. (1999), DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* (Oxford, England). 15(2): p. 174-175 (DOI:10.1093/bioinformatics/15.2.174).
 40. Kumar S., Tamura K., Nei. M. (1994), MEGA: molecular evolutionary genetics analysis software for microcomputers. *Bioinformatics*. 10(2): p. 189-191 (DOI:10.1093/bioinformatics/10.2.189).

Chapter 4

Comparative analyses, distribution, and characterization of microsatellites in Orf virus genome

4.1 Introduction

Contagious ecthyma or Orf is a zoonotic viral disease of sheep, goats, and other small ruminants, characterized by proliferative skin lesions in and around the oral cavity in the form of erythematous macule, papule, vesicle, pustule, and scabs. The causative agent is the Orf virus (ORFV), a member of the genus Parapoxvirus of the *Poxviridae* family. The virus is highly contagious, quite stable in the environment, and remains in the infectious form in wools or animal excreta for months to years [1]. The disease is manifested by proliferative lesions on the mouth and muzzle that usually get resolved in 1–2 months [2]. These facial scabs and oral lesions in lambs may interfere with suckling, while lesions on the udder may interfere in feeding neonates. Often foot lesions cause transient lameness in the infected animals, and together all these results in poor health and loss of body weight. Lesions progress through all clinical stages but are generally non-proliferative and usually resolve within 2–3 weeks. ORFV specific antibodies do not seem to confer protective immunity, although the IgG2 isotype is believed to provide some defense against ORFV infection [3]. As IgG2 is not secreted in the colostrum of ruminants, lamb and kids don't get the required protection [5]. However, Orf is normally non-fatal but often comes with high morbidity (up to 100%) that can be life-threatening to the neonates by interfering with the suckling of milk from the infected udder or predisposing the animals to

secondary bacterial or fungal infections [6]. For these reasons, the mortality rate may reach up to 15% [7]. There is increasing evidence of ORFV to cross-infect other species of animals such as camels, gazelles, reindeers, musk ox, and Japanese serows [3].

The virus can infect humans, particularly those who are closely associated with animal handling. Zoonosis occurs most frequently during lambing, shearing, docking, drenching, or slaughtering of affected animals [1, 3]. Orf infections in human appear in hand [8] but occasionally seen in the face [9], nose [10], axilla [11], scalp [12], genitals [13, 14], urethral [9], and pericanthal eyelid skin and the wound heals spontaneously. However, in immunosuppressive individuals, large-sized poorly healing lesions could remain for an extended period up to a couple of months [15]. This poses a significant health risk to animal-handlers and veterinarians who often get infected by direct contact and develop painful pustular lesions in the skins. Complications of Orf with secondary bacterial infections are potentially life-threatening and need urgent medical attention.

The ORFV is a classic epitheliotropic virus, having a double-stranded DNA genome with a higher (64%) GC content [16]. The genome consists of central conserved and terminal variable domains with sizes varying from 134 to 139 kbp having ~130 putative genes, 88 of which are conserved to Chordopoxviruses [17, 18]. Having such a devastating character, this virus has got less attention in terms of genomic information, which is evident from the availability of only eleven complete genome sequences worldwide. Although several conserved genomic regions such as, envelope protein B2L (ORFV011), F1L (ORFV059) and A32L (ORFV108)

were used for ORFV identification and phylogenetic tree construction [19]. Still, there is a lack of clarity regarding the real diversity of ORFV due to the absence of a reliable system for virus identification, which consists of hypermutable regions such as microsatellites rather than conventional conserved genes.

Simple sequence repeats (SSRs), also known as microsatellites, refer to mono-, di-, tri-, tetra-, penta- and hexanucleotide sequence units that are repeated in tandem in a genome [20]. Those short motifs of DNA are distributed ubiquitously in the genome of eukaryotes [21], and prokaryotes [22], and are regarded as the most variable type of DNA sequence within the viral genome [23, 24]. The microsatellites may be classified as either simple or compound, depending on the constituent of nucleotide sequences. The interruptions present in microsatellite will give rise to interrupted pure, compound, interrupted compound, complex and interrupted complex types. Two or more microsatellites resides directly adjacent to each other to form compound microsatellites by interruption of repeats [25]. Due to their unique characteristics, these SSRs play a major role in meiotic recombination [26, 27, 28], the evolution of species [29], genome mapping [30], differentiation of viral strains [31], studying population genetics [32] and secondary structure formation [33]. Many studies have highlighted the presence of microsatellite repeats in viruses, such as menovirus [34], vesicular stomatitis virus [35], hepatitis C virus [36], and human respiratory syncytial virus [37]. Here, we report for the first time a comparative analysis of microsatellites with respect to the abundance, distribution,

composition, and polymorphism of SSRs within ORFV through the *in silico* approach, followed by the development and characterization of thirteen microsatellites markers. Using these tools, we further tested its usefulness by screening viral genome from an ORFV outbreak and constructing a concatenated phylogenetic tree, which elucidated that the investigated virus is closely related to the Chinese isolate. These markers could be used as a tool for making multiplex PCR assays for virus identification, strain demarcation, and evolutionary analysis.

4.2 Results

4.2.1 Distribution of SSRs and cSSRs in ORFV genome

Our study revealed a large number of SSRs scattered throughout the ORFV genomes varying from 1038 to 1181 in number, with an average of 1092 per genome. The RA and RD ranged from 7.6-8.4 and 53.0-59.5, respectively, in the analyzed ORFV genomes. However, in other DNA viruses such as Human Papilloma Viruses (HPVs), the RA and RD ranged from 3.6-8.3 and 23.9-59.1 [47]. In the case of Herpesviruses, RA and RD occurred to be 4.1-13.3 and 26.9-102.9 [48]. On examining the SSR unit size classes, dinucleotide repeats were found to be most abundant (76.9%), followed by trinucleotide (17.7%) and mononucleotide repeats (4.9%) in all the genomes. Tetranucleotide and hexanucleotide repeats were the least in number and represented 0.4% and 0.2% within the ORFV genome, respectively. There were no SSRs with pentanucleotide repeats observed in the ORFV genome. Approximately 90% and 10% of microsatellite motifs were distributed within coding and noncoding regions.

Among the noncoding region, 4.8% are present in the UTR while 5.4% in the intergenic regions, where functional protein and hypothetical protein occupied 68.8% and 21%, respectively (Figure 4.1). The genome-wide scan revealed the presence of 83–107 cSSRs with an average of 93 occurrences per genome. In the case of compound microsatellite, the calculated RA and RD ranged from 0.6-0.8 and 12.1-17.0. However, in other DNA viruses such as HPVs, RA and RD exhibited 0-1.2 and 0-27.3, whereas, in Herpesviruses, the RA and RD occurred 0.1-1.8 and 2.2-35.1 [47, 48]. Approximately 89.5% and 10.5% of microsatellite motifs were distributed within coding or noncoding regions. Among the noncoding region, 5.0% were represented in the UTR while 5.5% in the intergenic region, where functional protein and hypothetical protein occupied 60.7% and 28.8%, respectively (Figure 4.2).

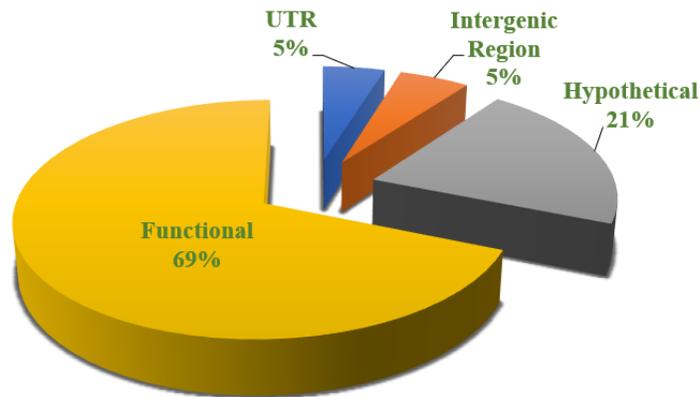


Figure 4.1: Pie-chart illustrating the percentage of differential distribution of SSR within coding (UTR and intergenic region) and non-coding (functional and hypothetical proteins) regions of ORFV genome.

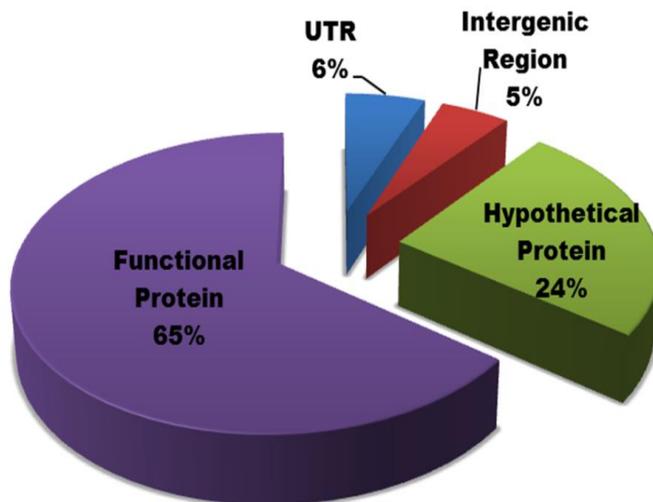


Figure 4.2: Pie-chart illustrating the percentage of differential distribution of cSSR within coding (UTR and intergenic region) and non-coding (functional and hypothetical proteins) regions of ORFV genome.

The percentage of individual microsatellites being part of compound microsatellite (cSSR%) ranged from 7.9-9.0 (Table 4.1). Based on the dMAX value, the maximum distance between any two adjacent microsatellites, and if the distance separating two microsatellites is less than or equivalent to dMAX, than microsatellites are classified as cSSR [49]. To determine the impact of dMAX, all the studied genome sequences were chosen to determine the variability of cSSR with increasing dMAX. The value of dMAX was set between 10 and 100 by Microsatellite Identification Search Analysis (MISA) [50]. Our analysis revealed an overall increase in the number of cSSRs with higher dMAX value and attained a plateau (Figure 4.3).

Sr. No .	Acc. no.	Names of the strains	Year of strain isolation	Size (bp)	Country	Host	GC content (%)	Total no of SSRs	RA	RD	Total no of cSSRs	cRA	cRD	% of cSSR
S1	AY386264	OV-SA00	2004	139962	USA	Goat	63.44	1181	8.43	59.5	107	0.76	16.98	9.06
S2	AY386263	OV-IA82	2004	137241	USA	Lamb	64.33	1089	7.93	55.66	98	0.67	14.51	8.99
S3	DQ184476	NZ2	2006	137820	New Zealand	Sheep	64.34	1082	7.85	55.42	95	0.68	14.11	8.78
S4	HM133903	D1701	2011	134038	Germany	Sheep	63.69	1038	7.74	54.34	83	0.61	12.13	7.99
S5	KF234407	NA11	2015	137080	China	Sheep	63.63	1049	7.65	53.54	87	0.63	12.78	8.29
S6	KP010353	YX	2015	138231	China	Goat	63.75	1099	7.95	55.4	90	0.65	12.89	8.18
S7	KP010354	GO	2018	139866	China	Goat	63.6	1114	7.96	55.61	97	0.69	13.81	8.7
S8	KP010355	NP	2015	132111	China	Goat	63.76	1054	7.97	56.02	86	0.65	12.8	8.15
S9	KP010356	SJ1	2015	139112	China	Goat	63.63	1126	8.09	57.01	99	0.71	13.74	8.79
S10	KY053526	OV-HN3/12	2012	136643	China	Sheep	63.67	1036	7.58	53.04	84	0.61	12.31	8.18
S11	MG712417	SY17	2016	140413	China	Sheep	63.81	1087	7.74	54.28	92	0.65	12.97	8.46

Table 4.1: Overview of microsatellites in ORFV complete genome sequences.

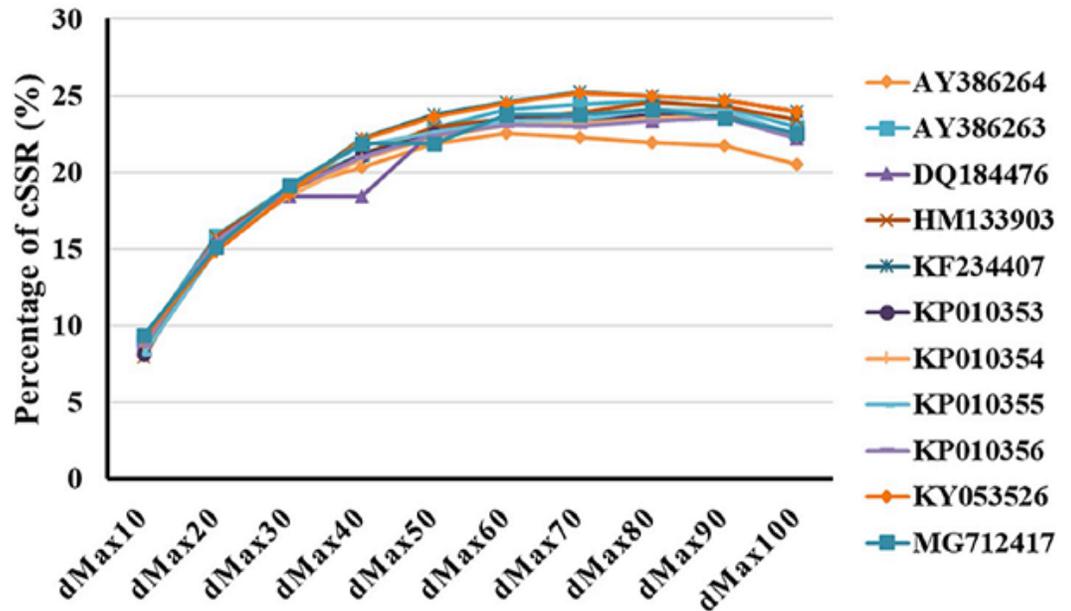


Figure 4.3: Frequency of cSSR in relation to varying dMAX (10–100) across eleven ORFVcomplete genomes represented

on the right side of the graph. A higher cSSR incidence was observed with increasing dMAX in the genomes.

4.2.2 Genomic parameters influencing SSR and cSSR distribution

We tested for the correlation between genome size and GC content with the incidence, RA, RD of SSR and cSSR. Except incidence ($R^2 = 0.6162$, $P > 0.05$), all other parameters such as RA and RD of SSRs had no correlation ($R^2 = 0.002374$, $P > 0.05$; $R^2 = 0.18$, $P < 0.05$) with the genome size and GC content ($R^2 = 0.09377$, $P < 0.05$, $R^2 = 0.00126$, $P > 0.05$; $R^2 = 0.08129$, $P < 0.05$). The regression analysis of cSSR showed significant correlation with the incidence ($R^2 = 0.6483$, $P > 0.05$) and RA ($R^2 = 0.4823$, $P > 0.05$) while displayed non-significant correlation with RD ($R^2 = 0.3759$, $P < 0.05$). On the contrary, the GC content was weakly correlated with the number ($R^2 = 0.02903$, $P > 0.05$), RD ($R^2 = 0.004839$, $P < 0.05$) and RA ($R^2 = 0.03917$, $P < 0.05$) of cSSR.

4.2.3 The frequency of classified repeat types

The overall frequency of mononucleotide repeats A/T (64.1%), dinucleotide repeat motif CG/GC (81.6%) were the most prevalent than poly G/C (35.9%), GA/TC (5.0%), AC/GT(4.5%), AG/CT (3.9%), CA/TG (3.6%) and AT/TA (1.4%), respectively. Analysis of the classified tri-repeat types revealed that the ORFV genome had 30 types of trinucleotide from which CGC/GCG, GCC/GGC, CAG/CTG, AGC/GCT, CCG/CGG were abundantly present, exhibiting 18.2%, 14.5%, 6.3%, 6.2%, and 6.3%, respectively. The most common tetra and hexanucleotide repeats were CGAG/CTCG

4.2.5 Identification of polymorphic microsatellite through *in silico* approach

For a polymorphic microsatellite, the length of the repeat block should be non-identical with that of the other sequences in the database, and this length difference must be a multiple of the repeat unit [20, 31, 52]. For the identification of polymorphic microsatellite, eleven strains of ORFV were used, where (AY386264) acted as the reference. A total thirteen number of polymorphic microsatellites were observed; among these, two were observed within the hypothetical protein, three in the intergenic regions, and the rest eight in the protein-coding/genic regions. The polymorphic genic region containing the microsatellites encodes several important proteins such as Ankyrin repeat protein (ANK protein), DNA-binding phosphoprotein, virion core protein, Granulocyte-Macrophage Colony Stimulating Factor (GM-CSF), Interleukin 10 protein (IL-10protein), Putative serine/threonine- protein kinase protein (Table 4.2). The Circos map provides a clear vision regarding the SSR and cSSR distribution and other related details in ORFV (OV-SA00) genome (Figure 4.5).

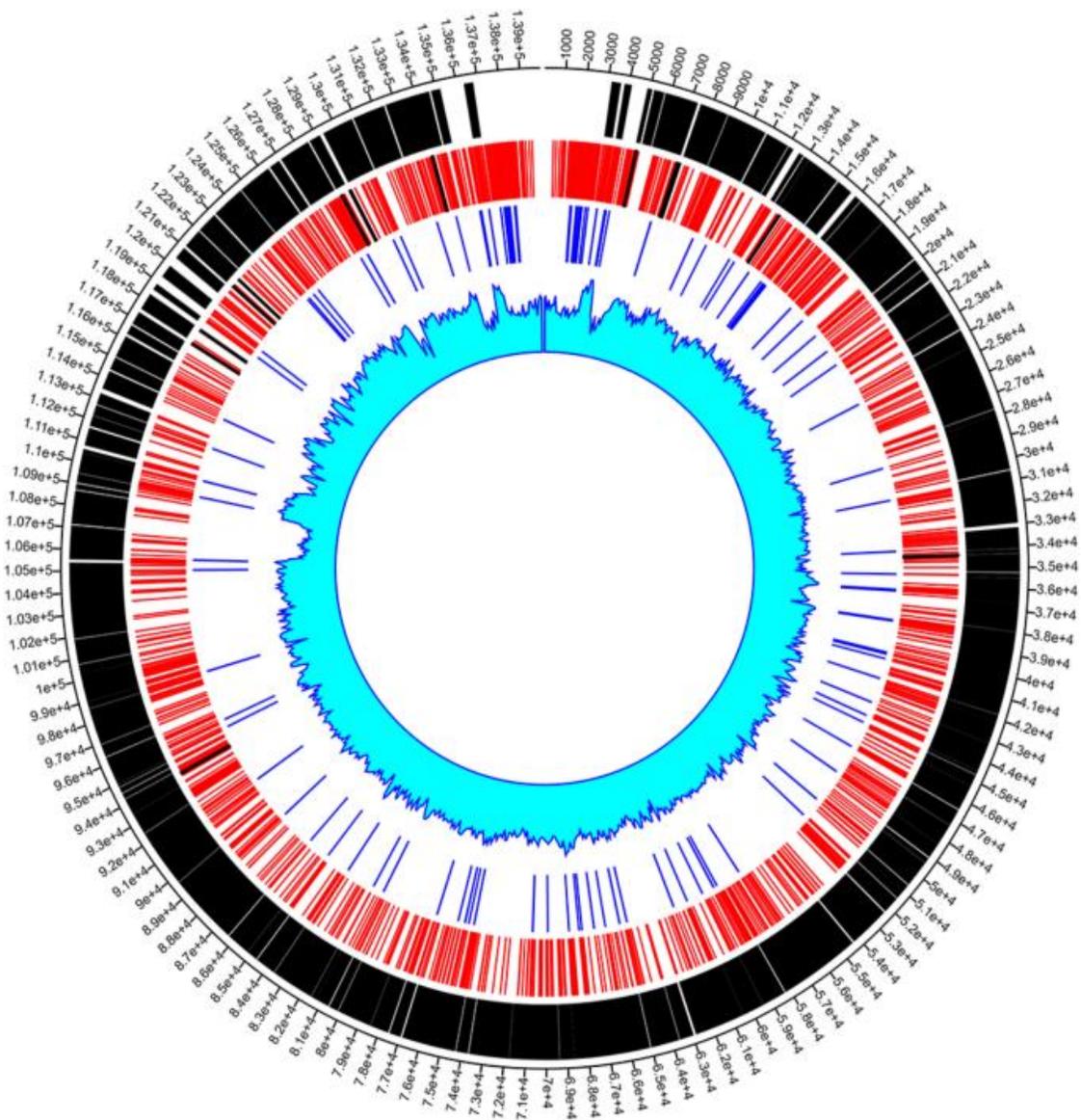


Figure 4.5: Circos Plot showing the Genome size, CDS, Distribution of SSR, selected SSR markers, cSSR, and GC content in ORFV (OV-SA00) genome. From outer track to inner track: Genome size, CDS, SSR, selected SSR markers (Black lines within the SSR), cSSR, and GC content.

Table 4.2: Characteristics of the 13 microsatellite markers developed for the ORFV.

Primer Name	Sequence	Expected size (bp)	Target repeat	Functional region of the genome	ORF	Position in genome	Temp (°C)	No. of variants
ORFV-SSR-1	F-CACCACCATTAACACCACCA R-AAAGGGTTCGCAAGTACACC	166	(CA) ₃	Hypothetical protein	ORF005	4974-4979	55	2
ORFV-SSR-2	F-GACCGTGGCGAGATCCAC R-CACCCTTATTGCCATTACAGC	159	(GGC) ₃	Ankyrin repeat protein	ORF008	7290-7298	55	2
ORFV-SSR-3	F-ATCTTTATGGGCGCTGAATG R-CCCAGTGTAGAGGCCAATTC	151	(A) ₇	Intergenic region		7406-7412	56	3
ORFV-SSR-4	F-ATGAGACAATGCAGACCAG R-GAGCAGACACTGCCTACGAC	130	(CG) ₃	Hypothetical protein	ORF015	13445-13450	58	2
ORFV-SSR-5	F-TCAAAGTCCTCGTCCGAGTT R-CACATTCACCGAGGAGCAG	168	(TAC) ₃	DNA-binding phosphoprotein	ORF032	34352-34360	56	2
ORFV-SSR-6	F-ATGACCTAGAGCCCGTGGAC R-GAGCAGGTCATTCGTGGAG	172	(GAG) ₃	Virion core protein	ORF088	93996-94004	55	2
ORFV-SSR-7	F-GCCGCCACTACTTCAGAAAC R-CTAGAGCCAGCGCAGGTACA	200	(T) ₆	Intergenic region		117434-117439	60	2
ORFV-SSR-8	F-TTTACGTGAAGGCGTTCCT R-TGAGGCACTTCTGGACATC	159	(A) ₆	GM-CSF/IL-2 inhibition factor-like protein	ORF117	118261-118266	58	2
ORFV-SSR-9	F-TTCCTAGGTGCGTTCAGAGG R-GAGCTGTCTGGGATCTCG	155	(CAC) ₃	Ankyrin repeat protein	ORF121	121158-121166	54	2
ORFV-SSR-10	F-TCACTACGAGACCCCTGACC R-AGTGCTTCATTGGGAAGTCG	164	(C) ₆	Ankyrin repeat protein	ORF121	121625-121630	61	2
ORFV-SSR-11	F-CACAGATGCGTATTGTGTTGAG R-TTCAGTTGGTCTTTCATCTGGA	156	(AGT) ₃	IL-10-like protein	ORF127	128736-128744	57	2
ORFV-SSR-12	F-AGTTATCGGTCCGATTCTCG R-GCGCAATACGAGAGTGAACA	150	(AGTTAC) ₃	Intergenic region		129259-129276	55	3
ORFV-SSR-13	F-GTTCTCCCGCTGGATAAATG R-CGAGGAAGACGTCGTACAGC	160	(CGC) ₃	Putative serine/threonine protein kinase	ORF130	134033-134041	55	2

4.2.6 Development and characterization of SSR markers

All clinical samples collected during the outbreak were found to be positive for ORFV tested by producing the desired PCR amplicon size of 140bp (Figure 4.6). We chose all thirteen polymorphic markers to validate in vitro. Hence, PCR was set with each primer sets to amplify DNA isolated from a positive clinical sample. The SSR name, primer sequences, expected size, targeted motif, functional region, protein motif position, gene, ORF number, and annealing temperature, were summarized in Table 2. All the SSR markers produced reliable and reproducible PCR products with the expected molecular size (Figure 4.7). The amplified SSRs were further characterized by sequencing, mapping with the GenBank database through BLASTn and BLASTx. The results of BLASTn alignment revealed a 100% of query coverage and a high identity percentage (91–100%) between the respective sequencing product and their equivalent genes from the published OV-SA00 isolate genome sequence. The results of BLASTx alignment revealed various degrees of query coverage (38–96%) and a high identity percentage (91–100%) with their equivalent amino acid sequences (Table 4.3). The concatenated phylogenetic tree showed the ORFV of our study closely related to Chinese isolate (MG712417) (Figure 4.8). We observed the presence of 2-3 alleles within ORFV genomes.

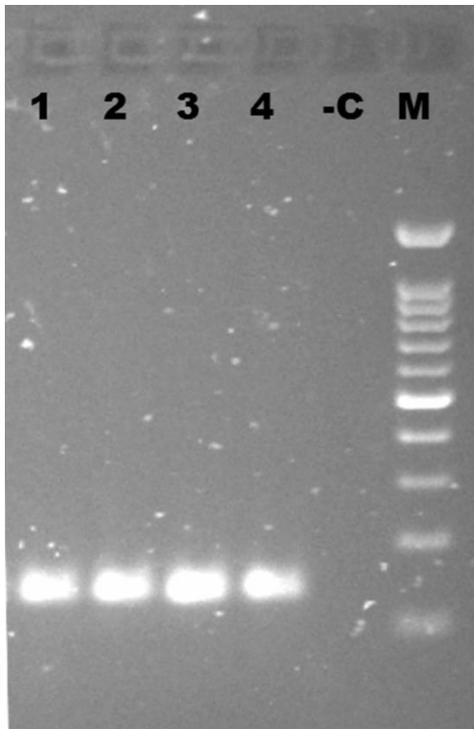


Figure 4.6: Clinical samples evaluation by universal ORFV primers. Electrophoresis gel showing the PCR amplicon of four suspected ORFV clinical samples collected from Black Bengal goats. M: 100 bp DNA ladder; -C: Negative controls (PCR using nuclease-free water as DNA template); 1-4: Clinical samples

SSR	BLASTn			BLASTx		
	Query cover	E value	Identity	Query cover	E value	Identity
ORFV-SSR-1	100%	6.00E-81	96%	52%	0.41	91%
ORFV-SSR-2	100%	5.00E-76	100%	81%	2.00E-19	100%
ORFV-SSR-3	100%	2.00E-50	91%	Intergenic	Intergenic	Intergenic
ORFV-SSR-4	100%	5.00E-60	100%	96%	1.00E-18	100%
ORFV-SSR-5	100%	6.00E-80	99%	55%	3.00E-12	100%
ORFV-SSR-6	100%	1.00E-67	95%	41%	6.00E-05	100%
ORFV-SSR-7	100%	2.00E-85	97%	Intergenic	Intergenic	Intergenic
ORFV-SSR-8	100%	5.00E-65	92%	65%	2.00E-18	100%
ORFV-SSR-9	100%	4.00E-67	97%	67%	5.00E-17	100%
ORFV-SSR-10	100%	6.00E-75	99%	71%	2.00E-07	100%
ORFV-SSR-11	100%	2.00E-74	99%	98%	9.00E-30	100%
ORFV-SSR-12	100%	2.00E-78	92%	Intergenic	Intergenic	Intergenic
ORFV-SSR-13	100%	3.00E-23	100%	38%	1.00E-15	100%

Table 4.3: Alignment of the 13 sequenced microsatellite markers (partial) against the complete genome present in the NCBI database.

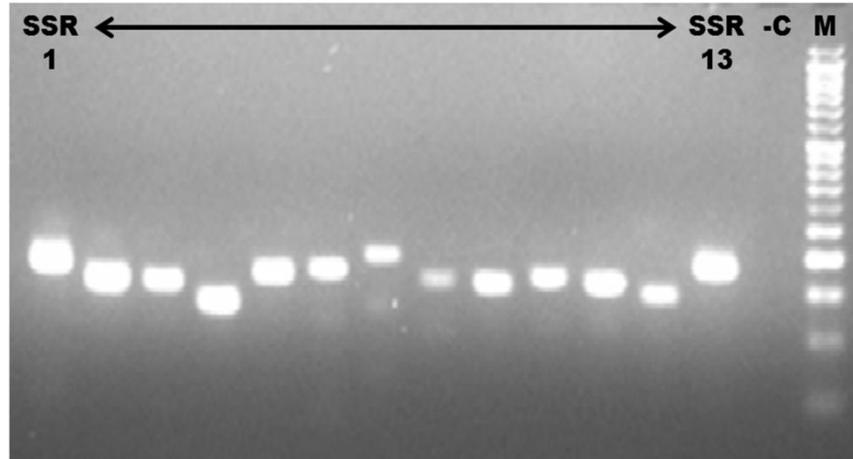


Figure 4.7: Clinical samples validation using SSR markers. Electrophoresis gel showing the PCR amplicon of the developed SSR markers in ORFV. SSR markers from SSR1 to SSR13; M: 50 bp DNA ladder; -C: Negative controls (PCR using nuclease-free water as DNA template).

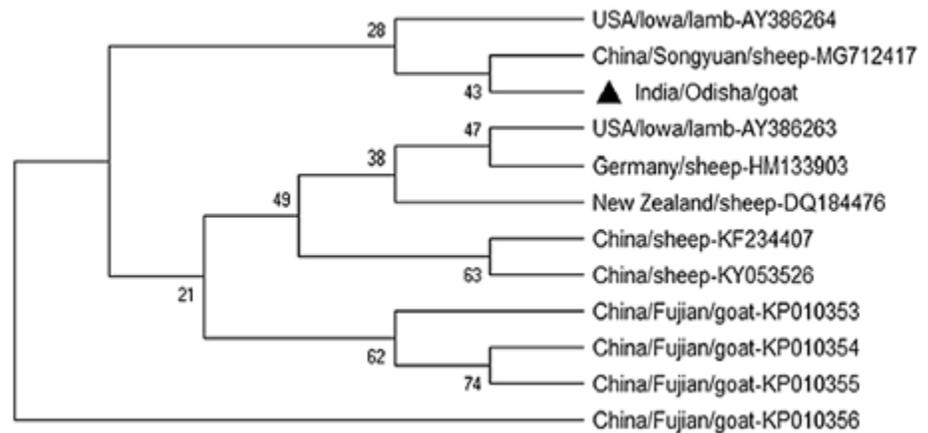


Figure 4.8: The concatenated phylogenetic tree was constructed using the bootstrap consensus tree building method of neighbor-joining with bootstrap value 500 using MEGA 5. The black triangle represents the ORFV isolates of the present investigation showing its relationship with eleven global strains.

4.3 Discussion

Microsatellites, otherwise known as Short Tandem Repeats (STRs) or Variable Number of Tandem Repeats (VNTRs), are being used to discriminate various viruses, such as human cytomegalovirus (hCMV) [23, 24], white spot syndrome virus (WSSV) [53-58], Herpes Simplex Virus Type 1 [31, 59], Herpes Simplex Virus type 2 [60], Herpesvirus 3 [61], Herpesvirus 6 [62], Adenovirus [63], Ostreidherpesvirus 1 [64, 65], Marek's disease virus 1 [66], and Spodopteralittoralis multiple nucleopolyhedrovirus (SpliMNPV) [67] due to its polymorphic in nature. To get insight into the microsatellite in ORFV, we have employed a comparative genomics approach for development and characterization through *in silico* and *in vitro* analysis and validated our findings using sample collected from the recent Orf outbreak for the first time.

The specific parameters, such as its incidence, RA, and RD of SSR and cSSR in ORFV genomes, show abundance variations compared to their genome size and GC content due to the heterogeneity of ORFVs. Until now, limited full-length ORF genomes exist in the database. Based on our analysis, we observed little variation in RA and RD in ORFV. However, in other viruses such as HPVs [47] and Herpes viruses [48], higher variation in RA and RD were reported. The large variation with the parameters was not observed in ORFV, probably due to a lack of enough size difference in the genome. However, few complete genome sequences are available for this virus, in comparison to HPV and herpes viruses, which act as a constraint to get the optimal range. Correlation analysis confirmed that incidence of both SSR and cSSR, RA of cSSR were dependent on genome size but independent of GC content, which was similar to that of HPV

[47], but opposite to HIV [68], potexvirus, carlavirus, and tobamovirus [69-71]. The distribution of microsatellite in the viral genome is pathogen-specific rather than host-specific. The increase of cSSR is predominant when dMAX approaches 10–90bp and further decreases with the increase of dmax (Figure 4.3). This may be due to the occurrence of SSR in the overlapping region of increasing dMAX. We have observed ORFV have more SSR within coding regions than noncoding regions in comparison with other DNA virus, such as herpes simplex virus. This might be due to higher relaxed selection pressure on coding regions in comparison to non-coding region in the respective virus.

The cSSR percentages of ORFV ranges from 7.9 -9.0 %, which is lower in comparison to HIV-1, 0–24.2% [68], gemini virus, 0-27.2% [72], herpes virus, 8.1-33.3% [48]. Generally, the number of compound microsatellites decreases with an increase in complexity [73]. Moreover, the lack of sufficient genomic resources from diverse geographical locations may contribute to a stagnant range of cSSR%. In ORFV, 22.1% of cSSRs were composed of similar motifs, probably contributed by genome duplication. Some study suggests that genome duplication may be helpful for the repeat tendency mechanism [74] which promotes the expansion of genome size such as yeast [75].

In ORFV genomes, the poly A/T repeats were significantly more prevalent than poly G/C repeats, similar to eukaryotic and prokaryotic genomes [21, 76]. The presence of mononucleotide repeats in Mengovirus and encephalomyocarditis virus affect virus growth in murine cell culture [77]. In the case of ORFV, its significance needs further validation. In this study, we also observed the

microsatellite having polymorphism in poly A/T (ORF117), poly C/G (ORF121), within the important immune-regulatory genes, such as granulocyte-macrophage colony-stimulating factor (GM-CSF) and ANK protein, respectively. GM-CSF secreted by a variety of cell types triggers neutrophil, monocyte, and eosinophil myelopoiesis and stimulates early events in immune responses, controlling the differentiation and function of antigen-presenting dendritic cells. IL-2 is a T-cell-derived lymphokine that stimulates T-cell and NK cell activation and proliferation and activated-B-cell proliferation [78, 79]. ANK protein leads to down-regulation of Hypoxia-Induced factor (HIF) activity and regulates energy metabolism, angiogenesis, the apoptotic cascade, the *NFκB* signaling pathway, and cell cycle regulation [80]. The functional effects of this polymorphism in these regions require further investigations.

Dinucleotide AT/TA is more prevalent in most of the ORFV genomes, similar to that of DNA viruses such as HPVs [47], Caulimoviruses, Geminiviruses [52, 81]. AT/TA repeat could form Z conformation or other alternative secondary DNA to facilitate the recombination activity [82, 83]. In our study, the polymorphism within dinucleotide (AC/CA)₃ and (CG/CG)₃ observed within the hypothetical protein. Dinucleotide repeats have the highest slippage rate as compared to any other type of repeats [81]. Among 257 viral genomes examined in a published study, the highest number of dinucleotide SSRs were found when compared to the other types [84]. Dinucleotide repeats are also speculated to be recombination hot spots [85, 86]. In the present study, the presence of higher dinucleotide repeats over tri-nucleotide repeats suggests a possible role of hosts in the evolution of dinucleotide repeats

within poxvirus genomes. Inconsistency frequency of SSR in different accession of the same virus may be attributed to instability because of a higher slippage rate [87].

Trinucleotide motif ATA/TAA/AAT or ATT/TTA/TAT were most prevalent in most genomes of poxvirus, whereas in other DNA virus GAG/AGA was most prevalent in HPVs and AAG/GAA in DNA A and caulimoviruses. The higher density of trinucleotide repeats was observed compared to any other repeat type within coding regions of eukaryotic and prokaryotic genomes [33]. Interestingly, dynamic mutations within trinucleotide repeat responsible for the development of some diseases in humans [88], as well as viral enzymes that interfere Pathogenicity of the Influenza Virus [89]. Our study revealed the presence of tri-nucleotide CGC/GCG and GCC/GGC repeats to be most prevalent than others. The trinucleotide polymorphism was observed in some immunoregulatory genes such as ANK protein (GGC/GCC)₃ (ORF008), IL-10 protein (AGT/ACT)₃ (ORF127) and structural genes virion core protein (GAG/CTC)₃, Putative serine/threonine-protein kinase (CGC/GCG)₃, which needs further functional evaluation.

Three polymorphic SSRs such as (A/T)₇, (T/A)₆, (AGTTAC/GTAACT)₃ were observed within non-coding regions. The microsatellite present within the non-coding regions was evolutionarily neutral and can be utilized as an excellent molecular marker [31]. Finally, we have characterized those polymorphic markers present at non-genic as well as coding (genic) regions. These genic microsatellites, however, may provide adaptive variation important to viral evolution and genetic variability, perhaps similar to the functionally important mononucleotide runs found in VSV

[35] and respiratory syncytial virus [37] and virulence of avian influenza virus encephalo-myocarditis virus [90, 91]. It is noteworthy to mention that, recently, microsatellite present in HSV-1 glycoprotein coding region US4 was useful for strain differentiation. The concatenated tree, which was constructed utilizing sequence information of characterized markers, confirmed that the ORFV of present study is closely related to Chinese isolate (MG712417). Our previous report, as well as several other studies, observed a similar pattern of relationship [19, 92]. We speculate that transboundary and cross-species transfer of ORFV isolates could have resulted in this, as India is geographically adjacent to China. It is interesting to observe the presence of a number of the alleles (2-3) within ORFV genomes indicate the existence of polymorphism within microsatellites, which could act as a useful tool to estimate the diversity [61]. Using a single repeated mononucleotide was able to follow the dynamics of transmission of a human adenovirus during an epidemic [63]. Therefore, microsatellites constitute a potential powerful tool for epidemiological studies of the transmission routes and evolution of ORFV and other related poxviruses. This study provides an important new type of molecular markers useful to investigate questions not only related to epidemiology but also for deciphering the diversity of the virus. However, the characterized microsatellites of the present study are not biased to the particular strain, which indicates the presence of recombinant strains circulating within the Indian subcontinent. This information is not concrete, which requires validation by several whole-genome sequence analysis of ORFV isolates of Indian origin. So far, our understanding of the functional and evolutionary role of

microsatellites in ORFV biology is limited, which needs further in-depth evaluation and possible implementation.

4.4 Materials and methods

4.4.1 Genome sequences

The publicly available eleven complete genome sequences of ORFV isolates obtained from the NCBI database (www.ncbi.nlm.nih.gov) were used for genome-wide in silico microsatellites analysis. To compare genomic sequences of different lengths, we calculated the Relative Density (RD) and Relative Abundance (RA) values. RD is defined as the total length (bp) contributed by each microsatellite per kilobase (kb) of sequence analyzed, whereas; RA is the number of microsatellites present per kb of the genome (kb). Among all the strains, we have chosen OV-SA00 (Acc. number: AY386264) as the reference to evaluate the polymorphism of microsatellites through in silico approach as well as the development of SSRs for Indian origin ORFV (Table 4.1).

4.4.2 Microsatellite identification, investigation, and statistical analysis

For identification of perfect mono, di, tri, tetra, penta, hexa as well as compound microsatellite, IMEx software [38] was utilized. Microsatellites from genomes were extracted using the 'Advance-Mode' of IMEx using the parameters previously used for RNA viruses [39, 40] and DNA viruses [41]. The parameters used were as follows: type of repeat: perfect; repeat size: all; minimum repeat number: 6, 3, 3, 3, 3, 3 for mono, di, tri, tetra, penta, and hexanucleotide repeats, respectively. The maximum distance allowed between any two SSRs (dMAX) is 10 nucleotides. Other parameters were

used as default. Compound microsatellites (cSSR) were not standardized in order to determine real composition.

4.4.3 Multiple sequence alignment and identification of polymorphic SSRs

The microsatellites of OV-SA00 were considered for the identification of polymorphic microsatellites as well as consensus motifs. Sequences were first transferred to BioEdit version 7.2.5 software [42] and aligned by CLUSTAL W [43] module and checked manually for the presence of polymorphism. The Circos plot was generated using the Circos software to map the genome size, CDS, SSR distribution, cSSR distribution, and GC content in ORFV (OV-SA00) genome.

4.4.4 Sample and outbreak data collection

The study did not involve experiments on live vertebrates. Rather, samples were collected from the diseased goats (showing the symptoms of Orf) those reported for veterinary care. Veterinary professionals did sample collection as routine. In October and November 2017, an outbreak of ORFV was noticed within Black Bengal goats in the Eastern-Indian state of Odisha with the geographical location, Cuttack (20.4625° N, 85.8830° E). Tissue samples in the form of scabs from four suspected goats were collected at both infective and recovery/convalescent phase and simultaneously treated for wounds with 2% boro glycerine and parenteral application of Enrofloxacin @ 5mg/kg IM (Figure 4.9). 5g of tissue samples were collected from each animal and subsequently dissolved in phosphate-buffered saline (PBS) of pH 7.2 added with antibiotics and antifungal supplements in a labeled sterile

tube. The homogenized samples were then treated with tissue lysis buffer containing proteinase K, and the mixture was incubated at 56° C overnight. Finally, the mixture was passed through a column, and DNA was purified from the column by using the standard phenol-chloroform method as described by Sambrook et al., 1989, and stored at -20° C until further use. The four suspected samples collected during these outbreaks produced the expected fragment size of 140bp using ORFV specific PCR primers orf1 and orf244 having nucleotide sequences Orf1: cgcagacgtggctgagtacgt and Orf2: tgagctggtggcgctgcct, which confirmed the presence of the virus.



Figure 4.9: ORFV infection in goat. Representative figure depicting clinical cases of ORFV infection in Black Bengal goat having proliferative lesions around the lip recorded in the study area.

4.4.5 Development of polymorphic SSRs

The polymorphic microsatellites identified through in silico approach were further validated through in vitro approach using one ORFV positive clinical sample. Motifs located within defined flanking regions were PCR amplified using specially designed SSR-PCR primer pairs by Primer3Plus web tool (<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/>). The primer length was kept between 18 and 22 bp with product size in the range of 130–200 bp. For proper annealing to the template DNA, the annealing temperature was adjusted between 54 and 61 °C. The thermal cycling conditions for all genes were as follows: initial denaturation step at 95 °C for 5 min, with 35 cycles of denaturation at 95 °C for 50s, with varying annealing temperature for each set of primers (55-61 °C) and extension step at 72 °C for 90s with a final extension at 72 °C for 7 min. PCR amplification was performed in a Thermal Cycler system 2720 (Applied Biosystems, USA). The amplified products were resolved by electrophoresis in a 3% agarose gel. The PCR amplified products, stained with ethidium bromide were visualized and photographed using a Gel Doc™ XR+ System with Image Lab™ Software (Bio-Rad®). Subsequently, the amplified products were purified using QIAquick® purification kit (QIAGEN, USA), and the purified fragments were sent for sequencing using a 3100 ABI sequencer (Applied Biosystems, USA) as described by Sanger et al., 1977 [45]. All sequences obtained were analyzed and verified twice in each direction.

4.4.6 Sequencing data analysis and phylogenetic tree construction

The sequencing results of the developed SSR markers were aligned by using discontinuous-MegaBLAST to identify specific regions among the reads (microsatellites) within the ORFV genome [46]. Next, the sequencing results were subjected to the BLASTx analysis, which compares translational products of the nucleotide query sequence to protein databases (<http://www.ncbi.nlm.nih.gov>). A concatenated phylogenetic tree was constructed using the bootstrap consensus tree building method of neighbor-joining with bootstrap value 500 through MEGA 5 to elucidate the genetic relationship of outbreak sample with the global strains of ORFV.

4.5 Conclusion

The study of microsatellites in the ORFV genome is the first step towards a better understanding of the nature, function, and evolutionary biology of the species. Our preliminary results can be considered as a useful tool in the study of ORFV strain demarcation, diversity estimation, and evolutionary analysis. Our next strategy is to characterize several ORFV strain complete genome from Indian origin through next-generation sequencing to get a better insight into genome organization, development a suitable multiplex panel, which can be utilized as an effective tool for virus identification, genotyping, and evolutionary analysis of that respective virus.

4.6 References

1. Haig D. M., Mercer A. A. (1998), *Ovine diseases*. *Orf*. *Veterinary research*. 29: p. 311-326
2. Gökce H. I., Genc O., Gökce G. (2005), *Sero-prevalence of contagious ecthyma in lambs and humans in Kars, Turkey*. *Turkish Journal of Veterinary and Animal Sciences*. 29: p. 95-101.
3. Hosamani M., Scagliarini A., Bhanuprakash V., McInnes C. J., Singh R. K. (2009), *Orf: an update on current research and future perspectives*. *Expert review of anti-infective therapy*. 7: p. 879-893 (DOI: 10.1586/eri.09.64).
4. Haig D. M., McInnes C. J. (2002), *Immunity and counter-immunity during infection with the parapoxvirus orf virus*. *Virus Research*. 88: p. 3-16, (DOI: 10.1016/s0168-1702(02)00117-x).
5. Haig, D.M., McInnes, C.J., Thomson, J., Wood, A., Bunyan, K. and Mercer, A. (1998), *The orf virus OV20. 0L gene product is involved in interferon resistance and inhibits an interferon-inducible, double-stranded RNA-dependent kinase*. *Immunology*. 93(3): p. 335 (DOI: 10.1046/j.1365-2567.1998.00438.x).
6. Pugh D. G., Baird, N. N. (2012), *Sheep & Goat Medicine-E-Book*. (Elsevier Health Science).
7. Bora, D.P., Barman, N.N., Das, S.K., Bhanuprakash, V., Yogisharadhya, R., Venkatesan, G., Kumar, A., Rajbongshi, G., Khatoon, E., Chakraborty, A. and Bujarbaruah, K.M. (2012), *Identification and phylogenetic analysis of orf viruses isolated from outbreaks in goats of Assam, a northeastern state of*

- India. *Virus Genes*. 45(1): p. 98-104 (DOI: 10.1007/s11262-012-0740-y).
8. Uzel, M., Sasmaz, S., Bakaris, S., Cetinus, E., Bilgic, E., Karaoguz, A., Ozkul, A. and Arican, O. (2005), A viral infection of the hand commonly seen after the feast of sacrifice: human orf (orf of the hand). *Epidemiology & Infection*. 133(4): p. 653-657 (DOI: 10.1017/s0950268805003778).
 9. Duchateau N. C., Aerts O., Lambert J. (2013), Autoinoculation with Orf virus (ecthyma contagiosum). *International Journal of Dermatology*. 53: p. e60-e62 (DOI:10.1111/j.1365-4632.2012.05622.x).
 10. Maki A., Jr. Hinsberg A., Percheson P., Marshall D. G. (1988), Orf: contagious pustular dermatitis. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne*. 139: p. 971-972
 11. Weide, B., Metzler, G., Eigentler, T.K., Fehrenbacher, B., Sönnichsen, K. and Garbe, C. (2009), Inflammatory nodules around the axilla: an uncommon localization of orf virus infection. *Clinical and Experimental Dermatology: Viewpoints in dermatology*. 34(2): p. 240-242 (DOI:10.1111/j.1365-2230.2007.02567.x).
 12. Glass J., Ghali F. E., Sinkre P., Ricotti C. A., Cockerell C. J. (2009), Acute onset of erythematous scalp nodules in a child—quiz case. *Archives of Dermatology*. 145: p. 1053-1058 (DOI: 10.1001/archdermatol.2009.184-a).

13. Kandemir H., Ciftcioglu M. A., Yilmaz E. (2008), Genital ORF. *European journal of dermatology* : EJD.18: p. 460-461, (DOI: 10.1684/ejd.2008.0444).
14. Vogel T. A., Schuttelaar M.L. A. (2013), Generalized orf superinfection in a child with atopic dermatitis. *European Journal of Dermatology*. 23: p. 538-539 (DOI: 10.1684/ejd.2013.2076).
15. Caravaglio J. V., Khachemoune A. (2017), Orf Virus Infection in Humans: A Review With a Focus on Advances in Diagnosis and Treatment. *J Drugs Dermatol*. 16: p. 684-689.
16. Cottone R. (1998), Analysis of genomic rearrangement and subsequent gene deletion of the attenuated Orf virus strain D1701. *Virus Research*. 56: 53-67, (DOI: 10.1016/s0168-1702(98)00056-2).
17. Delhon, G., Tulman, E.R., Afonso, C.L., Lu, Z., De la Concha-Bermejillo, A., Lehmkuhl, H.D., Piccone, M.E., Kutish, G.F. and Rock, D.L. (2004), Genomes of the parapoxviruses ORF virus and bovine papular stomatitis virus. *Journal of virology*, 78(1), p.168-177 (DOI: 10.1128/jvi.78.1.168-177.2004).
18. Mercer, A.A., Ueda, N., Friederichs, S.M., Hofmann, K., Fraser, K.M., Bateman, T. and Fleming, S.B. (2006), Comparative analysis of genome sequences of three isolates of Orf virus reveals unexpected sequence variation. *Virus research*. 116(1-2): p. 146-158 (DOI: 10.1016/j.virusres.2005.09.011).
19. Sahu, B. P., Majee, P., Sahoo A., Nayak D. (2019), Molecular characterization, comparative and evolutionary analysis of the recent Orf outbreaks among goats in the Eastern part of India (Odisha). *Agri*

- Gene. 12: p. 100088 (DOI: 10.1016/j.aggene.2019.100088).
20. Chen, M., Zeng, G., Tan, Z., Jiang, M., Zhang, J., Zhang, C., Lu, L., Lin, Y. and Peng, J. (2011), Compound microsatellites in complete *Escherichia coli* genomes. *FEBS letters*. 585(7): p. 1072-1076 (DOI: 10.1016/j.febslet.2011.03.005).
21. Tóth G., Gáspári Z., Jurka J.(2000), Microsatellites in different eukaryotic genomes: survey and analysis. *Genome research*, 10: p. 967-981 (DOI: 10.1101/gr.10.7.967).
22. Mrázek J., Guo X., Shah A. (2007), Simple sequence repeats in prokaryotic genomes. *Proceedings of the National Academy of Sciences*. 104: p. 8472 (DOI: 10.1073/pnas.0702412104).
23. Davis, C.L., Field, D., Metzgar, D., Saiz, R., Morin, P.A., Smith, I.L., Spector, S.A. and Wills, C. (1999), Numerous length polymorphisms at short tandem repeats in human cytomegalovirus. *Journal of virology*. 73(8): p. 6265-6270 (DOI: 10.1128/JVI.73.8.6265-6270.1999).
24. Walker, A., Petheram, S.J., Ballard, L., Murph, J.R., Demmler, G.J. and Bale, J.F. (2001), Characterization of human cytomegalovirus strains by analysis of short tandem repeat polymorphisms. *Journal of clinical microbiology*. 39(6): p. 2219-2226 (DOI: 10.1128/JCM.39.6.2219-2226.2001).
25. Chambers G. K., MacAvoy E. S. (2000), Microsatellites: consensus and controversy. *Comparative Biochemistry and Physiology Part B*:

- Biochemistry and Molecular Biology. 126: p. 455-476 (DOI: 10.1016/s0305-0491(00)00233-9).
26. Schultes N. P., Szostak J. W. (1991), A poly(dA.dT) tract is a component of the recombination initiation site at the ARG4 locus in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology*. 11: p. 322 (DOI: 10.1128/mcb.11.1.322).
27. Gendrel C. G., Boulet A., Dutreix M. (2000), (CA/GT)(n) microsatellites affect homologous recombination during yeast meiosis. *Genes & development*. 14: p. 1261-1268 (DOI: 10.1101/gad.14.10.1261).
28. Kirkpatrick D. T., Wang Y.H., Dominska M., Griffith J. D., Petes T. D. (1999), Control of Meiotic Recombination and Gene Expression in Yeast by a Simple Repetitive DNA Sequence That Excludes Nucleosomes. *Molecular and Cellular Biology*. 19: p. 7661 (DOI: 10.1128/mcb.19.11.7661).
29. Bowcock, A.M., Ruiz-Linares, A., Tomfohrde, J., Minch, E., Kidd, J.R. and Cavalli-Sforza, L.L. (1994), High resolution of human evolutionary trees with polymorphic microsatellites. *Nature*. 368(6470): p. 455-457. (DOI: 10.1038/368455a0)
30. Sahoo, L., Patel, A., Sahu, B.P., Mitra, S., Meher, P.K., Mahapatra, K.D., Dash, S.K., Jayasankar, P. and Das, P. (2015), Preliminary genetic linkage map of Indian major carp, *Labeo rohita* (Hamilton 1822) based on microsatellite markers. *Journal of genetics*. 94(2): p. 271-277 (DOI: 10.1007/s12041-015-0528-7).
31. Deback, C., Boutolleau, D., Depienne, C., Luyt, C.E., Bonnafous, P., Gautheret-Dejean, A., Garrigue, I. and

- Agut, H. (2009), Utilization of microsatellite polymorphism for differentiating herpes simplex virus type 1 strains. *Journal of clinical microbiology*. 47(3): p. 533-540 (DOI: 10.1128/JCM.01565-08).
32. Sahoo, L., Sahu, B.P., Das, S.P., Swain, S.K., Bej, D., Patel, A., Jayasankar, P. and Das, P. (2014), Limited genetic differentiation in *Labeo rohita* (Hamilton 1822) populations as revealed by microsatellite markers. *Biochemical Systematics and Ecology*. 57: p. 427-431 (DOI: 10.1016/j.bse.2014.09.014).
33. Li Y.C., Korol A. B., Fahima T., Nevo E. (2004), Microsatellites Within Genes: Structure, Function, and Evolution. *Molecular Biology and Evolution*. 21: p. 991-1007 (DOI: 10.1093/molbev/msh073).
34. Duke, G. M., Osorio J. E., Palmenberg A. C. (1990), Attenuation of Mengo virus through genetic engineering of the 5' noncoding poly(C) tract. *Nature*. 343: p. 474 (DOI: 10.1038/343474a0).
35. Barr J. N., Whelan S. P., Wertz G. W. (1997), cis-Acting signals involved in termination of vesicular stomatitis virus mRNA synthesis include the conserved AUAC and the U7 signal for polyadenylation. *Journal of Virology*. 71: p. 8718 (DOI: 10.1128/JVI.71.11.8718-8725.1997).
36. Yamada, N., Tanihara, K., Takada, A., Yoriuzzi, T., Tsutsumi, M., Shimomura, H., Tsuji, T. and Date, T. (1996), Genetic organization and diversity of the 3' noncoding region of the hepatitis C virus genome. *Virology*. 223: 255-261 (DOI: 10.1006/viro.1996.0476).

37. García-Barreno B., Delgado T., Melero J. A. (1994), Oligo(A) sequences of human respiratory syncytial virus G protein gene: assessment of their genetic stability in frameshift mutants. *Journal of Virology*. 68: p. 5460 (DOI: 10.1128/JVI.68.9.5460-5468.1994).
38. Mudunuri S. B., Nagarajaram H. A. (2007), IMEx: Imperfect Microsatellite Extractor. *Bioinformatics*. 23: p. 1181-1187, (DOI: 10.1093/bioinformatics/btm097).
39. Chen, M., Tan, Z., Jiang, J., Li, M., Chen, H., Shen, G. and Yu, R. (2009), Similar distribution of simple sequence repeats in diverse completed Human Immunodeficiency Virus Type 1 genomes. *FEBS Letters*. 583: p. 2959-2963 (DOI: 10.1016/j.febslet.2009.08.004).
40. Alam C. M., George B., Sharfuddin C., Jain S. K., Chakraborty S. (2013), Occurrence and analysis of imperfect microsatellites in diverse potyvirus genomes. *Gene*. 521: p. 238-244 (DOI: 10.1016/j.gene.2013.02.045).
41. Wu X., Zhou L., Zhao X., Tan, Z. (2014), The analysis of microsatellites and compound microsatellites in 56 complete genomes of Herpesvirales. *Gene*. 551: p. 103-109 (DOI: 10.1016/j.gene.2014.08.054).
42. Hall T. A. (2000), in *Nucleic acids symposium series*. 95-98 ([London]: Information Retrieval Ltd. p. c1979-c
43. Thompson J. D., Gibson T. J., Higgins D. G. (2002), Multiple Sequence Alignment Using ClustalW and ClustalX. *Current Protocols in Bioinformatics*. 00: p. 2.3.1-2.3.22 (DOI: 10.1002/0471250953.bi0203s00).

44. Torfason E. G., Guðnadóttir S. (2002), Polymerase chain reaction for laboratory diagnosis of orf virus infections. *Journal of Clinical Virology*. 24: p. 79-84 (DOI: 10.1016/s1386-6532(01)00232-3).
45. Sanger F., Nicklen S., Coulson A. R. (1977), DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*. 74: p. 5463 (DOI: 10.1073/pnas.74.12.5463).
46. Altschul S. F., Gish W., Miller W., Myers E. W., Lipman D. J. (1990), Basic local alignment search tool. *Journal of Molecular Biology*. 215: p. 403-410 (DOI: 10.1016/s0022-2836(05)80360-2).
47. Singh A. K., Alam C. M., Sharfuddin C., Ali S. (2014), Frequency and distribution of simple and compound microsatellites in forty-eight Human papillomavirus (HPV) genomes. *Infection, Genetics and Evolution*. 24: p. 92-98 (DOI: 10.1016/j.meegid.2014.03.010).
48. Ouyang, Q., Zhao, X., Feng, H., Tian, Y., Li, D., Li, M. and Tan, Z. (2012), High GC content of simple sequence repeats in Herpes simplex virus type 1 genome. *Gene*. 499(1): p. 37-40 (DOI: 10.1016/j.gene.2012.02.049).
49. Kofler R., Schlötterer C., Luschützky E., Lelley T. (2008), Survey of microsatellite clustering in eight fully sequenced species sheds light on the origin of compound microsatellites. *BMC Genomics*. 9: p. 612 (DOI: 10.1186/1471-2164-9-612).
50. Beier S., Thiel T., Münch T., Scholz U., Mascher M. (2017), MISA-web: a web server for microsatellite

- prediction. *Bioinformatics*. 33: p. 2583-2585, (DOI: 10.1093/bioinformatics/btx198).
51. Jakupciak J. P., Wells R. D. (1999), Genetic instabilities in (CTG.CAG) repeats occur by recombination. *The Journal of biological chemistry*. 274: p. 23468-23479 (DOI: 10.1074/jbc.274.33.23468)
52. George B., Gnanasekaran P., Jain S. K., Chakraborty S. (2014), Genome wide survey and analysis of small repetitive sequences in caulimoviruses. *Infection, Genetics and Evolution*. 27: p. 15-24, (DOI: 10.1016/j.meegid.2014.06.018).
53. Hoa, T.T., Hodgson, R.A., Oanh, D.T., Phuong, N.T., Preston, N.J. and Walker, P.J. (2005), Genotypic variations in tandem repeat DNA segments between ribonucleotide reductase subunit genes of white spot syndrome virus (WSSV) isolates from Vietnam. *Diseases in Asian Aquaculture V*. p. 339-351.
54. Kiatpathomchai, W., Taweetuntragoon, A., Jittivadhana, K., Wongteerasupaya, C., Boonsaeng, V. and Flegel, T.W. (2005), Target for standard Thai PCR assay identical in 12 white spot syndrome virus (WSSV) types that differ in DNA multiple repeat length. *Journal of Virological Methods*.130: 79-82 (DOI: 10.1016/j.jviromet.2005.06.006).
55. Pradeep B., Shekar M., Gudkovs N., Karunasagar I., Karunasagar I. (2008), Genotyping of white spot syndrome virus prevalent in shrimp farms of India. *Diseases of aquatic organisms*. 78: p. 189-198 (DOI: 10.3354/dao01878).
56. Pradeep B., Shekar M., Karunasagar I., Karunasagar I. (2008), Characterization of variable genomic regions

- of Indian white spot syndrome virus. *Virology*. 376: p. 24-30 (DOI: 10.1016/j.virol.2008.02.037).
57. Tan, Y., Xing, Y., Zhang, H., Feng, Y., Zhou, Y. and Shi, Z.L. (2009), Molecular detection of three shrimp viruses and genetic variation of white spot syndrome virus in Hainan Province, China, in 2007. *Journal of Fish Diseases*. 32: p. 777-784 (DOI: 10.1111/j.1365-2761.2009.01055.x).
58. Kang H. H., Lu C. P. (2007), Comparison of variable region genes of shrimp White Spot Syndrome Virus (WSSV) in different areas in China. *Bing du xue bao = Chinese journal of virology*. 23: p. 490-493.
59. Szpara, M.L., Gatherer, D., Ochoa, A., Greenbaum, B., Dolan, A., Bowden, R.J., Enquist, L.W., Legendre, M. and Davison, A.J. (2014), Evolution and diversity in human herpes simplex virus genomes. *Journal of virology*. 88: p. 1209-1227 (DOI: 10.1128/JVI.01987-13).
60. Burrel, S., Ait-Arkoub, Z., Voujon, D., Deback, C., Abrao, E.P., Agut, H. and Boutolleau, D. (2013), Molecular characterization of herpes simplex virus 2 strains by analysis of microsatellite polymorphism. *Journal of clinical microbiology*. 51(11): p. 3616-3623 (DOI: 10.1128/JCM.01714-13).
61. Avarre, J.C., Madeira, J.P., Santika, A., Zainun, Z., Baud, M., Cabon, J., Caruso, D., Castric, J., Bigarré, L., Engelsma, M. and Maskur, M., 2011. Investigation of Cyprinid herpesvirus-3 genetic diversity by a multi-locus variable number of tandem repeats analysis.

- Journal of Virological Methods. 173: p. 320-327 (DOI:10.1016/j.jviromet.2011.03.002).
62. Achour, A., Malet, I., Deback, C., Bonnafous, P., Boutolleau, D., Gautheret-Dejean, A. and Agut, H. (2009), Length variability of telomeric repeat sequences of human herpesvirus 6 DNA. *Journal of Virological Methods*.159: p. 127-130 (DOI:10.1016/j.jviromet.2009.03.002).
63. Houg, H.S.H., Lott, L., Gong, H., Kuschner, R.A., Lynch, J.A. and Metzgar, D. (2009), Adenovirus microsatellite reveals dynamics of transmission during a recent epidemic of human adenovirus serotype 14 infection. *Journal of clinical microbiology*. 47(7): p. 2243-2248 (DOI: 10.1128/JCM.01659-08).
64. Segarra, A., Pépin, J.F., Arzul, I., Morga, B., Faury, N. and Renault, T. (2010), Detection and description of a particular Ostreid herpesvirus 1 genotype associated with massive mortality outbreaks of Pacific oysters, *Crassostrea gigas*, in France in 2008. *Virus Research*. 153: p. 92-99 (DOI: 10.1016/j.virusres.2010.07.011).
65. Renault, T., Tchaleu, G., Faury, N., Moreau, P., Segarra, A., Barbosa-Solomieu, V. and Lapègue, S. (2014), Genotyping of a microsatellite locus to differentiate clinical Ostreid herpesvirus 1 specimens. *Veterinary research*. 45(1): p. 1-8 (DOI:10.1186/1297-9716-45-3).
66. Spatz S. J., Silva R. F. (2007), Polymorphisms in the repeat long regions of oncogenic and attenuated pathotypes of Marek's disease virus 1. *Virus Genes*. 35: p. 41-53 (DOI:10.1007/s11262-006-0024-5).

67. Atia M. A. M., Osman G. H., Elmenofy W. H. (2016), Genome-wide In Silico Analysis, Characterization and Identification of Microsatellites in *Spodoptera littoralis* Multiple nucleopolyhedrovirus (SpliMNPV). *Scientific Reports*. 6: p. 33741 (DOI:10.1038/srep33741).
68. Chen M., Tan Z., Zeng G., Zeng Z.(2012), Differential distribution of compound microsatellites in various Human Immunodeficiency Virus Type 1 complete genomes. *Infection, Genetics and Evolution*. 12: p. 1452-1457 (DOI: 10.1016/j.meegid.2012.05.006).
69. Alam C. M., Singh A. K., Sharfuddin C., Ali S. (2014), Incidence, complexity and diversity of simple sequence repeats across potexvirus genomes. *Gene*. 537: p. 189-196 (DOI:10.1016/j.gene.2014.01.007).
70. Alam C. M., Singh A. K., Sharfuddin C., Ali S. (2014), Genome-wide scan for analysis of simple and imperfect microsatellites in diverse carlaviruses. *Infection, Genetics and Evolution*. 21: p. 287-294 (DOI: 10.1016/j.meegid.2013.11.018).
71. Alam C. M., Singh A. K., Sharfuddin C., Ali S.(2013), In-silico analysis of simple and imperfect microsatellites in diverse tobamovirus genomes. *Gene*. 530: p. 193-200 (DOI: 10.1016/j.gene.2013.08.046).
72. George B., Mashhood Alam C., Jain S. K., Sharfuddin C., Chakraborty S. (2012), Differential distribution and occurrence of simple sequence repeats in diverse geminivirus genomes. *Virus Genes*. 45: p. 556-566 (DOI: 10.1007/s11262-012-0802-1).

73. Kruglyak S., Durrett R., Schug M. D., Aquadro C. F. (2000), Distribution and Abundance of Microsatellites in the Yeast Genome Can Be Explained by a Balance Between Slippage Events and Point Mutations. *Molecular Biology and Evolution*. 17: p. 1210-1219 (DOI:10.1093/oxfordjournals.molbev.a026404)
74. Fadda Z., Daròs J. A., Flores R., Duran-Vila N. (2003), Identification in eggplant of a variant of citrus exocortis viroid (CEVd) with a 96 nucleotide duplication in the right terminal region of the rod-like secondary structure. *Virus Research*. 97: 145-149 (DOI:10.1016/j.virusres.2003.08.002).
75. Liti G., Louis, E. J. (2005), YEAST EVOLUTION AND COMPARATIVE GENOMICS. *Annual Review of Microbiology*. 59: p. 135-153 (DOI: 10.1146/annurev.micro.59.030804.121400).
76. Karaoglu H., Lee C. M. Y., Meyer W. (2005), Survey of Simple Sequence Repeats in Completed Fungal Genomes. *Molecular Biology and Evolution*. 22: p. 639-649 (DOI:10.1093/molbev/msi057).
77. Martin L. R., Neal Z. C., McBride M. S., Palmenberg A. C. (2000), Mengovirus and encephalomyocarditis virus poly(C) tract lengths can affect virus growth in murine cell culture. *Journal of virology*. 74: p. 3074-3081 (DOI:10.1128/jvi.74.7.3074-3081.2000).
78. McNiece I. (1997), Interleukin-3 and the colony-stimulating factors. *Cytokines in health and disease*, 2nd ed. Marcel Dekker, Ann Arbor. Mich: p. 41-53.
79. Farner N. L. (1997), Interleukin-2: molecular and clinical aspects. *Cytokines in Health and Diseases*

80. Semenza G. L. (2003), Targeting HIF-1 for cancer therapy. *Nature Reviews Cancer*. 3: p. 721 (DOI: 10.1038/nrc1187).
81. Di Rienzo A. *et al.*, (1994), Mutational processes of simple-sequence repeat loci in human populations. *Proceedings of the National Academy of Sciences of the United States of America*. 91: p. 3166-3170 (DOI : 10.1073/pnas.91.8.3166).
82. Karlin S., Campbell A. M., Mrázek J. (1998), Comparative DNA analysis across diverse genomes. *Annu Rev Genet*. 32: p. 185-225, (DOI:10.1146/annurev.genet.32.1.185).
83. Li, D., Zhang, H., Peng, S., Pan, S. and Tan, Z. (2019), Conserved microsatellites may contribute to stem-loop structures in 5', 3' terminals of Ebolavirus genomes. *Biochemical and biophysical research communications*. 514(3): p.726-733 (DOI: 10.1016/j.bbrc.2019.04.192).
84. Zhao, X., Tian, Y., Yang, R., Feng, H., Ouyang, Q., Tian, Y., Tan, Z., Li, M., Niu, Y., Jiang, J. and Shen, G., 2012. Coevolution between simple sequence repeats (SSRs) and virus genome size. *BMC genomics*. 13: 435-435 (DOI:10.1186/1471-2164-13-435).
85. Treco D., Arnheim N. (1986), The evolutionarily conserved repetitive sequence d(TG.AC)_n promotes reciprocal exchange and generates unusual recombinant tetrads during yeast meiosis. *Molecular and cellular biology*. 6: p. 3934-3947 (DOI: 10.1128/mcb.6.11.3934).
86. Wahls W. P., Wallace L. J., Moore P. D. (1990), The Z-DNA motif d(TG)₃₀ promotes reception of

- information during gene conversion events while stimulating homologous recombination in human cells in culture. *Molecular and cellular biology*. 10: p. 785-793 (DOI: 10.1128/MCB.10.2.785).
87. Tachida H., Iizuka, M. (1992), Persistence of repeated sequences that evolve by replication slippage. *Genetics*. 131: p. 471-478.
88. Usdin K. (2008), The biological effects of simple tandem repeats: lessons from the repeat expansion diseases. *Genome research*. 18: p. 1011-1019 (DOI: 10.1101/gr.070409.107).
89. Steinhauer D. A. (1999), Role of hemagglutinin cleavage for the pathogenicity of influenza virus. *Virology*. 258: p. 1-20 (DOI: 10.1006/viro.1999.9716).
90. Hahn H., Palmenberg A. C. (1995), Encephalomyocarditis viruses with short poly(C) tracts are more virulent than their mengovirus counterparts. *Journal of Virology*. 69: p. 2697-91 (DOI: 10.1128/JVI.69.4.2697-2699.1995).
91. Perdue M. L., García M., Senne D., Fraire M. (1997), Virulence-associated sequence duplication at the hemagglutinin cleavage site of avian influenza viruses. *Virus Research*. 49: p. 173-186 (DOI: 10.1016/s0168-1702(97)01468-8).
92. Karki M., Kumar A., Arya S., Ramakrishnan M. A., Venkatesan G. (2019), Poxviral E3L ortholog (Viral Interferon resistance gene) of orf viruses of sheep and goats indicates species-specific clustering with heterogeneity among parapoxviruses. *Cytokine*. 120: p. 15-21 (DOI: 10.1016/j.cyto.2019.04.001).

Chapter 5

Recombination drives the emergence of Orf virus diversity: Evidence from the first complete genome of Indian Orf virus and comparative genomic analysis

5.1 Introduction

Orf or contagious ecthyma is otherwise known as contagious pustular dermatitis, is a neglected zoonotic disease caused by the Orf virus (ORFV), a member of the genus *parapoxvirus*. It primarily affects the sheep and goats, occasionally other ruminants and wild animals, as well as animal handlers, veterinarians, and para veterinarians [1, 2]. The virus is extremely contagious and can exist within the animals' wool and feces for several years [2]. The infection is exhibited with the presence of intensified skin wounds with exacerbated blisters throughout the buccal cavity, often leading to weight loss and anorexia. Ulcers are usually large, proliferative, have 2–4 mm raised crusts, and normally disappear within two months [3]. Severe symptoms are characterized by the expansion of lesions, pneumonia, and arthritis in goats [4]. Early-stage diagnosis may be difficult due to the asymptomatic nature of the pathology [4]. The morbidity rate often reaches up to 100%, resulting in emaciation in adults, and kids, thereby negatively affecting the herd economy [2]. Animal handlers are prone to the zoonotic potential of this virus. The Orfzoonoses are often manifested in the form of unbearable pustules on hands that can expand to other body parts such as genitals and face of affected persons [5-8].

The current report investigates the ORFV epidemic reported in the Black Bengal goats of India. The Black Bengal is an impressive, small body-sized goat breed with distinctive properties such as superb chevon quality, early breeding age, and higher reproductive rate. It performs well in moist, humid, and sub-humid hot climatic conditions [9]. In many tropical countries, sheep and goats seem to be an excellent source for quality meat, dairy products, and wool. The contagious ecthyma primarily caused by the incidence of ORFV followed by secondary bacterial and fungal invasion act as a major constrain for goat searing with a significant economic loss. ORFV outbreaks are being reported across many countries over the past years [10-12]. Since 1999, ORFV outbreaks are observed across different Indian states, majorly in Assam [13], Tripura [14], Odisha [15], Uttarakhand [16], Uttar Pradesh [17], Kashmir Himalayas [18], and Tamil Nadu [19].

The ORFV genome consists of a double-stranded DNA (dsDNA), accommodating almost 130 distinctive genes. Genes in the central region are relatively more conserved and involved in mature virion formation and virus replication. In contrast, genes in the terminal regions are more variable and are often attributed to the virulence, immune modulation, and high frequency of gene recombination [20, 21]. Despite its global distribution, only fourteen complete genomes information are available so far. Information regarding ORFV epidemiology, particularly in central India, is limited. The absence of a complete genome sequence of the Indian isolates makes it difficult to comprehend genetic analysis and thus hinders further functional studies. We, therefore, performed molecular detection of ORFV isolates prevalent in central India and thereby, for the first time, performed the complete

genome analysis of the circulating strain. Using next-generation sequencing (NGS) platform and comparative genomics approaches, we report recombination events along with the phylogenetic and evolutionary status of the identified Indian ORFV isolate.

5.2 Results

5.2.1 Clinical gross pathological changes

Affected animals showed characteristics of Orf lesions such as papules, pustules, and scabs around their lips. Infected goats showed ulcerated or proliferative lesions in the epidermis around the lips. Other pathological indications include anorexia, weight loss, and the appearance of proliferative papillomatous nodules. All infected animals got recovered within 21 to 30 days after the onset of clinical signs. The herd morbidity of the outbreak was recorded by about 20% (n=50) with no mortality. None of the animal handlers of the farm got infected with the virus.

5.2.2 Virus conformation through PCR and sequencing

The ORFV presence was first PCR-verified by four sets of primers targeting ORFV011, ORFV020, ORFV059, and ORFV108 genes. All ten samples produced the expected DNA size. Since samples were collected from an unvaccinated herd, we assume that there is a single strain circulating in the herd. Hence, we sequenced a representative sample and submitted the information to GenBank. Subsequently, gene-specific phylogenetic trees were constructed to infer the genetic relationship and transboundary potential of the circulating

strain by comparing within the country and with other global isolates (Figure 5.1 A-D).

Nucleotide BLAST and phylogenetic analysis based on these four genes confirmed that the present isolate has 99%-100% similarity with the earlier published Indian isolate Ind/Od/Kh/01/2016 [15]. At the global level, the maximum similarity was observed with China/NP/2011, China/FJ-YT2015, China/HuB13/2013, and USA/ORFD/2003 isolates. The concatenated phylogenetic tree analysis indicated that the present isolate Ind/MP/2017 (Ind/MP) is closely related to the China/GO/2012 (Chi/GO) (KP010354) isolate; thus it was further considered as the reference genome for mapping and mutations analysis (Figure 5.2)

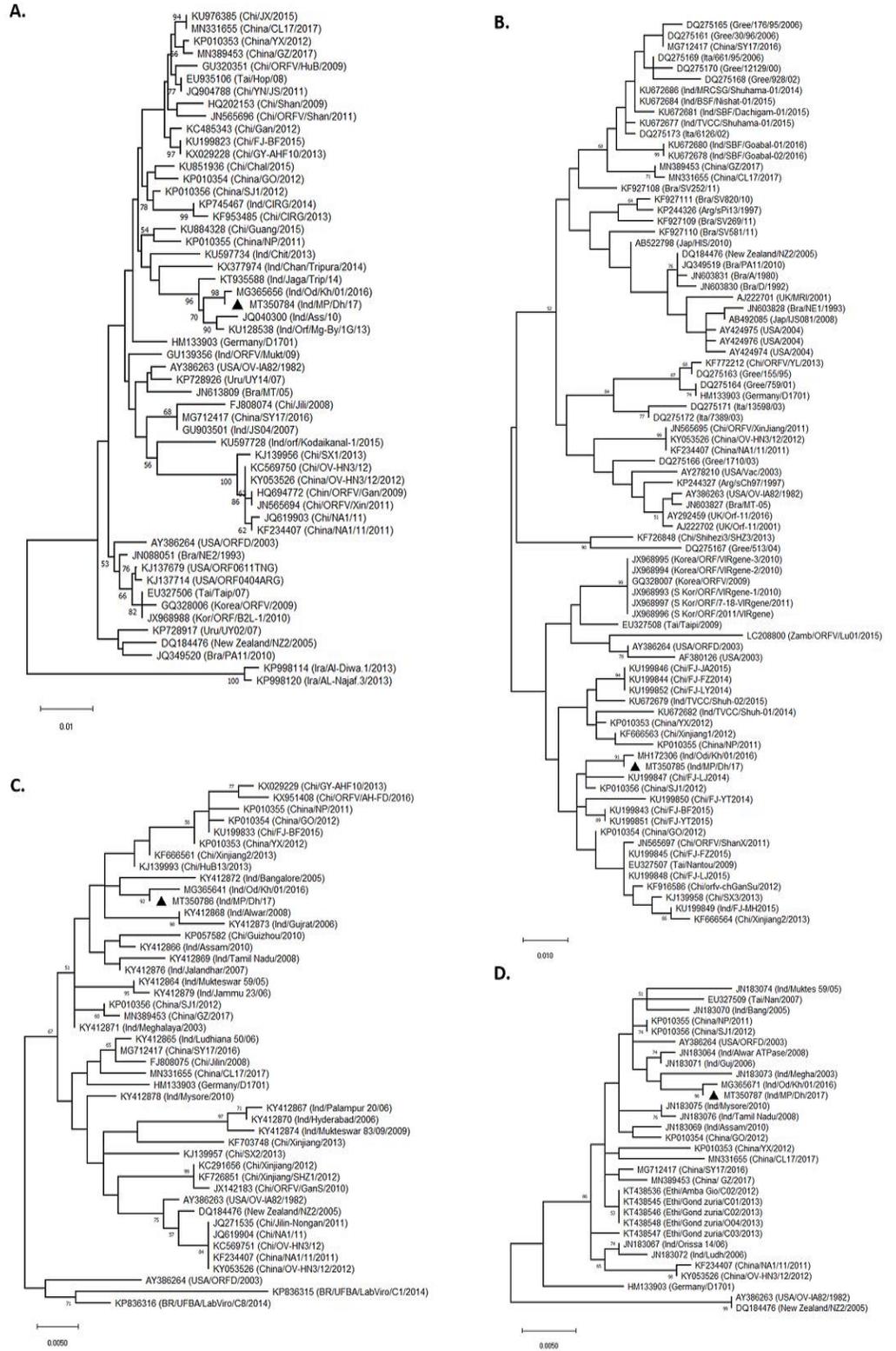


Figure 5.1 A-D: Phylogenetic trees constructed based on the four ORFV genes, namely ORFV011, ORFV020, ORFV059, and ORFV108. The phylogenetic relationship was constructed by the GTR model of the maximum-likelihood method using MEGA 6.0 software. Numbers at the branching points indicate the bootstrap support calculated for 1,000 replicates (A) ORFV011, (B) ORFV020, (C) ORFV059, and (D) ORFV108 based on partial nucleotide sequences. The black triangle represents the new isolate mentioned in our study.

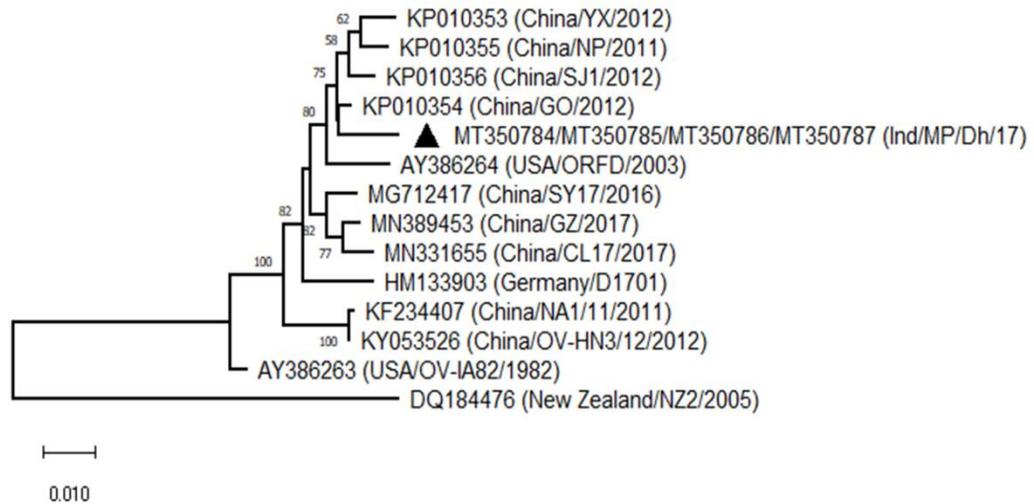


Figure 5.2: Construction of concatenated phylogenetic tree to determine reference genome. Concatenated DNA sequences comprising ORFV011, ORFV020, ORFV059, and ORFV108 genes of ORFV were used to construct a phylogenetic tree with bootstrap value 1000 using the GTR model of maximum likelihood method. The black triangle represents the ORFV isolate of the present investigation showing its close relationship with the strain KP010354, which is considered as the reference.

5.2.3 NGS output and structural analysis of ORFV

On average, ~2.31 GB of high-quality data was generated with 7,934,711 numbers of qualities reads by NextSeq 500 NGS platform. These sequences were aligned with Chi/GOstrain of ORFV, and subsequently, the reference-based assembly was done by using BWA MEM (version 0.7.17). The total length of the assembled genome exhibited 139,807bp in length, and the assigned NCBI accession number is MT332357. Like other PPV genomes, the genome possessed a high (63.7%) G+C content. The genome's left-most nucleotide was arbitrarily designated as base 1, and the starting point of the Inverted Terminal Repeats (ITRs), which spanned throughout ORFV001 and ORFV134 having a total length of 3,910bp. Each ITR is composed of a terminal *Bam*HI site. Telomere resolution motifs are composed of TAAAT, followed by a spacer sequence, ACCCGACC, and six T residues, which form the terminal hairpin loop (Figure 5.3). Using NCBI's ORF Finder tool and NCBI's BLAST (Basic Local Alignment Search Tool), we obtained 132 ORFs for a distinct set of genes. Among these, 65 genes orientated positively and rest 67 have a negative orientation (Figure 5.4). Genes were numbered according to the method described by Delhon *et al.* [22] and Mercer *et al.* [23] in which the newly recognized genes (12.5 and 107.5) were also observed. The BLASTN and BLASTP results for ORFs varied from 63.85% to 100% and 22.73% to 100% with the existing complete genomes of other ORFV GenBank data.



Figure 5.3: Analysis of ITR-ends of ORFV. (A) The left end (5') sequence alignment of 14 complete ORFV genomes consists of a terminal BamHI site (green box) along with telomere resolution motifs (red box) (ATTTTTT-N(8)-TAAAT). (B) The right end (3') sequence alignment of 14 complete ORFV genomes consists of a terminal BamHI site (green box) along with telomere resolution motifs (red box) (ATTTTTT-N(8)-TAAAT).

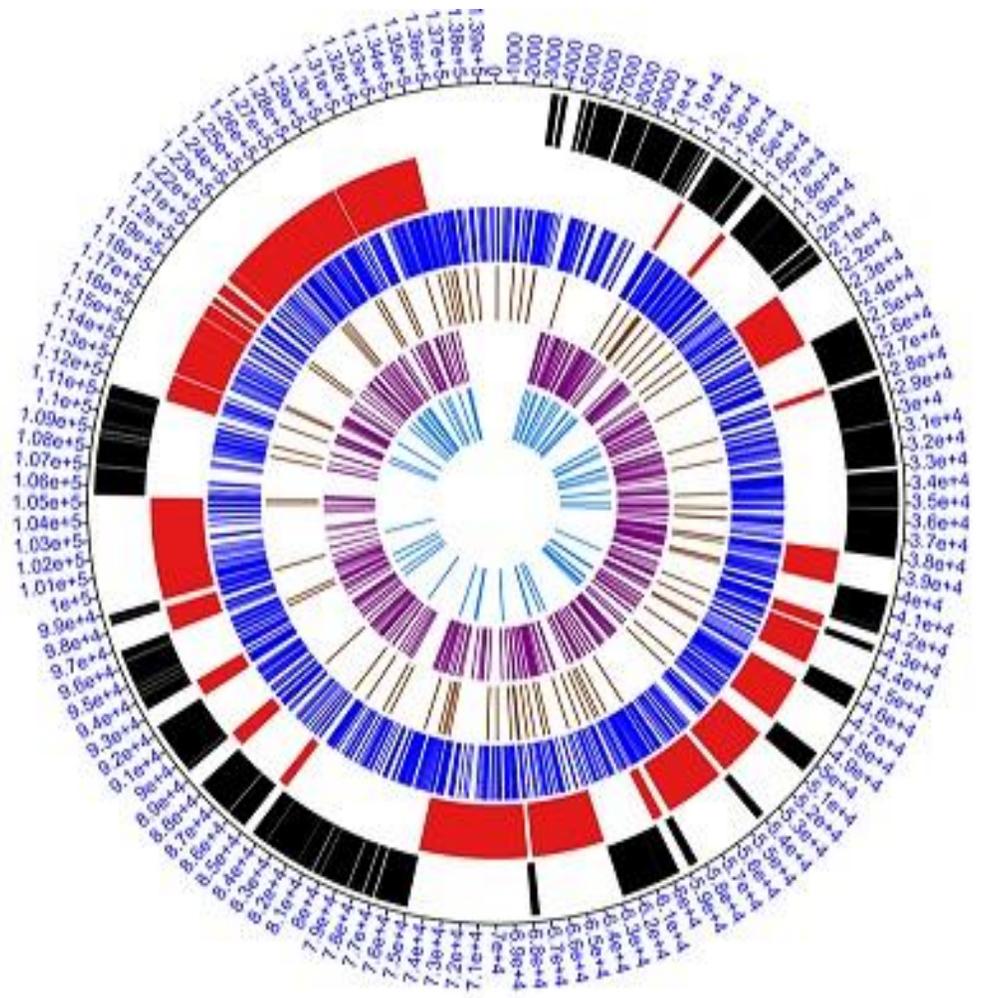


Figure 5.4: *Circos plot illustrating complete genome of the Ind/MP ORFV strain. Schematic presentation of the complete genome: from outer to inner track: ORFV genome size with 1Kb division, genes on a negative-strand (black), genes on a positive-strand (red), SSRs, cSSRs, Synonymous mutation, Nonsynonymous mutation. These mutations, along with SSRs and cSSR, were ubiquitously present throughout the genome irrespective of gene orientation.*

5.2.4. Mutations, DNA polymorphism, and evolutionary analysis

In comparison to the reference genome Chi/GO, our analysis showed nearly 488 unique mutations in the current isolate. Among all the predicted genes, 106 coding regions showed substitution mutations, while 26 genes lacked any mutation (Table 5.1). These observed mutations led to both synonymous and nonsynonymous amino acid substitutions. The highest number of synonymous and nonsynonymous amino acid substitutions was recorded in RNA helicase NPH-II, RNA-polymerase subunit RPO147, virion core protein P4a precursor and EEV maturation protein, Poly (A)-polymerase catalytic subunit PAPL, NF-kappa pathway inhibitor, DNA-binding protein, Ankyrin/F-box protein, respectively. Besides these, several other predicted proteins contained synonymous as well as non-synonymous mutations that support the notion of maintaining the heterogeneity and virulence of this pathogen. Using DnaSP, we observed 11,545 numbers of polymorphic sites within the ORFV genome. The nucleotide diversity (π) and haplotype diversity (Hd) were observed to be 0.02815 and 1.000, respectively. Selection pressure analysis ($\theta=dN/dS$), with a value of 0.02911 revealed that ORFV resides under purifying selection. Tajima's D test of neutrality resulted in a significant negative value (-0.14928), suggesting that this virus might be undergoing a period of evolutionary expansion.

Gene	Gene product	Number unique synonymous substitutions	Number unique non-synonymous substitutions	Non-synonymous substitutions
ORFV001	hypothetical protein	0	0	0
ORFV002	NF-kappa-p65 acetylation inhibitor	0	0	0
ORFV005	hypothetical protein	0	1	Lys ⁵⁰ →Gln
ORFV007	dUTPase	0	1	Thr ⁶ →Met
ORFV008	Ankyrin/F-box protein	10	2	Asn ¹⁸⁵ →Asp, Thr ⁴³⁹ →Ala
ORFV009	hypothetical protein	10	2	Val ¹⁶⁴ →Leu, Ala ³⁰⁸ →Thr
ORFV010	EEV maturation protein	10	3	Asn ⁴⁵ →Asp, Val ²⁹³ →Met, Thr ³⁷⁸ →Ala
ORFV011	EEV envelope phospholipase	6	1	Ser ²⁹ →Gly
ORFV012	hypothetical protein	3	0	
ORFV012.5	hypothetical protein	1	1	Val ¹⁴ →Leu
ORFV013	hypothetical protein	1	0	
ORFV014	RING-H2 motif protein	4	0	
ORFV015	hypothetical protein	4	1	Thr ⁴⁸³ →Ala
ORFV016	hypothetical protein	2	3	Ser ¹⁰⁵ →Gly, Ile ¹⁷⁶ →Met, Tyr ¹⁸⁷ →His
ORFV017	DNA-binding phosphoprotein	0	0	
ORFV018	Poly(A)-polymerase catalytic subunit PAPL	1	3	Ala ³⁶ →Glu, Cys ¹²¹ →Arg, Ile ³² →Val
ORFV019	hypothetical protein	11	0	
ORFV020	DsRNA-binding, interferon resistance	1	0	
ORFV021	RNA-polymerase subunit RPO30	0	0	
ORFV022	Pox virus E6 proteins	4	0	
ORFV023	Membrane protein	4	0	
ORFV024	NF-kappa pathway inhibitor	3	3	Phe ²⁰ →Ser, Ser ³⁴ →Gly, Lys ³⁹ →Glu
ORFV025	DNA-polymerase	10	2	Thr ⁶³⁷ →Ala, Ser ⁸²⁰ →Ala
ORFV026	ERV ALR-like protein (IMV redox protein)	0	0	
ORFV027	Virion core protein	0	1	Asp ⁷⁹ →Glu,
ORFV028	DNA-binding protein	8	3	Thr ¹²¹ →Ala, Ile ³³⁹ →Met, Ser ⁷⁰⁷ →Ala
ORFV029	hypothetical protein	8	2	Thr ²³ →Ile, Gln ⁵⁷ →Arg,
ORFV030	DNA-binding virion protein	7	0	
ORFV031	hypothetical protein	0	0	
ORFV032	DNA-binding phosphoprotein	5	1	Met ⁹⁷ →Ile

ORFV033	IMV membrane protein	3	0	
ORFV034	Telomere-binding protein	5	1	Leu ²⁷⁶ →Met
ORFV035	Virion core protease	6	0	
ORFV036	RNA helicase NPH-II	15	0	
ORFV037	Zn-protease, virion morphogenesis	6	0	
ORFV039	hypothetical protein	0	0	
ORFV038	Late transcription elongation factor	0	0	
ORFV040	Glutaredoxin-like protein	0	0	
ORFV041	hypothetical protein	7	1	Glu ³⁰³ →Ala
ORFV042	RNA-polymerase subunit RPO7	0	0	
ORFV043	hypothetical protein	2	0	
ORFV044	Virion core protein	1	0	
ORFV045	Late transcription factor VLTF-1	2	0	
ORFV046	Myristylated protein	5	3	Thr ¹²³ →Ala, Ala ²¹⁰ →Glu
ORFV047	Myristylated IMV envelope protein	5	0	
ORFV048	hypothetical protein	0	0	
ORFV049	hypothetical protein	2	3	Arg ²²⁰ →Ser, Val ¹⁹⁴ →Met, Ser ¹⁵⁸ →Gly
ORFV050	DNA-binding virion core protein VP8	2	2	Arg ¹⁶⁵ →Lys, Glu ²⁵⁶ →Asp
ORFV051	Membrane protein	2	1	Ser ⁴⁵ →Ala
ORFV052	IMV membrane protein	0	1	Val ⁹³ →Ala
ORFV053	Poly(A)-polymerase small subunit VP39 PAPS	4	0	
ORFV054	RNA-polymerase subunit RPO22	2	0	
ORFV055	Late membrane protein	3	0	
ORFV056	RNA-polymerase subunit RPO147	13	0	
ORFV057	Tyrosine phosphatase, virus assembly	2	0	
ORFV058	IMV, viral entry	2	0	
ORFV059	Immunodominant envelope protein	4	1	Thr ¹¹⁰ →Ala
ORFV060	RNA-polymerase associated protein RAP94	2	1	Ser ⁵⁴⁹ →Gly
ORFV061	Late transcription factor VLTF4	2	2	Asp ⁷⁰ →Glu, Asp ¹³⁴ →Glu
ORFV062	DNA topoisomerase type I	3	0	
ORFV063	hypothetical protein	2	0	
ORFV064	mRNA capping enzyme large subunit	11	0	
ORFV065	Virion protein	1	0	

ORFV066	Virion protein	1	0	
ORFV067	Uracil DNA glycosidase	1	1	Thr ⁸ →Pro
ORFV068	NTPase	2	2	
ORFV069	Early transcription factor	3	3	
ORFV070	RNA-polymerase subunit RPO18	0	2	Met ⁵² →Leu, Asp ¹⁶⁷ →Tyr
ORFV071	NPH-PPH downregulator (NTP pyrophosphohydrolase)	3	0	
ORFV072	Transcription termination factor NPH-I	5	0	
ORFV073	hypothetical protein	3	1	Gln ¹²⁰ →Arg
ORFV074	mRNA capping enzyme small subunit	6	0	
ORFV075	Rifampicin resistance protein	0	0	
ORFV076	Late transcription factor VLTF2	0	0	
ORFV077	Late transcription factor VLTF3	0	0	
ORFV078	Thioredoxin-like protein	0	0	
ORFV079	Virion core protein P4b precursor	8	0	
ORFV080	Virion core protein	2	1	Val ¹⁴⁵ →Ala
ORFV081	RNA-polymerase subunit RPO19	0	0	
ORFV082	hypothetical protein	5	0	
ORFV083	Early transcription factor	15	0	
ORFV084	Intermediate transcription factor VITF-3	1	0	
ORFV085	Late virion membrane protein	0	0	
ORFV086	Virion core protein P4a precursor	13	2	Ser ⁶¹¹ →Ala, Thr ²³⁰ →Ala
ORFV087	Virion formation	0	0	
ORFV088	Virion core protein	4	1	Gly ²⁰⁴ →Ser
ORFV089	Virion membrane protein	1	0	
ORFV090	IMV phosphorylated membrane protein	0	0	
ORFV091	IMV membrane protein	1	0	
ORFV092	hypothetical protein	1	1	Thr ⁶² →Ala
ORFV093	Myristylated protein	3	0	
ORFV094	phosphorylated IMV membrane protein	2	0	
ORFV095	DNA helicase	7	0	
ORFV096	Zn-finger protein	1	0	
ORFV098	hypothetical protein			

ORFV097	DNA-polymerase processivity factor	2	1	Ser ²³⁰ →Ala
ORFV099	Holliday junction resolvase	1	0	
ORFV100	Intermediate transcription factor VITF-3	1	0	
ORFV101	RNA-polymerase subunit RPO132	7	0	
ORFV102	A-type inclusion protein/fusion peptide hybrid	3	0	
ORFV103	A-type inclusion protein	0	0	
ORFV104	Viral fusion peptide	3	0	
ORFV105	IMV surface protein	2	0	
ORFV106	RNA-polymerase subunit RPO35	2	0	
ORFV107	Virion morphogenesis	0	0	
ORFV107.5	hypothetical protein	0	0	
ORFV108	DNA packaging protein ATPase	3	0	
ORFV109	EEV glycoprotein	0	0	
ORFV110	EEV glycoprotein	0	0	
ORFV111	hypothetical protein	2	2	Thr ⁸⁹ →Ala, Ala ¹²⁵ →Ser
ORFV112	chemokine binding protein	3	1	Asp ¹⁶⁷ →Glu
ORFV113	hypothetical protein	0	0	
ORFV114	hypothetical protein	7	0	
ORFV115	hypothetical protein	2	0	
ORFV116	hypothetical protein	0	1	Thr ¹³⁵ →Ala
ORFV117	GM-CSF IL-2 inhibition factor	4	0	
ORFV118	hypothetical protein	1	0	
ORFV119	hypothetical protein	1	0	
ORFV120	hypothetical protein	1	1	Phe ¹⁵⁸ →Leu
ORFV121	NF-kappa pathway inhibitor	4	1	Glu ²⁹³ →Asp
ORFV122	hypothetical protein	6	0	
ORFV123	Ankyri/F-box protein	5	3	Gln ¹⁷ →Arg, Val ¹⁵⁴ →Ala, Ile ²²⁸ →Met
ORFV124	hypothetical protein	6	5	Gly ²⁶ →Ser, Ala ⁸² →Val, Thr ³³¹ →Cys,
ORFV125	Apoptosis inhibitor	2	0	
ORFV126	Ankyrin/F-box protein	1	0	
ORFV127	IL-10-like protein	1	1	Ala ¹⁰ →Val
ORFV128	Ankyrin/F-box protein	7	2	Ala ²⁰⁹ →Asp, Leu ⁴²⁷ →Pro
ORFV129	Ankyrin/F-box protein	7	0	
ORFV130	Serine/threonine protein kinase	3	1	Glu ⁴⁰ →Asp

ORFV131	Membrane protein	2	1	Ser ¹⁷⁸ →Gly
ORFV132	VEGF-like protein	3	1	Leu ⁷ →Ile, Asp ⁸⁴ →Glu, Thr ⁹⁹ →Val
ORFV134	hypothetical protein	0	0	

Table 5.1: *Mutational analysis ORF isolates. The numbers of unique synonymous and unique nonsynonymous amino acid substitutions across all isolates sequenced for this study are presented in the table. Individual nonsynonymous substitutions are also shown in the column on the right of the table.*

5.2.5 Microsatellite and tandem repeat detection analysis

Our study revealed 1,108 and 94 numbers of SSRs and cSSR scattered throughout the ORFV genome. The ORFV genome is rich mostly with dinucleotide repeats (76.5%), followed by trinucleotide (18.14%), and mononucleotide repeats (5.14%). The hexanucleotide SSR repeats were most scarcely present and constituted only 0.18% of the ORFV genome (Figure 5.5). While the tetranucleotide and penta- nucleotide SSR repeats were observed in the ORFV genome. If we look into the microsatellite distribution among the coding and noncoding regions, around 89% SSR motifs were present within the coding and 11% in the noncoding regions, respectively. Focusing more into the SSR distribution in the noncoding region, 4% were distributed within the UTR and 6% in the intergenic regions. Among the coding region constituting the functional proteins and hypothetical proteins, 70% was occupied by the functional ones and 19% by the hypothetical proteins. In the case of cSSR, 88% of the motifs were present in the coding regions and 12% in the non-coding regions. Within the non-coding region, UTR comprised 5.0% of the

cSSRs, and intergenic regions comprised 6% of the total cSSR motifs. cSSR distribution in the coding region was categorized as 67% among the functional proteins and 21% among the hypothetical (predicted) proteins.

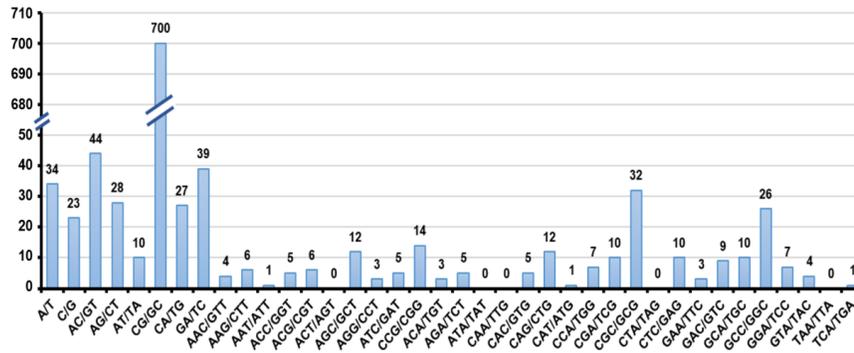


Figure 5.5: Classification of SSR distribution in ORFV genome. Distribution of different motifs of SSRs within ORFV genomes. Dinucleotide repeats were found to be the most abundant, followed by trinucleotide and mononucleotide repeats in the reported genome. Among the di and trinucleotide repeats, CG/GC and CGC/GCG having the highest number among all classified repeats.

5.2.6 Phylogenetic and recombination analysis

A phylogenetic tree based on the complete genomic sequences of 19 poxvirus strains was constructed to reveal the genetic relationship of the present strain with other PPVs and Orthopoxvirus. It revealed that six ORFV strains originating in goats and eight strains belonging to sheep formed two separate clades except for Ger/D1701 with 61-100% bootstrap support. The present ORFV strain showed a close relationship with Chi/GO and USA/ORFD isolates. Our analysis also showed that all ORFVs were more closely related to PPVs

(PCPVs and BPSV) than to Orthopoxvirus (MPVs) (Figure 5.6). We observed a total of 40 potential recombination events with significant *P*-values detected across the ORFV genomes (Table 5.2). A number of recombination events were overlapped within RNA-polymerase subunit (events 1, 17 and 37), Ankyrin/F-box protein (events 4, 5 and 6), hypothetical protein (events 11, 19, 25, 26 and 40) and UTR region (events 3, 9, 10, 14, 23, 31 and 38). Out of 40, potential recombination events, 21 events include Ind/MP as recombinant, minor, and major parental sequences. The recombinant Ind/MP (event 30, 31, 39) were observed within hypothetical proteins, UTR and A-type inclusion protein, using USA/OV-IA82, Chi/SY17, Chi/GZ, Chi/NA1/11 as a minor parent and Chi/NP, Chi/GO, Chi/YX, Chi/GOGer/D1701 as a major parent. The Ind/MP strain acts as minor and major parental sequences within a total of 12 and 6 events for the generation of recombinant sequences Chi/YX, Chi/CL17, Chi/GZ,NZ/NZ2, USA/OV-IA82, and Chi/YX, Chi/SJ1, Ger/D1701, Chi/NP, respectively.

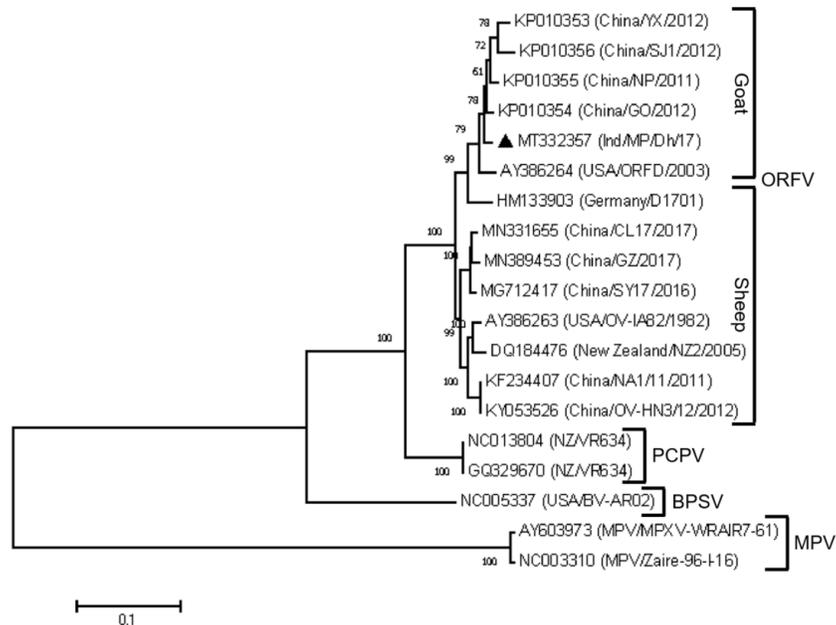


Figure 5.6: Comparison of PPVs by constructing Phylogenetic tree. Nineteen complete genome sequences, including the terminal repetitions, were aligned to construct a phylogenetic tree with bootstrap value 1000 using the GTR model of the maximum likelihood method. The black triangle represents the ORFV isolates of the present investigation showing its relationship with fourteen global strains.

Recombination event number	BP* start in Ind/MP ORFV	BP* end in Ind/MP ORFV	Recombinant gene in ORFV	Recombinant sequence(s)	Minor parental sequence(s)	Major parental sequence(s)	Detection methods	P-value
1	102311	105339	RNA-polymerase subunit RPO132	Chi/YX	Ger/D1701	Chi/NP	RGBMCST	6.93E-316
						Ind/MP		
						Chi/GO		
						USA/ORFD		
2	108034	109113	A-type inclusion protein	Chi/YX	NZ/NZ2	Ind/MP	RGBMCST	6.54E-36
					Chi/CL17	Chi/GO		
					Chi/SY17	USA/ORFD		
					Chi/GZ			
					USA/OV-IA82			
					Chi/NA1/11			
					Chi/OV-HN3/12			
3	1207	1668	UTR	Chi/CL17	Ind/MP	NZ/NZ2	RGBMCST	6.53E-68
				Chi/GZ	Chi/GO			
					Chi/NP	Chi/NA1/11		
					Chi/YX	Chi/OV-HN3/12		
					Chi/SJ1			
4	132207	134268	Ankyrin/F-box protein	Chi/CL17	Chi/GO	USA/OV-IA82	RGC	2.78E-38
				Chi/GZ	Ind/MP	Chi/NA1/11		
					Chi/NP	Chi/OV-HN3/12		
					Chi/YX	Chi/SJ1		
5	130127	131711	Ankyrin/F-box protein	Ger/D1701	Chi/GZ		RGBMCST	5.57E-11
					Chi/CL17			
					Chi/SY17			
					Chi/NA1/11			
					Chi/OV-HN3/12			
6	130025	130883	Ankyrin/F-box protein	USA/OV-IA82	Unknown (Chi/GZ)	NZ/NZ2	RGBMCST	9.78E-15
7	134408	134735	VEGF-like protein	Chi/GZ	Ind/MP	Chi/CL17	RGBMCST	7.42E-29
8	43796	46928	Virion core protein	Chi/CL17	Chi/GO	Chi/SY17	RGC	6.56E-10
9	6793	9301	UTR	Chi/CL17	Chi/YX	Chi/SY17	RGBMCS	7.68E-15

					Chi/GO			
10	6694	9265	UTR	Chi/GZ	Chi/YX	Chi/SY17	RGBMCST	7.86E-10
					Chi/GO			
11	119339	121370	hypothetical protein	Chi/CL17	Chi/YX	Chi/GZ	RGBMCST	9.86E-11
12	2839	5566	UTR	Ger/D1701	Chi/NA1/11	USA/ORFD	RBMC	9.36E-09
					USA/OV-IA82			
					NZ/NZ2			
					Chi/OV-HN3/12			
13	113590	114703	chemokine binding protein	Chi/GZ	USA/ORFD	Chi/CL17	RGBMCS	9.50E-20
					Ind/MP	Chi/SY17		
					Chi/GO			
					Chi/YX			
14	2381	2683	UTR	Chi/SJ1	Chi/GZ	Ind/MP	RBMCST	5.10E-18
					Chi/CL17			
					Chi/SY17			
15	81359	82959	Thioredoxin-like protein	Chi/CL17	Chi/YX	Chi/SY17	RGBMCST	5.93E-07
					Chi/GO			
16	44481	44785	RNA-polymerase subunit RPO7	USA/ORFD	Ger/D1701	Chi/YX	RGBMCST	7.95E-04
					Chi/GO			
					Chi/NP			
					USA/OV-IA82			
17	102268	105087	RNA-polymerase subunit RPO132	NZ/NZ2	Ger/D1701	Chi/NA1/11	RMCT	5.04E-09
					Ind/MP	Chi/OV-HN3/12		
18	38787	41928	RNA helicase NPH-II,Zn-protease	Ger/D1701	USA/OV-IA82	USA/ORFD	RBMC	5.91E-08
					NZ/NZ2	Ind/MP		
					Chi/GO			
					Chi/NP			
					Chi/SJ1			
19	13159	14477	hypothetical protein	Chi/GZ	Chi/YX	Chi/SY17	RGBMCST	5.06E-04
					Chi/GO			
					Chi/NP			
20	15309	16345	DNA-binding phosphoprotein	Chi/GZ	Chi/YX	Chi/CL17	RGBMCS	5.72E-14
					Chi/GO			
21	59685	60433	Immunodominant envelope protein	Ger/D1701	Chi/GZ	Chi/YX	GBMCS	6.45E-13
					Chi/CL17	Chi/NP		
					Chi/SY17			
22	109850	110870	Virion morphogenesis protein, hypothetical protein	Chi/NP	Chi/GZ	Chi/SJ1	RGBMCST	7.99E-08
					Chi/CL17	Ind/MP		
					Chi/SY17			
23	1672	2991	UTR	Chi/CL17	Chi/GZ	Chi/SY17	RGMCS	8.26E-

								09
24	29748	30996	DNA-binding protein	Chi/GZ	Chi/SJ1	Chi/CL17	RGBMCS	5.95E-04
						Chi/SY17		
25	112618	113071	hypothetical protein	Chi/GZ	Ind/MP	Chi/CL17	RGMST	5.64E-06
					Chi/GO			
					Chi/YX	Chi/OV-HN3/12		
26	119241	120473	hypothetical protein	Chi/GZ	Chi/SJ1	Chi/SY17	RBMCS	7.12E-07
27	26828	27986	DNA-polymerase	Chi/GZ	Chi/SJ1	Chi/SY17	RGBMCS	5.20E-04
			IMV redox protein		Ind/MP			
28	81188	82997	Thioredoxin-like protein	Chi/GZ	Chi/SJ1	Chi/SY17	RBMCS	7.99E-03
29	39578	40476	RNA helicase NPH-II	Chi/GZ	USA/ORFD	Chi/SY17	RGBMCS	5.03E-03
30	44373	45334	hypothetical protein	Ind/MP	USA/OV-IA82	Chi/NP	RGBMT	8.96E-09
					Chi/SY17	Chi/GO		
31	5293	6693	UTR	Ind/MP	Chi/GZ	Chi/YX	RGBMCST	5.07E-03
					USA/OV-IA82	Chi/GO		
32	126180	126569	Apoptosis inhibitor	Chi/CL17	Chi/SJ1	Chi/GZ	GBT	6.56E-05
					Chi/NP	Chi/SY17		
					Chi/YX			
33	117522	117772	GM-CSF/IL-2 inhibition factor	Chi/GZ	Ind/MP	Chi/CL17	GBT	3.59E-02
						Chi/SY17		
						NZ/NZ2		
34	97779	101729	Zn-finger protein	Chi/CL17	Chi/YX	Chi/OV-HN3/12	RBMCT	9.87E-03
			DNA-polymerase processivity factor	NZ/NZ2	Ind/MP	Chi/NA1/11		
			Holliday junction resolvase		Chi/GO			
					Chi/SJ1			
					Ger/D1701			
35	65150	67860	mRNA capping enzyme large subunit	USA/OV-IA82	Ind/MP	Chi/NA1/11	RBMT	5.94E-05
						Chi/OV-HN3/12		
36	24856	26088	DNA-polymerase	Chi/GZ	Ind/MP	NZ/NZ2	RBS	2.65E-02
37	105045	105144	RNA-polymerase subunit RPO132	Ger/D1701	Chi/NA1/11	Chi/NP	RGMC	7.09E-04
					Chi/OV-HN3/12	Ind/MP		
						Chi/GO		
						USA/ORFD		
38	7733	8471	UTR	USA/OV-IA82	Ger/D1701	Chi/NA1/11	RBT	8.35E-04
						Chi/OV-HN3/12		

39	105510	106144	A-type inclusion protein	Ind/MP	Chi/NA1/11	Ger/D1701	RBM	7.49E-03
				Chi/GO				
40	12729	14917	RING-H2 motif protein	NZ/NZ2	Ind/MP	Chi/NA1/11	RBT	5.64E-03
			hypothetical protein	USA/OV-IA82				

Table 5.2: Predicted potential recombination events between ORFV. Details recombination events detected between ORFV isolates were analyzed in this study using the RDP4 program. Detection method coding R, G, B, M, C, S, P, L, and T represents methods RDP, GENECONV, Bootscan, MaxChi, Chimaera, SiScan, PhylPro, LARD, and 3Seq, respectively. The P-value for the detection method is shown in bold. To emphasize the recombination potential of Indian isolate, (Ind/MP) has been bold. Ind/MP strain consists of 21 potential events in which it acts as recombinant, minor, and major parental sequences.

5.3 Discussion

Although several reports suggest ORFV prevalence in India, limited information exists on isolates from central India. We characterized the ORFV isolate from the central Indian state of Madhya Pradesh by utilizing conserved partial genes analysis of ORFV011, ORFV020, ORFV059, and ORFV108. We observed a moderate rate of morbidity (20%) and no mortality during this outbreak, similar to the previous report of ORFV outbreak in the North-Eastern state of Assam, India [13]. The sequence and phylogenetic analysis is revealed 99-100% nucleotide similarities with the ORFV isolate of Odisha, a state of Eastern India. The global comparison was consistent with the previous study showing its closeness with the Chinese ORFV strains [17]. This indicates the transboundary

potential of the present virus isolates to spread to the neighboring states or countries. We speculate that trade exchange and transport can be a potential route to establish the epidemiological linkage between the strains isolated from the two distinct geographical areas.

Despite the high number of outbreaks reported from India, no complete genome and associated genomic information are available. Ours is the first ORFV complete genome sequence report from India. NGS was performed with DNA isolated from the clinical samples, similar to the recently adopted strategy reported in other poxviruses, such as *seal parapoxvirus* (SePPV) [24], *canidalphaherpesvirus 1* (CHV-1) [25], *kangaroopox virus* (KPV) [25]. The Chi/GO strain was assigned as the reference genome based on constructing a concatenated phylogenetic tree using all the fourteen complete genomes. All these features were relatively the same as the previously published ORFV genome around the world [26, 27]. Like other poxviruses, ORFV genomes contain a large central coding region bound by two identical inverted terminal repeat (ITR) regions [28, 29]. The current isolate was observed to have the ITRs of 3910 bp length and span throughout the ORFV001 and ORFV134 with conserved telomere resolution sequences at both ends of its genome, consistent with the previous report [30].

Amino acid substitutions within the immune regulatory genes may have altered clinical manifestation and refined antiviral response evaluation [31, 32]. We observed that the highest number of nonsynonymous substitutions within immune regulatory genes such as NF-kappa pathway inhibitor, and Ankyrin/F-box protein, etc. These mutations are likely to

maintain the heterogeneity and mimicked the virulence of this pathogen. Selection pressure analysis ($\theta=dN/dS$) obtained the value 0.02911 suggested that ORFV be under purifying selection. A similar pattern of θ value was obtained from recently studied Avipoxvirus (APV) and ORFV partial genes, which ranged 0.065 -0.200 [15, 33, 34]. This confirmed dN and Ds impel selection pressure to alter the rate of evolution.

Due to its polymorphic nature, microsatellites, otherwise known as Short Tandem Repeats (STRs) or Variable Number of Tandem Repeats (VNTRs), are being used for strain demarcation and estimation of evolutionary distance. Such approaches are being in place for a number of viruses such as human cytomegalovirus (hCMV) [35], white spot syndrome virus (WSSV) [36], Herpes Simplex Virus type 2 [37], Adenovirus [38]. The distribution of classified repeats suggests that dinucleotide GC/CG is more prevalent in most of the ORFV genomes, similar to other DNA viruses like HPVs [39], Caulimoviruses [40], Geminiviruses [41]. Di-nucleotide repeat could form Z-conformation or other alternative secondary DNA to facilitate the recombination activity [42]. Here, the presence of higher dinucleotide relative to trinucleotide repeats suggests a possible host role in the evolution of dinucleotide repeats within poxvirus genomes. By using a single mononucleotide repeat, Houngh et al., could follow the transmission dynamics of a human adenovirus during an epidemic [38]. Therefore, these microsatellites could potentially be used as a powerful tool for epidemiological and evolutionary studies for ORFV.

The phylogenetic tree (of the complete genome) showed the six goat ORFVs and eight sheep ORFVs formed distinctly

separate branches except for Ger/D1701. Interestingly, this result slightly deviated from recently studied phylogenetic analysis based on the complete genome, which forms two host-specific clades [7, 27]. However, our analysis, in comparison to the previous study, showed an increase in the heterogeneity and inability to maintain the perfection of the host-specific clade. This kind of ambiguity was also observed during phylodynamic analysis of the parapoxvirus genus in Mexico (2007–2011), where Ger/D1701, with several other isolates, exhibited a separate clade rather than host-specific clade [34]. Viruses undergo genetic recombination to form new variants in the population by deleting many of their non-essential genes or by acquiring new host genes [20]. Viral genome sequencing elucidates that recombination plays a vital role in understanding human and animal pathogens' evolution, including Vaccinia and Variolaviruses [43, 44]. Remarkably, the comparative genomics approach of ORFV complete genome provides evidence of extensive recombination among the virus. Although several attempts were made to identify recombination events, the lack of complete genome sequences made this study exercise unsuccessful [26]. However, in this study, we identified forty potential recombination events where Ind/MP actively participated in more than 50% of events by forming recombinant as well as major and minor parents. Thus, the Ind/MP strain has the potential to evolve via homologous or non-homologous site-specific recombination and can act as a major or minor parent to form new variants.

5.4 Materials and Methods

5.4.1 Goat herds and tissue collection

The study area is located in the Dhar district of Madhya Pradesh, a central Indian state (75.30E, 22.59 N). Samples were collected from a private farm during January 2017. Samples (n=10) were collected from goats aged between one to eleven months, showing typical Orf skin lesions on their lips, and stored at -80°C for virus isolation and further analysis. Samples were collected from a single herd of unvaccinated goats housed together, and therefore likely to have the same viral strain.

5.4.2 DNA extraction and virus conformation

Total genomic DNA was isolated from the skin tissue according to the protocol described by Sarker et al. 2017 using the DNeasy Blood and tissue purification Kits (QIAGEN, Germany) [25]. A total of 25mg of tissue was aseptically dissected, and chopped and transferred into a microcentrifuge tube. After centrifugation for 2minutes at $800\times g$, the supernatants were subsequently filtered through $5\mu\text{m}$ centrifuge filters (Millipore) and used for viral DNA isolation. The filtrates were RNase A treated by incubating the tube at 56°C for 10min. The viral genomic nucleic (DNA) acids were subsequently extracted using Qiagen binding columns (QIAGEN, Germany).

The viral presence was confirmed by PCR utilizing four sets of primers targeting ORFV011, ORFV020, ORFV059, and ORFV108, commonly known as B2L, E3L, F1L, and A32L. The PCR amplified DNA was purified using the MiniElute gel extraction kit (QIAGEN, Germany) and sent for Sanger sequencing [45]. Prior to NGS, we analyzed four partial gene sequences to get a clear picture of the intra- and inter-strain relationships and constructed a gene-specific phylogenetic

tree. This exercise facilitated the understanding of the transboundary potential and helped us determine the reference genome for mapping, assembly, and mutational analysis. For this, we retrieved sequences of B2L (ORFV011), E3L (ORFV020), F1L (ORFV059), and A32L (ORFV108) genes from GenBank. All phylogenetic trees were constructed using the general-time-reversible (GTR) substitution model for the maximum likelihood (ML) phylogeny with 1,000 bootstrap values using MEGA 6.0 [25, 46]. We also constructed a concatenated phylogenetic tree by taking these genes along with the homologous sequences of complete genomes to determine the reference genome. Virus isolation was attempted by passing clinical samples in African green monkey kidney (Vero) cells and primary lamb testicle cell line [10, 47]. However, the virus could not be recovered until the sixth blind passage. So, we moved ahead with the NGS experiment by isolating viral DNA directly from the clinical samples [24, 25, 48, 49].

5.4.3 Library construction and Illumina NextSeq500 sequencing

The paired-end sequencing library was prepared from the QC-passed viral DNA sample using IlluminaTruSeq Nano DNA Library Prep Kit. Approximately 200 ng of QC-passed DNA was fragmented by Covaris M220 to generate a mean fragment distribution of 350bp. Covaris shearing generates dsDNA fragments with 3' or 5' overhangs. The fragments were then subjected to end-repair. This process converts the overhangs resulting from fragmentation into blunt ends using End Repair Mix. The 3' to 5' exonuclease activity of this mix removes the 3' overhangs, and the 5' to 3' polymerase activity fills in the 5' overhangs followed by adapter ligation to the

fragments. This strategy ensures a low rate of chimera (concatenated template) formation. The ligated products were size-selected using AMPure XP beads (Invitrogen™, USA). The size-selected products were PCR amplified following temperature cycling profile: 72 °C initial denaturation for 3 min; 95 °C denaturation for 30 s; 12 cycles of 95 °C for 10 s, 55 °C for 30 s, and 72 °C for 30 s; and 72 °C final extension step for 5min with the index primer as described in the kit protocol. Indexing adapters were ligated to the ends of the DNA fragments, preparing them for hybridization onto a flow cell. The PCR enriched library was analyzed on the 4200 Tape Station system (Agilent Technologies) using high sensitivity D1000 Screen tape as per the manufacturer's instructions. Cluster generation and sequencing of the pooled DNA-library was performed as paired-end on NextSeq500 sequencing chemistry according to the manufacturer's instructions.

5.4.4 Raw sequence data processing, mapping, assembly, and genome annotations

The sequenced raw data were processed to obtain high-quality clean reads using the Trimmomatic v0.38 platform to remove adapter sequences, ambiguous reads, and host sequences [50]. The high-quality reads were aligned to the reference genome using BWA MEM software (version 0.7.17) [51]. Consensus sequences were extracted using SAM tools [52]. The ORFV genome was annotated using GATU to capture all the potential open reading frames (ORFs) [53]. Intergenic regions were further checked for the presence of ORFs using the BWA MEM analysis tool (version 0.7.17). All the ORFs were subsequently extracted into a FASTA file, and similarity searches, including nucleotide (BLASTN) and protein (BLASTP), were performed. Annotation of ORFs as potential

genes were established based on shared significant sequence similarity to the known viral or cellular genes (BLAST E value $\leq e^{-4}$) or contained a putative conserved domain as predicted by BLASTP. The final ORFV annotations were evaluated with other globally available ORFV isolates to determine the correct methionine start site, assign the proper stop codons, assign the truncation, and validate overlaps.

5.4.5 Mutations, DNA polymorphism, and evolutionary analysis

Mutations were identified in the newly assembled genome by manually comparing with the reference genome, utilizing the BioEdit and ExPaSy tools. This led to the identification of unique nucleotide mutations with synonymous or nonsynonymous amino acid changes. Next, by taking into account all the fourteen available complete genome sequences, we investigated for identification of polymorphic sites and different evolutionary indices like haplotype (gene) diversity, nucleotide diversity (π), selection pressure (θ), Tajima's D or neutrality test utilizing DnaSP. The impact of selection pressure over ORFV was measured by considering the ratio ($\theta = dN/dS$) of nonsynonymous (dN) and synonymous (dS) substitution employing DnaSP with a bootstrap value of 100 replicates. The Tajima's D or neutrality test was evaluated using the number of polymorphic sites and the standard differences within the nucleotide number [54].

5.4.6 Detection of Simple Sequence Repeats (SSRs) or Microsatellites

Identification of perfect mono, di, tri, tetra, penta, hexa as well as compound microsatellites was made by IMEx software

[55]. Microsatellites from genomes were extracted using the ‘Advance-Mode’ of IMEx using the parameters previously used for RNA viruses [56] and DNA viruses [57]. The parameters used were as follows: type of repeat: perfect; repeat size: all; minimum repeat number: 6, 3, 3, 3, 3, 3 for mono, di, tri, tetra, penta, and hexanucleotide repeats, respectively. The maximum distance allowed between any two SSRs (dMAX) was 10 nucleotides. Other parameters were used as default. Compound microsatellites (cSSR) were not standardized in order to determine real composition.

5.4.7 Phylogenetic and recombination analysis

We retrieved fourteen available ORFV complete genome sequences for the GenBank database. Additionally, five sequences of Parapoxvirus (PPV) and Orthopoxvirus, consisting of two Pseudocowpoxvirus (PCPV), one Bovine papular stomatitis virus (BPSV), and two Monkeypox virus (MPV), respectively, were retrieved from the GenBank. These sequences were aligned by using MAFFT (version 7) [58] and then manually edited using the BioEdit (version 7.2) [59] tool. MEGA 6 was used to create a phylogenetic tree based on the GTR substitution model for the ML phylogeny with 1000 bootstrap [46]. To understand the source of genetic variation among all the ORFV complete genomes, we looked for evidence of recombination using the RDP, GENECONV, Bootscan, MaxChi, Chimaera, Siscan, PhylPro, LARD, and 3Seq methods contained in the RDP4 program [60]. Events that were detected by at least three of the aforesaid methods with significant p-values were considered plausible recombinant events.

5.5 Conclusion

In conclusion, we report the complete genome of circulating ORFV isolate from central India. Additionally, we estimated the transboundary and evolutionary potential of this isolate. Subsequently, by in-depth analysis through a comparative genomic approach, we propose future recombination events to drive ORFV evolution and generation of new isolates in the region and globally. We hope that the current genomic information would be greatly useful for further understanding of ORFV biology, epidemiology, and research carried in front of diagnosis and vaccine development.

5.6 References

1. Guo J., Rasmussen J., Wünschmann A., de La Concha-Bermejillo A. J. V. m. (2004), Genetic characterization of orf viruses isolated from various ruminant species of a zoo. 99: p. 81-92 (DOI: 10.1016/j.vetmic.2003.11.010).
2. Haig D. M., Mercer A. A., (1998), Ovine diseases. Orf. Veterinary research. 29: p. 311-326
3. Gökce H. I., Genc O., Gökce G. (2005), Seroprevalence of contagious ecthyma in lambs and humans in Kars, Turkey. Turkish Journal of Veterinary and Animal Sciences. 29: p. 95-101
4. De La Concha-Bermejillo A., Guo J., Zhang Z., Waldron D. J. J. (2003), o. v. d. i. Severe persistent orf in young goats. 15: p. 423-431 (DOI: 10.1177/104063870301500504).
5. Duchateau N. C., Aerts O., Lambert J. (2013), Autoinoculation with Orf virus (ecthyma contagiosum). International Journal of Dermatology.

- 53: p. e60-e62 (DOI:10.1111/j.1365-4632.2012.05622.x).
6. Rajkomar V., Hannah M., Coulson I., Owen C. J. C. (2016), dermatology, e. A case of human to human transmission of orf between mother and child. 41, p. 60-63 (DOI: 10.1111/ced.12697).
 7. Andreani, J., Fongue, J., Khalil, J.Y.B., David, L., Mougari, S., Le Bideau, M., Abrahão, J., Berbis, P. and La Scola, B. (2019), Human Infection with Orf Virus and Description of Its Whole Genome, France, 2017. *Emerging infectious diseases*. 25(12): p. 2197 (DOI: 10.3201/eid2512.181513).
 8. Turk B. G., Senturk B., Dereli T., Yaman B. (2014), A rare human-to-human transmission of orf. *Int J Dermatol*.53: p. e63-65 (DOI:10.1111/j.1365-4632.2012.05669.x).
 9. Acharya R., (1982), Sheep and goat breeds of India. (Food and Agriculture Organization of the United Nations).
 10. Kottaridi, C., Nomikou, K., Teodori, L., Savini, G., Lelli, R., Markoulatos, P. and Mangana, O. (2006), Phylogenetic correlation of Greek and Italian orf virus isolates based on VIR gene. *Veterinary microbiology*. 116(4): p. 310-316 (DOI: 10.1016/j.vetmic.2006.04.020).
 11. Billinis C., Mavrogianni V. S., Spyrou V., Fthenakis G. C. J. V. j. (2012), Phylogenetic analysis of strains of Orf virus isolated from two outbreaks of the disease in sheep in Greece. 9: p. 24 (DOI: 10.1186/1743-422X-9-24).

12. Li, H., Zhu, X., Zheng, Y., Wang, S., Liu, Z., Dou, Y., Li, H., Cai, X. and Luo, X. (2013), Phylogenetic analysis of two Chinese orf virus isolates based on sequences of B2L and VIR genes. *Archives of virology*. 158(7): p. 1477-1485 (DOI: 10.1007/s00705-013-1641-7).
13. Bora, D.P., Barman, N.N., Das, S.K., Bhanuprakash, V., Yogisharadhya, R., Venkatesan, G., Kumar, A., Rajbongshi, G., Khatoon, E., Chakraborty, A. and Bujarbaruah, K.M. (2012), Identification and phylogenetic analysis of orf viruses isolated from outbreaks in goats of Assam, a northeastern state of India. *Virus Genes*. 45(1): p. 98-104 (DOI: 10.1007/s11262-012-0740-y).
14. Venkatesan, G., De, A., Arya, S., Kumar, A., Muthuchelvan, D., Debnath, B.C., Dutta, T.K., Hemadri, D. and Pandey, A.B. (2018), Molecular evidence and phylogenetic analysis of orf virus isolates from outbreaks in Tripura state of North-East India. *VirusDisease*. 29(2): p. 216-220 (DOI: 10.1007/s13337-018-0442-8).
15. Sahu B. P., Majee P., Sahoo A., Nayak D. J. A. G. (2019), Molecular characterization, comparative and evolutionary analysis of the recent Orf outbreaks among goats in the Eastern part of India (Odisha). 12: p. 100088 (DOI: 10.1016/j.aggene.2019).
16. Venkatesan, G., Balamurugan, V., Bora, D.P., Yogisharadhya, R., Prabhu, M. and Bhanuprakash, V. (2011), Sequence and phylogenetic analyses of an Indian isolate of orf virus from sheep. *Vet Ital*. 47(3): p. 323-332 .

17. Kumar, N., Wadhwa, A., Chaubey, K.K., Singh, S.V., Gupta, S., Sharma, S., Sharma, D.K., Singh, M.K. and Mishra, A.K. (2014), Isolation and phylogenetic analysis of an orf virus from sheep in Makhdoom, India. *Virus Genes*. 48(2): p. 312-319 (DOI: 10.1007/s11262-013-1025-9).
18. Ahanger, S.A., Parveen, R., Nazki, S., Dar, Z., Dar, T., Dar, K.H., Dar, A., Rai, N. and Dar, P. (2018), Detection and phylogenetic analysis of Orf virus in Kashmir Himalayas. *Virus disease*. 29(3): p. 405-410 (DOI 10.1007/s13337-018-0473-1).
19. Nagarajan, G., Pourouchottamane, R., Reddy, G.M., Yogisharadhya, R., Sumana, K., Rajapandi, S., Murali, G., Thirumaran, S.M.K., Mallick, P.K. and Rajendiran, A.S. (2019), Molecular characterization of Orf virus isolates from Kodai hills, Tamil Nadu, India. *Veterinary world*. 12(7): p. 1022 (DOI: 10.14202/vetworld.2019.1022-1027).
20. Fleming S. B., Lyttle D. J., Sullivan J. T., Mercer A. A., Robinson, A. J. J. (1995), Genomic analysis of a transposition-deletion variant of orf virus reveals a 3.3 kbp region of non-essential DNA. 76: p. 2969-2978 (DOI: 10.1099/0022-1317-76-12-2969).
21. Cottone, R., Büttner, M., Bauer, B., Henkel, M., Hettich, E. and Rziha, H.J. (1998), Analysis of genomic rearrangement and subsequent gene deletion of the attenuated Orf virus strain D1701. *Virus research*. 56(1): p. 53-67. *Virus Research*. 56: p. 53-67 (DOI: 10.1016/s0168-1702(98)00056-2).
22. Delhon, G., Tulman, E.R., Afonso, C.L., Lu, Z., De La Concha-Bermejillo, A., Lehmkuhl, H.D., Piccone,

- M.E., Kutish, G.F. and Rock, D.L. (2004), Genomes of the parapoxviruses ORF virus and bovine papular stomatitis virus. *Journal of virology*. 78(1): p. 168-177 (DOI: 10.1128/jvi.78.1.168-177.2004).
23. Mercer, A.A., Ueda, N., Friederichs, S.M., Hofmann, K., Fraser, K.M., Bateman, T. and Fleming, S.B. (2006), Comparative analysis of genome sequences of three isolates of Orf virus reveals unexpected sequence variation. *Virus research*. 116(1-2): p. 146-158, (DOI: 10.1016/j.virusres.2005.09.011).
24. Günther, T., Haas, L., Alawi, M., Wohlsein, P., Marks, J., Grundhoff, A., Becher, P. and Fischer, N. (2017), Recovery of the first full-length genome sequence of a parapoxvirus directly from a clinical sample. *Scientific reports*. 7(1): p.1-8 (DOI: 10.1038/s41598-017-03997-y).
25. Sarker, S., Roberts, H.K., Tidd, N., Ault, S., Ladmore, G., Peters, A., Forwood, J.K., Helbig, K. and Raidal, S.R. (2017), Molecular and microscopic characterization of a novel Eastern grey kangaropox virus genome directly from a clinical sample. *Scientific reports*. 7(1): p. 1-13 (DOI: 10.1038/s41598-017-16775-7).
26. Chi, X., Zeng, X., Li, W., Hao, W., Li, M., Huang, X., Huang, Y., Rock, D.L., Luo, S. and Wang, S. (2015), Genome analysis of orf virus isolates from goats in the Fujian Province of southern China. *Frontiers in microbiology*. 6: p. 1135 (DOI: 10.3389/fmicb.2015.01135).
27. Zhong, J., Guan, J., Zhou, Y., Cui, S., Wang, Z., Zhou, S., Xu, M., Wei, X., Gao, Y., Zhai, S. and Song, D.

- (2019), Genomic characterization of two Orf virus isolates from Jilin province in China. *Virus genes*. 55(4): p. 490-501 (DOI: 10.1007/s11262-019-01666-y).
28. Mercer A. A., Fraser K., Barns G., Robinson A. J. J. V. (1987), The structure and cloning of orf virus DNA. 157: p. 1-12 (DOI: 10.1016/0042-6822(87)90307-2).
29. Fraser K. M., Hill D. F., Mercer A. A., Robinson A. J. J. V. (1990), Sequence analysis of the inverted terminal repetition in the genome of the parapoxvirus, orf virus. 176: p. 379-389 (DOI: 10.1016/0042-6822(90)90008-f).
30. Merchlinsky M. J. J. o. v. (1990), Mutational analysis of the resolution sequence of vaccinia virus DNA: essential sequence consists of two separate AT-rich regions highly conserved among poxviruses. 64: p. 5029-5035 (DOI: 10.1128/JVI.64.10.5029-5035.1990).
31. Filer, C.W., Ramji, J.V., Allen, G.D., Brown, T.A., Fowles, S.E., Hollis, F.J. and Mort, E.E. (1995), Metabolic and pharmacokinetic studies following oral administration of famciclovir to the rat and dog. *Xenobiotica*. 25(5): p. 477-490 (DOI: 10.3109/00498259509061867).
32. Thomasy S. M., Maggs D. J. J. V. o. (2016), A review of antiviral drugs and other compounds with activity against feline herpesvirus type 1. 19: p. 119-130 (DOI: 10.1111/vop.12375).
33. Le Loc'h, G., Bertagnoli S., Ducatez M. F. (2015), Time scale evolution of avipoxviruses. *Infection, Genetics and Evolution*. 35: p. 75-81 (DOI: 10.1016/j.meegid.2015.07.031).

34. Velazquez-Salinas, L., Ramirez-Medina, E., Bracht, A.J., Hole, K., Brito, B.P., Gladue, D.P. and Carrillo, C. (2018), Phylodynamics of parapoxvirus genus in Mexico (2007–2011). *Infection, Genetics and Evolution*. 65: p.12-14. 65 (DOI: 10.1016/j.meegid.2018.07.005).
35. Davis, C.L., Field, D., Metzgar, D., Saiz, R., Morin, P.A., Smith, I.L., Spector, S.A. and Wills, C. (1999), Numerous length polymorphisms at short tandem repeats in human cytomegalovirus. *Journal of virology*. 73(8): p. 6265-6270 (DOI: 10.1128/JVI.73.8.6265-6270.1999).
36. Pradeep B., Shekar M., Karunasagar I., Karunasagar I. (2008), Characterization of variable genomic regions of Indian white spot syndrome virus. *Virology*. 376: p. 24-30 (DOI: 10.1016/j.virol.2008.02.037).
37. Burrel, S., Ait-Arkoub, Z., Voujon, D., Deback, C., Abrao, E.P., Agut, H. and Boutolleau, D. (2013), Molecular characterization of herpes simplex virus 2 strains by analysis of microsatellite polymorphism. *Journal of clinical microbiology*. 51(11): p. 3616-3623 (DOI: 10.1128/JCM.01714-13).
38. Houg, H.S.H., Lott, L., Gong, H., Kuschner, R.A., Lynch, J.A. and Metzgar, D. (2009), Adenovirus microsatellite reveals dynamics of transmission during a recent epidemic of human adenovirus serotype 14 infection. *Journal of clinical microbiology*. 47(7): p. 2243-2248 (DOI: 10.1128/jcm.01659-08).
39. Singh A. K., Alam C. M., Sharfuddin C., Ali S. (2014), Frequency and distribution of simple and compound microsatellites in forty-eight Human

- papillomavirus (HPV) genomes. *Infection, Genetics and Evolution*. 24: p. 92-98 (DOI: 10.1016/j.meegid.2014.03.010).
40. George B., Gnanasekaran P., Jain S., Chakraborty S. J. I. (2014), *Genetics & Evolution*. Genome wide survey and analysis of small repetitive sequences in caulimoviruses. 27: p. 15-24 (DOI: 10.1016/j.meegid.2014.06.018).
41. George B., Alam C. M., Kumar R. V., Gnanasekaran P., Chakraborty S. J. V. (2015), Potential linkage between compound microsatellites and recombination in geminiviruses: Evidence from comparative analysis. 482: p. 41-50 (DOI: 10.1016/j.virol.2015.03.003).
42. Treco D., Arnheim N. (1986), The evolutionarily conserved repetitive sequence d(TG.AC)_n promotes reciprocal exchange and generates unusual recombinant tetrads during yeast meiosis. *Molecular and cellular biology*. 6: p. 3934-3947 (DOI: 10.1128/mcb.6.11.3934).
43. Smithson, C., Purdy, A., Verster, A.J. and Upton, C. (2014), Prediction of steps in the evolution of variola virus host range. *PloS one*. 9(3): p. e91520. (DOI: 10.1371/journal.pone.0091520).
44. Odom M. R., Hendrickson R. C., Lefkowitz E. J. J. V. r. (2009), Poxvirus protein evolution: family wide assessment of possible horizontal gene transfer events. 144: p. 233-249 (DOI: 10.1016/j.virusres.2009.05.006).
45. Sanger F., Nicklen S., Coulson A. R. (1977), DNA sequencing with chain-terminating inhibitors.

- Proceedings of the National Academy of Sciences. 74:
p. 5463 (DOI: 10.1073/pnas.74.12.5463).
46. Tamura, K., Stecher, G., Peterson, D., Filipowski, A. and Kumar, S. (2013), Molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol.* 30(12): p. 2725-2729 (DOI: 10.1093/molbev/mst197).
 47. Mazur C., Machado R. D. (1989), Detection of contagious pustular dermatitis virus of goats in a severe outbreak. *Veterinary Record.* 125: p. 419 (DOI: 10.1136/vr.125.16.419).
 48. Bennett, M., Tu, S.L., Upton, C., McArtor, C., Gillett, A., Laird, T. and O’Dea, M. (2017), Complete genomic characterisation of two novel poxviruses (WKPV and EKPV) from western and eastern grey kangaroos. *Virus research.* 242: p. 106-121 (DOI: 10.1016/j.virusres.2017.09.016).
 49. Burioli E., Prearo M., Houssin M. J. V. (2017), Complete genome sequence of Ostreid herpesvirus type 1 μ Var isolated during mortality events in the Pacific oyster *Crassostrea gigas* in France and Ireland. *Virus Research.* 242: p. 239-251 (DOI: 10.1016/j.virusres.2017.09.016).
 50. Bolger A. M., Lohse M., Usadel B. J. B. (2014), Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 30: p. 2114-2120 (DOI: 10.1093/bioinformatics/btu170).
 51. Li H., Durbin R. J. b. (2009), Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics.* 25, p. 1754-1760 (DOI: 10.1093/bioinformatics/btp324).
 52. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. (2009), The sequence alignment/map format and

- SAMtools. *Bioinformatics*. 25(16): p. 2078-2079 (DOI: 10.1093/bioinformatics/btp352).
53. Tcherepanov V., Ehlers A., Upton C. J. B. g. (2006), Genome Annotation Transfer Utility (GATU): rapid annotation of viral genomes using a closely related reference genome. *Bioinformatics*. 22(7): p. 150 (DOI: 10.1186/1471-2164-7-150).
54. Librado P., Rozas J. (2009), DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*. 25: p. 1451-1452, (DOI: 10.1093/bioinformatics/btp187)
55. Mudunuri S. B., Nagarajaram H. A. (2007), IMEx: Imperfect Microsatellite Extractor. *Bioinformatics*. 23: p. 1181-1187 (DOI: 10.1093/bioinformatics/btm097)
56. Chen, M., Tan, Z., Jiang, J., Li, M., Chen, H., Shen, G. and Yu, R., 2009. Similar distribution of simple sequence repeats in diverse completed Human Immunodeficiency Virus Type 1 genomes. *FEBS letters*. 583(17): p. 2959-2963 (DOI: 10.1016/j.febslet.2009.08.004).
57. Wu X., Zhou L., Zhao X., Tan Z. (2014), The analysis of microsatellites and compound microsatellites in 56 complete genomes of Herpesvirales. *Gene*. 551: p. 103-109 (DOI: 10.1016/j.gene.2014.08.054).
58. Katoh K., Standley D. M. (2013), MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*. 30: p. 772-780 (DOI: 10.1093/molbev/mst010).

59. Hall T., Biosciences I., Carlsbad C. J. G. B. B. (2011), BioEdit: an important software for molecular biology. 2: p. 60-61
60. Martin D. P., Murrell B., Golden M., Khoosal A., Muhire B. J. V. e. (2015), RDP4: Detection and analysis of recombination patterns in virus genomes. (DOI: 10.1093/ve/vev003).

Chapter 6

The emergence of subclade A1 and A3 Avipoxviruses in India

6.1 Introduction

The *Poxviridae* family comprises enveloped DNA viruses that have a characteristic brick-shaped or ovoid morphology. The *Poxviridae* family is divided into two subfamilies, *Entomopoxvirinae*, and *Chordopoxvirinae*, which infect insects and chordates, respectively [1, 2]. Within the *Chordopoxvirinae* subfamily, Avipoxvirus (APV) is the only characterized genus known to infect a wide range of avian species with a worldwide distribution [3, 4]. Infected birds manifest disease in two major forms. The first is a milder cutaneous form (dry form), characterized by proliferative nodular skin lesions on the featherless parts of the skin and is self-limiting. The second is a diphtheritic form, characterized by fibrino-necrotic lesions that appear in the upper respiratory and gastrointestinal tracts of birds, and is often fatal [2]. Avipoxvirus infection can cause economic losses to the poultry farming industry due to a reduction in egg production, reduced growth rates, blindness, and death. Viral transmission occurs through direct or, indirect contact, either passing through broken skin or more commonly, by insect bites [5].

APV infection can reduce breeding success and thus pose a risk to small and endangered populations, particularly to island bird species [6, 7, 8].

The Avipoxvirus genome consists of a fairly large double-stranded DNA sequence (130-375 kbp) that encodes approximately 150-330 proteins [9, 10]. The genome transcription is temporally regulated, where the early gene products assist genome replication, encode virulence factors such as ankyrin repeat gene family protein (ANK), interleukin binding protein, C4L/C10L-like gene family protein, and B-cell lymphoma protein (Bcl-2) and the late gene products, typically contribute to viral assembly [11, 12]. The diagnosis of Avipoxvirus is primarily performed by histopathological examination of characteristic skin lesions corroborated by virus isolation on the chorioallantoic membrane (CAM) of embryonated chicken eggs or cell cultures. Recently, molecular biological methods such as PCR have emerged as a gold standard technique for routine Avipoxvirus diagnosis. The highly conserved loci of the viral genome, such as the gene encoding the P4b core protein (FPV167) and the viral DNA polymerase (FPV094), are targeted for PCR-based diagnosis as well as phylogenetic analysis [13-17]. A phylogenetic tree constructed based on these loci revealed that Avipoxviruses cluster into three major clades, A, B, and C, with clades A and B being subdivided further into A1, A2, A3, A4, B1, and B2 [14, 2, 18]. Recently, two additional clades, D and E, were observed containing one and two isolates of APV, respectively; these findings motivated us to evaluate the presence of different viral clades and subclades circulating throughout India [9, 16, 18].

In India, the re-emergence of Avipoxviruses in various avian species, such as fowl, turkey, pigeon, duck, quail, peacock, golden pheasant, and silver pheasant, manifests with high morbidity and mortality [19-22]. In the majority of cases, diagnosis is based on histopathological examination of diseased birds rather than molecular detection. Despite having such high economic and ecological importance, Avipoxvirus has received limited attention regarding its epidemiology and preventive measures. In this study, Avipoxvirus DNA recovered from infected wild birds from the Eastern Indian state of Odisha during the years 2010–2018 were analyzed. The outbreaks were characterized by the presence of cutaneous lesions on featherless parts of the birds with varying rates of morbidity and mortality. Fragments of the viral P4b core protein and the DNA polymerase genes were PCR amplified and sequenced. Subsequently, phylogenetic trees were constructed to compare the sequences with those of previously reported global strains to better understand the emergence of circulating strains in India. This is the first molecular characterization of Avipoxvirus strains in the Eastern Indian state of Odisha that is based on the conserved DNA polymerase gene. Our results show the emergence of subclades A1 and A3 in Indian chickens and pigeons, respectively. The present study highlights the Avipoxvirus diversity, which could facilitate the development of strategies for disease prevention and possible eradication.

6.2 Results

6.2.1 Clinical pathology

A total of 1280 birds were investigated showing symptoms for APV infection, in which the morbidity ranged from 18-19.2% and 16.9-23% in chickens and pigeons, respectively (Figure 6.1). A high mortality rate of 47.3-52.7% in chickens and 39.13-92% in pigeons was observed. Geographical location/district-wise morbidity and mortality rates are presented in Table 6.2. Further histopathological investigation revealed the cause of death to be most commonly associated with anorexia and respiratory distress followed by weight loss. The highest morbidity and mortality was observed in the 0-8 weeks age group (Table 6.1), followed by the 9-20 weeks age group, and was lowest in adult birds.

Sr. No.	Disease	Host species	Geographic area from of outbreaks	Coordinates	Year of virus outbreak	P4b (Accession numbers)	DNA pol (Accession numbers)
1	Fowl pox	Chicken (Gallus gallus)	Subarnapur	19.17124°N,84.653809°E	2010	MH721409	MH830343
2	Fowl pox	Chicken (Gallus gallus)	Ganjam	19.388201°N,85.049004°E	2011	MH721410	MH830344
3	Fowl pox	Chicken (Gallus gallus)	Kundhamal	20.10807°N,84.041670°E	2012	MH721411	MH830345
4	Pigeon pox	Pigeon (Columba livia)	Dhenkanal	20.662901°N,85.597900°E	2010	MH721412	MH830346
5	Pigeon pox	Pigeon (Columba livia)	Bhadrak	21.058273°N,86.495842°E	2013	MH721413	MH830347
6	Pigeon pox	Pigeon (Columba livia)	Puri	19.80044°N,85.826752°E	2014	MH721414	MH830348
7	Pigeon pox	Pigeon (Columba livia)	Sambalpur	21.466871°N,83.981171°E	2016	MH721415	MH830349
8	Pigeon pox	Pigeon (Columba livia)	Khordha	20.1863°N,85.622597°E	2017	MH721416	MH830350
9	Pigeon pox	Pigeon (Columba livia)	Nabarangpur	19.150709°N,82.601067°E	2018	MH721417	MH830351

		livia)					
10	Pigeon pox	Pigeon (Columba livia)	Boudh	20.8382°N,84.322403°E	2018	MH721418	MH830352

Table 6.1: Details of Avipoxvirus-infected domestic birds and the location of the samples collected in this study.

Sr. No.	Disease	Geographic area/year of sampling	No. of host investigated				No. of host affected (morbidity %)				No. of host died (mortality %)			
			Chicks or Squabs (0-8W)	Growers (9-20 W)	Adults (>20 W)	Total	Chicks or Squabs (0-8W)	Growers (9-20 W)	Adults (>20 W)	Total	Chicks or Squabs (0-8W)	Growers (9-20 W)	Adults (>20 W)	Total
1	Fowl pox	Subarnapur/2010	113	77	10	200	25 (22.1)	10 (12.9)	1 (10.0)	36 (18.0)	18 (72.0)	1 (10.0)	0	19 (52.7)
2	Fowl pox	Ganjam/2011	79	41	20	130	20 (25.3)	5 (12.19)	0	25 (19.2)	12 (60.0)	0	0	12 (48.0)
3	Fowl pox	Kundhamal/2012	30	47	23	100	15 (50.0)	4 (8.5)	0 (8.6)	19 (19.0)	9 (60.0)	0	0	9 (47.3)
4	Pigeon pox	Dhenkanal/2010	64	36	70	170	25 (39.0)	7 (19.4)	6 (8.5)	38 (22.3)	18 (72.0)	2 (28.5)	0	20 (52.6)
5	Pigeon pox	Bhadrak/2013	35	55	10	100	10 (28.5)	10 (18.1)	3 (30.0)	23 (23.0)	6 (60.0)	3 (30.0)	0	9 (39.1)
6	Pigeon pox	Puri/2014	40	40	50	130	14 (35.0)	6 (15.0)	2 (4.0)	22 (16.9)	10 (71.4)	1 (16.6)	1 (50.0)	12 (54.5)
7	Pigeon pox	Sambalpur/2016	73	12	10	90	14 (19.1)	2 (16.6)	2 (20.0)	18 (20.0)	13 (92.8)	1 (50.0)	0	15 (83.3)
8	Pigeon pox	Khordha/2017	92	13	15	120	25 (27.1)	0	0	25 (20.8)	23 (92.0)	0	0	23 (92.0)

9	Pigeon pox	Nabarangpur/2018	77	23	15	115	20 (25.9)	0	0	20 (17.3)	18 (90.0)	0	0	18 (90.0)
10	Pigeon pox	Boudh/2018	69	37	19	125	12 (17.3)	8 (21.6)	3 (15.7)	23 (18.4)	9 (75.0)	2 (25.0)	0	11 (47.8)

Table 6.2: Table showing the characteristics of the birds included in the study with their corresponding morbidity and mortality percentage (W=Weeks).

6.2.2 Phylogenetic analysis of the P4b (FPV 167) and DNA polymerase (FPV094) gene

Initial PCR screening for positive samples was performed with primers targeting the P4b and DNA polymerase gene (Figure 6.2 A-B). Subsequently, loci comprising fragments of 368bp (P4b) and 520 bp (DNA polymerase) nucleotides were used for the phylogenetic analysis. For P4b gene phylogenetic tree construction, all 85 available Avipoxvirus isolates stemming from 49 different bird species were retrieved from GenBank, while for DNA polymerase, all 39 available sequences obtained from 23 different bird species were analyzed. Trees were constructed using three different methods, such as neighbor-joining (NJ), maximum likelihood (ML), and Bayesian, based on the alignment of the concatenated, P4b, and DNA polymerase gene sequences. Based on bootstrap and posterior probability values the Bayesian trees were considered to be the most reliable, followed by the NJ and ML trees (Figure 6.3). The concatenated tree topology exhibited the presence of three clades, A, B, and C followed by subclades A1-A7 and B1-B2. However, we have noticed a mixed subclade A2 and A3 during our investigation. To get a clear picture of subclade

distribution, we choose the Bayesian approach to construct a gene-specific phylogenetic tree using the P4b gene (Figure 6.4) and DNA polymerase gene (Figure 6.5). The P4b phylogenetic tree results showed the presence of five clades, A, B, C, D, and E, with subclades A1-A7 and B1-B3 among the global circulates (Figure 6.4), while phylogenetic tree based on DNA polymerase gene indicates the presence of four clades such as A, B, C, and E with subclade A1-A7 and B1-B2. These minor differences in the clade and subclade distribution between concatenated and gene-specific phylogeny were also reported by Gyuranecz et al., 2013 [14] while studying the worldwide phylogenetic relationship of APVs. Importantly, all isolates of this study grouped into clade A. In the subclade analysis, three samples of APVs (collected from infected chicken in this study) grouped into clade A1 along with some previously reported Indian isolates collected from infected golden pheasant (HM481402), sparrow (HM481407), chicken (KF548036), and peacock (HM481405). However, several other non-Indian isolates also grouped into subclade A1, including APVs isolated from chicken (Mozambique, Tanzania, China, Portugal, Nigeria, UK, France, Europe, Hungary, Egypt, and USA), turkey (Egypt, Iran, USA, and Italy), parrot (Chile), macQueens bustard (UAE), blue-eared pheasant (Hungary) and paradise shelduck, variable oystercatcher, black robin, shore plover from New Zealand. The percentage of nucleotide similarity of all isolates analyzed from different species was 99.80-100% identical to each other.

Samples collected from pigeons in this study grouped in subclade A2, along with other previously reported APVs from pigeons (MF496043 and

DQ873811) and chicken (DQ873810) in India. Several other global APV isolates collected from turkey (Mozambique, UK, and Italy), pigeon (UK, Egypt, and South Africa), dove (Hungary), Indian peafowl (Hungary), macQueens bustard (UAE), grey partridge (Italy), gyrfalcon (Italy), canary (Italy) along with the booted eagle, red kite, red-legged partridge from Spain were also grouped into the A2 subclade. The percentage nucleotide identity of each locus revealed that all isolates collected from different species were identical to each other in a species-specific manner, with 99.39-100% similarity.

The phylogenetic tree based on the DNA polymerase gene showed the presence of four clades, A, B, C, and E, with subclades A1-A7 and B1-B2 (Figure 6.5). The isolates in this study grouped into clade A. Further subclade analysis showed that Avipoxvirus isolates of chickens grouped into subclade A1 along with several other global isolates, including APVs isolated from chicken (Mozambique, Hungary, France, Brazil, and Egypt), turkey (USA), parrot (Chile), blue-eared pheasant (Hungary) with 99.84-100% nucleotide similarity. Interestingly, samples collected from infected pigeons grouped into subclade A3, along with previously reported global isolates such as those of great bustard (Spain), oriental turtle doves from South Korea, and feral pigeons of South Africa. The percentage nucleotide identity of each locus revealed that all isolates investigated here were identical to each other in a species-specific manner, with 99.08-100% similarity.

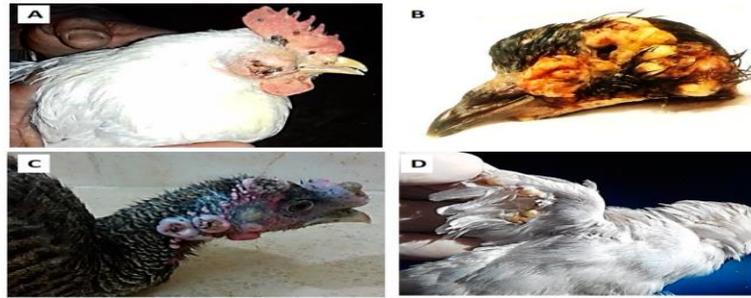


Figure 6.1: Figure illustrating clinical cases of FPV infection in chickens showing lesions around the beak, eye lids and featherless part of the body during the outbreak investigation.

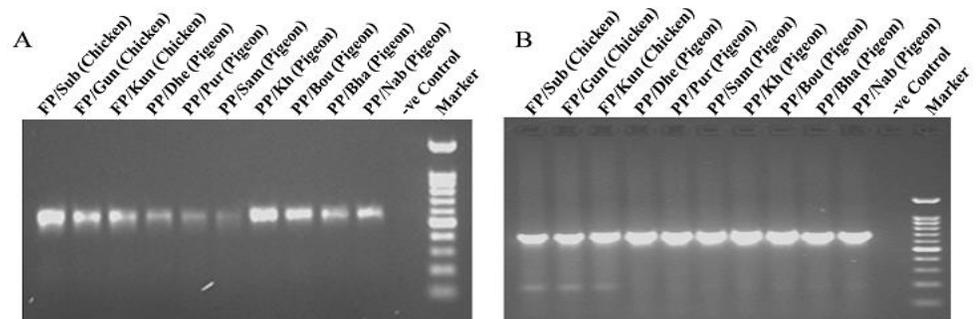


Figure 6.2 A-B: PCR confirmation of avipoxvirus genes: (a) The left panel shows the PCR amplification of the P4b gene from the ten APV isolates in this study. The last lane represents the marker (100 bp), while the second to last lane represents the negative control, i.e., no template. (b) The right panel shows the PCR amplification of the DNA polymerase gene from the ten APV isolates in this study. The last lane represents the marker (100 bp), while the second to last lane represents the negative control, i.e., no template.

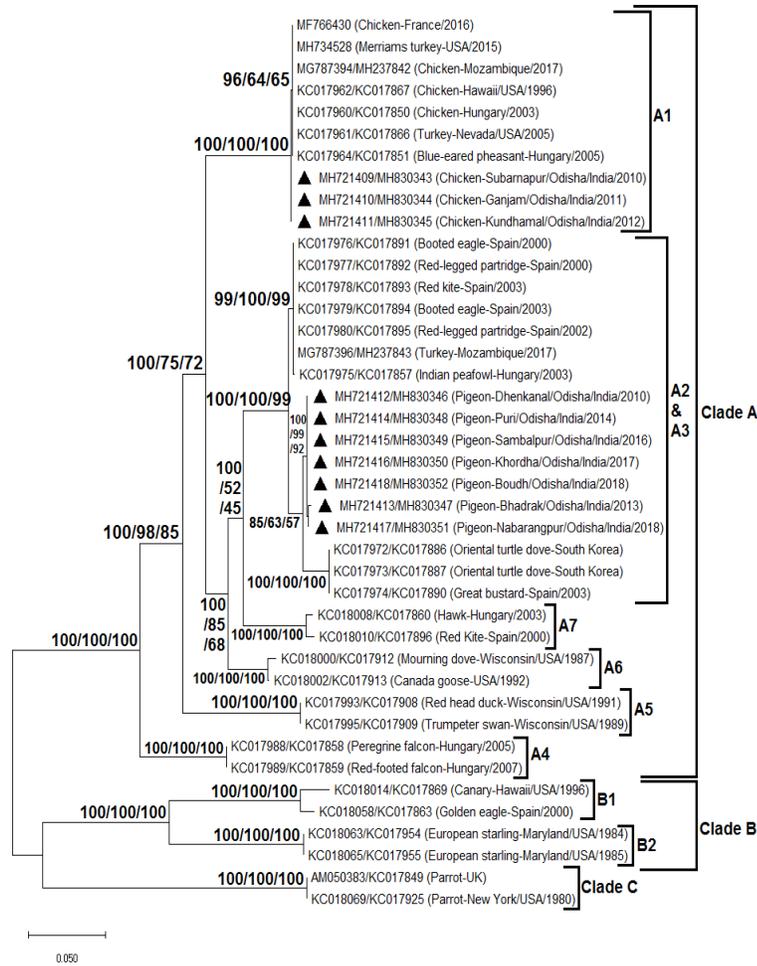


Figure 6.3: Phylogenetic tree construction for avipoxvirus isolates. Concatenated DNA sequences comprising P4b and DNA polymerase genes Avipoxviruses were used to construct a phylogenetic tree with posterior probability values of the Bayesian trees (1,000 replicates), neighbor-joining, and maximum likelihood bootstrap values (1,000 replicates). The left side of the branch points having some numeric, which represents the Bayesian posterior probability values/ neighbor-joining bootstrap values/ maximum likelihood bootstrap values and subclades, labeled according to the nomenclature of Mapaco et al. 2018 and Offerman et al., 2013. The number of substitutions per site was represented in the scale.

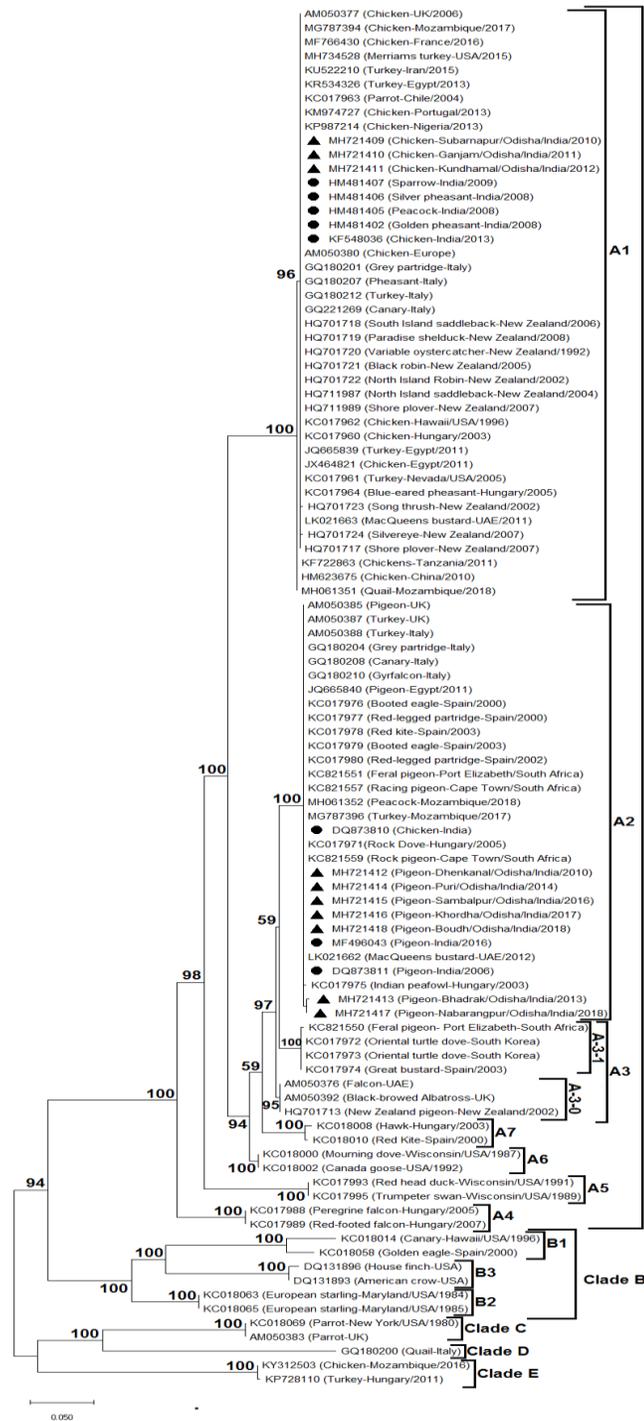


Figure 6.4: The phylogenetic tree of nucleotide sequences for the P4b core protein gene of the APV isolates. Black circles represent the APVs previously isolated in India; black triangles represent the isolates of the current study.

6.3 Discussion

Avipoxviruses often cross species barriers, which could be one of the reasons for the presence of these viruses in both domesticated and wild birds. Among the various fatal viral diseases, APV infection causes significant mortality in domesticated birds, second only to Newcastle disease virus infection, which causes the most fatalities [23]. In the present study, Avipoxviruses were detected in two different genera of birds from ten geographically distinct regions of Odisha over a period of eight years, and the pathology, morbidity, and mortality were assessed. Our study revealed that the morbidity/mortality percentage in APV infections is similar to the previous reports, in which rates were found to be more than 50% [24-26]. However, our results differed from the recently studied Avipox outbreak, in which the percentages varied from 17-18% and 41-45%, respectively [27]. The difference in severity may be due to the presence of higher chicks or squabs population (0-8W) within the study group [28, 29]. The highest incidence of morbidity and mortality was recorded in the younger population (0-8 weeks old chicks or squabs), followed by growers (9-20 W) and adults (>20 W). This is probably due to the absence of maternal antibodies in the chicks or squabs [30, 31] or due to overall underdeveloped immune systems. As most of the poultry farmers in this region did not adopt vaccination, this might have contributed towards higher susceptibility and severity of the disease [32]. The viral presence was confirmed by PCR amplification of two loci containing the P4b core protein and the DNA polymerase gene. BLAST and phylogenetic analysis revealed species-specific matches of 100% nucleotide similarity within the collected virus samples. This finding indicates that

geographically restricted isolates are stabilized over time. Earlier reports on Avipoxvirus classified them into three major clades, A, B, and C, with subclades A1, A2, A3, B1, and B2 [20]. However, recently, two important clades (D and E), as well as subclades A1-A7 and B1-B3, have emerged [9, 16-18]. Following this nomenclature, we observed that three FPV isolates in this study, along with similar Indian isolates, including strains isolated from endangered birds derived from GenBank, formed a robust A1 subclade. Other isolates from chicken (Mozambique, Tanzania, China, Portugal, Nigeria, UK, France, Europe, Hungary, Egypt, and USA), turkey (Egypt Iran, USA, and Italy), parrot (Chile), macQueens bustard (UAE), Blue-eared pheasant (Hungary) and paradise shelduck, variable oystercatcher, black robin, shore plover from New Zealand also grouped into this subclade. We observed that the hosts of the order Galliformes primarily harbored the A1 subclade, which is similar to the findings reported by Gyuranecz et al., 2013 [14] and Jarmin et al., 2006 [14, 2]. Although subclade A1 was previously reported in India [20], the host species were observed to be endangered wild birds sampled from the Indian states of Andhra Pradesh, Karnataka, and Maharashtra. In this study, for the first time, we report the presence of subclade A1 in domesticated and economically important bird species of India. We speculate that transboundary and cross-species transfer of APV isolates could have resulted into this, as Odisha is geographically adjacent to Andhra Pradesh. Similarly, it is interesting to note that the chicken isolates collected from India and China have 100% nucleotide similarity and are grouped into the same subclade, which may share a common origin. However, further in-depth genetic characterization is required to confirm

this origin as well as the similarity in virulence and mortality rates. The remaining seven isolates derived from pigeons exhibited 99.39-100% similarity with the virus isolates from a pigeon (UK, Egypt, and South Africa), turkey (Mozambique, UK, and Italy), dove (Hungary), Indian peafowl (Hungary), macQueens bustard (UAE), grey partridge (Italy), gyrfalcon (Italy), canary (Italy) along with the booted eagle, red kite, red-legged partridge from Spain, and grouped into subclade A2. Interestingly, our observation is consistent with the earlier report [33].

The DNA polymerase gene is used as a marker for the molecular characterization of APV isolates [14, 16, 17]. For this reason, we performed a phylogenetic analysis with the DNA polymerase gene for the first time with Indian isolates. All three FPV isolates in our study, along with other strains derived from chicken (Mozambique, Hungary, France, Brazil, and Egypt), turkey (USA), parrot (Chile), blue-eared pheasant (Hungary), formed a robust A1 subclade based on the polymerase gene sequence with 99.84-100% nucleotide similarity. In contrast, the pigeon pox isolates were grouped into subclade A3, along with those of great bustard (Spain), oriental turtle doves, and feral pigeons from South Korea and South Africa, respectively, with 99.08-100% nucleotide identity. Previously, Offerman et al., reported the presence of subclade A3 in African pigeons [33]. For the first time, the present study shows the existence of the A3 subclade in Indian pigeons. Several studies have shown that APVs from the same species of bird form different subclades [2]. Our study observed a similar pattern of subclade distribution, where the phylogenetic tree-based on the P4b and polymerase genes of pigeonpox virus from two different subclades, A2 and A3,

respectively. This type of ambiguity has been observed in many studies due to the complicated nature of the host range of APV [2]. APVs undergo genetic recombination during which many non-essential genes may be deleted, or new host genes may be acquired, leading to the generation of new variants in the population [14]. To investigate whether recombination plays a crucial role in APV evolution in India, we need to utilize genomic resources with complete genome sequence information from a large population. Currently, not a single complete APV genome has been reported from India, which restricts our ability to assess the evolution of APV in the Indian subcontinent.

6.4 Materials and methods

6.4.1 Sample and outbreak data collection

Skin or scab samples from the comb, eyelids, and beak from suspected Avipoxvirus-infected birds were collected from ten outbreaks, which occurred in diverse geographical locations in Odisha, India, during the years 2010-2018. In this surveillance, 1280 birds were investigated, followed by a collection of 249 suspected tissue samples from unvaccinated backyard farms having a cutaneous infection. The primary symptoms of infection comprised of dehydration, anorexia, emaciation, weight loss, stunted growth, decreased egg production, and in certain cases, birds faced blindness due to blepharitis and conjunctivitis. Infected tissue samples (5 g) were transferred to sterile phosphate-buffered saline (PBS), pH 7.2 supplemented with antibiotics and antifungal reagents, and stored at $-20\text{ }^{\circ}\text{C}$ until needed for further processing and laboratory analysis.

6.4.2 DNA isolation

Briefly, all suspected tissue samples were homogenized in 1 ml of sterile 0.1 M PBS and then treated with tissue lysis buffer containing proteinase K, and the mixture was incubated at 56°C overnight for cell lysis. Finally, the cell lysates were passed through a charged column (Qiagen, USA), and DNA was eluted from the column by using the previously described standard phenol-chloroform and stored at -20°C until further use [34].

6.4.3 PCR amplification

The P4b core protein fragment was amplified by the primers, For: 5'CAG CAG GTG CTA AAC AA3' and Rev: 5'CGG TAG CTT AAC GCC GAA TA3'; the DNA polymerase gene fragment amplification was performed with For: 5'CGC CGC ATC TAC TTA TC3', and Rev: 5'CCA CAC AGC GCC ATT CAT TA3' [14]. The PCR mixture (25 µL) contained 5 pmol of each primer, 2 mM dNTPs in 1x buffer supplemented with 2.5 U Taq polymerase as per the supplier's instructions (NEB). The thermal cycling conditions were as follows: initial denaturation at 95°C for 5 min, followed by 35 cycles of denaturation at 95 °C for 50 s, annealing at 55 °C for 60 s, extension at 72 °C for 90 s and a final extension at 72 °C for 7 min. PCR amplified products (5 µL) were loaded onto a 1.5% agarose gel. Samples collected from a single flock of chicken or pigeons housed together in a particular location during the outbreak were grouped together, which resulted in three FPV and seven PPV groups. PCR was done from the samples representative of the group, and respective amplicons were gel-purified by using a PureLink quick gel extraction kit (Invitrogen, USA) and sequenced commercially.

6.4.4 DNA sequence analysis of the P4b and DNA polymerase genes

Phylogenetic analyses were performed separately for the P4b and DNA polymerase gene sequences. DNA sequences were initially edited by BioEdit v7.0.5 and submitted to the GenBank database under the accession numbers listed in Table 6.1. The other isolates of APVs were downloaded from the GenBank database and then aligned using CLUSTAL X (Annexure B). Phylogenetic trees were constructed individually and for concatenated, P4b, and DNA polymerase gene sequences. Trees were generated using three methods: neighbour-joining (NJ), maximum likelihood (ML), and the Bayesian approach. To choose the best model of evolution, jModelTest was used [35]. According to Akaike's information criterion, the best suitable model for the P4b, and the concatenated sequences were noted to be transitional model TPM1uf+G, while that for DNA polymerase gene, a gamma distribution (GTR+G) was found to be the best. The gamma rates for the three gene sequences were as follows: concatenated = 0.3612, P4b = 0.3680 and polymerase = 0.8700. The NJ and ML analysis in MEGA 6.0 was done with a bootstrapped for 1000 with the respective model [36]. Bayesian analysis in MrBayes 3.1 was run for 1 to 2.5 million generations, with sampling at every 100th generation until model convergence was achieved [37]. Further, clade subdivisions were performed according to the methods described in previous reports [14, 17].

6.5 Conclusion

Since APVs have a wide host range, it is difficult to develop an effective preventive strategy because detailed knowledge of viral isolates is lacking. Our findings provide evidence of circulating APV strains in India, which may contribute to a better understanding of APV epidemiology to aid the development of effective vaccines in the near future.

6.6 References

1. Gubser C., Hué S., Kellam, P., Smith G. L. (2004), Poxvirus genomes: a phylogenetic analysis. *Journal of General Virology*. 85(1): p. 105-117 (DOI: 10.1099/vir.0.19565-0).
2. Jarmin S., Manvell R., Gough R. E., Laidlaw S. M., Skinner M. A. (2006)., Avipoxvirus phylogenetics: identification of a PCR length polymorphism that discriminates between the two major clades. *Journal of General Virology*. 87(8): p. 2191-2201 (DOI: 10.1099/vir.0.81738-0).
3. Bolte A. L., Meurer J., Kaleta E. F. (1999)., Avian host spectrum of avipoxviruses. *Avian Pathology*. 28(5): p. 415-432 (DOI: 10.1080/03079459994434).
4. Van Riper C., Forrester D. (2008), Avian pox. In H. D. Thomas NJ, Atkinson & CT (Eds.), *Infectious diseases of wild birds*: Blackwell Publishing Professional: p. 131-171.
5. Friend M., Franson J. C. (1999), *Field manual of wildlife diseases. General field procedures and diseases of birds*.
6. Kane O. J., Uhart M. M., Rago V., Pereda A. J., Smith J. R., Van Buren A., Boersma P. D. (2012), Avian pox

- in Magellanic penguins (*Spheniscus magellanicus*).
Journal of Wildlife Diseases. 48(3): p. 790-794. (DOI:
10.7589/0090-3558-48.3.790).
7. Kleindorfer S., Dudaniec, R. Y. (2006), Increasing prevalence of avian poxvirus in Darwin's finches and its effect on male pairing success. *Journal of Avian Biology*. 37(1): p. 69-76 (DOI:10.1111/j.0908-8857.2006.03503.x).
 8. Lachish S., Lawson B., Cunningham A. A., Sheldon B. C. (2012), Epidemiology of the emergent disease Paridae pox in an intensively studied wild bird population. *PLoS One*. 7(11): p. e38316. (DOI: 10.1371/journal.pone.0038316).
 9. Bányai K., Palya V., Dénes B., Glávits R., Ivanics É., Horváth B., Gyuranecz M. (2015), Unique genomic organization of a novel Avipoxvirus detected in turkey (*Meleagris gallopavo*). *Infection, Genetics and Evolution*. 35: p. 221-229 (DOI: 10.1016/j.meegid.2015.08.001).
 10. Tulman E., Afonso C., Lu Z., Zsak L., Kutish G., Rock, D. (2004), The genome of canarypox virus. *Journal of virology*. 78(1): p. 353-366 (DOI: 10.1128/jvi.78.1.353-366.2004).
 11. Afonso C., Tulman E., Lu Z., Zsak L., Kutish G., Rock D. (2000), The genome of fowlpox virus. *Journal of virology*. 74(8): p. 3815-3831 (DOI: 10.1128/jvi.74.8.3815-3831.2000).
 12. Offerman K., Carulei O., van der Walt A. P., Douglass N., Williamson, A.-L. (2014), The complete genome sequences of poxviruses isolated from a penguin and a pigeon in South Africa and comparison to other

- sequenced avipoxviruses. *BMC genomics*. 15(1): p. 463 (DOI: 10.1186/1471-2164-15-463).
13. Abdallah F. M., Hassanin, O. (2013), Detection and molecular characterization of avipoxviruses isolated from different avian species in Egypt. *Virus Genes*. 46(1): p. 63-70 (DOI: 10.1007/s11262-012-0821-y).
 14. Gyuranecz M., Foster J. T., Dán Á., Ip H. S., Egstad K. F., Parker P. G., Kreizinger Z. (2013), Worldwide phylogenetic relationship of avian poxviruses. *Journal of virology*. *JVI*: p. 03183-03112 (DOI: 10.1128/JVI.03183-12).
 15. Lecis R., Secci F., Antuofermo E., Nuvoli S., Scagliarini A., Pittau M., Alberti A. (2017), Multiple gene typing and phylogeny of avipoxvirus associated with cutaneous lesions in a stone curlew. *Veterinary research communications*. 41(2): p. 77-83 (DOI: 10.1007/s11259-016-9674-5).
 16. Mapaco L. P., Lacerda Z., Monjane I. V., Gelaye E., Sussuro A. H., Viljoen G. J., Achá S. J. (2017), Identification of clade E avipoxvirus, Mozambique, 2016. *Emerging infectious diseases*. 23(9): p. 1602 (DOI: 10.3201/eid2309.161981).
 17. Mapaco L. P., Lacerda Z., Monjane I. V., Sussuro A., Viljoen G. J., Cattoli G., Achá S. J. (2018), Molecular characterization of avipoxviruses circulating in Mozambique, 2016-2018. *Archives of virology*. p. 1-7 (DOI: 10.1007/s00705-018-3864-0).
 18. Manarolla G., Pisoni G., Sironi G., Rampin, T. (2010), Molecular biological characterization of avian poxvirus strains isolated from different avian species.

- Veterinary microbiology. 140(1-2): p. 1-8 (DOI: 10.1016/j.vetmic.2009.07.004).
19. Mohan M., Fernandez T. F. (2008), A case report of pigeon pox-histopathologic diagnosis. *Veterinary World*. 1(4): p. 117
 20. Pawar R. M., Bhushan S. S., Poornachandar A., Lakshmikantan U., Shivaji, S. (2011), Avian pox infection in different wild birds in India. *European Journal of Wildlife Research*. 57(4): p. 785-793 (DOI: 10.1007/s10344-010-0488-4).
 21. Singh A., Dash B., Kataria J., Dandapat S., Dhama K. (2003), Characterization of an Indian isolate of turkey pox virus. *INDIAN Indian journal of comparative microbiology immunology and infectious diseases*. 24(2): p. 149
 22. Yadav S., Dash B., Kataria J., Dhama K., Gupta S., Rahul S. (2007), Pathogenicity study of different avipoxviruses in embryonated chicken eggs and cell cultures. *Indian J. Vet. Pathol*. 31(1): p. 17-20.
 23. Dana S. S. (1998), Animal husbandry practices among Santal and Lodha tribes of Medinipur district of West Bengal. *Indian Veterinary Research Institute; Izatnagar*,
 24. Ivanov I., Kril A., Sainova I. (2001), Propagation of avian virus strains in the permanent duck embryo cell line DEC 99. *Exp. Pathol. Parasitol*. 5: p. 31-34.
 25. Skinner M. A. (2008), Poxviridae. In M. Pattison, P. F. McMullin, J. M. Bradbury, & D. J. Alexander (Eds.), *Poultry Disease (6th ed.)*: WB Saunders Company, London.

26. Tripathy D. N. (1993), Avipox viruses. . In J. B. McFerran & M. S. McNulty (Eds.), Virus infections of birds: Amsterdam (Netherlands) Elsevier. p. 1-15 (DOI: 10.1111/tbed.13413) .
27. Sharma, B., Nashiruddullah, N., Bhat, M.A., Taku, A., Roychoudhury, P., Ahmed, J.A., Sood, S. and Mehmood, S., 2019. Occurrence and phylogenetic analysis of avipoxvirus isolated from birds around Jammu. *Virus Disease*. 30(2): p. 288-293 (DOI: 10.1007/s13337-018-00507-0).
28. Pathak N. (2016). Prevalence, pathology and molecular diagnosis of pox in domestic birds. Ph. D. thesis submitted to Assam Agricultural University, Khanapara, Guwahati, India.
29. Pathak N., Goswami S., Sharma R., Mukit A. (2011), Incidence of duck diseases in Assam. *Indian Journal of Poultry Science*. 46(2): p. 255-257
30. Lawson B., Lachish S., Colvile K. M., Durrant C., Peck K. M., Toms M. P., Cunningham A. A. (2012), Emergence of a Novel Avian Pox Disease in British Tit Species. *PLoS One*. 7(11): p. e40176 (DOI: 10.1371/journal.pone.0040176).
31. Sawale G., Roshini S., Bulbule N., Chawak M., Kinge G. (2012), Pathology of fowl pox in chickens. *Indian Journal of Veterinary Pathology*. 36(1): p. 110-111
32. Adebajo M. C., Ademola S. I., Oluwaseun A. (2012), Seroprevalence of fowl pox antibody in indigenous chickens in Jos North and South council areas of Plateau State, Nigeria: Implication for vector vaccine. *ISRN veterinary science*, 2012 (DOI: 10.5402/2012/154971)

33. Offerman K., Carulei O., Gous T. A., Douglass N., Williamson A.L. (2013), Phylogenetic and histological variation in avipoxviruses isolated in South Africa. *J Gen Virol.* 94(Pt 10): p. 2338-2351 (DOI: 10.1099/vir.0.054049-0).
34. Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989), *Molecular cloning: a laboratory manual*: Cold spring harbor laboratory press.
35. Posada, D. (2008), jModelTest: Phylogenetic Model Averaging. *Molecular biology and evolution.* 25(7): p. 1253-1256 (DOI: 10.1093/molbev/msn083).
36. Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013), MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular biology and evolution.* 30(12): p. 2725-2729 (DOI: 10.1093/molbev/mst197).
37. Huelsenbeck, J. P., & Ronquist, F. (2001), MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics.* 17(8): p. 754-755. (DOI: 10.1093/bioinformatics/17.8.754).

Chapter 7

Genome-wide identification and characterization of microsatellite markers within the Avipox virus

7.1 Introduction

The *Poxvirus* family consists of enveloped DNA viruses having distinguished brick-shaped or ovoid architecture that replicates within the cytoplasm. The genus Avipox virus (APV), derived from the subfamily *Chordopoxvirinae*, infects more than 300 bird species globally [1, 2]. There are eight complete APV genomic sequences available in GenBank derived from different species, such as fowlpox virus (FWPV) [3], canarypox virus (CNPV) [4], pigeonpox virus (FeP2), penguinpox virus (PEPV) [5], turkeypox virus (TKPV) [6], shearwaterpox viruses (SWPV) [7], flamingopox virus (FGPV) [8], and magpiepox virus (MPPV) [9]. Two significant forms of infection usually are observed in the birds. The most common is the moderate cutaneous form (dry form), manifested by small rounded warts like skin lesions present over the body's unfeathered parts. The second is the diphtheritic form, which generally infects the upper portion of the respiratory and gastrointestinal systems with higher mortality rates [10]. Infection often shrinks egg production, results in stunted growth, anopia, fatality, and severely impact farm economics. Transmission occurs through direct contacts, such as passing respective pathogens through skin aberration or indirectly by insect bites [11]. With perspective to critically endangered and endangered bird species, APV infection can reduce breeding success and thus possess a risk of extinction [12].

The APV genome consists of a double-stranded DNA sequence ranged between 130-375 kbp, which encodes approximately 150-330 proteins [6]. The genome encodes early genes that regulate viral replications well as virulence factors. These include C-type lectin protein, B-cell lymphoma 2 (Bcl-2 protein), serpin, semaphorin, and the late gene expression such as virion assembly protein, putative IVM membrane protein maintain the viral assembly [5]. The histopathological characterization of skin lesions or cell culture methods used to conform to the virus's presence. Molecular detection includes the PCR technique utilizing conserved loci of the viral genome, namely, P4b core protein (FPV167) and the viral DNA polymerase (FPV094) [13]. As these genes were derived from the conserved region, these could not help to mitigate strain diversity.

Microsatellites refer to mono to hexanucleotide repeated sequence units ubiquitously distributed in eukaryotic [14] and prokaryotic genome [15], including that of viruses [16]. Depending on the nucleotide sequence composition, the SSRs termed as either simple or compound. The interruptions within the simple microsatellite lead to an interrupted pure, interrupted compound, complex, and interrupted complex. Two microsatellites close to one another generate compound microsatellites due to repeats' interruption [17]. The SSRs govern a vital role in the genome, as recombination during meiosis [18], species evolution [19], mapping of the genome [20], viral strains demarcation [21, 22], and population genetics analysis [23]. Evidence suggests the microsatellites are well present in the viral genome. These include human cytomegalovirus (hCMV) [24], white spot syndrome virus (WSSV) [25], Adenovirus [26], *Spodoptera littoralis* multiple

nucleopolyhedrovirus [27] (SpliMNPV), and Salmon gill poxvirus [28]. Here we report for the first time on comparative analysis, abundance, distribution, and composition of SSRs within APV through an *in-silico* approach. Following this development, we characterized forty compound microsatellite markers in two species. Thus these markers could be used as a tool for virus identification, strain demarcation, and the respective virus's evolutionary study.

7.2 Results

7.2.1 Distribution of SSRs and cSSR in APV Genome

During this study, we observed microsatellites spanned along the entire APV genomes with a ranged 1531-2473 and an average of 2082. The RA and RD varied from 6.87-8.12 and 45.8-53.58, respectively (Table 7.1). On examining the SSR unit size classes, di-nucleotide repeats were found to be most abundant (53.85%), followed by mononucleotide (34.20%) and trinucleotide (11.11%) repeats in all the genomes. Tetranucleotide, pentanucleotide, and hexanucleotide repeats were the least in number and represented 0.61%, 0.11%, and 0.11% within the APV genomes, respectively. Approximately 79.75% and 20.25% of microsatellite motifs were distributed within coding or noncoding regions, respectively. Among the noncoding region, 1.07% is present in UTR, while 19.18% in the intergenic region, where functional protein and hypothetical protein occupied 65.02% and 14.72%, respectively. The incidence of compound microsatellites number ranged from 135-198, with an average of 164 occurring in each genome. The compound microsatellite's RA and RD values ranged from 0.5-0.71 and 9.39-12.92, respectively. Approximately 73.12% and 26.88% of

microsatellite motifs were distributed within coding or noncoding regions, respectively. Among the noncoding region, 0.76% is present in UTR, while 26.12% in the intergenic region, where functional protein and hypothetical protein occupied 58.49% and 14.62%, respectively (Figure 7.1 A-B).

Sr. No.	Virus species	Acc. no.	Strain	Size (bp)	Country	GC (%)	Total no of SSR	RA	RD	Total no of C SSR	cRA	cRD	% of cSSR
S1	Fowlpox virus	MF766431	16069-Trachea	288539	France	30.83	2001	6.93	45.8	146	0.5	9.39	7.29
S2	Pigeonpox virus	NC024447	FeP2	282356	South Africa	29.52	2013	7.12	47.95	157	0.55	10.86	7.79
S3	Turkeypox virus	NC028238	TKPV-HU1124	188534	Hungary	29.78	1531	8.12	53.58	135	0.71	12.92	8.81
S4	Shearwaterpox virus	KX857215	SWPV-2	351108	Australia	30.23	2439	6.94	46.67	188	0.53	10.23	7.7
S5	Flamingopox virus	NC036582	FGPVKD09	293123	South Africa	29.46	2027	6.91	46.67	153	0.52	9.85	7.54
S6	Canarypox virus	MG760432	FMV15SnBuPox	359853	Canada	30.37	2473	6.87	46.24	198	0.55	10.31	8
S7	Magpiepox virus	MK903864	N/A	293226	Australia	29.59	2057	7.01	46.87	178	0.6	11.46	8.65
S8	Penguinpox virus	KJ859677	PSan92	306862	South Africa	29.5	2119	6.9	46.21	158	0.51	9.71	7.47

Table 7.1: Overview of SSRs within various species of APV complete genome sequences and their parameters obtained during this study.

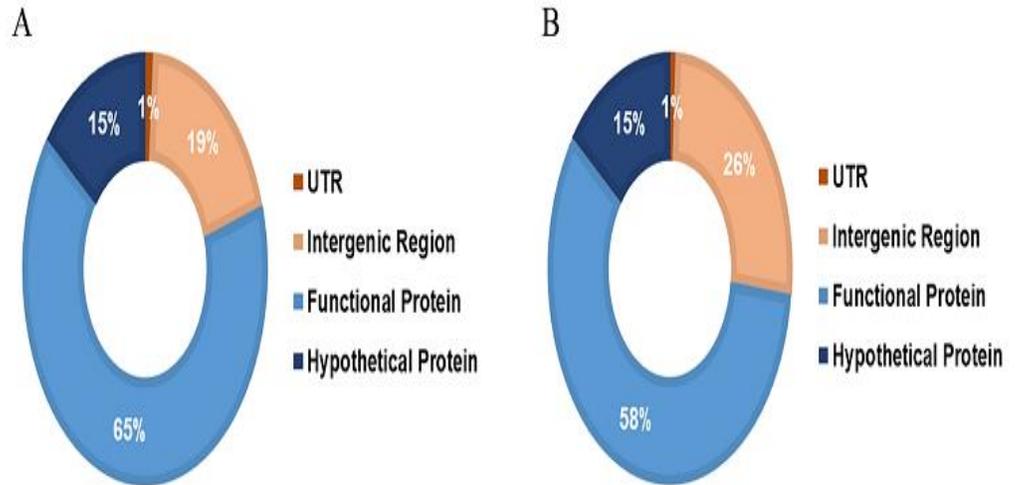


Figure 7.1 A-B: Distribution of SSR and cSSR within AVP genome. A. Graphical representation showing the distribution of SSRs and cSSR within the different genomic locations, including the noncoding and coding regions.

A varied range of cSSR% percentage observed within the APV genome ranges from 7.29-8.81 (Table 7.1). To reveal the effect of dMAX, all the APV genome sequences of this study utilized to calculate the difference of incidence cSSR with increasing dMAX. The dMAX was estimated from 10 and 100 by Microsatellite Identification Search Analysis (MISA) [29]. It resulted in an elevation in the incidence of compound microsatellites with an increase in dMAX value but decreased after a certain point (Figure 7.2).

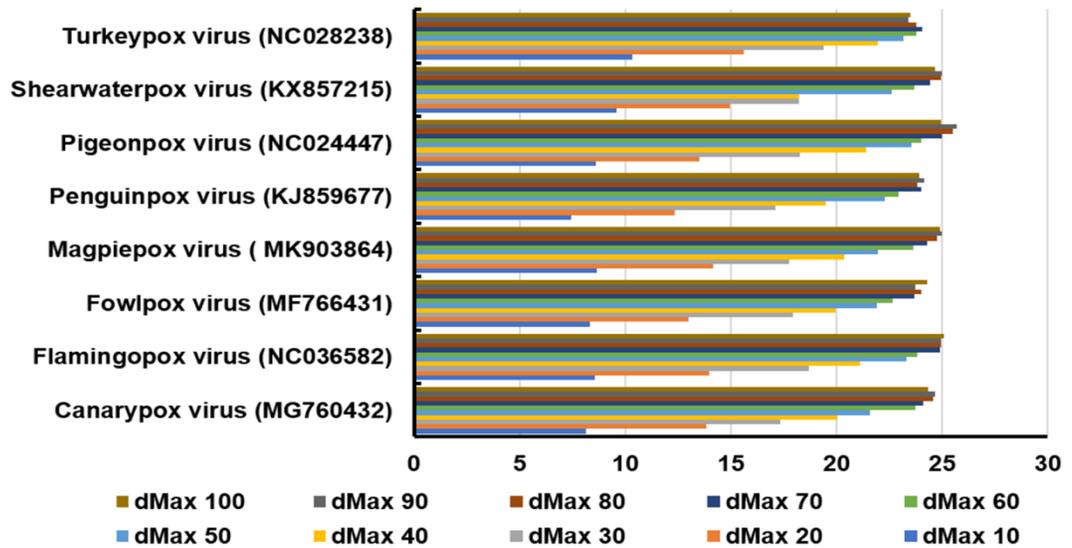


Figure 7.2: Effect of dMAX. Percentage of compound microsatellite differed with various dMAX values (10–100). The accession number of each species of the APV genus is mentioned on the right side of the graph. An increased incidence of compound microsatellite was observed with a higher value of dMAX.

7.2.2 Frequency of classified repeat types

The overall frequency of mononucleotide repeats was 34.20% for A/T, while that of C/G was 0.33%. In case of the dinucleotide repeat motif, AT/TA (43.80%) were the most prevalent than AG/CT repeats (5.55%), AC/GT(4.23%) and CG/GC (0.27%). Analysis of the classified tri-repeat types revealed that the APV genome had 10 types of trinucleotide from which AAT/ATT, ATC/ATG, AAG/CTT, ACT/AGT, AAC/GTT, ACC/GGT, AGG/CCT, AGC/CTG, ACG/CGT, and CCG/CGG were abundantly present exhibiting 5.90%, 1.60%, 1.44%, 1.02%, 0.54%, 0.18%, 0.15%, 0.14%, 0.13%, and 0.01% respectively. The prevalent tetranucleotide repeats were AAAT/ATTT (0.22%), AAGT/ACTT (0.06%),

AATT/AATT (0.06%), ACAT/ATGT (0.06%), AAAG/CTTT (0.05%), AACT/AGTT (0.05%), AGAT/ATCT (0.03%), ATCC/ATGG (0.02%), AAAC/GTTT (0.01%), AATC/ATTG (0.01%), AACC/GGTT (0.01%), AAGG/CCTT (0.01%), AATG/ATTC (0.01%), ACCT/AGGT (0.01%), and ATGC/ATGC (0.01%). The penta and hexanucleotide repeats present in the AVP genomes were AAAAT/ATTTT, AAATT/AATTT, AAAAG/CTTTT, AATAT/ATATT, AATGT/ACATT, AGGGG/CCCCT, AAAGAT/ATCTTT, AAATAT/ATATTT, AAATGG/ATTTCC, AAGAAT/ATTCTT, AAGATG/ATCTTC, AAGGGT/ACCCTT, AAGTAT/ACTTAT, AATCAT/ATGATT, AATGAT/ATCATT, ACGGAG/CCGTCT, AGATAT/ATATCT, AGGCGG/CCGCCT, ATATGC/ATATGC respectively but were very sparsely populated. However, the accession specific analysis illustrated that, the frequency of mono, di, tri repeats were varied from each other (Figure 7.3 A-C).

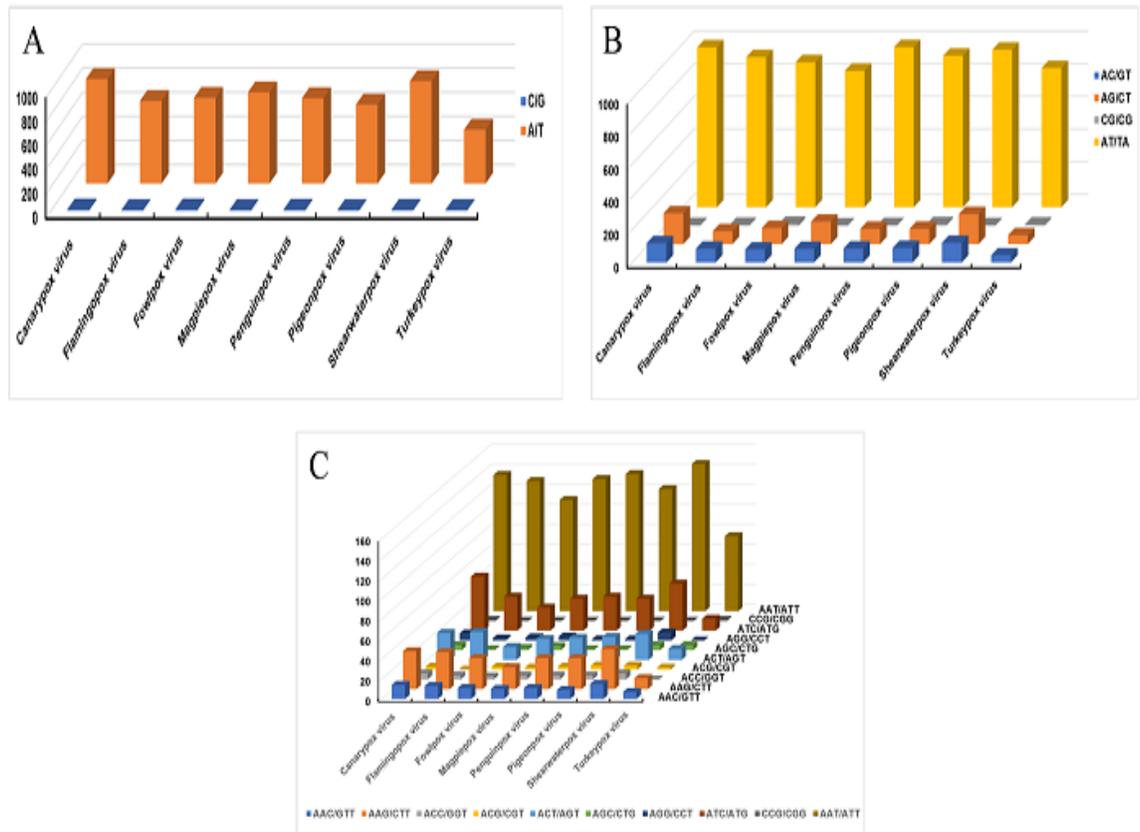


Figure 7.3 A-C: Distribution of classified microsatellites among APV genomes. (A) Distribution pattern of mononucleotide SSRs within APV genomes, (B) distribution pattern of dinucleotide SSRs within APV genomes, and (C) distribution pattern of trinucleotide SSRs within APV genomes.

7.2.3 Motif complexity of compound microsatellites

cSSRs usually consist of more than two individual microsatellites close together. According to their microsatellite constituents, these are classified as '2-microsatellite' or '3-microsatellite' having the pattern like m1-xn-m2/m1-xn-m2-xn-m3 [30]. The cSSRs of APV genomes majorly consists of two motifs, which contribute 99% of total

cSSR, followed by tri and tetra motifs. Evidence suggests that the cSSR, CTG-CAG composed of self-complementary motifs, has evolved through the process of recombination [31]. However, such a cSSR motif was not observed during our study, indicating the lack of correlation between recombination processes with the cSSR formation. Motifs with maximum representation within the genome have the pattern $[m1]n-xn-[m2]n$ termed as SSR-couples. We observed microsatellite couples, like $(TA)_{3-x}-(TC)_3, (TAA)_{3-x}-(TAT)_3, (AG)_{3-x}-(AT)_4, (TC)_{3-x}-(TA)_3, (AG)_{3-x}-(AT)_3, (GC)_{3-x}-(GA)_3, (AT)_{3-x}-(AG)_3, (TA)_{3-x}-(TAT)_3$ within the analyzed genome. Self-complementary cSSR like $(AT)_{3-x1}-(TA)_3, (AT)_{3-x0}-(TA)_3, (TA)_{3-x6}-(AT)_3, (TA)_{3-x5}-(AT)_3, (TA)_{3-x7}-(AT)_3, (TA)_{4-x0}-(AT)_3, (AT)_{3-x9}-(TA)_3, (AT)_{3-x6}-(TA)_3, (AG)_{3-x1}-(TC)_3$ observed within APV, those leads to the formation of secondary structure within the genome. Motif composition like $(CA)_n-(X)y-(CA)_z$ resembled duplication phenomena in which a similar motif is located on both ends of the spacer sequence. About 0.92% of the total cSSR were made up of duplicated sequences having the motif pattern $(TA)_{3-x5}-(TA)_3, (T)_{6-x2}-(T)_9, (AT)_{3-x1}-(AT)_3, (AT)_{3-x8}-(AT)_3, (T)_{6-x5}-(T)_6, (TA)_{3-x7}-(TA)_3, (A)_{6-x6}-(A)_6, (T)_{7-x2}-(T)_7, (ATC)_{3-x3}-(ATC)_{3-x3}-(ATC)_{3-x3}-(ATC)_3, (AG)_{3-x9}-(AG)_3, (A)_{10-x7}-(A)_6, (TCA)_{3-x3}-(TCA)_3, (T)_{6-x2}-(T)_9$. The Circos map generated here in Figure 7.4 clearly showcases the distribution pattern of the SSRs and cSSRs within the selected APV genomes.

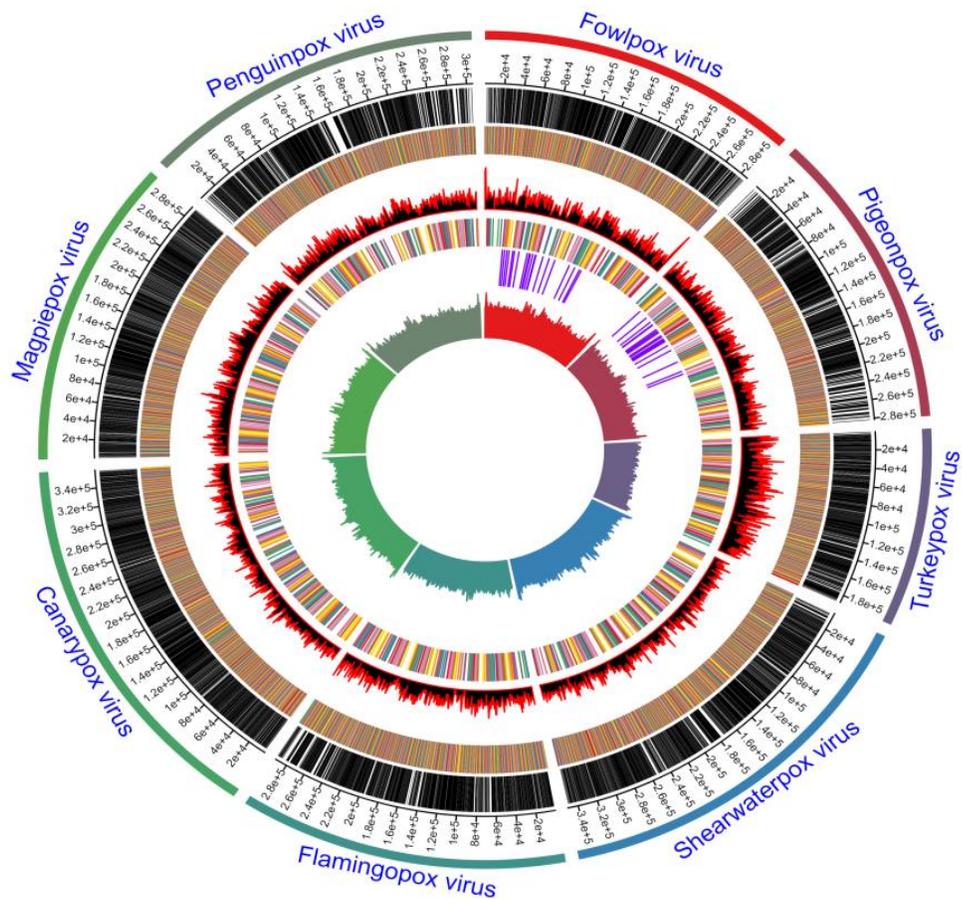


Figure 7.4: Circos map construction. Circos plot demonstrating the genome-wide study of SSRs and cSSRs across eight different species of Avipoxvirus. From outer to inner ring: Species, Genome size, Genes, SSRs, SSRs frequency/1kbs, cSSRs, developed cSSR markers, and GC content.

7.2.4 Development and characterization of cSSR markers

We have targeted the development and characterization of compound microsatellites during this study due to their higher polymorphic rate [32]. So, randomly twenty sets of cSSR

markers were chosen from FPV and PPV and validated in vitro by amplifying each primer set with the respective isolate collected during our previous study [13]. The compound microsatellites' name, sequence information of the primer, annealing temperature, expected product size, ORF, and selected motif summarized in Table 7.2. Each tested primer pair forms an amplicon having expected size bands within the agarose gel (Figure 7.5). These amplicons were further verified through nucleotide sequencing, followed by the utilization of BLASTn and BLASTx for mapping. BLASTn tool deciphered 66-100% coverage of query with a high percentage of identity (88-100%) within sequencing reads of the present study and the existing APV complete genome's respective genes. Intriguingly, we have identified PPV SSR 15 showing polymorphism with the other strain like penguinpox virus, pigeonpox virus, and flamingopox virus having an additional T (Figure 7.6). BLASTx analysis showed query coverage 44-100% with an elevated percentage of identity (62-100%) with the respective genes' amino acid sequences (Table 7.3). The concatenated phylogenetic tree built by using the neighbor-joining method with bootstrap value 1000 of MEGA 7 revealed that the FPV and PPV are closely related to each other and with 99% similarity Feral pigeon belonging to South Africa (NC024447) (Figure 7.7).

Name	Primer sequence (5-----3)	(Motif) length	Position	Gene	Expected Size (bp)
FPV-SSR1-F	TTCCGATACCATGG ATGAAC	(A)8-x1-(TA)3- x5-(A)6	31785- 31810	G protein-coupled receptor family protein	197
FPV-SSR1-R	TTTTTGTATAATTC GTCAAAGGTTTTT				
FPV-SSR2-F	TGATATACAACCC CCGAACC	(AT)3-x5-(T)7	35736- 35753	alkaline phosphodiesterase	212
FPV-SSR2-R	AAGAATGCATAAG CGATGAAGA				
FPV-SSR3-F	GCTACGTGAATAG GCGCTTT	(AT)3-x5-(T)7	40274- 40291	ankyrin repeat protein	228
FPV-SSR3-R	AAAAGACGATAAA GGATATACAGCA				
FPV-SSR4-F	CAACCTTCTGTTTC ATCTACGG	(AT)3-x1-(AT)3	44260- 44272	serpin family protein	243
FPV-SSR4-R	TTCTTCTTTATGCG CGTTGA				
FPV-SSR5-F	GCGATATTACACG GTATTACAAA	(GAT)3-x3- (TA)3	51043- 51060	semaphorin-like protein	247
FPV-SSR5-R	TGTTGTGAGCCAG AATGTCC				
FPV-SSR6-F	TAAACGATCCCTTC AGGAAA	(TA)3-x8-(TC)3	54222- 54241	rifampicin resistance protein N3L	214
FPV-SSR6-R	GCATTGAGAGTGG TAGGTCTAAA				
FPV-SSR7-F	TCCAAGGTAGCTA ATTGATGAGG	(AT)3-x1-(T)6	62434- 62444	NTPase	186
FPV-SSR7-R	TACGGGCAAGACA TTGACAG				
FPV-SSR8-F	CGGTGTTAGGTAC AAACTCCAT	(T)6-x9-(A)6	77928- 77948	N1R/p28 family protein	237
FPV-SSR8-R	AAGGGCATTGATG AGTCAAGA				
FPV-SSR9-F	TCATTGCGCTAACGA GGATCA	(AT)3-x4-(AT)3	81938- 81953	putative metalloprotease	179
FPV-SSR9-R	GCGCTTTTGGTATT CATGGT				
FPV-SSR10-F	TGAACGGCGTTAA GAACCTC	(AT)3-x2-(T)7	83774- 83788	RNA helicase NPH-II	216
FPV-SSR10-R	CGATTCCAAAGCG TATACAGA				
FPV-SSR11-F	CTATTGCGACGCTT CGTTTT	(GT)3-x8-(T)7	88186- 88206	thymidine kinase	201
FPV-SSR11-R	TATCGAACTCCATT CCGTGT				
FPV-SSR12-F	GCCAAAAACGGGT TACATTG	(T)7-x2-(T)7	93089- 93104	putative virion core protein	225
FPV-SSR12-R	GCATTACCGTGTGC TGAATG				
FPV-SSR13-F	TGATGCGACGGAA ACAGTTA	(AAT)3-x9-(T)6	94305- 94328	DNA polymerase	245
FPV-SSR13-R	TTACCGTGGCTTCC AATGTT				

FPV-SSR14-F	GTCTCAGTACCCG GCATCAT	(AT) ^{3-x8} -(TC) ³	107530- 107549	variola B22R family protein	217
FPV-SSR14-R	CATCGGGTATTATC GCGTCT				
FPV-SSR15-F	TCCATCATCTACCG AATACCAA	(T) ^{6-x5} -(A) ⁹	119237- 119256	poly(A) polymerase large subunit	193
FPV-SSR15-R	TTTCGCGTGTTTCT GTTTTG				
FPV-SSR16-F	TCGAGCGATATTCC TGTATCC	(AT) ^{3-x0} -(A) ⁸	130170- 130183	EEV maturation protein	237
FPV-SSR16-R	TCAGTGCCGGTTTT ACAATG				
FPV-SSR17-F	TGGGAGAAACTGT CTTGCAT	(GT) ^{3-x2} - (GAA) ⁴	158345- 158364	V-type Ig domain containing protein	190
FPV-SSR17-R	CCGTCGCATATATC CACTTG				
FPV-SSR18-F	CGCCGCTAATTAC ACAAAGA	(A) ^{6-x6} -(TA) ⁴	169249- 169268	RNA polymerase subunit RPO163	214
FPV-SSR18-R	CATGGGATTAGTTT CGTCCAA				
FPV-SSR19-F	CCAAGGGAGAAGA TTCTACAACA	(TA) ^{4-x1} -(AT) ³	178894- 178908	mRNA capping enzyme large subunit	209
FPV-SSR19-R	TTCAACGCTTCCGT TGAGAT				
FPV-SSR20-F	CCATCCTCAATTAG CGTTGT	(AT) ^{3-x3} -(AT) ³	182371- 182385	p28-like protein	234
FPV-SSR20-R	TGTAAAGCCTCAA ACCCACA				
PPV-SSR1-F	TCGTCCTTGTTATC GTAGGATTC	(AT) ^{3-x1} -(T) ⁶	16487- 16497	serpin family protein	238
PPV-SSR1-R	TCATTA AATTGATA ACAGCAATGAG				
PPV-SSR2-F	GCAGACATTTTGA GCCGTTA	(AC) ^{3-x5} - (CTT) ³	33608- 33627	ankyrin repeat protein	222
PPV-SSR2-R	AATTATACCATAC ATAATGGCCTGT				
PPV-SSR3-F	CGGATATCGCGGC AATAATA	(AT) ^{3-x4} -(T) ⁶	44579- 44594	alkaline phosphodiesterase	214
PPV-SSR3-R	ACCGCTATGGATC ACGGATA				
PPV-SSR4-F	TTTGGTACGATGGT TGATACAGA	(T) ^{8-x0} -(AT) ³	50775- 50788	serpin family protein	184
PPV-SSR4-R	TTTATGCGCGTTGG GTAAAT				
PPV-SSR5-F	CAGGAGGAGACAT AGAAGCTTTG	(AT) ^{3-x7} -(TG) ³	51898- 51916	DNA ligase	250
PPV-SSR5-R	AGAAATGCAATTC CTGATTGG				
PPV-SSR6-F	TGCACCGGAAGTA ATGAAAG	(TA) ^{3-x2} -(A) ⁶ - x8-(A) ⁷	62537- 62565	NPH-1 transcription termination factor	249
PPV-SSR6-R	GCCGACAAACATA TCAGTTTATTAT				
PPV-SSR7-F	GAGGATTTGTTACC AGTAAAGCA	(TA) ^{3-x2} -(T) ⁶	65599- 65612	V-type Ig domain protein	192

PPV-SSR7-R	GTAAAAGCAGCAC CACGTCA				
PPV-SSR8-F	CACAAAGAAATGA TCTTCTGTTATCTT	(TA) ^{3-x8} -(T) ⁶	67183- 67202	early transcription factor VETF	161
PPV-SSR8-R	CGGACGTTGTATTA GGAGAGC				
PPV-SSR9-F	TCGTATACGCCATT AAGAAATTG	(TAT) ^{3-x6} - (CA) ^{3-x5} -(CA) ³	70177- 70208	DNA replication protein	232
PPV-SSR9-R	GAGGAGAGTATCT AGTTTGGCTTAAA A				
PPV-SSR10-F	TTTTTGTGATGGGT TGCCTA	(AGT) ^{3-x6} - (TA) ³	77504- 77524	glutathione peroxidase	207
PPV-SSR10-R	GGACCAGGTACAT CTCCAGT				
PPV-SSR11-F	TTGATGACATTTGA TCATCGTT	(T) ^{6-x8} -(TA) ³	80295- 80314	virion protein	200
PPV-SSR11-R	CATCCAACGTTGTC TGTCCT				
PPV-SSR12-F	GGACAAGCCTTTA ATGGCAAC	(T) ^{6-x7} -(T) ⁶	84858- 84876	N1R/p28 family protein	242
PPV-SSR12-R	GGGTTTTTCCTAGC ATTGCAC				
PPV-SSR13-F	TATGCAAACGACG TGCATTC	(AT) ^{3-x4} -(TA) ³	87512- 87527	transforming growth factor B	218
PPV-SSR13-R	AAGAAGCTTTTAG GTGTTGGTGA				
PPV-SSR14-F	AATTTTCCTTGGTT GAAGAATG	(AT) ^{3-x4} -(AT) ³	89239- 89254	metalloprotease	230
PPV-SSR14-R	TTTGGATCTGTTTC TGGTGTA AAA				
PPV-SSR15-F	ACGTCCTCTACCGG GTTTCT	(T) ^{6-x7} -(T) ^{6-x7} - (AC) ³	90978- 91009	DNA/RNA helicase/NPH-11	246
PPV-SSR15-R	AAATATTCGTACG ATTCCAAAGC				
PPV-SSR16-F	AAACCAAAAACGG GCTACG	(T) ^{7-x2} -(T) ⁷	100399- 100414	sulfhydryl oxidase ERV1	230
PPV-SSR16-R	AAAGCATTACCGT GCTCTGAA				
PPV-SSR17-F	AAACCAAAAACGG GCTACG	(CA) ^{3-x8} -(T) ⁶ - x5-(AC) ³	111336- 111366	B22R family protein	230
PPV-SSR17-R	AAAGCATTACCGT GCTCTGAA				
PPV-SSR18-F	TGAACAACCAATA CTTTCATCGTT	(A) ^{6-x5} -(T) ⁶	136869- 136885	virion release protein	204
PPV-SSR18-R	TCAGAATTATTATT GGGGGTTTC				
PPV-SSR19-F	CTATTGGGCTCGGT AGTTGC	(GC) ^{3-x7} -(GA) ³	156830- 156848	DNA-binding virion core VP8	154
PPV-SSR19-R	CTCCCGCCGCTATT ATTTC				
PPV-SSR20-F	CGGACAGATTGTG TACGGTAA	(A) ^{6-x6} -(TA) ⁴	162754- 162773	RNA polymerase subunit RPO147	248

Table 7.2: The characteristics of forty compound SSR developed using a clinical sample.

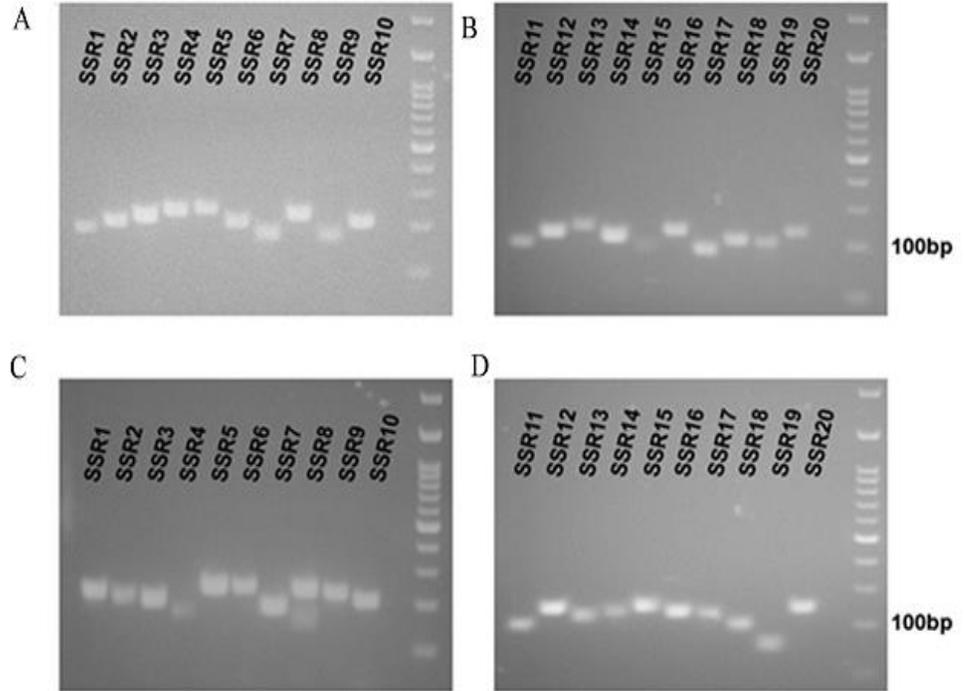


Figure 7.5: Microsatellite markers development using clinical samples of APV. Validation of compound microsatellites primers using a clinical sample. M:100 bp DNA ladder; -C: Negative controls (PCR using nuclease-free water as DNA template). (a) Fowlpox virus (FPV SSR1-FPV SSR10) (b) Fowlpox virus (FPV SSR11-FPV SSR20) (c) Pigeonpox virus (PPV SSR1-PPV SSR10) (d) Pigeonpox virus (PPV SSR11-PPV SSR20).

PPV-SSR5	100%	2.00E-69	93%	44%	1.00E-04	79%
PPV-SSR6	100%	1.00E-91	99%	65%	4.00E-19	100%
PPV-SSR7	100%	6.00E-61	99%	87%	1.00E-10	100%
PPV-SSR8	92%	8.00E-34	93%	64%	9.00E-08	100%
PPV-SSR9	100%	8.00E-87	100%	98%	8.00E-33	100%
PPV-SSR10	100%	2.00E-69	98%	80%	2.00E-19	91%
PPV-SSR11	100%	2.00E-69	99%	94%	3.00E-25	100%
PPV-SSR12	100%	2.00E-88	98%	88%	7.00E-23	81%
PPV-SSR13	100%	2.00E-74	98%	99%	1.00E-27	96%
PPV-SSR14	100%	1.00E-84	99%	97%	1.00E-29	95%
PPV-SSR15	100%	3.00E-61	91%	73%	6.00E-06	63%
PPV-SSR16	100%	2.00E-88	99%	80%	4.00E-28	100%
PPV-SSR17	100%	1.00E-84	99%	50%	1.00E-11	100%
PPV-SSR18	100%	5.00E-69	99%	93%	7.00E-21	96%
PPV-SSR19	100%	4.00E-43	99%	87%	3.00E-11	100%
PPV-SSR20	100%	4.00E-91	99%	99%	2.00E-32	94%

Table 7.3: Characterization of forty cSSR markers against the NCBI database.

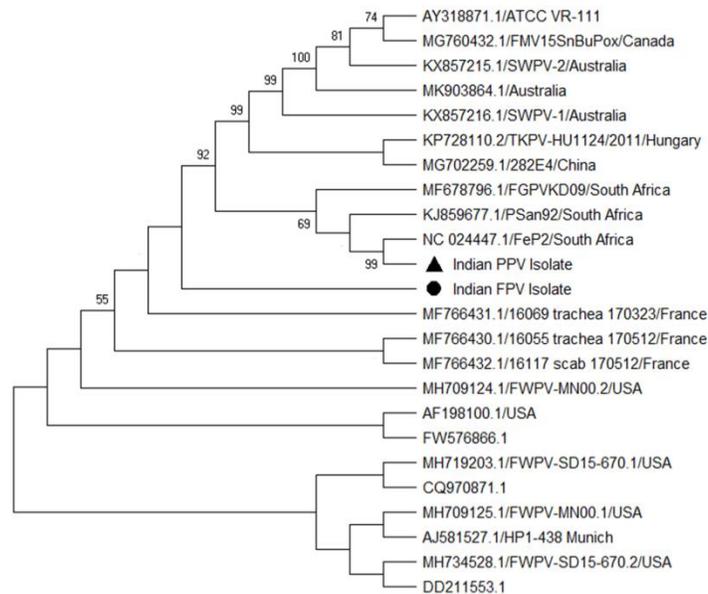


Figure 7.7: Construction of a concatenated phylogenetic tree. A consensus phylogenetic tree constructed using a concatenated sequence of microsatellites by the neighbor-joining method with bootstrap value 1000 of MEGA7. During

the study, the clinical isolate represented a black triangle and black circle showing its relationship and Feral pigeon strain.

7.3 Discussion

APV is regarded as a stable, extremely contagious virus, having transboundary potential that can profoundly impact livelihood. Once introduced to a new geographic area, APV can quickly spread among individuals lacking proper husbandry practices and elevated stocking densities that lead to close contact within infected and susceptible individuals. As continuously, the outbreak of APV reported from India in different birds from diverse geographical regions for decades [33, 13, 34]. That influences us to investigate the viral genome, followed by specific molecular markers to monitor epidemiological surveillance and prevent its spread.

Microsatellite markers act as a genomic fingerprint, well studied within the animal [35, 36, 37], plant [38, 39], yeast [40, 18, 41], bacteria [42, 43]. Although few reports suggest its existence within the virus long ago, it got less attention than other organisms [32]. Since the past decade, using the *in-silico* approach, molecular markers in many familiar viruses are extensively analyzed. These include Human Immunodeficiency Virus [44], Ebolavirus [45], Hepatitis C virus [46], Human papillomavirus [47], Caulimoviruses [48] as well as their application within Ostreid herpesvirus 1 [49], HSV-1 [50], Alphaherpesviruses [51], Cyprinid herpesvirus-3 [52], Herpes Simplex Virus 2 [21], pseudorabies virus [53], Salmon Gill Poxvirus [28], Orf virus [54]. This motivated us to perform in-depth *in-silico* analysis followed by their validation of these markers through a comparative genomics approach within APV for the first time.

The specific parameters, such as the range of RA, ranged from 6.87-8.12 and RD from 45.8-53.58 in analyzed APV genomes. In other poxviruses like HPV, RA, RD ranged from 3.62-8.34, 23.95-59.15, respectively, while in herpesviruses, the value obtained 4.13-13.31 and 26.92-102.91 [55]. During the analysis of cRA, cRD varies from 0.5-0.71, 9.39-12.92 respectively within analyzed APV genomes, whereas, in other DNA viruses such as HPV and herpesviruses it ranged from 0-1.26, 0-27.3 and 0.16-1.82, 2.21-35.10, respectively [55, 47]. A lower percentage of cSSR (7.29- 8.81%) obtained in comparison to the herpes virus, 8.12-33.31% [55], HPV, 0-15.15% [47]. The absence of major differences among the parameters indicating the similarity of the genomic complexity within the virus derived from the APVgenus. In APV, 0.92% of cSSRs are composed of similar motifs, probably contributed by genome duplication. Some studies suggest genome duplication acquisition of new or more complex transcription regulation mechanisms [56]. An increased percentage of cSSR observed with increasing dMAX until 90bp and further decrease (Figure 7.2). It might be possible due to an increase in the overlapping region of SSR with a higher dMAX value.

In APV genomes, the poly(A/T) repeats were significantly higher in all APV genome than poly(G/C) repeats, which expresses similar pattern in eukaryotic and prokaryotic genomes [57, 58]. Interestingly, we have observed a mononucleotide SSR showing polymorphism against SSR 15 with related species.

Di-nucleotide AT/TA is more prevalent in all analyzed AVP genomes, similar to that of DNA viruses such as HPV [47],

caulimovirus [59], geminiviruses [60]. Di-nucleotide repeats have the highest slippage rate than any other type of repetitions [61]. Among 257 viral genomes examined in a published study, the highest number of di-nucleotide SSRs found compared to the other types [62]. Di-nucleotide repeats speculated to be recombination hot spots [63, 64]. In the present study, the presence of more di-nucleotide repeats over tri-nucleotide repeats suggests a possible role of hosts in the evolution of di-nucleotide repeats within poxvirus genomes. Inconsistency frequency of SSR in different accessions of the same virus was attributed to genome instability because of a higher slippage rate [65]. Trinucleotide motif AAT/ATT followed by ATC/ATG, and AAG/CTT were most prevalent in most genomes of APV, whereas in other DNA virus GAG/AGA was most prevalent in HPV and AAG/GAA in DNA A and caulimovirus. The higher density of trinucleotide repeats observed compared to any other repeat type within the eukaryotic and prokaryotic genomes [66]. Interestingly, dynamic mutations within the trinucleotide repeat responsible for the development of diseases [67]. Intracellular mature virions (IMV), essential for virus replication in cell culture, predominantly exist within the intact cell's cytoplasm and are only released upon cell lysis. Microsatellites controlled several cellular functions, such as replication, recombination, repair mechanisms, and sequence diversity drives adaptive forces [68]. As a result of this, pathogens gain the ability to utilize SSRs to frustrate the host immune system to enhance their antigenic variability [62]. It would be interesting to decipher the polymorphism's functional outcome related to these genes within the APV genome.

Compared to the simple microsatellites, compound microsatellites have a higher slippage rate/polymorphism, which infer for strain identification and evolutionary estimation in several viruses such as Herpes simplex virus [21] and Adenovirus [26]. Recently, a compound microsatellite within begomovirus genomes suggests that the cSSRs significantly correlated with recombination hot spots [69]. These fundamental reasons influence us to target cSSR for development and characterization within the AVP genome. During our study, all the cSSR chose amplified perfectly, followed by dendrogram construction, suggesting that these markers can be an efficient tool for strain identification.

7.4 Materials and methods

7.4.1 Genome sequences

Eight complete genome sequences belonging to one strain of each species were selected randomly from APV from a publicly available database (www.ncbi.nlm.nih.gov). These sequences are used for genome-wide microsatellites identification through the in-silico approach. The relative density (RD) and relative abundance (RA) values estimated by the net acquisition of microsatellites (total bp) within the viral genome and the number of SSR present per kb within the viral genome (Table 7.1).

7.4.2 Microsatellite identification, investigation, and statistical analysis

The perfect mono to hexa SSRs and cSSR identified utilizing the IMEx server [70]. The ‘Advance-Mode’ of IMEx server used having parameters similar to the previous report for DNA viruses [13] and RNA viruses [59]. The parameters included:

repeat types: perfect; with minimum repeat number: 6, 3, 3, 3, 3, 3 for mono, di, tri, tetra, penta, and hexanucleotide repeats, respectively. The distance within two microsatellites (dMAX) was fixed to ten nucleotides during the analysis. The rest of the parameters are used as default. Similar parameters were implemented to decipher compound microsatellites (cSSR) composition.

7.4.3 Microsatellite PCR analysis and sequencing

A compound microsatellite was used as an excellent choice to utilize as a molecular marker due to the high polymorphic nature [17, 32]. So, twenty sets of compound microsatellites markers were randomly chosen from FPV and PPV. Primer3Plus web tool (<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/>) with the target to specific motif used to design the primer pairs having unique flanking regions. The primers' length kept between 18 and 22 bp with varied product sizes ranging from 154 to 250 bp. These cSSR primers further examined through an in-vitro approach using one fowlpox virus isolate (FPV/Sub-Od-4b/01) and one pigeon pox isolate (PPV/Dhe-Od-4b/01), recently reported from our study [13]. The PCR conditions for all cSSR were set as follows: initial denaturation step at 95°C for 5 min, 35 cycles denaturation 95 °C for 50 s, annealing (54-62°C) for each cSSR primer, extension step at 72 °C for 90 s and the final extension cycle at 72 °C for 7 min (Table 7.2). The amplicons were resolved through electrophoresis using a 3% agarose gel stained with ethidium bromide and run at 90 V followed by the image's capture using iBright CL750 Imaging System (Invitrogen, USA). The amplicons, further purified using PureLink™ PCR Purification Kit (Invitrogen, USA)

followed by nucleotide sequencing twice in each direction using 3100 ABI sequencers (Applied Biosystems, USA) [71]

7.4.4 Sequencing data analysis and construction of the phylogenetic tree

The sequencing result further validated comparing with the genome of APV present in the GenBank by using discontinuous-MegaBLAST [72] followed by polymorphism analysis. To decipher the nucleotide query's translational products, the sequencing product was subjected to the BLASTx analysis (<http://www.ncbi.nlm.nih.gov>). A consensus phylogenetic tree was constructed utilizing a concatenated sequence of microsatellites. The homologous sequence obtained during mapping, followed by the neighbor-joining method with bootstrap value 500 of MEGA 5 to reveal the genetic relation of those clinical samples with other APV. The Circos software was used to generate the Circos map depicting different parameters, including genome size, CDS, GC content, SSRs, and cSSRs distribution among the designated APV genomes.

7.5 Conclusion

In conclusion, the comprehensive analysis of AVP's SSRs is a first step to deciphering the type, distribution, and possible implications within viral species. During this study, the markers developed can be used as suitable multiplex PCR panels to solve virus genotype and evolutionary status by using geographically diverse clinical samples. This introductory study needs further evaluation by doing continuous outbreak surveillance, followed by constructing a

useful genotyping assay utilizing these microsatellites to evaluate clinical samples.

7.6 References

1. Bolte A.L., Meurer J., Kaleta E.F. (1999), Avian host spectrum of avipoxviruses. *Avian Pathology*. 28: p. 415-432 (DOI: 10.1080/03079459994434).
2. Van Riper C., Forrester D. (2008), Avian pox., in: Thomas NJ, H.D., Atkinson, CT (Eds.), *Infectious diseases of wild birds*. Blackwell Publishing Professional. p. 131-171 (DOI: 10.1111/tbed.13413).
3. Afonso C.L., Tulman E.R., Lu Z., Zsak L., Kutish G.F., Rock D.L. (2000), The genome of fowlpox virus. *Journal of virology*. 74: 3815-3831 (DOI: 10.1111/tbed.13413).
4. Tulman E., Afonso C., Lu Z., Zsak L., Kutish G., Rock D. (2004), The genome of canarypox virus. *Journal of virology*. 78: 353-366 (DOI: 10.1128/jvi.78.1.353-366.2004).
5. Offerman K., Carulei O., van der Walt A.P., Douglass N., Williamson A.-L. (2014), The complete genome sequences of poxviruses isolated from a penguin and a pigeon in South Africa and comparison to other sequenced avipoxviruses. *BMC genomics*. 15: p. 463 (DOI: 10.1186/1471-2164-15-463).
6. Bányai K., Palya V., Dénes B., Glávits R., Ivanics É., Horváth B., Farkas S.L., Marton S., Bálint Á., Gyuranecz M. (2015), Unique genomic organization of a novel Avipoxvirus detected in turkey (Meleagris

- gallopavo). *Infection, Genetics and Evolution*. 35: p. 221-229 (DOI: 10.1016/j.meegid.2015.08.001).
7. Sarker S., Das S., Lavers J.L., Hutton I., Helbig K., Imbery J., Upton C., Raidal S.R. (2017), Genomic characterization of two novel pathogenic avipoxviruses isolated from pacific shearwaters (*Ardenna* spp.). *BMC genomics*. 18: p. 298 (DOI: 10.1186/s12864-017-3680-z).
 8. Carulei O., Douglass N., Williamson A.L. (2017), Comparative analysis of avian poxvirus genomes, including a novel poxvirus from lesser flamingos (*Phoenicopterus minor*), highlights the lack of conservation of the central region. *BMC genomics*. 18: p. 947 (DOI 10.1186/s12864-017-4315-0).
 9. Sarker S., Batinovic S., Talukder S., Das S., Park F., Petrovski S., Forwood J.K., Helbig K.J., Raidal S.R. (2020), Molecular characterisation of a novel pathogenic avipoxvirus from the Australian magpie (*Gymnorhina tibicen*). *Virology*. 540: p. 1-16 (DOI: 10.1016/j.virol.2019.11.005).
 10. Jarmin S., Manvell R., Gough R.E., Laidlaw S.M., Skinner M.A. (2006) , Avipoxvirus phylogenetics: identification of a PCR length polymorphism that discriminates between the two major clades. *Journal of General Virology*. 87: p. 2191-2201 (DOI: 10.1099/vir.0.81738-0).
 11. Atkinson C.T., LaPointe D.A. (2009), Introduced avian diseases, climate change, and the future of Hawaiian honeycreepers. *Journal of Avian Medicine. Surgery* 23: p. 53-63 (DOI: 10.1647/2008-059.1).

12. Parker P.G., Buckles E.L., Farrington H., Petren K., Whiteman N.K., Ricklefs R.E., Bollmer J.L., Jiménez-Uzcátegui G. (2011), 110 years of Avipoxvirus in the Galapagos Islands. *Plos one*. 6: p. e15989 (DOI: 10.1371/journal.pone.0015989).
13. Sahu B.P., Majee P., Mishra C., Dash M., Biswal S., Sahoo N., Nayak D.J.T., Diseases E., (2020a), The emergence of subclades A1 and A3 avipoxviruses in India. 67: p. 510-517 (DOI: 10.1111/tbed.13413).
14. Sharma P.C., Grover A., Kahl G. (2007), Mining microsatellites in eukaryotic genomes. *Trends in biotechnology*. 25: p. 490-498 (DOI: 10.1016/j.tibtech.2007.07.013).
15. Sreenu V.B., Alevoor V., Nagaraju J., Nagarajaram H.A. (2003), MICdb: database of prokaryotic microsatellites. *Nucleic Acids Research*. 31: p. 106-108 (DOI: 10.1093/nar/gkg002).
16. Renault T., Tchaleu G., Faury N., Moreau P., Segarra A., Barbosa-Solomieu V., Lapègue S. (2014), Genotyping of a microsatellite locus to differentiate clinical Ostreid herpesvirus 1 specimens. *Veterinary Research*. 45: p. 3 (DOI: 10.1186/1297-9716-45-3).
17. Bull L.N., Pabón-Peña C.R., Freimer N.B. (1999), Compound microsatellite repeats: practical and theoretical features. *Genome Research*. 9: p. 830-838 (DOI: 10.1101/gr.9.9.830).
18. Gendrel C.G., Boulet A., Dutreix M. (2000), (CA/GT)_n microsatellites affect homologous recombination during yeast meiosis. *Genes Development*. 14: p. 1261-1268 (DOI: 10.1101/gad.14.10.1261).

19. Saeed A.F., Wang R., Wang S. (2016), Microsatellites in pursuit of microbial genome evolution. *Frontiers in microbiology*. 6: p. 1462 (DOI: 10.3389/fmicb.2015.01462).
20. Sahoo L., Patel A., Sahu B.P., Mitra S., Meher P.K., Mahapatra K.D., Dash S.K., Jayasankar P., Das P. (2015), Preliminary genetic linkage map of Indian major carp, *Labeo rohita* (Hamilton 1822) based on microsatellite markers. *Journal of Genetics*. 94: p. 271-277 (DOI: 10.1007/s12041-015-0528-7).
21. Burrel S., Ait-Arkoub Z., Voujon D., Deback C., Abrao E.P., Agut H., Boutolleau D. (2013), Molecular characterization of herpes simplex virus 2 strains by analysis of microsatellite polymorphism. *Journal of clinical microbiology*. 51: p. 3616-3623 (DOI: 10.1128/JCM.01714-13).
22. Deback C., Boutolleau D., Depienne C., Luyt C., Bonnafous P., Gautheret-Dejean A., Garrigue I., Agut H.J.J.o.c.m. (2009), Utilization of microsatellite polymorphism for differentiating herpes simplex virus type 1 strains. 47: p. 533-540 (DOI: 10.1128/JCM.01565-08).
23. Sahu B., Sahoo L., Joshi C., Mohanty P., Sundaray J., Jayasankar P., Das P. (2014), Isolation and characterization of polymorphic microsatellite loci in Indian major carp, *Catla catla* using next-generation sequencing platform. *Biochemical Systematics*. 57: p. 357-362 (DOI: 10.1016/j.bse.2014.09.010).
24. Pignatelli S., Monte P.D., Rossini G., Landini M.P. (2004), Genetic polymorphisms among human cytomegalovirus (HCMV) wild-type strains. *Reviews*

- in medical virology. 14: p. 383-410 (DOI: 10.1002/rmv.438).
25. Pradeep B., Shekar M., Karunasagar I., Karunasagar I. (2008), Characterization of variable genomic regions of Indian white spot syndrome virus. *J Virology*. 376: p. 24-30 (DOI: 10.1016/j.virol.2008.02.037).
 26. Hounq H.S., Lott L., Gong H., Kuschner R.A., Lynch J.A., Metzgar D. (2009), Adenovirus microsatellite reveals dynamics of transmission during a recent epidemic of human adenovirus serotype 14 infection. *J Clin Microbiol* 47: p. 2243-2248 (DOI: 10.1128/JCM.01659-08).
 27. Atia M.A.M., Osman G.H., Elmenofy W.H. (2016), Genome-wide In Silico Analysis, Characterization, and Identification of Microsatellites in *Spodoptera littoralis* Multiple nucleopolyhedrovirus (SpliMNPV). *Sci Rep*. 6: p. 33741-33741 (DOI: 10.1038/srep33741).
 28. Gulla S., Tengs T., Mohammad S.N., Gjessing M., Garseth Å.H., Sveinsson K., Moldal T., Petersen P.E., Tørud B., Dale O.B. (2020), Genotyping of Salmon Gill Poxvirus Reveals One Main Predominant Lineage in Europe, Featuring Fjord-and Fish Farm-Specific Sub-Lineages. *Frontiers in microbiology*. 11: p. 1071 (DOI: 10.3389/fmicb.2020.01071).
 29. Beier S., Thiel T., Münch T., Scholz U., Mascher M. (2017), MISA-web: a web server for microsatellite prediction. *Bioinformatics*. 33: p. 2583-2585 (DOI: 10.1093/bioinformatics/btx198).
 30. Kofler R., Schlötterer C., Luschützky E., Lelley T. (2008), Survey of microsatellite clustering in eight fully sequenced species sheds light on the origin of

- compound microsatellites. *BMC Genomics*. 9: p. 612 (DOI: 10.1186/1471-2164-9-612).
31. Jakupciak J.P., Wells R.D. (1999), Genetic instabilities in (CTG.CAG) repeats occur by recombination. *The Journal of biological chemistry*. 274: p. 23468-23479 (DOI: 10.1074/jbc.274.33.23468).
 32. Davis C.L., Field D., Metzgar D., Saiz R., Morin P.A., Smith I.L., Spector S.A., Wills C. (1999), Numerous Length Polymorphisms at Short Tandem Repeats in Human Cytomegalovirus. *Journal of Virology*. 73: p. 6265 (DOI: 10.1128/JVI.73.8.6265-6270.1999).
 33. Pawar R.M., Bhushan S.S., Poornachandar A., Lakshmikantan U., Shivaji S. (2011), Avian pox infection in different wild birds in India. *European Journal of Wildlife Research*. 57: p. 785-793 (DOI: 10.1007/s10344-010-0488-4).
 34. Sharma B., Nashiruddullah N., Bhat M.A., Taku A., Roychoudhury P., Ahmed J.A., Sood S., Mehmood S. (2019), Occurrence and phylogenetic analysis of avipoxvirus isolated from birds around Jammu. *Virus Disease*. p. 1-6. (DOI: 10.1007/s13337-018-00507-0).
 35. Bowcock A.M., Ruiz-Linares A., Tomfohrde J., Minch E., Kidd J.R., Cavalli-Sforza L.L. (1994), High resolution of human evolutionary trees with polymorphic microsatellites. *Nature*. 368: p. 455-457 (DOI: 10.1038/368455a0).
 36. Dib C., Fauré S., Fizames C., Samson D., Drouot N., Vignal A., Millasseau P., Marc S., Kazan J., Seboun E. (1996), A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature*. 380: p. 152-154 (DOI: 10.1038/380152a0).

37. Ellegren H. (2000), Heterogeneous mutation processes in human microsatellite DNA sequences. *J Nature genetics*. 24: p. 400-402 (DOI: 10.1038/74249).
38. Morgante M., Hanafey M., Powell W.J.N.g. (2002), Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. 30: p. 194-200 (DOI: 10.1038/ng822).
39. Varshney R.K., Graner A., Sorrells M.E. (2005), Genic microsatellite markers in plants: features and applications. *Trends in Biotechnology*. 23: 48-55 (DOI: 10.1016/j.tibtech.2004.11.005).
40. Field D., Wills C. (1998), Abundant microsatellite polymorphism in *Saccharomyces cerevisiae*, and the different distributions of microsatellites in eight prokaryotes and *S. cerevisiae*, result from strong mutation pressures and a variety of selective forces. *Proceedings of the National Academy of Sciences*. 95: 1647-1652 (DOI: 10.1073/pnas.95.4.1647).
41. Strand M., Prolla T.A., Liskay R.M., Petes T.D. (1993), Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature*. 365: p. 274-276 (DOI: 10.1038/365274a0).
42. Bayliss C.D., Dixon K.M., Moxon E.R. (2004). Simple sequence repeats (microsatellites): mutational mechanisms and contributions to bacterial pathogenesis. A meeting review. *FEMS Immunology Medical Microbiology*. 40: p. 11-19 (DOI: 10.1016/S0928-8244(03)00325-0).
43. Metzgar D., Thomas E., Davis C., Field D., Wills C.J.M.m. (2001), The microsatellites of *Escherichia*

- coli: rapidly evolving repetitive DNAs in a non-pathogenic prokaryote. 39: p. 183-190 (DOI: 10.1046/j.1365-2958.2001.02245.x).
44. Chen M., Tan Z., Jiang J., Li M., Chen H., Shen G., Yu R. (2009), Similar distribution of simple sequence repeats in diverse completed Human Immunodeficiency Virus Type 1 genomes. FEBS letters. 583: p. 2959-2963 (DOI: 10.1016/j.febslet.2009.08.004).
 45. Li D., Zhang H., Peng S., Pan S., Tan Z. (2019). Conserved microsatellites may contribute to stem-loop structures in 5', 3' terminals of Ebolavirus genomes. Biochemical biophysical research communications. 514: p. 726-733. (DOI: 10.1016/j.bbrc.2019.04.192).
 46. Chen M., Tan Z., Zeng G. (2011), Microsatellite is an important component of complete Hepatitis C virus genomes. Infection, Genetics Evolution. 11: p. 1646-1654 (DOI: 10.1016/j.meegid.2011.06.012).
 47. Singh A.K., Alam C.M., Sharfuddin C., Ali S. (2014a), Frequency and distribution of simple and compound microsatellites in forty-eight Human papillomavirus (HPV) genomes. Infection, Genetics and Evolution. 24: p. 92-98 (DOI: 10.1016/j.meegid.2014.03.010).
 48. George B., Gnanasekaran P., Jain S., Chakraborty S. (2014a), Genome wide survey and analysis of small repetitive sequences in caulimoviruses. Infection, Genetics and Evolution. 27, 15-24 (DOI: 10.1016/j.meegid.2014.06.018).
 49. Segarra A., Pépin J.F., Arzul I., Morga B., Faury N., Renault T. (2010). Detection and description of a particular Ostreid herpesvirus 1 genotype associated

- with massive mortality outbreaks of Pacific oysters, *Crassostrea gigas*, in France in 2008. *Virus research*. 153: p. 92-99 (DOI: 10.1016/j.virusres.2010.07.011).
50. Deback C., Luyt C., Lespinats S., Depienne C., Boutolleau D., Chastre J., Agut H. (2010), Microsatellite analysis of HSV-1 isolates: from oropharynx reactivation toward lung infection in patients undergoing mechanical ventilation. *Journal of clinical virology*. 47: p. 313-320 (DOI: 10.1016/j.jcv.2010.01.019).
51. Szpara M.L., Tafuri Y.R., Parsons L., Shamim S.R., Verstrepen K.J., Legendre M., Enquist L. (2011), A wide extent of inter-strain diversity in virulent and vaccine strains of alphaherpesviruses. *PLoS Pathogen*. 7: e1002282 (DOI: 10.1371/journal.ppat.1002282).
52. Avarre J.C., Madeira J.P., Santika A., Zainun Z., Baud M., Cabon J., Caruso D., Castric J., Bigarré L., Engelsma M. (2011), Investigation of Cyprinid herpesvirus-3 genetic diversity by a multi-locus variable number of tandem repeats analysis. *Journal of virological methods*. 173: p. 320-327 (DOI: 10.1016/j.jviromet.2011.03.002).
53. Ye C., Zhang Q.Z., Tian Z.J., Zheng H., Zhao K., Liu F., Guo J.C., Tong W., Jiang C.G., Wang S.J. (2015), Genomic characterization of emergent pseudorabies virus in China reveals marked sequence divergence: evidence for the existence of two major genotypes. *Virology*. 483: p. 32-43 (DOI: 10.1016/j.virol.2015.04.013).
54. Sahu B.P., Majee P., Singh R.R., Sahoo A., Nayak D. (2020b), Comparative analysis, distribution, and

- characterization of microsatellites in Orf virus genome. *Scientific Reports*. 10: p. 13852 (DOI: 10.1038/s41598-020-70634-6).
55. Ouyang Q., Zhao X., Feng H., Tian Y., Li D., Li M., Tan Z. (2012), High GC content of simple sequence repeats in Herpes simplex virus type 1 genome. *Gene*. 499: p. 37-40 (DOI: 10.1016/j.gene.2012.02.049).
56. Hancock J.M. (1993), Evolution of sequence repetition and gene duplications in the TATA-binding protein TBP (TFIID). *Nucleic acids research*. 21: p. 2823-2830 (DOI: 10.1093/nar/21.12.2823).
57. Karaoglu H., Lee C.M.Y., Meyer W. (2005), Survey of Simple Sequence Repeats in Completed Fungal Genomes. *Molecular Biology and Evolution*. 22: p. 639-649 (DOI: 10.1093/molbev/msi057).
58. Tóth G., Gáspári Z., Jurka J. (2000), Microsatellites in different eukaryotic genomes: survey and analysis. *Genome research*. 10: p. 967-981 (DOI: 10.1101/gr.10.7.967).
59. George B., Gnanasekaran P., Jain S.K., Chakraborty S. (2014b), Genome wide survey and analysis of small repetitive sequences in caulimoviruses. *Infection, Genetics and Evolution*. 27: p. 15-24 (DOI: 10.1016/j.meegid.2014.06.018).
60. George B., Mashhood Alam C., Jain S.K., Sharfuddin C., Chakraborty S. (2012), Differential distribution and occurrence of simple sequence repeats in diverse geminivirus genomes. *Virus Genes*. 45: p. 556-566 (DOI: 10.1007/s11262-012-0802-1).
61. Di Rienzo A., Peterson A.C., Garza J.C., Valdes A.M., Slatkin M., Freimer N.B. (1994), Mutational processes

- of simple-sequence repeat loci in human populations. Proceedings of the National Academy of Sciences of the United States of America. 91: p. 3166-3170 (DOI: 10.1073/pnas.91.8.3166).
62. Zhao X., Tian Y., Yang R., Feng H., Ouyang Q., Tian Y., Tan Z., Li M., Niu Y., Jiang J., Shen G., Yu R. (2012), Coevolution between simple sequence repeats (SSRs) and virus genome size. BMC genomics. 13: p. 435-435. (DOI: 10.1186/1471-2164-13-435).
63. Treco D., Arnheim N. (1986), The evolutionarily conserved repetitive sequence d(TG.AC)_n promotes reciprocal exchange and generates unusual recombinant tetrads during yeast meiosis. Molecular and cellular biology. 6: p. 3934-3947 (DOI: 10.1128/mcb.6.11.3934).
64. Wahls W.P., Wallace L.J., Moore P.D. (1990), The Z-DNA motif d(TG)₃₀ promotes reception of information during gene conversion events while stimulating homologous recombination in human cells in culture. Molecular and cellular biology. 10: p. 785-793 (DOI: 10.1128/mcb.10.2.785).
65. Katti M.V., Ranjekar P.K., Gupta V.S. (2001), Differential distribution of simple sequence repeats in eukaryotic genome sequences. J.M.b., evolution. 18: p. 1161-1167. (DOI: 10.1093/oxfordjournals.molbev.a003903).
66. Li Y.C., Korol A.B., Fahima T., Nevo E. (2004), Microsatellites Within Genes: Structure, Function, and Evolution. Molecular Biology and Evolution. 21: p. 991-1007 (DOI: 10.1093/molbev/msh073).

67. Usdin K. (2008), The biological effects of simple tandem repeats: lessons from the repeat expansion diseases. *Genome research*. 18: p. 1011-1019 (DOI: 10.1101/gr.070409.107).
68. Kashi Y., King D.G. (2006), Simple sequence repeats as advantageous mutators in evolution. *Trends in Genetics*. 22: p. 253-259 (DOI: 10.1016/j.tig.2006.03.005).
69. George B., Alam C.M., Kumar R.V., Gnanasekaran P., Chakraborty S.J.V. (2015), Potential linkage between compound microsatellites and recombination in geminiviruses: Evidence from comparative analysis. *Journal of Virology*. 89: p. 41-50 (DOI: 10.1128/JVI.01451-15).
70. Mudunuri S.B., Nagarajaram H.A. (2007), IMEx: Imperfect Microsatellite Extractor. *Bioinformatics*. 23: p. 1181-1187 (DOI: 10.1093/bioinformatics/btm097).
71. Sanger F., Nicklen S., Coulson A.R. (1977), DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*. 74: p. 5463 (DOI: 10.1073/pnas.74.12.5463).
72. Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. (1990), Basic local alignment search tool. *Journal of Molecular Biology*. 215: p. 403-410. (DOI: 10.1016/S0022-2836(05)80360-2).

Chapter 8

Complete genome analysis of Indian fowlpox virus

8.1 Introduction

Fowlpox virus (FPV) is a species that belongs to the Avipox virus genus (APV), infects chickens, and has a worldwide distribution, including India [1, 2]. It has a significant economic impact worldwide, with a drop in egg production and stunted growth rates in chickens. The disease is manifested by a high rate of mortality and morbidity [3]. The severity of this disease usually depends upon the susceptibility of the host, degree of virulence, and positions of the infection [4]. Two forms of infection are associated depending on the routes of infection. The cutaneous form is most common and visualized on beak or featherless area and form scab for secondary bacterial infection. The second type of infection is severe diphtheric, form involves droplet infection of the mucous membranes of the mouth and infects the respiratory system, followed by cause death by asphyxiation [5]. The disease spread by various means like viral vectors such as arthropods like mosquitoes, mites, flies, through the air from diseased birds and intake of contaminated food or water [6]. Although vaccination with live-attenuated viruses (FPV) is used to control this disease [7], the lack of efficacy and its immune response suggest improving proper vaccine candidates [8]. The virus accomplished abortive replication in mammalian cells and was therefore used as host range-restricted mammalian expression vectors [9].

The general structure of APVs is oval-shaped, covered with a lipid membrane, centrally located conserves regions, and variable terminal inverted repeats [10]. The genome contains

double-stranded DNA, ranged from 240-365 kb and low G + C content (30 to 40 %). The virus usually replicates within the cytoplasm of the infected cells and contains 259-359 genes. Depending upon the expression of the putative promoter of genes, those were classified as E, early; I, intermediate; L, late genes and control the virus biology in a synchronized manner.

This virus can persist in the hot, humid climate and withstand for a long period in the environment resulted in reinfection in the similar farms [11]. A number of reports suggest the presence of this virus throughout India and have a great impact on socio- economic status of the farmers [12, 1]. The lack of enough surveillance, as well as vaccine failure in India, indicates the absence of enough genomic resources of those corroborated the re-emergence of the viral species. Thus in the present study, for the first time, we have deciphered the complete genome sequence of a fowlpox virus of Indian origin through next generation sequencing platform and utilize a comparative genomics approach to decipher more about this virus biology.

8.2 Results

8.2.1 NGS output and general features of Indian FPV

The NextSeq 500 NGS generated a total of ~2.16 GB of quality data with 7,303,963 numbers of quality reads. The previously published complete genome isolate fowlpox virus (Acc. No- AF198100) was used as the reference genome. The sequence generated through NGS mapped with the reference and assembled. The newly assembled viral genome exhibited 260,066 bp in length. A lower percentage of GC content (28.5) was observed in this newly assembled genome, in comparison to other APV, such as Shearwater pox virus

(30.23), Canarypox virus (30.37), Turkeypox virus (29.78). The FPV viral genome encodes 255 putative genes. As seen in other poxviruses, the FPV genome contains a central coding region bounded by two identical inverted terminal repeat (ITR) regions of approximately 9350bp each end.

8.2.2 Phylogenetic and recombination analysis

The present FPV strain showed a close relationship with other FPV around the world (Acc. No-AF198100 and MF766431) . We observed a total of 10 potential recombination events with significant P-values detected across the APV genomes (Table 8.2). A number of recombination events were overlapped within intergenic (events 5, 7, 8, 9, and 10), hypothetical protein (events 1, 2, and 3), transcription factor VLTF (events 6), thymidylate kinase (events 4). Out of 10 potential recombination events, 3 events include FPV/Ind act as recombinant and major parental sequences.

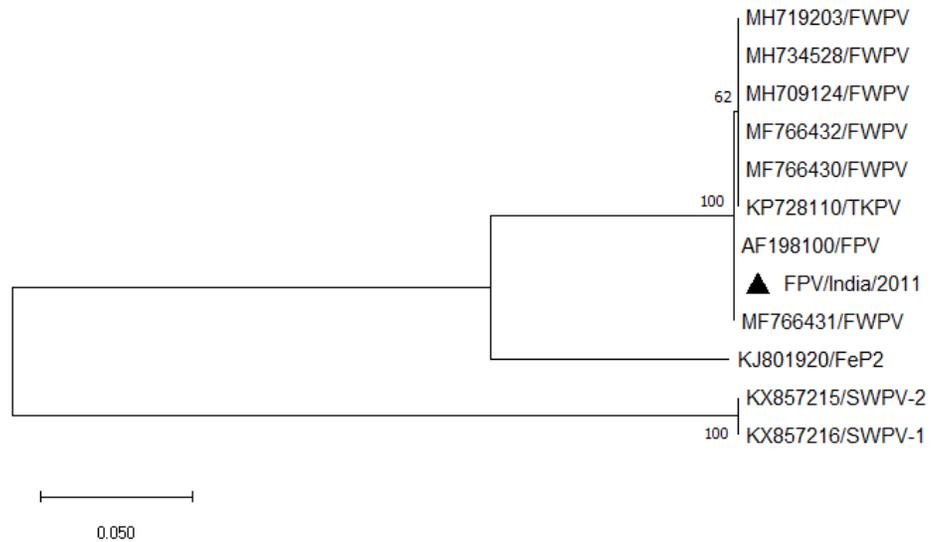


Figure 8.1: Phylogenetic relationship between Indian FPV and other APVs. The General Time Reversible (GTR) model of MEGA6 was used to create a phylogenetic tree based on the Maximum likelihood (ML) phylogeny with 500 bootstraps. The abbreviations and GenBank accession details for seven core protein sequences of the selected AVPs have used pigeonpox virus (FeP2; KJ801920), fowlpox virus (FWPV; AF198100, MF766430-32, MH709124, MH719203, and MH734528), shearwaterpox virus 1 (SWPV-1; KX857216), shearwaterpox virus 2 (SWPV-2; KX857215).

Recombination Event	Begin	End	Recombinant gene in PPV	Recombinant Sequence(s)	Minor Parental Sequence(s)	Major Parental Sequence(s)	Detection Methods
1	159807	163207	hypothetical protein	KX857215 (SWPV-2)	MF766430	KJ801920 (FeP2)	RGB MCS
2	151374	162969	hypothetical protein	KP728110	MF766430	KJ801920 (FeP2)	RGB MCS
3	157308	175063	hypothetical protein	MF766430	KX857215 (SWPV-2)	FPV/Ind	RGB MC
4	152172	173452	thymidylate kinase	MF766430	(MF766430)	KP728110	RGB MCS
5	160331	179564	Intergenic Region	KP728110	KX857215 (SWPV-2)	MF766430	RGB MCS
6	170353	185060	transcription factor VLTF-1	MF766430	KX857215 (SWPV-2)	KP728110	RBM CS
7	151329	181593	Intergenic Region	FPV/Ind	KX857215 (SWPV-2)	MF766430	RGB MS
8	180979	194098	Intergenic Region	KJ801920 (FeP2)	MF766430	KP728110	RGS
9	159642	165067	Intergenic Region	MF766430	KX857215 (SWPV-2)	FPV/Ind	RGB MCS

10	220 608	251 225	Intergen ic Region	KP7281 10	KX857 215 (SWPV -2)	MF76643 0	RGB MCS
----	------------	------------	--------------------------	--------------	------------------------------	--------------	------------

Table 8.1: Predicted potential recombination events between FPV/Ind. Details recombination events detected between FPV isolates were analyzed in this study using the RDP4 program. Detection method coding R, G, B, M, C, S, P, L, and T represents methods RDP, GENECONV, Bootscan, MaxChi, Chimaera, SiScan, PhylPro, LARD, and 3Seq, respectively. The P-value for the detection method is shown in bold.

8.3 Discussion

Despite a number of outbreaks reported from India, no complete genome and associated genomic information are available for fowlpox virus [1, 13]. In this study, we have elucidated the first complete genome of FPV directly from clinical samples using the NGS platform. Most of the features of this complete genome similar to the previously published FPV genome with some difference. Usually, all pox viruses retain a central coding region guarded by two identical inverted terminal repeat (ITR) regions. The present isolate was also bear the ITRs of 9350 bp and arranged in sense and antisense orientation.

The concatenated phylogenetic tree analysis demonstrated that the strain is closely related with the FPV strain of the USA (Figure 8.1). We speculate that trade exchange and transport can be a potential route to establish the epidemiological linkage between the strains isolated from the two distinct geographical areas. Although we have no evidence of the reservoir host of this FPV/Ind, mosquitoes are believed to play

a part in mechanical transmission within this wild bird population [14].

The comparative genomics of FPV with other APV complete genomes provides evidence of recombination among the virus. Although several attempts were made to identify recombination events, the lack of complete genome sequences made this study exercise unsuccessful. However, in this study, we identified ten potential recombination events where FPV/Ind actively participated in 30% of events by forming recombinant as well as major parents. Thus, the FPV/Ind strain has the potential to evolve via homologous or non-homologous site-specific recombination and can act as a major or minor parent to form new variants.

8.4 Materials and methods

8.4.1 Nucleic acid extraction

To decipher the complete genome sequence of Indian FPV, we have utilized the isolates FPV/Ganjam/Ind previously identified from our lab. The extraction of nucleic acid from infected scab samples was performed by using DNeasy blood and tissue purification Kits (QIAGEN, USA) with some modification established by Sarker et al., 2017. The quality and quantity of DNA checked with 1% agarose gel followed by nanodrop estimation. Finally, 250 ng of genomic DNA send for library construction for the respective next-generation sequencing platform.

8.4.2 Library construction and Illumina NextSeq500 sequencing

The genomic DNA revived earlier was QC was checked, followed by paired-end sequencing library using TruSeq Nano DNA Library Prep Kit. The fragmentation of that DNA occurred through the Covaris M220 shearing method to obtain the fragment of 350bp with 3' or 5' overhangs followed by end repair by using the End Repair Mix provided within the kit. Then adaptors were ligated to the fragments having a low rate of concatenated formation. The size selection of those ligated products conducted through AMPure XP beads (Invitrogen, USA). The selected ligation product further PCR amplified to enrich the amplicons by using the program 72° C initial denaturation for 3 min; 95 °C denaturation for 30 s; 12 cycles of 95 °C for 10 s, 55 °C for 30 s, and 72 °C for 30 s; and 72 °C final extension step for 5 min recommended by manufactures' protocol. The enriched genomic enriched library prepared through PCR amplification was further validated by the 4200 Tape Station system (Agilent Technologies). Cluster formation, followed by sequencing of the enriched library, was obtained through paired-end sequencing chemistry on the NextSeq500 sequencing platform.

8.4.3 Raw sequence data processing, mapping, assembly, and genome annotations

The raw data generated through the NGS platform was further cleaned by Trimmomatic v0.38 to eliminate noisy or ambiguous reads, adapter, and host sequence [15]. The quality reads further mapped to reference sequence utilizing BWA MEM software (version 0.7.17) [16]. The extraction of consensus reads Along with Denovo assembly was performed using SAM tools [17]. GATU was used to identify possible

Open reading frames (ORFs) as well as genome annotation [18]. To decipher the intergenic region along with the presence of ORFs within them BWA MEM analysis tool (version 0.7.17) was utilized. The presences of possible genes or ORFs were identified depending upon the significant sequence similarity with known cellular or viral proteins. However, the annotation was completed through the comparative genomics approach with the published reference genome to organize proper start and stop codons, overlap regions, site of truncation.

8.4.4 Phylogenetic and recombination analysis

To infer evolutionary relationship of present virus with other APVs, a phylogenetic tree was constructed. The representative complete genome of APVs was downloaded from GenBank. MAFFT (version 7) was used to align these complete genomes, followed by manually editing by BioEdit (version 7.2) tool [19]. The General Time Reversible (GTR) model of MEGA6 was used to create a phylogenetic tree based on for the Maximum likelihood (ML) phylogeny with 500 bootstraps. To get an insight into the role of recombination the evolution of FPV within the species of APVs, we verified all possible major parents, minor parents recombinant, and the region associated with the hotspot of recombination by using the RDP, GENECONV, Bootscan, MaxChi, Chimaera, Siscan, PhylPro, LARD and 3Seq methods of RDP4 program [20]. Parameters had a significant p-value within three of the methods referred to as a possible recombinant event.

8.5 Conclusion

In conclusion, we report the first complete genome of FPV isolate from India. Utilizing a comparative genomic approach, we deciphered the presence of recombination within Avipox viruses, and the presence isolate taking an active role during recombination events. The current genomic resource would be greatly useful for further understanding of FPV biology, epidemiology, and research carried in the front of diagnosis and vaccine development.

8.6 References

1. Sahu, B.P., Majee, P., Mishra, C., Dash, M., Biswal, S., Sahoo, N. and Nayak, D. (2020), The emergence of subclades A1 and A3 avipoxviruses in India. *Transboundary and Emerging Diseases*. 67(2): p. 510-517 (DOI: 10.1111/tbed.13413).
2. Gyuranecz, M., Foster, J.T., Dán, Á., Ip, H.S., Egstad, K.F., Parker, P.G., Higashiguchi, J.M., Skinner, M.A., Höfle, U., Kreizinger, Z. and Dorrestein, G.M. (2013), Worldwide phylogenetic relationship of avian poxviruses. *Journal of virology*. 87(9): p. 4938-4951 (DOI: 10.1128/JVI.03183-12).
3. Saif, Y.M., Barnes, H.J., Gilsson, J.R., Fadly, A.M., Mc Dongald, L.R. and Swayne, D.E. (2007), *Diseases of Poultry*. Iowa state university press.
4. Tripathy, D.N. and Reed, W.M. (2003), *Pox in: Diseases of Poultry*
5. Esteves, F.C., Marín, S.Y., Resende, M., Silva, A.S., Coelho, H.L., Barbosa, M.B., D'Aparecida, N.S., de Resende, J.S., Torres, A.C. and Martins, N.R., 2017.

- Avian pox in native captive Psittacines, Brazil (2015), *Emerging infectious diseases*. 23(1): p.154 (DOI: 10.3201/eid2301.161133).
6. Medina, F.M., Ramírez, G.A. and Hernández, A. (2004), Avian pox in white-tailed laurel-pigeons from the Canary Islands. *Journal of wildlife diseases*. 40(2): p. 351-355 (DOI: 10.7589/0090-3558-40.2.351).
 7. Hitchner, S.B. (1981) Canary pox vaccination with live embryo-attenuated virus. *Avian diseases*. p. 874-881 (DOI: 10.2307/1590062).
 8. Back, A., Soncini, R.A., Ruthes, O., Madureira Jr, S. and Flores, R. (1995), An atypical fowl pox outbreak in broilers in southern Brazil. *Avian diseases*. p. 902-906 (DOI: 10.2307/1592431).
 9. Taylor, J., Weinberg, R., Tartaglia, J., Richardson, C., Alkhatib, G., Briedis, D., Appel, M., Norton, E. and Paoletti, E. (1992), Nonreplicating viral vectors as potential vaccines: recombinant canarypox virus expressing measles virus fusion (F) and hemagglutinin (HA) glycoproteins. *Virology*. 187(1): p. 321-328 (DOI: 10.1016/0042-6822(92)90321-f).
 10. Afonso, C.L., Tulman, E.R., Lu, Z., Zsak, L., Kutish, G.F. and Rock, D.L., 2000. The genome of fowlpox virus. *Journal of virology*. 74(8): p. 3815-3831 (DOI: 10.1128/JVI.74.8.3815-3831.2000).
 11. Alehegn, E. and Mersha, C. (2012), A systematic review of serological and clinicopathological features and associated risk factors of avian pox (DOI: 10.5829/idosi.bjps.2014.3.3.8553).
 12. Sharma, B., Nashiruddullah, N., Bhat, M.A., Taku, A., Roychoudhury, P., Ahmed, J.A., Sood, S. and

- Mehmood, S. (2019), Occurrence and phylogenetic analysis of avipoxvirus isolated from birds around Jammu. *VirusDisease*. 30(2): p. 288-293 (DOI: 10.1007/s13337-018-00507-0).
13. Audarya, S.D., Riyesh, T., KumarN, C.D., Sikrodia, R., Sharda, R., Barua, S. and Garg, U.K. (2018), Molecular Diagnosis of a Cutaneous Form of Pox in Pigeons at Mhow in Madhya Pradesh, India. *International journal of current microbiology and applied sciences*. 72319: P. 7706 (DOI: 10.20546/ijcmas.2018.709.157).
 14. Annuar, B.O., Mackenzie, J.S. and Lalor, P.A. (1983), Isolation and characterization of avipoxviruses from wild birds in Western Australia. *Archives of virology*. 76(3): p. 217-229.
 15. Bolger A. M., LohseM., Usadel B. J. B. (2014), Trimmomatic: a flexible trimmer for Illumina sequence data. 30: p. 2114-2120 (DOI: 10.1093/bioinformatics/btu170)
 16. Li H., Durbin R. J. b. (2009), Fast and accurate short read alignment with Burrows–Wheeler transform. 25: p. 1754-1760 (DOI: 10.1093/bioinformatics/btp324).
 17. Li H. et al.(2009), The sequence alignment/map format and SAM tools. 25, p. 2078-2079 53 (DOI: 10.1093/bioinformatics/btp352).
 18. Tcherepanov V., Ehlers A., Upton C. J. B. g. (2006), Genome Annotation Transfer Utility (GATU): rapid annotation of viral genomes using a closely related reference genome. 7: p. 150 (DOI: 10.1186/1471-2164-7-150).

19. Hall T., Biosciences I., Carlsbad C. J. G. B. B. (2011), BioEdit: an important software for molecular biology. 2: p. 60-61
20. Martin D. P., Murrell B., Golden M., Khoosal A., Muhire B. J. V. e. (2015), RDP4: Detection and analysis of recombination patterns in virus genomes. (DOI: 10.1093/ve/vev003).

Chapter 9

Complete genome analysis of Indian pigeonpox virus

9.1 Introduction

The emergence of the pox virus has a serious health hazard to both domestic as well as free-ranging wild birds regardless of their age and sex due to its exorbitant contagious property, which influences both the economic and conservation aspects [1]. Recently, the International Union for Conservation of Nature (IUCN) reported the existence of more than 11,000 avian species, with 773 endangered and critically endangered species in 2016 [2]. Pox-like infections have been reported among wild and domestic bird species globally derived from 20 orders, which includes 76 families having 329 avian species across the globe, caused by a group of viruses named as family Avipox viruses (APVs) [3, 4]. Although antigenically and immunologically, these viruses were distinct from one another, the cross-relation within the broad host range arose the complications for viral strain demarcation distinguishable from each other complication arose to mitigate the viral strain due to broad host range. However, the virus nomenclature is usually determined according to the host from which the virus is derived. A few of the examples included the fowlpox virus (FWPV), turkeypox virus (TKPV), pigeonpox virus (PGPV), canarypox virus (CNPV), etc.

APVs, present throughout the globe, is the causative agent of highly transmissible avian pox disease, which spread by various means like viral vectors such as arthropods like mosquitoes, mites, flies, through the air from diseased birds and intake of contaminated food or water [5, 6]. The

incubation time generally differed from 4-10 days according to the host. The severity of this disease usually depends upon the susceptibility of the host, degree of virulence, and positions of the infection [7]. As this virus can resist the hot, humid climate and withstand for a long period in the environment, resulting in repetition of infection in similar farms frequently due to poor management practices [8]. The severity of the infection generally depends upon two forms of the disease. The first one is mild cutaneous lesions (dry form), rarely fatal in comparison to the second form, such as diphtheric (wet) form. The secondary bacterial and fungal infection increases the rate of morbidity and mortality (up to 60%) by forming the painful lesions around the eye, beak respiratory track that interfere with eating, respiration, and vision [9, 10]. This disease has a great impact on the socio-economic of small backyard farmers due to economic losses as the infection influence the drop in egg production, stunted growth, impaired fertility, and peak in mortality rate [9]. Within threatened or endangered species, poxvirus infection leads to a decrease their ability to care for the young ones and increases the chance of predation [10].

APVs belongs to DNA viruses having enveloped lipid membrane, oval-shaped with the particle size 270×350 nm and centrally located conserved regions and variable terminal inverted repeats [11]. The typical genome organization consists of double-stranded DNA, which is the largest among all poxviruses having genome size from 260-365 kb and low G + C content (30 to 40 %), that replicates within the cytoplasm of the infected cells and contains 259-359 genes, which is higher in comparison to other orthopoxviruses such as vaccinia, cowpox, and monkeypox. Depending upon the

expression of the putative promoter of genes, those were classified as E, early; I, intermediate; L, late genes and control the virus biology in a synchronized manner. Due to its large genome size with the stable central region and host range-restricted to avian species increase the efficacy to produce recombinant vaccines for foot and mouth disease virus, Influenza and HIV.

Although a number of reports suggest the emergence of Avipoxvirus worldwide [13, 14], including India causing infection in fowl, turkey, pigeon, duck, and quail with high morbidity and mortality, still a long way to move to decipher this complex range of viruses [15, 16, 17]. The lack of enough surveillance, as well as vaccine failure in India, indicates the absence of enough genomic resources of those corroborated the re-emergence of the viral species. To do so, we have taken the pigeonpox virus into our consideration as it is an infectious agent to Indian pigeons, which are utilized as a food source, sport (racing), show animals, and household pets. For the first time, we have deciphered the complete genome sequence of a pigeonpox virus of Indian origin through a next-generation sequencing platform and utilize a comparative genomics approach to decipher more about this virus biology.

9.2 Results

9.2.1 NGS output and general features of Indian PPV

The NextSeq 500 NGS generated a total of ~2.72 GB of quality data with 9,541,974 numbers of quality reads. The previously published complete genome isolate Pigeonpox virus FeP2 (Acc. No-NC_024447) was used as the reference genome, to which the sequence generated through NGS

mapped and assembled. The assembled genome of this novel virus exhibited 280058 bp in length. A lower percentage of GC content (29.5) was observed in this newly assembled genome, in comparison to fowlpox virus (30.83), shearwater pox virus (30.23), canarypox virus (30.37), turkeypox virus (29.78) but similar to other APV such as pigeonpox virus, Flamingo pox virus, magpiepox virus and penguinpox virus. The extreme left nucleotide was considered as base 1, and the beginning region of the Inverted Terminal Repeats (ITRs), which consists of a total length 4688 bp. The ITR spanned throughout PPV-001 to PPV-004a and PPV_256 to PPV_259. Using NCBI's The ORF Finder tool of NCBI revealed the presence of 270 ORFs, from which 11 have been annotated as truncated and one fragmented gene, respectively. The naming of the genes and their product were compiled based on the report described by Offerman et al., 2014 [19].

9.2.2 Phylogenetic and recombination analysis

The present PPV strain showed a close relationship with Pigeonpox virus FeP2 (Acc. No-NC_024447) (Figure 9.1). We observed a total of 21 potential recombination events with significant P-values detected across the ORFV genomes (Table 2). A number of recombination events were overlapped within intergenic (events 5, 7, 8, 9, 13, 16, and 20), hypothetical protein (events 1, 2, 3 17, and 18), transcription factor VLTF (events 6 and 12), ankyrin repeat protein (events 15 and 21). Out of 21 potential recombination events, 15 events include PPV/Pur as recombinant, minor, and major parental sequences. The recombinant PPV/Pur (event 7, 8) were observed within the intergenic region, using KX857215 (SWPV-2), (MK903864) (crusty tissue/MGP) as a minor

parent and MF766431 (16069_trachea_170323), KJ859677 (PSan92) as a major parent.

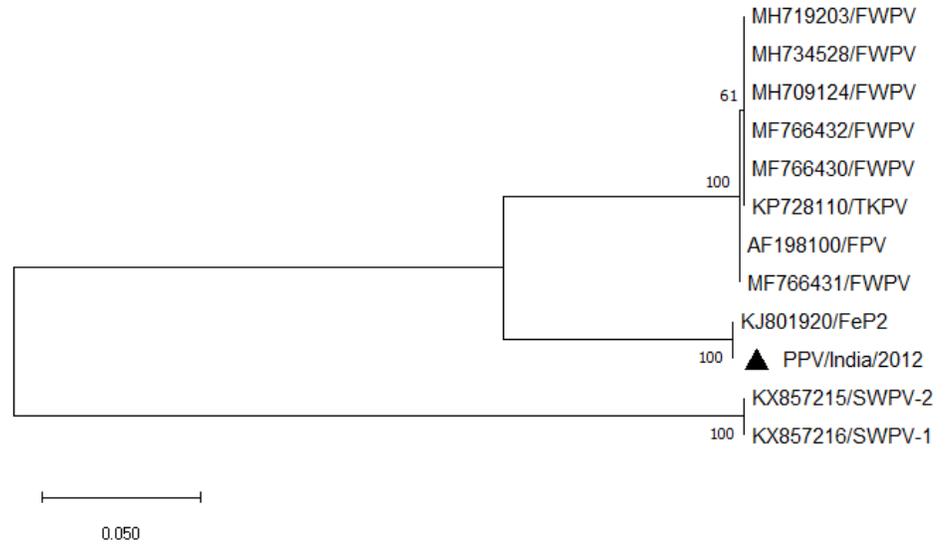


Figure 9.1: Phylogenetic relationship between Indian PPV and other AVPs. The General Time Reversible (GTR) model of MEGA6 was used to create a phylogenetic tree based on the Maximum likelihood (ML) phylogeny with 500 bootstraps. The abbreviations and GenBank accession details for seven core protein sequences of the selected AVPs have used pigeonpox virus (FeP2; KJ801920), fowlpox virus (FWPV; AF198100, MF766430-32, MH709124, MH719203, and MH734528), shearwaterpox virus 1 (SWPV-1; KX857216), shearwaterpox virus 2 (SWPV-2; KX857215).

Recombination Event	Begin	End	Recombinant gene in PPV	Recombinant Sequence(s)	Minor Parental Sequence(s)	Major Parental Sequence(s)	Detection Methods
1	147807	149207	hypothetical protein	KX857215 (SWPV-2)	MF766431 (16069_trachea_170323)	NC_024447 (FeP2)	RGBMCS
2	147774	149069	hypothetical protein	KJ859677 (PSan92)	MF766431 (16069_trachea_170323)	NC_024447 (FeP2)	RGBMCS
3	146908	147763	hypothetical protein	MF766431 (16069_trachea_170323)	KX857215 (SWPV-2)	lcl PPV/Pur	RGBMC
4	149972	150352	thymidylate kinase	MF766431 (16069_trachea_170323)	(MK903864) (crusty tissue/MGP)	KJ859677 (PSan92)	RGBMCS
5	149731	149864	Intergenic Region	KJ859677 (PSan92)	KX857215 (SWPV-2)	MF766431 (16069_trachea_170323)	RGBMCS
6	150353*	150660	transcription factor VLTF-1	MF766431 (16069_trachea_170323)	KX857215 (SWPV-2)	KJ859677 (PSan92)	RBMS
7	149229*	149493	Intergenic Region	lcl PPV/Pur	KX857215 (SWPV-2)	MF766431 (16069_trachea_170323)	RGBMS
8	149079*	149198*	Intergenic Region	NC_024447 (FeP2)	(MK903864) (crusty tissue/MGP)	KJ859677 (PSan92)	RGS
9	149642	149867	Intergenic Region	MF766431 (16069_trachea_170323)	KX857215 (SWPV-2)	lcl PPV/Pur	RGBMCS
10	216208	217925	A-type inclusion protein	NC_036582 (FGPVKD09)	Unknown (NC_028238) (TKPV-HU1124/2011)	MG760432 (FMV15SnBuPox)	RGBMCS
11	116559	116689	B22R family protein	KX857215 (SWPV-2)	MF766431 (16069_trachea_170323)	(MK903864) (crusty tissue/MGP)	RBMS
12	190192	190634	transcription factor VLTF-3	MF766431 (16069_trachea_170323)	(MK903864) (crusty tissue/MGP)	KJ859677 (PSan92)	RBMS
14	231177	231553	epidermal growth factor-like protein	MF766431 (16069_trachea_170323)	KJ859677 (PSan92)	Unknown (lcl PPV/Pur)	RBMC
15	11458	11573	ankyrin repeat protein	MF766431 (16069_trachea_170323)	(MK903864) (crusty tissue/MGP)	KJ859677 (PSan92)	RGBS
16	146746	146796	Intergenic Region	KJ859677 (PSan92)	KX857215 (SWPV-2)	MF766431 (16069_trachea_170323)	RBS
17	144885	145534	hypothetical protein	KJ859677 (PSan92)	Unknown (MF766431)	NC_024447 (FeP2)	RBM
18	4833	4895	hypothetical protein	MF766431 (16069_trachea_170323)	KX857215 (SWPV-2)	lcl PPV/Pur	RBS
19	100571	100962	DNA polymerase	MF766431 (16069_trachea_170323)	(MK903864) (crusty tissue/MGP)	NC_024447 (FeP2)	RBMS
20	149442	149469	Intergenic Region	KJ859677 (PSan92)	KX857215 (SWPV-2)	MF766431 (16069_trachea_170323)	RBS
21	270859	271942	ankyrin repeat protein	NC_036582 (FGPVKD09)	Unknown (NC_028238) (TKPV-HU1124/2011)	MG760432 (FMV15SnBuPox)	RMCS

Table 9.1: Predicted potential recombination events between PPV/Ind. Details recombination events detected between PPV isolates were analyzed in this study using the RDP4 program. Detection method coding *R, G, B, M, C, S, P, L*, and *T* represents methods *RDP, GENECONV, Bootscan, MaxChi, Chimaera, SiScan, PhylPro, LARD, and 3Seq*, respectively. The *P*-value for the detection method is shown in bold.

9.3 Discussion

Despite a number of outbreaks reported from India, no complete genome and associated genomic information are available [15, 16, 17, 18]. In this study, we have elucidated the first complete genome of PPV directly from clinical samples using the NGS platform. Most of the features of this complete genome similar to the previously published PPV genome, with some minor differences like the absence of one gene and shorted ITR [19]. Usually, all poxviruses retain a central coding region guarded by two identical inverted terminal repeat (ITR) regions. The present isolate was also bear the ITRs of 3910bp and arranged in sense and antisense orientation.

The phylogenetic tree indicates the transboundary potential of the present virus isolates to spread to a geographically distinct country. We speculate that trade exchange and transport can be a potential route to establish the epidemiological linkage between the strains isolated from the two distinct geographical areas.

The comparative genomics of PPV with other APV complete genomes provides evidence of recombination among the virus. Although several attempts were made to identify recombination events, the lack of complete genome sequences made this study exercise unsuccessful. However, in this study, we identified 21 potential recombination events where PPV/Ind actively participated in more than 40% of events by forming recombinant as well as major and minor parents. Thus, the strain has the potential to evolve via homologous or non-homologous site-specific recombination and can act as a major or minor parent to form new variants.

9.4 Materials and methods

9.4.1 Nucleic acid extraction

To decipher the complete genome sequence of Indian PPV, we have utilized the isolates PPV/Pur-Od-4b/01/Ind previously identified from our lab. The extraction of nucleic acid from infected scab samples was performed by using DNeasy blood and tissue purification kits (QIAGEN, USA) with some modifications established by Sarker et al., 2017 [20]. The quality and quantity of DNA checked with 1% agarose gel followed by nanodrop estimation. Finally, 250ng of genomic DNA send for library construction for respective next-generation sequencing platform.

9.4.2 Library construction and Illumina NextSeq500 sequencing

The genomic DNA revived earlier was QC checked, followed by paired-end sequencing library using TruSeq Nano DNA Library Prep kit. The fragmentation of that DNA occurred through Covaris M220 shearing method to obtain the fragment of 350bp with 3' or 5' overhangs followed by end repair by using End Repair Mix provided within the Kit. Then adaptors were ligated to the fragments having a low rate of concatenated formation. The size selection of those ligated products conducted through AMPure XP beads (Invitrogen, USA). The selected ligation product further PCR amplified to enrich the amplicons by using the program 72 °C initial denaturation for 3 min; 95 °C denaturation for 30 s; 12 cycles of 95 °C for 10 s, 55 °C for 30 s, and 72 °C for 30 s; and 72 °C final extension step for 5 min recommended by manufactures' protocol. The enriched genomic enriched library prepared through PCR amplification was further

validated by the 4200 Tape Station system (Agilent Technologies). Cluster formation, followed by sequencing of the enrich library was obtained through paired-end sequencing chemistry on NextSeq500 sequencing platform.

9.4.2 Raw sequence data processing, mapping, assembly, and genome annotations

The raw data generated through the NGS platform was further cleaned by Trimmomatic v0.38 to eliminate noisy or ambiguous reads, adapter and host sequence [21]. The quality reads further mapped to reference sequence utilizing BWA MEM software (version 0.7.17) [22]. The extraction of consensus reads along with Denovo assembly was performed using SAMtools [23]. GATU was used to identify possible Open reading frames (ORFs) as well as genome annotation [24]. To decipher the intergenic region along with the presence of ORFs within them BWA MEM analysis tool (version 0.7.17) was utilized. The presences of possible genes or ORFs were identified depending upon the significant sequence similarity with known cellular or viral proteins. However, the annotation was completed through the comparative genomics approach with the published reference genome to organize proper start and stop codons, overlap regions, site of truncation.

9.4.2 Phylogenetic and recombination analysis

To infer evolutionary relationship of present virus with other APVs, a phylogenetic tree was constructed. The representative complete genome of APVs was downloaded from GenBank. MAFFT (version 7) was used to align these complete genomes, followed by manually editing by BioEdit (version 7.2) tool [25]. The General Time Reversible (GTR) model of MEGA6 was used to create a phylogenetic tree based on the

Maximum likelihood (ML) phylogeny with 500 bootstraps [26]. To get an insight into the role of recombination for the evolution of PPV within the species of APVs, we verified all possible major parents, minor parents recombinant, and the region associated with the hotspot of recombination by using the RDP, GENECONV, Bootscan, MaxChi, Chimaera, Siscan, PhylPro, LARD, and 3Seq methods of RDP4 program [27]. Parameters with a significant p-value within three of the methods were referred as a possible recombinant event.

9.5 Conclusion

In conclusion, we report the first draft genome of PPV isolate from India. Utilizing, a comparative genomic approach, we deciphered the presence of recombination within Avipox viruses, and the presence isolate taking an active role during recombination events. The current genomic resource would be greatly useful for further understanding of PPV biology, epidemiology, and research carried in the front of diagnosis and vaccine development.

9.6 References

1. Murphy F.A., Gibbs E.P.J., Horzinek M.C., Studdert M.J. (1999), *Veterinary virology*. Elsevier
2. Bodewes R. (2018), Novel viruses in birds: Flying through the roof or is a cage needed?. *The Veterinary Journal*. 233: p. 55-62 (DOI: 10.1016/j.tvjl.2017.12.023).
3. Bolte A.L., Meurer J., Kaleta E.F. (1999), Avian host spectrum of avipoxviruses. *Avian Pathology*. 28(5): p. 415-432 (DOI: 10.1080/03079459994434).

4. Carulei O., Douglass N., Williamson A. L. (2017), Comparative analysis of avian poxvirus genomes, including a novel poxvirus from lesser flamingos (*Phoenicopterus minor*), highlights the lack of conservation of the central region. *BMC genomics*. 18(1): p. 947 (DOI: 10.1186/s12864-017-4315-0).
5. Medina F.M., Ramírez G.A., Hernández A. (2004), Avian pox in white-tailed laurel-pigeons from the Canary Islands. *Journal of wildlife diseases*. 40(2): p. 351-355 (DOI: 10.7589/0090-3558-40.2.351).
6. Adebajo M.C., Ademola S.I., Oluwaseun A. (2012), Seroprevalence of fowl pox antibody in indigenous chickens in Jos North and South Council Areas of Plateau State, Nigeria: Implication for vector vaccine. *ISRN veterinary science*. (DOI: 10.5402/2012/154971. Print 2012).
7. Tripathy D.N., Reed W.M., (2003), Pox in: *Diseases of Poultry*
8. Alehegn E., Mersha C. (2012), A systematic review of serological and clinicopathological features and associated risk factors of avian pox (DOI: 10.5829/idosi.bjps.2014.3.3.8553).
9. Tripathy D.N. (1993), Avipox viruses. *Virus infections of birds*. p.1-15
10. Skinner M.A. (2008), Poxviridae. In *Poultry Diseases* (pp. 333-339). WB Saunders
11. Afonso, C.L., Tulman, E.R., Lu, Z., Zsak, L., Kutish, G.F. and Rock, D.L. (2000), The genome of fowlpox virus. *Journal of virology*. 74(8): p. 3815-3831 (DOI: 10.1128/JVI.74.8.3815-3831.2000).

12. Manarolla G., Pisoni G., Sironi G., Rampin T. (2010), Molecular biological characterization of avian poxvirus strains isolated from different avian species. *Veterinary Microbiology*, 140(1–2): p. 1–8 (DOI: 10.1016/j.vetmic.2009.07.004).
13. Abdallah F. M., Hassanin O. (2013), Detection and molecular characterization of avipoxviruses isolated from different avian species in Egypt. *Virus Genes*. 46(1): p. 63–70 (DOI: 10.1007/s11262-012-0821-y).
14. Gyuranecz M., Foster J.T., Dán Á., Ip H.S., Egstad K.F., Parker P.G., Higashiguchi J.M., Skinner M.A., Höfle U., Kreizinger Z., Dorrestein G.M. (2013), Worldwide phylogenetic relationship of avian poxviruses. *Journal of virology*. 87(9): p. 4938-4951 (DOI: 10.1128/JVI.03183-12).
15. Pawar R.M., Bhushan S.S., Poornachandar A., Lakshmikantan U., Shivaji S. (2011), Avian pox infection in different wild birds in India. *European Journal of Wildlife Research*. 57(4): p.785-793 (DOI: 10.1007/s10344-010-0488-4).
16. Sharma B., Nashiruddullah N., Bhat M.A., Taku A., Roychoudhury P., Ahmed, J.A., Sood, S., Mehmood S. (2019), Occurrence and phylogenetic analysis of avipoxvirus isolated from birds around Jammu. *Virus Disease*. 30(2): p.288-293 (DOI: 10.1007/s13337-018-00507-0).
17. Sahu B.P., Majee P., Mishra C., Dash M., Biswal S., Sahoo N., Nayak D. (2020), The emergence of subclades A1 and A3 avipoxviruses in India. *Transboundary and Emerging Diseases*. 67(2): p. 510-517 (DOI: 10.1111/tbed.13413).

18. Audarya S.D., Riyesh T., Kumar N C.D., Sikrodia R., Sharda R., Barua S., Garg U.K. (2018), Molecular Diagnosis of a Cutaneous Form of Pox in Pigeons at Mhow in Madhya Pradesh, India. *International journal of current microbiology and applied sciences*. 72319, 7706 (DOI: 10.20546/ijcmas.2018.709.157).
19. Offerman K., Carulei O., van der Walt A.P., Douglass N., Williamson A.L. (2014), The complete genome sequences of poxviruses isolated from a penguin and a pigeon in South Africa and comparison to other sequenced avipoxviruses. *BMC genomics*. 15(1): p.463 (DOI: 10.1186/1471-2164-15-463).
20. Sarker, S., Roberts, H.K., Tidd, N., Ault, S., Ladmore, G., Peters, A., Forwood, J.K., Helbig, K. and Raidal, S.R. (2017), Molecular and microscopic characterization of a novel Eastern grey kangaropox virus genome directly from a clinical sample. *Scientific reports*. 7(1): p.1-13 (DOI: 10.1038/s41598-017-16775-7).
21. Bolger A.M., Lohse M., Usadel B. (2014), Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 30(15): p. 2114-2120 (DOI: 10.1093/bioinformatics/btu170).
22. Li H., Durbin R. (2009), Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 25(14): p. 1754-1760 (DOI: 10.1093/bioinformatics/btp324).

23. Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. (2009), The sequence alignment/map format and SAMtools. *Bioinformatics*. 25(16): p. 2078-2079 (DOI: 10.1093/bioinformatics/btp352).
24. Tcherepanov V., Ehlers A., Upton C. (2006), Genome Annotation Transfer Utility (GATU): rapid annotation of viral genomes using a closely related reference genome. *BMC genomics*. 7(1): p. 150 (DOI: 10.1186/1471-2164-7-150).
25. Hall T., Biosciences I., Carlsbad C. (2011), BioEdit: an important software for molecular biology. *GERF Bull Biosci*. 2(1): p. 60-61
26. Tamura K., Stecher G., Peterson D., Filipski A., Kumar S. (2013), MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular biology and evolution*. 30(12): p. 2725-2729 (DOI: 10.1093/molbev/mst197).
27. Martin D. P., Murrell B., Golden M., Khoosal A., Muhire B. J. V. e. (2015), RDP4: Detection and analysis of recombination patterns in virus genomes. (DOI: 10.1093/ve/vev003).

Chapter 10

A comprehensive analysis of Simple Sequence Repeats in Picornaviruses.

10.1 Introduction

Non-enveloped virions with icosahedral capsids characterize the Picornaviridae family members. The capsid contains a small, single, and positive-stranded RNA as the genome. According to the *International Committee on Taxonomy of Virus* (ICTV) 2020, the family consists of 63 genera and 147 species. The family contains a range of human and animal pathogens such as foot and mouth disease virus (FMDV), human enterovirus 71 (EV-71), poliovirus, etc. The Picornavirus genome range from 7 to 8.8 kb in size and share a typical pattern of gene arrangement. The 5' end of the genome is covalently linked to the small viral protein VPg (2KD size), and 3' end is polyadenylated [1, 2]. The 5' end of the genome is further anchored by untranslated region (5' UTR), and a 5' terminal domain; regulates viral replication and an internal ribosomal entry site (IRES) element that allows cap-independent translation activity [3]. The 3' UTR contains a pseudoknot structure involved in regulating replication and ends with a poly-A tail that is heterogeneous in length [4]. Following virus entry to the permissible host cell, the cap-independent translation of the viral genome results in a precursor polyprotein synthesis. The precursor polyprotein is further proteolytically cleaved to form several precursor molecules and 10–12 mature proteins [5]. These include structural proteins of the capsid (1A–1D or VP4–VP1) located at the amino-terminal portion of the polypeptide, followed by the nonstructural proteins (2A–2C and 3A–3D). In some

genera, the polyprotein starts with a leader (L) peptide followed by VP0 structural protein. Upon synthesis of the viral proteins, the genome replicated in complexes associated with cytoplasmic membranes [6]. During viral genome replication, viral RNA dependent RNA polymerase (RdRp) frequently switches the template, which results in "replicative recombination". It not only helps to maintain not only stability but also the variation of Picornavirus genomes. These recombination events, coupled with mutation, generate a large ensemble of genetically diverse genomes, contributing to the Picornavirus evolution [7, 8].

Simple sequence repeats (SSRs), also called microsatellites, are tandem repetitions of relatively short nucleotide motifs. They are distributed ubiquitously in the genome of eukaryotes and prokaryotes. SSRs are relatively rare in smaller viral genomes [9-11]. The microsatellites may be classified as either simple or compound, depending on the constituent of nucleotide sequences. Two or more microsatellites reside directly adjacent to each other form compound microsatellites, and often, they are interrupted by other repeats [12]. Compound microsatellites have been found in diverse taxa, including viruses, prokaryotes, and eukaryotes [13, 14]. In the human genome, ~10% of microsatellites are categorized as compound microsatellites. These contain some highly polymorphic compound repeats, such as (dC-dA)_n-(dG-dT)_n [15]. Eukaryotic such as *Homo sapiens*, *Macaca mulatta*, *Mus musculus*, *Rattus norvegicus*, *Ornithorhynchus anatinus*, *Gallus gallus*, *Danio rerio*, and *Drosophila melanogaster* genomes show a higher frequency (4–25%) of compound microsatellites [14]. In a similar line, HIV-I isolates show a

different pattern of compound microsatellites distribution, where the frequency varies from 0–24.24% for compound microsatellites [13, 16, 17]. Further, viral genome studies have unraveled many exciting facts about the distribution pattern, evolution, and regulatory role of microsatellites in the viral genome.

Length polymorphism of microsatellites may bring changes to DNA structure and influence protein functions [18]. This phenomenon could trigger diseases like Huntington's disease, dentatorubro-pallidoluysian atrophy, hereditary ataxias, and spinobulbar muscular atrophy [18, 19]. Microsatellite instability through recombination processes or DNA polymerase slippage during DNA replication causes significant genetic variation [20]. Some reports suggest that recombination events between homologous microsatellites generate compound microsatellites [21]. Microsatellites are regarded as recombination hot spots, where polymerase enzymes have a higher affinity for dinucleotide repeat sequences [22]. Due to their unique characteristics, SSRs plays a significant role in meiotic recombination [23, 24]. The evolutionary role of microsatellites in various lower order species, including that of viruses, has recently come to the picture [25, 20]. Studies have identified the presence of microsatellite repeats in many viruses. These include Menovirus [26], vesicular stomatitis virus [27], hepatitis C virus [28], and human respiratory syncytial virus [29]. In the present study, we have systematically analyzed the occurrence, size, density, and motif types of different simple and compound microsatellites prevalent in picornaviruses. We also show the correlation between different parameters influencing the distribution of these repeats. We have

identified the polymorphism of these microsatellites and hypothesized their functional significance in the viral genome. Further, the recombination breakpoints in the viral genome were identified following the microsatellites' presence at these sites.

10. 2 Results

10.2.1 Number, relative abundance, and relative density of SSR in picornaviruses

Genome-wide screening of 88 available Picornavirus genomes revealed a total of

2,488 SSRs distributed across the species (Table 10.1, Figure 10.1). On average, 30 SSRs were observed per genome. The least incidence frequency of 14 was observed in Avisivirus C (V11), and a maximum incidence frequency of 46 was found in SalivirusA (V80) (Table 10.2). The RD of SSRs was observed, ranging from 13.39 bp/kb in Avisivirus C (V11) to 45.02 bp/kb in Cardiovirus A (V13). Similarly, relative abundance varied from a minimum of 1.953 in Avisivirus C (V11) to a maximum of 5.763bp/kb in Parechovirus B (V69) (Table 10.2).

Sr. no.	Genus	Species	Organism	Acc. No.	Isolate/ Strain	Geo Location	Host	Isolation Source	Disease Caused
V1	Aalivirus	Aalivirus A	Duck aalivirus 1	NC_023985	GL/12	China	Anas platyrhynchos	Abdominal cavity	Unknown
V2	Ailurivirus	Ailurivirus A	Aimelvirus 1	MF327529	gpai001	China	Unknown	Faeces	Unknown
V3	Ampivirus	Ampivirus A	Ampivirus A1	NC_027214	NEWT/2013/HUN	Hungary	Newt	Faeces	Unknown
V4	Aphthovirus	Bovine rhinitis B virus	Bovine rhinitis B virus	NC_010354	EC11	United Kingdom	Bos taurus	Lungs	Respiratory disease in cattle
V5		Equine rhinitis A virus	Equine rhinitis A virus	KM269483	D1305-03	United Arab Emirates: Dubai	Camelus dromedarius	Unknown	Respiratory tract disease in horses
V6		Foot-and-mouth disease virus	Foot-and-mouth disease virus - type O	AF511039	Akesu/58	China	Cattle	Cattle epithelial blister	Foot and mouth disease
V7	Aquamavirus	Aquamavirus A	Seal picornavirus type 1	NC_009891	HO.02.21	Canada	Pusahispida	Oronasopharynx, lymph, lung	Unknown
V8	Avihepatovirus	Avihepatovirus A	Avihepatovirus A	KC893553	12-01	China	Anatidae	Unknown	Duck hepatitis disease
V9	Avisivirus	Avisivirus A	Turkey avisivirus	KC465954	turkey/M17-6-TuASV/2011/HUN	Hungary	Meleagris gallopavo	Faeces	Turkey viral hepatitis
V10		Avisivirus B	Chicken picornavirus 2	NC_024766	44C	Hong Kong	Gallus gallus	Tracheal and cloacal swab	Diarrhoea in chicken
V11		Avisivirus C	Chicken picornavirus 3	NC_024767	45C	Hong Kong	Gallus gallus	Tracheal and cloacal swab	Unknown
V12	Bopivirus	Bopivirus A	Bopivirus A	KM589358	TCH6	USA	Bos taurus	Unknown	Unknown
V13	Cardiovirus	Cardiovirus A	Encephalomyocarditis virus	X74312	emcv-pv21	Germany	Unknown	By plague purification from "M-variant"	Myocarditis and encephalitis
V14		Cardiovirus B	Human TMEV-like cardiovirus	GU595289	HTCV-UC6	USA	Homo sapiens	Faeces	Encephalitis
V15		Cardiovirus C	Boone cardiovirus 1	JQ864242	BCV-1	USA	Rattus norvegicus	Unknown	Unknown
V16	Cosavirus	Cosavirus A	Cosavirus A	NC_012800	HCoSV-A1	N/A	Homo sapiens	Faeces	Unknown
V17		Cosavirus B	Human cosavirus B	NC_012801	HCoSV-B1	Pakistan	Homo sapiens	Faeces	Acute flaccid paralysis, acute diarrhoea and acute gastroenteritis

V18		Cosavirus D	Cosavirus D	NC_012802	HCoSV-D1	Pakistan	Homo sapiens	Faeces	Unknown
V19		Cosavirus E	Cosavirus E	NC_012798	HCoSV-E1	Australia	Homo sapiens	Stool specimen	Unknown
V20	Dicipivirus	Cadicivirus A	Canine picodicrovirus	NC_021178	209	Hong Kong	Canis lupus familiaris	Faecal and urine specimen	Unknown
V21	Enterovirus	Enterovirus A	Coxsackievirus A8	KM609477	CVA8/SZ127/CHN/2012	China	Homo sapiens	Stool specimen	Aseptic meningitis
V22		Enterovirus B	Echovirus E14	KP289441	E14/P968/2013/China	China	Homo sapiens	Stool specimen	Hand, foot, mouth disease
V23		Enterovirus C	Enterovirus C	KP793687	Brunenders	N/A	Homo sapiens	Unknown	Acute flaccid paralysis
V24		Enterovirus D	Enterovirus D68	KT280503	2011-21186	China	Homo sapiens	Unknown	Acute haemorrhagic conjunctivitis
V25		Enterovirus E	Enterovirus E	KM667941	BEV/Egypt/2014	Egypt	Cattle	Faeces	Reproductive, respiratory, or enteric disease in cattle
V26		Enterovirus F	Bovine enterovirus type 2	HQ663846	BJ001	China	bovine	Faecal swabs	Respiratory disease
V27		Enterovirus G	Porcine enterovirus 9	HM131607	Ch-ah-f1	China	Swine	Faeces	Stillbirths, mummification of foetuses, embryonic death, and infertility in porcine
V28		Enterovirus H	Enterovirus H	NC_003988	N/A	N/A	Unknown	Unknown	Unknown
V29		Enterovirus I	Dromedary camel enterovirus 19CC	KP345887	19CC	United Arab Emirates: Dubai	Dromedary	Faeces	Unknown
V30		Enterovirus J	Enterovirus J	NC_010415	1631	N/A	Unknown	Unknown	Unknown
V31		Enterovirus K	Picornaviridae sp. rodent/Ee/PicoV/NX2015	NC_038989	rodent/Ee/PicoV/NX2015	China	Rodent	Pharyngeal and anal swab specimens	Unknown
V32		Enterovirus L	Enterovirus SEV-gx	NC_029905	SEV-gx	China	Macaca mulatta	Faeces	Unknown
V33		Rhinovirus A	Human rhinovirus A1	FJ445111	ATCC VR-1559	USA	Homo sapiens	Naso-pharyngeal washings	Common cold
V34		Rhinovirus B	Human	FJ445186	ATCC VR-	USA	Homo	Presumed	Common

			rhinovirus B27		1137		sapiens	from human throat washings	cold
V35		Rhinovirus C	Rhinovirus C	EF582387	026	Hong Kong	Homo sapiens	Nasopharyngeal aspirates of patients	Respiratory tract disease
V36	Erbovirus	Erbovirus A	Equine rhinitis B virus 2	AF361253	P313/75	N/A	Equus caballus, rabbit (Laboratory host)	Unknown	Respiratory infection
V37	Gallivirus	Gallivirus A	Turkey gallivirus	NC_018400	turkey/M176/2011/HUN	Hungary	Meleagris gallopavo	Faeces	Gastrointestinal Disease
V38	Harkavirus	Harkavirus A	Falcovirus A1	NC_026921	kestrel/VOVE0622/2013/HUN	Hungary	Falco tinnunculus	cloacal sample	Unknown
V39	Hepatovirus	Hepatovirus A	Hepatovirus A	AB279735	HAJ85-1	Japan: Yamana shi	Homo sapiens	serum	Hepatitis
V40		Hepatovirus B	Phopivirus	NC_027818	NewEngland_USA/2011	USA	Phocavitulina	lungs, livers, spleens, and oral mucosae	Unknown
V41		Hepatovirus C	Bat hepatovirus SMG18520Minmav2014	KT452742	SMG18520Minmav2014	Madagascar	Miniopterus cf. manavi	Tissue, Blood and faeces	Unknown
V42		Hepatovirus D	Rodent hepatovirus RMU101637Micarv2010	NC_028363	RMU101637Micarv2010	Germany	Microtus arvalis	Tissue, Blood and faeces	Unknown
V43		Hepatovirus F	Rodent hepatovirus KEF121Sigmass2012	NC_038315	KEF121Sigmass2012	Mexico	Sigmodon mscotensis	Tissue, Blood and faeces	Unknown
V44		Hepatovirus I	Shrew hepatovirus KS121232Sora2012	NC_028364	KS121232Sora2012	Germany	Sorex araneus	Tissue, Blood and faeces	Unknown
V45	Hunnivirus	Hunnivirus A	Bovine hungarovirus 1	NC_018668	BHUV1/2008/HUN	Hungary	Bos taurus	Faeces	Unknown
V46	Kobuvirus	Aichivirus A	Aichi virus 1	NC_001918	A846/88	N/A	Homo sapiens	Unknown	Gastroenteritis (the stomach flu)
V47		Aichivirus B	Aichivirus B	NC_004421	U-1	N/A	bovine	Unknown	Gastroenteritis (the stomach flu)
V48		Aichivirus C	Porcine kobuvirus	KJ452348	CH/DX/2012	China	pig	Faeces	Gastroenteritis (the stomach flu)

V49		Aichivirus D	Kobuvirus cattle/Kagoshima-1-22-KoV/2014/JPN	NC_027919	Kagoshima-1-22-KoV/2014/JPN	Japan: Kagoshima	Bos taurus	Faeces	Bovine diarrhoea
V50		Aichivirus E	Rabbit picornavirus	KT325852	Rabbit01/2013/HUN	Hungary	Oryctolagus cuniculus var. domestica	Faeces	Unknown
V51		Aichivirus F	Bat picornavirus	KJ641687	BtMf-PicoV/FJ2012	China	Miniopterus fuliginosus	Pharyngeal and anal swab specimens	Unknown
V52	Kunsagivirus	Kunsagivirus A	Kunsagivirus A	KC935379	roller/SZAL6-KuV/2011/HUN	Hungary	Coracias garrulus	Faeces	Unknown
V53		Kunsagivirus C	Bakunsa virus	NC_034206	baboon/M27-KuV/1986/TAN	Tanzania: Mikumi National Park	Papio cynocephalus	Unknown	Unknown
V54	Limnipivirus	Limnipivirus A	Bluegill picornavirus	JX134222	04-032	USA	Lepomis macrochirus	Unknown	Unknown
V55		Limnipivirus B	Carp picornavirus 1	NC_023162	F37/06	Germany	Cyprinus carpio	heart, brain and liver	Unknown
V56		Limnipivirus C	Fathead minnow picornavirus	NC_039212	FHMPV-1	USA	Pimephales promelas	kidney/spleen or entire viscera	Unknown
V57	Livupivirus	Livupivirus A	Livupivirus A	NC_032126	newt/II-5-Pilis/2014/HUN	Hungary	Lissotriton vulgaris	Faeces	Unknown
V58	Megrivirus	Megrivirus A	Megrivirus A	KC663628	LY	N/A	Anas platyrhynchos	Unknown	Unknown
V59		Megrivirus B	Picornavirus HK21	NC_038957	HK21	Hong Kong	pigeon	Faeces	Unknown
V60		Megrivirus D	Harrier picornavirus 1	NC_034617	harrier/MR-01/HUN/2014	Hungary	Circus aeruginosus	cloacal sample	Unknown
V61		Megrivirus E	Penguin megrivirus	NC_039004	KGI-Bel-P5/2015	N/A	Pygoscelis adeliae	Unknown	Unknown
V62		Melegrivirus A	Turkey hepatitis virus 2993D	NC_021201	124	USA	Turkey poultry	liver	Hepatitis
V63	Mischivirus	Mischivirus A	Miniopterus schreibersii picornavirus 1	JQ814851	Unknown	China	Miniopterus schreibersii	Unknown	Unknown
V64		Mischivirus C	African bat icavirus PREDICT-06105	NC_026470	PREDICT-06105	Democratic Republic of the Congo	Hipposideros gigas	oral swab	Unknown
V65	Mosavirus	Mosavirus A	Mosavirus A2	NC_023987	SZAL6-MoV/2011/HUN	Hungary	Coracias garrulus	Faeces	Unknown
V66	Orivirus	Orivirus A	Orivirus A	KM203656	chicken/Pf-CHK1/2013	Hungary	Gallus gallus domesticus	cloacal sample	Common cold

					/HUN		cus		
V67	Oscivirus	Oscivirus A	Oscivirus A1	GU182409	007167	Hong Kong	Oriental Magpie Robin	Unknown	Unknown
V68	Parechovirus	Parechovirus A	Human parechovirus 1	JX575746	CAU10-NN	South Korea	Homo sapiens	Unknown	Respiratory or gastrointestinal illness
V69		Parechovirus B	Ljungan virus	EF202833	87-012G	Sweden	Clethrionomy sglareolus	Unknown	Malformations, intrauterine foetal death
V70		Parechovirus C	Sebokele virus 1	NC_021482	1	Central African Republic	Hylomyscus, suckling newborn mice (Laboratory host)	Unknown	Unknown
V71	Pasivirus	Pasivirus A	Pasivirus A	LT898442	SPaV-A/GER/L00721/2014	N/A	Unknown	Unknown	Diarrhoea
V72	Passerivirus	Passerivirus A	Passerivirus A1	NC_014411	00356	Hong Kong	Pale Thrush	Unknown	Unknown
V73	Poecivirus	Poecivirus A	Poecivirus BCCH-449	KU977108	BCCH-449	USA: Alaska	Black-capped chickadee	beak	Avian keratin disorder (AKD)
V74	Potamipivirus	Potamipivirus A	Potamipivirus A	MK189163	TSPV	USA: Alaska	Gasterosteus-aculeatus	intestine	Unknown
V75	Rabovirus	Rabovirus A	Rabovirus A	NC_026314	Berlin/Jan2011/0572	Germany	Rattus norvegicus	Faeces	Unknown
V76	Rosavirus	Rosavirus A	Rosavirus A2	NC_024070	GA7403	Gambia	Homo sapiens	Faeces	Unknown
V77		Rosavirus B	Rosavirus B	NC_031105	RNCW0602091R	Hong Kong	Rattus norvegicus	Unknown	Multisystemic dissemination
V78		Rosavirus C	Rosavirus C	NC_031106	RATLC11A	Hong Kong	Rattusandam anensis	Unknown	Multisystemic dissemination
V79	Sakobuvirus	Sakobuvirus A	Feline sakobuvirus A	NC_022802	FFUP1	Portugal	Feliscatus	Unknown	Unknown
V80	Salivirus	Salivirus A	Salivirus A	NC_012986	02394-01	USA	Homo sapiens	stool specimen	Gastroenteritis
V81	Sapelovirus	Avian sapelovirus	Avian sapelovirus	NC_006553	TW90A	N/A	Unknown	Unknown	Unknown
V82		Sapelovirus A	Porcine sapelovirus 1	NC_003987	V13	N/A	Unknown	Unknown	Diarrhoea, polioencephalomyelitis, pneumonia
V83		Sapelovirus B	Simian sapelovirus 1	NC_004451	2383	N/A	Unknown	Unknown	Unknown
V84	Senecavirus	Senecavirus	Senecavirus A	KC667560	11-55910-3	Canada	swine	brain	Vesicular

		A							lesions in pigs
V85	Siccinivirus	Siccinivirus A	Siccinivirus Pf-CHK1/SiV	KT880665	Pf-CHK1/SiV	Hungary	Gallus gallusdomesticus	cloacal sample	Unknown
V86	Teschovirus	Teschovirus A	Teschovirus A	KC667562	13-3064-3	Dominican Republic	swine	spinal cord	Encephalomyelitis in pigs
V87	Torchivirus	Torchivirus A	Tortoise picornavirus	KM873611	14-04	Germany	Testudohermanni	Unknown	Unknown
V88	Tremovirus	Tremovirus A	Tremovirus A	KF979338	204C	Hong Kong	Gallus gallus	Unknown	Encephalomyelitis

Table 10.1: Overview of selected 88 species of Picornaviridae family. The dataset includes specification of the viral strain, specific host, geo-location, isolation source, and disease caused.

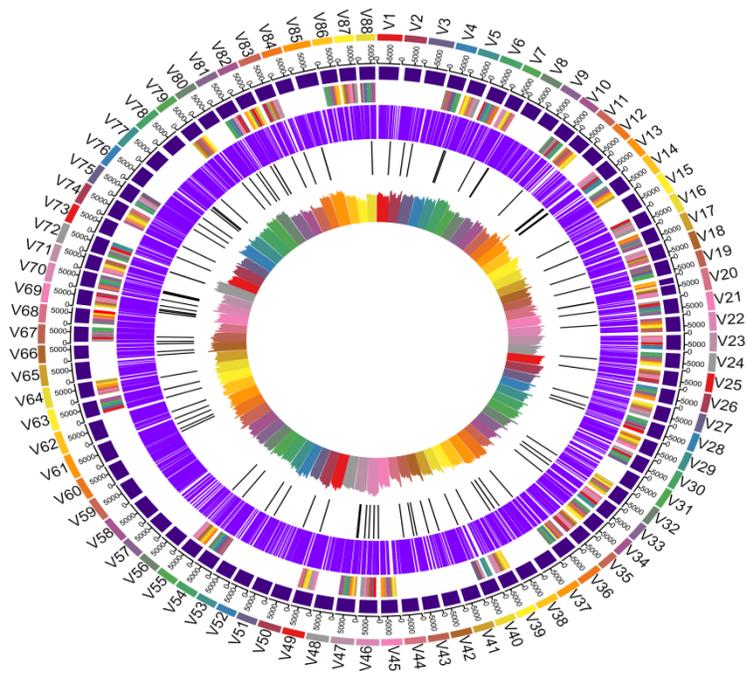


Figure 10.1: The Circos map is representing 88 picornavirus genomes. From the outer ring to inner: Selected picornavirus genomes, genome size, polyprotein gene, mature proteins, SSRs, cSSRs, and GC content.

10.2.2 Number, relative abundance, and relative density of cSSR in Picornaviruses

The genome-wide scan revealed 1–5 compound microsatellites (cSSRs) in each analyzed sequence. Interestingly, 26 genomes lacked any cSSR. A total of 100 cSSRs were observed in 88 genomes (Table 10.2). The RD of cSSRs changed drastically in selected genomes, ranging from 1.36 bp/kb in Erbovirus A (V36) to 26.84 bp/kb in Cardiovirus A (V13). Similarly, the RA varied from 0.108 bp/kb in Ampivirus A (V3) to 0.636 bp/kb in the genome of Cardiovirus A (V13). The percentage of individual microsatellites being part of compound microsatellite (cSSR%) ranged from 2.439 in GallivirusA (V37) to 15.15 in Cardiovirus A (V13) (Table 10.2).

Sr. No.	Acc. No.	Genome size (bp)	GC (%)	Total no of SSR	Total no of cSSR	RA (SSR)	RD (SSR)	RA (cSSR)	RD (cSSR)	% of cSSR
V1	NC_023985	8976	43.17	29	1	3.231	24.06	0.111	2.78	3.448
V2	MF327529	8029	45.19	29	1	3.612	24.909	0.125	2.49	3.448
V3	NC_027214	9246	45.14	31	1	3.353	21.84	0.108	2.16	3.226
V4	NC_010354	7556	45.55	32	0	4.235	33.24	0.000	0	0.000
V5	KM269483	7681	47.57	38	3	4.947	31.62	0.391	7.16	7.895
V6	AF511039	8147	53.38	24	0	2.946	20.6	0.000	0	0.000
V7	NC_009891	6718	43.68	35	1	5.210	39.27	0.149	2.23	2.857
V8	KC893553	7800	43.01	38	2	4.872	34.47	0.256	3.46	5.263
V9	KC465954	7532	44.96	23	0	3.054	21.1	0.000	0	0.000
V10	NC_024766	7310	48.52	23	0	3.146	19.69	0.000	0	0.000
V11	NC_024767	7167	45.21	14	1	1.953	13.39	0.140	1.95	7.143
V12	KM589358	7018	50.37	21	0	2.992	22.08	0.000	0	0.000
V13	X74312	7861	49.73	33	5	4.198	45.02	0.636	26.84	15.152
V14	GU595289	8047	43.85	20	0	2.485	16.89	0.000	0	0.000
V15	JQ864242	8530	47.74	29	1	3.400	26.49	0.117	1.87	3.448
V16	NC_012800	7632	43.75	28	0	3.669	27.64	0.000	0	0.000
V17	NC_012801	7205	43.78	27	1	3.747	27.48	0.139	4.99	3.704
V18	NC_012802	7215	42.47	27	1	3.742	25.502	0.139	2.91	3.703

V19	NC_012798	6580	41.93	26	1	3.951	29.93	0.152	1.82	3.846
V20	NC_021178	8785	41.58	33	2	3.756	27.66	0.228	5.69	6.061
V21	KM609477	7396	47.64	21	1	2.839	19.06	0.135	2.02	4.762
V22	KP289441	7451	47.6	27	1	3.624	24.15	0.134	1.87	3.704
V23	KP793687	7507	46.09	21	0	2.797	28.22	0.000	0	0.000
V24	KT280503	7332	41.8	24	1	3.273	18.67	0.136	2.04	4.167
V25	KM667941	7417	50.41	20	1	2.697	20.89	0.135	1.69	5.000
V26	HQ663846	7430	50.76	18	0	2.423	19.77	0.000	0	0.000
V27	HM131607	7390	45.27	19	1	2.571	17.18	0.135	2.57	5.263
V28	NC_003988	7374	42.93	26	2	3.526	29.96	0.271	3.25	7.692
V29	KP345887	7441	45.39	26	0	3.494	26.86	0.000	0	0.000
V30	NC_010415	7351	45.28	20	0	2.721	16.99	0.000	0	0.000
V31	NC_038989	7805	46.09	17	0	2.178	16.14	0.000	0	0.000
V32	NC_029905	7367	43.14	21	1	2.851	19.95	0.136	2.03	4.762
V33	FJ445111	7137	37.43	31	2	4.344	29.704	0.280	4.2	6.452
V34	FJ445186	7217	39.46	26	1	3.603	21.05	0.139	1.66	3.846
V35	EF582387	7086	42.98	27	0	3.810	22.45	0.000	0	0.000
V36	AF361253	8821	50.4	35	1	3.968	27.2	0.113	1.36	2.857
V37	NC_018400	8496	48.28	41	1	4.826	32.71	0.118	1.41	2.439
V38	NC_026921	8003	41.80	29	3	3.624	25.61	0.375	7.74	10.345
V39	AB279735	7478	38.2	22	1	2.942	22.45	0.134	2	4.545
V40	NC_027818	7475	36.80	16	1	2.140	17.25	0.134	1.605	6.250
V41	KT452742	7570	38.08	34	0	4.491	27.73	0.000	0	0.000
V42	NC_028363	7656	36.10	26	2	3.396	26.907	0.261	5.09	7.692
V43	NC_038315	7563	35.11	35	2	4.628	31.46	0.264	4.23	5.714
V44	NC_028364	7810	36.71	36	1	4.609	29.705	0.128	1.92	2.778
V45	NC_018668	7583	45.57	24	0	3.165	22.41	0.000	0	0.000
V46	NC_001918	8251	58.9	45	4	5.454	36.59	0.485	6.54	8.889
V47	NC_004421	8374	54.6	38	2	4.538	29.83	0.239	3.22	5.263
V48	KJ452348	8124	51.82	30	0	3.693	25.72	0.000	0	0.000
V49	NC_027919	8087	55.19	40	1	4.946	34.87	0.124	2.59	2.500
V50	KT325852	8364	55.17	35	1	4.185	29.41	0.120	2.51	2.857
V51	KJ641687	7435	49.37	18	0	2.421	24.209	0.000	0	0.000
V52	KC935379	7272	53.01	35	1	4.813	33.69	0.138	2.47	2.857
V53	NC_034206	7429	49.02	25	1	3.365	22.47	0.135	2.42	4.000
V54	JX134222	8050	44.01	36	0	4.472	32.79	0.000	0	0.000
V55	NC_023162	7697	45.01	20	1	2.598	25.98	0.130	2.2	5.000
V56	NC_039212	7746	45.82	18	1	2.324	15.87	0.129	1.93	5.556
V57	NC_032126	7768	46.77	24	0	3.090	20.85	0.000	0	0.000
V58	KC663628	9700	45.26	44	0	4.536	30.309	0.000	0	0.000

V59	NC_038957	9100	47.37	34	0	3.736	27.36	0.000	0	0.000
V60	NC_034617	8541	45.51	28	1	3.278	20.25	0.117	1.404	3.571
V61	NC_039004	9702	43.22	41	3	4.226	29.16	0.309	5.66	7.317
V62	NC_021201	9075	46.07	39	1	4.297	28.98	0.110	2.09	2.564
V63	JQ814851	8468	47.43	38	2	4.487	31.52	0.236	3.18	5.263
V64	NC_026470	8096	47.31	33	1	4.076	26.92	0.124	2.09	3.030
V65	NC_023987	8398	45.47	24	0	2.858	21.18	0.000	0	0.000
V66	KM203656	7037	49.8	29	3	4.121	26.57	0.426	7.1	10.345
V67	GU182409	7640	46.58	34	2	4.450	32.32	0.262	3.14	5.882
V68	JX575746	7348	39.75	24	0	3.266	23.38	0.000	0	0.000
V69	EF202833	7635	42.58	44	4	5.763	37.19	0.524	6.67	9.091
V70	NC_021482	7537	45.75	27	4	3.582	23.61	0.530	8.35	14.815
V71	LT898442	6917	43.02	27	0	3.903	25.01	0.000	0	0.000
V72	NC_014411	8035	57.92	26	2	3.236	23.26	0.249	3.23	7.692
V73	KU977108	7653	43.45	34	1	4.443	29.26	0.131	1.56	2.941
V74	MK189163	8532	43.33	30	1	3.516	27.89	0.117	2.1	3.333
V75	NC_026314	7853	43.23	32	1	4.074	28.77	0.127	2.54	3.125
V76	NC_024070	8930	51.41	40	2	4.479	28.55	0.224	4.25	5.000
V77	NC_031105	8938	50.92	34	3	3.803	24.83	0.335	7.04	8.824
V78	NC_031106	9112	50.98	33	1	3.622	26.44	0.110	1.86	3.030
V79	NC_022802	7807	55.51	40	1	5.124	35.34	0.128	1.53	2.500
V80	NC_012986	7989	56.67	46	2	5.758	39.04	0.250	3.37	4.348
V81	NC_006553	8289	42.69	35	2	4.222	34.97	0.241	2.89	5.714
V82	NC_003987	7491	41.03	26	2	3.471	22.01	0.267	3.6	7.692
V83	NC_004451	8126	40.38	29	1	3.569	23.12	0.123	1.84	3.448
V84	KC667560	7356	51.54	25	1	3.399	25.55	0.136	1.63	4.000
V85	KT880665	9883	53.81	35	2	3.541	22.35	0.202	2.42	5.714
V86	KC667562	7155	44.55	21	0	2.935	25.7	0.000	0	0.000
V87	KM873611	7093	36.04	30	0	4.230	30.17	0.000	0	0.000
V88	KF979338	6955	44.18	25	1	3.595	23.28	0.144	1.72	4.000

Table 10.2: Survey of various microsatellites (SSRs and cSSRs) with their relative abundance, relative density, and cSSR % in selected Picornaviridae family.

10.2.3 Effect of dMAX on cSSR incidence

The dMAX is the maximum distance between any two adjacent microsatellites. If the distance separating two

microsatellites is less than or equivalent to dMAX, these microsatellites are considered as cSSR [14]. To determine the impact of dMAX, five genome sequences such as bovine rhinitis B virus (V4), Cardiovirus A (V13), Enterovirus A (V21), Aichivirus A (V46), Salivirus A (V80) were chosen randomly to determine the variability of cSSR with increasing dMAX. It is noteworthy that the dMAX value in the IMEx software could only be set between 0 and 50 for analysis [30]. Our study revealed an overall increase in the number of cSSRs with higher dMAX value in all selected picornaviruses except for Enterovirus A (V21) (Figure 10.2).

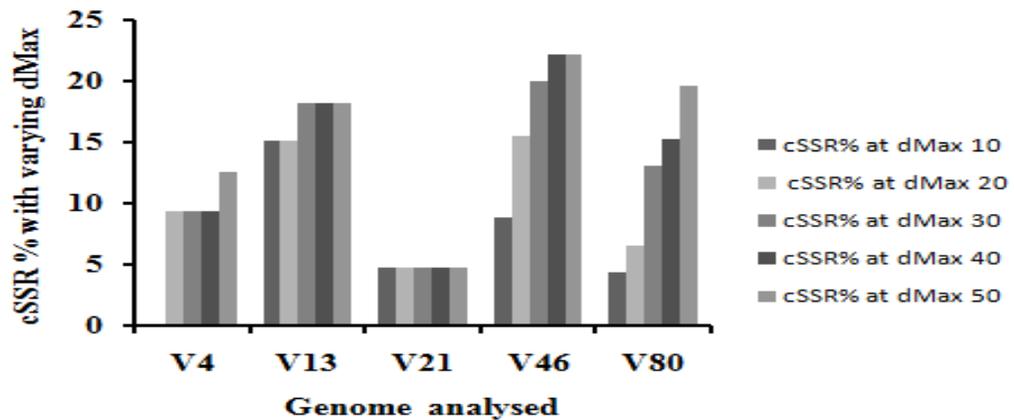


Figure 10.2: Effect of cSSR% on varying dMAX in five different species of Picornavirus: For this analysis, IMEx software was used, which allowed varying the dMAX value from 10 to 50.

10.2.4 Diversity of derived motifs

This study's derived microsatellites revealed the presence of mono to penta-nucleotide motif throughout the genome of the Picornaviridae family. Mono-nucleotide repeats were exhibited predominantly next to the dinucleotide motif. A 141 mono-nucleotide repeats stretch was observed in Cardiovirus

A (V13) (Acc. no. X74312). Poly (G/C) repeats were found to be more prevalent than poly (A/T) repeats. Within the dinucleotide repeats, CT/TC maintained the highest, whereas CG/GC repeats exhibited the lowest in terms of their distribution in the Picornaviridae. The CT/TC repeats were approximately 4.25 times more abundant than the least represented CG/GC repeats. The tri-nucleotide repeats were present abundantly next to mono-nucleotide repeats and contain 54 codon type repeats. Among them, GAA/AAG, and CAA/AAC represent the highest coding density, codes for glutamic acid/lysine and glutamine/asparagine, respectively. In this study, seventeen tetra-nucleotide and six pentanucleotide motifs were observed among the Picornaviruses. Among these, three tetra-nucleotide motifs such as CTCC, TCCT, and TTAG were localized in Aichivirus A (V46), and two pentanucleotide motifs such as GTTAA and TTAAG were localized in Aichivirus F (V51). Additionally, we have verified the distribution of SSRs throughout the non-genic and genic regions. The non-genic regions occupy 15.50% of SSR, and the rest 84.50% contribute toward the genic region. The 3D gene, which codes for the RdRp enzyme, contains maximum percentage (14.80%) followed by 2C (12.39%), then by VP3 (10.33%), which codes the protein for viral replication and capsid formation, respectively (Figure 10.3). We further examined the presence of mono-, di-, and tri-nucleotide repeats in the genic and non-genic regions depicted in Figure 10.4.

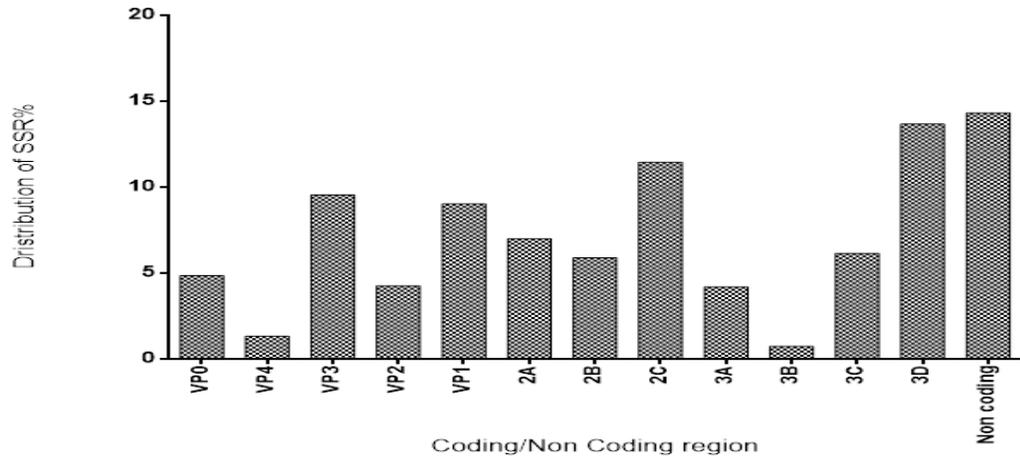


Figure 10.3: Differential distribution of SSR in the coding/noncoding region of Picornavirus.

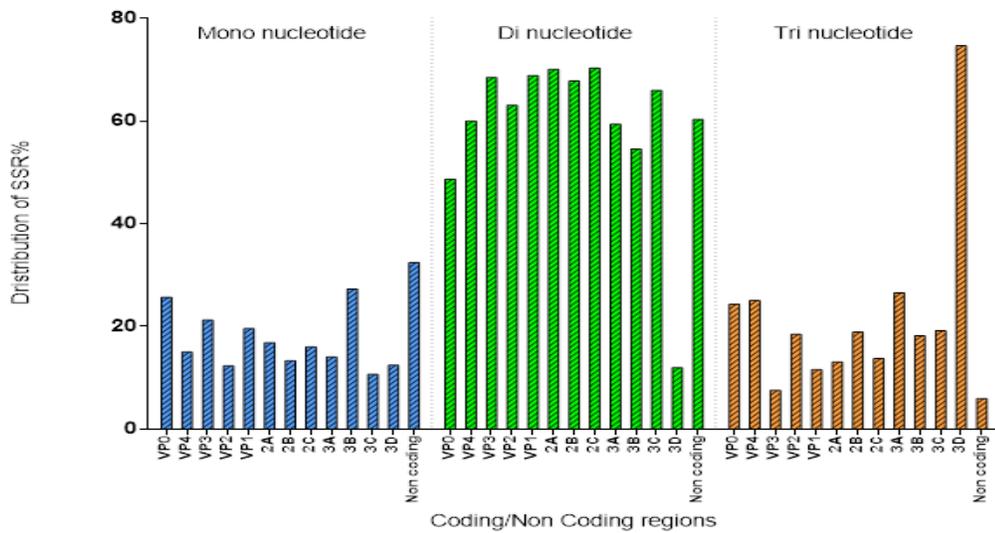


Figure 10.4: Distribution of mono-, di- and tri-nucleotide SSR motifs (%) across coding/noncoding regions of Picornavirus.

We have searched the biasness of the class of repeats to a genomic region. We found that the capsid protein VP4 is biased towards (A)₆ repeats, which code for lysine. While VP0, VP3, and 2A (codes for viral protease) were enriched with polyC, which codes for proline. We surveyed the conserved repeats at the 5' UTR and 3' UTR, which were

enriched with polyC and polyA, respectively. Except for VP0 (helping for RNA encapsidation), the gene enriched with dinucleotide repeats such as TC/CT, codes for serine/leucine, and rest of the regions were highly diverse with di- as well as tri-repeat types. The diversification of microsatellite indicates the rapid evolution of the species to adapt to a broad range of hosts.

10.2.5 Over and under-representation of mononucleotide repeat

We noticed a considerable variation in the numbers of mononucleotide repeats (≥ 6 nt), ranging from one (Enterovirus J, Enterovirus K, Hepatovirus D, Mosavirus A & Tremovirus A) to 22 (Parechovirus B). The ratio between the observed number of repeats and the expected number of repeats also varied considerably from -7.6407 (Rhinovirus A) to 14.8191 (Parechovirus B) (Table 10.3). Among the analyzed sequences, approximately 72.5% showed the under-represented distribution of MNRs. Broader host range and virus genome type might have influenced the ratio of the observed number of repeats to the expected number of repeats (O/E ratio) to some degree. The Z-scores revealed the statistical significance of mononucleotide repeat representation. If the observed value is more than the expected value, the Z score is assigned $Z > 0$ and vice versa. Thus, the greater the Z value, the higher is the statistical significance. Kunsagivirus A with the highest Z value of 4.85 denotes the most significant genome in terms of mononucleotide microsatellite evolution. Some of the viruses showed an over-representation of MNR loci compared to the expected value. This may arise due to the co-evolution of viral SSRs with that of the host genome.

Sl. No.	Observed Total	Real Expected	O/E	O-E	Square Root of E	Z score
V1	6	8.792490113	0.682401	-2.79249	2.965213333	-0.941750154
V2	2	5.853882275	0.341654	-3.85388	2.419479753	-1.592855766
V3	5	6.950827321	0.719339	-1.95083	2.636442171	-0.739946942
V4	6	6.079360927	0.986946	-0.07936	2.465636009	-0.032186798
V5	10	5.774709596	1.731689	4.22529	2.403062545	1.758293979
V6	3	6.305285261	0.475791	-3.30529	2.511032708	-1.31630514
V7	7	5.921120397	1.182209	1.07888	2.433335241	0.443374832
V8	7	7.148990812	0.979159	-0.14899	2.673759677	-0.055723337
V9	8	6.221224962	1.28592	1.778775	2.494238353	0.713153591
V10	5	6.34132674	0.788479	-1.34133	2.518199107	-0.532653171
V11	4	5.67221231	0.705192	-1.67221	2.381640676	-0.702126197
V12	3	6.75032156	0.444423	-3.75032	2.598138095	-1.44346506
V13	9	5.738697585	1.5683	3.261302	2.395557886	1.361395788
V14	4	7.032841022	0.56876	-3.03284	2.651950418	-1.143626593
V15	6	6.67352211	0.899075	-0.67352	2.583316107	-0.260719975
V16	4	5.34522324	0.748332	-1.34522	2.311973884	-0.581850534
V17	4	5.87240551	0.681152	-1.87241	2.423304667	-0.772666159
V18	4	5.95257341	0.671978	-1.95257	2.439789624	-0.800304006
V19	5	6.45245031	0.774899	-1.45245	2.540167378	-0.571793151
V20	3	6.026337571	0.497815	-3.02634	2.45485999	-1.232794368
V21	2	5.550081918	0.360355	-3.55008	2.355861184	-1.506914729
V22	3	5.596912349	0.53601	-2.59691	2.365779438	-1.097698419
V23	3	5.910549533	0.507567	-2.91055	2.431162177	-1.197184441
V24	2	7.243071394	0.276126	-5.24307	2.691295486	-1.948158952
V25	4	5.416504044	0.738484	-1.4165	2.327338403	-0.608636906
V26	2	5.437974326	0.367784	-3.43797	2.331946467	-1.474293846
V27	2	5.76785342	0.346749	-3.76785	2.401635572	-1.568869759
V28	7	6.788938974	1.031089	0.211061	2.605559244	0.081004117
V29	5	6.027594892	0.829518	-1.02759	2.455116065	-0.41855247
V30	1	5.98317411	0.167135	-4.98317	2.446052761	-2.037230835
V31	1	5.23123021	0.19116	-4.23123	2.287188276	-1.849970225
V32	6	6.2352211	0.962275	-0.23522	2.497042471	-0.09419988
V33	2	9.640702298	0.207454	-7.6407	3.104948035	-2.460814871
V34	4	8.380757236	0.477284	-4.38076	2.894953754	-1.513239108
V35	2	6.503150558	0.307543	-4.50315	2.550127557	-1.765853063
V36	7	6.446453953	1.085868	0.553546	2.538986797	0.218018482
V37	13	6.295819278	2.064862	6.704181	2.509147122	2.671896225
V38	5	9.532580223	0.524517	-4.53258	3.087487688	-1.468048032

V39	5	6.858401211	0.729033	-1.8584	2.618854943	-0.709623577
V40	2	6.87223112	0.291026	-4.87223	2.621494063	-1.858570343
V41	6	9.73778124	0.616157	-3.73778	3.120541818	-1.197798798
V42	1	5.73224502	0.174452	-4.73225	2.39421073	-1.976536552
V43	11	5.894062032	1.866285	5.105938	2.427768941	2.103139999
V44	11	6.23342312	1.76468	4.766577	2.496682423	1.909164272
V45	2	6.096045186	0.328082	-4.09605	2.469017049	-1.658978089
V46	14	8.570159884	1.633575	5.42984	2.927483541	1.854780749
V47	5	6.795861417	0.735742	-1.79586	2.606887304	-0.688891083
V48	4	6.034100908	0.662899	-2.0341	2.456440699	-0.828068395
V49	11	4.32408912	2.543888	6.675911	2.079444426	3.210430054
V50	10	5.98	1.672241	4.02	2.445403852	1.643900248
V51	2	6.34	0.315457	-4.34	2.517935662	-1.723634192
V52	17	5.559061712	3.05807	11.44094	2.357766255	4.85244806
V53	11	5.234853212	2.101301	5.765147	2.28798016	2.519753838
V54	6	5.711408548	1.050529	0.288591	2.38985534	0.120756871
V55	5	6.345861415	0.787915	-1.34586	2.519099326	-0.534262941
V56	2	6.36221342	0.314356	-4.36221	2.522342843	-1.729429221
V57	5	6.24513412	0.800623	-1.24513	2.499026635	-0.498247639
V58	9	5.2365712	1.718682	3.763429	2.288355567	1.644599666
V59	4	6.74127623	0.593359	-2.74128	2.596396778	-1.055800197
V60	7	6.2412311	1.121574	0.758769	2.498245604	0.303720699
V61	6	6.8751218	0.872712	-0.87512	2.622045347	-0.333755402
V62	2	6.5869412	0.303631	-4.58694	2.566503692	-1.787233431
V63	6	6.391546817	0.93874	-0.39155	2.528150869	-0.154874783
V64	5	6.57221347	0.760779	-1.57221	2.563632866	-0.613275595
V65	1	6.783046221	0.147426	-5.78305	2.604428195	-2.220466755
V66	7	6.546553953	1.069265	0.453446	2.558623449	0.177222657
V67	6	5.911408748	1.014986	0.088591	2.43133888	0.036437229
V68	3	8.356769708	0.35899	-5.35677	2.890807795	-1.853035583
V69	22	7.180837653	3.063709	14.81916	2.679708501	5.530139692
V70	4	6.3652329	0.628414	-2.36523	2.522941319	-0.937490255
V71	8	6.321524062	1.265518	1.678476	2.514264119	0.667581391
V72	9	7.821327733	1.1507	1.178672	2.796663679	0.421456565
V73	14	8.79219342	1.592322	5.207807	2.965163304	1.756330443
V74	6	5.920408748	1.013444	0.079591	2.433189008	0.032710674
V75	4	6.76219785	0.591524	-2.7622	2.600422629	-1.062211126
V76	14	8.2323145	1.700615	5.767686	2.869201021	2.010206137
V77	13	6.5312134	1.990442	6.468787	2.555623877	2.531196652
V78	10	5.3423145	1.871848	4.657686	2.311344738	2.015140979
V79	5	6.611529168	0.756255	-1.61153	2.571289398	-0.626739709

V80	9	7.205316295	1.249078	1.794684	2.684272023	0.668592337
V81	2	7.746547446	0.25818	-5.74655	2.783262015	-2.064680729
V82	3	7.789803711	0.385119	-4.7898	2.791021983	-1.716146895
V83	6	8.841380113	0.678627	-2.84138	2.973445832	-0.955584959
V84	6	5.434544771	1.104048	0.565455	2.331211009	0.242558579
V85	16	7.772379513	2.058572	8.22762	2.787898763	2.951190551
V86	3	6.026337571	0.497815	-3.02634	2.45485999	-1.232794368
V87	8	6.291234962	1.27161	1.708765	2.508233434	0.681262364
V88	1	5.968473567	0.167547	-4.96847	2.443045961	-2.033720874

Table 10.3: Mono-nucleotide microsatellite representation in Picornaviridae family.

10.2.6 Motif complexity and polymorphism of Picornavirus genome

Compound microsatellites (cSSRs) are termed by the presence of two or more adjacent individual microsatellites. The common patterns of cSSR are represented as m1-xn-m2, m1-xnm2-xn-m3 considered as '2-microsatellite', and '3-microsatellite', respectively [14]. The analysis of compound microsatellite complexity indicated that all surveyed genomes were rich in '2-microsatellite' cSSR followed by a single '3-microsatellite'. Among the 88 species of Picornaviridae analyzed, the cSSR composition showed a random distribution pattern throughout the genome. The non-coding regions consist of 19% of the total microsatellite, while the rest, 81%, are present in the coding regions. Within the coding regions, VP3 occupied 14% of cSSRs, followed by 3D (13.23%) and VP1 (10.29%) regions (Figure 10.5). The CTG-CAG compound microsatellite composed of self-complementary motifs has been proposed to be created by recombination [21]. However, our study showed no such motifs, suggesting these

compound microsatellites were not likely to be derived from recombination. The motif having the form $[m1]_n-x_n-[m2]_n$ can be termed as SSR-couples, such as the compound microsatellite $(A)_6-x_2-(TA)_3$ and $(A)_8-x_2-(TA)_5$. We identified three types of SSR couples, such as $(TG)-x-(CT)$, $(TC)-x-(C)$, $(CT)-x-(C)$ were presented twice in all analyzed genomes. Some of the self-complementary motifs have been observed in Picornaviridae $(GC)-x-(CG)$, $(TC)-x-(AG)$, $(GT)-x-(CA)$, and $(TC)-x-(AG)$, which plays an important role in secondary structure formation. Motif duplication is one of the phenomena in which similar motifs are located on both ends of the spacer sequence, for example, $(CA)_n-(X)_y-(CA)_z$. About 13.23% of the total cSSR were made up of duplicated sequences having the motif pattern $(C)-x-(C)$, $(TG)-x-(TG)$, $(AT)-x-(AT)$, $(TC)-x-(TC)$ and $(TG)-x-(TG)$.

To check polymorphism, the species with more than or equal to five strains were taken into consideration. Due to the limitation of sequence information, 12 out of 88 species got qualified for the detection of polymorphism. A sum of 141 sequences was analyzed to determine species-specific consensus microsatellite motifs. We unraveled that five species contained the species-specific consensus sequences presented in the leader and VP3 regions, presumably being the most conserved microsatellite within the genome (Table 10.4). These conserved microsatellites could be used as potential biomarkers of virus identifications and population genetics study.

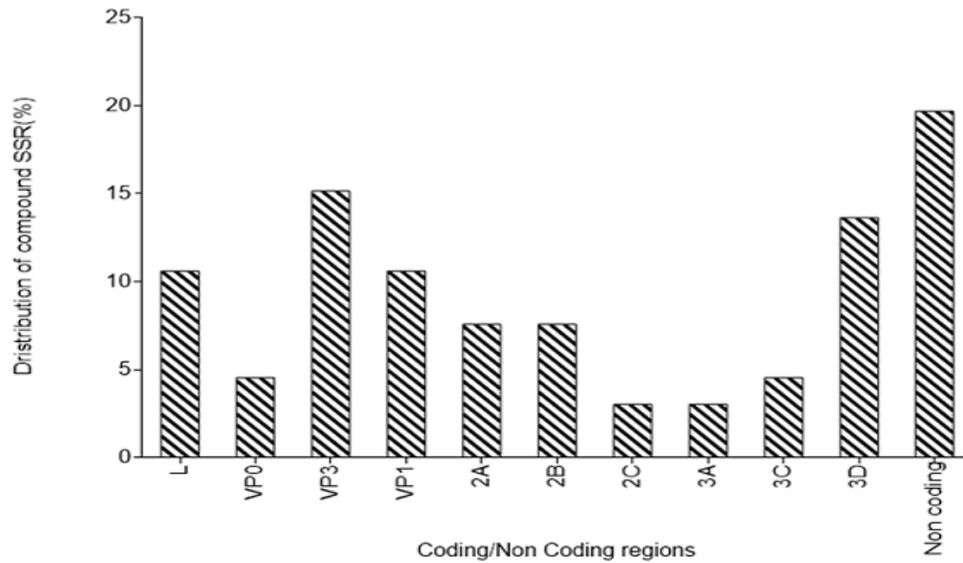


Figure 10.5: Differential distribution of cSSR in the coding/noncoding region of Picornavirus.

Species	Compound microsatellite	Start	End	Location	(a/b)	Consensus motif
Avihepatovirus A	(CT)3-x-1-(T)6	1740	1750	VP3	(17/9)	(CT)3-x-1-(T)6
	(TGA)3-x4-(GT)3	2152	2170	VP1	(17/5)	absent
Enterovirus A	(GC)3-x3-(CAA)3	2128	2145	VP3	(10/5)	(GC)3-x3-(CAA)3
Enterovirus B	(TG)3-x1-(TG)4	2229	2243	VP3	(15/2)	absent
Enterovirus D	(GAA)3-x5-(AT)3	6251	6270	3D	(17/16)	(GAA)3-x5-(AT)3
Rhinovirus B	(AT)3-x3-(AT)3	604	618	Non-coding	(12/5)	absent
Hepatovirus A	(TG)3-x-3-(GTG)3	5784	5795	3C	(21/1)	absent
Aichivirus A	(C)6-x1-(C)6	95	107	Non-coding	(6/2)	absent
	(C)6-x2-(C)6	2706	2719	VP-3	(6/2)	absent
	(GC)3-x0-(CG)3	5371	5382	2C	(6/2)	absent
	(C)6-x5-(TCCT)3	7619	7641	3D	(6/2)	absent
Parechovirus B	(GT)3-x6-(GA)3	1760	1777	VP3	(6/2)	absent
	(TC)3-x2-(CA)3	2881	2894	VP1	(6/2)	absent
	(T)8-x9-(TG)3	3681	3703	2B	(6/5)	(T)8-x9-(TG)3

	(A)7-x4-(TG)3	5964	5980	3C	(6/2)	absent
Salivirus A	(CT)3-x5-(TC)4	2449	2467	VP3	(9/2)	absent
	(CT)3-x3-(ACT)3	7201	7218	3D	(9/5)	(CT)3-x3-(ACT)3
Sapelovirus A	(C)6-x7-(CTA)3	438	459	Leader	(9/9)	(C)n-x7-(CTA)3 (n=6-7)
	(AG)3-x3-(GT)3	493	507	Leader	(9/8)	(AG)3-x3-(GT)3
Senecavirus A	(GT)3-x0-(TG)3	1205	1216	Leader	(19/19)	(GT)3-x0-(TG)3
Sicinivirus A	(GT)3-x7-(CA)3	3231	3249	VP3	(5/1)	absent
	(T)6-x6-(GT)3	3319	3336	VP3	(5/1)	absent
Tremovirus A	(TA)3-x5-(GT)3	4549	4565	3A	(5/1)	absent
a/b indicates the number of strains per species used/number of strains having cSSR						

Table 10.4: *The existence of the consensus motif in diverse species of Picornaviridae family.*

10.2.7 Recombination analysis

When the dataset of 141 complete genomes of 12 species was analyzed, three species were found lacking any recombination event. The rest of the species' recombination breakpoints were checked with major parent and minor parent information and their exact position in the genome. This analysis resulted in a total of 15 breakpoints, a majority of which associated with the structural proteins (VP4-VP1), followed by nonstructural capsid protein (2A-2C and 3A-3D). However, the least number of breakpoints were observed in the non-coding regions, followed by the leader sequence. Previous studies suggest that repetitive sequences are favored for the recombination and act as a hotspot, owing to the high affinity of recombinase enzymes towards dinucleotide repeats [22]. This prompted us to check the repeats within the breakpoints present in nine species. Our results show that out of 15 breakpoints, three were devoid of any microsatellite.

In contrast, others are rich in dinucleotide repeats, which contribute 92%, followed by 4% tri-nucleotide repeats and 4% mono-nucleotide repeats. Among the dinucleotide repeats, the

GT/TG prevailed at the highest (25%) followed by AG/GA (18%), and CT/TC occurred at the least (2.2%) frequency. However, these data were not conclusive but laid the foundation for further research to know the relationship between microsatellite and recombination hotspot in the Picornavirus family.

10.2.8 Genomic parameters influencing SSR and cSSR distribution

Regression analysis shows significant correlation of genome size ($R^2 = 0.258$; $P < 0.05$) and GC content ($R^2 = 0.055$; $P < 0.05$) with incidence of SSR. There was no significant correlation observed between genome size, relative abundance RA ($R^2 = 0.027$; $P > 0.05$) and relative density RD ($R^2 = 0.014$; $P > 0.05$). Similarly, GC content, relative abundance ($R^2 = 0.020$; $P > 0.05$) and relative density ($R^2 = 0.022$; $P > 0.05$) were also found to be non-significantly correlated (Figure 10.6 A-B). The regression analysis of cSSR revealed no significant correlation of incidence of cSSR with GC content ($R^2 = 0.015$; $P > 0.05$) and genome size ($R^2 = 0.031$; $P > 0.05$). Furthermore, genome size was also non-significantly correlated with cRA ($R^2 = 0.007$; $P > 0.05$), cRD ($R^2 = 0.007$; $P > 0.05$), percentage of cSSRs ($R^2 = 0.003$; $P > 0.05$). Similarly, no significant correlation was observed in case of cRA ($R^2 = 0.009$; $P > 0.05$) and cRD ($R^2 = 0.006$; $P > 0.05$) and percentage of cSSRs ($R^2 = 0.000$; $P > 0.05$) with GC content of cSSRs (Figure 10.7 A-B).

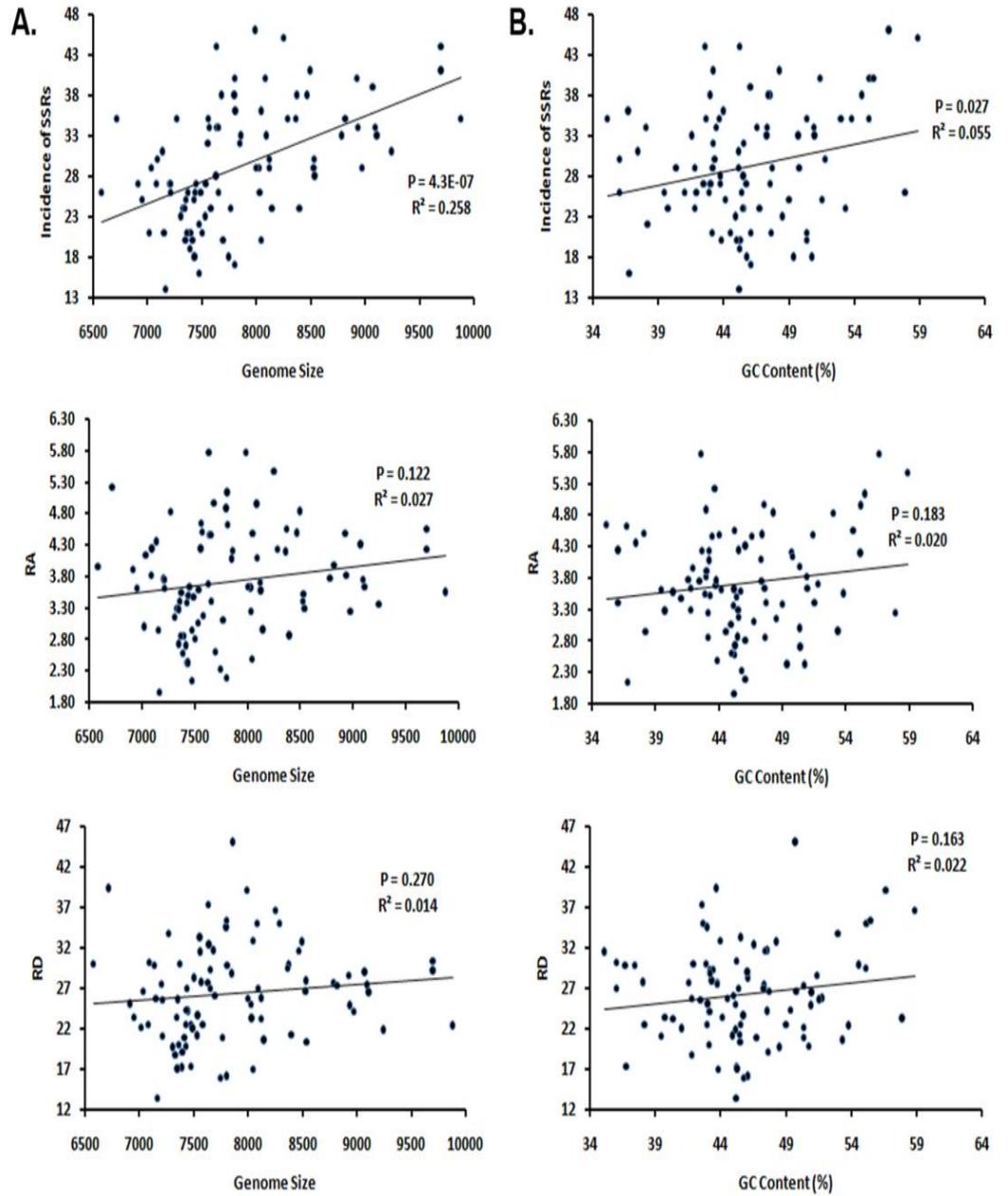


Figure 10.6 A-B: Regression analysis of genomic features, namely, the genome size (A) and GC content (B) with

Incidence, RA, and RD of SSRs.

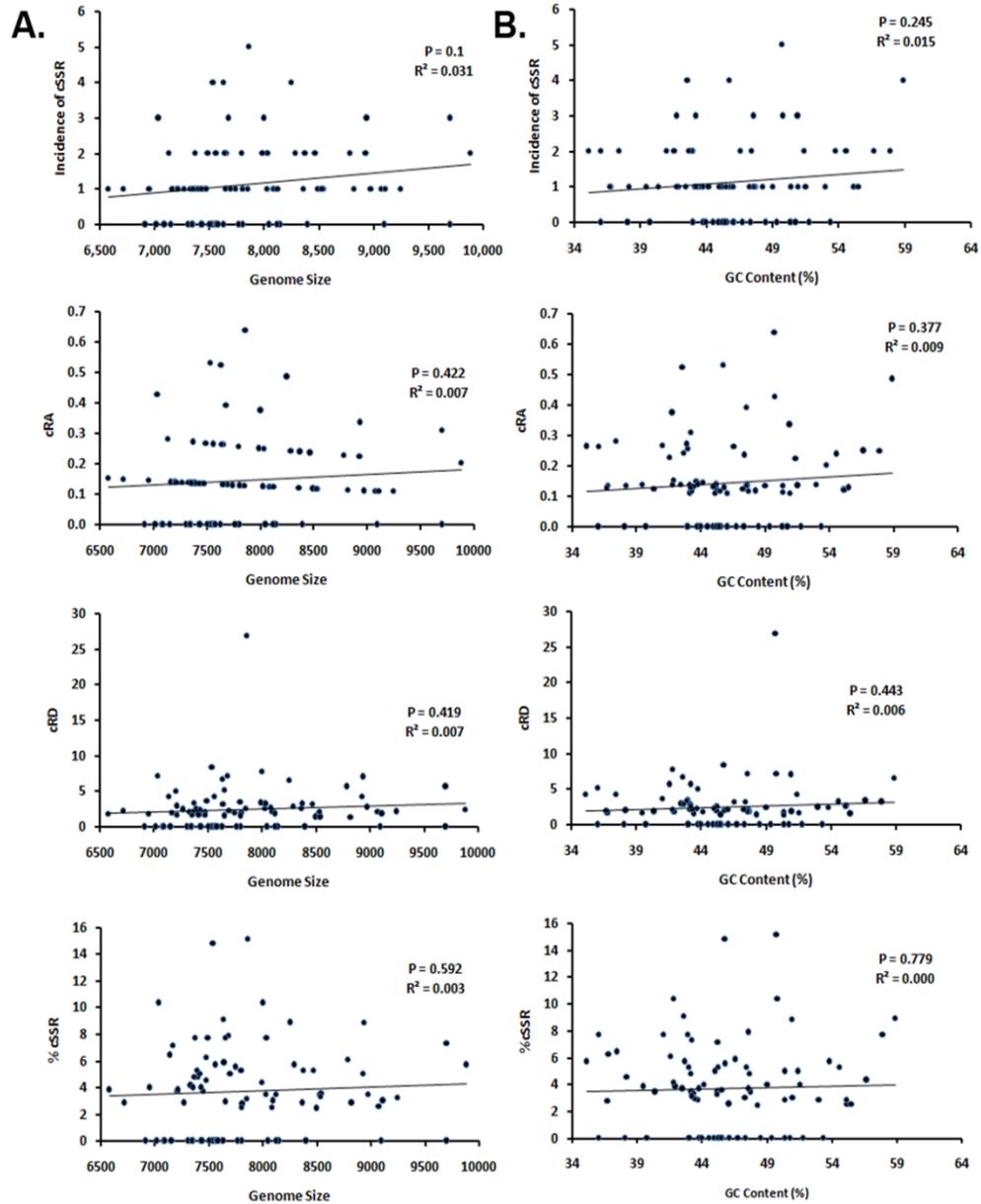


Figure 10.7 A-B: Regression analysis of genomic features, namely, the genome size (A) and GC content (B) with Incidence, RA, and RD of cSSRs.

10.3 Discussion

Picornaviridae, one of the largest RNA virus family's and consisting of the most diverse groups of species, is taken for study for the following reasons. Firstly, Picornaviruses have a broad host range, many of which cause serious health hazards to humans and animals. Secondly, these viruses bear higher mutability properties. The higher mutation leads to create polymorphism in the microsatellite. Thus, it can be an excellent tool to study population genetics, evolution, and immune evasion strategies. However, virtually no comprehensive study of microsatellite analysis has been done in any of these viruses. Therefore, the microsatellites' analysis and their properties provide an excellent resource for further study of virus biology in general. For this reason, we have analyzed 88 complete genome sequences of Picornaviruses. As genome size is relatively small, fewer microsatellite repeats (2488 SSRs and 100 cSSR) with incidence frequency between 14 to 46 were observed. This incidence of microsatellites was proportional to the genome size, as seen in the case of Tobamovirus, having a lower number of SSRs than potyviruses [17]. These observations come in line with the SSRs present in other viruses such as Flavivirus (27-67SSRs) [36] and HIV (22–48 SSRs) [13]. The RD of SSR varied from 13.39 bp/kb to 45.02 bp/kb on an average 26.29. Dinucleotide repeats were found to be more abundant than tri-nucleotide repeats in Picornaviruses. These indicate that dinucleotide repeats were more liable towards instability than the rest type of repeats, owing to a higher slippage rate. This type of diversification of motif might act as a molecular device for faster adaptation to environmental stresses in Picornavirus [32]. In viruses, microsatellites were mostly accumulated in

the coding regions, probably due to the high coding density of the viral genome [9]. In the higher strata, the presence of SSRs in protein-coding regions are associated with functions such as social behavior in voles, sporulation efficiency and cell adhesion in yeast, skeletal morphology in domestic dogs, adaptive divergence in barley and wheat populations [32]. The cSSR percentage of picornaviruses ranged from 0-15.15%, which is lower in comparison to HIV-1 (0-24.24%) [13], Geminivirus (0-27.27%) [33], Herpesvirus (8.12-33.31%) [34] and other eukaryotic genomes (4-25%) [14]. However, the picornaviruses cSSRs were in higher percentage compared to that of Flavivirus (0-12.82%) [36], Potexvirus (0-11.76%) [37], and prokaryotes such as *E. coli* (1.75-2.85%) [38]. It was observed that the cSSR percentage increased with an increase in dMAX. However, this increase was not uniform in the five species. This indicates microsatellite distribution in Picornavirus genomes probably has diverse patterns. Similar observations were also made in the genome of eukaryotes and prokaryotes [14]. Compound microsatellites present in Picornavirus have lower complexity as compared to the eukaryotes. The majority of cSSR consists of only two motifs. The largest cSSR was composed of three motifs. The prokaryotes show up to four cSSRs, while more than eight cSSRs are observed in eukaryotes. This indicates that the smaller genomes, the microsatellites are distributed far from one another. In general, the number of compound microsatellites decreases with an increase in complexity. Similar results have been demonstrated in HIV-1 genomes and *E. coli*, where smaller genome sizes and higher coding density probably were responsible for limiting the complexity of compound microsatellites. In Picornaviruses, 13.13% of

cSSRs were composed of similar motifs, probably the result of genome duplication. Some study suggests that genome duplication may be helpful for the repeat tendency mechanism, which promotes the expansion of genome size such as yeast [40, 41]. In the evolutionary process, the occurrence of repeated sequences having a harmful or even lethal effect on the host might have been eliminated as a result of negative selection. This notion very well supports the fact that genome sizes do not increase abruptly. On the other hand, the SSRs showing neutral effect could be selected and fixed by genetic drift. Some longer repeat sequences could bring new functions to the organism, which may be beneficial for the organism's survival and could have been positively selected over time. The instability within the repeats creates polymorphism, which could act as a double-edged sword to drive adaptive variation in the genome through insertion, deletion, and substitution mutation and evades the host immune system. We could very well propose that the polymorphism directly or indirectly is related to important phenomena such as the birth or death of microsatellite. Kelkar et al., 2008 [35] suggested that substitutions in short stretches of repeats be the predominant cause of births of microsatellites, whereas insertions and deletions lead to the death phase. Most of the microsatellites were found to be polymorphic owing to the higher mutability of Picornavirus. Each microsatellite type has its intrinsic properties, i.e., conserved microsatellites acting as a fingerprint in the viral genome to identify the viral strain. In contrast, polymorphic microsatellite can be of importance to study population genetic and complex evolution of the viruses.

10.4 Materials and methods

10.4.1 Genome sequences

The ICTV (2020) classification of the Picornaviridae family comprises 63 genera and 147 species (<http://www.ictvonline.org/virusTaxonomy>). In this study, we downloaded all 88 available complete genome sequences representing each species from GenBank (<http://www.ncbi.nlm.nih.gov>) (Table 10.1). To compare genomic sequences of different lengths, we calculated the relative density (RD) and relative abundance (RA) values. RD is defined as the total length (bp) contributed by each microsatellite per kb of sequence analyzed. In contrast, RA is the number of microsatellites present per kb of the genome (kb).

10.4.2 Microsatellite identification and investigation

The identification of perfect mono, di, tri, tetra, penta, hexa, and compound microsatellites, IMEx software, was used [30]. Microsatellite information was extracted by the ‘Advance-Mode’ of IMEx using the parameters previously utilized for RNA viruses [37, 38]. The parameters used are as follows: type of repeat: perfect; repeat size: all; minimum repeat number: 6, 3, 3, 3, 3, 3 for mono, di, tri, tetra, penta, and hexanucleotide repeats, respectively. The maximum distance allowed between any two SSRs (dMAX) was ten nucleotides. Other parameters were set as default. Compound microsatellites (cSSR) were not standardized to determine the original composition.

10.4.3 Estimation of the number of expected mononucleotide repeats

We randomly chose mono-nucleotide repeats to evaluate the over or under-represented motifs in the Picornavirus genome, compared the observed and expected number of microsatellites. The expected number of microsatellites composed of Mt (M is the motif of the microsatellite with repeat number of t, and its length is L) in a genome of length G was calculated using the formula given by de Wachter, 1981 [42]. The Z score indicates the statistical significance of the microsatellite representation [10].

$$\text{Exp (Mt)} = f (M)t [1- f (M)] [G^t (1- f (M)+2L)] \dots\dots\dots (1)$$

$$G^t = G-tL-2L+1 \dots\dots\dots (2)$$

$$Z = \frac{O-E}{\sqrt{E}} \dots\dots\dots (3)$$

Where Exp (Mt) represents the expected number of microsatellites, and f (M) represents the probability of microsatellites.

10.4.4 Multiple sequence alignment and identification of polymorphic cSSR

The genomes having cSSR only were considered for identifying polymorphic microsatellites and consensus motifs. Species having more than five strains were considered for this investigation. Sequences were first transferred to BioEdit version 7.2.5 software and aligned by the CLUSTAL W module with some manual editing [43]. Finally, the consensus

sequences were identified by the procedure followed by George et al., 2015 [33].

10.4.5 Recombination analysis

To elucidate the potential recombinant sequences and their parents (major and minor), the multiple sequence alignment files formed above were used as input for the RDP4 package [44]. RDP4 consists of seven recombination detection methods such as RDP [44], GENCONV [45] BOOTSCAN [46], MAXCHI [47], CHIMAERA [48], SiScan [49] and 3SEQ [50] used to identify recombinant strain within the alignment. A sequence is considered a potential recombinant only if detected as significant (with the p-value cutoff of 0.00001) by at least three methods stated above.

10.4.6 Statistical analysis

All statistical analysis was performed using Graphpad Prism version 5 (La Jolla, CA, USA). Linear regression was used to reveal the correlation between incidence, RA, RD of simple as well as compound microsatellites with genome size and GC content.

10.5 Conclusion

This study shows a clear picture of type, relative abundance, relative density, and ubiquitous distribution pattern of microsatellite in Picornaviruses. The higher level of microsatellite polymorphism among Picornaviruses could be an excellent resource to address not only population genetics and evolution related questions but also solve the complex mystery, such as birth or death of microsatellite. Breakpoints of these analyzed species are rich in dinucleotide repeats, which need further study to explore the relationship between

recombination hotspot and microsatellite in this family. Finally, we can postulate that in-depth experimental analysis of SSR and cSSR in the diverse viral genome will pave the way to reveal the complex biological features such as host-pathogen interaction and the emergence of new epidemic strains.

10.6 References

1. Flanagan, J.B., Petterson, R., Ambros, V., Hewlett, N., Baltimore, D. (1977), Covalent linkage of a protein to a defined nucleotide sequence at the 5'-terminus of virion and replicative intermediate RNAs of poliovirus. *Proceedings of the National Academy of Sciences*. 74: p. 961-965 (DOI: 10.1073/pnas.74.3.961).
2. Toyoda, H., Franco, D., Fujita, K., Paul, A.V., Wimmer, E. (2007), Replication of poliovirus requires binding of the poly (rC) binding protein to the cloverleaf as well as to the adjacent C-rich spacer sequence between the cloverleaf and the internal ribosomal entry site. *Journal of virology*. 81: p. 10017-10028 (DOI: 10.1128/JVI.00516-07).
3. Jacobson, S., Konings, D., Sarnow, P. (1993), Biochemical and genetic evidence for a pseudoknot structure at the 3'terminus of the poliovirus RNA genome and its role in viral RNA amplification. *Journal of virology*. 67: p. 2961-2971 (DOI: 10.1128/JVI.67.6.2961-2971.1993).
4. Jacobson, S., Konings, D., Sarnow, P. (1993), Biochemical and genetic evidence for a pseudoknot

- structure at the 3'terminus of the poliovirus RNA genome and its role in viral RNA amplification. *Journal of virology*. 67: p. 2961-2971 (DOI: 10.1128/JVI.67.6.2961-2971.1993).
5. Kitamura, N., Semler, B.L., Rothberg, P.G., Larsen, G.R., Adler, C.J., Dorner, A.J., Emini, E.A., Hanecak, R., Lee, J.J., van der Werf, S., Anderson, C.W., Wimmer, E. (1981), Primary structure, gene organization and polypeptide expression of poliovirus RNA. *Nature*. 291: p. 547-553 (DOI: 10.1038/291547a0).
 6. Caliguirri, L.A., Tamm, I. (1969), Membranous structures associated with translation and transcription of poliovirus RNA. *Science*. 166: p. 885-886 (DOI: 10.1126/science.166.3907.885).
 7. Agol, V.I. (2010), Picornaviruses as a model for studying the nature of RNA recombination, *The Picornaviruses*. American Society of Microbiology. pp. 239-252.
 8. Simmonds, P. (2010), Recombination in the evolution of picornaviruses, *The Picornaviruses*. American Society of Microbiology. pp. 229-237.
 9. Chen, M., Tan, Z., Jiang, J., Li, M., Chen, H., Shen, G., Yu, R. (2009), Similar distribution of simple sequence repeats in diverse completed Human Immunodeficiency Virus Type 1 genomes. *FEBS letters*. 583: p. 2959-2963 (DOI: 10.1016/j.febslet.2009.08.004).
 10. Mrázek, J., Guo, X., Shah, A. (2007), Simple sequence repeats in prokaryotic genomes. *Proceedings of the*

- National Academy of Sciences. 104: p. 8472-8477 (DOI: 10.1073/pnas.0702412104).
11. Tóth, G., Gáspári, Z., Jurka, J. (2000), Microsatellites in different eukaryotic genomes: survey and analysis. *Genome research*. 10: p. 967-981 (DOI: 10.1101/gr.10.7.967).
 12. Chambers, G.K., MacAvoy, E.S. (2000), Microsatellites: consensus and controversy. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*. 126: p. 455-476 (DOI: 10.1016/s0305-0491(00)00233-9).
 13. Chen, M., Tan, Z., Zeng, G., Zeng, Z. (2012), Differential distribution of compound microsatellites in various human immunodeficiency virus type 1 complete genomes. *Infection, Genetics and Evolution*. 12: p. 1452-1457 (DOI: 10.1016/j.meegid.2012.05.006).
 14. Kofler, R., Schlötterer, C., Luschützky, E., Lelley, T. (2008), Survey of microsatellite clustering in eight fully sequenced species sheds light on the origin of compound microsatellites. *BMC genomics*. 9: p. 612 (DOI: 10.1186/1471-2164-9-612).
 15. Weber, J.L. (1990), Informativeness of human (dC-dA) n·(dG-dT) n polymorphisms. *Genomics*. 7: p. 524-530 (DOI: 10.1016/0888-7543(90)90195-z).
 16. George, B., Alam, C.M., Jain, S., Sharfuddin, C., Chakraborty, S. (2012) Differential distribution and occurrence of simple sequence repeats in diverse geminivirus genomes. *Virus genes* 45, 556-566. (DOI: 10.1007/s11262-012-0802-1)

17. Zhao, X., Tan, Z., Feng, H., Yang, R., Li, M., Jiang, J., Shen, G., Yu, R. (2011), Microsatellites in different Potyvirus genomes: survey and analysis. *Gene*. 488: p. 52-56 (DOI: 10.1016/j.gene.2011.08.016).
18. Li, Y.-C., Korol, A.B., Fahima, T., Nevo, E. (2004), Microsatellites within genes: structure, function, and evolution. *Molecular biology and evolution*. 21: p. 991-1007 (DOI: 10.1093/molbev/msh073).
19. Usdin, K. (2008), The biological effects of simple tandem repeats: lessons from the repeat expansion diseases. *Genome research*. 18: p. 1011-1019 (DOI: 10.1101/gr.070409.107).
20. Ellegren, H. (2004), Microsatellites: simple sequences with complex evolution. *Nature reviews genetics*. 5: p. 435 (DOI: 10.1038/nrg1348).
21. Jakupciak, J.P., Wells, R.D. (1999) Genetic instabilities in (CTG· CAG) repeats occur by recombination. *Journal of Biological Chemistry*. 274: p. 23468-23479. (DOI: 10.1074/jbc.274.33.23468).
22. Biet, E., Sun, J.-S., Dutreix, M. (1999), Conserved sequence preference in DNA binding among recombination proteins: an effect of ssDNA secondary structure. *Nucleic Acids Research*. 27: p. 596-600. (DOI: 10.1093/nar/27.2.596).
23. Kirkpatrick, D.T., Wang, Y.-H., Dominska, M., Griffith, J.D., Petes, T.D., (1999), Control of meiotic recombination and gene expression in yeast by a simple repetitive DNA sequence that excludes nucleosomes. *Molecular and cellular biology*. 19: p. 7661-7671 (DOI: 10.1128/mcb.19.11.7661).

24. Schultes, N.P., Szostak, J.W. (1991), A poly (dA. dT) tract is a component of the recombination initiation site at the ARG4 locus in *Saccharomyces cerevisiae*. *Molecular and cellular biology*. 11: p. 322-328 (DOI: 10.1128/mcb.11.1.322).
25. Bagshaw, A.T., Pitt, J.P., Gemmell, N.J. (2008), High frequency of microsatellites in *S. cerevisiae* meiotic recombination hotspots. *BMC genomics*. 9: p. 49 (DOI: 10.1186/1471-2164-9-49).
26. Duke, G.M., Osorio, J.E., Palmenberg, A.C. (1990), Attenuation of Mengo virus through genetic engineering of the 5' noncoding poly(C) tract. *Nature*. 343: p. 474-476 (DOI: 10.1038/343474a0).
27. Barr, J.N., Whelan, S., Wertz, G.W. (1997), cis-Acting signals involved in termination of vesicular stomatitis virus mRNA synthesis include the conserved AUAC and the U7 signal for polyadenylation. *Journal of virology* 71. P. 8718-8725 (DOI: 10.1128/JVI.71.11.8718-8725.1997).
28. Yamada, N., Tanihara, K., Takada, A., Yorihuzi, T., Tsutsumi, M., Shimomura, H., Tsuji, T., Date, T. (1996), Genetic organization and diversity of the 3' noncoding region of the hepatitis C virus genome. *Virology*. 223: p. 255-261 (DOI: 10.1006/viro.1996.0476).
29. Garcia-Barreno, B., Delgado, T., Melero, J.A. (1994), Oligo (A) sequences of human respiratory syncytial virus G protein gene: assessment of their genetic stability in frameshift mutants. *Journal of virology*. 68: p. 5460-5468 (DOI: 10.1128/JVI.68.9.5460-5468.1994).

30. Mudunuri, S.B., Nagarajaram, H.A. (2007), IMEx: imperfect microsatellite extractor. *Bioinformatics*. 23: p. 1181-1187 (DOI: 10.1093/bioinformatics/btm097).
31. Katti, M.V., Ranjekar, P.K., Gupta, V.S. (2001), Differential distribution of simple sequence repeats in eukaryotic genome sequences. *Molecular biology and evolution*. 18: p. 1161-1167 (DOI: 10.1093/oxfordjournals.molbev.a003903).
32. Kashi, Y., King, D.G. (2006), Simple sequence repeats as advantageous mutators in evolution. *TRENDS in Genetics*. 22: p. 253-259 (DOI: 10.1016/j.tig.2006.03.005).
33. George, B., Alam, C.M., Kumar, R.V., Gnanasekaran, P., Chakraborty, S., 2015. Potential linkage between compound microsatellites and recombination in geminiviruses: Evidence from comparative analysis. *Virology*. 482: p. 41-50 (DOI: 10.1016/j.virol.2015.03.003).
34. Ouyang, Q., Zhao, X., Feng, H., Tian, Y., Li, D., Li, M., Tan, Z. (2012), High GC content of simple sequence repeats in Herpes simplex virus type 1 genome. *Gene*. 499: p. 37-40 (DOI: 10.1016/j.gene.2012.02.049).
35. Kelkar, Y.D., Tyekucheva, S., Chiaromonte, F., Makova, K.D. (2008), The genome-wide determinants of human and chimpanzee microsatellite evolution. *Genome research*. 18: p. 30-38 (DOI: 10.1101/gr.7113408).
36. Alam, C., Iqbal, A., Thadari, B., Ali, S. (2016), Imex based analysis of repeat sequences in flavivirus genomes, including dengue virus. *J Data Mining*

- Genomics Proteomics. 7: p. 2153-0602 (DOI:10.4172/2153-0602.1000187).
37. Alam, C.M., George, B., Sharfuddin, C., Jain, S., Chakraborty, S. (2013), Occurrence and analysis of imperfect microsatellites in diverse potyvirus genomes. *Gene*. 521: p. 238-244 (DOI: 10.1016/j.gene.2013.02.045).
 38. Chen, M., Zeng, G., Tan, Z., Jiang, M., Zhang, J., Zhang, C., Lu, L., Lin, Y., Peng, J. (2011), Compound microsatellites in complete *Escherichia coli* genomes. *FEBS letters*. 585: p. 1072-1076 (DOI 10.1016/j.febslet.2011.03.005).
 39. Kruglyak, S., Durrett, R., Schug, M.D., Aquadro, C.F. (2000), Distribution and abundance of microsatellites in the yeast genome can be explained by a balance between slippage events and point mutations. *Molecular Biology and Evolution*. 17: p. 1210-1219 (DOI: 10.1093/oxfordjournals.molbev.a026404).
 40. Fadda, Z., Daròs, J., Flores, R., Duran-Vila, N. (2003), Identification in eggplant of a variant of citrus exocortis viroid (CEVd) with a 96 nucleotide duplication in the right terminal region of the rod-like secondary structure. *Virus research*. 97: p. 145-149 (DOI: 10.1016/j.virusres.2003.08.002).
 41. Liti, G., Louis, E.J. (2005), Yeast evolution and comparative genomics. *Annu. Rev. Microbiol.* 59: p. 135-153 (DOI: 10.1146/annurev.micro.59.030804.121400).
 42. de Wachter, R. (1981), The number of repeats expected in random nucleic acid sequences and found

- in genes. *Journal of theoretical biology*. 91: p. 71-98 (DOI: 10.1016/0022-5193(81)90375-1).
43. Thompson, J.D., Gibson, T.J., Higgins, D.G. (2003), Multiple sequence alignment using ClustalW and ClustalX. *Current protocols in bioinformatics* 2.3: p. 1-2 (DOI: 10.1002/0471250953.bi0203s00).
 44. Martin, D., Murrell, B., Golden, M., Khoosal, A., Muhire, B. (2015), RDP4: detection and analysis of recombination patterns in virus genomes. *Virus Evol.* 1: vev003. (DOI: 10.1093/ve/vev003).
 45. Padidam, M., Sawyer, S., Fauquet, C.M. (1999), Possible emergence of new geminiviruses by frequent recombination. *Virology*. 265: p. 218-225 (DOI: 10.1006/viro.1999.0056).
 46. Martin, D., Posada, D., Crandall, K., Williamson, C. (2005), A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Research & Human Retroviruses*. 21: p. 98-102 (DOI: 10.1089/aid.2005.21.98).
 47. Smith, J.M. (1992), Analyzing the mosaic structure of genes. *Journal of molecular evolution*. 34: p. 126-129 (DOI: 10.1007/BF00182389).
 48. Posada, D., Crandall, K.A. (2001), Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proceedings of the National Academy of Sciences*. 98: p. 13757-13762 (DOI: 10.1073/pnas.241370698).
 49. Gibbs, M.J., Armstrong, J.S., Gibbs, A.J. (2000), Sister-scanning: a Monte Carlo procedure for assessing

- signals in recombinant sequences. *Bioinformatics*. 16: p. 573-582 (DOI: 10.1093/bioinformatics/16.7.573).
50. Boni, M.F., Posada, D., Feldman, M.W. (2007), An exact nonparametric method for inferring mosaic structure in sequence triplets. *Genetics*. 176: p. 1035-1047 (DOI: 10.1534/genetics.106.068874).

Chapter 11

Conclusion and scope for future work

11.1 Conclusion drawn from the present work

One health approach covers a wide range of viral diseases that have zoonotic potential as well as seriously impact the livelihood and health of companion animals. To address those problems, epidemiology plays a crucial role in deciphering the distribution, pattern of evolution, and transmission of diseases in populations. Virus transmission is corroborated by host behavior, ecological, and environmental factors. Viral epidemiology aims to study these factors and help to provide an effective direction to control and prevention strategies by investigating outbreak sources. During the outbreak investigation one can clarify the etiology of viral diseases, interaction of viruses with the environment, host susceptibility, modes of transmission, and the evaluation of vaccines and therapeutics on a large scale. For example, the recent pandemic of COVID-19 caused by SARS-CoV-2 affects millions of human beings throughout the world, with hundreds of thousands of fatalities. Due to ease of communication around the globe, this virus has spread all over the world from its epicenter, i.e., Wuhan city of China. Epidemiologist usually implements two approaches such as serological epidemiology and molecular epidemiology during the investigation of a virus outbreak. In serological epidemiology they target morbidity, mortality, age-specific prevalence patterns, possible transmission routes, and a new infections natural reservoir. Molecular epidemiology includes advances in rapid sequencing techniques that have allowed

comparison of virus genome sequences to answer many epidemiological questions such as geographical origin; reservoir animal species, recovery of the virus, the emergence of a new strain, possible origins, and the sensitivity of a new virus isolate to antiviral drugs, and monitor the spread of drug-resistant viruses in the community.

This chapter elucidated the epidemiological characteristics of a zoonotic Orf virus (ORFV), the causative agent of contagious ecthyma disease that primarily occurs within small ruminants and animal handlers. We surveyed the fifteen outbreak locations of eastern India during this study and collected clinical samples from infected goats. Total DNA was extracted and screened with the universal PCR primers to confirm the respective virus. Further amplification using gene-specific primers targeting fragments of ORFV011, ORFV020, ORFV059, ORFV108 genes, followed by sequencing and GenBank submission to get accession numbers. The comparative genomics approach was deployed to construct the gene-specific phylogenetic tree and evolutionary parameters of this virus, suggesting that the viral strain is closely related to Chinese isolates. The evolutionary study revealed the purifying selection is maintaining the heterogeneity within the viral strain.

Genome-wide in-silico investigation of microsatellites markers within the ORFV genome has been accomplished to decipher the type, distribution, and potential role in the genome evolution, during this study. We have scrutinized eleven ORFV strains during this study, which showed the existence of 1,036–1,181 microsatellites or Simple Sequence Repeats (SSRs) per strain. Further analysis deciphered

residence of 83–107 compound SSRs (cSSRs) per genome. Almost all compound microsatellites were made up of two SSRs, which showed unbiased to a particular motif. Motif duplication pattern, like, (C)-x-(C), (TG)-x-(TG), (AT)-x-(AT), (TC)- x-(TC) and self-complementary cSSRs, such as (GC)-x-(CG), (TC)-x-(AG), (GT)-x-(CA) and (TC)-x-(AG) were noticed during the study. Finally, in-silico polymorphism was evaluated, followed by in-vitro validation utilizing PCR and sequencing analysis. The thirteen polymorphic microsatellite markers developed during this investigation were characterized by mapping with the GenBank database. The outcome of this study suggests that these microsatellites could be a tool for strain demarcation, genetic diversity estimation, and evolutionary analysis of the virus.

We have explored the 2017 ORFV outbreak within goats in Madhya Pradesh, a central India state, during this study. The outbreak was distinguished by a moderate morbidity rate (up to 20%) with no mortality. Phylogenetic analysis deciphered the transboundary perspective of the virus by indicating its correlation with a distinct geographical location. We elucidated this viral strain's complete genome, named Ind/MP, and carry out a comparative genomic analysis. The genome was made up of 132 open reading frames (ORFs) flanked by inverted terminal repeats (ITRs) of 3,910 bp at both ends with terminal BamHI sites and conserved telomere resolution sequences. Population genetic indices like nucleotide diversity (π), selection pressure analysis ($\theta=dN/dS$), etc., indicate that the ORFV inhabit under purifying selection. A sum of forty recombination incidents was recognized. Ind/MP strain actively contributed to twenty-one events indicating this strain's potential for recombination to generate new variants.

In this chapter, we described the surveillance report of outbreaks of avipoxvirus (APV) infected domestic chickens and pigeons belongs to the Eastern Indian state of Odisha during the years 2010–2018. Analyzing typical lesions over the body, followed by molecular techniques, the overall morbidity was observed 18%–19.23% and 16.92%–23% in chickens and pigeons, respectively. PCR amplicons were intending the viral P4b core protein-coding gene and the DNA polymerase gene confirmed APV strains within ten birds. Phylogenetic analysis of two genes evident the circulating strains were members of APV clade A. The subclade analysis deciphered the introduction of A1 and A3 subclades in Indian chickens and pigeons, respectively. This study reveals APVs' presence in Eastern Indian birds (Odisha) and the polymerase gene's utilization for the first time to elucidate the spreading clades of Indian APVs.

In this chapter, we conducted a genome-wide analysis to decipher the type, distribution pattern of eight complete genomes derived from the Avipox virus genus, the causative agent of pox like lesions above 300 avian species and one of the major diseases for the extinction of endangered avian species as well a severe economic threat to livelihood. The *in-silico* screening deciphered the existence of 1531-2473 SSRs per strain. In the case of compound SSRs (cSSRs), the value opted for 83-107 per genome. The analysis of motif composition of cSSR revealed that most of the compound microsatellite was made up of two microsatellites, with some unique duplicated pattern of motif like, (TA)-x-(TA), (TCA)-x-(TCA), etc. and self-complementary motifs, as (TA)-x-(AT). Finally, we have validated forty sets of compound microsatellite markers through the in-vitro approach utilizing

clinical specimens, followed by mapping the sequencing product with the database through comparative genomics approaches. The microsatellites developed during this study could be as a multiplex panel for strain demarcation, genetic diversity estimation, and evolutionary analysis of the virus.

For the first time, we have deciphered the complete genome sequence of a fowlpox virus of Indian origin through a next-generation sequencing platform and utilize a comparative genomics approach to solve more about this virus biology. The NextSeq 500 NGS generated a total of ~2.16 GB of quality data with 7,303,963 numbers of quality reads. The previously published complete genome isolate fowlpox virus (Acc. No- AF198100) was used as the reference genome, to which the sequence generated through NGS mapped and assembled. The assembled genome of this novel virus exhibited 260,066 bp in length with a lower percentage of GC content (28.5) in comparison to fowlpox virus (30.83), shearwater pox virus (30.23), canarypox virus (30.37), and turkeypox virus (29.78). This information would help to know more about this virus biology and to develop effective vaccine candidates.

For the first time, we also deciphered the complete genome sequence of a pigeon pox virus of Indian origin through a next-generation sequencing platform and utilize a comparative genomics approach to solve more about this virus biology. The NextSeq 500 NGS generated a total of ~2.16 GB of quality data with 7,303,963 numbers of qualities reads. The previously published complete genome isolate fowlpox virus (Acc. No- NC_002188) was used as the reference genome, to which the sequence generated through NGS mapped and

assembled. The assembled genome of this novel virus exhibited 250048 bp in length. A lower percentage of GC content (29.5) was observed in this newly assembled genome, in comparison to fowlpox virus (30.83), shearwater pox virus (30.23), canarypox virus (30.37), and turkeypox virus (29.78) but similar to other APV such as pigeon pox virus, flamingo pox virus, magpiepox virus and penguinpox virus. The extreme left nucleotide was considered base 1, and the beginning region of the Inverted Terminal Repeats (ITRs), which consists of a total length of 4688 bp. The ITR spanned throughout PPV-001 to PPV-004a and PPV-256 to PPV-259. The genome information would pave the way to develop effective vaccine candidates for this virus.

Genome-wide identification of simple sequence repeats (SSRs) of Picornaviruses was carried out to investigate type, distribution, and potential role in genome evolution. Investigation on 88 Picornavirus species revealed the presence of 2,488 SSRs and 100 compound SSRs. Duplication pattern of motifs, like (C)-x-(C), (TG)-x-(TG), etc., and self-complementary motifs, such as (GC)-x-(CG), (TC)-x-(AG), etc. were observed in cSSR. Polymorphism analysis revealed that most of the cSSR were prone to instability, followed by consensus motifs. Finally, recombination analysis revealed that the breakpoints were rich in dinucleotide repeat, especially GT. However, further experimental validation is needed to elucidate the correlation between recombination hotspots and microsatellites.

11.2 Future scope of the present work

ORFVs have a wide host range with zoonotic potential accompanied with threat of serious economic loss to animal

husbandry globally. Although its presence is well documents in several reports, still a detailed information is lacking from India. In this study, we have investigated several Orf outbreaks followed by comparative genomics to decipher the population genetics of this virus. Then we have developed molecular markers as a diagnostic tool, which can be implemented for further study. Our preliminary results can be considered as a useful tool in the study of ORFV strain demarcation, diversity estimation, and evolutionary analysis. Finally, we have analyzed the complete genome of this virus directly from the clinical samples for the first time and found recombination modulating virus heterogeneity. We hope that the current genomic information would be greatly useful for further understanding of ORFV biology, epidemiology, and research carried in front of diagnosis and vaccine development.

Food security is always considered as a global concern. Livelihood based on poultry acts as a major portion of agricultural and economic importance around the world. Since APVs have a wide host range, including chicken, it is difficult to develop an effective preventive strategy because detailed knowledge of viral isolates is lacking. The outbreak investigation and phylogenetic analysis provide the emergence of a new subclade, suggest this virus evolves rapidly. During this study, the markers developed can be used as suitable multiplex PCR panels to solve virus genotype and evolutionary status by using geographically diverse clinical samples. The complete genome of the Indian fowlpox virus

could help to the understanding of FPV epidemiology and effective vaccine development.

The pigeon is considered as one of the companion animals. Pigeon pox virus usually infects the pigeon and declines the species. To decipher this virus biology, we report the first complete genome of PPV isolates from India. The comparative genomics deciphered the presence of recombination events within Avipox viruses. The current genomic resource would be greatly useful for further understanding of PPV biology. This study shows a clear picture of type, relative abundance, relative density, and ubiquitous distribution pattern of microsatellite in picornaviruses. The higher level of microsatellite polymorphism among picornaviruses could be an excellent resource to address not only population genetics and evolution related questions but also solve the complex mystery, such as the birth or death of microsatellite. Breakpoints of these analyzed species are rich in dinucleotide repeats, which need further study to explore the relationship between recombination hotspot and microsatellite in this family. Finally, we can postulate that in-depth experimental analysis of SSR and cSSR in the diverse viral genome will pave the way to reveal the complex biological features such as host-pathogen interaction and the emergence of new epidemic strains.

Annexure A: Published isolates of ORFV used for phylogenetic and population genetic analysis.

Sr. No.	Isolates name	Gene name	Animal species	Country	GenBank accession no	Reference
1	India 82/04	ORFV 011	goat	India	DQ263303	Hosamani,M.
2	India 59/05	ORFV 011	goat	India	DQ263304	Hosamani,M.
3	India 67/04	ORFV 011	sheep	India	DQ263305	Hosamani,M.
4	India 79/04	ORFV 011	sheep	India	DQ263306	Hosamani,M.
5	Nantou	ORFV 011	goat	Taiwan	DQ904351	Chan,K.W.
6	Taiping	ORFV 011	goat	Taiwan	EU327506	Chan,K.W.
7	Hoping	ORFV 011	goat	Taiwan	EU935106	Chan,K.-W.
8	Muk/2000	ORFV 011	goat	India	HM466933	Venkatesan, G.
9	ORFV Assam/09	ORFV 011	Capra hircus	India	JN846834	Bora,D.P.
10	Assam/10	ORFV 011	Capra hircus	India	JQ040300	Bora,D.P.
11	CIRG/2013	ORFV 011	Capra hircus	India	KF953485	Kumar,N.
12	Jaganathpara/Tripura/14	ORFV 011	Capra hircus	India	KT935588	Venkatesan, G.
13	Baulia Basti/Tripura/14	ORFV 011	Capra hircus	India	KT935589	Venkatesan, G.
14	Basudebpara/Tripura/14	ORFV 011	Capra hircus	India	KT935590	Venkatesan, G.
15	Orf/As-Hj/13G	ORFV 011	Capra hircus	India	KU128530	Venkatesan, G.
16	Orf/As-By/17G	ORFV 011	Capra hircus	India	KU128531	Venkatesan, G.
17	Orf/As-Lk/24G	ORFV 011	Capra hircus "	India	KU128532	Venkatesan, G.
18	Orf/As-Lk/41G	ORFV 011	Capra hircus	India	KU128533	Venkatesan, G.
19	Orf/Dri-As/63G	ORFV 011	Capra hircus	India	KU128534	Venkatesan, G.
20	Orf/As-Sag/8G	ORFV 011	Capra hircus	India	KU128535	Venkatesan, G.
21	ORf/Mg-By/5G	ORFV 011	Capra hircus	India	KU128536	Venkatesan, G.
22	Orf/As-Lk/32G	ORFV 011	Capra hircus	India	KU128537	Venkatesan, G.
23	Orf/Mg-By/1G	ORFV 011	Capra hircus	India	KU128538	Venkatesan, G.
24	Orf/As-Lk/15G	ORFV 011	Capra hircus	India	KU128539	Venkatesan, G.
25	Orf/As-Nb/2G	ORFV 011	Capra hircus	India	KU128540	Venkatesan, G.

26	orf/Kodaikanal-1	ORFV 011	sheep	India	KU597 728	Sumana,K.
27	orf/Kodaikanal-2	ORFV 011	sheep	India	KU597 729	Sumana,K.
28	orf/Kodaikanal-3	ORFV 011	sheep	India	KU597 730	Sumana,K.
29	Chitradurgar/Goat- 20	ORFV 011	goat	India	KU597 734	Sumana,K.
30	ORFV/Goat/India/ ICAR Goa	ORFV 011	goat	India	KX129 982	Shivasharan appa,N.
31	South Chandrapur/Tripur a/2014	ORFV 011	Capra hircus	India	KX377 974	Venkatesan, G.
32	ORFV Mukteswar/09	ORFV 011	sheep	India	GU139 356	Venkatesan, G.
33	CIRG/2014	ORFV 011	Goat	India	KP745 467	Kumar,N.
34	Jilin	ORFV 011	sheep	China	FJ8080 74	Zhao,K.
35	South Korea	ORFV 011	goat	Korea	GQ328 006	Oem,J.K.
36	ORFV/HuB/2009/ China	ORFV 011	goat	China	GU320 351	Zhang,K.S.
37	NZ2	ORFV 011	Ovis aries	China	GU903 501	Liu,Y.
38	Shanxi	ORFV 011	goat	China	HQ202 153	Shi,X.T.
39	ORFV/GanSu/200 9/China	ORFV 011	sheep	China	HQ694 772	Cai,X.
40	ORFV/LiaoNing/2 010/China	ORFV 011	goat	China	HQ694 773	Zhang,H.
41	ORFV/XinJiang/2 011/China	ORFV 011	sheep	China	JN565 694	Li,H.
42	ORFV/ShanXi/201 1/China	ORFV 011	goat	China	JN565 696	Liu,Z.
43	NA1/11	ORFV 011	sheep	China	JQ619 903	Li,W.
44	HuB/XN	ORFV 011	goat	China	JQ904 786	Zhang,K.
45	AnH/FD	ORFV 011	goat	China	JQ904 787	Zhang,K.
46	YN/JS	ORFV 011	goat	China	JQ904 788	Zhang,K.
47	China Vaccine	ORFV 011	goat	China	JQ904 789	Zhang,K.
48	HuB/XN 2	ORFV 011	goat	China	JQ904 790	Zhang,K.
49	JS/FX	ORFV 011	goat	China	JQ904 791	Zhang,K.
50	SC/JY	ORFV 011	goat	China	JQ904 792	Zhang,K.
51	GX/YB	ORFV 011	goat	China	JQ904 793	Zhang,K.
52	SD/DY	ORFV 011	sheep	China	JQ904 794	Zhang,K.
53	JL/TL	ORFV 011	sheep	China	JQ904 795	Zhang,K.
54	SC/NC	ORFV 011	goat	China	JQ904 796	Zhang,K.
55	CQ/WZ	ORFV 011	goat	China	JQ904 797	Zhang,K.

56	SC/YT	ORFV 011	goat	China	JQ904 798	Zhang,K.
57	NX/YC	ORFV 011	sheep	China	JQ904 799	Zhang,K.
58	ORF/2010/B2L-1	ORFV 011	goat	South Korea	JX968 988	Oem,J.-K.
59	ORF/2010/B2L-2	ORFV 011	goat	South Korea	JX968 989	Oem,J.-K.
60	ORF/2010/B2L-3	ORFV 011	goat	South Korea	JX968 990	Oem,J.-K.
61	ORF/2011/B2L	ORFV 011	goat	South Korea	JX968 991	Oem,J.-K.
62	ORF/2011/7-18- B2L	ORFV 011	goat	South Korea	JX968 992	Oem,J.-K.
63	Gansu	ORFV 011	Hom o sapie ns	China	KC485 343	Zhang,K.
64	OV-HN3/12	ORFV 011	sheep	China	KC569 750	Li,W.
65	Xinjiang1	ORFV 011	goat	China	KF666 560	Yang,H.
66	Xinjiang2	ORFV 011	goat	China	KF666 565	Yang,H.
67	Xinjiang	ORFV 011	goat	China	KF703 747	Yang,H.
68	YL-3	ORFV 011	goat	China	KF772 211	Liu,F.
69	SX1	ORFV 011	goat	China	KJ139 956	Shang,C.
70	Hub13	ORFV 011	goat	China	KJ139 994	Chi,X.
71	ORFV/Shaanxi/20 15/China	ORFV 011	goat	China	KU194 469	Fu,M.Z.
72	FJ-BF2015	ORFV 011	goat	China	KU199 823	Lin,Y.S.
73	FJ-FZ2014	ORFV 011	goat	China	KU199 824	Lin,Y.S.
74	FJ-FZ2015	ORFV 011	goat	China	KU199 825	Lin,Y.S.
75	FJ-JA2015	ORFV 011	goat	China	KU199 826	Lin,Y.S.
76	FJ-LJ2014	ORFV 011	goat	China	KU199 827	Lin,Y.S.
77	FJ-LJ2015	ORFV 011	goat	China	KU199 828	Lin,Y.S.
78	FJ-LY2014	ORFV 011	goat	China	KU199 829	Lin,Y.S.
79	FJ-MH2015	ORFV 011	goat	China	KU199 830	Lin,Y.S.
80	FJ-YT2014"	ORFV 011	goat	China	KU199 831	Lin,Y.S.
81	FJ-YT2015	ORFV 011	goat	China	KU199 832	Lin,Y.S.
82	OV/HLJ/04	ORFV 011	goat	China	KU523 790	Lin,Y.S.
83	Chaling	ORFV 011	black goat	China	KU851 936	Wang,B.
84	Guangdong	ORFV 011	black goat	China	KU884 328	Wang,B.

85	Hengyang	ORFV 011	black goat	China	KU884 329	Wang,B.
86	JX	ORFV 011	black goat	China	KU976 385	Wang,B.
87	LYJ	ORFV 011	black goat	China	KU976 387	Wang,B.
88	LYZ	ORFV 011	black goat	China	KU976 388	Wang,B.
89	SY	ORFV 011	black goat	China	KU976 389	Wang,B.
90	YZ	ORFV 011	black goat	China	KU976 390	Wang,B.
91	ZZ	ORFV 011	black goat	China	KU976 391	Wang,B.
92	NX	ORFV 011	black goat	China	KU976 392	Wang,B.
93	LD	ORFV 011	black goat	China	KU976 393	Wang,B.
94	GY-AHF10	ORFV 011	Capra hircus (goat)	China	KX029 228	Wang,B.
95	Al-Diwaniya.1	ORFV 011	camel	Iraq	KP998 114	Abd Al- Ameer
96	Al-Diwaniya.2	ORFV 011	camel	Iraq	KP998 115	Abd Al- Ameer
97	AL-Najaf.1	ORFV 011	camel	Iraq	KP998 118	Abd Al- Ameer
98	AL-Najaf.2	ORFV 011	camel	Iraq	KP998 119	Abd Al- Ameer
99	AL-Najaf.3	ORFV 011	camel	Iraq	KP998 120	Abd Al- Ameer
10 0	AL-Samawa.2	ORFV 011	camel	Iraq	KP998 122	Abd Al- Ameer
10 1	AL-Samawa.3	ORFV 011	camel	Iraq	KP998 123	Abd Al- Ameer
10 2	ORF0611TNG	ORFV 011	goat	USA	KJ137 679	Velazquez- Salinas
10 3	ORF1205WAG	ORFV 011	goat	USA	KJ137 680	Velazquez- Salinas,L.
10 4	ORF0403ING	ORFV 011	goat	USA	KJ137 681	Velazquez- Salinas,L.
10 5	ORF0903GAG	ORFV 011	goat	USA	KJ137 682	Velazquez- Salinas,L.
10 6	ORF0903GAG1	ORFV 011	goat	USA	KJ137 683	Velazquez- Salinas,L.
10 7	ORF0507VRS3	ORFV 011	goat	USA	KJ137 684	Velazquez- Salinas,L.
10 8	ORF0507VRS2	ORFV 011	goat	USA	KJ137 685	Velazquez- Salinas,L.
10 9	ORF1107CHG	ORFV 011	goat	USA	KJ137 686	Velazquez- Salinas,L.
11 0	ORF1108EMS	ORFV 011	sheep	USA	KJ137 687	Velazquez- Salinas,L.
11 1	ORF0908SNG1	ORFV 011	goat	USA	KJ137 688	Velazquez- Salinas,L.
11 2	ORF1210GRG1	ORFV 011	goat	USA	KJ137 689	Velazquez- Salinas,L.
11 3	ORF1010CMS1	ORFV 011	sheep	USA	KJ137 690	Velazquez- Salinas,L.
11 4	ORF1010OAS1	ORFV 011	sheep	USA	KJ137 691	Velazquez- Salinas,L.

11 5	ORF1011GRG1	ORFV 011	goat	USA	KJ137 692	Velazquez- Salinas,L.
11 6	ORF1111GRG1	ORFV 011	goat	USA	KJ137 693	Velazquez- Salinas,L.
11 7	ORF1111SNS1	ORFV 011	sheep	USA	KJ137 694	Velazquez- Salinas,L.
11 8	ORF0907CPS1	ORFV 011	sheep	USA	KJ137 695	Velazquez- Salinas,L.
11 9	ORF0907CPS2	ORFV 011	sheep	USA	KJ137 696	Velazquez- Salinas,L.
12 0	ORF0208CAS	ORFV 011	sheep	USA	KJ137 697	Velazquez- Salinas,L.
12 1	ORF1008OAS	ORFV 011	sheep	USA	KJ137 698	Velazquez- Salinas,L.
12 2	ORF0408VRS	ORFV 011	sheep	USA	KJ137 699	Velazquez- Salinas,L.
12 3	ORF1010OAS2	ORFV 011	sheep	USA	KJ137 700	Velazquez- Salinas,L.
12 4	ORF0810SOS	ORFV 011	sheep	USA	KJ137 701	Velazquez- Salinas,L.
12 5	ORF0810CMS	ORFV 011	sheep	USA	KJ137 702	Velazquez- Salinas,L.
12 6	ORF0810CMS	ORFV 011	sheep	USA	KJ137 703	Velazquez- Salinas,L.
12 7	ORF1108OAS	ORFV 011	sheep	USA	KJ137 704	Velazquez- Salinas,L.
12 8	ORF0408PBS	ORFV 011	sheep	USA	KJ137 705	Velazquez- Salinas,L.
12 9	ORF0806AKG	ORFV 011	goat	USA	KJ137 706	Velazquez- Salinas,L.
13 0	ORF0604WVG	ORFV 011	goat	USA	KJ137 707	Velazquez- Salinas,L.
13 1	ORF0604KYG	ORFV 011	goat	USA	KJ137 708	Velazquez- Salinas,L.
13 2	ORF1007PRG	ORFV 011	goat	USA	KJ137 709	Velazquez- Salinas,L.
13 3	ORF0406NMG	ORFV 011	goat	USA	KJ137 710	Velazquez- Salinas,L.
13 4	ORF1209MSG	ORFV 011	goat	USA	KJ137 711	Velazquez- Salinas,L.
13 5	ORF0403TNG	ORFV 011	goat	USA	KJ137 712	Velazquez- Salinas,L.
13 6	ORF0603VAG	ORFV 011	goat	USA	KJ137 713	Velazquez- Salinas,L.
13 7	ORF0404ARG	ORFV 011	goat	USA	KJ137 714	Velazquez- Salinas,L.
13 8	ORF0806NYG	ORFV 011	goat	USA	KJ137 715	Velazquez- Salinas,L.
13 9	UY01/07	ORFV 011	sheep	Uruguay	KP728 916	Olivero,N.
14 0	UY02/07	ORFV 011	sheep	Uruguay	KP728 917	Olivero,N.
14 1	UY03/07	ORFV 011	sheep	Uruguay	KP728 918	Olivero,N.
14 2	UY05/04	ORFV 011	sheep	Uruguay	KP728 919	Olivero,N.
14 3	UY07/79	ORFV 011	sheep	Uruguay	KP728 920	Olivero,N.
14 4	UY08/79	ORFV 011	sheep	Uruguay	KP728 921	Olivero,N.

145	UY09/79	ORFV 011	sheep	Uruguay	KP728 922	Olivero,N.
146	UY11/07	ORFV 011	sheep	Uruguay	KP728 923	Olivero,N.
147	UY12/07	ORFV 011	sheep	Uruguay	KP728 924	Olivero,N.
148	UY13/08	ORFV 011	sheep	Uruguay	KP728 925	Olivero,N.
149	UY14/07	ORFV 011	sheep	Uruguay	KP728 926	Olivero,N.
150	UY15/09	ORFV 011	sheep	Uruguay	KP728 927	Olivero,N.
151	UY16/09	ORFV 011	sheep	Uruguay	KP728 928	Olivero,N.
152	UY17/09	ORFV 011	sheep	Uruguay	KP728 929	Olivero,N.
153	UY18a/09	ORFV 011	sheep	Uruguay	KP728 930	Olivero,N.
154	UY18b/09	ORFV 011	sheep	Uruguay	KP728 931	Olivero,N.
155	UY19/10	ORFV 011	sheep	Uruguay	KP728 932	Olivero,N.
156	UY20/10	ORFV 011	sheep	Uruguay	KP728 933	Olivero,N.
157	UY21/10	ORFV 011	sheep	Uruguay	KP728 934	Olivero,N.
158	UY23/11	ORFV 011	sheep	Uruguay	KP728 935	Olivero,N.
159	UY24/11	ORFV 011	sheep	Uruguay	KP728 936	Olivero,N.
160	NE2	ORFV 011	goat	Brazil	JN088 051	Abrahao,J.S.
161	D	ORFV 011	sheep	Brazil	JN088 052	Abrahao,J.S.
162	MT-05	ORFV 011	sheep	Brazil	JN613 809	Abrahao,J.S.
163	PA11	ORFV 011	Capra hircus	Brazil	JQ349 520	de Oliveira
164	NE1	ORFV 011	goat	Brazil	JN613 810	Abrahao,J.S.
1	FJ-JA2015 VIR	ORFV 20	goat	China	KU199 846	Lin,Y.S.
2	FJ-BF2015 VIR	ORFV 20	goat	China	KU199 843	Lin,Y.S.
3	FJ-LY2014 VIR	ORFV 20	goat	China	KU199 852	Lin,Y.S.
4	FJ-YT2015 VIR	ORFV 20	goat	China	KU199 851	Lin,Y.S.
5	FJ-YT2015 VIR	ORFV 20	goat	China	KU672 683	Lin,Y.S.
6	FJ-FZ2014 VIR	ORFV 20	goat	China	KU199 844	Lin,Y.S.
7	FJ-LJ2014 VIR	ORFV 20	goat	China	KU199 847	Lin,Y.S.
8	TVCC/Shuhama-02	ORFV 20	Sheep	India	KU672 679	Ahanger,S.A
9	FJ-MH2015	ORFV 20	goat	China	KU199 849	Lin,Y.S.
10	SX3	ORFV 20	goat	China	KJ139 958	Shang,C.

11	orfv-chGanSu	ORFV 20	goat	China	KF916 586	Cheng,W.
12	Nantou	ORFV 20		Taiwan	EU327 507	Chan,K.W.
13	TVCC/Shuhama- 01/Goat	ORFV 20	goat	India	KU672 682	Ahanger,S.A.
14	FJ-YT2014	ORFV 20	goat	China	KU199 850	Lin,Y.S.
15	ORFV/ShanX/201 1/China	ORFV 20	goat	China	JN565 697	Liu,Z.
16	FJ-FZ2015	ORFV 20	goat	China	KU199 845	Lin,Y.S.
17	Xinjiang2	ORFV 20	goat	China	KF666 564	Yang,H.
18	Taiping	ORFV 20	goat	China	EU327 508	Chan,K.W.
19	FJ-LJ2015	ORFV 20	goat	China	KU199 848	Lin,Y.S.
20	ORFV/2009/Korea	ORFV 20	goat	Korea	GQ328 007	Oem,J.K.
21	Xinjiang1	ORFV 20	goat	China	KF666 563	Yang,H.
22	ORF/2010/VIRgen e-3	ORFV 20	goat	Korea	JX968 995	Oem,J.-K.
23	ORF/2010/VIRgen e-2	ORFV 20	goat	Korea	JX968 994	Oem,J.-K.
24	ORF/2010/VIRgen e-1	ORFV 20	goat	Korea	JX968 993	Oem,J.-K.
25	SV269/11	ORFV 20	Ovis aries	Brazil	KF927 109	Martins,M.
26	SV252/11	ORFV 20	Ovis aries	Brazil	KF927 108	Martins,M.
27	ORF/2011/7-18	ORFV 20	goat	Korea	JX968 997	Oem,J.-K.
28	ORF/2011/VIR	ORFV 20	goat	Korea	JX968 996	Oem,J.-K.
29		ORFV 20	vacci ne strain	USA	AY278 210	Guo,J.
30	MRCSG/Shuhama -02/Sheep	ORFV 20	sheep	India	KU672 687	Ahanger,S.A.
31	MRCSG/Shuhama -01/Sheep	ORFV 20	sheep	India	KU672 686	Ahanger,S.A.
32	BSF/Nishat- 02/Sheep	ORFV 20	sheep	India	KU672 685	Ahanger,S.A.
33	BSF/Nishat- 01/Sheep	ORFV 20	sheep	India	KU672 684	Ahanger,S.A.
34	TVCC/Shuhama- 01/Sheep	ORFV 20	sheep	India	KU672 677	Ahanger,S.A.
35	HIS	ORFV 20	Japan ese serow s	Japan	AB522 798	Inoshima,Y.
36	Orf-11	ORFV 20		UK	AY292 459	McInnes,C.J.
37	SBF/Dachigam- 01/Sheep	ORFV 20	sheep	India	KU672 681	Ahanger,S.A.
38	sCh97	ORFV 20	sheep	Argentina	KP244 327	Peralta,A.
39	sPi13	ORFV 20	sheep	Argentina	KP244 326	Peralta,A.

40	IJS081	ORFV 20	Japan ese serow s	Japan	AB492 085	Inoshima, Y.
41	Orf-11	ORFV 20		UK	AJ222 702	McInnes, C.J.
42	isolate D	ORFV 20	sheep	Brazil	JN603 830	Abrahao, J.S.
43	NE1	ORFV 20	goat	Brazil	JN603 828	Abrahao, J.S.
44	MT-05	ORFV 20	sheep	Brazil	JN603 827	Abrahao, J.S.
45	6126/02	ORFV 20		Italy	DQ275 173	Kottaridi, C.
46	SV820/10	ORFV 20	Ovis aries	Brazil	KF927 111	Martins, M.
47	ORFV/XinJiang/2 011/China	ORFV 20	sheep	China	JN565 695	Li, H.
48	Orf virus	ORFV 20	takin	USA	AY424 976	Guo, J.
49	Orf virus	ORFV 20	sheep	USA	AY424 975	Guo, J.
50	Orf virus	ORFV 20	Goat	USA	AF380 126	Guo, J.
51	MRI Scab	ORFV 20	goat	USA	AJ222 701	McInnes, C.J.
52	isolate 176/95	ORFV 20	goat	Greece	DQ275 165	Kottaridi, C.
53	30/96 interferon	ORFV 20	goat	Kottaridi, C.	DQ275 161	Kottaridi, C.
54	SBF/Goabal- 01/Sheep	ORFV 20	sheep	India	KU672 680	Ahanger, S.A.
55	SBF/Goabal- 02/Sheep	ORFV 20	sheep	India	KU672 678	Ahanger, S.A.
56	SV581/11	ORFV 20	Ovis aries	Brazil	KF927 110	Martins, M.
57		ORFV 20	musk ox	USA	AY424 974	Guo, J.
58	PA11	ORFV 20	Capra hircus	Brazil	JQ349 519	de Oliveira
59	12129/00	ORFV 20	goat	Italy	DQ275 170	Kottaridi, C.
60	661/95	ORFV 20	goat	Italy	DQ275 169	Kottaridi, C.
61	1710/03	ORFV 20	goat	Greece	DQ275 166	Kottaridi, C.
62	ORFV/YL/2013/C hina	ORFV 20		China	KF772 212	Liu, F.
63	A	ORFV 20	goat	Brazil	JN603 831	Abrahao, J.S.
64	7389/03	ORFV 20	goat	Italy	DQ275 172	Kottaridi, C.
65	155/95	ORFV 20	goat	Greece	DQ275 163	Kottaridi, C.
66	513/04	ORFV 20	goat	Greece	DQ275 167	Kottaridi, C.
67	928/02	ORFV 20	goat	Greece	DQ275 168	Kottaridi, C.
68	ORFV/2015/Zamb ia/Lu01	ORFV 20	Capra hircus	Zambia:L usaka	LC208 800	Simulundu, E.

69	13598/03	ORFV 20	goat	Italy	DQ275 171	Kottaridi,C.
70	759/01	ORFV 20	goat	Greece	DQ275 164	Kottaridi,C.
71	Shihezi3/SHZ3	ORFV 20	goat	China	KF726 848	Yang,H.
1	Feidong	ORFV 059	goat	China	KX951 408	Yang,H.
2	Xinjiang	ORFV 059	goat	China	KF703 748	Yang,H.
3	SX2	ORFV 059	goat	China	KJ139 957	Shang,C.
4	Guizhou	ORFV 059	goat	China	KP057 582	Liu,A.
5	Xinjiang/SHZ1	ORFV 059	goat	China	KF726 851	Yang,H.
6	ORFV/GanS/2010/ China	ORFV 059	unkn own	China	JX142 183	Jia,H.
7	Xinjiang	ORFV 059	sheep	China	KC291 656	He,Z.
8	HuB13	ORFV 059	goat	China	KJ139 993	Chi,X.
9	Xinjiang2	ORFV 059	goat	China	KF666 561	Yang,H.
10	Jilin	ORFV 059	sheep	China	FJ8080 75	Zhao,K.
11	FJ-YT2015	ORFV 059	goat	China	KU199 842	Lin,Y.S.
12	FJ-YT2014	ORFV 059	goat	China	KU199 841	Lin,Y.S.
13	FJ-MH2015	ORFV 059	goat	China	KU199 840	Lin,Y.S.
14	FJ-LY2014	ORFV 059	goat	China	KU199 839	Lin,Y.S.
15	FJ-LJ2015	ORFV 059	goat	China	KU199 838	Lin,Y.S.
16	FJ-LJ2014	ORFV 059	goat	China	KU199 837	Lin,Y.S.
17	FJ-JA2015	ORFV 059	goat	China	KU199 836	Lin,Y.S.
18	FJ-FZ2015	ORFV 059	goat	China	KU199 835	Lin,Y.S.
19	FJ-FZ2014	ORFV 059	goat	China	KU199 834	Lin,Y.S.
20	FJ-BF2015	ORFV 059	goat	China	KU199 833	Lin,Y.S.
21	NA1/11	ORFV 059	sheep	China	JQ619 904	Li,W.
22	Jilin-Nongan	ORFV 059	sheep	China	JQ271 535	Wang,G.
23	OV-HN3/12	ORFV 059	sheep	China	KC569 751	Li,W.
24	GY-AHF10	ORFV 059	goat	China	KX029 229	Wang,G.
25	Jammu 23/06	ORFV 059	goat	India	KY412 879	Yogisharadh ya,R.
26	Mysore	ORFV 059	sheep	India	KY412 878	Yogisharadh ya,R.
27	India 79/04	ORFV 059	sheep	India	KY412 877	Yogisharadh ya,R.

28	Jalandhar	ORFV 059	goat	India	KY412 876	Yogisharadh ya,R.
29	Orissa 14/06	ORFV 059	goat	India	KY412 875	Yogisharadh ya,R.
30	Mukteswar 83/09	ORFV 059	sheep	India	KY412 874	Yogisharadh ya,R.
31	Gujrat	ORFV 059	goat	India	KY412 873	Yogisharadh ya,R.
32	Bangalore	ORFV 059	goat	India	KY412 872	Yogisharadh ya,R.
33	Meghalaya	ORFV 059	goat	India	KY412 871	Yogisharadh ya,R.
34	Hyderabad	ORFV 059	goat	India	KY412 870	Yogisharadh ya,R.
35	Tamil Nadu	ORFV 059	goat	India	KY412 869	Yogisharadh ya,R.
36	Alwar	ORFV 059	goat	India	KY412 868	Yogisharadh ya,R.
37	Palampur 20/06	ORFV 059	goat	India	KY412 867	Yogisharadh ya,R.
38	Assam 2010	ORFV 059	goat	India	KY412 866	Yogisharadh ya,R.
39	Ludhiana 50/06	ORFV 059	goat	India	KY412 865	Yogisharadh ya,R.
40	Mukteswar 59/05	ORFV 059	goat	India	KY412 864	Yogisharadh ya,R.
41	Mukteswar Vaccine	ORFV 059	goat	India	KY412 880	Yogisharadh ya,R.
42	Bra/UFBA/LabVir o/C1/2014	ORFV 059	goat	Brazil	KP836 315	Tigre D.M.
43	Bra/UFBA/LabVir o/C8/2014	ORFV 059	goat	Brazil	KP836 316	Tigre D.M.
1	Gondar_zuria/O04 /2013	ORFV 108	sheep	Ethiopia	KT438 548	Gelaye,E.
2	Gondar_zuria/C03/ 2013	ORFV 108	sheep	Ethiopia	KT438 547	Gelaye,E.
3	Orf virus isolate Gondar_zuria/C02/ 2013	ORFV 108	goat	Ethiopia	KT438 546	Gelaye,E.
4	Gondar_zuria/C01/ 2013	ORFV 108	goat	Ethiopia	KT438 545	Gelaye,E.
5	ATARC/O02/2010	ORFV 108	sheep	Ethiopia	KT438 542	Gelaye,E.
6	ATARC/O02/2008	ORFV 108	sheep	Ethiopia	KT438 541	Gelaye,E.
7	ATARC/O01/2010	ORFV 108	sheep	Ethiopia	KT438 540	Gelaye,E.
8	ATARC/O01/2008	ORFV 108	sheep	Ethiopia	KT438 539	Gelaye,E.
9	Amba_Giorgis/O0 3/2012	ORFV 108	sheep	Ethiopia	KT438 538	Gelaye,E.
10	Amba_Giorgis/O0 1/2012	ORFV 108	sheep	Ethiopia	KT438 537	Gelaye,E.
11	Amba_Giorgis/C0 2/2012	ORFV 108	sheep	Ethiopia	KT438 536	Gelaye,E.
12	Adet/O01/2012	ORFV 108	sheep	Ethiopia	KT438 531	Gelaye,E.
13	Tamil Nadu	ORFV 108	sheep	India	JN183 076	Yogisharadh ya,R.
14	Mysore	ORFV 108	sheep	India	JN183 075	Yogisharadh ya,R.

15	Mukteswar 59/05	ORFV 108	goat	India	JN183 074	Yogisharadha, R.
16	Meghalaya	ORFV 108	goat	India	JN183 073	Yogisharadha, R.
17	Ludhiana	ORFV 108	goat	India	JN183 072	Yogisharadha, R.
18	Gujarat	ORFV 108	goat	India	JN183 071	Yogisharadha, R.
19	Bangalore	ORFV 108	goat	India	JN183 070	Yogisharadha, R.
20	Assam	ORFV 108	goat	India	JN183 069	Yogisharadha, R.
21	Orissa 14/06	ORFV 108	goat	India	JN183 067	Yogisharadha, R.
22	Alwar	ORFV 108	goat	India	JN183 064	Yogisharadha, R.
23	Nantou	ORFV 108	goat	Taiwan	EU327 509	Chan, K.W.

Annexure B: Published isolates of APV used for phylogenetic analysis.

Virus name	Gene	Host	GenBank accession number	Country	Authors
Fowlpox virus	P4b core protein	Chicken	AM050377	UK	Jarmin et al., 2006
Fowlpox virus	P4b core protein	Chicken	MG787394	Mozambique	Mapaco et al., 2018
Fowlpox virus	P4b core protein	Chicken	MF766430	France	Croville et al., 2018
Fowlpox virus	P4b core protein	Merriam's Turkey	MH734528	USA	Joshi et al., 2018
Fowlpox virus	P4b core protein	Turkey	KU522210	Iran	Norouzian et al., 2016
Turkeypox virus	P4b core protein	Turkey	KR534326	Egypt	Mahmoud et al., 2015
Avipox	P4b core	Parrot	KC017963	Chile	Gyuranecz et al.,

virus	protein				2013
Avipox virus	P4b core protein	Chicken	KM974727	Portugal	Henriques et al., 2016
Fowlpox virus	P4b core protein	Chicken	KP987214	Nigeria	Meseko et al., 2016
Fowlpox virus	P4b core protein	Sparrow	HM481407	India	Pawar et al., 2011
Fowlpox virus	P4b core protein	Silver pheasant	HM481406	India	Pawar et al., 2011
Fowlpox virus	P4b core protein	Peacock	HM481405	India	Pawar et al., 2011
Fowlpox virus	P4b core protein	Golden pheasant	HM481402	India	Pawar et al., 2011
Fowlpox virus	P4b core protein	Chicken	KF548036	India	Roy et al., 2013
Fowlpox virus	P4b core protein	Chicken	AM050380	Europe	Jarmin et al., 2006
Fowlpox virus	P4b core protein	Grey partridge	GQ180201	Italy	Manarolla et al., 2010
Fowlpox virus	P4b core protein	Pheasant	GQ180207	Italy	Manarolla et al., 2010
Turkeypox virus	P4b core protein	Turkey	GQ180212	Italy	Manarolla et al., 2010
Fowlpox virus	P4b core protein	Canary	GQ221269	Italy	Manarolla et al., 2010
Avipox virus	P4b core protein	South Island saddleback	HQ701718	New Zealand	Ha et al., 2011
Avipox virus	P4b core protein	Paradise shelduck	HQ701719	New Zealand	Ha et al., 2011

virus					
Avipox virus	P4b core protein	Variable oystercatcher	HQ701720	New Zealand	Ha et al., 2011
Avipox virus	P4b core protein	Black robin	HQ701721	New Zealand	Ha et al., 2011
Avipox virus	P4b core protein	North Island Robin	HQ701722	New Zealand	Ha et al., 2011
Avipox virus	P4b core protein	North Island saddleback	HQ711987	New Zealand	Ha et al., 2011
Avipox virus	P4b core protein	Shore plover	HQ711989	New Zealand	Ha et al., 2011
Fowlpox virus	P4b core protein	Chicken	KC017960	Hungary	Gyuranecz et al., 2013
Fowlpox virus	P4b core protein	Chicken	JX464821	Egypt	Abdallah et al., 2013
Turkeypox virus	P4b core protein	Turkey	JQ665839	Egypt	Abdallah et al., 2013
Turkeypox virus	P4b core protein	Turkey	KC017961	Nevada/USA	Gyuranecz et al., 2013
Fowlpox virus	P4b core protein	Chicken	KC017962	Hawaii/USA	Gyuranecz et al., 2013
Avipox virus	P4b core protein	Blue-eared pheasant	KC017964	Hungary	Gyuranecz et al., 2013
Avipox virus	P4b core protein	Song thrush	HQ701723	New Zealand	Ha et al., 2011
Fowlpox virus	P4b core protein	MacQueen's bustard	LK021663	UAE	Le Loc'h et al., 2014
Avipox	P4b core	Silver eye	HQ701724	New Zealand	Ha et al., 2011

virus	protein				
Avipox virus	P4b core protein	Shore plover	HQ701717	New Zealand	Ha et al., 2011
Fowlpox virus	P4b core protein	Chicken	KF722863	Tanzania	Masola et al., 2013
Fowlpox virus	P4b core protein	Chicken	HM623675	China	Li et al., 2010
Fowlpox virus	P4b core protein	Quail	MH061351	Mozambique	Mapaco et al., 2018
Pigeonpox virus	P4b core protein	Pigeon	AM050385	UK	Jarmin et al., 2006
Turkeypox virus	P4b core protein	Turkey	AM050387	UK	Jarmin et al., 2006
Turkeypox virus	P4b core protein	Turkey	AM050388	Italy	Jarmin et al., 2006
Fowlpox virus	P4b core protein	Grey partridge	GQ180204	Italy	Manarolla et al., 2010
Fowlpox virus	P4b core protein	Canary	GQ180208	Italy	Manarolla et al., 2010
Fowlpox virus	P4b core protein	Gyrfalcon	GQ180210	Italy	Manarolla et al., 2010
Pigeonpox virus	P4b core protein	Pigeon	JQ665840	Egypt	Abdallah et al., 2013
Avipox virus	P4b core protein	Booted eagle	KC017976	Spain	Gyuranecz et al., 2013
Avipox virus	P4b core protein	Red-legged partridge	KC017977	Spain	Gyuranecz et al., 2013
Avipox	P4b core protein	Red kite	KC017978	Spain	Gyuranecz et al., 2013

virus					
Avipox virus	P4b core protein	Booted eagle	KC017979	Spain	Gyuranecz et al., 2013
Avipox virus	P4b core protein	Red-legged partridge	KC017980	Spain	Gyuranecz et al., 2013
Pigeonpox virus	P4b core protein	Feral pigeon	KC821551	Port Elizabeth/South Africa	Offerman et al., 2013
Fowlpox virus	P4b core protein	Racing pigeon	KC821557	Cape Town/South Africa	Offerman et al., 2013
Fowlpox virus	P4b core protein	Peacock	MH061352	Mozambique	Mapaco et al., 2018
Fowlpox virus	P4b core protein	Turkey	MG787396	Mozambique	Mapaco et al., 2018
Avipox virus	P4b core protein	Rock pigeon	KC821559	Cape Town/South Africa	Gyuranecz et al., 2013
Pigeonpox virus	P4b core protein	Pigeon	MF496043	India	Sharma et al., 2019
Fowlpox virus	P4b core protein	MacQueen's bustard	LK021662	UAE	Le Loc'h et al., 2014
Fowlpox virus	P4b core protein	Chicken	DQ873810	India	Yadav et al., 2006
Avipox virus	P4b core protein	Rock-dove	KC017971	Hungary	Gyuranecz et al., 2013
Fowlpox virus	P4b core protein	Pigeon	DQ873811	India	Yadav et al., 2006
Avipox	P4b core	Indian	KC017975	Hungary	Gyuranecz et al.,

virus	protein	peafowl			2013
Pigeonpox virus	P4b core protein	Feral pigeon	KC821550	South Africa	Offerman et al., 2013
Avipox virus	P4b core protein	Oriental turtle-dove	KC017972	South Korea	Gyuranecz et al., 2013
Avipox virus	P4b core protein	Oriental turtle-dove	KC017973	South Korea	Gyuranecz et al., 2013
Avipox virus	P4b core protein	Great bustard	KC017974	Spain	Gyuranecz et al., 2013
Avipox virus	P4b core protein	Falcon	AM050376	UAE	Jarmin et al., 2006
Avipox virus	P4b core protein	Black-browed Albatross	AM050392	UK	Jarmin et al., 2006
Avipox virus	P4b core protein	New Zealand pigeon	HQ701713	New Zealand	Ha et al., 2011
Avipox virus	P4b core protein	Hawk	KC018008	Hungary	Gyuranecz et al., 2013
Avipox virus	P4b core protein	Red Kite	KC018010	Spain	Gyuranecz et al., 2013
Avipox virus	P4b core protein	Mourning dove	KC018000	Wisconsin/USA	Gyuranecz et al., 2013
Avipox virus	P4b core protein	Canada goose	KC018002	Wisconsin/USA	Gyuranecz et al., 2013
Avipox virus	P4b core protein	Red head duck	KC017993	Wisconsin/USA	Gyuranecz et al., 2013

virus					
Avipox virus	P4b core protein	Trumpeter swan	KC017995	Wisconsin/USA	Gyuranecz et al., 2013
Avipox virus	P4b core protein	Peregrine falcon	KC017988	Hungary	Gyuranecz et al., 2013
Avipox virus	P4b core protein	Red-footed falcon	KC017989	Hungary	Gyuranecz et al., 2013
Avipox virus	P4b core protein	Canary	KC018014	Hawaii/USA	Gyuranecz et al., 2013
Avipox virus	P4b core protein	Golden eagle	KC018058	Spain	Gyuranecz et al., 2013
Avipoxvirus	P4b core protein	House finch	DQ131896	USA	Adams et al., 2005
Avipox virus	P4b core protein	American crow	DQ131893	USA	Adams et al., 2005
Avipox virus	P4b core protein	European starling	KC018063	Maryland/USA	Gyuranecz et al., 2013
Avipoxvirus	P4b core protein	European starling	KC018065	Maryland/USA	Gyuranecz et al., 2013
Avipox virus	P4b core protein	Parrot	KC018069	New York/USA	Gyuranecz et al., 2013
Avipox virus	P4b core protein	Parrot	AM050383	UK	Jarmin et al., 2006
Turkeypox	P4b core	Chicken	KY312503	Mozambique	Mapaco et al., 2017

virus	protein				
Turkeypox virus	P4b core protein	Turkey	KP728110	Hungary	Banyai et al., 2015
Fowlpox virus	P4b core protein	Quail	GQ180200	Italy	Manarolla et al., 2010
Avipox virus	DNA polymerase	Blue-eared pheasant	KC017851	Hungary	Gyuranecz et al., 2013
Fowlpox virus	DNA polymerase	Chicken	KC017850	Hungary	Gyuranecz et al., 2013
Turkeypox virus	DNA polymerase	Turkey	KC017866	Nevada/USA	Gyuranecz et al., 2013
Fowlpox virus	DNA polymerase	Chicken	KC017867	Hawaii/USA	Gyuranecz et al., 2013
Fowlpox virus	DNA polymerase	Chicken	MF766430	France	Croville et al., 2018
Fowlpox virus	DNA polymerase	Merriam's Turkey	MH734528	USA	Joshi et al., 2018
Fowlpox virus	DNA polymerase	Chicken	MK651846	Brazil	Chacon et al., 2019 (Unpublished)
Fowlpox virus	DNA polymerase	Chicken	MH237842	Mozambique	Mapaco et al., 2018
Avipox virus	DNA polymerase	Parrot	KC017883	Chile	Gyuranecz et al., 2013
Avipox virus	DNA polymerase	Chicken	MH754152	Egypt	Lebdah et al., 2019 (Unpublished)
Avipox virus	DNA polymerase	Indian peafowl	KC017857	Hungary	Gyuranecz et al., 2013
Avipox	DNA	Red-legged	KC017892	Spain	Gyuranecz et al.,

virus	polymerase	partridge			2013
Avipox virus	DNA polymerase	Red kite	KC017893	Spain	Gyuranecz et al., 2013
Avipox virus	DNA polymerase	Booted eagle	KC017894	Spain	Gyuranecz et al., 2013
Avipox virus	DNA polymerase	Red-legged partridge	KC017895	Spain	Gyuranecz et al., 2013
Turkeypox virus	DNA polymerase	Turkey	MH237843	Mozambique	Mapaco et al., 2018
Avipox virus	DNA polymerase	Quail	MH237849	Mozambique	Mapaco et al., 2018
Avipox virus	DNA polymerase	Booted eagle	KC017891	Spain	Gyuranecz et al., 2013
Pigeonpox virus	DNA polymerase	Feral pigeon	NC024447	South Africa	Offerman et al., 2018
Pigeonpox virus	DNA polymerase	Feral pigeon	KJ801920	South Africa	Offerman et al., 2014
Avipox virus	DNA polymerase	Oriental turtle-dove	KC017886	South Korea	Gyuranecz et al., 2013
Avipox virus	DNA polymerase	Oriental turtle-dove	KC017887	South Korea	Gyuranecz et al., 2013
Avipox virus	DNA polymerase	Great bustard	KC017890	Spain	Gyuranecz et al., 2013
Avipox virus	DNA polymerase	Mourning dove	KC017912	Wisconsin/USA	Gyuranecz et al., 2013
Avipox	DNA	Canada	KC017913	Wisconsin/USA	Gyuranecz et al.,

virus	polymerase	goose			2013
Avipox virus	DNA polymerase	Hawk	KC017860	Hungary	Gyuranecz et al., 2013
Avipox virus	DNA polymerase	Red kite	KC017896	Spain	Gyuranecz et al., 2013
Avipox virus	DNA polymerase	Red duck	KC017908	USA	Gyuranecz et al., 2013
Avipox virus	DNA polymerase	Trumpeter swan	KC017909	Wisconsin/USA	Gyuranecz et al., 2013
Avipox virus	DNA polymerase	Peregrine falcon	KC017858	Hungary	Gyuranecz et al., 2013
Avipox virus	DNA polymerase	Red-footed falcon	KC017859	Hungary	Gyuranecz et al., 2013
Avipox virus	DNA polymerase	Canary	KC017869	Hawaii/USA	Gyuranecz et al., 2013
Avipox virus	DNA polymerase	Golden eagle	KC017863	Spain	Gyuranecz et al., 2013
Avipox virus	DNA polymerase	European starling	KC017954	Maryland/USA	Gyuranecz et al., 2013
Avipox virus	DNA polymerase	European starling	KC017955	Maryland/USA	Gyuranecz et al., 2013
Avipox virus	DNA polymerase	Parrot	KC017849	UK	Gyuranecz et al., 2013
Avipox virus	DNA polymerase	Parrot	KC017925	New York/USA	Gyuranecz et al., 2013

virus					
Turkeypox virus	DNA polymerase	Turkey	KP728110	Hungary	Banyai et al., 2015
Turkeypox virus	DNA polymerase	Turkey	KY312501	Mozambique	Mapaco et al., 2017