ASSESSMENT OF GENETIC RELATEDNESS TOOLS FOR ANCIENT DNA USING PEDIGREE SIMULATIONS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

ŞEVVAL AKTÜRK

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
MOLECULAR BIOLOGY AND GENETICS

AUGUST 2023

Approval of the thesis:

ASSESSMENT OF GENETIC RELATEDNESS TOOLS FOR ANCIENT DNA USING PEDIGREE SIMULATIONS

submitted by **ŞEVVAL AKTÜRK** in partial fulfillment of the requirements for the degree of **Master of Science** i**n Molecular Biology and Genetics, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. Mesut Muyan
Head of the Department, **Biological Sciences** _____

Prof. Dr. Mehmet Somel
Supervisor, **Biological Sciences, METU** _____

Examining Committee Members:

Assist. Prof. Dr. Aybar Can Acar
Health Informatics, METU _____

Prof. Dr. Mehmet Somel
Biological Sciences, METU _____

Assist. Prof. Dr. Gülşah Merve Kılınç
Bioinformatics, Hacettepe University _____

Assoc. Prof. Dr. Torsten Günther
Organismal Biology, Uppsala University _____

Assist. Prof. Dr. Burçak Otlu
Health Informatics, METU _____

Date: 24.08.2023

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Surname: Şevval Aktürk

Signature :

ABSTRACT

ASSESSMENT OF GENETIC RELATEDNESS TOOLS FOR ANCIENT DNA USING PEDIGREE SIMULATIONS

Aktürk, Şevval
M.S., Department of Molecular Biology and Genetics
Supervisor: Prof. Dr. Mehmet Somel

August 2023, 109 pages

Understanding the biological relationships among individuals retrieved in archaeological excavations has drawn attention for a long time in the ancient DNA field. The increasing number of SNP genotype or whole-genome sequence data has recently opened up the opportunity to estimate genetic relatedness using genome-wide markers. However, next-generation sequencing of ancient genomes at low depth causes low precision in genotype calling due to missing alleles, post-mortem damage, and fragmentation. To overcome these challenges, different tools have been developed, including algorithms relying on genotype likelihoods and population allele frequencies (e.g., NgsRelate and lcMLkin) or tools comparing genotype mismatch rates between a pair (e.g., READ). To systematically evaluate the reliability of these three most commonly used tools in the presence of a limited number of SNPs and inbreeding, we used ancient genome data produced from simulated pedigrees. Our results show that related pairs can be accurately classified as first-degree, even down to 1K shared SNPs, with 85% and 96% F_1 scores using READ and NgsRelate or lcMLkin, respectively. Distinguishing unrelated pairs from close relatives down to third-degree is possible with high accuracy (F_1 = 99%) at 5K shared SNPs using NgsRelate and

lcMLkin. Further, with 10K shared SNPs, NgsRelate and lcMLkin outperform READ in the differentiation of third-degree from second-degree relatedness with a 95% F_1 and 96% F_1 (80% for READ). Last, inbreeding (e.g., first cousin mating) leads to the overestimation of kinship coefficients. These results are promising but also call for exploring novel approaches for kinship estimation with ultra-low coverage genomes.

Keywords: Kinship coefficient, Ancient DNA, Simulation, Pedigree, Relatedness

ÖZ

SOYAĞACI SİMÜLASYONLARI KULLANILARAK GENETİK AKRABALIK ARAÇLARININ ANTİK DNA İÇİN DEĞERLENDİRMESİ

Aktürk, Şevval
Yüksek Lisans, Moleküler Biyoloji ve Genetik Bölümü
Tez Yöneticisi: Prof. Dr. Mehmet Somel

Ağustos 2023 , 109 sayfa

Arkeolojik kazılarda elde edilen bireyler arasındaki biyolojik yakınlığın anlaşılması, uzun zamandır antik DNA alanında dikkat çekmektedir.Tekil nukleotid polimorfizm (TNP) genotip ve tüm genom dizi verisindeki artış, yakın zamanda genom çapında belirteçlerin kullanılmasıyla genetik akrabalığı tahmin etme olanağı sağlamıştır.Halbuki, eksik aleller ile birlikte ölüm sonrası hasar ve parçalanma yüzünden antik genomların düşük derinlikte yeni nesil dizilenmesi, genotip çağırmada doğruluğun azalmasına yol açar. Bu zorlukların üstesinden gelebilmek için genotip olabilirliğine ve popülasyon alel frekanslarına dayanan (örn., NgsRelate ve lcMLkin) ya da bireyler arasındaki genotip uyumsuzluğu oranını karşılaştıran (örn., READ) araçlar dahil olmak üzere farklı araçlar geliştirilmiştir. Yaygın olarak kullanılan bu iki aracın (NgsRelate ve READ), kısıtlı TNP sayıları ve akraba evliliği varlığında sistematik bir şekilde güvenilirliklerini test etmek için simüle edilmiş soyağaçlarından üretilen antik genom verisini kullandık. Sonuçlarımız gösterdi ki bin TNPe kadar bile, READ ve NgsRelate (ya da lcMLkin) kullanarak akraba çiftler sırasıyla %85 and %96 F_1 skor ile doğru bir şekilde birinci derece olarak sınıflandırılabilir. Akraba olmayan çiftlerin,

üçüncü dereceye kadar yakın akrabalardan yüksek doğrulukta ayırt edilmesi, paylaşılan beş bin TNPte NgsRelate ve lcMLkin ile mümkündür (F_1 = %99). Ayrıca, ortak on bin TNP ile NgsRelate ve lcMLkin, READ'i, ikinci derece akrabalığın üçüncü dereceden ayrıştırılmasında %95 ve %96 F_1 ile geçer (READ için %80). Son olarak, akraba evliliği (örn. kuzen çiftleşmesi) akrabalık katsayısının olduğundan fazla tahmin edilmesine yol açar. Bu sonuçlar gelecek vadediyor fakat aynı zamanda oldukça düşük kapsamlı genomlar ile akrabalık katsayısı tahmini için yeni yaklaşımların araştırılmasını gerektirir.

Anahtar Kelimeler: Akrabalık katsayısı, Antik DNA, Simülasyon, Soyağacı, Akrabalık

To my family

ACKNOWLEDGMENTS

I want to thank Mehmet Somel and Mehmet Çetin for entrusting me at the start of this journey. As someone new to bioinformatics, Mehmet Çetin kindly guided me through building pipelines and handling errors. The support and encouragement of my supervisor, Mehmet Somel, played a significant role in shaping my scientific viewpoint. With his endless passion and diligence, he helped me move this research forward from start to finish.

I owe special thanks to the kinship team of our group, Igor Mapelli, Merve Nur Güler, and Kanat Gürün, who made precious contributions to this study. Their valuable suggestions and efforts greatly impacted the study throughout the entire process. Specifically, I am deeply thankful to Merve Nur Güler. She provided significant assistance in generating and analyzing data and ensuring the results were valuable.

The invaluable assistance of Ekin Sağlıcan, Kıvılcım Başak Vural, and Seda Çokoğlu was integral to the completion of this research.

I extend my appreciation to Nefize Ezgi Altınışık and Dilek Koptekin. I am inspired to keep moving forward by their dedication and expertise. Whenever I had questions or needed help, they were always kind enough to give me the necessary answers. The feedback I received from them helped me make progress with my research. Also, I want to express my admiration for the humble and friendful individuals comprising the Comparative and Evolutionary Biology (CompEvo) team. Their unwavering dedication to promoting innovative ideas and driving our research forward deserves recognition, especially in challenging circumstances. I am grateful to all members for valuing cooperation and assisting my research.

I want to acknowledge the support from the Scientific and Technical Research Council of Turkey (TUBITAK, no. 122Z433 to Yılmaz Selim Erdal and 2210-A National MSc Scholarship Program).

I cannot imagine this journey without the heartfelt support, faith, and patience of my dear mom and dad, Ferah and Rahmi. I am fortunate to have a wonderful family that prioritizes my well-being.

I am also grateful for my dear friend Yaren Canten, who has always been there for me in times of trouble with her kind and generous heart.

I immensely appreciate Dr. Ali Bayramoğlu for introducing me to the field of contextual behavioral science. He opened my eyes to the idea that our actions are not the outcomes of our thoughts and emotions. Thanks to this science, I now understand that being adaptable in our behaviors is more important than what goes on in our minds. He was a great help to me during the last stages of my thesis, keeping me motivated and focused.

My final gratitude goes to Arda Sevkar for his unwavering emotional and analytical support. During my toughest moments, he always comes through with practical solutions that have been consistently helpful. His companionship had been precious to me.

TABLE OF CONTENTS

ABSTRACT v

ÖZ vii

ACKNOWLEDGMENTS x

TABLE OF CONTENTS xii

LIST OF TABLES xvi

LIST OF FIGURES xxi

LIST OF ABBREVIATIONS xxv

CHAPTERS

1 INTRODUCTION 1

1.1 Genetic Relatedness 1

1.2 Ancient DNA 4

1.3 Kinship Estimation Methods 6

1.3.1 Method-of-moment Based Approaches 6

1.3.1.1 READ 6

1.3.1.2 TKGWV2 7

1.3.2 Maximum likelihood-based methods 7

1.3.2.1 NgsRelate and lcMLkin 7

1.3.3 KIN 8

1.3.4 ancIBD . . . 8
1.3.5 Genotype Calls and Genotype Likelihoods . . . 9
1.3.6 Limitations of the Existing Methods for Ancient DNA . . . 10
1.3.7 Kinship Studies With Ancient Genomes . . . 11
1.3.8 Research Objectives . . . 13
1.4 Structure of the Thesis . . . 13
2 MATERIAL AND METHODS . . . 15
2.1 Pedigree Simulations . . . 15
2.2 Ancient Sequence Simulations . . . 17
2.3 Preprocessing of Simulated Ancient Genomes . . . 18
2.4 Genotyping and Downsampling . . . 19
2.5 Genetic Relatedness Estimation . . . 20
2.5.1 READ . . . 20
2.5.2 NgsRelate . . . 21
2.5.3 lcMLkin . . . 22
2.6 Classification of Kinship Coefficient Estimates . . . 23
2.6.1 Classification Metrics . . . 25
2.7 Statistical Tests On Kinship Coefficient Estimates . . . 27
2.7.1 Binomial Test . . . 27
2.7.2 Linear Mixed Effect Model . . . 28
2.7.3 Levene's Test . . . 31
3 RESULTS . . . 33
3.1 Simulated Ancient Data . . . 33

3.2 The Distribution of Kinship Coefficient Estimates 34
3.2.1 First-degree Relatedness 34
3.2.2 Second-degree Relatedness 36
3.2.3 Third-degree Relatedness 37
3.2.4 Unrelatedness 38
3.3 Comparison of Kinship Coefficient Estimates Using Linear Mixed Effect Model 40
3.3.1 First-degree Relatedness 41
3.3.2 Second-degree Relatedness 42
3.3.3 Third-degree Relatedness 43
3.3.4 Unrelatedness 44
3.3.5 Overview of Linear Mixed Effect Model Results 45
3.4 Bias in Kinship Coefficient Estimates 46
3.5 The Variations of Kinship Coefficient Estimates 47
3.6 The Classification and Accuracy of Kinship Coefficient Estimates . . 49
3.6.1 First-degree Relatedness 50
3.6.2 Second-degree Relatedness 50
3.6.3 Third-degree Relatedness 51
3.6.4 Unrelatedness 51
3.6.5 Geometric Mean for the Classification 51
3.6.6 F-score of Kinship Coefficient Estimates 54
3.7 Kinship Inference In the Presence of Inbreeding 55
3.7.1 First-degree Relatedness With One Inbred Offspring 56

3.7.2 Second-degree Relatedness With One Inbred Offspring 57

3.7.3 Overview of Inbreeding Results 61

4 DISCUSSION 63

REFERENCES 69

APPENDICES

A BASH SCRIPT TO PRODUCE BINARY GL FILE 79

B LMM RESULTS 81

C LEVENE'S TEST RESULTS 97

D BINOMIAL TEST RESULTS 99

LIST OF TABLES

TABLES

Table 1.1 Cotterman and kinship coefficients for relationship types in the absence of inbreeding ($\theta = k_1/4 + k_2/2$) (Weir et al., 2006). 3

Table 2.1 Description of simulated pairs. The number of pedigrees represents the number of distinct pedigrees in each run. 17

Table 2.2 Theoretical kinship coefficient values and classification cutoffs. The cutoffs are calculated using arithmetic or geometric mean of theoretical values for identical, first-, second-, third-degree, and unrelated classes. . . 24

Table 2.3 The number of pairs and replications used in the LMMs for each relationship category. PO: parent-offspring, S: siblings, HS: half-siblings, AV: avuncular, GPC: grandparent-grandchild, FC: first cousins, GAV: grand avuncular, GGPC: great-grandparent-great-grandchild, U: unrelated. . . . 31

Table 3.1 Levene's test results of kinship coefficient estimates from READ, NgsRelate, and lcMLkin for variable relatedness degrees (first, second, and third) and SNP counts (n=5 per count). The values show the p-values from Levene's test. The values with a star represent significant p-values (<0.05). PO: parent-offspring, S: siblings, HS: half-siblings, AV: avuncular, GPC: grandparent-grandchild, FC: first cousins, GAV: grand avuncular, GGPC: great-grandparent-great-grandchild. 48

Table 3.2 Kinship coefficient estimation summary of READ, NgsRelate, and lcMLkin for parent-offspring pairs in the presence of inbreeding. The total number of pairs is n=240 for each SNP count. Sd: standard deviation. . . . 58

Table 3.3 Kinship coefficient estimation summary of READ, NgsRelate, and lcMLkin for parent-offspring pairs in the presence of inbreeding. The total number of pairs is n=240 for each SNP count. Sd: standard deviation. . . . 60

Table B.1 Variances of random effects and residuals estimated using LMM for each relationship type. 81

Table B.2 LMM summary for parent-offspring (PO) pairs. 82

Table B.3 LMM summary for sibling (S) pairs. 83

Table B.4 LMM summary for half-sibling (HS) pairs. 84

Table B.5 LMM summary for grandparent-grandchild (GPC) pairs. 85

Table B.6 LMM summary for avuncular (AV) pairs. 86

Table B.7 LMM summary for first cousin (FC) pairs. 87

Table B.8 LMM summary for grand avuncular (GAV) pairs. 88

Table B.9 LMM summary for great-grandparent-great-grandchild (GGPC) pairs. 89

Table B.10 LMM summary for unrelated (UNR) pairs. 90

Table B.11 LMM Model 1 (no interaction) using absolute mean differences for parent-offspring pairs. 91

Table B.12 LMM Model 1 (no interaction) using absolute mean differences for sibling pairs. 91

Table B.13 LMM Model 1 (no interaction) using absolute mean differences for half-sibling pairs. 92

Table B.14 LMM Model 1 (no interaction) using absolute mean differences for grandparent-grandchild pairs. 92

Table B.15 LMM Model 1 (no interaction) using absolute mean differences for avuncular pairs. 93

Table B.16 LMM Model 1 (no interaction) using absolute mean differences for first cousin pairs. 93

Table B.17 LMM Model 1 (no interaction) using absolute mean differences for great-grandparent-great-grandchild pairs. 94

Table B.18 LMM Model 1 (no interaction) using absolute mean differences for grand avuncular pairs. 94

Table B.19 LMM Model 1 (no interaction) of parent-offspring pairs in the presence of inbreeding. 95

Table B.20 LMM Model 1 (no interaction) using absolute mean differences for parent-offspring pairs in the presence of inbreeding. 95

Table B.21 LMM Model 1 (no interaction) of grandparent-grandchild pairs in the presence of inbreeding. 96

Table B.22 LMM Model 1 (no interaction) using absolute mean differences for grandparent-grandchild pairs in the presence of inbreeding. 96

Table C.1 Levene's test results of kinship coefficient estimates from READ, NgsRelate, and lcMLkin for parent-offspring and grandparent-grandchild pairs with one inbred individual and SNP counts (n=5 per count). The values show the p-values from Levene's test and the stars represent significant p-values (<0.05). 98

Table D.1 Binomial test results of kinship coefficient estimates among READ, NgsRelate, and lcMLkin for parent-offspring pairs. The binomial shows the p-values. The success rate represents the rate of the estimates above the corresponding theoretical value. 100

Table D.2 Binomial test results of kinship coefficient estimates among READ, NgsRelate, and lcMLkin for sibling pairs (n=48). The binomial shows the p-values. The success rate represents the rate of the θ estimates above the corresponding theoretical value (n=5 replicate per SNP count). 101

Table D.3 Binomial test results of kinship coefficient estimates among READ, NgsRelate, and lcMLkin for half-sibling pairs (n=48). The binomial shows the p-values. The success rate represents the rate of the θ estimates above the corresponding theoretical value (n=5 replicate per SNP count). 102

Table D.4 Binomial test results of kinship coefficient estimates among READ, NgsRelate, and lcMLkin for avuncular pairs(n=48). The binomial shows the p-values. The success rate represents the rate of the θ estimates above the corresponding theoretical value (n=5 replicate per SNP count). 103

Table D.5 Binomial test results of kinship coefficient estimates among READ, NgsRelate, and lcMLkin for grandparent-grandchild pairs(n=48). The binomial shows the p-values. The success rate represents the rate of the θ estimates above the corresponding theoretical value (n=5 replicate per SNP count). 104

Table D.6 Binomial test results of kinship coefficient estimates among READ, NgsRelate, and lcMLkin for first cousin pairs(n=48). The binomial shows the p-values. The success rate represents the rate of the θ estimates above the corresponding theoretical value (n=5 replicate per SNP count). 105

Table D.7 Binomial test results of kinship coefficient estimates among READ, NgsRelate, and lcMLkin for great-grandparent-great-grandchild pairs(n=48). The binomial shows the p-values. The success rate represents the rate of the θ estimates above the corresponding theoretical value (n=5 replicate per SNP count). 106

Table D.8 Binomial test results of kinship coefficient estimates among READ, NgsRelate, and lcMLkin for grand avuncular pairs(n=48). The binomial shows the p-values. The success rate represents the rate of the θ estimates above the corresponding theoretical value (n=5 replicate per SNP count). . 107

Table D.9 Binomial test results of kinship coefficient estimates among READ, NgsRelate, and lcMLkin for parent-offspring pairs in the presence of inbreeding (n=48). The binomial shows the p-values. The success rate represents the rate of the θ estimates above the corresponding theoretical value (n=5 replicate per SNP count). 108

Table D.10Binomial test results of kinship coefficient estimates among READ, NgsRelate, and lcMLkin for grandparent-grandchild pairs in the presence of inbreeding(n=48). The binomial shows the p-values. The success rate represents the rate of the θ estimates above the corresponding theoretical value (n=5 replicate per SNP count). 109

LIST OF FIGURES

FIGURES

Figure 3.1 Kinship coefficient estimates of simulated first-degree pairs, (A) parent-offspring, and (B) siblings. The points represent the kinship coefficient estimated by READ, NgsRelate, and lcMLkin for one pair of individuals sharing 1K, 5K, 10K, 20K, or 50K SNPs. For each SNP subset and each relationship type, the total number of simulated pairs is 240. Horizontal lines show the theoretical kinship coefficients. The boxplots, jitter-added points, and density plots show the distribution of the same sample of 240 points. 35

Figure 3.2 Kinship coefficient estimates of simulated second-degree pairs,(A) half-siblings, (B) avuncular, and (C) grandparent-grandchild. The points represent the kinship coefficient estimated by READ, NgsRelate, and lcMLkin for one pair of individuals sharing 1K, 5K, 10K, 20K, or 50K SNPs.For each SNP subset and each relationship type, the total number of simulated pairs is 240. Horizontal lines show the theoretical kinship coefficients. The boxplots, jitter-added points, and density plots show the distribution of the same sample of 240 points. 37

Figure 3.3 Kinship coefficient estimates of simulated third-degree pairs, (A) first cousins, (B) grand avuncular, and (C) great-grandparent-great-grandchild. The points represent the kinship coefficient estimated by READ, NgsRelate, and lcMLkin for one pair of third-degree related individuals sharing 1K, 5K, 10K, 20K, or 50K SNPs.For each SNP subset and each relationship type, the total number of simulated pairs is 240. Horizontal lines show the theoretical kinship coefficients. The boxplots, jitter-added points, and density plots show the distribution of the same sample of 240 points. 39

Figure 3.4 Kinship coefficient estimates of simulated unrelated pairs. The points represent the kinship coefficient estimated by READ, NgsRelate, and lcMLkin for one pair of individuals sharing 1K, 5K, 10K, 20K, or 50K SNPs.For each SNP subset, the total number of simulated pairs is 240. Horizontal lines show the theoretical kinship coefficients. The boxplots, jitter-added points, and density plots show the distribution of the same sample of 240 points. 40

Figure 3.5 The summary of linear mixed effect model with the interaction of SNP counts and genetic relatedness estimation tools for first-degree relatedness. The mean of estimated kinship coefficients for each overlapping SNP count is described as points with distinctive colors. The vertical lines show one standard error, estimated using all pairs (n=48) and replicates (n=5 per pair). The red dashed line represents the theoretical kinship coefficient value for first-degree relatedness. 42

Figure 3.6 The summary of linear mixed effect model with the interaction of SNP counts and genetic relatedness estimation tools for second-degree relatedness. The mean of estimated kinship coefficients for each overlapping SNP count is described as points with distinctive colors. The vertical lines show one standard error, estimated using all pairs (n=48) and replicates (n=5 per pair). The red dashed line represents the theoretical kinship coefficient value for second-degree relatedness. . . . 43

Figure 3.7 The summary of linear mixed effect model with the interaction of SNP counts and genetic relatedness estimation tools for third-degree relatedness. The mean of estimated kinship coefficients for each overlapping SNP count is described as points with distinctive colors. The vertical lines show one standard error, estimated using all pairs (n=48) and replicates (n=5 per pair). The red dashed line represents the theoretical kinship coefficient value for third-degree relatedness. 44

Figure 3.8 The summary of linear mixed effect model with the interaction of SNP counts and genetic relatedness estimation tools for unrelated pairs. The mean of estimated kinship coefficients for each overlapping SNP count is described as points with distinctive colors. The vertical lines show one standard error, estimated using all pairs (n=48) and replicates (n=5 per pair). The red dashed line represents the theoretical kinship coefficient value for unrelated pairs. 45

Figure 3.9 The relative frequency of pairs assigned to first-, second-, and third-degree related and unrelated categories by READ, NgsRelate, and lcMLkin. The kinship coefficient estimates from these tools were classified using the arithmetic mean of theoretical kinship coefficients. Colors refer to the assigned relatedness degree. The frequencies of pairs assigned to each category are indicated as percentages inside the bars (only for categories with frequency >5%). 52

Figure 3.10 FPR, FNR, TPR, and TNR for READ, NgsRelate, and lcMLkin using arithmetic mean to classify kinship coefficient estimates of first-, second-, and third-degree related and unrelated pairs. The colors represent the count of SNPs shared between individuals. 53

Figure 3.11 True positive rate (TPR, or sensitivity) and false positive rates (FPR, i.e., 1 – specificity) for READ, NgsRelate, and lcMLkin using two different classification methods, arithmetic and geometric mean. Colors refer to the relatedness degree and line types represent the classification method and classification metric. 53

Figure 3.12 Pairwise F – score (F_1) of READ, NgsRelate, and lcMLkin for first-, second-, and third-degree related and unrelated pairs with 1K, 5K, 10K, 20K, and 50K shared SNPs. Colors refer to the degree of relatedness. 54

Figure 3.13 Inbreeding pedigree simulation. Grandparents of the offspring are simulated as siblings and parents are simulated as first cousins. The circles and squares represent the female and male individuals. Diamonds show an individual with unspecified sex. Double horizontal lines connecting two simulated individuals (parents) represent mating between relatives. 55

Figure 3.14 Kinship coefficient estimates for parent-offspring pairs in the presence of inbreeding by READ, NgsRelate, and lcMLkin. Horizontal lines show the theoretical kinship coefficients. The boxplots, jitter-added points, and density plots show the distribution of the same sample of 240 points. θ from NgsRelate was calculated ignoring the inbreeding-related Jacquard coefficients: $\theta = J_7/2 + J_8/4$. 57

Figure 3.15 Kinship coefficient estimates for grandparent-grandchild pairs in the presence of inbreeding by READ, NgsRelate, and lcMLkin. Horizontal lines show the theoretical kinship coefficients. The boxplots, jitter-added points, and density plots show the distribution of the same sample of 240 points. θ from NgsRelate was calculated ignoring the inbreeding-related Jacquard coefficients: $\theta = J_7/2 + J_8/4$. 59

LIST OF ABBREVIATIONS

IBD	Identical By Descent
IBS	Identical By State
NGS	Next Generation Sequencing
SNP	Single Nucleotide Polymorphism
WGS	Whole Genome Sequencing
STR	Short Tandem Repeats
aDNA	Ancient DNA
bp	Base pair
Mbps	Megabase pairs
PMD	Post-Mortem Damage
TSI	Tuscany
GL	Genotype Likelihood
MAF	Minor Allele Frequency
AAF	Alternative Allele Frequency
HMM	Hidden Markov Model
LMM	Linear Mixed Effect Model
TPR	True Positive Rate
FPR	False Positive Rate
TNR	True Negative Rate
FNR	False Negative Rate
F1	F-score

CHAPTER 1

INTRODUCTION

Understanding the traditions of prehistoric and historic groups in their social and cultural settings in the light of their kin relations has been first underlined by an anthropologist, Lewis Henry Morgan, in the mid-nineteenth century (Morgan, 1871). This initial attempt paved the way for more methodological investigations of kinship among anthropologists. The genealogical method developed by W.H.R. Rivers in the 19th century (Rivers, 2012) allowed studying and interpreting families and their ancestor-descendant relationships as a part of sociocultural organizations. In an archeological context, besides tracing demographic patterns between groups, kin relations within particular sites and periods also became an apparent clue to answer how ancient communities were constructed. By analyzing skeletal and material remains from burial sites, anthropologists and archeologists could hypothesize the social and biological relationships within groups. However, even though these findings could give insights into possible relatedness between individual pairs in specific archeological contexts, the actual truth could only be gathered through genetic analysis, which became widely available only over the last decade.

1.1 Genetic Relatedness

Genetic relatedness between a pair can be defined as having one or more recent common ancestors relative to other individuals in a group. This makes the genomes of the two individuals look more similar to each other compared to the rest. To define this similarity, shared segments transmitted from a recent common ancestor are said to be identical by descent (IBD) (Goudet et al., 2018; Speed & Balding, 2015; Weir

et al., 2006). The IBD patterns for two diploid individuals with four alleles at any site of their genomes can be classified into 15 unique states (Brustad et al., 2021; Weir et al., 2006). Ignoring the parental origin of the alleles, the relative frequencies of each remaining nine distinct states are described as condensed Jacquard coefficients (J_1, J_2,..., J_9) (Jacquard, 1972). Each state's frequencies depend on the degree of relatedness between individuals and the possible inbreeding in their descendants. If the individuals are not inbred, Jacquard coefficients of six different IBD states become zero. The remaining three positive coefficients (J_9, J_8, and J_7) are called Cotterman coefficients and are mostly denoted as k_0, k_1, and k_2, respectively (Cotterman, 1941). The k_m quantity represents the probability that two diploid individuals share m alleles that are IBD.

Using these Jacquard coefficients, IBD-based summary statistics are used to estimate relatedness between individuals. The most commonly used one is the kinship coefficient (θ), i.e., the probability of randomly choosing two alleles from two individuals being IBD (Brustad et al., 2021; Goudet et al., 2018; Speed & Balding, 2015; Vai et al., 2020; Weir et al., 2006). θ can be defined in terms of Jacquard coefficients as:

$$J_1 + 0.5(J_3 + J_5 + J_7) + 0.25J_8 \tag{1.1}$$

Alternatively, it can be reduced to k_1/4 + k_2/2 if the individuals are outbred (i.e., when J_1, J_3, J_5 are null). The more related two individuals are, the higher the kinship coefficient is. For instance, the expected fraction of the genomes of unrelated individuals sharing 1 or 2 alleles of IBD is 0 (k_1 = 0 and k_2 = 0) since they do not carry similar DNA sequences revealing inheritance from a recent common ancestor. Hence, their kinship coefficient becomes zero. For a sibling pair, this theoretical expectation equals 0.25 with varying IBD state probabilities: k_1 = 0.25, k_2 = 0.5, and k_0 = 0.25. Although Cotterman coefficients can differ between distinct relationship types, the kinship coefficient stays the same for those with the same relatedness degree. Parent-offspring and sibling pairs, for example, have a one-in-four probability of randomly sampling alleles IBD (θ = 0.25), but their genome-wide pattern of allele sharing is not the same. Half of the offspring's genome comes from a mother, and the other half comes from a father. Thus, the expected k_1 for parent-offspring relatives is 1. This apparent difference makes using the Cotterman coefficients rather than the

kinship coefficients obligatory to differentiate between parent-offspring and sibling pairs (Weir et al., 2006).

Table 1.1: Cotterman and kinship coefficients for relationship types in the absence of inbreeding ($\theta = k_1/4 + k_2/2$) (Weir et al., 2006).

Relationship	**Degree**	**k0**	**k1**	**k2**	θ
Identical Twins	Identical Twins	0.00	0.00	1.00	0.50
Parent-offspring	First	0.00	1.00	0.00	0.25
Siblings	First	0.25	0.50	0.25	0.25
Grandparent- grandchild/ Half-siblings/ Avuncular	Second	0.50	0.50	0.00	0.125
First cousins/ Grand avuncular/ Great-grandparent- grandchild	Third	0.75	0.25	0.00	0.0625
Unrelated	Unrelated	1.00	0.00	0.00	0.00

The kinship coefficient can also be expressed as the number of parent-offspring lineages separating two individuals (Seguin-Orlando, 2021; Weir et al., 2006). Knowing the position of individuals in the pedigree allows inference of relatedness between them by directly calculating the probability of alleles IBD. However, pedigree information is not always available, so the relationships cannot always be deduced this way. In that case, the IBD properties of the individuals inferred from genetic data can be used to resolve genetic relatedness. However, for a given pair of individuals, the IBD status for each locus cannot be observed directly by looking at their DNA sequences. Instead, the identity-by-state (IBS) status, whether the sampled alleles seem identical or not, can be used as a proxy for IBD.

Today it is possible to measure the allelic states of loci across the genome to and determine IBS using various approaches; these include genotyping microsatellite (STR) loci, genotyping single nucleotide polymorphisms (SNPs) using microarrays, or genotyping SNPs and other polymorphisms using shotgun whole genome sequencing (WGS) or reduced representation sequencing (e.g., exomes or RAD tags). Using these methods, genotyping data can be produced in relatively short periods of time and high quantities; the resulting data can be used to determine the IBS status of alleles across the genome and used to determine the relationship between samples. Before the availability of high-density SNP data, microsatellites, i.e., repetitive DNA segments mostly known as short tandem repeats (STR), were the most com-

monly studied markers in the mid-1990s due to their vast variable occurrence between genomes with multiple alleles at one locus (Seguin-Orlando, 2021; Weir et al., 2006). Although STRs are still generally preferred in the forensic identification of possible relatives, their usage in much more complex relatedness inference has few advantages over SNPs due to their high mutation rate and consequent risk of homoplasy, i.e., parallel mutations that create the same alleles in unrelated individuals. The relatively stable genetic nature of SNPs and the ability to profile them in vast amounts with next-generation sequencing technologies have thus recently led to the development of computational tools and methods using SNPs instead of STRs for population genetic and kinship inference analysis.

1.2 Ancient DNA

The emergence of second-generation high-throughput sequencing in the late 2000s enabled the massive sequencing of DNA samples in short fragments in a parallelized manner and, thus, the detection and recovery of rare variants in large numbers. This revolution also accelerated ancient DNA (aDNA) research, which had, until that point, primarily relied on targeted short DNA fragment sequencing from mitochondrial genomes. Ancient DNA is the preserved DNA from archeological remains and samples, bringing the past into the present and enabling studies on organisms' lives and environments. Its extraction and sequencing are not as easy as modern samples, as the quality and quantity of aDNA depend on how much time has passed since the organism's death and its preservation conditions in this duration. Fortunately, numerous studies showed that aDNA could be substantially extracted from variable samples, such as bones, teeth, calculus, and sediments (Rizzi et al., 2012; Seguin-Orlando, 2021). Notably, two million-year-old ancient DNA (the oldest aDNA record) has recently been obtained from frozen soil in North Greenland (Kjær et al., 2022).

One of the well-studied characteristics of aDNA is fragmentation. After the death of the organisms, the DNA backbone is broken due to a hydrolytic degradation reaction (Orlando et al., 2021). This reaction starts the fragmentation process of ancient DNA, and the fragmentation continues due to additional environmental stress. If the archeological samples are located in a hot and humid environment, the remaining DNA is

broken into pieces faster. Other potential factors that contribute to degradation are pH concentration and microbial organism activities. This degradation process results in recovered aDNA fragment lengths typically shorter than 100 bp. Recently, DNA recovered from a 4000-year-old hair sample from Sudan has been found to have an average read length of 30bp (Wang et al., 2022).

Another unique property of ancient DNA that differentiates it from modern DNA is post-mortem damage (PMD). This type of damage emerges due to the deamination of cytosines (C), especially at a higher rate at the end of the DNA fragments (Peyrégne & Prüfer, 2020; Skoglund et al., 2014). The deamination converts the cytosines to uracil, but they are misread as thymines (T) during sequencing. Hence, the sequencing data of ancient samples contains C-to-T substitutions. Like C-to-T, substitution from guanine (G) to adenine (A) also occurs in the complementary DNA strand.

Most of the original DNA of the archeological samples is lost due to ancient DNA decay. However, remaining endogenous and external DNA from variable sources, such as microbes or modern humans, could be recovered (Orlando et al., 2021; Peyrégne & Prüfer, 2020; Rizzi et al., 2012). Microbial activities in the environment and inattentive handling of the archeological samples during the excavation and lab analysis can introduce contamination. If effective measures are not taken before and during wetlab stages, trying to eliminate the contamination makes ancient DNA analysis even more challenging.

These obstacles to the extraction and analysis of aDNA result in decreased quantity and quality, reflected in the low coverage of the sequencing data (generally less than 1x) or even the inability to obtain it. For most laboratories, even though the archeological samples are well-preserved or endogenous DNA content detected in the preliminary experiments is enough to study relevant questions, the high cost and the time needed to sequence samples to increase genomic depth is not feasible. The significant drawback of using low-quality sequence data is that genotyping diploid SNP loci is not possible for most sites when the coverage is <1x. A popular way to solve this is to apply pseudo-haploidization, i.e., randomly selecting only one sequencing read for each predefined SNP site (Orlando et al., 2021). This way, genotype data is artificially produced as haploid; therefore, each locus is represented by only one al-

lele. This practice is commonly used in population genetic and biological relatedness analyses.

1.3 Kinship Estimation Methods

Using NGS data and a number of bioinformatics methods and software, the efforts have been focused on the assessment of the genetic relatedness between individuals using pairwise IBD patterns. For that purpose, various tools such as PLINK (Purcell et al., 2007), GERMLINE (Gusev et al., 2009), and KING (Manichaikul et al., 2010) have been developed to identify IBD segments of high-quality genomes. Other methods, like ERSA (Huff et al., 2011), directly infer relatedness degrees of pairs with the prerequisite of IBD previously defined segments. However, they are not suitable for sparse NGS data. Recently, alternative methods have been proposed to deal with low-coverage genomic data. These methods can be classified into two broad groups : (i) method-of-moment based ; (ii) maximum likelihood (ML) based.

1.3.1 Method-of-moment Based Approaches

Method-of-moment-based approaches are able to estimate genetic relatedness directly by comparing the parameters for observed data, such as the mean or variance, with theoretically expected parameters under a statistical model (Speed & Balding, 2015). Unlike ML-based methods, they are computationally straightforward to apply, and therefore, they can work efficiently with large datasets.

1.3.1.1 READ

A unique example of this approach is READ (Kuhn et al., 2018). It uses the pairwise mismatches between a pair of pseudo-haploid genomes to uncover familial relationships up to second-degree relatedness and works with coverages as low as 0.1x. It operates by averaging out the genome-wide proportions of mismatching alleles in non-overlapping windows of 1 Mbps, the pairwise mismatch rate for a pair (P_0). To eliminate population-specific diversities, this value is then normalized using a P_0 of

an unrelated pair representing a baseline mismatch rate for the population under study. The normalization value can be calculated using the mean, median, or maximum of P_0 of all samples expecting to represent a P_0 of an average unrelated pair. It is also possible to apply a user-defined P0 value to normalize pairwise mismatches.

1.3.1.2 TKGWV2

Another example of method-of-moment-based approaches is TKGWV2 (Fernandes et al., 2021). Like READ, it has been developed to work with low-coverage pseudo-haploid genomes (0.02x) and detect genetic relatedness up to a second-degree level. However, its working principle is quite different as background allele frequencies are incorporated into the formula calculating pairwise genetic relatedness using overlapping SNPs, denoted as R_{xy} (Oliehoek et al., 2006; Queller & Goodnight, 1989). TKGWV2 does not require a normalization step using an unrelated sample, so it provides an opportunity to identify the relationship of a single pair of ancient individuals.

1.3.2 Maximum likelihood-based methods

ML methods use a likelihood function to estimate the probability of IBD given IBS. This likelihood function describes the probability of Jacquard coefficients or Cotterman coefficients corresponding to a particular allelic state (see Section 1.1 and Table 1.1) given background allele frequencies for a pair of individuals. The ML-based approaches estimate these coefficients maximizing the likelihood of observed genotype data.

1.3.2.1 NgsRelate and lcMLkin

ML methods such as NgsRelate (Korneliussen & Moltke, 2015) and lcMLkin (Lipatov et al., 2015) estimate the Cotterman coefficients (denoted as R) maximizing the likelihood function using the Expectation Maximization algorithm. Moreover, instead of called genotypes, both methods incorporate genotype likelihoods summed for each possible genotype for each locus per individual while estimating R. In this way, they

consider uncertainties in called genotypes as shallow sequencing, post-mortem damage, contamination, or errors introduced in sequencing or mapping can create biases. This makes NgsRelate and lcMLkin suitable to infer lower-degree relatedness down to the fifth degree. They can work with relatively low coverage NGS data, 0.1-4x. Besides Cotterman coefficients, NgsRelate version 2 (Hanghøj et al., 2019) estimates Jacquard coefficients reflecting 9 allelic IBS states (see Section 1.1) to be able to identify relationships in the presence of inbreeding. Both tools apply a minor allele frequency filter for the shared SNPs as a default, only keeping SNPs with MAF > 0.05; this allows for determining IBD probabilities based on more informative alleles represented more in a population. While NgsRelate can work with BAM or VCF files, including genotype likelihoods, lcMLkin can only take VCF files as input.

1.3.3 KIN

Recently, a more sophisticated Hidden Markov Model-based method, KIN (Popli et al., 2023), has been introduced. The properties of KIN that distinguish it from the above-mentioned tools are (i) the ability to differentiate between parent-offspring and sibling pairs; (ii) the incorporation of runs of homozygosity (ROH) to relatedness estimation; (iii) contamination correction. Also, unlike most other methods, and similar to READ, it does not depend on population allele frequencies. KIN further estimates pairwise differences in genomic windows directly from read data (BAM files) with a minimum 0.05x depth of coverage. Additionally, it incorporates the probability of window-based ROH tracts in each individual estimated by an $ROH - HMM$ model while fitting an IBD sharing pattern of pairs to the predefined relatedness models (unrelated, $5th$ degree, $4th$ degree, $3rd$ degree, $2nd$ degree, $1st$ degree and identical) provided by the $KIN - HMM$ model. Then, KIN assigns the most likely relationship degree for a pair with the highest likelihood.

1.3.4 ancIBD

The tools used to infer genetic relatedness between ancient samples are not limited to the software mentioned so far. Lately, an IBD segment detection-based kinship es-

timation tool, ancIBD, has been proposed (https://pypi.org/project/ancIBD/). Instead of estimating the probability of sharing 0, 1, or 2 alleles IBD (k0, k1, k2) for each locus like NgsRelate and lcMLkin, ancIBD estimates the length and number of IBD segments (or haplotypes) shared between two individuals. For the detection of IBD segments, the algorithm relies on HMM which takes input as genotype likelihoods phased and imputed by $GLIMPSE$ software (Rubinacci et al., 2021) using haplotypes from a modern reference panel (1000 Genomes Project) (Auton et al., 2015). It categorizes the estimates into four groups : >8, >12, >16, and >20 cM in length. Based on these estimates, ancIBD can classify genetic relatedness up to the sixth or seventh degree using genomes with a minimum 0.25x depth of coverage (Villalba-Mouco et al., 2023).

1.3.5 Genotype Calls and Genotype Likelihoods

Kinship estimation tools from ancient DNA also vary in terms of the type of input data they use: (i) genotype calls (either diploid or pseudohaploid) ; (ii) genotype likelihoods. Calling genotypes for aDNA can involve biases because of the properties of the ancient genome data, such as post-mortem damage, contamination, and missingness. For this reason, genotyping is usually limited to ascertained SNPs. Also, shallow sequencing can increase the uncertainties in diploid genotype calls. For example, heterozygous positions could not be captured accurately if the quantity of the ancient data is low, e.g., <2x. The most accepted approach to relatively eliminate these difficulties is producing pseudo-haploid genomes by randomly selecting one allele per sample and SNP locus. Thereby, each site is represented with only one allele, and heterozygous sites are neglected. This also allows equally treating all genomes in a sample irrespective of the differences in their coverage. The performance of some of the kinship estimators mentioned, such as READ and TKGWV2, relies on pseudo-haploid calls.

A distinct way to overcome genotype uncertainties is estimating genotype likelihoods, which are the probability of read data given genotype values for each site for each individual (Korneliussen & Moltke, 2015; Nielsen et al., 2011). It would be expected that genotype likelihoods are most helpful when coverages are modest, e.g., between

1x and 10x. The genetic relatedness estimation tools such as NgsRelate and lcMLkin take advantage of using genotype likelihoods instead of genotype calls while estimating the kinship coefficient.

1.3.6 Limitations of the Existing Methods for Ancient DNA

Despite the abundance of approaches in genetic relatedness estimation for ancient genomes, in practice, their ability to estimate familial relationships can be limited. One problem involves the requirement of background allele frequencies (for NgsRelate, lcMLkin, and TKGWV2). It is essential to use the most relevant and contemporary population as a reference while performing kinship analysis with such methods. One of the reasons for this potential problem is that the population used to estimate allele frequencies may not reflect the allele frequencies of the original population that was sampled due to genetic drift or population structure.

Meanwhile, when applying methods using the sample rather than allele frequencies for normalization (e.g., READ or KIN), the number of samples sequenced may not be enough to estimate background mismatch rates for unrelated pairs. The problem is also exacerbated when the sample is not genetically homogeneous but includes distinct groups.

Another limitation is that the sparse nature of ancient DNA data induces uncertainties in genotype calls and lowers the number of shared SNPs between genomes, which in turn causes genetic relatedness to be randomly over or underestimated. It is also inevitable to obtain wrong genotype calls due to PMD, contamination, and sequencing or mapping errors from ancient samples, which can result in biased estimations. For instance, low levels of residual PMD in the data would lead to underestimation of the kinship coefficients. Only KIN produces a solution for the errors introduced by contamination while identifying kin relationships.

More generally, biological processes such as inbreeding, admixture, or population structure will also affect the genetic kinship analysis. These processes can create genetically more similar or dissimilar pair of genomes. Most of the available strategies to infer relatedness assumes that the studied population is outbred and homogenous

(under Hardy-Weinberg equilibrium). However, this is not always the case for ancient samples. NgsRelate (version 2) and KIN have been designed to account for inbreeding, according to their own publications. The other methods, however, would be expected to overestimate the direct relationship degree under inbreeding (e.g., might infer first-degree relationships for second-degree relatives if the parents are under consanguinity). Meanwhile, there exist methods for estimating kinships in populations with admixed ancestry, such as REAP (Thornton et al., 2012) and RelateAdmix (Moltke & Albrechtsen, 2014), although these have not been tested using low coverage data. Recently, NGSremix (Nøhr et al., 2021) have been proposed to quantify genetic relatedness between admixed individuals using genotype likelihoods unlike REAP and RelateAdmix. None of these methods have been developed to work with the samples from different subpopulations. Software such as SEEKIN (Dou et al., 2017) and KING-robust (Manichaikul et al., 2010) are claimed to allow population structure for kinship estimation. However, only SEEKIN can deal with sparse sequencing data as low as 0.15x depth of coverage.

1.3.7 Kinship Studies With Ancient Genomes

Beyond doubt, biological kinship in the prehistoric or historic societies has become one of the powerful concepts to address the social structures of ancient populations. Revealing familial links can give insights into the sociocultural repertoire of the populations, such as inheritance systems, social status, and migration patterns. The recent approach of combining multilevel evidence from different disciplines such as anthropology, archaeology, and genetics paves the way for the identification of kinship practices of ancient populations. For instance, familial relationships were inferred for Late Neolithic to Bronze Age populations of Central, Southwestern and Southeastern Europe ($\sim$2880-1300 BCE) in the recent studies (Furtwängler et al., 2020; Mittnik et al., 2019; Schroeder et al., 2019; Villalba-Mouco et al., 2023; Žegarac et al., 2021) using different kinds of genetic relatedness estimation tools such as READ, lcMLkin, KING, or IBSrelate (Waples et al., 2019). In these studies, they analysed the genetic kinship between high-quality genomes from more than 60 (Furtwängler et al., 2020; Mittnik et al., 2019; Villalba-Mouco et al., 2023) or 10 ancient samples (Schroeder et al., 2019; Žegarac et al., 2021). These studies reported first-, second-, and third-

degree relationships established predominantly via males. Also, they point out that female offspring were found less compared to male offspring (Mittnik et al. 2019; Zegarac et. al. 2021), and the shared parents of the half-siblings were frequently males (Furtwängler et al. 2020; Villalba-Mouco et al 2022). Because of these observations, it was hypothesized that these societies involved female exogamy and patrilocality traditions. This line of evidence was also supported by the analysis of strontium isotopes, with females being found more often to be non-local than males in these sites and high mtDNA haplogroup diversity indicating possible incorporation of females with diverse background to these sites (Mittnik et al., 2019; Žegarac et al., 2021). The existence of shared grave goods found in the burials of closely related individuals but not in the ones of unrelated pairs (Mittnik et al., 2019) and the positioning of the individuals based on their biological relationships in the graves (Schroeder et al., 2019; Žegarac et al., 2021) indicate that familial links shaped the social structure of the ancient populations.

Although the picture from Neolithic to Bronze Age Europe appears relatively homogeneous, with a clear bias towards family-based structure and patrilocality, other work has shown that this was not necessarily a universal pattern. The evidence comes from two studies (Altınışık et al., 2022; Yaka et al., 2021) that investigated kinship practices in Neolithic sites of Anatolia, namely Boncuklu, Aşıklı, Çayönü, Barcın, and Çatalhöyük (c. 8500 - 6000 BC). The genomic coverage of ancient samples was relatively low compared to the aforementioned studies (< 0.3x). For this reason, they used the minority of the genomes with sufficient coverage to determine genetic kin using READ, lcMLkin, and NgsRelate (and TKGWV2 for Altınışık et al.,2022) in parallel. The relationships between samples were tested using READ, lcMLkin, and NgsRelate in parallel. Since the genome coverages are not suitable enough to reliably estimate background allele frequencies for the analysis with NgsRelate and lcMLkin, they estimated the allele frequencies of external Holocene populations of nearby regions. In contrast to patrilineal lineages observed in Neolithic to Bronze Age settlements of Europe, in the early Neolithic sites of Boncuklu, Aşıklı, and Çayönü, populations did not show any clear sign for female exogamy as related adult females were found buried together. Further, while the frequencies of genetic relatives identified among co-buried individuals (buried inside the same buildings) in these sites

were relatively high, in the late Neolithic settlements of Barcın and Çatalhöyük, individuals co-buried in the same buildings (mostly subadults) were frequently unrelated. These results point out that patrilocality is not a standard sociocultural practise to be followed in ancient populations. Also, this indicates that the social structure of the populations is not always shaped by familial links.

1.3.8 Research Objectives

The main goal of this study is to systematically compare the performances of the three most commonly used genetic relatedness estimation tools for ancient genomes: READ, NgsRelate, and lcMLkin. While READ is preferably applied due to its easy use requiring only genotype information in most of the aDNA studies, NgsRelate and lcMLkin are mostly used to further reveal relationships up to the fifth degree, especially when a relevant population is available for the allele frequency estimation. Recently, a study benchmarked these approaches using ancient and modern genomes with varying coverages (2.12x - 0.02x) (Marsh et al., 2023). However, the accuracies and sensitivities of READ, NgsRelate, and lcMLkin with low-coverage ancient samples comprising lower degree of relatedness (second or third-degree), limited number of SNPs, or inbreeding have not been systematically evaluated.

To explore their performances, we simulated ancient genomes showing varying degrees of relatedness, creating first-, second-, and third-degree related simulated pairs. Also, we aimed to investigate their dependency on the number of shared SNPs (1K, 5K, 10K, 20K, and 50K) to precisely estimate the relationships between individuals. Furthermore, we simulated parent-offspring and grandparent-grandchild pairs with one inbred individual whose parents are simulated as third-degree related (i.e., first cousins) to determine what extent these genetic relatedness estimation tools are affected by inbreeding.

1.4 Structure of the Thesis

This thesis consists of Chapter 1 (p. 1) Chapter 2 (p. 15) Chapter 3 (p. 33) Chapter 4, and (p. 63).

CHAPTER 2

MATERIAL AND METHODS

2.1 Pedigree Simulations

The goal of this study is to determine how common kinship estimation tools perform on ultra-low coverage ancient genome data. To assess this most effectively, we simulated ancient genome data representing pairs of individuals with known relationships. Specifically, we applied pedigree simulation software $PED - SIM$ ($v1.3$) (Caballero et al., 2019) to produce genotypes from pedigrees of various relationship degrees and types separately, including first-, second-, and third-degree relatedness without inbreeding, as well as first-degree and second-degree relatedness with first cousin mating. $PED - SIM$ creates individual genotypes based on user-specified pedigrees starting from founder genotypes given as input and using a recombination map.

We created founder genotype data from scratch as follows: We chose SNPs with minor allele frequencies (MAF) equal to or higher than 0.01 of the modern-day Tuscany (TSI) sample (n=112) from the 1000 Genomes Project (Auton et al., 2015), and used for the founder data generation. For the resulting 8,677,101 biallelic SNPs, we recorded the reference or alternative alleles at each SNP position as observed in the TSI dataset and calculated the alternative allele frequency (AAF) per SNP in TSI. We then created the genotypes of each founder by randomly choosing, for each SNP independently, the alternative or reference allele with probability AAF and 1-AAF, respectively, and repeating this twice to create a diploid genotype per founder. Note that this method of creating diploid genotypes eliminates any background relatedness among founders as well as any homozygosity tracts within founder genomes. We re-

peated the creation of founder data 12 times (or runs), each time producing different sets of founders.

We generated 120 unrelated founders (10 for each run, each with n=12) used for first-degree and 240 unrelated founders (20 for each run, each with n=12) for second- and third-degree pedigree simulations. We simulated genomes with X chromosomes to reflect the possible variation in kinship coefficient estimations generated by recombination differences between males and females. A sex-specific genetic map (Bherer et al., 2017) with linear interpolation and crossover interference model (Housworth & Stahl, 2003) was used to simulate genotype data of related individuals with autosomal and X chromosomes. We randomly assign the sex of the descendants in the $PED - SIM$ parameter file (def file). In this way, we simulated distinct pedigrees with the same structure but varied in the sex of the individuals.

We simulated n=160 pedigrees representing first-degree relatedness, n=640 second-degree pedigrees, and n=1200 pedigrees composed of third-degree related pairs (total n=2,000). The founders of each pedigree and simulated individuals from distinct pedigrees were considered unrelated. Alternatively, first-degree and second-degree pedigrees (parent-offspring and grandparent-grandchild relationships) were simulated in the presence of first-cousin mating (i.e., the parents of an offspring or a grandchild are first cousins, respectively). The latter was produced to explore the possible effect of inbreeding on kinship coefficient estimation. In all the analyses, we used the same number of pairs from each relationship type (n=48).

Igor Mapelli conducted the founder genotype data production and pedigree simulations using $PED - SIM$.

Table 2.1: Description of simulated pairs. The number of pedigrees represents the number of distinct pedigrees in each run.

Relationship	**Degree**	**Number of Pedigrees**	**Number of Pairs**
Parent-offsprings	First	2	48
Siblings	First	3	48
Half-siblings	Second	6	48
Grandparent- grandchild	Second	4	48
Avuncular	Second	8	48
First Cousins	Third	10	48
Great Grandparent- Great Grandchild	Third	8	48
Grand Avuncular	Third	16	48
Unrelated	Unrelated	3	29706
Parent-offsprings (inbreeding)	First	8	48
Grandparent - grandchild (inbreeding)	Second	4	48

2.2 Ancient Sequence Simulations

To evaluate the performances of READ (Kuhn et al., 2018), NgsRelate (Korneliussen & Moltke, 2015), and lcMLkin (Lipatov et al., 2015) in genetic relatedness estimation, we used forty-eight unique simulated pairs from each relationship type: first-degree, second-degree, and third-degree relatives (number of individuals n = 72 for parent-offspring, grandparent-grandchild, and great-grandparent - great-grandchild pairs, and n= 96 for siblings, half-siblings, first cousins, avuncular, and grand avuncular pairs) (Note that the choice of 72 vs. 96 was arbitrary). A pair of individuals of distinct pedigrees within or between runs are considered unrelated. All possible sex combinations of a simulated pair specific to each type of relationship were included in the analysis.

Our aim is to examine how kinship estimation is affected by low SNP numbers, and we want to speed up downstream simulations. To achieve this, we used an in-house bash script to downsample the genome-wide autosomal genotype data of all pairs of simulated individuals to the same randomly chosen subset of autosomal 200,000 biallelic SNPs.

We next used the $gargammel$ software (Renaud et al., 2017) to simulate ancient

DNA-like read data, as we aim to compare and contrast three genetic relatedness estimation tools that are commonly and especially used for ancient DNA. The ancient read simulator cuts a given FASTA file into variable lengths, which reflect the distribution of actual ancient read lengths, and adds post-mortem DNA damage (PMD). Then, it adds Illumina adapters to the ends of the reads. Finally, sequencing errors and quality scores are introduced, producing ancient-looking FASTQ files. To generate input FASTA files for $gargammel$, for each individual separately (two files for each individual), at each SNP position, we inserted alternative alleles according to their genotype into the human reference genome (GRCh37) via $vcftools\ consensus$ (Danecek et al., 2011). We then cut the FASTA files into 100 bp sequence intervals surrounding each SNP (50 bp on each side) using $bedtools\ getfasta$ (Quinlan & Hall, 2010). We did this because the most time-consuming step of the dataset preparation is mapping a vast number of genomes to the reference genome; by limiting the number of reads to those surrounding the variable sites, we could significantly reduce the computation time. The read size distribution file, another input file for $gargammel$, was generated using the $rnorm$ function in R (R Core Team, 2022) with a mean of 66.2 bps and a median of 61 bps, ranging between 35 bps and 119 bps. We specified the deamination patterns (-$damage$ 0.024, 0.36, 0.009, 0.55) using Briggs model parameters (Briggs et al., 2007). We set the final depth of coverage of the samples as 5x. Given these inputs, we finally generated simulated ancient read data with 5x depth of coverage without any present-day human or microbial contamination.

2.3 Preprocessing of Simulated Ancient Genomes

We processed the $gargammel$-simulated read data following the same procedure as applied to ancient genome sequencing libraries in our group and other research teams (e.g., Altınışık et al., 2022; Koptekin et al., 2023; Yaka et al., 2021) Firstly, we removed the adapters from the simulated ancient reads and then merged the paired-end reads only if they had at least 11 base pairs (bps) length overlap using $AdapterRemoval$ $v2.3.1$ (Schubert et al., 2016). Secondly, the generated single-end ancient reads were mapped to a human reference genome (hs37d5) using BWA software (version 0.7.15) (Li & Durbin, 2009) with the "-aln" option, and parameters are set

to "-l 16500", "-n0.01" and "-o 2". We eliminated the reads with a minimum of 10% mismatches to the human reference genome. Finally, the remaining reads were trimmed ten bps from both ends to remove C-to-T substitutions (or reciprocally G-to-A) resulting from post-mortem damage of aDNA using $bamUtil$ software with the "$trimBAM$" option ((https://github.com/statgen/bamUtil).

Kıvılcım Başak Vural conducted the preprocessing, including mapping and trimming.

2.4 Genotyping and Downsampling

After the simulation of ancient reads and mapping back to the human reference genome, we randomly downsampled the BAM files of all simulated individuals from 5x to 1x coverage using $Picard\ Tools\ DownsampleSam$ (Picard Toolkit, 2019). Because our goal is to study the performance of the kinship coefficient estimation (θ) by READ, NgsRelate, and lcMLkin on low-depth ancient data, most of our analyses involve subsamples of the 1x data (only 1 read per SNP).

We next performed pseudo-haploid genotyping from simulated ancient genomes with 1x depth of coverage. Pseudo-haploidization is a regular step in most aDNA genome studies (see Section 1.3.5). Pseudo-haploidization was done using $SAMtools\ v.1.9$ $mpileup$ (Li et al., 2009) and $PileupCaller\ v1.4.0.5$ with the "–$randomHaploid$" parameter (https://github.com/stschiff/sequenceTools). First, to generate text pileup files for all BAM files, we used the random subset of 200,000 autosomal SNPs that we had selected earlier (see Section 2.1). Mapping and base quality filters were set to Phred score >30 in $SAMtools\ v.1.9\ mpileup$. Second, the output pileup files were given as input to $pileupCaller$ software to produce pseudo-haploid genotype data by randomly sampling one read and recording its allele at each SNP. Third, the output files were converted to $packedped$ format using $ADMIXTOOLS\ convertf$ package (Patterson et al., 2012) with parameter "-p" and then to transpose ped/fam format using $PLINK\ v1.9$ (Purcell et al., 2007). Last, we retained only non-missing genotype calls for each pair of individuals using PLINK v1.9 with the option "-$geno$ 0". This reduced the number of SNPs from 200,000 to an average of 77,000 for 1x depth of coverage. Missing genotype calls in low-coverage ancient genomes led to a

considerable decrease in the number of SNPs.

To explore the lower limits of using ancient genomes for genetic relatedness estimation, we randomly took subsets of 1,000, 5,000, 10,000, 20,000, and 50,000 SNPs shared between each simulated pair. This randomized downsampling was repeated five times for each subset. This allowed us to study how much kinship coefficient estimates vary depending on the set of variants used for the analysis. We note that the terms, repeats or runs, used for the downstream analysis refer to this repeated downsampling (n=5).

2.5 Genetic Relatedness Estimation

2.5.1 READ

The first software we investigated was READ, a non-parametric genetic relatedness estimation algorithm. READ compares pseudo-haploid genotypes between pairs and calculates the proportion of mismatch positions, i.e., the pairwise mismatch rate ($P0$), in non-overlapping windows of 1 Mbps. READ then calculates the genome-wide average $P0$ per pair and normalizes this using a $P0$ value corresponding to an average unrelated pair. This can be either the mean, maximum, or median (default) of all $P0$ values in a sample, assuming the average pair is unrelated, or it may be a user-specified value. Note that READ does not use population allele frequencies in $P0$ estimation.

We ran READ with pseudo-haploid genotype data of the simulated individuals pairwisely using default parameters (see Section 2.4). We combined the individual pair results from READ and prepared one dataset for each relationship type. Further, we removed the pairs composed of individuals from different pedigrees which randomly share the same founders to prevent a possible bias in the normalization with $P0$ estimates. As our simulated samples were mainly composed of unrelated individuals (see Table 2.1), we preferred to apply the median $P0$ value for normalization. The kinship coefficient (θ) estimate for each related and unrelated pair was calculated using the

formula:

$$\theta = 1 - (P0pair/P0unrelated) \tag{2.1}$$

This θ estimation approach can give negative results when a pair shares fewer alleles IBS than the ones of an average unrelated pair (Konovalov & Heg, 2008), suggesting a non-kin relationship. Thus, we set the negative θ estimates to 0.

2.5.2 NgsRelate

Unlike READ, NgsRelate (version 2) uses maximum likelihood (ML) for estimating genetic relatedness given genotype information and population allele frequencies. NgsRelate further relies on genotype likelihoods (GL) to account for the uncertainty in low-coverage ancient data. NgsRelate uses an expectation-maximization algorithm to estimate nine condensed Jacquard coefficients (J_1, J_2,..., J_9) given GL and population allele frequencies; these coefficients are then used for the direct calculation of kinship:

$$\theta = J_1 + 0.5(J_3 + J_5 + J_7) + 0.25J_8 \tag{2.2}$$

The inbreeding coefficient F of individuals 1 and 2 is calculated using these Jacquard coefficients (Hedrick et al., 2015) :

$$F_1 = J_1 + J_2 + J_3 + J_4 \tag{2.3}$$

$$F_2 = J_1 + J_2 + J_5 + J_6 \tag{2.4}$$

For NgsRelate analysis, we used the $ANGSD$ program (Korneliussen et al., 2014) to calculate the GL for each individual separately using the $gargammel$-produced BAM files ($-gl$ 2). We limited GL calculation to 200K autosomal SNPs using the "$sites$" parameter for every individual. This analysis left us with 199,730 SNPs passing $ANGSD$ default filters (base quality > 13 and mapping quality > 10). The beagle text output file of $ANGSD$ ($-doGlf$ 2) was manipulated to generate a GL file containing only two individuals with their shared SNPs. We eliminated pairwise missing SNPs by keeping only sites with GL values not equal to 0.33 for neither genotypes for both individuals.Next, we randomly downsampled the shared SNPs between every pair of individuals to 1,000, 5,000, 10,000, 20,000, and 50,000, five times each,

using an in-house bash script. Then, every pair's GL files with five different SNP subsets were converted to the binary GL file format NgsRelate accepts (see Appendix A). The background allele frequency files for corresponding SNPs were prepared using their MAF of the 1000 Genomes TSI sample with n=112 individuals (see Section 2.1). As the autosomal bi-allelic variants with MAF < 0.01 were excluded from the simulations, the MAF threshold of NgsRelate was set to 0 with the option " -l " (as the default is 0.05). The output file produced by NgsRelate for each pair includes a θ value corresponding to a kinship coefficient estimate. We used this estimated value for subsequent analysis.

Merve Nur Güler conducted the kinship analysis with NgsRelate.

2.5.3 lcMLkin

Another relatedness estimation software using genotype likelihood and population allele frequencies is lcMLkin (Lipatov et al., 2015). Assuming a non-inbred population (unlike NgsRelate) and biallelic loci in linkage equilibrium, lcMLkin estimates the maximum likelihood of Cotterman coefficients also using the Expectation-Maximization (EM) algorithm and determines the coefficient of relatedness as r = k_1/2 + k_2. Like NgsRelate, the uncertainty in genotype calls of low-coverage NGS data is modeled by summing log-likelihood values of every possible genotype for each site across the genome.

In the first version of lcMLkin, target bam files were converted to VCF format composed of genotype calls and likelihoods using a Python script provided ($SNPbam2-vcf.py$). This script is also implemented for all genotyped samples' background allele frequency estimation. The recently updated version of lcMLkin (Žegarac et al., 2021) does not provide a genotyping script but allows linkage disequilibrium (LD) filtering, fulfilling the assumption of independence of target SNPs. Also, in the newer version, allele frequency estimation is not just restricted to the samples tested. Any relevant population's allele frequencies can be estimated using the "–$freq$" option of $PLINK\ v1.9$ incorporated into the software for later calculating the IBD probability.

We prepared input VCF files for each pair to run lcMLkin v2.1 implemented for

Python3 (https://github.com/altinisik/lcMLkin-v2.1). For that purpose, we used $bcf-tools\ mpileup$ and $call$ packages (Li, 2011) to estimate the genotype likelihoods of each individual using BAM files for the 200K SNP set with the mapping and base quality filter parameters "-$q10$" and "-$Q13$", respectively. These thresholds were selected based on the default filters of $ANGSD$ to estimate GLs for NgsRelate analysis. In this way, we aimed to render the kinship coefficient estimate results from lcMLkin comparable with the estimates from NgsRelate. Besides the VCF files of target samples, lcMLkin requires the genotype data of the selected background population for allele frequency estimation. This genotype data is provided in $PLINK$ format ($bed/bim/fam$) with an argument "-p". We prepared this genotype data using the 200K autosomal SNPs (MAF > 0.01) chosen from the n=112 TSI sample defined earlier (see Section 2.1). We changed the default allele frequency thresholds integrated into the lcMLkin python script from minimum 0.05 and maximum 0.95 to minimum 0.01 and maximum 0.99. We filtered out missing (non-shared) SNPs from VCF files using an in-house bash script to collect only overlapping SNPs between each simulated pair for the subsequent random downsampling step. After that, we randomly selected 1K, 5K, 10K, 20K, and 50K shared SNPs between pairs of samples, independently five times each, and generated downsampled VCF files using $bcftools\ view$ (Li, 2011) with the "-R" parameter. As the LD pruning application of lcMLkin removes closely linked SNPs from the relatedness analysis, we modified the script such that downsampled SNPs are not filtered by LD. This was done for simplicity to ensure we use the same number of SNPs in each trial and also given that with even 50K SNPs across the genome, the linkage between neighboring SNPs will be minimal.

The relatedness coefficient (r) is represented with the "PI_HAT" estimate in the output files of lcMLkin. We calculated the kinship coefficient value as $\theta = r/2$ for the following analysis.

2.6 Classification of Kinship Coefficient Estimates

To systematically test the reliability and robustness of kinship coefficient estimates by READ, NgsRelate, and lcMLkin on ancient samples, we categorized each simulated

pair into one of five relationship categories, i.e., identical twins, first-, second-, or third-degree related, or unrelated, using their θ estimates. Here we used two assessment criteria. The first criterion we investigated was the arithmetic mean (or average) of theoretical kinship coefficient values. For this purpose, we calculated the mid-point of expected kinship coefficient values of two relatedness degrees (Table 2.2). The mid-point cutoffs calculated for identical twins, first-, second-, and third-degree were 0.375 (average of identical and first-degree expectations), 0.1875 (average of first- and second-degree expectations), 0.09375 (average of first-and second-degree expectations), and 0.03125 (average of second-and third-degree expectations). For instance, in pairs with $\theta \geq 0.375$, the simulated pair is classified as identical twins, while pairs with $0.1875 > \theta > 0.09375$ would be assigned as second-degree, etc. READ and TKGWV2 also use this mid-point cutoff approach to designate kinship estimates to the appropriate relatedness categories.

Table 2.2: Theoretical kinship coefficient values and classification cutoffs. The cutoffs are calculated using arithmetic or geometric mean of theoretical values for identical, first-, second-, third-degree, and unrelated classes.

Relationship	**Theoretical Value**	**Classification**	
		Arithmetic Mean	**Geometric Mean**
Identical	0.5	>0.375	>0.354
First- Degree	0.25	0.1875 - 0.374	0.1768 – 0.353
Second-Degree	0.125	0.09375 - 0.1874	0.08839 - 0.1767
Third-Degree	0.0625	0.03125 - 0.09374	0.03078 - 0.08838
Unrelated	0	0 – 0.03124	0 - 0.03077

The second classification criterion we explored was the geometric mean of theoretical kinship coefficient values. The geometric mean defines the average value of the set of the numbers under study based on their products, and it is always less than the arithmetic mean, being closer to the lower value when two values are used. As θ values decrease with lower degrees of relatedness in a non-linear fashion (see Figures 3.2, and 3.3), we decided to test if using the geometric mean may improve the accuracy of kinship type classification. We calculated the geometric mean of expected kinship

coefficient values in R using the formula:

$$(x_1 \cdot x_2 \cdot ... \cdot x_n)^{\frac{1}{n}} \tag{2.5}$$

The geometric mean cutoffs for identical twins, first and second-degree, are 0.35355, 0.17677, and 0.08838, respectively (Table 2.2). For the third degree, we determined the threshold using theoretical kinship coefficients of the third-degree related and unrelated pairs, 0.0625 and 0.0, respectively. As zero values cannot be tolerated while calculating geometric mean, we applied a modified geometric mean for third-degree cutoff using the $splicejam$ (version 0.0.63.900) package in R (Ward & Farris, 2023). In this way, we derived the third-degree threshold as 0.03078. With a similar logic behind the mid-point approach, if either of the tools estimated the kinship coefficient of a simulated pair above one of these threshold values (0.35355, 0.17677, 0.08838, and 0.03078) and below its upper threshold, we classified this pair in the corresponding relatedness class. For example, pairs are assigned to the first-degree related category only if the estimated kinship coefficient value is within the range of 0.35355 and 0.17677.

2.6.1 Classification Metrics

To compare and contrast the classification performance of the relatedness methods, we created a confusion matrix using either the arithmetic or geometric mean criteria. For this, we used the $confusionMatrix$ function of the $caret$ $v3.5$ package (Kuhn & Max, 2008) developed for R. Based on the estimated values; this creates a matrix summarizing predictions across a reference or known set of values. In this study, the known values correspond to the relationship categories we simulated (first-, second-, third-degree related, and unrelated). By comparing predicted values with actual ones, the function also calculates classification metrics describing the tested tools' inference performance, such as sensitivity (true positive rate, recall, or TPR) and specificity (true negative rate or TNR). Sensitivity describes how frequently the classifier approach can correctly predict the interested class (i.e., the positive cases) (TP/TP+FN). Conversely, the correct classification rate of the classes of non-interest (negative cases) refers to the model's specificity (TN/TN+FP). Regarding our classifier approach relying on either the arithmetic or geometric mean, these metrics were

defined for each class separately, i.e., each relatedness category (first-, second-, third-degree related, and unrelated) had a unique set of these accuracy metrics. Thus, the rate of correctly classified first-, second-, third-degree related, or unrelated pairs represents the sensitivity. To calculate the true negative rate of the class under examination, the accurate assignment of other relationship classes to their classes (not to the positive class) was considered. While producing a confusion matrix and calculating classification metrics in a multi-class scenario like this, it is important to keep the balance between classes, i.e., an equal number of samples for each class. In our study, the first-degree class includes 96 pairs in total, and it has the lowest number of pairs compared to second (n=144 total, n=48 for each type), third (n=144 total, n=48 for each type), and unrelated (n=29,706) classes (see Table 2.1). For this reason, we randomly selected only 96 second- and third-degree related and unrelated pairs using the "sample" function of R without replacement while keeping the number of each relationship type the same for second- and third-degree pairs (n=32). After that, we prepared three different datasets for our tools, consisting of classified estimates based on either arithmetic or geometric mean and their actual classes. We separately applied the confusion matrix function to the datasets for each shared SNP number (1K, 5K, 10K, 20K, and 50K).

The metrics we used for benchmarking READ, NgsRelate, and lcMLkin are the true positive rate (TPR), true negative rate (TNR), false positive rate (FPR), false negative rate (FNR), precision, and the F-score (F1). The false positive and negative rates refer to the rate of incorrect assignments of positive and negative classes by the classifier model, respectively. These metrics provide insight into the model's potential for making incorrect decisions and the level of trust we can place in its results. Alternatively, another metric, precision, focuses on models' predictive ability to avoid false positive predictions. Among all the instances predicted as positive by the model, the proportion of the actual positive predictions describes the precision (TP/TP+FP). It differs from recall in using false positive instances instead of false negatives for the calculation. Each described classification metric uncovers distinct aspects of the model's performance, so it is essential to use these metrics to explore the differences between models. To summarise the overall power of the model for the identification of the classes, the precision and recall are combined in a score called an F-score

(or F1 score). As the arithmetic mean is so sensitive to large differences between values and deviates towards higher values, the harmonic mean of precision and recall is traditionally used to calculate the F-score. In this way, a boost in the F-score, which signifies enhanced classification power, is only possible simultaneously with high levels of precision and recall. The formula of the F-score is

$$F1 = 2 \cdot (precision \cdot recall)/(precision + recall) \tag{2.6}$$

The F-score results are distributed between 0 and 1. A rise in F-score is represented by higher values close to 1. Using the same confusion matrix results, we separately calculated the F-score of three software for each class (first-, second-, third-degree related, and unrelated) and shared SNP counts (1K, 5K, 10K, 20K, and 50K). To understand how often the three software correctly identified the estimates, we also determined the relative frequency of both true and false predictions for each class and SNP count. Additionally, we categorized the false predictions according to their inferred classes using the same confusion matrix again.

2.7 Statistical Tests On Kinship Coefficient Estimates

2.7.1 Binomial Test

In this study, we are interested not only in evaluating the classification capability of the relatedness tools but also in determining if there exist any systematic and significant deviations between point estimates and theoretical kinship coefficients. Estimates may randomly vary from the expected by random noise caused by missing data as well as the randomness of meiotic processes (recombination and segregation). However, if a kinship inference method displays deviations from the theoretical expectations on average, this would be considered as a bias. Such deviation can result from low-quality ancient genome data characteristics, such as post-mortem damage, sequencing error, or reference biases (Günther & Nettelblad, 2019).

By performing the binomial test, we investigated possible biases in kinship coefficient estimates of READ, NgsRelate, and lcMLkin. The binomial test is a method to test if the observed proportion of one category of a binary variable significantly differs

from the expected or theoretical proportion. Essentially, it is a statistical technique used to test the significance of a null hypothesis under the assumption of independent observations with equal probabilities.

Our null hypothesis was that the distribution of estimated kinship coefficient values is equally divided between higher or lower values than theoretical kinship coefficients for each relatedness class. If we assume that none of the benchmarked approaches are biased, it is expected that their point estimates are equal to the theoretical expectations on average for each class. Higher or lower estimations are assumed to be equally likely. Hence, the binomial tests for READ, NgsRelate, and lcMLkin can reveal whether the deviations from theoretical kinship coefficient values (or bias) are significant or not.

To apply binomial tests for each relationship class and SNP count, we used $binom.test$ function from the $stats$ package of R (R Core Team, 2022). Firstly, because of the independence assumption, we prepared a separate dataset for each replicate of each pair and SNP number. Secondly, we divided the point estimates of each relationship degree into two groups: "success" for higher and "fail" for lower values. Lastly, we defined our expected proportion for the success group under the null hypothesis, p = 0.5. Accordingly, the null hypothesis cannot be rejected if the estimates are distributed randomly around the theoretical value. We thus performed 25 two-sided binomial tests across eight classes (related and unrelated groups) and five SNP sets.

2.7.2 Linear Mixed Effect Model

After individually evaluating possible biases in the three genetic relatedness tools, we studied their differences using a broader approach. We thus directly compared the kinship coefficient estimates from READ, NgsRelate, and lcMLkin to determine if their point kinship coefficient estimates, on average, are significantly distinguishable from each other. As we performed three different analyses on the same pair of samples, we applied statistical methods accounting for the dependency between measurements, such as repeated measures ANOVA. Alternatively, a linear mixed effect model (random-effect model or multi-level model) can be preferred to study the effect of software choice on kinship coefficient estimation of repeated samples. The

linear mixed effect model is an effective technique that can accommodate the variability attributed to the samples selected, which are related or clustered. This model combines both fixed and random effects, providing an expansion to traditional linear models. Linear models, also known as linear regression models, examine the effect of an unknown parameter or explanatory variable on a response variable. The fixed factors are explanatory variables specifically investigated for their impacts on the observations. For instance, sex or treatment factors can be fixed effects with limited levels. The primary focus is on understanding the distinctions among these levels. Meanwhile, random effects merely represent random samples from a population and are not the ones that are relevant for elucidating the underlying association between the explanatory variables and response measurements. Depending on the clustering nature of the data, they can be designed in a crossed or nested way. If the clustering is hierarchical, lower-level variables only belong to one of the higher categorical levels; the variables are considered nested. However, if each combination of lower and higher level variables is possible, the random effects become crossed. By utilizing a linear mixed effect model, variations are considered and regulated via random effects while investigating the linear relationship between levels of fixed effects. It is possible to introduce multiple fixed effects to the model, enabling separate or combined analysis with the interaction term. Their interaction can be studied to understand how fixed effects impact the response variable in relation to another one.

A basic linear regression is modeled with a fixed intercept and slope to predict the response value using the explanatory variable. A residual error is also included in the model, indicating variation the linear model cannot explain. In the linear mixed effect model, the linear prediction formula is altered such that the intercept and slope are no longer fixed. This means that the intercept and slope of the model are uniquely defined for each grouping sample. This is achieved through the incorporation of random effects into the model. The intercept or the slope is allowed to vary with the levels of the random effects using random intercept or slope model, respectively. The random intercept model suggests that the baseline values of the response variable may differ among groups of random effects, but the relationship between the response and explanatory variables (or slope) should remain constant. However, this relationship is not fixed in the random slope model and can vary between different groups. To

estimate the parameters of the models, the maximum likelihood (ML) or restricted maximum likelihood (REML) methods can be used. The ML approach maximizes the probability of observing the data given the parameters. However, the variances estimated for random effects using this approach are negatively biased when the number of samples is small. To remove this bias, REML is preferred.

In our linear mixed effect model, the fixed effects were (a) the type of genetic relatedness estimation tools we used, i.e., READ, NgsRelate, and lcMLkin, and (b) SNP counts shared between simulated individuals (1K, 5K, 10K, 20K, and 50K). We explored their effect on kinship coefficient estimations, defined as a response variable. Also, we extended our model with the crossed random effects represented by repeated simulated pairs of samples (see below).

To fit our simulated data to a linear mixed effect model, we used the $lmer$ function in R. We repeated the analysis with each relationship type separately (Table 2.3). Moreover, we again removed the same simulated individuals within the parent-offspring, grandparent-grandchild, and great–grandparent–great–grandchild, and unrelated groups not to add another factor to our model. In this way, we kept only 24 pairs from these relatedness groups. We added the random effect component to our model as " 1|pairs + 1|replicates". This phrase implies that kinship coefficient estimates for the same pairs and the same replicate may deviate from the average in the same manner, which would be expected. This representation also describes that these random effects are crossed. We also tested the random slope model by adding the expressions "(1+Software|pairs)" or "(1+Software+Snp_Count|pairs)". The variable "Software" refers to the tools used for kinship analysis, and "Snp_Count" indicates the number of shared SNPs between individuals.

The linear mixed effect models (LMMs) used in the analysis are as follows :

Model 1 :

$$lmer(\theta \sim Software + SNP_Count + (1|pairs) + (1|replicates)) \tag{2.7}$$

Model 2 :

$$lmer(\theta \sim Software * SNP_Count + (1|pairs) + (1|replicates)) \tag{2.8}$$

Model 3 :

$$lmer(\theta \sim Software + SNP_Count + (1 + Software|pairs) \tag{2.9}$$

Model 4 :

$$lmer(\theta \sim Software + SNP_Count + (1 + Software + SNP_Count|pairs) \tag{2.10}$$

Table 2.3: The number of pairs and replications used in the LMMs for each relationship category. PO: parent-offspring, S: siblings, HS: half-siblings, AV: avuncular, GPC: grandparent-grandchild, FC: first cousins, GAV: grand avuncular, GGPC: great-grandparent-great-grandchild, U: unrelated.

	PO	**S**	**HS**	**AV**	**GPC**	**FC**	**GAV**	**GGPC**	**U**
pairs	24	48	48	48	24	48	48	24	48
replicate	5	5	5	5	5	5	5	5	5

Additionally, we applied the same linear mixed effect model (Model 1) to the absolute differences between the θ estimate of a pair and group mean as a response variable. In this way, we investigated the variations of θ estimates based on the possible deviations from the group means while accounting for the variances between pairs.

2.7.3 Levene's Test

For a comprehensive analysis of kinship coefficient estimates, it is essential to compare not only their means across different methods but also study the variability within each set of estimates. This can lead to a better understanding of which approach is more precise in genetic relatedness estimates. Various tests compare variances between groups, such as the Fisher's F test, Levene's test, Brown-Forsythe test, or Fligner-Killeen test. Unlike others, the Fisher's F test requires that the data per group should have a normal distribution. Indeed, the F test is too sensitive to depart from normality. When the data is not normally distributed, it is better to choose alternative tests that are less affected by deviations from normality. We decided to perform Levene's test to explore the homogeneity of variances between the estimates of the

tools. Levene's test is used to decide the variances of sample groups significantly differ from each other. Unlike the Fisher's F test, the test is particularly helpful when dealing with non-normally distributed samples, as it is less susceptible to deviations from a normal distribution. Using the sum of the mean of differences between each sample and its group mean, called absolute deviation, its test statistic W is calculated.

In order to compare variances using Levene's test, we utilized the $leveneTest$ function in R. We first divided the estimates from READ, NgsRelate, and lcMLkin into groups based on SNP counts and runs. Then, we applied Levene's test separately to each group using their kinship coefficient estimates.

CHAPTER 3

RESULTS

3.1 Simulated Ancient Data

To systematically investigate the performances of READ, NgsRelate, and lcMLkin on kinship estimation on low-coverage ancient genomes, using $PED - SIM$, we simulated ancient human genome sequencing data for pairs of individuals with varying relatedness levels (first-, second-, third-degree related, and unrelated). This included 480 outbred individual pairs, as well as 96 pairs of first- or second-degree pairs of individuals where the parents of the younger individual are third-degree relatives (first cousins). To test the power of the relatedness estimation tools concerning the number of shared SNPs between related or unrelated individuals, we randomly downsampled simulated ancient genomes from 5x to 1x coverage, and then, we randomly selected 1000, 5000, 10,000, 20,000, and 50,000 SNPs overlapping between each pair of individuals. This random selection of SNPs was performed five times independently for each pair and shared SNP count to study the variability of kinship coefficient estimates across distinct SNP subsets.

Four additional features of our simulation may be important to note here. First, we only used SNPs with a minor allele frequency of 0.01 or higher in the background population. Second, we produced data as similar as possible to ancient DNA reads, e.g., bearing PMD patterns, and processed it using standard pipelines. Third, the simulated genome data did not include any background relatedness other than the explicitly modeled inbreeding. Fourth, we also used population allele frequencies directly without estimating them from a sample.

3.2 The Distribution of Kinship Coefficient Estimates

Our benchmarking approach is based on a genetic relatedness quantity, the kinship coefficient (θ), which is the probability of two randomly chosen alleles from a pair of individuals being IBD. Although READ does not directly estimate the kinship coefficient for samples given, estimated normalized pairwise mismatch values ($P0$) can be used to calculate the coefficient using formula 2.1 (see Section 2.5.1.1). On the other hand, NgsRelate and lcMLkin provide a direct estimation of the kinship coefficient.

The genetic relatedness between an equal number of pairs of individuals (n = 48) was analyzed using the three tools for each overlapping SNP number and for nine types of relationship, i.e., parent-offspring, siblings, grandparent-grandchild, half-siblings, avuncular, grand avuncular, first cousins, great-grandparent-great-grandchild, and unrelated.

3.2.1 First-degree Relatedness

NgsRelate and lcMLkin differ from READ in estimating kinship for parent-offspring data, not in the average estimate but in their spread (Figure 3.1). The spread gets narrower for all three software, as expected, through increasing shared SNP counts from 5K to 50K. Interestingly, the majority of point estimates of NgsRelate were under the theoretical value of 0.25 (63%, 70%, 80%, and 93%, respectively). Similarly, even with the presence of 50K shared SNPs, over 50% of the lcMLkin estimates fell below the theoretical threshold. These results point out that at least 5K shared SNPs are sufficient to correctly infer a relationship as a parent-child by READ, NgsRelate, and lcMLkin. Additionally, lcMLkin has a lower tendency to deviate from the anticipated kinship coefficient.

Regardless of using READ or NgsRelate, if the low number of SNPs (i.e., 1K SNPs) were shared and between individual pairs, simulated sibling estimates showed a bias toward slightly lower values. This bias increased in the presence of more SNPs used by NgsRelate (median $\theta < 0.24$). For 20K and 50K SNPs, we observed two modes in the kinship estimate distribution, above and below the expected threshold. We

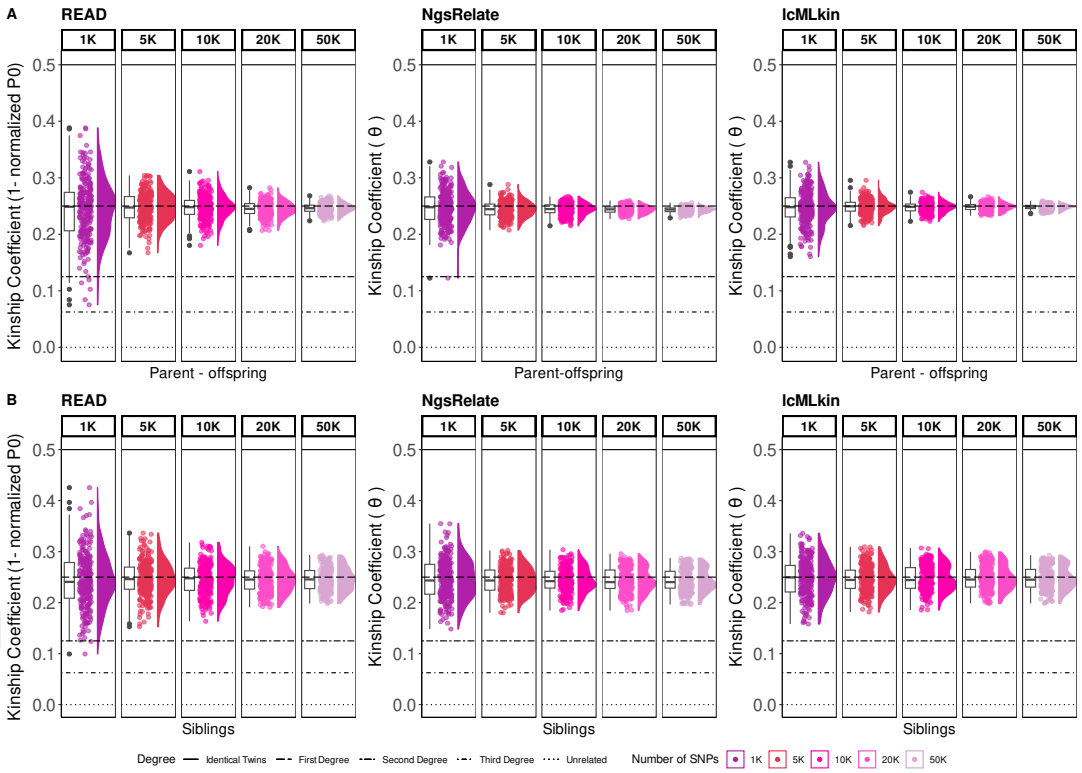


Figure 3.1: Kinship coefficient estimates of simulated first-degree pairs, (A) parent-offspring, and (B) siblings. The points represent the kinship coefficient estimated by READ, NgsRelate, and lcMLkin for one pair of individuals sharing 1K, 5K, 10K, 20K, or 50K SNPs. For each SNP subset and each relationship type, the total number of simulated pairs is 240. Horizontal lines show the theoretical kinship coefficients. The boxplots, jitter-added points, and density plots show the distribution of the same sample of 240 points.

detected a similar underestimation for kinship coefficient estimates of sibling pairs while using lcMLkin. Notably, lcMLkin did not exhibit any bias on average when using 1K shared SNPs between siblings.

Even though the variation of θ estimates is high, the median is close to the theoretical kinship coefficient value for first-degree relatedness using lcMLkin. We also observed that the θ distributions for sibling pairs were wider than those for parent-offspring pairs. For parent-offspring pairs, the variation gradually decreased as the number of SNPs increased from 1K to 50K. However, there was weaker or even no visible change in variation observed for sibling pairs in all three software, likely because

the variation among estimates is driven by biological variation rather than sampling error. Moreover, we found that even with a high number of SNPs (> 10K), NgsRelate still tends to underestimate the kinship coefficient. As a result, among the tools that we tested for first-degree relatedness, NgsRelate appears to be the most negatively biased.

3.2.2 Second-degree Relatedness

Near-zero values were observed with individuals sharing 1K SNPs while using each tool (Figure 3.2). The variation was substantially reduced with >5K SNPs, and the point estimates reached the theoretical threshold ($\theta = 0.125$) with lcMLkin regardless of the relationship type. Grandparent-grandchild pairs showed the largest decline in the median with NgsRelate, but the greatest variation with READ. Half-siblings and avuncular median values were close to each other using NgsRelate, but half-sibling estimates for each SNP count except 50K were slightly higher than avuncular estimates. Furthermore, they were higher than half-sibling estimates from READ (except 5K and 50K common SNPs). Similar to the relationship between half-siblings and avuncular pair estimates from NgsRelate, half-sibling estimates by READ were closer to the theoretical kinship coefficient than avuncular estimates. Sharing less than 10K SNPs between second-degree related pairs of individuals made READ more vulnerable to deviations from expectation as variation increases dramatically compared to NgsRelate. Notably, overlapping 50K SNPs mostly leads to a higher decrease in the estimates for NgsRelate.

These findings suggest that half-sibling relationships can be most precisely estimated among all second-degree relatives using any of the three tools. Furthermore, we observed a lower tail of the kinship coefficient estimates for second-degree relatives when the pairs shared less than 20K SNPs and were analyzed using READ, with a proportion of estimates closer to the expected θ value for third-degree relatedness ($\theta = 0.0625$). Additionally, the results suggest that the lcMLkin may be the most accurate option for predicting second-degree relationships (see below). However, there was no clear and consistent separation between READ and NgsRelate median estimates.

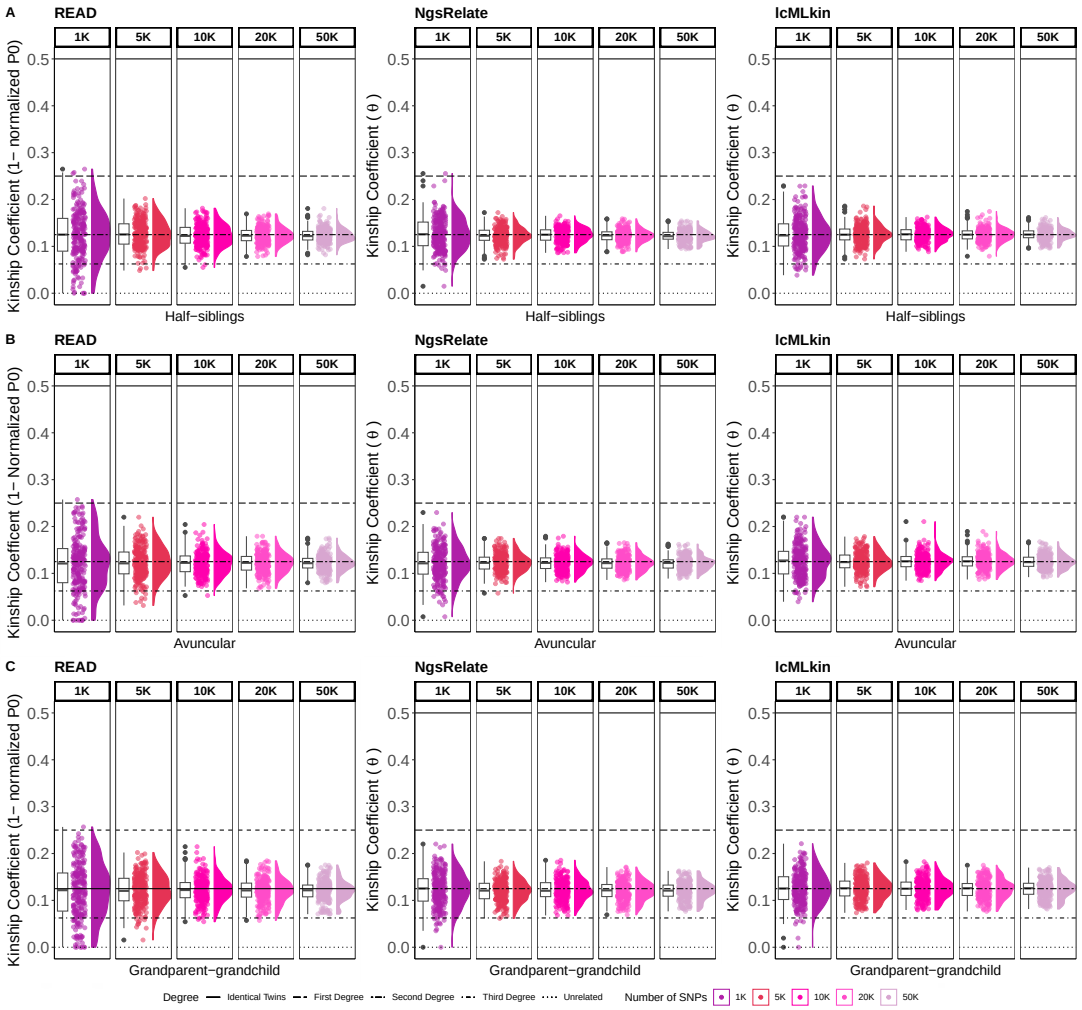


Figure 3.2: Kinship coefficient estimates of simulated second-degree pairs,(A) half-siblings, (B) avuncular, and (C) grandparent-grandchild. The points represent the kinship coefficient estimated by READ, NgsRelate, and lcMLkin for one pair of individuals sharing 1K, 5K, 10K, 20K, or 50K SNPs.For each SNP subset and each relationship type, the total number of simulated pairs is 240. Horizontal lines show the theoretical kinship coefficients. The boxplots, jitter-added points, and density plots show the distribution of the same sample of 240 points.

3.2.3 Third-degree Relatedness

When we ran the analysis with simulated third-degree relatives (i.e., first cousins, grand avuncular, and great-grandparent-great-grandchild), differences among the three tools became more apparent (Figure 3.3). We note that READ was initially presented

by its authors (Kuhn et al., 2018) as a tool for inferring second-degree or higher relationships; nevertheless, we decided to study its performance in detecting third-degree relatives by calculating θ for each pair (see Section 2.5.1.1) and contrasting this with the estimates of NgsRelate and lcMLkin.

We found that θ values calculated for third-degree pairs spread substantially towards the expected value for unrelated pairs ($\theta = 0$) and reciprocally spread up to the second-degree expected value. While a median of point estimates from first cousins was slightly lower than the theoretical threshold for NgsRelate and READ, great-grandparent - great-grandchild pair estimates from all software were above the expected value, even with the maximum number of SNPs used. The first cousin kinship coefficients were initially overestimated when using lcMLkin with a low number of SNPs (1K and 5K). These values became underestimated as the number of SNPs increased. Nonetheless, with only 20K shared SNPs lcMLkin appeared to be able to estimate third-degree relatedness relatively accurately. We also observed that the median estimates from lcMLkin for grand avuncular pairs sharing 1K, 20K, or 50K SNPs approached the theoretical limit. However, this was not true for the median estimates of READ and NgsRelate. These estimates fell below the expected threshold of third-degree relatedness, with the minimum value being 0.0604 for NgsRelate (with 50K SNPs) and 0.0583 for READ (with 10K SNPs).

Overall, these results indicate that the performance of the tools is apparently sensitive to the third-degree relatedness type analyzed. While great-grandparent-great-grandchild pairs are represented with higher median estimates, first cousins are represented with lower estimates across nearly all trials. An increase in SNP counts makes a positive contribution to lcMlkin to predict the relationship between grand avuncular accurately, but less so for READ and NgsRelate.

3.2.4 Unrelatedness

We also analyzed unrelated pairs, which are simulated individuals from distinct founder pairs (Figure 3.4.). Their kinship coefficient estimations were gathered around zero, which meets the theoretical expectation. We found that the pairwise estimates with 1K SNPs showed a skewed distribution overlapping with the second-degree expected

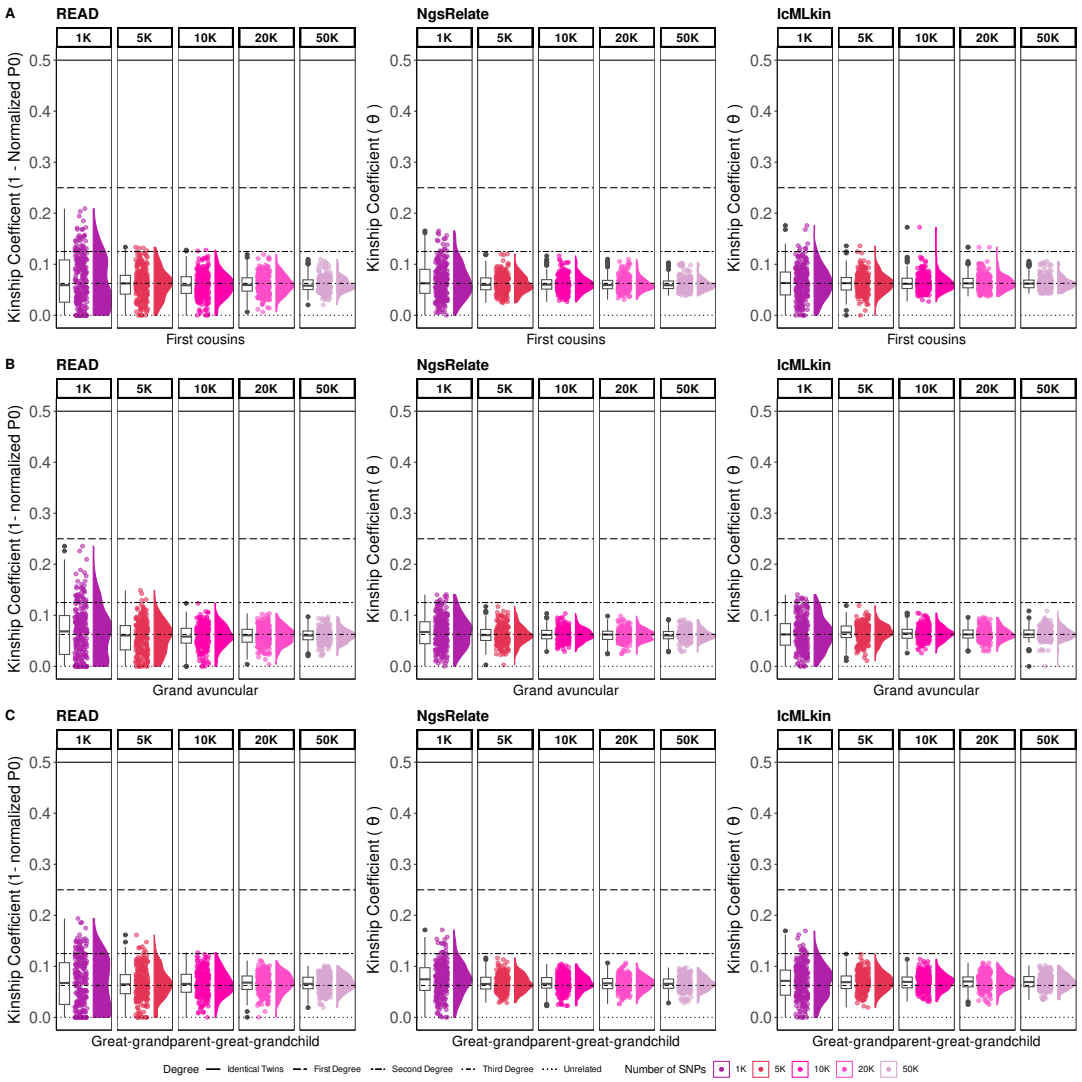


Figure 3.3: Kinship coefficient estimates of simulated third-degree pairs, (A) first cousins, (B) grand avuncular, and (C) great-grandparent-great-grandchild. The points represent the kinship coefficient estimated by READ, NgsRelate, and lcMLkin for one pair of third-degree related individuals sharing 1K, 5K, 10K, 20K, or 50K SNPs.For each SNP subset and each relationship type, the total number of simulated pairs is 240. Horizontal lines show the theoretical kinship coefficients. The boxplots, jitter-added points, and density plots show the distribution of the same sample of 240 points.

value for READ and lcMLkin and with the third-degree expected value for NgsRelate. Notably, the medians of kinship coefficient values estimated by NgsRelate and lcMLkin (>1K SNPs) were slightly higher than the unrelated expected value ($\theta > 0.0018$ and $\theta > 0.0024$, respectively). Unlike estimates of relative pairs, the variabilities of

θ estimates for unrelated pairs were generally low, except estimates from 1K shared SNPs.

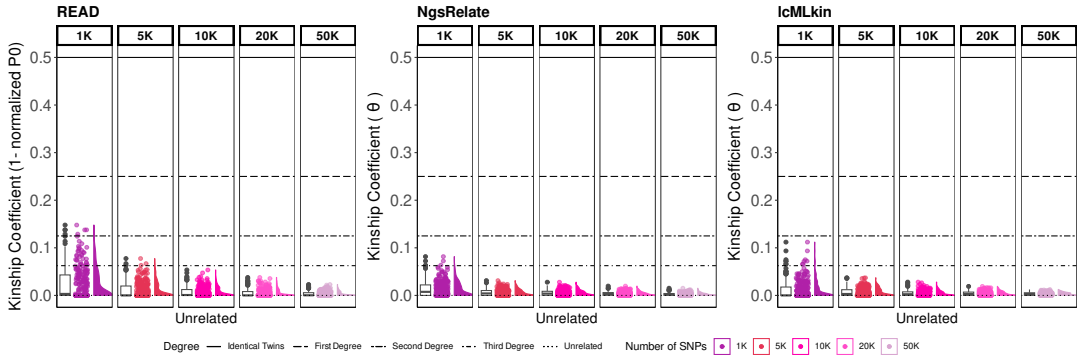


Figure 3.4: Kinship coefficient estimates of simulated unrelated pairs. The points represent the kinship coefficient estimated by READ, NgsRelate, and lcMLkin for one pair of individuals sharing 1K, 5K, 10K, 20K, or 50K SNPs.For each SNP subset, the total number of simulated pairs is 240. Horizontal lines show the theoretical kinship coefficients. The boxplots, jitter-added points, and density plots show the distribution of the same sample of 240 points.

3.3 Comparison of Kinship Coefficient Estimates Using Linear Mixed Effect Model

To further explore whether these differences between READ, NgsRelate, and lcMLkin θ estimates are statistically significant and which tool outperforms the others, we applied a linear mixed effect model test, basically fitting a linear model to the θ estimates (the response variable) with fixed and random effects. We defined the fixed effects as the genetic relatedness estimation tools and shared SNP counts since these factors have the potential to change the response variable in a systematic manner. Because we have analyzed the same pairs repeatedly, using three different tools to measure their kinship coefficients, the pairs, and runs were considered as clustered grouping variables and included as random effects in the model. This mixed effect model can thus control the variations in θ values at both the pair and run levels. We first fit a random intercept model to our data (Model 1, see formula 2.7 and Table 2.3), allowing a change in the intercepts of the levels of the random effects. Further,

we tested the interaction of the fixed effects (Model 2, see formula 2.8 and Table 2.3), i.e., the relatedness tools and SNP counts, to uncover the sensitivity of the different approaches to the number of shared SNPs. Besides the random intercept model, we also performed a random slope model (Models 3 and 4, see formula 2.9 and 2.10 and Table 2.3) to account for the differences between the impacts of software used or SNP counts shared on distinct pairs or runs.

To keep the model as simple as possible and to avoid including one more random effect in the model, we removed the pairs containing the same individuals from parent-offspring, grandparent-grandchild, great-grandparent-great-grandchild, and unrelated categories. Thus, half of the pairs (n=24) remained and were used for model fitting just for these categories (see Table 2.3).

We separately applied the models to the relationship classes. Our reference categories of the model were selected as lcMLkin and 50K shared SNPs.

3.3.1 First-degree Relatedness

For first-degree related pairs, the model overall suggests that NgsRelate and READ have significantly lower means of kinship coefficient estimates than lcMLkin (Figure 3.5, Appendix B). The lowest estimate belonged to NgsRelate. However, according to a model, the shared SNP counts do not have any significant influence on the mean estimates compared to estimates with 50K SNPs. Notably, the pairs as a grouping variable can explain some part of the variations in the θ estimates for this model, but that is not true for repeat (see Appendix B). Only for siblings, the variance among the pairs is higher than the residual variance. The variance for the other random effect repeats was equal to 0. This indicates that repeated estimation of kinship coefficients selecting random SNPs overlapping between individuals does not contribute to the variability. Alternatively, when we allowed interactions between SNP counts and software, the tools did not differ from each other any longer, likely because of over-fitting (Appendix B). Thus, we can conclude that the model cannot estimate a significant difference between these tools for a reference group, 50K SNPs. With the interaction model, we observed that for some of the analyses, the bias of θ estimates from expected values differs among tools and SNP counts. However, the interaction

between SNP counts and target tools did not significantly contribute to the kinship coefficient estimations in most of the analyses. In other words, the differences between genetic relatedness estimators for first-degree relatedness are not solely dependent on the number of SNPs shared between simulated individuals.

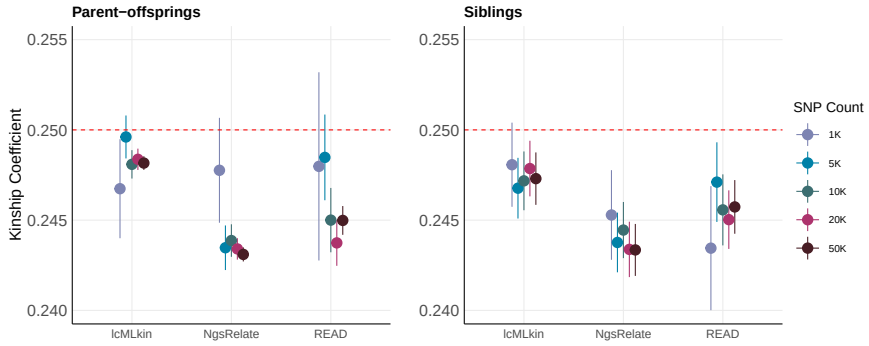


Figure 3.5: The summary of linear mixed effect model with the interaction of SNP counts and genetic relatedness estimation tools for first-degree relatedness. The mean of estimated kinship coefficients for each overlapping SNP count is described as points with distinctive colors. The vertical lines show one standard error, estimated using all pairs (n=48) and replicates (n=5 per pair). The red dashed line represents the theoretical kinship coefficient value for first-degree relatedness.

3.3.2 Second-degree Relatedness

For the second-degree related pairs, except for half-sibling pairs, the tools differentiate significantly from each other in their mean estimates (Figure 3.6, Appendix B). Grandparent-grandchild's fitted intercept fell below the theoretical threshold (0.122), while half-siblings met the expectation (0.125). The model predicted a slightly overestimated intercept for avuncular pairs analyzed with lcMLkin (0.126). Like the model with first-degree relatives, the mean estimates from lcMLkin have the highest values, while READ gives the lowest estimates. With this model, we again could not find any significant relationship between the number of shared SNPs and mean kinship coefficient estimates for second-degree relatives. Also, the results again suggest that repeat (random SNP subsets) is not a significant random effect. Similar to the interaction model with first-degree relatives, the interaction term did not show

any possibly significant relationship between SNP counts and tools (relative to 50K SNPs estimates from lcMLkin) for second-degree pairs (Appendix B). For the half-siblings, we could not detect any significance between mean estimates from the three tools, demonstrating an equal power for kinship estimation tools to distinguish half-siblings.

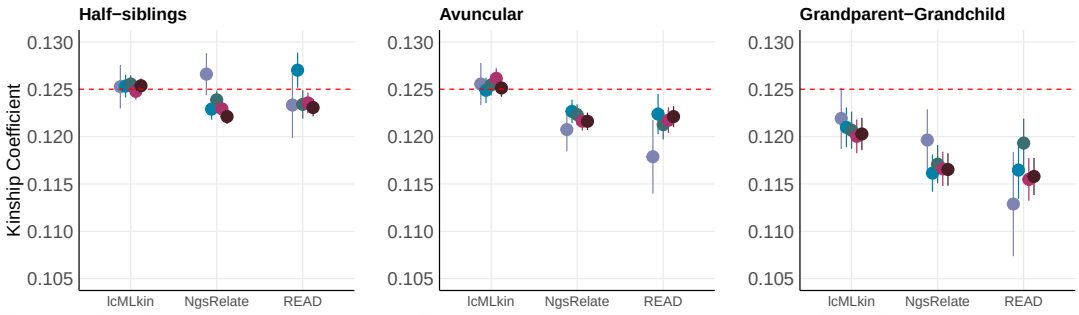


Figure 3.6: The summary of linear mixed effect model with the interaction of SNP counts and genetic relatedness estimation tools for second-degree relatedness. The mean of estimated kinship coefficients for each overlapping SNP count is described as points with distinctive colors. The vertical lines show one standard error, estimated using all pairs (n=48) and replicates (n=5 per pair). The red dashed line represents the theoretical kinship coefficient value for second-degree relatedness.

3.3.3 Third-degree Relatedness

The random intercept model considering all the estimates from distinct SNP counts (Model 1, see formula 2.7 and Table 2.3) predicts that NgsRelate and lcMLkin behave similarly for third-degree relatedness, but both differ from READ (Figure 3.7). Regardless of the relationship type, the difference between estimates from READ and the others was 0.003 on average (Appendix B). The model showed that the estimates from READ were biased towards lower values when we used first cousins and grand avuncular relatives; however, READ θ estimates best approximated the theoretical expectation for the great-grandparent-great-grandchild relationship. Again, pairs were identified as a significant random effect, but random SNP subsets did not (see Appendix B). The random intercept model fit also confirms that using only 1K shared SNPs to estimate θ between third-degree related individuals tends to increase

the estimates by about 0.004 relative to the reference category for SNP counts, which is 50K. We also tested the dependency of estimates from each software on the number of SNPs by integrating interaction terms into the random intercept model (Model 2) (Appendix B). We discovered that using only 1K, 5K, or 10K SNPs resulted in a substantial discrepancy between READ and NgsRelate or lcMLkin. Additionally, all the tools estimated kinship coefficient values similarly on average for first cousins and great-grandparent-great-grandchild relationships sharing low to high number of SNPs (except for READ underestimating θ using 10K SNPs for first cousins).

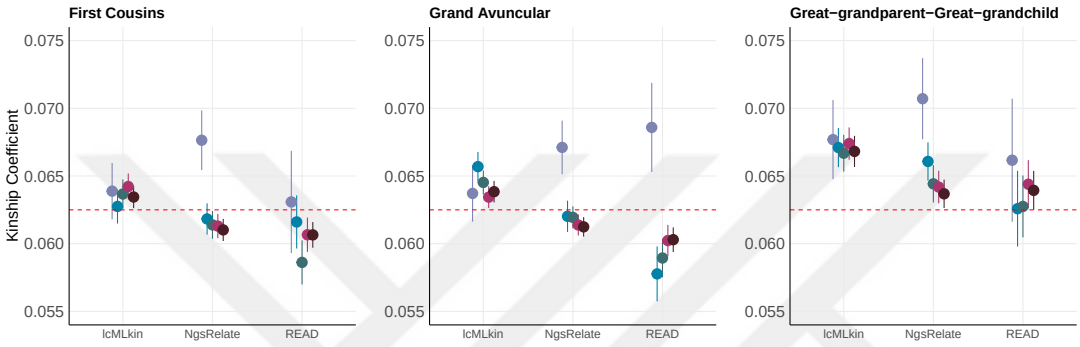


Figure 3.7: The summary of linear mixed effect model with the interaction of SNP counts and genetic relatedness estimation tools for third-degree relatedness. The mean of estimated kinship coefficients for each overlapping SNP count is described as points with distinctive colors. The vertical lines show one standard error, estimated using all pairs (n=48) and replicates (n=5 per pair). The red dashed line represents the theoretical kinship coefficient value for third-degree relatedness.

3.3.4 Unrelatedness

We analyzed the θ estimates of unrelated pairs using the same linear mixed effect models (Figure 3.8, Appendix B). According to the Model 1 (see formula 2.7 and Table 2.3), the mean of θ estimates was 0.002. This slight increase could be due to noise in the data, whether biological (e.g. meiosis) or technical (e.g. reference bias). lcMLkin and NgsRelate predicted the θ similarly on average, but READ gives higher estimates. Using 1K SNPs for the kinship analysis significantly increased the θ estimates for all tools. From lower to higher shared SNP counts, the mean estimates

started to be closer to the theoretical value. With the maximum SNP count (50K), we obtained the lowest mean estimate from NgsRelate.

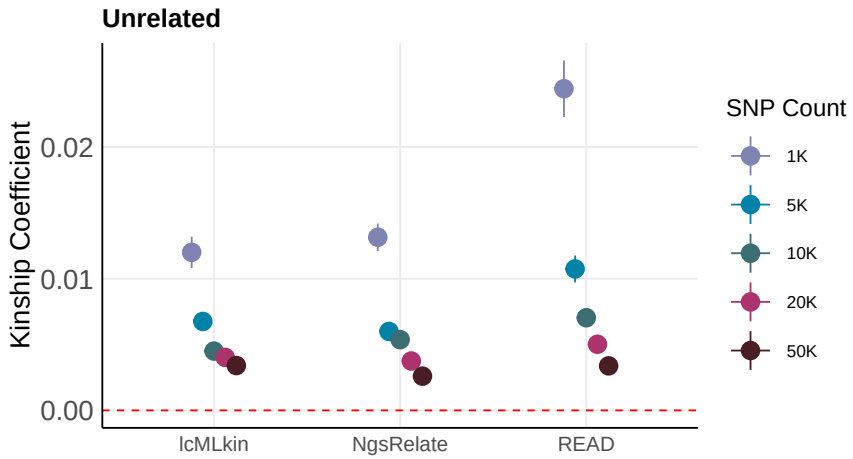


Figure 3.8: The summary of linear mixed effect model with the interaction of SNP counts and genetic relatedness estimation tools for unrelated pairs. The mean of estimated kinship coefficients for each overlapping SNP count is described as points with distinctive colors. The vertical lines show one standard error, estimated using all pairs (n=48) and replicates (n=5 per pair). The red dashed line represents the theoretical kinship coefficient value for unrelated pairs.

Integrating random slope to the models (Model 3 and 4, see formula 2.9 and 2.10) was unsuccessful because the models did not converge. The underlying reason could be overfitting as the number of parameters increased dramatically.

3.3.5 Overview of Linear Mixed Effect Model Results

The above-described linear mixed effect models with random intercepts showed us that methodological differences between READ, NgsRelate, and lcMLkin significantly affect their average kinship coefficient estimates. We observed that lcMLkin outperforms other methods by producing pairwise estimations closer to the theoretical values across most conditions. While NgsRelate showed a negative bias in analyzing first-degree relationships, READ had the lowest mean values for second and third-degree relatives, except for half-siblings. Given all the estimates from all the

SNP counts, NgsRelate only matched lcMLkin in mean estimates for third-degree relationships. When we used models with the interaction (Model 2, formula 2.8), we found that for the categories of siblings, avuncular, and grand-avuncular relationships, shared SNP counts had an influence on the mean estimates of READ, NgsRelate, and lcMLkin (see Appendix B). On the other hand, kinship tools showed no interaction with SNP counts across other types of relationships, namely half-siblings, grandparents and grandchildren, first cousins, and great-grandparents and great-grandchildren. In general, we were unable to identify a clear correlation between tool performance, measured as average proximity to the expected value, and the number of overlapping SNPs between simulated individuals.

3.4 Bias in Kinship Coefficient Estimates

If READ, NgsRelate, and lcMLkin were unbiased methods, their θ estimates should center around the expected kinship coefficient values. Depending on the kinship inference methodology or the nature of the data (i.e., low-coverage ancient genomes), the observed estimates can diverge from the expected. Conceivably, a systematic bias could result from the software itself being tendential to either SNP counts or relationship degree/type. We used the binomial test to assess whether these tools significantly underestimated or overestimated the simulated pairs' kinship coefficient values. For sibling pairs, a number of SNPs from 1K to 20K did not yield a significant underestimation or overestimation by READ and NgsRelate (see Appendix D). However, two runs with 20K SNPs and three runs with 50K SNPs of NgsRelate analysis showed a significant bias toward lower values ($p < 0.03$ and success rate < 34%). This is in line with the kinship coefficient distribution at 50K SNPs from NgsRelate, showing a bimodal pattern with a higher peak below the theoretical threshold (Figure 3.1). The same pattern was observed with the first-cousin pairs used by READ for 1K shared SNPs (four runs out of five with $p < 0.03$ and success rate < 34%). This was compatible with the previous observation that most of kinship coefficient values from READ at 1K SNPs were under the theoretical value (> 60% for each run). For first cousins, some of the repeated analysis of 20K and 50K SNPs also showed a significant underestimation with NgsRelate ($p < 0.03$). Apart from these biased estimates, both tools

resulted in unbiased estimations regardless of SNP count and relatedness degree in most of the runs. Hence, it appears that neither READ nor NgsRelate were statistically superior to one another in estimating an actual value for variable relationship kinds. In contrast, the estimates from lcMLkin did not show any significant bias in any condition.

3.5 The Variations of Kinship Coefficient Estimates

From first-degree relatedness down to third-degree, repeating recombination and Mendelian segregation events in every generation makes IBD probability and relatedness prediction challenging. To investigate this, we performed a linear mixed effect model test again, but at this time, we compared the groups based on their absolute differences between estimates and mean values (see Appendix B). By analyzing these differences through a linear mixed-effect model, we can study the variations in kinship coefficient estimations associated with different pairs of individuals. The difference in the estimates of simulated sibling pairs did not significantly differ between lcMLkin and NgsRelate, but the mean estimates of NgsRelate and lcMLkin were significantly lower than READ. The underlying effect was caused by 1K and 5K estimates from READ. Similarly, lcMLkin and NgsRelate could not be differentiated from each other in their absolute mean differences in parent-offspring estimates. The estimates from READ significantly had higher absolute mean differences compared to others, again caused by using 1K or 5K SNPs. For second-degree pairs, we observed consistent results between relationship types. Like first-degree results, the estimates from READ varied more compared to NgsRelate and lcMLkin, mostly due to using 1K, 5K, or 10K SNPs. Also, the model indicated that overlapping SNP counts less than 20K used for lcMLkin had a significant variation difference relative to SNP counts more than 20K. We observed a similar pattern for a model fit on third-degree related pairs. We conducted a further analysis using the same model but included the interaction of fixed effects. The high variance of estimates from READ (relative to lcMLkin estimates with 50K SNPs) was significantly represented with 1K, 5K, and 10K overlapping SNP counts between first cousins and grand avuncular.

These differences showed a deviation from the mean; therefore, it also reflects the

Table 3.1: Levene's test results of kinship coefficient estimates from READ, NgsRelate, and lcMLkin for variable relatedness degrees (first, second, and third) and SNP counts (n=5 per count). The values show the p-values from Levene's test. The values with a star represent significant p-values (<0.05). PO: parent-offspring, S: siblings, HS: half-siblings, AV: avuncular, GPC: grandparent-grandchild, FC: first cousins, GAV: grand avuncular, GGPC: great-grandparent-great-grandchild.

SNP Count	**Repeat**	**First Degree**		**Second Degree**			**Third Degree**		
		PO	S	HS	AV	GPC	FC	GAV	GGPC
1K	1	0.03*	0.69	<0.01*	0.09	0.09	<0.01*	<0.01*	0.03*
	2	0.01*	<0.01*	<0.01*	0.09	0.01*	0.01*	<0.01*	0.08
	3	<0.01*	0.06	<0.01*	<0.01*	0.01*	<0.01*	0.02*	0.04*
	4	0.12	<0.01*	<0.01*	<0.01*	<0.01*	<0.01*	<0.01*	0.01*
	5	0.03*	0.40	0.52	<0.01*	0.01*	<0.01*	<0.01*	0.02*
5K	1	<0.01*	0.08	<0.01*	<0.01*	0.48	<0.01*	<0.01*	0.02*
	2	0.01*	0.14	0.01*	<0.01*	0.01*	<0.01*	<0.01*	0.04*
	3	0.01*	0.41	<0.01*	<0.01*	0.12	0.03*	0.01*	0.01*
	4	<0.01*	0.07	<0.01*	<0.01*	0.35	<0.01*	<0.01*	<0.01*
	5	<0.01*	0.19	<0.01*	0.02*	0.16	0.02*	<0.01*	0.16
10K	1	<0.01*	0.56	<0.01*	<0.01*	0.84	<0.01*	0.02*	<0.01*
	2	<0.01*	0.18	<0.01*	0.08	0.20	0.16	<0.01*	0.48
	3	0.06	0.04*	<0.01*	0.01*	0.61	0.01*	<0.01*	0.06
	4	<0.01*	0.09	<0.01*	0.29	0.36	<0.01*	<0.01*	<0.01*
	5	0.08	0.72	<0.01*	0.01*	0.21	<0.01*	<0.01*	0.65
20K	1	<0.01*	0.28	0.16	0.04*	0.68	0.02*	<0.01*	0.33
	2	0.03*	0.84	0.02*	0.70	0.52	0.01*	<0.01*	0.17
	3	0.03*	0.45	0.06	0.17	0.79	0.11	0.38	0.05
	4	<0.01*	0.65	0.08	0.01*	0.21	0.13	<0.01*	0.10
	5	<0.01*	0.86	0.72	0.04*	0.52	0.13	0.02*	0.04*
50K	1	<0.01*	0.91	0.74	0.66	0.88	0.35	0.55	0.26
	2	<0.01*	0.93	0.26	0.94	0.96	0.34	0.13	0.31
	3	<0.01*	0.98	0.48	0.60	0.99	0.88	0.12	0.06
	4	<0.01*	0.95	0.41	0.33	0.92	0.87	0.60	0.36
	5	<0.01*	0.99	0.42	0.36	0.81	0.37	0.33	0.14

variation in kinship coefficient estimations by three methods. Hence, we also directly compared their variance levels using the paired Levene's test. As we saw in Section 3.3 and Figures 3.1-3.3, the variances of estimates were significantly different among tools for most relatedness types and shared SNP counts (Table 3.1). Siblings were an interesting exception, with no difference among tools and SNP counts. This is likely due to the variation of IBD patterns across genomes of sibling pairs. This was also true for grandparent-grandchild pairs except for 1K ($p < 0.15$). Yet, the gap between variances was closed in all the relationships analyzed using 50K SNPs. The most distinct differences in the estimated relationships were found in parent-offspring pairs, followed by other second and third-degree pairs.

Overall, these tests suggested that READ, NgsRelate, and lcMLkin differ from each other in terms of the mean of their estimates or deviation from expectations under some conditions. All tools performed relatively well at higher SNP counts and relationship kinds, even with third-degree relatives. However, the point estimates from READ, NgsRelate, and lcMLkin had different levels of variability. Hence, we hypothesized that the classification of simulated pairs to their known relationship degrees could vary in accuracy between these approaches.

3.6 The Classification and Accuracy of Kinship Coefficient Estimates

Since for almost each SNP set, kinship coefficient estimates from the three tools were found dispersed over a wide range; we require an interval considering theoretical limits to classify pairs into relatedness classes. Like READ, we first used the arithmetic mean of kinship coefficient expectations for all three tools to set the boundaries between each relationship degree (see Table 2.2). To evaluate the performances of READ, NgsRelate, and lcMLkin, we measured their effectiveness by comparing predicted kinship coefficient estimates with the simulated (actual) estimates for every pair. We defined true positive rate (TPR), false positive rate (FPR), true negative rate (TNR), and false negative rate (FNR) for each relatedness category and the number of overlapping SNPs. An imbalance in the size of each class can create biases in the classification measurements. Hence, we kept the number of pairs in each relationship degree the same (n=96 for first-, second-, and third-degree relatedness and unrelated

categories).

Alternatively, we also tested whether the geometric mean of theoretical kinship coefficients can be better thresholds than the arithmetic mean for classification (Figure 3.4). We reasoned that because the geometric mean is closer to lower values and given the left-skew observed for at least some of the θ distributions (Figures 3.2 and 3.3), the geometric mean may perform better (see Methods Section 2.6).

3.6.1 First-degree Relatedness

Results for first-degree relatives (i.e., parent-offspring (n=240) and siblings (n=240)) showed that almost all the pairs with 5K overlapping SNP numbers were assigned correctly for all software (98% for READ and 100% for NgsRelate and lcMLkin) (Figure 3.8), with high TPR and low FPR values (98% TPR and 0.3% FPR for READ, and 100% TPR and 0% FPR for NgsRelate and lcMLkin) (Figure 3.9). At 1K shared SNPs, NgsRelate and lcMLkin could still accurately assign first-degree relatives with 96% TPR and 1% FPR, but the accuracy decreased for READ, with 85% TPR and 5% FPR. As shown in Figure 3.1, the variation of kinship coefficient estimates from siblings was higher than the ones from parent-offspring pairs. Hence, the variable estimates of sibling pairs could explain the decrease in correct assignments when using <10K SNPs. Overall, we concluded that up to 20K SNPs, NgsRelate, and lcMLkin slightly outperform READ. Yet, with 20K or 50K SNPs, the power of all three tools to infer first-degree relatedness became equalized.

3.6.2 Second-degree Relatedness

The more distant the relatedness is, the lower the power of relatedness estimation analyzing the simulated second (n=480) and third-degree related pairs (n=480). NgsRelate and lcMLkin achieved a high classification accuracy for second-degree relatives (89% TPR with >5K SNPs) as opposed to READ (66% TPR with >5K SNPs). For 1K SNPs, the power of the tools declines (40% TPR for READ, 60% for NgsRelate, and 63% for lcMLkin). For all tools, from the lowest to the highest SNP numbers, second-degree pairs are primarily misclassified as third-degree in decreasing percent-

ages. Alternatively, simulated pairs with the lowest SNPs (1K and 5K) were incorrectly assigned to the first-degree and unrelated classes. These results indicate that if the samples used for kinship analysis share a low number of SNPs (< 5K), tools' kinship coefficient estimates for second-degree relatedness could be unreliable.

3.6.3 Third-degree Relatedness

Misclassification tends to increase with the analysis of simulated third-degree relatives compared to other classes for each SNP count. READ assigned most of the third-degree pairs (60% of pairs) as second-degree or unrelated at 1K SNPs (40% TPR). Similarly, we observed that the recall of the NgsRelate and lcMLkin classification declined by under 70%. The actual third-degree pairs were wrongly predicted as second-degree or unrelated, reflected by an FNR value above 34%. More than 1K SNPs shared between third-degree relatives were enough to achieve higher correct identification rates by NgsRelate and lcMLkin (> 89% TPR). Also, kinship analysis with NgsRelate and lcMLkin for third-degree relationships still overtakes the one with READ.

3.6.4 Unrelatedness

Classification generally worked well with unrelated pairs using any of the tools (with TPR > 86% for READ and > 99% for NgsRelate and lcMLkin) for 5K SNPs. Using 1K SNPs, TPR values went down to 71% for READ and to 87% for NgsRelate and lcMLkin). However, the classification accuracy of unrelated pairs would likely worsen if we had included more distant relatives in the simulations.

3.6.5 Geometric Mean for the Classification

Specifically for first-degree, second-degree, and unrelated classes, geometric mean has slightly higher power to differentiate between categories and correctly assign them regardless of which tool was used. However, the classification accuracy of READ, NgsRelate, and lcMLkin for third-degree relatives decreases when we use

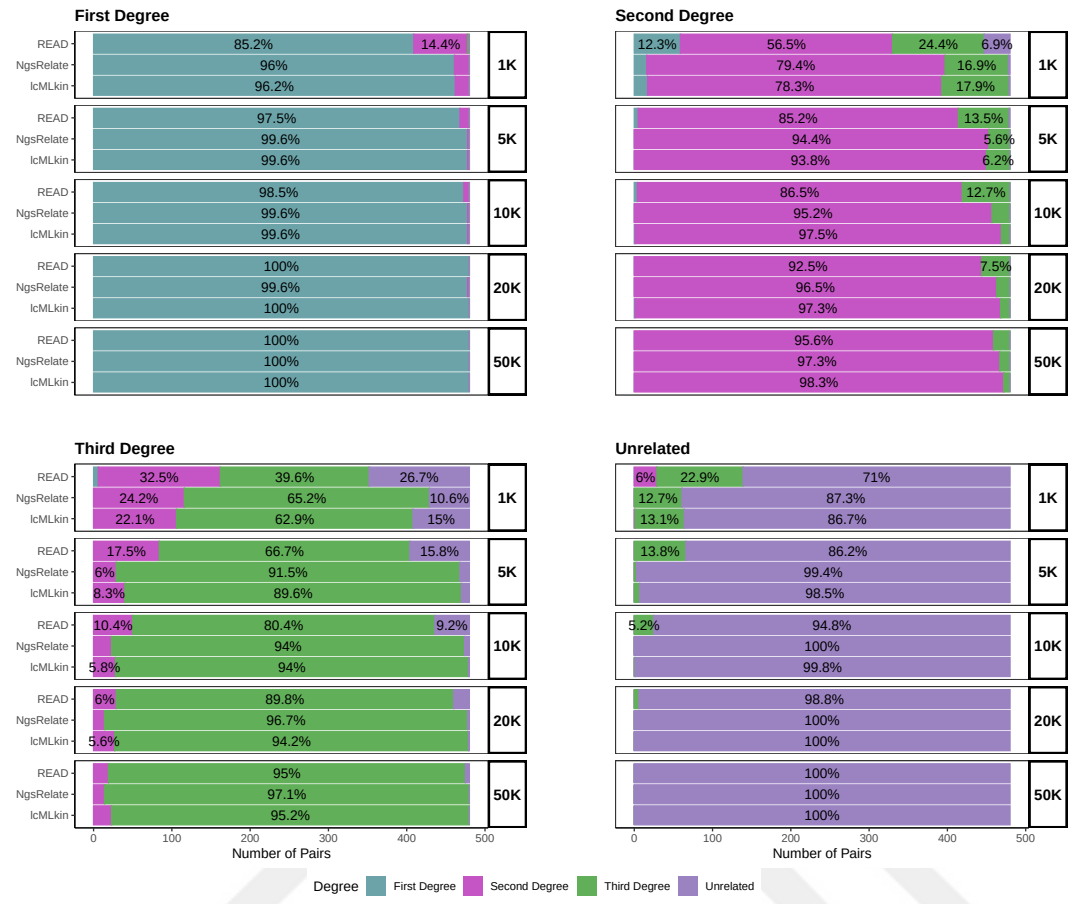


Figure 3.9: The relative frequency of pairs assigned to first-, second-, and third-degree related and unrelated categories by READ, NgsRelate, and lcMLkin. The kinship coefficient estimates from these tools were classified using the arithmetic mean of theoretical kinship coefficients. Colors refer to the assigned relatedness degree. The frequencies of pairs assigned to each category are indicated as percentages inside the bars (only for categories with frequency >5%).

geometric instead of the arithmetic mean. This drop in TPR was not dependent on how much SNP pairs share. Even with 50K SNPs, the difference between TPR values was 4% for READ, 2% for NgsRelate, and 3% for lcMLkin. The reason for this decrease was the actual third-degree related pairs being classified as second-degree more with the geometric mean.

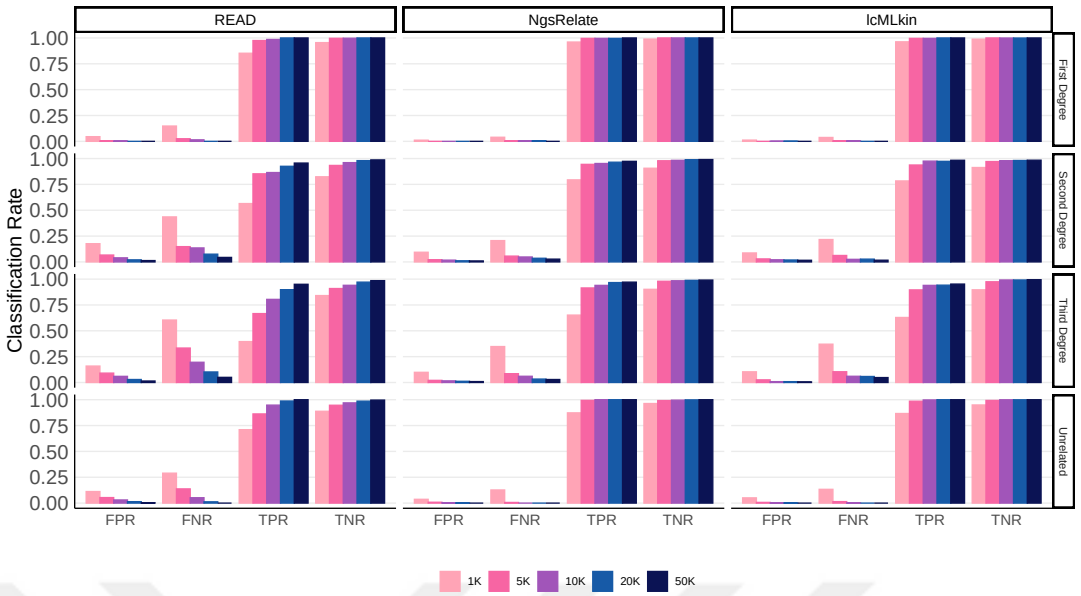


Figure 3.10: FPR, FNR, TPR, and TNR for READ, NgsRelate, and lcMLkin using arithmetic mean to classify kinship coefficient estimates of first-, second-, and third-degree related and unrelated pairs. The colors represent the count of SNPs shared between individuals.

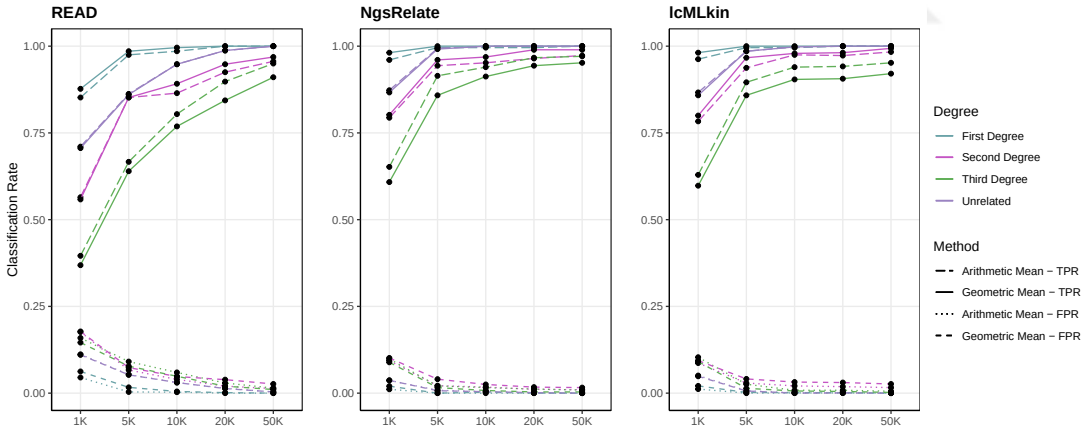


Figure 3.11: True positive rate (TPR, or sensitivity) and false positive rates (FPR, i.e., 1 – specificity) for READ, NgsRelate, and lcMLkin using two different classification methods, arithmetic and geometric mean. Colors refer to the relatedness degree and line types represent the classification method and classification metric.

3.6.6 F-score of Kinship Coefficient Estimates

Another classification metric, F-score (F_1), can also be used to test the overall accuracy power of READ, NgsRelate, and lcMLkin (see Section 2.6.1). In line with our previous observations, sharing a low number of SNPs between pairs caused a sharp decline in correct predictions by kinship estimation tools. Only first-degree simulations yielded above 80% F_1 for all software with 1K SNPs, but unrelated pairs' accuracy was above 80% only with NgsRelate and lcMLkin for 1K SNPs. We found that F_1 from READ for second-degree and third-degree relatives remained under 60%. Up to 20K SNPs, the power of the READ was worse than NgsRelate and lcMLkin for all the relatedness degrees. NgsRelate and lcMLkin estimates similarly showed high F_1 (>80%) except the second and third-degree estimates for 1K SNPs.

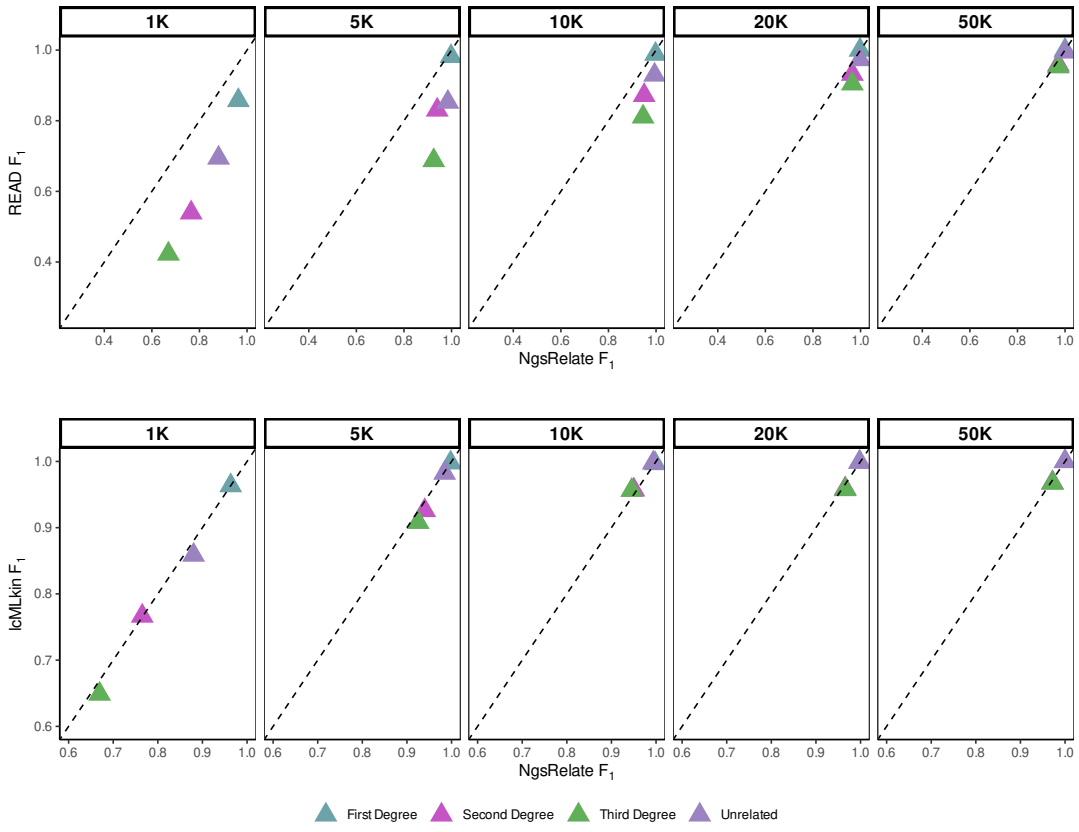


Figure 3.12: Pairwise F – score (F_1) of READ, NgsRelate, and lcMLkin for first-, second-, and third-degree related and unrelated pairs with 1K, 5K, 10K, 20K, and 50K shared SNPs. Colors refer to the degree of relatedness.

3.7 Kinship Inference In the Presence of Inbreeding

To illustrate how genetic relatedness estimation tools are affected by inbreeding in the samples analyzed, we also simulated relative pairs in the presence of inbreeding. We simulated first-cousin parents and then created parent-offspring and grandparent-grandchild pairs (Figure 3.6). While READ and lcMLkin are not specialized for inbreeding scenarios, NgsRelate version 2 has been specifically developed to infer relatedness even if the population is not outbred. It considers inbreeding using Jacquard coefficients representing IBD states of inbred individual pairs.

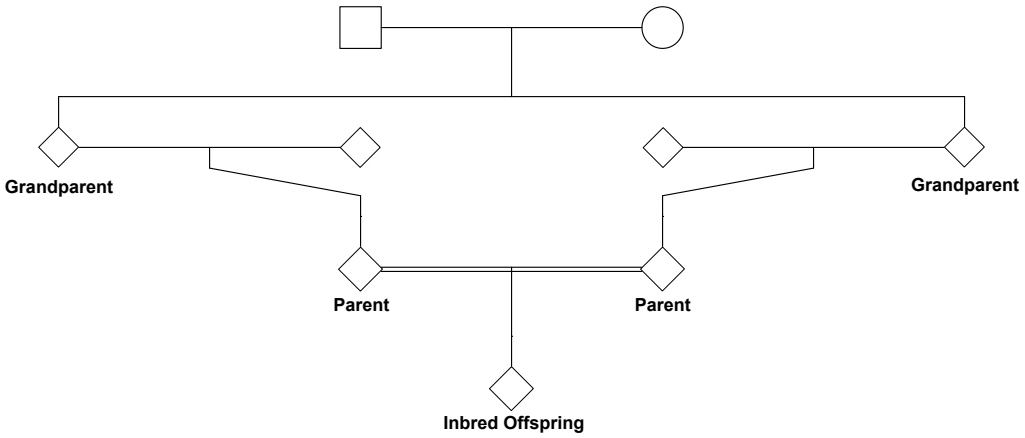


Figure 3.13: Inbreeding pedigree simulation. Grandparents of the offspring are simulated as siblings and parents are simulated as first cousins. The circles and squares represent the female and male individuals. Diamonds show an individual with unspecified sex. Double horizontal lines connecting two simulated individuals (parents) represent mating between relatives.

As NgsRelate estimates considering inbreeding, the formula (see formula 2.2 and Section 2.5.2) contains not only J_7 and J_8 but also J_3 and J_5 coefficients related to inbreeding. Therefore, when there is an inbreeding between pairs, θ estimates increase. To study whether NgsRelate accurately reveals the relationship type of pairs with one inbred individual, we calculated θ of the pairs eliminating inbreeding-related

coefficients like the one used in outbred populations, $\theta = J_7/2 + J_8/4$.

3.7.1 First-degree Relatedness With One Inbred Offspring

When we analyze the relationship between related parents (i.e., first cousins) and their child using NgsRelate, we found that it does not overestimate the coefficient from lowest SNP counts (1K) to highest (50K) (Figure 3.13). The mean of the estimated coefficients was slightly underestimated up to 50K SNPs ($\theta = 0.23$ for 1K SNPs and $\theta = 0.24$ for 5K, 10K, and 50K SNPs). Between variable SNP counts, the range of the θ values became narrower (Table 3.2). However, even with 50K SNP counts, the median of the θ values estimated by lcMLkin was 0.27, and by READ was 0.28. To evaluate whether these high values systematically produced significant bias for each tool, we tested their deviation from the equal probability of estimating above or below the expected threshold using a binomial test (see Appendix D). Due to the independence assumption among samples, we removed the repeated individuals (i.e., inbred children) from the analysis. We identified a persistently higher number of kinship coefficient values estimated by READ and lcMLkin above the threshold. The test showed that the deviation of θ estimates from the theoretical value of first-degree relatedness was significant for READ and lcMLkin. An increase in SNP counts between individuals was an ineffectual factor in eliminating the bias.

To statistically compare their impact to estimate kinship coefficients of relationships with inbred individuals, we performed a linear mixed model (see Appendix B). The kinship coefficient values were determined as the response variable. We also selected software type, and SNP counts as fixed effects and pairs as random effects for the model. The reference level was lcMLkin and 50K SNP count. We found that the difference in the means of the coefficient estimates from READ, NgsRelate, and lcMLkin was a significantly different, lower mean estimate for NgsRelate and a higher mean estimate for READ. For 10K, 20K, and 50K SNPs, READ and lcMLkin mean estimates were similar.

Alternatively, we contrasted their variation using Levene's test (see Appendix C). Up to 10K overlapping SNP numbers, the variances between NgsRelate, lcMLkin, and READ estimates were not significantly equal ($p < 0.044$). This points out that the vari-

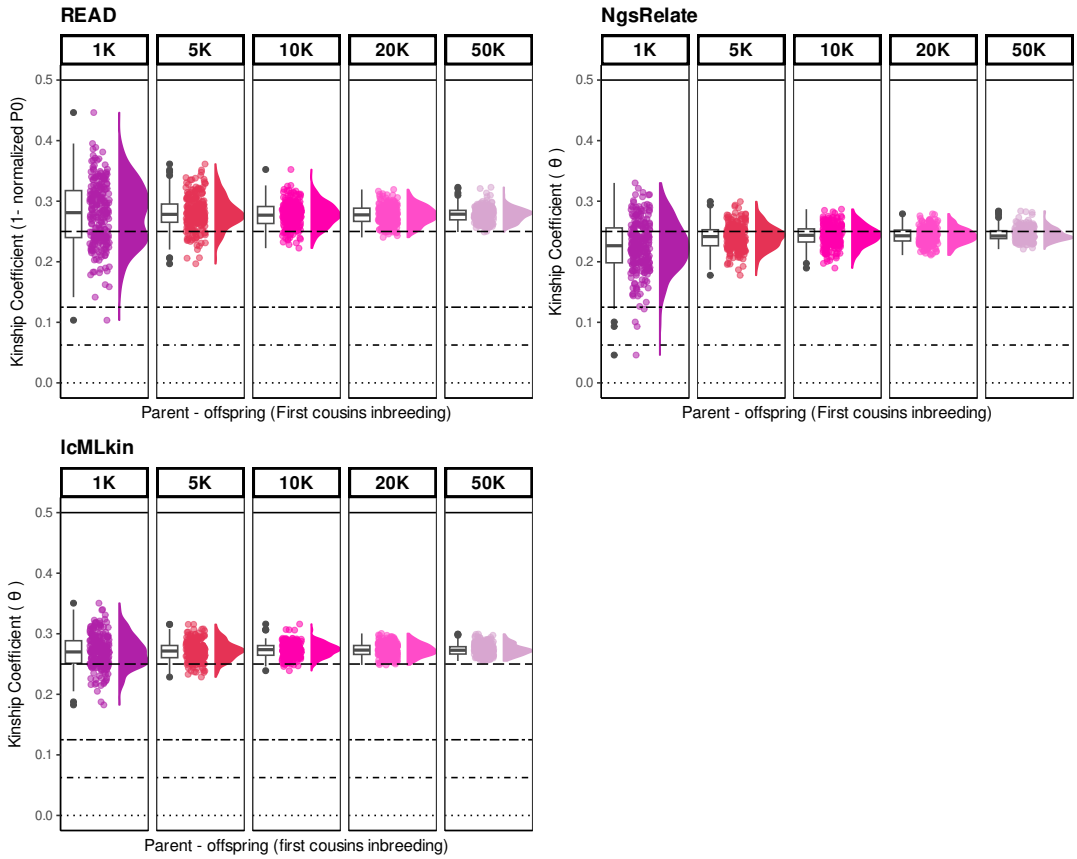


Figure 3.14: Kinship coefficient estimates for parent-offspring pairs in the presence of inbreeding by READ, NgsRelate, and lcMLkin. Horizontal lines show the theoretical kinship coefficients. The boxplots, jitter-added points, and density plots show the distribution of the same sample of 240 points. θ from NgsRelate was calculated ignoring the inbreeding-related Jacquard coefficients: $\theta = J_7/2 + J_8/4$.

ation differences between tools depend on the number of SNPs overlapped between individuals. We also tested their absolute mean differences using a linear mixed-effect model. The results showed that except for 50K SNPs, the estimates' differences from the mean for READ were significantly higher than others for lcMLkin and NgsRelate for all of the SNP counts.

3.7.2 Second-degree Relatedness With One Inbred Offspring

We further assessed their performance on the kinship coefficient estimation under an inbreeding scenario using a more distant relationship, grandparent-grandchild, in

Table 3.2: Kinship coefficient estimation summary of READ, NgsRelate, and lcMLkin for parent-offspring pairs in the presence of inbreeding. The total number of pairs is n=240 for each SNP count. Sd: standard deviation.

SNP Count	Software	Minimum	Maximum	Median	Mean	Sd
1K	READ	0.104	0.447	0.282	0.280	0.052
	NgsRelate	0.0459	0.330	0.226	0.226	0.0435
	lcMLkin	0.182	0.350	0.270	0.270	0.027
5K	READ	0.198	0.362	0.279	0.281	0.027
	NgsRelate	0.178	0.299	0.242	0.241	0.0199
	lcMLkin	0.228	0.316	0.272	0.271	0.016
10K	READ	0.223	0.353	0.277	0.278	0.021
	NgsRelate	0.190	0.287	0.244	0.243	0.0173
	lcMLkin	0.239	0.316	0.274	0.273	0.012
20K	READ	0.241	0.320	0.278	0.278	0.016
	NgsRelate	0.211	0.279	0.243	0.243	0.0132
	lcMLkin	0.248	0.300	0.273	0.273	0.011
50K	READ	0.249	0.323	0.279	0.278	0.012
	NgsRelate	0.221	0.284	0.243	0.245	0.011
	lcMLkin	0.255	0.300	0.273	0.274	0.010

which grandparents are simulated as siblings, and therefore, the grandchild's parents are simulated as first cousins. READ and lcMLkin estimated the coefficients above 0.19 for all SNP counts, while the estimates from NgsRelate were lower than 0.16 (Figure 3.14 and Table 3.3). The theoretical and estimated θ difference was above 0.03 for all tools. This overestimation was supported by a significantly higher number of estimates above the theoretical threshold across all SNP counts for lcMLkin (binomial test $p < 0.002$, see Appendix D). The significant majority of the estimates from READ and NgsRelate were also above the threshold ($p < 0.0003$ and $p < 0.007$, respectively), excluding the estimates from 1K SNPs. Indeed, the median estimates from READ were higher than the ones from NgsRelate and lcMLkin. According to the linear mixed effect analysis, NgsRelate has the significantly lowest mean estimate

(see Appendix B). Notably, the tools disagreed only on the variances of estimates from 1K and 5K SNP subsets (Levene's test $p < 0.04$, see Appendix C). Similar to absolute mean difference analysis with parents and their inbred children, READ significantly differed from other tools due to the high differences in their estimates from the mean.

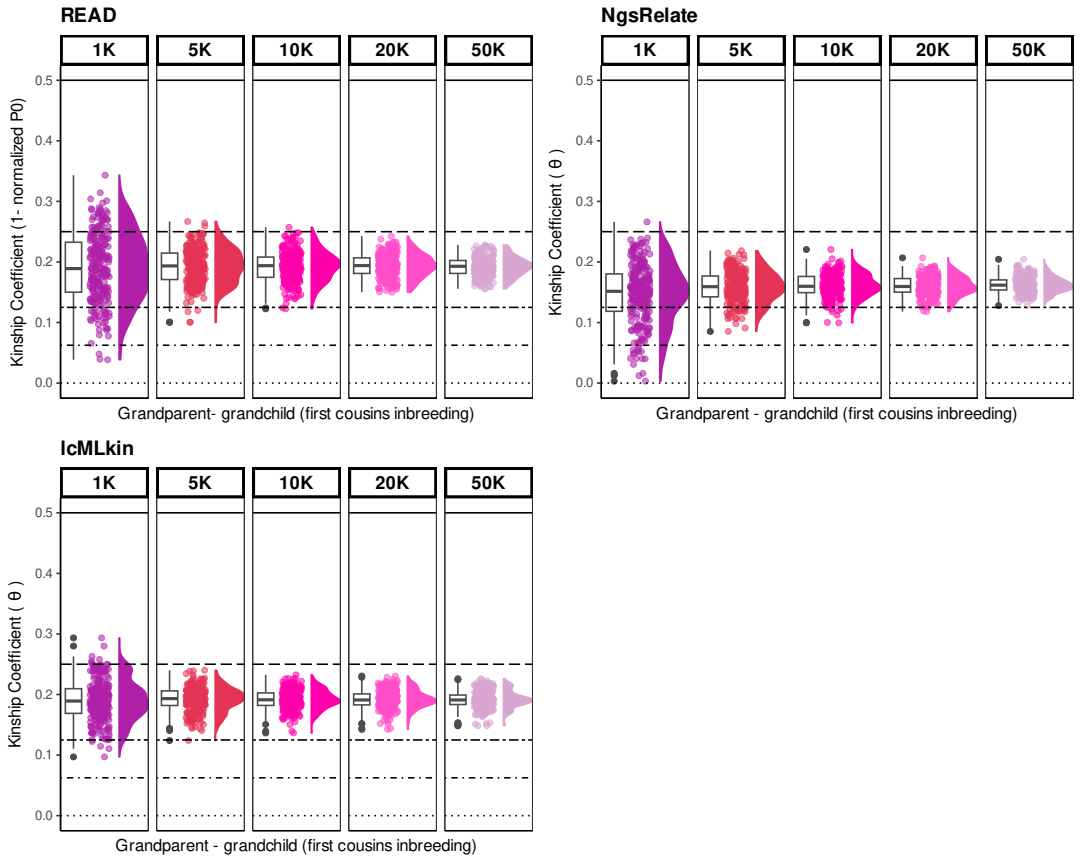


Figure 3.15: Kinship coefficient estimates for grandparent-grandchild pairs in the presence of inbreeding by READ, NgsRelate, and lcMLkin. Horizontal lines show the theoretical kinship coefficients. The boxplots, jitter-added points, and density plots show the distribution of the same sample of 240 points. θ from NgsRelate was calculated ignoring the inbreeding-related Jacquard coefficients: $\theta = J_7/2 + J_8/4$.

Alternatively, we investigated their classification accuracies in the presence of inbreeding using these simulated parent-offspring and grandparent-grandchild pairs. We used the same classification thresholds calculated from arithmetic means. We observed that all the pairwise estimates of parent-offsprings pairs from READ and

Table 3.3: Kinship coefficient estimation summary of READ, NgsRelate, and lcMLkin for parent-offspring pairs in the presence of inbreeding. The total number of pairs is n=240 for each SNP count. Sd: standard deviation.

SNP Count	Software	Minimum	Maximum	Median	Mean	Sd
1K	READ	0.039	0.343	0.189	0.191	0.057
	NgsRelate	0.00316	0.266	0.152	0.147	0.0499
	lcMLkin	0.097	0.294	0.189	0.190	0.034
5K	READ	0.100	0.267	0.194	0.193	0.030
	NgsRelate	0.0853	0.219	0.159	0.158	0.0254
	lcMLkin	0.124	0.240	0.193	0.192	0.020
10K	READ	0.123	0.257	0.194	0.193	0.025
	NgsRelate	0.0994	0.221	0.160	0.161	0.0193
	lcMLkin	0.136	0.232	0.191	0.192	0.016
20K	READ	0.150	0.243	0.194	0.194	0.019
	NgsRelate	0.118	0.207	0.160	0.160	0.0163
	lcMLkin	0.142	0.230	0.191	0.192	0.015
50K	READ	0.156	0.228	0.193	0.192	0.016
	NgsRelate	0.128	0.204	0.162	0.162	0.0129
	lcMLkin	0.148	0.226	0.191	0.191	0.015

NgsRelate sharing 5K to 50K SNPs were accurately identified. Similarly, except for one pair sharing 1K SNPs with a kinship coefficient value of 0.1825, all the estimates from lcMLkin were assigned to first-degree category. In contrast, the relationship between grandparents and their inbred grandchild could not be identified reliably by READ and lcMLkin, even with the highest number of SNPs. Most of the estimates of READ (60%) and of lcMLkin (60%) exceeded the expected second-degree range. However, NgsRelate accurately identifies grandparent-grandchild pairs from 1K to 50K SNPs with increasing frequencies, 61%- 96%. Except for estimates with 1K SNPs by READ and NgsRelate, all the false negatives were caused by the pairs misclassified as first-degree. At 1K SNPs, 3% and 18% of misclassified pairs were inferred as third-degree related by READ and NgsRelate, respectively.

3.7.3 Overview of Inbreeding Results

Overall, inbreeding causes a significant overestimation of the kinship coefficient. This is expected as recent inbreeding increases the probability of two alleles for a locus being IBD transmitted to the inbred individual. Even though NgsRelate version 2 is specialized for inbreeding scenarios, we found that it is still biased when the parents are third-degree related to each other. Using a MAF cutoff above 1% and allele frequencies from the founders themselves did not guarantee unbiased estimates for NgsRelate. While NgsRelate performed relatively well and reliably infer the relationships of first-degree pairs with one inbred child, their reliability substantially decreased if the target relationship included an inbred grandchild whose parents are first cousins.

CHAPTER 4

DISCUSSION

In this study, we tested the performances of genetic kin detection software, READ, NgsRelate, and lcMLkin, the most preferred tools in ancient DNA studies, to compare their reliabilities on kinship coefficient estimations in a wide range of shared SNPs (from 1K to 50K) using simulated pedigrees. Recently, a study examined the tools' effectiveness, with varying coverage levels (Marsh et al., 2023). Despite an increase in false positive rates, they observed that READ is capable of functioning efficiently at 0.05x coverage. Conversely, their results showed that both NgsRelate and lcMLkin exhibit diminished performance levels below 1x coverage. However, no study has systematically explored their performance based on different relationship types (from first- to third-degree relatedness) with or without inbreeding using simulated using sparse ancient sequencing data. Firstly, we systematically analyzed whether their kinship coefficient estimates from READ, NgsRelate, and lcMLkin meet the theoretical expectations or deviate from these using a linear mixed effect model and binomial test for each simulated pedigree. Secondly, we investigated the variances of kinship coefficient estimates between simulated relationships or shared SNP counts using Levene's test. Thirdly, we compared their power (i.e., TPR, TNR, FPR, FNR, and F1) based on arithmetic and geometric mean. Finally, we explored their response to inbreeding using simulated first-degree and second-degree related pairs with one inbred individual (first-cousin mating). In brief, we evaluated which tool is most reliable in a variety of conditions for revealing genetic kin between low-coverage ancient genomes based on these results.

Kinship coefficient estimate distributions of simulated parent-offspring pairs showed that the variation of kinship coefficient estimates gradually decreased from the low-

est SNP count (1K) to the highest SNP count (50K) for all three tools, as expected. READ estimates for parent-offspring pairs varied more than others, a pattern supported by Levene's test results showing significant variance differences among tools for each SNP count. However, there were no significant differences in estimate variances among the tools for sibling pairs (Table 3.1).

We observed higher variation for all SNP counts between sibling pairs than parent-offspring pairs. This pattern is expected as genome-wide IBD sharing varies for every sibling pair. They share 0, 1, or 2 alleles IBD across loci in different proportions, represented with, on average, 0.5 k_1, 0.25 k_0, and k_2 (see Table 1.1). Therefore, the observed variation can be explained by biological variation.

According to the linear mixed effect model (see Section 3.3.1 and Figure 3.5), the tool converging most towards the expected θ estimates for first-degree relatedness was lcMLkin. The results showed that analysis with NgsRelate and READ had significantly lower mean θ values compared to analysis with lcMLkin for sibling pairs. This underestimation is mostly observed when the pairs sharing 50K SNPs and 1K SNPs were used with NgsRelate and READ, respectively. Similarly, NgsRelate and READ estimated the θ of parent-offspring pairs lower than lcMLkin on average, but only NgsRelate estimates were significantly different. This significant decrease again was caused by the estimates from pairs sharing 50K SNPs. We would expect slight variations between θ estimates due to sampling error or recombination, but there is a systematic decrease with 50K SNPs when we used NgsRelate. The underlying reason could be PMD or sequencing or mapping errors. PMD profiles of ancient genomes (i.e., C to T at 5' end and G to A transitions at 3' end of the reads) or errors introduced during sequencing or mapping processes can make the genetically related genomes dissimilar, so eventually, their shared IBD proportions could be lower than expected. Before conducting a kinship analysis, it is recommended to filter sites with transitions to determine if PMD is responsible for underestimation. However, this can reduce the number of shared SNPs among individuals, so it is necessary to further investigate its potential impact on kinship coefficient estimations.

With the linear mixed effect model, we also tested whether the variation among θ estimates is affected by the individual pairs used. For all the relatedness types, we

found that θ estimates vary by pairs. This could be the direct or indirect outcome of (i) the random nature of recombination and Mendelian inheritance (meiosis) and (ii) the linkage among SNPs. These factors can increase the biological noise in the data for θ estimation. For instance, SNPs in close proximity are not independent of each other when they are in linkage disequilibrium (LD), more likely to be transmitted together to the next generations. This can create higher IBD proportions between pairs.

In contrast to the significant influence of pairs used on θ estimates repeated random subsampling of shared SNPs between individual pairs (n=5) did not significantly affect the θ estimates, revealed by the linear mixed effect model. Their estimated variation was equal to zero. This result suggests that the different sets of SNPs with MAF > 0.01 give approximately the same information related to IBD sharing.

Except for the estimates between half-sibling pairs, when the second-degree related pairs were analyzed using READ and NgsRelate, their mean estimates were significantly lower than the ones using lcMLkin. The highest decline in the mean estimates was observed using 50K SNPs for NgsRelate and READ, and 1K SNPs for READ. However, these estimates were not significantly different than the ones using other SNP counts based on the reference level, and analysis with 50K SNPs using lcMLkin. These results were also in line with the binomial test, indicating no significant deviation in θ estimates from expected values for any second-degree relatedness type for any SNP count (see Appendix D). Therefore, we can suggest that variable SNP counts from low to high do not have a significant effect on mean θ estimates.

Like first-degree relatedness, for second-degree relatedness, the linear mixed effect model suggests that the most accurate tool in terms of its mean estimate is lcMLkin. However, the model predicts the mean estimate of grandparent-grandchild pairs for all three software was lower than the theoretical expectation (intercept θ = 0.1203). Furthermore, the mean estimates of half-sibling pairs were not significantly different than the expected value for any of the tools (intercept θ = 0.1244). Given all the estimates from second-degree related pairs, we can summarize that the mean θ estimates are ranked from highest to lowest: avuncular > half-siblings > grandparent-grandchild for lcMLkin mean estimates, but half-siblings > avuncular > grandparent-grandchild

for READ and NgsRelate. Two possible factors can influence these changes in the θ estimates of second-degree relatedness: the number of (i) meiosis; and (ii) shared ancestors between individuals. Half-sibling and grandparent-grandchild pairs are separated by two meioses. Also, the number of recent common ancestors for half-siblings and grandparent-grandchild pairs is one (common parent) and zero, respectively. However, avuncular pairs are separated by three meiosis, and they share two recent ancestors (grandparents of the younger individual), which eventually creates shorter but more IBD segments. The significant bias towards lower values can be an indirect product of these factors, but their influence was not clear given the inconsistency of the pattern of mean θ estimates among tools and the random nature of meiosis.

Analyzing third-degree relatedness using a linear mixed effect model suggests that among kinship tools, θ estimates of lcMLkin were the closest to the theoretical θ with first cousins and grand avuncular pairs (see Section 3.3.3 and Figure 3.7). Unlike first- or second-degree relatedness results, lcMLkin and NgsRelate were indistinguishable in their mean θ estimates for these third-degree related pairs. However, READ estimates were significantly lower than the others.

These variance differences among kinship estimation tools, especially higher variance seen with READ, affect their classification rates for first-, second-, and third-degree and unrelated pairs. As expected, the accuracies of the three software increased dramatically from 1K shared SNPs to 50K shared SNPs, reflected by higher TPR and TNR and lower FPR and FNR (see Figure 3.10). Furthermore, first-degree related and unrelated pairs provide the highest values for all the tools in terms of their specificity and sensitivity. Even with 1K SNPs, NgsRelate and lcMLkin still perform well with 96% TPR and 1% FPR, but this was not the case for READ which performed much worse. For each SNP count, all the tools wrongly assigned the minority of the second-and third-degree related pairs as third-degree and second-degree or unrelated, respectively. Regardless of which tool is used, accurate assignment of second-and third-degree related pairs was possible when using more than 1K and 5K SNPs, respectively.

To compare the overall power of the kinship tools based on their precision and re-

call, we calculated F1 values for each relatedness degree and SNP count (Figure 3.12). According to F1 values, the power of lcMLkin and NgsRelate to correctly infer relatedness of all kinds is similar. Their F1 scores were higher than 90% for >5K SNPs. Only with 1K SNPs and third-degree relatedness, the F1 values of lcMLkin and NgRelate decreased down to 65%. Notably, we observed that lcMLkin and NgsRelate outperformed READ for all degrees and SNP counts less than 20K SNPs. Meanwhile, even though READ is generally used in ancient DNA studies to infer relatedness up to second-degree, we confirmed that its discriminative power for third-degree relatedness also reached 80% F1 beyond 5K shared SNPs.

We also performed classification with geometric mean rather than arithmetic mean. Using a lower cutoff for classification with geometric mean slightly increased the sensitivity and decreases the specificity for first- and second-degree relatedness but reciprocally affected the third-degree relatedness (Figure 3.11). The estimates of simulated third-degree related pairs were assigned as second-degree due to the lower threshold. Therefore, we do not recommend applying classification cutoff based on the geometric mean of theoretical θ when the goal is to discriminate the second- and third-degree relatedness.

In the presence of inbreeding, sharing more alleles IBD, θ of an individual pair increases because the Jacquard coefficients corresponding to the six different allelic IBD states are no longer equal to zero (see Section 1.1). NgsRelate version 2 gives θ estimates by incorporating these extra coefficients (see Section 2.5.2) To explore the sensitivity of these estimates, we calculated θ only using J_7 and J_8 estimates of NgsRelate, which are not related to inbreeding ($\theta = J_7/2 + J_8/4$). For the relationship between parents (first cousins) and their inbred offspring pairs, θ estimates were close to the theoretical value of first-degree relatedness in non-inbred populations (mean θ = 0.24). Thus, we can conclude that NgsRelate estimates θ based on J_7 and J_8 without any bias, which helps to distinguish these pairs as first-degree related. In contrast, with the more complex scenario (i.e., grandparents are siblings) and lower degree of relatedness (grandparent-grandchild), NgsRelate tended to overestimate the mean θ by 0.03.

Above, I discussed that READ performs worse than the other two tools in accuracy,

especially when using low SNP counts. There is a possibility that the usage of the allele frequencies from the founders contributes more to the higher power and lower variance of NgsRelate and lcMLkin compared to READ. In our simulation, allele frequencies of our background population (TSI) perfectly match the frequencies of the sampled population. As the performances of maximum likelihood-based methods are highly dependent on allele frequency estimates, the higher sensitivity and specificity of the θ estimates from NgsRelate and lcMLkin could be the direct reflection of this perfect background. For this study, we did not explore whether noise in population allele frequency estimates or the use of estimates from a proxy population might affect the results of NgsRelate and lcMLkin. This is an open question and requires further research.

Similarly, we used SNPs with 0.01 or higher MAF in the reference population. Ignoring less represented alleles with low frequencies in the sampled population (e.g., MAF > 0.05) could potentially increase accuracy as the infrequent alleles could actually be the result of sequencing or mapping errors. Again, this characteristic of our simulations could contribute to the unbiased estimations of the tools in most of the analyses with varying SNP counts. Given that, applying higher thresholds for MAF (5% or more) in kinship detection may improve the estimations up to some level. However, this would also lower power, and whether an optimal solution exists has not been explored yet, but would be beneficial to study in the future.

Another feature of our pedigree simulations is that the genomes of the founders do not show background relatedness, i.e., shared short IBD segments between individuals of the particular population transmitted through distant common ancestors. Therefore, when applied to real ancient samples, the results of the θ estimates from READ, NgsRelate, and lcMLkin, can be different from what we observed in this study. For this reason, we recommend that the same analysis could be extended to include simulations mimicking the biological background of the populations directly. We suggest referring to my colleague Merve Nur Güler's master's thesis for a more in-depth examination of realistic population simulation with background relatedness.

REFERENCES

Altınışık, N. E., Kazancı, D. D., Aydoğan, A., Gemici, H. C., Erdal, Ö. D., Sarialtun, S., Vural, K. B., Koptekin, D., Gürün, K., Sağlıcan, E., Fernandes, D., Çakan, G., Koruyucu, M. M., Lagerholm, V. K., Karamurat, C., Özkan, M., Kılınç, G. M., Sevkar, A., Sürer, E., … Somel, M. (2022). A genomic snapshot of demographic and cultural dynamism in Upper Mesopotamia during the Neolithic Transition. Science Advances, 8(44), 3609. https://doi.org/10.1126/sciadv.abo3609

Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Bentley, D. R., Chakravarti, A., Clark, A. G., Donnelly, P., Eichler, E. E., Flicek, P., Gabriel, S. B., Gibbs, R. A., Green, E. D., Hurles, M. E., Knoppers, B. M., Korbel, J. O., Lander, E. S., Lee, C., Lehrach, H., … Schloss, J. A. (2015). A global reference for human genetic variation. Nature 2015 526:7571, 526(7571), 68–74. https://doi.org/10.1038/nature15393

Bherer, C., Campbell, C. L., & Auton, A. (2017). Refined genetic maps reveal sexual dimorphism in human meiotic recombination at multiple scales. Nature Communications 2017 8:1, 8(1), 1–9. https://doi.org/10.1038/ncomms14994

Briggs, A. W., Stenzel, U., Johnson, P. L. F., Green, R. E., Kelso, J., Prüfer, K., Meyer, M., Krause, J., Ronan, M. T., Lachmann, M., & Pääbo, S. (2007). Patterns of damage in genomic DNA sequences from a Neandertal. Proceedings of the National Academy of Sciences of the United States of America, 104(37), 14616–14621. https://doi.org/10.1073/PNAS.0704665104

Brustad, H. K., Vigeland, M. D., & Egeland, T. (2021). Pairwise relatedness testing in the context of inbreeding: expectation and variance of the likelihood ratio. International Journal of Legal Medicine, 135(1), 117–129.

Caballero, M., Seidman, D. N., Qiao, Y., Sannerud, J., Dyer, T. D., Lehman, D. M., Curran, J. E., Duggirala, R., Blangero, J., & Carmi, S. (2019). Crossover interference

and sex-specific genetic maps shape identical by descent sharing in close relatives. PLoS Genetics, 15(12), e1007979–e1007979.

Cotterman, C. W. (1941). Relatives and human genetic analysis. The Scientific Monthly, 53(3), 227–234. 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. (2011). The variant call format and VCFtools. Bioinformatics, 27(15), 2156–2158. https://doi.org/10.1093/BIOINFORMATICS/BTR330

Dou, J., Sun, B., Sim, X., Hughes, J. D., Reilly, D. F., Tai, E. S., Liu, J., & Wang, C. (2017). Estimation of kinship coefficient in structured and admixed populations using sparse sequencing data. PLOS Genetics, 13(9), e1007021–e1007021. https://doi.org/10.1371/JOURNAL.PGEN.1007021

Fernandes, D. M., Cheronet, O., Gelabert, P., & Pinhasi, R. (2021). TKGWV2: an ancient DNA relatedness pipeline for ultra-low coverage whole genome shotgun data. Scientific Reports 2021 11:1, 11(1), 1–9. https://doi.org/10.1038/s41598-021-00581-3

Furtwängler, A., Rohrlach, A. B., Lamnidis, T. C., Papac, L., Neumann, G. U., Siebke, I., Reiter, E., Steuri, N., Hald, J., Denaire, A., Schnitzler, B., Wahl, J., Ramstein, M., Schuenemann, V. J., Stockhammer, P. W., Hafner, A., Lösch, S., Haak, W., Schiffels, S., & Krause, J. (2020). Ancient genomes reveal social and genetic structure of Late Neolithic Switzerland. Nature Communications 2020 11:1, 11(1), 1–11. https://doi.org/10.1038/s41467-020-15560-x

Goudet, J., Kay, T., & Weir, B. S. (2018). How to estimate kinship. Molecular Ecology, 27(20), 4121–4135. https://doi.org/10.1111/MEC.14833

Gusev, A., Lowe, J. K., Stoffel, M., Daly, M. J., Altshuler, D., Breslow, J. L., Friedman, J. M., & Pe'er, I. (2009). Whole population, genome-wide mapping of hidden relatedness. Genome Research, 19(2), 318–326.

Günther, T., & Nettelblad, C. (2019). The presence and impact of reference bias on population genomic studies of prehistoric human populations. PLOS Genetics, 15(7), e1008302. https://doi.org/10.1371/JOURNAL.PGEN.1008302

Hanghøj, K., Moltke, I., Andersen, P. A., Manica, A., & Korneliussen, T. S. (2019). Fast and accurate relatedness estimation from high-throughput sequencing data in the presence of inbreeding. GigaScience, 8(5), 1–9. https://doi.org/10.1093/GIGASCIENCE/GIZ034

Hedrick, P. P. W., Lacy, R. C., & Baker, C. S. (2015). Measuring Relatedness between Inbred Individuals. Journal of Heredity, 106(1), 20–25. https://doi.org/10.1093/JHERED/ESU072

Housworth, E. A., & Stahl, F. W. (2003). Crossover interference in humans. American Journal of Human Genetics, 73(1), 188–197. https://doi.org/10.1086/376610

Huff, C. D., Witherspoon, D. J., Simonson, T. S., Xing, J., Watkins, W. S., Zhang, Y., Tuohy, T. M., Neklason, D. W., Burt, R. W., Guthery, S. L., Woodward, S. R., & Jorde, L. B. (2011). Maximum-likelihood estimation of recent shared ancestry (ERSA). Genome Research, 21(5), 768–774. https://doi.org/10.1101/gr.115972.110

Jacquard, A. (1972). The genotypic distribution of relatives when consanguinity is present. Annals of Human Genetics, 36(2), 233–236. https://doi.org/10.1111/J.1469-1809.1972.TB00772.X

Kjær, K. H., Winther Pedersen, M., De Sanctis, B., De Cahsan, B., Korneliussen, T. S., Michelsen, C. S., Sand, K. K., Jelavić, S., Ruter, A. H., & Schmidt, A. (2022). A 2-million-year-old ecosystem in Greenland uncovered by environmental DNA. Nature, 612(7939), 283–291.

Konovalov, D. A., & Heg, D. (2008). TECHNICAL ADVANCES: A maximum-likelihood relatedness estimator allowing for negative relatedness values. Molecular Ecology Resources, 8(2), 256–263. https://doi.org/10.1111/J.1471-8286.2007.01940.X

Koptekin, D., Yüncü, E., Rodríguez-Varela, R., Altınışık, N. E., Psonis, N., Kashuba, N., Yorulmaz, S., George, R., Kazancı, D. D., Kaptan, D., Gürün, K., Vural, K. B., Gemici, H. C., Vassou, D., Daskalaki, E., Karamurat, C., Lagerholm, V. K., Erdal, Ö. D., Kırdök, E., … Somel, M. (2023). Spatial and temporal heterogeneity in human mobility patterns in Holocene Southwest Asia and the East Mediterranean. Current

Biology: CB, 33(1), 41-57.e15. https://doi.org/10.1016/J.CUB.2022.11.034

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics, 15(1), 1–13. https://doi.org/10.1186/S12859-014-0356-4/TABLES/4

Korneliussen, T. S., & Moltke, I. (2015). NgsRelate: a software tool for estimating pairwise relatedness from next-generation sequencing data. Bioinformatics, 31(24), 4009–4011. https://doi.org/10.1093/BIOINFORMATICS/BTV509

Kuhn, J. M. M., Jakobsson, M., & Günther, T. (2018). Estimating genetic kin relationships in prehistoric populations. PLOS ONE, 13(4), e0195491. https://doi.org/10.1371/JOURNAL.PONE.0195491

Kuhn, & Max. (2008). Building Predictive Models in R Using the caret Package. Journal of Statistical Software, 28(5), 1–26. https://doi.org/10.18637/jss.v028.i05

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics (Oxford, England), 27(21), 2987–2993. https://doi.org/10.1093/BIOINFORMATICS/BTR509

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics, 25(14), 1754–1760. https://doi.org/10.1093/BIOINFORMATICS/BTP324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics (Oxford, England), 25(16), 2078–2079. https://doi.org/10.1093/BIOINFORMATICS/BTP352

Lipatov, M., Sanjeev, K., Patro, R., & Veeramah, K. R. (2015). Maximum Likelihood Estimation of Biological Relatedness from Low Coverage Sequencing Data. https://doi.org/10.1101/023374

Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., & Chen, W.-M. (2010). Robust relationship inference in genome-wide association studies. Bioinfor-

matics, 26(22), 2867–2873.

Marsh, W. A., Brace, S., & Barnes, I. (2023). Inferring biological kinship in ancient datasets: comparing the response of ancient DNA-specific software packages to low coverage data. BMC Genomics, 24(1). https://doi.org/10.1186/S12864-023-09198-4

Mittnik, A., Massy, K., Knipper, C., Wittenborn, F., Friedrich, R., Pfrengle, S., Burri, M., Carlichi-Witjes, N., Deeg, H., Furtwängler, A., Harbeck, M., von Heyking, K., Kociumaka, C., Kucukkalipci, I., Lindauer, S., Metz, S., Staskiewicz, A., Thiel, A., Wahl, J., … Krause, J. (2019). Kinship-based social inequality in Bronze Age Europe. Science, 366(6466), 731–734. https://doi.org/10.1126/SCIENCE.AAX6219

Moltke, I., & Albrechtsen, A. (2014). RelateAdmix: a software tool for estimating relatedness between admixed individuals. Bioinformatics, 30(7), 1027–1028.

Morgan, L. H. (1871). Systems of Consanguinity and affinity of the human family (Issue 218). Smithsonian Institution.

Nielsen, R., Paul, J. S., Albrechtsen, A., & Song, Y. S. (2011). Genotype and SNP calling from next-generation sequencing data. Nature Reviews. Genetics, 12(6), 443–451. https://doi.org/10.1038/NRG2986

Nøhr, A. K., Hanghøj, K., Garcia-Erill, G., Li, Z., Moltke, I., & Albrechtsen, A. (2021). NGSremix: a software tool for estimating pairwise relatedness between admixed individuals from next-generation sequencing data.
G3 Genes|Genomes|Gentics, 11(8), 1–8.
https://doi.org/10.1093/G3JOURNAL/JKAB174

Oliehoek, P. A., Windig, J. J., Van Arendonk, J. A. M., & Bijma, P. (2006). Estimating Relatedness Between Individuals in General Populations With a Focus on Their Use in Conservation Programs. Genetics, 173(1), 483.
https://doi.org/10.1534/GENETICS.105.049940

Orlando, L., Allaby, R., Skoglund, P., Der Sarkissian, C., Stockhammer, P. W., Ávila-Arcos, M. C., Fu, Q., Krause, J., Willerslev, E., & Stone, A. C. (2021). Ancient DNA analysis. Nature Reviews Methods Primers, 1(1), 1–26.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., & Reich, D. (2012). Ancient admixture in human history. Genetics, 192(3), 1065–1093. https://doi.org/10.1534/GENETICS.112.145037/-/DC1

Peyrégne, S., & Prüfer, K. (2020). Present-Day DNA Contamination in Ancient DNA Datasets. Bioessays, 42(9), 2000081.

Picard Toolkit. (2019). GitHub repository. Broad Institute. Available Online at: Http://Broadinstitute. Github. Io/Picard.

Popli, D., Peyrégne, S., & Peter, B. M. (2023). KIN: a method to infer relatedness from low-coverage ancient DNA. Genome Biology, 24(1), 1–22. https://doi.org/10.1186/S13059-023-02847-7/FIGURES/6

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., De Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics, 81(3), 559–575. https://doi.org/10.1086/519795

Queller, D. C., & Goodnight, K. F. (1989). Estimating Relatedness Using Genetic Markers. Evolution, 43(2), 258–275. https://doi.org/10.1111/J.1558-5646.1989.TB04226.X

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics, 26(6), 841–842. https://doi.org/10.1093/BIOINFORMATICS/BTQ033

R Core Team. (2022). R: A Language and Environment for Statistical Computing. https://www.R-project.org/

Renaud, G., Hanghøj, K., Willerslev, E., & Orlando, L. (2017). gargammel: a sequence simulator for ancient DNA. Bioinformatics (Oxford, England), 33(4), 577–579. https://doi.org/10.1093/BIOINFORMATICS/BTW670

Rivers, W. H. R. (2012). Kinship and Social Organisation. Routledge Revivals. https://doi.org/10.4324/9780203814659

Rizzi, E., Lari, M., Gigli, E., De Bellis, G., & Caramelli, D. (2012). Ancient DNA

studies: new perspectives on old samples. Genetics Selection Evolution, 44(1), 1–19.

Rubinacci, S., Ribeiro, D. M., Hofmeister, R. J., & Delaneau, O. (2021). Efficient phasing and imputation of low-coverage sequencing data using large reference panels. Nature Genetics 2021 53:1, 53(1), 120–126. https://doi.org/10.1038/s41588-020-00756-0

Schroeder, H., Margaryan, A., Szmyt, M., Theulot, B., Włodarczak, P., Rasmussen, S., Gopalakrishnan, S., Szczepanek, A., Konopka, T., Jensen, T. Z. T., Witkowska, B., Wilk, S., Przybyła, M. M., Pospieszny, Ł., Sjögren, K. G., Belka, Z., Olsen, J., Kristiansen, K., Willerslev, E., … Allentoft, M. E. (2019). Unraveling ancestry, kinship, and violence in a Late Neolithic mass grave. Proceedings of the National Academy of Sciences of the United States of America, 166(22), 10705–10710. https://doi.org/10.1073/PNAS.1820210116

Schubert, M., Lindgreen, S., & Orlando, L. (2016). AdapterRemoval v2: Rapid adapter trimming, identification, and read merging. BMC Research Notes, 9(1). https://doi.org/10.1186/s13104-016-1900-2

Seguin-Orlando, A. (2021). Statistical methods for kinship inference amongst ancient individuals. In H. Reyes - Centeno & K. Harvati (Eds.), Ancient Connections in Eurasia (pp. 101–127). Tübingen: Kerns Verlag.
https://doi.org/10.51315/9783935751377.006

Skoglund, P., Northoff, B. H., Shunkov, M. V., Derevianko, A. P., Pääbo, S., Krause, J., & Jakobsson, M. (2014). Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. Proceedings of the National Academy of Sciences of the United States of America, 111(6), 2229–2234.
https://doi.org/10.1073/PNAS.1318934111

Speed, D., & Balding, D. J. (2015). Relatedness in the post-genomic era: is it still useful? Nature Reviews Genetics, 16(1), 33–44.

Thornton, T., Tang, H., Hoffmann, T. J., Ochs-Balcom, H. M., Caan, B. J., & Risch, N. (2012). Estimating kinship in admixed populations. American Journal of Human Genetics, 91(1), 122–138. https://doi.org/10.1016/J.AJHG.2012.05.024

Vai, S., Amorim, C. E. G., Lari, M., & Caramelli, D. (2020). Kinship Determination in Archeological Contexts Through DNA Analysis. Frontiers in Ecology and Evolution, 8, 513005. https://doi.org/10.3389/FEVO.2020.00083/BIBTEX

Villalba-Mouco, V., van de Loosdrecht, M. S., Rohrlach, A. B., Fewlass, H., Talamo, S., Yu, H., Aron, F., Lalueza-Fox, C., Cabello, L., Cantalejo Duarte, P., Ramos-Muñoz, J., Posth, C., Krause, J., Weniger, G. C., & Haak, W. (2023). A 23,000-year-old southern Iberian individual links human groups that lived in Western Europe before and after the Last Glacial Maximum. Nature Ecology & Evolution 2023 7:4, 7(4), 597–609. https://doi.org/10.1038/s41559-023-01987-0

Wang, K., Bleasdale, M., Le Moyne, C., Freund, C., Krause, J., Boivin, N., & Schiffels, S. (2022). 4000-year-old hair from the Middle Nile highlights unusual ancient DNA degradation pattern and a potential source of early eastern Africa pastoralists. Scientific Reports, 12(1), 1–9.

Waples, R. K., Albrechtsen, A., & Moltke, I. (2019). Allele frequency-free inference of close familial relationships from genotypes or low-depth sequencing data. Molecular Ecology, 28(1), 35–48. https://doi.org/10.1111/MEC.14954

Ward, J. M., & Farris, S. (2023). splicejam: Analysis and Visualization of Gene Splice Variants and Transcriptome Data. http://github.com/jmw86069/splicejam

Weir, B. S., Anderson, A. D., & Hepler, A. B. (2006). Genetic relatedness analysis: modern data and new challenges. Nature Reviews. Genetics, 7(10), 771–780. https://doi.org/10.1038/NRG1960

Yaka, R., Mapelli, I., Kaptan, D., Doğu, A., Chyleński, M., Erdal, Ö. D., Koptekin, D., Vural, K. B., Bayliss, A., Mazzucato, C., Fer, E., Çokoğlu, S. S., Lagerholm, V. K., Krzewińska, M., Karamurat, C., Gemici, H. C., Sevkar, A., Dağtaş, N. D., Kılınç, G. M., … Somel, M. (2021). Variable kinship patterns in Neolithic Anatolia revealed by ancient genomes. Current Biology, 31(11), 2455-2468.e18. https://doi.org/10.1016/J.CUB.2021.03.050

Žegarac, A., Winkelbach, L., Blöcher, J., Diekmann, Y., Krečković Gavrilović, M., Porčić, M., Stojković, B., Milašinović, L., Schreiber, M., Wegmann, D., Veeramah, K.

R., Stefanović, S., & Burger, J. (2021). Ancient genomes provide insights into family structure and the heredity of social status in the early Bronze Age of southeastern Europe. Scientific Reports, 11(1). https://doi.org/10.1038/S41598-021-89090-X

APPENDIX A

BASH SCRIPT TO PRODUCE BINARY GL FILE

The script provided below takes the text file ($angsd.beagle.gz$) of GLs generated by $ANGSD$ and convert it to binary file ($angsd.glf.gz$) which is used later for the analysis with NgsRelate.

```
#!/bin/bash
zcat angsd.beagle.gz |
tail -n +2 |
perl -an -e 'for($i=3; $i<=$#F; $i++)print(pack("d",($F[$i]==0 ? -inf : log($F[$i]))))'
> angsd.glf.gz
```

APPENDIX B

LMM RESULTS

Table B.1: Variances of random effects and residuals estimated using LMM for each relationship type.

Model 1 Random Effects									
	PO	**S**	**HS**	**AV**	**GPC**	**FC**	**GAV**	**GGPC**	**U**
Pairs	<0.0001	0.0005	0.0002	0.0002	0.0004	0.0002	0.0001	0.0002	<0.0001
Residual	0.0005	0.0004	0.0004	0.0005	0.0005	0.0005	0.0004	0.0004	0.0002

Table B.2: LMM summary for parent-offspring (PO) pairs.

PO LMM	Fixed Effects	Estimate	Standard Error	t	Pr (>|t|)
Model 1	(Intercept)	0.2474	0.0015	163.95	<0.01*
	NgsRelate	-0.0039	0.0012	-3.11	<0.01*
	READ	-0.0022	0.0012	-1.73	0.08
	20K	-0.0002	0.0016	-0.15	0.88
	10K	0.0002	0.0016	0.15	0.88
	5K	0.0018	0.0016	1.10	0.27
	1K	0.0021	0.0016	1.29	0.20
Model 2	(Intercept)	0.2482	0.0021	118.99	<0.01*
	NgsRelate	-0.0051	0.0028	-1.82	0.07
	READ	-0.0032	0.0028	-1.14	0.25
	20K	0.0002	0.0028	0.07	0.94
	10K	-0.0001	0.0028	-0.03	0.98
	5K	0.0014	0.0028	0.52	0.60
	1K	-0.0014	0.0028	-0.51	0.61
	NgsRelate: 20K	0.0001	0.0039	0.02	0.98
	READ: 20K	-0.0015	0.0039	-0.37	0.71
	NgsRelate: 10K	0.0008	0.0039	0.21	0.83
	READ: 10K	0.0001	0.0039	0.02	0.98
	NgsRelate: 5K	-0.0011	0.0039	-0.27	0.78
	READ: 5K	0.0020	0.0039	0.52	0.60
	NgsRelate: 1K	0.0061	0.0039	1.54	0.12
	READ: 1K	0.0044	0.0039	1.12	0.26

Table B.3: LMM summary for sibling (S) pairs.

S LMM	Fixed Effects	Estimate	Standard Error	t	Pr (>|t|)
Model 1	(Intercept)	0.2473	0.0034	73.24	<0.01*
	NgsRelate	-0.0034	0.0008	-4.13	<0.01*
	READ	-0.0021	0.0008	-2.51	0.01*
	20K	0.0000	0.0011	-0.04	0.97
	10K	0.0003	0.0011	0.26	0.80
	5K	0.0004	0.0011	0.40	0.69
	1K	0.0001	0.0011	0.13	0.89
Model 2	(Intercept)	0.2473	0.0035	70.53	<0.01*
	NgsRelate	-0.0039	0.0018	-2.15	0.03*
	READ	-0.0016	0.0018	-0.85	0.39
	20K	0.0006	0.0018	0.30	0.76
	10K	-0.0001	0.0018	-0.07	0.95
	5K	-0.0005	0.0018	-0.29	0.77
	1K	0.0008	0.0018	0.42	0.67
	NgsRelate:20K	-0.0005	0.0026	-0.20	0.84
	READ:20K	-0.0013	0.0026	-0.49	0.63
	NgsRelate:10K	0.0012	0.0026	0.47	0.64
	READ:10K	0.0000	0.0026	-0.02	0.99
	NgsRelate:5K	0.0009	0.0026	0.36	0.72
	READ:5K	0.0019	0.0026	0.73	0.46
	NgsRelate:1K	0.0012	0.0026	0.45	0.65
	READ:1K	-0.0031	0.0026	-1.18	0.24

Table B.4: LMM summary for half-sibling (HS) pairs.

HS LMM	Fixed Effects	Estimate	Standard Error	t	Pr (>|t|)
Model 1	(Intercept)	0.1244	0.0020	61.77	<0.01*
	NgsRelate	-0.0016	0.0009	-1.83	0.07
	READ	-0.0012	0.0009	-1.38	0.17
	20K	0.0002	0.0011	0.22	0.83
	10K	0.0008	0.0011	0.69	0.49
	5K	0.0016	0.0011	1.41	0.16
	1K	0.0015	0.0011	1.39	0.16
Model 2	(Intercept)	0.1254	0.0022	55.81	<0.01*
	NgsRelate	-0.0033	0.0019	-1.69	0.09
	READ	-0.0023	0.0019	-1.20	0.23
	20K	-0.0006	0.0019	-0.31	0.76
	10K	0.0002	0.0019	0.10	0.92
	5K	0	0.0019	-0.02	0.98
	1K	-0.0001	0.0019	-0.05	0.96
	NgsRelate:20K	0.0014	0.0027	0.52	0.60
	READ:20K	0.0011	0.0027	0.40	0.69
	NgsRelate:10K	0.0016	0.0027	0.58	0.56
	READ:10K	0.0001	0.0027	0.05	0.96
	NgsRelate:5K	0.0008	0.0027	0.30	0.76
	READ:5K	0.0040	0.0027	1.47	0.14
	NgsRelate:1K	0.0046	0.0027	1.68	0.09
	READ:1K	0.0004	0.0027	0.13	0.90

Table B.5: LMM summary for grandparent-grandchild (GPC) pairs.

GPC LMM	Fixed Effects	Estimate	Standard Error	t	Pr (>|t|)
Model 1	(Intercept)	0.1203	0.0041	29.21	<0.01*
	NgsRelate	-0.0036	0.0013	-2.81	<0.01*
	READ	-0.0048	0.0013	-3.76	<0.01*
	20K	-0.0002	0.0016	-0.10	0.92
	10K	0.0015	0.0016	0.91	0.36
	5K	0.0003	0.0016	0.20	0.84
	1K	0.0006	0.0016	0.37	0.71
Model 2	(Intercept)	0.1203	0.0044	27.50	<0.01*
	NgsRelate	-0.0038	0.0028	-1.32	0.19
	READ	-0.0045	0.0028	-1.57	0.12
	20K	-0.0002	0.0028	-0.09	0.93
	10K	0.0004	0.0028	0.14	0.89
	5K	0.0007	0.0028	0.24	0.81
	1K	0.0016	0.0028	0.58	0.56
	NgsRelate:20K	0.0003	0.0040	0.08	0.94
	READ:20K	-0.0001	0.0040	-0.02	0.99
	NgsRelate:10K	0.0002	0.0040	0.04	0.97
	READ:10K	0.0031	0.0040	0.78	0.44
	NgsRelate:5K	-0.0011	0.0040	-0.27	0.79
	READ:5K	0.0000	0.0040	-0.01	1.00
	NgsRelate:1K	0.0015	0.0040	0.36	0.72
	READ:1K	-0.0046	0.0040	-1.13	0.26

Table B.6: LMM summary for avuncular (AV) pairs.

AV LMM	Fixed Effects	Estimate	Standard Error	t	Pr (>|t|)
Model 1	(Intercept)	0.1256	0.0024	53.14	<0.01*
	NgsRelate	-0.0036	0.0009	-3.98	<0.01*
	READ	-0.0044	0.0009	-4.76	<0.01*
	20K	0.0002	0.0012	0.18	0.85
	10K	0.0001	0.0012	0.05	0.96
	5K	0.0004	0.0012	0.30	0.76
	1K	-0.0016	0.0012	-1.32	0.19
Model 2	(Intercept)	0.1251	0.0026	48.31	<0.01*
	NgsRelate	-0.0035	0.0020	-1.72	0.09
	READ	-0.0030	0.0020	-1.47	0.14
	20K	0.0010	0.0020	0.49	0.62
	10K	0.0003	0.0020	0.15	0.88
	5K	-0.0002	0.0020	-0.12	0.91
	1K	0.0004	0.0020	0.20	0.84
	NgsRelate:20K	-0.0010	0.0029	-0.35	0.73
	READ:20K	-0.0014	0.0029	-0.48	0.63
	NgsRelate:10K	0.0004	0.0029	0.13	0.89
	READ:10K	-0.0012	0.0029	-0.40	0.69
	NgsRelate:5K	0.0013	0.0029	0.45	0.66
	READ:5K	0.0005	0.0029	0.18	0.86
	NgsRelate:1K	-0.0013	0.0029	-0.44	0.66
	READ:1K	-0.0047	0.0029	-1.61	0.11

Table B.7: LMM summary for first cousin (FC) pairs.

FC LMM	Fixed Effects	Estimate	Standard Error	t	Pr (>|t|)
Model 1	(Intercept)	0.0629	0.0021	29.53	<0.01*
	NgsRelate	-0.0010	0.0009	-1.07	0.28
	READ	-0.0027	0.0009	-2.99	<0.01*
	20K	0.0004	0.0012	0.31	0.76
	10K	-0.0005	0.0012	-0.42	0.68
	5K	0.0004	0.0012	0.31	0.76
	1K	0.0032	0.0012	2.74	0.01*
Model 2	(Intercept)	0.0634	0.0024	26.80	<0.01*
	NgsRelate	-0.0024	0.0020	-1.22	0.22
	READ	-0.0028	0.0020	-1.40	0.16
	20K	0.0008	0.0020	0.39	0.70
	10K	0.0002	0.0020	0.11	0.91
	5K	-0.0007	0.0020	-0.35	0.73
	1K	0.0004	0.0020	0.22	0.83
	NgsRelate:20K	-0.0005	0.0028	-0.17	0.86
	READ:20K	-0.0008	0.0028	-0.27	0.79
	NgsRelate:10K	0.0002	0.0028	0.05	0.96
	READ:10K	-0.0023	0.0028	-0.80	0.43
	NgsRelate:5K	0.0015	0.0028	0.53	0.59
	READ:5K	0.0017	0.0028	0.58	0.56
	NgsRelate:1K	0.0062	0.0028	2.19	0.03*
	READ:1K	0.0020	0.0028	0.71	0.48

Table B.8: LMM summary for grand avuncular (GAV) pairs.

GAV LMM	Fixed Effects	Estimate	Standard Error	t	Pr (>|t|)
Model 1	(Intercept)	0.0633	0.0018	35.25	<0.01*
	NgsRelate	-0.0015	0.0008	-1.77	0.08
	READ	-0.0031	0.0008	-3.64	<0.01*
	20K	-0.0001	0.0011	-0.10	0.92
	10K	0.0000	0.0011	0.01	0.99
	5K	0.0000	0.0011	0.03	0.97
	1K	0.0047	0.0011	4.28	<0.01*
Model 2	(Intercept)	0.0638	0.0020	31.26	<0.01*
	NgsRelate	-0.0026	0.0019	-1.38	0.17
	READ	-0.0035	0.0019	-1.88	0.06
	20K	-0.0004	0.0019	-0.21	0.83
	10K	0.0007	0.0019	0.36	0.72
	5K	0.0018	0.0019	0.98	0.33
	1K	-0.0001	0.0019	-0.07	0.95
	NgsRelate:20K	0.0005	0.0027	0.20	0.84
	READ:20K	0.0003	0.0027	0.13	0.90
	NgsRelate:10K	0.0000	0.0027	0.01	0.99
	READ:10K	-0.0020	0.0027	-0.76	0.45
	NgsRelate:5K	-0.0011	0.0027	-0.40	0.69
	READ:5K	-0.0044	0.0027	-1.64	0.10
	NgsRelate:1K	0.0060	0.0027	2.25	0.02*
	READ:1K	0.0084	0.0027	3.16	<0.01*

Table B.9: LMM summary for great-grandparent-great-grandchild (GGPC) pairs.

GGPC LMM	Fixed Effects	Estimate	Standard Error	t	Pr (>|t|)
Model 1	(Intercept)	0.0663	0.0028	23.67	<0.01*
	NgsRelate	-0.0013	0.0012	-1.13	0.26
	READ	-0.0032	0.0012	-2.71	0.01*
	20K	0.0005	0.0015	0.34	0.74
	10K	-0.0002	0.0015	-0.13	0.90
	5K	0.0004	0.0015	0.29	0.77
	1K	0.0034	0.0015	2.23	0.03*
Model 2	(Intercept)	0.0668	0.0031	21.47	<0.01*
	NgsRelate	-0.0031	0.0026	-1.19	0.23
	READ	-0.0029	0.0026	-1.10	0.27
	20K	0.0006	0.0026	0.22	0.83
	10K	-0.0001	0.0026	-0.05	0.96
	5K	0.0003	0.0026	0.11	0.91
	1K	0.0009	0.0026	0.33	0.74
	NgsRelate:20K	-0.0001	0.0037	-0.02	0.98
	READ:20K	-0.0001	0.0037	-0.03	0.97
	NgsRelate:10K	0.0009	0.0037	0.23	0.81
	READ:10K	-0.0011	0.0037	-0.28	0.78
	NgsRelate:5K	0.0021	0.0037	0.57	0.57
	READ:5K	-0.0016	0.0037	-0.44	0.66
	NgsRelate:1K	0.0062	0.0037	1.66	0.10
	READ:1K	0.0014	0.0037	0.37	0.71

Table B.10: LMM summary for unrelated (UNR) pairs.

UNR LMM	Fixed Effects	Estimate	Standard Error	t	Pr (>|t|)
Model 1	(Intercept)	0.002	0.0006	2.95	<0.01*
	NgsRelate	0.00	0.0005	0.07	0.95
	READ	0.004	0.0005	7.74	<0.01*
	20K	0.0011	0.0007	1.71	0.09
	10K	0.0025	0.0007	3.78	<0.01*
	5K	0.0047	0.0007	7.08	<0.01*
	1K	0.0134	0.0007	20.16	<0.01*
Model 2	(Intercept)	0.0034	0.0008	4.04	<0.01*
	NgsRelate	-0.0008	0.0011	-0.71	0.48
	READ	0.0000	0.0011	-0.02	0.98
	20K	0.0006	0.0011	0.55	0.58
	10K	0.0011	0.0011	0.97	0.33
	5K	0.0034	0.0011	2.95	<0.01*
	1K	0.0086	0.0011	7.56	<0.01*
	NgsRelate:20K	0.0005	0.0016	0.33	0.74
	READ:20K	0.0010	0.0016	0.63	0.53
	NgsRelate:10K	0.0017	0.0016	1.04	0.30
	READ:10K	0.0026	0.0016	1.59	0.11
	NgsRelate:5K	0.0000	0.0016	0.03	0.98
	READ:5K	0.0040	0.0016	2.49	0.01*
	NgsRelate:1K	0.0019	0.0016	1.21	0.23
	READ:1K	0.0124	0.0016	7.74	<0.01*

Table B.11: LMM Model 1 (no interaction) using absolute mean differences for parent-offspring pairs.

Model 1 Fixed Effects				
	Estimate	**Standard Error**	**t**	**Pr (>|t|)**
(Intercept)	0.0010	0.0009	1.15	0.25
NgsRelate	0.0009	0.0008	1.15	0.25
READ	0.0099	0.0008	12.77	<0.01*
20K	0.0023	0.0010	2.27	0.02*
10K	0.0050	0.0010	4.96	<0.01*
5K	0.0094	0.0010	9.37	<0.01*
1K	0.0263	0.0010	26.23	<0.01*

Table B.12: LMM Model 1 (no interaction) using absolute mean differences for sibling pairs.

Model 1 Fixed Effects				
	Estimate	**Standard Error**	**t**	**Pr (>|t|)**
(Intercept)	0.0170	0.0016	10.98	<0.01*
NgsRelate	0.0003	0.0006	0.44	0.66
READ	0.0046	0.0006	7.66	<0.01*
20K	0.0012	0.0008	1.51	0.13
10K	0.0032	0.0008	4.16	<0.01*
5K	0.0044	0.0008	5.58	<0.01*
1K	0.0150	0.0008	19.20	<0.01*

Table B.13: LMM Model 1 (no interaction) using absolute mean differences for half-sibling pairs.

Model 1 Fixed Effects				
	Estimate	**Standard Error**	**t**	**Pr (>|t|)**
(Intercept)	0.0076	0.0010	7.62	<0.01*
NgsRelate	-0.0003	0.0006	-0.46	0.65
READ	0.0072	0.0006	12.81	<0.01*
20K	0.0013	0.0007	1.76	0.08
10K	0.0043	0.0007	5.90	<0.01*
5K	0.0072	0.0007	9.90	<0.01*
1K	0.0229	0.0007	31.56	<0.01*

Table B.14: LMM Model 1 (no interaction) using absolute mean differences for grandparent-grandchild pairs.

Model 1 Fixed Effects				
	Estimate	**Standard Error**	**t**	**Pr (>|t|)**
(Intercept)	0.0129	0.0020	6.44	<0.01*
NgsRelate	-0.0002	0.0009	-0.20	0.84
READ	0.0076	0.0009	8.68	<0.01*
20K	0.0016	0.0011	1.41	0.16
10K	0.0034	0.0011	3.04	<0.01*
5K	0.0056	0.0011	4.97	<0.01*
1K	0.0187	0.0011	16.53	<0.01*

Table B.15: LMM Model 1 (no interaction) using absolute mean differences for avuncular pairs.

Model 1 Fixed Effects				
	Estimate	**Standard Error**	**t**	**Pr (>|t|)**
(Intercept)	0.0092	0.0011	8.08	<0.01*
NgsRelate	-0.0001	0.0006	-0.18	0.86
READ	0.0082	0.0006	13.03	<0.01*
20K	0.0020	0.0008	2.42	0.02*
10K	0.0035	0.0008	4.28	<0.01*
5K	0.0074	0.0008	9.07	<0.01*
1K	0.0219	0.0008	26.84	<0.01*

Table B.16: LMM Model 1 (no interaction) using absolute mean differences for first cousin pairs.

Model 1 Fixed Effects				
	Estimate	**Standard Error**	**t**	**Pr (>|t|)**
(Intercept)	0.0095	0.0011	8.76	<0.01*
NgsRelate	0.0000	0.0006	-0.01	0.99
READ	0.0087	0.0006	14.65	<0.01*
50K	-0.0022	0.0008	-2.90	<0.01*
10K	0.0025	0.0008	3.22	<0.01*
5K	0.0053	0.0008	6.99	<0.01*
1K	0.0209	0.0008	27.31	<0.01*

Table B.17: LMM Model 1 (no interaction) using absolute mean differences for great-grandparent-great-grandchild pairs.

Model 1 Fixed Effects				
	Estimate	**Standard Error**	**t**	**Pr (>|t|)**
(Intercept)	0.0082	0.0012	70.50	<0.01*
NgsRelate	-0.0005	0.0007	-0.73	0.47
READ	0.0083	0.0007	11.14	<0.01*
20K	0.0015	0.0010	1.53	0.13
10K	0.0038	0.0010	4.01	<0.01*
5K	0.0052	0.0010	5.39	<0.01*
1K	0.0199	0.0010	20.77	<0.01*

Table B.18: LMM Model 1 (no interaction) using absolute mean differences for grand avuncular pairs.

Model 1 Fixed Effects				
	Estimate	**Standard Error**	**t**	**Pr (>|t|)**
(Intercept)	0.0070	0.0009	8.16	<0.01*
NgsRelate	-0.0004	0.0005	-0.82	0.41
READ	0.0081	0.0005	15.15	<0.01*
20K	0.0017	0.0007	2.38	0.02*
10K	0.0031	0.0007	4.42	<0.01*
5K	0.0078	0.0007	11.22	<0.01*
1K	0.0209	0.0007	30.17	<0.01*

Table B.19: LMM Model 1 (no interaction) of parent-offspring pairs in the presence of inbreeding.

Model 1 Fixed Effects				
	Estimate	**Standard Error**	**t**	**Pr (>|t|)**
(Intercept)	0.2764	0.0022	123.33	<0.01*
NgsRelate	-0.0332	0.0013	-25.39	<0.01*
READ	0.0066	0.0013	5.01	<0.01*
20K	-0.0016	0.0017	-0.95	0.34
10K	-0.0004	0.0017	-0.25	0.81
5K	-0.0008	0.0017	-0.46	0.65
1K	-0.0067	0.0017	-3.95	<0.01*

Table B.20: LMM Model 1 (no interaction) using absolute mean differences for parent-offspring pairs in the presence of inbreeding.

Model 1 Fixed Effects				
	Estimate	**Standard Error**	**t**	**Pr (>|t|)**
(Intercept)	0.0044	0.0010	4.36	<0.01*
NgsRelate	0.0044	0.0008	5.48	<0.01*
READ	0.0098	0.0008	12.19	<0.01*
20K	0.0023	0.0010	2.20	0.03*
10K	0.0038	0.0010	3.67	<0.01*
5K	0.0073	0.0010	7.04	<0.01*
1K	0.0230	0.0010	22.24	<0.01*

Table B.21: LMM Model 1 (no interaction) of grandparent-grandchild pairs in the presence of inbreeding.

Model 1 Fixed Effects				
	Estimate	**Standard Error**	**t**	**Pr (>|t|)**
(Intercept)	0.1938	0.0030	64.98	<0.01*
NgsRelate	-0.0331	0.0015	-22.64	<0.01*
READ	-0.0008	0.0015	-0.53	0.59
20K	-0.0006	0.0019	-0.30	0.76
10K	0.0002	0.0019	0.10	0.92
5K	-0.0006	0.0019	-0.34	0.74
1K	-0.0094	0.0019	-4.97	<0.01*

Table B.22: LMM Model 1 (no interaction) using absolute mean differences for grandparent-grandchild pairs in the presence of inbreeding.

Model 1 Fixed Effects				
	Estimate	**Standard Error**	**t**	**Pr (>|t|)**
(Intercept)	0.0069	0.0013	5.17	<0.01*
NgsRelate	0.0042	0.0009	4.75	<0.01*
READ	0.0081	0.0009	9.16	<0.01*
20K	0.0019	0.0011	1.70	0.09
10K	0.0052	0.0011	4.51	<0.01*
5K	0.0093	0.0011	8.11	<0.01*
1K	0.0280	0.0011	24.50	<0.01*

APPENDIX C

LEVENE'S TEST RESULTS

Table C.1: Levene's test results of kinship coefficient estimates from READ, NgsRelate, and lcMLkin for parent-offspring and grandparent-grandchild pairs with one inbred individual and SNP counts (n=5 per count). The values show the p-values from Levene's test and the stars represent significant p-values (<0.05).

SNP Count	**Parent-offspring**					**Grandparent-grandchild**				
	1st	2nd	3rd	4th	5th	1st	2nd	3rd	4th	5th
1K	<0.01*	0.02	0.02	<0.01*	0.01	0.59	0.02	0.04	0.15	0.03
5K	<0.01*	0.01	0.66	0.12	0.04	0.01	0.10	<0.01*	<0.01*	0.32
10K	0.11	0.03	0.23	0.45	<0.01*	0.22	0.01	0.53	0.24	0.82
20K	0.26	0.18	0.21	0.19	0.12	0.25	0.31	0.72	0.39	0.75
50K	0.37	0.67	0.50	0.68	0.73	0.32	0.60	0.99	0.98	0.89

APPENDIX D

BINOMIAL TEST RESULTS

Table D.1: Binomial test results of kinship coefficient estimates among READ, NgsRelate, and lcMLkin for parent-offspring pairs. The binomial shows the p-values. The success rate represents the rate of the estimates above the corresponding theoretical value.

Software	**SNP Count**	**Binomial**					**Success Rate**				
		1st	2nd	3rd	4th	5th	1st	2nd	3rd	4th	5th
READ	1K	0.839	0.541	0.839	0.541	0.839	54.2	41.7	54.2	58.3	54.2
NgsRelate		0.839	1.000	0.541	0.541	1.000	45.8	50.0	41.7	58.3	50.0
lcMLkin		0.839	0.839	0.152	1.000	1.000	54.2	45.8	33.3	50.0	50.0
READ	5K	0.839	1.000	0.839	0.541	0.541	45.8	50.0	45.8	58.3	41.7
NgsRelate		0.064	0.007	0.064	0.541	0.307	29.2	20.8	29.2	41.7	37.5
lcMLkin		0.839	1.000	0.152	0.839	0.064	45.8	50.0	66.7	45.8	70.8
READ	10K	0.541	0.541	0.307	0.152	0.023*	41.7	41.7	37.5	33.3	25.0
NgsRelate		0.064	0.152	0.064	0.007*	0.002*	29.2	33.3	29.2	20.8	16.7
lcMLkin		0.839	0.064	0.152	0.839	0.541	45.8	29.2	66.7	54.2	41.7
READ	20K	0.064	0.023*	0.307	0.541	0.002*	29.2	25.0	37.5	41.7	16.7
NgsRelate		0.002*	<0.01*	0.064	<0.01*	<0.01*	16.7	8.3	29.2	12.5	12.5
lcMLkin		0.839	0.839	1.000	0.541	0.839	45.8	54.2	50.0	41.7	45.8
READ	50K	0.541	0.064	0.064	0.007	0.064	41.7	29.2	29.2	20.8	29.2
NgsRelate		<0.01*	<0.01*	<0.01*	<0.01*	<0.01*	0.0	4.2	12.5	8.3	4.2
lcMLkin		0.839	0.064	0.064	1.000	0.541	45.8	29.2	29.2	50.0	41.7

Table D.2: Binomial test results of kinship coefficient estimates among READ, NgsRelate, and lcMLkin for sibling pairs (n=48). The binomial shows the p-values. The success rate represents the rate of the θ estimates above the corresponding theoretical value (n=5 replicate per SNP count).

Software	**SNP Count**	**Binomial**					**Success Rate**				
		1st	2nd	3rd	4th	5th	1st	2nd	3rd	4th	5th
READ	1K	0.885	0.059	0.665	0.312	0.471	47.9	35.4	45.8	41.7	43.8
NgsRelate		0.111	0.665	0.471	0.885	0.111	37.5	54.2	43.8	52.1	37.5
lcMLkin		0.665	0.885	1.000	0.471	0.471	45.8	52.1	50.0	56.3	43.8
READ	5K	0.471	0.312	1.000	0.885	0.471	43.8	41.7	50.0	47.9	43.8
NgsRelate		0.471	0.111	0.471	0.059	0.665	43.8	37.5	43.8	35.4	45.8
lcMLkin		0.193	0.471	0.471	0.312	0.193	39.6	43.8	43.8	41.7	39.6
READ	10K	0.885	0.665	0.471	0.665	0.665	47.9	45.8	43.8	45.8	54.2
NgsRelate		0.193	0.193	0.193	0.193	0.111	39.6	39.6	39.6	39.6	37.5
lcMLkin		0.471	0.665	0.193	0.193	0.193	43.8	45.8	39.6	39.6	39.6
READ	20K	0.665	0.471	0.665	0.059	0.059	45.8	43.8	45.8	35.4	35.4
NgsRelate		0.029*	0.059	0.111	0.013*	0.193	33.3	35.4	37.5	31.3	39.6
lcMLkin		0.312	0.885	0.471	0.193	0.111	41.7	47.9	43.8	39.6	37.5
READ	50K	0.193	0.111	0.193	0.471	0.471	39.6	37.5	39.6	43.8	43.8
NgsRelate		0.029*	0.013*	0.059	0.029*	0.059	33.3	31.3	35.4	33.3	35.4
lcMLkin		0.312	0.665	0.193	0.111	0.312	41.7	45.8	39.6	37.5	41.7

Table D.3: Binomial test results of kinship coefficient estimates among READ, NgsRelate, and lcMLkin for half-sibling pairs (n=48). The binomial shows the p-values. The success rate represents the rate of the θ estimates above the corresponding theoretical value (n=5 replicate per SNP count).

Software	**SNP Count**	**Binomial**					**Success Rate**				
		1st	2nd	3rd	4th	5th	1st	2nd	3rd	4th	5th
READ	1K	1.000	1.000	1.000	0.885	0.885	50.0	50.0	50.0	47.9	52.1
NgsRelate		0.665	0.312	0.665	0.312	0.665	54.2	58.3	54.2	41.7	45.8
lcMLkin		0.312	0.312	0.665	0.885	0.471	58.3	41.7	54.2	47.9	43.8
READ	5K	0.471	0.471	1.000	0.885	0.665	56.3	43.8	50.0	47.9	54.2
NgsRelate		0.665	0.193	0.665	0.471	1.000	45.8	39.6	45.8	43.8	50.0
lcMLkin		1.000	0.471	0.885	0.885	0.665	50.0	56.3	47.9	47.9	54.2
READ	10K	0.471	0.665	0.471	1.000	0.665	43.8	45.8	43.8	50.0	45.8
NgsRelate		0.312	0.471	0.471	0.885	0.471	41.7	43.8	56.3	47.9	56.3
lcMLkin		0.471	0.471	0.312	0.471	1.000	43.8	56.3	58.3	56.3	50.0
READ	20K	0.312	0.312	0.312	0.885	0.193	41.7	58.3	41.7	47.9	39.6
NgsRelate		0.471	0.029*	0.111	0.312	0.885	43.8	33.3	37.5	41.7	47.9
lcMLkin		0.312	0.665	0.665	1.000	0.665	58.3	45.8	54.2	50.0	45.8
READ	50K	0.111	0.312	0.111	0.312	0.885	37.5	41.7	37.5	41.7	47.9
NgsRelate		0.111	0.193	0.193	0.059	0.471	37.5	39.6	39.6	35.4	43.8
lcMLkin		0.665	0.885	1.000	0.885	0.885	54.2	52.1	50.0	52.1	52.1

Table D.4: Binomial test results of kinship coefficient estimates among READ, NgsRelate, and lcMLkin for avuncular pairs(n=48). The binomial shows the p-values. The success rate represents the rate of the θ estimates above the corresponding theoretical value (n=5 replicate per SNP count).

Software	**SNP Count**	**Binomial**					**Success Rate**				
		1st	2nd	3rd	4th	5th	1st	2nd	3rd	4th	5th
READ	1K	0.471	0.029*	0.885	0.312	0.885	43.8	33.3	47.9	58.3	47.9
NgsRelate		0.193	0.471	0.059	0.885	0.665	39.6	43.8	35.4	47.9	54.2
lcMLkin		0.665	0.111	0.111	0.471	0.665	54.2	62.5	62.5	43.8	45.8
READ	5K	0.471	0.885	0.665	0.471	0.111	43.8	47.9	54.2	43.8	37.5
NgsRelate		0.665	0.111	0.471	0.885	0.471	45.8	37.5	43.8	52.1	43.8
lcMLkin		0.665	0.665	0.665	0.312	0.665	54.2	45.8	45.8	41.7	54.2
READ	10K	0.665	0.885	0.665	0.029*	0.471	45.8	52.1	54.2	33.3	43.8
NgsRelate		0.111	0.885	0.885	1.000	0.193	37.5	52.1	47.9	50.0	39.6
lcMLkin		0.885	0.471	0.471	0.059	0.471	52.1	56.3	56.3	64.6	43.8
READ	20K	0.059	0.471	0.665	0.885	1.000	35.4	43.8	45.8	52.1	50.0
NgsRelate		0.059	0.471	0.059	0.312	0.111	35.4	43.8	35.4	41.7	37.5
lcMLkin		0.665	0.885	0.885	1.000	0.471	54.2	52.1	52.1	50.0	56.3
READ	50K	0.193	0.312	0.471	0.111	0.193	39.6	41.7	43.8	37.5	39.6
NgsRelate		0.193	0.193	0.885	0.111	0.193	39.6	39.6	47.9	37.5	39.6
lcMLkin		1.000	1.000	0.885	1.000	0.665	50.0	50.0	47.9	50.0	45.8

Table D.5: Binomial test results of kinship coefficient estimates among READ, NgsRelate, and lcMLkin for grandparent-grandchild pairs(n=48). The binomial shows the p-values. The success rate represents the rate of the θ estimates above the corresponding theoretical value (n=5 replicate per SNP count).

Software	**SNP Count**	**Binomial**					**Success Rate**				
		1st	2nd	3rd	4th	5th	1st	2nd	3rd	4th	5th
READ	1K	0.541	0.152	0.064	0.839	0.839	41.7	33.3	29.2	45.8	45.8
NgsRelate		0.541	0.541	0.839	0.307	0.023*	58.3	41.7	45.8	62.5	25.0
lcMLkin		0.839	0.541	1.000	0.064	0.541	45.8	41.7	50.0	29.2	58.3
READ	5K	0.152	0.152	0.541	1.000	0.839	33.3	33.3	41.7	50.0	45.8
NgsRelate		0.152	0.064	0.152	0.023*	0.541	33.3	29.2	33.3	25.0	41.7
lcMLkin		0.541	1.000	0.839	0.541	1.000	41.7	50.0	45.8	41.7	50.0
READ	10K	0.541	1.000	0.541	0.541	0.152	41.7	50.0	41.7	41.7	33.3
NgsRelate		0.152	0.307	0.007*	0.023*	0.541	33.3	37.5	20.8	25.0	41.7
lcMLkin		0.839	0.307	0.307	0.064	0.541	45.8	37.5	37.5	29.2	41.7
READ	20K	0.307	0.064	0.152	0.839	0.064	37.5	29.2	33.3	45.8	29.2
NgsRelate		0.152	0.152	1.000	0.307	0.307	33.3	33.3	50.0	37.5	37.5
lcMLkin		0.839	0.839	0.839	1.000	0.307	45.8	45.8	45.8	50.0	37.5
READ	50K	0.023*	0.064	0.307	0.064	0.152	25.0	29.2	37.5	29.2	33.3
NgsRelate		0.541	0.307	0.307	0.152	0.307	41.7	37.5	37.5	33.3	37.5
lcMLkin		0.839	0.839	0.541	0.541	0.839	54.2	54.2	41.7	41.7	45.8

Table D.6: Binomial test results of kinship coefficient estimates among READ, NgsRelate, and lcMLkin for first cousin pairs(n=48). The binomial shows the p-values. The success rate represents the rate of the θ estimates above the corresponding theoretical value (n=5 replicate per SNP count).

Software	**SNP Count**	**Binomial**					**Success Rate**				
		1st	2nd	3rd	4th	5th	1st	2nd	3rd	4th	5th
READ	1K	0.312	0.001*	0.006*	0.029*	0.029*	41.7	25.0	29.2	33.3	33.3
NgsRelate		0.885	1.000	0.312	0.059	0.665	47.9	50.0	41.7	35.4	45.8
lcMLkin		0.471	0.471	0.193	0.312	0.193	56.3	56.3	39.6	41.7	39.6
READ	5K	0.665	0.471	0.471	0.111	1.000	54.2	43.8	56.3	37.5	50.0
NgsRelate		0.111	0.471	0.193	0.312	0.665	37.5	43.8	39.6	41.7	45.8
lcMLkin		0.665	0.665	0.665	1.000	0.885	45.8	54.2	54.2	50.0	47.9
READ	10K	0.885	0.471	0.471	0.312	0.312	47.9	43.8	43.8	41.7	41.7
NgsRelate		0.471	0.193	0.665	0.665	0.665	43.8	39.6	45.8	45.8	45.8
lcMLkin		1.000	0.665	0.665	0.665	0.471	50.0	45.8	45.8	54.2	43.8
READ	20K	0.312	0.471	0.665	0.193	0.111	41.7	43.8	54.2	39.6	37.5
NgsRelate		0.312	0.013*	0.312	0.013*	0.111	41.7	31.3	41.7	31.3	37.5
lcMLkin		0.885	1.000	1.000	1.000	1.000	52.1	50.0	50.0	50.0	50.0
READ	50K	0.111	0.312	0.111	0.059	0.193	37.5	41.7	37.5	35.4	39.6
NgsRelate		0.111	0.111	0.059	0.029*	0.029*	37.5	37.5	35.4	33.3	33.3
lcMLkin		0.665	0.885	0.885	0.471	0.665	45.8	52.1	52.1	43.8	45.8

Table D.7: Binomial test results of kinship coefficient estimates among READ, NgsRelate, and lcMLkin for great-grandparent-great-grandchild pairs(n=48). The binomial shows the p-values. The success rate represents the rate of the θ estimates above the corresponding theoretical value (n=5 replicate per SNP count).

Software	**SNP Count**	**Binomial**					**Success Rate**				
		1st	2nd	3rd	4th	5th	1st	2nd	3rd	4th	5th
READ	1K	1.000	0.152	0.541	0.023*	1.000	50.0	33.3	41.7	25.0	50.0
NgsRelate		1.000	1.000	0.839	0.839	<0.01*	50.0	50.0	45.8	54.2	87.5
lcMLkin		0.541	1.000	0.307	0.839	0.839	58.3	50.0	62.5	45.8	45.8
READ	5K	0.307	0.839	0.541	0.541	0.541	37.5	45.8	58.3	41.7	58.3
NgsRelate		0.152	0.839	0.064	0.152	0.541	66.7	45.8	70.8	66.7	58.3
lcMLkin		0.839	0.007*	0.839	0.152	0.839	54.2	79.2	45.8	66.7	54.2
READ	10K	0.839	0.839	0.152	0.307	0.839	54.2	45.8	33.3	62.5	54.2
NgsRelate		0.839	0.541	0.307	0.307	1.000	54.2	58.3	62.5	62.5	50.0
lcMLkin		0.152	0.152	1.000	0.152	0.541	66.7	66.7	50.0	66.7	58.3
READ	20K	0.307	0.541	0.541	0.541	0.541	62.5	58.3	58.3	58.3	58.3
NgsRelate		0.541	0.307	0.541	0.307	0.541	58.3	62.5	58.3	62.5	58.3
lcMLkin		0.307	0.307	0.152	1.000	0.064	62.5	62.5	66.7	50.0	70.8
READ	50K	0.541	0.541	0.307	0.307	0.541	58.3	58.3	62.5	62.5	58.3
NgsRelate		0.541	0.541	0.541	0.839	0.152	58.3	58.3	58.3	54.2	66.7
lcMLkin		0.152	0.064	0.307	0.152	0.307	66.7	70.8	62.5	66.7	62.5

Table D.8: Binomial test results of kinship coefficient estimates among READ, NgsRelate, and lcMLkin for grand avuncular pairs(n=48). The binomial shows the p-values. The success rate represents the rate of the θ estimates above the corresponding theoretical value (n=5 replicate per SNP count).

Software	**SNP Count**	**Binomial**					**Success Rate**				
		1st	2nd	3rd	4th	5th	1st	2nd	3rd	4th	5th
READ	1K	0.006*	0.111	1.000	0.665	0.006*	29.2	37.5	50.0	45.8	29.2
NgsRelate		0.885	0.665	0.885	0.193	0.006*	52.1	54.2	52.1	39.6	70.8
lcMLkin		0.312	0.665	0.193	0.471	0.885	58.3	45.8	39.6	43.8	47.9
READ	5K	0.665	0.885	0.029*	0.193	0.665	54.2	52.1	33.3	39.6	54.2
NgsRelate		0.111	0.665	0.471	0.471	0.665	37.5	54.2	43.8	43.8	45.8
lcMLkin		0.111	0.193	0.013*	0.029*	0.885	62.5	60.4	68.8	66.7	52.1
READ	10K	0.665	0.665	0.193	0.885	0.193	45.8	45.8	39.6	47.9	39.6
NgsRelate		0.111	0.665	0.885	0.312	1.000	37.5	54.2	52.1	41.7	50.0
lcMLkin		0.885	0.312	0.193	0.312	0.665	52.1	58.3	60.4	58.3	54.2
READ	20K	0.111	0.665	0.471	0.885	0.885	37.5	54.2	43.8	47.9	52.1
NgsRelate		0.665	0.312	0.885	0.885	0.471	54.2	41.7	47.9	52.1	43.8
lcMLkin		0.471	0.665	0.471	0.885	0.885	56.3	45.8	56.3	47.9	47.9
READ	50K	0.193	1.000	0.665	0.312	0.471	39.6	50.0	45.8	41.7	43.8
NgsRelate		0.193	0.193	0.193	0.312	0.193	39.6	39.6	39.6	41.7	39.6
lcMLkin		0.885	0.312	1.000	0.885	0.665	52.1	58.3	50.0	47.9	54.2

Table D.9: Binomial test results of kinship coefficient estimates among READ, NgsRelate, and lcMLkin for parent-offspring pairs in the presence of inbreeding (n=48). The binomial shows the p-values. The success rate represents the rate of the θ estimates above the corresponding theoretical value (n=5 replicate per SNP count).

Software	**SNP Count**	**Binomial**					**Success Rate**				
		1st	2nd	3rd	4th	5th	1st	2nd	3rd	4th	5th
READ	1K	0.541	0.541	0.064	0.307	0.064	58.33	58.3	70.83	62.50	70.83
NgsRelate		0.064	0.307	0.064	0.002	0.064	29.17	37.5	29.17	16.67	29.17
lcMLkin		0.007	<0.01*	0.007	0.023	<0.01*	79.17	95.8	79.17	75.00	87.50
READ	5K	<0.01*	<0.01*	<0.01*	<0.01*	<0.01*	87.50	87.5	87.50	87.50	91.67
NgsRelate		0.023	0.064	0.152	0.064	0.064	25.00	29.1	33.33	29.17	29.17
lcMLkin		<0.01*	<0.01*	<0.01*	<0.01*	<0.01*	95.83	100.0	87.5	95.83	95.83
READ	10K	<0.01*	0.002	<0.01*	<0.01*	0.007	87.5	83.3	95.8	100.0	79.2
NgsRelate		0.307	0.839	0.152	0.307	0.307	37.5	54.2	33.3	37.5	37.5
lcMLkin		<0.01*	<0.01*	<0.01*	<0.01*	<0.01*	100.0	100.0	95.8	100.0	100.0
READ	20K	<0.01*	<0.01*	<0.01*	<0.01*	<0.01*	100.0	95.8	100.0	100.0	91.7
NgsRelate		0.307	0.064	0.541	0.307	0.307	37.5	29.2	41.7	37.5	37.5
lcMLkin		<0.01*	<0.01*	<0.01*	<0.01*	<0.01*	100.0	95.8	100.0	100.0	100.0
READ	50K	<0.01*	<0.01*	<0.01*	<0.01*	<0.01*	100.0	100.0	100.0	100.0	100.0
NgsRelate		0.307	0.152	0.152	0.541	0.023	37.5	33.3	33.3	41.7	25.0
lcMLkin		<0.01*	<0.01*	<0.01*	<0.01*	<0.01*	100.0	100.0	100.0	100.0	100.0

Table D.10: Binomial test results of kinship coefficient estimates among READ, NgsRelate, and lcMLkin for grandparent-grandchild pairs in the presence of inbreeding(n=48). The binomial shows the p-values. The success rate represents the rate of the θ estimates above the corresponding theoretical value (n=5 replicate per SNP count).

Software	**SNP Count**	**Binomial**					**Success Rate**				
		1st	2nd	3rd	4th	5th	1st	2nd	3rd	4th	5th
READ	1K	0.002	0.307	0.064	0.307	0.023	83.3	62.5	70.8	62.5	75.0
NgsRelate		0.023	0.541	0.064	0.152	0.307	75.0	58.3	70.8	66.7	62.5
lcMLkin		<0.01*	0.002	<0.01*	<0.01*	<0.01*	91.7	83.3	100.0	91.7	95.8
READ	5K	<0.01*	<0.01*	<0.01*	<0.01*	<0.01*	100.0	95.8	100.0	87.5	95.8
NgsRelate		<0.01*	0.007	<0.01*	<0.01*	<0.01*	87.5	79.2	87.5	91.7	91.7
lcMLkin		<0.01*	<0.01*	<0.01*	<0.01*	<0.01*	100.0	100.0	100.0	100.0	100.0
READ	10K	<0.01*	<0.01*	<0.01*	<0.01*	<0.01*	100.0	95.8	100.0	95.8	100.0
NgsRelate		<0.01*	<0.01*	<0.01*	<0.01*	<0.01*	100.0	100.0	100.0	95.8	95.8
lcMLkin		<0.01*	<0.01*	<0.01*	<0.01*	<0.01*	100.0	100.0	100.0	100.0	100.0
READ	20K	<0.01*	<0.01*	<0.01*	<0.01*	<0.01*	100.0	100.0	100.0	100.0	100.0
NgsRelate		<0.01*	<0.01*	<0.01*	<0.01*	<0.01*	100.0	100.0	95.8	95.8	100.0
lcMLkin		<0.01*	<0.01*	<0.01*	<0.01*	<0.01*	100.0	100.0	100.0	100.0	100.0
READ	50K	<0.01*	<0.01*	<0.01*	<0.01*	<0.01*	100.0	100.0	100.0	100.0	100.0
NgsRelate		<0.01*	<0.01*	<0.01*	<0.01*	<0.01*	100.0	100.0	100.0	100.0	100.0
lcMLkin		<0.01*	<0.01*	<0.01*	<0.01*	<0.01*	100.0	100.0	100.0	100.0	100.0