

**EVALUATION OF CELLULOSE-MATRIX DNA
EXTRACTION AND *DADAPP* FOR NON-INVASIVE
PATERNITY ANALYSIS AND LEARNING IN WILD
BABOONS**

DAVID KIRAGU MWAURA

**MASTER OF SCIENCE IN
MOLECULAR BIOLOGY AND BIOINFORMATICS**

**JOMO KENYATTA UNIVERSITY
OF
AGRICULTURE AND TECHNOLOGY**

2026

Evaluation of Cellulose-Matrix DNA Extraction and *DadApp* for Non-Invasive Paternity Analysis and Learning in Wild Baboons

David Kiragu Mwaura

A Thesis submitted in Partial Fulfilment of the Requirements for the Degree of Master of Science in Bioinformatics and Molecular Biology of the Jomo Kenyatta University of Agriculture and Technology

2026

DECLARATION

This thesis is my original work and has not been presented for a degree in any other University

Signature.....Date.....

David Kiragu Mwaura

This thesis has been submitted for examination with our approval as University Supervisors

Signature.....Date.....

Prof. Jenny Tung, PhD

Max Planck Institute for Evolutionary Anthropology, Germany

Signature.....Date.....

Dr. Mercy Akinyi, PhD

KEMRI, Kenya

Signature.....Date.....

Dr. Daniel Kiboi, PhD

JKUAT, Kenya

DEDICATION

I dedicate this work to the Almighty God for it is through Him in me that made all this a success. Glory and honour to Him.

To my beloved family and dear friends, I dedicate this work to you. Your unwavering support, prayers, advice and concerns formed a strong foundation through which I sourced courage to go through this process till its completion with utmost success. Wishing you all divine favour.

ACKNOWLEDGEMENT

First and foremost, I express my deepest gratitude to the Lord God for His firm guidance and protection that sustained me throughout the course of this research. His grace has been my constant source of strength.

I am sincerely thankful to my supervisors: Prof. Jenny Tung, Max Planck Institute for Evolutionary Anthropology (Germany), Dr. Mercy Akinyi, Kenya Institute of Primate Research (Kenya) and Dr. Daniel Kiboi, Jomo Kenyatta University of Agriculture and Technology (Kenya) for continuous availability for consultation, invaluable feedback, expert supervision and material provisions that enabled me to complete this research project.

Dr. Jordan Anderson, Duke University, for his kind encouragement and support during my challenging days in coding and troubleshooting. I am greatly indebted to you.

My special appreciation to Prof. Susan Albert, Co-director of Amboseli Baboon Research Project for granting me a fully funded scholarship for my master's program. I also recognize the Amboseli Baboon Research Project Team, Dr. Tim Wango, Vivian Oudu, Mututua Raphael, Ben Oyath, Kinyua Warutere, Gideon Marinka, Siodi Long'ida, and Alex Meloimet for all the field logistics, field samples collection and transportation.

Special mention to all my friends, colleagues and technical staff at the Animal Science Department, Kenya Institute of Primate Research for your technical support and cooperation.

Finally, I extend heartfelt thanks to my beloved parents and siblings for their enduring prayers and unwavering encouragement, which was instrumental in achieving this milestone.

Glory and praise be to God forever. Amen

TABLE OF CONTENTS

DECLARATION.....	ii
DEDICATION.....	iii
ACKNOWLEDGEMENT	iv
TABLE OF CONTENTS.....	v
LIST OF TABLES	x
LIST OF FIGURES	xi
LIST OF APPENDICES	xii
ABBREVIATIONS AND ACRONYMS	xiii
ABSTRACT	xiv
CHAPTER ONE	1
INTRODUCTION.....	1
1.1 Background Information	1
1.2 Statement of the Problem	4
1.3 Justification of the Study	4
1.4 Objectives of the Study	5
1.4.1 General Objective	5

1.4.2 Specific Objectives	5
1.5 Research Questions	6
CHAPTER TWO	7
LITERATURE REVIEW.....	7
2.1 Introduction to Non-Invasive Genotyping	7
2.1.1 History of Non-Invasive Genotyping	8
2.1.2 Application of Non-Invasive Genotyping.....	9
2.2 Processes Involved in PCR-Based Non-Invasive Genotyping.....	10
2.2.1 Non-Invasive Sampling	10
2.2.2 DNA Extraction Methods	12
2.2.3 Target Gene Amplification	15
2.2.4 Non-Invasive Genotyping.....	15
2.3 Development of Non-Invasive Genotyping.....	17
2.4 Statistical and Software Development in Parentage Analysis.....	19
CHAPTER THREE	21
MATERIALS AND METHODS	21
3.1 Study Site	21

3.2 Sample Collection	22
3.3 Evaluation of Whatman DNA Extraction Method	24
3.3.1. Whatman-Based DNA Extraction	24
3.3.2 Bead Clean Up	25
3.3.3 Column - Based DNA Extraction	25
3.3.4 Multiplex PCR	26
3.3.5 Agarose Gel Electrophoresis	28
3.4 Implementation of a User-Friendly Shiny-Based Web Application for Determining Paternity Using Microsatellite or SNP Markers	28
3.4.1 Designing a Likelihood/LOD-Based Paternity Assignment Model that Uses Likelihood Ratio to Assign Paternity Using R Programming Language.....	28
3.4.2 Implementation of a Shiny Web Application that Uses a Likelihood Approach to Assign Paternity	32
3.4.3 Suitability of <i>DadApp</i> . A Case Study for Determining Paternity in a Wild Baboon Population in Amboseli Baboon Research Project (ABRP).....	33
3.5 Evaluation of Developed Shiny Web Application as an Active Learning Tool for Both Molecular and Computational Aspects of Paternity	34

CHAPTER FOUR.....	37
RESULTS	37
4.1 Sample Collection	37
4.2 Evaluation of Whatman DNA Extraction Method	37
4.3 Shiny Web Application Development - <i>DadApp</i>	40
4.3.1 <i>DadApp</i> Paternity Assignment Aligns with Real Data of Wild Baboons.....	42
4.4 <i>DadApp</i> as a Potential Interactive Educational tool	42
CHAPTER FIVE.....	48
DISCUSSION	48
5.1 Cellulose Based Extraction.....	48
5.2 <i>DadApp</i> Shiny Web Application as a Tool for Assigning Paternity in Wild Baboon Populations.....	50
5.2. <i>DadApp</i> as a potential active educational tool	53
CHAPTER SIX	56
CONCLUSIONS AND RECOMMENDATIONS.....	56
6.1 Conclusion.....	56
6.1 Recommendations	57
REFERENCES.....	58

APPENDICES76

LIST OF TABLES

Table 3.1: Primers that Target Microsatellite Loci Specific for Paternity Studies in Wild Baboons' Population as Described in Buchan et al., (2005)	26
Table 3.2: Caporaso Primer that Targets 16S rRNA Subunit of Bacterial DNA Used a Positive Control for PCR Amplification as Described by Caporaso et al (2012)	27
Table 3.3: Likelihood Ratio for Various Trio Paternity Combinations	31
Table 3.4: Likelihood Ratio for Various Duo Paternity Combinations	32
Table 4.1: Baboon Identification and Demographic Information for Sampled Individuals from Europa Troop Found in ABRP	37
Table 4.2: Optimized Touchdown PCR Thermal Conditions for Amplification of Microsatellite Loci Specific for Paternity Studies in Wild Baboons.....	38

LIST OF FIGURES

Figure 2.1: Genotyping Sampling Method	8
Figure 2.2: Overview of Paramagnetic Bead-Based DNA Extraction Workflow	13
Figure 3.1: Representation of the Study Site	22
Figure 3.2: Collection of Freshly Deposited Faecal Samples From Known Baboons Following Identification by ABRP Field Observers	23
Figure 3.3: Flow Chart Illustrating Steps in the Likelihood/LOD-Based Paternity Assignment Model of the Shiny Web Application	33
Figure 4.1: An Agarose Gel Image of PCR Amplicons.....	39
Figure 4.2: An Agarose Gel Image Showing PCR Amplicons.....	40
Figure 4.3: Represents the User Interface of the Shiny Application	41
Figure 4.4: Bar Plot Illustrating the Number of Participants	43
Figure 4.5: Bar Plot Depicting Individual Pre- and Post-Test Means with Standard Deviation, Measured before and after DadApp Exposure	44
Figure 4.6: Within-Individual Comparison of Pre-Test and Post-Test Scores following Interaction with <i>DadApp</i>	45
Figure 4.7: Bar Plots Illustrating Changes in Participants' Self-Reported Interest and Confidence across Stages of the Study	46

LIST OF APPENDICES

Appendix I: Ethics Approval Letter Used to Assess DadApp as an Active Educational Tool.....	76
Appendix II: Report of how Suitability of <i>DadApp</i> was Assessed Using Real Data from ABRP	78
Appendix III: DadApp Workshop Starter Pack Used for the Assessment of <i>DadApp</i> an Active Learning Tool.....	92

ABBREVIATIONS AND ACRONYMS

ABRP	Amboseli Baboon Research Project
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
FACS	Fluorescence-Activating Cell Sorting
FTA	Flinders Technology Associates
GUI	Graphical User Interface
IgG	Immunoglobulin G
LOD	Log-likelihood ratio
mtDNA	Mitochondrial deoxyribonucleic acid
PCR	Polymerase Chain Reaction
RAD-Seq	Restriction site Associated DNA Sequencing
rRNA	Ribosomal Ribonucleic acid
SNP	Single Nucleotide Polymorphism
T_m	Melting Temperature
USA	United States of America
μl	Microlitre

ABSTRACT

Since the 1990s, non-invasive genetic sampling has transformed research in genetics, evolutionary biology and ecological dynamics in natural animal populations. However, existing approaches are labour- and time-intensive relative to their limited data output, necessitating further optimization. Developing new laboratory methods, analytical tools and software can achieve this goal. This study evaluated the effectiveness of cellulose-matrix based nucleic acid extraction, developed a Shiny-based web application for non-invasive genotyping and paternity analysis in wild baboons and tested the suitability of the application as an educational tool in both molecular and computational aspects of parentage analysis. Faecal samples were collected from thirteen known individuals in one baboon troop from the Amboseli ecosystem in Kenya. Faecal samples were processed using two variant protocols of the cellulose extraction method (single Whatman and four Whatman) and a modified column-based extraction method as a positive control. While the cellulose extraction method failed to amplify DNA for paternity analysis using microsatellite primers, it showed potential for DNA extraction for bacterial identification using the 16S rRNA gene. Additionally, a Shiny web application named '*DadApp*' <https://kiragu-mwaura.shinyapps.io/dadapp/> was developed for paternity analysis using R programming language. Tested with validated non-invasive data from the ABRP, *DadApp* achieved a 100% success rate in correctly assigning paternity in confirmed triad relationships. In every case, the top log-likelihood ratio (LOD) scoring candidate father was consistent with the paternity assignment in the ABRP pedigree, which had been established through independent verification methods. Further, *DadApp* as an active educational tool, significantly improved participants' knowledge of paternity concepts, with participants' mean scores improving from 70.54 (range 45-90) on pre-test to 97.86 (range 80-100) on post-test. The difference was significant based on a paired Wilcoxon Signed-Rank test ($p = 3.75 \times 10^{-6}$), indicating enhanced learning outcomes. Additionally, participants' interest and confidence in paternity inference were significantly enhanced after a practical session with *DadApp* (p -values = 9.485×10^{-5} and 8.103×10^{-4} , respectively). This research highlights the potential, despite limited success, of the cellulose extraction method in non-invasive genotyping and demonstrates the effectiveness of *DadApp* for paternity assignment and education.

CHAPTER ONE

INTRODUCTION

1.1 Background Information

More than two decades ago, non-invasive genotyping was first employed within wild animal populations. This method evolved to avoid invasive approaches associated with obtaining large biopsies or blood from the animal of interest (Waits & Paetkau, 2005). Since then, multiple classes of analyses have used non-invasive genotype data to solve research questions in ecology, heredity/breeding, animal behaviour, evolution and conservation (Zemanova, 2021). These include studies on population composition (Natesh *et al.*, 2019), migration patterns (Schenker *et al.*, 2020), mating patterns, animal behaviour (Bischof *et al.*, 2016), social structure (Rodgers *et al.*, 2015), identification of adaptive genetic variants, and social demographics of endangered animal species (Banks & Piggott, 2022).

Nonetheless, non-invasive approaches face technical challenges such as poor DNA integrity or low DNA concentration, cross-contamination, limited amplification efficiency (Zemanova, 2020), genotyping errors like allelic dropout, and the need for a sophisticated laboratory set up with skilled manpower (Beja-Pereira *et al.*, 2009). Although typical non-invasive molecular genetic workflows require DNA re-extraction from multiple samples and sample re-genotyping, they can still result in amplification and genotyping failures (Hoffman & Amos, 2005). These techniques also experience challenges in scaling to large sample sizes because they can be costly and tedious (Beja-Pereira *et al.*, 2009). Further, non-invasive genotyping datasets still remain small compared to the resulting extensive data now achievable through standard high-throughput sequencing. Therefore, there is a deep need to update methods for generating and analysing genetic data from non-invasive samples due to their outstanding potential in ecological monitoring and conservation genetics.

In recent times, substantial progress has been made in accommodating these challenges associated with non-invasive genotyping. Numerous studies have underscored the significance of appropriate sample collection and optimized laboratory protocols. For example, new methodologies and protocols have emerged to adapt to the low integrity or low concentrated DNA samples, such as targeted sequencing using methylation-based enrichment techniques and hybridization probes (Zemanova, 2021). Other optimized laboratory protocols include the development of enrichment strategies for faecal-derived endogenous DNA (Perry *et al.*, 2010) and whole genome resequencing performed on faecal DNA following host cell isolation via fluorescence-activating cell sorting (FACS) (Orkin *et al.*, 2018). The use of GT-seq, a massively multiplexed two-step PCR reaction for amplifying SNP loci, is associated with a low error rate after additional optimization for poor quality and/or quantity DNA (Campbell *et al.*, 2015).

One notable advancement involves integrating computational approaches for paternity analysis such as WHODAD, with methodologies tailored to extract endogenous DNA from non-invasively acquired specimens (Snyder-Mackler *et al.*, 2016). This protocol, which incorporates biotinylated RNA baits, is economical and has convenient sample input requirements, countering the challenges experienced by non-invasive genotyping. It demonstrates remarkable efficiency in capturing complex, non-redundant DNA, thereby reducing the demand for thermal amplification. As a result, the genotyping data obtained from this method is nearly on par with that derived from blood samples.

Ecological scientists aim to maximize genetic data from non-invasive samples, making them comparable to data obtained by the current standard high-throughput sequencing approaches using invasive samples. The inspiration is caused by the limitation associated with non-invasively collected genotype data such as inability to carry out analyses that may require gene- or region- specific information such as haplotype-based selection analysis and ancestry-based mapping (Snyder-Mackler *et al.*, 2016). Despite these advancements, field-based protocols for non-invasive genotyping data in resource-limited settings are still needed.

However, not all applications in ecological monitoring require this advanced development. Species detection (Kilpatrick *et al.*, 2013), population size estimation (Aleix-Mata *et al.*, 2019), diet analysis (King & Schoenecker, 2019), health monitoring (Baek *et al.*, 2020), sexing and the construction of basic pedigree (Zemanova, 2021) can be achieved with simpler methods. Therefore, there is a need for a cost effective, field-based protocol that produces sufficient genetic data for commonly used applications like sample identification, paternity testing, sexing, and individual identification.

Primates serve as a bridge between understanding natural ecosystems variability and the impacts of human expansion. They act as “sentinels” for ecosystem health, providing comprehensive data on how changing environments affect large mammals over time (Fischer *et al.*, 2019). Long-running studies and associated datasets such as those collected from the Amboseli Baboon Research Project in Kenya, contain data over 50 years on baboon population dynamics, health and environmental interactions (Alberts & Altmann, 2012). Such datasets can be valuable for testing and refining protocols or tools, as well as evaluating how they can be applied to address specific research questions.

Optimized laboratory protocols for non-invasive genotyping are more likely to be broadly adopted if appropriate analysis tools are available for the resulting data. Currently, several programs have been developed for parentage analysis and establishing parent-offspring relationships, such as CERVUS (Kalinowski *et al.*, 2007), COLONY (O. R. Jones & Wang, 2010) and WHODAD (Snyder-Mackler *et al.*, 2016). However, all these programs require sophisticated expertise, are not future-proof or scalable and are limited to specific operating systems. They also limit the user with data visualization options, featured graphical user interfaces (GUIs), and customized capabilities (A. G. Jones *et al.*, 2010). An R Shiny app for paternity analysis, for example, would offer a user-friendly interface, seamless data analysis, customizability, real-time results, reproducibility, collaboration, and integration with other R packages (Jia *et al.*, 2022). These advantages can enhance the accessibility, efficiency and usability of paternity assignment, making it accessible to a wider scope of users and supporting more effective and transparent research practices.

1.2 Statement of the Problem

DNA sequence analysis has become a vital tool for ecological monitoring and conservation of wildlife (Waits & Paetkau, 2005). However, developing countries face significant challenges due to limited resources for generating and analysing genetic data. This often necessitates sending samples to developed countries for analysis, which presents logistical challenges, may compromise sample integrity and limits opportunities for technology transfer. To address this, there is a pressing need for efficient, field-based, low-cost methods for genetic analysis that can provide sufficient genetic information sufficient to monitor wild animal populations. Additionally, a flexible and user-friendly analytical software that is inexpensive to develop and easy to use is crucial to encourage adoption in resource-limited settings.

The desired non-invasive genotyping protocol should combine affordable nucleic acid extraction with a field-based method for generating genotype data. Cellulose-based DNA extraction is an efficient and economically feasible advancement that significantly reduces equipment needs (Zou *et al.*, 2018) . In this study, DNA extraction using Whatman No.1 was assessed to determine whether it can produce amplifiable DNA suitable for paternity analysis. Additionally, a user-friendly R Shiny web application was developed and tested for SNP-based paternity analysis. This application was also validated against known paternity information from ABRP (Alberts & Altmann, 2012). Finally, the use of the Shiny web application was assessed as an active learning tool in understanding paternity concepts.

1.3 Justification of the Study

In recent years, non-invasive genotyping has become an important component of ecological and conservation research due to its potential as a method for collecting data from a wide range of non-invasive sample types without capturing, handling, or otherwise disturbing animals in their natural environments (Zemanova, 2020, 2021). Despite its

widespread use, non-invasive genotyping remains constrained by several limitations, including poor-quality DNA samples, low PCR amplification success rates, elevated genotyping error rates, and substantial financial, logistical, and labour costs (Waits & Paetkau, 2005).

To address these limitations, this study evaluates the performance of a cellulose capture-based DNA extraction method that requires minimal laboratory infrastructure, using faecal samples collected from a wild baboon population. In parallel, the study develops a user-friendly Shiny-based web application, *DadApp*, designed to perform paternity assignment using SNP markers. The Shiny app was validated for its analytical performance using independently derived parent –offspring relationships from a long-term monitored baboon population in the Amboseli ecosystem. Finally, the study assesses the suitability of the application as an active educational tool in teaching molecular and computational principles underlying parentage analysis, particularly in settings concepts outside learning institutions.

1.4 Objectives of the Study

1.4.1 General Objective

To evaluate a cellulose-matrix DNA extraction approach and develop and validate a Shiny-based *DadApp* workflow for non-invasive paternity analysis and paternity-inference education in wild baboons.

1.4.2 Specific Objectives

- i. To evaluate cellulose capture-based DNA extraction in the field using stool samples from a wild baboon population.
- ii. To develop and validate a Shiny-based web application for paternity assignment using SNP data from the Amboseli wild baboon population.

- iii. To assess the suitability of *DadApp* as an active-learning tool for teaching molecular and computational concepts in paternity inference.

1.5 Research Questions

- i. How effective is cellulose capture-based DNA extraction for obtaining high-quality genetic material from faecal samples collected from a wild baboon population?
- ii. Does a Shiny-based web application accurately predict paternity assignment using SNP markers and independently derived genetic data from a wild baboon population?
- iii. How suitable is the Shiny-web application as an educational tool for paternity inference?

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction to Non-Invasive Genotyping

Non-invasive genotyping refers to genetic analysis techniques that avoid harming, capturing, or significantly disturbing the behaviour or fitness of the target organism from which the sample is obtained (Taberlet, 1996). Instead, this approach relies on samples such as faeces, saliva, sloughed and shed hair, feathers, skin, urine collected from the bedding material, scent marks, eggshells, and menstrual fluids (Taberlet & Luikart, 1999; Waits & Paetkau, 2005). Figure 2.1A shows different non-invasive genotyping sampling methods (Emami-Khoyi *et al.*, 2021).

In contrast, invasive genotyping involves obtaining DNA through methods that are more intrusive and often require capturing or handling the organism. Common invasive techniques include drawing blood, taking tissue biopsies, or extracting internal bodily fluids. While these methods can provide high-quality DNA and comprehensive genetic information, they can be stressful, harmful, and sometimes fatal to the subject, especially in sensitive or rare species (Zemanova, 2020). Figure 2.1B shows an example of an invasive genotyping sampling method (thailandanimalwelfare.org).

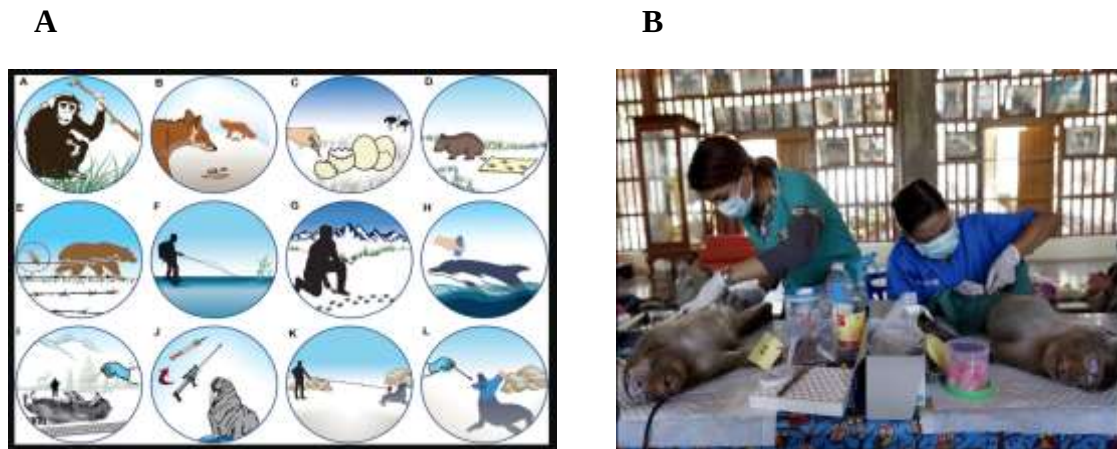


Figure 2.1: Genotyping Sampling Method

A: Major methods of **non-invasive sampling**: a) Chewed sticks b) faeces c) eggshells, d) adhesive tape trap e) hair snare f) filtering of water g) snow imprints, h) blowhole exhalation i) body parts j) non-lethal remote dart biopsy k) hair striping disc l) oral swabbing

B: Major method of **invasive sampling**

2.1.1 History of Non-Invasive Genotyping

Traditionally, genetic analysis of the natural populations required the collection of invasive biopsies. Such specimens were considered valuable resources by biologists studying evolution, ecological distribution and systematics. Before the 1990s, obtaining these tissue biopsies was challenging especially in elusive, secretive animal species and those that live in areas of low population density (Taberlet, 1996). For example, some of the motivating problems that inspired the development of non-invasive molecular genetics were to understand the population genetic composition of endangered animal species such as Arabian white oryx, Père David's elaphure fawn and square-lipped rhinos (Woodruff & Gall, 1992).

In the 1990s, non-invasive sampling was established as a collection approach that could be used without capturing, disturbing, or significantly impacting on the behaviour or health of the target species (Taberlet, 1996).

Unfortunately, non-invasively collected samples are characterised by low quantities of host DNA, which is often of poor quality due to fragmentation (Waits & Paetkau, 2005). The obtained genotype data were used to describe the population genetic composition of the threatened San Clement Island loggerhead shrike (Mundy *et al.*, 1997) and European bears (Taberlet *et al.*, 1997) using feathers and shed hair respectively. Non-invasively generated genetic data was also used to establish phylogenetic surveys of gibbons (Woodruff, 1993) and European bears (Taberlet *et al.*, 1997).

2.1.2 Application of Non-Invasive Genotyping

The development of PCR facilitated the amplification of low-abundance DNA extracted from samples obtained non-invasively. This advancement revolutionised forensics and conservation genetics in both wild and captive animal populations. (Taberlet, 1996). PCR-based non-invasive genotyping has since then been used by conservationists, ecologists, and geneticists to monitor ecological systems and answer research questions pertaining to an animal population of interest.

Scientists have used non-invasive genetics to identify the existence of elusive and endangered animal species such as the brown bear (Bellemain & Taberlet, 2004), Amur tiger (Baek *et al.*, 2020), coyote (Adams *et al.*, 2003) and rocky wallaby (Piggott *et al.*, 2006). Non-invasive genotyping has also been used to characterize the population structures of animal species such as the European wolf (Scandura, 2005), Iberian lynx (Palomares *et al.*, 2002), sage-grouse (Bush *et al.*, 2005), brown bear (Solberg *et al.*, 2006), several seal species (Casper *et al.*, 2007) and the greater flamingo (Balkiz *et al.*, 2007)

Genetic diversity, hybridisation and mating systems within the wild animal population has been determined using non-invasive samples (Waits & Paetkau, 2005). Examples of such wild animal species include the savannah baboon (Tung *et al.*, 2008), dhole (Iyengar *et al.*, 2005), wood grouse (Regnaut *et al.*, 2006) and the Nelson's bighorn sheep (*Ovis canadensis*) (Luikart *et al.*, 2008). Identification of rare species within animal populations of interest have also been made possible through non-invasive genotyping.

Examples of such animal populations include multiple butterfly species (Feinstein, 2004), the North American bullfrog (Dejean *et al.*, 2012), the imperial eagle (Rudnick *et al.*, 2008) and wolverine (Ulizio *et al.*, 2006).

2.2 Processes Involved in PCR-Based Non-Invasive Genotyping

PCR- based non-invasive genotyping is a comprehensive approach that consists of non-invasive sampling of subjects of interest, DNA extraction, target gene amplification and genotyping (Zemanova, 2021).

2.2.1 Non-Invasive Sampling

Non-invasive sampling was established as a collection approach for obtaining samples that avoids capturing, disturbing, or significantly impacting on the behaviour or condition of the target species (Taberlet & Luikart, 1999). Such samples include: faeces, saliva, skin, scent marks, eggshells, menstrual fluids, feathers, urine from bedding, sloughed and shed hair (Beja-Pereira *et al.*, 2009).

Hair and faecal samples are the most frequently collected non-invasive sample types, owing to their accessibility and reliable extraction of high-quality DNA (Beja-Pereira *et al.*, 2009). In hair samples, DNA is typically extracted from cellular material attached to the hair root (Broquet *et al.*, 2006). For faecal samples, endogenous DNA is isolated from the outer layer of faecal specimens, which is enriched with sloughed host intestinal epithelial cells (Ramón-Laca *et al.*, 2015). Additionally, faecal specimens are easily

accessible and can be collected while fresh when target animals are observed, to prevent contamination. Trained dogs have also been used in faecal sample detection and collection (Kerley & Salkina, 2007).

Limitations of non-invasive samples are frequently influenced by the source of the specimen (the animal species of interest) and the environmental conditions under which the sample was collected. For example, an animal's diet may contain secondary compounds that have inhibitory effects on PCR, especially when faecal samples are used (Broquet *et al.*, 2006).

In addition, environmental factors such as ultraviolet radiation, temperature, and humidity can compromise sample integrity (Waits & Paetkau, 2005), resulting in high levels of sample degradation.

Other common challenges associated with non-invasive sampling include low DNA quality and quantity, genotyping errors caused by allelic dropout, difficulties in standardising protocols and increased cost and labour requirements due to extensive quality control measures and the need for multiple replicates. To mitigate these challenges, researchers have often used pilot studies to assess potential limitations and employed subsequent optimised protocols.

2.2.1.1 Preservation of Non-Invasive Samples

The preservation of collected samples is also a crucial step that can influence the success of subsequent genotyping attempts (Nsubuga *et al.*, 2004). There are diverse methods of preserving the collected samples and researchers often need to conduct a pilot studies to determine the most robust technique for a particular sample (Beja-Pereira *et al.*, 2009). However, there are three main principles underlying distinct preservation approaches, all of which aim to deactivate nuclease enzymes involved in DNA degradation. These principles are: (i) dehydration of samples using desiccants (e.g., silica gel, ethanol) or

physical methods (e.g., oven heating, lyophilisation and vacuum spinning); (ii) chelation of divalent cations through agents such as EDTA or ion-exchange resins; and (iii) inhibition of nuclease enzymes by subjecting the sample to low temperatures e.g. freezing (Beja-Pereira *et al.*, 2009).

2.2.2 DNA Extraction Methods

Genomic DNA purification techniques enable the isolation and purification of DNA from non-invasive biological samples with high levels of sensitivity, flexibility and efficiency (Ramón-Laca *et al.*, 2015). The two main types of DNA extraction protocols are: the traditional liquid-based DNA isolation method and the solid-based isolation method.

Liquid-based DNA isolation methods are those that rely on phenol/chloroform, hexadecyltrimethylammonium bromide, and/or guanidinium salts break down cell membranes and nuclear envelopes to release DNA, separating it from other cellular components and purifying DNA for further analysis (Ali *et al.*, 2017). In contrast, commercial solid phase extraction methods utilise the affinity of nucleic acids on solid silica matrices in the presence of chaotropic salts. These kits typically employ sequential washing and centrifugation steps to remove contaminants, followed by elution of purified using a low-salt buffers (Ali *et al.*, 2017). The main bottlenecks of both solid-based and liquid-based isolation methods that involve complex procedures, are time-consuming, require qualified technicians, and require major laboratory equipment which requires an electrical supply. Therefore, they cannot be easily performed outside a modernized laboratory or in a field-based or resource-limited setting (Ali *et al.*, 2017).

2.2.2.1 Development of DNA Extraction Techniques

Over the recent past, there have been several advancements in DNA extraction techniques. Notably, one example specific to a non-invasive sample includes use of paramagnetic beads as illustrated in Figure 2.2. The paramagnetic bead technique, an example of solid-

phase extraction, uses a wide range of modified magnetic of bead surfaces in the process of nucleic acid purification. Magnetic separation is employed to capture the beads allowing contaminant removal through multiple washes. These beads remove the need for centrifugation and electrical equipment, which is quite helpful in field-based set ups (Saiyed *et al.*, 2006).

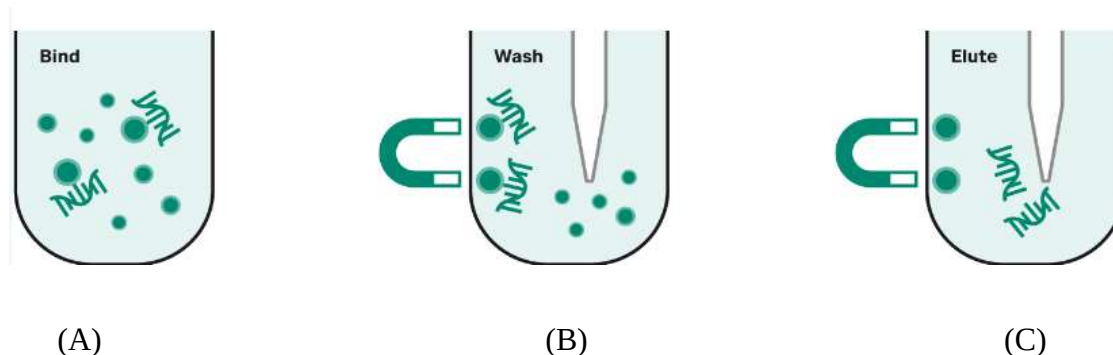


Figure 2.2: Overview of Paramagnetic Bead-Based DNA Extraction Workflow

- A. **binding stage**, which involves magnetic particles binding to target molecules,
- B. **wash stage**, which involves magnetic particles are captured with a magnet and the remainder of the sample is washed away and
- C. **elution stage**, where target molecules are released from magnetic particles for further analysis.

Recent studies have also reported a rapid nucleic acid isolation method utilising various membrane types (Tan *et al.*, 2024). Examples of used membranes include: Flinders Technology Associates (FTA) cellulose-based matrix cards (GE Healthcare, USA), aluminium oxide, the, and fused silicon dioxide-based filters (GE Healthcare, USA) (Tan *et al.*, 2024). However, there is still a need for a simpler and more rapid nucleic acid purification protocol for field set-ups.

2.2.2.1.1 Cellulose based DNA extraction

The desired purification protocol should be less complex, require minimal equipment and avoid the need for electrical appliances to capture and purify nucleic acids. As reported

by Zou *et al.* 2018, a modified cellulose-based DNA extraction method could meet these criteria, since it is inexpensive, disposable, portable and easily modified to suit the researchers' demand.

This technique has been influenced by the need for a method that is effective in a resource-limited setting, easily modified to suit researchers' demand, disposable, inexpensive, portable and field-based. Cellulose-based DNA binding is one of the protocols that has proven to be efficient and effective in extracting amplifiable DNA. DNA isolation can be achieved from a wide range of animals, microbial or plant samples within a short period of time, without the need for complex equipment or highly technical procedures (Tan *et al.*, 2024).

An example of a cellulose-based DNA-binding protocol is the use of Whatman No.1, which has proven to be a simple nucleic acid purification method. Apart from its use in nucleic acid storage and transport, Whatman No. 1 has a high efficiency in binding and purifying nucleic acid away from PCR inhibitors. The use of Whatman No.1 for purifying nucleic acid involves a series of incubation periods and washing stages.

Purified and amplified nucleic acid using Whatman No.1 is comparable to amplification results obtained from direct addition of isolated DNA into a PCR reaction (Zou *et al.*, 2018).

The mechanism underlying the DNA capture and release process during DNA purification is based on differences in kinetics during attraction and expulsion between DNA and the charged cellulose paper. The difference in kinetics occurs when the nucleic acid binds to the cellulose Whatman paper and when released from the cellulose paper. An advantage associated with Zou *et al.* (2018) model is the fact that PCR inhibitors are not purified together with DNA. However, it is still unclear whether these inhibitors fail to attach to the cellulose matrix or are rapidly eliminated during the washing stage.

2.2.3 Target Gene Amplification

Non-invasive biological specimens such as stool samples and hair contain DNA of low quality and quantity which may lead to suboptimal amplification outcomes and erroneous genotyping calls (Beja-Pereira *et al.*, 2009). Therefore, DNA extracted from non-invasive samples is required to undergo PCR to generate large amount of genotype data and prevent genotyping example, such as dropout and false alleles (Arandjelovic *et al.*, 2009; Taberlet, 1996).

For effective use of template DNA extracted from non-invasive samples, multiplex PCR can be employed to target multiple loci within a single reaction. This approach uses multiple primer sets targeting distinct loci, thereby enabling efficient amplification in a single-tube assay. However, multiplex PCR can be labour-intensive to optimise and may increase the risk of contamination due to the added complexity of the procedure. Additionally, multiplex PCR still requires optimisation of compatible primers based on their annealing temperature and size ranges. Another associated disadvantage is that it is highly affected by the presence of PCR inhibitors, which may also affect PCR efficiency (Arandjelovic *et al.*, 2009; Taberlet, 1996).

2.2.4 Non-Invasive Genotyping

Genotyping noninvasively collected samples has traditionally relied on identifying alleles based on fragment size analysis (Arandjelovic & Vigilant, 2018). More recently, however, non-invasive genotyping has taken advantage of newer, high-throughput sequencing techniques that produce genome-scale genotype data or capture approaches that enrich specific regions of interest. Examples of key molecular markers used in non-invasive genotyping include microsatellites, mitochondrial genome and single-nucleotide variants (Byrne *et al.*, 2021).

2.2.4.1. Mitochondrial DNA (mtDNA) Genotyping

Traditionally, mitochondrial DNA was the most popular molecular marker used by population geneticists, conservation geneticists, and molecular systematists (Moritz, 1994). mtDNA is readily amplified because it exists in high copy numbers within cells. Its gene content is also conserved across animal species with minimal duplicates, an absence of introns, and very short intergenic regions. This conservation has been a useful characteristic that has enabled tracing back phylogenies and also geographical movement of maternal lineages (Kivisild, 2015). mtDNA is also highly variable in the hypervariable control regions, which experience high rates of mutation (De Groot *et al.*, 2016).

A limiting factor of mtDNA markers is their maternal mode of inheritance, which can lead to a biased perception of a population's evolutionary history. Additionally, the use of mtDNA in animals is challenged by the widespread presence of mitochondrial pseudogenes integrated into nuclear genomes of most species (Bensasson, 2001). One remedy for the challenges associated with mtDNA is complement the obtained mtDNA data obtained with other markers that are biparentally inherited or the patrilineal Y-chromosome, microsatellites and single nucleotide polymorphisms (De Groot *et al.*, 2016).

2.2.4.2. Microsatellite Genotyping

Microsatellites are stretches of DNA consisting of 2-8 base pair sequences repeated several times in tandem. They are highly polymorphic and are used in population genetics to establish reliable individual profiles from non-invasive samples (De Groot *et al.*, 2016). Other notable features include: their codominant nature, widespread distribution throughout the genome, ability to be accurately typed by size, and the small amount of DNA required for genotypic analysis (De Groot *et al.*, 2016; Zhang & Hewitt, 2003). Microsatellite profiles have enabled the determination of different species identity, demographic information, hybridisation patterns and familial lineage in European wolves

(De Groot *et al.*, 2016). Some of the limitations associated with these markers are difficulties in combining microsatellite-derived datasets from different laboratories.

Other bottlenecks include unclear evolutionary relationships among microsatellite alleles, uncertain neutrality of microsatellites and inconsistent mutation rates between varieties of microsatellite alleles and among organisms (Zhang & Hewitt, 2003).

2.2.4.3. Single Nucleotide Polymorphism (SNPs) Genotyping

SNPs are widely utilized genetic markers for obtaining insights into wild animal populations (De Groot *et al.*, 2016). A SNP represents a genomic locus at which single nucleotide varies among individuals within a population. These markers are genetically stable and are abundant in both coding and non-coding regions. They tend to be bi-allelic, meaning that two allelic variants segregate within a population (Von Thaden *et al.*, 2017).

Other features that have made SNPs popular in the wild animal population include their high statistical power and genotyping resolution. SNP datasets are comparable between different laboratories thereby enabling exchange of dataset (De Groot *et al.*, 2016). One of the associated challenges that this marker encounters are that a large number of SNP loci is required in a single study because of its binary nature.

This is because the large number in a study enables SNPs to achieve statistical power comparable to that of microsatellites (Carroll *et al.*, 2018)

2.3 Development of Non-Invasive Genotyping

Over the recent past, there has been significant development of non-invasive genotyping through the use of high-throughput methods such as large genome-scale techniques that enable large-scale genetic studies in natural animal populations. Perry *et al* (2010) developed an optimized capture method complemented with a massively parallel sequencing technology that enabled the sequencing of large amounts (megabases) of

target nuclear DNA from chimpanzee fecal samples. Another advancement was the development of a genome-wide capture protocol tailored for faecal samples, utilizing biotinylated RNA probes to enrich target DNA and reduce contamination from non-target materials (Snyder-Mackler *et al.*, 2016). Additionally, the researchers created WHODAD software for paternity assessment and pedigree determination.

Due to the challenges associated with hybridisation, Chiou & Bergey, (2018) developed a Fecal-Seq method that utilises methyl complex (made up of Methyl-CpG-binding domain proteins, Fc tail of human IgG1 and Protein A immunoprecipitation beads) that selectively attach to and enrich DNA with high CpG-methylation densities. Natesh *et al.*, (2019) introduced MPCRseq, an affordable, rapid and easy genotyping approach that leverages next-generation sequencing of multiplex PCR amplicons derived from non-invasive specimens such as stool, animal hair, and processed biological samples. This method generates enough SNP genotype data for population differentiation, individual identification and close relatives of endangered wild tigers.

Most advancements have been effective in generating large-scale genomic data from non-invasive samples. Nevertheless, Natesh *et al.*, (2019), demonstrates the usefulness of small-scale genomic data. Such small-scale genomic data are particularly relevant in developing countries, where high rates of wildlife population decline and extinction necessitate ongoing monitoring of ecological parameters that do not require large quantities of genetic data (Newmark & Hough, 2000).

Therefore, it is in these developing countries characterised as resource-limited areas that are in urgent need for a point-of-need, cost-effective, field-based protocol for genotyping animal populations to assess the effects of challenges facing this population.

2.4 Statistical and Software Development in Parentage Analysis

With increased genotyping success of non-invasive samples, there is also need for the development and improvement of software and/or statistical methods that will accommodate associated inadequacy of non-invasive genetic data. Over the years, there have been continuous improvements in methodologies used in parentage analysis such as availability of molecular markers, advancements in statistical theories and software development.

Examples of such software include: WHODAD (Snyder-Mackler *et al.*, 2016), GERUD (A. G. Jones, 2005), KINGROUP (Konovalov *et al.*, 2004), FAMOZ (Gerber *et al.*, 2003), PATRI (Signorovitch & Nielsen, 2002), PARENTE (Cercueil, 2002), PAPA (Duchesne *et al.*, 2002), CERVUS (Slate *et al.*, 2000; Marshall *et al.*, 1998) and PROBMAX (Danzmann, 1997). These software differ significantly based on the parentage analysis methods implemented in their algorithms. According to Jones *et al.*, (2010), there are six main techniques: categorical allocation, exclusion, full probability parentage analysis, parental reconstruction, fractional allocation and sibship reconstruction. Other factors that differentiate the parentage software are the programming language, type of input datasets, ability to accommodate genotyping and scoring errors and the format of output files (A. G. Jones *et al.*, 2010).

CERVUS, one of the most widely used software for paternity analysis, employs a categorical allocation approach (Jones *et al.*, 2010). This method relies on the calculation of a log-likelihood ratio (LOD score), which quantifies the likelihood that a candidate individual is the true parent of a given offspring relative to the likelihood of being an unrelated individual within the population (Marshall *et al.*, 1998).

CERVUS computes scores for all possible parent-offspring combinations and assigns parentage based on these scores (Slate *et al.*, 2000). The software also utilises simulation-

based methods to define critical values for the variation in LOD scores between the top two candidate parents, thereby enhancing confidence in the assignment of paternity.

Some desirable merits of CERVUS are its ability to identify and estimate the frequencies of null alleles. The estimation of null alleles is conducted through loci that do not conform to the Hardy-Weinberg equilibrium. Other advantages include the simulation of inbreeding, assignment of individual parent pairs and support for an open population framework, allowing for the possibility that a subset of true parents may be absent from the sampled individuals. (Slate *et al.*, 2000).

However, in parent –pair analyses, the parental sex may be either known or unknown. Another limitation of the software is its assumption that all loci exhibit uniform error rates and that genotyping errors occur randomly. In reality, genotyping error rates can vary significantly across loci, posing challenges to the validity of the assessment. (A. G. Jones *et al.*, 2010). CERVUS also limits the users in terms of data visualization options, featured graphical user interface (GUI), the ability to customise the server codes to fit their own need and is intended for users with significant coding experience to understand the mechanism and workflow of the application.

In this study, a shiny web application was developed that addresses some of the limitations experienced by CERVUS. This includes a user-friendly interface, a wide variety of graphical representations, ability to share codes and or reproduce the codes to suit the needs of different researchers. Given that Shiny web applications are developed using a popular R programming language, they are easily adaptable, particularly in countries with limited technological resources, hence they can be used as active learning tools.

CHAPTER THREE

MATERIALS AND METHODS

3.1 Study Site

The study investigated free-ranging baboon population inhabiting the Amboseli ecosystem in southern Kenya, including areas within and surrounding Amboseli National Park (Alberts & Altmann, 2012) as shown in Fig. 3 .1. The study site lies at approximately 2°40'S, 37°15'E and at an altitude of around 1,100m above sea level, north and west of Mount Kilimanjaro. The habitat is characterised by semi-arid, short-grass savannah with Acacia woodlands and swamps. The region experiences a strongly seasonal climate, with a prolonged dry season (June to October) and a wetter period (November to May). Mean annual rainfall is relatively low, averaging about 355 mm, and variability in precipitation frequently results in drought conditions. Temperatures are generally highest during February and March, often exceeding 30 °C, while cooler conditions occur during the June–August dry season (Alberts & Altmann, 2012).

The study population currently consists of approximately 300 individually recognized baboons distributed across five study groups namely Hokey, Yoda, Acacia, Narasha and Europa which vary in size between 10-120 individuals. These groups have been continuously monitored by the ABRP since 1971. Demographic data, such as births and deaths, along with behavioural observations of the population, are collected and recorded daily by experienced field observers. The wild baboon population mainly consists of yellow Savanna baboons (*Papio cynocephalus*), with evidence of genetic admixture with closely related olive baboon (*Papio anubis*) (Alberts & Altmann, 2001; Tung *et al.*, 2008).

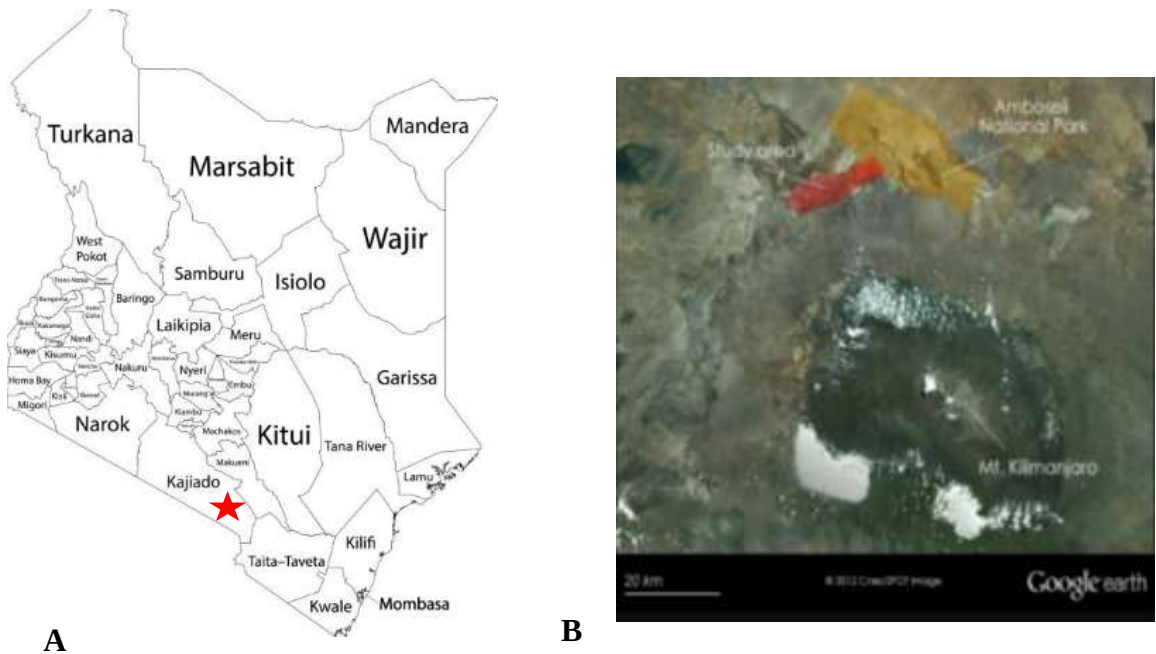


Figure 3.1: Representation of the Study Site

- A) Red Star represents the location of Amboseli study site on a map of Kenya
- B) Google map presenting the study area in relation to Mt. Kilimanjaro and Amboseli National Park.

3.2 Sample Collection

Faecal samples were collected using a non-invasive approach from one troop of wild baboons, known as the Europa group. At the time of sampling, Europa group (n=13 individuals) was the smallest of the five troops monitored by the ABRP. This made it convenient for evaluating the effectiveness of the proposed protocols in correctly inferring pedigree structure, as faecal samples could be readily obtained from all members of the troop for subsequent analysis. The Amboseli baboons are well-habituated to field-observers; therefore, only freshly deposited faecal samples were collected from animals that were individually recognised and tracked through ABRP long-term field records and observer-based identification. (Alberts & Altmann, 2012) as shown in Figure 3.2.

The outermost lining of the faecal bolus was targeted since it is estimated to contain the most host epithelial cells, sourced from the gut lining (Snyder-Mackler *et al.*, 2016).

Whatman swabs were carefully coated with faecal material and immersed in AL buffer, which primarily consists of guanidine hydrochloride and chaotropic maleic acid. In addition, approximately 2.5 g of faeces from each individual was collected and immediately mixed into 1ml of 95% ethanol as explained in sections 3.2.1 and 3.2.3 respectively.



Figure 3.2: Collection of Freshly Deposited Faecal Samples From Known Baboons Following Identification by ABRP Field Observers

Faecal samples preserved in ethanol were transported at 4°C from the field in a cold box to the ABRP camp daily for temporary storage and thereafter to the Kenya Institute of Primate Research (KIPRE) laboratories for further sample processing after 5 days. All materials and equipment used were sterilised using ethanol, with careful handling to prevent cross contamination. The tweezers were thoroughly sterilized using alcohol swabs before using again on a different sample.

3.3 Evaluation of Whatman DNA Extraction Method

For optimization purposes, three variants of the DNA extraction protocol were tested using opportunistically selected faecal samples from 13 individuals:

- i. collection of a single Whatman swab followed by immediate immersion in extraction buffer;
- ii. collection of four Whatman swabs followed by immediate immersion in the extraction buffer;
- iii. collection of the leading outer layer of the faecal bolus followed by immediate immersion in 95% ethanol for column-based DNA extraction.

The third approach adheres to the standardised faecal collection and DNA extraction protocol routinely employed in previous studies (Charpentier *et al.*, 2012; Buchan *et al.*, 2003) and therefore serves as a positive control.

3.3.1. Whatman-Based DNA Extraction

Swabbing of the outer lining of the faecal bolus was performed and the Whatman No. 1 filter paper(s) (GE Healthcare, USA) were placed into an extraction tube containing 200 µl of AL buffer. Using sterilised tweezers, the filter paper(s) were then transferred to a wash tube containing 200 µl of Tris buffer and incubated at room temperature for 1 minute. After incubation, the Whatman filter papers were transferred into an elution tube containing a homogeneous mixture of 50 µl nuclease-free water supplemented with 1 µl of 10 mM dNTPs to enhance DNA stability and support downstream PCR amplification. This was followed by a second 1-minute incubation. The filter papers were subsequently discarded and with DNA containing eluate was kept at -20°C for downstream processing (Zou *et al.*, 2018).

3.3.2 Bead Clean Up

About 35 µl of room temperature AMPure XP bead solution was added to the DNA eluate from the three protocols in a standard 1.5 ml microtube and thoroughly mixed to achieve a homogeneous solution. The resulting mixture was allowed to settle at ambient temperature for 5 minutes to enable DNA binding to the magnetic beads. After incubation, the tube was set on a magnetic rack for 3 minutes for the formation of bead-DNA complex and solution clarification.

The supernatant was gently aspirated, and the bead-bound DNA was washed twice with 200 µl of freshly prepared 70% ethanol. Each wash was followed by careful removal of the ethanol to avoid disturbing the bead-bound DNA pellet. To elute DNA, the pellet was dissolved in 13.5 µl of elution buffer, mixed thoroughly and left at room temperature for 5 minutes.

The tube was transferred to the magnetic rack for 2 minutes to allow bead separation, resulting in a clear DNA eluate. A volume of 12.5 µl of the clear DNA eluate is moved into a labelled 1.5 ml microtube and stored at -20°C for subsequent processing (Saiyed *et al.*, 2006).

3.3.3 Column - Based DNA Extraction

The Whatman-based DNA extraction method was evaluated in comparison with a conventional column-based DNA extraction protocol, specifically the QIAamp Stool Mini Kit (QIAGEN, Hilden, Germany), using non-invasively collected faecal samples stored in 95% ethanol. Extractions proceeded following the manufacturer's instructions, except for four modifications (Charpentier *et al.*, 2012; Buchan *et al.*, 2003). First 500 mg of faecal sample was used instead of 200 mg. Second, InhibitEX tablets were omitted; instead, 1.2 ml of the resulting supernatant was evenly divided between two sterile 2 ml microcentrifuge tubes, each preloaded with 25 µl of Proteinase K (Buchan *et al.*, 2003).

Third, 1.6 ml of ASL buffer was utilized during the lysis step, replacing the standard 1.2ml volume.

Fourth, the incubation during the lysis step was performed overnight as opposed to the standard 10 minutes incubation stated in the QIAGEN protocol in order to enhance cell disruption and improve DNA yield from low-biomass faecal samples. No modifications were made to the lysis buffer composition (Charpentier *et al.*, 2012).

3.3.4 Multiplex PCR

DNA extracted from each of the three protocols was subjected to multiplex PCR amplification targeting microsatellite loci. Specifically, multiplex microsatellite PCR was performed using a primer cocktail comprising five pairs of baboon-specific primers, each designed to target a distinct microsatellite locus previously used in paternity analysis within the Amboseli baboon study population (Buchan *et al.*, 2005). Specifically, we used 5 tetranucleotide primers AGAT006, D1s1656, D4s243, D8s1106 and D11s2002 as described by Buchan *et al.*, (2005) as listed in Table 3.1.

Table 3.1: Primers that Target Microsatellite Loci Specific for Paternity Studies in Wild Baboons' Population as Described in Buchan et al., (2005)

Target Locus	Locus size(bp)	Primer Sequence
AGAT006	131-181	5' AGTGGATCGATAGATTGACAGATG 3' 3' TCAGGTGACAGCCAAGTCAATTCA 5'
D1s1656	130-167	5' GTGTTGCTCAAGGGTCAACT 3' 3' GAGAAATAGAATCACTAGGGAACC 5'
D4s243	155-179	5' TCAGTCTCTCTTTCTCCTTGCA 3' 3' TAGGAGCCTGTGGTCCTGTT 5'
D8s1106	128-161	5' TTGTTTACCCCTGCATCACT 3' 3' TTCTCAGAATTGCTCATAGTGC 5'
D11s2002	128-161	5' CATGGCCCTTCTTTTCATAG 3' 3' AATGAGGTCTTACTTTGTTGCC 5'

A total reaction volume of 24 µl was used for each PCR setup, comprising of nuclease-free water (18.15 µl), PCR buffer (2.5 µl), magnesium chloride (0.75 µl), 10mM dNTPs (0.5 µl), primer cocktail (1.0 µl), DNA template (1.0 µl), and 0.1 µl of Invitrogen™ Platinum™ Taq DNA polymerase (Thermo Scientific Inc., USA).

A positive control PCR amplification was conducted alongside the multiplex microsatellite PCR. For this control, primers developed by Caporaso *et al.* (2012) were employed to target the V4 hypervariable region of the 16S rRNA subunit, specifically the conserved flanking site defined by the 515F/806R primer pair. The loci targeted in the Caporaso reaction were amplified under the same optimised thermal conditions used for the multiplex PCR in Table 4.2 (Refer to Chapter 4 Results).

Table 3.2: Caporaso Primer that Targets 16S rRNA Subunit of Bacterial DNA Used as a Positive Control for PCR Amplification as Described by Caporaso et al (2012)

Candidate Gene	Locus size(bp)	Primer Sequences
515F/806R	300	5' GTGCCAGCMGCCGCGGTAA 3' 3'TAATCTWTGGGVHCATCAGG 5'

Thus, these primers amplify bacterial DNA sequences present within the faecal DNA extract. Since over 95% of extracted DNA in these samples is believed to be bacterial, not endogenous baboon DNA, this PCR reaction serves as a quality control to verify the presence of amplifiable DNA (Snyder-Mackler *et al.*, 2016). Successful amplification with the 16S rRNA primers, coupled with failure of amplification with the validated microsatellite primer cocktail, would suggest that although DNA extraction was effective, the quantity or quality of baboon-specific DNA was insufficient for reliable amplification. Subsequently, each 24 µl PCR amplicon was run on a 1.5% agarose gel for visualisation as a quality check for successful amplification as described in section 3.3.3.

3.3.5 Agarose Gel Electrophoresis

A 1.5% (w/v) agarose gel was prepared by dissolving 2.1 g of agarose powder in 140 mL 1X of TAE buffer, then heating the mixture in a microwave for 30 seconds. After cooling, ethidium bromide was added into the solution prior to casting the gel in a tray fitted with combs, and the gel was allowed to solidify. Once polymerised, the gel was placed in an electrophoresis tank with 1X TAE buffer. An aliquot of 6 μ L of each PCR amplicon was mixed to homogeneity with 1 μ L of loading dye and loaded into the wells. Electrophoresis was carried out at 100 V for 1 hour. DNA bands were visualized using UV transilluminator and gel images were captured using a smartphone camera. Fragment sizes were estimated by comparison with a standardized 100 bp DNA ladder (New England Biolabs, USA; catalogue number N0551S/L).

3.4 Implementation of a User-Friendly Shiny-Based Web Application for Determining Paternity Using Microsatellite or SNP Markers

R (<https://www.R-project.org/>), an extensively used programming language for statistical computing and data visualisation in the biological sciences (R Core Team, 2018), was used to implement an interactive Shiny-based online dashboard via the *Shiny* R-package (Chang *et al.*, 2015).

3.4.1 Designing a Likelihood/LOD-Based Paternity Assignment Model that Uses Likelihood Ratio to Assign Paternity Using R Programming Language

First, likelihood/LOD-based paternity assignment model was developed using the R programming language to calculate the paternity likelihood score using the logarithm of odds (LOD) score in accordance with Hardy-Weinberg equilibrium as illustrated and used in other tools such as CERVUS (Slate *et al.*, 2000). The simulation model encoded genotype calls for biallelic SNPs as 0, 1, or 2 where 0 represented homozygous reference genotype, 1 as heterozygous and 2 as homozygous alternate (Danecek *et al.*, 2011).

The likelihood method utilises likelihood ratio (LOD) scores to assess whether the currently assigned male has a higher probability of being the true father than a randomly selected potential male from the population. This principle was described by Marshall *et al.*, (1998) who defined likelihood ratio as the probability of observing the genetic data assuming that the candidate male is the true father, divided by the probability of observing the same data assuming the candidate male is not related to the offspring (A. G. Jones *et al.*, 2010).

To calculate the LOD score, the following formula was used

$$\text{LOD} = \log_{10}(L(H_2)/L(H_1))$$

Where L represents likelihood, H_1 represents the hypothesis that the candidate male is the true father of the target offspring. H_2 represents the hypothesis that the candidate male is unrelated to the target offspring. H_1 and H_2 utilizes the probabilities based on Hardy Weinberg equilibrium and Mendelian segregation, of offspring's genotype. This calculation uses the assumption that the genotype of the father or both parents (father and mother) are of co-dominant autosomal locus as stated Marshall *et al.*, (1998).

A positive LOD score signifies high probability that the putative male is the true father of the target offspring whereas a negative LOD score signifies a low probability that the putative male is the true father of the target offspring. After every run, once all LOD scores of every probable fathers have been calculated, the real father is characterised as the putative father with highest LOD score with a statistical confidence determined by subtracting the LOD score of the second assigned putative father from LOD score of the assigned male.

The model calculates LOD score for two scenarios: i) The presence of maternal genotyping dataset; ii) absence of the genotype of the mother. Below are the formulae of calculating LOD scores for the two scenarios:

Presence of mother's genotype: **Formulae 1**

$$L(H_1, H_2 \mid g_m, g_f, g_o) = \frac{T(g_o \mid g_m, g_f)P(g_m)P(g_f)}{T(g_o \mid g_m)P(g_m)P(g_f)} = \frac{T(g_o \mid g_m, g_f)}{T(g_o \mid g_m)}$$

where:

- g_m = maternal genotype
- g_f = paternal genotype
- g_o = offspring genotype
- $T(\cdot)$ = transmission probability
- $P(\cdot)$ = genotype probability frequency in the population.

Absence of mother's genotype: **Formulae 2**

$$L(H_1, H_2 \mid g_f, g_o) = \frac{T(g_o \mid g_f)P(g_f)}{P(g_f)} = T(g_o \mid g_f)$$

where:

- g_f = paternal genotype
- g_o = offspring genotype
- $T(\cdot)$ = transmission probability
- $P(g_f)$ = genotype frequency probability in the population.

In both formula 1 and 2, the numerator corresponds to the first hypothesis, H1, which assumes the putative male is the true father whereas the denominator corresponds to the alternative hypothesis, H2, in which the putative male is unrelated to the target offspring. g_o , g_f , g_m represents the genotypes of the target offspring, alleged father and mother respectively.

The transition probabilities, $T(g_o \mid g_m)$, $T(g_o \mid g_f)$, $T(g_o \mid g_m, g_f)$, represents the Mendelian likelihood of observing the target offspring's genotype (g_o), given the mother genotype alone (g_m), the alleged father's alone (g_f) or both parents (g_m, g_f) respectively. These

Mendelian probabilities are derived using the Hardy Weinberg equilibrium and below is a list as stated by Marshall *et al.*, (1998).

Where PP is the homozygous reference QQ represents the homozygous alternate, RR represents genotypes other than P and Q. p,q,r denotes the frequency of homozygous reference, homozygous alternate and genotype other than P and Q. It is a similar table but more detailed to tables in (Brenner, 1997) and (Marshall *et al.*, 1998).

Table 3.3: Likelihood Ratio for Various Trio Paternity Combinations

Offspring genotype	Probable father's genotype	Mother's genotype	Transition Probability	Probability of offspring when mother's genotype is present	Likelihood
PP(p)	PP(p)	PP(p)	1.00	p	1/p
PP(p)	PQ(pq)	PP(p)	0.50	p	0.5p
PP(p)	PP(p)	PQ(pq)	0.50	0.5p	1/b
PP(p)	PQ(pq)	PQ(pq)	0.25	0.5p	0.5p
PQ(pq)	PP(p)	RR(r)	1.00	p	1/p
PQ(pq)	PQ(pq)	RR(r)	0.50	p	0.5p
PQ(pq)	PP(p)	RQ(rq)	0.50	0.5p	1/p
PQ(pq)	PQ(pq)	RQ(rq)	0.25	0.5p	0.5p
PQ(pq)	PP(p)	PQ(pq)	0.50	(p+q)/2	1/(p+q)
PQ(pq)	PR(pr)	PQ(pq)	0.25	(p+q)/2	0.5(p+q)
PQ(pq)	PQ(pq)	PQ(pq)	0.50	(p+q)/2	1/(p+q)
PQ(pq)	QQ(q)	PP(p)	1.00	p	1/p
PQ(pq)	QQ(q)	PQ(pq)	0.50	(p+q)/2	1/(p+q)
PQ(pq)	QR(qr)	PP(p)	0	0.5p	impossible
QQ(q)	QQ(q)	PP(p)	0	0p	impossible
QQ(q)	QR(qr)	PP(p)	0	0p	impossible
QQ(q)	QR(qr)	PQ(pq)	1	0.5p	0.5p
QQ(q)	QR(qr)	QQ(q)	1	0.5p	0.5p

Table 3.4: Likelihood Ratio for Various Duo Paternity Combinations

Offspring genotype	Probable father's genotype	Transition Probability	Probability of offspring when father's genotype is present	Likelihood
PP(p)	PP(p)	1p	p^2	1/p
PP(p)	PQ(pq)	0.50p	p^2	0.5p
PQ(pq)	PP(p)	1q	2pq	0.5p
PQ(pq)	PR(pr)	0.5q	2pq	0.25p
PQ(pq)	PQ(pq)	$(p+q)/2$	2p+q	$(p+q)/4pq$
QQ(qq)	RR(rr)	0	0	impossible

To test the simulation model, during each single run, a true father amongst the 100 potential fathers were determined and verified using independent methods. A total of 100 simulations were performed and each iteration began by randomly assigning genotypes 0, 1 and 2 to a cohort of 100 independent males and 100 independent females. In the first scenario, the first male and female were designated as the true father and mother respectively, whereas in the second scenario, only the putative fathers' genotypes were considered.

After developing the likelihood/LOD-based paternity assignment model, the accuracy of paternity assignment was assessed by measuring the proportion of correctly assigned paternity cases in scenarios with and without maternal genotype calls. Additionally, the impact of genotype call was assessed by examining how varying the number of loci influenced the accuracy of paternity assignment, both in scenarios that either include and excluded maternal genotype calls.

3.4.2 Implementation of a Shiny Web Application that Uses a Likelihood Approach to Assign Paternity

Using the simulation model, a Shiny App was developed to calculate allele frequencies at each locus and determine the probability of paternity for every potential father, given the genotype calls of the offspring of interest.

The app first computes the probability of paternity at each locus and then derives the overall likelihood for each potential father by multiplying these probabilities across all loci. This likelihood is stored in a variable called *likelihood_loci*.

To account for Hardy Weinberg equilibrium, the app calculates the probability of observing the offspring's genotype if a randomly selected male is the biological father. These probabilities are stored in a variable called *hardy_loci*, and their product is saved as *hardy_likelihood*

Finally, the likelihood ratio (LOD scores) for each potential father is obtained by dividing *likelihood_loci* by *hardy_likelihood*. LOD scores offer an estimate of how much more likely a specific male is to be the biological father of a target offspring than a randomly selected male from the population, as described in Figure 3.3.

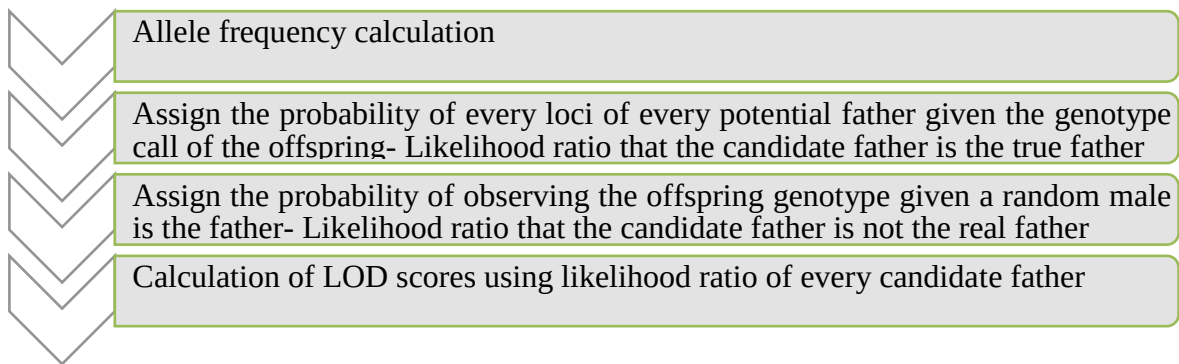


Figure 3.3: Flow Chart Illustrating Steps in the Likelihood/LOD-Based Paternity Assignment Model of the Shiny Web Application

3.4.3 Suitability of *DadApp*. A Case Study for Determining Paternity in a Wild Baboon Population in Amboseli Baboon Research Project (ABRP)

Because the cellulose-based extraction workflow did not yield reliable host microsatellite amplification, *DadApp* was validated using an independent curated ABRP genotype dataset with known paternity relationships from the wild baboon population inhabiting the

Amboseli ecosystem, as described by Vilgalys *et al.*, (2022). Since 1971, ABRP has conducted nearly daily monitoring of the demography and behaviour of individually recognised baboons within this population (Alberts & Altmann, 2012). The population mainly consists of yellow baboons with documented admixture from the closely related olive baboon (Tung *et al.*, 2008).

The data were generated through whole-genome resequencing, enabling the estimation of both local and global ancestry for each distinct baboon using a composite likelihood method (Vilgalys *et al.*, 2022). The dataset comprised 99 randomly sampled biallelic loci for 30 individuals (10 randomly sampled pedigrees of both dyads and triads each). These genotypes of the 99 loci of each baboon were encoded as 0 for homozygous reference, 1 for heterozygous, 2 for homozygous alternate genotypes.

3.5 Evaluation of Developed Shiny Web Application as an Active Learning Tool for Both Molecular and Computational Aspects of Paternity

Finally, an evaluation was conducted to determine the effectiveness of an R shiny web application as an interactive educational tool. The evaluation focused on its potential for teaching key concepts in genetic data analysis, Mendelian inheritance, and paternity inference. An intervention-based design was employed in which the intervention's efficacy (the use of the Shiny web application as an educational tool) was assessed through comparisons of pre- and post-intervention tests and surveys. This was carried out through a workshop study which was conducted at the Kenya Institute of Primate Research (KIPRE), Kenya from 4th March 2024 to 25th March 2024 (Refer to Appendix III). The assessment and survey were reviewed and approved (Protocol No. P689/09/2023) by the Ethics and Research Committee of Kenyatta National Hospital-University of Nairobi.

The inclusion criteria were as follows:

1. Participants provided consent to the study

2. Participants should have had no research or formal expertise in genetic reconstruction of kinship or paternity.
3. Participants should have had an interest in behavioural or ecological monitoring, and/or wildlife management.
4. Participants should have at least completed or be currently enrolled in an undergraduate course in ecology, conservation, biology, or genetics.
5. Participants were computer literate.

The exclusion criteria include:

1. Participant did not consent to the study
2. Participants were an expert in genetic reconstruction of kinship or paternity.

Participants began with a one-hour, lecture-style introduction to paternity and kinship inference. Immediately afterwards, they completed a 30-minute pre-test to assess their baseline understanding of the concepts. Following the pre-test, participants engaged in a one-hour interactive session using the Shiny web application. The session included guided questions designed to help participants explore various scenarios related to paternity inference.

The interactive session took place in a classroom setting, with participants working in groups of three to six, each using their own laptop. Immediately after interacting with the app, participants completed a 30-minute post-test. In this phase, participants were encouraged to discuss and reflect on the material with their group members.

To evaluate perceived changes in understanding and interest, participants also completed an online retrospective survey. The survey included 5-point Likert scale questions assessing self-reported confidence in genetics, Mendelian inheritance, and paternity inference (e.g. 1-Not Confident, 2-Somewhat Confident, 3-Don't Know, 4-Confident, 5-

Extremely Confident) and interest and enthusiasm in the topic (e.g. 1-Very Disinterested, 2-Disinterested, 3-Neutral/No opinion, 4-Interested, 5-Very Interested).

Quantitative analysis involved comparing the mean scores and distribution of responses before and after the intervention. The Wilcoxon signed-ranks test was employed to assess whether the variance in individual scores was statistically significant. Paired t-tests were applied to evaluate differences in pre- and post-test scores and to determine the effectiveness of the interactive learning component.

In addition to analysing pre- and post-test performance, the study considered participant gender (male, female or prefer not to say) and educational background. Educational levels were categorised into three groups: undergraduate (currently enrolled in or having completed undergraduate studies), MSc-level (enrolled in or completed a Master's program), and PhD-level (enrolled in or completed a PhD). These educational background categories were selected to reflect distinct levels of formal educational attainment within the study population, enabling consistent grouping for comparative analysis.

To explore how participants' characteristics might influence learning outcomes, a linear regression model was used to test if educational background, self-reported interest or confidence influenced the percentage change in test scores from pre-test to post-test.

CHAPTER FOUR

RESULTS

4.1 Sample Collection

Freshly deposited faecal samples were collected in December 2019 exclusively from thirteen individually recognized baboons by the ABRP field observers (shown in Table 4.1), from the Europa troop at the time of sampling.

Table 4.1: Baboon Identification and Demographic Information for Sampled Individuals from Europa Troop Found in ABRP

Animal I.D	Sex	Age Classes (as per the date of sampling)	Date of Sampling
Vogue	Female	Adult	2 nd December 2019
Vuyani	Male	Juvenile	2 nd December 2019
Tigo	Female	Adult	2 nd December 2019
Rapel	Male	Sub-adult	2 nd December 2019
Vyema	Male	Juvenile	4 th December 2019
Telly	Female	Adult	4 th December 2019
Vakko	Male	Adult	4 th December 2019
Toyota	Male	Infant	5 th December 2019
Vomer	Male	Infant	5 th December 2019
Vugu	Female	Adult	5 th December 2019
Veja	Female	Infant	5 th December 2019
Ujumbe	Male	Sub- adult	6 th December 2019
Ayok	Male	Adult	7 th December 2019

4.2 Evaluation of Whatman DNA Extraction Method

Four faecal DNA samples were opportunistically selected from donors, namely Vogue, Vugu, Telly and Tigo for the subsequent molecular assays to provide a representative subset for preliminary optimisation and assessment of assay performance across individuals. Furthermore, DNA extracted from baboon blood obtained from previous

studies (referred to as BabDNA) and DNA extracted from a bacterial colony of *Bacillus* species (referred to as BactDNA) were employed as additional reaction controls.

All loci were targeted using specific primers described in materials and methods chapter 3 (refer to Table 3.2) and amplified using an optimized touchdown PCR assay as described below in Table 4.2.

Table 4.2: Optimized Touchdown PCR Thermal Conditions for Amplification of Microsatellite Loci Specific for Paternity Studies in Wild Baboons

Program	Duration	Cycles
Initial Denaturation	94 °C for 2minutes	1
Denaturation	94 °C for 30 seconds	16
Annealing	66 °C -50 °C (the temperature was decreased by 1°C between consecutive steps)	
Extension	72 °C for 1 minute	
Denaturation	94 °C for 30 seconds	29
Annealing	50 °C for 30 seconds	
Extension	72 °C for 1 minute	
Final Extension	72 °C for 1 minute	1
Holding stage	4 °C	

The positive test control characterized by the Caporaso primer run was successful, as visible bands were observed in all DNA extracts including (1X) Whatman, four (4X) Whatman and column-extracted DNA (Lane 4 – Lane 15). However, the (4X) Whatman extracted DNA was characterized by faint bands (Lanes 6, 9, 12, 15). No visible band was observed in the no-template control indicating that no contamination occurred (Lane 2). Representative gel image of DNA extracts for the samples are shown in Figures 4.1



Figure 4.1: An Agarose Gel Image of PCR Amplicons

An agarose gel image of PCR amplicons generated using Caporaso primers (515F/806R) targeting the bacterial 16S rRNA gene, with an expected amplicon size of approximately 291bp. Clear bands at the expected size were interpreted as successful amplification of bacterial DNA. Lane identities were as follows: (1- Baboon DNA, 2- No Template Control (NTC), 3- Bacteria DNA, 4-6- Telly, 7-9- Vugu, 10-12- Vogue, 13-15- Tigo)

Within each individual, the lanes represented column extraction, 1X Whatman extraction, and 4X Whatman extraction, respectively. The absence of amplification in the NTC confirmed the lack of contamination, while amplification in the positive controls confirmed successful PCR performance.

With regard to the microsatellite reaction, only one visible band was observed within the target region (100-200bp) that was DNA extracted from baboon's blood. The remaining of the DNA extracts showed primer dimers which were below 100 bp. Representative gel images of DNA extracts for the samples are shown in Figures 4.2.



Figure 4.2: An Agarose Gel Image Showing PCR Amplicons

An agarose gel image showing PCR amplicons generated using the Microsatellite cocktail primers set targeting baboon microsatellite loci. Clear bands at the expected amplicon sizes were interpreted as successful amplification of the host DNA. Lane identities were as follows: 1- Bacteria DNA, 2- No Template Control (NTC), 3- Baboon DNA positive control, 4-6- Telly, 7-9- Vugu, 10-12- Vogue, 13-15- Tigo)

Within each individual, the lanes represented column extraction, 1X Whatman extraction, and 4X Whatman extraction, respectively. The absence of amplification in the NTC confirmed the lack of contamination, while amplification in the positive controls confirmed successful PCR performance.

4.3 Shiny Web Application Development - *DadApp*

Using a likelihood/LOD-based paternity assignment model, an R shiny web app (named “*DadApp*”) which is diverse with several variations, easy to use, and is easily shareable thereby enhancing reproducibility was developed. *DadApp* is publicly available via Posit’s shinyapps.io platform and can be accessed at <https://kiragu-mwaura.shinyapps.io/dadapp/>. The user interface of *DadApp* is shown in Figure 4.3.

Additional annotated source code and instructions for execution is accessible through the Github repository: https://github.com/KIRAGU-MWAURA/DadApp_Shiny_Web_App.

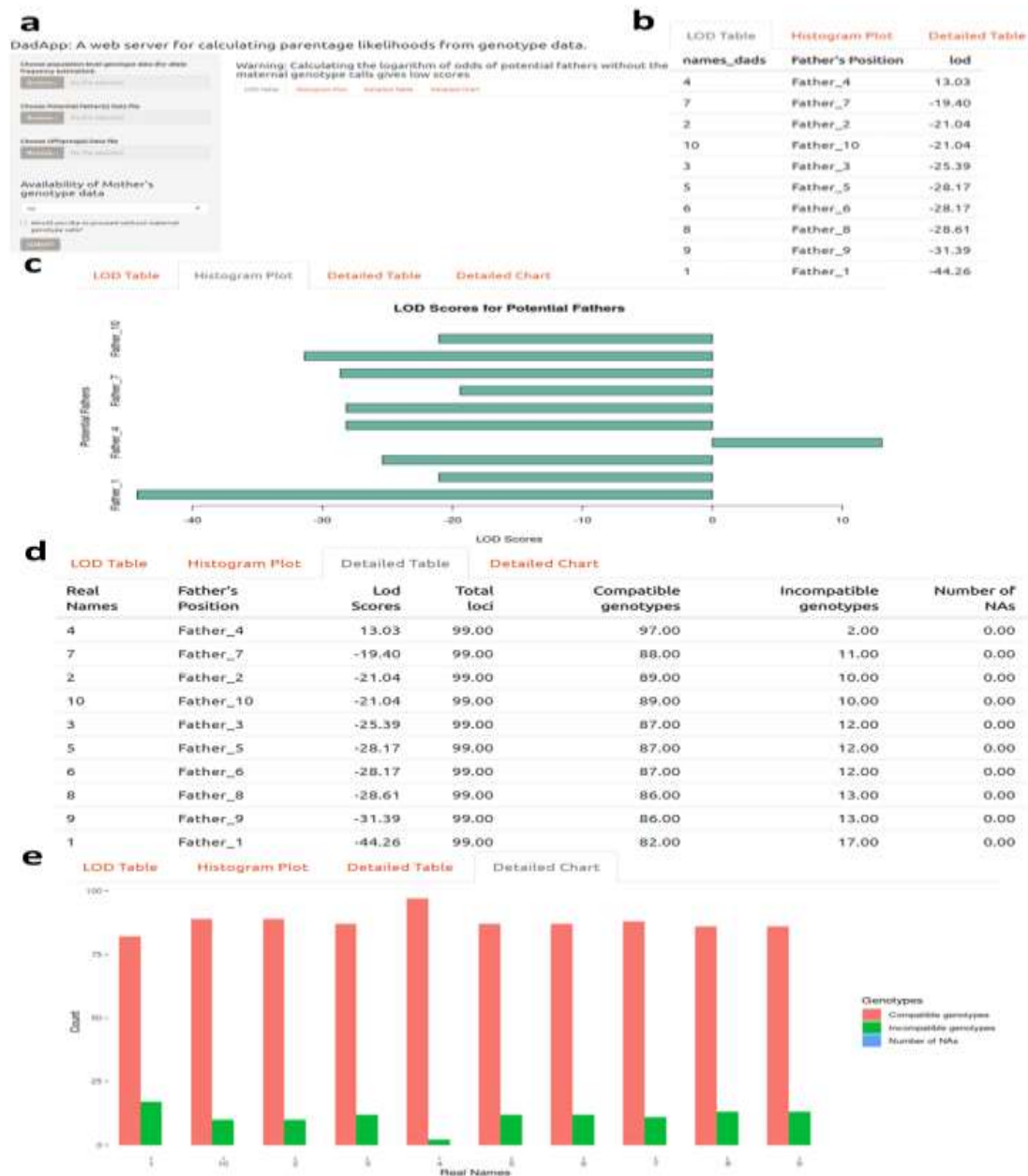


Figure 4.3: Represents the User Interface of the Shiny Application

a) The default user interface of *DadApp* upon launch;

- b) Tabular output listing candidate fathers with their associated LOD scores;
- c) Bar plot of the LOD scores of candidate fathers
- d) Detailed tabular output listing candidate fathers with their associated LOD scores, total number of loci used, compatible genotypes, incompatible genotypes, and loci with missing data;
- e) Bar plot summarising the number of compatible genotypes, incompatible genotypes, and loci with missing data (denoted as NA) for each candidate father.

4.3.1 *DadApp* Paternity Assignment Aligns with Real Data of Wild Baboons

DadApp was able to assign correctly 100% of all the 20 paternal individuals in 20 confirmed triad relationships. These assignments had an average of 12.08 statistical confidence which is calculated by subtracting the LOD score of the second most probable father from LOD score of the assigned father (Refer to Appendix II).

In all instances, the top-ranking LOD scores matched the paternity assignments established in the ABRP pedigree, which was based on independent analyses of 6-14 microsatellite markers (Alberts *et al.*, 2006). This demonstrates that *DadApp* is an accurate implementation of the CERVUS algorithm within an R Shiny web application, making it a viable alternative tool for basic paternity assignment using compatible real-world datasets.

4.4 *DadApp* as a Potential Interactive Educational tool

Twenty-eight (28) participants of different self-identified genders and self-reported educational backgrounds (Figure 4.4) who met the selection criteria were voluntarily recruited into the study through direct outreach and referrals.

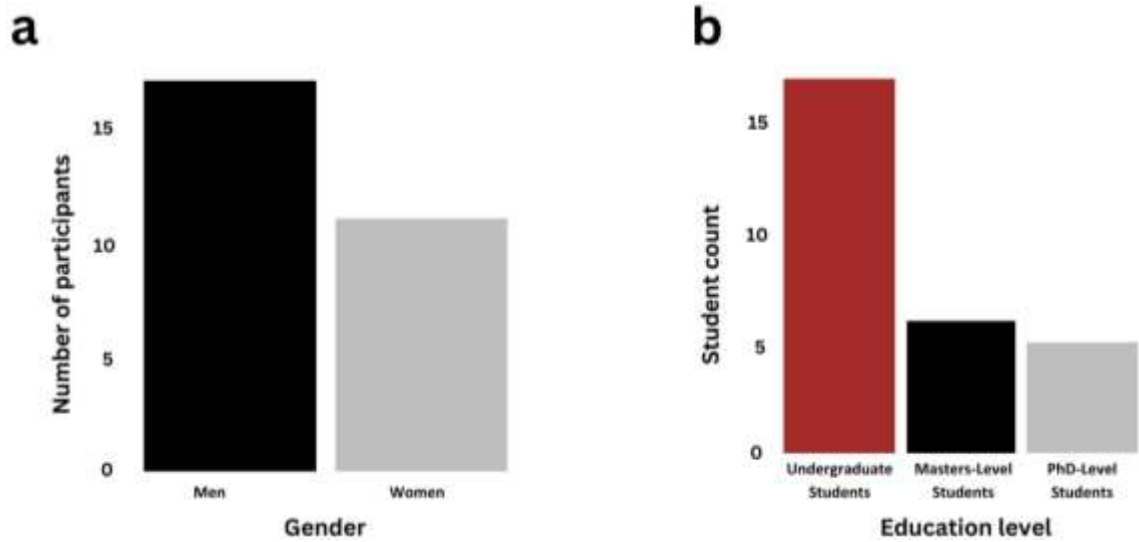


Figure 4.4: Bar Plot Illustrating the Number of Participants

a) Self-identified gender and

b) Self-reported education level.

An evaluation was conducted to determine whether *DadApp* could serve as an interactive educational tool for teaching key concepts in Mendelian inheritance and paternity inference outside of the context of a formal educational setting.

The mean scores of 28 participants (with range in parentheses) was pre-test 70.54% (45-90) and post-test 97.86% (80-100) as shown in Figure 4.5. A significant increase in scores from pre-test to post-test was observed within individuals (paired Wilcoxon signed-rank test $p = 3.75 \times 10^{-6}$)(see Figure 4.6).

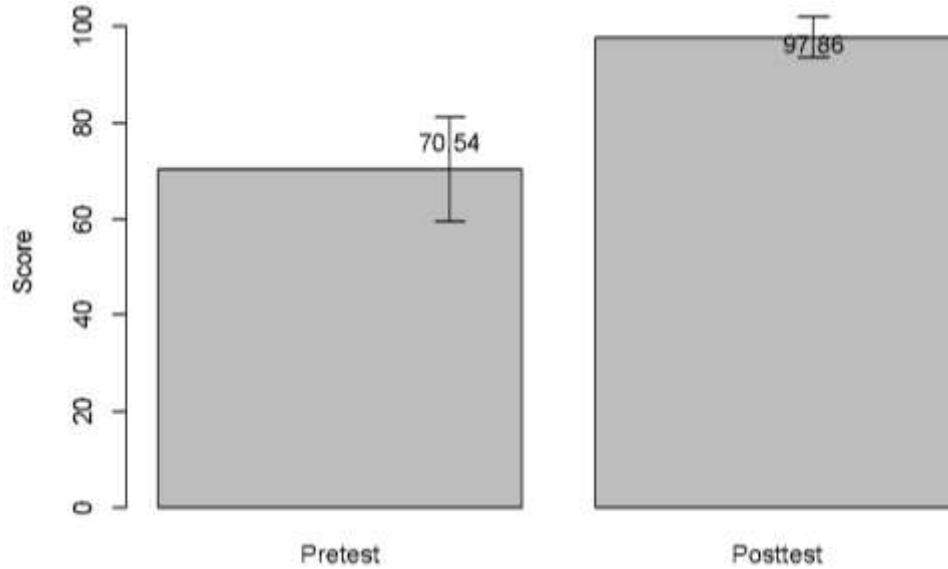


Figure 4.5: Bar Plot Depicting Individual Pre- and Post-Test Means with Standard Deviation, Measured before and after DadApp Exposure

Self-reported educational level showed no significant influence on pre-test performance (ANOVA: $df = 2$, $F\text{-value} = 2.037$, $p\text{-value} = 0.152$) nor post-test performance (ANOVA: $df = 2$, $F\text{-value} = 0.584$, $p\text{-value} = 0.565$). Similarly, neither self-reported confidence (linear regression: $\beta_{\text{confidence}} = -14.5147$, $p\text{-value} = 0.652$) nor interest (linear regression: $\beta_{\text{interest}} = -2.452414$, $p\text{-value} = 0.9152$) at the start of the study and after the introductory respectively did not significantly predicted the percentage change in the test scores.

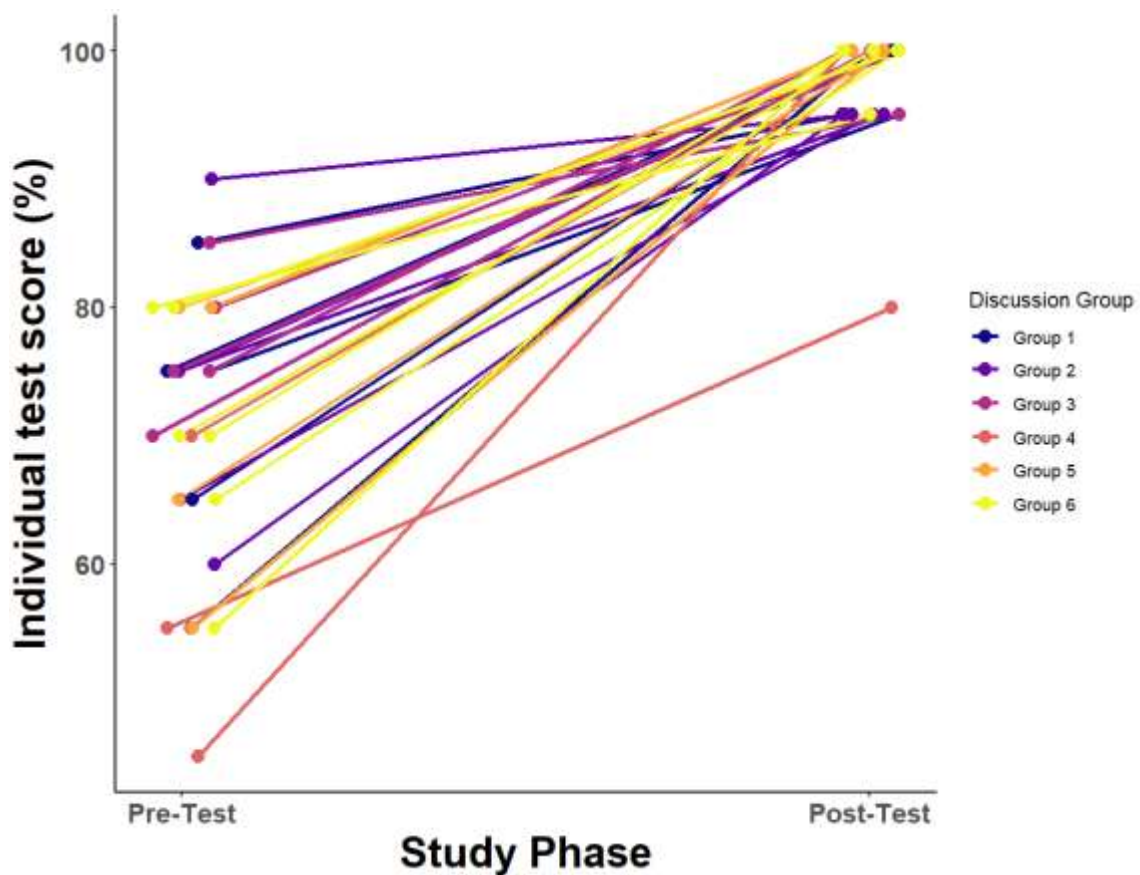


Figure 4.6: Within-Individual Comparison of Pre-Test and Post-Test Scores following Interaction with *DadApp*.

Each line represents a participant ($n=28$) and connects paired scores (pre- and post-test). Colours represent different discussion groups during the active learning and post-test phase. Data points were jittered within pre-test and post-test values to avoid overlap.

Participants completed a retrospective post-engagement survey to assess their experiences during interaction with *DadApp*. Participants' ratings to most survey questions spanned the entire 5-point Likert scale, indicating variability in participants' perceptions. No statistically significant difference in self-reported interest was observed between the retrospective ratings before and after the introductory lecture (mean interest rating before

lecture = 3.046 ± 1.48 s.d; after the lecture = 3.038 ± 1.40 s.d.; p-value=0.6114; paired t-test) as shown in Figure 4.7a.

In contrast, a highly significant increase in self-reported interest was observed following the interaction with the web app compared to the post-lecture ratings (p-value = 9.485×10^{-5} ; paired t-test, mean = 4.746 ± 0.489 s.d; Figure 4.7a). Also, participant confidence increased significantly between the post-lecture stage (mean= 4.038 ± 0.958 s.d.) and after interaction with the web app (mean = 4.769 ± 0.815 s.d.; paired t-test p = 8.103×10^{-4} ; Figure 4.7b)

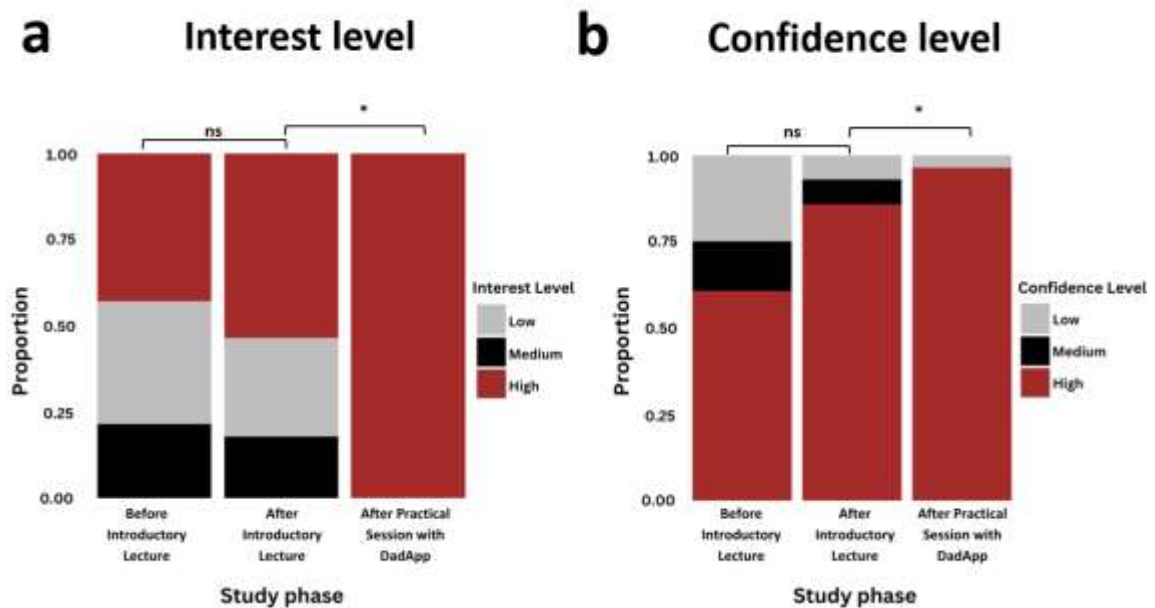


Figure 4.7: Bar Plots Illustrating Changes in Participants' Self-Reported Interest and Confidence across Stages of the Study

a) Change in participants' self-reported interest levels: non-significant difference (ns) after the introductory lecture, with a significant increase (p-value = 9.485×10^{-5}) after hands-on practical session with *DadApp*).

b) Change in participants' self-reported confidence levels: non-significant difference after the introductory lecture (ns), with a significant increase (p-value = 8.103×10^{-4}) after hands-on practical session with *DadApp*.

CHAPTER FIVE

DISCUSSION

5.1 Cellulose Based Extraction

Faecal DNA-based sampling is one of the continuing advancements in the non-invasive genotyping of wild animal populations (Zemanova, 2021). This technique has been previously used in characterizing several population parameters including estimation of population size, species identification, migration routes, sex determination, paternity, hybridization and dietary composition (Zemanova, 2020). Given prospects, there is a continuous need to develop faecal DNA sampling protocols to fit different researcher's needs and demands.

In this study, we tested whether swabbing the outer lining of faecal bolus using cellulose based Whatman No.1 paper discs can extract PCR-ready DNA from non-invasively collected faecal samples. We further tested whether the proposed protocol can be utilised to correctly infer pedigree structure of a well-studied baboon population. Two variant protocols were tested 1) using a single Whatman No.1 paper disc and 2) Using four Whatman No.1 paper discs. The main difference of the two variants was the number of cellulose discs which is hypothesized to increase the DNA binding surface area with downstream effects on the amount of purified DNA (Zou *et al.*, 2018). In this case, we hypothesized a positive correlation between the cellulose disks and purified DNA. This cellulose-based extraction method was compared to a gold standard protocol of column-based extraction method QIAamp Stool Mini kit QIAGEN, Hilden, Germany) from faecal samples in 95% ethanol (with modifications) (Charpentier *et al.*, 2012; Buchan *et al.*, 2003).

Unfortunately, the Whatman No.1 paper disc extraction method did not yield PCR ready DNA that can be used in pedigree construction (See Figure 4.1 samples labelled as 1X).

There could be several reasons as to why there was no success with amplification of paternity-specific genes from DNA extracted from faecal samples using Whatman No.1 paper disk. First, approximately 1% of the DNA obtained from faecal-derived samples originates from the donor animal itself, while the majority is microbial DNA (Snyder-Mackler *et al.*, 2016; Perry *et al.*, 2010). Extracted DNA endogenous to the host is often in trace quantities such that they are often difficult to quantify using conventional methods such as spectrophotometer (Beja-Pereira *et al.*, 2009).

Modification of PCR such as cell enrichment method, LAMP, real-time quantitative PCR and use of molecular probes are known to enhance recovery of high molecular weight DNA which were not applied in this study. Hence, this advantage can be shared with non-invasively collected samples to counter challenges associated with quantities and qualities of DNA extracted. Unfortunately, all these enrichment advancements on nucleic acid purification requires experienced personnel, expensive and advanced machineries (Valencia *et al.*, 2013) which cannot be conducive and effective in resource limited or field-based settings. Touchdown PCR was the only enrichment method that was used in this study. As described by Zou *et al.*, (2018), the major stages that can enhance the quality and yield of DNA using Whatman cards - or that warrants further advancements- is pre-treatment of cards or sample, using stronger wash buffers or robust methods, the introduction of heat elution, DNA amplification or enrichment, and the integration of this method with other DNA extraction techniques.

The PCR effectiveness was assessed using primers developed by Caporaso *et al.*, (2012) that targeted the V4 hypervariable region of the 16S rRNA subunit and it was successful. The success of the 16S rRNA reaction justifies that there is potential in the cellulose-based extraction method as an alternative DNA extraction protocol. This method has demonstrated utility in diverse applications, including microbiome profiling (Moldovan *et al.*, 2020), blood pathogen detection (Song & Gyarmati, 2022), and antimicrobial resistance surveillance (Benčič *et al.*, 2024) Especially, where there is need for a

simplistic, relatively cheap, field-based, point of need protocol particularly in resource limited settings whose research requires limited target sequence data.

The 16S rRNA reaction results correlates with the finds in Zou *et al.*, (2018) where they used the same protocol in extracting PCR amplifiable DNA from blood and faecal samples.

5.2 *DadApp* Shiny Web Application as a Tool for Assigning Paternity in Wild Baboon Populations

The increasing success of genotyping non-invasively collected samples necessitates the development of software and statistical methods capable of addressing the limitations of non-invasive genetic data (Beja-Pereira *et al.*, 2009). Parentage analysis methodologies have undergone continuous improvement over the years, incorporating advancements in molecular markers, statistical theory, and software development. Several software programs are available for this purpose, including WHODAD (Snyder-Mackler *et al.*, 2016), GERUD (A. G. Jones, 2005), KINGROUP (Konovalov *et al.*, 2004), FAMOZ (Gerber *et al.*, 2003), PATRI (Signorovitch & Nielsen, 2002), PARENTE (Cercueil, 2002), PAPA (Duchesne *et al.*, 2002), CERVUS (Slate *et al.*, 2000; Marshall *et al.*, 1998) and PROBMAX (Danzmann, 1997). These programs vary significantly based on their algorithms and techniques for parentage analysis, such as categorical allocation, parental reconstruction, exclusion, full probability parentage analysis, fractional allocation and sibship reconstruction (A. G. Jones *et al.*, 2010).

In this study, we re-implemented CERVUS software in R (<https://www.R-project.org/>), one of the most widely used statistical programming languages (R Core Team, 2018), utilizing the *shiny* package (Chang *et al.*, 2017). The R Shiny web app (named *DadApp*) uses categorical allocation to assess paternity and conduct pedigree reconstruction. *DadApp* was validated using a subset of the whole genome resequencing data of wild baboon population (Vilgalys *et al.*, 2022). It calculates LOD scores for potential parent-

offspring pairings and utilizes simulation to determine critical values for confident paternity assignment. *DadApp* model simulation presented showed an increase in accuracy in paternal assignment with increase in the number of loci which represents the amount of genetic information of a specific population. The two scenarios 1.) Presence of maternal genotype 2.) Absence of maternal genotype were simulated as illustrated in Appendix II. This finding correlates with what was found in the study of assigning paternity of 364 red deer (*Cervus elaphus*) using a panel of 84 microsatellite markers where they identified 172 paternal relationships using both relaxed and strict confidence levels (Walling *et al.*, 2010). Due to the technicalities associated with real genotyping data, in *DadApp*, we added filters that clean uploaded datasets by removing non-genotyped loci in fathers', mother's and offspring's genotype dataset. Additionally, the app removes loci with minor allele frequency of less than 0.05 which characterizes rare variants in the population (Kalinowski *et al.*, 2007).

Whole-genome resequenced data in Vilgalys *et al.*, (2022) used in assessing the impact of admixture of wild baboon population was drawn for validation of *DadApp* app. From the study, they estimated both global and local ancestry of every baboon individual in the study via a composite likelihood method that utilizes genotype likelihood across genomic windows as opposed to genotypes at specific variants. In this study, 10 dyads and 10 triad confirmed relationships that are father-offspring and father-mother-offspring respectively were retrieved from Vilgalys *et al.*, (2022) study. First, the ability of *DadApp* to correctly assign the confirmed relationships. A 100% success rate, characterized by 20 out of 20 confirmed relationships being correctly assigned, demonstrated the app's accuracy in assigning paternity. All these assignments had an average of 12.08 statistical confidence which is derived from subtracting the LOD score of the second probable father from the LOD score of the assigned father as calculated by Kalinowski *et al.*, (2007).

CERVUS, a widely used program, applies categorical parentage assignment approach based on likelihood ratios (LOD scores). It also estimates null alleles and accommodates open systems where some parents may be missing from the sample (Slate *et al.*, 2000). However, it assumes all loci have the same error rate and lacks certain data visualization options and a user-friendly graphical interface. It is better suited for users with coding experience to customize the software according to their needs (A. G. Jones *et al.*, 2010).

Comparatively, *DadApp*, a re-implementation of CERVUS using R programming language exhibits the benefits associated with R which include a user-friendly interface, a wide variety of graphical representation, ability to share codes and or reproduce the codes to suit any researcher's demands. Additionally, this web application has the capacity for the user to edit and implement error rates associated with the genotyped data in both population-based genotyped data or in individual genotype dataset. In this study, 0.01% was used as the genotyping error as used in Marshall *et al.*, (1998) in both creation of simulation models and in assigning paternity using the ABRP real study case.

Genotyping error, for example false exclusion, is famously prone to affect relationship assessment (Kalinowski *et al.*, 2007). Contamination, allelic dropout, microsatellite stutter, null alleles, or human mistake are all potential causes of such incorrect exclusion. The issue is particularly pernicious, regardless of the number of loci genotyped, the genuine father can still be excluded. Therefore, *DadApp* model employs adjusted likelihood equations that incorporated an error rate of 0.01 as a parameter for inference. This value can be modified based on user's preference or tailored to the characteristic of the available dataset. Similar likelihood equations with an incorporated error rate of 0.01 is described in (Kalinowski *et al.*, 2007). Even though incorporating error rate lowers the LOD scores it increases the accuracy of paternity assignment especially with increase in genetic data (increase in number of loci).

In this study, the app did not have any mismatch as experienced in the study by Slate *et al* (2000) which had 75.1% and 91.9% at confidence intervals of 80% and 95% respectively. The differences could be associated with the number of paternal relationships that were screened, in our study we only screened 20 confirmed relationships using 100 SNPs whereas in the study by Slate *et al.*, (2000) they screened 172 paternities using 84 polymorphic microsatellite loci. Therefore, the differences in correct assignment could be associated with the number of loci used, which in our study we have shown that increase in the number of loci has a positive impact on proportion of correct paternity assignment.

Generally, these findings equate *DadApp* as a potential paternity tool that utilizes a popular R programming language which claims the benefits of a wide array of graphical representation, ease in reproducibility and diverse with regards to package which other software's are limited with (Jia *et al.*, 2022; Kasprzak *et al.*, 2021).

5.2. *DadApp* as a potential active educational tool

Active learning approaches, which integrate hands-on, interactive experience into classical theory-based instruction, have gained recognition as a highly effective pedagogical approach in education (Lombardi *et al.*, 2021; Neyhart & Watkins, 2020; Chi & Wylie, 2014; Freeman *et al.*, 2014a). Examples of such approaches used in active learning in the study of genetics include use of clicker questions, peer discussions, peer instruction and other group activities (Kober *et al.*, 2020; Owens *et al.*, 2020; Vickrey *et al.*, 2015). Moreover, shiny web apps possess a dynamic quality that enables them to complement and enhance other educational materials (Neyhart & Watkins, 2020). The application of this strategy, whether through student-led discussions or problem-solving exercises, has demonstrated its ability to enhance student performance and reduce attrition rates, especially in the science field (Neyhart & Watkins, 2020; Freeman *et al.*, 2014a)

This study provides a first evaluation of the impact of Shiny applications on student comprehension and active interaction in paternity within a non-conventional educational context in the developing world.

The findings show that Shiny apps can effectively support active learning methods in settings with limited access to conventional learning methods and few innovative teaching approaches. Although this study was conducted at a suburban research institute, the approach is readily transferable to field-based research environment.

Evidently, exposure to *DadApp* as an instructional tool had varied impact on the participants' interest and confidence in paternity-related concepts who had varying educational backgrounds. Notably, calculations of likelihood ratio- a core parameter in paternal inference- were ranked as the most engaging and confidence-boosting topic. This may be the case reflecting on the workshop's emphasis on the mathematical foundations of paternal inference. Therefore, resulting in the most consistent and positive changes in participants' confidence and interest across different academic levels. Similarly, this observation aligns with existing research that showed active learning strategies are effective for teaching mathematically intensive topics in biology (Scheiner, 2023; Martinez *et al.*, 2014; Jeltsch *et al.*, 2013). However, to ascertain these results, a classroom testing of this application can be done in an educational setting with participants of similar educational background.

Participants also demonstrated improved post-test performance compared to their pre-test results, regardless of their academic level or gender. This findings mirrors Moore *et al.*, (2022), who reported increased understanding and interest in ecological modelling and forecasting among both undergraduates and graduate students following the use of an R shiny app-based module. Numerous studies have similar findings indicating that technology-enhanced and collaborative learning environments improve students outcomes (Qureshi *et al.*, 2023; Herrera-Pavo, 2021). Our study extends these findings to

the field of molecular primatology within this study system, marking the first instance of such evidence from a primate range country.

Importantly, participants reported increased confidence and interest in kinship analysis and paternity inference after question-guided interaction with *DadApp*. These findings supports other prior research that demonstrated positive impact of interactive, application-based learning on student engagements. For instance, Neyhart & Watkins (2020) found increased interest in statistical genetics following the use of the *qShiny* app, and Moore *et al.*, (2022) documented similar benefits in self-reported confidence with their R shiny ecological forecasting tool. Broadly, literature supports the idea that interactivity, user-friendliness, and real-world relevance in educational apps enhance both motivation and engagement (González *et al.*, 2018; Clark & Mayer, 2016; Zhang, 2005).

Despite the promising results, the study had several limitations, the most notable being that participants completed the post-test in small groups, potentially inflating scores and obscuring individual comprehension. This group setting, while a confounding factor, also suggested that peer discussion could enhance the effectiveness of active learning tools (Freeman *et al.*, 2014). The study's short-term nature of the assessment limits insight into long-term retention or sustained interests. Additionally, while educational apps require significant instructional effort to implement, this burden is increasingly offset by the availability of web-based tools and ready-made lesson plans (Jia *et al.*, 2022).

Overall, this study validates that incorporating active learning tools can improve student's outcome, confidence and interest, even in non-conventional classroom environment within developing countries.

CHAPTER SIX

CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusion

Advancements in point-of-need, field-based DNA extraction can be highly helpful, especially in resource-limited countries. This study shows that the cellulose-based extraction method has been a potential to be an alternative DNA extraction method in microbial studies; however, it is not effective in successfully extracting DNA of the individual from the host's faecal samples.

In this study, *DadApp*, a shiny web application, was developed and shown to assign paternity accurately in the wild baboon population using non-invasive genotyped data. The app filters out non-genotyped loci, loci that are not present in offspring's genotyped data, rare variant loci characterised by <0.5 allele frequency and also incorporates an error rate which can be set by the user depending on the uploaded whole population genotyped dataset.

Additionally, *DadApp* has shown potential as an active learning tool for teaching paternity concepts outside the context of an educational setting. Participants demonstrated a significant increase in interest and understanding in concepts regardless of their academic qualifications.

This is the first study that has assessed the use of a Shiny web application as an educational tool in developing nations in a non-educational context. Overall, this study contributes to assessing the potential of both cellulose based DNA extraction and *DadApp* as developments of non-invasive genotyping tools.

6.1 Recommendations

1. Further optimization can be done on the cellulose-based extraction method. For example, utilizing probe-based enrichment methods
2. Validate the Shiny Web application using more SNP or microsatellite genetic data from other wild animal populations
3. Use the Shiny Web application as an active learning tool in a structured learning context to assess individual impact of participants with comparable educational backgrounds, using stratified or controlled study design to ensure balanced comparison across groups.

REFERENCES

- Adams, J. R., Kelly, B. T., & Waits, L. P. (2003). Using faecal DNA sampling and GIS to monitor hybridization between red wolves (*Canis rufus*) and coyotes (*Canis latrans*). *Molecular Ecology*, 12(8), 2175–2186. <https://doi.org/10.1046/j.1365-294X.2003.01895.x>
- Alberts, S. C., & Altmann, J. (2001). Immigration and hybridization patterns of yellow and anubis baboons in and around Amboseli, Kenya. *American Journal of Primatology*, 53(4), 139–154. <https://doi.org/10.1002/ajp.1>
- Alberts, S. C., & Altmann, J. (2012). The Amboseli Baboon Research Project: 40 Years of Continuity and Change. In P. M. Kappeler & D. P. Watts (Eds.), *Long-Term Field Studies of Primates* (pp. 261–287). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-22514-7_12
- Aleix-Mata, G., Adrados, B., Boos, M., Marty, E., Mourieres, P., Tucac, G., Thion, N., Mossoll-Torres, M., Pérez, J. M., & Sánchez, A. (2019). Comparing methods for estimating the abundance of Western Capercaillie *Tetrao urogallus* males in Pyrenean leks: Singing counts versus genetic analysis of non-invasive samples. *Bird Study*, 66(4), 565–569. <https://doi.org/10.1080/00063657.2020.1720594>
- Ali, N., Rampazzo, R. D. C. P., Costa, A. D. T., & Krieger, M. A. (2017). Current Nucleic Acid Extraction Methods and Their Implications to Point-of-Care Diagnostics. *BioMed Research International*, 2017, 1–13. <https://doi.org/10.1155/2017/9306564>
- Arandjelovic, M., Guschanski, K., Schubert, G., Harris, T. R., Thalmann, O., Siedel, H., & Vigilant, L. (2009). Two-step multiplex polymerase chain reaction improves the speed and accuracy of genotyping using DNA from noninvasive and museum

samples. *Molecular Ecology Resources*, 9(1), 28–36.
<https://doi.org/10.1111/j.1755-0998.2008.02387.x>

Arandjelovic, M., & Vigilant, L. (2018). Non-invasive genetic censusing and monitoring of primate populations. *American Journal of Primatology*, 80(3), e22743.
<https://doi.org/10.1002/ajp.22743>

Baek, H. E., Bandivadekar, R. R., Pandit, P., Mah, M., Sehgal, R. N. M., & Tell, L. A. (2020). TaqMan quantitative real-time PCR for detecting Avipoxvirus DNA in various sample types from hummingbirds. *PLOS ONE*, 15(6), e0230701.
<https://doi.org/10.1371/journal.pone.0230701>

Balkiz, Ö., Dano, S., Barbraud, C., Tekin, S., Özesmi, U., Dündar, M., & Béchet, A. (2007). Sexing Greater Flamingo Chicks from Feather Bulb DNA. *Waterbirds*, 30(3), 450–453.
[https://doi.org/10.1675/1524-4695\(2007\)030\[0450:SGFCFF\]2.0.CO;2](https://doi.org/10.1675/1524-4695(2007)030[0450:SGFCFF]2.0.CO;2)

Banks, S. C., & Piggott, M. P. (2022). Non-invasive genetic sampling is one of our most powerful and ethical tools for threatened species population monitoring: A reply to Lavery et al. *Biodiversity and Conservation*, 31(2), 723–728.
<https://doi.org/10.1007/s10531-022-02377-x>

Beja-Pereira, A., Oliveira, R., Alves, P. C., Schwartz, M. K., & Luikart, G. (2009). Advancing ecological understandings through technological transformations in noninvasive genetics. *Molecular Ecology Resources*, 9(5), 1279–1301.
<https://doi.org/10.1111/j.1755-0998.2009.02699.x>

Bellemain, E., & Taberlet, P. (2004). Improved noninvasive genotyping method: Application to brown bear (*Ursus arctos*) faeces. *Molecular Ecology Notes*, 4(3), 519–522. <https://doi.org/10.1111/j.1471-8286.2004.00711.x>

- Benčič, A., Toplak, N., Koren, S., Bogožalec Košir, A., Milavec, M., Tomič, V., Lužnik, D., & Dreo, T. (2024). Metrological evaluation of DNA extraction method effects on the bacterial microbiome and resistome in sputum. *mSystems*, 9(9), e0073524. <https://doi.org/10.1128/msystems.00735-24>
- Bensasson, D. (2001). Mitochondrial pseudogenes: Evolution's misplaced witnesses. *Trends in Ecology & Evolution*, 16(6), 314–321. [https://doi.org/10.1016/S0169-5347\(01\)02151-6](https://doi.org/10.1016/S0169-5347(01)02151-6)
- Bischof, R., Gregersen, E. R., Brøseth, H., Ellegren, H., & Flagstad, Ø. (2016). Noninvasive genetic sampling reveals intrasex territoriality in wolverines. *Ecology and Evolution*, 6(5), 1527–1536. <https://doi.org/10.1002/ece3.1983>
- Brenner, C. H. (1997). Symbolic Kinship Program. *Genetics*, 145(2), 535–542. <https://doi.org/10.1093/genetics/145.2.535>
- Broquet, T., Ménard, N., & Petit, E. (2006). Noninvasive population genetics: A review of sample source, diet, fragment length and microsatellite motif effects on amplification success and genotyping error rates. *Conservation Genetics*, 8(1), 249–260. <https://doi.org/10.1007/s10592-006-9146-5>
- Buchan, J. C., Alberts, S. C., Silk, J. B., & Altmann, J. (2003). True paternal care in a multi-male primate society. *Nature*, 425(6954), 179–181. <https://doi.org/10.1038/nature01866>
- Bush, K. L., Vinsky, M. D., Aldridge, C. L., & Paszkowski, C. A. (2005). A comparison of sample types varying in invasiveness for use in DNA sex determination in an endangered population of greater Sage-Grouse (*Centrocercus urophasianus*). *Conservation Genetics*, 6(5), 867–870. <https://doi.org/10.1007/s10592-005-9040-6>

- Byrne, M. S., Peralta, D. M., Ibañez, E. A., Nardelli, M., & Túnez, J. I. (2021). Non-invasive Sampling Techniques Applied to Conservation Genetic Studies in Mammals. In M. Nardelli & J. I. Túnez (Eds.), *Molecular Ecology and Conservation Genetics of Neotropical Mammals* (pp. 63–83). Springer International Publishing. https://doi.org/10.1007/978-3-030-65606-5_4
- Campbell, N. R., Harmon, S. A., & Narum, S. R. (2015). Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. *Molecular Ecology Resources*, 15(4), 855–867. <https://doi.org/10.1111/1755-0998.12357>
- Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S. M., Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert, J. A., Smith, G., & Knight, R. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME Journal*, 6(8), 1621–1624. <https://doi.org/10.1038/ismej.2012.8>
- Carroll, E. L., Bruford, M. W., DeWoody, J. A., Leroy, G., Strand, A., Waits, L., & Wang, J. (2018). Genetic and genomic monitoring with minimally invasive sampling methods. *Evolutionary Applications*, 11(7), 1094–1119. <https://doi.org/10.1111/eva.12600>
- Casper, R. M., Jarman, S. N., Deagle, B. E., Gales, N. J., & Hindell, M. A. (2007). Detecting prey from DNA in predator scats: A comparison with morphological analysis, using *Arctocephalus* seals fed a known diet. *Journal of Experimental Marine Biology and Ecology*, 347(1–2), 144–154. <https://doi.org/10.1016/j.jembe.2007.04.002>
- Cercueil, A. (2002). PARENTE: Computer Program for Parentage Analysis. *Journal of Heredity*, 93(6), 458–459. <https://doi.org/10.1093/jhered/93.6.458>

- Chang, W., Cheng, J., Allaire, J., Xie, Y., & McPherson, J. (2015). *Package 'shiny'* [Computer software]. . <http://shiny.rstudio.com>
- Charpentier, M. J. E., Fontaine, M. C., Cherel, E., Renoult, J. P., Jenkins, T., Benoit, L., Barthès, N., Alberts, S. C., & Tung, J. (2012). Genetic structure in a dynamic baboon hybrid zone corroborates behavioural observations in a hybrid population. *Molecular Ecology*, 21(3), 715–731. <https://doi.org/10.1111/j.1365-294X.2011.05302.x>
- Chi, M. T. H., & Wylie, R. (2014). The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educational Psychologist*, 49(4), 219–243. <https://doi.org/10.1080/00461520.2014.965823>
- Chiou, K. L., & Bergey, C. M. (2018). Methylation-based enrichment facilitates low-cost, noninvasive genomic scale sequencing of populations from feces. *Scientific Reports*, 8(1), 1975. <https://doi.org/10.1038/s41598-018-20427-9>
- Clark, R. C., & Mayer, R. E. (Eds.). (2016). *e-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning* (1st ed.). Wiley. <https://doi.org/10.1002/9781119239086>
- C. W., Wehausen, J. D., Palsbøll, P. J., & Mccullough, D. R. (2010). Using Genetic Tools to Track Desert Bighorn Sheep Colonizations. *The Journal of Wildlife Management*, 74(3), 522–531. <https://doi.org/10.2193/2008-448>
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>

- Danzmann, R. G. (1997). PROBMAX: A Computer Program for Assigning Unknown Parentage in Pedigree Analysis From Known Genotypic Pools of Parents and Progeny. *Journal of Heredity*, 88(4), 333–333. <https://doi.org/10.1093/oxfordjournals.jhered.a023112>
- De Groot, G. A., Nowak, C., Skrbinišek, T., Andersen, L. W., Aspi, J., Fumagalli, L., Godinho, R., Harms, V., Jansman, H. A. H., Liberg, O., Marucco, F., Mysłajek, R. W., Nowak, S., Pilot, M., Randi, E., Reinhardt, I., Śmietana, W., Szewczyk, M., Taberlet, P., ... Muñoz-Fuentes, V. (2016). Decades of population genetic research reveal the need for harmonization of molecular markers: The grey wolf *Canis lupus* as a case study. *Mammal Review*, 46(1), 44–59. <https://doi.org/10.1111/mam.12052>
- Dejean, T., Valentini, A., Miquel, C., Taberlet, P., Bellemain, E., & Miaud, C. (2012). Improved detection of an alien invasive species through environmental DNA barcoding: The example of the American bullfrog *Lithobates catesbeianus*. *Journal of Applied Ecology*, 49(4), 953–959. <https://doi.org/10.1111/j.1365-2664.2012.02171.x>
- Duchesne, P., Godbout, M., & Bernatchez, L. (2002). PAPA (package for the analysis of parental allocation): A computer program for simulated and real parental allocation. *Molecular Ecology Notes*, 2(2), 191–193. <https://doi.org/10.1046/j.1471-8286.2002.00164.x>
- Emami-Khoyi, A., Agnew, T. W., Adair, M. G., Murphy, E. C., Benmazouz, I., Monsanto, D. M., ... Ross, J. G. (2021). A New Non-invasive Method for Collecting DNA From Small Mammals in the Field, and Its Application in Simultaneous Vector and Disease Monitoring in Brushtail Possums. *Frontiers in Environmental Science*, 9. <https://doi.org/10.3389/fenvs.2021.701033>

- Feinstein, J. (2004). DNA sequence from butterfly frass and exuviae. *Conservation Genetics*, 5(1), 103–104. <https://doi.org/10.1023/B:COGE.0000014058.34840.94>
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014a). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23), 8410–8415. <https://doi.org/10.1073/pnas.1319030111>
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014b). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23), 8410–8415. <https://doi.org/10.1073/pnas.1319030111>
- Gerber, S., Chabrier, P., & Kremer, A. (2003). FAMOZ: A software for parentage analysis using dominant, codominant and uniparentally inherited markers. *Molecular Ecology Notes*, 3(3), 479–481. <https://doi.org/10.1046/j.1471-8286.2003.00439.x>
- González, J. A., López, M., Cobo, E., & Cortés, J. (2018). Assessing Shiny apps through student feedback: Recommendations from a qualitative study. *Computer Applications in Engineering Education*, 26(5), 1813–1824. <https://doi.org/10.1002/cae.21932>
- Goymann, W. (2005). Noninvasive Monitoring of Hormones in Bird Droppings: Physiological Validation, Sampling, Extraction, Sex Differences, and the Influence of Diet on Hormone Metabolite Levels. *Annals of the New York Academy of Sciences*, 1046(1), 35–53. <https://doi.org/10.1196/annals.1343.005>

- Herrera-Pavo, M. Á. (2021). Collaborative learning for virtual higher education. *Learning, Culture and Social Interaction*, 28, 100437. <https://doi.org/10.1016/j.lcsi.2020.100437>
- Hoffman, J. I., & Amos, W. (2005). Microsatellite genotyping errors: Detection approaches, common sources and consequences for paternal exclusion. *Molecular Ecology*, 14(2), 599–612. <https://doi.org/10.1111/j.1365-294X.2004.02419.x>
- Iyengar, A., Babu, V. N., Hedges, S., Venkataraman, A. B., Maclean, N., & Morin, P. A. (2005). Phylogeography, genetic structure, and diversity in the dhole (*Cuon alpinus*). *Molecular Ecology*, 14(8), 2281–2297. <https://doi.org/10.1111/j.1365-294X.2005.02582.x>
- Jeltsch, F., Blaum, N., Brose, U., Chipperfield, J. D., Clough, Y., Farwig, N., Geissler, K., Graham, C. H., Grimm, V., Hickler, T., Huth, A., May, F., Meyer, K. M., Pagel, J., Reineking, B., Rillig, M. C., Shea, K., Schurr, F. M., Schröder, B., ... Zurell, D. (2013). How can we bring together empiricists and modellers in functional biodiversity research? *Basic and Applied Ecology*, 14(2), 93–101 <https://doi.org/10.1016/j.baae.2013.01.001>
- Jia, L., Yao, W., Jiang, Y., Li, Y., Wang, Z., Li, H., Huang, F., Li, J., Chen, T., & Zhang, H. (2022). Development of interactive biological web applications with R/Shiny. *Briefings in Bioinformatics*, 23(1), bbab415. <https://doi.org/10.1093/bib/bbab415>
- Jones, A. G. (2005). gerud 2.0: A computer program for the reconstruction of parental genotypes from half-sib progeny arrays with known or unknown parents. *Molecular Ecology Notes*, 5(3), 708–711. <https://doi.org/10.1111/j.1471-8286.2005.01029.x>

- Jones, A. G., Small, C. M., Paczolt, K. A., & Ratterman, N. L. (2010). A practical guide to methods of parentage analysis. *Molecular Ecology Resources*, 10(1), 6–30. <https://doi.org/10.1111/j.1755-0998.2009.02778.x>
- Jones, O. R., & Wang, J. (2010). COLONY: A program for parentage and sibship inference from multilocus genotype data. *Molecular Ecology Resources*, 10(3), 551–555. <https://doi.org/10.1111/j.1755-0998.2009.02787.x>
- Kalinowski, S. T., Taper, M. L., & Marshall, T. C. (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology*, 16(5), 1099–1106. <https://doi.org/10.1111/j.1365-294X.2007.03089.x>
- Kasprzak, P., Mitchell, L., Kravchuk, O., & Timmins, A. (2021). *Six Years of Shiny in Research—Collaborative Development of Web Tools in R* (Version 1). arXiv. <https://doi.org/10.48550/ARXIV.2101.10948>
- Kerley, L. L., & Salkina, G. P. (2007). Using Scent-Matching Dogs to Identify Individual Amur Tigers from Scats. *The Journal of Wildlife Management*, 71(4), 1349–1356. <https://doi.org/10.2193/2006-361>
- Kilpatrick, H. J., Goodie, T. J., & Kovach, A. I. (2013). Comparison of live-trapping and noninvasive genetic sampling to assess patch occupancy by New England cottontail (*Sylvilagus transitionalis*) rabbits. *Wildlife Society Bulletin*, 37(4), 901–905. <https://doi.org/10.1002/wsb.330>
- King, S. R. B., & Schoenecker, K. A. (2019). Comparison of Methods To Examine Diet of Feral Horses from Noninvasively Collected Fecal Samples. *Rangeland Ecology & Management*, 72(4), 661–666. <https://doi.org/10.1016/j.rama.2019.02.005>
- Kivisild, T. (2015). Maternal

ancestry and population history from whole mitochondrial genomes. *Investigative Genetics*, 6(1), 3. <https://doi.org/10.1186/s13323-015-0022-2>

Kober, S. E., Wood, G., Kiili, K., Moeller, K., & Ninaus, M. (2020). Game-based learning environments affect frontal brain activity. *PLOS ONE*, 15(11), e0242573. <https://doi.org/10.1371/journal.pone.0242573>

Konovalov, D. A., Manning, C., & Henshaw, M. T. (2004). kingroup: A program for pedigree relationship reconstruction and kin group assignments using genetic markers. *Molecular Ecology Notes*, 4(4), 779–782. <https://doi.org/10.1111/j.1471-8286.2004.00796.x>

Lombardi, D., Shipley, T. F., Astronomy Team, Biology Team, Chemistry Team, Engineering Team, Geography Team, Geoscience Team, and Physics Team, Bailey, J. M., Bretones, P. S., Prather, E. E., Ballen, C. J., Knight, J. K., Smith, M. K., Stowe, R. L., Cooper, M. M., Prince, M., Atit, K., Uttal, D. H., LaDue, N. D., McNeal, P. M., Ryker, K., St. John, K., Van Der Hoeven Kraft, K. J., & Docktor, J. L. (2021). The Curious Construct of Active Learning. *Psychological Science in the Public Interest*, 22(1), 8–43. <https://doi.org/10.1177/1529100620973974>

Luikart, G., Zundel, S., Rioux, D., Miquel, C., Keating, K. A., Hogg, J. T., Steele, B., Foresman, K., & Taberlet, P. (2008). Low Genotyping Error Rates and Noninvasive Sampling in Bighorn Sheep. *The Journal of Wildlife Management*, 72(1), 299–304. <https://doi.org/10.2193/2006-006>

Marshall, T. C., Slate, J., Kruuk, L. E. B., & Pemberton, J. M. (1998). Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology*, 7(5), 639–655. <https://doi.org/10.1046/j.1365-294x.1998.00374.x>

- Martinez, D., Alenya, G., Jimenez, P., Torras, C., Rossmann, J., Wantia, N., Aksoy, E. E., Haller, S., & Piater, J. (2014). Active learning of manipulation sequences. *2014 IEEE International Conference on Robotics and Automation (ICRA)*, 5671–5678. <https://doi.org/10.1109/ICRA.2014.6907693>
- Moldovan, O. T., Baricz, A., Szekeres, E., Kenesz, M., Hoaghia, M. A., Levei, E. A., Mirea, I. C., Năstase-Bucur, R., Brad, T., Chiciudean, I., & Banciu, H. L. (2020). Testing Different Membrane Filters for 16S rRNA Gene-Based Metabarcoding in Karstic Springs. *Water*, *12*(12), 3400. <https://doi.org/10.3390/w12123400>
- Moore, T. N., Thomas, R. Q., Woelmer, W. M., & Carey, C. C. (2022). Integrating Ecological Forecasting into Undergraduate Ecology Curricula with an R Shiny Application-Based Teaching Module. *Forecasting*, *4*(3), 604–633. <https://doi.org/10.3390/forecast4030033>
- Moritz, C. (1994). Applications of mitochondrial DNA analysis in conservation: A critical review. *Molecular Ecology*, *3*(4), 401–411. <https://doi.org/10.1111/j.1365-294X.1994.tb00080.x>
- Mundy, N. I., Winchell, C. S., & Woodruff, D. S. (1997). Genetic differences between the endangered San Clemente Island loggerhead shrike *Lanius ludovicianus mearnsi* and two neighbouring subspecies demonstrated by mtDNA control region and cytochrome *b* sequence variation. *Molecular Ecology*, *6*(1), 29–37. <https://doi.org/10.1046/j.1365-294X.1997.00149.x>
- Natesh, M., Taylor, R. W., Truelove, N. K., Hadly, E. A., Palumbi, S. R., Petrov, D. A., & Ramakrishnan, U. (2019). Empowering conservation practice with efficient and economical genotyping from poor quality samples. *Methods in Ecology and Evolution*, *10*(6), 853–859. <https://doi.org/10.1111/2041-210X.13173>

- Newmark, W. D., & Hough, J. L. (2000). Conserving Wildlife in Africa: Integrated Conservation and Development Projects and Beyond. *BioScience*, 50(7), 585. [https://doi.org/10.1641/0006-3568\(2000\)050\[0585:CWIAIC\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2000)050[0585:CWIAIC]2.0.CO;2)
- Neyhart, J. L., & Watkins, E. (2020). An active learning tool for quantitative genetics instruction using *R* and *shiny*. *Natural Sciences Education*, 49(1), e20026. <https://doi.org/10.1002/nse2.20026>
- Nsubuga, A. M., Robbins, M. M., Roeder, A. D., Morin, P. A., Boesch, C., & Vigilant, L. (2004). Factors affecting the amount of genomic DNA extracted from ape faeces and the identification of an improved sample storage method. *Molecular Ecology*, 13(7), 2089–2094. <https://doi.org/10.1111/j.1365-294X.2004.02207.x>
- Orkin, J. D., Montague, M. J., Tejada-Martinez, D., De Manuel, M., Del Campo, J., Hernandez, S. C., Di Fiore, A., Fontseré, C., Hodgson, J. A., Janiak, M. C., Kuderna, L. F. K., Lizano, E., Martin, M. P., Niimura, Y., Perry, G. H., Valverde, C. S., Tang, J., Warren, W. C., De Magalhães, J. P., ... Melin, A. D. (2018). *The evolution of ecological flexibility, large brains, and long lives: Capuchin monkey genomics revealed with fecalFACS*. Genomics. <https://doi.org/10.1101/366112>
- Owens, D. C., Sadler, T. D., Barlow, A. T., & Smith-Walters, C. (2020). Student Motivation from and Resistance to Active Learning Rooted in Essential Science Practices. *Research in Science Education*, 50(1), 253–277. <https://doi.org/10.1007/s11165-017-9688-1>
- Palomares, F., Godoy, J. A., Piriz, A., O'Brien, S. J., & W. E. Johnson. (2002). Faecal genetic analysis to determine the presence and distribution of elusive carnivores: Design and feasibility for the Iberian lynx. *Molecular Ecology*, 11(10), 2171–2182. <https://doi.org/10.1046/j.1365-294X.2002.01608.x>

- Perry, G. H., Marioni, J. C., Melsted, P., & Gilad, Y. (2010). Genomic-scale capture and sequencing of endogenous DNA from feces: GENOMIC-SCALE SEQUENCING OF FECAL DNA. *Molecular Ecology*, 19(24), 5332–5344. <https://doi.org/10.1111/j.1365-294X.2010.04888.x>
- Piggott, M. P., Banks, S. C., Stone, N., Banffy, C., & Taylor, A. C. (2006). Estimating population size of endangered brush-tailed rock-wallaby (*Petrogale penicillata*) colonies using faecal DNA. *Molecular Ecology*, 15(1), 81–91. <https://doi.org/10.1111/j.1365-294X.2005.02783.x>
- Qureshi, M. A., Khaskheli, A., Qureshi, J. A., Raza, S. A., & Yousufi, S. Q. (2023). Factors affecting students’ learning performance through collaborative learning and engagement. *Interactive Learning Environments*, 31(4), 2371–2391. <https://doi.org/10.1080/10494820.2021.1884886>
- R Core Team, R. C. T. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>. [Computer software].
- Ramón-Laca, A., Soriano, L., Gleeson, D., & Godoy, J. A. (2015). A simple and effective method for obtaining mammal DNA from faeces. *Wildlife Biology*, 21(4), 195–203. <https://doi.org/10.2981/wlb.00096>
- Regnaut, S., Christe, P., Chapuisat, M., & Fumagalli, L. (2006). Genotyping faeces reveals facultative kin association on capercaillie’s leks. *Conservation Genetics*, 7(5), 665–674. <https://doi.org/10.1007/s10592-005-9097-2>
- Rodgers, T. W., Giacalone, J., Heske, E. J., Janečka, J. E., Jansen, P. A., Phillips, C. A., & Schooley, R. L. (2015). Socio-spatial organization and kin structure in ocelots from integration of camera trapping and noninvasive genetics. *Journal of Mammalogy*, 96(1), 120–128. <https://doi.org/10.1093/jmammal/gyu012>

- Rudnick, J. A., Katzner, T. E., Bragin, E. A., & DeWoody, J. A. (2008). A non-invasive genetic evaluation of population size, natal philopatry, and roosting behavior of non-breeding eastern imperial eagles (*Aquila heliaca*) in central Asia. *Conservation Genetics*, 9(3), 667–676. <https://doi.org/10.1007/s10592-007-9397-9>
- Saiyed, Z. M., Bochiwal, C., Gorasia, H., Telang, S. D., & Ramchand, C. N. (2006). Application of magnetic particles (Fe₃O₄) for isolation of genomic DNA from mammalian cells. *Analytical Biochemistry*, 356(2), 306–308. <https://doi.org/10.1016/j.ab.2006.06.027>
- Scandura, M. (2005). Individual sexing and genotyping from blood spots on the snow: A reliable source of DNA for non-invasive genetic surveys. *Conservation Genetics*, 6(5), 871–874. <https://doi.org/10.1007/s10592-005-9041-5>
- Scheiner, T. (2023). Shifting the ways prospective teachers frame and notice student mathematical thinking: From deficits to strengths. *Educational Studies in Mathematics*, 114(1), 35–61. <https://doi.org/10.1007/s10649-023-10235-y>
- Schenker, L., Bollmann, K., Rehnus, M., Brodbeck, S., & Gugerli, F. (2020). Hare's affairs: Lessons learnt from a noninvasive genetic monitoring for tracking mountain hare individuals. *Ecology and Evolution*, 10(18), 10150–10166. <https://doi.org/10.1002/ece3.6676>
- Signorovitch, J., & Nielsen, R. (2002). PATRI—paternity inference using genetic data. *Bioinformatics*, 18(2), 341–342. <https://doi.org/10.1093/bioinformatics/18.2.341>
- Slate, J., Marshall, T., & Pemberton, J. (2000). A retrospective assessment of the accuracy of the paternity inference program CERVUS. *Molecular Ecology*, 9(6), 801–808. <https://doi.org/10.1046/j.1365-294x.2000.00930.x>

- Snyder-Mackler, N., Majoros, W. H., Yuan, M. L., Shaver, A. O., Gordon, J. B., Kopp, G. H., Schlebusch, S. A., Wall, J. D., Alberts, S. C., Mukherjee, S., Zhou, X., & Tung, J. (2016). Efficient Genome-Wide Sequencing and Low-Coverage Pedigree Analysis from Noninvasively Collected Samples. *Genetics*, 203(2), 699–714. <https://doi.org/10.1534/genetics.116.187492>
- Solberg, K. H., Bellemain, E., Drageset, O.-M., Taberlet, P., & Swenson, J. E. (2006). An evaluation of field and non-invasive genetic methods to estimate brown bear (*Ursus arctos*) population size. *Biological Conservation*, 128(2), 158–168. <https://doi.org/10.1016/j.biocon.2005.09.025>
- Song, Y., & Gyarmati, P. (2022). Rapid DNA visual detection of polymicrobial bloodstream infection using filter paper. *Scientific Reports*, 12(1), 4515. <https://doi.org/10.1038/s41598-022-08487-4>
- Taberlet, P. (1996). Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research*, 24(16), 3189–3194. <https://doi.org/10.1093/nar/24.16.3189>
- Taberlet, P., Camarra, J. -J., Griffin, S., Uhrès, E., Hanotte, O., Waits, L. P., Dubois-Paganon, C., Burke, T., & Bouvet, J. (1997). Noninvasive genetic tracking of the endangered Pyrenean brown bear population. *Molecular Ecology*, 6(9), 869–876. <https://doi.org/10.1111/j.1365-294X.1997.tb00141.x>
- Taberlet, P., & Luikart, G. (1999). Non-invasive genetic sampling and individual identification. *Biological Journal of the Linnean Society*, 68(1–2), 41–55. <https://doi.org/10.1111/j.1095-8312.1999.tb01157.x>
- Tan, J., Zhu, C., Li, L., Wang, J., Xia, X.-H., & Wang, C. (2024). Engineering Cell Membranes: From Extraction Strategies to Emerging Biosensing Applications.

Analytical Chemistry, 96(20), 7880–7894.
<https://doi.org/10.1021/acs.analchem.3c01746>

Tung, J., Charpentier, M. J. E., Garfield, D. A., Altmann, J., & Alberts, S. C. (2008). Genetic evidence reveals temporal change in hybridization patterns in a wild baboon population. *Molecular Ecology*, 17(8), 1998–2011.
<https://doi.org/10.1111/j.1365-294X.2008.03723.x>

Ulizio, T. J., Squires, J. R., Pletscher, D. H., Schwartz, M. K., Claar, J. J., & Ruggiero, L. F. (2006). The Efficacy of Obtaining Genetic-Based Identifications from Putative Wolverine Snow Tracks. *Wildlife Society Bulletin*, 34(5), 1326–1332.
[https://doi.org/10.2193/0091-7648\(2006\)34\[1326:TEOOGI\]2.0.CO;2](https://doi.org/10.2193/0091-7648(2006)34[1326:TEOOGI]2.0.CO;2)

Valencia, C. A., Pervaiz, M. A., Husami, A., Qian, Y., & Zhang, K. (2013). A Review of DNA Enrichment Technologies. In C. A. Valencia, M. A. Pervaiz, A. Husami, Y. Qian, & K. Zhang, *Next Generation Sequencing Technologies in Medical Genetics* (pp. 25–32). Springer New York. https://doi.org/10.1007/978-1-4614-9032-6_3

Vickrey, T., Rosploch, K., Rahmanian, R., Pilarz, M., & Stains, M. (2015). Research-Based Implementation of Peer Instruction: A Literature Review. *CBE—Life Sciences Education*, 14(1), es3. <https://doi.org/10.1187/cbe.14-11-0198>

Vilgalys, T. P., Fogel, A. S., Anderson, J. A., Mututua, R. S., Warutere, J. K., Siodi, I. L., Kim, S. Y., Voyles, T. N., Robinson, J. A., Wall, J. D., Archie, E. A., Alberts, S. C., & Tung, J. (2022). Selection against admixture and gene regulatory divergence in a long-term primate field study. *Science*, 377(6606), 635–641.
<https://doi.org/10.1126/science.abm4917>

Von Thaden, A., Cocchiararo, B., Jarausch, A., Jüngling, H., Karamanlidis, A. A., Tiesmeyer, A., Nowak, C., & Muñoz-Fuentes, V. (2017). Assessing SNP

genotyping of noninvasively collected wildlife samples using microfluidic arrays. *Scientific Reports*, 7(1), 10768. <https://doi.org/10.1038/s41598-017-10647-w>

Waits, L. P., & Paetkau, D. (2005). NONINVASIVE GENETIC SAMPLING TOOLS FOR WILDLIFE BIOLOGISTS: A REVIEW OF APPLICATIONS AND RECOMMENDATIONS FOR ACCURATE DATA COLLECTION. *Journal of Wildlife Management*, 69(4), 1419–1433. [https://doi.org/10.2193/0022-541X\(2005\)69\[1419:NGSTFW\]2.0.CO;2](https://doi.org/10.2193/0022-541X(2005)69[1419:NGSTFW]2.0.CO;2)

Walling, C. A., Pemberton, J. M., Hadfield, J. D., & Kruuk, L. E. B. (2010). Comparing parentage inference software: Reanalysis of a red deer pedigree. *Molecular Ecology*, 19(9), 1914–1928. <https://doi.org/10.1111/j.1365-294X.2010.04604.x>

Woodruff, D. S. (1993). Non-invasive genotyping of primates. *Primates*, 34(3), 333–346. <https://doi.org/10.1007/BF02382629>

Woodruff, D. S., & Gall, G. A. E. (1992). Genetics and conservation. *Agriculture, Ecosystems & Environment*, 42(1–2), 53–73. [https://doi.org/10.1016/0167-8809\(92\)90019-8](https://doi.org/10.1016/0167-8809(92)90019-8)

Zemanova, M. A. (2020). Towards more compassionate wildlife research through the 3Rs principles: Moving from invasive to non-invasive methods. *Wildlife Biology*, 2020(1). <https://doi.org/10.2981/wlb.00607>

Zemanova, M. A. (2021). Noninvasive Genetic Assessment Is an Effective Wildlife Research Tool When Compared with Other Approaches. *Genes*, 12(11), 1672. <https://doi.org/10.3390/genes12111672>

- Zhang, D. (2005). Interactive Multimedia-Based E-Learning: A Study of Effectiveness. *American Journal of Distance Education*, 19(3), 149–162. https://doi.org/10.1207/s15389286ajde1903_3
- Zhang, D., & Hewitt, G. M. (2003). Nuclear DNA analyses in genetic studies of populations: Practice, problems and prospects. *Molecular Ecology*, 12(3), 563–584. <https://doi.org/10.1046/j.1365-294X.2003.01773.x>
- Zou, Y., Mason, M. G., Wang, Y., Wee, E., Turni, C., Blackall, P. J., Trau, M., & Botella, J. R. (2018). Correction: Nucleic acid purification from plants, animals and microbes in under 30 seconds. *PLOS Biology*, 16(5), e1002630. <https://doi.org/10.1371/journal.pbio.1002630>

APPENDICES

Appendix I: Ethics Approval Letter Used to Assess DadApp as an Active Educational Tool



UNIVERSITY OF NAIROBI
FACULTY OF HEALTH SCIENCES
P O BOX 19676 Code 00202
TELEGRAMS: versity
Tel: (254-020) 2720200 Ext 44355



KENYATTA NATIONAL HOSPITAL
P O BOX 20723 Code 00202
Tel: 726300-0
Fax: 725272
Telegrams: MEDSUP, Nairobi

KNH-UoN ERC
Email: uonkh_erc@uonbi.ac.ke
Website: <http://www.erc.uonbi.ac.ke>
Facebook: https://www.facebook.com/uonkh_erc/
Twitter: @UONKH_ERC https://twitter.com/UONKH_ERC

Ref: KNH-ERC/A/06

4th January 2024

Dr. Mercy Akinyi
Principal Investigator
Animal Science Department
Institute of Primate Research
Email: Akinyimercy@gmail.com



Dear Dr. Akinyi

ETHICAL APPROVAL-RESEARCH PROPOSAL: ASSESSING DADIAPP; A SHINY WEB APPLICATION THAT ASSIGNS PATERNITY IN WILD ANIMAL POPULATION, AS AN EDUCATIONAL TOOL IN ECOLOGY (P689/09/2023)

This is to inform you that KNH-UoN ERC has reviewed and approved your above research proposal. Your application approval number is **P689/09/2023**. The approval period is 4th January 2024- 3rd January 2025.

This approval is subject to compliance with the following requirements:

- Only approved documents including (informed consents, study instruments, MTA) will be used.
- All changes including (amendments, deviations, and violations) are submitted for review and approval by KNH-UoN ERC.
- Death and life threatening problems and serious adverse events or unexpected adverse events whether related or unrelated to the study must be reported to KNH-UoN ERC 72 hours of notification.
- Any changes, anticipated or otherwise that may increase the risks or affected safety or welfare of study participants and others or affect the integrity of the research must be reported to KNH-UoN ERC within 72 hours.
- Clearance for export of biological specimens must be obtained from relevant institutions.
- Submission of a request for renewal of approval at least 60 days prior to expiry of the approval period. Attach a comprehensive progress report to support the renewal.
- Submission of an executive summary report within 90 days upon completion of the study to KNH-UoN ERC.

Prior to commencing your study, you will be expected to obtain a research license from National Commission for Science, Technology and Innovation (NACOSTI) <https://research-portal.nacosti.go.ke> and also obtain other clearances needed.

Yours sincerely,



PROF. BEATRICE K.M. AMUGUNE
SECRETARY, KNH- UoN ERC

c.c. The Dean, Faculty of Health Sciences, UoN
The Senior Director, CS, KNH
The Chairperson, KNH- UoN ERC
The Assistant Director, Health Information Dept., KNH
Co-Investigators: Dr. David Kiragu, Mweura, Animal Science Department, Institute of Primate Research
Dr. Daniel Kiboi, Dept. of Biochemistry, JKUAT
Prof. Jonny Tsag, Dept of Primate Behavior and Evolution, Max Planck Institute of Evolutionary Anthropology, Germany.

Appendix II: Report of how Suitability of *DadApp* was Assessed Using Real Data from ABRP

USE OF ABRP'S DATASET CHARACTERIZED WITH FEW INCOMPATIBLE GENOTYPES.

A dataset generated by Arielle was used to assess the accuracy of *DadApp* paternity assignment. The dataset was made up of 10 potential fathers, 10 mothers and 10 offspring each with 100 genotyped loci.

From the dataset, below are the confirmed relationship.

Fathers	Mothers	Offspring
AMB_118	AMB_238	AMB_308
AMB_189	AMB_111	AMB_108
AMB_083	AMB_100	AMB_101
AMB_140	AMB_238	AMB_279
AMB_104	AMB_179	AMB_175
AMB_056	AMB_097	AMB_118
AMB_216	AMB_336	AMB_025
AMB_010	AMB_286	AMB_064
AMB_272	AMB_350	AMB_048
AMB_031	AMB_111	AMB_116

Quick summary:

- The script correctly assigned **10 out of 10** relationships.
- Overall trend detected, the more compatible genotype the higher the LOD scores as expected.

- Assigned fathers who are the real father have no incompatible genotypes in all cases.
- In dad's raw data, they all have 100 loci, but after filters (allele frequency not equal to 1, more than 0 and MAF), 95 loci remained which was used for calculation of LOD scores. The 95 loci are denoted as total loci in below results.
- The statistical confidence (LOD score of the first assigned father minus LOD score of the second assigned father) of every case was quite high
- Real fathers are highlighted in green and the tabular results are arranged in descending order using the LOD scores.
- AMB_118 was in two confirmed relationships, one as the offspring and another as the father.

Below is every single trio run and their results

1.

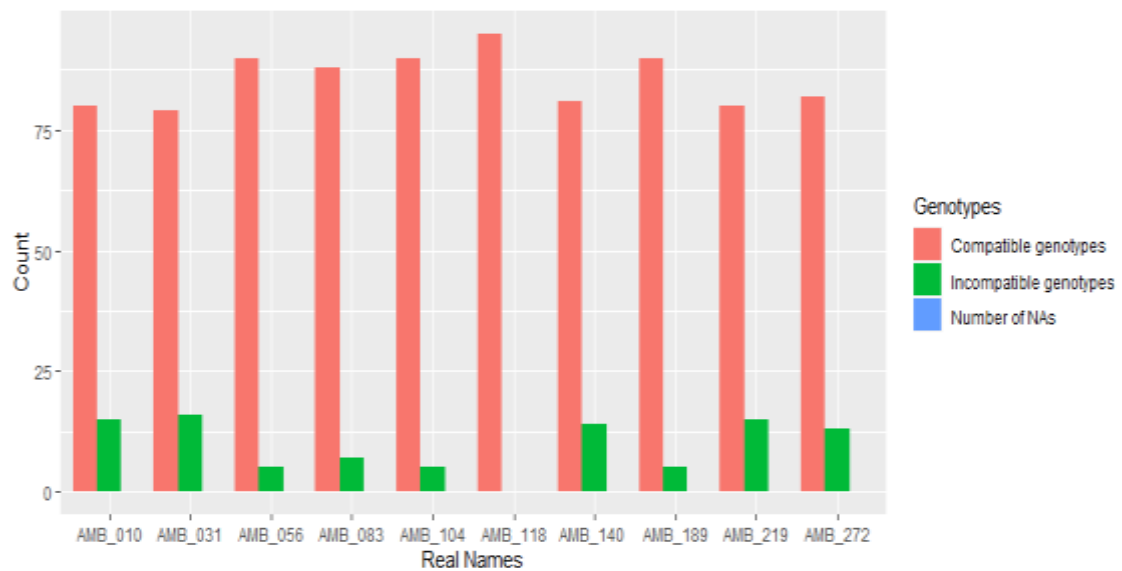
Fathers	Mothers	Offspring
AMB_118	AMB_238	AMB_308

Tabular results:

Real Name	LOD	Total loci	Compatible genotype	Incompatible genotype	No. of NAs
AMB_118	8.599	95	95	0	0
AMB_189	-2.002	95	90	5	0
AMB_083	-2.604	95	90	5	0
AMB_140	-2.905	95	90	5	0
AMB_104	-4.798	95	88	7	0
AMB_056	-16.282	95	80	15	0
AMB_219	-16.691	95	81	14	0

AMB_010	-16.798	95	82	13	0
AMB_272	-18.884	95	79	16	0
AMB_031	-20.196	95	80	15	0

Graphical representation:



2.

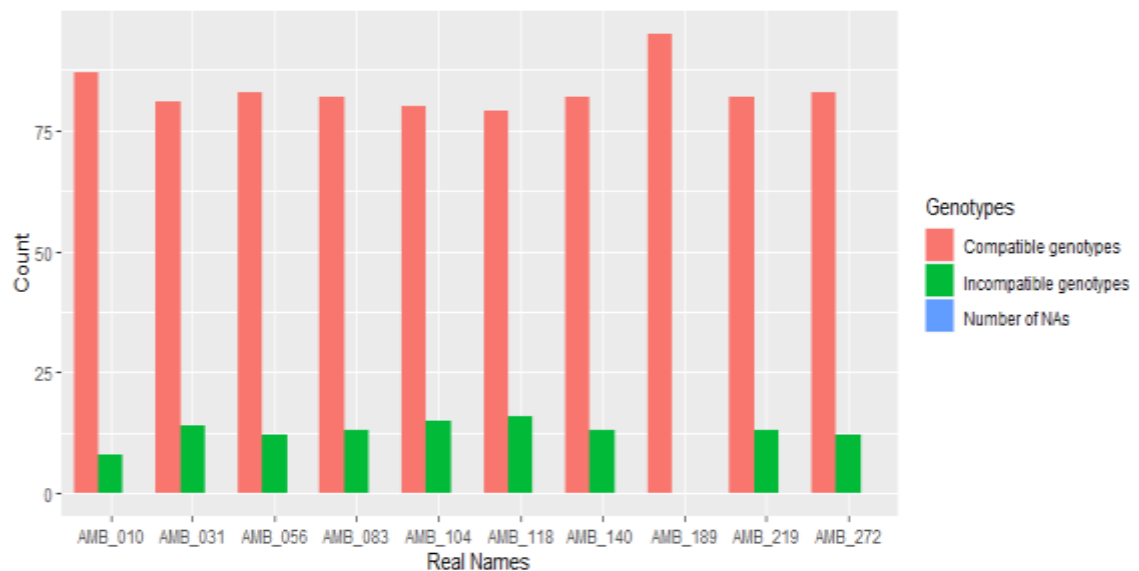
Offspring	Mothers	Fathers
AMB_105	AMB_111	AMB_189

Tabular results:

Real Name	LOD	Total loci	Compatible genotype	Incompatibl e genotype	No. of NAs
AMB_189	12.662	95	95	0	0
AMB_010	-4.842	95	87	8	0
AMB_140	-11.531	95	82	13	0

AMB_219	-12.133	95	82	13	0
AMB_272	-12.240	95	83	12	0
AMB_056	-13.143	95	83	12	0
AMB_083	-13.638	95	82	13	0
AMB_031	-14.434	95	81	14	0
AMB_104	-18.241	95	80	15	0
AMB_118	-19.04	95	79	16	0

Graphical Representation:



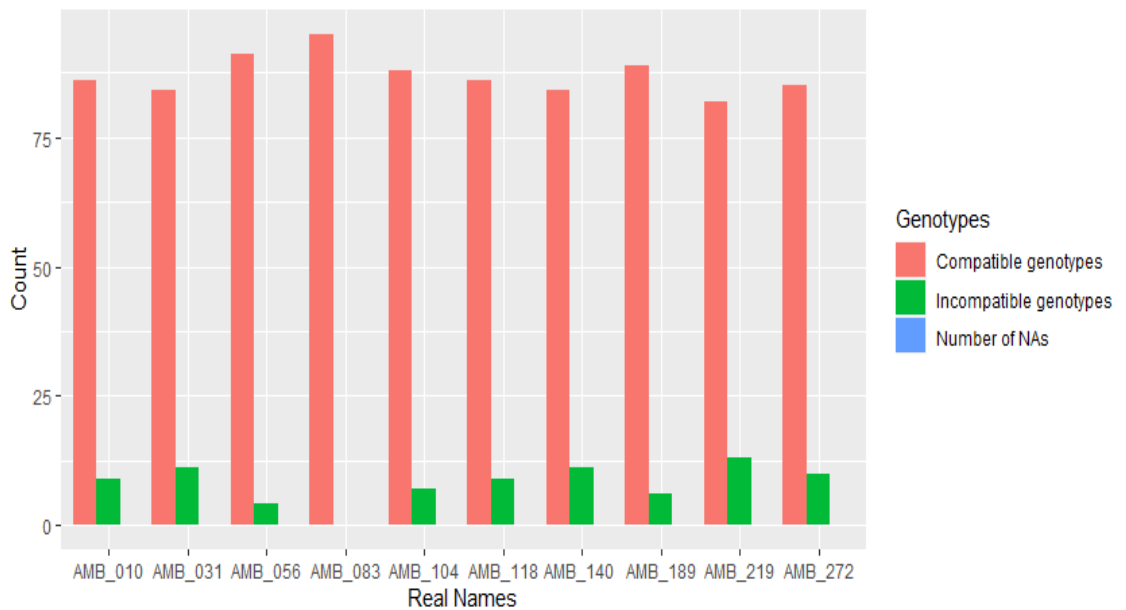
3.

Offspring	Mothers	Fathers
AMB_101	AMB_100	AMB_083

Tabular results:

Real Name	LOD	Total loci	Compatible genotype	Incompatible genotype	No. of NAs
AMB_083	10.386	95	95	0	0
AMB_056	3.289	95	91	4	0
AMB_189	-0.710	95	89	6	0
AMB_104	-3.011	95	88	7	0
AMB_118	-4.603	95	86	9	0
AMB_010	-7.312	95	86	9	0
AMB_140	-7.700	95	84	11	0
AMB_272	-7.807	95	85	10	0
AMB_031	-9.205	95	84	11	0
AMB_219	-10.796	95	82	13	0

Graphical Representation:



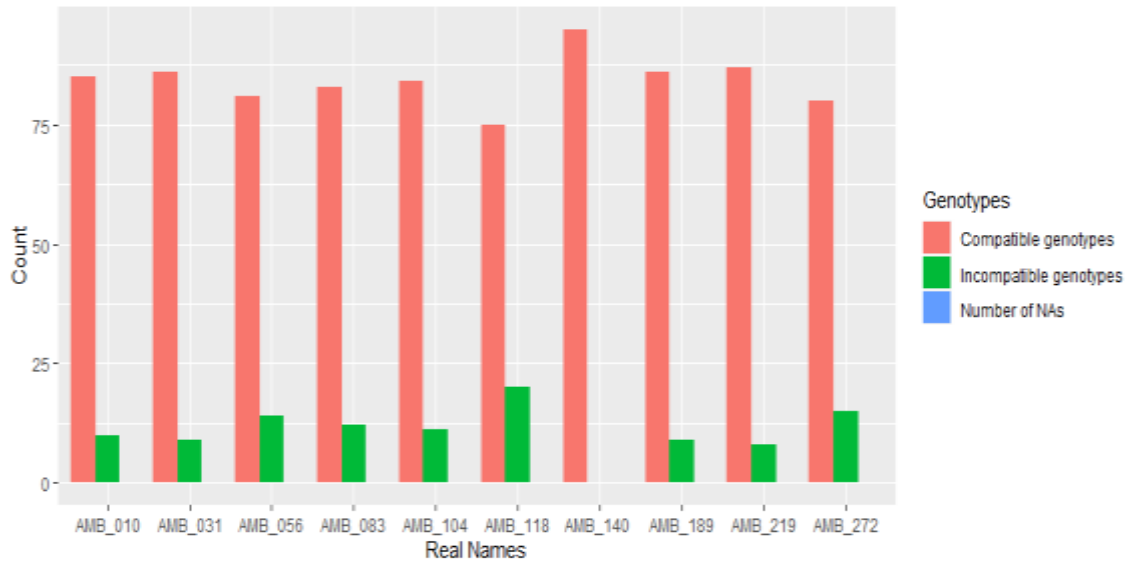
4.

Offspring	Mothers	Fathers
AMB_279	AMB_238	AMB_140

Tabular results:

Real Name	LOD	Total loci	Compatible genotype	Incompatibl e genotype	No. of NAs
AMB_140	17.021	95	95	0	0
AMB_219	3.127	95	87	8	0
AMB_189	-0.979	95	86	9	0
AMB_031	-1.280	95	86	9	0
AMB_010	-1.775	95	85	10	0
AMB_083	-4.270	95	83	12	0
AMB_104	-4.377	95	84	11	0
AMB_056	-10.076	95	81	14	0
AMB_272	-12.377	95	80	15	0
AMB_118	-19.065	95	75	10	0

Graphical Representation:



5.

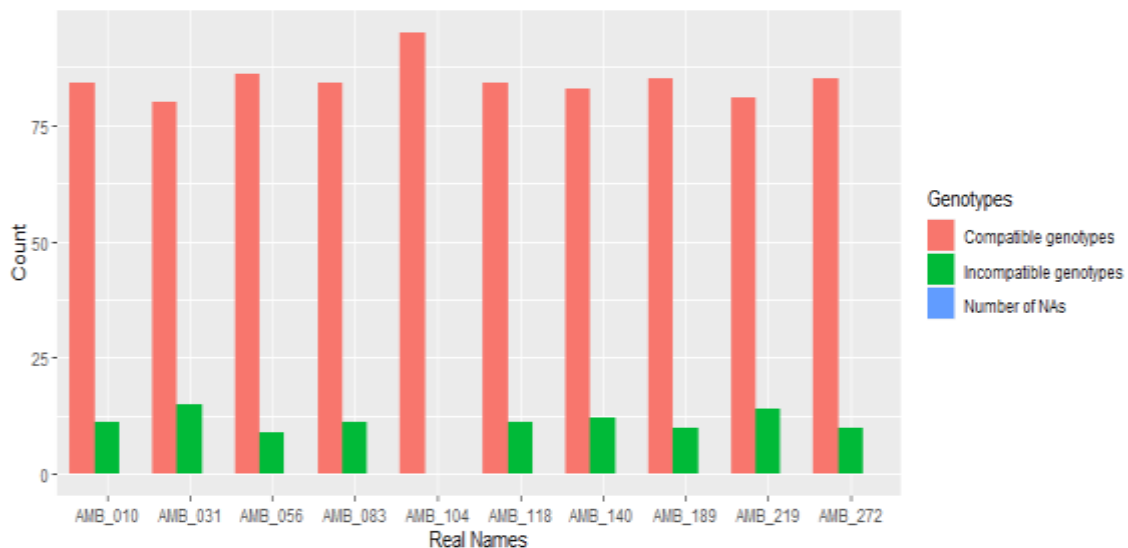
Offspring	Mothers	Fathers
AMB_175	AMB_179	AMB_104

Tabular results:

Real Name	LOD	Total loci	Compatible genotype	Incompatible genotype	No. of NAs
AMB_104	10.195	95	95	0	0
AMB_189	-5.590	95	85	10	0
AMB_056	-6.299	95	86	9	0
AMB_272	-6.494	95	85	10	0
AMB_118	-7.590	95	84	11	0
AMB_083	-8.493	95	84	11	0
AMB_140	-9.891	95	83	12	0

AMB_010	-9.998	95	84	11	0
AMB_219	-12.988	95	81	14	0
AMB_031	-13.784	95	80	15	0

Graphical Representation:



6.

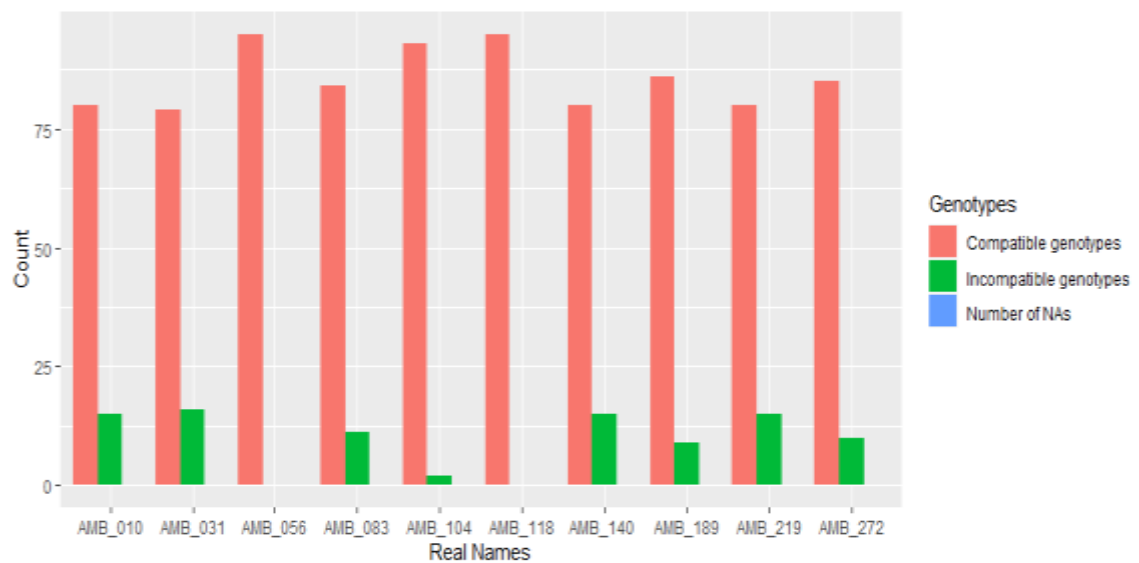
Offspring	Mothers	Fathers
AMB_118	AMB_097	AMB_056

Tabular results:

Real Name	LOD	Total loci	Compatible genotype	Incompatible genotype	No. of NAs
AMB_118	15.126	95	95	0	0
AMB_056	10.000	95	95	0	0
AMB_104	5.407	95	93	2	0
AMB_189	-7.389	95	86	9	0

AMB_272	-8.787	95	85	10	0
AMB_083	-9.883	95	84	11	0
AMB_219	-15.476	95	80	15	0
AMB_140	-16.981	95	80	15	0
AMB_031	-18.078	95	79	16	0
AMB_010	-19.088	95	80	15	0

Graphical Representation:



7.

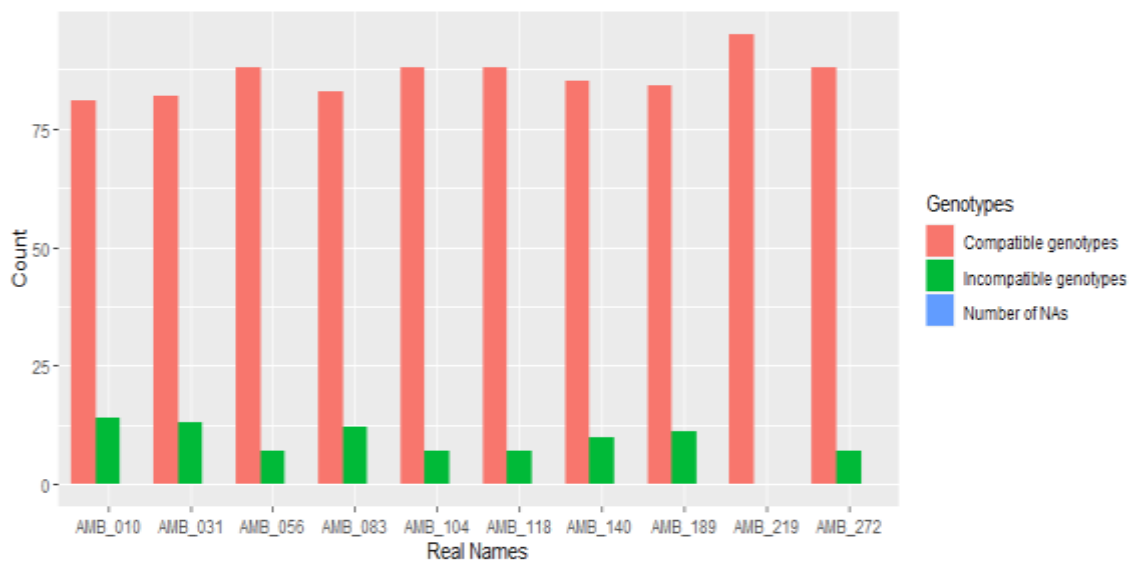
Offspring	Mothers	Fathers
AMB_025	AMB_336	AMB_219

Tabular results:

Real Name	LOD	Total loci	Compatible genotype	Incompatibl e genotype	No. of NAs

AMB_219	12.944	95	95	0	0
AMB_118	-2.259	95	88	7	0
AMB_056	-3.162	95	88	7	0
AMB_272	-4.065	95	88	7	0
AMB_104	-4.366	95	88	7	0
AMB_140	-7.055	95	85	10	0
AMB_189	-9.958	95	84	11	0
AMB_031	-10.044	95	82	13	0
AMB_083	-10.754	95	83	12	0
AMB_010	-15.356	95	81	14	0

Graphical Representation:



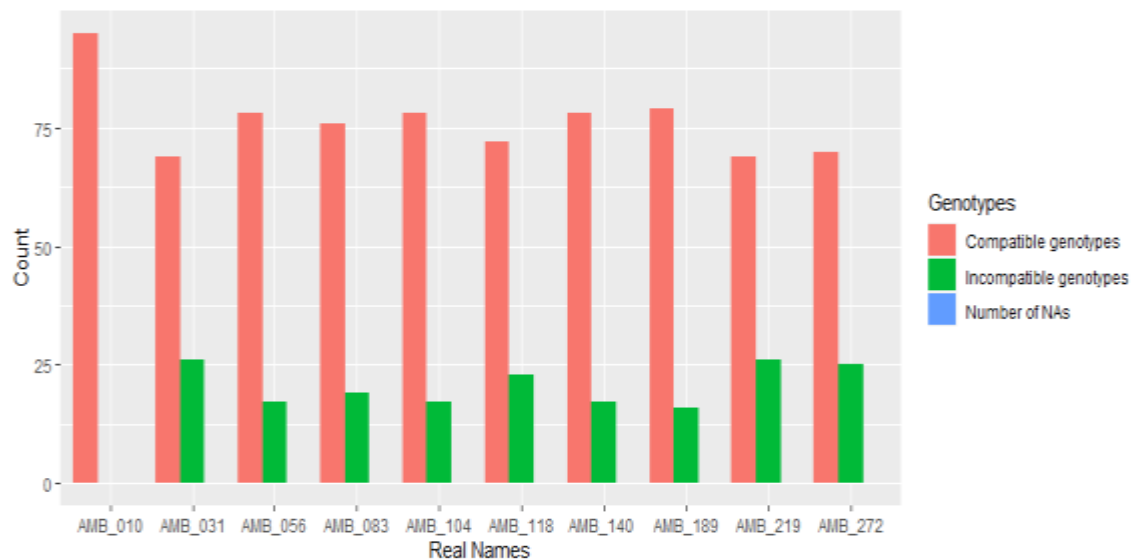
8.

Offspring	Mothers	Fathers
AMB_064	AMB_289	AMB_010

Tabular results:

Real Name	LOD	Total loci	Compatible genotype	Incompatible genotype	No. of NAs
AMB_010	21.558	95	95	0	0
AMB_189	-8.033	95	79	16	0
AMB_056	-10.334	95	78	17	0
AMB_140	-11.237	95	78	17	0
AMB_104	-12.442	95	78	17	0
AMB_083	-12.529	95	76	19	0
AMB_118	-19.927	95	72	23	0
AMB_031	-23.518	95	69	26	0
AMB_219	-24.120	95	69	26	0
AMB_272	-24.830	95	70	25	0

Graphical representation:



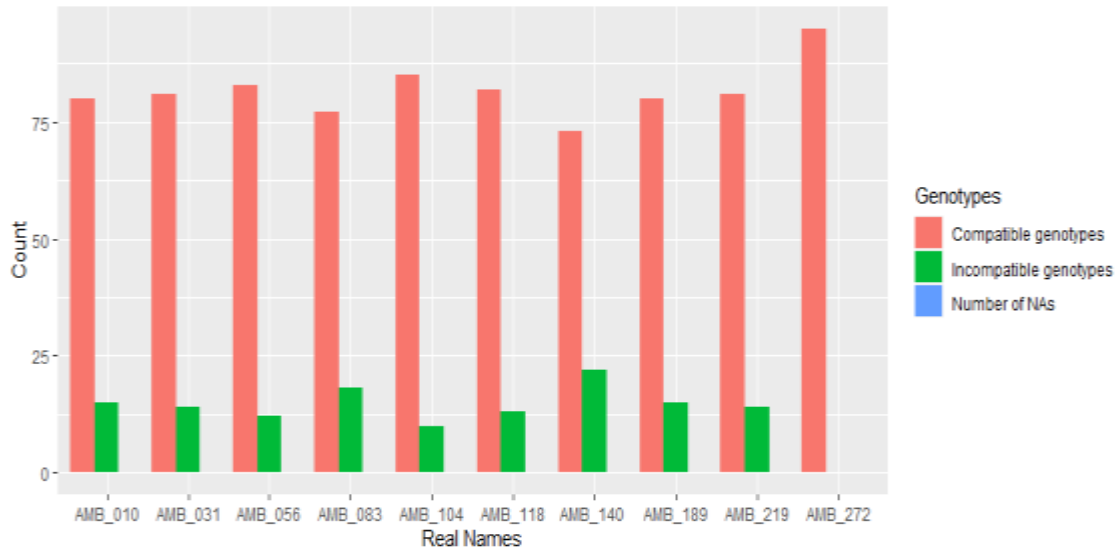
9.

Offspring	Mothers	Fathers
AMB_048	AMB_350	AMB_272

Tabular results:

Real Name	LOD	Total loci	Compatible genotype	Incompatibl e genotype	No. of NAs
AMB_272	16.767	95	95	0	0
AMB_104	-4.738	95	85	10	0
AMB_118	-6.222	95	82	13	0
AMB_056	-7.533	95	83	12	0
AMB_219	-8.222	95	81	14	0
AMB_031	-9.125	95	81	14	0
AMB_189	-11.727	95	80	15	0
AMB_010	-12.329	95	80	15	0
AMB_083	-16.824	95	77	18	0
AMB_140	-23.018	95	73	22	0

Graphical Representation:



10.

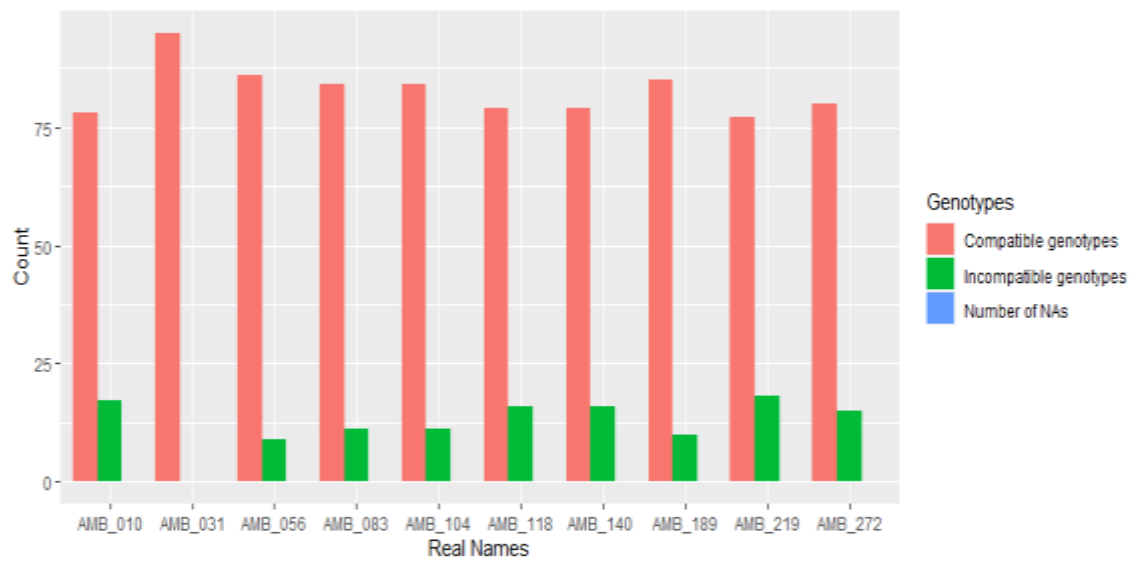
Offspring	Mothers	Fathers
AMB_116	AMB_111	AMB_031

Tabular results:

Real Name	LOD	Total loci	Compatible genotype	Incompatible genotype	No. of NAs
AMB_055	15.778	95	95	0	0
AMB_142	-3.921	95	85	10	0
AMB_292	-4.931	95	86	9	0
AMB_150	-6.824	95	84	11	0
AMB_104	-6.824	95	84	11	0
AMB_168	-13.211	95	79	16	0
AMB_119	-14.824	95	80	15	0

AMB_047	-15.619	95	79	16	0
AMB_315	-17.211	95	77	18	0
AMB_315	-17.617	95	78	17	0

Graphical Representation:



Appendix III: DadApp Workshop Starter Pack Used for the Assessment of *DadApp* an Active Learning Tool

DADAPP WORKSHOP COURSE PACK

PARTICIPANT INFORMATION AND CONSENT FORM

SAMPLE ADULT CONSENT

FOR ENROLLMENT IN THE STUDY

Title of Study: Assessing *DadApp*, a Shiny Web Application that Assigns Paternity in Wild Animal Population, as an Educational Tool in Ecology.

Principal Investigator and institutional affiliation:

Dr. Mercy Akinyi, Animal Science Department, Institute of Primate Research

Email Address: akinyimercy@gmail.com

Tel No: +254722394428

Co-Investigators and institutional affiliation:

David Kiragu Mwaura, Institute of Primate Research

Email Address: davkmwaura@gmail.com

Tel No: +254723124258

Dr. Daniel Kiboi, Biochemistry Department, Jomo Kenyatta University of Agriculture and Technology

Email Address: dmuthui@jkuat.ac.ke

Tel No: +254721249977

Prof. Jenny Tung, Department of Primate Behavior and Evolution, Max Planck Institute for Evolutionary Anthropology, Germany.

Email Address: jtung@eva.mpg.de

Tel No: +49 341 3550-230

Introduction:

The researchers listed above are conducting a study on the usefulness of an interactive educational tool for understanding how genetics can be used to study kin relationships. The purpose of this consent form is to give you the information you will need to help you decide whether or not to be a participant in the study. Feel free to ask any questions about the purpose of the research, what happens if you participate in the study, the possible risks and benefits, your rights as a volunteer, and anything else about the research or this form that is not clear. When we have answered all your questions to your satisfaction, you may decide to participate in the study or not. This process is called 'informed consent'. Once you understand and agree to be in the study, I will request you to sign your name on this form. You should understand the general principles which apply to all participants in a research study: i) Your decision to participate is entirely voluntary ii) You may withdraw from the study at any time and do not have to give a reason for your withdrawal iii) Refusal to participate in the research will not affect your role as a student or staff member at IPR or at any other institution with which you are affiliated. We will give you a copy of this form for your records.

May I continue? YES / NO

This study has approval by The Kenyatta National Hospital-University of Nairobi Ethics and Research Committee protocol No. **P689/09/2023**

WHAT IS THIS STUDY ABOUT?

The researchers listed above are interviewing individuals who are computer literate; have no research or formal expertise in genetic reconstruction of kinship or paternity; have an interest in behavioural or ecological monitoring, and/or wildlife management; and have completed or are currently enrolled in an undergraduate course in ecology, conservation, biology, or genetics. The purpose of the interview is to find out if *DadApp*, a Shiny web application that implements a commonly used pedigree inference and paternity assignment tool, improves user understanding of Mendelian inheritance, genetic kinship and relatedness, and paternity inference, and is a useful active learning tool for teaching these concepts. Participants in this research study will be asked questions about Mendelian inheritance, genetic kinship and relatedness, and paternity inference.

There will be approximately **30 volunteer** participants in this study. We are asking for your consent to consider participating in this study.

WHAT WILL HAPPEN IF YOU DECIDE TO BE IN THIS RESEARCH STUDY?

If you agree to participate in this study, the following things will happen:

You will be given a pre-test to gauge your baseline knowledge of Mendelian inheritance, kinship, and genetic paternity after an initial introductory lecture on the same concepts. Thereafter, you will be asked to interact with simulations and visualisations of *DadApp* that implement the theoretical concepts of paternity for a designated period of 1 hour. Subsequently, you will take a post-test that the researchers will use to measure knowledge or understanding gained through the active learning component. Finally, you will complete a survey to assess your perception of the web application's usability, effectiveness, and its impact on learning.

Six small focus groups of 5 participants each will be conducted to facilitate in-depth discussion on your experience with the web application. This will provide insight into user preferences, challenges, and suggestions for improvement.

ARE THERE ANY RISKS, HARMS, OR DISCOMFORTS ASSOCIATED WITH THIS STUDY?

Research using human participants has the potential to introduce psychological, social, emotional and physical risks. The risks involved in this study are minimal, but include the following:

One potential risk of being in the study is loss of privacy. We will keep all written and verbal responses to the training as confidential as possible. We will use a code number to identify you in a password-protected computer database and will keep all of our paper records in a locked file cabinet. However, no system of protecting your confidentiality can be absolutely secure, so it is still possible that someone could find out you were in this study and could find out information about you.

Also, answering questions in the test setting may be uncomfortable for you. If there are any questions you do not want to answer, you can skip them. You have the right to refuse to answer test questions or questions asked during focus groups.

ARE THERE ANY BENEFITS BEING IN THIS STUDY?

You may benefit by developing a better understanding of Mendelian inheritance, kinship, and relatedness, with the potential to apply this information towards your research interests or to better interpret scientific studies and/or popular science that makes use of these concepts.

You will also receive hands-on training using an increasingly popular type of open-source web application, similar to other Shiny apps you may encounter in future activities.

WILL BEING IN THIS STUDY COST YOU ANY MONEY?

No, participation in this study will not cost you any money.

WILL YOU GET A REFUND FOR ANY MONEY SPENT AS PART OF THIS STUDY?

It is not anticipated that participants will spend money as part of this study.

WHAT IF YOU HAVE QUESTIONS IN FUTURE?

If you have further questions or concerns about participating in this study, please call or send a text message to the study staff at the number provided at the bottom of the consent form.

For more information about your rights as a research participant you may contact the Secretary/Chairperson, Kenyatta National Hospital-University of Nairobi Ethics and Research Committee Telephone No. 2726300 Ext. 44102 email uonknh_erc@uonbi.ac.ke.

WHAT ARE YOUR OTHER CHOICES?

Your decision to participate in research is voluntary. You are free to decline participation in the study and you can withdraw from the study at any time without injustice or loss of any benefits.

CONSENT FORM (STATEMENT OF CONSENT)

Participant's statement

I have read this consent form or had the information read to me. I have had the chance to discuss this research study with a study Principal Investigator, **Dr. Mercy Akinyi**. I have had my questions answered in a language that I understand. The risks and benefits have been explained to me. I understand that my participation in this study is voluntary and that I may choose to withdraw any time. I freely agree to participate in this research study.

I understand that all efforts will be made to keep information regarding my personal identity confidential.

By signing this consent form, I have not given up any of the legal rights that I have as a participant in a research study.

I agree to participate in this research study: **Yes** **No**

I agree to provide contact information for follow-up: **Yes** **No**

Participant **printed** **name:**

Participant signature / Thumb stamp _____ **Date** _____

Researcher's statement

I, the undersigned, have fully explained the relevant details of this research study to the participant named above and believe that the participant has understood and has willingly and freely given his/her consent.

Researcher's Name: _____ **Date:** _____ **Signature** _____

Role in the study: _____

For more information contact

Dr. Mercy Akinyi via email **akinyimercy@gmail.com** or Mobile No: +254722394428

Or

David Kiragu Mwaura via email **davkmwaura@gmail.com** or Mobile No: +254723124258

DadApp Workshop Timetable

Date: March 4, 2024

TIME(EAT)	ACTIVITY	RESPONSIBILITY
10:30am – 10:45am	Arrival and Registration	David Mwaura (JKUAT & KIPRE)
10:45am – 11:00am	Signing of consent forms, Introduction to researchers of the project	Dr. Mercy Akinyi (KIPRE) Dr. Dan Kiboi (JKUAT)

	Introduction to the workshop	
11:00am – 12:00am	PART One: Introductory lecture on Mendelian inheritance, kinship, and genetic paternity inference	Prof. Jenny Tung (Max Planck Institute for Evolutionary Anthropology, Germany)
12:00am – 12:30 am	Answering Pre-test by participants	
12:30am- 1:30am	Lunch Break	
1:30am-2:00pm	PART Two: Introduction to <i>DadApp</i> Shiny App	David Mwaura(KIPRE)
2:00pm-3:15pm	Interactive session with <i>DadApp</i> by Participants	David & Dr. Akinyi (KIPRE)
3:15pm-3:30pm	Coffee Break	
3:30pm-3:45pm	Answering Post-test by Participants	David & Dr. Akinyi (KIPRE)
3:45pm- 4:00pm	Filling survey	David & Dr. Akinyi (KIPRE)
4:00pm-4:10pm	Closing Remarks	Dr. Akinyi & Dr. Kiboi, Prof. Tung

Material and Equipment

- Note book and Pen (To be provided)
- Laptops with internet access

NOTE:

- All participants will be encouraged to have laptops
- During the lecture sessions (Introductory lecture & Introduction to *DadApp*), participants will be encouraged to sit alone
- Answering the pre-test, interactive session with *DadApp*, and post-test, participants will work in pairs
- Participants will be fill the survey individually.

Part One: Introductory Lecture Learning Outcomes**Overview**

The lecture will cover principles of Mendelian inheritance and kinship, including use of these concepts for genetic inference. Examples of applied relevance will be drawn from research on humans and non-human primates.

Learning Outcomes

Upon successful completion of the introductory lecture, the student will be able to:

1. Recognize and explain the principles of Mendel's Laws of Inheritance.
2. Demonstrate Mendel's Law of Segregation by using a Punnett Square
3. Understand the types of genetic markers commonly used in kinship inference and genetic data analysis, including how allele frequencies are calculated and how the accuracy of inference scales with increasing amounts of data.
4. Provide examples in which kinship inference is useful in basic research

Additional Resources and Study Materials

For more information please find below the following references:

1. Jones, A. G., Small, C. M., Paczolt, K. A., & Ratterman, N. L. (2010). A practical guide to methods of parentage analysis. *Molecular Ecology Resources*, 10(1), 6–30. <https://doi.org/10.1111/j.1755-0998.2009.02778.x>
2. Marshall, T. C., Slate, J., Kruuk, L. E. B., & Pemberton, J. M. (1998). Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology*, 7(5), 639–655. <https://doi.org/10.1046/j.1365-294x.1998.00374.x>
3. DNA Paternity Test <https://my.clevelandclinic.org/health/diagnostics/10119-dna-paternity-test>
4. Fundamentals of Genetics <https://catdir.loc.gov/catdir/samples/cam041/2003048463.pdf>

Pre-test:

Please read the following questions and answer appropriately.

Question 1: What role does statistical analysis play in paternity inference?

- a) It evaluates the physical resemblance between the alleged mother and child.
- b) It measures the likelihood of paternity based on genetics.
- c) It measures the emotional connection between the father and child.
- d) It assesses the flow of genetic markers from offspring to parents.

Question 2: What of the following statements is FALSE?

- a) Heterozygotes are individuals with dissimilar allele on homologous chromosomes
- b) An organism with two identical alleles of a gene in their genome is called a homozygote
- c) A recessive trait is a genetic characteristic that is expressed only when an individual carries two copies of the allele
- d) Dominant traits are masked by recessive traits unless an individual inherits recessive allele from both parents

Question 3: As minor allele frequencies (MAF) increase in a dataset, how might this impact the accuracy of genetic inference?

- a) Accuracy remains constant regardless of MAF
- b) Accuracy increases
- c) Identifies rare variants in the population which negatively impacts paternal inference
- d) None of the above

Question 4: If the frequency of the B allele in a finite population is given by p and the frequency of the alternative allele b is given by q, which of the following statements is false:

- a) If B and b are the only possible alleles at this locus, $p = 1 - q$
- b) If two Bb heterozygotes mate, their offspring are expected to follow the ratio 1 BB: 2Bb : 1bb
- c) Under random mating, the frequency of heterozygotes in the population is given by $2pq$.

- d) If B is selectively advantageous, b will definitely increase from one generation to the next generation

Question 5: What is the significance of having a shared genotype between an alleged father and a child in paternity testing?

- a) It increases the allele frequency of the shared genetic marker
- b) It increases the probability of paternity
- c) It decreases the likelihood ratio of the tested relationship
- d) It indicates an error in the testing process

Question 6: What information, in addition to genetic marker data, is ideally required for a more accurate paternity inference of an offspring?

- a) Only the genetic profiles of the alleged father and child
- b) Genetic profiles of the mother, alleged father, and child
- c) Only the genetic history of the alleged father
- d) The allele frequency of the maternal lineage only

Question 7: Why is it advisable to use multiple genetic markers rather than relying on a single marker for paternity testing?

- a) Single markers provide more accurate results
- b) Multiple markers increase the complexity of the analysis

- c) Allele frequency of single markers are the same in different population and geographical area
- d) Multiple markers enhance the reliability and accuracy of results

Question 8: Why is it important to consider the population's allele frequencies when performing paternity testing?

- a) Allele frequency determines the genotypic frequency of the offspring only
- b) It helps identify potential mutations in both offspring and father's DNA
- c) It provides a baseline for calculating likelihood ratios for patriarchal inference
- d) It influences the child's height and weight

Question 9: By itself, the equation: $p + q = 1$ tells us that

- a) The frequencies of alleles in a population are not changed by random mating
- b) The sum of the frequencies of two alternate alleles at a locus must total 1.0
- c) Genotype frequencies can be calculated if allele frequencies are known
- d) There are only two possible genotypes in this population
- e) A population that does not conform to this equation is evolving

Question 10: You are a researcher about to embark on paternity inference. You calculate that you can afford to genotype 20 loci from a set of 40 known biallelic sites, in a fixed number of samples. To increase the power and accuracy of paternity assignment, you choose:

- a) The set of sites with the highest minor allele frequencies.
- b) A set of 20 sites at random.
- c) A set of sites that equally represent the range of minor allele frequencies in the possible set.
- d) The set of sites with the lower minor allele frequencies.

Highlight where the following statements are TRUE or FALSE

Question 11: True or False: A high allele frequency indicates that this particular allele is less common in the population.

Question 12: True or False: Calculating the likelihood ratio of potential fathers without maternal genotype calls results to low score of paternal inference

Question 13: True or False: Increase in amount of genetic information has negative impact on the statistical confidence on paternal assignment

Question 14: True or False: Allele frequencies is the same between different populations and geographical regions.

Question 15: True or False: multivariants are less informative in distinguishing individuals within a population.

Question 16: True or False: In paternity testing, the comparison of the child's genotype with those of both the mother and potential fathers helps determine the likelihood of biological parentage.

Question 17: True or False: Low allele frequency suggests that this particular allele is rare in the population.

Question 18: True or False: The accuracy of paternal assignment is positively influenced by the quantity of compatible genotypes shared between the alleged father and offspring.

Question 19: True or False: If both parents are homozygous for a specific locus, their offspring will also be homozygous for that locus.

Question 20: True or False: Allele frequency depends on the total number of alleles present, and multiple sets of genotype frequencies can be produced from the same allele frequency.

PART TWO: *DadApp* Shiny App Training

Overview

This session will cover the principle of Shiny Apps and introduce a specific app, *DadApp*, as a tool for understanding paternity inference.

Learning Outcomes

Upon successful completion of the introductory lecture, the participant will:

1. Develop introductory familiarity with Shiny Apps
2. Learn how to calculate allele frequencies using *DadApp*
3. Understand how to representing Mendelian inheritance in *DadApp*

4. Be able to use *DadApp* to assign paternity or identify most likely parent-offspring pairs using a comparison of statistical likelihoods

Additional Resources and Materials

For more information please find below the following references:

1. Welcome to Shiny <https://shiny.posit.co/r/getstarted/shiny-basics/lesson1/index.html>
2. *DadAPP* <https://zenodo.org/records/8171383>
3. Jones, A. G., Small, C. M., Paczolt, K. A., & Ratterman, N. L. (2010). A practical guide to methods of parentage analysis. *Molecular Ecology Resources*, 10(1), 6–30. <https://doi.org/10.1111/j.1755-0998.2009.02778.x>
4. Marshall, T. C., Slate, J., Kruuk, L. E. B., & Pemberton, J. M. (1998). Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology*, 7(5), 639–655. <https://doi.org/10.1046/j.1365-294x.1998.00374.x>

WORKSHEET on *DadApp* Training

Instruction on uploading files:

- a) Using a laptop connected to internet search the following link in your browser <https://kiragu-mwaura.shinyapps.io/dadapp>.

Or click the link [dadapp](https://kiragu-mwaura.shinyapps.io/dadapp)

The underlying code for *DadApp* is also publicly available on GitHub at

https://github.com/KIRAGU-MWAURA/DadApp_Shiny_Web_App

Or click the link [rawcode](#)

DadApp: A web server for calculating parentage likelihoods from genotype data.

Choose population-level genotype data (for allele frequency estimation)

Browse... No file selected

Choose Potential Father(s) Data file

Browse... No file selected

Choose GMapping() Data file

Browse... No file selected

Availability of Mother's genotype data

no

☐ Would you like to proceed without maternal genotype data?

Submit

Warning: Calculating the logarithm of odds of potential fathers without the maternal genotype calls gives low scores.

100 Sites Haplotype Plot Detailed Table Detailed Chart

Warning: Calculating the logarithm of odds of potential fathers without the maternal genotype calls gives low scores.

File selection dialog showing the 'Test Data' folder containing files like 'Back_genome101', 'mother101', 'mother102', 'Offspring', and 'whole_genome101'.

DadApp: A web server for calculating parentage likelihoods from genotype data.

Choose population-level genotype data (for allele frequency estimation):

Choose Potential Father(s) Data file:

Choose Offspring(s) Data file:

Availability of Mother's genotype data:

☐ Would you like to proceed without maternal genotype calls?

Warning: Calculating the logarithm of odds of potential fathers without the maternal genotype calls gives low scores

[L200 Table](#) [Histogram Plot](#) [Detailed Table](#) [Detailed Chart](#)

- c) Upload the fathers' genotype dataset as shown in B and successful upload will look like below

Choose population-level genotype data (for allele frequency estimation):

Choose Potential Father(s) Data file:

Choose Offspring(s) Data file:

Availability of Mother's genotype data:

☐ Would you like to proceed without maternal genotype calls?

Warning: Calculating the logarithm of odds of potential fathers without the maternal genotype calls gives low scores

[L200 Table](#) [Histogram Plot](#) [Detailed Table](#) [Detailed Chart](#)

- d) In cases where the mother's genotype is absent, select no for 'Availability of Mother's genotype data' and highlight/tick 'Would you like to proceed without maternal genotype data' as shown below

DadApp: A web server for calculating parentage likelihoods from genotype data.

Choose population-level genotype data (for allele frequency estimation):

Choose Potential Father(s) Data file:

Choose Offspring(s) Data file:

Availability of Mother's genotype data:

☒ Would you like to proceed without maternal genotype calls?

Warning: Calculating the logarithm of odds of potential fathers without the maternal genotype calls gives low scores

[LOD Table](#) [Heterozygote Plot](#) [Detailed Table](#) [Detailed Chart](#)

- e) In cases where the mother's genotype is present, select yes for '*Availability of Mother's genotype data*' and upload the mother's genotype dataset as shown below

DadApp: A web server for calculating parentage likelihoods from genotype data.

Choose population-level genotype data (for allele frequency estimation):

Choose Potential Father(s) Data file:

Choose Offspring(s) Data file:

Availability of Mother's genotype data:

Choose mother(s) Data file if present:

[LOD Table](#) [Heterozygote Plot](#) [Detailed Table](#) [Detailed Chart](#)

- f) Click '*SUBMIT*' once you have uploaded every dataset as per project's case requirement for results

Exercise

Dataset that will be used in this exercise is available on GitHub through the following [link](https://github.com/KIRAGU-MWAURA/DadApp_Shiny_Web_App/tree/main/Trial%20Data):

https://github.com/KIRAGU-MWAURA/DadApp_Shiny_Web_App/tree/main/Trial%20Data

1. In this question, we will evaluate how including mother's genotype data influences the accuracy in paternity assignment. We will consider two scenarios and record their LOD scores:

(i) In the absence of the mother's genotype

a) Assigned Father's ID.....LOD scores
.....

(ii) When the mother's genotype is known.

a) Assigned Father's ID.....LOD scores
.....

What impact does the inclusion of the mother's genotype have on the likelihood ratio for determining the true father of a specific offspring?

- A) The likelihood ratio is unaffected by the mother's genotype.
- B) The likelihood ratio increases when the mother's genotype is considered.
- C) The likelihood ratio decreases when the mother's genotype is considered.
- D) The mother's genotype is the sole determinant of the likelihood ratio

E) The father's genotype is the sole determinant of the likelihood ratio

2. In this exercise, we will assess the impact of decrease in the amount of genetic data available on paternity assignment, including the LOD scores of the assigned true father. We will upload datasets under three scenarios and record the LOD scores:

i) Fathers', mother's and offspring datasets with 100 loci typed

**a) Assigned Father's ID..... LOD scores
.....**

ii) Fathers', mother's and offspring's datasets with 50 loci typed

a) Assigned Father's ID..... LOD scores

iii) Fathers', mothers' and offspring's datasets with 25 loci typed

**a) Assigned Father's ID.....LOD scores
.....**

How does a reduction in the number of genotyped loci impact the accuracy of paternity assignment? Which of the following statements is true?

A) It does not affect the accuracy of paternity determination.

B) Decreasing the number of genotyped loci increases the LOD score of the actual father of the offspring of interest and the precision of identifying the actual father.

C) A decrease in the number of genotyped loci may lead to inconclusive results in determining the actual father.

D) A decrease in the number of genotyped loci increases the number of compatible loci between the true parents' genotype and the offspring.

E) A decrease in the number of genotyped loci may reduce false positive and false negative paternity assignment due to a reduction in coincidental genotype matches or mismatches.

3. In this question, we will assess the impact of increased genotyping error rates in the data set on the associated LOD scores of the true father and the overall assignment of paternity. We will upload datasets in three scenarios

i) Fathers' datasets with 10% error rate

a) Assigned Father's ID.....LOD scores
.....

ii) Fathers' datasets with 30% error rate

a) Assigned Father's ID..... LOD scores
.....

iii) Fathers' datasets with 50% error rate

a) Assigned Father's ID..... LOD scores
.....

In the context of paternity testing, how does an increase in the error rate impact the likelihood ratio of the actual father of an offspring of interest in a population?"

A) An increase in error rate does not affect the likelihood ratio.

B) An increase in error rate decreases the likelihood ratio.

C) An increase in error rate increases the likelihood ratio.

D) The likelihood ratio is independent of the error rate in paternity testing

Post-Test

Highlight where the following statements are TRUE or FALSE

Question 1: True or False: If both parents are homozygous for a specific locus, their offspring will also be homozygous for that locus

Question 2: True or False: Loci of rare variants are less informative in distinguishing individuals within a population.

Question 3: True or False: A high allele frequency indicates that a particular allele is less common in the population.

Question 4: True or False: The accuracy of paternal assignment is positively influenced by the quantity of compatible genotypes shared between the alleged father and offspring.

Question 5: True or False: Calculating the likelihood ratio of potential fathers without maternal genotype calls results to low score of paternal inference

Question 6: True or False: Allele frequency depends on the total number of alleles present, and multiple sets of genotype frequencies can be produced from the same allele frequency

Question 7: True or False: Allele frequencies is the same between different populations and geographical regions.

Question 8: True or False: In paternity testing, the comparison of the child's genotype with those of both the mother and potential fathers helps determine the likelihood of biological parentage.

Question 9: True or False: An increase in the amount of genetic information has a negative impact on the statistical confidence on paternal assignment

Question 10: True or False: Low allele frequency suggests that this particular allele is rare in the population.

Please read the following questions and answer appropriately:

Question 11: As minor allele frequencies (MAF) increase in a dataset, how might this impact the accuracy of genetic inference?

- a) Accuracy remains constant regardless of MAF
- b) Accuracy increases
- c) Identifies rare variants in the population which negatively impact paternal inference
- d) None of the above

Question 12: Which of the following statements is FALSE?

- a) Heterozygotes are individuals with dissimilar alleles on homologous chromosomes
- b) An organism with two identical alleles of a gene in their genome is called a homozygote
- c) A recessive trait is a genetic characteristic that is expressed only when an individual carries two copies of the allele
- d) Dominant traits are masked by recessive traits unless an individual inherits a recessive allele from both parents

Question 13: Why is it advisable to use multiple genetic markers rather than relying on a single marker for paternity testing?

- a) Single markers provide more accurate results
- b) Multiple markers increase the complexity of the analysis
- c) Allele frequency of single markers are the same in different population and geographical area
- d) Multiple markers enhance the reliability and accuracy of results

Question 14: You are a researcher about to embark on paternity inference. You calculate that you can afford to genotype 20 loci from a set of 40 known biallelic sites, in a fixed number of samples. To increase the power and accuracy of paternity assignment, you choose:

- a) The set of sites with the highest minor allele frequencies.
- b) A set of 20 sites at random.
- c) A set of sites that equally represent the range of minor allele frequencies in the possible set.
- d) The set of sites with the lower minor allele frequencies

Question 15: What information, in addition to genetic marker data, is ideally required for a more accurate paternity inference of an offspring?

- a) Only the genetic profiles of the alleged father and child
- b) Genetic profiles of the mother, alleged father, and child

- c) Only the genetic history of the alleged father
- d) The allele frequency of the maternal lineage only

Question 16: Why is it important to consider the population's allele frequencies when performing paternity testing?

- a) Allele frequency determines the genotypic frequency of the offspring only
- b) It helps identify potential mutations in both offspring and the father's DNA
- c) It provides a baseline for calculating likelihood ratios for patriarchal inference
- d) It influences the child's height and weight

Question 17: If the frequency of the B allele in a finite population is given by p and the frequency of the alternative allele q is given by y , which of the following statements is false:

- e) If B and b are the only possible alleles at this locus, $q = 1 - p$
- f) If two Bb heterozygotes mate, their offspring are expected to follow the ratio 1 BB: 2Bb: 1bb
- g) Under random mating, the frequency of heterozygotes in the population is given by $2pq$
- h) If B is selectively advantageous, b will definitely increase from one generation to the next generation

Question 18: What role does statistical analysis play in paternity inference?

- a) It evaluates the physical resemblance between the alleged mother and child.

- b) It measures the likelihood of paternity based on genetics.
- c) It measures the emotional connection between the father and child.
- d) It assesses the flow of genetic markers from offspring to parents.

Question 19: What is the significance of having a shared genetic allele between an alleged father and a child in paternity testing?

- a) It increases the allele frequency of the shared genetic marker
- b) It increases the probability of paternity
- c) It decreases the likelihood ratio of the tested relationship
- d) It indicates an error in the testing process

Question 20: By itself, the equation $p + q = 1$ tells us that

- a) The frequencies of alleles in a population are not changed by random mating
- b) The sum of the frequencies of two alternate alleles at a locus must total 1.0
- c) Genotype frequencies can be calculated if allele frequencies are known
- d) There are only two possible genotypes in this population
- e) A population that does not conform to this equation is evolving

Workshop Survey

Disclaimer:

- The survey is anonymous

- The survey should be filled in individually

Find the link below to the [survey](#)

<https://forms.gle/smtfwC3e2S5WSoYU9>

QR code to the survey Google form

