Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

Investigating the impact of genetic linkage between sheep flocks on the accuracy and outcomes of genetic evaluations

A thesis presented in partial fulfilment of the requirements for the degree of

Master of Science

in

Animal Science

At Massey University, Palmerston North, New Zealand



Jessica Louise Smith

2026

Abstract

Genetic connectedness is essential for accurate comparison of estimated breeding values (EBVs) across flocks, ensuring that animals raised in different environments can be evaluated on a common genetic scale. This study assessed the impact of alternative connectedness criteria and data structures on across-flock genetic evaluations using facial eczema (FE) and growth datasets from the nProve database as examples of sparsely recorded and well recorded traits, respectively. The FE analysis focused on blood concentration of the liver enzyme gamma-glutamyl transferase response measured 21 days post-challenge with the fungal toxin sporidesmin, while weaning weight was used as the growth trait. Connectedness analyses were conducted using the current nProve methodology in SAS and the Genetic Connectedness Analysis package in R. Prediction error variance of differences and coefficient of determination metrics were used to assess across-flock evaluation accuracy. The effects of varying the evaluation data window (years of data included in connectedness assessment), sequential removal of highly connected flocks, and masking of prolific sires were investigated. Genetic parameters and EBVs were estimated using single-trait animal models fitted in ASReml, and EBV correlations across scenarios were used to assess the stability of genetic evaluations. Extending the data window increased connectedness in the sparsely recorded FE dataset but had little effect on the growth dataset. Removal of highly connected flocks slightly reduced connectedness for FE, while masking prolific sires had minimal impact in both datasets. Coefficient of determination values increased with the inclusion of additional historical data, reflecting improved evaluation accuracy, before stabilising after approximately five to six years of data. Estimated breeding values and animal rankings derived from the 3-year data window, representing the current nProve connectedness window, differed from those obtained using the full dataset. As the evaluation window expanded from one to eleven years, correlations between EBVs from consecutive data windows increased, with Spearman rank correlations rising from 0.57 to 0.88 for FE and from 0.54 to 0.88 for growth. However, gains in correlation plateaued after approximately five to seven years of data. These findings indicate that while the inclusion of additional historical data generally improves across-flock evaluation accuracy, benefits diminish beyond approximately five years of records for both growth and FE traits. Conservative short-term connectedness assessment windows may unnecessarily restrict flock participation in across-flock evaluations, whereas extending the evaluation window could provide a better balance between accuracy, connectedness, and computational efficiency. The

limited impact of flock removal and sire masking demonstrates that current genetic evaluations are not overly dependent on individual flocks or sires.

Acknowledgements

Completing this thesis would not have been possible without the support, guidance, and encouragement of many people throughout my time at Massey University. Firstly, I would like to thank my supervisors, Dr. Nick Sneddon, Dr. Neville Jopson, Professor Nicolas Lopez-Villalobos, Dr Natalia Martin, Dr. Danitsja van der Linden, and Dr. Jason Archer. Having the opportunity to work alongside so many knowledgeable and passionate people within the industry was an incredibly valuable experience.

Nick, thank you for going above and beyond with your support and guidance throughout this project. Your calm and positive approach made even the most daunting stages of postgraduate study feel manageable, and I have greatly appreciated your willingness to always make time to answer questions and offer reassurance when needed. Thank you also for the opportunity to have this work presented at the World Congress on Genetics Applied to Livestock Production.

Nicolas and Natalia, thank you for your thoughtful feedback, for sharing your expertise, and for challenging my thinking throughout this research. Neville, thank you for your guidance with the statistical components of this project, for helping me understand the nProve code, and for directing the work in a way that best addressed these important industry questions. Jason and Dani, thank you for proposing the research questions that shaped this thesis, and for your enthusiasm and passion for supporting students like myself entering the industry.

I would also like to thank Beef + Lamb New Zealand Genetics for their financial support throughout my studies and for providing access to the nProve data used in this research.

I am incredibly grateful to Livestock Improvement Corporation for awarding me the Brian Aspin Memorial Scholarship, as well as the C. Alma Baker Trust for their generous scholarship support. A huge thank you also goes to Massey University for the Charles Elgar Scholarship, Helen E. Akers Postgraduate Scholarship, Millar Massey Buchanan Scholarship, John Alexander Milne Scholarship, Sinclair Cummings Veterinary and Animal Sciences Scholarship, and the Peter Densem Postgraduate Scholarship.

To the friends that I met during my time at Massey, Dorothy, Emma, Leith and Leo, thank you for making my time in Palmerston North far more enjoyable. The laughs, shared baking, and random debates in the postgraduate study room provided many welcomed distractions, and I am excited to see where life takes each of you in the future.

A big thank you to Regan for always being by my side, putting up with my stress, and giving me some “practical experience” in the sheep industry during university breaks spent on the farm. Finally, I would like to thank my parents for their ongoing love and support, and their willingness to pretend they understood my thesis updates each week. I would not be where I am today without your guidance, encouragement, and confidence in me every step of the way.

Table of Contents

Abstract ..... *i*

Acknowledgements ..... *iii*

List of Figures ..... *vii*

List of Tables ..... *ix*

List of Abbreviations ..... *xi*

Chapter 1: Introduction *1*

Chapter 2: Literature Review *3*

2.1 Introduction 3

2.2 Overview of New Zealand Sheep Breeds 4

2.3 The Evolution of New Zealand's Sheep Industry 5

2.4 Land Use and Farming Systems in New Zealand's Sheep Industry 6

2.5 Challenges and Pressures Facing Sheep Production 7

2.6 Future Directions in Breeding 9

2.7 Key Principles in Progeny Testing 11

2.8 Progeny Test Methodology and Design 13

2.9 Evolution of Sheep Genetic Evaluation in New Zealand 15

2.10 Impact and Future of Progeny Testing 20

2.11 Connectedness 22

2.12 Sire Attribution 26

2.13 Future Directions: Strengthening Genetic Connectedness 28

2.14 Identifying Optimal Levels of Connectedness 28

2.15 Conclusions 30

2.16 Research Question 31

Chapter 3: Methods and Methodology *33*

3.1 Data 33

3.2 nProve Connectedness Analysis....34
3.3 Accuracy of Genetic Evaluations using GCA Connectedness Metrics....37
3.4 Estimation of Genetic Parameters and Breeding Values....38
Chapter 4: Results....43
4.1 Descriptive Statistics of Phenotypic Datasets....43
4.2 nProve Connectedness....44
4.3 Accuracy of Genetic Evaluations for Facial Eczema using GCA Connectedness Metrics....55
4.4 Genetic Parameter Estimates and Breeding Value Evaluation....56
Chapter 5: Discussion and Conclusions....67
5.1 nProve Connectedness....67
5.2 Accuracy of Genetic Evaluations....69
5.3 Genetic Parameter Estimates and Breeding Value Evaluation....70
5.4 Implications....74
5.5 Limitations....75
5.6 Conclusions....77
References....79
Appendix 1: Dendrograms of genetic connectedness among flocks....89
Appendix 2: Paper for World Congress on Genetics Applied to Livestock Production....91

List of Figures

Figure 2.1: Timeline of key milestones in the evolution of sheep genetic evaluation in New Zealand, progressing from the early days (left) through to the present (right). 15

Figure 4.1: Base dendrograms illustrating genetic connectedness among flocks for (A) facial eczema and (B) growth, estimated using the nProve method. 45

Figure 4.2: Dendrograms illustrating genetic connectedness among flocks for facial eczema using the nProve method under varying evaluation window scenarios. (A) 2-year window (2023–2024 progeny birth years); (B) 4-year window (2021–2024 progeny birth years); (C) 5-year window (2020–2024 progeny birth years); and (D) 6-year window (2019–2024 progeny birth years). 46

Figure 4.3: Dendrograms illustrating genetic connectedness among flocks for growth using the nProve method under varying evaluation window scenarios. (A) 2-year window (2023–2024 progeny birth years); (B) 4-year window (2021–2024 progeny birth years); (C) 5-year window (2020–2024 progeny birth years); and (D) 6-year window (2019–2024 progeny birth years). 48

Figure 4.4: Dendrograms illustrating changes in flock connectedness for facial eczema (FE) following the removal of key flocks using the nProve method. (A) Flock F1 removed; (B) Flock F6 removed; and (C) Flock F18 removed. 51

Figure 4.5: Dendrograms illustrating changes in flock connectedness for growth following the removal of key flocks using the nProve method. (A) Flock F67 removed; (B) Flock F65 removed; (C) Flock F72 removed; and (D) Flock F62 removed. 53

Figure 4.6: Effect of data window length on across-flock evaluation accuracy for facial eczema, measured using the average prediction error variance of differences ($PEVD_{GrpAve}$) and coefficient of determination (CD_{GrpAve}). The x-axis shows the earliest birth year included in the evaluation. 56

List of Tables

Table 3.1: Evaluation window scenarios used throughout this study and the corresponding progeny birth years included in each scenario. 35

Table 4.1: Summary of the phenotypic datasets generated using the Geprep preprocessing programme for facial eczema and growth. For each year, the number of animals and flocks contributing data, as well as the mean and standard deviation (SD) of gamma-glutamyl transferase response measured 21 days post-challenge (log GGT21) for facial eczema and weaning weight (WWT) for growth are shown. A dash (-) indicates years with no available data for the trait. Total represents all available records from 1995 to 2025 for each dataset. . 43

Table 4.2: Number and percentage of flocks identified as connected (flocks that met the connectedness criteria) or unconnected, as inferred from the dendrograms generated under each evaluation window scenario in the nProve connectedness analysis for facial eczema. 47

Table 4.3: Number and percentage of flocks identified as connected (flocks that met the connectedness criteria) or unconnected, as inferred from the dendrograms generated under each evaluation window scenario in the nProve connectedness analysis for growth. 49

Table 4.4: Number and percentage of flocks identified as connected (flocks that met the connectedness criteria) or unconnected, as inferred from the dendrograms generated after sequential removal of key flocks in the nProve connectedness analysis for facial eczema. ... 52

Table 4.5: Number and percentage of flocks identified as connected (flocks that met the connectedness criteria) or unconnected, as inferred from the dendrograms generated after sequential removal of key flocks in the nProve connectedness analysis for growth. 54

Table 4.6: Number and percentage of flocks identified as connected (flocks that met the connectedness criteria) or unconnected, as inferred from the dendrograms generated after sequential masking of prolific sires in the nProve connectedness analysis for facial eczema. 55

Table 4.7: Number and percentage of flocks identified as connected (flocks that met the connectedness criteria) or unconnected, as inferred from the dendrograms generated after sequential masking of prolific sires in the nProve connectedness analysis for growth. 55

Table 4.8: Additive genetic variance (σ_a^2), residual variance (σ_e^2), heritability (h^2), and standard error of heritability (SE(h^2)) for gamma-glutamyl transferase response measured 21 days post-challenge (GGT21) under data depth, flock removal, and sire masking scenarios. The whole population scenario used the full pedigree and phenotypic dataset. Flock removal and sire masking analyses used phenotypic datasets restricted to the 3-year reference window (2022–2024 birth years). 57

Table 4.9: Pearson (below diagonal) and Spearman rank (above diagonal) correlation coefficients between estimated breeding values (EBVs) for facial eczema across 12 analysis scenarios, including data depth, flock removal, and sire masking. Spearman correlations reflect the degree of re-ranking between scenarios. All correlations were statistically significant ($P < 0.0001$); p-values are not presented. 61
Table 4.10: Additive genetic variance (σ_a^2), dam variance (σ_d^2) residual variance (σ_e^2), heritability (h^2), and standard error of heritability ($SE(h^2)$) for weaning weight (WWT) under data depth, flock removal, and sire masking scenarios. The whole population scenario used the full pedigree and phenotypic dataset. Flock removal and sire masking analyses used phenotypic datasets restricted to the 3-year reference window (2022–2024 birth years). 62
Table 4.11: Pearson (below diagonal) and Spearman rank (above diagonal) correlation coefficients between estimated breeding values (EBVs) for growth across 13 analysis scenarios, including data depth, flock removal, and sire masking. Spearman correlations reflect the degree of re-ranking between scenarios. All correlations were statistically significant ($P < 0.0001$); p-values are not presented. 66

List of Abbreviations

BLUP	Best Linear Unbiased Prediction
CD	Coefficient of determination
CPT	Central Progeny Test
EBV	Estimated breeding value
FE	Facial eczema
GCA	Genetic Connectedness Analysis
GGT21	Gamma-glutamyl transferase response measured 21 days post-challenge
NZGE	New Zealand Genetic Evaluation
PEVD	Prediction error variance of differences
SIL	Sheep Improvement Limited
ssGBLUP	Single-step genomic Best Linear Unbiased Prediction
WWT	Weaning weight

Chapter 1: Introduction

As global agricultural landscapes evolve, genetic improvement offers a lever for enhancing the sustainability and profitability of New Zealand's sheep sector. By accelerating the rate of genetic gain, farmers can respond more effectively to emerging challenges while also unlocking new opportunities for value creation. In the face of fluctuating wool prices (Ministry for Primary Industries 2026), changing climate conditions, and rising animal welfare expectations, the ability to make accurate data-driven selection decisions is more critical now than ever before. Whether the goal is faster growth, improved reproductive efficiency or enhanced tolerance to disease and environmental stresses, achieving these goals requires robust genetic evaluation systems. These systems must enable breeders and farmers to confidently identify and select animals with superior genetic merit.

The New Zealand Genetics Evaluation system (NZGE) supports much of the national selection process. It uses single-step genomic Best Linear Unbiased Prediction (ssGBLUP) to estimate breeding values (EBVs) and ranks animals based on their genetic merit (Lee et al. 2021). These EBVs guide selection decisions and ultimately influence the rate of genetic gain in the national flock. However, their accuracy depends on high-quality performance data, including results from progeny test programmes which assess the genetic merit of sires across multiple flocks (Yardibi et al. 2023).

The effectiveness of across-flock EBV comparison hinges on the concept of genetic connectedness, which refers to the degree of genetic linkage between animals in different flocks (Kuehn et al. 2007). Connectedness ensures that animals raised in different environments can be fairly compared on a common genetic scale (Brown et al. 2007). Although sire sharing is the primary method used to establish connectedness, uncertainty remains around the quantity and pattern of linking required to support reliable across-flock evaluations. For example, it is unclear how many sires must be shared between flocks to achieve sufficient connectedness, or how long historical connections can persist before re-linking is required.

These uncertainties have practical consequences for industry genetic evaluations. nProve, formerly known as Sheep Improvement Limited (SIL), the tool widely used by New Zealand sheep farmers to guide ram selection (Beef + Lamb New Zealand Genetics 2020), currently adopts a relatively conservative connectedness criterion, assessing linkage based on the number

of shared progeny from link sires within a rolling 3-year window (Holmes et al. 2015). This approach imposes criteria that may limit the number of flocks recognised as sufficiently connected for public reporting, potentially reducing the availability of across-flock comparisons for some flocks with meaningful genetic links.

This thesis investigates the level of genetic connectedness required to produce accurate and unbiased EBVs in the New Zealand sheep industry. Using data from the nProve database, this study evaluates how the depth of data influences connectedness and across-flock evaluation accuracy, and examines the impact of flock removal and sire masking. By providing evidence-based recommendations, this research aims to strengthen the robustness of across-flock genetic evaluations and determine whether less conservative connectedness criteria could maintain valid genetic comparisons while enabling a greater number of flocks to be included in reliable across-flock comparisons.

Chapter 2: Literature Review

2.1 Introduction

New Zealand's sheep industry has historically played a huge part in the success of the national economy through returns from both wool and meat. In fact, sheep had once dominated the New Zealand farming landscape (Morris 2013), but new challenges are facing the industry that are contributing to a decline in total sheep numbers. Traditional dual-purpose breeds once valued for both wool and meat may no longer be suited to an evolving market shaped by shifting consumer expectations, climate pressures, and economic realities. As wool prices have declined, farmers have increasingly prioritised lamb production, focussing on traits related to growth and reproduction. To ensure continued viability of the sheep sector, breeding objectives must expand to include traits such as disease tolerance, low environmental footprint, and welfare-related traits (Johnson et al. 2021).

The genetic evaluation of livestock plays a pivotal role in advancing these goals by providing breeders with tools to make informed selection decisions based on estimated breeding values (EBVs). High genetic merit sires are especially important in driving productivity and profitability improvements as they can pass on desirable traits to a large number of offspring (Young and Amer 2009). Progeny tests have long been a cornerstone of accurate genetic evaluations. By assessing the performance of an animal's offspring, progeny testing enables more accurate EBVs, which is especially important for traits where sires cannot be directly assessed through their own performance (Johansson and Robertson 1952) or traits that are difficult or expensive to measure directly, such as those related to reproduction or carcass quality. In New Zealand, the establishment of the Central Progeny Test (CPT) in 2001 marked a significant milestone in the genetic improvement for the national flock. The CPT enabled robust comparisons across flocks, contributing to long-term genetic gain (McLean et al. 2006).

The reliability of across-flock evaluations depends on sufficient genetic connectedness between flocks, often established through the use of shared sires. Connectedness ensures that animals raised in different management units can be evaluated on a common scale (Kuehn et al. 2007). The effectiveness of selection can be compromised by poor connectedness or errors in pedigree recording, particularly incorrect sire attribution.

This review analyses current literature on the structure of New Zealand's sheep industry, the role of progeny testing in genetic improvement, and the importance of genetic connectedness for robust across-flock comparisons. It examines how progeny testing has supported genetic progress to date and considers its future potential in maintaining a resilient and competitive sheep sector. Additionally, it explores how connectedness is measured, the consequences of incorrect sire attribution and strategies for enhancing the reliability of across-flock evaluations.

2.2 Overview of New Zealand Sheep Breeds

Due to the diverse climate, landscape and production systems found throughout New Zealand, a wide range of sheep breeds have been utilised. The industry has a significant proportion of composite and crossbred animals that are better suited to the environment and New Zealand farmer's commercial needs (Brito et al. 2017). Initially, Merinos imported from Australia in the 1800s were favoured due to their fine wool but were not suitable for meat production or the heavy soils and wet conditions often found on New Zealand farms (Cranston et al. 2023). Farmers then looked to dual-purpose breeds suitable for both meat and wool production, which led to the importation of British breeds including English Leicester, Lincoln, Romney, Cheviot, and Border Leicester. Many of these breeds were used for crossing to produce new breeds which farmers still use today including the Corriedale, Perendale and Coopworth (Johnson et al. 2021). In the 1970s the price of wool began to decline (Cocks and Brown 2005), driving the shift from dual-purpose sheep to a focus on lamb production.

Romney was originally favoured as a dual-purpose breed, and still remains the most predominant breed making up around 50% of New Zealand's breeding ewe population (Beef + Lamb New Zealand Economic Service 2020). Composite breeds now make up a large portion of the rest of the national flock as breeders look to take advantage of desired traits present in different breeds. Incorporating the genetics from breeds such as Texel, Finnish Landrace and East Friesian into traditional Romney, Coopworth or Perendale breeds has allowed breeders to combine the strengths of multiple breeds to improve production efficiency and overall flock performance (Cranston et al. 2023). New dairy breeds including Zealandia™ and Southern Cross™ have also been developed and registered in New Zealand, utilising imported genetics from Europe (McCoard et al. 2023).

2.3 The Evolution of New Zealand's Sheep Industry

The success of New Zealand's sheep industry is largely built upon its capacity to export products to large overseas markets. Historically, wool was the industry's primary commodity, with exports to the United Kingdom driving a profitable and robust market. The development of meat transportation to international markets later introduced an additional revenue stream, enhancing the viability of small sheep farms (Cranston et al. 2023). While the relative value of sheep products has shifted overtime, New Zealand remains a dominant player in the global sheep export market. In 2023, New Zealand supplied around 28% of the world's fresh or chilled sheep meat exports and 31% of degreased or carbonized wool exports by quantity (Food and Agriculture Organisation of the United Nations 2023). Key export markets for sheep products include the European Union, China and the United States (Ministry for Primary Industries 2024a). Lamb meat now constitutes the industry's primary export, with revenue from lamb exports reaching $3.179 billion and mutton exports totalling $407 million in 2024, positioning New Zealand as the largest exporter of sheep meat worldwide (Morris and Kenyon 2014; Ministry for Primary Industries 2024a). The total sheep population within New Zealand has declined from peaking at 70 million in 1982, to just 23.3 million in 2024 (Beef + Lamb New Zealand Economic Service 2024). Despite sheep numbers declining, sheep meat volumes have not experienced a corresponding decrease and continue to remain reasonably stable (Johnson et al. 2021). Total revenue from wool exports reached $448 million for 2024, the highest in the past five years (Ministry for Primary Industries 2024a).

Wool prices are largely influenced by which class of wool it corresponds to: coarse, medium or fine. Majority of wool produced in New Zealand is coarse and is primarily used for interior furnishings; however, finer wool such as that from the Merino breed commands a higher value (Johnson et al. 2021). Shifting away from coarse wool in breeds favoured for meat-production characteristics and breeding in a finer wool phenotype could offer economic benefits. Another potential alternative is developing a breed that produces little to no wool, as many farmers do not achieve a positive net return from coarse wool production. The skin of sheep, also referred to as 'pelts' have long been a by-product in New Zealand lamb production, generating no return for farmers. As such, little interest has been shown in pelts from both researchers and sheep breeders (Näsholm and Eythorsdottir 2011).

New Zealand's commercial dairy sheep industry is relatively new, and as early selection efforts were primarily focused on lamb production, very little was done on selecting for persistent

milk production. Consequently, traditional sheep breeds tend to reduce milk yield after their lactation peak, making it challenging to achieve a profitable return for farmers (Peterson 2017). Growing demand for sheep dairy products has driven the establishment of sheep dairy farms, which can be broadly categorised into small-scaled family-owned operations or larger commercial-scale operations (McCoard et al. 2023). Whilst small-scale operations typically cater to the domestic market, large-scale operations have facilitated industry growth, particularly through research and breeding programmes. The number of commercially milked ewes in New Zealand is now estimated at approximately 30,000 (McCoard et al. 2023).

2.4 Land Use and Farming Systems in New Zealand's Sheep Industry

The types of land used for farming sheep in New Zealand are highly diverse, with variations in sheep breeds and sheep performance across the different types. Intensive sheep farming is commonly practiced on lowland areas, where high pasture growth rates, and where necessary, irrigation support high levels of animal production per hectare (Morris 2009). This includes increased lamb growth rates and greater kilograms of carcass meat produced per hectare. These farms typically support a much higher stocking rate, often exceeding 10 stock units per hectare, compared to hill or high-country systems (Cranston et al. 2023).

Hill or high-country farms are characterised by steep terrain, lower soil fertility, and thus breeds used for these systems are selected for their ability to withstand challenging environmental conditions. Many of these farms carry fine wool sheep, such as Merino, resulting in a greater proportion of their gross income being derived from wool sales compared to intensive lowland farms (Morris 2009; Cranston et al. 2023).

Stud farms also play an important part of New Zealand's sheep industry with a focus on genetic improvement and the development of superior breeding stock. This enables the breeding of high-quality rams that are sold to commercial farms to improve their productivity and profitability (Cranston et al. 2023).

Sheep farming holds a comparative advantage over other livestock systems due to the ability of sheep to thrive and forage in rugged terrain where cattle performance is often limited (Morris 2009). In addition, sheep production typically has a lower environmental footprint than dairy or beef production systems. This is attributed to reduced soil damage, lower water requirements, and a more uniform distribution of urine, which minimises localised nutrient

loading and associated environmental impacts (Johnson et al. 2021). As a result, sheep farming remains a viable strategy for utilising less arable land, contributing to both environmental conservation and economic viability in such regions. However, advances in innovation and technology have enabled land previously used for sheep farming to be converted for dairy production through the application of irrigation and fertiliser. This shift is largely driven by the higher productivity and increased saleable output that dairy farming offers (Copland and Stevens 2012). Moreover, government initiatives aimed at increasing carbon sequestration have driven the conversion of portions of sheep farms, or in some cases entire farms, into forestry blocks (Orme 2025). These land-use changes have significantly contributed to the decline in sheep numbers.

2.5 Challenges and Pressures Facing Sheep Production

Whilst the shift in land use poses significant challenges for sheep farmers, numerous other factors are forcing farmers to re-evaluate their production systems. Among these, climate change is a major driver. For example, the severe drought in the 2023–2024 season (NIWA 2024) was identified as the primary contributor to a 3% decline in sheep numbers over this period as farmers were required to reduce stock numbers to match the limited feed supply (Ministry for Primary Industries 2024a). Rising global temperatures are also impacting pasture growth patterns as well as animal health. Heat stress in livestock has gained increasing attention as rising temperatures are expected to expose sheep to greater levels of thermal stress, negatively affecting their performance and productivity (Ministry for Primary Industries 2024b). Studies have shown that exposure to temperatures exceeding 32°C significantly impacts reproduction performance of both ewes and rams (van Wettere et al. 2021). However, there is limited research on the extent to which heat stress affects wool or meat-producing sheep breeds, highlighting a gap in the literature.

Alongside these pressures, disease presents a significant challenge for sheep farmers nationwide, with climate and geographic location largely influencing the predominant diseases that require management for a specific farm. Current management strategies often rely heavily on chemical inputs for disease treatment and control, which conflicts with societal expectations for more sustainable and natural farming practices (Johnson et al. 2021). Several of these increasingly prevalent diseases are discussed in more detail below.

Facial eczema (FE) is a fungal disease that predominantly affects sheep in the warmer regions of New Zealand. It arises when animals ingest sporidesmin, a toxin produced by the fungus *Pithomyces chartarum*, leading to liver damage, reduced feed intake, and overall production losses (Morris et al. 2004; Jordan 2020). Current control strategies such as zinc supplementation, pasture spraying with fungicides, and pasture management are relatively effective but can be costly and labour-intensive (McRae et al. 2021). Despite these efforts, FE continues to impose a substantial economic burden on New Zealand's livestock industries through lost fertility and production. The total economic cost to the livestock sector has been estimated at approximately NZD$332 million per year (McClay 2024). This challenge is expected to intensify as climate change is predicted to increase the distribution of the causative fungus (Phillips et al. 2023).

Gastrointestinal parasites also pose a significant challenge to sheep farmers nationwide, with anthelmintic resistance increasingly limiting the effectiveness of control strategies (Brosnahan et al. 2025). This results in reduced animal performance and ongoing economic losses, stemming from reduced productivity and costs associated with parasite management (Johnson et al. 2021).

Flystrike also remains a persistent issue for the sheep industry. The condition occurs when flies lay eggs on soiled or damp wool, with the developing larvae feeding on the sheep's tissue (Pickering 2013; Cranston et al. 2023). Management practices such as insecticide application and shearing are commonly used to control its impact.

Another significant challenge facing sheep farmers is the increasing societal expectation to uphold high standards of animal welfare, particularly regarding issues such as flystrike, tail docking, and parasite management, while also adopting best practices and achieving environmental sustainability. Farmers are required to adapt their operations to comply with stricter regulations and meet evolving consumer demands, ensuring they retain their social licence to operate (Beban et al. 2024). This is critical for the long-term viability of the sheep industry, as it facilitates access to domestic and international markets. Booth et al. (2024) reported that the characteristics deemed most important for defining a 'good farmer' were largely aligned between farmers and the general public, particularly regarding animal welfare, environmental stewardship, ethical farm management, and compliance with regulations. Increasingly, transparency is a key expectation for consumers who are more interested in understanding where their food has come from and how it is produced (Johnson et al. 2021).

Whilst it is often stigmatised that the media amplifies negative portrayals of the agricultural sector, Booth et al. (2024) also highlighted that farmers are generally regarded as more trustworthy than the mainstream media as a source of reliable information. This trust presents an opportunity for farmers to engage with the public to leverage their creditability and strengthen societal support and understanding of agricultural practices.

2.6 Future Directions in Breeding

Future sheep breeding programmes are expected to address the complex challenges confronting the industry. Breeding programmes must focus on reducing the environmental footprint of sheep farming, enhancing disease resistance and improving animal welfare. This section reviews these three critical areas, discussing the role of genetics in fostering a more sustainable, profitable and socially responsible sheep industry.

2.6.1 Reducing Environmental Impact

Environmental sustainability has become a critical focus in modern agriculture. Genetics should be utilised to better align the sheep industry with societal environmental goals. Among these efforts is the reduction of greenhouse gas emissions. Agriculture remains the single largest contributor to New Zealand's greenhouse gas emissions, with methane comprising a substantial proportion of agriculture emissions. Sheep and beef cattle together accounted for approximately 50% of enteric methane emissions from the agriculture sector in 2019 (Ministry for the Environment 2021). While the decline in sheep numbers has led to a reduction in the absolute emissions released by the sheep industry (Vibart et al. 2022), further progress can be achieved through genetic improvements. Methane emissions per unit of dry matter intake have been shown to be heritable, presenting an opportunity to breed low-methane-emitting sheep (Jonker et al. 2018). The other main greenhouse gas from the agricultural sector is nitrous oxide which predominantly originates from the urine and dung excreted by livestock (Johnson et al. 2021). While no studies have yet explored whether genetic variation for nitrous oxide emissions exists in sheep and whether it could be targeted through selective breeding, this could be an interesting area for future research.

2.6.2 Managing Disease Risk

Genetic improvement provides an opportunity to reduce the impact of several major disease challenges affecting the New Zealand sheep industry. Compared with ongoing chemical or

management-based interventions, breeding for disease resilience offers a more permanent and cumulative long-term solution.

Breeding for enhanced FE tolerance is one promising strategy, as moderate to high heritability estimates have been reported for this trait (Morris et al. 1995; McRae et al. 2016). However, FE tolerance is not widely measured because current tools, such as the RamGuard™ service, require dosing rams with the sporidesmin toxin to assess their response, a process that can be costly and ethically challenging (Amyes and Hawkes 2014).

Similarly, genetic variation in resistance to gastrointestinal parasites has been identified across several parasite species, offering a promising avenue for breeding more resistant sheep. Progress has already been observed in certain flocks, where faecal egg count has been used as an indicator trait for selecting resistance (Bishop and Morris 2007). This approach is likely to gain increasing importance as we see internal parasites continue to develop resistance to anthelmintic drenches (Pickering et al. 2015).

Genetic approaches are also being explored to reduce susceptibility to flystrike. This is reflected in the inclusion of dag score, a key indicator trait associated with flystrike risk, within the national genetic evaluation system (Johnson et al. 2021). Additionally, research has identified quantitative trait loci associated with the dagginess trait, highlighting the potential for genomic tools to accelerate genetic gains in flystrike resistance (Pickering et al. 2015).

2.6.3 Improving Animal Welfare

Societal expectations are likely to see an increased emphasis put on ethical farming practices and enhanced livestock well-being, with breeding expected to play a role in meeting these demands. Tail docking, a routine practice among New Zealand sheep farmers, is expected to be challenged in the future due to animal welfare concerns. Breeding for shorter tails offers a permanent solution to improve animal health by reducing the incidence of dags and flystrike (Johnson et al. 2021). Unlike New Zealand sheep breeds, many international breeds exhibit naturally shorter tails. James (2006) reported that tail length is often a quantitative trait with moderate to high heritability. Similarly, research has shown that breech bareness is heritable and is associated with lower dag scores, indicating a reduced risk of breech strike (Scobie et al. 2007). Selecting for belly and breech bareness represents another opportunity to mitigate flystrike and enhance sheep welfare.

While many of the traits important for improved resilience, welfare, and environmental sustainability already exist in New Zealand's sheep breeds, their introgression and widespread adoption will require substantial time and investment. The importation of overseas genetics remains an option, though it comes with challenges such as strict biosecurity regulations and disease risk (Johnson et al. 2021).

2.7 Key Principles in Progeny Testing

Achieving genetic progress in any trait relies upon accurate identification of animals with superior genetic merit. The following section outlines the key principles of progeny testing, a fundamental approach used to evaluate genetic merit and support effective selection decisions in breeding programmes.

2.7.1 Evaluating Genetic Merit through Progeny Performance

A progeny test is a structured approach used to evaluate an animal's genetic merit by assessing the performance of a reasonably large number of its offspring, and has been used across many species for extended periods of time (McLean 1948). By observing the expression of inherited traits in an animal's progeny, breeders can distinguish genetic influences from environmental effects with greater accuracy. Most traits of economic importance in livestock production systems are quantitative in nature, meaning that they are controlled by many (often unknown) genes each contributing a small effect (Johansson and Robertson 1952). Therefore, the heritability of a trait becomes a key consideration in genetic evaluation as it reflects the proportion of phenotypic variation that is attributable to the genetic differences among animals (Lush 1940). Understanding heritability is essential for predicting breeding values and understanding the extent to which a trait will be passed on to the offspring when an animal is selected for breeding (Johansson and Robertson 1952). Without progeny records, an individual's EBV from a single observation of a given trait is estimated as their phenotypic deviation from the population mean, multiplied by the heritability of that trait (Falconer and Mackay 1996). The accuracy of that prediction is the square root of the heritability (Falconer and Mackay 1996), meaning that lowly heritable traits have low accuracy without additional records or progeny.

2.7.2 Across-Flock Comparisons in Progeny Testing

A key challenge faced by early sheep breeders was the difficultly of reliably comparing EBVs of rams across different flocks. This was largely due to the inability to disentangle genetic influences from environmental and management effects, which limited the accuracy of selection decisions and ultimately slowed the rate of genetic gain (Young and Newman 2009). To address this, the industry moved towards across-flock comparisons, which required genetic linkages between flocks to enable meaningful comparisons across breeding groups. The two methods used to achieve this included direct comparisons in central testing stations, and the use of reference sires across multiple flocks with statistical adjustments to account for management unit effects (Kuehn et al. 2007). Central testing involves raising animals from different breeders under identical conditions to remove environmental and management unit bias. Alternatively, reference sires can be used across flocks to create genetic connectedness, allowing for reliable benchmarking of sires from different breeding groups (Clarke 2003).

2.7.3 Within-Flock Progeny Testing

While across-flock comparisons have driven industry-wide genetic improvement, within-flock progeny testing has also played a vital role in improving genetic evaluations at the flock level. These tests evaluate sires within a single environment and allow breeders to estimate genetic merit more precisely within a closed population. An early example is the study by McLean (1948) in which progeny testing was used to assess Southdown rams by mating them to comparable groups of Corriedale ewes. The lambs were all raised together under uniform conditions so that researchers could evaluate heritable differences in traits based on within-flock performance comparisons. Such tests offer valuable information for selection decisions in flocks that are not genetically connected to broader breeding populations. However, their utility is limited due to the lack of comparability across flocks. As Notter (1998) noted, this limitation was a driving force behind the reassessment of the U.S National Sheep Improvement Program in 1993 which led to the development of across-flock genetic evaluations to improve the accuracy and scope of genetic merit estimation. In New Zealand, the establishment of Sheep Improvement Limited (SIL) enabled breeders to compare animals across flocks more effectively, overcoming a key limitation of within-flock evaluations and facilitating greater rates of genetic gain (Newman et al. 2000; Newman 2003).

2.8 Progeny Test Methodology and Design

Progeny testing methodologies vary depending on the breeding objective, traits of interest and logistical constraints, but generally follow a structured approach to ensure reliable genetic evaluation. Key components include experimental design, statistical modelling, and interpretation of genetic evaluation metrics.

2.8.1 Experimental Design

Designing a progeny test requires balancing several factors, including the number of offspring recorded per sire, the mating structure, and the logistical (and financial) capacity of the testing programme (Rae 1948). Progeny group size directly affects the accuracy with which a sire's genetic merit can be estimated. Larger groups increase the reliability of EBVs, but testing excessively large progeny groups can reduce overall genetic gain by limiting the number of sires evaluated when total test capacity is fixed (Rae 1948). Early work by Rae (1948) demonstrated the principle of diminishing returns, showing that accuracy gains plateaued once approximately 14–15 daughters per sire were evaluated. Modelling by Dekkers et al. (1996) in the context of dairy artificial insemination (AI) companies indicated optimal progeny group size of around 57–61 daughters per bull under high-intensity commercial selection. These studies highlighted that the optimal progeny group size is not fixed, but depend on the programme's structure, economic constraints, and selection objectives, which are all factors that differ between cattle AI systems and dispersed progeny test programmes in sheep.

Another key consideration in progeny test design is the choice of mating method. Artificial insemination remains limited in its application within the New Zealand sheep industry, largely due to practical constraints including cost, labour requirements and relatively low conception rates compared with natural mating (Blair and Garrick 2007). However, AI has proven to be valuable in allowing more precise allocation of sires across different flocks to strengthen genetic links for across-flock comparisons (Notter 1998).

2.8.2 Statistical Models

Statistical modelling plays a central role in progeny testing with Best Linear Unbiased Prediction (BLUP) long serving as the standard method for predicting EBVs in the New Zealand sheep industry (Lee et al. 2021). This method incorporates phenotypic performance data, pedigree records, and fixed effects into mixed model equations to estimate EBVs. The basic BLUP model is:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}$$

Where **y** is the vector of observed phenotypes, **b** is the vector of fixed effects, **u** is vector of breeding values, and **X** and **Z** are the incidence matrices that relate observations to fixed and random effects, respectively (Henderson 1975).

Under this model, breeding values are random effects distributed with expectation E(u) = 0, and variance Var(u)=$\mathbf{A}\sigma_a^2$, where **A** is the numerator relationship matrix derived from pedigree information, and σ_a^2 is the additive genetic variance. Residual effects are assumed to be independently distributed with expectation E(e) = 0, and variance Var(e)=$\mathbf{I}\sigma_e^2$, where **I** is the identity matrix, and σ_e^2 is the residual variance. The model further assumes that breeding values and residual effects are uncorrelated, such that $\mathrm{Cov}(u, e) = 0$ (Henderson 1985).

Management units can be fitted as fixed effects to allow for valid comparisons of animals across different flocks, farms or groups (Kuehn et al. 2007). More recently, the introduction of genomic information through single-step genomic BLUP (ssGBLUP) has improved the accuracy of EBVs, particularly for young animals or those with limited progeny data. Unlike traditional BLUP, ssGBLUP replaces the relationship matrix with a matrix that integrates both pedigree and genomic information (Lee et al. 2021).

2.8.3 Genetic Evaluation Metrics

Progeny test results are interpreted through genetic evaluation metrics with EBVs serving as the primary tool for quantifying an animal's genetic merit. Estimated breeding values predict an animal's ability to pass on desirable traits to its offspring, allowing farmers to rank animals and make informed selection decisions that will accelerate the rate of genetic gain (Yardibi et al. 2023). Progeny testing can significantly improve the accuracy of EBVs by providing reliable phenotypic data from multiple offspring, which will reduce the influence of environmental variation. Blair and Garrick (2007) highlighted that achieving high selection accuracy where an individual's phenotype reliably represents its genotype for quantitative traits, requires extensive progeny testing. Thus, the effectiveness of EBVs can depend on how well progeny testing is integrated into structured breeding programmes.

2.9 Evolution of Sheep Genetic Evaluation in New Zealand

Sheep genetic evaluation in New Zealand has progressed from informal, visual selection of breeding stock to complex, national-scale genomic evaluations (Figure 2.1) (Morris 1980; Lee et al. 2021). These advances have been driven by the need to improve production traits and to provide reliable information for ram breeders and commercial farmers, supported by the development of statistical methods and computational power that have enabled the analysis of industry scale datasets.

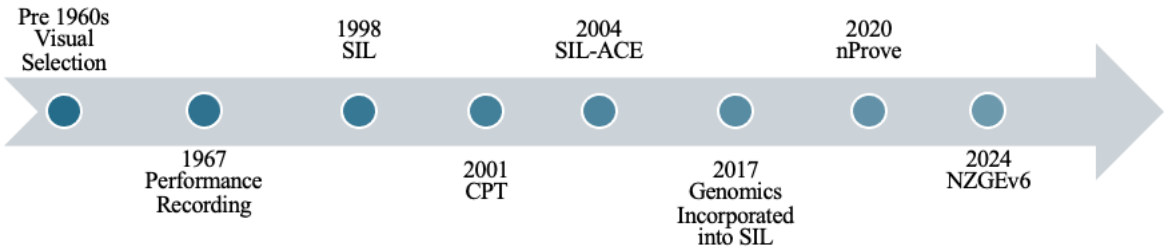


Figure 2.1: Timeline of key milestones in the evolution of sheep genetic evaluation in New Zealand, progressing from the early days (left) through to the present (right).

2.9.1 Visual Selection (Pre 1960s)

In the early stages of sheep evaluation in New Zealand, farmers relied almost entirely on phenotypic assessment to select breeding stock. Selection focused on traits that could be evaluated on appearance such as body size, fleece weight, and conformation (Blair et al. 1984). With Romney sheep dominating hill country systems and Merinos common in high country regions, preferred animals were those that appeared robust, carried a dense fleece, and showed signs of good fertility (Cranston et al. 2023). However, the effectiveness of this approach was limited, as environmental influences often masked genetic differences. Production traits of economic importance such as lambing percentage, lamb survival, and growth rate, could not be reliably improved through visual assessment alone. Due to their low heritability, appearance was a poor predictor of genetic merit leading to slow rates of genetic progress during this period (Cruickshank 2003).

2.9.2 Performance Recording (1967)

The establishment of the National Flock Recording Scheme (NFRS) in 1967 marked the beginning of objective performance recording in New Zealand sheep breeding (Cruickshank 2003). The scheme enabled the systematic collection of fertility data, such as the number of lambs reared per ewe and the total weight of lamb weaned, alongside fleece traits, including

weight, character and count (Clarke 1967). Unlike visual selection, performance recording provided breeders with quantitative data that more accurately reflected genetic merit. This allowed selection decisions to be guided by measurable outcomes including fertility records, wool production data, and combined indices such as the fertility and fleece weight index (Clarke 1967). By the 1970s, ram breeders increasingly incorporated performance recording into their selection programmes to supply commercial farmers with rams that offered higher reproductive output and greater lamb production (Craw 2009). This marked a transition from selecting animals primarily on breed type or appearance toward identifying individuals with superior performance records, enabling faster genetic improvement in fertility and growth. However, following the introduction of the NFRS, several recording schemes continued to operate independently, resulting in fragmented data and inconsistent genetic analyses (Geenty 2000).

2.9.3 Establishment of Sheep Improvement Limited (1998)

The establishment of SIL in 1998 represented a major advancement in New Zealand's sheep evaluation systems. Sheep Improvement Limited was developed to manage a centralised national database that standardised genetic evaluations across the industry, delivering consistent, accurate, and accessible outputs for farmers and breeders (Cruickshank 2003). SIL provided a system to deliver a consistent national genetic evaluation system, but the breeding values were on a within-flock basis or on an across-flock basis for the breed-based sire reference programmes present at that time (Newman et al. 2000). The reliability and ease of use encouraged wider adoption of performance recording.

The creation of SIL was intended to improve decision-making across the sheep industry. For commercial ram buyers, SIL indexes provided a simple practical tool for comparing rams and identifying those with superior overall genetic merit for traits of economic importance. For ram breeders, the system produced EBVs, enabling more accurate within-flock selection based on objective genetic evaluations (Newman et al. 2000). This was particularly important at a time when wool returns were declining, and the economic focus of the industry had shifted towards meat production traits such as reproductive efficiency, carcass quality, and lamb growth. While Romney remained the predominant breed in commercial flocks, composite breeds were beginning to increase in popularity (Cranston et al. 2023). By supplying both buyers and breeders with performance data, SIL supported the production of rams capable of improving

commercial flock productivity while also strengthening the genetic base of breeding nuclei (Cruickshank 2003).

2.9.4 Establishment of the Central Progeny Test (2001)

The CPT was launched in 2001 as a key initiative to advance New Zealand's sheep evaluation framework. The underlying aim of the CPT was to create genetic connections across both flocks and breeds to enable a single across-breed genetic analysis for the New Zealand sheep industry (Beef + Lamb New Zealand 2012). It comprised of a national progeny testing programme in which selected rams from as many breeds and sire-reference groups as possible were mated to ewes at designated sites, with their offspring evaluated across a wide range of performance traits. Funded initially by Alliance Group Ltd, the first matings were conducted in 2002 where both terminal and dual-purpose rams were tested primarily for growth rate and carcass merit (Beef + Lamb New Zealand 2012). From 2004 onwards, assessments of maternal traits were also incorporated for dual-purpose rams, broadening the scope of evaluation to encompass overall productivity (Beef + Lamb New Zealand 2012). Early mating designs aimed to generate a minimum of 20 lambs per sire, and the resulting performance data was analysed using BLUP to generate EBVs for the tested sires. Results from CPTs conducted between 2002 to 2005 revealed considerable genetic variation among sires, with a difference of approximately $5 in progeny returns per sire based on the meat and growth index (McLean et al. 2006).

The CPT was designed to improve carcass value in sires and to establish genetic linkages across flocks and breeding programmes, thereby strengthening the accuracy of across-flock genetic evaluations (McLean et al. 2006). By benchmarking rams from previously independent breeding operations, the CPT helped create the connectedness required for robust national genetic evaluations (Beef + Lamb New Zealand Genetics 2015). Progeny from CPT sires were monitored for three years across a wide range of traits, with the resulting data contributed directly into the New Zealand Genetic Evaluation (NZGE) (Beef + Lamb New Zealand Genetics 2019). Outputs were expressed as standardised indexes, including the New Zealand Maternal Worth (NZMW) index which is used when breeding replacement stock, incorporating capped reproduction, survival, growth, and adult size. Whereas the New Zealand Terminal Worth (NZTW) index focuses on terminal sire performance, namely survival, growth, and meat yield (Beef + Lamb New Zealand Genetics 2019). Indexes may also be tailored to include additional traits of industry relevance, such as resistance to FE.

The Beef + Lamb New Zealand CPT was expanded in 2019 to investigate genotype by environment interactions through the establishment of three 'hub' and two 'Next Generation' sites (Beef + Lamb New Zealand Genetics 2019). Hub sites were used to strengthen genetic connectedness across the national flock by evaluating rams from a range of flocks and breeds (Beef + Lamb New Zealand Genetics 2019). In contrast, Next Generation sites focused on specific breed types or production challenges relevant to local environments. These sites fostered stronger genetic links between breeding operations and offered ram breeders an opportunity to test their rams under commercial farming conditions (Beef + Lamb New Zealand Genetics 2019). An important addition to the CPT has been the Low Input Sheep Progeny Test which was developed to evaluate sheep genetics that perform well with minimal intervention, reflecting the growing market demand for more sustainable and ethically produced meat (Beef + Lamb New Zealand Genetics 2023).

The Next Generation and Low Input programmes were completed in 2024. The core CPT programme is now run on two Hub sites, one in the South Island and one in the North Island, with the aim of maintaining and strengthening across breed genetic connections to support the national across-breed analysis (N Jopson pers. comm).

2.9.5 Introduction of Sheep Improvement Limited Advanced Central Evaluation (2004)

The expansion of progeny testing through the CPT built on existing sire referencing practices used by breeders and further strengthened the genetic connectedness necessary for national across-flock genetic evaluations (Sise et al. 2012). This framework provided the foundation for the SIL Advanced Central Evaluation (SIL-ACE), the first iteration of the NZGE and the next major step in national sheep evaluation. Introduced in 2004, SIL-ACE marked the beginning of large-scale evaluations across flocks and breeds in New Zealand, representing a significant shift from earlier within-flock assessments (Young and Newman 2009). This made it possible for breeders to benchmark rams against a much wider population, allowing direct genetic comparisons and more informed selection decisions.

Evaluations were conducted using genetically connected flocks, ensuring that EBVs reflected true genetic differences rather than environmental or management effects. Animals were required to meet minimum progeny or sibling measurement criteria for each trait to be included, and only well-connected sires were reported in leader lists for various indexes and traits. These evaluations were carried out routinely with results published on the SIL website (Young and Newman 2009). This SIL-ACE framework has since developed into the current NZGE which

now operates as a single weekly evaluation incorporating all active flocks recorded on SIL/nProve. By integrating performance and pedigree data from across the national flock, NZGE produces standardised EBVs and indexes that underpin selection decisions (Jopson et al. 2015).

2.9.6 Incorporation of Genomics into Sheep Improvement Limited (2017)

Genomic information was first incorporated into the SIL evaluation system in 2017 using a multi-step approach (Beef + Lamb New Zealand 2024). Using this method, pedigree and genomic analyses were performed separately and the resulting information was then combined to enhance breeding value predictions (Lee et al. 2021). Since 2018, Beef + Lamb New Zealand has adopted a ssGBLUP methodology in their NZGE which simultaneously incorporated pedigree, performance, and genomic data into a single analysis (Lee et al. 2021). Incorporating genomic data improved the accuracy of EBVs beyond what was possible with traditional pedigree and performance records alone. This gain in accuracy was particularly valuable for traits that are difficult or costly to measure (e.g. FE or carcass quality), expressed in only one sex (e.g. maternal traits), expressed later in life (e.g. reproduction), or only once in life (e.g. carcass traits). Genomic prediction also enabled reliable evaluation of young animals before phenotypic information is available, shortening the generation interval (Kaseja et al. 2023). The use of genomic EBVs in breeding programmes provides the basis for genomic selection and has the potential to substantially accelerate genetic gain across the national flock (Scholtens et al. 2021).

2.9.7 Launch of nProve (2020)

Beef + Lamb New Zealand Genetics launched nProve in 2020 to make the outputs of NZGE more accessible to commercial farmers (Beef + Lamb New Zealand Genetics 2020). Powered by NZGE, the tool draws directly on EBVs and indexes from the national evaluation to provide a farmer-facing interface for identifying ram breeders offering genetics aligned with specific production goals. The tool employed interactive filters including buttons and slider scales through which farmers can specify desired ram attributes such as growth, fertility or maternal performance. Based on these selections, nProve generated customised lists of breeders whose rams best match the chosen objectives (Rural News 2020). By streamlining access to genetic information and replacing earlier SIL tools such as FlockFinder and RamFinder, nProve represents an important development in integrating national genetic evaluations into practical on-farm ram selection (Beef + Lamb New Zealand Genetics 2020).

2.9.8 Genetic Evaluation Version 6 (2024)

While nProve focused on making existing evaluations more accessible to farmers, the launch of NZGEv6 was a major technical upgrade to the underlying evaluation system itself. In 2024, Beef + Lamb New Zealand Genetics rolled out Genetic Evaluation Version 6 (NZGEv6 comprising of GTABLUP), designed to support wider use of genomic information and to strengthen the evaluation system's capacity to include new traits (Beef + Lamb New Zealand Genetics 2024; Rural News 2025). This system has expanded capacity and enabled the inclusion of millions of genotypes from multiple breeds into a single evaluation. Integration of large and diverse datasets improved the precision and scope of breeding value predictions (Beef + Lamb New Zealand 2024). In addition, NZGEv6 also enhanced the evaluation of novel or less commonly recorded traits. Even when very few animals have data for these traits, the system can still make useful predictions allowing farmers to begin incorporating these traits into breeding objectives (Beef + Lamb New Zealand 2024). However, the accuracy and effectiveness of these national evaluations depend on high-quality performance data, much of which is generated through structured progeny test programmes.

2.10 Impact and Future of Progeny Testing

2.10.1 Impact on Genetic Gain

The implementation of progeny test programmes, such as the CPT, has played a pivotal role in accelerating genetic gain across the New Zealand sheep industry. A study by Amer (2009) quantified annual rates of genetic progress over four distinct periods between 1990 and 2006. The period from 2004 to 2006 followed the joint introduction of the CPT and the SIL-ACE genetic evaluation and showed a marked increase in genetic progress. For dual purpose flocks, annual improvement in the overall index rose from 54 to 84 cents, while terminal sire flocks saw a gain from 35 to 48 cents per year (Amer 2009). A subsequent study by Sise et al. (2012) expanded this analysis using a larger dataset and confirmed the association between the introduction of CPT and a period of accelerated genetic gain. However, this study reported slightly lower average progress rates for the post-CPT period compared to those reported by Amer (2009), likely reflecting the inclusion of additional flocks with lower average genetic merit in the expanded dataset.

Recognition of progeny testing as a powerful tool for genetic improvement dates to the 1930's (Lush 1935). A study by Rae (1984) indicated that incorporating progeny test data into selection decisions for lean meat traits enhanced genetic progress, achieving gains of 40–50% higher compared to individual-based selection. This highlights the advantage gained from evaluating sires based on their offspring's performance, especially for traits that are difficult or expensive to measure directly. More recent work expanded on this principle using advanced statistical and genomic tools such as genomic BLUP. Bolormaa et al. (2017) reported that incorporating a progeny test measure known as daughter trait deviations into genomic evaluations for reproduction traits improved the accuracy of EBVs compared to using ewe records only.

2.10.2 Challenges and Future Directions for Progeny Testing

Building on these gains will require scientists and breeders to address several ongoing challenges and adapt to emerging consumer demands. One notable limitation of progeny testing is the time lag between mating and the availability of performance data to inform selection decisions. While reducing the generation interval is a key strategy for accelerating genetic gain, it often comes at the cost of reduced accuracy (Meuwissen et al. 2013). As such, the annual rate of genetic improvement through progeny testing is constrained by the age at which reliable progeny performance data can be obtained for a given sire (Rae 1948). Another barrier is the high cost of progeny testing, particularly for smaller breeding operations. Expenses include mating costs such as AI procedures or the management of natural mating, as well as costs related to ewe management, monitoring of lambing, and the collection of performance records. Livestock maintenance costs to feed and manage stock until EBVs can be obtained represent a financial burden (König et al. 2009; Kariuki et al. 2017).

To address the limitations of time and cost inherent in a traditional progeny test, there is growing interest in the use of genomic tools such as single nucleotide polymorphism (SNP) genotyping and genomic selection. These methods offer potential to significantly reduce generation intervals while maintaining or improving accuracy (Schaeffer 2006). Studies from the dairy industry have indicated that genomic selection using breeding values predicted directly from an individual's DNA, can lower the cost per unit of genetic gain, primarily by enabling earlier and more accurate selection of high-merit sires (König et al. 2009; Kariuki et al. 2017). Genomic information is already being integrated into genetic evaluations in the New Zealand sheep industry using ssGBLUP for across-flock comparisons (Lee et al. 2021).

However as Lee et al. (2021) notes, genomic improvement relies on a strong foundation of phenotypic and pedigree data. Therefore, robust progeny testing schemes will likely remain an essential component of genetic improvement efforts, working alongside genomic technologies to drive future progress in the sheep industry. The effectiveness of progeny test results depends on the data collected but also on the degree of genetic connectedness between flocks. Ensuring sufficient connectedness is essential for accurate across-flock comparisons, which underpin both traditional and genomic breeding value estimation (Špehar et al. 2025).

2.11 Connectedness

2.11.1 Definition and Levels of Genetic Connectedness

Genetic connectedness refers to the degree to which animals in different management units are genetically linked, enabling fair and accurate comparisons of their genetic merit (Kuehn et al. 2007). In practice, connectedness is often achieved using common sires, either by mating the same sires across flocks or by evaluating their progeny within a shared environment (McLean et al. 2006). Structured progeny test programmes, such as the CPT and the more recent Low Input Sheep Progeny Test, use both approaches. Sires are sourced from different flocks and are either evaluated together under standardised conditions at hub sites, or mated across multiple environments including Next Generation sites to build genetic links which facilitate across-flock genetic evaluations (McLean et al. 2006; Beef + Lamb New Zealand Genetics 2023).

Connectedness can occur at several levels such as across flocks, years, ewe age groups, or management mobs, with each contributing to the robustness of genetic comparisons (Beef + Lamb New Zealand Genetics 2016). Across-year connectedness refers to the genetic linkage between animals born in different years to account for year-to-year environmental variation and enabling reliable comparisons over time. Similarly, across ewe age group or across management mob connectedness reflects the ability to fairly compare animals managed in different age brackets or farm subgroups which may experience different environmental conditions or selection pressures (Beef + Lamb New Zealand Genetics 2016). Genetic connectedness is typically assessed using statistical measures that reflect how precisely animals from different groups can be compared (Kennedy and Trus 1993). Within nProve, connectedness is currently evaluated according to the number of shared progeny produced by link sires across flocks within a rolling 3-year period (Holmes et al. 2015).

2.11.2 Importance and Impact of Connectedness

Maintaining strong genetic connectedness is essential to support accurate and unbiased across-flock genetic evaluations (Brown et al. 2007). Where connectedness is low, comparisons between animals from different management units can be influenced by environmental or management effects, leading to biased EBVs. As genetic variance between flocks has been suggested to be as large as within-flock variation, it is critical to either model these biases directly or apply tools that can quantify the risk of comparing animals across units (Kuehn et al. 2007). Improving connectedness offers a means to reduce such bias. For example, Kuehn et al. (2007) found that when at least one third of the breeding females in a flock are mated to common reference sires, bias in across-flock EBV comparisons was reduced to nearly optimal levels. Similarly, Lewis et al. (1999) demonstrated that increasing the number of reference sires from one to three improved connectedness by 59%, enhancing EBV precision and reducing deviations between true and estimated breeding values.

In addition, increased connectedness can facilitate gene flow between flocks. This allows superior genetics from higher performing populations to disseminate into lower performing ones, gradually reducing differences in genetic merit across the wider population (Hanocq et al. 1996). Another important implication is that many evaluation systems require sufficient genetic linkage between flocks before animals can be jointly evaluated, ensuring that across-flock rankings are robust (Mathur 2005). This principle underpins New Zealand's current nProve evaluation system, where breeding values and selection indices are only displayed to the public via nProve for flocks that are sufficiently connected (Newman 2017). Consequently, ensuring adequate connectedness is not only important for reducing bias, but is also a prerequisite for participation in national across-flock evaluations that guide selection and genetic gain.

2.11.3 Measures of Connectedness

Several statistical measures have been developed to quantify the degree of genetic connectedness between management units. One of the most widely accepted metrics is the prediction error variance of differences (PEVD) between EBVs, which reflects the uncertainty in comparing genetic merit across flocks (Kennedy and Trus 1993). While the prediction error variance (PEV) of an individual animal provides insight into the reliability of its EBV, it does not adequately capture the uncertainty associated with comparing animals across different

units, making PEVD a more informative measure (Kuehn et al. 2007). A lower PEVD indicates stronger connectedness and more reliable across-flock comparisons.

Another useful metric is the prediction error correlation (r_{ij}) which assesses the correlation between prediction errors of genetic values across flocks (Lewis et al. 1999). Greater r_{ij} values indicate stronger connectedness and more consistent EBV comparisons. The coefficient of determination (CD) is also widely used and reflects the proportion of true genetic difference that can be reliably predicted, or equivalently, the squared correlation between true and estimated breeding values (Laloë 1993). In modern mixed-model evaluations, CD extends naturally to across-flock comparisons by quantifying the information available for predicting genetic differences between units. Unlike PEVD, which focuses on the absolute uncertainty in EBV differences, CD incorporates the structure of the relationship matrix and accounts for variability in relatedness among animals, providing a bounded (0–1) measure of reliability (Yu et al. 2017).

In addition, tools such as gene flow analysis and genetic drift variances are also used to assess connectedness. Gene flow tracks the transfer of genetic material between flocks via shared ancestry, whereas genetic drift variance captures divergence in genetic means caused by random mating and small population sizes (Kennedy and Trus 1993).

These statistical approaches have been applied in both simulation and real-world studies. For example, Kuehn et al. (2008) evaluated sire referencing schemes across 15 flocks and demonstrated that connectedness increased as selected sires were mated across units, with improvements reflected in higher r_{ij} values and declining PEVD. When flock r_{ij} reached 0.10, only 10% of the original bias in EBV comparisons remained, indicating this level may serve as a potential benchmark for minimising bias (Kuehn et al. 2008). Gene flow analysis, as described by Škorput et al. (2018), was used to quantify connectedness among conservation pig herds by tracing the movement of genetic material between groups using the matrix product X'ZTQ, which links animals to their herd of origin and tracks the transmission of genes across generations. Herds with higher levels of gene flow to other groups were identified as key contributors to population-wide genetic linkage (Škorput et al. 2018). Additionally, Yu et al. (2017) demonstrated that genomic relationship matrices strengthened connectedness in CD compared with pedigree-based evaluations, highlighting the value of genomic information for improving the reliability of across-flock EBV comparisons.

2.11.4 The GCA R Package

Although a range of statistical measures exist to quantify genetic connectedness, their application has historically been limited by the lack of user-friendly software. The Genetic Connectedness Analysis (GCA) R package was developed to address this gap by providing a comprehensive and accessible platform for quantifying genetic connectedness using either pedigree- or genome-based relationship matrices (Yu and Morota 2021). The GCA package implements a wide suite of statistics derived from PEV and variance of unit effect estimates (VE), which quantify the connectedness between units and the precision of comparisons between units, including the PEVD, CD, r_{ij}, variance of differences in unit effects (VED), coefficient of determination of VE (CDVE), and connectedness rating (CR). Metrics based on PEV can be summarised at different levels including within units, between units, or using contrast vectors, offering flexibility in how connectedness is evaluated (Yu and Morota 2021). The practical value of these tools has been demonstrated in recent across-flock studies. For example, Špehar et al. (2025) used the GCA package to quantify pedigree-based connectedness among dairy sheep flocks and reported a PEVD of 0.81. Their analyses illustrated how connectedness metrics could be incorporated into population-level evaluations by exploring relationships with parameters such as effective population size. The study also provided preliminary evidence that higher levels of genetic connectedness may contribute positively to the maintenance of genetic diversity, although further research is required to confirm this relationship.

2.11.5 Practical Considerations and Guidelines for Connectedness

Practical guidelines have been established to ensure sufficient connectedness for across-flock genetic evaluations. In New Zealand, SIL recommends that at least one link sire should be used in small flocks, and two or more in larger flocks each year to create genetic links across breeding groups. These sires should have 25 or more progeny measured per flock for the traits of interest (Beef + Lamb New Zealand Genetics 2016). Mathur (2005) reported that typically around 15% of progeny from well-connected sires within a flock is sufficient to achieve the minimum degree of connectedness. These recommendations provide a practical benchmark for designing mating schemes in progeny test and dispersed breeding programmes. In Australia, Sheep Genetics provides similar guidance, with breeders encouraged to ensure that at least one in every ten sires used has also been used in another Sheep Genetics flock. It is also recommended to obtain at least 30 recorded progeny per link sire for traits of interest (Sheep Genetics 2025).

Despite these guidelines, there has been very limited evidence-based research to determine the precise number of link sires required to achieve sufficient connectedness. Improving connectedness can also present practical and economic challenges, particularly in sheep breeding systems. Increasing the number of shared sires across flocks often requires physically relocating rams between farms for natural mating, creating logistical, financial and biosecurity challenges. However, continually increasing genetic connectedness is not always optimal, as excessive sire sharing could lead to flocks becoming genetically similar, reducing genetic diversity and increasing the risk of inbreeding (Yu et al. 2018). Therefore, it is important to identify an optimal level of connectedness that balances the need for accurate EBV comparisons, preservation of genetic diversity across the national flock, and the practical costs associated with sire sharing strategies.

2.12 Sire Attribution

2.12.1 Causes, Prevalence and Implications of Incorrect Sire Attribution

While the inclusion of link sires is critical for establishing genetic connectedness across flocks, the accuracy of across-flock evaluations can be compromised by pedigree recording errors, particularly incorrect sire attribution. Misidentification of a lamb's biological sire in livestock breeding programmes can lead to biased EBVs and inflated estimates of connectedness, ultimately slowing genetic progress (Gizaw et al. 2022). This issue is especially problematic in natural mating systems where errors are more likely due to multi-sire mating and inadequate record keeping. In practice, achieving 100% pedigree accuracy is challenging in sheep breeding as it is not feasible to monitor every mating or lambing event. While some errors arise from mis-mothering at lambing, the majority originate during mating with common causes including rams jumping fences, gates being accidentally left open, or unintended inclusion of the wrong rams in mating groups (Beef + Lamb New Zealand Genetics 2016).

Research quantifying incorrect sire attribution rates in New Zealand sheep systems is limited. A study by Crawford et al. (1993) utilised polymorphic genetic markers to assess pedigree accuracy across four New Zealand flocks. Error rates measured on a per relationship basis reflect the proportion of incorrect sire-offspring or dam-offspring assignments. These rates ranged from 5% in some research flocks to as low as 0.4% in a well-managed progeny test flock. The authors proposed that the cause of that variation in error rates was influenced by the

intensity of stock management, with higher error rates associated with lower management input, larger paddock sizes and fewer staff (Crawford et al. 1993). Sire misidentification has been more extensively studied in New Zealand's dairy sector due to its greater economic scale and more structured recording systems. In progeny test herds managed by Livestock Improvement Corporation between 2005 and 2007, sire misidentification rates were estimated at approximately 5% (Winkelman 2013). These rates were taken from sire proving scheme herds, so are expected to be lower than what is observed in commercial settings due to more rigorous management and stricter recording protocols.

Sire misidentification can have significant consequences for genetic evaluations. Errors in pedigree information can reduce the accuracy of EBVs, lower heritability estimates, and compromise the reliability of genetic parameter estimates (Senneke et al. 2004; Dechow et al. 2008). Such effects are especially problematic in progeny test herds where accurate pedigree recording is crucial for evaluating the genetic merit of young sires (Dechow et al. 2008). Geldermann et al. (1986) demonstrated that reducing misidentification allows for more accurate ranking of breeding candidates and can enhance the efficiency of progeny testing schemes either by testing more sires with the same resources or by maintaining selection accuracy with fewer progeny per sire. A study by Israel and Weller (2000) quantified the impact of sire misidentification using simulation studies finding that a 10% error rate in paternity assignment reduced annual genetic gain by 4.3%. This demonstrates that even modest levels of pedigree error can bias EBVs and diminish genetic progress.

2.12.2 Current Approaches and Challenges in Managing Sire Attribution

Several strategies are used in New Zealand to manage sire attribution in sheep flocks. A performance recording guide by Beef + Lamb New Zealand Genetics (2016) outlines three options for recording parentage, each varying in accuracy and practicality. The highest level of accuracy is provided by DNA parentage testing, followed by mating records combined with tagging at birth, and lastly, mating records combined with pregnancy scanning and observation of mothering of lambs (Beef + Lamb New Zealand Genetics 2016). Whole flock DNA parentage testing involves collecting DNA samples from all potential sires at mating and from lambs at docking. This approach helps detect unexpected sires and improves pedigree accuracy (Beef + Lamb New Zealand Genetics 2016). To support the adoption of DNA-based pedigree recording, Clarke et al. (2014) developed a high-throughput SNP assay tailored for the New Zealand sheep industry. The 84 parentage SNP panel achieved an average sire assignment rate

of 99% without any dam information, showing the assay's potential and suitability for paternity testing. The other two methods involve keeping detailed records of matings as well as tagging lambs at birth where they are matched to their dam based on observation, or observing ewe-lamb behaviour between birth and tailing (Sheep Genetics 2025). The use of DNA-based parentage became available for New Zealand's sheep industry around the year 2000, and its adoption has since accelerated, driven mostly by large-scale breeders (Crawford et al. 2007). Although many sheep farmers still rely on conventional parentage recording methods due to cost constraints or infrastructure limitations, DNA technology has become a core component of structured breeding programmes such as the Beef + Lamb CPT and Low Input Progeny Test (Beef + Lamb New Zealand Genetics 2015, 2023).

2.13 Future Directions: Strengthening Genetic Connectedness

While the theoretical importance of genetic connectedness is well recognised, there remains a critical need for evidence-based research within New Zealand sheep production systems to define optimal levels of connectedness. Additionally, quantifying the impact of insufficient linkage on EBV accuracy and genetic gain is essential. Current guidelines provided by SIL, such as the recommendation to use two or more link sires in larger flocks, lack clear justification and specificity (Beef + Lamb New Zealand Genetics 2016). For example, there is little evidence indicating the level of connectedness that can be achieved under this recommendation and how this varies across different flock sizes and traits with differing recording intensity and heritability. The integration of genomic information could present an opportunity to produce stronger and more reliable estimates of connectedness across flocks, particularly where pedigree data is incomplete or poorly recorded (Yu et al. 2018). This is largely because genomics can identify hidden genetic connections between flocks that are not evident from pedigree records (Momen and Morota 2018). As genomic technologies become increasingly accessible and affordable, their adoption could be a key next step to facilitate fairer across-flock comparisons and enhance EBV accuracy.

2.14 Identifying Optimal Levels of Connectedness

To quantify optimal thresholds of genetic connectedness, it is useful to focus on traits that are particularly relevant to New Zealand sheep breeding. Two contrasting examples are growth

and FE tolerance. Growth traits are relatively inexpensive to measure and widely recorded across flocks, providing strong connectedness. In contrast, FE tolerance is more difficult and costly to measure (Tipa 2017), resulting in lower levels of connectedness across flocks.

2.14.1 Growth

Growth characteristics are central within New Zealand sheep breeding objectives. From an industry perspective, heavier lambs at weaning provide strong economic returns. Modelling by Moloney et al. (2023) demonstrated that increasing weaning weight by 10, 20, and 30% improved cash operating surplus from NZD $291/ha to $342, $392 and $444/ha, respectively. As carcass weight (CW) is also directly tied to market returns, selection for growth traits remains central to improving the profitability of sheep farming enterprises.

In the NZMW index, growth is represented by direct weaning weight (WWT), maternal weaning weight (WWTM), and CW, while in the NZTW index, growth is represented by WWT and CW only (Beef + Lamb New Zealand Genetics 2022). Weaning weight is particularly important as an indicator of lamb growth. Typically measured at 10–12 weeks of age, the weight of a lamb reflects both the lamb's own genetic potential for growth and the ewe's maternal contribution and is therefore used to estimate direct WWT and WWTM breeding values (Beef + Lamb New Zealand Genetics 2016). The best practice guidelines from SIL recommend that all live lambs in a mob are weighed at weaning, and birth date and mob codes are also recorded to account for non-genetic influences (Beef + Lamb New Zealand Genetics 2016). Estimates from New Zealand dual purpose flocks show moderate heritability for both direct and maternal components (Pickering et al. 2012), confirming that genetic gain is possible through selection on lamb and ewe performance. By separating direct and maternal effects, SIL breeding objectives help balance improvements in lamb growth with ewe efficiency and flock feed demands, supporting sustainable genetic progress (Beef + Lamb New Zealand Genetics 2016).

2.14.2 Facial Eczema

Although selection for FE tolerance has been shown to be both feasible and effective (Morris et al. 1989), achieving substantial resistance at the national flock level would require widespread uptake of such breeding programmes. However, FE tolerance has limited phenotypic recording despite being a serious health and production issue in New Zealand.

One practical tool for breeders is the RamGuard™ FE Tolerance Testing Service, which supports breeders in selecting rams with increased resistance to FE (McRae et al. 2021). Rams are challenged with a controlled dose of the toxin sporidesmin and ranked based on their gamma-glutamyl transferase response measured 21 days post-challenge (GGT21). Lower GGT21 values indicate greater FE resistance, as they reflect less liver damage caused by the toxin (Beef + Lamb New Zealand Genetics 2025). The service costs approximately NZD $250–$300 per ram, representing a significant expense for breeders undertaking large-scale testing (Beef + Lamb New Zealand 2019). Results are reported in a breeder's FE certificate which includes each ram's individual GGT21 response and the average tolerance level of the flock, allowing commercial farmers to make informed purchasing decisions (Amyes and Hawkes 2014). Breeders that have used RamGuard™ for eight to nine years have, on average, doubled or tripled the base level of tolerance levels in their flocks (Beef + Lamb New Zealand 2019). Although complete immunity is not achieved, these improvements enable animals to graze safely on pastures with higher spore counts, reducing the risk of FE. Commercial farmers can rapidly enhance the FE tolerance of their flocks by sourcing rams from these tested lines. However, the overall success of this strategy will depend on greater industry adoption and recognition of its long-term value (Amyes and Hawkes 2014).

2.15 Conclusions

Genetic improvement has long been fundamental to the progress in New Zealand's sheep industry, and its importance continues to grow as the sector adapts to evolving pressures. The NZGE system, managed by SIL, underpins much of the genetic selection in the industry by employing ssGTABLUP to predict breeding values and identify rams capable of driving significant genetic gain (Lee et al. 2021). Successful genetic progress relies on the availability of genetic variation for specific traits as this provides the foundation for selection. Genomic selection offers opportunities to accelerate genetic gain by improving the accuracy of breeding value predictions and allows animals to be selected at an earlier age (Bora 2023).

Progeny testing continues to play a critical role in supporting accurate evaluations by providing data that enables breeders to disentangle genetic merit from environment and management effects. Structured programmes, like New Zealand's CPT, have been instrumental in establishing genetic connectedness across the national flock, accelerating improvements in key traits such as growth, reproduction and carcass quality. As the industry faces evolving

economic, environmental, and social challenges, the focus of genetic improvement is beginning to shift towards welfare and disease resistance traits (Johnson et al. 2021). This shift is exemplified by initiatives like the Low Input Sheep Progeny Test which reflects the industry's evolving priorities. Despite the cost and time lag challenges, continued investment in progeny test programmes remains vital for driving sustained genetic gain.

Progeny test programmes foster the genetic connectedness required for accurate across-flock comparisons. While structured progeny test programmes and the strategic use of link sires have improved connectedness, challenges remain in optimising connectedness levels and managing the practical limitations of sire sharing. Furthermore, the accuracy of genetic evaluations is compromised by incorrect sire attribution, particularly in natural mating systems. Although advancements in DNA parentage testing have significantly improved pedigree accuracy, uptake across the industry is variable due to economic and logistical constraints. Future research should focus on quantifying optimal thresholds of genetic connectedness and refining guidelines for the use of link sires, with particular attention to improving the robustness of across-flock genetic evaluations.

2.16 Research Question

Therefore, the research question is:

- Is there an optimal level of genetic connectedness required between flocks in a dispersed progeny evaluation to ensure unbiased and accurate breeding value estimation in the New Zealand sheep industry?

The objectives of this study were to:

- Determine how variations in the connectedness evaluation window affect the level of genetic connectedness and the number of flocks eligible for across-flock evaluations.
- Evaluate the impact of removing highly connected flocks or masking prolific sires on the linkage structure and connectedness among the remaining flocks.
- Assess how applying less conservative connectedness criteria affects the accuracy of across-flock EBVs.

Chapter 3: Methods and Methodology

3.1 Data

Two primary datasets were used for the connectedness analyses: a facial eczema (FE) dataset and a growth dataset. Both were provided as data downloads from the nProve database in the form of .csv files, containing animal identification, pedigree, and performance trait records. Key columns included birth flock, birth year, animal tag, sire and dam identifiers and relevant phenotypic measurements.

The FE dataset consisted of records for 2,019,387 animals born between 1995 and 2025, representing 305 distinct flocks. The growth dataset was provided as a subset of an nProve download and contained records for 501,222 animals, also born between 1995 and 2025, representing 119 distinct flocks.

Both datasets were processed using SAS software 9.4 (SAS Institute Inc 2023) and the nProve genetic engine preprocessing programme, 'Geprep', which prepares datasets for the ssGBLUP analysis including producing a pedigree file and trait datasets that have been filtered, scaled and transformed (as appropriate) for analysis. Traits are grouped for analysis into collections of traits that are associated for specific components of animal performance (e.g. growth traits, meat traits, wool traits and FE tolerance traits). These trait files are used for the downstream connectedness analyses. Three key datasets are produced which are required for the nProve genetic engine and connectedness analyses:

1. Pedigree dataset: animal IDs with corresponding sire and dam IDs.
2. Unique identification file (UID): animal ID, flock of origin, birth year, and ear tag.
3. Phenotypic dataset: trait-specific performance records.

For the analysis of the FE trait, the phenotypic dataset contained records for gamma-glutamyl transferase response measured 21 days post-challenge (GGT21) for 48,058 animals represented by 5,855 sires across 92 different flocks. The records for GGT21 were log-transformed prior to analysis. The FE data was extracted in October 2025; therefore, phenotypic records for animals born in 2024 may be incomplete because additional GGT21 recording may have occurred throughout 2025, and no records were available for animals born in 2025. For the growth analysis, the phenotypic dataset contained weaning weight (WWT) records for 375,709 animals, represented by 4,393 sires across 101 different flocks. The growth data was extracted

in January 2026; therefore, records for animals born in 2025 may be incomplete as some animals had not yet reached weaning age.

3.2 nProve Connectedness Analysis

Connectedness analysis was performed in SAS 9.4 (SAS Institute Inc, 2023) using the current nProve connectedness procedure. Following preprocessing in Geprep, the first step involved running the programme SILLinkage, which generates a parameter file defining the evaluation window, specifying the number of years of historical data (up to and including the analysis end year) used to compute across-flock linkages. Connectedness was then computed using the nProve connectedness programme, which evaluates the degree of across-flock linkage within a rolling 3-year window based on the number of progeny from link sires shared between pairs of flocks, producing a distance matrix for all pairs of flocks in the analysis.

The workflow begins by importing and cleaning the pedigree, UID, and trait files. These datasets are merged to produce a dataset where each animal has a unique internal identifier linked to its flock, contemporary group, sex, and trait validity. The evaluation window is then applied to define the number of years of historical data included in the analysis. Contemporary groups are constructed at both within-flock and across-flock levels. Progeny counts per sire and per flock group are summarised to identify link sires, defined as sires with progeny recorded in more than one flock group. These link sires form the basis for evaluating across-flock genetic connectedness. Parent-group progeny tables are then generated and used to compute a distance-based connectedness measure, where values closer to 0 indicate stronger genetic linkage between flocks. The resulting lower-triangular distance matrix represents the degree of connectedness among all flock groups included in the analysis. Hierarchical clustering using single-linkage (Proc Cluster) is applied to this distance matrix to produce a connectedness dendrogram, with branch lengths reflecting relative connectedness. For this study, connectedness values were scaled to range from 3 to -2, with values above 0 indicating flocks that met the connectedness criteria and values below 0 indicating flocks that did not (N Jopson pers. comm). This dendrogram represents the base scenario, corresponding to the standard nProve methodology currently used in industry evaluations. Subsequent analyses in this study, involving modifications to the dataset or amount of historical data used, were compared against this base scenario to assess their effect on connectedness estimates.

In addition to the distance matrix and dendrogram, the nProve programme also generates a ‘traffic-light’ connectedness report. This report summarises rolling 3-year connectedness

trends for each flock, classifying the strength of linkage (strong, moderate, weak, or missing) based on predefined thresholds, and will flag potential future loss of connectedness.

3.2.1 Varying Evaluation Window Length

One component of the nProve connectedness analysis is the length of the evaluation window, which determines the number of years of historical progeny records included in the analysis. The default window varies by trait, reflecting differences in data recording and trait availability over time. For both FE and growth, nProve currently uses a 3-year evaluation window, which includes progeny born between 2022 and 2024 (Table 3.1).

Table 3.1: Evaluation window scenarios used throughout this study and the corresponding progeny birth years included in each scenario.

Evaluation window	Progeny birth years included
1-year window	2024 only
2-year window	2023–2024
3-year window (reference)	2022–2024
4-year window	2021–2024
5-year window	2020–2024
6-year window	2019–2024
7-year window	2018–2024
8-year window	2017–2024
9-year window	2016–2024
10-year window	2015–2024
11-year window	2014–2024

To evaluate the sensitivity of connectedness to data depth, the evaluation window was systematically adjusted. The default 3-year window was modified to include one, two, or three additional years of data (4, 5, and 6-year windows respectively), as well as one or two fewer years of data (2 and 1-year windows respectively). These adjustments were made directly within the SILLinkage programme by editing the macro variable that defines the amount of historical data included in the analysis in the two locations where it is defined. For FE analysis, the macro variable “lagyr_fe” was modified, while for the growth the macro variable “lagyr_g” was modified. Following each adjustment, the connectedness analysis was rerun in SAS, and the resulting distance matrices and dendrograms were compared to assess how including more or less historical data affected the apparent level of connectedness.

3.2.2 Removal of Flocks

To investigate the influence of highly connected flocks on the overall genetic linkage structure, analyses were performed in which selected flocks were sequentially removed from each dataset. Flocks were selected for removal based on two criteria: they exhibited high connectedness values in the base scenario, indicating a central role in linking multiple breeding groups, and were known to have historically contributed substantially to connectedness networks through their participation in the Central Progeny Test programme. Threshold values of approximately >1.5 for FE and >2.0 for growth were used to identify relatively highly connected flocks within each base dendrogram.

Based on these criteria, flocks F1, F6, and F18 were selected for the FE analysis, with connectedness values of approximately 2.2, 2.4, and 1.6, respectively. For the growth analysis, flocks F67, F65, F72, and F62 were selected, with connectedness values of approximately 2.7, 2.4, 2.2, and 2.2, respectively. Flock identifiers (e.g. F1 and F6) were anonymised codes, with new codes corresponding to the original flock numbers to ensure confidentiality.

Each of these flocks were removed one by one from the phenotypic data file to assess the effect of their exclusion on the pattern and degree of connectedness. This was implemented in SAS by inserting a conditional deletion statement within the relevant import step of the nProve connectedness programme. Specifically, flock removals were applied within the ImportFEczema step for the FE dataset and within the ImportGrowth step for the growth dataset. After each removal, the distance matrix was recalculated, and changes in the magnitude and structure of flock relationships were evaluated using the resulting dendrograms.

3.2.3 Masking of Sires

Another aim of this study was to assess the influence of highly prolific sires on flock connectedness. Prolific sires were selected to test whether the most widely used sires within each dataset were an important source of genetic linkage across flocks. Prolific sires were identified based on the number of progeny with phenotypic records (GGT21 or WWT) within each dataset, following merging of pedigree and phenotypic information in SAS. For the FE analysis, the three most prolific sires (S1, S2 and S3) were selected, having sired 184, 178, and 154 progeny, respectively. For the growth analysis three different prolific sires (S4, S5 and S6) were selected, having sired 1,320, 1,038, and 993 progeny, respectively. Sire identifiers (e.g. S1 and S2) were anonymised codes, with new codes used to replace the original sire identification numbers to maintain confidentiality.

Sire masking was implemented during the Geprep preprocessing stage in SAS. At the beginning of the Geprep script, a switch (%let exclsires = TRUE;) enables the masking of specified sires from the dataset. When activated, the programme reads a dataset (excl_sires) that lists the flock of origin, birth year, tag, and a deletion flag (delflag =1) for each sire to be masked. When the nProve download is read in, the flag instructs the programme to remove sire records from all progeny associated with the listed sire from the analysis. The progeny records themselves (their GGT21 or WWT records) are retained in the dataset. However, they appear without sire in any other flock in which they were used. This method effectively breaks all genetic links created through those sires, thereby eliminating their contribution to across-flock connectedness. Following the masking of each sire, the connectedness analysis was re-run to generate an updated distance matrix, and the impact on flock connectedness was evaluated using the resulting dendrograms.

3.3 Accuracy of Genetic Evaluations using GCA Connectedness Metrics

Connectedness analysis was conducted using the Genetic Connectedness Analysis (GCA) R package v0.1.0 (Yu and Morota 2019). For each analysis, pedigree data and unique identification files were imported and processed in RStudio v2025.09.1+401 (Posit team 2025) using the data.table v1.17.8 (Barrett et al. 2025) and LaF v0.8.6 (van der Laan 2024) packages. A pruned pedigree was generated using the AGHmatrix package v2.1.4 (Amadeu et al. 2023), from which a numerator relationship matrix (A matrix) was constructed. Connectedness among flocks was assessed based on pairwise prediction error variance of differences (PEVD) using the gca() function, assuming variance components of $\sigma^2_a = 0.3321$ and $\sigma^2_e = 0.4779$ derived from the SIL Mix99 analysis.

To assess the impact of the number of years of data on the coefficient of determination (CD) and PEVD for the accuracy of genetic evaluations across flocks, the GCA connectedness analysis was re-run in RStudio using progressively larger subsets of the FE phenotypic dataset. The range of birth years included was incrementally expanded from two years of phenotypic data (2023–2024 birth years) up to a maximum of eleven years (2014–2024 birth years). For each scenario, the same birth-year threshold was applied to both the phenotypic records and the pedigree, such that the phenotypic dataset was filtered to include only animals with FE records within the specified window, and the pedigree was re-pruned to retain these animals and their ancestors prior to construction of the numerator relationship matrix.

Across-flock evaluation accuracy was quantified using PEVD and CD. Pairwise PEVD values between flocks were first calculated using the GCA framework and then summarised as the overall average PEVD ($PEVD_{GrpAve}$) across all flock comparisons. The corresponding overall average CD (CD_{GrpAve}) was also calculated. For each data window scenario, the overall $PEVD_{GrpAve}$, CD_{GrpAve}, the number of phenotypic records, and the number of animals retained in the pruned pedigree were recorded.

The same analysis was not undertaken for the growth dataset. The substantially larger number of WWT records per year resulted in considerably greater computational requirements for repeated GCA analyses across multiple data window scenarios. Due to the available computational resources, the assessment of genetic evaluation accuracy using PEVD and CD was restricted to the FE dataset.

3.4 Estimation of Genetic Parameters and Breeding Values

Genetic parameters for FE and growth were estimated using GGT21 and WWT records, respectively. Single-trait linear mixed models were fitted by restricted maximum likelihood in ASReml version 4.2 (Gilmour et al. 2021). Pedigree and phenotypic datasets for both traits had previously been generated using SAS 9.4 and the nProve genetic engine preprocessing programme (Geprep), as described in Section 3.1. The pedigree dataset contained animal, sire, and dam IDs, and the phenotypic dataset contained the relevant performance records and associated explanatory variables.

For the FE analysis, the phenotypic dataset included GGT21 performance records, animal ID, flock, contemporary group, and sex. In this study, flock referred to the farm code, whereas a management group referred to sheep within the same farm and year that were managed together as a single mob. Contemporary group was defined as birth year and management group. For the growth analysis, the phenotypic dataset contained WWT performance records along with animal ID, dam ID, flock, contemporary group, sex, birth rearing rank, age of dam, and birth day deviation (defined as the deviation from the average birth date for the flock and year). Contemporary group was defined as the combination of birth year, weaning weight mob, and birth date deviation mob (early, middle or late). For both analyses, flock and contemporary group were combined into a single fixed effect (flock-contemporary group) to account for environmental effects specific to each flock-management grouping.

The pedigree files were processed in ASReml using the !make option to construct the numerator relationship matrix (A matrix), allowing estimation of additive genetic variance under an animal model. The following model was fitted:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{e}$$

Where **y** is the vector of phenotypic observations (GGT21 or WWT), **b** is the vector of fixed effects, **a** is the vector of random additive genetic effects for each animal, and **e** is the vector of residual errors. **X** and **Z** are incidence matrices relating observations to the corresponding fixed and random effects, respectively (Henderson 1975).

For the FE analysis, fixed effects included flock-contemporary group and sex, and the additive genetic effect of Animal ID was fitted as a random effect. For the growth analysis, fixed effects included flock-contemporary group, sex, birth rearing rank, age of dam, and birth day deviation. Random effects included the additive genetic effect of the animal and a random dam effect to account for maternal influences.

Heritability was calculated as:

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2}$$

Where σ_a^2 is the additive genetic variance and σ_e^2 is the residual variance. Estimated breeding values (EBVs) for all animals, along with solutions for fixed effects, were extracted from the ASReml solutions file. This analysis represents the whole population scenario for both the FE and growth datasets, where the full pedigree and phenotypic datasets were used without modification. This scenario provides estimates of genetic parameters using all available population records.

A series of sensitivity analyses was conducted to assess how estimates of genetic parameters responded to changes in the data, including variations in evaluation window length, the removal of highly connected flocks, and the masking of highly prolific sires. Unless stated otherwise, the same single-trait animal models and variance component estimation procedures described above were used.

3.4.1 Varying Evaluation Window Length

To investigate the influence of the evaluation window on genetic parameter estimates, additional analyses were conducted for both the FE and growth datasets in which the

phenotypic datasets were restricted to different birth-year windows. Six scenarios were defined based on the number of consecutive birth years included, while the original pedigree file generated from Geprep remained complete for all analyses.

The reference scenario comprised a 3-year window (2022–2024 birth years), consistent with the data window currently used by nProve. Additional scenarios consisted of a 1, 5, 7, 9 and 11-year window. For each scenario, the phenotypic datasets were restricted to animals meeting the specified birth year criteria specified in Table 3.1, while maintaining the full pedigree structure for construction of the relationship matrix. The single-trait animal models were re-run for each dataset using the same fixed and random effect structures described above for GGT21 and WWT. Variance components, heritability estimates, and EBVs were extracted from the ASReml output files to assess the stability of genetic parameter estimates across different levels of temporal data inclusion.

3.4.2 Removal of Flocks

To evaluate whether genetic parameter estimates were sensitive to the contribution of individual flocks, additional analyses were conducted for both the FE and growth datasets in which selected flocks were removed from the phenotypic dataset. The flocks removed in these analyses were those identified in Section 3.2.2 (FE: F1, F6, and F18; growth: F67, F65, F72, and F62).

Animals belonging to each flock were sequentially excluded from the phenotypic datasets using data filtering in R v4.5.1 (R Core Team 2025). Following flock removal, the resulting datasets were restricted to a 3-year birth window (2022–2024), consistent with the data window currently used by nProve. Each reduced dataset was then exported for analysis in ASReml. The single-trait animal models were re-run using the same fixed and random effect structures described above for GGT21 and WWT. Variance components, heritability estimates, and EBVs were extracted from the ASReml output files to assess the stability of genetic parameter estimates following the removal of major contributing flocks relative to the 3-year reference scenario.

3.4.3 Masking of Sires

To evaluate the influence of masking highly prolific sires on genetic parameter estimates, additional analyses were conducted for both the FE and growth datasets in which selected sires were sequentially masked in the pedigree. The sires masked in these analyses were those

identified in Section 3.2.3 as the most prolific within each dataset (FE: S1, S2 and S3; growth: S4, S5 and S6).

For each sire-masking scenario, the target sire ID in the pedigree was replaced with the ASReml missing parent indicator ('.') in R, such that the sire was treated as an unknown parent. This effectively removed the genetic contribution of the sire from the pedigree. A new pedigree file was then exported and processed in ASReml to reconstruct the numerator relationship matrix. For these analyses, the phenotypic datasets were restricted to a 3-year birth window (2022–2024), consistent with the reference data window currently used by nProve. The single-trait animal models were subsequently re-run using the same fixed and random effect structures described above for GGT21 and WWT. Variance components, heritability estimates, and updated EBVs were extracted from the ASReml output files to evaluate changes in genetic parameter estimation following sire masking relative to the 3-year reference scenario.

3.4.4 Correlation and Re-ranking of Breeding Values

To evaluate the consistency of EBVs across scenarios, correlation and re-ranking analyses were performed for both the FE and growth datasets. EBVs and their associated standard errors were extracted from the ASReml solution files for each scenario. For each trait, these outputs were imported into R and filtered to retain EBVs corresponding to individual animals (Animal ID solutions), excluding fixed effect solutions. Datasets from all scenarios were then merged by animal ID to create a single dataset containing EBVs for each animal across all analyses (14 scenarios for growth and 13 scenarios for FE).

The merged datasets were exported as .csv files and analysed in SAS. Pearson correlation coefficients were calculated to assess the strength of the linear relationship between EBVs across scenarios, while the Spearman rank correlation coefficients were used to evaluate the degree of re-ranking. Correlations were computed among all scenario combinations using Proc Corr, with separate analyses conducted for Pearson and Spearman correlations. Correlation coefficients were summarised in matrix form, with Pearson correlations presented in the lower triangle and Spearman rank correlations in the upper triangle.

Chapter 4: Results

4.1 Descriptive Statistics of Phenotypic Datasets

The phenotypic datasets generated using Geprep are summarised in Table 4.1. The growth dataset included fewer flocks per birth year as the growth analyses used a subset of the nProve download due to computational constraints, whereas the facial eczema (FE) analyses used a full dataset. However, the growth dataset contained more animals per flock, resulting in a much larger phenotypic dataset with a greater number of animals contributing records across all years compared to FE. Mean weaning weight (WWT) was scaled to 30kg within each contemporary group, and the standard deviation was relatively stable across years. Mean values for log-transformed gamma-glutamyl transferase response measured 21 days post-challenge (GGT21) for FE showed some fluctuation between years as GGT21 is log transformed for analysis but not scaled. No FE records were available for animals born in 2025.

Table 4.1: Summary of the phenotypic datasets generated using the Geprep preprocessing programme for facial eczema and growth. For each year, the number of animals and flocks contributing data, as well as the mean and standard deviation (SD) of gamma-glutamyl transferase response measured 21 days post-challenge (log GGT21) for facial eczema and weaning weight (WWT) for growth are shown. A dash (-) indicates years with no available data for the trait. Total represents all available records from 1995 to 2025 for each dataset.

	Facial eczema				Growth			
Birth year	n (animals)	n (flocks)	Mean GGT21	SD GGT21	n (animals)	n (flocks)	Mean WWT	SD WWT
2014	2580	52	4.54	1.08	17016	25	30.00	4.96
2015	3831	49	4.44	1.02	15925	24	29.99	4.96
2016	1953	54	4.54	0.96	17013	23	30.00	5.13
2017	3110	57	4.51	0.96	16609	29	29.99	4.95
2018	3469	51	4.46	1.04	16370	27	30.00	4.88
2019	1667	52	4.59	0.98	17763	30	30.00	4.80
2020	3095	45	4.44	1.03	18414	29	30.00	4.86
2021	3834	46	4.43	0.98	18417	26	29.99	4.99
2022	1851	44	4.60	1.02	17711	23	30.00	5.01
2023	1962	45	4.61	1.11	19078	24	30.00	4.84
2024	1118	25	4.79	1.08	19678	22	30.00	5.07
2025	-	-	-	-	11134	13	30.00	5.03
Total	48058	92	4.57	1.06	375709	101	30.00	4.97

4.2 nProve Connectedness

Genetic connectedness among flocks assessed using the nProve method is shown in Figure 4.1A for the FE dataset and in Figure 4.1B for the growth dataset. Flocks are listed along the vertical axis on the left of each dendrogram, while the horizontal axis represents the degree of genetic linkage based on the number of shared progeny from link sires. Flocks that are more strongly connected to other flocks in the evaluation are positioned towards the top of the dendrogram, whereas those with weaker or no genetic links appear lower down. Branch length serves as an indicator of linkage strength, such that shorter branches between any two groups reflect stronger genetic links arising from shared sires. The scale ranges from 3 to -2, with positive values (nodes positioned to the left of 0) representing flocks that met the connectedness criteria and negative values (nodes to the right of 0) representing flocks that did not. Based on this threshold, 24 of the 49 flocks included in the FE evaluation were classified as connected, with all flocks positioned above F25 forming the connected cluster, while the remaining 25 flocks were unconnected. In contrast 17 of the 23 flocks included in the growth evaluation were classified as connected, with all flocks positioned above F75 forming the connected cluster and the remaining 6 flocks classified as unconnected. Overall, the growth dataset showed a greater proportion of connected flocks than the FE dataset. The connected cluster also displayed generally shorter branch lengths, indicating stronger genetic linkage among flocks in the growth evaluation.

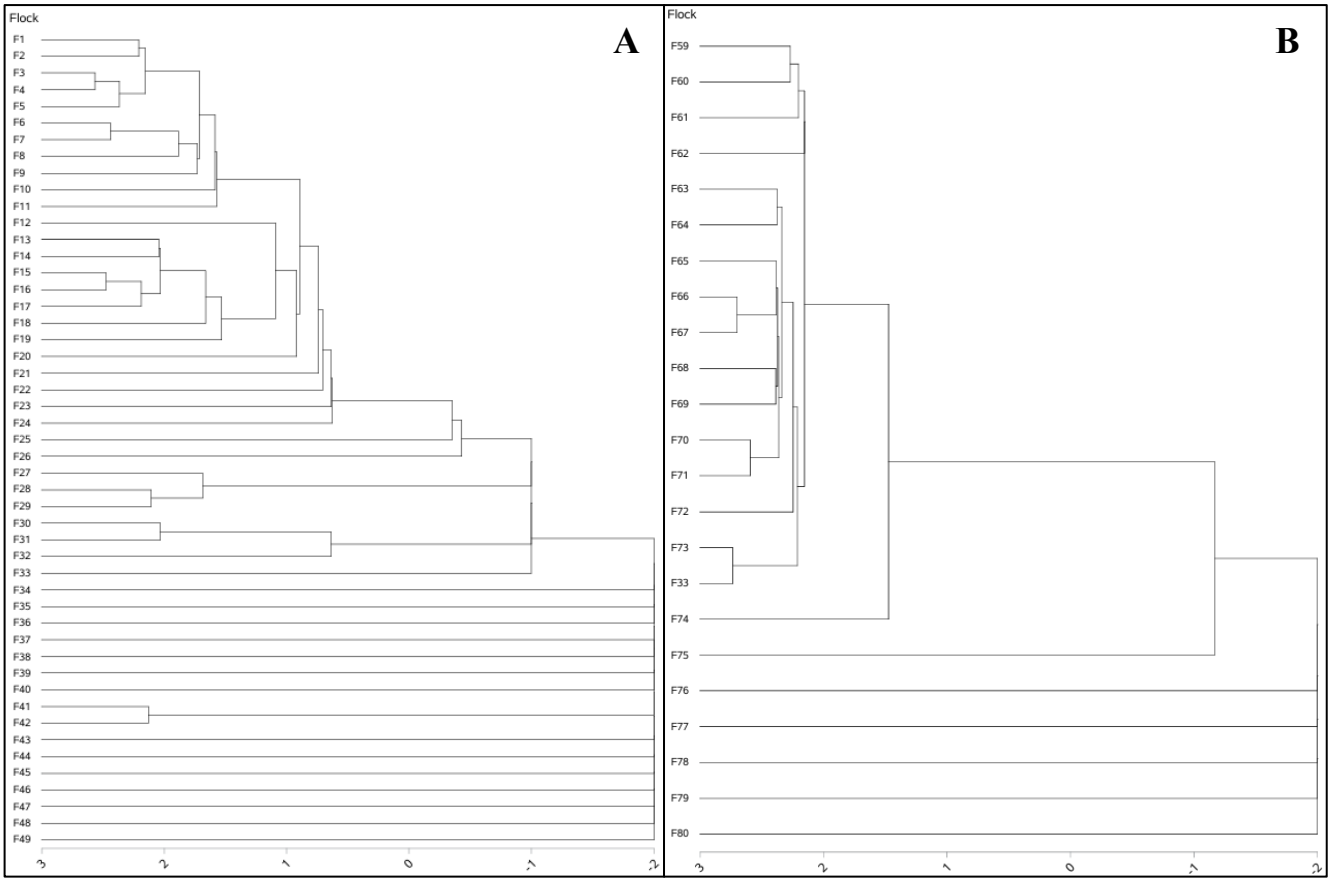


Figure 4.1: Base dendrograms illustrating genetic connectedness among flocks for (A) facial eczema and (B) growth, estimated using the nProve method.

4.2.1 Varying Evaluation Window Length

4.2.1.1 Facial Eczema

The influence of evaluation window length on flock connectedness for the FE dataset is illustrated in Figure 4.2. Reducing the window from 3 years (Figure 4.1A) to 2 years (removal of one year of data; Figure 4.2A) resulted in a decline in overall connectedness. Many branches extended further to the right and several flocks, such as F9, moved beyond the 0 threshold, becoming unconnected. Further reduction to a 1-year window (removal of two years of data) made this pattern more pronounced, with a noticeable decline in connectedness. Increasing the window to 4 years (one additional year of data; Figure 4.2B) had the opposite effect, with previously unconnected flocks such as F27 now meeting the connectedness criteria. Extending the window further to 5 and 6 years (Figure 4.2C and 4.2D, respectively) strengthened linkage, causing shorter branches and nodes shifting left along the scale, resulting in more flocks forming the main connected cluster. Flock F36 illustrates this effect, moving from an

unconnected status in the base and reduced datasets to progressively moving higher in the connected cluster as additional pedigree data was incorporated.

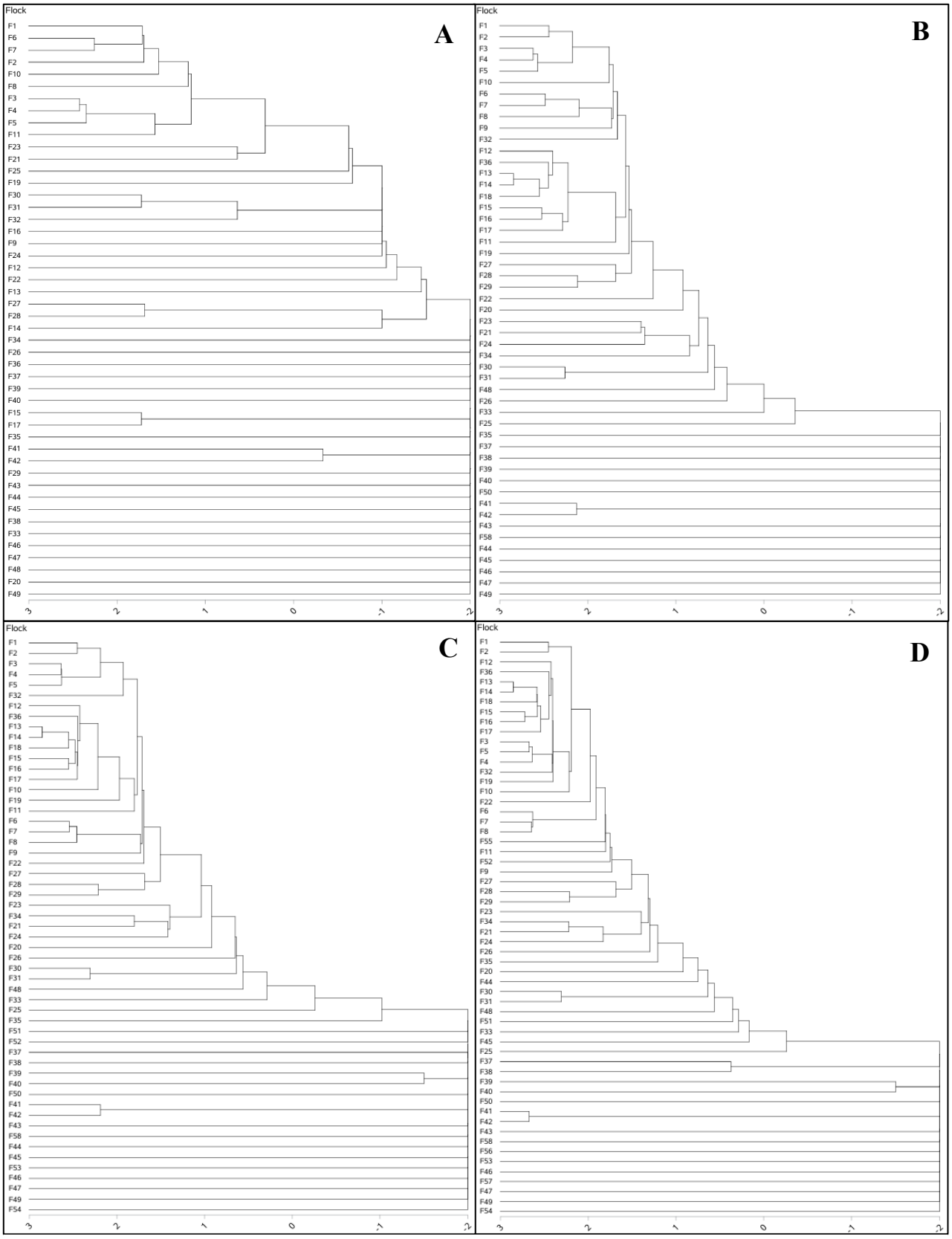


Figure 4.2: Dendrograms illustrating genetic connectedness among flocks for facial eczema using the nProve method under varying evaluation window scenarios. (A) 2-year window (2023–2024 progeny birth years); (B) 4-year window (2021–2024 progeny birth years); (C) 5-year window (2020–2024 progeny birth years); and (D) 6-year window (2019–2024 progeny birth years).

The effect of changing the evaluation window on the number of connected flocks for FE, as inferred from the dendrograms, is summarised in Table 4.2. The total number of flocks included in the FE analysis increased with longer windows due to the inclusion of additional pedigree records. Consistent with the visual patterns observed in the dendrograms, both the number and proportion of flocks meeting the connectedness criteria increased as more years of data were added. When one year of data was removed (2-year window), 25% of flocks were connected, falling further to 21.7% when two years of data were removed (1-year window). In contrast, the base scenario (3-year window) showed nearly half of all flocks (49%) as connected. Extending the window to 4 years increased this proportion to 66.7%, with improvements also seen with a 5-year window (63.6%) and a 6-year window (70.7%).

Table 4.2: Number and percentage of flocks identified as connected (flocks that met the connectedness criteria) or unconnected, as inferred from the dendrograms generated under each evaluation window scenario in the nProve connectedness analysis for facial eczema.

Evaluation Window	Total Flocks (n)	Connected Flocks (n)	Unconnected Flocks (n)	% Connected
1-year	46	10	36	21.7
2-years	48	12	36	25.0
3-years (base)	49	24	25	49.0
4-years	51	34	17	66.7
5-years	55	35	20	63.6
6-years	58	41	17	70.7

4.2.1.2 Growth

The influence of changing the evaluation window length on flock connectedness is illustrated in Figure 4.3 for the growth dataset. While extending the window increased the total number of flocks included in the analysis through additional pedigree records, the number of flocks meeting the connectedness criteria remained comparatively stable across the scenarios. Most flocks consistently formed a main connected cluster situated above the 0 threshold, with only a small subset remaining unconnected.

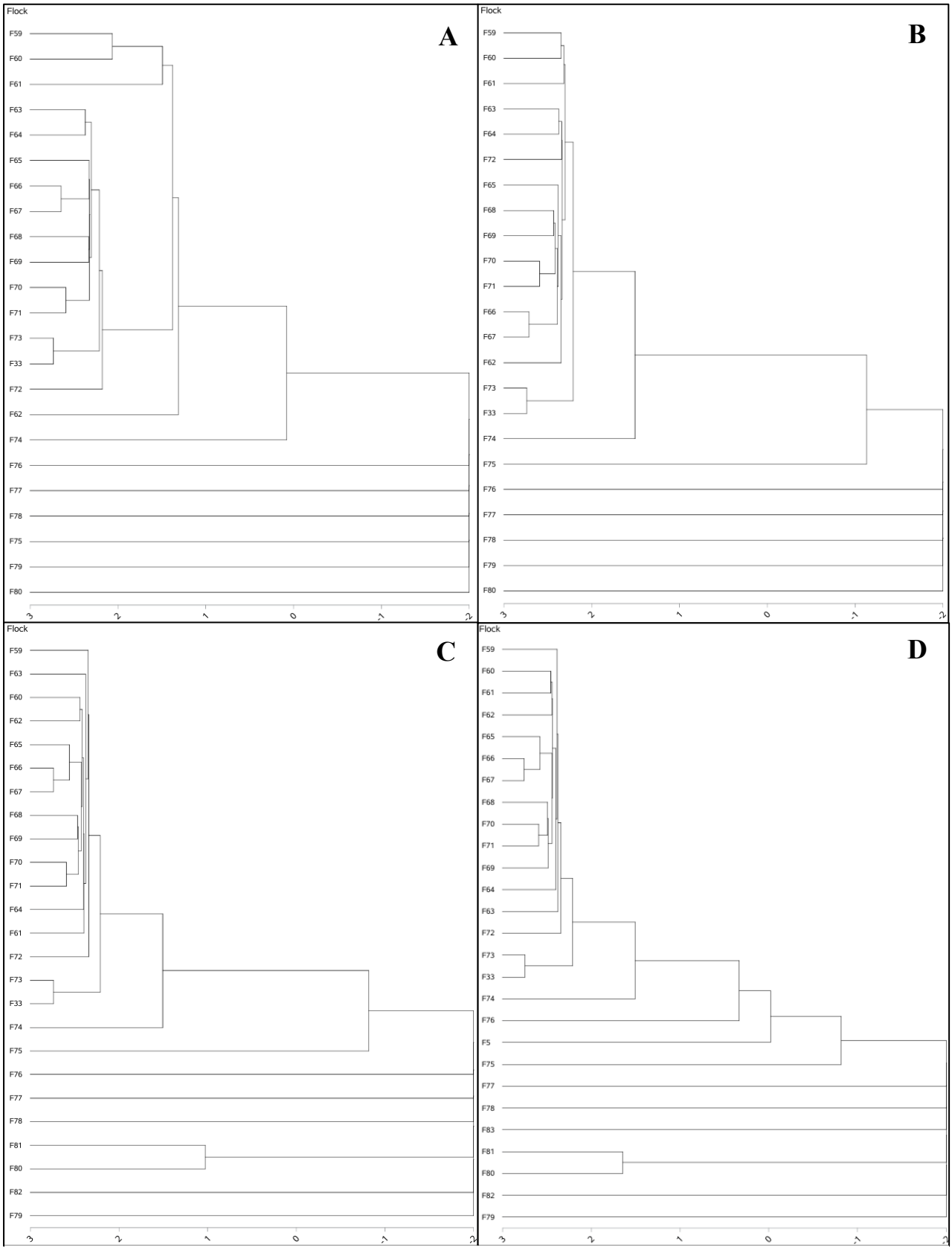


Figure 4.3: Dendrograms illustrating genetic connectedness among flocks for growth using the nProve method under varying evaluation window scenarios. (A) 2-year window (2023–2024 progeny birth years); (B) 4-year window (2021–2024 progeny birth years); (C) 5-year window (2020–2024 progeny birth years); and (D) 6-year window (2019–2024 progeny birth years).

Reducing the evaluation window from the base scenario of 3 years (Figure 4.1B) to 2 years (removal of one year of data; Figure 4.3A) or 1 year (removal of two years of data) caused only minor shifts in connectedness patterns. Although some branches extended slightly further right, indicating weaker genetic linkage, there was very little movement of flocks between connected and unconnected clusters. However, some individual flocks showed sensitivity to shorter data windows. For example, flock F59 remained strongly connected under the base and extended window scenarios but lost all connectedness when two years of data were removed, indicating that its linkage to other flocks was dependent on pedigree connections established in earlier years. Increasing the data window beyond the base scenario (Figure 4.3B–D) strengthened linkage within the connected cluster, with branch lengths becoming progressively shorter and nodes shifting left along the scale. However, these improvements were largely increases in linkage strength rather than changes in connectedness status.

As summarised in Table 4.3, the proportion of flocks meeting the connectedness criteria remained consistently high, ranging from 68.0% to 73.9%. The base scenario (3-year window) produced 73.9% connected flocks, identical to the one-year reduced and one-year extended scenarios (2-year and 4-year windows, respectively). When two years of data were removed (1-year window), the proportion decreased slightly to 68.2%. Extending the window to incorporate two and three additional years of data (5-year and 6-year windows, respectively) resulted in connectedness proportions of 68.0% and 70.4%, respectively. These results indicate that, unlike the FE dataset where longer evaluation windows increased the number of connected flocks, growth evaluations were characterised by a stable connectedness structure, in which extended windows primarily strengthened linkage within the existing connected cluster rather than expanding it.

Table 4.3: Number and percentage of flocks identified as connected (flocks that met the connectedness criteria) or unconnected, as inferred from the dendrograms generated under each evaluation window scenario in the nProve connectedness analysis for growth.

Evaluation Window	Total Flocks (n)	Connected Flocks (n)	Unconnected Flocks (n)	% Connected
1-year	22	15	7	68.2
2-years	23	17	6	73.9
3-years (base)	23	17	6	73.9
4-years	23	17	6	73.9
5-years	25	17	8	68.0
6-years	27	19	8	70.4

4.2.2 Flock Removal

4.2.2.1 Facial Eczema

The effect of sequentially removing flocks on the pattern of genetic connectedness is shown in Figure 4.4 for the FE dataset. Removal of highly connected flocks had a noticeable impact on dendrogram structure, with shifts in the relative order of flocks and changes in connectedness status. Following the removal of flocks F1 and F6 (Figure 4.4A and 4.4B, respectively), multiple flocks that previously met the connectedness criteria in the base scenario immediately moved beyond the 0 threshold, becoming unconnected. For example, in the base dendrogram, flock F6 was part of a connected cluster with three other flocks (F7, F8 and F9). Following its removal, flock F9 dropped markedly and became unconnected. This pattern is consistent with the relative high connectedness value of 0.127 between flocks F9 and F6 observed in the base distance matrix, highlighting the importance of F6 in maintaining this genetic linkage. In contrast, the removal of flock F18 (Figure 4.4C) had minimal impact on dendrogram structure or flock ordering, indicating that F18 was less central within the network of genetic linkages.

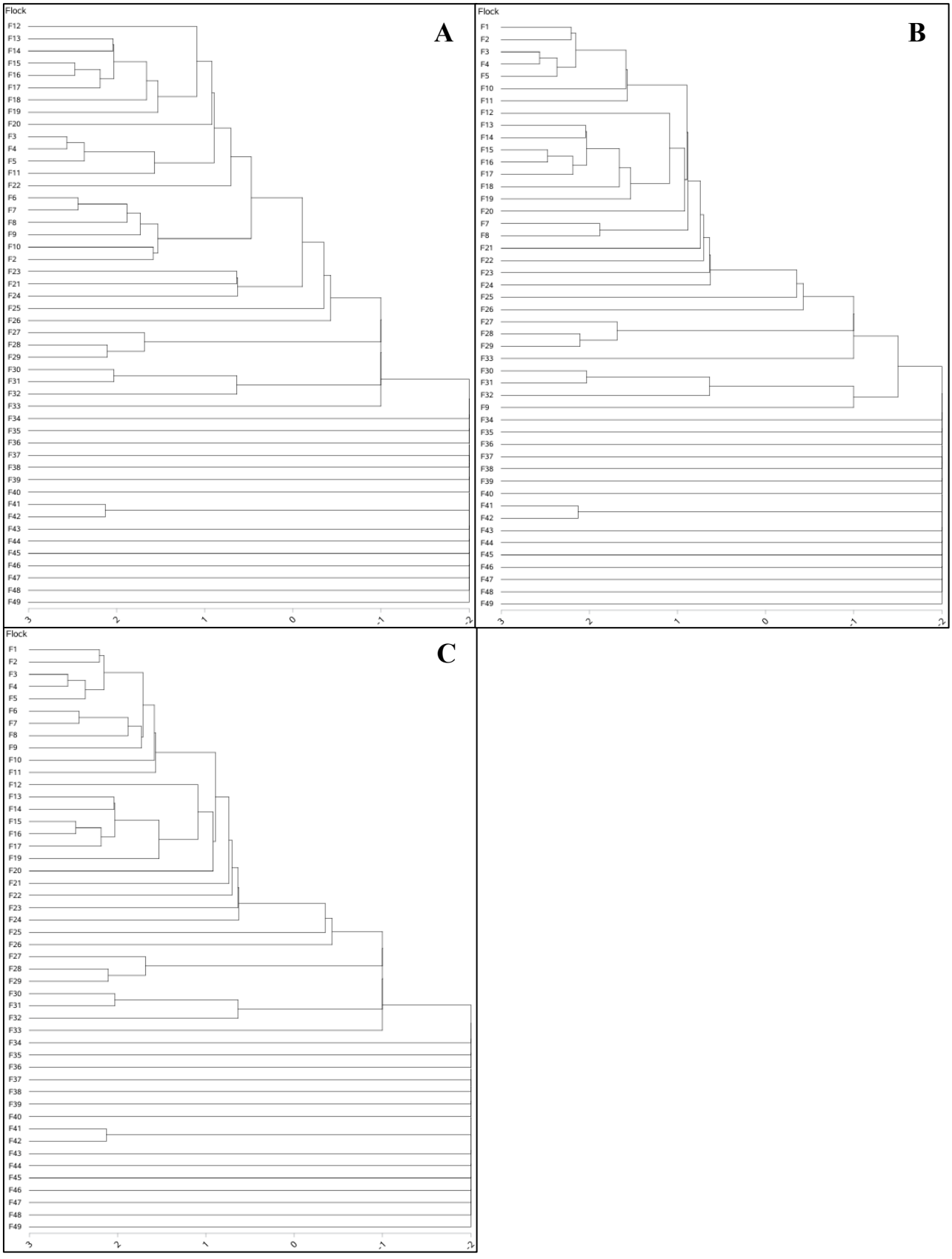


Figure 4.4: Dendrograms illustrating changes in flock connectedness for facial eczema following the removal of key flocks using the nProve method. (A) Flock F1 removed; (B) Flock F6 removed; and (C) Flock F18 removed.

The effect of flock removal on the number of connected flocks for FE is summarised in Table 4.4. Relative to the base scenario, removing flock F1 reduced the proportion of connected flocks from 49.0% to 41.7%, reflecting its central role in maintaining linkage within the industry. Similarly, removal of flock F6 decreased the percentage of connected flocks to 45.8%, while exclusion of flock F18 resulted in 47.9% of flocks being connected.

Table 4.4: Number and percentage of flocks identified as connected (flocks that met the connectedness criteria) or unconnected, as inferred from the dendrograms generated after sequential removal of key flocks in the nProve connectedness analysis for facial eczema.

Scenario	Total Flocks (n)	Connected Flocks (n)	Unconnected Flocks (n)	% Connected
Base (all flocks)	49	24	25	49.0
Flock F1 removed	48	20	28	41.7
Flock F6 removed	48	22	26	45.8
Flock F18 removed	48	23	25	47.9

4.2.2.2 Growth

The effect of sequentially removing flocks on the pattern of genetic connectedness is shown in Figure 4.5 for the growth dataset. The dataset was highly robust to the sequential removal of key flocks (F67, F65, F72, and F62). Across all removal scenarios, there was minimal change in dendrogram structure or branch lengths, and no remaining flocks shifted between connected and unconnected clusters. In each scenario the number of connected flocks decreased by one relative to the base analysis, corresponding only to the exclusion of the removed flock itself. The proportion of connected flocks remained unchanged at 72.7% compared with 73.9% in the base analysis (Table 4.5), indicating that genetic linkage within the growth dataset was widely distributed and not dependent on any single flock.

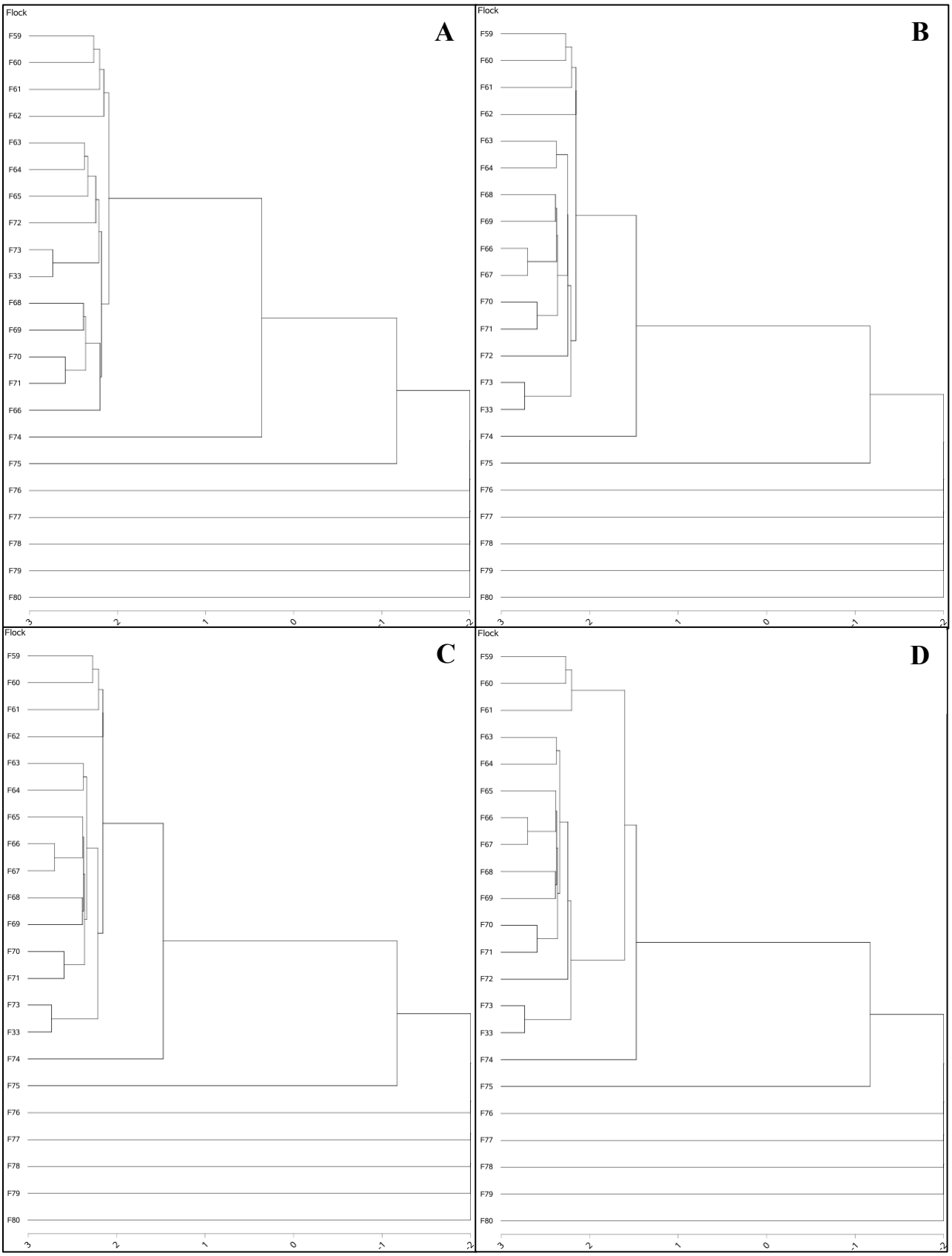


Figure 4.5: Dendrograms illustrating changes in flock connectedness for growth following the removal of key flocks using the nProve method. (A) Flock F67 removed; (B) Flock F65 removed; (C) Flock F72 removed; and (D) Flock F62 removed.

Table 4.5: Number and percentage of flocks identified as connected (flocks that met the connectedness criteria) or unconnected, as inferred from the dendrograms generated after sequential removal of key flocks in the nProve connectedness analysis for growth.

Scenario	Total Flocks (n)	Connected Flocks (n)	Unconnected Flocks (n)	% Connected
Base (all flocks)	23	17	6	73.9
Flock F67 removed	22	16	6	72.7
Flock F65 removed	22	16	6	72.7
Flock F72 removed	22	16	6	72.7
Flock F62 removed	22	16	6	72.7

4.2.3 Sire Masking

Masking the three most prolific sires had minimal impact on the structure of the dendrograms in either dataset. For the FE dataset, masking of sires S1, S2, and S3 produced dendrograms that were nearly identical to the base scenario, with only minor changes in the ordering of a small number of flocks. As summarised in Table 4.6, across all analyses the same 24 flocks met the connectedness criteria and the same 25 flocks remained unconnected as in the base scenario. Although five flocks changed their relative position within the dendrograms, these changes were minor and did not alter the broader clustering structure. Similarly, for the growth dataset, masking of sires S4, S5, and S6 resulted in dendrograms that showed minimal deviation from the base scenario. Only slight changes in branch lengths were observed for some connections following the masking of sire S4. As shown in Table 4.7, the same 17 flocks met the connectedness criteria and the same 6 flocks remained unconnected, consistent with the base analysis.

Overall, the stability of the dendrogram structures and the unchanged connectedness status following the masking of each sire indicates that no single sire contributed substantially to the overall network of genetic links in either dataset.

Table 4.6: Number and percentage of flocks identified as connected (flocks that met the connectedness criteria) or unconnected, as inferred from the dendrograms generated after sequential masking of prolific sires in the nProve connectedness analysis for facial eczema.

Scenario	Total Flocks (n)	Connected Flocks (n)	Unconnected Flocks (n)	% Connected
Base (all sires)	49	24	25	49.0
Sire S1 masked	49	24	25	49.0
Sire S2 masked	49	24	25	49.0
Sire S3 masked	49	24	25	49.0

Table 4.7: Number and percentage of flocks identified as connected (flocks that met the connectedness criteria) or unconnected, as inferred from the dendrograms generated after sequential masking of prolific sires in the nProve connectedness analysis for growth.

Scenario	Total Flocks (n)	Connected Flocks (n)	Unconnected Flocks (n)	% Connected
Base (all sires)	23	17	6	73.9
Sire S4 masked	23	17	6	73.9
Sire S5 masked	23	17	6	73.9
Sire S6 masked	23	17	6	73.9

4.3 Accuracy of Genetic Evaluations for Facial Eczema using GCA Connectedness Metrics

Across-flock evaluation accuracy for FE measured using prediction error variance of differences (PEVD) and coefficient of determination (CD) showed little change as the data window increased (Figure 4.6). Using a 3-year dataset consistent with the data window currently applied in nProve for assessing connectedness (earliest birth year = 2022), PEVD was 0.564 and CD was 0.286. Across the full range of data windows evaluated (2–11 years), the number of phenotypic records increased from 3,080 to 28,470, and the number of animals in the pedigree increased from 7,257 to 66,073. The CD increased steadily by 0.117 (from 0.259 to 0.376), indicating improved across-flock evaluation accuracy, whereas PEVD fluctuated within a narrow range (0.504–0.587). Both metrics showed evidence of plateauing after approximately 5 to 6 years of data (earliest birth year = 2019–2020).

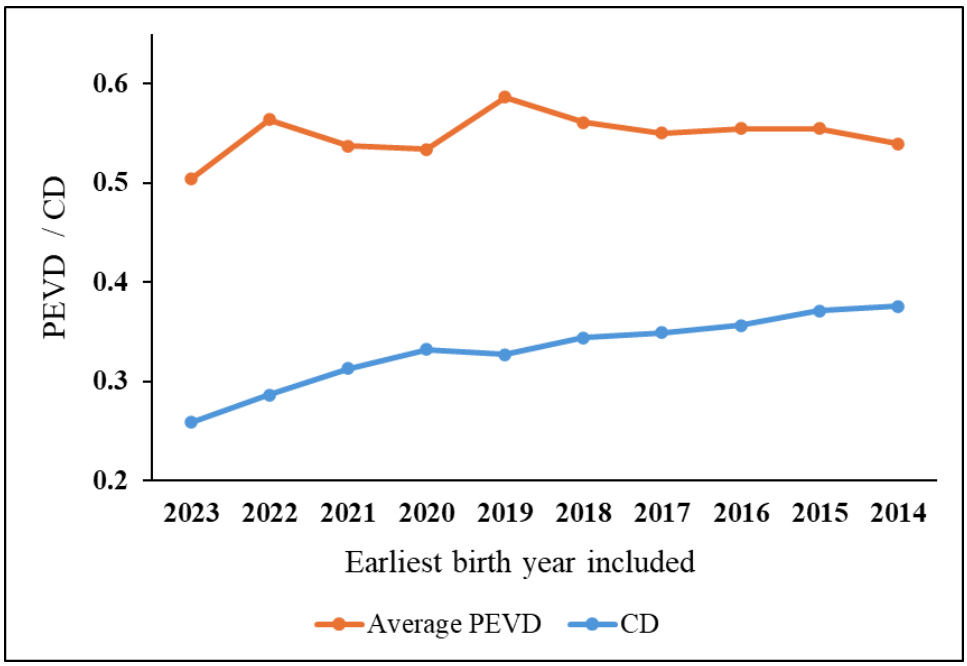


Figure 4.6: Effect of data window length on across-flock evaluation accuracy for facial eczema, measured using the average prediction error variance of differences ($PEVD_{GrpAve}$) and coefficient of determination (CD_{GrpAve}). The x-axis shows the earliest birth year included in the evaluation.

4.4 Genetic Parameter Estimates and Breeding Value Evaluation

4.4.1 Facial Eczema

4.4.1.1 Variance Components and Heritability for GGT21

Phenotypic Data Depth

The whole population analysis for FE using the full pedigree (2,182,618 animals) and phenotypic dataset (48,058 records) produced an estimated heritability of 0.44 (SE = 0.014), with additive genetic and residual variances of 0.36 and 0.46, respectively (Table 4.8). Restricting the phenotypic dataset to a 3-year data window (2022–2024 birth years), corresponding to the data window currently used in the nProve evaluation, reduced the number of records to 4,931. This 3-year reference scenario produced a heritability estimate of 0.45 (SE = 0.047), with additive genetic and residual variances of 0.33 and 0.41. respectively. This scenario was subsequently used as a reference for all further sensitivity analyses.

Table 4.8: Additive genetic variance (σ_a^2), residual variance (σ_e^2), heritability (h^2), and standard error of heritability (SE(h^2)) for gamma-glutamyl transferase response measured 21 days post-challenge (GGT21) under data depth, flock removal, and sire masking scenarios. The whole population scenario used the full pedigree and phenotypic dataset. Flock removal and sire masking analyses used phenotypic datasets restricted to the 3-year reference window (2022–2024 birth years).

Scenario	σ_a^2	σ_e^2	h^2	SE(h^2)
Whole population	0.36	0.46	0.44	0.014
1-year window	0.61	0.35	0.64	0.108
3-year window (reference)	0.33	0.41	0.45	0.047
5-year window	0.26	0.51	0.33	0.027
7-year window	0.25	0.48	0.34	0.022
9-year window	0.24	0.46	0.35	0.019
11-year window	0.28	0.48	0.37	0.017
Flock F1 removed	0.34	0.42	0.45	0.048
Flock F6 removed	0.34	0.36	0.48	0.050
Flock F18 removed	0.33	0.41	0.45	0.047
Sire S1 masked	0.33	0.41	0.45	0.047
Sire S2 masked	0.33	0.41	0.45	0.047
Sire S3 masked	0.33	0.41	0.45	0.047

Restricting the dataset further to a single birth year (2024) increased the additive genetic variance to 0.61 and decreased the residual variance to 0.35, resulting in a higher heritability estimate of 0.64 (SE = 0.108). This estimate was associated with greater error, reflecting the limited number of records. Including additional years of phenotypic data (5, 7, 9 and 11-year windows) resulted in additive genetic variances ranging from 0.24 to 0.28 and residual variances between 0.46 to 0.51. The resulting heritability estimates gradually increased from 0.33 to 0.37, remaining slightly lower than that observed in the 3-year reference scenario. Standard errors decreased progressively with increasing dataset size, reaching 0.017 in the 11-year scenario, indicating improved precision with greater phenotypic data depth. Overall, analyses based on limited phenotypic data produced higher but less precise heritability estimates, whereas including multiple years of data produced more consistent and reliable parameter estimates.

Flock Removal

To assess the influence of individual flocks on genetic parameter estimation, analyses were repeated following the sequential removal of three flocks (F1, F6, and F18) from the FE phenotypic dataset restricted to the 3-year data window (2022–2024 birth years) (Table 4.8).

Removal of flocks F1 and F6 reduced the number of phenotypic records, from 4,931 in the 3-year reference scenario to 4,719 and 4,543 records, respectively. Removal of flock F18 had no effect on dataset size because no phenotypic records from this flock were present within the 3-year window. Therefore, the high connectedness value that resulted in F18 being selected for removal was likely driven by historical records outside the 3-year window. Variance components and heritability estimates obtained following removal of flocks F1 and F6 were largely consistent with those from the 3-year reference scenario. As expected, removal of flock F18 produced identical parameter estimates to the 3-year reference analysis. These results indicate that exclusion of individual flocks had minimal impact on the estimation of genetic parameters for GGT21 and that parameter estimates were robust to the removal of individual flocks within the 3-year data window.

Sire Masking

The influence of highly prolific sires on genetic parameter estimation was assessed by sequentially masking three sires (S1, S2, and S3) from the pedigree in the analysis based on the 3-year phenotypic data window (2022–2024 birth years) (Table 4.8). Variance components and heritability estimates obtained following masking of each sire were consistent with those from the 3-year reference scenario, indicating that the exclusion of individual highly prolific sires from the pedigree had negligible impact on the estimation of genetic parameters for GGT21 within the 3-year data window.

4.4.1.2 Correlation and Re-ranking of Breeding Values for GGT21

Phenotypic Data Depth

Correlation coefficients between estimated breeding values (EBVs) derived from the data window scenarios and the whole population scenario for FE showed a clear increase with increasing data window length. Pearson correlations ranged from 0.46 for the 1-year window to 0.87 for the 11-year window, reflecting progressively stronger agreement in EBV values as additional historical data were incorporated. Spearman rank correlations followed a similar pattern, increasing from 0.40 to 0.78, indicating improved consistency in animal ranking with greater data depth. The 3-year window, corresponding to the current nProve connectedness assessment period, showed moderate correlations with the whole population (Pearson = 0.59;

Spearman = 0.48), demonstrating some divergence in both EBV values and ranking relative to analyses using all available data.

To evaluate the stability of EBVs around the current 3-year connectedness window, correlations between each data depth scenario and the 3-year window were examined (Table 4.9). The highest correlations were observed for the 5-year window (Pearson = 0.83; Spearman = 0.75), while both shorter (1-year: Pearson = 0.66; Spearman = 0.57) and longer windows (11-year: Pearson = 0.66; Spearman = 0.53) showed reduced agreement. This pattern indicates that deviations from the 3-year data window are associated with changes in EBV estimates and some degree of re-ranking among animals.

Correlations between consecutive data windows were examined to assess the incremental impact of increasing data depth on EBVs for the FE dataset. The transition from 1 to 3 years showed a moderate correlation (Pearson = 0.66; Spearman = 0.57), while extending to 5 years increased correlations (Pearson = 0.83; Spearman = 0.75), indicating improved consistency in EBV values and rankings. High correlations were observed from 5 to 7 years (Pearson = 0.90; Spearman = 0.80). Subsequent increases from 7 to 9 years and 9 to 11 years resulted in only minor changes, indicating that EBVs stabilise as data depth increases, with limited additional benefit beyond approximately 5 to 7 years.

Flock removal and Sire masking

To assess the sensitivity of EBVs to the removal of influential flocks and masking of sires, correlations between each scenario and the 3-year reference dataset were examined for FE (Table 4.9). Pearson correlations were consistently high across all scenarios, ranging from 0.93 to 1.00. Similarly, Spearman rank correlations ranged from 0.94 to 1.00, indicating minimal changes in both EBV values and animal ranking following the removal of individual flocks or masking of prolific sires.

The removal of flock F6 resulted in the lowest correlations (Pearson = 0.93; Spearman = 0.94), demonstrating a small degree of deviation in EBV estimates and minor re-ranking of animals relative to the 3-year reference. In contrast, the removal of flock F18 produced correlations of 1.00 for both Pearson and Spearman measures, as expected given that no phenotypic records from this flock were included within the 3-year data window. Similarly, all sire masking scenarios produced correlations of 1.00 for both Pearson and Spearman measures, indicating no observable change in EBV values or rankings. Overall, EBVs for the FE dataset appear

highly robust to the exclusion of individual flocks and sires, with negligible impact on both EBV magnitude and animal ranking under a 3-year data window.

Table 4.9: Pearson (below diagonal) and Spearman rank (above diagonal) correlation coefficients between estimated breeding values (EBVs) for facial eczema across 12 analysis scenarios, including data depth, flock removal, and sire masking. Spearman correlations reflect the degree of re-ranking between scenarios. All correlations were statistically significant ($P < 0.0001$); p-values are not presented.

	1-year window	3-year window	Flock F1 removed	Flock F6 removed	Flock F18 removed	Sire S1 masked	Sire S2 masked	Sire S3 masked	5-year window	7-year window	9-year window	11-year window
1-year window	1.00	0.57	0.56	0.54	0.57	0.57	0.57	0.57	0.55	0.49	0.50	0.44
3-year window	0.66	1.00	0.97	0.94	1.00	1.00	1.00	1.00	0.75	0.63	0.60	0.53
Flock F1 removed	0.64	0.99	1.00	0.91	0.97	0.97	0.97	0.97	0.74	0.65	0.61	0.54
Flock F6 removed	0.65	0.93	0.92	1.00	0.94	0.94	0.94	0.94	0.73	0.60	0.57	0.51
Flock F18 removed	0.66	1.00	0.99	0.93	1.00	1.00	1.00	1.00	0.75	0.63	0.60	0.53
Sire S1 masked	0.66	1.00	0.99	0.93	1.00	1.00	1.00	1.00	0.75	0.63	0.60	0.53
Sire S2 masked	0.66	1.00	0.99	0.93	1.00	1.00	1.00	1.00	0.75	0.63	0.60	0.53
Sire S3 masked	0.66	1.00	0.99	0.93	1.00	1.00	1.00	1.00	0.75	0.63	0.60	0.53
5-year window	0.62	0.83	0.82	0.79	0.83	0.83	0.83	0.83	1.00	0.80	0.74	0.66
7-year window	0.57	0.74	0.74	0.71	0.74	0.74	0.74	0.74	0.90	1.00	0.87	0.78
9-year window	0.55	0.71	0.71	0.68	0.71	0.71	0.71	0.71	0.86	0.94	1.00	0.88
11-year window	0.51	0.66	0.66	0.64	0.66	0.66	0.66	0.66	0.80	0.88	0.94	1.00

4.4.2 Growth

4.4.2.1 Variance Components and Heritability for Weaning Weight

Phenotypic Data Depth

The whole population analysis for growth using the full pedigree (568,247 animals) and phenotypic dataset (375,709 records) produced an estimated heritability of 0.20 (SE = 0.0045), with additive genetic, dam, and residual variances of 3.97, 3.00, and 13.00, respectively (Table 4.10). Restricting the dataset to a 3-year data window (2022–2024 birth years), corresponding to the current nProve evaluation window, reduced the number of records to 56,469. This scenario produced a heritability estimate of 0.32 (SE = 0.0168), with additive genetic, dam, and residual variances of 6.75, 4.03, and 10.01, respectively. These values were used as the reference for all subsequent sensitivity analyses.

Table 4.10: Additive genetic variance (σ_a^2), dam variance (σ_d^2) residual variance (σ_e^2), heritability (h^2), and standard error of heritability (SE(h^2)) for weaning weight (WWT) under data depth, flock removal, and sire masking scenarios. The whole population scenario used the full pedigree and phenotypic dataset. Flock removal and sire masking analyses used phenotypic datasets restricted to the 3-year reference window (2022–2024 birth years).

Scenario	σ_a^2	σ_d^2	σ_e^2	h^2	SE(h^2)
Whole population	3.97	3.00	13.00	0.20	0.0045
1-year window	9.55	6.91	5.78	0.43	0.0372
3-year window (reference)	6.75	4.03	10.01	0.32	0.0168
5-year window	5.99	3.72	10.82	0.29	0.0115
7-year window	5.16	3.50	11.41	0.26	0.0091
9-year window	5.14	3.45	11.72	0.25	0.0079
11-year window	5.07	3.35	11.90	0.25	0.0070
Flock F67 removed	6.20	4.07	10.26	0.30	0.0169
Flock F65 removed	6.68	4.06	10.05	0.32	0.0174
Flock F72 removed	6.34	4.16	10.02	0.31	0.0172
Flock F62 removed	6.44	3.97	10.24	0.31	0.0171
Sire S4 masked	6.74	4.03	10.01	0.32	0.0168
Sire S5 masked	6.75	4.03	10.01	0.32	0.0168
Sire S6 masked	6.74	4.03	10.01	0.32	0.0168

Using only the 2024 birth year WWT records resulted in higher estimates of additive genetic variance (9.55) and dam variance (6.91), while the residual variance was lower (5.78), resulting in a heritability estimate of 0.43 (SE = 0.0372). As expected, this estimate showed substantially

greater uncertainty due to the limited number of records. Inclusion of additional years of phenotypic data (5, 7, 9 and 11-year windows) gradually decreased additive genetic variance (from 5.99 to 5.07) and dam variance (from 3.72 to 3.35), while residual variance increased (from 10.82 to 11.90). Heritability estimates ranged from 0.25 to 0.29, decreasing with increasing data depth and stabilising at 0.25 in the 9 and 11-year windows. Standard errors decreased to 0.0070 in the 11-year data window scenario, indicating improved precision with increased data depth. These findings indicate that analyses based on very limited phenotypic data may overestimate heritability, whereas inclusion of multiple years provides more reliable and precise estimates of variance components and heritability.

Flock Removal

To assess the influence of individual flocks on genetic parameter estimation for WWT, four flocks (F67, F65, F72, and F62) were sequentially removed from the phenotypic dataset restricted to the 3-year birth window (2022–2024) (Table 4.10). Removal of each flock resulted in reductions in the number of phenotypic records from 56,469 in the 3-year reference scenario to between 52,174 and 54,344 records. Variance components and heritability estimates were largely consistent with those from the 3-year reference scenario, indicating that removal of individual flocks had minimal impact on the estimation of genetic parameters for WWT within the 3-year data window.

Sire Masking

The influence of highly prolific sires (S4, S5, and S6) was evaluated by sequentially masking each sire in the pedigree while using the phenotypic dataset restricted to the 3-year data window (2022–2024 birth years) (Table 4.10). Across all sire masking scenarios, additive genetic variance, dam variance, residual variance and heritability estimates remained largely unchanged from the 3-year reference scenario. These results indicate that exclusion of highly prolific sires had minimal impact on the estimation of genetic parameters for WWT using a 3-year data window.

4.4.2.2 Correlation and Re-ranking of Breeding Values for Weaning Weight

Phenotypic Data Depth

Correlation coefficients between EBVs derived from the data window scenarios and the whole population scenario increased progressively as the length of the data window expanded. Pearson correlations rose from 0.37 for the 1-year window to 0.76 for the 11-year window, indicating increasing agreement in EBV values as more historical data were included. A similar trend was observed for Spearman rank correlations, which increased from 0.37 to 0.73 across the same range of scenarios, indicating improved consistency in animal ranking with greater data depth. The 3-year window, corresponding to the current nProve connectedness evaluation window, showed moderate correlations with the whole population (Pearson = 0.47; Spearman = 0.40), indicating deviation in both EBV magnitude and ranking relative to analyses using all available data.

When compared to the 3-year reference scenario, Pearson correlations were highest for the 5-year window (0.80) and decreased for both shorter (1-year = 0.65) and longer windows (11-year = 0.62) (Table 4.11). A similar pattern was observed for Spearman rank correlations, which were highest for the 5-year window (0.74) and declined for both shorter (1-year = 0.54) and longer windows (11-year = 0.51). This indicates that deviations from the 3-year data window are associated with changes in EBV values and potential re-ranking of animals.

Analysis of correlations between consecutive data windows further illustrated the effect of incrementally increasing data depth. The correlation between the 1-year and 3-year windows was moderate (Pearson = 0.65; Spearman = 0.54), indicating that the inclusion of two additional years of data improved agreement in both EBV values and animal ranking. Expanding the window from 3 to 5 years resulted in higher correlations (Pearson = 0.80; Spearman = 0.74), showing that EBVs became more consistent relative to the current 3-year reference scenario. Further extension from 5 to 7 years produced very high correlations (Pearson = 0.90; Spearman = 0.87), with EBV values and rankings remaining largely stable. Subsequent increases from 7 to 9 years and 9 to 11 years led to only minimal changes, indicating diminishing returns from the inclusion of additional data beyond approximately 5 to 7 years.

Flock removal and Sire masking

The sensitivity of EBVs to the removal of influential flocks and masking of sires was evaluated by examining correlations between each scenario and the 3-year reference dataset for growth

(Table 4.11). Pearson correlations remained high across all scenarios, ranging from 0.96 to 1.00. Spearman rank correlations showed a similar pattern, varying between 0.95 and 1.00, indicating that both EBV values and animal rankings were largely unaffected by the exclusion of individual flocks or sires.

Among the flock removal scenarios, the exclusion of flock F72 resulted in the lowest correlations (Pearson = 0.96; Spearman = 0.95), reflecting a small degree of variation in EBV estimates and minor re-ranking of animals relative to the 3-year reference. Removal of flocks F67, F65, and F62 produced correlations closer to 1, indicating minimal impact on EBV values and rankings.

All sire masking scenarios showed near-complete agreement with the reference dataset (Pearson = 1.00; Spearman = 0.99-1.00), demonstrating negligible influence of individual sires on EBV estimates under the 3-year data window. Overall, EBVs for the growth dataset demonstrate a high level of robustness to the removal of individual flocks and masking of prolific sires.

Table 4.11: Pearson (below diagonal) and Spearman rank (above diagonal) correlation coefficients between estimated breeding values (EBVs) for growth across 13 analysis scenarios, including data depth, flock removal, and sire masking. Spearman correlations reflect the degree of re-ranking between scenarios. All correlations were statistically significant ($P < 0.0001$); p-values are not presented.

	1-year window	3-year window	Flock F67 removed	Flock F65 removed	Flock F72 removed	Flock F62 removed	Sire S4 masked	Sire S5 masked	Sire S6 masked	5-year window	7-year window	9-year window	11-year window
1-year window	1.00	0.54	0.52	0.53	0.54	0.54	0.54	0.54	0.53	0.54	0.49	0.44	0.42
3-year window	0.65	1.00	0.96	0.96	0.95	0.98	1.00	1.00	0.99	0.74	0.66	0.57	0.51
Flock F67 removed	0.63	0.98	1.00	0.93	0.91	0.95	0.96	0.96	0.95	0.73	0.64	0.55	0.50
Flock F65 removed	0.64	0.98	0.96	1.00	0.91	0.95	0.96	0.96	0.96	0.75	0.66	0.56	0.50
Flock F72 removed	0.63	0.96	0.93	0.93	1.00	0.94	0.95	0.95	0.94	0.70	0.62	0.54	0.50
Flock F62 removed	0.65	0.99	0.97	0.97	0.94	1.00	0.98	0.98	0.98	0.72	0.65	0.56	0.51
Sire S4 masked	0.65	1.00	0.98	0.98	0.96	0.99	1.00	0.99	0.99	0.74	0.67	0.57	0.52
Sire S5 masked	0.65	1.00	0.98	0.98	0.96	0.99	1.00	1.00	0.99	0.73	0.66	0.57	0.51
Sire S6 masked	0.64	1.00	0.97	0.98	0.95	0.98	1.00	1.00	1.00	0.72	0.65	0.56	0.51
5-year window	0.58	0.80	0.79	0.79	0.76	0.78	0.80	0.79	0.78	1.00	0.87	0.72	0.63
7-year window	0.54	0.74	0.73	0.73	0.70	0.73	0.74	0.73	0.73	0.90	1.00	0.86	0.76
9-year window	0.50	0.67	0.65	0.66	0.63	0.66	0.67	0.67	0.66	0.78	0.90	1.00	0.88
11-year window	0.45	0.62	0.60	0.61	0.59	0.61	0.62	0.62	0.62	0.70	0.83	0.92	1.00

Chapter 5: Discussion and Conclusions

This study investigated how changes to inclusion criteria influenced genetic evaluation structure, flock connectedness, and the accuracy of the resulting across-flock genetic comparisons within New Zealand's sheep industry. The objectives were; to determine how variation in the depth of data influences levels of connectedness; assess the impact of removing highly connected flocks or masking prolific sires on connectedness; and evaluate how the application of a less conservative connectedness criteria would affect the reliability of across-flock connectedness within the evaluation. The analyses demonstrated that the optimal level of connectedness is not fixed, but instead depends on the trait being evaluated as well as the structure and depth of the available data.

5.1 nProve Connectedness

A high level of connectedness was observed for the growth analysis, reflecting the widespread routine recording of weaning weight (WWT) across flocks, which resulted in large, well-linked datasets. In contrast, facial eczema (FE) tolerance measured through gamma-glutamyl transferase response measured 21 days post-challenge (GGT21) is costly and logistically challenging to measure (Amyes and Hawkes 2014), leading to sparse, uneven data recording. Consequently, connectedness for FE was weaker at the same data depth, and estimated breeding values (EBVs) are more reliant on limited genetic links and historical data. This increases sensitivity to data structure and reduces overall confidence in across-flock comparisons.

Despite the importance of connectedness in genetic evaluation systems, there is limited published research evaluating how factors such as the depth of data, or the removal of flocks or sires, influence connectedness in breeding populations. Results from this study indicated that increasing the data window i.e. the number of years of data included, improves connectedness for sparsely recorded traits such as FE, highlighting the importance of historical records in establishing genetic links. This showed that short evaluation windows may underestimate connectedness. This is supported by Holmes et al. (2015) who reported that the SIL/nProve connectedness measure can underestimate true genetic connectedness by ignoring additional relationships captured through more complete pedigree structures. The results from

Holmes et al. (2015) demonstrated that incorporating more extensive pedigree and genomic relationship information increased estimated connectedness and reduced the number of flocks classified as isolated, indicating that meaningful genetic links extend beyond the short-term window. As a result, underestimation of connectedness under the current nProve framework may prevent some flocks from meeting the connectedness requirements for public reporting of across-flock breeding values and selection indices in nProve.

However, incorporating all available pedigree information without accounting for the age of records may introduce its own bias (Holmes et al. 2015), as genetic links derived from distant ancestors can inflate estimates of connectedness despite contributing little to the exchange of recent genetic information. Because relationship matrices weight historical and recent connections equally, this can overstate the level of meaningful connectedness between flocks (Kennedy and Trus 1993; Holmes et al. 2015). Future research could investigate incorporating age-based weightings into connectedness calculations, allowing recent genetic linkages to contribute more strongly than historical ones. This highlights a trade-off in the selection of evaluation windows, where shorter timeframes may exclude relevant genetic information, whilst longer evaluation windows may incorporate historical connections that are irrelevant to current breeding populations. In contrast, the analysis of growth data showed little sensitivity to evaluation window length, indicating that sufficient connectedness for routinely recorded traits can be achieved using relatively recent WWT data.

In addition to the influence of the evaluation window, connectedness was affected by the underlying structure of the population. The reduction in connectedness for FE following flock removal indicated that the genetic network for this trait was more dependent on a smaller number of contributing flocks. This highlights a potential vulnerability whereby the loss of key flocks could reduce the strength of genetic links across the evaluation. In contrast, the stability of connectedness for growth following flock removal indicated a more evenly distributed network, making it more resilient to changes in industry participation.

The lack of impact from sire masking on connectedness for either trait indicated that genetic linkage was not overly reliant on a small number of influential sires but instead reflects a more distributed pattern of genetic exchange across flocks. This is desirable as it indicates that connectedness is relatively robust and less vulnerable to the loss of individual animals.

5.2 Accuracy of Genetic Evaluations

Beyond describing patterns of connectedness, it is important to assess how changes in evaluation window length and the resulting changes in connectedness influence the accuracy of across-flock genetic evaluations. However, there is limited published research evaluating how changes in the evaluation window length influence connectedness metrics such as the coefficient of determination (CD) and prediction error variance of differences (PEVD), particularly under commercial sheep breeding conditions. Results from this study's Genetic Connectedness Analysis (GCA) provided practical insight into how data depth affects the reliability of EBV comparisons in nProve populations.

Extending the evaluation window from two to eleven years of FE data showed how the inclusion of each additional birth year influences the reliability of across-flock comparisons. The CD increased steadily as additional years of data were included, indicating improved reliability of across-flock EBV comparisons. Similarly, a simulation study by Yu et al. (2018) demonstrated that increasing connectedness, indicated by higher CD values resulting from a greater proportion of linked individuals, was associated with increased prediction accuracy. Prediction accuracy was defined as the correlation between true and estimated breeding values. However, improvements in CD are not expected to continue indefinitely. In this study, CD plateaued around the 2019–2020 birth years (corresponding to a 5 to 6-year data window). This plateau likely reflected the saturation of informative genetic links between flocks, where additional records contributed limited new relationship information. A similar pattern was observed in a simulation study where CD increased up to a point but subsequently declined as genetic variability between groups was reduced, reflecting the penalisation of overly similar populations (Yu et al. 2018).

In contrast, PEVD fluctuated across the data windows. This difference likely reflects the distinct properties of the two metrics. The PEVD quantifies the absolute uncertainty associated with comparing breeding values, as it is directly derived from prediction error variances and covariances and is not scaled by genetic variability (Yu and Morota 2021). As a result, it may be more sensitive to changes in data structure. Whereas CD represents the relative reliability of comparisons by scaling prediction error variance relative to genetic variance and relationships among individuals (Yu and Morota 2021), resulting in a more stable trend. Increases in PEVD observed when extending the data window (e.g. for the 2022 and 2019 birth years) coincided with the inclusion of birth years with lower levels of GGT21 recording. Reduced recording in these years decreases the amount of phenotypic information available

and limits the number of genetic links between flocks which the cohort is contributing, thereby reducing the accuracy of EBV estimation and constraining improvements in connectedness. Effectively, the inclusion of sparsely recorded years lowers the average information content contributed per year. This is similar to the effect that would be seen by dropping out a year of data from the analysis. In a related sheep study, EBV accuracy was quantified by comparing EBVs estimated under reduced-recording scenarios with those obtained using a full dataset, demonstrating an asymptotic improvement in accuracy as the proportion of phenotypic records increased (Craig et al. 2015).

In this study, a 3-year phenotypic data window resulted in a CD of 0.286 for FE, indicating a moderate level of connectedness where across-flock comparisons are possible, although there remained scope for improvement. Previously, Fouilloux et al. (2008) reported that a CD of 0.40 in a large-scale beef cattle evaluation was sufficient to classify herds as connected for reliable comparison and genetic evaluations. Although such thresholds are context-dependent, the lower CD observed here indicates that connectedness between flocks may be below levels typically considered optimal. Extending the data window to five years increased CD to 0.332, indicating improved connectedness, although still below the 0.40 threshold reported by Fouilloux et al. (2008) for sufficiently connected herds.

5.3 Genetic Parameter Estimates and Breeding Value Evaluation

5.3.1 Variance Components and Heritability

Variations in estimated genetic parameters across phenotypic data windows highlights the influence of the depth of data on both the magnitude and precision of heritability estimates for FE and growth. Analyses based on limited data (e.g. a single birth year) produced inflated heritability estimates with large standard errors, whereas estimates derived from larger data windows were lower, more stable, and estimated with greater precision.

Heritability estimates for both GGT21 (FE) and WWT (growth) were generally reduced when the phenotypic data window was expanded beyond the 3-year reference scenario, stabilising at lower values beyond the 5-year window. This reduction relative to the 3-year estimate likely reflects changes in the relative contribution of additive and residual variance components. Expanding the data window introduced a wider range of environmental conditions, which likely increased the contribution of residual environmental variation relative to changes in

additive genetic variation, resulting in lower heritability estimates (Charmantier and Garant 2005). Differences in measurement scale and trait architecture likely contributed to the smaller variance components observed for GGT21 compared to WWT, as GGT21 was constrained to a much narrower range of values.

For growth, the whole population analysis produced the lowest heritability estimate, but also the lowest standard error. This likely reflected the inclusion of older WWT records dating back to 1995, where measurement techniques, parentage recording systems and management practices were less consistent. Historical developments in New Zealand sheep performance recording indicate substantial improvements in data collection accuracy and standardisation over time (Cooke et al. 2015). The inclusion of such historical data may therefore introduce phenotypic noise, inflating residual variance and reducing heritability estimates. This interpretation is supported by a study reporting that removal of older phenotypic records could reduce prediction bias and maintain or even improve EBV accuracy (Granado-Tajada and Ugarte 2024), indicating that not all data contributes equally to genetic evaluations. However, the impact of historical data is highly dependent on the structure and characteristics of the dataset, including trait definition, recording consistency, and data availability over time. In contrast, the whole population heritability estimate for FE was similar to that obtained from the 3-year data window, likely reflecting both the limited availability of early records and the more consistent nature of FE tolerance phenotyping. This consistency is supported by the use of the RamGuard™ system, where data is collected under controlled challenge conditions with veterinary oversight (Beef + Lamb New Zealand Genetics 2025), likely reducing the influence of historical data on parameter estimation.

Heritability estimates for FE in this study were consistent with previously reported values. For example, estimates of 0.44 (± 0.03) for GGT21 have been reported using a RamGuard™ dataset (McRae et al. 2016), aligning closely with the whole population (0.44) and 3-year (0.45) estimates obtained in the present study. These close estimations reflect the use of comparable data sources, as the FE analyses in this study were based on the full nProve database download. However, heritability estimates for WWT were greater than those reported in previous studies, where estimates are approximately 0.14–0.15 (Wuliji et al. 2011; Pickering et al. 2012), compared with 0.20 for the whole population and 0.32 for the 3-year window in this study. This difference likely reflects variation in data structure and analytical approaches between studies. For example, Wuliji et al. (2011) analysed a relatively small, closed Romney population under selection. Reduced genetic diversity and long-term selection in this

population likely reduced the magnitude of additive genetic variance for WWT, resulting in lower heritability estimates. Although Pickering et al. (2012) also used data from the Sheep Improvement Limited (SIL) database, their analysis incorporated a much larger and more complete representation of the national population with records for more than two million animals spanning a wider range of flocks, breeds, and environmental conditions, than the narrower subset of SIL data used in this study. This broader population structure captured greater environmental and pedigree complexity, which may alter the separation of genetic and residual variance, potentially contributing to the lower heritability estimates.

Misidentification of sires has previously been shown to bias estimates of heritability downward, due to incorrect assignment of paternal half-sib relationships and the resulting underestimation of between sire variance (Van Vleck 1970). However, this represents a scenario in which pedigree relationships are incorrectly specified. In this study, we instead assessed the sensitivity of genetic parameter estimates to the intentional masking of highly prolific sires within an otherwise complete pedigree, rather than pedigree misidentification. Under this framework, genetic parameter estimates remained robust within the 3-year data window for both FE and growth. The minimal changes observed in variance components and heritability indicate that no single flock or sire disproportionately influenced the estimation of genetic parameters.

5.3.2 Correlation and Re-ranking of Breeding Values

Variation in EBV correlations across phenotypic data windows further demonstrated the importance of the depth of data for the stability and reliability of genetic evaluations. Estimated breeding values derived from limited data (e.g. a single birth year) presented low to moderate correlations with the whole population analysis, indicating differences in both EBV magnitude and animal ranking. As the whole population analysis utilised a full pedigree and phenotypic dataset, it provided the closest approximation to true breeding values against which reduced data scenarios can be compared. As additional years of data were included, correlations increased progressively, demonstrating improved consistency with the full-population evaluation. This supports the expectation that increasing the depth of data improves the stability of EBVs by providing more information from related animals (Yang and Su 2016). Consistent with this, a simulation study in sheep showed that the accuracy of EBVs increased progressively as phenotypic and pedigree data accumulated overtime (Casellas and Piedrafita 2015). In that study, accuracy was quantified as the Pearson correlation between true

(simulated) breeding values and EBVs from Best Linear Unbiased Prediction (BLUP) evaluations, and was shown to increase rapidly in early years of data collection before reaching a plateau.

The moderate correlations observed between the 3-year reference scenario and the whole population scenario indicated that EBVs derived from the 3-year window may not fully capture the information available in the complete dataset, potentially leading to more variable estimates of genetic merit. Analysis of correlations between consecutive data windows further highlighted diminishing returns associated with increasing the depth of data. The greatest improvements in correlation were observed when expanding from shorter windows (1 to 3 years, and 3 to 5 years), whereas beyond approximately 5 to 7 years, correlations plateaued and additional data resulted in only minor changes to EBVs and animal ranking. This indicates that once a sufficient volume of data captures key sources of variation and genetic linkage, further increases in data depth provide limited additional benefit.

Sensitivity analyses for both growth and FE showed that EBVs were highly robust to the removal of individual flocks and masking of highly prolific sires under the 3-year data window. High correlations across all removal scenarios indicated that no single flock or sire had a disproportionate influence on EBV estimates or animal ranking. This showed that the flocks within the datasets are well connected, with genetic information distributed across multiple flocks. Slightly lower correlations following the removal of specific flocks (e.g. F6 for FE and F72 for growth) implied that these flocks contributed relatively more information, although their overall influence remains limited. While there is limited research directly evaluating the impact of removing individual sires, a simulation study has shown that EBV accuracy can decline substantially when sire information is unavailable within the pedigree (Casellas and Piedrafita 2015). However, this scenario involved the complete absence of sire identification across the pedigree rather than the masking of individual sires. This weakened the relationship structure used in BLUP evaluations, reducing the accuracy of genetic comparisons between animals, which in turn reduced expected genetic gain. In contrast, the minimal impact observed here indicated that in large, well-connected populations with largely complete pedigrees, genetic evaluations are robust to loss of individual contributors.

5.4 Implications

Overall, the findings of this study have important implications for the design of genetic evaluations such as those used in nProve. Sparsely recorded traits such as GGT21 (FE) rely more heavily on historical data to establish genetic links, making evaluations more sensitive to data availability and structure. In contrast, routinely recorded traits such as WWT (growth) achieve sufficient connectedness for stable and robust EBVs using fewer years of data. While increasing the depth of data generally improves connectedness, EBV stability, and the precision of genetic parameter estimates, the results of this study indicate that benefits diminish beyond approximately five years of data. This indicates that incorporating all available historical data may not be necessary. Instead, an intermediate data window may provide an optimal balance between EBV accuracy and computational efficiency.

There is limited published research evaluating how evaluation window length influences connectedness and EBV accuracy in commercial sheep populations. This study is among the first to investigate these relationships within the nProve framework. The current 3-year connectedness window may be overly conservative, as it potentially excludes older but still informative and relevant pedigree and performance links that contribute meaningfully to genetic connectedness between flocks. Ignoring these historical connections may limit the public reporting of across-flock breeding values for some flocks in nProve, despite still sharing relevant genetic links through common ancestors within a slightly longer timeframe. Extending nProve's current 3-year connectedness evaluation window to approximately five years for both FE and growth could therefore enable greater flock participation, improve connectedness, and provide a practical balance between improving CD, maintaining PEVD, increasing EBV accuracy, and keeping the evaluation computationally efficient.

These findings also have practical implications for how new traits should be incorporated into the nProve evaluation system. For traits recently introduced into the recording framework, such as emerging health, welfare or efficiency traits, a minimum period of data accumulation is required before reliable across-flock evaluations can be undertaken. Based on the patterns observed in this study, a minimum of three years of performance data is likely necessary before stable and reliable EBVs can be produced. Shorter windows would risk weak connectedness, inflated heritability estimates, and a reduced EBV accuracy. Once a trait reaches this minimum threshold, a rolling 5-year evaluation window appears to provide an optimal balance between capturing informative pedigree and performance links, maintaining adequate connectedness, and avoiding diminishing returns. Implementing such a framework would ensure that nProve

can deliver robust across-flock comparisons for new traits while maintaining computational efficiency and avoiding unnecessary delays in trait deployment.

5.5 Limitations

Several limitations should be considered when interpreting the findings of this study. First, the growth analyses were based on a subset of the full nProve database rather than the complete national dataset. Therefore, it may not fully capture the extent of sire usage, flock relationships, and historical pedigree depth present in the full evaluation system. As a result, estimates of connectedness and EBV stability for growth may be somewhat conservative relative to what would be observed using the entire dataset.

A further limitation relates to the imbalance in dataset size and structure between the FE and growth analyses. The phenotypic dataset used for connectedness analyses for FE (48,058 GGT21 records) was much smaller than the growth dataset (375,709 WWT records). Due to computational constraints associated with the large number of WWT records per year, the analysis of genetic evaluation accuracy using PEVD and CD could not be replicated for the growth dataset, restricting this assessment to the FE dataset only. Based on preliminary analyses for growth using phenotypic data windows of up to three years (unpublished results), a similar pattern to FE would be expected, with CD increasing as the data window was extended before reaching a plateau. Differences in dataset size reduced comparability in terms of statistical power and network density, meaning that some of the observed differences in sensitivity to evaluation window length, flock removal and sire masking may reflect structural imbalances rather than inherent differences between traits. At the same time, this contrast in dataset scale underpins the rationale for including both traits in the study with FE and growth representing two extremes of recording intensity within the industry, allowing the analyses to illustrate how connectedness and EBV stability behave under very different data environments. Nevertheless, caution is needed when generalising findings from one trait to the other or to traits with intermediate recording patterns.

The study assumed that flock behaviour remains constant under different evaluation window policies. In practice, extending the connectedness evaluation window from three to five years may alter farmer behaviour. If breeders perceive that they have a longer period before connectedness deteriorates, they are likely to reduce the frequency with which they share rams across flocks. This behavioural response could weaken future connectedness, potentially

offsetting some of the benefits identified in this study. Any change to the connectedness evaluation window should therefore be accompanied by clear communication that the purpose of extending the window is to incorporate informative historical genetic links that contribute meaningful connectedness between flocks, rather than to reduce the frequency of genetic exchange between flocks. The long-term relevance of the results therefore depends on how breeders respond to changes in the nProve approach to measuring connectedness, an aspect that was beyond the scope of the current analysis.

The recommendation to extend the evaluation window to five years applies to the current nProve connectedness method. This method is intentionally conservative and relies solely on counts of shared progeny from link sires within a rolling 3-year window, meaning it captures only a subset of the genetic links present in the population. Although this study used the GCA framework to evaluate changes in evaluation accuracy, we were unable to quantify how closely the nProve connectedness measure aligns with connectedness derived from the GCA framework. Correlating flock-level connectedness values from both methods would provide valuable insight into whether they identify similar patterns of linkage or respond differently to changes in the depth of data, flock removal, or sire masking. Establishing the degree of agreement between these approaches would strengthen confidence in the recommended data window settings and clarify whether conclusions are robust across alternative connectedness methodologies.

In addition to the limitations of the nProve method itself, the connectedness metrics used within the GCA framework should also be considered. While PEVD and CD are widely accepted indicators of across-flock evaluation accuracy, they represent only a subset of the statistical approaches available. As outlined in the literature, the GCA package supports a broader suite of PEV-based and VE-based metrics, each capturing slightly different aspects of the information structure among flocks (Yu and Morota 2021). Because this study focused on PEVD and CD, it is possible that the sensitivity of connectedness to evaluation window length may vary under alternative metrics. Future work incorporating a wider range of connectedness statistics would provide a more comprehensive assessment of the robustness of these conclusions.

Several additional analyses were beyond the scope of this project but may have provided valuable insight into the drivers of flock connectedness. For example, systematically removing individual birth years to identify which cohorts contribute most strongly to across-flock linkage

could help determine whether connectedness is driven disproportionately by specific recording periods or changes in industry participation. Such analyses may also reveal whether genetic exchange between flocks has occurred consistently over time or has been concentrated within particular periods. These unexamined aspects limit the ability to identify the underlying factors contributing to connectedness and highlight opportunities for future research.

5.6 Conclusions

Genetic connectedness is fundamental to the success of across-flock genetic evaluations in the New Zealand sheep industry. As the industry increasingly relies on genetic evaluation systems such as nProve to drive selection decisions and accelerate genetic gain, having a robust approach to measuring connectedness is becoming increasingly important.

This study demonstrated that the optimal level of connectedness was dependent on the quality and structure of the available data. Traits with extensive routine recording, such as growth, maintained relatively stable connectedness under shorter evaluation windows. In contrast, connectedness for sparsely recorded traits such as FE was more sensitive to changes in data depth and relied more heavily on historical pedigree and performance records to maintain genetic links between flocks.

Increasing the depth of data generally improved connectedness, EBV stability, and the precision of genetic parameter estimates. However, the benefits diminished beyond approximately five years of data, indicating that incorporating all historical records would provide limited additional value. Connectedness and EBV estimation were relatively robust to the removal of highly connected flocks and the masking of prolific sires, indicating that genetic exchange within the evaluated populations is not overly dependent on a small number of contributors.

Overall, the findings indicate that the current 3-year nProve connectedness evaluation window may be overly conservative, potentially preventing some flocks with meaningful genetic connections from being recognised as sufficiently connected for public reporting of across-flock breeding values. Extending the evaluation window to approximately five years would likely provide a more effective balance between increasing the number of flocks eligible for public reporting, enabling reliable across-flock comparisons, and maintaining computational

efficiency. In addition, the results highlight the importance of allowing sufficient data accumulation before implementing across-flock evaluations for newly recorded traits.

This research contributes to the limited body of literature investigating connectedness within commercial sheep breeding populations and provides practical insight into how evaluation window length influences the reliability of genetic evaluations in New Zealand sheep. The findings have direct implications for the future development of nProve and related genetic evaluation systems, supporting more flexible and evidence-based approaches to measuring connectedness that may ultimately improve access to reliable across-flock comparisons, selection accuracy, and long-term rates of genetic gain.

References

Amadeu RR, Garcia AAF, Munoz PR, Ferrão LFV. 2023. AGHmatrix: genetic relationship matrices in R. Bioinformatics 39: btad445. https://doi.org/10.1093/bioinformatics/btad445

Amer PR. 2009. Rates of genetic progress being achieved throughout the New Zealand ram breeding industry. Proceedings of the New Zealand Society of Animal Production 69: 155-157.

Amyes NC, Hawkes AD. 2014. Ramguard - increasing the tolerance to facial eczema in New Zealand sheep. Proceedings of the New Zealand Society of Animal Production 74: 154-157.

Barrett T, Dowle M, Srinivasan A, Gorecki J, Chirico M, Hocking T, Schwendinger B, Krylov I. 2025. data.table: Extension of `data.frame` (R package version 1.17.8), CRAN. https://CRAN.R-project.org/package=data.table

Beban A, Korson C, Reid J, Procter J, Halley J, Mackenzie K. 2024. Building a place-based social licence to operate. https://doi.org/10.57935/AGR.26001745.

Beef + Lamb New Zealand. 2012. Central Progeny Test Results 2011-2012. https://www.sil.co.nz/files/1487641543_10th.pdf [8 June 2026]

Beef + Lamb New Zealand. 2019. Facing up to Facial Eczema.

Beef + Lamb New Zealand. 2024. New Zealand Genetic Engine Version 6 launched. https://beeflambnz.com/news/new-zealand-genetic-engine-version-6-launched [8 June 2026]

Beef + Lamb New Zealand Economic Service. 2020. Compendium of New Zealand Farm Facts 44th Edition P20001. https://beeflambnz.com/knowledge-hub/PDF/farm-facts-2020.pdf [8 June 2026]

Beef + Lamb New Zealand Economic Service. 2024. Stock Number Survey as at 30 June 2024 P24003. https://beeflambnz.com/knowledge-hub/PDF/stock-number-survey-30-june-2024.pdf [8 June 2026]

Beef + Lamb New Zealand Genetics. 2015. Central Progeny Test Results 2014-2015. https://www.sil.co.nz/files/1487641521_13th.pdf [8 June 2026]

Beef + Lamb New Zealand Genetics. 2016. Best Practice in Performance Recording for SIL Genetic Evaluations. https://www.sil.co.nz/files/1489435846_Best%20Practice%20Guide%20Dec%202016_WEB.pdf [8 June 2026]

Beef + Lamb New Zealand Genetics. 2019. B+LNZ Genetics Sheep Progeny Test HUB Annual report, April 2019. https://www.sil.co.nz/files/1561346416_HUB%20Annual%20report-Mar%202019%20V4.pdf [8 June 2026]

Beef + Lamb New Zealand Genetics. 2020. B+LNZ Genetics launches revolutionary tool. https://www.blnzgenetics.com/news/media-releases/blnz-genetics-launches-revolutionary-too [8 June 2026]

Beef + Lamb New Zealand Genetics. 2022. Sheep Industry Breeding Objectives 2022. https://www.sil.co.nz/files/1661310874929.pdf [8 June 2026]

Beef + Lamb New Zealand Genetics. 2023. Low Input Sheep Progeny Test Public Report January 2023. https://www.sil.co.nz/files/Low%20Input%20Public%20Report%20-%20Jan23%20FINAL.pdf [8 June 2026]

Beef + Lamb New Zealand Genetics. 2024. New Zealand Genetic Evaluation (NZGE) upgrade (V6) - 2024 https://www.sil.co.nz/files/1731892529222.pdf [8 June 2026]

Beef + Lamb New Zealand Genetics. 2025. Facial Eczema Tolerance. https://www.sil.nz/files/174674064762.pdf [8 June 2026]
Bishop SC, Morris CA. 2007. Genetics of disease resistance in sheep and goats. Small ruminant research 70: 48-59. https://doi.org/10.1016/j.smallrumres.2007.01.006
Blair HT, Garrick DJ. 2007. Application of new technologies in sheep breeding. New Zealand Journal of Agricultural Research 50: 89-102. https://doi.org/10.1080/00288230709510285
Blair HT, Garrick DJ, Rae AL, Wickham GA. 1984. Selection response in New Zealand Romney sheep: 1. Selection for wool-free faces. New Zealand Journal of Agricultural Research 27: 329-336. https://doi.org/10.1080/00288233.1984.10430632
Bolormaa S, Brown DJ, Swan AA, van der Werf JHJ, Hayes BJ, Daetwyler HD. 2017. Genomic prediction of reproduction traits for Merino sheep. Animal genetics 48: 338-348. https://doi.org/10.1111/age.12541
Booth PL, Espig M, Kirk N, Sinner J, Campbell H, Edwards P, Kannemeyer R, Dear C. 2024. Perceptions of the 'good farmer'and social licence to operate in Aotearoa New Zealand. Journal of the Royal Society of New Zealand 55: 1589–1609. https://doi.org/10.1080/03036758.2024.2351910
Bora SK. 2023. Applications of Genomic Selection in Animal Breeding: Challenges and Opportunities. World News of Natural Sciences 48: 107-117.
Brito LF, McEwan JC, Miller SP, Pickering NK, Bain WE, Dodds KG, Schenkel FS, Clarke SM. 2017. Genetic diversity of a New Zealand multi-breed sheep population and composite breeds' history revealed by a high-density SNP chip. BMC Genetics 18: 25. https://doi.org/10.1186/s12863-017-0492-8
Brosnahan CL, Warburton D, Cotter N, Tanner JC, Greer AW. 2025. Trialling the SmartWorm® application in New Zealand sheep farms. International Journal for Parasitology: Drugs and Drug Resistance 29: 100616. https://doi.org/10.1016/j.ijpddr.2025.100616
Brown DJ, Huisman AE, Swan AA, Graser HU, Woolaston RR, Ball AJ, Atkins KD, Banks RG. 2007. Genetic evaluation for the Australian sheep industry. Proceedings of the Association for the Advancement of Animal Breeding and Genetics 17: 187-194.
Casellas J, Piedrafita J. 2015. Accuracy and expected genetic gain under genetic or genomic evaluation in sheep flocks with different amounts of pedigree, genomic and phenotypic data. Livestock Science 182: 58-63. https://doi.org/10.1016/j.livsci.2015.10.014
Charmantier A, Garant D. 2005. Environmental quality and evolutionary potential: lessons from wild populations. Proceedings: Biological Sciences 272: 1415-1425. https://doi.org/10.1098/rspb.2005.3117
Clarke EA. 1967. Performance recording of sheep. Proceedings of the New Zealand Society of Animal Production 27: 29-45.
Clarke JN. 2003. The benefits of across-flock analysis. Proceedings of the New Zealand Society of Animal Production 63: 204-208.
Clarke SM, Henry HM, Dodds KG, Jowett TWD, Manley TR, Anderson RM, McEwan JC. 2014. A high throughput single nucleotide polymorphism multiplex assay for parentage assignment in New Zealand sheep. PloS one 9: e93392. https://doi.org/10.1371/journal.pone.0093392
Cocks J, Brown C. 2005. Long-term analysis of price, productivity, and profitability trends in New Zealand agriculture. Proceedings of the New Zealand Grassland Association 67: 9-17.
Cooke AC, Davys JW, Alder G, Young MJ. 2015. Innovations in sheep performance recording in New Zealand. ICAR Technical Series No. 19: 137-142.

Copland RJ, Stevens DR. 2012. The changing face of southern New Zealand farming: opportunities of land use change. Proceedings of the New Zealand Grassland Association 74: 1-6. https://doi.org/10.33584/jnzg.2012.74.2876

Craig HJB, Jopson NB, Young MJ. 2015. Change in accuracy of estimated breeding values for VIAscan lean-meat yield when estimated from half-sib records. Proceedings of the Association for the Advancement of Animal Breeding and Genetics 21: 413-416.

Cranston L, Corner-Thomas R, Ridler A, Kenyon P, Greer A. 2023. Sheep Production. In: Martin N, Sneddon N ed. Livestock Production in New Zealand. Revised ed, Massey University Press. Pp. 84-125

Craw H. 2009. What will influence the future direction of New Zealand sheep genetics? Kellogg Rural Leaders Programme report series. https://hdl.handle.net/10182/5923 [8 June 2026]

Crawford AM, Anderson RM, McEwan KM. 2007. Uptake of DNA testing by the livestock industries of New Zealand. Proceedings of the New Zealand Society of Animal Production 67: 168-174.

Crawford AM, Tate ML, McEwan JC, Kumaramanickavel G, McEwan KM, Dodds KG, Swarbrick PA, Thomson P. 1993. How reliable are sheep pedigrees? Proceedings of the New Zealand Society of Animal Production 53: 363-370. https://doi.org/10.13140/2.1.3680.5762

Cruickshank GJ. 2003. Structure and role of SIL in sheep genetic improvement in New Zealand. Proceedings of the New Zealand Society of Animal Production 63: 209-211.

Dechow CD, Norman HD, Zwald NR, Cowan CM, Meland OM. 2008. Relationship between individual herd-heritability estimates and sire misidentification rate. Journal of Dairy Science 91: 1640-1647. https://doi.org/10.3168/jds.2007-0534

Dekkers JCM, Vandervoort GE, Burnside EB. 1996. Optimal size of progeny groups for progeny-testing programs by artificial insemination firms. Journal of dairy science 79: 2056-2070. https://doi.org/10.3168/jds.S0022-0302(96)76579-7

Falconer DS, Mackay TFC. 1996. Introduction to Quantitative Genetics. 4th ed, Longman Scientific & Technical

Food and Agriculture Organisation of the United Nations. 2023. FAOSTAT: Food and agriculture data – Crops and Livestock Products. https://www.fao.org/faostat/en/#data/TCL [8 August 2025]

Fouilloux M-N, Clément V, Laloë D. 2008. Measuring connectedness among herds in mixed linear models: from theory to practice in large-sized genetic evaluations. Genetics Selection Evolution 40: 145-159. https://doi.org/10.1186/1297-9686-40-2-145

Geenty KG. 2000. Sheep industry vision and SIL. Proceedings of the New Zealand Society of Animal Production 60: 180-183.

Geldermann H, Pieper U, Weber WE. 1986. Effect of misidentification on the estimation of breeding value and heritability in cattle. Journal of Animal Science 63: 1759-1768. https://doi.org/doi.org/10.2527/jas1986.6361759x

Gilmour AR, Gogel BJ, Cullis BR, Welham SJ, Thompson R. 2021. ASReml User Guide Release 4.2 Functional Specification, VSN International Ltd

Gizaw S, Abebe A, Goshme S, Getachew T, Bisrat A, Abebe A, Besufikad S. 2022. Evaluating the accuracy of smallholder farmers' sire identification for introducing genetic evaluation in community-based sheep breeding programs. Livestock Science 255: 104804. https://doi.org/10.1016/j.livsci.2021.104804

Granado-Tajada I, Ugarte E. 2024. Impact of truncating historical data on prediction ability of dairy sheep selection candidates. animal 18: 101245. https://doi.org/10.1016/j.animal.2024.101245

Hanocq E, Boichard D, Foulley J-L. 1996. A simulation study of the effect of connectedness on genetic trend. Genetics Selection Evolution 28: 67-82. https://doi.org/10.1051/gse:19960104
Henderson CR. 1975. Best linear unbiased estimation and prediction under a selection model. Biometrics 31: 423-447. https://doi.org/10.2307/2529430
Henderson CR. 1985. Best linear unbiased prediction of nonadditive genetic merits in noninbred populations. Journal of Animal Science 60: 111-117. https://doi.org/10.2527/jas1985.601111x
Holmes JB, Auvray B, Newman S-A, Dodds KG, Lee MA. 2015. A comparison of genetic connectedness measures using data from the NZ sheep industry. Proceedings of the Association for the Advancement of Animal Breeding and Genetics 21: 469-472.
Israel C, Weller JI. 2000. Effect of misidentification on genetic gain and estimation of breeding value in dairy cattle populations. Journal of Dairy Science 83: 181-187. https://doi.org/10.3168/jds.S0022-0302(00)74869-7
James PJ. 2006. Genetic alternatives to mulesing and tail docking in sheep: a review. Australian Journal of Experimental Agriculture 46: 1-18. https://doi.org/10.1071/EA05100
Johansson I, Robertson A. 1952. Progeny testing in the Breeding of Farm Animals. Proceedings of the British Society of Animal Production 1953: 79-105. https://doi.org/10.1017/S0369852100001160
Johnson PL, Newman S-AN, McRae KM, van der Weerden TJ, Brown M, Scobie DR. 2021. A review of the current sheep industry in New Zealand and opportunities for change to meet future challenges. New Zealand Journal of Animal Science and Production 81: 1-15.
Jonker A, Hickey SM, Rowe SJ, Janssen PH, Shackell GH, Elmes S, Bain WE, Wing J, Greer GJ, Bryson B, MacLean S, Dodds KG, Pinares-Patiño CS, Young EA, Knowler K, Pickering NK, McEwan JC. 2018. Genetic parameters of methane emissions determined using portable accumulation chambers in lambs and ewes grazing pasture and genetic correlations with emissions determined in respiration chambers. Journal of Animal Science 96: 3031-3042. https://doi.org/10.1093/jas/sky187
Jopson NB, Linscott EM, Young MJ, Dodds KG, Clarke SM, Rowe SJ, Byrne TJ, Alder GF. 2015. Contract session: Application of genetic technologies to the New Zealand sheep industry. Proceedings of the New Zealand Society of Animal Production 75: 184-187.
Jordan TW. 2020. The cellular and molecular toxicity of sporidesmin. New Zealand Veterinary Journal 68: 203-213. https://doi.org/10.1080/00480169.2020.1729268
Kariuki CM, Brascamp EW, Komen H, Kahi AK, van Arendonk JAM. 2017. Economic evaluation of progeny-testing and genomic selection schemes for small-sized nucleus dairy cattle breeding programs in developing countries. Journal of Dairy Science 100: 2258-2268. https://doi.org/10.3168/jds.2016-11816
Kaseja K, Mucha S, Smith E, Yates J, Banos G, Conington J. 2023. Including genotypic information in genetic evaluations increases the accuracy of sheep breeding values. Journal of Animal Breeding and Genetics 140: 462-471. https://doi.org/10.1111/jbg.12771
Kennedy BW, Trus D. 1993. Considerations on genetic connectedness between management units under an animal model. Journal of Animal Science 71: 2341-2352. https://doi.org/10.2527/1993.7192341x
König S, Simianer H, Willam A. 2009. Economic evaluation of genomic breeding programs. Journal of Dairy Science 92: 382-391. https://doi.org/10.3168/jds.2008-1310
Kuehn LA, Lewis RM, Notter DR. 2007. Managing the risk of comparing estimated breeding values across flocks or herds through connectedness: a review and application. Genetics Selection Evolution 39: 225. https://doi.org/10.1186/1297-9686-39-3-225

Kuehn LA, Notter DR, Nieuwhof GJ, Lewis RM. 2008. Changes in connectedness over time in alternative sheep sire referencing schemes. Journal of Animal Science 86: 536-544. https://doi.org/10.2527/jas.2007-0256

Laloë D. 1993. Precision and information in linear models of genetic evaluation. Genetics Selection Evolution 25: 557-576. https://doi.org/10.1186/1297-9686-25-6-557

Lee M, Newman S-A, Dodds K, Nilforooshan M, Auvray B, McIntyre S, Campbell D. 2021. Genomic evaluations of sheep in New Zealand. New Zealand Journal of Animal Science and Production 81: 106-111.

Lewis RM, Crump RE, Simm G, Thompson R. 1999. Assessing connectedness in across-flock genetic evaluations. Proceedings of the British Society of Animal Science 1999: 121-121. https://doi.org/10.1017/S1752756200002763

Lush JL. 1935. Progeny Test and Individual Performance as Indicators of an Animal's Breeding Value. Journal of Dairy Science 18: 1-19. https://doi.org/10.3168/jds.S0022-0302(35)93109-5

Lush JL. 1940. Intra-sire correlations or regressions of offspring on dam as a method of estimating heritability of characteristics. Proceedings of the American Society of Animal Production 33: 293-301. https://doi.org/10.2527/jas1940.19401293x

Mathur PK. 2005. Importance of 'connectedness' between herds for effective across herd genetic evaluation. Journal of South China Agricultural University 26: 61-68.

McClay T. 2024. Government partnership to tackle $332m facial eczema problem. https://www.beehive.govt.nz/release/government-partnership-tackle-332m-facial-eczema-problem [8 June 2026]

McCoard S, Stevens D, Selbie D, Day L, Young W, Bekhit AE-D, Samuelsson L. 2023. Supporting the growth of the dairy sheep industry in New Zealand–industry update and review of a programme linking industry and science. New Zealand Journal of Agricultural Research 68: 183-217. https://doi.org/10.1080/00288233.2023.2272594

McLean JW. 1948. Progeny Testing in Sheep - The inheritance of birth weight, growth rate and cannon bone length. Proceedings of the New Zealand Society of Animal Production 8: 86-96.

McLean NJ, Jopson NB, Campbell AW, Knowler K, Behrent M, Cruickshank G, Logan CM, Muir PD, Wilson T, McEwan JC. 2006. An evaluation of sheep meat genetics in New Zealand: the central progeny test (CPT). Proceedings of the New Zealand Society of Animal Production 66: 368-372.

McRae KM, Cullen NG, Amyes NC, Johnson PL. 2016. Brief Communication: An update on genetic parameters for facial eczema tolerance in sheep. Proceedings of the New Zealand Society of Animal Production 76: 43-45.

McRae KM, Rowe SJ, Johnson PL, Baird HJ, Cullen NG, Bixley MJ, Plowman JE, Deb-Choudhury S, Brauning R, Amyes NC, Dodds KG, Newman S-AN, McEwan JC, Clarke SM. 2021. Genomic tools for the identification of loci associated with facial eczema in New Zealand sheep. Genes 12: 1560. https://doi.org/10.3390/genes12101560

Meuwissen T, Hayes B, Goddard M. 2013. Accelerating improvement of livestock with genomic selection. Annual Reveiw of Animal Biosciences 1: 221-237. https://doi.org/10.1146/annurev-animal-031412-103705

Ministry for Primary Industries. 2024a. Situation and Outlook for Primary Industries: December 2024. https://www.mpi.govt.nz/dmsdocument/66648-Situation-and-Outlook-for-Primary-Industries-SOPI-December-2024 [8 June 2026]

Ministry for Primary Industries. 2024b. The risk and impact of heat stress on sheep welfare on-farm Information Paper No. 2024/05. https://www.mpi.govt.nz/dmsdocument/69633-The-risk-and-impact-of-heat-stress-on-sheep-welfare-on-farm [8 June 2026]

Ministry for Primary Industries. 2026. Situation and Outlook for Primary Industries: June 2026. https://www.mpi.govt.nz/dmsdocument/72076-Situation-and-Outlook-for-Primary-Industries-SOPI-June-2026 [14 June 2026]
Ministry for the Environment. 2021. New Zealand's Greenhouse Gas Inventory 1990–2019 Publication number: ME 1559. https://environment.govt.nz/assets/Publications/New-Zealands-Greenhouse-Gas-Inventory-1990-2019-Volume-1-Chapters-1-15.pdf [8 June 2026]
Moloney AJ, Tozer PR, Morris ST, Kenyon PR. 2023. Bigger lambs or more lambs: The conundrum for New Zealand lamb producers. Livestock Science 270: 105204. https://doi.org/10.1016/j.livsci.2023.105204
Momen M, Morota G. 2018. Quantifying genomic connectedness and prediction accuracy from additive and non-additive gene actions. Genetics Selection Evolution 50: 45. https://doi.org/10.1186/s12711-018-0415-9
Morris CA. 1980. Some benefits and costs of genetic improvement in New Zealand's sheep and beef cattle industry: I. The annual selection response expected within closed flocks or herds. New Zealand Journal of Experimental Agriculture 8: 331-340. https://doi.org/10.1080/03015521.1980.10426283
Morris CA, Towers NR, Campbell AG, Meyer HH, Wesselink C, Wheeler M. 1989. Responses achieved in Romney flocks selected for or against susceptibility to facial eczema, 1975–87. New Zealand Journal of Agricultural Research 32: 379-388. https://doi.org/10.1080/00288233.1989.10421756
Morris CA, Towers NR, Hohenboken WD, Maqbool N, Smith BL, Phua SH. 2004. Inheritance of resistance to facial eczema: a review of research findings from sheep and cattle in New Zealand. New Zealand Veterinary Journal 52: 205-215. https://doi.org/10.1080/00480169.2004.36431
Morris CA, Towers NR, Wheeler M, Wesselink C. 1995. Selection for or against facial eczema susceptibility in Romney sheep, as monitored by serum concentrations of a liver enzyme. New Zealand Journal of Agricultural Research 38: 211-219. https://doi.org/10.1080/00288233.1995.9513121
Morris ST. 2009. Economics of sheep production. Small Ruminant Research 86: 59-62. https://doi.org/https://doi.org/10.1016/j.smallrumres.2009.09.019
Morris ST. 2013. Sheep and beef cattle production systems. Ecosystems services in New Zealand: 79-84.
Morris ST, Kenyon PR. 2014. Intensive sheep and beef production from pasture - A New Zealand perspective of concerns, opportunities and challenges. Meat Science 98: 330-335. https://doi.org/10.1016/j.meatsci.2014.06.011
Näsholm A, Eythorsdottir E. 2011. Characteristics and utilization of sheep pelts. Small Ruminant Research 101: 182-187. https://doi.org/10.1016/j.smallrumres.2011.09.038
Newman S-A. 2003. Making better use of the SIL genetic engine. Proceedings of the New Zealand Society of Animal Production 63: 194-196.
Newman S-A. 2017. Reporting by connectedness on across flock reports. https://www.sil.co.nz/files/1503864890806.pdf [9 June 2026]
Newman S-A, Dodds KG, Clarke JN, Garrick DJ, McEwan JC. 2000. The sheep improvement limited (SIL) genetic engine. Proceedings of the New Zealand Society of Animal Production 60: 195-197. https://doi.org/10.13140/2.1.3932.0965
NIWA. 2024. Aotearoa New Zealand Climate Summary: Summer 2023-24. https://niwa.co.nz/sites/default/files/2024-03/Climate_Summary_Summer_2024_Final.pdf [9 June 2026]

Notter DR. 1998. The U.S. National Sheep Improvement Program: across-flock genetic evaluations and new trait development. Journal of Animal Science 76: 2324-2330. https://doi.org/10.2527/1998.7692324x

Orme P. 2025. Update of land-use change from pastoral farming to large-scale forestry for 1/01/2024–31/03/2025. https://beeflambnz.com/knowledge-hub/PDF/land-use-change-pastoral-farming-large-scale-forestry.pdf [12 June 2026]

Peterson SW. 2017. Producing quality milk from sheep. In: Greyling J ed. Achieving sustainable production of sheep, Burleigh Dodds Science Publishing. Pp. 95-134

Phillips C, Johnson T, Tomasetto F, McRae K, van der Weerden T. 2023. Predicting facial eczema risks in a changing New Zealand climate. Journal of New Zealand Grasslands 85: 63–73. https://doi.org/10.33584/jnzg.2023.85.3650

Pickering NK. 2013. Genetics of flystrike, dagginess and associated traits in New Zealand dual-purpose sheep. PhD thesis, Massey University, Palmerston North, New Zealand.

Pickering NK, Auvray B, Dodds KG, McEwan JC. 2015. Genomic prediction and genome-wide association study for dagginess and host internal parasite resistance in New Zealand sheep. BMC Genomics 16: 958. https://doi.org/10.1186/s12864-015-2148-2

Pickering NK, Dodds KG, Blair HT, Hickson RE, Johnson PL, McEwan JC. 2012. Genetic parameters for production traits in New Zealand dual-purpose sheep, with an emphasis on dagginess. Journal of Animal Science 90: 1411-1420. https://doi.org/10.2527/jas.2011-4163

Posit team. 2025. RStudio: Integrated Development Environment for R. Boston, MA, Posit Software, PBC. http://www.posit.co/

R Core Team. 2025. R: A Language and Environment for Statistical Computing. Vienna, Austria, R Foundation for Statistical Computing. https://www.R-project.org/

Rae AL. 1948. Some aspects of progeny testing in sheep. Proceedings of the New Zealand Society of Animal Production 8: 97-108.

Rae AL. 1984. Development of selection programs for increasing lean meat production in sheep. Proceedings of 4th Conference of the Australian Association of Animal Breeding and Genetics: 3-7.

Rural News. 2020. B+LNZ Genetics launches ram breeding tool. Rural News. https://ruralnewsgroup.co.nz/rural-news/rural-general-news/b-lnz-genetics-launches-ram-breeding-tool [9 June 2026]

Rural News. 2025. NZ genetic engine version 6 launched. Rural News. https://www.ruralnewsgroup.co.nz/rural-news/rural-farm-health/nz-genetic-engine-version-6-launched#:~:text=Genetics%20is%20estimated%20to%20have,stud%20and%20commercial%20sheep%20breeders [9 June 2026]

SAS Institute Inc. 2023. SAS Software (Version 9.4_M6). Cary, NC, USA, SAS Institute Inc. https://www.sas.com

Schaeffer LR. 2006. Strategy for applying genome-wide selection in dairy cattle. Journal of Animal Breeding and Genetics 123: 218-223. https://doi.org/10.1111/j.1439-0388.2006.00595.x

Scholtens M, Lopez-Villalobos N, Lehnert K, Snell R, Garrick D, Blair HT. 2021. Advantage of including genomic information to predict breeding values for lactation yields of milk, fat, and protein or somatic cell score in a New Zealand dairy goat herd. Animals 11: 24. https://doi.org/10.3390/ani11010024

Scobie DR, O'Connell D, Morris CA, Hickey SM. 2007. A preliminary genetic analysis of breech and tail traits with the aim of improving the welfare of sheep. Australian Journal of Agricultural Research 58: 161-167. https://doi.org/10.1071/AR05444

Senneke SL, MacNeil MD, Van Vleck LD. 2004. Effects of sire misidentification on estimates of genetic parameters for birth and weaning weights in Hereford cattle. Journal of Animal Science 82: 2307-2312. https://doi.org/10.2527/2004.8282307x

Sheep Genetics. 2025. The Breeders Guide to Quality Assurance. https://www.sheepgenetics.org.au/globalassets/sheep-genetics/resources/breeders-guide/the-breeders-guide-to-quality-assurance.pdf [9 June 2026]

Sise JA, Byrne TJ, Jopson NB, Young MJ, Amer PR. 2012. A comparative analysis of genetic trends within the New Zealand sheep industry. Proceedings of the New Zealand Society of Animal Production 72: 159-162.

Škorput D, Špehar M, Luković Z. 2018. Connectedness between contemporary groups in Black Slavonian pig. Livestock Science 216: 6-8. https://doi.org/10.1016/j.livsci.2018.06.012

Špehar M, Ramljak J, Držaić V, Mioč B, Barać Z, Kasap A. 2025. Relationship of Genetic Connectedness Among Flocks with Effective Population Size in Dairy-Oriented Pag Sheep. Agriculture 15: 474. https://doi.org/10.3390/agriculture15050474

Tipa P. 2017. Breed for FE tolerance best answer. Rural News. https://www.ruralnewsgroup.co.nz/rural-news/rural-farm-health/breed-for-fe-tolerance-best-answer#:~:text=However%2C%20there%20is%20no%20way,to%20support%20that%20RamGuard%20tool. [9 June 2026]

van der Laan J. 2024. LaF: Fast Access to Large ASCII Files (R package version 0.8.6), CRAN. https://CRAN.R-project.org/package=LaF

Van Vleck LD. 1970. Misidentification in estimating the paternal sib correlation. Journal of Dairy Science 53: 1469-1474. https://doi.org/10.3168/jds.S0022-0302(70)86416-5

van Wettere WHEJ, Kind KL, Gatford KL, Swinbourne AM, Leu ST, Hayman PT, Kelly JM, Weaver AC, Kleemann DO, Walker SK. 2021. Review of the impact of heat stress on reproductive performance of sheep. Journal of Animal Science and Biotechnology 12: 26. https://doi.org/10.1186/s40104-020-00537-z

Vibart RE, Rennie G, Hutchinson K, Burtt A, Chrystal J, Dynes R. 2022. Greenhouse gas emissions from New Zealand sheep and beef farms. Journal of New Zealand Grasslands 83: 252-232. https://doi.org/10.33584/jnzg.2021.83.3501

Winkelman AM. 2013. Effect of daughter misidentification on dairy sire evaluation. Proceedings of the Association for the Advancement of Animal Breeding and Genetics 20: 25-28.

Wuliji T, Dodds KG, Andrews RN, Turner PR. 2011. Selection response to fleece weight, wool characteristics, and heritability estimates in yearling Romney sheep. Livestock Science 135: 26-31. https://doi.org/10.1016/j.livsci.2010.06.003

Yang H, Su G. 2016. Impact of phenotypic information of previous generations and depth of pedigree on estimates of genetic parameters and breeding values. Livestock Science 187: 61-67. https://doi.org/10.1016/j.livsci.2016.03.001

Yardibi F, Chen C, Fırat MZ, Karacaören B, Süzen E. 2023. The trend of breeding value research in animal science: bibliometric analysis. Archives Animal Breeding 66: 163-181. https://doi.org/10.5194/aab-66-163-2023

Young MJ, Amer PR. 2009. Rates of genetic gain in New Zealand sheep. Proceedings of the Association for the Advancement of Animal Breeding and Genetics 18: 422-425.

Young MJ, Newman S-AN. 2009. SIL-ACE - Increasing access to genetic information for sheep farmers. Proceedings of the New Zealand Society of Animal Production 69: 153-154.

Yu H, Morota G. 2019. GCA: Genetic Connectedness Analysis (R package version 0.1.0), GitHub repository: QGresources. https://github.com/QGresources/GCA

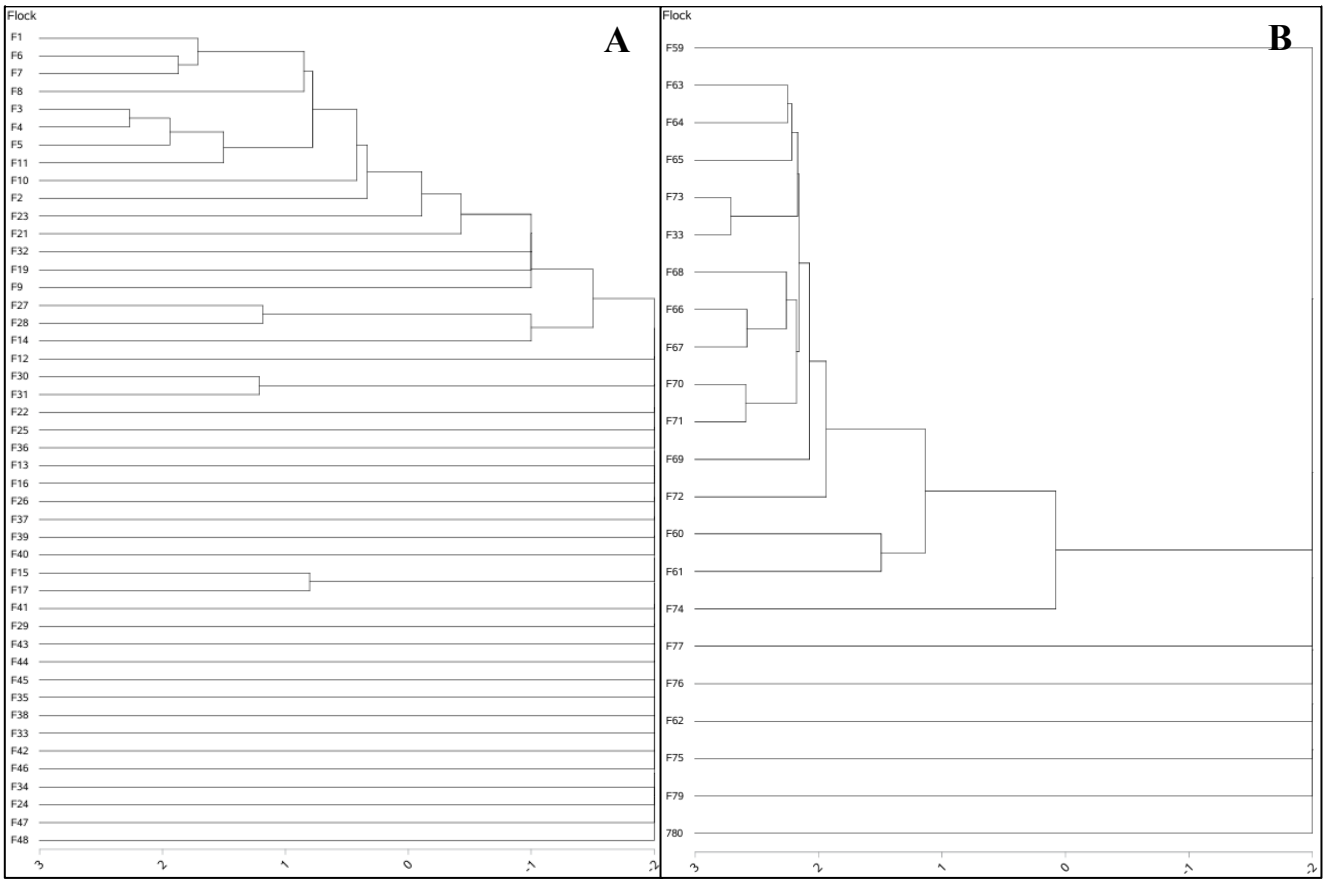
Yu H, Morota G. 2021. GCA: an R package for genetic connectedness analysis using pedigree and genomic data. BMC Genomics 22: 119. https://doi.org/https://doi.org/10.1186/s12864-021-07414-7

Yu H, Spangler ML, Lewis RM, Morota G. 2017. Genomic relatedness strengthens genetic connectedness across management units. G3: Genes, Genomes, Genetics 7: 3543-3556. https://doi.org/10.1534/g3.117.300151

Yu H, Spangler ML, Lewis RM, Morota G. 2018. Do stronger measures of genomic connectedness enhance prediction accuracies across management units? Journal of Animal Science 96: 4490-4500. https://doi.org/10.1093/jas/sky316

Appendix 1: Dendrograms of genetic connectedness among flocks

Dendrograms showing genetic connectedness among flocks for (A) facial eczema and (B) growth using the nProve method with a 1-year evaluation window (removal of two years of data).



Appendix 2: Paper for World Congress on Genetics Applied to Livestock Production

Investigating the Impact of Genetic Linkage Between Sheep Flocks on the Accuracy and Outcomes of Genetic Evaluations

J.L. Smith[1], N.W. Sneddon[1*], N.B. Jopson[2], N.P. Martín[1], N. López-Villalobos[1], and D. van der Linden[3]

[1]Massey University, Palmerston North, New Zealand; [2]AbacusBio Limited, Dunedin, New Zealand; [3]Beef + Lamb New Zealand Genetics, Dunedin, New Zealand; [*]Presenting and corresponding author: N.W.Sneddon@massey.ac.nz

Introduction

Genetic gain in the New Zealand sheep industry relies on the accurate estimation of breeding values (BVs) for different traits and the ability to compare them reliably across flocks. The effectiveness of across-flock BV comparison hinges on genetic connectedness, defined as the degree of genetic linkage between animals in different flocks (Kuehn *et al*., 2007). Sire sharing is the primary mechanism used to establish this linkage, enabling animals raised in different environments to be evaluated on a common genetic scale (Brown *et al.*, 2007). Despite the importance of linkages, there are few evidence-based guidelines specifying: the number of linkages required, the effectiveness of historical links, and thresholds to support accurate comparisons of BVs across flocks. In New Zealand, the nProve evaluation applies a relatively conservative connectedness criterion, assessing linkage based on the number of shared progeny from link sires within a rolling three-year window (Newman, 2003). Although this approach ensures robust benchmarking, it may restrict the number of connected flocks. This could be of advantage for traits that are difficult to measure. For example, facial eczema tolerance (FET) has limited phenotypic recording despite being a serious health and production issue in New Zealand. Facial eczema is a fungal disease caused by ingestion of the toxin sporidesmin (Morris

et al., 2004). Phenotyping for FET is expensive and difficult to measure due to the controlled nature of the challenge and risk to animals, limiting recording to a subset of animals and number of participating flocks (B+LNZ Genetics, 2025).

This study evaluates how changes to inclusion criteria and data composition influences connectedness statistics. By identifying the thresholds for connectedness that support reliable across-flock comparisons, this research aims to inform the progeny test design and evaluate whether current connectedness criteria could be relaxed while still maintaining valid genetic benchmarking.

Materials & Methods

Data for this study were obtained from the nProve database, which included FET, pedigree, and identification records for all animals in the system. Data processing and filtering were carried out in SAS 9.4 using the standard nProve preprocessing programme to generate pedigree, unique identification, and trait files for connectedness assessment.

Connectedness was first evaluated using the current nProve methodology, which computes across-flock linkage based on the number of progeny from link sires shared between each pair of flocks within a three-year window appropriate to each trait. This method constructed a flock-by-flock connectedness matrix in which linkage strength reflects the total number of progeny from link sires used across both flocks. This matrix was transformed into a distance matrix and visualised using hierarchical clustering, producing a dendrogram that identifies groups of connected and unconnected flocks for this purpose, any flocks with connectedness value below 0 are considered unconnected. The second approach used the Genetic Connectedness Analysis package in R (Yu and Morota, 2021). A pruned pedigree was generated for animals contributing FET data in the most recent three years, and a corresponding numerator relationship matrix was constructed.

Three different analyses were conducted to assess how changes to inclusion criteria and data composition influence observed connectedness. First, the assessed data window was varied (from 1 year to 6 years). Second, highly connected flocks were removed from analysis, one by one, to evaluate the stability of across-flock linkage and the dependence of the system on key connecting flocks (flocks F1, F6 and F18, with connectedness values >2.2, >2.4 and >1.6,

respectively). Third, the three most prolific sires were identified and removed one by one to assess how the loss of major link sires altered flock connectedness. For each modified scenario, connectedness matrices and dendrograms were generated using the nProve procedure and compared with the baseline connectedness structure.

The Genetic Connectedness Analysis (GCA) package in R was also used to quantify across-flock connectedness under the different scenarios tested using Best Linear Unbiased Prediction (BLUP) equations and relationship matrices (Yu and Morota, 2021). To assess the impact of the number of years of data on the coefficient of determination (CD) and prediction error variance of differences (PEVD) for the accuracy of genetic evaluations across flocks, the GCA connectedness analysis was re-run in RStudio using progressively larger subsets of the FET dataset. The model fitted included the fixed effects of flock and sex, σ_a^2= 0.3321 and σ_e^2= 0.4779. The range of birth years included was incrementally expanded from two years of phenotypic data (earliest birth year included = 2023) up to ten years (earliest birth year included = 2015). Across-flock evaluation accuracy was quantified using PEVD and CD. Pairwise PEVD values between flocks were first calculated using the GCA framework and then summarised as the overall average PEVD ($PEVD_{GrpAve}$) across all flock comparisons. The corresponding overall average CD (CD_{GrpAve}) was also calculated, as the variance of true differences between flocks computed from σ_a^2 and the mean of pairwise relationships within and between flocks derived from the pedigree matrix. For each scenario, overall $PEVD_{GrpAve}$, CD_{GrpAve}, the number of phenotypic records, and the number of animals retained in the pruned pedigree were recorded.

Results and Discussion

Genetic connectedness among flocks, as assessed using the nProve method, is shown in Figure 1. The number and percentage of connected flocks are presented in Table 1. Clusters with shorter branch lengths indicate stronger genetic linkage between flocks. The scale ranges from 3 to -2, with positive values meeting connectedness criteria, and negative values do not. Based on this threshold, 24 of 49 flocks were classified as connected for the base scenario, forming a cluster above flock F25, with the remaining 25 flocks unconnected.

Table 1. Number and percentage of flocks identified as connected (flocks that met the connectedness criteria, connectedness >0) or unconnected, as inferred from the dendrograms generated under each scenario in the nProve connectedness analysis for facial eczema (FET).

Scenario	Total Flocks (n)	Connected Flocks (n)	Unconnected Flocks (n)	% Connected
Base (3y window)	49	24	25	49.0
1y window	46	10	36	21.7
2y window	48	12	36	25.0
4y window	51	34	17	66.7
5y window	55	35	20	63.6
6y window	58	41	17	70.7
Base - Flock 1	48	20	28	41.7
Base - Flock 6	48	22	26	45.8
Base - Flock 18	48	23	25	47.9
Base - Prolific Sires	49	24	25	49.0

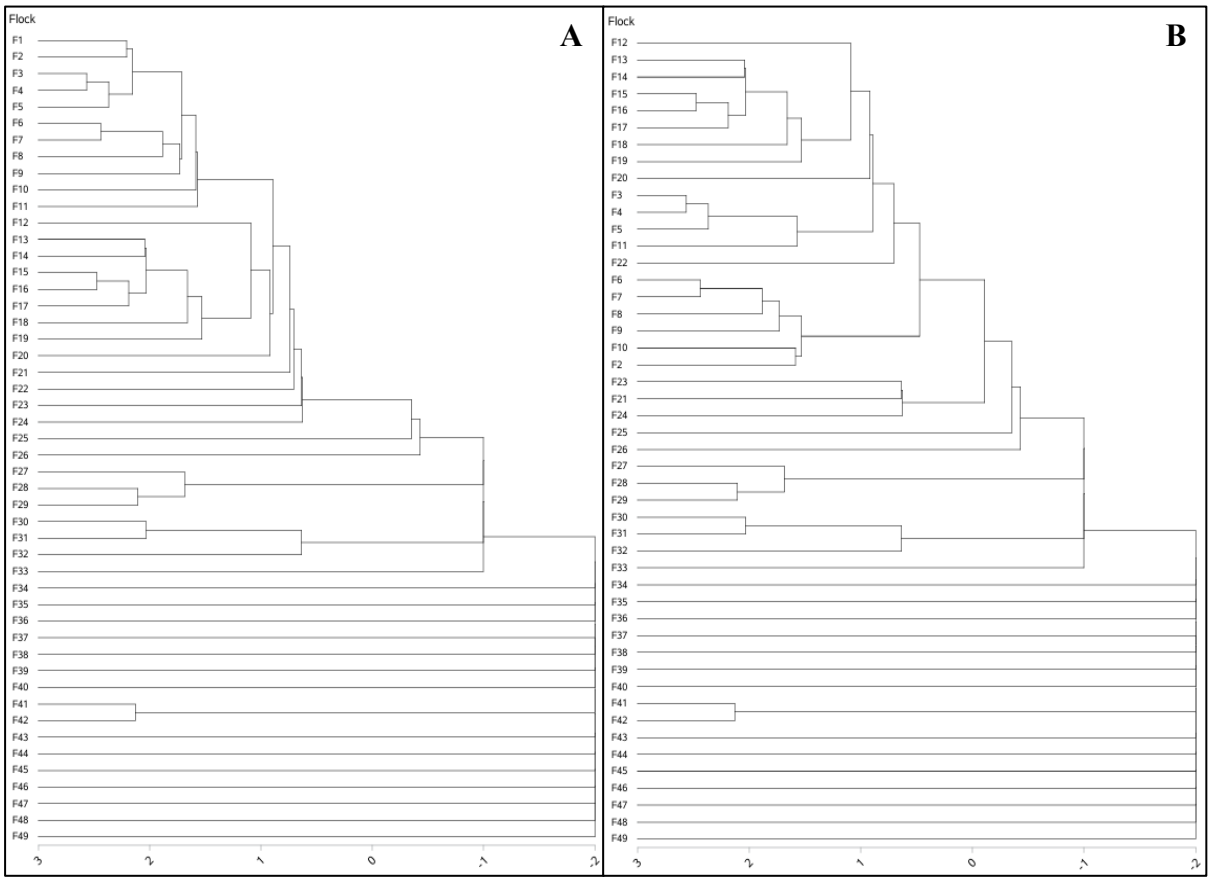


Figure 1. Dendrograms illustrating genetic connectedness among flocks for facial eczema. Clusters with shorter branch lengths indicate stronger genetic linkage between flocks using the nProve method (A) Base scenario (3-year window, no flocks or sires removed); (B) Flock F1 removed but same inclusion criteria as the base scenario.

Effect of data-window on connectedness. The current recommendation is to restrict across-flock linkage assessments to a 3 to 5 year window (Newman, 2003); however, an expansion of this range has not previosusly been evaluated in relation to connectedness or across-flock accuracy. The number of connected flocks increased as the data window was extended (Table 1). Reducing the window from the current 3-year standard data window to a 2- and 1-year window resulted in a decline in overall connectedness. This reduction was reflected in the proportion of connected flocks, which fell from 49.0% to 25.0% and 21.7% for 2- and 1-year windows, respectively. In contrast, extending the window to 4-, 5-, and 6-years increased connectedness, with 70.7% of flocks connected when a 6-year window was used.

Flock and sire removal. As this has not been evaluated previously, sequential removal of highly connected flocks altered the dendrogram structure and changed the connectedness status of multiple flocks, with some immediately losing any connection to other flocks within the evaluation as shown in Figure 1. Removal of flock F1 reduced the proportion of connected flocks from 49.0% to 41.7%, while removal of flock F6 reduced connectedness from 49.0% to 45.8% (Table 1). In contrast, removing flock F18 had minimal influence on dendrogram structure or overall connectivity. The removal of the three most prolific sires in the FET dataset also produced negligible changes, with no flock shifting in or out of the connected status, indicating that no individual sire maintained across-flock connectivity in this evaluation.

Effect of data-window on PEVD and CD. Across-flock evaluation accuracy for FET measured using PEVD and CD showed little change as the data window increased (Figure 2). Using a 3-year dataset consistent with the data window currently applied in nProve for assessing connectedness (earliest birth year = 2022), PEVD was 0.564 and CD was 0.286. Including older data increased the number of phenotypic records from 4,931 to 25,890, and the number of animals in the pedigree from 12,076 to 60,081 with 10 years of data. CD increased smoothly by 0.112 over the 10-year window (to 0.371), indicating improved across-flock accuracy, whereas PEVD fluctuated within a narrow range (0.504 – 0.587) across the range of

data windows evaluated. Both metrics showed evidence of plateauing after approximately 5 to 6 years of data (earliest birth year = 2019-2020).

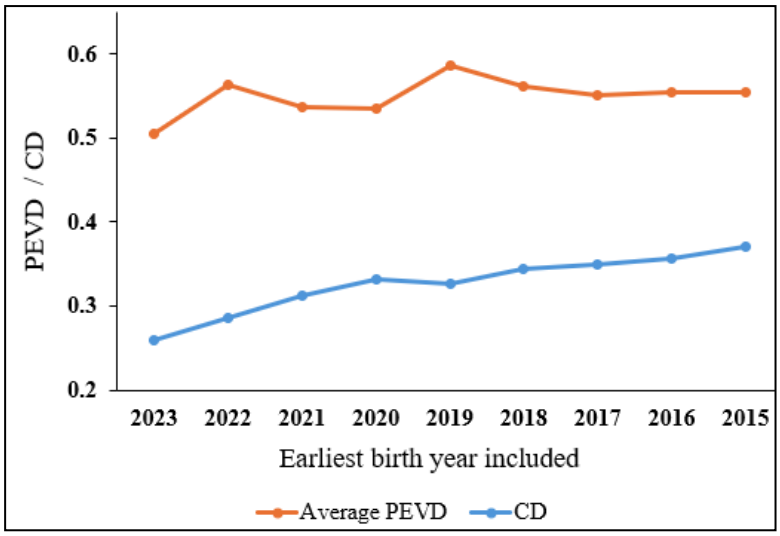


Figure 2. Effect of data-window length on across-flock evaluation accuracy for facial eczema (FET), measured using the average prediction error variance of differences ($PEVD_{GrpAve}$) and the coefficient of determination (CD_{GrpAve}). The x-axis shows the earliest birth year included in the evaluation.

Conclusions

Extending the assessed data window beyond the current 3-year standard increased the number of connected flocks by capturing additional historic genetic links. However, it was unclear whether the inclusion of these historical connections would improve the validity of across-flock BV comparisons. Analysis of evaluation accuracy using the GCA method demonstrated that extending the data window did not improve PEVD; however, CD increased, indicating improved reliability of across-flock comparisons. These findings suggest that for FET, extending the nProve evaluation to a 5-year window could enable greater flock participation and provide a practical balance between improving CD, maintaining PEVD, and keeping the evaluation manageable in terms of computational time and effort. Whether these findings extend to traits with higher levels of phenotypic recording remains to be evaluated.

References

B+LNZ Genetics. (2025). Facial Eczema Tolerance: Technical Note. https://www.sil.nz/files/174674064762.pdf

Brown D., Huisman A., Swan A., Graser H., Woolaston R., Ball, A., Atkins, K., and Banks, R. (2007) Proceedings of the Association for the Advancement of Animal Breeding and Genetics, 17: 187-194. https://www.aaabg.org/livestocklibrary/2007/brown187.pdf

Kuehn, L.A., Lewis, R.M., and Notter, D.R. (2007) Genetics Selection Evolution 39: 225. doi.org/10.1186/1297-9686-39-3-225

Morris, C., Towers, N., Hohenboken, W., Maqbool, N., Smith, B., and Phua, S. (2004) New Zealand Veterinary Journal 52: 205-215. doi.org/10.1080/00480169.2004.36431

Newman, S.-A. (2003). Proceedings of the New Zealand Society of Animal Production, 63: 194-196.

Yu, H., and Morota, G. (2021) BMC Genomics, 22: 119. doi.org/10.1186/s12864-021-07414-7