Genetic Analysis of Milk Production Traits in Relation to Heat Tolerance in Australian Dairy Cows

by

Evans Kiptoo Cheruiyot

(MSc in Animal Genetics & Breeding)

Thesis

submitted in total fulfilment of the requirements for the degree of

Doctor of Philosophy

School of Applied Systems Biology

College of Science, Health and Engineering

La Trobe University

Victoria, Australia

November 2021

Contents

Abbreviations

AD	Additive genetic variance
AI	Artificial insemination
AUD	Australian dollar
BL	Broken line model
BOM	Bureau of Meteorology
BPI	Balanced performance index
CCI	Comprehensive climate index
CSIRO	Commonwealth Scientific & Industrial Research
DHLI	Dairy heat load index
DIM	Days in milk
DTD	Daughter trait deviations
ES	Environmental sensitivity
FAETH	Function-And-Evolutionary Trait Heritability
FAO	Food and agricultural organisation
FDR	False discovery rate
FYint	Fat yield intercept
$G \times E$	Genotype by environment interactions
GBLUP	Genomic best linear unbiased prediction
GEBV	Genomic estimated breeding value
GRM	Genomic relationship matrix
GS	Genomic selection
GWAS	Genome-wide association study
h^2	Heritability
HD	High density
HLI	Heat load index
HT-ABV	Australian heat tolerance breeding value
HTD	Herd-test-day
HTFYslope	Heat tolerance fat yield slope
HTMYslope	Heat tolerance milk yield slope
HTPYslope	Heat tolerance protein yield slope
HWI	Health weighted index
IPCC	Intergovernmental Panel on Climate Change

KEGG	Kyoto encyclopedia of genes and genomes
LCT	Lower critical temperature
LD	Linkage disequilibrium
MACE	Multiple across-country evaluation
MAF	Minor allele frequency
MIR	Mid-infrared
MT	Multi-trait analysis
MYint	Milk yield intercept
NEBAL	Negative energy balance
NEFA	Non-esterified fatty acid
PC	Principal component
PE	Permanent environmental variance
PYint	Protein yield intercept
QTL	Quantitative trait loci
RH	Relative humidity
RM	Reaction norm model
ROS	Reactive oxygen species
RR	Respiratory rate
RT	Rectal temperature
SCC	Somatic cell count
SD	Standard deviation
SE	Standard error
SNPs	Single nucleotide polymorphisms
TCA	Tricarboxylic acid cycle
TD	Trait deviation
THI	Temperature-humidity index
TMR	Total mixed ration
TNZ	Thermoneutral zone
UCT	Upper critical temperature
UMD	University of Maryland
USD	US dollar
WGS	Whole-genome sequencing
WS	Wind speed

Abstract

Heat stress from rising global temperatures is an issue of growing concern worldwide, posing serious harmful consequences to humans and compromising production and reproduction in livestock, resulting in multi-billion-dollar economic losses. High-producing dairy cattle are particularly prone to thermal stress because they generate heat from the fermentation of additional dry matter during lactation; therefore, they are good model species for studying heat tolerance. While animal responses to increasing temperatures are known to vary substantially within and between species, the genetic control of resilience to heat is still not well characterised in animals. Therefore, the purpose of this thesis was to investigate if Australian dairy cattle exhibit significant genetic variation in response to changes in temperature and humidity – that is genotype $\times$ environment interaction (G $\times$ E) and dissect the genetic basis underlying thermal stress with a view to improve the genomic prediction of this economically important trait.

A large dataset of test-day milk production records (test-day milk, fat, and protein yields) of over 40,000 animals (Holsteins, Jersey, and Holstein-Jersey crossbreds) with imputed-whole genome sequences (~15 million SNPs) was used in the studies described. Since milk production in dairy cows often decreases under thermal stress, heat-tolerance phenotypes were defined as the rate of milk yield decline with an increase in the temperature-humidity index (THI), where THI is a value that combines temperature and humidity values obtained from weather stations near dairy herds in Australia. Statistical analyses using multi-trait and reaction norm models identified some re-ranking among the bulls for heat tolerance milk production traits (slopes traits) at the upper extreme of THI trajectory (warmer environments). However, the extent of re-ranking was not considered to be large enough (i.e., genetic correlations were higher than 0.80) to justify forming separate breeding programmes for colder and warmer environments. However, the genetic differences in environmental response to heat among Australian dairy animals allow farmers to select animals that are best suited to their production environment. A large genome-wide association (GWAS) (using ~ 30,000 Holstein cows) revealed promising candidate variants and genes underlying the response to thermal stress in dairy cattle, including 61 potential functional variants at genomic sites highly conserved across 100 vertebrate species. Intriguingly, the candidate genes underlying variation in thermotolerance in animals are enriched in the nervous system (*ITPR1, ITPR2,* and *GRIA4*) and neuroactive ligand-receptor interaction functions (*NPFFR2, CALCR,* and *GHR*). A suite of top candidate sequence variants from GWAS (ranging between ~2,000 and 9,000) were pre-

selected to confirm their relevance in thermotolerance and then tested in genomic evaluations of this trait. It was found that the accuracy of genomic prediction for heat tolerance improved when the pre-selected set of sequence variants were added to the industry-standard 50k SNP panel and analysed using non-linear Bayesian models (BayesR and BayesRC methods) with an average increase in accuracy of around 5% units and values ranging from 0.1% to 10% units depending on the prediction scenario. However, prediction accuracy decreased in some cases (particularly in Jerseys) when a set of sequence variants discovered from Holsteins were analysed alongside 50k SNPs compared to using only the 50k array.

Overall, the results of this thesis provide new discoveries about the genetic basis and biology of thermal stress. The mechanisms by which animals use to minimise the effects of heat are useful to understand given the effect of increasing global temperatures on performance, reproduction, and ultimately the welfare of farm animals. The genetic variants discovered in this thesis provide functional information and can be imputed into the industry SNP panels, or used to develop custom SNP arrays or similar sequence-based approaches that can help to drive the genetic improvement for heat tolerance.

Statement of authorship

Except where reference is made in the text of the thesis, this thesis contains no material published elsewhere or extracted in whole or in part from a thesis accepted for the award of any other degree or diploma.

No other person's work has been used without due acknowledgment in the main text of the thesis.

This thesis has not been submitted for the award of any degree or diploma in any other tertiary institution.

Evans Kiptoo Cheruiyot

La Trobe University, Australia
November 2021

Acknowledgements

It is with great joy and humility to reflect and express my profound gratitude to all the people and institutions who contributed to making this thesis a reality. To my supervisors: Prof Jennie Pryce, Prof Ben Cocks, and Dr Haile Mekonnen, I say a big thank you for your massive support, guidance, and mentorship over the last three and half years, which allowed me to flourish in my academic experience. I will forever be thankful for the numerous catch-up meetings we had both physically and virtually (due to COVID-19 pandemic lockdowns), the insightful thoughts which shaped my thesis, and for being more than a friend.

Special thanks to Dr Thuy Nguyen for being my supervisor during the first year of my PhD – your guidance, support, and encouragement gave me positive thrust to start and finish my PhD within the timeframe. Thank you so much Dr Iona Macleod for being part of my supervisory team during the second year of my PhD until I finished. The fact that you agreed to spare your busy time to help me complete Chapters 4 and 5 of this thesis is something I will always be thankful for. To Prof Raphael Mrode, accept my profound thanks for your support and mentorship from the start of my MSc and into the PhD and for hosting me during my brief study at ILRI (Kenya). Thank you, Dr Ruidong Xiang, Dr Mary Abdelsayed, Dr Bolormaa Sunduimijid, and Dr Majid Khansefid, your contributions to the success of this thesis in terms of analysis and guidance is highly appreciated.

I am very grateful to La Trobe University Full-Fee PhD Scholarship. I am also thankful to DairyBio [funded by Dairy Australia (Melbourne, Australia), the Gardiner Foundation (Melbourne, Australia), and Agriculture Victoria (Melbourne, Australia)] for the financial support to my PhD project. Thank you, Dr Janet Wheeler (post-graduate coordinator La Trobe University, school of applied Systems Biology) and Kendra Whiteman (student coordinator Agriculture Victoria Research), for facilitating my PhD study.

To my wonderful PhD colleagues at Computational Biology Group (AgriBio): Jigme Dorji, Beth Scott, Caeli Richardson, Ee Cheng, Boris Sepulveda, Michele Axford, Cheng Li, and Yvonne Ogaji, you guys are amazing for making the going bearable, particularly during the tough moments of COVID-19 lockdowns. I would also like to thank other friends for their support and encouragement during my studies: Dr David Thuo, Huan Zhao, Nga Nguyen, and many more, including those whom we played soccer together (please don't feel left out).

Finally, my hearty appreciation goes to my family – my biggest fans – for their steadfast support throughout my academic journey. I will end by dedicating this thesis to my grandfather, who sadly passed away the same day I completed the first draft of this thesis.

Preface

This thesis consists of 6 chapters. Chapter 1 introduces the research topic (heat tolerance), discusses its relevance, highlights the specific objectives, and provides a brief description of each thesis chapter. Chapter 2 covers published literature on the research topic, while Chapter 3 to 5 presents the original PhD research work. These research chapters (i.e., Chapter 3 to 5) have been published in peer-reviewed journals. The formatting style of the research chapters follows the respective journal specifications. For each of the research chapters, I (the student) did the main role of designing the study, cleaning, and analysing the phenotype and genotype data, interpreting the results, and writing the first draft of each manuscript. My supervisors and co-authors provided advice on the design, analysis, interpretation of the results, and contributed to the writing of manuscripts. Chapter 6 is my own thinking where I discuss the key findings from the three research chapters and how they are thematically linked in a broader context relevant to the dairy industry. Also, in Chapter 6, I discuss the limitations of this thesis work and suggest future research directions. Finally, the three peer-reviewed conference materials, which resulted from this thesis, are provided in Appendix 1 to 3. Formatting styles of the conference materials follow the guidelines from the respective conference proceedings.

Chapter 1: General introduction

1.1. **Background**

Heat stress from rising global temperatures is an issue of growing concern worldwide, affecting humans, wildlife, and livestock. Recent reports (e.g., Li et al., 2020a) show that many people are now affected by heat, and this has risen by more than two-fold when compared to the pre-industrial climates (i.e., 95 versus 275 million people), with future predictions showing that over 1 billion people will experience an even greater impact of heat within the next 50 years (Xu et al., 2020). In the dairy industry, potential heat stress calculated from temperature and humidity data above specific comfort zones has been increasing worldwide (Silanikove and Koluman, 2015, Polsky and von Keyserlingk, 2017) making this a growing issue that could compromise production (reduced growth, milk, etc.) and reproduction, and sometimes death in extreme cases. Dairy cattle are especially prone to environmental heat stress because they generate high metabolic heat loads from the fermentation of additional dry matter during lactation. A substantial annual economic loss due to heat stress alone in the dairy industry was estimated to be up to USD 897 million in the USA (St-Pierre et al., 2003), up to AUD 300 million in Australia's 1.4 million cow population (DairyBio, 2018), and up to £33 million in the South-West region of the UK (Fodor et al., 2018) – home to around 0.5 million dairy cattle (Defra, 2021).

While heat stress is an issue to many dairy industries all around the world, Australia faces two main unique challenges: 1) Australian dairy system is highly heterogeneous, with diverse climatic conditions ranging from the warmer northeast regions to the cooler southeast regions of the country. As such, dairy cows experience varying magnitudes of heat stress (both temporal and seasonal), with some dairying regions, including Queensland, New South Wales, and Western and South Australia, under relatively high heat loads extending up to half of the year, while temperatures are moderate in Victoria and lowest in Tasmania, the southernmost part of the country (Nguyen et al., 2016); and 2) Australian dairy farms mostly graze cows outdoors on pasture, with limited management measures to combat heat stress, in contrast to many European countries, the USA, and Canada, where cows are generally housed indoors and fed total mixed rations (TMR). Under these conditions, genotype $\times$ environment interaction (G $\times$ E) or the re-ranking of genotypes due to heat stress across different production environments may be an issue for dairy farmers in Australia.

G × E due to heat stress could exacerbate over the coming decades because more climate perturbations characterised by increased temperatures and extreme hot days are expected to continue across Australia (BOM, 2020; Figure 2-2). Also, BOM and CSIRO (2015) projects that even more temperate dairying regions, such as Gippsland, a major contributor of national milk production, will see a rise in the annual average temperature of up to 2.6 °C by 2070, meaning that dairy animals in those regions will experience more severe hot days and longer warm spells. Nguyen et al. (2017) found that the ability of dairy cattle in Australia to tolerate heat has been declining over the past years at a rate of 0.3%/yr, in part, due to continued selection for milk production traits. This mirrors more recent data for heat tolerance from DataGene – the organisation responsible for routine genetic evaluation in Australia (DataGene, 2021), with a steeper genetic decline of heat tolerance following the introduction of genomic selection in 2010 (Figure 1-1). Regardless, the dairy industry will still have to deal with the double challenge of increasing production, even more, to feed a growing population while coping with the effects of heat stress and ever-changing production environments.
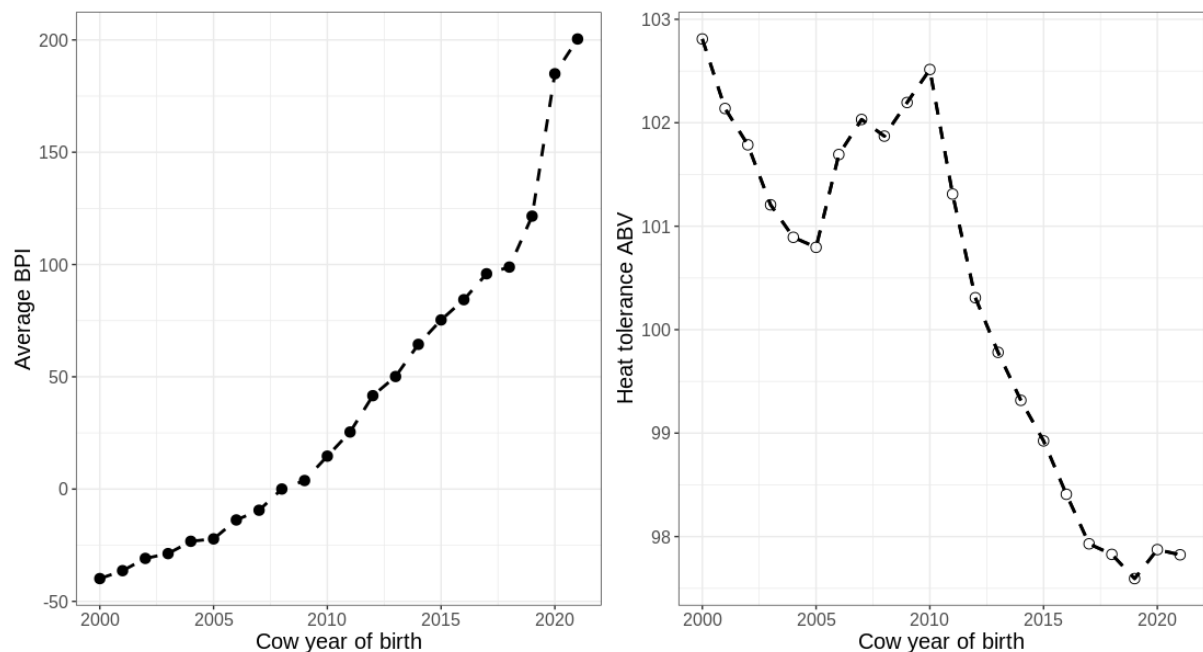
Figure 1-1 Genetic trend of Australian average balanced performance index (BPI) and heat tolerance breeding values (ABVg) over the past years for Holstein cows (Date source; DataGene, https://datagene.com.au/; accessed September, 2021).

With predominantly pasture-based dairy systems in Australia, the implementation of short-term management solutions to combat heat stress (e.g., shades, sprinklers, and fans) can be challenging and cost-prohibitive for most dairy farms. Therefore, as global temperatures

continue to rise, long-term solutions such as genetic selection of animals that can maintain performance under heat stress is advantageous since it is permanent and cumulative and can be used alongside management strategies. To date, most research studies to combat heat stress have focused on the aforementioned non-genetic interventions, including shade, sprinklers, and feed modifications (e.g., Cool Cows initiative; https://coolcows.dairyaustralia.com.au/ and Feeding Cool Cows project; https://dairyfeedbase.com.au/feeding-cool-cows/, both of which are being implemented in Australia). It is possible to select for heat tolerance since the ability to tolerate heat varies between individuals, with heritabilities of this complex trait being large enough for genetic selection to be effective (Ravagnolo and Misztal, 2000, Nguyen et al., 2016). However, there is still a big gap in our understanding on the genetic aspects that confer thermotolerance to animals and is now an active area of research in many countries worldwide, of which Australia is at the forefront of genomic evaluations following the development and release of the first breeding values for heat tolerance in 2017 (Nguyen et al., 2016, Nguyen et al., 2017). The overall goal of this thesis work is to expand our knowledge on the genetic features that contribute to thermotolerance, considering the harmful effects of rising temperatures on humans and animal production and welfare.

1.2. **Objectives of the study**

This thesis focuses on three broad aspects of heat tolerance with the following specific objectives aligning to the research chapters:

1) To investigate the magnitude and relevance of G × E due to heat stress for milk production traits (milk, fat, protein yield) in the Australian dairy cattle.
2) Perform a genome-wide association study (GWAS) to identify genetic markers and elucidate biological mechanisms underlying thermotolerance in animals.
3) Test the added benefits of using pre-selected whole-genome sequence variants from GWAS for the genomic prediction of heat tolerance.

1.3. **Thesis outline**

Heat tolerance is no doubt an area of great relevance worldwide, considering the harmful effects of warming climates on different species, including humans, livestock, and plants. This thesis research was motivated by the fact that there is still a huge gap in our knowledge regarding the genetic features that confer thermotolerance in animals. Chapter 1 of this thesis introduces the research topic and its relevance within the global and local (i.e., Australian industry) context, focusing on the dairy industry and finally concludes by

highlighting specific objectives of the study. Chapter 2 covers the review of published literature on heat tolerance, including the non-genetic and genetic aspects for combating thermal stress in livestock. In Chapter 3, the magnitude and relevance of G × E due to heat stress for production traits (milk, fat, and proteins) in the Australian dairy cattle is investigated. This was crucial aimed at addressing the growing concern about the possibility of significant re-ranking of dairy sires in terms of their performance in colder and hotter environments (in part because of climate change and continued selection of high-yielding dairy animals). Motivated by the fact that a substantial genetic variation for heat tolerance exists in the Australian dairy cattle (discussed in Chapter 3), GWAS analyses were conducted (in Chapter 4) to pinpoint specific genetic markers and biological features controlling this trait. Compared to earlier GWAS studies on heat tolerance, this thesis work benefited from a large sample size and high-resolution genotypes (~15 million SNPs), meaning that the candidate causal mutations for heat tolerance were mapped with reasonably high confidence. It was then a logical next step to conduct genomic prediction analyses, as described in Chapter 5, to validate and demonstrate the added benefit of pre-selected sequence variants from GWAS for improving prediction accuracy of heat tolerance, which is critical for driving the genetic improvement for this trait. Finally, in Chapter 6, I provide a general perspective of the findings with respect to breeding for thermotolerance in cattle and how the research chapters are thematically linked while highlighting some limitations of the study and concluded by suggesting future research directions. An overview of this thesis is shown in Figure 1-2.
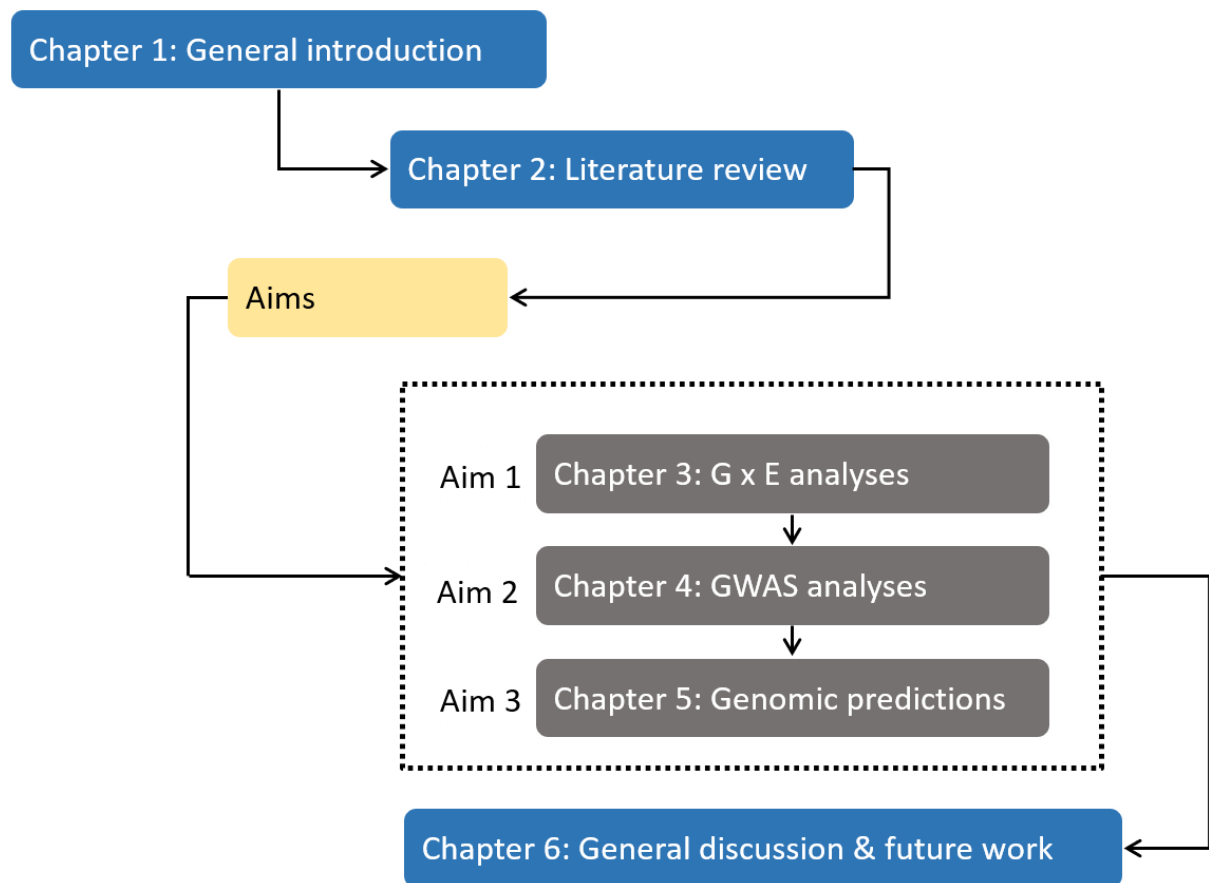
Figure 1-2 Thesis outline with published research chapters shown in grey boxes and other chapters in blue boxes.

Chapter 2: Literature review

2.1. **Introduction**

Over 6 billion people around the world (i.e., > 80% of the global population) consume nutrient-rich milk and milk products on regular basis, with the demand expected to increase in the coming decades in line with the growing per capita income and human population – 7.6 billion people now versus 9.8 billion people in 2050 (FAO, 2019). The need to increase milk production to satisfy the demand has largely shaped the breeding objectives of the global dairy industry over the past years. As such, remarkable progress has been achieved in increasing milk production per cow, in part, through genetic selection [approximately 30%, Cole and VanRaden (2018); Figure 2-1] and the estimated world milk production per cow per lactation has more than doubled over the past few decades (Oltenacu and Broom, 2010). In Australia, the annual milk yield per cow per lactation has grown over the last four decades from approximately 3,500 litres per cow per year in the 1980s to around 7,000 litres in 2015/16 (DataGene, 2021; https://datagene.com.au/, Figure 2-1). Globally, it is projected that milk production needs to double by mid-century to meet the demands of the growing population (Britt et al., 2018). However, rising temperatures from global warming is now threatening the progress of dairy production, which requires proactive and multi-pronged measures to prevent current and future economic losses and contribute to feeding the burgeoning population.

This review covers the genetic aspects of heat tolerance in relation to milk production traits (milk, fat, and proteins) in dairy cattle. First, the problem of climate change and heat stress in the dairy industry is discussed, focusing on the Australian dairy perspective. Next, methods for quantifying heat stress (thermal indices) in the livestock; the impacts of heat stress in dairy cattle, and the physiological mechanisms causing milk decline in heat-stressed cows are reviewed. This is followed by management (shade, sprinklers, feed, etc.) and genetic approaches used to combat heat stress: using adapted breeds, crossbreeding, or genetic selection. Furthermore, published GWAS studies for heat tolerance traits are reviewed. Finally, G × E due to heat stress in the Australian dairy system are discussed followed by a summary and research questions.
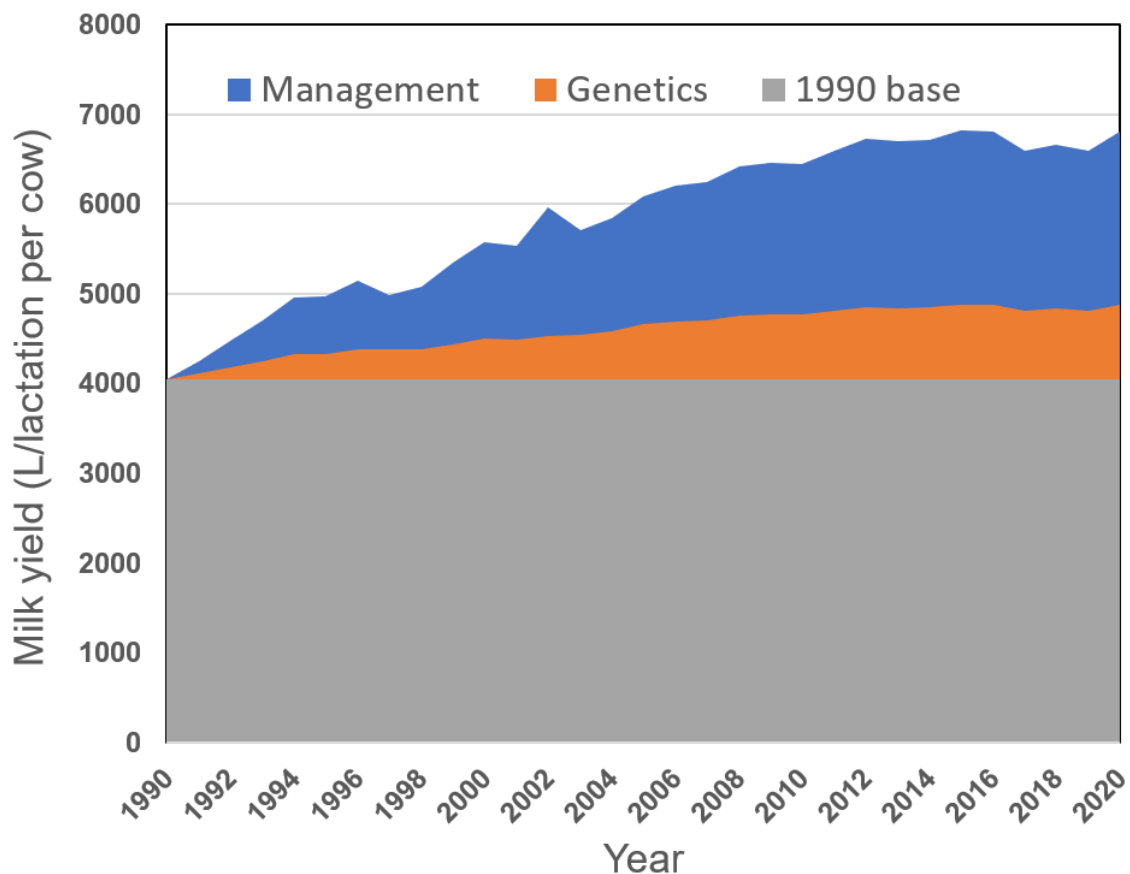
Figure 2-1 The annual increase in milk yield (in litres) per cow per lactation attributed to management (blue) and genetic (orange) improvement in Australian Holstein and Jerseys herd-recordings between 1990 and 2020 [Data source: DataGene Ltd, Melbourne, Australia (https://datagene.com.au/); accessed, October 2021].

## 2.2. **The problem of climate change and heat stress**

Concern about the impacts of warming climates on livestock production used to be an issue mainly in the tropics; but it is now being felt in temperate zones, affecting countries including Australia, Canada, the northern United States, and parts of Europe (Renaudeau et al., 2012, Silanikove and Koluman, 2015, Polsky and von Keyserlingk, 2017). Globally, the weather patterns have changed relative to the pre-industrial climates characterised by an increase in extreme heatwaves, heavy precipitation, and droughts in some regions of the world (IPCC, 2014). Across Australia, the number of hot days per year [defined as the temperature exceeding 39 ^{o}C by the Bureau of Meteorology (BOM)] has been increasing over the years, with very high monthly maximum temperatures occurring in the recent decade (2005 – 2019; Figure 2-2) (BOM, 2020).

Dairy farmers in Australia are already feeling the impacts of climate changes, with dairy cows experiencing heat stress that varies widely across dairy regions ranging from around 300 per year (Queensland) to about 60 days (Tasmania) (Nguyen et al., 2016). The impact of heat stress is expected to increase as the production environments continue to warm across Australia (Dairy Australia, 2020), which calls for proactive measures to minimise both current and future long-term risks to the dairy industry.
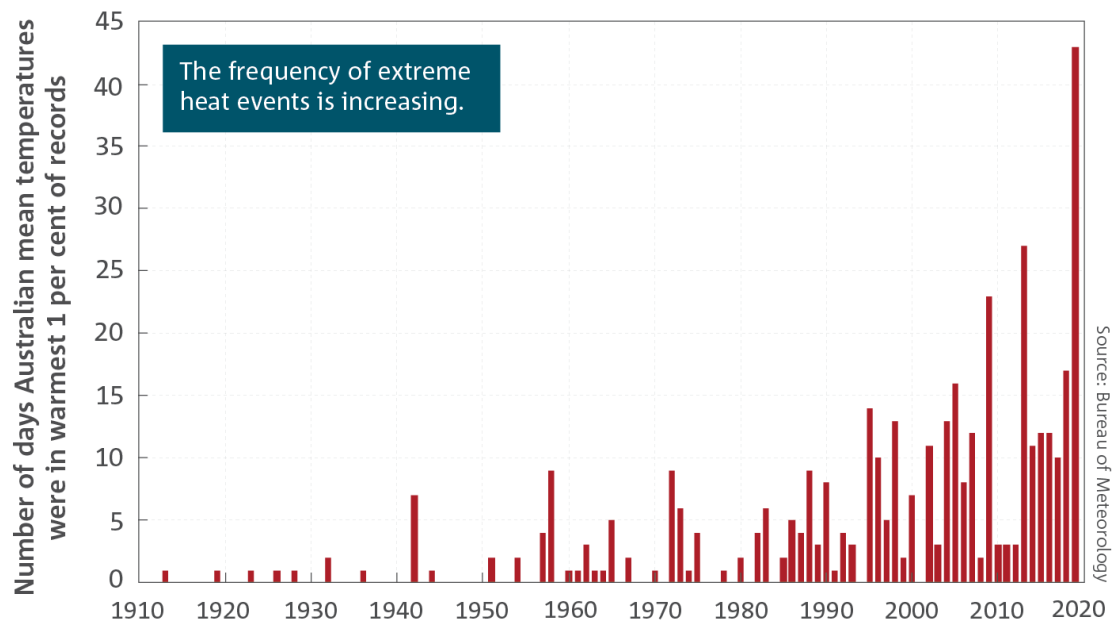
Figure 2-2 Number of days each year where the Australian area-averaged daily mean temperature for each month is extreme (BOM, 2020).

Dairy cattle and other homeotherms maintain stable body temperature within a specific range called thermoneutral zone (TNZ) (Figure 2-3), defined as the optimum thermal environment in which health and productivity thrive (Ames, 1980). It is also defined as the zone of minimal heat production at normal temperature (Kadzere et al., 2002). The TNZ in dairy animals varies widely depending on many factors (e.g., age, breed, feed intake, temperature acclimation, level of production, coat insulation, etc. (Yousef, 1985)), usually ranging from 5 to around 25 °C (Kadzere et al., 2002). Heat stress begins when the ambient temperature rises above the upper critical temperatures of an animal, which triggers physiological and behavioural responses to maintain homeothermy.
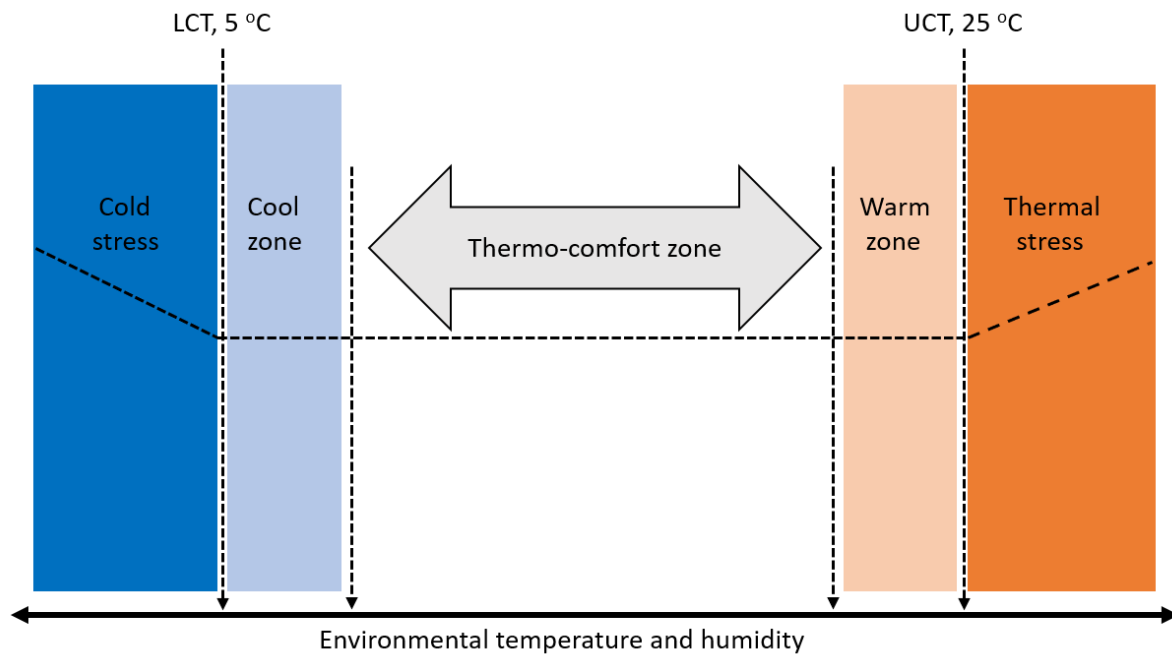
Figure 2-3 Schematic of thermoneutral zone (TNZ), lower critical temperature (LCT), and upper critical temperature (UCT) of a dairy cow [adapted from Kadzere et al. (2002)].

2.3. **Measuring heat stress and environmental heat load in dairy cattle**

Many indices have also been developed to account for the thermal flow mechanisms (ambient temperature, relative humidity, solar radiation, wind speed, precipitation) on the animal and then used to define thresholds beyond which heat stress begins. Due to limited publicly available data on solar radiation and wind speed, most studies have focused on ambient temperature and relative humidity to quantify environmental heat load. The most frequently used environmental heat load index in dairy cattle is the temperature-humidity index (THI), originally developed to measure discomfort in humans (Thom, 1959). THI is a single value that combines ambient temperature and relative humidity. Several THI models (Table 1) have been proposed based on the relative weights given to the individual components: 1) dry bulb temperature (T_{db}) (i.e., air temperature) 2) dew point temperature (T_{dp}) or relative humidity (RH), which measures the amount of moisture in the air at a given temperature. The T_{dp} is often preferred to RH as a better reflection of air saturation (Wood, 1970), calculated as follows:

Dew point $(T) = (237.3 \times B)/ (1.0 + B)$ where, $B = (\ln(RH/100) + ((17.27 \times T_{db})/ (237.3 + T_{db})))/17.27$ and RH = relative humidity.

Table 2.1 THI models proposed for calculations of thermal comfort in cattle.

	Model	Reference
1.	$THI = [0.4 \times (T_{db} + T_{wb})] \times 1.8 + 32 + 15$	(Thom, 1959)
2.	$THI = T_{dp} + (0.36 \times T_{dp}) + 41.2$	(Yousef, 1985)
3.	$THI = (0.35 \times T_{db} + 0.65 \times T_{wb}) \times 1.8 + 32$	(Bianca, 1962)
4.	$THI = (0.55 \times T_{db} + 0.2 \times T_{dp}) \times 1.8 + 32 + 17.5$	(NRC, 1971)
5.	$THI = (T_{db} + T_{wb}) \times 0.72 + 40.6$	(NRC, 1971)
6.	$THI = (1.8 \times T_{dp} + 32 - (0.55 - 0.0055\ RH) \times (1.8 \times T_{dp} - 26)$	(NRC, 1971)
7.	$THI = 1.8 \times T_{dp} - (1 - RH)(T_{dp} - 14.3) + 32$	(Kibler, 1964)

THI – Temperature-humidity index; $\mathbf{T_a}$ – dry bulb temperature or air temperature; $\mathbf{T_{dp}}$ – dew point temperature; T_{wb} – wet bulb temperature; **RH** – relative humidity.

Since climate factors (ambient temperature, relative humidity, solar radiation, etc.) vary widely across regions and countries, some studies have tested the ability of different THI models to estimate milk production losses due to heat stress. For example, in the USA, Bohmanova et al. (2007) compared the suitability of seven THI models to estimate milk losses in some regions in the USA (Athens, Georgia, and Phoenix, Arizona) and concluded that models which give more emphasis to humidity (e.g., Model 3) are best for humid environments, while models giving greater weights to temperature should be appropriate for drier climates (e.g., Model 6). Wang et al. (2018) reviewed 16 thermal indices for dairy cattle including THI models on the suitability to capture thermal stress in cows and concluded that they vary significantly given different climate variables used to construct them; thus, they should be chosen carefully for different production context.

Various approaches have been used to calculate THI in dairy cattle. Some studies compute based on average daily air temperature and humidity (Bouraoui et al., 2002) whereas others use daily maximum and minimum humidity (Ravagnolo and Misztal, 2002, Vitali et al., 2009, Lambertz et al., 2014). Other studies use hourly air temperature and humidity to obtain hourly THI then average them to get overall daily THI value (e.g., Nguyen et al., 2016) while

some calculate separate average daily THI and maximum THI then combine to get composite THI (e.g., Bohmanova et al., 2007, Lambertz et al., 2014).

Studies usually average the mean of THI for several days before test day to capture the effects of cumulated heat loads on performance. According to West (2003), THI values 2 days before test day have significant effect on milk yield. Hayes et al. (2003) found that THI for 3 days preceding the test day has significant effect on herd-test day milk yield for Australian dairy cows. Numerous studies have used 3-day THI values in their analyses (e.g., Bohmanova et al., 2007, Brügemann et al., 2011). Lambertz et al. (2014) showed that 3-day THI is associated with the increase in the somatic cell count. Cumulative THI 4 to 5 days before the day of insemination was reported to have significant negative affect on non-return rate in Italian Holstein cows (Biffani et al., 2016). Hill and Wall (2015) found that the cumulative heat loads up to a week before the test day better captures the variations for milk yield and fat content.

THI is generally considered the best index for measuring heat loads (Dikmen and Hansen, 2009). However, the main limitation of THI is that it fails to account for other important weather elements that contribute to heat stress in animals, such as the wind speed and solar radiation. Consequently, THI has been extended to incorporate wind speed and solar radiation (Mader and Davis, 2002, Mader et al., 2006, Gaughan et al., 2008). Gaughan et al. (2008) formulated a heat load index (HLI) which combines the effect of relative humidity, black globe temperature (BG > 25) and wind speed (WS): $HLI_{BG>25} = 8.62 + (0.38 \times RH) + (1.55 \times BG) - (0.5 \times WS) + e^{(2.4-WS),}$ and $HLI_{BG<25} = 10.66 + (0.28 \times RH) + (1.3 \times BG) - WS$, where e = base of natural logarithm.

Mader et al. (2006) developed a THI adjusted for WS and solar radiation (RAD): $THI_{adj} = [4.51 + THI - (1.992 \times WS) + (0.0068 \times RAD)]$, where $THI = [0.8 \times T_{dp}] + [(RH/100) \times (T_{dp} - 14.4)] + 46.4$. In their study, Hammami et al. (2013) showed that using a measure of THI that accounts for WS and RAD are best suited for temperate climates. More recently, (Lees et al., 2018) developed a dairy heat load index (DHLI) for lactating dairy cows which incorporates BG temperature and relative humidity; where BG is a single value that includes the combined effects of ambient temperature, relative humidity, solar radiation, and wind speed. Overall, despite the development of new indices, THI variants are still widely used, perhaps due to the lack or limited availability of other climatic parameters (wind speed, solar radiation, etc.).

Figure 2-4 shows equivalent temperature and humidity values representing different THI values computed based on. THI of 60, for example, is equivalent to ~20°C and 30%

humidity based on the model. Much higher THI values are expected when calculations are based on a model which gives more weight to relative humidity (e.g., Model 1 above; Thom, 1959). Notably, THI values have been used over the years as a valuable indicator for guiding farm management decisions during hot and humid weather conditions. For example, weather safety indices (LCI, 1970) are classified as [Normal ≤ 74; Alert $75 - 78$; Danger $79 - 83$; Emergency ≥ 84] based on THI values from (Thom, 1959), as shown in Figure 2-4. De Rensis et al. (2015) defined THI of < 68 (normal or comfortable), $68 \leq \text{THI} \leq 74$ (mild heat stress at which production begins to decline), and ≥ 75 (severe heat stress resulting in drastic production decline). Similarly, Armstrong (1994) defined THI < 71, $72 \leq \text{THI} \leq 79$, $80 \leq \text{THI} \leq 90$, and THI > 90 as thermoneutral, mild, moderate, and severe heat stress, respectively. Given the significance of these THI groups in livestock management, it is paramount to have the most accurate THI values (calculated based on suitable THI model).

A number of studies around the world (e.g., Hayes et al., 2003, Boonkum et al., 2011, Brügemann et al., 2012) around the world have reported different THI thresholds at which milk production starts to decline in dairy cattle, in part, because of several factors: 1) differences in the model used to calculate THI; 2) differences in breeds; 3) the level of production among study dairy animals 4) differences in regional climatic conditions, and 5) "acclimatisation" or adjustment to the climate – a process which does not involve genetic changes (Carabaño et al., 2019). Moreover, it could be due to the heat abatement practices and acclimation [defined as coordinated phenotypic response to the environment (Carabaño et al., 2019)], which can influence the onset of stress (Zimbelman et al., 2009) and, thus, the estimated THI threshold. For example, Freitas et al. (2006) in their study reported a relatively higher THI threshold for cows in some regions in the USA which had more efficient heat abatement strategies. Overall, the heterogeneity in the calculation of THI values is an issue of concern to estimating heat tolerance parameters. This calls for a unified calculation approach to allow comparability of results across studies. However, this is not trivial considering that specific models may not be applicable to wide range of environmental conditions and farming systems. THI values based on model 2 [Table 2; i.e., (Yousef, 1985)] were used in all analyses in this thesis, which is comparable to previous studies for heat tolerance in Australia (Nguyen et al., 2016). Nevertheless, given the drawback of the THI model used in these studies (i.e., does not account for airspeed and solar radiation), it is paramount to re-evaluate a suitable model (s) for the Australian context since dairy herds are predominantly kept on pasture, meaning that solar radiation contributes to heat stress.

Temperature (°C)	Relative Humidity (%)											
	5	10	15	20	30	40	50	60	70	80	90	100
20	51	55	57	59	62	64	66	67	69	70	71	71
21	52	56	59	61	64	66	67	69	70	71	72	73
22	53	58	60	62	65	67	69	70	72	73	74	74
23	55	59	62	64	67	69	70	72	73	74	75	76
24	56	60	63	65	68	70	72	73	75	76	77	77
25	57	62	64	66	69	72	73	75	76	77	78	79
26	59	63	66	68	71	73	75	76	77	79	80	81
27	60	65	67	69	72	74	76	78	79	80	81	82
28	61	66	69	71	74	76	78	79	80	82	83	84
29	63	67	70	72	75	77	79	81	82	83	84	85
30	64	69	71	74	77	79	81	82	83	85	86	87
31	66	70	73	75	78	80	82	84	85	86	87	88
32	67	71	74	76	79	82	84	85	86	88	89	90
33	68	73	76	78	81	83	85	87	88	89	90	91
34	70	74	77	79	82	85	87	88	89	91	92	93
35	71	76	79	81	84	86	88	90	91	92	93	94
36	72	77	80	82	85	88	89	91	92	94	95	96
37	74	78	81	83	87	89	91	92	94	95	96	97
38	75	80	83	85	88	90	92	94	95	97	98	99
39	76	81	84	86	90	92	94	95	97	98	99	100
40	78	83	86	88	91	93	95	97	98	100	101	102
41	79	84	87	89	92	95	97	98	100	101	102	103
42	80	85	88	91	94	96	98	100	101	103	104	105
43	82	87	90	92	95	98	100	101	103	104	105	106

Figure 2-4 Temperature-humidity index (THI) chart based on (Yousef, 1985). Heat stress risk categories are defined as [Normal ≤ 74; Alert 75 – 78; Danger 79 – 83; Emergency ≥ 84], as described by Hahn et al. (2009).

2.4. **Impacts of heat stress**

Heat stress in animals has complex and multifaceted consequences. Elevated temperatures above the TNZ of the dairy cow triggers physiological changes leading to increased core body temperature, respiratory rate, panting heart rate, and sweating, as well as hormonal changes (Kadzere et al., 2002, Das et al., 2016). Consequently, these can lead to suppressed feed intake, altered rumen function, and udder health often resulting in decreased quantity and quality, reproduction, and general deterioration of cow welfare and death in some extreme cases (Das et al., 2016, Polsky and von Keyserlingk, 2017, Pragna et al., 2017). Specifically, heat stress alters mammary gland function and integrity leading to reduced milk

yield (Tao et al., 2018). In addition, an elevated temperature predisposes the cow's udder to infection by bacteria such as streptococci which subsequently leads to decreased milk quality and composition (Pragna et al., 2017).

The rate of milk yield decline after THI exceeds the comfort threshold (i.e., under heat stress) has been quantified by several studies in livestock. In dairy cattle, Sánchez et al. (2009a), Bernabucci et al. (2010), and Bohmanova et al. (2007) found milk decay of up to -0.15 kg, -0.27 kg, and -0.6 kg per THI unit above the comfort threshold, respectively. Similarly, a milk decay estimate of -0.2 kg/unit THI was reported for dairy cattle in the Georgia-USA (Ravagnolo and Misztal, 2000). In New Zealand, Bryant et al. (2007b) reported a milk solids reduction of more than -10 g/day per unit increase of THI above the threshold of 68, 69, and 75 in Holsteins, Jersey, and H × F crossbreds. Fat and protein yield decline of up 8.6 g/unit THI above the threshold was reported in dairy sheep (Finocchiaro et al., 2005). While it has been widely documented that heat-stress lowers milk production differently within and between breeds, there is a lack of information on genetic aspects of the recovery period of production to baseline after heat stress and is worth exploring to identify resilient animals.

Apart from milk production, heat stress can adversely affect reproductive performance in dairy animals. For example, elevated heat stress can compromise female fertility due to reduced duration and intensity of oestrous, altered follicular development, and impairment of embryonic development (Jordan, 2003). In males, heat stress can negatively affect the quality and integrity of semen, resulting in reduced fertility and overall reproductive performance (Das et al., 2016, Schüller et al., 2016). Several have reported seasonal variability in the reproductive performance of dairy cattle. In Australia, Talukder et al. (2015) found that dairy cows calving in autumn are 43% more likely to conceive than cows calving in summer (under heat stress). Similarly, in the southern USA, De Vries and Risco (2005) found a 15.8% pregnancy rate in winter compared to 5.6% in summer.

2.5. Physiological mechanisms for milk production decline

Understanding how heat stress alters the physiology of dairy cows to impact productivity is critical in developing mitigation strategies (genetic and management) to prevent losses and improve the health and well-being of animals. Lactating dairy cows are sensitive to changes in the environment which can lower their milk yield by up to 40% under heat stress conditions (West, 2003). Multiple mechanisms have been proposed for this decline. Rhoads et al. (2010) found that depressed feed intake that often occurs during heat stress explains around

35% to 50% of milk yield decline while the cause of the remaining portion of milk drop is still not well understood, but inconclusive evidence suggest that it is related, in part, to post-absorptive nutrient utilizations. The depressed feed intake in heat-stressed dairy cows and other ruminants is a survival strategy to reduce the metabolic heat load generated during milk production. Beside the reduced feed intake, heat-stressed cow alters eating behaviour characterised by smaller meals and more lying time, which predisposes them to rumen acidosis (Bernabucci et al., 2010) when compared to cows in thermoneutral conditions.

Like other mammals, heat-stressed dairy cows use energy to maintain homeostasis during heat stress, predominantly by evaporative heat loss via panting and sweating (Fuquay, 1981). As such, it has been hypothesised that the nutrients required for production are diverted to maintain homeostasis contributing, in part, to the milk decline in heat-stressed animals (Baumgard et al., 2011, Baumgard and Rhoads Jr, 2013). Controlled studies (e.g., Garner et al., 2016) show that heat-tolerant animals have superior thermoregulation ability than heat-susceptible animals suggesting less energy expenditure for homeostasis and thereby less reduction in their milk yield. These authors found that heat-tolerant cows are more efficient at dissipating heat via vasodilation (measured by skin surface temperature); therefore, they were able to maintain relatively lower core body temperature compared to the heat-susceptible cows. Similar findings were reported in the USA for slick (more tolerant) versus non-slick (less heat-tolerant) crossbred Holstein cows in which the latter cows exhibited lower body temperature and less milk yield decline contributed, in part, by superior heat dissipation via sweating (Dikmen et al., 2014). Overall, these studies suggest that animals that spend less energy on their thermoregulation have less reduction in milk yield and would be more suited for future warming climates.

Depressed feed intake coupled with increased energy needed for homeostasis predisposes dairy animals to the negative energy balance (NEBAL), typical to what often occur in early lactation cows (Bernabucci et al., 2010). The effects of heat stress on plasma non-esterified fatty acid (NEFA) associated with depressed feed intake are conflicting among studies with some reporting increased NEFA (e.g., Garner et al., 2020), while others indicate a decrease or no change in NEFA (e.g., O'brien et al., 2010, Wheelock et al., 2010). Unaltered NEFA is partly related to the inability of heat-stressed cows to employ 'glucose sparing' mechanisms, such that adipose tissue is not mobilised to generate NEFA (Rhoads et al., 2010, Baumgard and Rhoads Jr, 2013). Another physiological adjustment arising from NEBAL

during heat stress is related to the altered carbohydrate metabolism that is mediated, in part, by accentuated glucose uptake driven by elevated hepatic insulin activity (Wheelock et al., 2010, Baumgard and Rhoads Jr, 2013). The lipid composition profile of milk is also altered under acute heat stress in cows leading to reduced abundance of short- and medium-chain fatty acids and the relative increased abundance of long-chain fatty acids (Liu et al., 2017).

Milk production in dairy cattle and other mammals depends largely on the number and efficacy of the synthetic capacity of the mammary gland epithelial cells. Exposure of dairy cows to heat stress before and during lactation negatively affects mammary gland epithelial cell development, physiology, and integrity impacting milk production [reviewed by (Tao et al., 2018)]. Dado-Senn et al. (2018) conducted RNA-Seq analysis and identified over 3,000 candidate genes and pathways involved in mammary gland development under heat stress, including upregulation of cell death, cytoskeleton degradation, and immune response. Recent evidence shows that heat stress in dry and pregnant cows has negative carry-over effects on the lifetime performance of their progeny (Laporta et al., 2020), implying severe economic cost to the dairy industry. Moreover, heat stress alters hormonal (insulin, prolactin, glucocorticoids, etc.) functions and reduces the rate of blood flow across the mammary gland leading to the decreased supply of pre-cursor nutrients (amino acids, lipids, glucose, etc.) to the mammary gland needed for milk synthesis (Bernabucci et al., 2010).

The nervous system in mammals connects the internal and external environment and thereby plays a crucial role in regulating core body temperature (Nakamura and Morrison, 2008). This process is initiated by the activation of transient receptor located in the nerve endings of the thermosensory organs in the dermis and epidermis of the peripheral organs. This signal then travels (afferent pathways) to the pre-optic area of the hypothalamus and finally to the anterior part of the hypothalamus (Collier and Gebremedhin, 2015). This region of contains thermosensitive neurons that act as the control centre (or the thermostat) of the brain (Collier and Gebremedhin, 2015). This thermosensory signal is interpreted and an appropriate message is sent (efferent pathways) to trigger various behavioural responses such as shivering, painting, peripheral vasodilation, and hormonal production (Boulant, 2000, Collier and Gebremedhin, 2015).

The mechanism by which the nervous system is coordinated in heat-stressed dairy cows has not been widely studied. It is hypothesised that the nervous system is employed during chronic heat stress to coordinate a cascade of metabolic and hormonal processes associated

with responses including affecting milk production, such as growth factors, insulin, serotonin, thyroid, prolactin, and mineralocorticoids associated with milk synthesis (Bernabucci et al., 2010). The acute or short-term response to heat stress in animals is mediated by nervous thermoregulatory systems, which trigger the release of hormones (e.g., catecholamines and glucocorticoids) while the chronic phase of heat stress (long-term or recurrent exposure to heat) is driven by homeorhetic regulators of the endocrine system (resulting in acclimation or new physiological state) both of which alters metabolism and energy balance; thus, contributing to milk decline (Collier et al., 2017). Some transcriptomic studies (e.g., Kim et al., 2017) have found evidence of genes and pathways related to the nervous system (neuroactive ligand-receptor interaction) in avian species under heat stress conditions. Pegolo et al. (2018) found that genes associated with milk proteins in dairy cattle are involved with neuronal and hormonal signalling. More studies are needed to understand the genetic basis and biological pathways underpinning the neuronal system in dairy cattle under heat stress conditions, which may encourage new interventions to minimise production losses.

2.6. **Management solutions to combat heat stress**

Various short-term heat stress mitigation strategies have been implemented with some degree of success, including environmental modifications to prevent or limit heat exposure (e.g., the use of shades, sprinklers, and fans) and nutritional interventions aimed at increasing feed intake and decreasing metabolic heat production (Renaudeau et al., 2012; Fournel et al., 2017).

In Australia, the Cool Cows program (https://coolcows.dairyaustralia.com.au/) provides dairy farmers with practical solutions to effectively deal with heat stress, particularly using shades, sprinklers, and fans. The cost-benefit analyses of various heat abatement strategies are also provided in the tool. In Australia, Nidumolu et al. (2010) demonstrated that shade is more effective as a cooling strategy for ameliorating heat than sprinklers. These authors found that the mean THI for shade and spray treatment over the 182 days were 70.1 and 73.4, respectively, suggesting that using spray likely adds more relative humidity. The study also found that animals left in the open without any adaptation exhibited the greatest milk loss; thus, emphasizing the importance of adopting mitigation strategies. While such adaptation strategies are generally effective in combating heat stress, the additional production costs associated with implementation will eventually have implications on the net marginal income in the dairy farms. It is estimated that the energy expenditure associated with cooling cows

costs an additional USD 0.86 per weight of milk in the warmest regions of the USA (Key et al., 2014). Apart from the costs, the large volume of water required for cooling further casts doubt on the future sustainability of cooling strategies, especially in regions where water availability is limited or restricted (Polsky and von Keyserlingk, 2017). Since dairying in Australia is predominantly seasonal pasture-based, the implementation of such interventions can be challenging and cost-prohibitive for farmers with limited budgets. However, with no alternative option, some farmers are currently implementing such cooling strategies in many countries, including some dairy regions in Australia such as New South Wales and Queensland. To facilitate management decisions, intelligent tools that provides heat alerts (https://dairy.katestone.com.au/) have been developed for use by Australian dairy farmers.

Various nutritional strategies to cope with heat stress include ration modification, restricted feeding, and shifting feeding time (West, 2003, Wheelock et al., 2010). The general goal of these strategies is to maintain homeostasis and prevent nutrient deficiencies (Mader, 2003, Das et al., 2016). Since heat stress reduces feed intake, a common strategy in ration formulation is to reduce fibre content and increase concentrates while taking care not to induce rumen acidosis (Baumgard et al., 2014). The most common and critical well-known strategy in managing heat-stressed cows is to provide adequate, cold, and clean water.

While heat abatement solutions have been quite successful, it is expected that production will continue to be negatively affected by chronic heat (Renaudeau et al., 2012). Furthermore, as water scarcity continues to become a challenge and the cost of cooling increases, long-term selective breeding for heat tolerance will be inevitable. In the following sections genetic solutions to improve heat tolerance are reviewed.

2.7. **Genetics of heat tolerance to combat heat stress**

Heat tolerance is the ability of an animal to maintain thermal stability at high temperatures and humidity (Carabaño et al., 2019), largely governed by the relationship between heat dissipation capacity and metabolic heat production. The main biological determinants of heat tolerance in animals include the relative body surface area (Berman, 2003), sweating rate (determined by the morphology, density, and water transfer capacity of sweat glands) (Bru et al., 1987), coat characteristics (length, thickness, colour, weight per unit surface, etc.) (Olson et al., 2003, Dikmen et al., 2008, Collier and Collier, 2011) as well as the rate of metabolic heat production and dissipation (e.g., lactating versus non-lactating cows) (Bernabucci et al., 1999, Farooq et al., 2010). The genetic variants underpinning these

biological and morphological aspects in animals are still poorly understood. As the climate gets warmer, the selection of animals with superior genetic traits for heat tolerance is gaining increasing interest globally using several approaches, which I will now discuss.

2.7.1. **Use of adapted breeds, crossbreeding or gene editing**

Considerable genetic variation exists for heat tolerance between and within individuals of a breed (Ravagnolo and Misztal, 2002, McManus et al., 2009, Renaudeau et al., 2012, Gantner et al., 2017). The literature is rich, showing that the adapted tropical cattle have the inherent ability to resist environmental stressors (heat, disease, parasites, poor nutrition, etc.), which can be exploited to improve resilience in temperate breeds. For example, the tropical local *Bos indicus* breeds (Figure 2-5) have superior heat tolerance ability than temperate *Bos taurus* breeds (Hansen, 2004). Besides the morphological adaptations (sweating capacity, hair coat characteristics, tissue insulation), the heat tolerance ability of tropical *Bos indicus* is associated with small body size and low production level compared to *Bos taurus* (Figure 2-5) breeds (Berman, 2011, Collier and Collier, 2011). The low productivity of these *indicus* breeds is considered unsuitable for high production in temperate climates; but can be very important for successful breeding in harsh and hot climatic conditions (Renaudeau et al., 2012). It is believed that *Bos indicus* breeds acquired thermotolerance genes after sustained exposure to hot and harsh environments following domestication (Hansen, 2004, Collier and Collier, 2011). Thus, the adaptive genes for heat tolerance can be identified and introduced to crossbreds, or in high-producing heat susceptible breeds (Mcmanus et al., 2014) either using crossbreeding strategies or genome editing.

Figure 2-5 *Bos taurus* (Jersey and Holstein), *Bos indicus* (Boran), and composite breed (Senepol). Photo credit: Holstein and Jersey – https://domesticanimalbreeds.com; Senepol – https://learnnaturalfarming.com/senepol-cattle-breed/; Boran – https://www.boran.org.za.

Unlike *Bos indicus* breeds, *Bos taurus* breeds are well known for high milk production but generally more susceptible to heat stress, thus less adapted to harsh and hotter climatic conditions. Variations in heat tolerance also exist between *Bos taurus* cattle breeds. For example, greater heat tolerance in Jerseys than in Holsteins has been documented, although Holsteins generally produce more milk (Bryant et al., 2007b). These researchers also found that the New Zealand Holsteins appear to exhibit higher reductions in milk yield in hotter climates than Jerseys or crossbreds, indicating lower heat adaptation for Holsteins.

Crossbreeding can be exploited to introgress genes from breeds adapted to heat stress to high producing yet heat susceptible animals. Crossbreeding exploits breed complementarity and heterosis, and is beneficial for low to moderately heritable traits, such as heat tolerance (Bourdon, 2000). Recently, Buckley et al. (2014) presented a comprehensive review of past crossbreeding practices in dairy cattle. The authors noted favourable animal performance in crossbreds for fertility and survival traits. It has been shown that crosses of *Bos indicus* breeds with *Bos taurus* are more heat tolerant compared to pure *Bos taurus* breeds (Gaughan et al., 1999, Mcmanus et al., 2014). Among *Bos taurus* breeds, several studies have reported greater heat tolerance for Holstein x Jersey crosses than pure Holstein breeds (Bru et al., 1987, Muller and Botha, 1993, Bryant et al., 2007b, Vermunt and Tranter, 2011, Smith et al., 2013). While these crossbreeding studies have generally demonstrated superior performance for crossbreds

compared to adapted breeds, the adaptive potential to heat stress often reduces in crossbreds (Mcmanus et al., 2014). The percentage of adapted breed genetic background required to benefit from heterosis advantage for heat tolerance in crossbreds while optimising productivity remains unknown. Regarding production traits, a study by Hickson et al. (2014) reported heavier birth weight for calves born from Angus crossbred cows compared to purebred Angus cows suggesting, in part, the effect of heterosis in crossbred cows.

Perhaps a well-known successful application of crossbreeding to improve thermotolerance in Holsteins was carried out in the USA by crossing Holsteins with the Senepol cattle (Dikmen et al., 2014). The Senepol breed (Figure 2-5) is predominantly *Bos taurus* ancestry and originates from the island of St Croix in the Caribbean (Padda, 1999). The individuals of this breed have short sleek hair coat that is known to be a major determinant for their inherent thermotolerance. The pioneering work of Olson et al. (2003) identified a major gene responsible for the sleek coat in the Senepol cattle, referred to as the SLICK gene. The causal mutation for the SLICK gene has been mapped to the prolactin receptor gene (*PRLR*) on chromosome 20 (Littlejohn et al., 2014). These authors described SLICK phenotype to be associated with frameshift mutation of single base deletion in exon 10 of *PRLR* gene which causes premature stop codon and loss of carboxyl-terminus of the amino acid from the long isoform of the receptor (Figure 2-6). Crossbreeding work of Dikmen et al. (2014) focused on introducing the thermotolerance slick *PRLR* variant of the Senepol cattle into Holsteins. The inheritance of SLICK haplotype in crosses was confirmed by an association study with the Senepol population (Dikmen et al., 2014). These authors reported lower rectal temperature in heat-stressed slick-haired crossbred cows that inherited the SLICK haplotype compared to non-slick cows. In addition, the slick-haired crossbred cows showed a relatively high sweating rate and less reduction in milk yield during heat stress, indicating superior thermotolerance ability compared to non-slick cows (Dikmen et al., 2014). Similar work is currently underway in New Zealand and Puerto Rico to introduce the SLICK gene to Holstein dairy cattle (Davis et al., 2017, Hansen, 2020). Overall, the above studies demonstrate the potential to improve heat tolerance using either between-breed selection or crossbreeding. However, as to which approach is more suitable for fast-tracking the genetic gains for heat tolerance in livestock is debatable. Perhaps, combining various approaches is more effective.
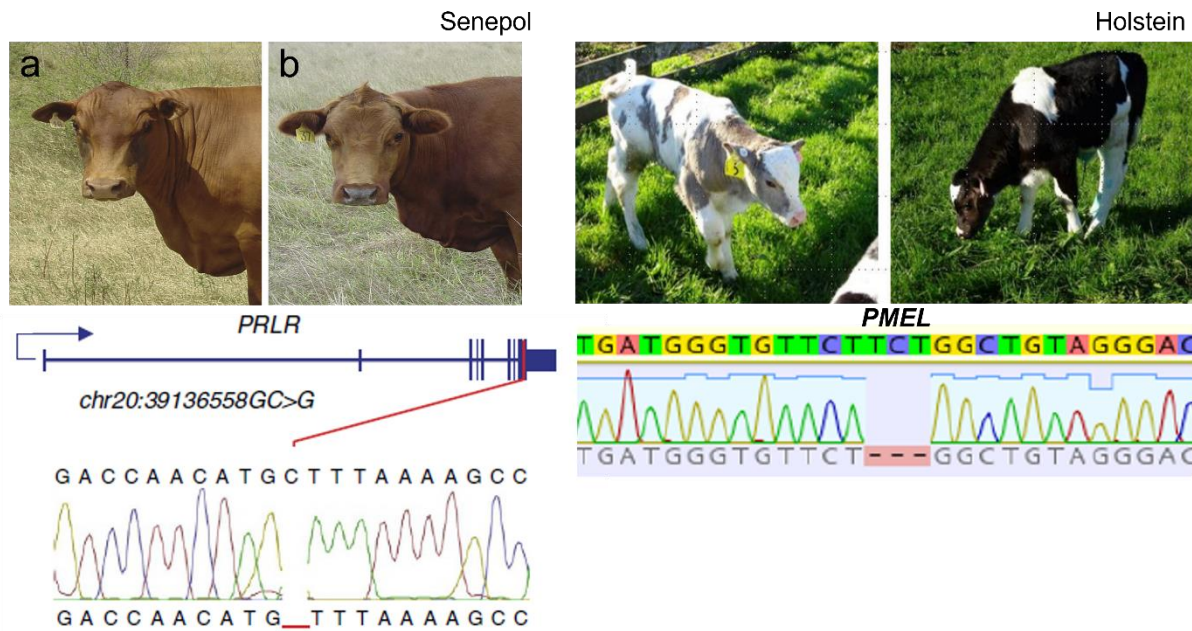
Figure 2-6 Left picture: SLICK (a) cow and non-SLICK (b) Senepol cows showing causal mutation at *PRLR* gene in chromosome 20 (Littlejohn et al., 2014). Right plot: wild-type (black and wild calf) and *PMEL* mutant Holstein calf (light skin pigmentation) from gene-editing involving three bp deletion in chromosome 5 at 57 Mb (Laible et al., 2020).

Besides crossbreeding, gene editing is another way to modify or introduce thermotolerance genes into heat-susceptible animals. Gene editing tool has been successfully used to improve growth, diseased resistance, and welfare (e.g., hornless cattle) in farmed animals (see comprehensive review by Tait-Burkard et al., 2018). The utility of this method to thermotolerance traits in animals is currently at the early stages in the dairy industry and is attractive because it could potentially facilitate a more rapid and targeted transfer of the desirable genes instead of wholescale gene transfer via crossbreeding (Hansen, 2020). Therefore, it may be more suitable for modifying traits controlled by few alleles with large effects. Recent work in New Zealand (Laible et al., 2020) has attempted to lighten the coat color of Holsteins (Figure 2-6) through gene editing of the pre-melanosomal protein 17 gene (*PMEL*) aimed to reduce absorption of solar radiation under hot weather conditions. Although this study was successful in producing *PMEL*-mutant calves (Figure 2-6) they, unfortunately, died within four weeks of birth, meaning that the benefits of this gene-editing with respect to performance and heat tolerance could not be assessed.

Since heat tolerance is antagonistically correlated with the level of milk production (Ravagnolo and Misztal, 2000), it may be more strategic to use gene editing to improve heat

tolerance while safeguarding milk yield. However, since heat tolerance is a highly polygenic trait (i.e., controlled by many genes or alleles), it may be cumbersome to simultaneously edit multiple alleles. Therefore, genome editing needs to be incorporated with other genetic solutions such as genomic selection. Notably, genetic targets associated with heat tolerance for editing are still largely unknown in cattle and other animals.

2.7.2. **Selection of heat tolerance within breeds and measurable phenotypes**

There is a strong desire to improve heat tolerance within breeds for two main reasons. First, milk production is the main driver of dairy profitability worldwide. Therefore, improving heat tolerance in high-producing breeds such as Holsteins helps to maintain milk production. Second, breeding in the dairy industry is mostly based on the within-breed selection principle mainly on sire-pathway. Notably, the genetic variation that exists within breeds provides an excellent opportunity to improve heat tolerance while maintaining milk volume. This was demonstrated by the work of Nguyen et al. (2016) who identified heat-tolerant individuals within Holstein cows that were confirmed in a later controlled study to exhibit better thermoregulation abilities and lower reduction of their milk yield under heat stress compared to the heat susceptible cows (Garner et al., 2016). While it is widely known that inherent genetic differences (coat color, body size, sweating capacity, hair coat characteristics, tissue insulation, production level, etc.) confer thermotolerance between breeds (e.g., *Bos indicus* versus *Bos taurus* breeds), the genetic mechanisms that allow individuals within a breed to differ in thermotolerance are still poorly understood.

It is necessary to define a measurable trait for heat tolerance to identify and select heat-tolerant animals. Measures of body temperature (rectal, vaginal, rumen temperature, etc.) (e.g., Dikmen et al., 2008, Otto et al., 2019), performance (production, reproduction, health) (e.g., Ravagnolo and Misztal, 2000, Nguyen et al., 2016) and physiological (respiratory rate, sweating rate) (e.g., Dikmen et al., 2015) traits under heat stress conditions have been used to describe heat tolerance in dairy cattle and other livestock species. Previous studies, e.g., (Dikmen et al., 2008, Garner et al., 2016) have consistently demonstrated that heat tolerance animals can maintain lower core body temperature during heat stress due to their superior thermoregulatory abilities (Figure 2-7). This trait is often considered as a 'gold standard' measure of heat tolerance in cattle (Carabaño et al., 2019).
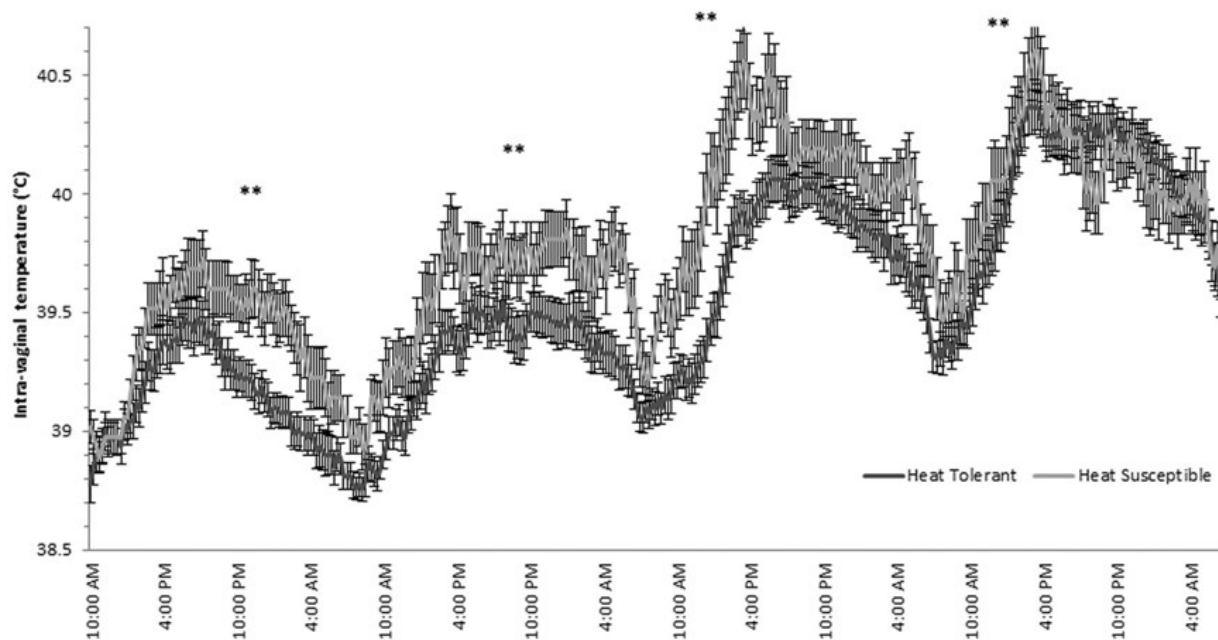
Figure 2-7 Intra-vaginal temperature for heat tolerant versus heat susceptible cows measured at different time points under experimental conditions with heat chambers in Australia [Source; Garner et al. (2016)].

The decline of milk, fat, and protein yield as a function of the environmental heat load is the most frequently used proxy of heat tolerance (e.g., Ravagnolo et al., 2000, Bohmanova et al., 2007, Aguilar et al., 2009, Sánchez et al., 2009a, Boonkum et al., 2011, Brügemann et al., 2011) as illustrated in Figure 2-8. The smaller rate of decline in performance with increasing heat loads is a characteristic of greater heat tolerance (Ravagnolo and Misztal, 2000, Ravagnolo et al., 2000). Also, the higher the THI threshold at which the performance begins to decline indicates greater heat tolerance (Sánchez et al., 2009b). Notably, the rate of milk decline for each animal (Figure 2-8) was used throughout this thesis to describe heat tolerance.
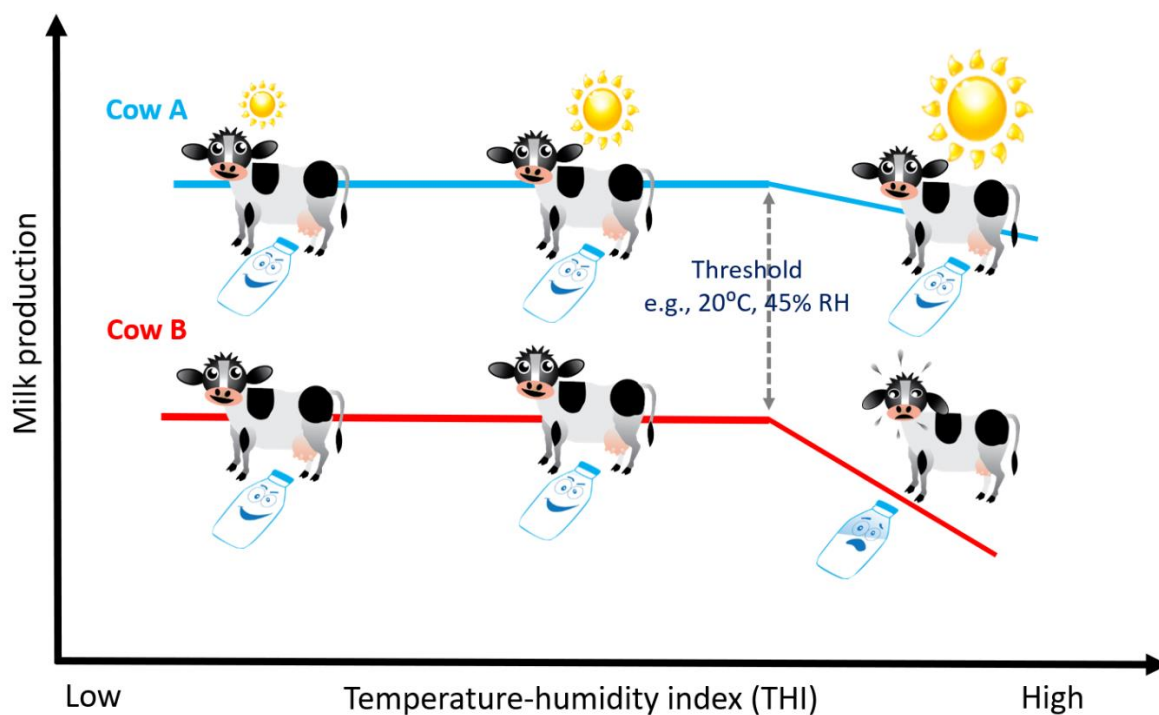
Figure 2-8 Describing heat tolerance. Cow A and B produce comparable quantity of milk at thermoneutral conditions (i.e., at low THI). As the temperature and humidity increases, the milk yield at first remains unaffected up to a given point called a threshold at which the milk begins to decline for both cows, but the rate of decline (indicated by slopes) is larger for Cow B than Cow A. Therefore, Cow A is considered more tolerant to heat than cow B. This thesis used the slope traits for milk, fat, and protein yield to define heat tolerance (image credit: Dr Thuy Nguyen; DataGene Ltd, Melbourne, Australia).

Modelling heat stress based on the decline in production traits has allowed the use of large datasets available from milk recordings that are combined with climate data from public weather stations (e.g., Ravagnolo and Misztal, 2000, Nguyen et al., 2016). However, such phenotypes do not fully capture heat tolerance, particularly when milk reduction due to heat stress is estimated from less frequently recorded test day milk production. Furthermore, test day variation in the physiological status of the animals and possible differences in acclimation, if not accounted for can result in biased estimates of the responses to heat stress among the animals. The heritability estimates of heat tolerance measured based on the rate of milk decline under heat stress in dairy cattle is low (~0.1) to moderate (~0.30) (Ravagnolo and Misztal, 2000, Brügemann et al., 2011, Carabaño et al., 2014, Nguyen et al., 2016), while that of rectal temperature under heat stress conditions is 0.17 (Dikmen et al., 2012).

Heat stress triggers various physiological conditions resulting in up- or down-regulation of metabolites in plasma and milk that can be used as biomarkers of heat stress. Several plasma biomarkers have been identified, including urea, insulin, and glucose (Wheelock et al., 2010), *HSP70* (Gaughan et al., 2013), NEFA (O'brien et al., 2010) as well as cortisol (Pereira et al., 2008). While these candidate biomarkers are generally associated with heat stress, it is not yet conclusive as to which set of biomarkers are most suitable for identifying thermotolerant animals. Furthermore, anecdotal reports are available on whether these plasma biomarkers are better indicators of heat stress compared to conventional measures of milk decline.

Apart from plasma, milk biomarkers are considered as most promising indicators of heat stress since they are non-invasive to the animals and large-scale on-farm data can be obtained. Hammami et al. (2015) and Liu et al. (2017) found that among the lipids, C18:1 *cis*-9 fatty acids and Lysophosphatidylcholine polar lipids are the most sensitive to heat stress and can be used as potential biomarkers for heat tolerance. In the metabolomic work of Tian et al. (2016), 10 potential heat stress milk biomarkers were reported, including the C18:1 *cis*-9 fatty acids.

With the recent advancement in sensor technology, exciting opportunities are emerging to both scale up measurable data and to capture novel traits for heat tolerance and other complex traits in real-time. Novel wearable sensor devices are now available that can measure heat stress based on body temperature (e.g., Kou et al., 2017) and breathing dynamics (Bar et al., 2019) on an almost continuous basis. Some studies have reported a good agreement between the data from wearable devices and those from vaginal or rectal measurements [see review by Koltes et al. (2018)]. However, since animals differ in how they respond to heat, it is important to rigorously validate wearable measuring devices to minimise risks (Pryce et al; unpublished). Ultimately, the availability of large-scale data from sensor devices in the foreseeable future may facilitate a comprehensive assessment and a more accurate selection of heat-tolerant animals.

2.8. Genomic selection

Genomic selection (GS) is a revolutionary tool that uses genomic markers to estimate breeding values of animals, first described by Meuwissen et al. (2001). GS is based on the basic principle that the existence of linkage disequilibrium (LD) between one or several markers (e.g., SNPs) and the quantitative trait loci (QTL) can be used to explain variations in quantitative traits. In practice, GS involves deriving prediction equation from a reference

population with both phenotypic (e.g., milk production, fertility, conformation traits, etc.) and genotypic information (e.g., the standard 50k array) (Figure 2-9). This prediction equation is generated from the reference population and used to estimate breeding values for each animal by combining marker genotypes (often coded as 0, 1, and 2 based on the minor allele dosage) and their effects. This prediction equation can be used to compute breeding values from another set of genotyped animals with no phenotypic records. These animals are then ranked based on their genomic breeding values (GEBVs) to select the best animals for breeding.

The use of GS has been a breakthrough of the recent decade, particularly in overcoming the limitations of conventional breeding techniques. Specifically, GS is advantageous for several reasons: 1) it allows selection of animals early in life, thereby reducing generational interval; thus, greater genetic gain (García-Ruiz et al., 2016), 2) it is not limited to sex 3) it is advantageous for a hard-to-measure or expensive-to-measure traits (Eggen, 2012) and 4) it provides more accurate breeding values and reduces the cost of proving bulls (Schaeffer, 2006, VanRaden et al., 2009). The generation interval (the average age of the parents at birth of their progeny) in dairy cattle (as described in the following genetic response model) has been dramatically reduced by more than half after the introduction of GS: [($\Delta G = r * \sigma_g * i/L$), where G = annual change of trait's genetic merit, r = accuracy of selection (correlation between breeding values and estimated breeding values), σ_g = amount of genetic variation, i = selection intensity, L = generation interval]. This has allowed rapid genetic improvement, especially for complex traits.
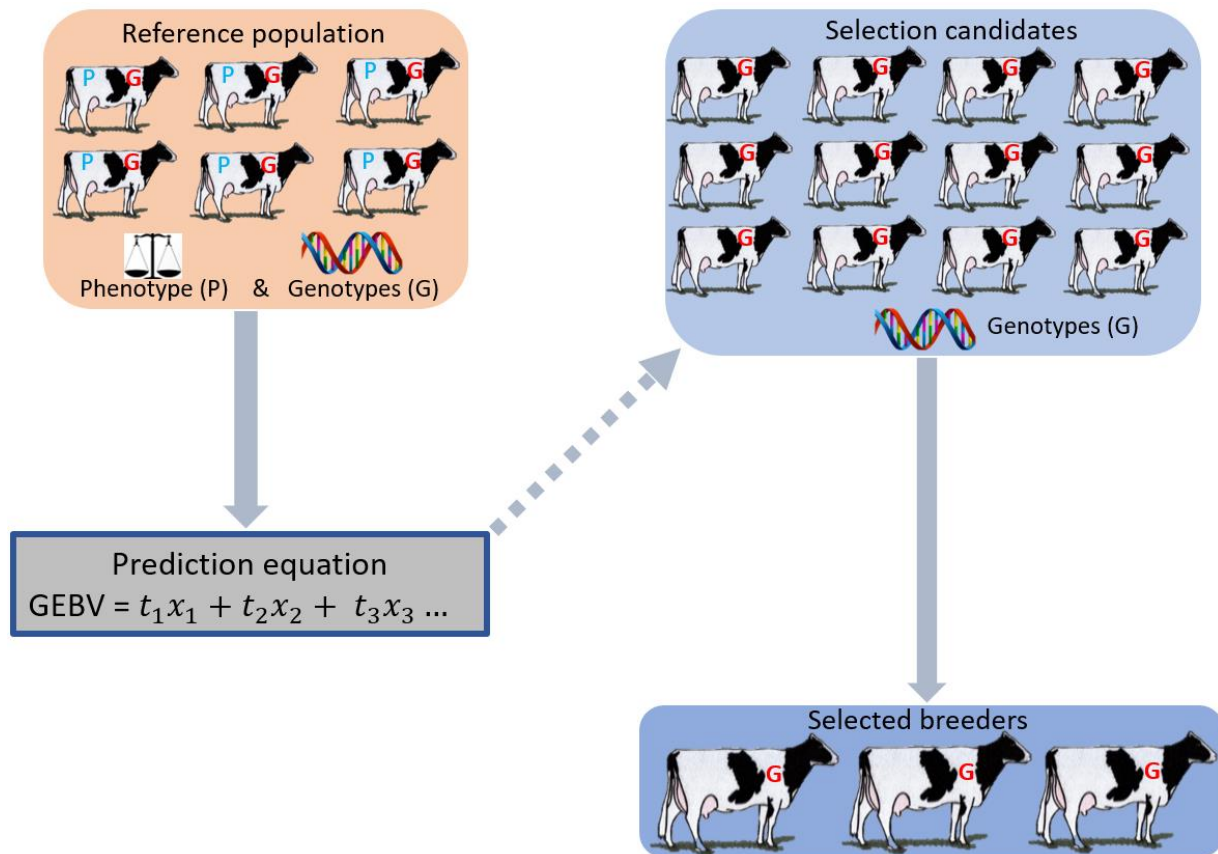
Figure 2-9 Genomic selection involves deriving prediction equation from a reference population with both phenotypic and genotypic information. This prediction equation can be used to compute breeding values from another set of genotyped animals with no phenotypic records. These animals are then ranked based on their genomic breeding values (GEBVs) to select the best animals for breeding.

Fertility and feed conversion efficiency are among the traits that are difficult or expensive to measure, which have benefited considerably from GS. In Australia, promising estimates of the accuracy of genomic prediction of feed efficiency have been achieved and are now incorporated in the national selection indices (Pryce et al., 2012, Pryce et al., 2018). Also, favourable accuracies of GEBV for fertility traits have been reported in dairy cattle (Wiggans et al., 2011, Zhang et al., 2014). Besides fertility and feed efficiency traits, genomic selection for heat tolerance in the livestock industry has received growing attention in recent years, driven by the desire to cope with the effects of global warming.

Recently, the potential of genomic selection for heat tolerance in Australian dairy cattle was demonstrated by (Nguyen et al., 2016, 2017). In that work, heat tolerance was estimated by using test-day records for milk, fat, and protein yield that were matched with meteorological data from public weather stations near dairy herds. Heat tolerance was described as the rate of

decline in milk yield (also defined as heat tolerance slope traits) with increasing variability of heat stress using reaction norm models. With the slope traits (milk, fat, and protein), GEBV were calculated using a prediction equation that was derived from the medium-density (50k) SNP markers. Then, GEBV for three slope traits (milk, fat, and protein) were combined to obtain Australian Heat tolerance breeding value (HT GEBV) for use by farmers (Nguyen et al., 2017). These breeding values were standardised to have a mean of 100, such that if, for example, a bull has a value of 105, then its daughters are expected to be 5% more tolerant to heat than daughters of the average bull and vis versa for the bull with HT GEBV value of 95. The study reported promising accuracies ranging from around 40% to 60% in Holsteins and Jerseys (Nguyen et al., 2016). GEBVs for heat tolerance are now available to Australian dairy farmers to choose animals that can withstand heat stress (Nguyen et al., 2017, Pryce et al., 2018). Australian national selection index (Balanced Performance Index, BPI) favours production and does not currently include heat tolerance traits. However, it is worth considering this trait in the selection index to improve resilience.

While the accuracies of GEBVs reported by (Nguyen et al., 2016) are encouraging and adequate, higher values are desirable such as those for production traits with estimates as high as 0.80 (e.g., recent data from DataGene; https://datagene.com.au/; accessed October 2021). Even a small increase in prediction accuracy is a bonus to the dairy industry since genetic improvement is linearly correlated with the selection accuracy component of the mathematical formula for determining genetic change (described earlier). One way to boost the prediction accuracy of heat tolerance is to enlarge the size of the reference population used for estimating SNP effects and then develop a prediction equation for calculating GEBVs. Initiatives are now being undertaken in Australia to increase the size of the female reference population through the Genomic Information Nucleus (Ginfo) project (Pryce et al., 2018). Another way to increase the accuracy of prediction is to leverage the availability of dense molecular makers. This option is strongly influenced by research, thanks to the availability of tens of thousands of genetic variants following recent advancements in genotyping technologies. To maximise benefits from the genetic variants, thousands of animals with phenotypes must be genotyped at reasonable depth so that the variants are in strong or perfect LD with the causal mutations for a trait. It is still quite expensive to sequence thousands of animals to the whole-genome sequence level. However, the 1000 Bull Genome project (Hayes and Daetwyler, 2019) is currently an excellent resource to impute animals with phenotypes that were genotyped with

lower density SNP chip to a whole-genome sequence level. To facilitate imputation, the Run7 of the 1000 Bull Genome project currently hosts a collection of reference datasets for over 3,000 cattle from multiple collaborators around the world (Hayes and Daetwyler, 2019).

Several studies have tested the utility of imputed whole-genome sequence in improving the accuracy of prediction for various traits in cattle (Calus et al., 2016), sheep (Moghaddar et al., 2019), and avian species (Heidaritabar et al., 2016). However, these studies have found little or no appreciate gains in accuracy when lower-density SNP arrays (50k or HD panels) were replaced with the full set of imputed whole-genome sequence variants (WGS). Besides, using a full set of WGS data for genomic predictions come with a huge computational cost and may not be feasible in some cases. An alternative approach in which a substantial increase in the prediction estimates has been reported in both simulation and real studies is by carefully selecting causal or predictive sequence variants and then adding them to the lower density SNP panels (50k or HD panels) (Brøndum et al., 2015, van den Berg et al., 2016, Al Kalaldeh et al., 2019, Moghaddar et al., 2019). For example, an increase in prediction accuracies of up 25% has been reported for some complex traits such as parasite resistance in sheep based on this approach (Al Kalaldeh et al., 2019). A critical first step in this approach is to perform an association or fine-mapping study to identify and prioritise genetic variants for a trait (s). The following sections review previous genome-wide association studies (GWAS) to identify genetic variants associated with heat tolerance in dairy cattle.

2.9. Genome-wide association studies (GWAS) for heat tolerance traits

GWAS exploits linkage disequilibrium (LD) to identify genetic markers that tag causal mutations for a trait. While numerous GWAS have been conducted for various traits in livestock, specific studies aimed at searching for genes that contribute to heat tolerance in animals are very scarce, implying that the genetic architecture of this trait is still not well characterised.

Hayes et al. (2009) performed GWAS to identify QTL associated with milk yield response to heat stress in Australian Holstein-Friesian bulls (n = 781). The study estimated slope traits based on sires' daughter yield deviation (n = 62,343) as a function of THI. The study reported 42 significant SNPs associated with response to heat stress (THI slope), of which one SNP (BFGL-NGS-30169) on chromosome 29 at 48329079 was independently validated in Holstein and Jersey breeds. Moreover, these authors reported three significant SNPs that were validated only in Holstein cattle: BFGL-NGS-139 (BTA 8), BFGL-NGS-89500 (BTA 10), and

BFGL-BAC-38208 (BTA 25). The candidate region at BTA 29 harbours the *FGF4* gene associated with thermotolerance (Hayes et al., 2009). In a later study in Australia, Wang et al. (2017) found at least 14 candidate genes associated with milk production (milk, fat, and protein) decline under heat stress, including *YBEY, SERPINE2, HSF1, STIP1,* and *CACNA1D* genes.

Dikmen et al. (2013) performed GWAS using rectal temperature (RT) for 1,440 Holstein cows in the USA and THI. Analysis was performed based on the SNP windows ranging from 2 to 10 adjacent SNPs. The largest proportion of SNP variance (0.07% - 0.44%) explained by markers was located between 28,877,547 and 28,907,154 bp on BTA 24 (Dikmen et al., 2013). This region harbours the *U1* spliceosomal RNA (U1) and *NCAD* cadherin-2 genes. Other top-ranking SNPs associated with RT based on 5-SNP window analysis reported by these authors include SNPs on BTA 5 at position 89512928; BTA 26 at position 20259486; BTA 16 at 35230105 and BTA 4 at 64386271.

In a follow-up study, Dikmen et al. (2015) investigated whether they could replicate candidate thermotolerant SNPs identified in the previous GWAS studies (Hayes et al., 2009, Dikmen et al., 2013). Also, several well-known candidate thermotolerant SNP genes (*ATPA1A, HSP70A, HSP90AB, PPARA*) were investigated. They used three heat tolerance traits: rectal temperature, respiratory rate, and sweating rate. The study confirmed six candidate SNPs from previous GWAS and identified new putative markers (n = 20) associated with heat tolerance traits. For example, the BFGL-NGS-30169 SNP validated by Hayes et al. (2009) in Jerseys and Holsteins was found to be significantly associated with sweating rate. The small number of heat tolerance SNPs replicated in this study was attributed to the low reliability of the current GWAS studies, small sample sizes used in the earlier GWAS, and low heritability of heat tolerance traits.

Howard et al. (2014) performed a GWAS in crossbred steers and heifers to identify genomic regions associated with heat and cold stress. Heat tolerance trait was defined as averages of hourly tympanic and vaginal temperature recordings during summer and winter. Comprehensive Climate Index (CCI) as a measure of heat load was derived based on the recording of ambient temperature, relative humidity, wind speed, and solar radiation. The study identified several genes and functional pathways associated with heat and cold stress, including apoptosis (*RIPK1*), pentose phosphate pathway (*FBP1* and *FBP2*), ion regulation (*PRKCB* and *CACNG3*), body weight and feed intake (*NBEA*), heat shock protein response (*HSPH1*),

generation of precursor metabolites and energy (*COX7C*), (Howard et al., 2014). Notably, these authors found that different sets of genetic markers appear to control the extremes of temperatures (cold versus heat), indicating the possibility of breeding for robustness against heat stress.

Macciotta et al. (2017) followed a similar approach as Hayes et al. (2009) to identify SNP associated with heat tolerance. Heat tolerant traits were derived from daughter trait deviations (DTD) of 423 bulls based on the principal component analysis of milk yield traits. The first principal component (PC) was interpreted as the intercept and the second component was the slope as in Hayes et al. (2009) (i.e., the rate of milk decline with increasing variability in heat stress). The authors reported 8 significant SNPs, of which 2 (ARS-BFGL-NGS-29678 and Hapmap32110-BTA-153952) in BTA 6 and 26 were associated with the response of milk yield to heat stress. One SNP (ARS-BFGL-NGS-19275) associated with the response of protein content to heat stress was mapped to BTA 6. None of the significant SNPs identified in previous studies were replicated in this study.

Recently, Otto et al. (2019) conducted a GWAS in Gir × Holstein crossbred cattle in Brazil using measures of rectal temperature to describe heat tolerance and the 50k SNP panel. These authors reported several genes for heat tolerance, including *LIF* (Leukemia inhibitory factor), *OSM* (oncostatin M), *DGCR8* (*DGCR8* microprocessor complex subunit), and *TXNRD2* (thioredoxin reductase 2) genes. Some of these genes are involved in the regulation of heat shock proteins (HSP) genes and oxidative stress in animals. In another study in the Florida-US, Sigdel et al. (2019) used the rate of milk decay as a function of THI to describe heat tolerance as in (Hayes et al., 2009) and the 50k SNP panel in their GWAS and found at least three candidate genes including *MAPK8IP1, HSF1,* and *CDKN1B* in BTA 5, 14, 15, respectively. *HSF1* gene has been repeatedly linked with thermal stress in various studies in cattle (Li et al., 2011, Wang et al., 2017, Rong et al., 2019, Garner et al., 2020). Also, Sigdel and colleagues conducted a GWAS on reproductive performance (conception rates) in US dairy cows and found several candidate genes involved with fertility functions under heat stress, including *BRWD1, EXD2, ADAM20, EPAS1, TAOK3*, and *NOS1* (Sigdel et al., 2020). More recently, Luo et al. (2021) did a GWAS for morning and afternoon rectal temperature in Chinese Holstein cattle and found 10 significant SNPs mapping on chromosomes 3, 4, 8, 13, 14, and 29. Two candidate genes mapping to some of these SNPs (*FAM107B* and *PHRF1*)

were subsequently validated in an mRNA expression analysis of peripheral blood mononuclear cells.

Besides GWAS above, several genome-wide scan of signatures of selection (**SS**) have been conducted to find genes for thermotolerance in cattle (Taye et al., 2017, Li et al., 2020b, Freitas et al., 2021). Taye et al. (2017) did SS in African indigenous breeds that are adapted to hot tropical climates and identified over 200 significant genes associated with heat tolerance. Some of the genes were reported to be involved in functions such as oxidative stress response, heat shock response, and sweat gland development. Similarly, Li et al. (2020b) performed SS in zebu breed (Dehong humped cattle) that is adapted to hot tropical conditions of southwestern China and detected several genes and pathways that contribute to heat tolerance: heat shock (*HSF1*) and oxidative stress response (*PLCB1, PLCB4*), coat color (*RAB31*), feed intake (*ATP8A1, SHC3*) and reproduction (*TP63, MAP3K13, PTPN4, PPP3CC, ADAMTSL1, SS18L1, OSBPL2, TOX, RREB1, GRK2*). Recently, Freitas et al. (2021) conducted SS in the various cattle breeds including Chinese local breeds and found multiple genes and pathways related to heat tolerance such as heat-shock proteins, oxygen transport, anatomical traits, mitochondrial DNA maintenance, metabolic activity, feed intake, carcass conformation, fertility, and reproduction. Overall, the large number of genes and diverse biological pathways identified in the above studies suggest complexity and polygenic nature of heat tolerance trait.

The SLICK locus, responsible for sleek hair coat in the Senepol cattle, has been strongly associated with thermotolerance (Olson et al., 2003, Dikmen et al., 2008, Dikmen et al., 2014). Previous association analysis revealed strong causal single mutations in prolactin (*PLR*) at the 10th exon of the prolactin receptor (*PRLR*) genes (Littlejohn et al., 2014). In a recent GWAS of the Limonero cattle, additional truncation mutations in the 11th exon of the *PRLR* was identified as responsible for the slick coat (Porto-Neto et al., 2018). The authors noted that these mutations could not explain all the variations in the slick coats of the study individuals suggesting the possibility more causal mutations. Apart from the SLICK gene polymorphisms, there are currently no other genetic variants that have been described to have clear and beneficial effects on heat tolerance in dairy cattle.

In summary, several key points emerge from the reviewed GWAS studies. Firstly, it is obvious there are limited GWAS studies for heat tolerance traits, most of which have been performed on Holstein breeds. Secondly, GWAS for heat tolerance seems to be underpowered as evidenced by the lack of repeatability of the results in the studies reviewed. Among the over

500 candidate variants for heat tolerance reported in various, very few could be confirmed in some later studies. Poor replication of these GWAS studies for heat tolerance could be related to several factors: 1) variation in sample sizes and genetic difference across populations used (e.g., the LD level); 2) multiple correction tests which limit the number of significant detections; 3) failure of the markers to capture genetic variation in the trait. For heat tolerance, a common problem that can likely contribute to poor repeatability is the heterogeneity of the methods as well as definitions of the traits used in GWAS studies. Variations in the calculations of environmental heat load (reviewed in section 2.3) can also have impact on the outcome of the GWAS for heat tolerance. The differences in the choice of the heat tolerance trait is also an issue in GWAS studies (Carabaño et al., 2017). For example, Hayes et al. (2009) and Macciotta et al. (2017) evaluated heat stress based on the decay of milk production whereas other GWAS studies have often used direct measurements of body temperature (e.g., Howard et al., 2014, Dikmen et al., 2015). Nevertheless, using a large sample size and high-resolution sequence variants has a potential to find important genetic variants or SNPs for heat tolerance. As sample size increases, the loci significantly associated with complex traits are expected to increase (Wood et al., 2014). In addition, the increased availability of high-resolution whole genome sequencing data in dairy cattle is likely to improve power and precision of understanding the genetic architecture of heat tolerance, as demonstrated for complex traits in humans such as Parkinson disease (Maraganore et al., 2005).

2.10. G × E in the context of heat tolerance

Genotype by environment interactions (G × E) exists when a trait value or the performance of genotypes varies across different environments (Falconer and MacKay, 1996). In this case, heat tolerance can present challenges in optimizing breeding across different climatic conditions or heat stress levels. For example, to optimise production across different production environments, the warmer subtropical regions of Australia may require a different set of heat-tolerant genotypes compared to cooler regions, such as Tasmania.

G × E for heat tolerance is of great importance both at regional and global dairy industries owing, in part, to the current extensive use of very few elite sires across diverse farming systems and climatic conditions within and between countries, facilitated mainly by Artificial insemination (AI). In Australia, (one of the largest countries in the world), dairying is distributed across a wide range of environments characterised by different factors, including production systems (e.g., pasture-based, and intensive), feeding level, herd (composition and

size), and climatic variables (e.g., regional differences in ambient temperature and relative humidity). Dairy farming is mostly located in temperate coastal areas predominantly around Victoria. Most dairy herds (~70 – 75%) are reared on pasture, with varying level of supplementation (Dairy Australia, 2017). The dairy herd of around 1.74 million cows is mainly composed of a wide array of breeds (Dairy Australia, 2017). The cow populations are composed of the Holsteins (~ 60%), Jersey (~ 11%), and their crosses (~ 5%) (DataGene, 2016, Dairy Australia, 2017). The national herd recorded cows is concentrated in Victoria's dairying region (63%), especially along coastal regions (Figure 2-10), while the rest come from New South Wales (14%), Queensland (7%), South Australia (7%), Tasmania (5%) and Western Australia (4%) (DataGene, 2016). There is considerable variation in the climatic conditions among the dairying regions in Australia (i.e., subtropical versus temperate regions), with dairy cows experiencing a varying level of heat stress in all regions (Figure 2-10). Generally, the average ambient temperature is the highest in Queensland and lowest in Tasmania. Seasonal variations in weather conditions are characterised by hot summers and moderate winters, meaning that animals are exposed to different environments during the year (Figure 2-10). Daily average temperature and humidity can vary substantially within and between dairy regions with temperatures in Northern Victoria, for example, ranging from -5 to $38\,^{\circ}C$ (Nguyen et al., 2016). Considering these factors shaping then Australian dairy landscape, $G \times E$ due to heat stress may be relevant and needs to be investigated and accounted for in breeding programmes.
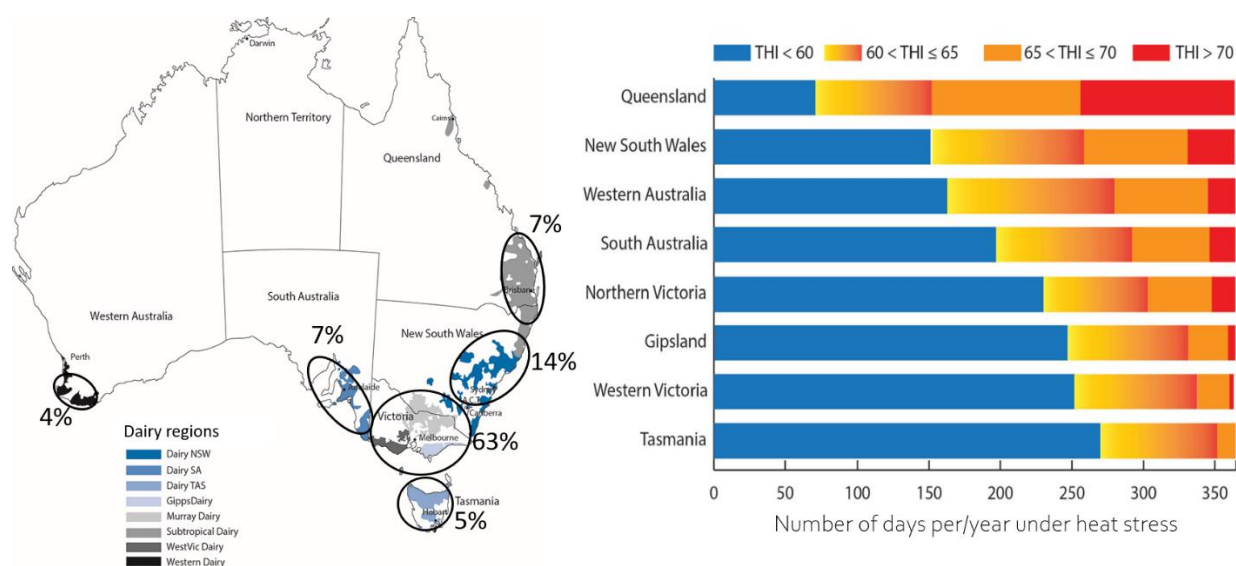

Figure 2-10 Average number of days per year with different range of temperature-humidity index (THI): THI < 60, 60 < THI ≤ 65, 65 < THI ≤ 70, and THI ≥ 65 (Nguyen et al., 2016) across different

dairying regions in Australian (Dairy Australia, 2020). The percentage values represent the proportion of herds in major dairy regions based on data from DataGene (DataGene, 2016).

The genotypes (G) compared in previous G × E studies for heat tolerance in dairy cattle included breeds and individuals within breeds (Hayes et al., 2003, Bryant et al., 2007a, Bohmanova et al., 2008, Haile-Mariam et al., 2008, Carabaño et al., 2014, Hammami et al., 2015, Santana Jr et al., 2017) and SNP genotypes (Hayes et al., 2009). The commonly used environmental descriptor (E) in G × E studies for heat tolerance is the THI.

The magnitude of G × E is often quantified by genetic correlations of a trait(s) between environments, with values ranging from zero to unity. The higher the genetic correlations, the lower the G × E and vis versa. Typically, production traits have a higher genetic correlation (i.e., smaller G × E) than functional traits such as fertility (Mulder et al., 2006). G × E can take 2 major forms (Figure 2-11): a) rank-change G × E, where genotypes are ranked differently in different environments b) scaling G × E, where the genotypic performance varies across different environments, but no rank change occurs (Falconer and MacKay, 1996). The presence of rank-change G × E is challenging in animal breeding if the objective is to optimise genetic improvement across diverse production environments. As a guiding principle, if the genetic correlation between two environments is lower than 0.80 then it may be necessary to have separate breeding programmes for each environment (Mulder et al., 2006). In the case of heat stress, for example, one environment could be Tasmania (with relatively lower heat loads) and another environment could be Queensland (with relatively higher heat loads).
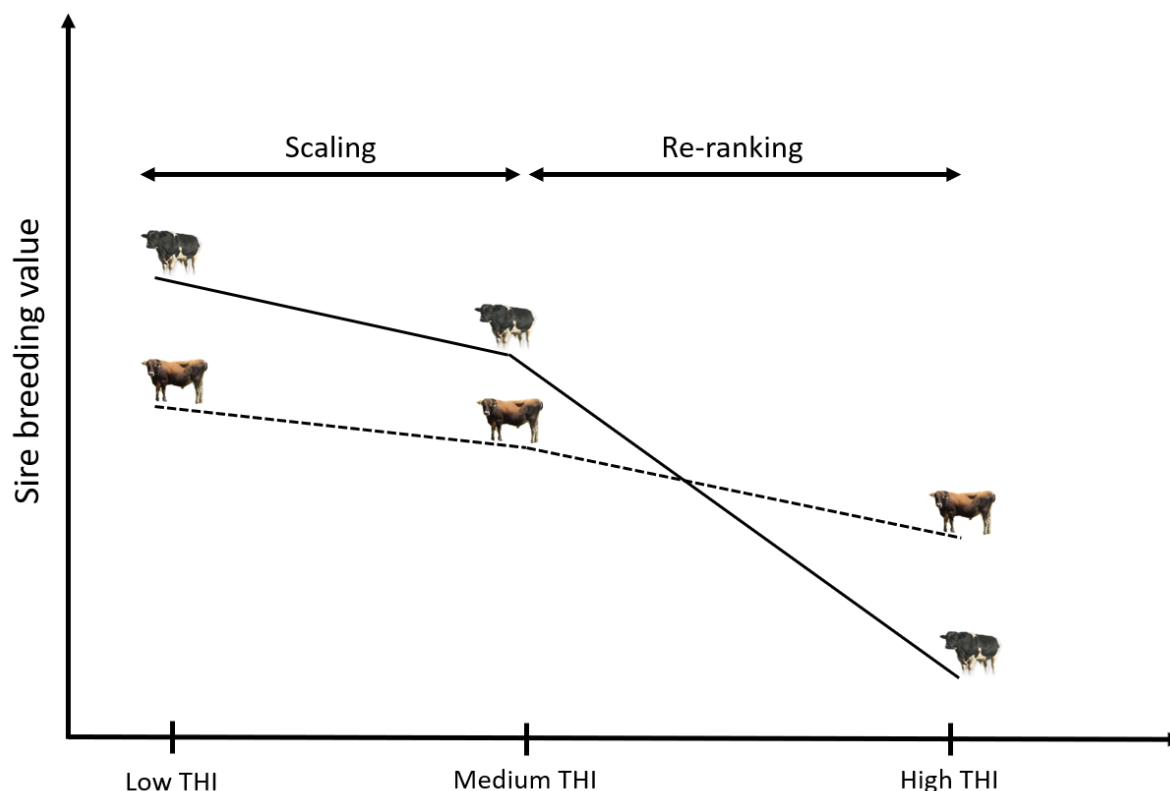
Figure 2-11 Schematic representation of scaling and re-ranking G × E for two sires across three different temperature-humidity index (THI) environments.

G × E resulting from heat stress for various traits has been previously investigated in Australia. Hayes et al. (2003) found minimal G × E, with the genetic correlation of 0.90 for protein yield at the 5th and 95th percentiles of THI in Australian Holstein-Friesian cows (these percentiles represent low and high heat environments). Similarly, Haile-Mariam et al. (2008) reported considerable G × E for fertility traits, with genetic correlations of 0.79 between the 5th and 95th percentile of THI. These studies indicate a lack of significant re-ranking of Australian sires resulting from heat stress for the traits investigated. However, larger effects of G × E in dairy cattle have been reported in other countries. For example, studies in Spain and Belgium observed significantly large G × E causing rank-change of sires for fat yield, protein yield, and somatic cell count (SCC), with genetic correlations < 0.70 between hot and cold temperatures (Carabaño et al., 2014, Hammami et al., 2015).

The Australian dairy industry, through the dairy genetic evaluation unit of DataGene, was the first country in the world to release the genomic breeding values for heat tolerance in dairy cattle following the work of Nguyen et al. (2016) and Garner et al. (2016) (discussed earlier). The question now remains whether these genotypes will perform optimally across

different production environments, given the diverse climatic conditions in Australia. If significant $G \times E$ exists, then the genetic gain that will be achieved will be lower-than-expected when selection is largely based on an environment where heat stress is less challenging.

While some $G \times E$ studies on heat tolerance have been carried out in Australia (Hayes et al., 2003, Haile-Mariam et al., 2008), there are several other compelling reasons to warrant additional studies. Firstly, the previous $G \times E$ work on heat tolerance in Australia and other countries have focused on Holsteins and to a very limited extent on Jersey and other cattle breeds, presumably due to insufficient data. Secondly, although the effects of heat stress on fertility traits are well documented (reviewed by Das et al., 2016), the differences in the response to the environments among different genotypes remains poorly understood. Thirdly, besides the availability of GEBV for heat tolerance, it is possible that more genotypes (sires) and data per sires across different environments have now grown to facilitate better and more accurate estimates of $G \times E$ with greater reliability. The greater the sample size, the greater the likelihood of including unusual reaction norms, hence larger $G \times E$. It is also likely that a larger magnitude of $G \times E$ has occurred following continued selection for production traits. Fourth, the availability of novel phenotypes, such as mid-infrared predicted milk biomarkers (Liu et al., 2017), may provide an opportunity to scale up measurements of heat stress, thus better estimates of $G \times E$ for heat tolerance. Moreover, the increasing trend of warming environments from climate change suggest that $G \times E$ for heat tolerance may have grown over the last decade in Australia.

2.11. **Statistical models used to quantify heat tolerance and $G \times E$**

The models used for describing heat tolerance are the multi-trait model and the reaction norm model. The choice of the two models is often dictated by the environment descriptors. The review of each model follows.

2.11.1. **Multi-trait model to estimate $G \times E$**

Multi-trait models describe the genotype as a function of a discrete environmental descriptor. In this model, traits in different environments are considered different but correlated traits (Falconer, 1952). A typical example of multi-trait model analysis is the Interbull concept of expressing many traits evaluated in different countries as separate but genetically correlated traits (De Jong and Bijma, 2002). Several studies have used multi-trait models for investigating heat tolerance where the environmental descriptors are based on THI as follows: THI at the 5th and 95th percentile (Hayes et al., 2003) or based on THI groups (no HS (THI < 72) and moderate

HS (THI $\geq$ 79) (Hagiya et al., 2017)). Also, the environmental descriptor can be grouped into geographic regions defined by different climatic conditions.

A simple multi-trait mixed model for two environments (two traits) can be represented as an equation as follows: $y_i = Xb_i + Zu_i + e_i$, and in matrix notation:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix},$$

where y_1 and y_2 are traits records in environment 1 and 2, respectively; b_1 and b_1 are vectors for fixed effects; u_1 and u_2 are vectors for random effects in environment 1 and 2, respectively; X_1, X_2, Z_1, and Z_2 are incidence matrices; e_1 and e_2 are vectors for random residuals. The drawback to the model is that it groups an arbitrary number of environments which must be limited to meet computational demands required for estimating (co)variance components and the convergence of the parameters.
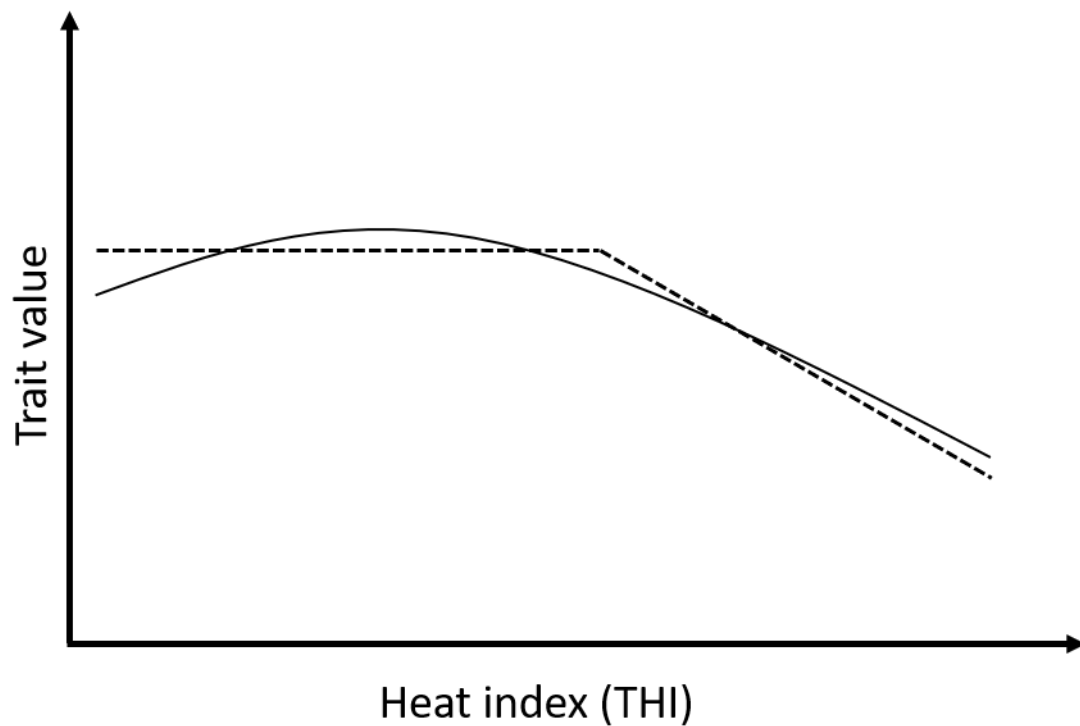
2.11.2. **Reaction norm models for quantifying heat tolerance and G $\times$ E**

The reaction norm model, or the norm of reaction, is the most common and widely used model for investigating heat tolerance and G $\times$ E in dairy cattle and other livestock species. In these models, the pattern of the phenotypic expression of a genotype (e.g., milk production traits across different environments) is regressed as a function of the environmental heat load (e.g., THI) (Ravagnolo and Misztal, 2000, Ravagnolo et al., 2000). Due to its simplicity, THI is used in most studies as an indicator of environmental heat load.

Dairy animals often reduce their production performance when THI increases. As such, studies have proposed various reaction norm models to quantify heat tolerance by combining milk phenotypes and THI. The commonly used model defines the THI threshold at which production begins to decline and the rate of linear decline (Ravagnolo and Misztal, 2000), referred to as the broken line model (BL) (Figure 2-12). For this model, and in the case of heat tolerance, the intercept solutions traits represent the level of milk production while the slopes (are the rate of milk decline) are the heat tolerance values for the animals. This model is appealing due to its direct biological interpretation of heat tolerance; that is, the heat-tolerant animal will have a higher threshold and the smaller the rate of decline of milk traits (slopes). Most studies using BL often assume a common threshold for all individuals and quantifies heat tolerance based on the slope parameter. However, there is a concern of possible variation in thresholds among individuals, which can bias the estimates when the common threshold is

assumed (Sánchez et al., 2009a). Moreover, the BL model assumption of the linearity of the production decline after the threshold is unrealistic in some cases (Bernabucci et al., 2014). Sánchez et al. (2009b) proposed a hierarchical model which accounts for variation in individual thresholds and the slope. The authors found that the individual THI slope and threshold are strongly and positively correlated (> 0.95) in terms of environmental and genetic aspects, implying that selection based on them will lead to change in the counterpart in the same direction (i.e., the higher the threshold leads to greater sensitivity (slope)) (Sánchez et al., 2009a). The drawback to this model is its complexity resulting in slow mixing and convergence rates (Sánchez et al., 2009a). The authors found that constraining the threshold to a common estimated value is still required to obtain reasonable estimates for genetic variation for heat tolerance.

The Legendre polynomials functions in random regression models have also been applied to describe the trajectory pattern of heat tolerance traits to increasing heat load (Figure 2-12). Brügemann et al. (2012) used a third-order Legendre polynomial to describe the effect of heat load on protein yield in dairy cows. Similarly, Carabaño et al. (2014) used cubic Legendre polynomial to describe heat stress for production traits and somatic cell count. Unlike BL models, parameters for Legendre polynomial reaction norms lack direct biological interpretation, except for the intercept. In addition, the drawback to Legendre polynomials is the lack of asymptotes which can lead to an unrealistic fitting at the extremes of the x-axis, especially when the data is sparse at the extremes of the distribution as reported in the work of Brügemann et al. (2013). Unexpected trajectories at the extremes of the environmental descriptor due to few data points is also an issue when using BL models (e.g., Misztal et al., 2000).

Figure 2-12 Schematic of broken line model (dashed line) and higher order Legendre Polynomial (solid line) reaction norm models for individual genetic response to heat stress (adapted from Carabaño et al., 2017).

In addition to quantifying heat tolerance (defined as the slope of the decay in production under thermal heat load), reaction norm models have been used to describe G × E for heat tolerance across environments. If the reaction norm of an individual is horizontal, then the performance does not vary across environments and is said to have low sensitivity to the environment examined (Rauw and Gomez-Raya, 2015, Friggens et al., 2017). Conversely, if the slope of the reaction norm is not constant, then it implies that the performance of the genotypes changes according to the environmental conditions; the steeper slope for an individual is indicative of greater environmental sensitivity or G × E. Daughters of sires with steep slope are likely to produce less under heat stress conditions and therefore can be used in regions with the consistently low heat load.

2.12. **Summary**

Heat stress is an issue of growing concern for livestock industries worldwide. Heat in dairy cows negatively impacts productivity, and welfare, causing huge losses running into millions of dollars in the livestock industries (St-Pierre et al., 2003). Climate changes characterised by severe heat events are predicted to continue in the coming decades (BOM, 2020), meaning that even more economic losses are likely to occur in the future if proactive and long-term solutions are not sought to minimize the impacts of the warming climates.

Australia's unique dairy landscape characterised by diverse climate conditions and highly heterogeneous farming systems suggests that G × E due to heat stress may be an issue of concern. G × E occurs when the performance of animal differs in different environments and presents a challenge when the breeding goal is to optimise genetic gain across environments. In addition, Australian dairy herds are predominantly reared on pasture, meaning that management strategies such as shades and sprinklers may be cumbersome and costly to implement for some farmers. Therefore, genetic selection for heat-tolerant animals offers a promising and long-term solution in addition to management strategies. Australia is currently at the forefront globally regarding estimating the genetic merit of heat tolerance aimed at helping farmers to cope with the warming climates while maintaining productivity (Nguyen et al., 2016). However, more work is still needed to improve the current genomic evaluations for heat tolerance in dairy cattle. One way to do this is to identify specific genetic variants that confer thermotolerance to animals and include them in the genetic evaluation SNP panels or introduce them to heat-susceptible animals through approaches such as gene editing. Except for the mutations in the SLICK locus in the Senepol cattle (Olson et al., 2003, Dikmen et al., 2014, Littlejohn et al., 2014), other variants with clear effects on heat tolerance in other dairy breeds, e.g., Holsteins and Jersey have not been well characterised and tested in genomic predictions.

Therefore, the following three general hypotheses were formulated to investigate the above research questions:

1) "that G × E due to heat stress for production traits in Australian dairy cattle is not substantial to warrant forming separate genetic evaluations for colder and warmer environments."

2) "that the large phenotype datasets and extensive genomic information available for Australian dairy cattle provide sufficient power to pinpoint specific candidate causal variants and biological mechanisms that confer thermotolerance to animals."

3) "that the candidate genetic variants identified in 'hypothesis 2' above are beneficial for improving genomic prediction for heat tolerance in dairy cattle".

2.13. **References**

Aguilar, I., I. Misztal, and S. Tsuruta. 2009. Genetic components of heat stress for dairy cattle with multiple lactations. Journal of Dairy Science 92(11):5702-5711.

Al Kalaldeh, M., J. Gibson, N. Duijvesteijn, H. D. Daetwyler, I. MacLeod, N. Moghaddar, S. H. Lee, and J. H. van der Werf. 2019. Using imputed whole-genome sequence data to improve the accuracy of genomic prediction for parasite resistance in Australian sheep. Genetics Selection Evolution 51(1):32.

Ames, D. 1980. Thermal environment affects production efficiency of livestock. BioScience 30(7):457-460.

Armstrong, D. 1994. Heat stress interaction with shade and cooling. Journal of Dairy Science 77(7):2044-2050.

Bar, D., M. Kaim, I. Flamenbaum, B. Hanochi, and R. L. Toaff-Rosenstein. 2019. Accelerometer-based recording of heavy breathing in lactating and dry cows as an automated measure of heat load. Journal of Dairy Science 102(4):3480-3486.

Baumgard, L., M. Abuajamieh, S. Stoakes, M. Sanz-Fernandez, J. Johnson, R. Rhoads, and M. Eastridge. 2014. Feeding and managing cows to minimize heat stress. Proceedings of the 23rd Tri-State Dairy Nutrition Conference, Fort Wayne, Indiana, USA, p14-16.

Baumgard, L., J. Wheelock, S. Sanders, C. Moore, H. Green, M. Waldron, and R. Rhoads. 2011. Postabsorptive carbohydrate adaptations to heat stress and monensin supplementation in lactating Holstein cows. Journal of Dairy Science 94(11):5620-5633.

Baumgard, L. H. and R. P. Rhoads Jr. 2013. Effects of heat stress on postabsorptive metabolism and energetics. Annual Review Animal Biosciences 1(1):311-337.

Berman, A. 2011. Invited review: Are adaptations present to support dairy cattle productivity in warm climates? Journal of Dairy Science 94(5):2147-2158.

Bernabucci, U., S. Biffani, L. Buggiotti, A. Vitali, N. Lacetera, and A. Nardone. 2014. The effects of heat stress in Italian Holstein dairy cattle. Journal of Dairy Science 97(1):471-486.

Bernabucci, U., N. Lacetera, L. H. Baumgard, R. P. Rhoads, B. Ronchi, and A. Nardone. 2010. Metabolic and hormonal acclimation to heat stress in domesticated ruminants. Animal 4(7):1167-1183.

Bianca, W. 1962. Relative Importance of Dry-and Wet-bulb Tem peratures in Causing Heat Stress in Cattle. Nature 195:251-252.

Biffani, S., U. Bernabucci, A. Vitali, N. Lacetera, and A. Nardone. 2016. Effect of heat stress on nonreturn rate of Italian Holstein cows. Journal of Dairy Science 99(7):5837-5843.

Bohmanova, J., I. Misztal, and J. Cole. 2007. Temperature-humidity indices as indicators of milk production losses due to heat stress. Journal of Dairy Science 90(4):1947-1956.

Bohmanova, J., I. Misztal, S. Tsuruta, H. Norman, and T. Lawlor. 2008. Genotype by environment interaction due to heat stress. Journal of Dairy Science 91(2):840-846.

BOM. 2020. State of the Climate 2020. http://www.bom.gov.au/state-of-the-climate/australias-changing-climate.shtml.

BOM and CSIRO. 2015. Climate-ready Victoria: Gippsland. https://www.climatechange.vic.gov.au/__data/assets/pdf_file/0021/60744/Gippsland.pdf.

Boonkum, W., I. Misztal, M. Duangjinda, V. Pattarajinda, S. Tumwasorn, and J. Sanpote. 2011. Genetic effects of heat stress on milk yield of Thai Holstein crossbreds. Journal of Dairy Science 94(1):487-492.

Boulant, J. A. 2000. Role of the preoptic-anterior hypothalamus in thermoregulation and fever. Clinical Infectious Diseases 31(Supplement_5):S157-S161.

Bourdon, R. M. 2000. Understanding animal breeding. Prentice Hall Upper Saddle River, NJ.

Britt, J., R. Cushman, C. Dechow, H. Dobson, P. Humblot, M. Hutjens, G. Jones, P. Ruegg, I. Sheldon, and J. Stevenson. 2018. Invited review: Learning from the future—A vision for dairy farms and cows in 2067. Journal of Dairy Science 101(5):3722-3741.

Brøndum, R., G. Su, L. Janss, G. Sahana, B. Guldbrandtsen, D. Boichard, and M. Lund. 2015. Quantitative trait loci markers derived from whole genome sequence data increases the reliability of genomic prediction. Journal of Dairy Science 98(6):4107-4116.

Bru, J., P. Berbigier, and S. Sophie. 1987. Estimation of sweat rate and thermal tolerance of pure Creole and of Limousin×Creole crossbred growing bulls in Guadeloupe (French West Indies). International Journal of Biometeorology 31(1):77-84.

Brügemann, K., E. Gernand, U. König von Borstel, and S. König. 2012. Defining and evaluating heat stress thresholds in different dairy cow production systems. Archives Animal Breeding 55(1):13-24.

Brügemann, K., E. Gernand, U. Von Borstel, and S. König. 2011. Genetic analyses of protein yield in dairy cows applying random regression models with time-dependent and temperature x humidity-dependent covariates. Journal of Dairy Science 94(8):4129-4139.

Brügemann, K., E. Gernand, U. von Borstel, and S. König. 2013. Application of random regression models to infer the genetic background and phenotypic trajectory of binary conception rate by alterations of temperature×humidity indices. Livestock Science 157(2-3):389-396.

Bryant, J., N. Lopez-Villalobos, J. Pryce, C. Holmes, D. Johnson, and D. Garrick. 2007a. Environmental sensitivity in New Zealand dairy cattle. Journal of Dairy Science 90(3):1538-1547.

Bryant, J., N. López-Villalobos, J. Pryce, C. Holmes, and D. Johnson. 2007b. Quantifying the effect of thermal environment on production traits in three breeds of dairy cattle in New Zealand. New Zealand Journal of Agricultural Research 50(3):327-338.

Buckley, F., N. Lopez-Villalobos, and B. Heins. 2014. Crossbreeding: implications for dairy cow fertility and survival. Animal 8(s1):122-133.

Calus, M. P., A. C. Bouwman, C. Schrooten, and R. F. Veerkamp. 2016. Efficient genomic prediction based on whole-genome sequence data using split-and-merge Bayesian variable selection. Genetics Selection Evolution 48(1):1-19.

Carabaño, M., K. Bachagha, M. Ramón, and C. Díaz. 2014. Modeling heat stress effect on Holstein cows under hot and dry conditions: Selection tools. Journal of Dairy Science 97(12):7889-7904.

Carabaño, M., M. Ramón, C. Díaz, A. Molina, M. Pérez-Guzmán, and J. Serradilla. 2017. BREEDING AND GENETICS SYMPOSIUM: breeding for resilience to heat stress effects in dairy ruminants. A comprehensive review. Journal of Animal Science 95(4):1813-1826.

Carabaño, M. J., M. Ramón, A. Menéndez-Buxadera, A. Molina, and C. Díaz. 2019. Selecting for heat tolerance. Animal Frontiers 9(1):62-68.

Cole, J. and P. VanRaden. 2018. Symposium review: Possibilities in an age of genomics: The future of selection indices. Journal of Dairy Science 101(4):3686-3701.

Collier, R. J. and J. Collier. 2011. Environmental physiology of livestock. Wiley-Blackwell, New Jersey.

Collier, R. J. and K. G. Gebremedhin. 2015. Thermal biology of domestic animals. Annual Review of Animal Biosciences 3(1):513-532.

Collier, R. J., B. J. Renquist, and Y. Xiao. 2017. A 100-Year Review: Stress Physiology including heat stress. Journal of Dairy Science 100(12):10367-10380.

Dado-Senn, B., A. L. Skibiel, T. F. Fabris, Y. Zhang, G. E. Dahl, F. Peñagaricano, and J. Laporta. 2018. RNA-Seq reveals novel genes and pathways involved in bovine mammary involution during the dry period and under environmental heat stress. Scientific Reports 8(1):1-11.

Dairy Australia. 2017. Strategic plan 2016/17 to 2018/19. https://www.dairyaustralia.com.au/-/media/dairyaustralia/documents/about-dairy-australia/key-publications/201610-da-3-year-strategic-plan.ashx?la=en&hash=7CF049652AD10BE3EB04B7B51EAA6CB739795746.

Dairy Australia. 2020. Climate Change Strategy 2020–25. https://www.dairyaustralia.com.au/land-water-and-climate/climate-change-and-dairy#.YG_JUPkzYX8.

DairyBio. 2018. Annual progress report 2016–17. http://dairybio.com.au/wp-content/uploads/2018/03/DairyBio-Annual-Progress-Report-2016-17.pdf.

Das, R., L. Sailo, N. Verma, P. Bharti, J. Saikia, Imtiwati, and R. Kumar. 2016. Impact of heat stress on health and performance of dairy animals: A review. Veterinary World 9(3):260-268.

DataGene. 2016. Australian dairy herd improvement report. https://datagene.com.au/sites/default/files/Upload%20Files/Australian%20Dairy%20Herd%2 0Improvement%20%20Report%202016.pdf.

Davis, S., R. Spelman, and M. Littlejohn. 2017. Breeding and genetics symposium: Breeding heat tolerant dairy cattle: The case for introgression of the "slick" prolactin receptor variant into Bos taurus dairy breeds. Journal of Animal Science 95(4):1788-1800.

De Jong, G. and P. Bijma. 2002. Selection and phenotypic plasticity in evolutionary biology and animal breeding. Livestock Production Science 78(3):195-214.

De Rensis, F., I. Garcia-Ispierto, and F. López-Gatius. 2015. Seasonal heat stress: Clinical implications and hormone treatments for the fertility of dairy cows. Theriogenology 84(5):659-666.

De Vries, A. and C. A. Risco. 2005. Trends and seasonality of reproductive performance in Florida and Georgia dairy herds from 1976 to 2002. Journal of Dairy Science 88(9):3155-3165.

Defra. 2021. Agricultural Facts England Regional Profiles. Deapartment of Environment Food & Agricultural Affairs, UK.

Dikmen, S., E. Alava, E. Pontes, J. Fear, B. Dikmen, T. Olson, and P. Hansen. 2008. Differences in thermoregulatory ability between slick-haired and wild-type lactating Holstein cows in response to acute heat stress. Journal of Dairy Science 91(9):3395-3402.

Dikmen, S., J. Cole, D. Null, and P. Hansen. 2012. Heritability of rectal temperature and genetic correlations with production and reproduction traits in dairy cattle. Journal of Dairy Science 95(6):3401-3405.

Dikmen, S., J. B. Cole, D. J. Null, and P. J. Hansen. 2013. Genome-wide association mapping for identification of quantitative trait loci for rectal temperature during heat stress in Holstein cattle. PloS ONE 8(7):e69202.

Dikmen, S. and P. Hansen. 2009. Is the temperature-humidity index the best indicator of heat stress in lactating dairy cows in a subtropical environment? Journal of Dairy Science 92(1):109-116.

Dikmen, S., F. Khan, H. Huson, T. Sonstegard, J. Moss, G. Dahl, and P. Hansen. 2014. The SLICK hair locus derived from Senepol cattle confers thermotolerance to intensively managed lactating Holstein cows. Journal of Dairy Science 97(9):5508-5520.

Dikmen, S., X. z. Wang, M. S. Ortega, J. Cole, D. Null, and P. Hansen. 2015. Single nucleotide polymorphisms associated with thermoregulation in lactating dairy cows exposed to heat stress. Journal of Animal Breeding and Genetics 132(6):409-419.

Eggen, A. 2012. The development and application of genomic selection as a new breeding paradigm. Animal Frontiers 2(1):10-15.

Falconer, D. and T. MacKay. 1996. Introduction to quantitative genetics. Pearson Education, India.

Falconer, D. S. 1952. The problem of environment and selection. The American Naturalist 86(830):293-298.

FAO. 2019. Climate change and the global dairy cattle sector – The role of the dairy sector in a low-carbon future. http://www.fao.org/3/CA2929EN/ca2929en.pdf.

Finocchiaro, R., J. Van Kaam, B. Portolano, and I. Misztal. 2005. Effect of heat stress on production of Mediterranean dairy sheep. Journal of Dairy Science 88(5):1855-1864.

Fodor, N., A. Foskolos, C. F. Topp, J. M. Moorby, L. Pásztor, and C. H. Foyer. 2018. Spatially explicit estimation of heat stress-related impacts of climate change on the milk production of dairy cows in the United Kingdom. PloS ONE 13(5):e0197076.

Freitas, M., I. Misztal, J. Bohmanova, and R. Torres. 2006. Regional differences in heat stress in US Holsteins. Proceedings of the 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, Minas Gerais, Brazil, 13-18 August, 2006. Instituto Prociência, p1-11.

Freitas, P. H., Y. Wang, P. Yan, H. R. Oliveira, F. S. Schenkel, Y. Zhang, Q. Xu, and L. F. Brito. 2021. Genetic Diversity and Signatures of Selection for Thermal Stress in Cattle and Other Two Bos Species Adapted to Divergent Climatic Conditions. Frontiers in Genetics 12:102.

Friggens, N., F. Blanc, D. Berry, and L. Puillet. 2017. Deciphering animal robustness. A synthesis to facilitate its use in livestock breeding and management. Animal 11(12):2237-2251.

Fuquay, J. 1981. Heat stress as it affects animal production. Journal of Animal Science 52(1):164-174.

Gantner, V., T. Bobic, R. Gantner, M. Gregic, K. Kuterovac, J. Novakovic, and K. Potocnik. 2017. Differences in response to heat stress due to production level and breed of dairy cows. International Journal of Biometeorology 61(9):1675-1685.

García-Ruiz, A., J. B. Cole, P. M. VanRaden, G. R. Wiggans, F. J. Ruiz-López, and C. P. Van Tassell. 2016. Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. Proceedings of the National Academy of Sciences 113(28):E3995-E4004.

Garner, J., A. Chamberlain, C. Vander Jagt, T. Nguyen, B. Mason, L. Marett, B. Leury, W. Wales, and B. Hayes. 2020. Gene expression of the heat stress response in bovine peripheral white blood cells and milk somatic cells in vivo. Scientific Reports 10(1):1-12.

Garner, J., M. Douglas, S. O. Williams, W. Wales, L. Marett, T. Nguyen, C. Reich, and B. Hayes. 2016. Genomic selection improves heat tolerance in dairy cattle. Scientific Reports 6(1):1-9.

Gaughan, J., S. Bonner, I. Loxton, and T. Mader. 2013. Effects of chronic heat stress on plasma concentration of secreted heat shock protein 70 in growing feedlot cattle. Journal of Animal Science 91(1):120-129.

Gaughan, J., T. Mader, S. Holt, M. Josey, and K. Rowan. 1999. Heat tolerance of Boran and Tuli crossbred steers. Journal of Animal Science 77(9):2398-2405.

Gaughan, J., T. L. Mader, S. Holt, and A. Lisle. 2008. A new heat load index for feedlot cattle. Journal of Animal Science 86(1):226-234.

Hagiya, K., K. Hayasaka, T. Yamazaki, T. Shirai, T. Osawa, Y. Terawaki, Y. Nagamine, Y. Masuda, and M. Suzuki. 2017. Effects of heat stress on production, somatic cell score and conception rate in Holsteins. Animal Science Journal 88(1):3-10.

Hahn, G. L., J. B. Gaughan, T. L. Mader, and R. A. Eigenberg. 2009. Thermal indices and their applications for livestock environments. Proceedings of the Livestock energetics and thermal environment management. American Society of Agricultural and Biological Engineers, p113-130.

Haile-Mariam, M., M. Carrick, and M. Goddard. 2008. Genotype by environment interaction for fertility, survival, and milk production traits in Australian dairy cattle. Journal of Dairy Science 91(12):4840-4853.

Hammami, H., J. Bormann, N. M'hamdi, H. H. Montaldo, and N. Gengler. 2013. Evaluation of heat stress effects on production traits and somatic cell score of Holsteins in a temperate environment. Journal of Dairy Science 96(3):1844-1855.

Hammami, H., J. Vandenplas, M.-L. Vanrobays, B. Rekik, C. Bastin, and N. Gengler. 2015. Genetic analysis of heat stress effects on yield traits, udder health, and fatty acids of Walloon Holstein cows. Journal of Dairy Science 98(7):4956-4968.

Hansen, P. 2004. Physiological and cellular adaptations of zebu cattle to thermal stress. Animal Reproduction Science 82:349-360.

Hansen, P. J. 2020. Prospects for gene introgression or gene editing as a strategy for reduction of the impact of heat stress on production and reproduction in cattle. Theriogenology 154:190-202.

Hayes, B., M. Carrick, P. Bowman, and M. Goddard. 2003. Genotype×environment interaction for milk production of daughters of Australian dairy sires from test-day records. Journal of Dairy Science 86(11):3736-3744.

Hayes, B. J., P. J. Bowman, A. J. Chamberlain, K. Savin, C. P. Van Tassell, T. S. Sonstegard, and M. E. Goddard. 2009. A validated genome wide association study to breed cattle adapted to an environment altered by climate change. PloS ONE 4(8):e6676.

Hayes, B. J. and H. D. Daetwyler. 2019. 1000 bull genomes project to map simple and complex genetic traits in cattle: applications and outcomes. Annual Review of Animal Biosciences 7:89-102.

Heidaritabar, M., M. P. Calus, H. J. Megens, A. Vereijken, M. A. Groenen, and J. W. Bastiaansen. 2016. Accuracy of genomic prediction using imputed whole-genome sequence data in white layers. Journal of Animal Breeding and Genetics 133(3):167-179.

Hickson, R., N. Lopez-Villalobos, P. Kenyon, and S. Morris. 2014. Breed effects and heterosis for productivity traits at first calving of Angus, Holstein Friesian, Jersey and crossbred beef cows. Animal Production Science 54(9):1381-1387.

Hill, D. L. and E. Wall. 2015. Dairy cattle in a temperate climate: the effects of weather on milk yield and composition depend on management. Animal 9(1):138-149.

Howard, J. T., S. D. Kachman, W. M. Snelling, E. J. Pollak, D. C. Ciobanu, L. A. Kuehn, and M. L. Spangler. 2014. Beef cattle body temperature during climatic stress: a genome-wide association study. Internaional Journal Biometeorology 58(7):1665-1672.

IPCC. 2014. Climate change 2014: synthesis report. Contribution of Working Groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change. Ipcc, p151.

Kadzere, C., M. Murphy, N. Silanikove, and E. Maltz. 2002. Heat stress in lactating dairy cows: a review. Livestock Production Science 77(1):59-91.

Key, N., S. Sneeringer, and D. Marquardt. 2014. Climate change, heat stress, and U.S. dairy production. USDA-ERS Economic Research Report, 175.

Kibler, H. 1964. Environmental physiology and shelter engineering. LXVII. Thermal effects of various temperature-humidity combinations on Holstein cattle as measured by eight physiological responses. Research Bulletin Missouri Agricultural Experiment Station, 862.

Kim, J.-M., K.-S. Lim, M. Byun, K.-T. Lee, Y.-r. Yang, M. Park, D. Lim, H.-H. Chai, H.-T. Bang, and J. Hwangbo. 2017. Identification of the acclimation genes in transcriptomic responses to heat stress of White Pekin duck. Cell Stress and Chaperones 22(6):787-797.

Koltes, J. E., D. A. Koltes, B. E. Mote, J. Tucker, and D. S. Hubbell III. 2018. Automated collection of heat stress data in livestock: new technologies and opportunities. Translational Animal Science 2(3):319-323.

Kou, H., Y. Zhao, K. Ren, X. Chen, Y. Lu, and D. Wang. 2017. Automated measurement of cattle surface temperature and its correlation with rectal temperature. PloS ONE 12(4):e0175377.

Laible, G., S.-A. Cole, B. Brophy, S. Leath, S. Jivanji, M. D. Littlejohn, and D. N. Wells. 2020. Holstein Friesian dairy cattle edited for diluted coat color as adaptation to climate change. bioRxiv [preprint].

Lambertz, C., C. Sanker, and M. Gauly. 2014. Climatic effects on milk production traits and somatic cell score in lactating Holstein-Friesian cows in different housing systems. Journal of Dairy Science 97(1):319-329.

Laporta, J., F. Ferreira, V. Ouellet, B. Dado-Senn, A. Almeida, A. De Vries, and G. Dahl. 2020. Late-gestation heat stress impairs daughter and granddaughter lifetime performance. Journal of Dairy Science 103(8):7555-7568.

LCI, 1970. Patterns of transit losses. Livestock Conservation. Inc., Omaha, NE.

Lees, J., A. Lees, and J. Gaughan. 2018. Developing a heat load index for lactating dairy cows. Animal Production Science 58(8):1387-1391.

Li, D., J. Yuan, and R. B. Kopp. 2020a. Escalating global exposure to compound heat-humidity extremes with warming. Environmental Research Letters 15(6):064003.

Li, Q.-L., Z.-H. Ju, J.-M. Huang, J.-B. Li, R.-L. Li, M.-H. Hou, C.-F. Wang, and J.-F. Zhong. 2011. Two novel SNPs in HSF1 gene are associated with thermal tolerance traits in Chinese Holstein cattle. DNA and Cell Biology 30(4):247-254.

Li, R., C. Li, H. Chen, R. Li, Q. Chong, H. Xiao, and S. Chen. 2020b. Genome-wide scan of selection signatures in Dehong humped cattle for heat tolerance and disease resistance. Animal Genetics 51(2):292-299.

Littlejohn, M. D., K. M. Henty, K. Tiplady, T. Johnson, C. Harland, T. Lopdell, R. G. Sherlock, W. Li, S. D. Lukefahr, and B. C. Shanks. 2014. Functionally reciprocal mutations of the prolactin signalling pathway define hairy and slick cattle. Nature Communications 5(1):1-8.

Liu, Z., V. Ezernieks, J. Wang, N. W. Arachchillage, J. Garner, W. Wales, B. Cocks, and S. Rochfort. 2017. Heat stress in dairy cattle alters lipid composition of milk. Scientific Reports 7(1):961.

Luo, H., X. Li, L. Hu, W. Xu, Q. Chu, A. Liu, G. Guo, L. Liu, L. F. Brito, and Y. Wang. 2021. Genomic analyses and biological validation of candidate genes for rectal temperature as an indicator of heat stress in Holstein cattle. Journal of Dairy Science 104(4):4441-4451.

Macciotta, N., S. Biffani, U. Bernabucci, N. Lacetera, A. Vitali, P. Ajmone-Marsan, and A. Nardone. 2017. Derivation and genome-wide association study of a principal component-based measure of heat tolerance in dairy cattle. Journal of Dairy Science 100(6):4683-4697.

Mader, T. L. 2003. Environmental stress in confined beef cattle. Journal of Animal Science 81(14_suppl_2):E110-E119.

Mader, T. L. and M. Davis. 2002. Wind speed and solar radiation corrections for the temperature-humidity index. Proceedings of American Meteorological Society, 15th Conference of Biometiorology and Aerobiology, Joint with 16th International Congress on Biometiorology, p154-157.

Mader, T. L., M. Davis, and T. Brown-Brandl. 2006. Environmental factors influencing heat stress in feedlot cattle. Journal of Animal Science 84(3):712-719.

Maraganore, D. M., M. De Andrade, T. G. Lesnick, K. J. Strain, M. J. Farrer, W. A. Rocca, P. K. Pant, K. A. Frazer, D. R. Cox, and D. G. Ballinger. 2005. High-resolution whole-genome association study of Parkinson disease. The American Journal of Human Genetics 77(5):685-693.

McManus, C., E. Prescott, G. Paludo, E. Bianchini, H. Louvandini, and A. Mariante. 2009. Heat tolerance in naturalized Brazilian cattle breeds. Livestock Science 120(3):256-264.

Mcmanus, C. M., H. Louvandini, T. P. Paim, and F. E. M. Bernal. 2014. Factors affecting heat tolerance in crossbred cattle in central Brazil. Ciência Animal Brasileira 15(2):152-158.

Meuwissen, T., B. Hayes, and M. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157(4):1819-1829.

Misztal, I., T. Strabel, J. Jamrozik, E. Mäntysaari, and T. Meuwissen. 2000. Strategies for estimating the parameters needed for different test-day models. Journal of Dairy Science 83(5):1125-1134.

Moghaddar, N., M. Khansefid, J. H. van der Werf, S. Bolormaa, N. Duijvesteijn, S. A. Clark, A. A. Swan, H. D. Daetwyler, and I. M. MacLeod. 2019. Genomic prediction based on selected variants from imputed whole-genome sequence data in Australian sheep populations. Genetics Selection Evolution 51(1):72.

Mulder, H., R. Veerkamp, B. Ducro, J. Van Arendonk, and P. Bijma. 2006. Optimization of dairy cattle breeding programs for different environments with genotype by environment interaction. Journal of Dairy Science 89(5):1740-1752.

Muller, C. and J. Botha. 1993. Effect of summer climatic conditions on different heat tolerance indicators in primiparous Friesian and Jersey cows. South African Journal of Animal Science 23(3):98-103.

Nakamura, K. and S. F. Morrison. 2008. A thermosensory pathway that controls body temperature. Nature Neuroscience 11(1):62-71.

Nguyen, T. T., P. J. Bowman, M. Haile-Mariam, G. J. Nieuwhof, B. J. Hayes, and J. E. Pryce. 2017. Implementation of a breeding value for heat tolerance in Australian dairy cattle. Journal of Dairy Science 100(9):7362-7367.

Nguyen, T. T., P. J. Bowman, M. Haile-Mariam, J. E. Pryce, and B. J. Hayes. 2016. Genomic selection for tolerance to heat stress in Australian dairy cattle. Journal of Dairy Science 99(4):2849-2862.

Nidumolu, U., S. Crimp, D. Gobbett, A. Laing, M. Howden, and S. Little. 2010. Heat Stress in dairy cattle in northern Victoria: responses to a changing climate. CSIRO Climate Adaptations Flagship Work Paper 10, p1-72.

NRC. 1971. A Guide to Environmental Research on Animals. National Academy of Science, Washington DC.

O'brien, M., R. Rhoads, S. Sanders, G. Duff, and L. Baumgard. 2010. Metabolic adaptations to heat stress in growing cattle. Domestic Animal Endocrinology 38(2):86-94.

Olson, T., C. Lucena, C. Chase Jr, and A. Hammond. 2003. Evidence of a major gene influencing hair length and heat tolerance in Bos taurus cattle. Journal of Animal Science 81(1):80-90.

Oltenacu, P. A. and D. M. Broom. 2010. The impact of genetic selection for increased milk yield on the welfare of dairy cows. Animal Welfare 19(1):39-49.

Otto, P. I., S. E. Guimarães, L. L. Verardo, A. L. S. Azevedo, J. Vandenplas, C. A. Sevillano, D. B. Marques, A. P. Maria de Fatima, C. de Freitas, and R. S. Verneque. 2019. Genome-wide association studies for heat stress response in Bos taurus× Bos indicus crossbred cattle. Journal of Dairy Science 102(9):8148-8158.

Padda, D. 1999. Development of Senepol cattle: A collaborative research story. Proceedings International Senepol Research Symposium Kingshill.

Pegolo, S., N. Mach, Y. Ramayo-Caldas, S. Schiavon, G. Bittante, and A. Cecchinato. 2018. Integration of GWAS, pathway and network analyses reveals novel mechanistic insights into the synthesis of milk proteins in dairy cows. Scientific Reports 8(1):1-15.

Pereira, A. M., F. Baccari, E. A. Titto, and J. A. Almeida. 2008. Effect of thermal stress on physiological parameters, feed intake and plasma thyroid hormones concentration in Alentejana, Mertolenga, Frisian and Limousine cattle breeds. International Journal Biometeorology 52(3):199-208.

Polsky, L. and M. A. von Keyserlingk. 2017. Invited review: Effects of heat stress on dairy cattle welfare. Journal of Dairy Science 100(11):8645-8657.

Porto-Neto, L. R., D. M. Bickhart, A. J. Landaeta-Hernandez, Y. T. Utsunomiya, M. Pagan, E. Jimenez, P. J. Hansen, S. Dikmen, S. G. Schroeder, and E. -S. Kim. 2018. Convergent Evolution of Slick Coat in Cattle through Truncation Mutations in the Prolactin Receptor. Frontiers in Genetics 9:57.

Pragna, P., P. Archana, J. Aleena, V. Sejian, G. Krishnan, M. Bagath, A. Manimaran, V. Beena, E. Kurien, and G. Varma. 2017. Heat stress and dairy cow: impact on both milk yield and composition. Int. Journal of Dairy Science 12(1):1-11.

Pryce, J., J. Arias, P. Bowman, S. Davis, K. Macdonald, G. Waghorn, W. Wales, Y. Williams, R. Spelman, and B. Hayes. 2012. Accuracy of genomic predictions of residual feed intake and 250-day body weight in growing heifers using 625,000 single nucleotide polymorphism markers. Journal of Dairy Science 95(4):2108-2119.

Pryce, J., T. Nguyen, M. Axford, G. Nieuwhof, and M. Shaffer. 2018. Symposium review: Building a better cow—The Australian experience and future perspectives. Journal of Dairy Science 101(4):3702-3713.

Rauw, W. M. and L. Gomez-Raya. 2015. Genotype by environment interaction and breeding for robustness in livestock. Frontiers in Genetics 6:310.

Ravagnolo, O. and I. Misztal. 2000. Genetic component of heat stress in dairy cattle, parameter estimation. Journal of Dairy Science 83(9):2126-2130.

Ravagnolo, O. and I. Misztal. 2002. Effect of heat stress on nonreturn rate in Holstein cows: Genetic analyses. Journal of Dairy Science 85(11):3092-3100.

Ravagnolo, O., I. Misztal, and G. Hoogenboom. 2000. Genetic component of heat stress in dairy cattle, development of heat index function. Journal of Dairy Science 83(9):2120-2125.

Renaudeau, D., A. Collin, S. Yahav, V. De Basilio, J. Gourdine, and R. Collier. 2012. Adaptation to hot climate and strategies to alleviate heat stress in livestock production. Animal 6(5):707-728.

Rhoads, M., J. Kim, R. J. Collier, B. Crooker, Y. Boisclair, L. Baumgard, and R. Rhoads. 2010. Effects of heat stress and nutrition on lactating Holstein cows: II. Aspects of hepatic growth hormone responsiveness. Journal of Dairy Science 93(1):170-179.

Rong, Y., M. Zeng, X. Guan, K. Qu, J. Liu, J. Zhang, H. Chen, B. Huang, and C. Lei. 2019. Association of HSF1 Genetic Variation with Heat Tolerance in Chinese Cattle. Animals 9(12):1027.

Sánchez, J., I. Misztal, I. Aguilar, B. Zumbach, and R. Rekaya. 2009a. Genetic determination of the onset of heat stress on daily milk production in the US Holstein cattle. Journal of Dairy Science 92(8):4035-4045.

Sánchez, J. P., R. Rekaya, and I. Misztal. 2009b. Model for fitting longitudinal traits subject to threshold response applied to genetic evaluation for heat tolerance. Genet. Sel. Evol. 41(1):10.

Santana Jr, M., A. Bignardi, G. Stefani, and L. El Faro. 2017. Genetic component of sensitivity to heat stress for nonreturn rate of Brazilian Holstein cattle. Theriogenology 98:101-107.

Schaeffer, L. 2006. Strategy for applying genome-wide selection in dairy cattle. Journal of Animal Breeding and Genetics 123(4):218-223.

Schüller, L.-K., O. Burfeind, and W. Heuwieser. 2016. Effect of short-and long-term heat stress on the conception risk of dairy cows under natural service and artificial insemination breeding programs. Journal of Dairy Science 99(4):2996-3002.

Sigdel, A., R. Abdollahi-Arpanahi, I. Aguilar, and F. Peñagaricano. 2019. Whole genome mapping reveals novel genes and pathways involved in milk production under heat stress in US Holstein cows. Frontiers in Genetics 10:928.

Sigdel, A., L. Liu, R. Abdollahi-Arpanahi, I. Aguilar, and F. Peñagaricano. 2020. Genetic dissection of reproductive performance of dairy cows under heat stress. Animal Genetics.

Silanikove, N. and N. Koluman. 2015. Impact of climate change on the dairy industry in temperate zones: predications on the overall negative impact and on the positive role of dairy goats in adaptation to earth warming. Small Ruminant Research 123(1):27-34.

Smith, D. L., T. Smith, B. Rude, and S. Ward. 2013. Comparison of the effects of heat stress on milk and component yields and somatic cell score in Holstein and Jersey cows. Journal of Dairy Science 96(5):3028-3033.

St-Pierre, N., B. Cobanov, and G. Schnitkey. 2003. Economic losses from heat stress by US livestock industries. Journal of Dairy Science 86:E52-E77.

Tait-Burkard, C., A. Doeschl-Wilson, M. J. McGrew, A. L. Archibald, H. M. Sang, R. D. Houston, C. B. Whitelaw, and M. Watson. 2018. Livestock 2.0–genome editing for fitter, healthier, and more productive farmed animals. Genome Biology 19(1):1-11.

Talukder, S., P. Celi, K. Kerrisk, S. Garcia, and N. K. Dhand. 2015. Factors affecting reproductive performance of dairy cows in a pasture-based, automatic milking system research farm: a retrospective, single-cohort study. Animal Production Science 55(1):31-41.

Tao, S., R. Orellana, X. Weng, T. Marins, G. Dahl, and J. Bernard. 2018. Symposium review: The influences of heat stress on bovine mammary gland function. Journal of Dairy Science 101(6):5642-5654.

Taye, M., W. Lee, K. Caetano-Anolles, T. Dessie, O. Hanotte, O. A. Mwai, S. Kemp, S. Cho, S. J. Oh, and H. K. Lee. 2017. Whole genome detection of signature of positive selection in African cattle reveals selection for thermotolerance. Animal Science Journal 88(12):1889-1901.

Thom, E. C. 1959. The discomfort index. Weatherwise 12(2):57-61.

Tian, H., N. Zheng, W. Wang, J. Cheng, S. Li, Y. Zhang, and J. Wang. 2016. Integrated Metabolomics Study of the Milk of Heat-stressed Lactating Dairy Cows. Scientific Reports 6:24208.

van den Berg, I., D. Boichard, B. Guldbrandtsen, and M. S. Lund. 2016. Using sequence variants in linkage disequilibrium with causative mutations to improve across-breed prediction in dairy cattle: a simulation study. G3: Genes, Genomes, Genetics 6(8):2553-2561.

VanRaden, P., C. Van Tassell, G. Wiggans, T. Sonstegard, R. Schnabel, J. Taylor, and F. Schenkel. 2009. Invited review: Reliability of genomic predictions for North American Holstein bulls. Journal of Dairy Science 92(1):16-24.

Vermunt, J. J. and B. P. Tranter. 2011. Heat stress in dairy cattle–a review, and some of the potential risks associated with the nutritional management of this condition. Proceedings review of AVA QLD Division Conference 25-27/3/10. Australian Veterinary Association, p212-221.

Vitali, A., M. Segnalini, L. Bertocchi, U. Bernabucci, A. Nardone, and N. Lacetera. 2009. Seasonal pattern of mortality and relationships between mortality and temperature-humidity index in dairy cows. Journal of Dairy Science 92(8):3781-3790.

Wang, T., Y.-P. P. Chen, I. M. MacLeod, J. E. Pryce, M. E. Goddard, and B. J. Hayes. 2017. Application of a Bayesian non-linear model hybrid scheme to sequence data for genomic prediction and QTL mapping. BMC genomics 18(1):618.

Wang, X., B. S. Bjerg, C. Y. Choi, C. Zong, and G. Zhang. 2018. A review and quantitative assessment of cattle-related thermal indices. Journal of Thermal Biology 77:24-37.

West, J. 2003. Effects of heat-stress on production in dairy cattle. Journal of Dairy Science 86(6):2131-2144.

Wheelock, J., R. Rhoads, M. VanBaale, S. Sanders, and L. Baumgard. 2010. Effects of heat stress on energetic metabolism in lactating Holstein cows1. Journal of Dairy Science 93(2):644-655.

Wiggans, G., P. VanRaden, and T. Cooper. 2011. The genomic evaluation system in the United States: Past, present, future. Journal of Dairy Science 94(6):3202-3211.

Wood, A. R., T. Esko, J. Yang, S. Vedantam, T. H. Pers, S. Gustafsson, A. Y. Chu, K. Estrada, Z. Kutalik, and N. Amin. 2014. Defining the role of common variation in the genomic and biological architecture of adult human height. Nature Genetics 46(11):1173-1186.

Xu, C., T. A. Kohler, T. M. Lenton, J.-C. Svenning, and M. Scheffer. 2020. Future of the human climate niche. Proceedings of the National Academy of Sciences 117(21):11350-11355.

Yousef, M. K. 1985. Stress physiology in livestock. Basic principles. CRC press.

Zhang, Y., D. Johnston, S. Bolormaa, R. Hawken, and B. Tier. 2014. Genomic selection for female reproduction in Australian tropically adapted beef cattle. Animal Production Science 54(1):16-24.

Zimbelman, R., R. Rhoads, M. Rhoads, L. Baumgard, and R. Collier. 2009. A re-evaluation of the impact of temperature humidity index (THI) and black globe humidity index (BGHI) on milk production in high producing dairy cows. Proceedings of Western Dairy Management Conference Reno, NV. USA, p113-125.

Chapter 3: Genotype-by-environment (temperature-humidity) interaction of milk production traits in Australian Holstein cattle

E. K. Cheruiyot,[1,2] T. T. T. Nguyen,[1] M. Haile-Mariam,[1*] B. G. Cocks,[1,2] M. Abdelsayed,[3] and J. E. Pryce[1,2]

[1]Agriculture Victoria Research, Department of Jobs, Precincts and Regions, Bundoora, VIC 3083, Australia

[2]School of Applied Systems Biology, La Trobe University, Bundoora, VIC 3083, Australia

[3]Datagene Ltd., AgriBio, Centre for AgriBioscience, Bundoora, VIC 3083, Australia

Genotype-by-environment (temperature-humidity) interaction of milk production traits in Australian Holstein cattle

E. K. Cheruiyot,[1,2] ® T. T. T. Nguyen,[1] M. Haile-Mariam,[1]* ® B. G. Cocks,[1,2] M. Abdelsayed,[3] and J. E. Pryce[1,2] ®
[1]Agriculture Victoria Research, Department of Jobs, Precincts and Regions, Bundoora, VIC 3083, Australia
[2]School of Applied Systems Biology, La Trobe University, Bundoora, VIC 3083, Australia
[3]Datagene Ltd., AgriBio, Centre for AgriBioscience, Bundoora, VIC 3083, Australia

ABSTRACT

Dairying in Australia is practiced in highly diverse climatic conditions and production systems, which means that re-ranking of genotypes could occur across environments that vary in temperature and humidity—that is, genotype-by-environment interactions (G × E) may exist. The objective of this study was to investigate G × E due to heat stress with respect to milk production traits in Australian Holsteins. A total of 6.7 million test-day milk yield records for first, second, and third lactations from 491,562 cows and 6,410 sires that had progeny in different climatic environments were included in the analysis. The environmental gradient used was the temperature-humidity index (THI) calculated from climate data from 163 Australian public weather stations between 2003 and 2017. Data were analyzed using univariate reaction norm (RM) sire model, and the results were compared with multi-trait model (MT). The MT analysis treated test-day yields at 5th percentile (THI = 61; i.e., thermoneutral conditions), 50th percentile (THI = 67; i.e., moderate heat stress conditions), and 95th percentile (THI = 73; i.e., high heat stress conditions) of the trajectory of THI as correlated traits. A THI series of 61, 67, and 73, for example, is equivalent to average temperature and relative humidity of approximately 20°C and 45%, 25°C and 45%, and 31°C and 50%, respectively. We observed some degree of heterogeneity of additive (AG) and permanent environmental (PE) variance over the trajectory THI from RM analysis, with estimates decreasing at higher THI values more steeply for PE than for AG variance. The genetic correlations of the tests between the 5th and 95th percentiles of THI for milk, protein, and fat yield from RM were 0.88 ± 0.01 (standard error), 0.79 ± 0.01, and 0.86 ± 0.01, respec-

tively, whereas the corresponding estimates from MT were 0.86 ± 0.02, 0.84 ± 0.03, and 0.87 ± 0.03. We observed lower genetic correlations between the 5th and 95th percentiles of THI for milk tests from recent years (i.e., 2009 and 2017) compared with earlier years (i.e., 2003 and 2008), which suggests that the level of G × E is increasing in the studied population and should be monitored especially in anticipation of future expected increase in daily average temperature and frequency of heat events. Overall, our results indicate presence of G × E at the upper extreme of the trajectory of THI, but the current extent of sire re-ranking may not justify providing separate genetic evaluations for different levels of heat stress. However, variations observed in the sire sensitivity to heat stress suggest that dairy herds in high heat load conditions could benefit more from using heat-tolerant or resilient sires.
Key words: genotype by environment, heat tolerance, reaction norm, multi-trait model, dairy cattle

INTRODUCTION

Concern about effects of heat stress on livestock production largely used to be an issue mainly in the tropics, but it has now expanded to temperate zones, affecting countries including Australia, Canada, the northern United States, and parts of Europe (Renaudeau et al., 2012; Polsky and von Keyserlingk, 2017). The global livestock industry has to face the double challenge of increasing production to feed a growing population, while dealing with the challenges of changing production environments (Gerber et al., 2013; Polsky and von Keyserlingk, 2017). Heat stress affects livestock production and reproduction, leading to substantial economic losses (West, 2003). Therefore, identifying appropriate genotypes to cope with changing environments is of vital importance, especially in the current scenario of global warming.

Several studies have explored the genetic basis for heat tolerance in dairy cattle and other livestock species [reviewed by Carabaño et al. (2017)]. Ravagnolo

et al. (2000) introduced the use of climate data from weather stations in combination with test-day production data to explore the genetic components of heat stress. This approach has been preferred in many dairy studies because of the availability of large-scale climate data from public weather stations and routine milk recordings. The widely used variable for quantifying the external heat load of animals is temperature-humidity index (**THI**), which is a metric combining ambient temperature and relative humidity. Ravagnolo and Misztal (2000) developed a reaction norm (**RM**) model, which assumes a range of heat load index delimited by a threshold value, beyond which production begins to decrease linearly with increasing THI. The RM approach is appealing because it allows environmental sensitivity (**ES**) of animals to be characterized based on the slope coefficient of the reaction norms (de Jong and Bijma, 2002), such that if the slope value of the RM is zero or close to zero, or if the trend of the RM is consistent across the environmental gradient, an animal is considered to be resilient; otherwise it is labeled "plastic" or sensitive to environmental changes (de Jong and Bijma, 2002; Berghof et al., 2019). These RM models have been widely used to evaluate genetic components of heat tolerance in dairy cattle in various countries, including the United States (Bohmanova et al., 2005; Aguilar et al., 2009), Spain (Carabaño et al., 2014), Italy (Bernabucci et al., 2014), and Germany (Hammami et al., 2015).

In Australia, Nguyen et al. (2016) applied the RM approach in Holsteins and Jersey dairy cattle, which culminated in the development and release of genomic estimated breeding values for heat tolerance to the dairy industry in 2017 (Nguyen et al., 2017; Pryce et al., 2018). The fact that dairying in Australia is practiced in a wide range of climatic conditions and production systems means that the importance of genotype-by-environment interaction (G × E) due to heat stress requires routine monitoring, particularly because of increased global warming. The G × E refers to the change in performance or change in re-ranking of animals in different environments (Falconer and MacKay Longman, 1996). If G × E exists, then animals are expected to re-rank in different environments, which may warrant adapting genetic evaluations for heat tolerance. Information on the extent of G × E due to heat stress —that is, the performance of different genotypes across environments—such as the THI, could be useful for selection to improve thermotolerance.

Australian weather conditions vary considerably among dairying regions. Nguyen et al. (2016) provided an overview of heat stress effects in these regions. Queensland (the northeast region of the country) and Tasmania (the southernmost region of the country) are the regions with the highest and the lowest heat load among dairying regions, with approximately 300 and approximately 100 d per year having THI values beyond the comfort level for dairy cattle, respectively (Nguyen et al., 2016). Also, temporal climatic profiles vary considerably, with average daily temperature in Northern Victoria, for example, ranging from −5 to 38°C (Nguyen et al., 2016). Climate perturbations in the form of increased temperature and frequency of heat events are expected to worsen over the next decades because of global warming (CSIRO and BoM, 2018), implying that genotype-by-climate interactions could become more relevant.

The previous study in Australia using test-day milk records from 1998 to 2001 has revealed evidence of G × E for production traits due to heat stress in Holsteins (Hayes et al., 2003). In a later study, Haile-Mariam et al. (2008) also noted evidence of G × E for fertility and production traits in Australian dairy cattle. The work of Hayes et al. (2003) and Haile-Mariam et al. (2008) considered only first-lactation test-day or whole-lactation records. Greater G × E might be expected in later lactations due to relatively greater milk yield in multiparous cows (Bernabucci et al., 2014), especially when considering a larger data set. In addition, the genetic merit for heat tolerance in Australian Holstein and Jersey cattle has been declining over the years due to selection for production traits (Nguyen et al., 2016; Nguyen et al., 2017). This implies that the extent of re-ranking across heat stress environments in these cattle populations may have increased. Furthermore, Australian dairying is predominantly pasture-based, with limited heat stress adaptation measures, unlike those, for example, in the United States, where extensive managerial and environmental strategies are used to reduce the effects of heat stress. Therefore, the Australian dairy population provides a different perspective in investigating the importance of G × E due to heat stress.

The aim of this study was to investigate the magnitude of G × E across heat stress environments (as measured by THI) and to characterize ES in Australian Holstein cows using test-day milk yield records in combination with temperature-humidity data from public weather stations from 2003 to 2017. A secondary objective of the study was to investigate changes in G × E over time by comparing estimates of the genetic correlations for test-day records from different timescales. We applied RM to the data and compared the results with a multi-trait (**MT**) model, wherein test-day yields measured at different THI values were treated as correlated traits.

Table 1. Overall characteristics of first-, second-, and third-parity data used in the study (SD in parentheses)

Item	Parity 1	Parity 2	Parity 3
Number of herds	2,200	2,090	1,867
Number of herd test dates	99,246	101,492	87,587
Number of records	3,118,223	2,195,283	1,408,885
Number of cows	491,562	368,953	243,278
Number of sires	6,410	6,118	5,622
Average number of cows per herd test date	31.42	21.63	16.08
Average number of daughters per sire	76.69	60.31	43.27
Mean test-day milk yield (L)	22.32 (6.69)	26.42 (8.73)	28.12 (9.25)
Mean test-day fat yield (kg)	0.84 (0.25)	1.00 (0.30)	1.07 (0.33)
Mean test-day protein yield (kg)	0.72 (0.22)	0.86 (0.27)	0.92 (0.28)

MATERIALS AND METHODS

Climate Data

The climate data included hourly dry bulb and dew point temperatures and relative humidity obtained from the Bureau of Meteorology (Melbourne, Australia) for 163 weather stations in Australia from 2003 to 2017. The locations of the weather stations and dairy herds from which the study data were obtained have been presented previously in Nguyen et al. (2016). The pairwise distances between herds and weather stations were calculated from geographical coordinates, as detailed in Nguyen et al. (2016). The 5th, 50th, and 95th percentiles of distances between weather stations were 4, 23, and 60 km, respectively (Nguyen et al., 2016). Test-day milk records were assigned to the nearest weather station. Hourly THI for each weather station was calculated using the following formula (Yousef, 1985) and averaged for 24 h to obtain the daily THI:

$$THI = T_{db} + (0.36\,T_{dp}) + 41.2,$$

where T_{dp} is dew point temperature (°C) and $T_{dp} = (237.3b)/(1.0 - b)$, where b = [log (RH/100.0) + $(17.27\,T_{db})/(237.3 + T_{db})]/17.27$, and RH = relative humidity, T_{db} = dry bulb temperature (°C). The daily THI on the test day and d 1, 2, 3, and 4 before test day were then averaged and assigned to the respective test-day records.

Test-Day Data

Milk, protein, and fat yield data were obtained from DataGene (DataGene Ltd., Melbourne, Australia). These data sets consisted of 6.6, 4.7, and 3.0 million records for first, second, and third lactations, respectively, for Holstein cows calving between 2003 and 2017. Data editing for the first lactation was as follows: (1) tests <5 or >305 DIM and test-day records with less than

10 cows were removed; (2) sires with daughters in fewer than 2 herds and herds using fewer than 2 sires were excluded; (3) only cows with at least 4 herd-test records within the first lactation were retained for analyses. Tests for the second and third lactations were then selected if the cows were present in the first lactation. The remaining data set included 11.2 million records for 823,055 cows and 6,615 sires from 3,732 herds. Due to computational limitations, roughly 60% of the data could be used for analysis. Therefore, a random sample of 2,200 herds was selected for analysis. The final data set comprised 6.7 million records for 491,562 cows and 6,410 sires (Table 1). The pedigree for this data included parents up to 15 generations.

Reaction Norm Analysis

In this study, we used RM models because they are better suited to continuous environmental descriptors (in this case, THI) and also facilitate distinguishing between individuals that are less or more affected by environmental changes. Due to computational limitations, univariate sire models were used instead of animal models. The combined data for first, second, and third lactation was fitted as follows:

$$y_{ijklm} = \mu + HTD_i + YS_j + PAR_k + \sum_{n=1}^{3} A_n X_n + \sum_{n=1}^{8} PAR_k D_n Z_n$$
$$+ \sum_{n=1}^{2} P_n T_n + \sum_{n=0}^{1} S_{ln} Q_n + \sum_{n=0}^{1} C_{mn} W_n + e_{ijklm},$$

where y_{ijklm} is yield of milk in liters, fat or protein in kg from the ith herd test day, jth year season of calving, lth sire, and mth cow in kth parity; μ is the intercept; HTD_i is the effect of the ith herd test day; YS_j is the effect of the jth year season of calving; PAR_k is the effect of kth parity; X_n is the nth-order Legendre polynomials corresponding to age on day of test; A_n is a fixed regression coefficient of traits on age at test; Z_n is the

nth-order Legendre polynomials corresponding to DIM at test; D_n is a fixed regression coefficient of traits on DIM nested within parity; P_n is the nth-order Legendre polynomial corresponding to THI; T_n is a fixed regression coefficient of traits on THI; S_{ln} is an RM coefficient on THI for the lth sire; C_{mn} is the RM coefficient on THI for the mth cow to account for repeated records of cows; Q_n and W_n are either the intercept (n = 0) or slope solution (n = 1) for THI for sires and cows, respectively; and e_{ijkm} is the vector of residual effects. To compare our results with a previous study in Australia (Hayes et al., 2003), we also fitted the same model to first-lactation data separately, excluding the effect of parity.

Before analysis, a small proportion (0.004%) of tests with THI values of above 75 were arbitrarily given a value of 75. This was done to avoid possible artifacts of variance estimation using RM models, which might lead to unexpected trajectories at the extremes of the environmental descriptor due to few data points (Misztal et al., 2000). Milk yield traits in Australia begin to decline at THI > 60 (Hayes et al., 2003; Nguyen et al., 2016). Therefore, the THI threshold was set at 60 in this study (i.e., if THI < 60, then THI = 60). The variance or covariance structure for additive sire effects was the following:

$$Var(S) = \begin{bmatrix} S_0 \\ S_1 \end{bmatrix} = \begin{bmatrix} \mathbf{A}\sigma_{S_0}^2 & \mathbf{A}\sigma_{S_0 S_1} \\ \mathbf{A}\sigma_{S_0 S_1} & \mathbf{A}\sigma_{S_1}^2 \end{bmatrix},$$

where $\mathbf{A}$ is the relationship matrix among sires constructed from pedigree data; S_0 and S_1 are the intercept and slope for sires; and $\sigma_{S_0}^2$, $\sigma_{S_0 S_1}$, $\sigma_{S_0 S_1}$, and $\sigma_{S_0}^2$ are (co)variance for sire effects on environmental descriptor. The (co)variance structure for the PE effect was as follows:

$$Var(C) = \begin{bmatrix} C_0 \\ C_1 \end{bmatrix} = \begin{bmatrix} \mathbf{I}\sigma_{C_0}^2 & \mathbf{I}\sigma_{C_0 C_1} \\ \mathbf{I}\sigma_{C_0 C_1} & \mathbf{I}\sigma_{C_1}^2 \end{bmatrix},$$

where $\mathbf{I}$ is the identity matrix for each test-day record; C_0 and C_1 are the intercept and slope for cows; and $\sigma_{C_0}^2, \sigma_{C_0 C_1}, \sigma_{C_0 C_1}, \sigma_{C_1}^2$ are (co)variance for cow effects on environmental descriptor (i.e., THI).

Heterogeneous error variance was modeled for 10 DIM intervals over a lactation (DIM = 5 to 30, 31 to 60, 61 to 90, 91 to 120, 121 to 150, 151 to 180, 181 to 210, 211 to 240, 241 to 270, and 271 to 305), assuming the following variance structure:

$$Var(e) = \mathrm{var} \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_{30} \end{bmatrix} = \begin{bmatrix} \sigma_{e_1}^2 & 0 & 0 & 0 \\ 0 & \sigma_{e_2}^2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \sigma_{e_{30}}^2 \end{bmatrix},$$

where $\sigma_{e_1}^2, \sigma_{e_2}^2, \ldots, \sigma_{e_{30}}^2$ represent error variances (30×30 error matrix for first, second, and third lactations). All analyses were performed using ASREML version 4.2 (Gilmour et al., 2015).

Calculation of Genetic Parameters

Additive genetic 16×16 (co)variance $\widehat{\mathbf{G}}$ matrixes for sires along the THI trajectory (i.e., $60 \leq \text{THI} \leq 75$) were calculated as follows:

$$\widehat{\mathbf{G}} = 4 \times \mathbf{\Phi}\,\mathrm{var}\left(\hat{\mathbf{S}}\right)\mathbf{\Phi}',$$

where $\mathbf{\Phi}$ is the 16×2 matrix of Legendre polynomial function for THI and $\hat{\mathbf{S}}$ is a 2×2 sire (co)variance matrix. Similarly, the PE (co)variances matrix $\widehat{\mathbf{PE}}$ were calculated from (2×2) $\widehat{\mathbf{C}}$ cow (co)variance matrix. The residual (co)variance matrix $\widehat{\mathbf{R}}$ was a 16×16 identity matrix; the diagonal elements were an average of 10 estimates of variances for first parity (estimated from 10 DIM intervals over the lactation), or 30 estimates of variances for 3 parity analyses. The phenotypic (co)variance matrix $\widehat{\mathbf{P}}$ was obtained by linearly summing additive, permanent, and residual variances $\left(\widehat{\mathbf{P}} = \widehat{\mathbf{G}} + \widehat{\mathbf{PE}} + \widehat{\mathbf{R}}\right)$.

The genetic correlations were calculated as $rg_{i,j} = g_{i,j} / \sqrt{g_{i,i} \times g_{j,j}}$, where i and j are the genetic (co)variances for yield at THI = 60 and THI $\geq$ 60, respectively, and their corresponding approximate standard errors are computed as proposed by Fischer et al. (2004) and expounded in Su et al. (2007), as follows:

$$SE^2 = rg^2 \begin{bmatrix} \dfrac{var(g_{i,i})}{4(g_{i,i})^2} + \dfrac{var(g_{j,j})}{4(g_{j,j})^2} + \dfrac{var(g_{i,j})}{(g_{i,j})^2} \\[2ex] + \dfrac{cov(g_{i,i},\ g_{j,j})}{2g_{i,i}g_{j,j}} - \dfrac{cov(g_{i,i},\ g_{i,j})}{g_{i,i}g_{i,j}} - \dfrac{cov(g_{j,j},\ \sigma_{i,\ j})}{g_{j,j}g_{i,j}} \end{bmatrix},$$

where $g_{i,i}$, $g_{j,j}$, and $g_{i,j}$ are the elements of $(16 \times 16) \times (16 \times 16)$ $\widehat{\mathbf{G}}$ matrix with their (co)variances $\widehat{\mathbf{VG}}$ approximated from the observed inverse of the average information matrix.

The heritability at the ith THI was calculated as $h_i^2 = \dfrac{4 \times g_{i,i}}{p_{i,i}}$, with their corresponding approximate standard errors computed as

$$SE^2 = \left(h_i^2\right)^2 \left[4\frac{var\left(g_{i,i}\right)}{\left(g_{i,i}\right)^2} + \frac{var\left(p_{i,i}\right)}{\left(p_{i,i}\right)^2} - 2\frac{cov\left(g_{i,i},\; p_{i,i}\right)}{4 g_{i,i} p_{i,i}} \right],$$

where $g_{i,i}$, $p_{i,i}$, $var(g_{i,i})$, and $var(p_{i,i})$ are the diagonal elements of $\widehat{\mathbf{G}}$, $\widehat{\mathbf{P}}$, $\widehat{\mathbf{VG}}$, and $\widehat{\mathbf{VP}}$ matrices with (16×16) $\times$ (16×16) dimensions, respectively; $\widehat{\mathbf{VP}} = \widehat{\mathbf{VG}} + \widehat{\mathbf{VPE}} + \widehat{\mathbf{VR}}$. The diagonal elements of $\widehat{\mathbf{VR}}$ were an average of 10 estimates of variances from $\widehat{\mathbf{R}}$ matrix.

The EBV along the THI trajectory for the ith sire was calculated as $EBV_i = \phi_j \times \hat{a}_i{}'$, where $\hat{a}_i{}'$ is the vector of estimated RM coefficients for the slope and intercept for sire i, and ϕ_j is the vector of Legendre polynomials evaluated at THI j. To examine changes in performance, the EBV for sires with more than 100 daughters with yield records (n = 617) were estimated and represented along the THI trajectory. Environmental sensitivity of sires was characterized according to Mattar et al. (2011), based on the absolute values of the slope of the RM: $\left|S_i\right| < \sigma_s$, resilient sires; $\sigma_s \leq \left|S_i\right| \leq 2\sigma_s$, sensitive sires; and $\left|S_i\right| \geq 2\sigma_s$, extremely sensitive sires.

Multi-Trait Analysis

To compare the results of RM, we performed an MT analysis for 3 parity data by considering tests at 5th $(60 < \text{THI} \leq 61)$, 50th $(66 < \text{THI} \leq 67)$, and 95th $(72 < \text{THI} \leq 74)$ percentiles of THI distribution as different but correlated traits. Hereafter, yields at 5th, 50th, and 95th percentiles are indicated as THI = 61, 67, and 73, respectively. The proportion of tests at the 5th, 50th, and 95th percentile THI points were 59% (n = 373,564), 31% (n = 192,571), and 10% (n = 61,548), respectively. A small proportion of cows (0.007%, 0.008%, and 0.005%) had repeated records in at least 1 of the 3 THI percentiles for parity 1, 2, or 3, respectively. An initial model considering PE effects, aimed at accounting for the repeated records, failed to converge, most likely due to a few repeated data. Therefore, only 1 record was randomly selected at each THI point for all the cows with repeated records and included in the final analysis.

The following MT model was fitted to the data:

$$y_{tijkl} = \mu + HTD_{ti} + YS_{tj} + PAR_{tk} + \sum_{n=1}^{3} A_{tn} X_{tn}$$

$$+ \sum_{n=1}^{8} PAR_{tk} D_{tn} Z_{tn} + S_{tl} + e_{tijkl},$$

where y_{tijkl} is the nth observation for the tth trait (i.e., yields at 5th, 50th, and 95th THI percentiles); μ_t is the overall mean for the tth trait. Fixed effects were the same as described in the RM model, except that the fixed regression of the environment descriptor (THI) was excluded; and S_{tl} is the random effect of the lth sire for the tth trait. The (co)variance structure for the 3 traits was as follows:

$$\text{var}\begin{bmatrix} S_1 \\ S_2 \\ S_3 \\ e_1 \\ e_2 \\ e_3 \end{bmatrix} = \begin{bmatrix} \mathbf{A}\sigma_{S_1}^2 & \mathbf{A}\sigma_{S_1 S_2} & \mathbf{A}\sigma_{S_1 S_3} & 0 & 0 & 0 \\ \mathbf{A}\sigma_{S_2 S_1} & \mathbf{A}\sigma_{S_2}^2 & \mathbf{A}\sigma_{S_2 S_3} & 0 & 0 & 0 \\ \mathbf{A}\sigma_{S_3 S_1} & \mathbf{A}\sigma_{S_3 S_2} & \mathbf{A}\sigma_{S_3}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{e_1}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{e_2}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_{e_3}^2 \end{bmatrix},$$

where S_1, S_2, and S_3 are the genetic sire effects for the cows with records in the 5th, 50th, and 95th percentiles of THI distribution, respectively; $\mathbf{A}$ is the relationship matrix for sires constructed from pedigree; $\sigma_{S_1}^2$, $\sigma_{S_2}^2$, $\sigma_{S_3}^2$, $\sigma_{S_1 S_2}$, $\sigma_{S_1 S_3}$, $\sigma_{S_2 S_1}$, $\sigma_{S_2 S_3}$, and $\sigma_{S_3 S_1}$ are the sire genetic (co)variances for the milk records at the 3 THI percentiles; and $\sigma_{e_1}^2$, $\sigma_{e_2}^2$, and $\sigma_{e_3}^2$ are the residual variances for the 3 traits.

Changes in Genetic Correlations over Time

To see whether G × E has changed over time, we divided data from first-parity test-day yields into 2 groups: (1) tests from 2003 to 2008 and (2) tests from 2009 to 2017. We then compared the sizes of genetic correlations in these subsets. The 2 timescales were arbitrarily chosen to ensure that the number of records in each subset was similar. Unlike the RM above, all the first-parity data that remained after editing were included, to ensure sufficient records in each subset. Only first-lactation data were used to minimize possible bias of selection and to be comparable to previous work in Australia (Hayes et al., 2003). Subset 1 (2003 to 2008) comprised 2.8 million records of 444,818 cows and 4,225 sires, whereas subset 2 (2009 to 2017) com-

prised 2.7 million records of 427,800 cows and 4,668 sires. The same RM, as described earlier, was applied to each subset separately, excluding the effect of parity.

RESULTS

Climate Data and Phenotypes

The distribution of THI is presented in Figure 1. Approximately 45% of the milk records in this study had a THI of greater than 60, which is the threshold determined as being indicative of heat stress (Hayes et al., 2003; Nguyen et al., 2016). The average test-day yields increased from the first to the third lactation for all milk traits (Table 1). The average yields for multi-parity data used in the MT analysis are relatively greater at the 95th percentile of THI (Table 2), but with larger standard deviations.

Additive Genetic and Permanent Environmental Variances

Sire parameter estimates resulting from the RM model are given in Table 3. As expected, the correlations between the intercept and the slope of the reaction norms were negative for all traits: milk (−0.10), protein (−0.05), and fat (−0.26). Additive genetic (**AG**) variance along the trajectory of THI for milk, protein, and fat yields from RM and MT are presented in Figure 2. The estimates for AG variance were higher for the MT than for the RM analysis for all the milk traits, particularly at higher THI values (i.e., 95th percentile; Figure 2). The estimates and trends for the AG variance across the THI are similar for both analyses. In RM, the AG variance for milk and protein yield decreased with increasing THI values up to approximately THI = 70, beyond which it increases slightly. For fat yield, the AG variance increased marginally at THI > 72. The estimates for AG variance ranged from 3.06 to 3.65 (L^2), 0.002 to 0.003 (kg^2), and 0.003 to 0.004 (kg^2) for milk, protein, and fat, respectively. In MT, estimates decreased from the 5th (THI = 61) to the 50th percentile (THI = 67) and appear to markedly increase at the 95th percentile, with conspicuously larger standard errors at the 95th compared with those for the 5th and 50th percentiles. The estimates of AG variance from MT models are given in Table 2.

The patterns of permanent environmental (**PE**) variances for milk traits from RM (Figure 2) are similar to the pattern of AG variance, with estimates decreasing with increasing THI, up to around THI = 70, beyond which it increases. The PE component was consider-
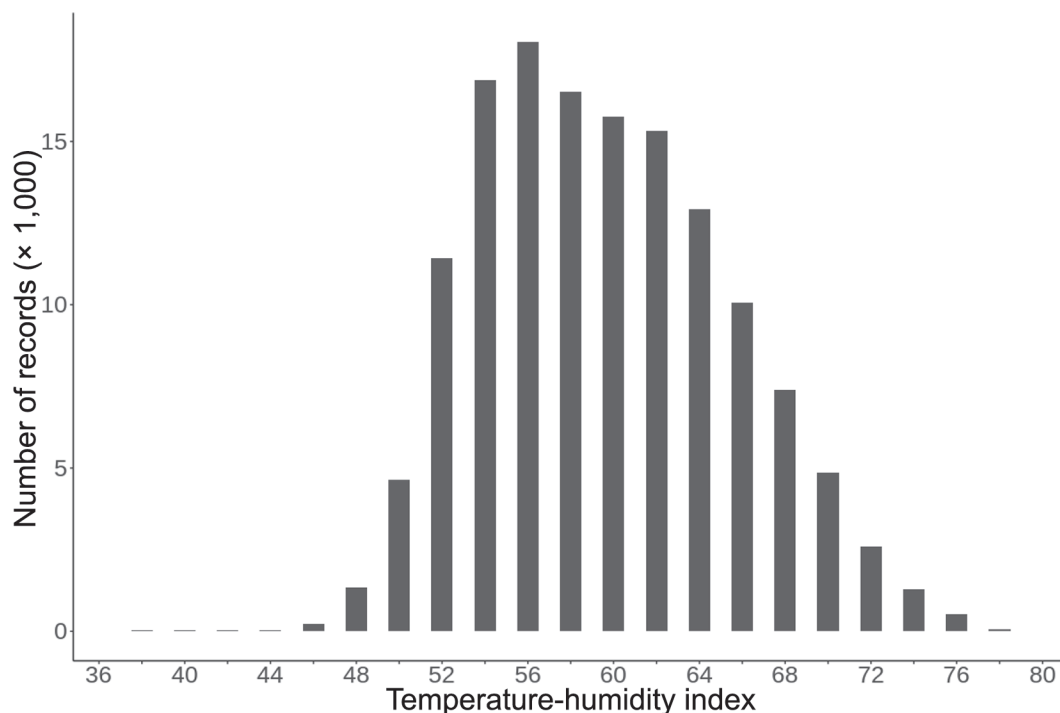
Figure 1. Distribution of temperature-humidity index (THI) values in the data used in this study: THI < 60, 55%; 60 ≤ THI ≤ 65, 26%; 65 < THI ≤ 70, 14%; and THI > 70, 5%.

Table 2. Average milk yield traits and temperature-humidity index (THI), additive genetic variance (AG), heritabilities (h^2), and genetic correlations from the multi-trait model for milk yields at 5th (THI = 61), 50th (THI = 67), and 95th (THI = 73) percentiles of THI (SE estimates in parentheses)

Item	5th $\bar{X}^1$	5th AG	5th h^2	50th $\bar{X}$	50th AG	50th h^2	95th $\bar{X}$	95th AG	95th h^2	5th vs. 50th	50th vs. 95th	5th vs. 95th
Milk	23.84 (8.35)	3.79 (0.16)	0.18 (0.01)	23.43 (7.72)	3.70 (0.17)	0.20 (0.01)	24.63 (9.19)	4.84 (0.34)	0.21 (0.01)	0.98 (0.01)	0.92 (0.02)	0.86 (0.02)
Protein	0.77 (0.25)	0.0027 (0.0001)	0.14 (0.01)	0.74 (0.23)	0.0026 (0.0001)	0.15 (0.01)	0.76 (0.27)	0.0032 (0.0002)	0.16 (0.01)	0.97 (0.01)	0.90 (0.02)	0.84 (0.03)
Fat	0.91 (0.29)	0.0044 (0.0002)	0.12 (0.01)	0.90 (0.28)	0.0041 (0.0002)	0.12 (0.01)	0.92 (0.34)	0.0048 (0.0004)	0.11 (0.01)	0.97 (0.01)	0.87 (0.03)	0.87 (0.03)
THI	61.50			67.48			73.84					

Genetic correlation columns: 5th vs. 50th, 50th vs. 95th, 5th vs. 95th.

[1]$\bar{X}$ = mean value of the trait and THI at 5th, 50th, and 95th percentiles.

ably higher than the AG variance component, with estimates that were more than 2-fold higher, particularly at lower THI values. For example, at the 5th percentile (THI = 61), the estimated PE variance for milk yield was 8.38 (L^2), whereas the AG variance was 3.59 (L^2). The changes in estimated PE variance along the THI trajectory were greatest for protein (34%), followed by fat (25%) and milk yield (23%), relative to estimates from RM analysis between THI = 61 and THI = 70 (i.e., the interval between which estimates decrease). The PE estimates ranged from 6.60 to 8.58 (L^2), 0.005 to 0.007 (kg^2), and 0.008 to 0.011 (kg^2) for milk, protein, and fat yield, respectively; note that the PE variance was not fitted in MT analysis.

Heritabilities

Heritability estimates along the THI trajectory from RM and MT analyses are presented in Figure 3. For both analyses, heritability estimates are highest for milk followed by protein and fat yields, respectively. In RM, heritability estimates decrease with increasing THI across the entire trajectory of fat, but it increases at THI > 70 for milk and protein yield, with a more pronounced increase for protein relative to milk yield. The steeper line curve for fat yield indicates larger decreasing estimates. The heritability estimates for milk, protein, and fat yield over THI trajectory (i.e., $60 \leq THI \geq 75$) ranged from 0.13 ± 0.01 to 0.14 ± 0.009, 0.10 ± 0.007 to 0.11 ± 0.01, and 0.07 ± 0.007 to 0.09 ± 0.006, respectively. The standard errors for these estimates (Figure 3) are consistent across the THI trajectory, with a slight increase at the extremes. When considering only first-parity data in RM analysis, heritability estimates increased marginally for milk (0.14 ± 0.01 to 0.16 ± 0.01), protein (0.10 ± 0.01 to 0.12 ± 0.01), and fat (0.08 ± 0.01 to 0.11 ± 0.01).

In contrast, heritability estimates from MT were greater than those from RM, with values at 5th and 95th percentiles of 0.18 ± 0.007 and 0.21 ± 0.01, 0.14 ± 0.006 and 0.16 ± 0.01, and 0.11 ± 0.01 and 0.12 ± 0.006 for milk, protein, and fat yields, respectively (Table 2). Unlike RM, the estimates for all the milk traits from MT at higher THI (i.e., 95th) are associated with large standard errors (Figure 3).

Genetic Correlations

The genetic correlation estimates for milk production traits between THI = 60 and those beyond THI = 60 are similar for both RM and MT analyses, with slightly larger estimates for protein and fat yields from MT at higher THI values (Figure 4). The genetic correlations between tests at THI = 60 versus THI >

70

Table 3. Sire variance estimates of intercept $\left(\sigma_{S_0}^2\right)$ and slope $\left(\sigma_{S_1}^2\right)$, covariance between slope and intercept $\left(\sigma_{s_0s_1}\right)$, correlation between slope and intercept $\left(r_{s_0s_1}\right)$, residual variance $\left(\sigma_e^2\right)$, and heritability of intercept $\left(h_{S_0}^2\right)$ and slope $\left(h_{S_1}^2\right)$ from reaction norm analysis for multi-parity data for milk, protein, and fat yields (SE in parentheses)

Trait	$\sigma_{S_0}^2$	$\sigma_{S_1}^2$	$\sigma_{s_0s_1}$	$r_{s_0s_1}$	σ_e^2	$h_{S_0}^2$	$h_{S_1}^2$
Milk (L)	1.555	0.057	−0.031	−0.103	15.86	0.21	0.014
	(0.05)	(0.005)	(0.01)	(0.01)	(0.061)	(0.01)	(0.001)
Protein (kg)	0.001	0.00007	−0.00001	−0.049	0.032	0.15	0.017
	(0.00003)	(0.00001)	(0.00001)	(0.071)	(0.001)	(0.01)	(0.001)
Fat (kg)	0.0015	0.00007	−0.00009	−0.263	0.016	0.12	0.009
	(0.0001)	(0.00001)	(0.00001)	(0.015)	(0.0001)	(0.005)	(0.001)

70 were lowest for protein, followed by milk and fat yields, respectively. The correlation estimates for tests between the 5th and the 95th percentiles of THI from RM were 0.88 ± 0.01, 0.79 ± 0.02, and 0.86 ± 0.02 for milk, protein, and fat yield. These were in good general agreement with estimates from MT, which were 0.86 ± 0.02, 0.84 ± 0.03, and 0.87 ± 0.03, respectively (Table 2 and Figure 4).
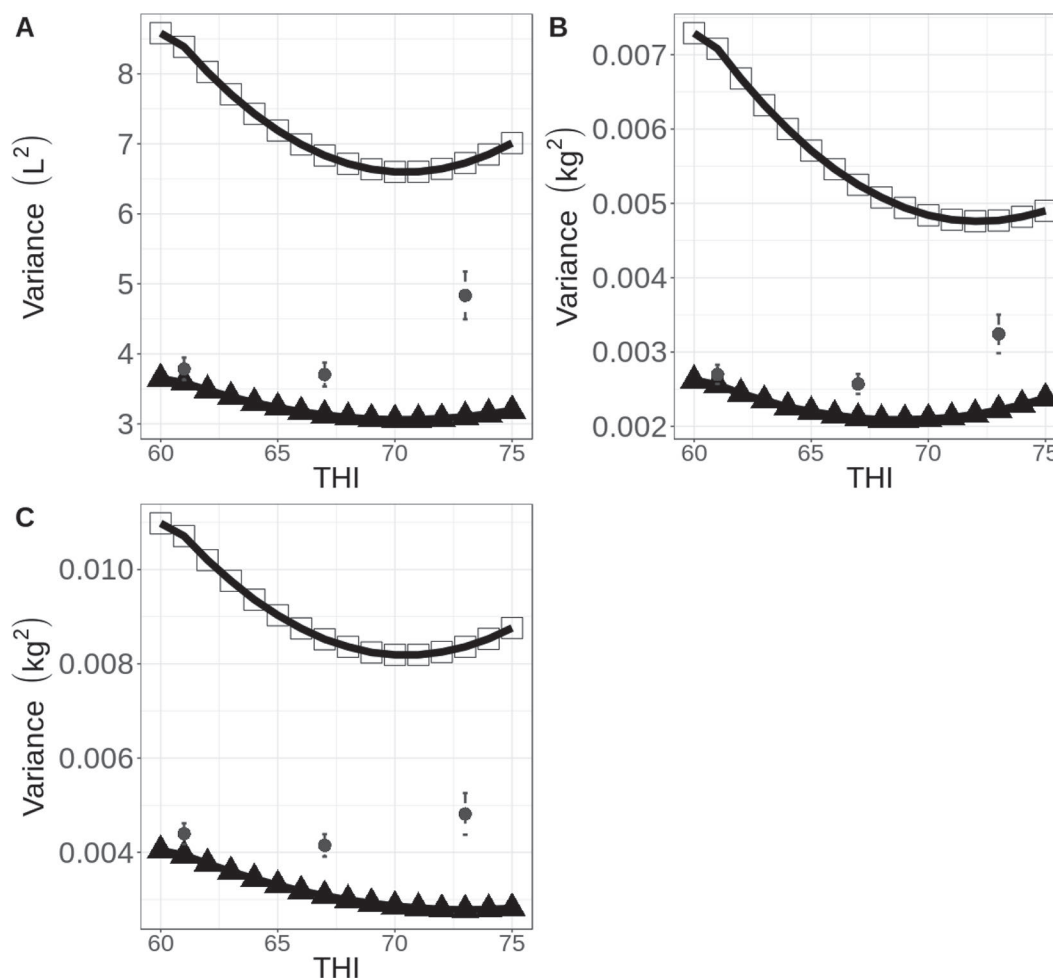
Figure 2. Additive genetic (▲) and permanent environmental (□) variance along the trajectory of the temperature-humidity index (THI) for milk (A), protein (B), and fat (C) yields from reaction norm analysis. Dot points (●) represent additive genetic variances and respective SE (vertical dashed lines) at 5th (THI = 61), 50th (THI = 67), and 95th (THI = 73) THI percentiles from the multi-trait analysis.

Changes in Correlations over Time

The genetic correlations from RM for the data split into 2 subsets are shown in Figure 5. Subset 1 were test-day records from 2003 to 2008, and subset 2 were test-day records from 2009 to 2017. The correlation estimates are comparatively lower for more recent data (subset 2) than for older data (subset 1) at extreme THI values for all the milk traits, with the largest differences observed for protein, followed by fat and milk yields (Figure 5). The correlation estimates for milk, protein, and fat yields between the 5th and 95th THI points for subset 1 (2003 to 2008) were 0.89 ± 0.01, 0.84 ± 0.02, and 0.86 ± 0.02; whereas those for subset 2 (2009 to 2017) were 0.87 ± 0.01, 0.79 ± 0.02, and 0.81 ± 0.02, respectively.

Quantifying Resilience to Heat Stress

Figure 6 shows the EBV (i.e., the slope of the reaction norms) over the trajectory of THI for a random sample of 10 sires with more than 100 daughters with records. The reaction norms are very similar for all milk traits. Two groups of sires can be identified based on the magnitude of their EBV at thermoneutral conditions (THI = 60) versus heat stress conditions (THI = 75). The first group (shown in gray) are sires with above-average EBV at THI = 60 and smaller EBV at THI = 75. The second group (black) are sires with above-average EBV at THI = 60 and stable EBV (i.e., their EBV do not substantially change at THI = 60 and THI = 75).

The correlation between EBV for slope and intercept of the reaction norms for milk, protein, and fat yield were −0.20, −0.39, and −0.25, respectively. The slope of the reaction norms (i.e., ES) for sires with at least 100 daughters with records in multi-parity ranged from −11.69 to 5.84 (SD = 2.08), −3.19 to 1.96 (SD = 0.82), and −2.92 to 2.29 (SD = 0.75) for milk, protein, and fat yields, respectively. Of these, the percentage of resilient sires (i.e., sires with EBV with consistent slope) were 65%, 57%, and 64%, respectively (Figure 7). Greater proportions of sensitive sires were found for more recent milk tests (2009 to 2017) than for earlier tests (2003 to 2008) particularly for milk (30 vs. 23%) and protein yield (29 vs. 26%) (Figure 8).

DISCUSSION

In this study we used test-day milk production records from across Australia to estimate G × E due to heat stress—that is, the changes in the relative performance of different genotypes (sires) across THI environments (i.e., heat stress conditions). Australia is a large country, with dairy herds dispersed throughout the country and in 8 distinct dairying regions (Dairy Australia, 2016). Quantifying the extent of re-ranking due to heat stress in Australian dairy cows is important for 2 main reasons. First, Australia's dairy herds are kept in highly diverse climatic conditions, ranging from the warmer northeast regions to the cooler southeast regions of the country. As such, dairy cows experience varying magnitudes of heat stress (both temporal and seasonal), with some dairying regions, including Queensland, New South Wales, and Western and South Australia, under relatively high heat loads extending up to half of the year, whereas temperatures are moderate in Victoria and lowest in Tasmania, the southernmost part of the country (Nguyen et al., 2016). Second, Australian dairy farms are mostly kept outdoors and follow pasture-based systems, with limited management measures to alleviate heat stress, in contrast to many European countries, the United States, and Canada, where cows are generally housed indoors and fed TMR. Thus, Australia is likely to be a good case study for examining G × E due to heat stress in dairy cows.

Variance Components

AG Variance. In our analyses, we used RM in addition to the MT model for comparison purposes. We found similar estimates for AG variance from both approaches at lower THI values, but differences between estimates increased at higher levels of THI (i.e., 95th percentile). The contradictory results at extreme trajectories were expected because of the sparsity of data, which perhaps led to poor fitting in both analyses. The general trend of AG variance was similar in both analyses and consistent with other studies that reported a typical U-shaped variance pattern for production traits due to heat stress (Ravagnolo et al., 2000; Aguilar et al., 2009; Brügemann et al., 2011). The AG variance for all the milk traits from RM was relatively stable across the THI trajectory (Figure 2), suggesting that a similar response to selection is expected regardless of the THI environment.

PE Variance. As with the AG variance, we also observed a declining trend for PE variance for all the milk traits across THI environment, which is in line with those reported for protein yields in German Holsteins (Brügemann et al., 2011) and milk yields in Iranian Holstein cattle (Santana Jr. et al., 2015). The higher PE estimates than AG variance from RM is consistent with other studies (Aguilar et al., 2009; Brügemann et al., 2011), indicating a greater effect of heat stress on production traits.

We used the first-order polynomial function to model changes in PE as well as AG variance across THI trajectory in RM analysis. Applying higher-order regression, such as quadratic regression, might have allowed a better fit to the data (Strabel et al., 2005), but this was computationally impractical due to the large data size used in this study. Nonetheless, similar studies in Australia (Hayes et al., 2003; Nguyen et al., 2016) have successfully applied first-order polynomial functions to evaluate heat tolerance for production traits in Holsteins and Jersey dairy cattle.

Heritabilities. We used the average value of heterogeneous variances over the DIM to estimate heritability in RM, whereas separate residual variances for the 3 "traits" (i.e., yields at 5th, 50th, and 75th THI percentiles) were used in MT. We found contradicting

trends for heritability estimates across the THI trajectory from the 2 analyses, which may be attributable, at least in part, to the way the heritability estimates were calculated. An alternative to RM, which perhaps could have allowed more comparable results, would have been to model (co)variance components for the interactions between DIM and THI, as in Bohlouli et al. (2019), and then estimate residual variances and heritabilities within DIM × THI combinations, but unfortunately this was not computationally feasible in our study.

Our heritability estimates from RM for both first-parity and multi-parity data were almost 2-fold lower than those reported in a previous study in Australia (Hayes et al., 2003). In that work, the authors applied the same RM as in our study and found estimates for first-parity milk, protein, and fat yields at 5th and 95th
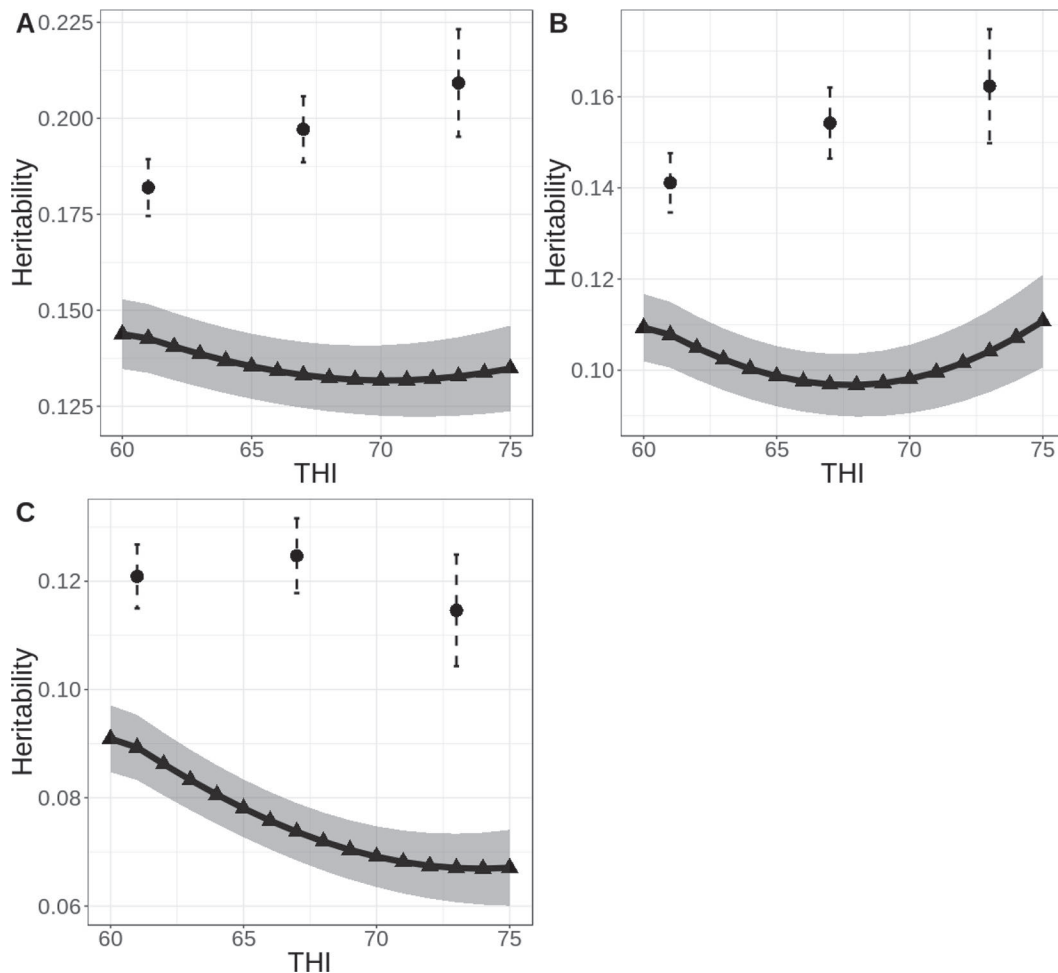
Figure 3. Heritabilities for milk (A), protein (B), and fat (C) yields across the temperature-humidity index (THI) obtained from the reaction norm (RM) and multi-trait (MT) analyses. Lines represent estimates obtained from RM and their approximate SE (shaded area). Dot points (●) are estimates and their SE (dashed vertical lines) at 5th (THI = 61), 50th (THI = 67), and 95th (THI = 73) percentiles from the MT.

THI percentiles of 0.26 and 0.24, 0.21 and 0.19, and 0.19 and 0.16, respectively. Although our estimates from MT were greater than those from RM, they are still much lower than those found by Hayes et al. (2003). Moreover, our estimates are lower than the heritability estimates for heat tolerance for first-lactation Australian Holsteins reported by Nguyen et al. (2016) for milk (0.19), protein (0.17), and fat (0.17) yields, which used a genomic relationship matrix to derive heritabilities. A relatively wide range of heritability estimates from 0.10 to 0.35 (Carabaño et al., 2014), 0.14 to 0.31 (Bohlouli et al., 2013), and 0.10 to 0.24 (Aguilar et al., 2009) contrasts with the narrow range for all the milk traits observed here, which can be attributed in part to the narrow range of the available THI scale (i.e., THI = 60 to THI = 75) as well as to the small change in genetic variance observed across the trajectory of THI.

Genetic Correlations and Changes over Time

The correlation estimates of yield traits at 5th versus 95th THI percentiles from both RM and MT analyses were 0.80 or more but are lower than those reported by Hayes et al. (2003), which considered only first-parity data. The correlations between the 5th and the 95th percentiles reported by Hayes et al. (2003) for milk, protein, and fat yield were 0.94, 0.92, and 0.90, respectively. Our results confirm previous work in Australia (Hayes et al., 2003; Haile-Mariam et al., 2008), which found that G × E exists at extreme THI values (i.e., THI > 70) for milk traits, but we observed greater G × E in the study population. In addition, we observed the greatest G × E for protein yield, whereas Hayes et al. (2003) found that fat yield had the greatest G × E. When considering only first-parity data in our
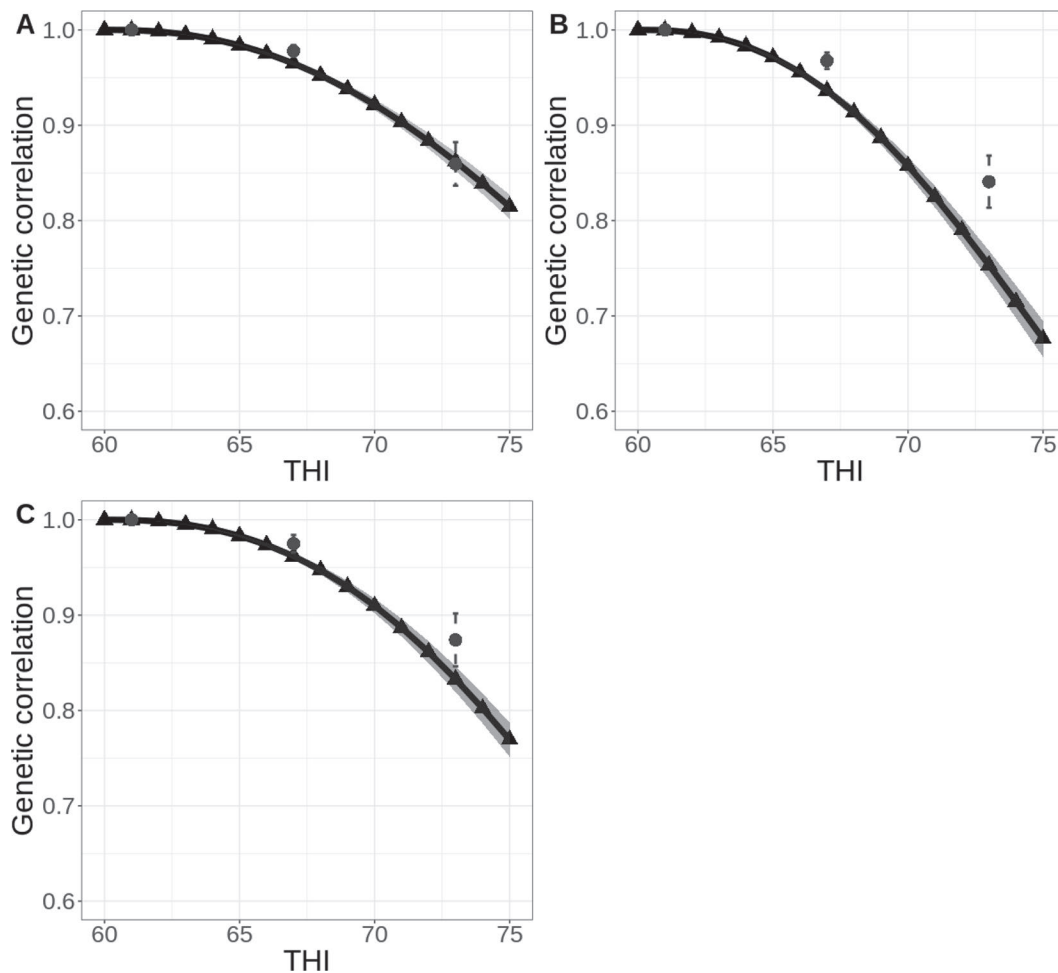
Figure 4. Genetic correlations from multi-trait (MT) and reaction norm (RM) analysis for milk (A), protein (B), and fat (C) yields from analysis of whole data set. Dot points (●) represent genetic correlations and their SE (vertical lines) for yields between 5th [temperature-humidity index (THI) = 61] and 50th (THI = 67), 5th and 95th (THI = 73), and 50th and 95th percentiles of the THI from the MT. Lines represent genetic correlations for yields between THI at 60 and all THI values ≥60 from RM. Shaded areas are SE from RM.

RM analysis, as in Hayes et al. (2003), the correlations changed slightly with estimates for the 5th versus the 95th percentile of 0.89 (milk), 0.84 (protein), and 0.85 (fat), which further confirms greater G × E in the data used in our study. Later-parity cows appear to be more susceptible to heat stress than first-parity cows (Aguilar et al., 2009; Bernabucci et al., 2014), and greater G × E was expected when analyzing data sets that include multiple lactations in our study.

Despite large variability in climatic differences in Australia, most dairying is concentrated in the state of Victoria, which typically experiences moderate heat loads; around two-thirds of the Australian national dairy herd data comes from this state (DataGene, 2016; Nguyen et al., 2016). For example, of the total national herd statistics for 2015 to 2016, the number of herds in Victoria were 1,756 (approx. 65% of the national herd

data records for 2015 to 2016), as opposed to 183 (approx. 7%) from Queensland (DataGene, 2016). This is reflected in our study, in which the milk tests obtained under high heat loads (i.e., THI > 70) accounted for approximately 10% of the total records, which means that the effect of the upper extreme of the THI trajectory in proportion to the total data is small.

However, our results demonstrate that G × E due to heat stress in Australia dairy cows is becoming more important. This is evidenced by the decrease in the genetic correlations among milk traits over time, with greater decreases for protein (0.84 ± 0.016 vs. 0.79 ± 0.02) and fat (0.86 ± 0.02 vs. 0.81 ± 0.02) than for milk yield (0.89 ± 0.01 vs. 0.87 ± 0.01). The increasing G × E in this study could be associated, at least in part, to the greater sensitivity to heat stress in the study population. Nguyen et al. (2017) noted that the genetic
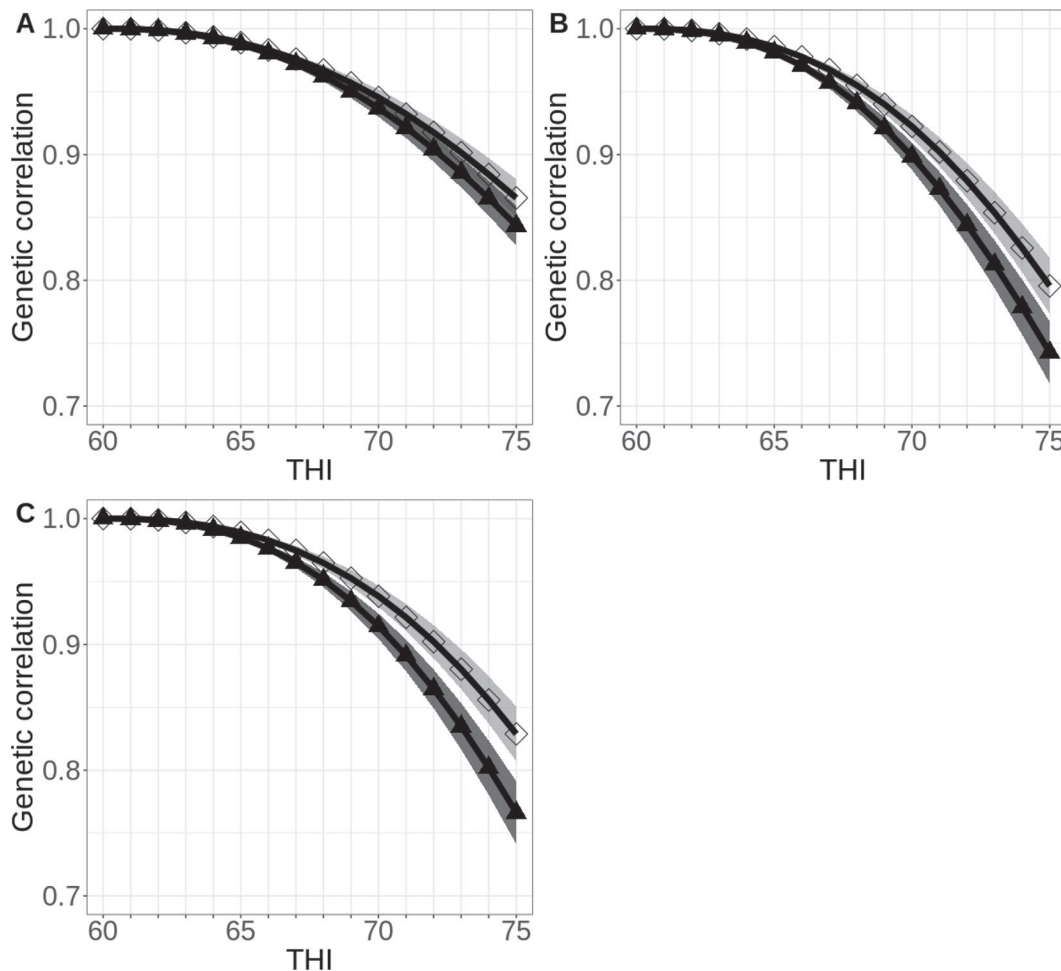

Figure 5. Genetic correlations from reaction norm analysis for milk (A), protein (B), and fat (C) yields from 2003 to 2008 (□) and from 2009 to 2017 (▲). Lines represent genetic correlations for milk traits between temperature-humidity index (THI) at 60 and all THI ≥60. Shaded areas are approximate SE.

merit for heat tolerance has been declining over the years at a rate of 0.3%/yr in Australian Holstein and Jersey cattle due to continued selection for production traits. Therefore, routine monitoring of G × E is highly recommended, given that climatic perturbations in Australia in terms of increased intensity and frequency of heat events are anticipated to worsen in the next decades (CSIRO and BoM, 2018).

Quantifying Resilience to Heat Stress

Resilient animals are minimally affected by environmental disturbances (Colditz and Hine, 2016; Berghof et al., 2019), and their performances are expected to be comparatively consistent under different heat stress conditions. Our results suggest that the proportion of resilient sires has decreased in recent years (Figure 8),

confirming the increasing relevance of G × E in the study population. For example, the proportion of resilient sires in milk yield tests from 2003 to 2008 was 69%, versus 63% for tests between 2009 and 2017. As noted previously, the increasing G × E could be partly associated with the decline in genetic merit for heat tolerance over the years, resulting from more emphasis on production traits in the breeding goals (Nguyen et al., 2016, 2017). In contrast to resilient sires, the performance of the daughters of sensitive sires fluctuates, and they are likely to produce less under heat stress conditions and may be best suited for milk production in regions with consistently low heat loads. If they are to perform optimally and consistently under high heat load conditions, then a more controlled environment is necessary, such as the provision of shade and diets designed to lower core body temperature.
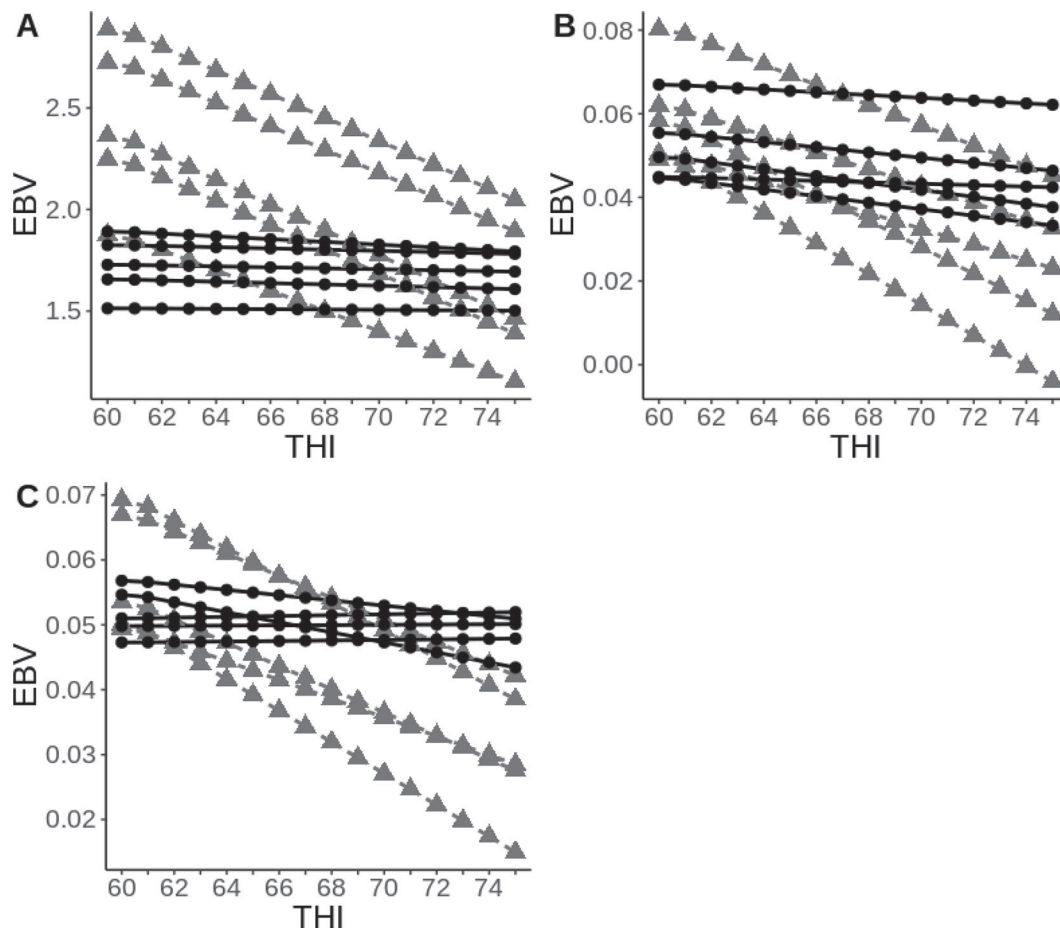

Figure 6. Estimated breeding values (i.e., the slope of reaction norms) for milk (A), protein (B), and fat (C) yields over the temperature-humidity index (THI) for a sample of 10 sires with over 100 daughters with yield records. Gray lines (▲) represent sires with above-average EBV at thermoneutral conditions (THI = 60) and smaller EBV at heat stress conditions (THI = 75), whereas black lines (●) are sires with above-average and stable EBV (i.e., their EBV do not substantially change at THI = 60 and THI = 75).

Implications

Genomic estimated breeding values for heat tolerance were released to the Australian dairy industry in 2017 for selection of animals that are more tolerant to heat stress (Nguyen et al., 2017; Pryce et al., 2018). Our results (genetic correlations ≥0.80) indicate that optimal genetic gain for heat tolerance can be achieved with the current genetic evaluations. Although these results do not justify separate breeding programs, re-ranking can still occur among some sires because of the heterogeneity of genetic variance and heritabilities observed across heat loads (Figures 2 and 3). Moreover, the differences in the patterns of reaction norms (i.e., the slope of EBV) for some sires (Figure 6) suggest that some farmers should be aware that re-ranking may occur, especially in regions such as Northern Victoria and Queensland, which experience broad ranges of weather conditions, with average daily temperature up to 38°C (Nguyen et al., 2016). It may be more beneficial to use resilient sires in such regions. Several approaches are available to incorporate G × E into the genetic evaluation provided to the dairy industry. For example,

Kolmodin et al. (2002) proposed a unique ranking of sires according to the environment in which they will perform, so that farmers can choose bulls that best fit their production environments. Similarly, Bryant et al. (2006) suggested the development of customized EBV for specific environments, and (Nguyen et al., 2017) proposed a selection tool that could balance the profit index and heat tolerance. More recently, some authors have demonstrated opportunities for including resilience in breeding goals (Mulder, 2016; Friggens et al., 2017; Berghof et al., 2019). Berghof et al. (2019) proposed using the slope of the reaction norms as an indicator trait for resilience to heat stress and calculating economic values for selection index based on the reduced labor costs associated with the management of heat-stressed animals. Our study shows the existence of genetic variation in sire sensitivity to heat stress, which could be exploited in the selection of resilient animals or to optimize breeding programs for different environments. Currently, economic drivers in the Australian national selection index (Balanced Performance Index; BPI) favors production traits, which account for about 50% of the selection pressure, and does not include
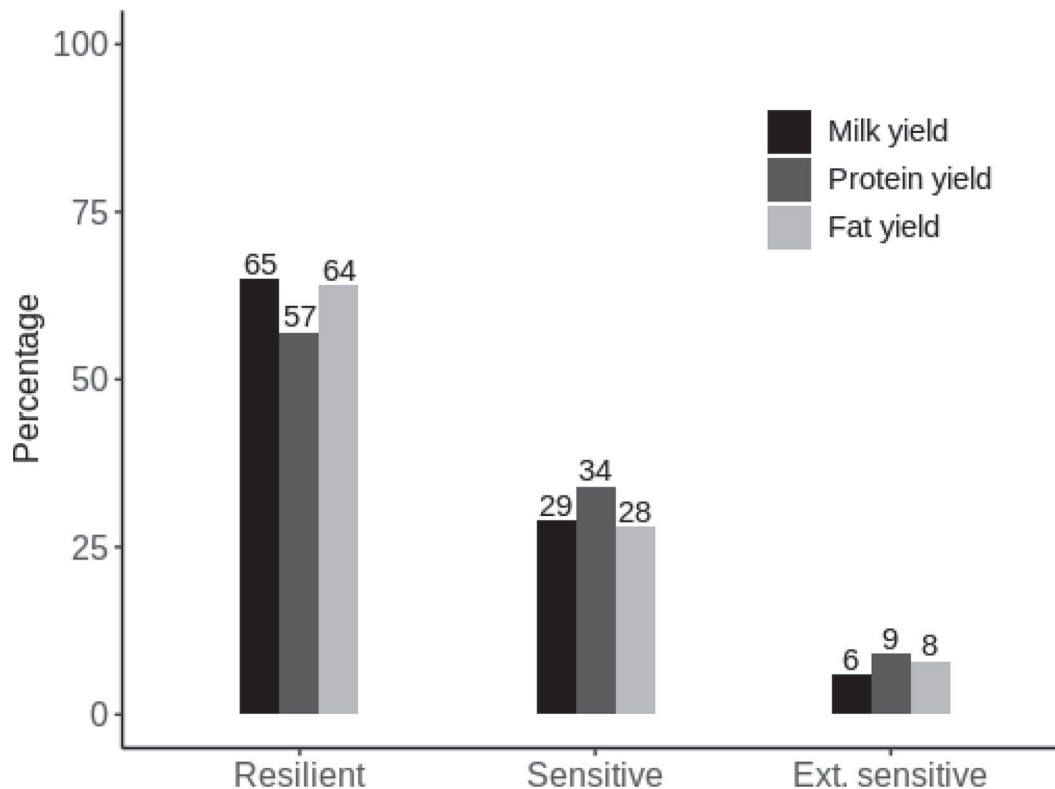
Figure 7. Proportion of resilient, heat sensitive, and extremely (ext.) sensitive sires for milk, protein, and fat yields, grouped based on the absolute values of the slope (S) of the reaction norms, from analysis of whole data set: $|S_i| < \sigma_S$, resilient sires; $\sigma_S \leq |S_i| < 2\sigma_S$, sensitive sires; and $|S_i| \geq 2\sigma_S$, extremely sensitive sires. Values on the tops of the bars are actual percentages.

heat tolerance (Byrne et al., 2016). However, it is worth considering whether heat tolerance should be included in the breeding goal to minimize G × E and thereby improve resilience.

In this study, we considered only the magnitude of re-ranking of genotypes across climate covariate (THI) with respect to production traits. It is also important to consider other traits, because sire selection indices in Australia are designed to include different aspects of farm profitability, including production, fertility, health, functional, and type as well as efficiency traits (Byrne et al., 2016). Selecting for resilience would be advantageous if the desire is to simultaneously achieve an optimal level of ES for multiple traits. Therefore, it would also be useful to quantify ES in other traits, such as fertility, health, functional, and efficiency traits. In this regard, G × E for fertility traits in relation to heat

stress merits investigation and will be a subject of a subsequent study.

CONCLUSIONS

Our results demonstrate that G × E exists at extreme heat stress conditions (genetic correlations for all the milk traits were ≥0.80) but currently, the extent of re-ranking for most of the sires may not justify separate genetic evaluations for high heat stress environments. In addition, we observed lower correlation estimates for first-parity milk tests from recent years (i.e., 2009 and 2017) than for milk tests from earlier years (i.e., 2003 and 2008), which suggests that the level of G × E is increasing in the study population and should be regularly monitored, especially considering the anticipated increase in climate changes. The reaction norms for sire
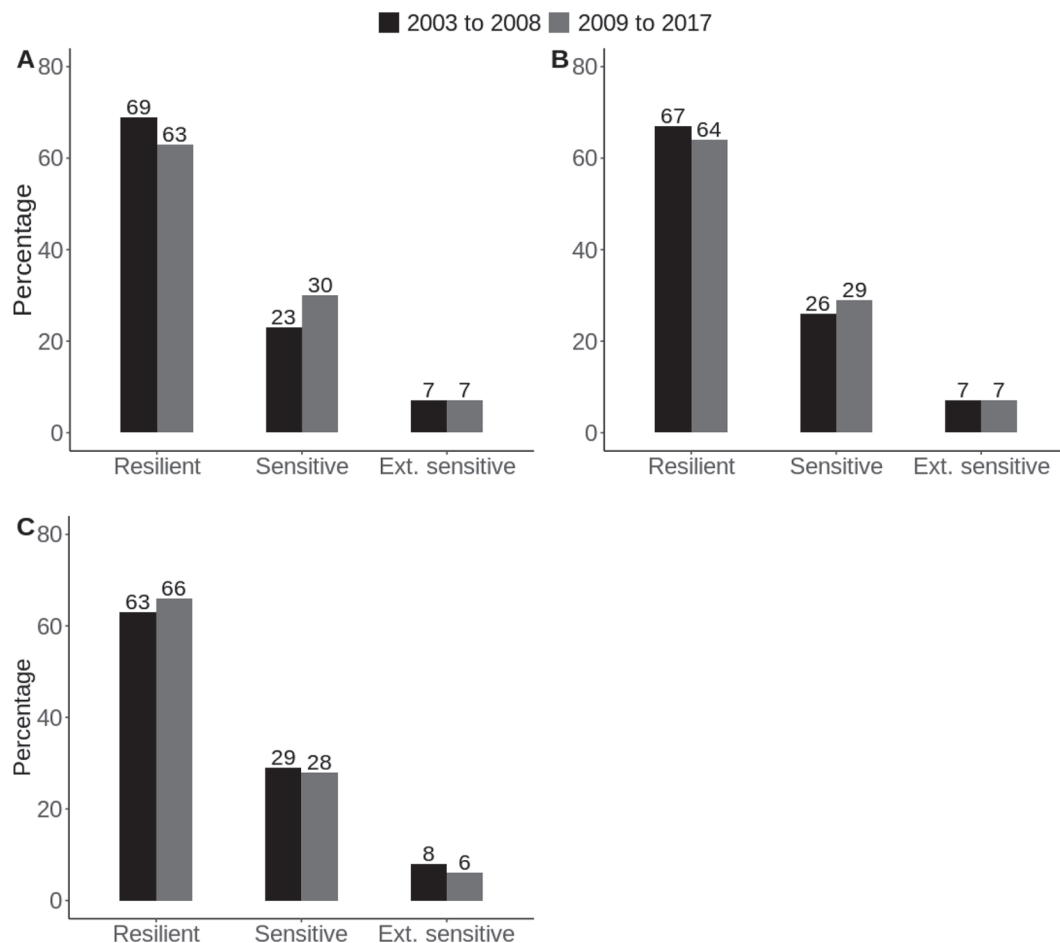
Figure 8. Proportion of resilient, heat sensitive, and extremely (ext.) sensitive sires for milk (A), protein (B), and fat yields (C) of first-parity milk tests from 2003 to 2008 and 2009 to 2017. Sires are grouped based on absolute values of the slope (S) of the reaction norms: $|S_i| < \sigma_S$, resilient sires; $\sigma_S \leq |S_i| < 2\sigma_S$, sensitive sires, and $|S_i| \geq 2\sigma_S$, extremely sensitive sires. Values on the tops of the bars are actual percentages.

EBV along the trajectory of THI indicates that genetic variations exist, which can be used to select animals that perform optimally in different environments.

ACKNOWLEDGMENTS

REFERENCES

Aguilar, I., I. Misztal, and S. Tsuruta. 2009. Genetic components of heat stress for dairy cattle with multiple lactations. J. Dairy Sci. 92:5702–5711. https://doi.org/10.3168/jds.2008-1928.

Berghof, T. V. L., M. Poppe, and H. A. Mulder. 2019. Opportunities to improve resilience in animal breeding programs. Front. Genet. 9:692. https://doi.org/10.3389/fgene.2018.00692.

Bernabucci, U., S. Biffani, L. Buggiotti, A. Vitali, N. Lacetera, and A. Nardone. 2014. The effects of heat stress in Italian Holstein dairy cattle. J. Dairy Sci. 97:471–486. https://doi.org/10.3168/jds.2013-6611.

Bohlouli, M., S. Alijani, S. Naderi, T. Yin, and S. König. 2019. Prediction accuracies and genetic parameters for test-day traits from genomic and pedigree-based random regression models with or without heat stress interactions. J. Dairy Sci. 102:488–502. https://doi.org/10.3168/jds.2018-15329.

Bohlouli, M., J. Shodja, S. Alijani, and A. Eghbal. 2013. The relationship between temperature-humidity index and test-day milk yield of Iranian Holstein dairy cattle using random regression model. Livest. Sci. 157:414–420. https://doi.org/10.1016/j.livsci.2013.09.005.

Bohmanova, J., I. Misztal, S. Tsuruta, H. Norman, and T. Lawlor. 2005. National genetic evaluation of milk yield for heat tolerance of United States Holsteins. Interbull Bull. 160.

Brügemann, K., E. Gernand, U. Von Borstel, and S. König. 2011. Genetic analyses of protein yield in dairy cows applying random regression models with time-dependent and temperature × humidity-dependent covariates. J. Dairy Sci. 94:4129–4139. https://doi.org/10.3168/jds.2010-4063.

Bryant, J., N. López-Villalobos, J. Pryce, C. Holmes, and D. Johnson. 2006. Reaction norms used to quantify the responses of New Zealand dairy cattle of mixed breeds to nutritional environment. N. Z. J. Agric. Res. 49:371–381. https://doi.org/10.1080/00288233.2006.9513727.

Byrne, T. J., B. F. S. Santos, P. R. Amer, D. Martin-Collado, J. E. Pryce, and M. Axford. 2016. New breeding objectives and selection indices for the Australian dairy industry. J. Dairy Sci. 99:8146–8167. https://doi.org/10.3168/jds.2015-10747.

Carabaño, M. J., K. Bachagha, M. Ramón, and C. Díaz. 2014. Modeling heat stress effect on Holstein cows under hot and dry conditions: Selection tools. J. Dairy Sci. 97:7889–7904. https://doi.org/10.3168/jds.2014-8023.

Carabaño, M. J., M. Ramón, C. Díaz, A. Molina, M. D. Pérez-Guzmán, and J. M. Serradilla. 2017. Breeding and genetics symposium: breeding for resilience to heat stress effects in dairy ruminants. A comprehensive review. J. Anim. Sci. 95:1813–1826. https://doi.org/10.2527/jas2016.1114.

Colditz, I. G., and B. C. Hine. 2016. Resilience in farm animals: Biology, management, breeding and implications for animal welfare. Anim. Prod. Sci. 56:1961–1983. https://doi.org/10.1071/AN15297.

CSIRO and BoM. 2018. State of the climate 2018, CSIRO. Accessed Aug. 24, 2018. https://www.csiro.au∼/media/OnA/Files/State-of-the-Climate-2018-CSIRO-BOM-Dec2018.pdf.

Dairy Australia. 2016. Australian dairy industry in focus—2016. Accessed Jul. 12, 2018. https://www.dairyaustralia.com.au/-/media/dairyaustralia/documents/about-dairy-australia/key-publications/2016-in-focus.pdf.

DataGene. 2016. Australian dairy herd improvement report. Accessed Jun. 17, 2018. https://datagene.com.au/sites/default/files/Upload%20Files/Australian%20Dairy%20Herd%20Improvement%20%20Report%202016.pdf.

de Jong, G., and P. Bijma. 2002. Selection and phenotypic plasticity in evolutionary biology and animal breeding. Livest. Prod. Sci. 78:195–214. https://doi.org/10.1016/S0301-6226(02)00096-9.

Falconer, D. S., and T. F. C. MacKay Longman. 1996. Introduction to Quantitative Genetics. 4th ed. D. S. Falconer and T. F. C. Mackay Longman, ed. Pearson Education Limited, Essex, UK.

Fischer, T. M., A. R. Gilmour, and J. H. J. van der Werf. 2004. Computing approximate standard errors for genetic parameters derived from random regression models fitted by average information REML. Genet. Sel. Evol. 36:363. https://doi.org/10.1186/1297-9686-36-3-363.

Friggens, N. C., F. Blanc, D. P. Berry, and L. Puillet. 2017. Deciphering animal robustness. A synthesis to facilitate its use in livestock breeding and management. Animal 11:2237–2251. https://doi.org/10.1017/S175173111700088X.

Gerber, P. J., H. Steinfeld, B. Henderson, A. Mottet, C. Opio, J. Dijkman, A. Falcucci, and G. Tempio. 2013. Tackling climate change through livestock: A global assessment of emissions and mitigation opportunities. Food and Agriculture Organization of the United Nations (FAO), Rome, Italy.

Gilmour, A., B. Gogel, B. Cullis, S. Welham, and R. Thompson. 2015. ASReml User Guide Release 4.1 Structural Specification. VSN International Ltd., Hemel Hempstead, UK.

Haile-Mariam, M., M. Carrick, and M. Goddard. 2008. Genotype by environment interaction for fertility, survival, and milk production traits in Australian dairy cattle. J. Dairy Sci. 91:4840–4853. https://doi.org/10.3168/jds.2008-1084.

Hammami, H., J. Vandenplas, M.-L. Vanrobays, B. Rekik, C. Bastin, and N. Gengler. 2015. Genetic analysis of heat stress effects on yield traits, udder health, and fatty acids of Walloon Holstein cows. J. Dairy Sci. 98:4956–4968. https://doi.org/10.3168/jds.2014-9148.

Hayes, B. J., M. Carrick, P. Bowman, and M. E. Goddard. 2003. Genotype × environment interaction for milk production of daughters of Australian dairy sires from test-day records. J. Dairy Sci. 86:3736–3744. https://doi.org/10.3168/jds.S0022-0302(03)73980-0.

Kolmodin, R., E. Strandberg, P. Madsen, J. Jensen, and H. Jorjani. 2002. Genotype by environment interaction in Nordic dairy cattle studied using reaction norms. Acta Agric. Scand. A Anim. Sci. 52:11–24. https://doi.org/10.1080/09064700252806380.

Mattar, M., L. Silva, M. Alencar, and F. Cardoso. 2011. Genotype × environment interaction for long-yearling weight in Canchim cattle quantified by reaction norm analysis. J. Anim. Sci. 89:2349–2355. https://doi.org/10.2527/jas.2010-3770.

Misztal, I., T. Strabel, J. Jamrozik, E. A. Mäntysaari, and T. H. E. Meuwissen. 2000. Strategies for estimating the parameters needed for different test-day models. J. Dairy Sci. 83:1125–1134. https://doi.org/10.3168/jds.S0022-0302(00)74978-2.

Mulder, H. A. 2016. Genomic selection improves response to selection in resilience by exploiting genotype by environment interactions. Front. Genet. 7. https://doi.org/10.3389/fgene.2016.00178.

Nguyen, T. T., P. J. Bowman, M. Haile-Mariam, G. J. Nieuwhof, B. J. Hayes, and J. E. Pryce. 2017. Implementation of a breeding value for heat tolerance in Australian dairy cattle. J. Dairy Sci. 100:7362–7367. https://doi.org/10.3168/jds.2017-12898.

Nguyen, T. T., P. J. Bowman, M. Haile-Mariam, J. E. Pryce, and B. J. Hayes. 2016. Genomic selection for tolerance to heat stress in Australian dairy cattle. J. Dairy Sci. 99:2849–2862. https://doi .org/10.3168/jds.2015-9685.

Polsky, L., and M. A. von Keyserlingk. 2017. Invited review: Effects of heat stress on dairy cattle welfare. J. Dairy Sci. 100:8645–8657. https://doi.org/10.3168/jds.2017-12651.

Pryce, J. E., T. T. T. Nguyen, M. Axford, G. Nieuwhof, and M. Shaffer. 2018. Symposium review: Building a better cow—The Australian experience and future perspectives. J. Dairy Sci. 101:3702–3713. https://doi.org/10.3168/jds.2017-13377.

Ravagnolo, O., and I. Misztal. 2000. Genetic component of heat stress in dairy cattle, parameter estimation. J. Dairy Sci. 83:2126–2130. https://doi.org/10.3168/jds.S0022-0302(00)75095-8.

Ravagnolo, O., I. Misztal, and G. Hoogenboom. 2000. Genetic component of heat stress in dairy cattle, development of heat index function. J. Dairy Sci. 83:2120–2125. https://doi.org/10.3168/jds .S0022-0302(00)75094-6.

Renaudeau, D., A. Collin, S. Yahav, V. De Basilio, J. Gourdine, and R. Collier. 2012. Adaptation to hot climate and strategies to alleviate heat stress in livestock production. Animal 6:707–728. https://doi .org/10.1017/S1751731111002448.

Santana, M. L. Jr., R. J. Pereira, A. B. Bignardi, A. E. Vercesi Filho, A. Menendez-Buxadera, and L. El Faro. 2015. Detrimental effect of selection for milk yield on genetic tolerance to heat stress in purebred Zebu cattle: genetic parameters and trends. J. Dairy Sci. 98:9035–9043. https://doi.org/10.3168/jds.2015-9817.

Strabel, T., J. Szyda, E. Ptak, and J. Jamrozik. 2005. Comparison of random regression test-day models for Polish Black and White cattle. J. Dairy Sci. 88:3688–3699. https://doi.org/10.3168/jds .S0022-0302(05)73055-1.

Su, G., M. S. Lund, and D. Sorensen. 2007. Selection for litter size at day five to improve litter size at weaning and piglet survival rate1. J. Anim. Sci. 85:1385–1392. https://doi.org/10.2527/jas.2006-631.

West, J. W. 2003. Effects of heat-stress on production in dairy cattle. J. Dairy Sci. 86:2131–2144. https://doi.org/10.3168/jds.S0022 -0302(03)73803-X.

Yousef, M. K. 1985. Stress Physiology in Livestock. Volume I. Basic Principles. CRC Press, Boca Raton, FL.

ORCIDS

E. K. Cheruiyot ● https://orcid.org/0000-0002-6403-7967

M. Haile-Mariam ● https://orcid.org/0000-0001-5476-7475

J. E. Pryce ● https://orcid.org/0000-0002-1397-1282

Chapter 4: New loci and neuronal pathways for resilience to heat stress in cattle

Evans K. Cheruiyot[1,2], Mekonnen Haile-Mariam[2*], Benjamin G. Cocks[1,2], Iona M. MacLeod[2], Ruidong Xiang[2,3], Jennie E. Pryce[1,2]

[1]School of Applied Systems Biology, La Trobe University, Bundoora, Victoria 3083, Australia

[2]Agriculture Victoria Research, AgriBio, Centre for AgriBiosciences, Bundoora, Victoria 3083, Australia

[3]Faculty of Veterinary & Agricultural Science, The University of Melbourne, Parkville 3052, Victoria, Australia

scientific reports

OPEN

New loci and neuronal pathways for resilience to heat stress in cattle

Evans K. Cheruiyot[1,2], Mekonnen Haile-Mariam[2✉], Benjamin G. Cocks[1,2], Iona M. MacLeod[2], Ruidong Xiang[2,3] & Jennie E. Pryce[1,2]

While understanding the genetic basis of heat tolerance is crucial in the context of global warming's effect on humans, livestock, and wildlife, the specific genetic variants and biological features that confer thermotolerance in animals are still not well characterized. We used dairy cows as a model to study heat tolerance because they are lactating, and therefore often prone to thermal stress. The data comprised almost 0.5 million milk records (milk, fat, and proteins) of 29,107 Australian Holsteins, each having around 15 million imputed sequence variants. Dairy animals often reduce their milk production when temperature and humidity rise; thus, the phenotypes used to measure an individual's heat tolerance were defined as the rate of milk production decline (slope traits) with a rising temperature–humidity index. With these slope traits, we performed a genome-wide association study (GWAS) using different approaches, including conditional analyses, to correct for the relationship between heat tolerance and level of milk production. The results revealed multiple novel loci for heat tolerance, including 61 potential functional variants at sites highly conserved across 100 vertebrate species. Moreover, it was interesting that specific candidate variants and genes are related to the neuronal system (*ITPR1*, *ITPR2*, and *GRIA4*) and neuroactive ligand–receptor interaction functions for heat tolerance (*NPFFR2*, *CALCR*, and *GHR*), providing a novel insight that can help to develop genetic and management approaches to combat heat stress.

Heat stress from rising global temperatures is an issue of growing importance across tropical and temperate zones affecting humans, livestock, wildlife, and plants. A recent study[1] indicates that many people are now exposed to harmful heat, and this has risen by more than twofold when compared to the pre-industrial climates (i.e., 95 vs. 275 million people), with future projections showing that over 1 billion people will experience an even greater impact of heat within the next 50 years[2]. In livestock, the annual temperature–humidity values that rise above thresholds considered to be comfortable have been increasing in many regions including Australia, the USA, Canada, and parts of Europe[3,4], making heat stress a multimillion-dollar issue in the livestock industry that compromises production (reduced growth, milk, eggs, etc.) and reproduction leading to economic losses[5].

The thermoregulatory capacities of mammals and plants to cope with extreme heat have been studied for decades. Genetic variation of thermoregulation during heat stress exists within species, including cattle breeds, with the literature indicating that tropical breeds, such as Zebu (*Bos indicus*), have a better tolerance to temperature and humidity than cattle from temperate zones (e.g., Holsteins), in part, due to the lower productivity of Zebu cattle[6]. Temperate breeds also show genetic variation in heat tolerance; for example, New Zealand Holsteins appear to exhibit higher reductions in milk yield in hotter climates than Jerseys or crossbreds[7]. While it is not fully understood why animals differ in their thermotolerance, it is hypothesised to be due to a myriad of biological mechanisms; including cellular, morphological (coat color, coat length, etc.), behavioural (e.g., feed and water intake, standing and lying time), as well as neuro-endocrine systems. See comprehensive review by[8] for more information. Notably, the molecular basis for differences in these adaptive responses within various mammalian species is still largely unknown.

Dairy cattle particularly Holsteins are excellent and convenient model for enhancing our knowledge on the molecular aspects of heat tolerance in mammals for two main reasons: (1) large phenotype datasets needed to study heat tolerance, as well as extensive genomic information, are available; (2) they have been genetically selected mainly for high milk production over many years, offering an opportunity to understand the genetic basis for coping with both environmental and elevated metabolic-heat stress associated with increased milk production.

[1]School of Applied Systems Biology, La Trobe University, Bundoora, VIC 3083, Australia. [2]Agriculture Victoria Research, Centre for AgriBiosciences, AgriBio, Bundoora, VIC 3083, Australia. [3]Faculty of Veterinary and Agricultural Science, The University of Melbourne, Parkville, VIC 3052, Australia. ✉email: Mekonnen.HaileMariam@agriculture.vic.gov.au

1

83

Trait	Milk intercept traits[a]		Heat tolerance milk slope traits[b]	
	AG ± SE	h² ± SE	AG ± SE	h² ± SE
Milk yield (liters)	381.0 (14.36)	0.36 (0.01)	4.80 (0.20)	0.23 (0.01)
Fat yield (kg)	38.78 (1.64)	0.30 (0.01)	0.56 (0.02)	0.21 (0.01)
Protein yield (kg)	18.60 (0.89)	0.24 (0.01)	0.47 (0.02)	0.20 (0.01)

Table 1. Additive genetic variance (AG) and genomic heritability (h²) for milk intercept and heat tolerance slope traits estimated for 29,107 cows based on 50 k SNP panel. *SE* Standard errors. [a]Represents the level of milk production of cows. [b]Heat tolerance proxy-phenotypes.

The development of methods to describe heat tolerance in cattle has been an active research area for many years. Measuring changes in core body temperature (e.g., rectal, vaginal, rumen temperature, etc.), thermal indices (e.g., temperature–humidity index (THI)) are some of the ways to assess thermal adaptations and performance in animals. Ravagnolo et al.[9] pioneered using daily milk yield and temperature–humidity data to measure variability in the rate of decline in milk yield associated with variability in response to heat stress. This method has been widely adopted due to the availability of large datasets from routine recording in dairy farms, e.g.,[3]. Heat tolerance in dairy cattle measured using rectal temperatures or the rate of milk yield decline is partly under genetic control, having a low (0.1) to moderate heritability (0.30)[3,9,10], which makes it amenable to selection. As such, considerable research has been undertaken to provide breeding solutions for heat stress, which is already a feature of dairy cattle breeding programmes in some parts of the world, e.g., Australia[3]. Identifying specific genetic variants that increase tolerance to heat may help to improve dairy breeding programmes in addition to improving our knowledge of the thermal biology in other mammals. However, except for mutations in the SLICK locus[11], the identification of the specific genetic variants for heat tolerance in cattle and other species has, in most cases, remained elusive, in part due to many reasons, including the sample size used in past studies[12–15].

Having a large sample size is particularly important for identifying rare causal variants with medium-sized effects and common variants with small effects. As sample size increases, the loci significantly associated with complex traits are expected to increase, as demonstrated for the human height[16]. Several selection signature work e.g.,[17,18] and genome-wide association studies (GWAS) using Single nucleotide polymorphisms (SNPs) have been conducted over the last decade to identify candidate causal genes for various heat tolerance traits (rectal temperature, heart rate, sweating rate, rate of milk yield decline, etc.) in dairy cattle[12–15] and pigs[19]. However, these GWAS were underpowered, with the largest sample size to d te of around 5000 animals[12,13]. These studies have also used standard industry SNP panels of random genome-wide markers, either 50 k or 600 k SNPs, leading to inconsistencies and poor replication of the results. Although these studies have identified multiple significant variants associated with heat stress in animals, none were established to be causal mutations.

Here, we performed a GWAS using milk production records of 29,107 Holstein cows, each having over 15 million sequence variants that were imputed from various lower density SNP chips to whole-genome sequence using a reference dataset of sequences from the Run7 of 1000 Bull Genome Project[20]. The specific aims of the current study were to: (1) perform single-trait GWAS to identify genomic variants associated with sensitivity of milk traits (milk, protein, and fat) to heat stress; (2) combine single-trait GWAS results in a multi-trait meta-analysis to boost the power and identify pleiotropic variants associated with all the milk traits; and (3) conduct post-GWAS pathway analysis using the list of candidate genes identified in single-trait GWAS and meta-analysis to elucidate biological mechanisms underlying heat tolerance.

Results

Descriptive statistics and genomic heritability of the study phenotypes.

The average yield and their corresponding standard deviation (in brackets) of milk (in liters), fat (kg), and proteins (kg) used in our study was 25.85 (8.19), 0.98 (0.30), and 0.85 (0.26), respectively. The heat tolerance proxy-phenotypes (i.e., slope traits) and intercepts (representing level of milk production) that were derived from the milk traits are in Table 1. The slope traits derived from the milk, fat, and protein yield using reaction norm models on a function of the temperature–humidity index (THI) were defined as follows: heat tolerance milk (HTMYslope), fat (HTFYslope), and protein (HTPYslope) yield slope traits, respectively. On the other hand, the intercept solutions from the reaction norm models – representing the level of milk production were defined as milk (MYint), fat (FYint), and protein (PYint) yield intercept traits. The values for slopes (no units) for HTMYslope, HTFYslope, and HTPYslope ranged between [− 36.80 to 27.17], [− 11.39 to 9.0] and [− 8.91 to 9.31], with values at 25% and 75% quartiles of [− 0.98 and 0.90], [− 2.76 and 2.66], and [− 1.03 and 0.95], respectively. Note that the values for milk, fat, and protein yield have been scaled by a factor of 10, 100 and 100, respectively (see "Methods"). The genomic heritability estimates for the intercept traits were high [0.36 ± 0.01 (MYint), 0.30 ± 0.01 (FYint), 0.24 ± 0.01 (PYint)] compared to slope traits [0.23 ± 0.01 (HTMYslope), 0.21 ± 0.01 (HTFYslope), 0.20 ± 0.01 (HTPYslope)] (Table 1). The phenotypic correlations between the intercept and slope traits were high, with values of -0.71 (MYint versus HTMYslope), -0.77 (FYint versus HTFYslope), and -0.83 (PYint versus HTPYslope), suggesting that lower producing cows have a smaller reduction in their yield as the THI increases. The Pearson correlations of slope solutions from the reaction norm model were 0.90 (HTMYslope versus HTPYslope), 0.56 (HTMYslope versus HTFYslope) and 0.62 (HTPYslope versus HTFYslope).

	Single-trait GWAS for intercept traits			Single-trait GWAS for slope traits			Multi-trait meta-analysis[a]	
	MYint	FYint	PYint	HTMYslope	HTFYslope	HTPYslope	Meta intercept	Meta slope
Significant (p < 1E–05)	9344 [72]	7844 [28]	4195 [49]	6061 [37]	8684 [21]	2998 [30]	51,568 [100]	40,220 [65]
FDR < 0.10	16,469 [188] (p = 1E–04)	11,469 [98] (p = 7E–05)	5285 [78] (p = 3E–05)	9172 [109] (p = 6E–05)	12,619 [98] (p = 8E–05)	3310 [51] (p = 2E–05)	108,934 [347] (p = 7E–04)	77,499 [293] (p = 5E–04)

Table 2. Number of SNPs identified at a *p* value of < 1E–05 (significant), and false discovery rate (FDR < 0.10) for QTL discovery cows (N = 29,107) based on 15 million imputed-whole genome sequence variants. MYint, Milk yield intercept; FYint, Fat yield intercept; PYint, protein yield intercept; HTMYslope, heat tolerance milk yield slope; HTFYslope, heat tolerance fat yield slope; HTPYslope, heat tolerance protein yield slope. [a]Multi-trait meta-analysis of single-trait GWAS was performed for intercept and slope traits following[21]. Values in square brackets are the number of lead SNPs defined as the top significant SNP within 5 Mb non-overlapping windows across the chromosome. The FDR was calculated following the method described by[21], where the *p* value in the brackets represents the cut-off threshold equivalent to FDR < 0.10 for each trait.

Single-trait GWAS for intercept and slope traits.

The number of significant SNPs was generally higher for intercept than slope traits at the *p* value thresholds tested (Table 2). At a stringent *p* value of < 1E–05, the false discovery rate (FDR) varied between 0.02 and 0.03 for intercept and 0.02 and 0.05 for slope traits. The number of significant independent QTL (based on the number of 5 Mb non-overlapping windows across the chromosome with at least one significant SNP) ranged from 28 to 72 for intercept traits and from 21 to 37 for slope traits. At a relaxed cut-off threshold, where the FDR was < 0.10, the number of significant QTLs from single-trait GWAS ranged from 78 to 188 (intercept traits) and from 51 to 109 (slope traits).

The number of significant (*p* < 1E–05) QTLs (i.e., 5 Mb windows) varied across the three slope traits with greater overlap between HTMYslope and HTPYslope (13 QTLs; 20.6%) compared to HTMYslope and HTFYslope (3 QTLs; 4.8%) (Supplementary Fig. S3). The overlaps were based on whether the lead SNPs (most significant) within QTLs between traits were close (within 1 Mb). Surprisingly, none of the candidate QTLs overlapped between HTFYslope and HTPYslope. The effects of the lead SNPs within QTLs that overlapped between HTMYslope and HTPYslope were generally in the same direction.

Multi-trait meta-analysis of GWAS to detect variants with pleiotropic effects.

Meta-analysis of GWAS results could increase the power of detecting informative variants[21, 22]. Compared to single-trait GWAS, the number of significant independent QTLs (based on 5 Mb windows with at least one significant SNP) was much higher for a multi-trait meta-analysis (Fig. 1; Table 2). At FDR < 0.10, the number of significant independent QTLs from multi-trait meta-analysis was 347 and 293 for intercept and slope traits, respectively (Table 2). At *p* < 1E–05, the number of significant QTLs was 100 (meta-analysis of intercept traits) and 65 (meta-analysis of slope traits). Of the significant QTLs (*p* < 1E–05; N = 65) for meta-analysis of slope traits, 35% (N = 23) overlapped with the candidate QTLs for single-trait GWAS analysis based on whether the lead SNP (most significant) within overlapping QTLs were close (within 1 Mb).

Lead SNPs detected using single-trait GWAS and meta-analysis of slope traits.

The lead SNPs were defined as the most significant SNPs within an independent QTL (i.e., the most significant SNP chosen within 5 Mb windows across the chromosome). Detailed annotation of all the lead SNPs for single slope traits and the meta-analysis (N = 118) detected at the most stringent *p* value cut-off (*p* < 1E–05) are in the Supplementary Table S2.

About half the lead SNPs (51%) for slopes were in relatively low LD (r² < 0.5) with nearby (within 1 Mb region) lead SNPs for intercepts, indicating that they are not strongly associated with the level of milk production. Some lead SNPs mapped within or close to several candidate genes, which have been linked to environmental stress or heat tolerance in animals in previous studies, including *REG3A*[23], *NPFFR2*[24], and *CLSTN2*[25]. Several other lead SNPs mapped close to novel candidate genes that, to our knowledge, have not been described for thermotolerance in previous studies.

However, the remaining lead SNPs (49%) for slopes were in medium to strong LD (r² > 0.50) with nearby (within 1 Mb) lead SNPs identified for intercept traits (Supplementary Fig. S4), suggesting that they affect both traits, which was expected due to the negative phenotypic correlation between heat tolerance and milk production, with estimates in this study of around -0.80. The most significant lead SNPs for heat tolerance (slope traits) that were strongly (LD; r² > 0.8) associated with the level of milk production (intercept traits) mapped close to or are within genomic loci previously reported to have pleiotropic effects on bovine milk production traits, including the *DGAT1*[26,27], *MGST1*[28], and *GHR* gene[29].

Conditional GWAS for slope traits on either the lead SNPs or the intercept traits.

We performed two conditional GWAS for slope traits to confirm whether the top hits (lead SNPs) detected in the first-round of GWAS for the slope traits were in fact discoveries of heat tolerance rather than indicators of milk yield (as the intercept and slope traits are genetically correlated). Of interest was the conditional GWAS analysis on chromosome 14, since the highly significant QTL around 0.5 Mb harbours the DGAT1 gene and the HSF1 (heat shock factor 1) gene, for which the latter has been linked to thermotolerance in Holstein cattle in different coun-
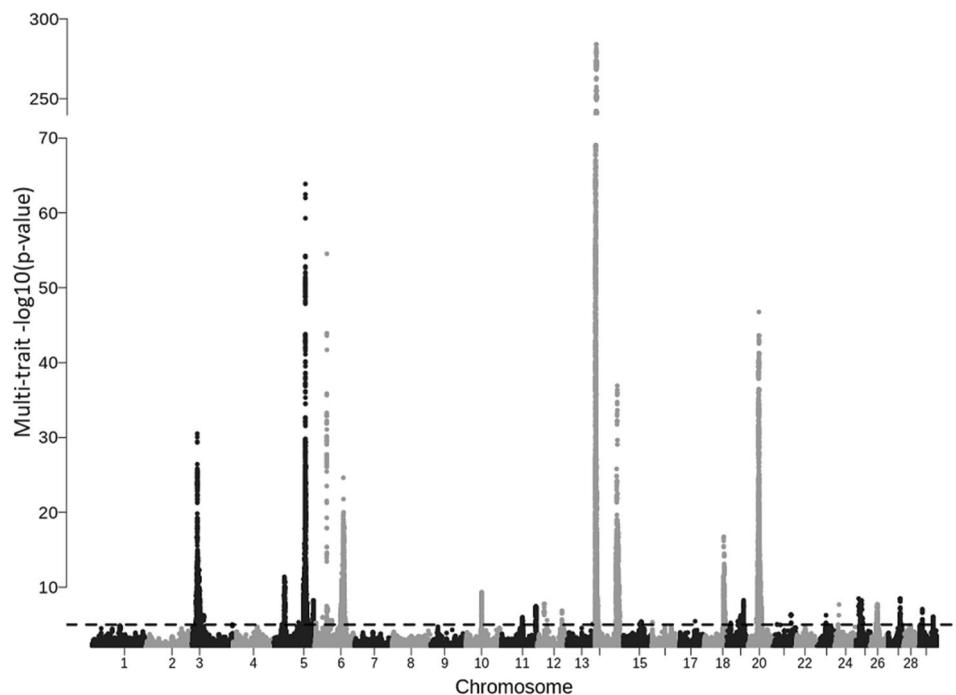
Figure 1. Manhattan plot of *p* values obtained from combining single-trait GWAS results for milk yield slope traits.

tries, including Australia[13], and the USA[14]. Notably, the lead SNPs from the first-round of GWAS for HTMYslope and HTFYslope (Chr14:581,569) and HTPYslope (Chr14:555,701) traits were upstream to *SLC52A2* and a synonymous variant in the *CPSF1* genes, respectively.

Figure 2 shows conditional GWAS results for chromosome 14 (around the region which showed the strongest signal in the first-round of GWAS for the slope traits—here, the conditional analyses were for slope traits on either the lead SNP or the intercept trait. In both approaches, we found that most of the SNPs were no longer significant after conditional analysis. This was the case for HTMYslope and HTPYslope slope traits, suggesting that these SNPs were possibly tagging the lead SNPs for slope traits. The lead SNP was in strong LD ($r^2 > 0.8$) with several other variants around this QTL spanning over 10 genes (Fig. 2), including variants in the HSF1 (heat shock factor 1) gene, which implies that any variant (s) around this region are possible causal mutations for heat tolerance. Nonetheless, the complex LD within this QTL region makes it difficult to pinpoint a putative causal variant (s) for heat tolerance.

Notably, even after fitting the lead SNP in a conditional GWAS analysis, there were still other somewhat significant ($p < 1E–05$; Fig. 2) SNPs remaining for the HTFYslope trait (though not very strong signals; Fig. 2), suggesting that they could be other QTLs for heat tolerance, which were not captured by the lead SNPs identified in the study.

Although the two conditional GWAS strategies (i.e., conditioning slopes on either lead SNP or intercept traits) were generally comparable regarding the strength of the GWAS signals (Fig. 2), we observed a significant (Student's *t* test; $p < 0.001$) difference in the distribution of the GWAS *p* values across slope traits. This is, in part, due to the difference in the two conditional GWAS approaches regarding the covariate fitted in the linear model. We also observed similar findings for the conditional GWAS analysis on chromosome 20 (Supplementary Fig. S5).

By conducting a conditional analysis of slope traits on the intercepts, we detected multiple additional QTL signals (lead SNPs) across the genome at $p < 1E–05$ (Supplementary Fig. S6). However, most of these lead SNPs were associated with a large FDR > 0.10—FDR for each SNP computed following Storey and Tibshirani[30]. Of the few candidate variants (all of which were detected from HTFYslope traits) with FDR < 0.10, the strongest GWAS signal was in BTA 14–1.7 Mb, of which the lead SNP (Chr14:1,726,184) mapped to the downstream region of JRK (Jrk helix-turn-helix protein). Notably, this gene was found to regulate behavioural rhythms in *Drosophila* flies, which is crucial for adaptive response to environmental changes such as temperature variations[31].

When combining conditional GWAS results for slope traits (conditioning on the intercept traits) in the meta-analysis approach, we detected 40 lead SNPs ($p < 1E–05$), all of which associated with low FDR < 0.10 (Supplementary Fig. S7 and Table S3). The mean LD between these 40 lead SNPs and the lead SNPs detected for intercept traits was very low ($r^2 < 0.20$), confirming that the conditional analysis was successful in identifying additional candidate variants for heat tolerance (besides the QTL detected from the first-round of GWAS) that are not strongly associated with the level of milk production. The most significant lead SNP (Chr14:531,267; $p = 9.04E–12$) mapped to the upstream region of the SLC39A4 gene, a member of the solute carrier family, required for intestinal zinc uptake.

4

86
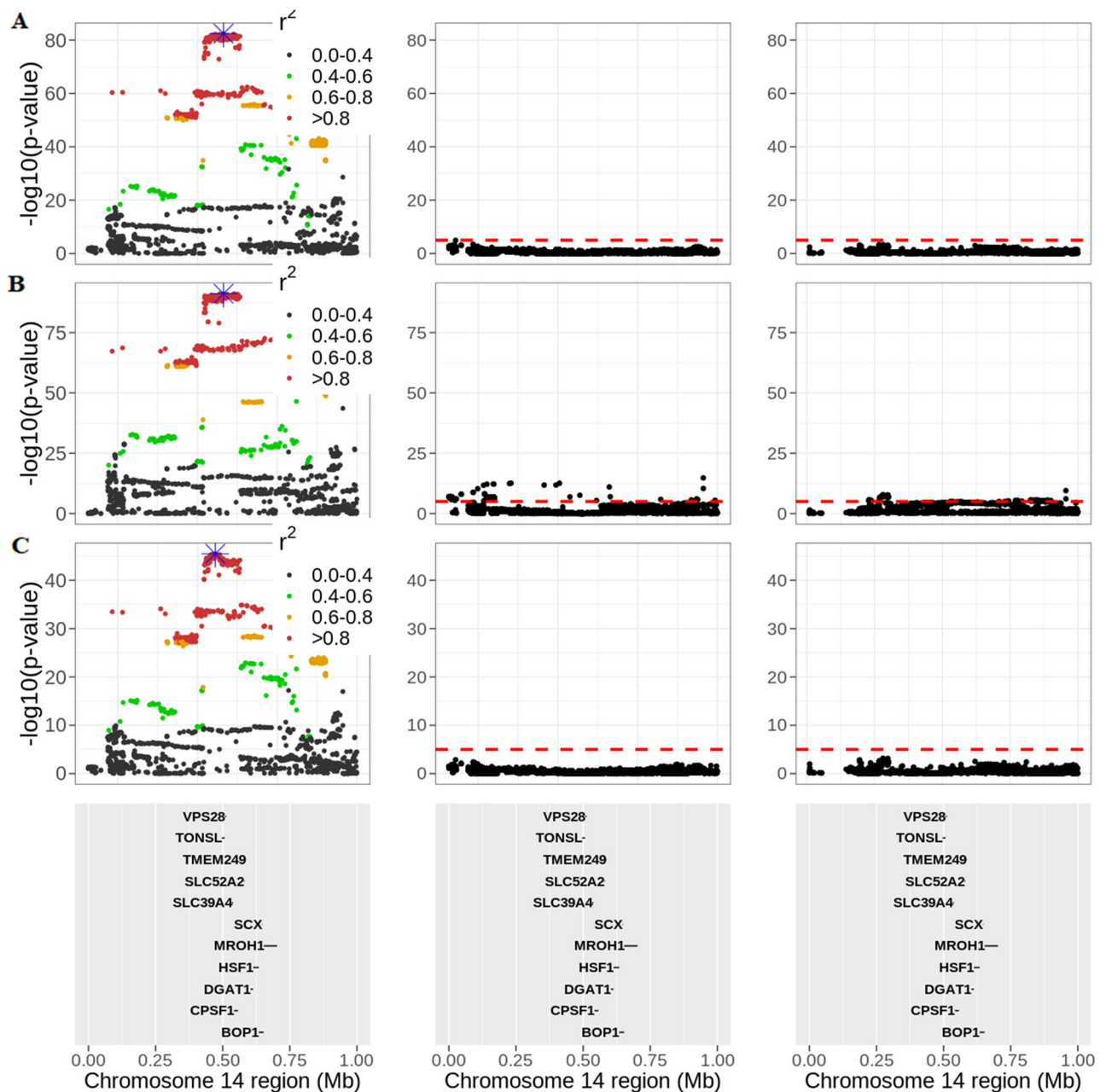
Figure 2. QTL discovery on chromosome 14 at 0 to 1 Mb for heat tolerance milk (HTMYslope; **A**), fat (HTFYslope; **B**), and protein (HTPYslope; **C**) yield slope traits. The three panels represent the GWAS *p* values before conditional analysis (right panel), after conditioning slope traits on the lead SNP (highlighted in blue) defined as the most significant SNP (middle panel), and after conditioning slope traits on the intercept traits (left panel), respectively. The red horizontal dashed line is the GWAS cut-off of $p < 1E{-}05$. The strength of LD (r^2) between the lead SNP (blue color) and all the other SNPs are color-coded accordingly.
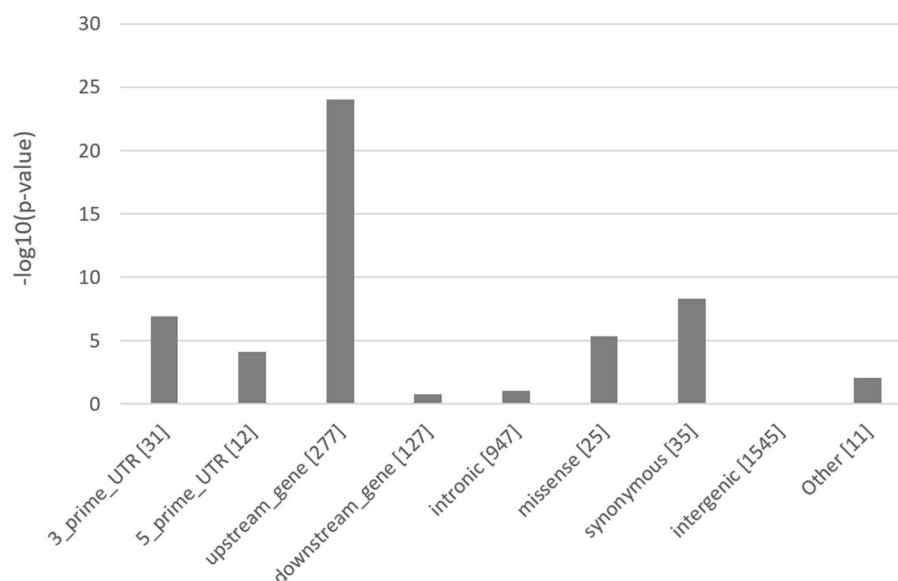
Candidate causal variants for heat tolerance across all analyses. The candidate causal variants for heat tolerance were defined as the lead SNP (most significant SNP within 5 Mb QTL window) plus other significant SNPs in strong LD ($r^2 > 0.8$) with the lead SNP, 500 kb up or downstream of the chromosome. We identified a total of 3010 candidate causal variants for heat tolerance (slope traits) across all the analyses: single-trait GWAS; a meta-analysis of single-trait GWAS results; and meta-analysis of conditional GWAS results for slope traits, most of which were intergenic (N = 1545; 51%) followed by intronic (N = 947; 32%) and upstream (N = 277, 9%) variants (Fig. 3 and Table S1). At least 25 candidate SNPs were missense variants, most (N = 13) of which were in chromosome 14, including two variants (Chr14:615,597 and Chr14:616,087) mapping to HSF1 (heat shock factor 1) gene.

The candidate causal variants for heat tolerance are highly enriched ($p = 8.54E{-}25$) in the upstream gene regions (Fig. 4), which agrees with GWAS for quantitative traits in humans[32], suggesting that they perhaps play
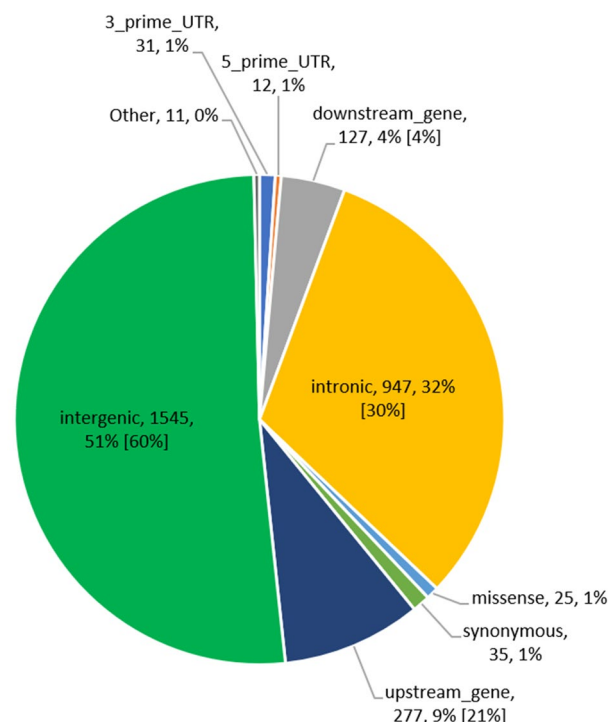
Figure 3. Proportion of candidate causal variants for heat tolerance within different functional classes identified from (**a**) single-trait GWAS, (**b**) meta-analysis, and (**c**) meta-analysis of conditional GWAS results for slope traits. Values in brackets are the proportions of all variants used in the study (~15 million SNPs). Functional classes without values in brackets were represented by a small (<1%) proportion of SNPs in the study dataset.

Figure 4. Enrichment of the candidate causal variants for heat tolerance across functional classes. The values in brackets are the number of variants within each class. The class "Other" includes variants with very small proportions of candidate variants (frameshift, stop-codon, splice variants, etc.).

a functional role in regulating gene expression. As expected, most candidate variants have modifier SnpEff[33] predicted impact (Table S5). Two candidate causal mutations detected from the meta-analysis of slope traits have a high SnpEff predicted impact: (a) a stop-gain mutation (Chr5:31,184,185) causing a premature stop codon in the *LALBA* (lactalbumin alpha) gene and (b) a frameshift mutation (Chr29:41,139,622) in *STX5* (syntaxin-5) gene. The two candidate mutations appear to have a stronger effect on milk production compared to heat tolerance. This

1.46E-04 [12]	4.12E-03 [11]	6.91E-04 [15]	KEGG: Neuroactive ligand-receptor interaction
2.77E-02 [5]		4.0E-03 [8]	KEGG: Glutamatergic synapse
	6.49E-03 [4]	1.87E-03 [5]	KEGG: Citrate cycle (TCA cycle)

SS-slope genes | Meta-slope genes | All-slope genes

Figure 5. Enriched Kyoto Encyclopedia of Gene and Genomes (KEGG) pathways obtained from candidate gene-list for slope traits detected at false discovery rate (FDR < 0.10). SS-slope genes–gene-list from single-trait GWAS; Meta-slope genes–gene-list from multi-trait meta-analysis of slope traits; All-slope genes–combined gene-list from single-trait and meta-analysis. Cells are color-coded according to the strength of the significance for each pathway. Values in brackets are the number of genes within each pathway.

is evidenced by a smaller ($p = 1.39E{-}19$) p value for the stop-gain mutation (Chr5:31,184,185) observed in the meta-analysis of intercept traits compared to the meta-analysis of slope traits ($p = 4.08{-}12$). Similarly, the p value for the frameshift mutation (Chr29:41,139,622) in the *STX5* gene was smaller ($p = 2.06E{-}16$) for the meta-analysis of intercept traits than the meta-analysis of slope traits ($p = 5.06E{-}06$). None of these two candidate stop-gain mutations were significant ($p < 0.05$) following conditional GWAS for slope traits on intercept traits (Table S5).

Using data from[34], which documented over 300 k sequence variants in cattle at highly evolutionarily conserved genome regions across 100 vertebrates (conservation/PhastCon scores > 0.9; see "Methods"), we identified 61 potential functional variants for heat tolerance at these conserved sites in our study (Table S4). However, the candidate causal mutations for heat tolerance are not enriched ($p = 1.0$) in the conserved regions of the genome.

Table 3 provides a short list of putative causal variants (upstream and missense) for heat tolerance that overlap at genomic sites highly conserved across vertebrates. Some of the candidate genes flanking these variants have been reported to be involved with cell survival under stress in animals, e.g., *SCD*[35], *KIAA1324*[36], and *TONSL*[14]. The *SCD* (stearoyl-CoA desaturase) gene encode fatty acid metabolic enzyme and perhaps is required for metabolic homeostasis during heat stress in mammals. Other putative candidate genes for heat tolerance include *KIFC2*, *VPS13B*, and *USP3*. For example, Fang et al.[37] demonstrated that the *USP3* gene, a member of the ubiquitin-specific proteases (USPs) family, is required for eliminating misfolded proteins under heat stress conditions in Yeast.

Pathway enrichment analysis. We generated a list of candidate genes mapping within or near lead SNPs detected at FDR < 0.10 for each trait for the pathway enrichment analyses. We found that the candidate gene-list for slope traits were highly enriched for the KEGG pathways related to the neuronal system (neuroactive ligand–receptor interaction and glutamatergic synapse) and metabolism system (citrate cycle) (Fig. 5). Interestingly, the heat tolerance candidate gene-list (N = ~400 genes) identified from various analyses (single-trait GWAS, meta-analysis, and conditional analysis) were consistently significantly enriched for a neuroactive ligand–receptor interaction pathway comprising of 15 genes (*CALCR, PTGER2, THRB, GRIK2, NPY2R, F2RL1, GRIN2A, NR3C1, CHRM3, GRM8, GRM7, GRID2, NPFFR2, MC4R, GHR*). A total of 8 genes were enriched ($p = 4.0E{-}03$) in the glutamatergic synapse pathway (*GRIN2A, GRM7, GRM8, ITPR1, ITPR2, SLC17A6, GRIK2, GRIA4*). The citrate cycle pathway was also enriched ($p = 1.87E{-}03$), comprising of 5 candidate genes for heat tolerance (*ACLY, PDHA2, MDH1, SUCLG2, PCK1*).

We also analysed a smaller set of genes (N = ~230) with the strongest ($p < 1E{-}05$) evidence of association for heat tolerance, separately (that is, the gene-list underlying the candidate causal variants defined as the lead SNP (most significant) within an independent QTL plus other significant SNPs in strong LD ($r^2 > 0.80$) with the lead SNP, 500 kb up or downstream), to see enriched biological pathways. Interestingly, we observed enrichment ($p = 0.02$) of the genes in the neuroactive ligand–receptor interaction pathway, which provides strong support that this neuronal pathway is relevant for heat tolerance comprising of 8 genes (*GHR, NPFFR2, P2RY8, GRIN2A, CHRM1, THRB, CALCR, F2RL1*).

When examining the candidate gene-list from single-trait GWAS analyses for slope traits separately, the neuroactive ligand–receptor interaction pathway was overrepresented for candidate gene-list for HTMYslope ($p = 3.19E{-}04$) and HTPYslope ($p = 7.79E{-}03$) traits (Fig. 6). On the other hand, gene-list for HTFYslope were enriched ($p = 1.55E{-}02$) for the axon guidance pathway comprising four genes (*ABLIM2, ABLIM3, NTN1, ROBO1*) and metabolic ($p = 0.06$) pathways.

To further test whether the neuronal pathways are real and not an artifact of our analyses for heat tolerance traits (slopes), we performed enrichment analyses for the significant candidate gene-list for intercepts traits (level of milk production traits). In the candidate gene-list for intercept traits, we found no evidence for enrichment ($p < 0.05$) in any neuronal pathways; thus, providing further favourable support that neuronal pathways are relevant for heat tolerance in mammals.

SNP	Chr	BP	Annotation	VEP_impact	Gene	HTMYslope	HTFYslope	HTPYslope	Meta-analysis[a]	Conditional analysis[b]	p value[c]
rs209684414	3	34,215,670	Upstream	MODIFIER	KIAA1324				✓		5.98E−07
rs207668220	3	34,273,899	Upstream	MODIFIER	C3H1orf194				✓		6.5E−07
rs210324395	10	46,503,113	Upstream	MODIFIER	USP3				✓		8.66E−09
rs210468775	10	46,505,212	Upstream	MODIFIER	USP3				✓		1.89E−09
rs207681599	14	432,274	Upstream	MODIFIER	LRRC14	✓	✓	✓	✓	✓	1.46E−52|1.13E−63|1.13E−28|8.8E−243|7.92E−08
rs136474298	14	471,951	Missense	MODERATE	KIFC2	✓	✓	✓	✓	✓	6.11E−53|4.83E−63|5.59E−29|2.5E−242|7.77E−08
rs207886320	14	479,761	Upstream	MODIFIER	KIFC2	✓	✓	✓	✓	✓	6.18E−52|4.17E−64|4.55E−28|3.9E−242|7.85E−08
rs137472016	14	494,621	Upstream	MODIFIER	TONSL	✓	✓	✓	✓	✓	6.28E−53|8.39E−63|5.94E−29|9.1E−242|7.87E−08
rs445616049	14	64,454,721	Missense	MODERATE	VPS13B	✓			✓		4.18E−08|1.75E−32
rs41946451	20	37,085,370	Missense	MODERATE	CPLANE1					✓	1.14E−07
rs41255693	26	21,272,422	Missense&splice	MODERATE	SCD		✓		✓		4.11E−07|1.85E−08

Table 3. Upstream and missense candidate causal variants for heat tolerance (slope) traits at genomic sites that are highly conserved (conservation score > 0.9) across 100 vertebrate species*. *Conservation scores (PhastCon score) of variants at conserved genomic sites were computed by[34] based on the LiftOver (https://genome.ucsc.edu/cgi-bin/hgLiftOver) human sites calculated across 100 vertebrate species; single-trait GWAS for heat tolerance milk (HTMYslope), fat (HTFYslope), and protein (HTPYslope) yield slope traits. [a]Meta-analysis combining single-trait GWAS results for slope traits. [b]Meta-analysis combining single-trait conditional GWAS results for slope traits. [c]For each analysis that identified this variant as significant.
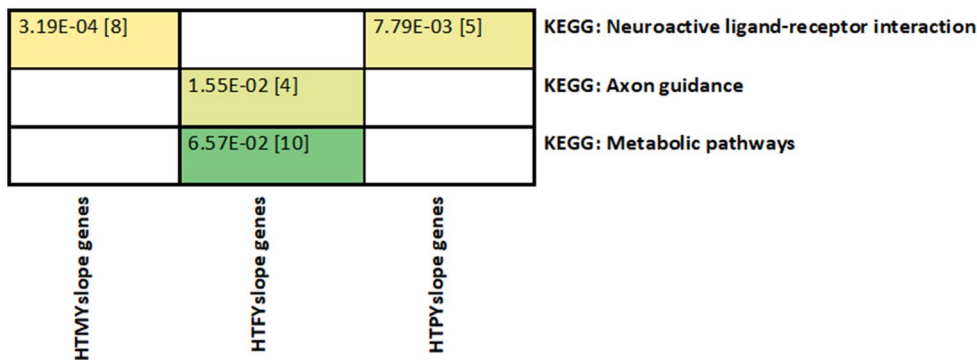

Figure 6. Enriched Kyoto Encyclopedia of Gene and Genomes (KEGG) pathways obtained from our gene-list for single-trait GWAS analysis of slope traits. HTMYslope (heat tolerance milk yield slope); HTFYslope (heat tolerance fat yield slope); and HTPYslope (heat tolerance protein yield slope). Cells are color-coded according to the strength of the significance for each pathway. Values in brackets are the number of genes within each pathway.

Discussion

In this study, we performed a GWAS using a large sample size of Australian dairy cows (N = 29,107) with milk production records and imputed sequence data (~ 15 million SNPs) to identify candidate causal variants and functional genes and pathways associated with heat tolerance. Australia's dairy cattle are uniquely placed for studying heat tolerance in mammals for two main reasons: (1) they are subjected to a wide range of seasonal climatic variations across diverse dairying regions spanning one of the geographically largest countries in the world, and (2) Australia's dairying is predominantly pasture-based with limited heat stress mitigation measures in contrast with those, for example, in North America, where extensive managerial strategies are used more to reduce thermal stress. Overall, we have identified novel candidate causal variants in the neuronal pathways that contribute significantly to heat tolerance in animals.

We leveraged two statistical approaches to identify genetic loci and pathways for heat tolerance: single-trait GWAS linear models and multi-trait meta-analysis. Single-trait GWAS is based on regressing phenotypes on each SNP one at a time. On the other hand, a meta-analysis that combines results of the single-trait GWAS allowed

90
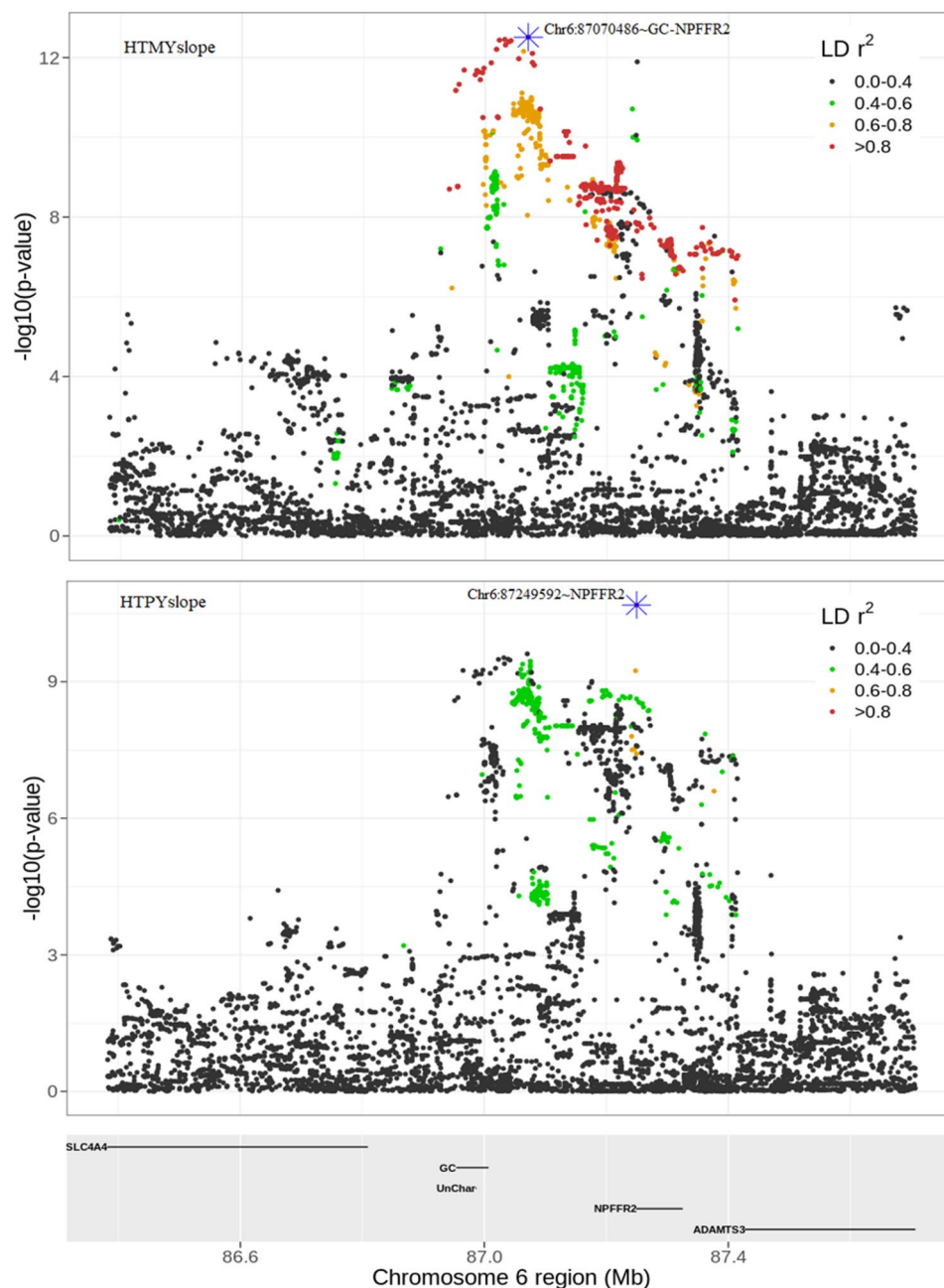
Figure 7. QTL discovery for heat tolerance milk (HTMYslope) and protein (HTPYslope) yield slope traits around the *NPFFR2* gene in bovine chromosome 6.

us to discern putative pleiotropic genetic variants for heat tolerance. Consequently, we identified multiple novel loci for heat tolerance, including 61 potential functional variants at genomic sites highly conserved across 100 vertebrates (Table 3 and Table S4), which could be valuable for fine-mapping and genomic prediction. Studies in humans[38] and cattle[34] have demonstrated that the conserved genomic sites have strong enrichment of trait heritability. Moreover, the results revealed specific candidate causal variants and genes related to neuronal functions for heat tolerance in animals, which we now discuss in more detail.

Heat stress responses are complex adaptations in animals involving many biological pathways, including the nervous system, which connects the internal and external environment to maintain stable core body temperature[39]. Among the candidate gene-list that contribute significantly to heat tolerance in the study animals (Holstein cows), the neuroactive ligand–receptor interaction and glutamatergic synapse pathways (Fig. 5), as components of the nervous system, were highly enriched ($p < 1E{-}03$) biological features.

At least two candidate variants in the intronic region of *ITPR2* (Chr5:83,330,185; $p = 1.3E{-}05$) and *GRIA4* (Chr15:2,461,074; $p = 5.8E{-}05$) genes in the glutamatergic synapse pathway could be potential targets for resilience to environmental stress in animals. *ITPR2* gene was associated with heat stress in the US Holsteins[14] or

sweating rate in humans and mice[40], while the *GRIA4* gene has been linked to thermoregulation in the Siberian cattle[41]. Another candidate variant (Chr22:21,783,956) detected for heat tolerance milk ($p = 3.87E–05$) and protein (7.15E–05) yield slope traits mapped to the intronic region of *ITPR1*—a gene associated with environmental adaptation in the domestic yak[42]. These three lead SNPs for slope traits overlapped with those for intercept traits, with opposing effect direction, suggesting that selecting for these variants may negatively impact milk production.

Previous studies show that the neuroactive ligand–receptor interaction is involved in maintaining energy homeostasis during heat stress in ducks[43]. As protein production is the most valuable output from dairy farms, the focus of breeding programs has been traits associated with yield, with the average milk volume per cow/year almost doubling within the past three decades in Australia[44]. The environmental heat stress, coupled with the elevated metabolic-induced thermogenesis, means that the genetic and cellular reprogramming of pathways such as the nervous system may be necessary to regulate a cascade of hormonal processes such as growth factors, insulin, serotonin, thyroid, prolactin, and mineralocorticoids associated with milk synthesis[45]. We identified 15 genes (FDR < 0.10) associated with the neuroactive ligand–receptor interaction, which could be relevant for metabolic homeostasis in cattle during thermal stress, of which three candidate genes (*GHR, NPFFR2,* and *CALCR*) showed the strongest evidence ($p < 1E–05$).

Here we discuss the evidence for each of these three candidate genes:

(1) Zhang et al.[24] demonstrated that the *NPFFR2* (neuropeptide FF receptor-2) gene, which is mainly expressed by neurons in the brain, plays a crucial role in regulating diet-induced thermogenesis and bone homeostasis in mice. In this study, two lead SNPs (Chr6:87,070,486 and Chr6:87,249,592), detected from single-trait GWAS for HTMYslope and HTPYslope ($p < 1E–05$) mapped to the intergenic and intronic regions of *NPFFR2* gene in BTA 6, respectively. Physiological studies suggest that NPFF family genes regulate feeding behaviour and energy expenditure in mammals reviewed in[46]. During heat events, dairy cattle typically reduce their dry matter intake by up to 30%, perhaps as part of an adaptive mechanism to depress metabolic heat production[47]. Other studies, e.g.,[48] show that inhibition of NPFF receptors induces hypothermia in mice. A recent review by Nguyen et al.[49] indicates that NPFF and its receptors have many promising therapeutic applications including pain, cardiovascular, and feeding regulations in mammals. By examining the genomic region around the *NPFFR2* gene (Fig. 7), it is more likely that the two lead SNPs within this QTL represent separate candidate causal mutations since they are not in strong LD. Interestingly, although the lead SNP (Chr6:87,070,486) for slope trait overlapped with the lead SNP detected for the milk yield intercept (MYint), we observed stronger evidence for the slope (HTMYslope; $p = 3.05E–13$) than the intercept (MYint; $p = 4.19E–10$), suggesting that this SNP is a good candidate for heat tolerance. Besides, this lead SNP (Chr6:87,070,486) remained significant ($p = 6.36E–06$) following single-trait conditional GWAS analysis for HTMYslope trait (conditioning slopes on the intercept traits) as well as in the meta-analysis of single-trait conditional GWAS results for slope traits ($p = 3.74E–06$).

(2) Calcitonin receptors regulate daily body temperature rhythm in mammals and insects and are essential for maintaining homeostasis[50]. In this study, the lead SNP (Chr4:10,815,768) was intronic in the *CALCR* (calcitonin receptor) gene, perhaps indicating that it could be relevant for animals experiencing recurrent or chronic stress, such as in Australian seasonal summers. The strong GWAS signal around this QTL (Supplementary Fig. S8) suggests that the *CALCR* gene likely harbours causal mutations affecting heat tolerance. Dairy cattle employ various adaptive behavioural strategies during heat stress such as reduced feed intake, increased volume, and frequency of water intake, increased standing time, shade seeking, and grazing at cooler day time. We think that *CALCR* is likely involved with some of these heat-stress adaptive behaviours in dairy cattle. Future studies are needed to confirm this, particularly by combining production traits with other relevant behavioural phenotypes such as panting scores from high-throughput recording devices, e.g., activity-based collars.

(3) The expression of the *GHR* (growth hormone receptor) gene is down-regulated during heat stress in livestock, including dairy cows[51] and avian species[43]. The adaptive physiological significance of this down-regulation is not well understood, and it is partly independent of the nutritional level of the animal[51]. In this study, the lead SNP (Chr20:32,103,408; $p = 2.01E–08$) identified only in one slope trait (HTMYslope) based on significant cut-off of $p < 1E–05$ mapped to intronic region of *GHR* gene (Supplementary Fig. S9). However, we found a stronger signal after combining the GWAS results for all the slope traits in a meta-analysis with the lead SNP (Chr20:32,201,287; $p = 1.7E–47$) mapping to the intergenic (~ 22 kb) region of the *GHR* gene, which confirms the pleiotropic effect of this QTL[22]. Also, we observed no significant SNP ($p < 1E–05$) around this QTL following single-trait conditional analyses, but a somewhat strong signal emerged when we combined single-trait conditional GWAS results in the meta-analysis, for which the lead SNP (Chr20:32,226,298; $p = 5.35E–07$) mapped to the intergenic region (~ 47 kb) of GHR. This further supports a possible second QTL that is independent of the level of milk production and shows pleiotropy for the heat tolerance traits. Other published GWAS have also reported an association of the GHR gene with milk production in heat-stressed cows[14] and respiratory rates in pigs during heat stress[19]. Several studies have also implicated the *GHR* polymorphisms to milk production in cattle, e.g., Chr20:31,888,449 phenylalanine-to-tyrosine missense mutation[29]. This mutation was not in strong LD ($r^2 > 0.8$) with the lead SNP detected for slope traits in our study. Taken together, polymorphisms around the *GHR* gene could be candidate targets for improving thermotolerance in livestock, although with possible antagonistic effect on milk production considering, for example, the opposing effect direction observed for the lead SNP (Chr20:32,103,408) within this QTL on the slope (HTMYslope) and intercept (MYint) traits.

There is general agreement that heat stress decreases milk yield (milk, proteins, fat, etc.) in dairy cattle. However, the genetic and biological basis for this reduction is still unclear. Evidence suggests that the reduced feed intake in heat-stressed dairy cows partially explains (35–50%) reduced milk yield and composition[47]. The molecular control and pathways for individual milk traits during heat stress are scarce and inconclusive. In this study, the QTLs detected for the heat tolerance traits varied across the three milk traits (HTMYslope, HTFYslope, HTPYslope), suggesting that they are, in part, regulated by different genes in heat-stressed cows. The greater overlap of candidate genes observed for HTMYslope and HTPYslope traits was expected due to their relatively high correlation (0.90) compared to HTMYslope and HTFYslope (0.56) or HTPYslope and HTFYslope (0.62). These correlations appear to mirror the proportions of SNPs with the same or inconsistent effect direction observed for significant SNPs between slope traits. Considering that heat stress alters carbohydrate, lipid, and amino acid metabolism[52], the large proportion of SNPs with inconsistent effect direction, particularly between HTPYslope and HTFYslope, suggest that these traits are somewhat differently regulated in heat-stressed dairy cows.

Several pair-fed studies suggest that pathways related to the mammary gland protein synthesis govern protein production under heat stress in dairy cows, in part, via reduced amino acid supply to the mammary gland, e.g.,[53,54]. We found that the candidate genes for HTMYslope and HTPYslope traits were overrepresented ($p < 0.005$) in the neuroactive ligand–receptor interaction pathway. This agrees with Pegolo et al.[55] that genes associated with milk proteins are involved in neuronal signaling pathways in dairy cattle. However, it remains unclear how this pathway is regulated during heat stress conditions in dairy cows to impact protein production.

On the other hand, the molecular pathways for fat production under heat stress conditions have not been widely studied. Some studies e.g.,[56] suggest that the reduced activation of PPAR (peroxisome proliferator-activated receptor) signaling pathways leads to decreased expression of genes associated with fat metabolism. Candidate genes for HTFYslope identified in this study are associated with the KEGG term "metabolic pathways" (Fig. 6). Five candidate genes (*DMGDH, PDHA2, UGP2, MDH1, PRDX6, NDUFA13*) within this pathway may be involved with alleviating oxidative stress in heat-stressed cows. In line with these findings, we found that the candidate genes for heat tolerance (Fig. 5) are overrepresented in the citrate cycle/TCA pathway, which is central to mitochondrion energetics, and might serve to reduce substrate oxidation and reactive oxygen species (ROS) production, thus preventing cellular damage during heat stress.

Notably, our pathway results are perhaps not directly comparable to most previous work in which the study cows were subjected to short-term acute heat stress under experimental conditions e.g.,[56] whereas the current work mimics recurrent or chronic stress that dairy cows experience during summer seasons in Australia. The effects of heat stress in livestock depend on its duration and severity, with the most recent work in Arabian camels somatic cells showing that acute heat stress elevates the expression of heat shock proteins and DNA repair enzymes while chronic heat leads to changes in cell integrity and reduction of total protein levels, metabolic enzymes, and cytoskeletal proteins[57]. Our candidate QTLs are particularly important since it provides novel insights into the molecular aspect of chronic stress considering that the study animals are predominantly reared under outdoor conditions with limited heat stress mitigations. Future studies are required to confirm if these QTLs are involved with recurrent chronic stress in other animal species.

We could not replicate most of the candidate genes with published GWAS results for heat tolerance in cattle, likely for several reasons. First, all comparable earlier studies were much smaller (< 5,000 animals) and therefore were under-powered, and the marker density used was typically 50 k or 600 k SNP array e.g.,[13,14,58]. As expected, we observed that our sequence variants showed markedly higher significance levels than the 50 k SNP array and increased the number of significant peaks across the genome (Supplementary Fig. S1). Second, the trait used to define heat tolerance in this study (i.e., the rate of milk yield decline under heat stress) differs from many other studies e.g.,[12] which used measures of core body temperature in their GWAS. Given that heat tolerance is a complex trait involving a wide array of adaptative responses (behavioural, physiological, cellular, etc.), different QTLs may be captured by different traits used in GWAS. Notably, although heat tolerance traits (slopes) used in earlier studies in Australia[13,58] were comparable to those used in our study, we could not confirm most candidate genes (except HSF1 gene). This is likely due to the reduced power in earlier GWAS studies (they used smaller sample size and 50 k or 600 k SNP data). Third, differences in the patterns of LD among study populations used and imputation quality may have implications on GWAS, particularly in the detection of putative causal mutations[59]. Here we explored QTLs for heat tolerance in purebred Holstein cows, while some other studies, e.g.,[15] have used crossbred cattle. Collectively, these factors likely impacted the replication of previous GWAS candidate genes for heat tolerance.

Although we detected multiple candidate causal variants for heat tolerance in this study, it appears that larger sample size (we used N = 29,107) would be beneficial considering the polygenic architecture of this trait. Larger sample size is required to detect causal variants with very small effects and the effects of rare causal variants[16]. For example, many of the lead SNPs (most significant) for heat tolerance were tagged by none or very few significant SNPs (Table S1), which may be false-positive variants passing the GWAS cut-off ($p < 1E-05$).

With the increasing availability of high-throughput data from automatic sensor devices such as activity-based collars or tags, it is now feasible to obtain large-scale data for thousands of animals; if genotyped, it would allow a comprehensive genetic evaluation of heat tolerance from a wide array of phenotypes e.g., mid-infrared (MIR) predicted traits from milk recording data[60]. Furthermore, we used conditional analyses of slope traits in a bid to separate production and heat tolerance genes. It may be useful to consider alternative heat tolerance traits in the future GWAS (besides milk decays) that are independent of production, such as those derived from milk yield based on the principal components (PC)[61] or eigen-functions[62]. Overall our results support the highly polygenic nature of heat tolerance characterised by multiple small-effect variants, suggesting that this trait is more amenable to genomic selection tools such as those currently implemented in the Australian dairy industry[3, 63] rather than approaches that exploit few QTLs with large effects. The significant variants detected in this study will be tested in a follow-up study to assess their benefits in the genomic prediction of heat tolerance in dairy cattle.

93

In conclusion, we performed GWAS for heat tolerance using large sample size and genotype dataset for dairy cattle. The increased sample size and high-resolution SNP data in our study compared to previous reports allowed us unprecedented power and precision of the GWAS to pinpoint multiple putative causal mutations, including 61 potential functional variants at genomic sites highly conserved across 100 vertebrate species. Also, results indicate that different genes and pathways, in part, regulate different milk production traits (milk, fat, and proteins) in heat-stressed dairy cows with a substantial overlap of genes for heat tolerance milk and protein yield. Overall, the results revealed the importance of variation in genes related to the neuronal functions for heat tolerance in mammals, which is of interest for future research towards understanding and managing heat stress for warm climates and particularly in view of the anticipated rising global temperatures.

Materials and methods

Animals and phenotypes.   No live animals were used in this study. Phenotypes used for GWAS were part of our previous study[64] obtained from DataGene (DataGene Ltd., Melbourne, Australia; https://datagene.com.au/)—the organisation responsible for genetic evaluation of dairy animals in Australia. The phenotypes were test-day milk, fat, and protein yield for Holstein dairy cows collected from dairy herds that were matched with climate data (daily temperature and humidity) obtained from weather stations across Australia's dairying regions. The distribution of dairy herds and weather stations; and the calculation of environmental covariate (i.e., temperature–humidity index (THI)) used here were described in our earlier studies[3,64].

Calculation of heat tolerance phenotypes for cows.   The dataset used to calculate heat tolerance phenotypes for cows was similar to that used by[64], comprising a total of 424,846 test-day milk records for first, second and third lactations from 312 herds and 15,906 herd-test days (HTD) collected over 15 years (2003–2017). A summary of the final dataset is given in Table 1. The rate of decline (slope) in milk, fat, and protein yield due to heat stress events was estimated using a reaction norm models[64]. In these models, data on milk, fat, or protein yield were adjusted for the fixed effects, including herd test day, year season of calving, parity, Legendre polynomials (order 3) on the cow age on the day of test, and the Legendre polynomials (order 8) on the interaction between parity and DIM. Random effects fitted in the model included a random regression on a linear orthogonal polynomial of THI, where the intercept represents the level of mean milk yield and the linear component represents the change in milk yield (slope) due to heat stress for each cow and a residual term. In the model, the threshold of THI was set to 60 following[65]. The analyses to derive trait deviation (TD) which represents a phenotype adjusted for all fixed effects (i.e., the mean/intercept and slope for each cow) were conducted using ASReml v4.2[66]. To facilitate convergence, milk, fat and protein yield traits were scaled by a factor of 10, 100, and 100, respectively. The description of heat tolerance traits (i.e., slopes) used in this study are comparable to those used in previous GWAS in Australia[13,58].

We refer to milk intercept traits as [MYint (i.e., milk yield intercept), FYint (i.e., fat yield intercept), and PYint (i.e., protein yield intercept)] and the slopes traits as [HTMYslope (i.e., heat tolerance milk yield slope), HTFYslope (i.e., heat tolerance fat yield slope), and HTPYslope (i.e., heat tolerance protein yield slope)], respectively.

Genotypes.   Two genotype datasets were analysed for 29,107 Holstein cows with the above phenotypes: 50 k SNP chip and 15,098,486 imputed whole-genome sequence variants (WGS). Most of the cows were originally genotyped with a custom low-density 10 k SNP panel or a standard medium density 50 k SNP array (BovineSNP50k BeadChip: Illumina Inc). The low-density genotypes were imputed to the 50 k array using a reference set of approximately 14,000 animals with real 50 k genotypes, with approximately 7,000 SNPs of the low-density SNP panel overlapping the 50 k SNP array. The 50 k genotypes were then imputed to the high-density Bovine SNP array (HD: BovineHD BeadChip, Illumina Inc) using a reference set of 2,700 animals with real HD genotypes. All SNP BeadChip genotypes were first converted to the ARS-UDC1.2 reference genome (https://www.ncbi.nlm.nih.gov/assembly/GCF_002263795.1/)[67] positions from reference genome UMD3.1 and imputed using Fimpute3[68]. The WGS was imputed from the HD genotypes using a reference set of 3,090 *Bos taurus* sequences in the Run7 of the 1000 Bull Genome Project (http://1000bullgenomes.com/)[20] aligned to the ARS-UCD1.2 reference genome. Only bi-allelic sequence variants with a minor allele count (≥ 4) and GATK[69] quality tranche 99.0 or better were retained for imputation. Pre-imputation, we also removed sequence variants from the imputation reference that had a higher than expected proportion of heterozygous calls (> 0.5) if these variants fell in a 500 kb window enriched for variants showing excessive heterozygosity (as a proxy to indicate regions where WGS mapping/alignment may be poor). A total of 31,994,954 sequence variants remained for imputation. Minimac3[70] was used for WGS imputation, having first pre-phased both the HD genotypes and the WGS reference using Eagle v2[71]. For the analysis, we retained only the variants with Minimac3 imputation accuracy, $R^2 > 0.4$ and MAF > 0.005 (N = 15,098,486 sequence variants).

Single-trait GWAS and multi-trait meta-analysis.   A genome-wide association analysis (GWAS) using a mixed linear model was used to test associations between individual SNP and cows' slope [HTMYslope, HTPYslope and HTFYslope] and intercept [MYint, FYint, PYint] traits using GCTA software[72]. Because phenotypes were TD already adjusted for nongenetic effects, for each autosomal SNP i with minor allele frequency (MAF) > 0.005, the fitted model per trait was,

$$y = mean + x\beta + g + \varepsilon,$$

where y was the vector of TD (intercept or slope traits) for cows (n = 29,107), β was the allele substitution effect of SNP i, x was the vector of genotype dosages (0, 1, or 2) for SNP i, g was the vector of polygenic effect

94

with $g \sim N(0, \text{GRM}\sigma_g^2)$ and ε was a vector of the residual effect with $e \sim N(0, I\sigma_e^2)$, where I was an $n \times n$ identity matrix. The variance of y was $\text{var}(y) = \text{GRM}\sigma_g^2 + I\sigma_e^2$ where GRM is the genomic relationship matrix between cows, and σ_g^2 and σ_e^2 were the genetic and residual variances. For animal j and k relationship was calculated using GCTA[72] as follows:

$$A_{jk} = \frac{1}{N} \sum_{i=1}^{N} \frac{(x_{ij} - 2p_i)(x_{ik} - 2p_i)}{2p_i(1 - p_i)}$$

where A_{jk} are the off-diagonal elements of GRM for animal j and k; N = total number of SNPs from 50 k SNP array data (MAF > 0.005; 45,504 SNPs); x_{ij} and x_{ik} are the number of copies for reference allele for the ith SNP; and p_i is the allele frequency for ith SNP.

Genomic heritability was calculated for each trait using variance component estimates from –reml option of GCTA for 50 k SNP array (45,504 SNPs) data of cows (N = 29,107): $h^2 = \sigma_g^2 / (\sigma_g^2 + \sigma_e^2)$.

To increase the power of GWAS and identify pleiotropic variants, we next combined single-trait GWAS results obtained above in a multi-trait meta-analysis following[21]. The multi-trait chi-squared (χ^2) statistics for ith SNP was calculated separately for intercept [MYint, FYint, PYint] and slope [HTMYslope, HTFYslope, and HTPYslope] traits as follows:

$$\chi^2 = t_i' V^{-1} t_i$$

where t_i is the vector of 3×1 vector of signed t-values (i.e., b/se) of ith SNP for either intercept or slope traits; and V^{-1} is the inverse of 3×3 correlation matrix of the signed t-values calculated based on all pairs for the intercept or slope traits. The significance of χ^2 value for ith SNP was calculated based on chi-squared distribution with 3 degrees of freedom—that is number of traits for either intercept or slope traits.

Conditional GWAS analysis.   Next, we performed two conditional GWAS strategies of slope traits using GCTA software[72] to test somewhat different hypotheses:

Conditional analysis of slope traits on lead SNP (i.e., most significant SNP within a chromosome from first-round GWAS)—aimed at identifying additional or secondary putative causal variants beside those detected from first-round GWAS. We performed a conditional analysis strategy on two chromosomes (BTA 14 and BTA 20), which showed the strongest GWAS signal for slope traits in the first-round GWAS (Supplementary Figs. S1 and S2) and are known to harbour QTLs with major effects on milk production (i.e., BTA14 ~ *DGAT1* and BTA20 ~ *GHR* gene).

Conditional analysis of slope traits on intercept traits—aimed at identifying QTLs for heat tolerance that are independent, or not also strongly associated with the level of milk production. We fitted the intercept traits of MYint, FYint, and PYint, as a covariate in the linear model when analysing the HTMYslope, HTFYslope, or HTPYslope, respectively. To increase the power of GWAS, we then combined conditional GWAS results for the three slope traits [HTMYslope, HTFYslope, and HTPYslope] in a multi-trait meta-analysis following[21] as described earlier.

Identifying candidate causal variants.   We used the following criteria to select candidate variants ($p < 1.0\text{E}{-}05$) from the three analytical approaches (single-trait GWAS, meta-analysis, conditional analysis).

1. For each trait, select all SNPs with $p < 1\text{E}{-}05$ (FDR < 0.10).
2. Split each chromosome (N = 1…29) into 5 Mb non-overlapping windows from the start to the distal end of the chromosome.
3. Within the ith 5 Mb window, select the most significant SNP (i.e., the SNP with the smallest p value below the threshold of $p < 1\text{E}{-}05$) defined as the 'lead SNP'. We chose this arbitrary 5 Mb window size to obtain a small set of significant lead SNPs representing independent QTL (that is, not in linkage disequilibrium) for further detailed examination.
4. Calculate the LD between each lead SNP and all the other SNPs within 500 kb up and downstream of the lead SNP using Plink v1.9[73].
5. For each lead SNP, extract all the significant SNPs ($p < 1\text{E}{-}05$) in strong LD ($r^2 > 0.80$) with the lead SNP within 500 kb up or down downstream – to account for the fact that the lead SNP (most significant) is not necessarily the causal variant.

Annotation of sequence variants and enrichment analysis.   Annotation of all variants (~ 15 million SNPs) was performed using SnpEff[33] tool. Using the annotation, we grouped the candidate causal variants for heat tolerance (slopes) into 9 classes (intergenic, intronic, missense, upstream, downstream, 3_prime_UTR, synonymous, 5_prime_UTR, and Other) and performed enrichment analysis using phyper in R v3.61[74]. The class "Other" comprised variants including 5_prime_UTR_premature/_start_codon_gain, frameshift, missense&splice, splice&intron, stop_gained, etc. Supplementary Table S1 provides the number of candidate causal variants for heat tolerance within the 9 classes.

Candidate variants at conserved genomic sites.   We identified candidate causal variants for heat tolerance at highly conserved genomic sites using data from[34]. Briefly, these authors documented over 300 k sequence

95

variants at conserved sites in cattle based on the LiftOver (https://genome.ucsc.edu/cgi-bin/hgLiftOver) human sites with conservation scores (PhastCon score) > 0.9 calculated across 100 vertebrate species (see https://www.pnas.org/content/pnas/suppl/2019/09/07/1904159116.DCSupplemental/pnas.1904159116.sapp.pdf for more details).

Pathway enrichment analysis. We generated candidate gene-list mapping near or underlying lead SNPs (most significant SNPs within 5 Mb QTL windows) identified at FDR < 0.10 cut-off threshold from both single-trait and multi-trait analyses of intercept or slope traits. For intergenic lead SNPs, we selected the closest gene on either side of the SNP. We chose this cut-off (FDR < 0.10) instead of a more stringent $p < 1E–05$ to include genes associated with smaller effects while guarding against false positives. We then performed the Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis using DAVID[75].

We also performed enrichment test separately for the gene-list associated with potential major effects on heat tolerance identified across all analyses (i.e., gene-list with the strongest ($p < 1E–05$) evidence of association defined as the candidate causal variants (i.e., lead SNP + other significant SNPs in strong LD ($r^2 > 0.80$) with the lead SNP within 500 kb up or downstream passing the cut-off p value of $1 < 1E–05$). For all the analyses, we considered functional pathways with Fisher's $p < 0.05$ as significantly enriched.

Data availability

Positions and annotations for all the lead SNPs (most significant SNPs) with $p < 1E–5$ are in Tables S1–S3. Data-Gene (DataGene Ltd., Melbourne, Australia; https://datagene.com.au/) are the custodians of the raw phenotype and genotype data of Australian farm animals. Research related requests for access to the data may be accommodated on a case-by-case basis.

References

1. Li, D., Yuan, J. & Kopp, R. B. Escalating global exposure to compound heat-humidity extremes with warming. *Environ. Res. Lett.* **15**, 064003 (2020).
2. Xu, C., Kohler, T. A., Lenton, T. M., Svenning, J.-C. & Scheffer, M. Future of the human climate niche. *Proc. Natl. Acad. Sci.* **117**, 11350–11355 (2020).
3. Nguyen, T. T., Bowman, P. J., Haile-Mariam, M., Pryce, J. E. & Hayes, B. J. Genomic selection for tolerance to heat stress in Australian dairy cattle. *J. Dairy Sci.* **99**, 2849–2862 (2016).
4. Polsky, L. & von Keyserlingk, M. A. Invited review: Effects of heat stress on dairy cattle welfare. *J. Dairy Sci.* **100**, 8645–8657 (2017).
5. St-Pierre, N., Cobanov, B. & Schnitkey, G. Economic losses from heat stress by US livestock industries. *J. Dairy Sci.* **86**, E52–E77 (2003).
6. Hansen, P. Physiological and cellular adaptations of zebu cattle to thermal stress. *Anim. Reprod. Sci.* **82**, 349–360 (2004).
7. Bryant, J., López-Villalobos, N., Pryce, J., Holmes, C. & Johnson, D. Quantifying the effect of thermal environment on production traits in three breeds of dairy cattle in New Zealand. *N. Z. J. Agric. Res.* **50**, 327–338 (2007).
8. Sejian, V., Bhatta, R., Gaughan, J., Dunshea, F. & Lacetera, N. Adaptation of animals to heat stress. *Animal* **12**, s431–s444 (2018).
9. Ravagnolo, O., Misztal, I. & Hoogenboom, G. Genetic component of heat stress in dairy cattle, development of heat index function. *J. Dairy Sci.* **83**, 2120–2125 (2000).
10. Dikmen, S., Cole, J., Null, D. & Hansen, P. Heritability of rectal temperature and genetic correlations with production and reproduction traits in dairy cattle. *J. Dairy Sci.* **95**, 3401–3405 (2012).
11. Littlejohn, M. D. *et al.* Functionally reciprocal mutations of the prolactin signalling pathway define hairy and slick cattle. *Nat. Commun.* **5**, 1–8 (2014).
12. Dikmen, S., Cole, J. B., Null, D. J. & Hansen, P. J. Genome-wide association mapping for identification of quantitative trait loci for rectal temperature during heat stress in Holstein cattle. *PLoS One* **8**, e69202 (2013).
13. Wang, T. *et al.* Application of a Bayesian non-linear model hybrid scheme to sequence data for genomic prediction and QTL mapping. *BMC Genomics* **18**, 618 (2017).
14. Sigdel, A., Abdollahi-Arpanahi, R., Aguilar, I. & Peñagaricano, F. Whole genome mapping reveals novel genes and pathways involved in milk production under heat stress in US Holstein cows. *Front. Genet.* **10**, 928 (2019).
15. Otto, P. I. *et al.* Genome-wide association studies for heat stress response in Bos taurus× Bos indicus crossbred cattle. *J. Dairy Sci.* **102**, 8148–8158 (2019).
16. Wood, A. R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46**, 1173–1186 (2014).
17. Taye, M. *et al.* Whole genome detection of signature of positive selection in African cattle reveals selection for thermotolerance. *Anim. Sci. J.* **88**, 1889–1901 (2017).
18. Freitas, P. H. *et al.* Genetic diversity and signatures of selection for thermal stress in cattle and other two bos species adapted to divergent climatic conditions. *Front. Genet.* **12**, 102 (2021).
19. Kim, K.-S. *et al.* Characterization of the acute heat stress response in gilts: III. Genome-wide association studies of thermotolerance traits in pigs. *J. Anim. Sci.* **96**, 2074–2085 (2018).
20. Hayes, B. J. & Daetwyler, H. D. 1000 bull genomes project to map simple and complex genetic traits in cattle: Applications and outcomes. *Annu. Rev. Anim. Biosci.* **7**, 89–102 (2019).
21. Bolormaa, S. *et al.* A multi-trait, meta-analysis for detecting pleiotropic polymorphisms for stature, fatness and reproduction in beef cattle. *PLoS Genet* **10**, e1004198 (2014).
22. Xiang, R., van den Berg, I., MacLeod, I. M., Daetwyler, H. D. & Goddard, M. E. Effect direction meta-analysis of GWAS identifies extreme, prevalent and shared pleiotropy in a large mammal. *Commun. Biol.* **3**, 88. https://doi.org/10.1038/s42003-020-0823-6 (2020).
23. Carvalheiro, R. *et al.* Unraveling genetic sensitivity of beef cattle to environmental variation under tropical conditions. *Genet. Sel. Evol.* **51**, 29 (2019).
24. Zhang, L. *et al.* Diet-induced adaptive thermogenesis requires neuropeptide FF receptor-2 signalling. *Nat. Commun.* **9**, 1–19 (2018).
25. Joutsen, J. *et al.* Heat shock factor 2 protects against proteotoxicity by maintaining cell-cell adhesion. *Cell Rep.* **30**, 583–597. e586 (2020).

26. Grisart, B. *et al.* Genetic and functional confirmation of the causality of the DGAT1 K232A quantitative trait nucleotide in affecting milk yield and composition. *Proc. Natl. Acad. Sci.* **101**, 2398–2403 (2004).
27. Winter, A. *et al.* Association of a lysine-232/alanine polymorphism in a bovine gene encoding acyl-CoA: Diacylglycerol acyltransferase (DGAT1) with variation at a quantitative trait locus for milk fat content. *Proc. Natl. Acad. Sci.* **99**, 9300–9305 (2002).
28. Littlejohn, M. D. *et al.* Sequence-based association analysis reveals an MGST1 eQTL with pleiotropic effects on bovine milk composition. *Sci. Rep.* **6**, 25376 (2016).
29. Blott, S. *et al.* Molecular dissection of a quantitative trait locus: A phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition. *Genetics* **163**, 253–266 (2003).
30. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci.* **100**, 9440–9445 (2003).
31. Menegazzi, P., Yoshii, T. & Helfrich-Förster, C. Laboratory versus nature: The two sides of the Drosophila circadian clock. *J. Biol. Rhythms* **27**, 433–442 (2012).
32. Sveinbjornsson, G. *et al.* Weighting sequence variants based on their annotation increases power of whole-genome association studies. *Nat. Genet.* **48**, 314 (2016).
33. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* **6**, 80–92 (2012).
34. Xiang, R. *et al.* Quantifying the contribution of sequence variants with regulatory and evolutionary significance to 34 bovine complex traits. *Proc. Natl. Acad. Sci.* **116**, 19398–19408 (2019).
35. Dado-Senn, B. *et al.* RNA-Seq reveals novel genes and pathways involved in bovine mammary involution during the dry period and under environmental heat stress. *Sci. Rep.* **8**, 1–11 (2018).
36. Deng, L., Feng, J. & Broaddus, R. The novel estrogen-induced gene EIG121 regulates autophagy and promotes cell survival under stress. *Cell Death Dis.* **1**, e32–e32 (2010).
37. Fang, N. N., Zhu, M., Rose, A., Wu, K.-P. & Mayor, T. Deubiquitinase activity is required for the proteasomal degradation of misfolded cytosolic proteins upon heat-stress. *Nat. Commun.* **7**, 1–16 (2016).
38. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228 (2015).
39. Nakamura, K. & Morrison, S. F. A thermosensory pathway that controls body temperature. *Nat. Neurosci.* **11**, 62–71 (2008).
40. Klar, J. *et al.* Abolished InsP 3 R2 function inhibits sweat secretion in both humans and mice. *J. Clin. Investig.* **124**, 4773–4780 (2014).
41. Igoshin, A. V. *et al.* Genome-wide association study and scan for signatures of selection point to candidate genes for body temperature maintenance under the cold stress in Siberian cattle populations. *BMC Genet.* **20**, 26 (2019).
42. Basang, W. & Zhu, Y. Whole-genome analysis identifying candidate genes of altitude adaptive ecological thresholds in yak populations. *J. Anim. breed. Genet.* **136**, 371–377 (2019).
43. Kim, J.-M. *et al.* Identification of the acclimation genes in transcriptomic responses to heat stress of White Pekin duck. *Cell Stress Chaperones* **22**, 787–797 (2017).
44. Pryce, J., Nguyen, T., Axford, M., Nieuwhof, G. & Shaffer, M. Symposium review: Building a better cow—The Australian experience and future perspectives. *J. Dairy Sci.* **101**, 3702–3713 (2018).
45. Bernabucci, U. *et al.* Metabolic and hormonal acclimation to heat stress in domesticated ruminants. *Animal Int. J. Anim. Biosci.* **4**, 1167 (2010).
46. Bechtold, D. A. & Luckman, S. M. The role of RFamide peptides in feeding. *J. Endocrinol.* **192**, 3–15 (2007).
47. Wheelock, J., Rhoads, R., VanBaale, M., Sanders, S. & Baumgard, L. Effects of heat stress on energetic metabolism in lactating Holstein cows. *J. Dairy Sci.* **93**, 644–655 (2010).
48. Wang, Y.-Q. *et al.* Neuropeptide FF receptors antagonist, RF9, attenuates opioid-evoked hypothermia in mice. *Peptides* **29**, 1183–1190 (2008).
49. Nguyen, T., Marusich, J., Li, J.-X. & Zhang, Y. Neuropeptide FF and its receptors: therapeutic applications and ligand development. *J. Med. Chem.* **63**, 12387 (2020).
50. Goda, T. *et al.* Calcitonin receptors are ancient modulators for rhythms of preferential temperature in insects and body temperature in mammals. *Genes Dev.* **32**, 140–155 (2018).
51. Rhoads, M. *et al.* Effects of heat stress and nutrition on lactating Holstein cows: II. Aspects of hepatic growth hormone responsiveness. *J. Dairy Sci.* **93**, 170–179 (2010).
52. Baumgard, L. H. & Rhoads, R. P. Jr. Effects of heat stress on postabsorptive metabolism and energetics. *Annu. Rev. Anim. Biosci.* **1**, 311–337 (2013).
53. Cowley, F., Barber, D., Houlihan, A. & Poppi, D. Immediate and residual effects of heat stress and restricted intake on milk protein and casein composition and energy metabolism. *J. Dairy Sci.* **98**, 2356–2368 (2015).
54. Gao, S. *et al.* The effects of heat stress on protein metabolism in lactating Holstein cows. *J. Dairy Sci.* **100**, 5040–5049 (2017).
55. Pegolo, S. *et al.* Integration of GWAS, pathway and network analyses reveals novel mechanistic insights into the synthesis of milk proteins in dairy cows. *Sci. Rep.* **8**, 1–15 (2018).
56. Gao, S. *et al.* Heat stress negatively affects the transcriptome related to overall metabolism and milk protein synthesis in mammary tissue of midlactating dairy cows. *Physiol. Genomics* **51**, 400–409 (2019).
57. Saadeldin, I. M. *et al.* Thermotolerance and plasticity of camel somatic cells exposed to acute and chronic heat stress. *J. Adv. Res.* **22**, 105–118 (2020).
58. Hayes, B. J. *et al.* A validated genome wide association study to breed cattle adapted to an environment altered by climate change. *PloS One* **4**, e6676 (2009).
59. Pausch, H. *et al.* Evaluation of the accuracy of imputed sequence variant genotypes and their utility for causal variant detection in cattle. *Genet. Sel. Evol.* **49**, 24 (2017).
60. van den Berg, I. *et al.* The use of milk mid-infrared spectroscopy to improve genomic prediction accuracy of serum biomarkers. *J. Dairy Sci.* **104**, 2008–2017 (2021).
61. Macciotta, N. *et al.* Derivation and genome-wide association study of a principal component-based measure of heat tolerance in dairy cattle. *J. Dairy Sci.* **100**, 4683–4697 (2017).
62. Carabaño, M., Bachagha, K., Ramón, M. & Díaz, C. Modeling heat stress effect on Holstein cows under hot and dry conditions: Selection tools. *J. Dairy Sci.* **97**, 7889–7904 (2014).
63. Nguyen, T. T. *et al.* Implementation of a breeding value for heat tolerance in Australian dairy cattle. *J. Dairy Sci.* **100**, 7362–7367 (2017).
64. Cheruiyot, E. K. *et al.* Genotype-by-environment (temperature–humidity) interaction of milk production traits in Australian Holstein cattle. *J. Dairy Sci.* **103**, 2460–2476. https://doi.org/10.3168/jds.2019-17609 (2020).
65. Hayes, B., Carrick, M., Bowman, P. & Goddard, M. Genotype × environment interaction for milk production of daughters of Australian dairy sires from test-day records. *J. Dairy Sci.* **86**, 3736–3744 (2003).
66. Gilmour, A., Gogel, B., Cullis, B., Welham, S. & Thompson, R. ASReml user guide release 4.2 functional specification. *VSN International Ltd, Hemel Hempstead, HP1 1ES, UK* www.vsni.co.uk *(2015)*.
67. Rosen, B. D. *et al.* De novo assembly of the cattle reference genome with single-molecule sequencing. *GigaScience* **9**, giaa021 (2020).

68. Sargolzaei, M., Chesnais, J. P. & Schenkel, F. S. A new approach for efficient genotype imputation using information from relatives. *BMC Genomics* **15**, 478 (2014).
69. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
70. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* **44**, 955–959 (2012).
71. Loh, P.-R. *et al.* Reference-based phasing using the Haplotype Reference Consortium panel. *Nat. Genet.* **48**, 1443 (2016).
72. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: A tool for genome-wide complex trait analysis. *Am. J. Human Genet.* **88**, 76–82 (2011).
73. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Human Genet.* **81**, 559–575 (2007).
74. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/. (2018).
75. Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44 (2009).

Acknowledgements

Author contributions

J.E.P., H.M., I.M.M., and E.K.C., conceived and designed the study; I.M.M. assisted with preparation and imputation of genotype data; E.K.C., I.M.M., M.H., J.E.P., B.G.C., R.X., contributed to the formal data analysis; E.K.C. wrote the first draft; all authors, reviewed and approved the final manuscript for publication.

Funding

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-95816-8.

Correspondence and requests for materials should be addressed to M.H.-M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Chapter 5: Functionally prioritised whole-genome sequence variants improve the accuracy of genomic prediction for heat tolerance

Evans K. Cheruiyot[1,2]*, Mekonnen Haile-Mariam[2]*, Benjamin G. Cocks[1,2], Iona M. MacLeod[2], Raphael Mrode[3,4], Jennie E. Pryce[1,2]

[1]School of Applied Systems Biology, La Trobe University, Bundoora, Victoria 3083, Australia

[2]Agriculture Victoria Research, AgriBio, Centre for AgriBiosciences, Bundoora, Victoria 3083, Australia

[3]International Livestock Research Institute, Nairobi, Kenya

[4]Scotland's Rural College, Edinburgh, United Kingdom

Please note that the formatting style of this chapter follows the journal specifications

GSE **G**enetics **S**election **E**volution

Check for updates

RESEARCH ARTICLE

Functionally prioritised whole-genome sequence variants improve the accuracy of genomic prediction for heat tolerance

Evans K. Cheruiyot[1,2], Mekonnen Haile-Mariam[2]* , Benjamin G. Cocks[1,2], Iona M. MacLeod[2], Raphael Mrode[3,4] and Jennie E. Pryce[1,2]

Abstract

Background: Heat tolerance is a trait of economic importance in the context of warm climates and the effects of global warming on livestock production, reproduction, health, and well-being. This study investigated the improvement in prediction accuracy for heat tolerance when selected sets of sequence variants from a large genome-wide association study (GWAS) were combined with a standard 50k single nucleotide polymorphism (SNP) panel used by the dairy industry.

Methods: Over 40,000 dairy cattle with genotype and phenotype data were analysed. The phenotypes used to measure an individual's heat tolerance were defined as the rate of decline in milk production traits with rising temperature and humidity. We used Holstein and Jersey cows to select sequence variants linked to heat tolerance. The prioritised sequence variants were the most significant SNPs passing a GWAS p-value threshold selected based on sliding 100-kb windows along each chromosome. We used a bull reference set to develop the genomic prediction equations, which were then validated in an independent set of Holstein, Jersey, and crossbred cows. Prediction analyses were performed using the BayesR, BayesRC, and GBLUP methods.

Results: The accuracy of genomic prediction for heat tolerance improved by up to 0.07, 0.05, and 0.10 units in Holstein, Jersey, and crossbred cows, respectively, when sets of selected sequence markers from Holstein cows were added to the 50k SNP panel. However, in some scenarios, the prediction accuracy decreased unexpectedly with the largest drop of − 0.10 units for the heat tolerance fat yield trait observed in Jersey cows when 50k plus pre-selected SNPs from Holstein cows were used. Using pre-selected SNPs discovered on a combined set of Holstein and Jersey cows generally improved the accuracy, especially in the Jersey validation. In addition, combining Holstein and Jersey bulls in the reference set generally improved prediction accuracy in most scenarios compared to using only Holstein bulls as the reference set.

Conclusions: Informative sequence markers can be prioritised to improve the genomic prediction of heat tolerance in different breeds. In addition to providing biological insight, these variants could also have a direct application for developing customized SNP arrays or can be used via imputation in current industry SNP panels.

*Correspondence: Mekonnen.HaileMariam@agriculture.vic.gov.au
[2] Agriculture Victoria Research, AgriBio, Centre for AgriBiosciences, Bundoora, VIC 3083, Australia
Full list of author information is available at the end of the article

Background

Heat tolerance is the ability of an animal to maintain its production and reproduction levels under hot and humid conditions. With increasing global warming effects on animal production, there is worldwide growing desire to breed for resilience to heat, in part, to meet the demand

Cheruiyot *et al. Genetics Selection Evolution*   (2022) 54:17

Page 2 of 18

of the increasing human population while coping with the challenges of hot and ever-changing production environments [1]. Dairy cows are often prone to heat stress due to the elevated metabolic heat of lactation. Temperature and humidity levels exceeding the thresholds that are considered as comfortable for the dairy cows and other farm animals can compromise production (reduced milk, growth, etc.), reproduction (e.g., reduced conception rates), and welfare (increased thirst and hunger), leading to substantial economic losses [2].

Considerable research has been conducted in many countries to assess heat tolerance and performance in farm livestock, including measuring changes in core body temperatures (e.g., rectal, vaginal, rumen, etc.) and thermal indices [e.g., temperature–humidity index (THI)] [3]. To study the effect of THI on milk production of dairy cows, Ravagnolo et al. [4] introduced a method in which daily milk records are merged with temperature-humidity data to measure the rate of milk decline associated with changes in heat stress. This method has been widely adopted in many countries [5–7] due to the availability of extensive test-day milk records from dairy farms and climate data from weather stations.

In Australia, Nguyen et al. [7] used test-day milk records (milk, fat, and protein yield) and climate data collected from across Australia's dairying regions to evaluate heat tolerance in dairy cattle, which culminated in the release to the dairy industry [through DataGene Ltd; (https://datagene.com.au/)] of the first genomic breeding values for this trait in 2017, with an average reliability of 38%. While current prediction estimates are promising, even a smaller lift in reliability is economically important to the wider industry since the genetic improvement is linearly related to the selection intensity, accuracy of estimated breeding values (EBV), genetic variation and is inversely proportional to the generation interval [8, 9]. The accuracy of prediction is the only component that is influenced by research in different ways to drive genetic improvement for a given trait whereas the other components (selection intensity, genetic variation, and generation cycle) are largely controlled by breeding companies and farmers.

Besides increasing the size of the reference population, one way to boost the accuracy of prediction is to increase the density of markers used in genomic predictions. However, replacing single nucleotide polymorphism (SNP) panels by the full set of whole-genome sequence variants has, in most cases, yielded limited, or no appreciable increase in the accuracy of prediction for various traits in cattle [10], sheep [11], and avian species [12]. Alternatively, a substantial increase in accuracy of prediction has been realized by augmenting standard industry SNP panels (e.g., a 50k SNP array) with a small set of informative or causal mutations for a trait [11, 13–15]. To fully maximize predictions, this approach requires a careful selection of informative markers. Thanks to the 1000 Bull Genomes project [16], it is now possible to use this sequence database to impute genotypes to the whole-genome sequence. This may facilitate a more accurate selection of highly informative variants for genomic predictions, especially for complex traits such as heat tolerance. Specifically, having a large sample size and high-resolution genotypes can help to identify many putative causal variants with medium- and small-sized effects.

In addition to sample size, the composition of the population used for discovering informative variants can have an impact on the genomic predictions of a trait. Several studies e.g., [17–19] have reported that the mapping precision of the causal variants underlying traits is improved in multi-breed compared to single-breed genome-wide association studies (GWAS), especially for quantitative trait loci (QTL) that segregate across breeds [19]. In a simulated study, van den Berg et al. [14] demonstrated that using variants that are close to the causal mutations can improve genomic predictions. With real data, Raymond et al. [20] found that the accuracy of prediction for stature increased when candidate variants that were discovered from a meta-GWAS of 17 cattle populations were used. In sheep, Moghaddar et al. [11] reported an enhanced accuracy of prediction for various production traits when they used pre-selected variants from the QTL discovery set that comprised multiple breed compositions. Besides these studies and several others that used single-breed sets to discover variants for traits, e.g., [15, 21], there is still a dearth of information on the value of variants that are discovered from multi-breed populations in genomic predictions. Notably, it is critical to ensure that the population(s) used to discover informative sequence variants for a trait is (are) independent of that used to train subsequent genomic predictions to avoid bias, as demonstrated by [22].

The main objective of this study was to quantify the accuracy of prediction of heat tolerance in Holsteins when sets of selected sequence markers from a GWAS based on a large sample of Holstein cows were added to the standard-industry 50k SNP panel that is routinely used for genomic evaluations in Australia. The selected variants are likely linked to causal mutations that underpin the genetic basis for heat tolerance [23] and, therefore, could enable more accurate and sustained genomic selection for heat tolerance. In addition, we investigated the accuracy of prediction when informative sequence markers discovered in Holstein cows are used in the genomic predictions of numerically smaller breeds, including Jersey and crossbred cattle. Moreover, we investigated the gain in accuracy of prediction when

Cheruiyot *et al. Genetics Selection Evolution*   (2022) 54:17

Page 3 of 18

using informative markers discovered in a combined set of Holstein and Jersey cows (i.e., a multi-breed QTL discovery set). Finally, we compared the gain in accuracy when single-breed (Holstein bulls) versus multi-breed (Holstein + Jersey bulls) reference sets are used in the genomic predictions.

Methods

Phenotypes

The phenotypes were obtained from DataGene (Data-Gene Ltd., Melbourne, Australia; https://datagene.com.au/), and included test-day milk, fat, and protein yield for Holstein, Jersey, and Holstein–Jersey crossbred cows collected from dairy herds between 2003 and 2017 that were combined with climate data (daily temperature and humidity) obtained from weather stations across Australia's dairying regions. The distribution of dairy herds and weather stations, data filtering, and the calculation of environmental covariate [i.e., temperature–humidity index (THI)] used in this work were described previously [23, 24].

The rate of decline (slope) in milk, fat, and protein yields due to heat stress events was estimated using reaction norm models described by [24]. Briefly, data on milk, fat, or protein yields were adjusted for the fixed effects, including herd-test-day, year-season of calving, parity, Legendre polynomials (order 3) on the cow age on the day of the test, and the Legendre polynomials (order 8) on the interaction between parity and DIM. The number of records (tests) per Holstein bull (N = 3323) ranged from 4 to 263,067 and that per Jersey bull (N = 852) ranged from 5 to 54,242. The number of daughters per Holstein bull ranged from 1 to 18,613 with an average of 149, and that per Jersey bull ranged from 1 to 3169 with an average of 88.8. The random non-genetic effect fitted in the model included a random regression on a linear orthogonal polynomial of THI, where the intercept represents the level of mean milk yield, and the linear component represents the change in milk yield (slope) due to heat stress for each cow, and a residual term. The reaction norm models [4] were used in the analyses with the THI threshold set at 60 (i.e., if THI < 60, then THI = 60) based on previous work in Australia [7, 25] showing that milk yield traits begin to decline at this THI threshold. The analyses to derive trait deviations (TD), which represent phenotypes adjusted for all fixed effects (i.e., the slope for each cow) were conducted using ASReml v4.2 [26]. Slope solutions (i.e., TD) for each bull's daughters were averaged to obtain heat tolerance slope traits for bulls and were equivalent to daughter trait deviations (DTD). As in [27], the DTD in this study should be treated as approximations equivalent to the averages of daughter phenotypes since the models did not include pedigree

data. Notably, the derivation of intercept and slope traits in this initial step was necessary because of the computational resources required to fit complex models and the large sample size in our study. From here on, the slope traits derived from milk, fat, and protein yield records are referred to as heat tolerance milk (HTMYslope), fat (HTFYslope), and protein (HTPYslope).
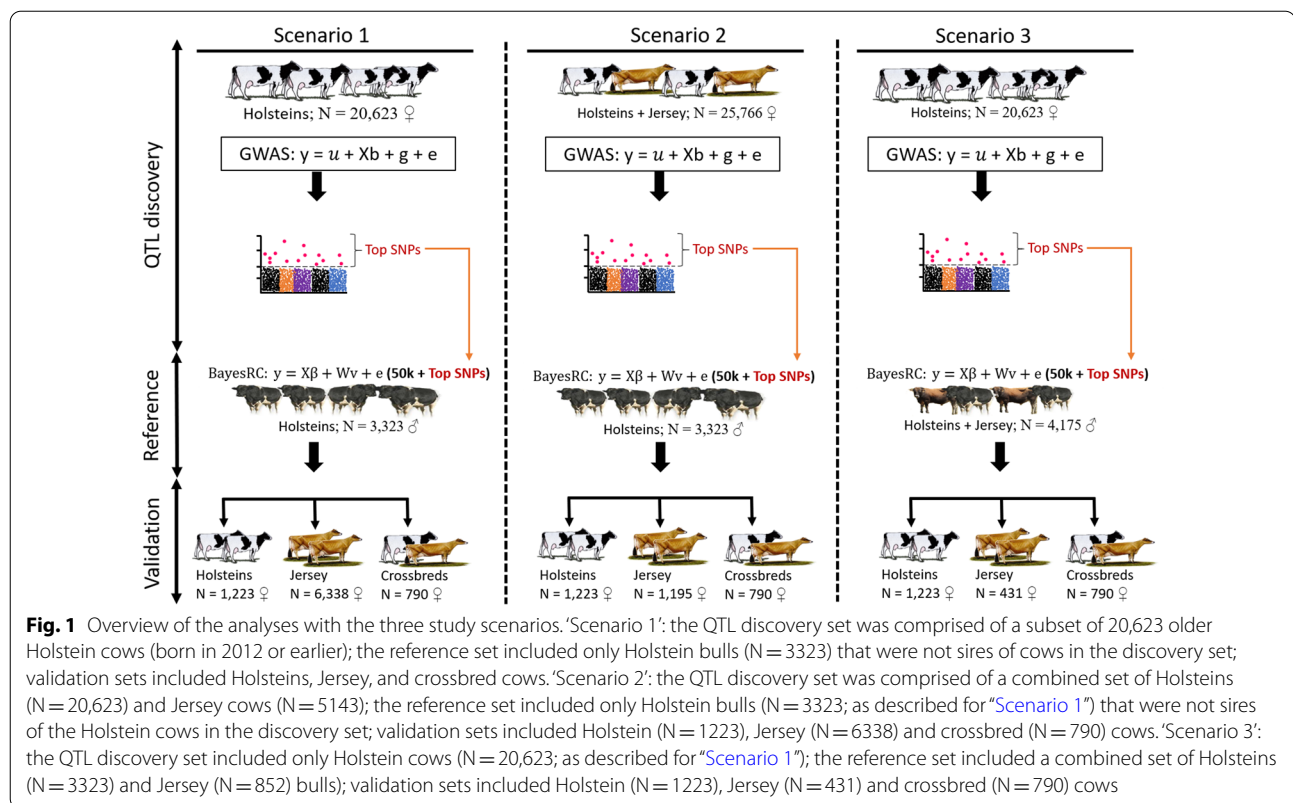
Genotypes

Two genotype datasets were prepared for the above cows and bulls with heat tolerance phenotypes: the standard 50k SNP chip (i.e., Illumina 50k Bovine Bead Chip used in previous work in Australia [7] and 15,098,486 imputed whole-genome sequence variants (WGS). The WGS was imputed [28] using the genomic sequence data from Run7 of the 1000 Bull Genome Project based on the ARS-UCD1.2 reference genome (http://1000bullgenomes.com/), and variants were filtered on the estimated imputation accuracy ($R^2 > 0.4$) and minor allele frequency (MAF > 0.005). The detailed imputation procedure is described in [23].

Study design: discovery, reference, and validation datasets

The animals with genotypes and heat tolerance phenotypes included Holsteins (29,107 ♀/3323 ♂), Jerseys (6338 ♀/1364 ♂), and Holstein–Jersey crossbreds (790 ♀/0 ♂). These animals were split into three independent groups to achieve the specific objectives: (i) a QTL discovery set that was used to discover informative sequence markers for heat tolerance, (ii) a reference set that was used to develop genomic prediction equations, and (iii) independent validation sets that were used to assess genomic prediction accuracy. The validation sets included three breed subsets: Holstein, Jersey, and crossbred cows. Across all the prediction scenarios, we ensured that the QTL discovery set used in the GWAS was independent of the reference set used in genomic predictions to minimise bias in the predictions [22]. The different sets of animals used for each group (QTL discovery, reference, and validation) are described schematically in Fig. 1, with a more detailed description in the following paragraphs.

Scenario 1

Scenario 1 aimed at testing the value of pre-selected sequence variants from Holstein cows in the genomic prediction of the same breed as well as in the prediction of other numerically smaller breeds, including Jersey and crossbred cows: (i) a QTL discovery set that included 20,623 Holstein cows born in 2012 or earlier; (ii) a reference set that included 3323 Holstein bulls with none of these bulls siring the cows in the discovery set to ensure the independence of the phenotypes between the two datasets; and (iii) three validation sets, i.e. (a) that was comprised of 1223 younger Holstein cows (born in 2013

Cheruiyot *et al. Genetics Selection Evolution*   (2022) 54:17

Page 4 of 18

Fig. 1 Overview of the analyses with the three study scenarios. 'Scenario 1': the QTL discovery set was comprised of a subset of 20,623 older Holstein cows (born in 2012 or earlier); the reference set included only Holstein bulls (N = 3323) that were not sires of cows in the discovery set; validation sets included Holsteins, Jersey, and crossbred cows. 'Scenario 2': the QTL discovery set was comprised of a combined set of Holsteins (N = 20,623) and Jersey cows (N = 5143); the reference set included only Holstein bulls (N = 3323; as described for "Scenario 1") that were not sires of the Holstein cows in the discovery set; validation sets included Holstein (N = 1223), Jersey (N = 6338) and crossbred (N = 790) cows. 'Scenario 3': the QTL discovery set included only Holstein cows (N = 20,623; as described for "Scenario 1"); the reference set included a combined set of Holsteins (N = 3323) and Jersey (N = 852) bulls; validation sets included Holstein (N = 1223), Jersey (N = 431) and crossbred (N = 790) cows

or later), which were not daughters of the Holstein bulls used in the reference set, (b) that included 6338 Jersey cows, and (c) that included 790 crossbred cows. Each of the three validation sets was randomly split into two subsets of approximately equal size (see Additional file 1: Table S1) to facilitate the calculation of standard errors of prediction.

Scenario 2

Scenario 2 aimed at testing whether pre-selected informative markers from a multi-breed population improves the accuracy of predictions compared to pre-selected markers from the single-breed QTL discovery set: (i) a QTL discovery set that included older cows i.e. Holstein (N = 20,623 ♀; born in 2012 or earlier) and Jersey (N = 5143 ♀; calved for the first time in 2014 or earlier); (ii) a reference set that comprised Holstein bulls (N = 3323); (iii) three validation sets, i.e. (a) Holstein cows (N = 1223 ♀; as described for "Scenario 1"), (b) Jersey cows (N = 1195; younger cows that calved for the first time in 2014 or later); and (c) crossbred cows (N = 790; as for "Scenario 1"). Each validation set was randomly split into two subsets (see Additional file 1: Table S1), and these were the same subsets used in "Scenario 1" for Holsteins and crossbreds.

Scenario 3

Scenario 3 aimed at testing the accuracy of prediction when using a multi-breed reference set as follows: (i) a QTL discovery set of Holstein cows (N = 20,623; born in 2012 or earlier as described for "Scenario 1", i.e., the single-breed discovery set); (ii) a reference set that consisted of a multi-breed set of Holstein bulls (N = 3323 ♂; as for "Scenario 1") and Jersey bulls (N = 852 ♂); and (iii) three validation sets, i.e. (a) Holstein cows (N = 1223; as for "Scenarios 1 and 2"), (b) Jersey cows (N = 431) that were not daughters of the bulls used in the multi-breed reference set; and (c) crossbred cows (N = 790; as for "Scenarios 1 and 2"). Validation sets were split into two subsets, and for Holstein and crossbred validation they were the same subsets as in "Scenarios 1 and 2".

QTL discovery and selection of informative markers ('top SNPs')

To identify informative sequence variants for heat tolerance traits (using the "discovery" sets described above), we performed a GWAS using mixed linear models to test associations between individual SNPs and cow's slope traits using the GCTA software [29]. The details of the GWAS for the Holstein discovery set are described in [23]. Briefly, a linear model was fitted to cow's (N = 20,623

Cheruiyot *et al. Genetics Selection Evolution*   (2022) 54:17

Page 5 of 18

Holsteins) slopes for production trait (HTMYslope, HTFYslope, and HTPYslope that were pre-adjusted for the nongenetic effects described by [24]), for each autosomal SNP (~15 million SNPs). The model included a genomic relationship matrix (GRM) constructed based on the 50k SNP genotype data of the cows. The same model was used when performing GWAS for the multi-breed (Holstein and Jersey cows; N = 25,766) QTL discovery set except that an additional binary covariate was fitted to account for breed effect.

To increase the power of GWAS to identify pleiotropic variants for heat tolerance from the three slope traits, we combined the above single-trait GWAS results in a multi-trait meta-GWAS (following methods described in [30]), and described for the Holstein data set in [23].

Using either the single-trait or multi-trait GWAS results, we selected informative variants defined as 'top SNPs' for each slope trait as follows:

1. Moving along each chromosome in 100-kb sliding windows, we chose the single most significant SNP from within the 100-kb window and then moved 50 kb along the chromosome to the next 100-kb window. This was repeated starting from the proximal to the distal end of each chromosome, as in [11]. To be selected, the SNP had to pass a GWAS threshold of $-\log_{10}(\text{p value}) \geq 3$. In addition, we tested a more relaxed GWAS threshold of $-\log_{10}(\text{p value}) \geq 2$ to determine if it could help the capture of variants with much smaller effect sizes for heat tolerance in addition to those with large effects [23].
2. Among each set of selected 'top SNPs', we removed one SNP of any pair in strong linkage disequilibrium (LD) ($r^2 > 0.95$) using the PLINK software [31], with the [− indep-pairwise 50 5 0.95] option, where LD is calculated within 50-SNP sliding windows, each time sliding five SNPs along the chromosome.

Genomic prediction using BayesR and BayesRC methods
We used BayesR [17, 32] to calculate genomic breeding values (GBV) for each cow in the validation set based on the standard-industry 50k SNP data. BayesR assumes one class of SNPs that are modelled as a mixture of four normal distributions corresponding to zero-, small-, medium- and large-sized effects [17, 32]. Currently, the Australian dairy industry uses the standard 50k SNP panel for routine genomic evaluations; thus, it served as the benchmark to test the added value of selected sequence variants (i.e., 'top SNPs'). Furthermore, the standard-industry 50k SNP panel includes a set of variants that were not selected intentionally for heat tolerance, which was ideal for our study.

The BayesR model fitted 42,572 variants (SNPs with a MAF > 0.005) from the 50k SNP panel using bulls (N = 3323) as reference set:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Wv} + \mathbf{e}, \tag{1}$$

where $\mathbf{y}$ is a vector of heat tolerance slope phenotypes (HTMYslope, HTFYslope, and HTPYslope) or intercept (i.e., mean yield) traits (MYint, FYint, and PYint); $\mathbf{X}$ is a design matrix; $\mathbf{\beta}$ is a vector of fixed effect solutions; $\mathbf{W}$ is a centred design matrix of SNP genotypes; $\mathbf{v}$ is a vector of SNP effects, modelled to have four possible normal distributions: $\mathbf{v} \sim N(\mathbf{0}, \mathbf{I}\sigma_i^2)$, where $\sigma_i^2 = \{0.0, 0.0001 * \sigma_v^2, 0.001 * \sigma_v^2, 0.01 * \sigma_v^2\}$, corresponding to zero-, small-, medium- and large-sized effects, respectively with σ_v^2 the additive genetic variance; $\mathbf{e}$ is a vector of residual errors $N(\mathbf{0}, \mathbf{E}\sigma_e^2)$, with $\mathbf{E}$ a diagonal matrix calculated as $diag(1/w_i)$, with w_i being a weighting factor for bull i calculated based on the available number records following [33]:

$$w_{bull_i} = \frac{1 - h^2}{ch^2 + \frac{(4 - h^2)}{p}}, \tag{2}$$

where h^2 is the heritability; c is the proportion of the genetic variance that is not accounted by the SNPs ($c = 0.2$); and p is the number of daughters for each bull.

The same model (Eq. 1) was used when analysing the multi-breed reference population (Holstein and Jersey; N = 4175), except that a binary covariate was fitted to account for the breed effect. To account for polygenic effects, we tested models with or without pedigree relationships, which yielded correlation estimates of SNP effects close to 1.0. Therefore, based on these preliminary analyses, we decided not to include pedigree data in the subsequent models.

To calculate GBV using a combined set of 50k SNPs and the pre-selected SNPs from GWAS (i.e., 50k + 'top SNPs') for the validation cows, we used the BayesRC method [34]. BayesRC is an extension of BayesR in which two or more classes of SNP effects are modelled: the SNPs within each class are fitted as a mixture of four normal distributions as in BayesR so that the mixture distribution can differ for each SNP class. In our study, the SNPs from the standard 50k array (42,572 SNPs) were allocated to class I and the pre-selected 'top SNPs' from GWAS to a separate class II. Class I variants are considered as a random set from the 50k array (as indicated earlier), while Class II variants ('top SNPs') may be enriched with causal and/or highly predictive mutations for heat tolerance.

For both BayesR and BayesRC models, we performed five Markov chain Monte Carlo (MCMC) replicate chains, each with 40,000 iterations, of which 20,000

were discarded as burn-in for all the traits. These iterations gave stable convergence across the five replicates. The results from these replicates were averaged to get the final estimate. To facilitate the calculation of standard errors, we randomly split the validation cows into two subsets of approximately equal size (see Additional file 1: Table S2) and then performed analyses (i.e., the BayesR and BayesRC) for each subset, separately.

For each analysis (described above), the accuracy of prediction was calculated as described in [11]: $Accuracy = \frac{r_{GBV,phen}}{\sqrt{h^2}}$, where $r_{GBV,phen}$ is the correlation of GBV with TD phenotypes (slope or intercept traits); (h^2 is the genomic heritability of the trait computed from 50k SNP data based on 29,107 Holstein cows). These heritability estimates used to calculate prediction accuracies are in Additional file 1: Tables S2 and S3. The corresponding standard errors of the accuracies were estimated as: $SE = SD/\sqrt{N}$, where N is the number of random validation subsets ($N = 2$); SD is the standard deviation of the accuracies of prediction calculated from the two validation sets per breed (i.e., Holstein, Jersey, and crossbred cows). The dispersion bias of the accuracy of prediction for different traits was assessed as the regression coefficient of the TD phenotypes on the GBV in the validation set and their corresponding standard errors calculated as described for the *SE* of the accuracies of prediction above. The regression coefficient = 1.0 indicates no dispersion bias, whereas values > 1.0 or < 1.0 indicate that the GBV are subject to deflation or inflation, respectively.

Results

Genomic heritability

Genomic heritability estimates based on 29,107 Holstein cows using the 50k SNP array were similar for all the slope (heat tolerance) traits (see Additional file 1: Table S2). The genomic heritability estimates based on Jersey cows (N = 6338) were comparable to those based on Holstein cows with values of 0.26 ± 0.02, 0.23 ± 0.02, and 0.25 ± 0.02 for the HTMYslope, HTFYslope and HTPYslope traits, respectively (see Additional file 1: Table S2). However, the values for crossbred cows (N = 790) were estimated with large standard errors [0.58 ± 0.10 (HTMYslope); 0.34 ± 0.11 (HTFYslope); 0.51 ± 0.10 (HTPYslope)], which is most likely due to the small sample size used. In contrast, the genomic heritability estimates for intercept traits were relatively larger than those for heat tolerance traits (see Additional file 1: Tables S2 and S3). In this study, we computed the accuracy of genomic predictions across all validation sets using the heritability estimates from Holstein cows (N = 29,107) that were estimated with the smallest standard errors.

Pre-selection of heat tolerance SNPs (i.e., top SNPs)

Single-breed (Holstein cows) QTL discovery set

Table 1 includes the number of selected informative sequence variants for heat tolerance defined as 'top SNPs' from single-trait GWAS and multi-trait meta-analyses of the Holstein cow discovery set (i.e., the single-breed discovery set; see "Methods" section—"Scenario 1"). Using a more stringent GWAS cut-off threshold of $-\log10(p\text{-value}) \geq 3$ resulted in about a fivefold smaller number of selected 'top SNPs' than a comparatively relaxed GWAS cut-off of $-\log10(p\text{-value}) \geq 2$. The numbers of selected 'top SNPs' at a $-\log10(p\text{-value}) \geq 2$ from single-trait GWAS (after pruning pairs of markers in strong LD, $r^2 > 0.95$) were equal to 9207 (HTMYslope), 9352 (HTFYslope), and 9633 (HTPYslope), and the numbers of those selected at a $-\log10(p\text{-value}) \geq 3$ were equal to 1654 (HTMYslope), 1708 (HTFYslope) and 1624 (HTPYslope) (Table 1). The largest number of 'top SNPs' was obtained for HTPYslope, followed by HTFYslope and HTMYslope (Table 1). Although the number of variants that passed the GWAS cut-off was largest for HTPYslope, the strength of the GWAS signal (peak) across the genome (see Additional file 2: Figs. S1 and S2) was relatively weak for this trait compared to the other traits (i.e., HTMYslope and HTFYslope).

A large proportion (> 50%) of the selected 'top SNPs' had a lower MAF compared to the SNPs in the 50k panel (see Additional file 2: Fig. S3). Compared to single-trait GWAS, and as expected, fewer 'top SNPs' were selected from the multi-trait meta-analyses of slope traits at a more stringent $[-\log10(p\text{-value}) \geq 3$; N = 2365 SNPs] than at a relaxed GWAS cut-off $[-\log10(p\text{-value}) \geq 2$; N = 9090 SNPs] (Table 1). Comparatively, a slightly larger number of 'top SNPs' was selected across intercept traits than across heat tolerance traits (see Additional file 1: Table S4).

Table 1 Number of informative markers for heat tolerance defined as 'top SNPs' selected from single-trait GWAS and multi-trait meta-analyses of heat tolerance slope traits of Holstein discovery cow set (N = 20,623)

Trait	'Top SNPs' ($-\log10(p\text{-value}) \geq 2$)	'Top SNPs' ($-\log10(p\text{-value}) \geq 3$)
HTMYslope	9207 (51,750)	1654 (44,219)
HTFYslope	9352 (51,894)	1708 (44,277)
HTPYslope	9633 (52,168)	1624 (44,190)
Meta-GWAS	9090 (51,636)	2365 (44,929)

Markers were selected based on the GWAS cut-off thresholds of $-\log10(p\text{-value}) \geq 2$ and $-\log10(p\text{-value}) \geq 3$. The values in brackets are the final number of SNPs after adding selected 'top SNPs' to the 50k SNP data used in the BayesRC analyses (i.e., 42,572 SNPs + top SNPs). Traits are defined as heat tolerance milk (HTMYslope), fat (HTFYslope) and protein (HTPYslope) yield slope traits

Cheruiyot *et al. Genetics Selection Evolution*   (2022) 54:17

Page 7 of 18

The proportion of phenotypic variance accounted for by the 'top SNPs' at a GWAS p-value cut-off of $(-\log10(\text{p-value}) \geq 2$; Table 1) varied across traits and populations. In general, the 'top SNPs' for HTMYslope explained a relatively larger variance compared to the 'top SNPs' for HTFYslope and HTPYslope across the studied scenarios. In the Holstein validation set, variance estimates for HTMYslope, HTFYslope and HTPYslope were 0.24 ± 0.05, 0.22 ± 0.05, and 0.21 ± 0.05, respectively. In the Jersey validation set, variance estimates explained by the 'top SNPs' for HTMYslope, HTFYslope and HTPYslope were 0.23 ± 0.02, 0.18 ± 0.02 and 0.22 ± 0.02, respectively. The variance estimates in the crossbred validation set were 0.55 ± 0.10, 0.24 ± 0.10, 0.37 ± 0.10, for HTMYslope, HTFYslope and HTPYslope, respectively. The large standard error for the variance estimates in crossbreds is likely due to their small sample size (N = 790).

Multi-breed (Holstein + Jersey cows) QTL discovery set

When Holstein cows (N = 20,623) were combined with Jersey cows (N = 5143) in the QTL discovery set (i.e., the multi-breed QTL discovery set, see "Methods" section— "Scenario 2"), we found a smaller number of selected 'top SNPs' (after pruning pairs of markers in strong LD, $r^2 > 0.95$) from the single-trait GWAS at $-\log10(\text{p-value}) \geq 2$ [HTMYslope = 6132; HTFYslope = 6286; HTPYslope = 6422] compared to those from the single-breed QTL discovery set at the same significance cut-off (described above). However, when compared to the single-breed GWAS (only Holstein cows), using a multi-breed QTL discovery set (Holsteins + Jersey cows) increased the strength of the GWAS signals in some genomic regions (e.g., on *Bos taurus* chromosome (BTA) 14 near the *DGAT1* gene) (see Additional file 2: Figs. S4 and S5).

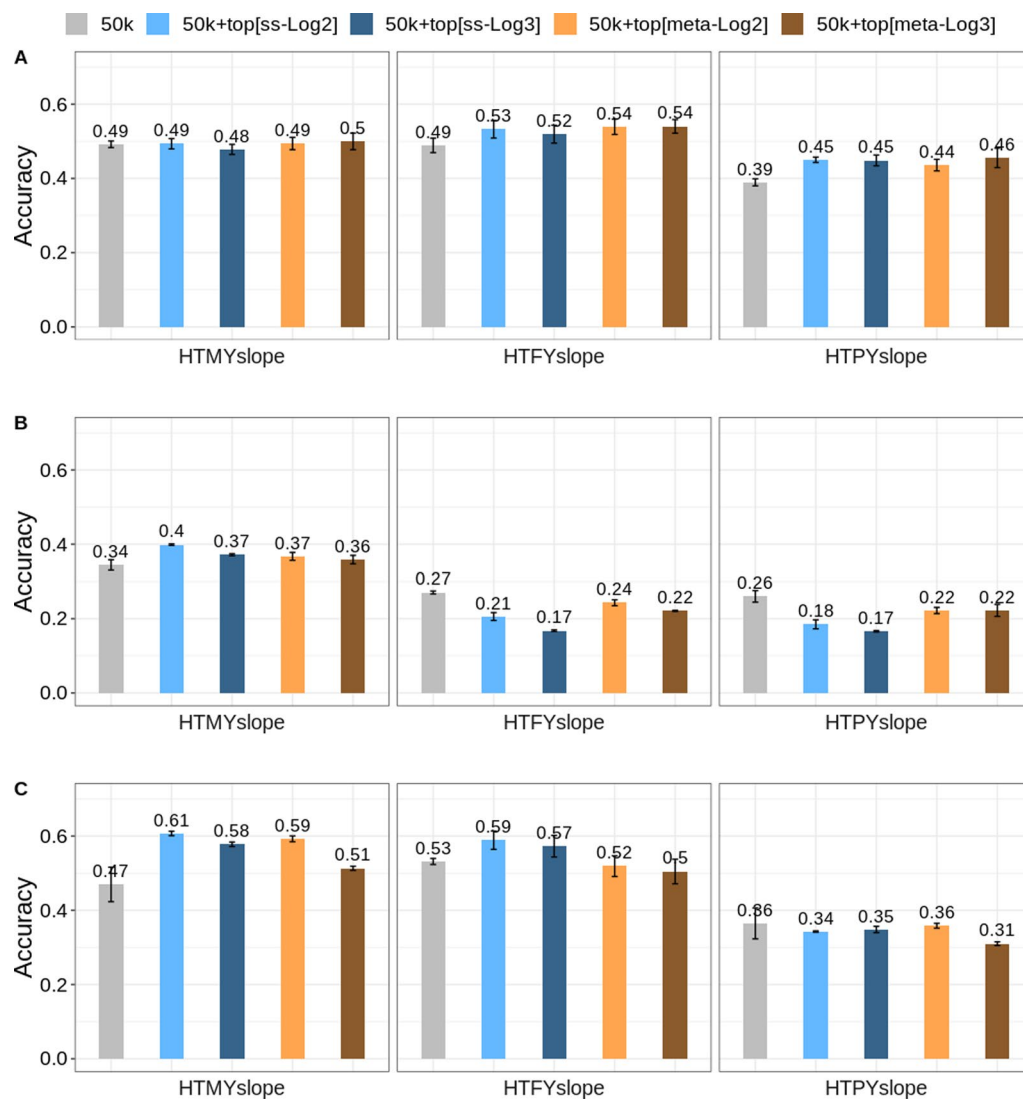### Genomic prediction using selected SNPs from the single-breed discovery set ('Scenario 1')

Figure 2 shows the accuracy of predictions when the selected 'top SNPs' from a single-breed (Holstein cows; N = 20,623) QTL discovery set were added to the standard 50k SNP array and analysed using the BayesRC model. For this comparison, the reference set included only Holstein bulls (N = 3323) and the validation set included Holstein (N = 1223), Jersey (N = 6338) and crossbred (N = 790) cows. The gain in accuracy for the different traits and models varied across the three validation sets. The increase in the accuracy of prediction was generally consistent for HTMYslope across most of the different scenarios (50k + 'top SNPs') tested, but not for HTFYslope and HTPYslope, particularly in the Jersey validation set. In general, the increase in accuracy of prediction ranged from 0.001 to 0.09, with the largest

increase (0.09) observed for HTMYslope in the crossbred validation set. In the Holstein validation set, the accuracy of prediction across all scenarios hardly changed for HTMYslope ranging from -0.01 to 0.008 units, whereas the changes for HTFYslope and HTPYslope ranged from 0.03 to 0.05 and from 0.04 to 0.06, respectively (Fig. 2). For the intercept traits in the Holstein validation set, fitting 50k + 'top SNPs' (in the BayesRC model) generally increased the accuracy of prediction compared to BayesR (using only 50k SNPs) by up to 0.04, 0.03 and 0.05 for FYint, PYint and MYint, respectively (see Additional file 2: Fig. S6).

For the crossbred validation set, the change in the accuracy of prediction from BayesRC (50k + 'top SNPs') over BayesR (fitting only 50k SNPs) ranged from -0.004 to 0.09, from -0.06 to 0.02, and from -0.04 to 0.009 for HTMYslope, HTFYslope, and HTPYslope, respectively (Fig. 2). Similarly, compared to BayesR (using only 50k SNPs), the accuracy of prediction for the intercept traits in crossbreds hardly changed across most prediction scenarios when fitting 50k + 'top SNPs' (BayesRC), with changes ranging from -0.01 to 0.01, from -0.02 to 0.02, and from -0.04 to -0.02 for MYint, FYint, and PYint, respectively (see Additional file 2: Fig. S6).

In the Jersey validation set (using 50k + 'top SNPs' in the BayesRC), we observed that the accuracy for HTMYslope increased compared to that of BayesR (using only 50k SNPs) across all prediction scenarios, with changes ranging from 0.01 to 0.05 units. However, the accuracy of prediction decreased considerably for HTFYslope (-0.10) and HTPYslope (-0.09) when the 'top SNPs' from Holstein cows were used in Jerseys with a slightly larger decrease in accuracy of prediction when using 'top SNPs' from the single-trait GWAS than those from the multi-trait meta-analysis (Fig. 2). Similarly, compared to BayesR using only 50k SNPs, the accuracy of prediction dropped for intercept traits when fitting the 'top SNPs' from the Holstein cow discovery set in Jerseys with changes ranging from -0.10 to -0.03 (FYint) and from -0.09 to -0.04; (PYint), while the accuracy of prediction for MYint from BayesRC increased from 0.02 to 0.06 compared to BayesR (see Additional file 2: Fig. S6).

Across all prediction scenarios (Fig. 2), using 'top SNPs' from the relaxed GWAS cut-off value of $[-\log10(\text{p-value}) \geq 2$ ($\sim$9000 SNPs) in the BayesRC model did not yield a substantial difference in accuracy of prediction compared to those based on the 'top SNPs' from a more stringent GWAS threshold $[-\log10(\text{p-value}) \geq 3$ ($\sim$2000 SNPs). The change in accuracy of prediction across all validation sets and traits ranged from -0.07 to 0.09 units and from -0.10 to 0.06 units when the 'top SNPs' from relaxed and more stringent GWAS cut-off p-values were added to the 50k SNP panel (BayesRC)

Cheruiyot *et al. Genetics Selection Evolution* (2022) 54:17

Page 8 of 18

Fig. 2 Accuracy of genomic predictions (Holstein only reference) using either 50k SNP data (colored grey) or 50k + a range of 'top SNPs' sets (selected from the Holstein QTL discovery set). The 'top SNPs' were selected from single-trait GWAS (colored blue) and multi-trait meta-analysis (colored orange) at a less stringent cut-off threshold of $-\log10(\text{p-value}) \geq 2$ [~9000 SNPs] and at a more stringent p-value of $-\log10(\text{p-value}) \geq 3$ [~2000 SNPs]. Accuracy of predictions are provided for three cow validation sets: **a** Holsteins, N = 1223), **b** Jersey, N = 6338, and **c** Holstein–Jersey crossbreds, N = 790. The traits analysed are heat tolerance milk (HTMYslope), fat (HTFYslope), and protein (HTPYslope) yield slopes. The genomic predictions were generated using either BayesR (50k SNP set) or BayesRC (50k + top SNPs). Vertical lines represent the standard errors calculated from two random validation subsets

compared to the results from BayesR (using only 50k) (Fig. 2). In general, using 'top SNPs' from the more stringent GWAS cut-off in the BayesRC model yielded a larger dispersion bias than the 'top SNPs' from the relaxed GWAS cut-off threshold for heat tolerance slope traits (see Additional file 2: Fig. S7). However, for intercept traits, the BayesRC model using 50k + top SNPs showed little or no increase in the dispersion bias (see Additional file 2: Fig. S8).

Moreover, there was no substantial difference in accuracy from BayesRC when using the 'top SNPs' from single-trait GWAS versus the 'top SNPs' from multi-trait meta-GWAS of slope traits across different prediction scenarios (Fig. 2). The change in accuracy based on the selected 'top SNPs' from the single-trait GWAS ranged from 0.002 (HTMYslope) to 0.06 (HTPYslope), from −0.05 (HTFYslope) to 0.02 (HTMYslope), and from −0.06 (HTFYslope) to 0.07 (HTMYslope) in Holsteins, Jerseys, and crossbred validation sets, respectively. These

changes are comparable to those obtained using 'top SNPs' from the meta-analysis of slope traits with changes ranging from −0.01 (HTMYslope) to 0.06 (HTPYslope), from −0.10 (HTFYslope) to 0.05 (HTMYslope), and from −0.006 (HTPYslope) to 0.09 (HTMYslope) in Holsteins, Jersey, and crossbred validation sets, respectively. Since the results were comparable when using 'top SNPs' from either relaxed or stringent GWAS cut-off values, we hereafter, only report the results based on the 'top SNPs' from the single-trait GWAS at the relaxed cut-off threshold (i.e., −log10(p-value) ≥ 2).

The dispersion bias across heat tolerance traits in the Holstein validation set ('Scenario 1') showed that the GBV were deflated (see Additional file 2: Fig. S7). In contrast, the predictions were less biased (i.e., regression coefficient values closer to 1.0) for the intercept traits, particularly for MYint and PYint in the Holstein validation set (see Additional file 2: Fig. S8). In the Jersey validation set, the dispersion bias for HTFYslope showed that the GBV were inflated (see Additional file 2: Fig. S7). Also, the GBV were inflated in the Jersey validation set for HTPYslope when the 'top SNPs' were added to the 50k SNP array and analysed using BayesRC [i.e., 1.11 in the BayesR model versus 0.79 in the BayesRC model; (see Additional file 2: Fig. S7)]. The dispersion bias for the Jersey validation set was inflated across intercept traits (see Additional file 2: Fig. S8). The predictions were extremely deflated in the crossbreds, particularly for HTMYslope (bias > 1.7), which is likely due to the small sample size and population used. The dispersion bias for heat tolerance traits was even more pronounced when the selected 'top SNPs' were added to the 50k SNP data using the BayesRC model compared to the estimates using the BayesR model and only the 50k SNP data (see Additional file 2: Fig. S7).

Genomic prediction using selected SNPs from multi-breed discovery set ('Scenario 2')

Figure 3 shows the change in accuracy of prediction (based on the BayesRC) when the selected 'top SNPs' (GWAS cut-off of −log10(p-value) ≥ 2) from the multi-breed (Holstein + Jersey cows) QTL discovery set were added to the 50k SNP array for which the reference set consisted only of Holstein bulls. In general, the change in accuracy of prediction across all traits and validation sets ranged from −0.05 (HTPYslope) in Jersey to 0.11 (HTMYslope) in crossbred cows. In the Holstein validation set (N = 1223), the accuracy of prediction increased across all traits with the greatest increase for HTPYslope (0.03) followed by HTFYslope (0.02) and HTMYslope (0.005), respectively. In this validation set, the dispersion bias was higher than 1.0 across all traits, indicating deflated GBV. The bias decreased slightly for HTMYslope

but increased for HTPYslope and HTFYslope when the 'top SNPs' were fitted in the BayesRC model (Fig. 3).

In the Jersey validation set (N = 1195), the change in accuracy of prediction (based on the BayesRC model) was not consistent across traits (Fig. 3). When using the selected 'top SNPs' from the multi-breed QTL discovery set, the accuracy of prediction increased for HTMYslope (0.03) and HTFYslope (0.02) but decreased for HTPYslope (−0.05). These values contrast with those obtained using selected 'top SNPs' from the single-breed QTL discovery set (only Holsteins; see "Methods", "Scenario 1"), where we found a change in accuracy of 0.09, 0.04, and 0.01 for HTMYslope, HTFYslope, and HTPYslope, respectively, when using a smaller subset of Jersey cows (i.e., N = 1195) instead of 6338 cows (as in "Scenario 1"). Unlike in "Scenario 1" where we found that GBV were inflated across heat tolerance traits (see Additional file 2: Figure S7) in Jerseys, the predictions were generally close to 1.0 in "Scenario 2", particularly for HTMYslope (Fig. 3).

In the crossbreds (N = 790), using 'top SNPs' discovered in the multi-breed (Holsteins + Jersey cows) set (based on the BayesRC models) yielded a larger (0.11 units) change in accuracy of prediction than with BayesR (using only 50k SNPs) for HTMYslope compared to a drop in accuracy from BayesRC over BayesR of −0.005, and −0.03 units for HTFYslope and HTPYslope, respectively (Fig. 3). Comparatively, using the 'top SNPs' from the single-breed (only Holsteins) QTL discovery set in crossbreds ('Scenario 1') yielded a change in accuracy from BayesRC over BayesR of 0.09, 0.02, and −0.006 for HTMYslope, HTFYslope, and HTPYslope, respectively. As in "Scenario 1", the dispersion bias in crossbreds for HTMYslope was extreme (> 1.7) compared to the other traits. In this crossbred validation set ('Scenario 2'), the bias increased more for HTMYslope but decreased for HTFYslope and HTPYslope when fitting the selected 'top SNPs' in BayesRC (Fig. 3).

Genomic prediction using multi-breed reference set ('Scenario 3')

When we used a multi-breed (Holstein + Jersey bulls) reference set in which the 'top SNPs' were only from the Holstein cow QTL discovery set (see "Methods" section, "Scenario 3"), we found a consistent increase in the accuracy of prediction in most cases (Fig. 4). The accuracy of prediction decreased only for HTMYslope (− 0.06) and HTPYslope (− 0.002) in the Jersey validation set for this scenario. The change in accuracy of prediction from BayesRC over BayesR for HTMYslope, HTFYslope, and HTPYslope were: [− 0.01, 0.05, and 0.05], [− 0.06, − 0.002, and 0.01], and [0.10, 0.03, and 0.04] in the Holstein (N = 1223), Jersey (N = 431) and crossbred (N = 790) cow validation sets, respectively (Fig. 4). These

Cheruiyot *et al. Genetics Selection Evolution*   (2022) 54:17

Page 10 of 18
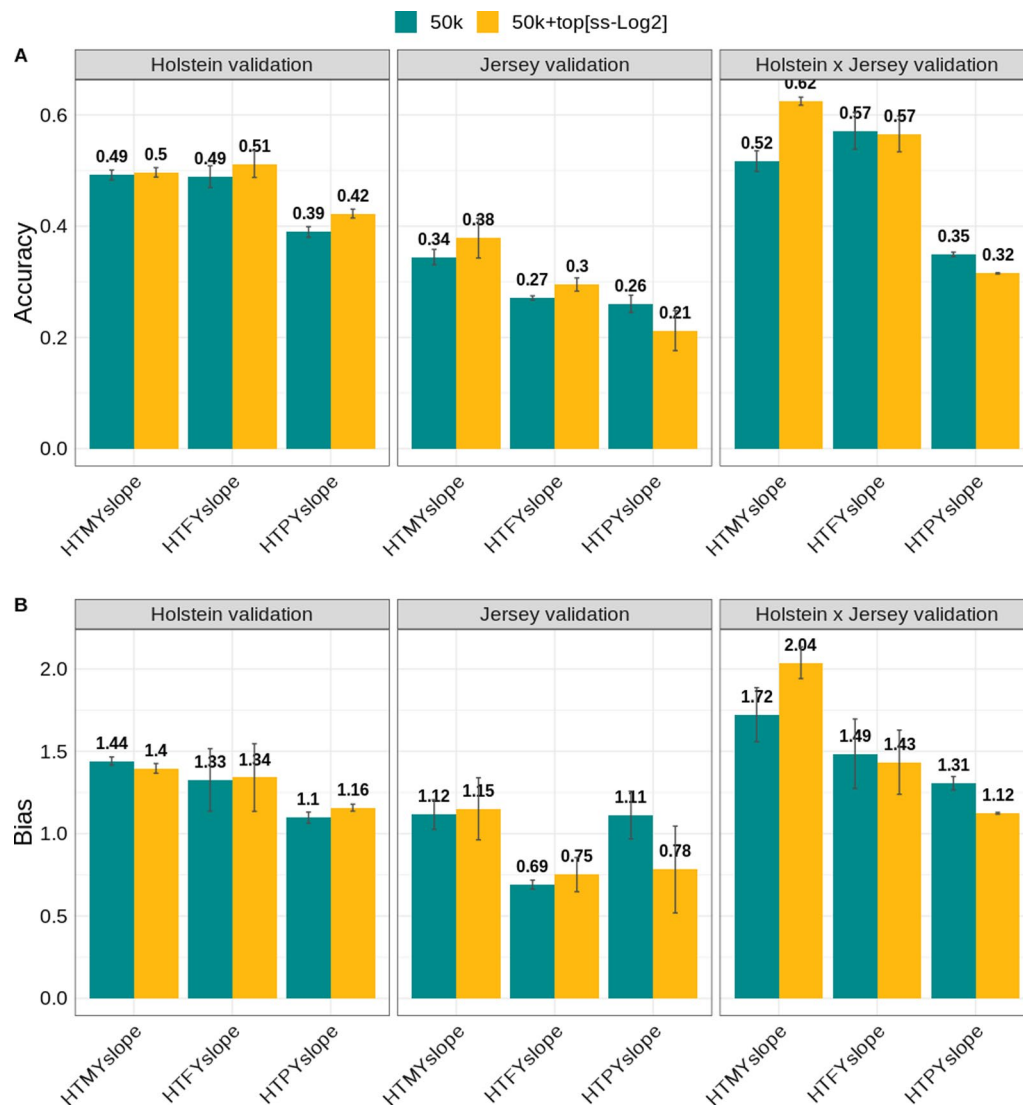
Fig. 3 Accuracy and dispersion bias in Holstein (N = 1223), Jersey (N = 1195) and crossbred (N = 790) cows when using 50k + 'top SNPs' selected from the multi-breed (Holstein + Jersey) QTL discovery set. Holstein bulls (N = 3323) were used as the reference set for genomic predictions. The 'top SNPs' were selected based on a single-trait GWAS cut-off of [− log10(p-value) ≥ 2]. The traits analysed are heat tolerance milk (HTMYslope), fat (HTFYslope), and protein (HTPYslope) yield. Vertical lines represent the standard errors calculated from two random validation subsets

changes in accuracy of prediction are slightly larger compared to those found when using a single-breed reference set (Fig. 2; "Scenario 1") from BayesRC over BayesR, with values for HTMYslope, HTFYslope and HTPYslope of [0.001, 0.04, and 0.06], [0.05, − 0.06 and − 0.07], [0.09, 0.02, and − 0.006] in the Holstein (N = 1223), Jersey (N = 6338), and crossbred (N = 790) validation sets, respectively. To be more comparable, when considering only a subset of Jersey cows (N = 431) in the validation set where the reference set consisted of a single breed (only Holstein bulls; "Scenario 1'), we found a change in accuracy from BayesRC over BayesR of − 0.02, 0.03,

and − 0.06 for HTMYslope, HTFYslope, and HTPYslope, respectively. Compared to estimates from the "Scenario 1" and "Scenario 2" analyses above, we observed the smallest bias (i.e., values around 1.0) when using the multi-breed reference set in the Holstein validation set. However, in the Jersey validation set, we found extreme bias (> 2.0) for HTPYslope, whereas the bias was smaller for HTFYslope. In the crossbreds, the bias was large for HTMYslope (> 1.5) and HTFYslope (> 1.3), whereas we observed a small bias (values closer to 1.0) for the HTPYslope trait.

Cheruiyot *et al. Genetics Selection Evolution*   *(2022) 54:17*

Page 11 of 18
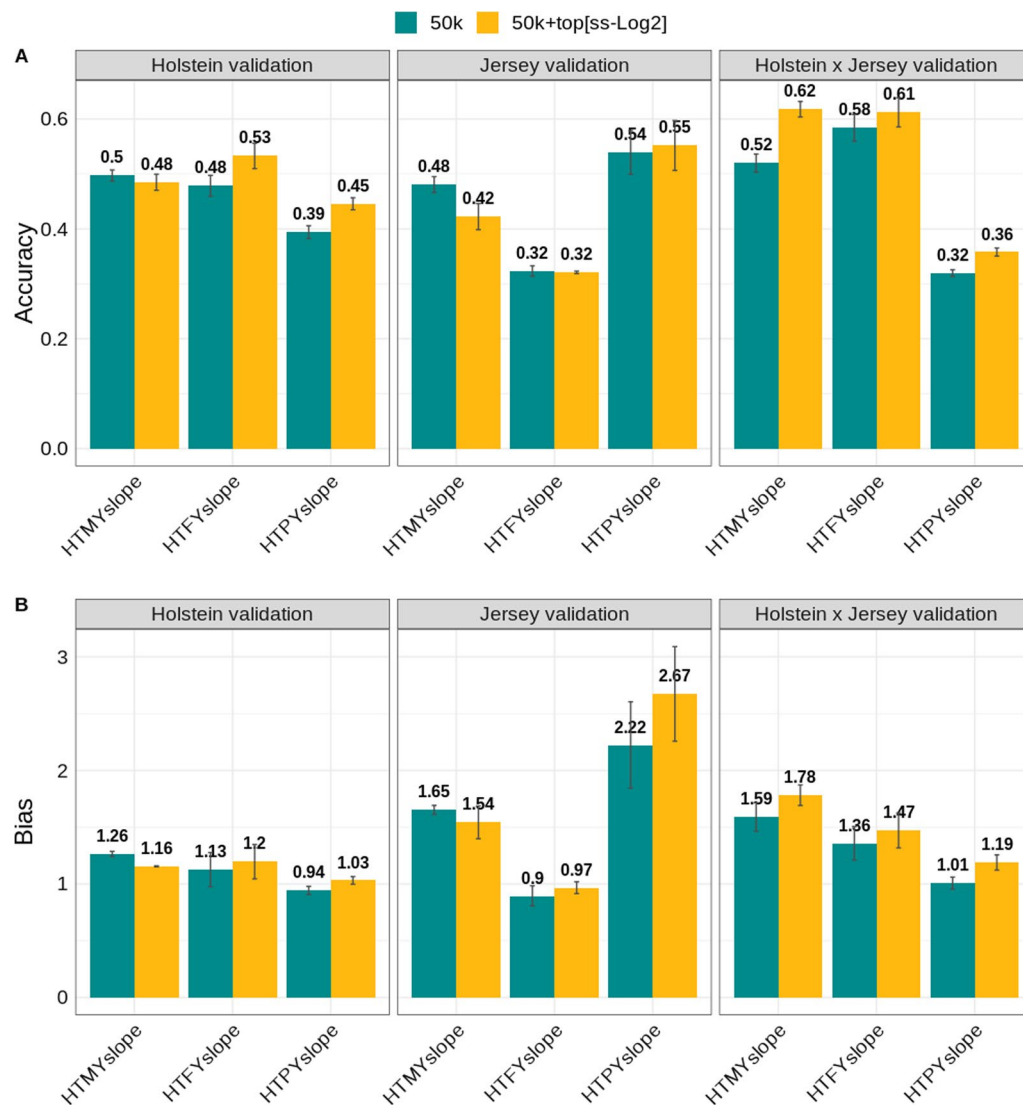
Fig. 4 Accuracy and bias of genomic predictions in Holstein (N = 1223), Jersey (N = 431) and crossbred (N = 790) cows when using the multi-breed reference set (Holstein and Jersey bulls; N = 4175). The selected 'top SNPs' used in the BayesRC were from the Holstein cow discovery set (N = 20,623) based on the single-trait GWAS cut-off of [− log10(p-value) ≥ 2]. The traits analysed are heat tolerance milk (HTMYslope), fat (HTFYslope), and protein (HTPYslope) yield slopes. Vertical lines represent the standard errors calculated from two random validation subsets

BayesR versus BayesRC methods

To test whether allocating selected informative markers to a separate SNP class (see "Methods") in BayesRC can show added benefit in our study, we combined 50k + 'top SNPs' from a single-breed (Holsteins) QTL discovery set and re-calculated GBV using BayesR where all SNPs were allocated to a single class. The total number of 50k + 'top SNPs' used in BayesR and BayesRC was 51,750, 51,894, and 52,168, for HTMYslope, HTFYslope, and HTPYslope traits, respectively (Table 1). The accuracy of prediction (±SE) was slightly higher for two of the three traits from BayesRC [0.49 ± 0.01 (HTMYslope); 0.53 ± 0.02

(HTFYslope); 0.45 ± 0.007 (HTPYslope)] compared to BayesR [0.51 ± 0.01 (HTMYslope); 0.51 ± 0.02 (HTFYslope); 0.44 ± 0.01 (HTPYslope)]. These results suggest that allocating SNPs to different classes in BayesRC, yields marginal benefit in the prediction of heat tolerance traits over BayesR. Moreover, there was little difference in the regression coefficient of predictions (±SE) between BayesRC [1.42 ± 0.002 (HTMYslope); 1.31 ± 0.08 (HTFYslope); 1.21 ± 0.02 (HTPYslope)] and BayesR [1.33 ± 0.005 (HTMYslope); 1.32 ± 0.17 (HTFYslope); 1.21 ± 0.02 (HTPYslope)].

Cheruiyot *et al. Genetics Selection Evolution*   (2022) 54:17

Page 12 of 18

Discussion

In this paper, we present a genomic prediction analysis of heat tolerance traits using a large sample size of over 40,000 cattle, comprising Holstein, Jersey, and crossbred individuals. The primary objective was to investigate if selected sequence variants from a GWAS in Holstein cattle benefits genomic prediction of heat tolerance phenotypes in the same breed (i.e., within-breed prediction). The hypothesis is that the selected variants are linked to causal mutations that underpin the genetic basis of heat tolerance, and thus could enable more accurate and sustained genomic selection for this trait. In addition, we also tested the value of pre-selected variants from Holstein cattle for the genomic prediction of breeds with numerically smaller sample sizes, such as Jersey and crossbreds. Furthermore, we investigated the benefits of using informative markers from a multi-breed (Holstein + Jersey cows) QTL discovery set for genomic prediction of heat tolerance. Overall, our results show that we can increase the accuracy of prediction of heat tolerance by up to 0.10 unit in some scenarios when pre-selected sequence variants are added to the standard-industry 50k SNP panel. However, the change in the accuracy of prediction when using pre-selected sequence variants in BayesRC (i.e., 50k + top SNPs) varied considerably across traits and prediction scenarios.

We used the BayesR and BayesRC methods to test different prediction scenarios. For BayesR, when only 50k SNP data were used, we found a high accuracy of prediction in Holsteins and crossbreds compared to Jerseys. We expected a lower accuracy in Jerseys because we used Holstein bulls as a reference set for genomic predictions (see "Methods", "Scenario 1"). These breeds are genetically divergent and may differ regarding the linkage disequilibrium of variants with causal mutations, they may not share all the same causal variants, or some variant effects may differ between these breeds [35]. As such, when we combined Holstein and Jersey bulls in the reference set (multi-breed reference set; see "Methods", "Scenario 3") and performed analysis using BayesR (without pre-selected 'top SNPs'), we found a substantial improvement in the accuracy of prediction across all traits for Jerseys which is consistent with the multi-breed genomic predictions reported in previous studies e.g., [17, 35].

For the BayesRC model, where 50k + selected 'top SNPs' were fitted in the analysis, we found a consistent increase in the accuracy of prediction across traits when using the 'top SNPs' that were selected from the Holstein discovery set for the prediction of the Holstein validation cows (i.e., within-breed QTL discovery and validation set; see "Methods", "Scenario 1"). Similarly, the use of 'top SNPs' from the Holstein discovery set in crossbred cattle based on BayesRC performed reasonably well, as

expected since our crossbred cows share a similar genetic background with Holsteins (i.e., there were mostly F_1 and backcrosses to Holstein). The gain in accuracy of prediction for Holsteins and crossbreds likely benefited, in part, from a powerful GWAS QTL discovery set (we used a sample size of 20,623 Holstein cows, each having around 15 million imputed sequence variants) and the methodology used for genomic prediction. To date, comparable GWAS have used a sample size of at most 5000 individuals e.g., [5] to search for variants linked to heat tolerance in dairy cattle. We expect an even greater increase in accuracy of prediction in the future with larger sample sizes for GWAS to increase the power of QTL discovery.

However, the genomic predictions in Jerseys performed rather poorly, particularly for HTFYslope and HTPYslope, with accuracies decreasing when the selected 'top SNPs' from the Holstein discovery set were added to the 50k SNP set and used in BayesRC. Given that Holstein and Jersey are genetically divergent breeds, using informative QTL from Holstein in a Jersey validation may have introduced noise in the genomic predictions since the common QTL may not be tracked across these breeds. Also, the drop in accuracy could be due to the non-additive genetic effects (i.e., dominance and epistasis) between Holstein and Jersey. Simulation studies e.g., [36] found that the additive genetic correlations between divergent populations can drop to values as low as 0.45 if reasonably large epistatic interactions exist among loci, which can impact genomic predictions across populations.

However, it is rather unclear why the accuracy of prediction increased for HTMYslope in Jerseys but not for HTFYslope and HTPYslope when using selected 'top SNPs' from Holsteins. One reason could be due to a difference in genetic architecture of these traits. One way to explain this result is to examine the direction of effect for the SNPs between populations. For example, by imposing the GWAS cut-off p-value of 0.001 in both Holstein bulls and Jersey cows, we found that 72% (N = 774) and 71% (N = 524) of the effects for the significant SNPs for HTMYslope and HTFYslope, respectively, were in the same direction. Comparatively, we found a larger proportion of significant SNPs (GWAS p-value < 0.001) having the same direction of effects for HTMYslope (85%; N = 420) and HTFYslope (95%; N = 1240) between Holstein bulls versus Holstein cows (i.e., within-breed comparison) (see Additional file 1: Table S5). Besides the direction of effects for the SNPs between populations, a smaller number of 'top SNPs' for HTMYslope was discovered from the GWAS in Holstein cattle at the relaxed cut off (p < 0.01) (Table 1) compared to HTPYslope and HTFYslope, suggesting that HTMYslope is controlled by relatively few QTLs with large effects compared to

the other traits. This is supported by the strength [based on the magnitude of $-\log10$(p-value)] and the number of significant GWAS signals across the genome based on the Manhattan plot (see Additional file 2: Figure S1). For HTMYslope, we observed four strong peaks on four chromosomes (i.e., BTA5, 6, 14, and 20). This contrasts with the HTFYslope trait for which we observed multiple clear GWAS signals across the genome (see Additional file 2: Figure S1). Moreover, these results are consistent with the evidence that the 'top SNPs' for the HTMYslope trait explained a relatively larger proportion of phenotypic variance compared to the 'top SNPs' for other traits across prediction scenarios.

By comparing the GWAS in Holsteins (N=20,623) and Jersey (N=6338) cows, we found the greatest overlap of top significant SNPs (i.e., top SNPs that were at least within 1-Mb regions in both breeds) for HTMYslope mapping to the genomic regions showing strong signals on BTA5, 14, 20, and 25. This overlap explains, in part, the greater consistency of the increase in accuracy of prediction for HTMYslope than for HTFYslope and HTPYslope. In this context, our findings are in line with those of [32], who reported that only a fraction of the QTL for milk yield segregate across Holstein and Jersey cattle. Overall, these results suggest that breed × SNP interactions exist, meaning that the informative markers obtained from Holstein are of little or no value for the prediction in Jersey. These findings have implications in the genomic prediction of complex traits such as heat tolerance since it is not unusual for one country to incorporate genetic variants discovered in an independent study from another country in their genomic evaluations, e.g., a meta-analysis of SNP effects from multiple countries using SNP-multiple across country evaluation (MACE) [37]. In addition, the results in this study seem to indicate that HTMYslope could be a more reliable indicator trait of heat tolerance and could be given greater weight in the selection index that incorporates heat tolerance, although further work is needed to confirm this. Currently, the Australian dairy industry gives more economic weight to HTPYslope (6.92) than to HTMYslope (−0.10) or HTFYslope (1.79) in the calculation of heat tolerance genomic breeding values based on weights for milk production traits [38, 39].

Previous research studies in cattle e.g., [18, 19] have reported that the mapping of putative causal mutations is more precise when using multi-breed populations in GWAS and have proposed pathways that underpin heat tolerance [23]. In this study, we found some improvement in the predictions, especially in Jersey, when using 'top SNPs' from a discovery set of combined Holstein and Jersey cows (i.e., the multi-breed QTL discovery set). For example, the accuracy of prediction increased by 0.03

for HTFYslope when using 'top SNPs' selected from the multi-breed discovery set in Jersey compared to a drop of 0.06 when the 'top SNPs' from the Holstein QTL discovery set (single breed) was used in BayesRC (Fig. 3). In principle, combining divergent breeds in the QTL discovery set may help to break long-range LD, such that the selected 'top SNPs' are closer to the causal mutations [17] than when a single-breed QTL discovery set is used. For example, the top significant SNP on BTA14 mapped to the upstream region of the *SLC52A2* gene and in an intron of the *HSF1* gene when using the single-breed and multi-breed QTL discovery sets, respectively (see Additional file 2: Figure S9). The *HSF1* gene is associated with thermotolerance in dairy cattle [5, 6, 23]. The smaller number of 'top SNPs' detected in our study with the multi-breed than with the within-breed QTL discovery set is consistent with the work of [19] and is attributed, in part, to the causal variants not all segregating across the Holstein and Jersey breeds.

However, we could still see a decrease in accuracy of prediction (−0.05) for HTPYslope when using the 'top SNPs' from a multi-breed discovery set in Jersey, although not as high as that (−0.08 units) found when using the 'top SNPs' from the single-breed (Holsteins) discovery set. As discussed earlier, one reason for the observed poor prediction for these traits in Jerseys could be partly due to the breed × SNP interactions or non-additive epistatic interactions among loci across breeds. Notably, our multi-breed QTL discovery set was highly dominated by Holstein individuals which explains, in part, the limited gain in accuracy when the selected 'top SNPs' from the multi-breed discovery set were used in the Jerseys. Besides, we used Holstein bulls as a reference set in genomic predictions in Jerseys. Since these breeds are divergent, a better approach to improve predictions in Jerseys would have been to use 'top SNPs' from a multi-breed or within-breed (Jersey) QTL discovery set and a reference set of the same breed (Jersey) or multi-breed set. However, compared to Holstein, the smaller number of Jersey individuals in our study means that it was not possible to split the Jersey dataset to obtain independent subsets with sufficient power for use in the QTL discovery and reference set for genomic predictions. This implies that there may be more room for improvement in accuracy of prediction for Jerseys when more animals with phenotype and genotype data are available in the future.

We compared the added value of informative markers (i.e., 'top SNPs') from single-trait GWAS versus multi-trait meta-GWAS in the genomic predictions. The aim of the meta-analysis of slopes was to increase the power of GWAS and obtain a set of 'top SNPs' with putative pleiotropic effects for heat tolerance phenotypes. There

is a recent trend towards developing custom SNP arrays that include variants with pleiotropic effects across multiple traits [40, 41]. In this study, we found a comparable increase in accuracy of prediction when we used 'top SNPs' from single-trait GWAS or from the meta-analysis, although the accuracy of prediction varied considerably across traits and validation sets used (Fig. 2). Our recent work [23] suggests that heat tolerance traits (milk, fat, and protein slopes) are regulated differently in heat-stressed dairy cows. As such, we think that the relatively lower accuracy realized from using selected 'top SNPs' from the meta-GWAS of slope traits in some scenarios (e.g., HTMYslope across the three validation sets; Fig. 2) could be due to the possible inclusion of non-causal 'top SNPs' in genomic prediction, which arose from combining SNP effects for different heat tolerance phenotypes. However, we observed a smaller drop in accuracy of prediction when using 'top SNPs' from the meta-GWAS compared to 'top SNPs' from the single-trait GWAS in Jerseys from the BayesRC over BayesR (Fig. 2).

In general, we demonstrated an increase in the accuracy of prediction of heat tolerance when informative sequence markers were added to the 50k SNP panel by up to 0.07, 0.05, and 0.10 units in Holstein, Jersey, and crossbred cows in some cases, respectively. Our findings are within the range of those reported for complex traits in cattle e.g., [42] and sheep e.g., [11, 13]. For example, Al Kalaldeh et al. [13] reported an increase in accuracy of prediction by 0.09 units for parasitic resistance in Australian sheep, while de Las Heras-Saldana et al. [42] found an increase of up to 0.06 units for carcass traits in cattle. These results indicate that informative markers can be prioritised, especially for the development of customized SNP arrays [41]. Adding informative variants for heat tolerance to the custom SNP panels as in [41] ensures that higher accuracies of prediction are achieved, which will help to drive genetic gain for this trait. Moreover, we expect that the genetic prediction of this trait would be sustained over generations when informative variants that are closer to the causal mutations are included in the custom SNP panels, as demonstrated by Khansefid et al. [43]. These authors found that using the custom XT_50k SNP panel, which contains prioritised sequence markers, gave a consistent and superior accuracy of prediction (compared to standard SNP panels) in crossbred cows (crossbreds represents "more distant relationships or many generations"). However, caution is needed when using pre-selected sequence variants from Holsteins in the prediction of Jerseys, considering that we found a decrease in accuracy, in most cases, when using the 'top SNPs' from the Holstein discovery set in Jerseys (Fig. 2). This agrees with the simulation work by [14] who

reported that the decrease in accuracy of predictions across-breeds depended on the distance between causal mutations and the markers.

Some studies, e.g., [11, 13], using genomic best linear unbiased prediction (GBLUP), have reported increased accuracy of prediction when fitting pre-selected sequence variants from GWAS together with standard 50k SNPs compared to fitting only standard 50k SNPs, especially when modelling separate GRM for 50k SNPs and pre-selected SNPs. To compare our results (from the BayesR and BayesRC analyses), we used GBLUP on the Holstein validation set (as in "Scenario 1"; see "Methods") to fit pre-selected heat tolerance SNPs ('top SNPs') either as one GRM (i.e., combined set of 50k+top SNPs) or separate GRM (i.e., 2 GRM) for 50k and 'top SNPs'. Although the accuracy of prediction increased when fitting two GRM compared to fitting only one GRM in the GBLUP model, BayesR and BayesRC outperformed GBLUP for the prediction of HTMYslope and HTFYslope but not for that of HTPYslope (see Additional file 1: Table S6) and see Fig. 2. This is comparable to the work of [44] who reported better predictions for milk yield and fat yield traits from Bayesian models than GBLUP models in Danish cattle.

In this study, a sizable proportion of the selected 'top SNPs' for heat tolerance (slopes) overlapped with the selected 'top SNPs' for the intercept traits: 11% (HTMYslope), 17% (HTFYslope) and 21% (HTPYslope) (see Additional file 1: Table S7). Notably, when assuming QTL windows of 1 Mb, more than 90% of the selected 'top SNPs' for heat tolerance traits fell within the same windows with those for intercept traits, which is consistent with the high (− 0.80) phenotypic correlations between these traits. In our recent work [23], we demonstrated, through conditional GWAS analyses, that the top GWAS hits/signal for heat tolerance are also important for milk production traits (i.e., intercept). Therefore, a key question is whether using the selected sequence variants for heat tolerance in genomic evaluations can impact milk production. We investigated this assuming (1) selection is for milk production traits (or Australian Selection Index (ASI), i.e., traits are weighted according to the way Australian farmers are paid for milk, fat, and protein) and (2) selection is for the balanced performance index (BPI) which includes production and functional traits [39]. To see the impact of using pre-selected SNPs in genomic evaluations of Holsteins, the correlation between EBV for heat tolerance (estimated with only 50k SNPs or 50k+selected sequence variant from GWAS— 'top SNPs') and ASI or BPI values were used. The correlation estimates (see Additional file 1: Table S8) suggest that adding the pre-selected SNPs for heat tolerance

Cheruiyot *et al. Genetics Selection Evolution*   (2022) 54:17

Page 15 of 18

(including those that overlapped with intercept traits) to the standard-industry 50k array has little to no impact on the ASI and BPI. However, we observed a favourable correlation between heat tolerance and BPI when pre-selected 'top SNPs' from HTMYslope are added into the 50k array (i.e., 0.06 (50k) versus 0.10 (50k + meta-GWAS top SNPs; see Additional file 1: Table S8). These results are comparable to those of [7], who found that the current selection practices in Australia based on BPI will lead to a negligible decrease in heat tolerance (measured as the rate of decline in yield).

In addition, some reports e.g., [3, 4] have raised concerns that selection for heat tolerance may negatively impact the progress for milk production due to a strong phenotypic correlation of about -0.80 between these traits [23]. Notably, the effects of all the overlapping SNPs for HTMYslope (see Additional file 1: Table S7) were in the same direction with those for MYint, whereas the effects of the overlapping SNPs for HTFYslope and HTPYslope were in opposite directions with their corresponding intercept traits (i.e., FYint and PYint). However, the overlap of top SNPs for MYint and HTMYslope is only 11% (see Additional file 1: Table S7).

Most of our dispersion bias of prediction for heat tolerance traits from BayesR and BayesRC were deflated. However, we also observed inflated predictions, in some cases, especially in Jerseys. In all our Bayesian analyses, we used only bulls in the reference population and only cows in the validation of genomic predictions. As such, the smaller variance of bull phenotypes resulting from averaging daughter slope solutions (see "Methods") explains, in part, the observed bias, especially in the Holstein cow validation set. To test this, we split Holstein cows into reference (older cows) and independent validation (young cows) sets. Consequently, we found that the GBV were inflated, which supports our hypothesis. Nevertheless, the magnitude of bias observed in this study may not be a big issue in the genomic evaluations of heat tolerance, where breeding values are calculated jointly based on bull and cow phenotypes using different weightings according to the amount of information [7, 38].

By comparing the Bayesian (BayesR and BayesRC) versus the GBLUP models (fitting either 1 or 2 GRM as described earlier), we found slightly less biased predictions from the former than the latter models (see Additional file 1: Table S6). This was expected since the Bayesian models simultaneously account for all the markers in the analysis and assume different distributions of SNP effects. However, recent studies in sheep [11] and cattle [43] have reported no difference in dispersion bias between the BayesR or emBayesR versus GBLUP models. We also assessed the dispersion bias of prediction for heat tolerance traits from the GBLUP

and BayesR models using the linear regression (LR) method described by Legarra and Reverter [45]. We did this by first estimating SNP effects from: (1) the full (N = 3323 ♂) Holstein reference set, and (2) a randomly selected reduced (N = 1662 ♂; 50%) reference set. We found no dispersion bias when regressing the GBV in the Holstein validation cows (N = 1223) generated from the full reference bull set on the GBV from the reduced reference bull set (see Additional file 1: Table S9). This suggests that the SNP effects from these reference sets are robust in terms of genomic predictions.

The fact that the dispersion bias of prediction, in most cases, was more pronounced when the selected 'top SNPs' were added to the 50k SNP array and analysed with BayesRC is consistent with some previous studies [20, 21], which is likely due to a phenomenon called the "Beavis effect" [46] that originates from the overestimation of the effect size of the pre-selected variants. The lower bias found when fitting the selected 'top SNPs' from the stringent GWAS cut-off than from the relaxed GWAS cut-off does not agree with the results of Veerkamp et al. [21], who reported a larger bias when markers were strongly pre-selected. Here, we used the Bayesian approach (BayesRC), while Veerkamp et al. [21] applied GBLUP. In our study, fitting separate GRM for the 50k and the selected 'top SNPs' (i.e., two GRM) in the GBLUP models reduced the dispersion bias compared to fitting only one GRM for the 50k + top SNPs (see Additional file 1: Table S6).

In this study, we investigated the utility of pre-selected sequence variants in the genomic prediction of heat tolerance for milk production traits (milk, fat, and protein yield). It is also worthwhile to investigate the added value of prioritised sequence variants for heat tolerance on other traits that are affected by heat stress (e.g., fertility) because there are likely to be benefits from achieving higher systemic heat tolerance across multiple traits. This added value could be significant since the economic selection indices, e.g. for the Australian dairy industry, are formulated to capture different aspects of farm profitability, including production, fertility, health, functional, and type as well as feed efficiency traits [39]. Selecting for thermotolerance would be advantageous if the goal is to simultaneously achieve an optimal level of heat tolerance for multiple traits [24]. Therefore, further studies are needed to investigate the benefits of sequence variants in improving heat tolerance with respect to other traits that are likely to be affected by heat and humidity, such as fertility and health traits.

Conclusions

Our results show that the accuracy of genomic prediction for heat-tolerance milk yield traits (milk, fat, and protein) can be improved when the selected sequence variants linked to heat tolerance are added to the standard 50k SNP panel, with values ranging from 0.01 to 0.10 units depending on the prediction scenario. However, when predicting across breeds, adding informative sequence markers from the Holstein cow discovery set to the standard 50k SNP array (i.e., 50k + top SNPs from GWAS) decreased the accuracy of prediction in Jerseys compared to using only 50k SNP set, especially for the heat tolerance fat and protein yield traits. We observed improved predictions, particularly in the Jersey validation when using pre-selected markers from the multi-breed (Holstein + Jersey cows) SNP discovery set, where the reference population used included Holstein and Jersey bulls (i.e., the multi-breed reference set). Prioritised sequence markers from single-trait GWAS yielded greater accuracy than those from the multi-trait meta-analysis of slope traits. Overall, the results show that sequence variants can be prioritised to improve the accuracy of heat tolerance and has a direct application in the development of custom SNP arrays.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12711-022-00708-8.

Additional file 1: Table S1. Number (included in brackets), breed, and sex of animals with phenotypes and genotypes used for different study objectives. **Table S2.** Additive genetic variance (AG) and genomic heritability (h^2) estimates for heat tolerance (slope) traits based on 50k SNP data for Holstein cows (N = 29,107), Holstein bulls (N = 3323), Jersey cows (N = 6338) and crossbred cows (N = 790). **Table S3.** Additive genetic variance (AG) and genomic heritability (h^2) estimates for milk intercept traits based on 50k SNP data for Holstein cows (N = 29,107), Holstein bulls (N = 3323), Jersey cows (N = 6338) and crossbred cows (N = 790). **Table S4.** Number of informative markers defined as 'top SNPs' selected from single-trait GWAS and multi-trait meta-analyses of intercept traits in the Holstein discovery cow set (N = 20,623). **Table S5.** Number of SNPs with the same effect direction at different GWAS p-value cut-off and their corresponding false discovery rate (FDR) between Holstein bulls (N = 3323) versus Holstein cows (N = 1223) and Holstein bulls (N = 3323) versus Jersey cows (N = 6338). The false discovery rate effect direction (FDR ED) was computed following [39] and the conventional false discovery rate (FDR) was calculated following [47]. **Table S6.** Accuracy and the dispersion bias of predictions (in brackets) from non-linear Bayesian methods (BayesR and BayesRC) versus GBLUP models. **Table S7.** Number of overlapping selected 'top SNPs' between intercept and slope traits and the proportion of SNPs with the sample effect direction detected from the Holstein cow discovery set at a GWAS p-value cut-off of 0.001. **Table S8.** Correlation between estimated breeding values for heat tolerance (based on BayesR and BayesRC) versus Australian selection index and balanced performance index values for the Holstein validation cows (N = 1122) used in this study. **Table S9.** Correlation of GBV between whole (GBV_{S_W}) versus "partial" data (GBV_{S_P}), and the dispersion bias of predictions ($b_{w,p}$) in the Holstein validation cows (N = 1223) using a linear regression method [45] based on GBLUP and BayesR models.

Additional file 2: Figure S1. Manhattan plot of p values from single-trait GWAS results of heat tolerance milk (A), fat (B), protein (C) yield slope traits for the Holstein cow discovery set (N = 20,623). The dashed line represents p-value cut-off = 0.001. **Figure S2.** QQ-plot for heat tolerance milk (HTMYslope), fat (HTFYslope), and protein (HTPYslope) from GWAS of the Holstein cow discovery set (N = 20,623). **Figure S3.** Manhattan plot of p values from single-trait GWAS results of heat tolerance milk (A), fat (B), protein (C) yield slope traits for combined set of Holstein and Jersey cow discovery set (N = 25,766). The dashed line represents p-value cut-off = 0.001. **Figure S4.** QQ-plot for heat tolerance milk (HTMYslope), fat (HTFYslope), and protein (HTPYslope) from GWAS using a combined set of Holsteins + Jersey cows (N = 25,766). **Figure S5.** Minor allele frequency (MAF) distribution of the 50k SNP data and the selected 'top SNPs' (most significant) from the imputed-whole genome sequence variants. **Figure S6.** Accuracy of genomic predictions in Holsteins (A; N = 1223), Jersey (B; N = 6338), and Holstein–Jersey crossbreds (C; N = 790) validation cows for milk (MYint), fat (FYint) and protein (PYint) yield intercept traits from different SNP sets based on the BayesR and BayesRC methods: (a) standard 50k SNP array (50k; colored grey) (b) 50k + top SNPs selected from single-trait GWAS (colored blue) and multi-trait meta-analysis (colored orange) at a less stringent cut-off threshold of [− log10(p-value) ≥ 2] and a more stringent p-value of [− log10(p-value) ≥ 3]. The top SNPs were selected from GWAS of Holstein cows (N = 20,623). Vertical lines represent standard errors calculated from three (Holsteins) and two (Jersey) random validation subsets. **Figure S7.** Bias of genomic predictions in Holsteins (A; N = 1223), Jersey (B; N = 6338), and Holstein–Jersey crossbreds (C; N = 790) validation cows for milk (MYint), fat (FYint), and protein (PYint) yield intercept traits from different SNP sets based on the BayesR and BayesRC methods. **Figure S8.** Bias of genomic predictions in Holsteins (A; N = 1223), Jersey (B; N = 6338), and Holstein-Jersey crossbreds (C; N = 790) validation cows for heat tolerance milk (HTMYslope), fat (HTFYslope), and protein (HTPYslope) yield slope traits from different SNP sets based on the BayesR and BayesRC methods. **Figure S9.** QTL discovery using the single-breed (Holstein cows; left) and the across-breed (Holsteins + Jersey cows; right) discovery set.

Authors' contributions

JEP, HM and IMM, conceived the study, designed, and supervised the analyses. IMM assisted in the preparation and imputation of genotype data. EKC performed association and genomic prediction analyses and wrote the first draft. All authors contributed to the formal data analysis, interpretation of results, and discussions. All authors read and approved the final manuscript.

Availability of data and materials

DataGene (DataGene Ltd., Melbourne, Australia; https://datagene.com.au/) are the custodians of the raw phenotype and genotype data of Australian dairy cows. Research-related requests for access to the data may be accommodated on a case-by-case basis.

Cheruiyot *et al. Genetics Selection Evolution*   (2022) 54:17

Page 17 of 18

Declarations

Ethics approval and consent to participate
The data used in this study were obtained for routine genetic evaluations by DataGene Ltd (Melbourne, Australia) and conform with the Australian dairy industry guidelines for data collection from commercial dairy farms.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]School of Applied Systems Biology, La Trobe University, Bundoora, VIC 3083, Australia. [2]Agriculture Victoria Research, AgriBio, Centre for AgriBiosciences, Bundoora, VIC 3083, Australia. [3]International Livestock Research Institute, Nairobi, Kenya. [4]Scotland's Rural College, Edinburgh, UK.

References

1. Polsky L, von Keyserlingk MA. Invited review: Effects of heat stress on dairy cattle welfare. J Dairy Sci. 2017;100:8645–57.
2. St-Pierre N, Cobanov B, Schnitkey G. Economic losses from heat stress by US livestock industries. J Dairy Sci. 2003;86:E52-77.
3. Carabaño MJ, Ramón M, Díaz C, Molina A, Pérez-Guzmán MD, Serradilla JM. Breeding and genetics symposium: breeding for resilience to heat stress effects in dairy ruminants. A comprehensive review. J Anim Sci. 2017;95:1813–26.
4. Ravagnolo O, Misztal I, Hoogenboom G. Genetic component of heat stress in dairy cattle, development of heat index function. J Dairy Sci. 2000;83:2120–5.
5. Wang T, Chen Y-P, MacLeod IM, Pryce JE, Goddard ME, Hayes BJ. Application of a Bayesian non-linear model hybrid scheme to sequence data for genomic prediction and QTL mapping. BMC Genomics. 2017;18:618.
6. Sigdel A, Abdollahi-Arpanahi R, Aguilar I, Peñagaricano F. Whole genome mapping reveals novel genes and pathways involved in milk production under heat stress in US Holstein cows. Front Genet. 2019;10:928.
7. Nguyen TT, Bowman PJ, Haile-Mariam M, Pryce JE, Hayes BJ. Genomic selection for tolerance to heat stress in Australian dairy cattle. J Dairy Sci. 2016;99:2849–62.
8. Schaeffer L. Strategy for applying genome-wide selection in dairy cattle. J Anim Breed Genet. 2006;123:218–23.
9. Schefers JM, Weigel KA. Genomic selection in dairy cattle: integration of DNA testing into breeding programs. Anim Front. 2012;2:4–9.
10. Calus MP, Bouwman AC, Schrooten C, Veerkamp RF. Efficient genomic prediction based on whole-genome sequence data using split-and-merge Bayesian variable selection. Genet Sel Evol. 2016;48:49.
11. Moghaddar N, Khansefid M, van der Werf JH, Bolormaa S, Duijvesteijn N, Clark SA, et al. Genomic prediction based on selected variants from imputed whole-genome sequence data in Australian sheep populations. Genet Sel Evol. 2019;51:72.
12. Heidaritabar M, Calus MP, Megens HJ, Vereijken A, Groenen MA, Bastiaansen JW. Accuracy of genomic prediction using imputed whole-genome sequence data in white layers. J Anim Breed Genet. 2016;133:167–79.
13. Al Kalaldeh M, Gibson J, Duijvesteijn N, Daetwyler HD, MacLeod I, Moghaddar N, et al. Using imputed whole-genome sequence data to improve the accuracy of genomic prediction for parasite resistance in Australian sheep. Genet Sel Evol. 2019;51:32.
14. van den Berg I, Boichard D, Guldbrandtsen B, Lund MS. Using sequence variants in linkage disequilibrium with causative mutations to improve across-breed prediction in dairy cattle: a simulation study. G3 (Bethesda). 2016;6:2553–61.
15. Brøndum R, Su G, Janss L, Sahana G, Guldbrandtsen B, Boichard D, et al. Quantitative trait loci markers derived from whole genome sequence data increases the reliability of genomic prediction. J Dairy Sci. 2015;98:4107–16.
16. Hayes BJ, Daetwyler HD. 1000 bull genomes project to map simple and complex genetic traits in cattle: applications and outcomes. Annu Rev Anim Biosci. 2019;7:89–102.
17. Kemper KE, Reich CM, Bowman PJ, Vander Jagt CJ, Chamberlain AJ, Mason BA, et al. Improved precision of QTL mapping using a nonlinear Bayesian method in a multi-breed population leads to greater accuracy of across-breed genomic predictions. Genet Sel Evol. 2015;47:29.
18. Sanchez M-P, Govignon-Gion A, Croiseau P, Fritz S, Hozé C, Miranda G, et al. Within-breed and multi-breed GWAS on imputed whole-genome sequence variants reveal candidate mutations affecting milk protein composition in dairy cattle. Genet Sel Evol. 2017;49:68.
19. Raven L-A, Cocks BG, Hayes BJ. Multibreed genome wide association can improve precision of mapping causative variants underlying milk production in dairy cattle. BMC Genomics. 2014;15:62.
20. Raymond B, Bouwman AC, Schrooten C, Houwing-Duistermaat J, Veerkamp RF. Utility of whole-genome sequence data for across-breed genomic prediction. Genet Sel Evol. 2018;50:27.
21. Veerkamp RF, Bouwman AC, Schrooten C, Calus MP. Genomic prediction using preselected DNA variants from a GWAS with whole-genome sequence data in Holstein–Friesian cattle. Genet Sel Evol. 2016;48:95.
22. MacLeod I, Bolormaa S, Schrooten C, Goddard M, Daetwyler H. Pitfalls of pre-selecting subsets of sequence variants for genomic prediction. Proc Assoc Advmt Anim Breed Genet. 2017;22:141–4.
23. Cheruiyot EK, Haile-Mariam M, Cocks BG, MacLeod IM, Xiang R, Pryce JE. New loci and neuronal pathways for resilience to heat stress in cattle. Sci Rep. 2021;11:16619.
24. Cheruiyot EK, Nguyen TTT, Haile-Mariam M, Cocks BG, Abdelsayed M, Pryce JE. Genotype-by-environment (temperature-humidity) interaction of milk production traits in Australian Holstein cattle. J Dairy Sci. 2020;103:2460–76.
25. Hayes B, Carrick M, Bowman P, Goddard M. Genotype× environment interaction for milk production of daughters of Australian dairy sires from test-day records. J Dairy Sci. 2003;86:3736–44.
26. Gilmour A, Gogel B, Cullis B, Welham S, Thompson R. ASReml user guide release 4.2 functional specification. Hemel Hempstead: VSN International Ltd; 2015.
27. Garner J, Douglas M, Williams SO, Wales W, Marett L, Nguyen T, et al. Genomic selection improves heat tolerance in dairy cattle. Sci Rep. 2016;6:34114.
28. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet. 2012;44:955–9.
29. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88:76–82.
30. Bolormaa S, Pryce JE, Reverter A, Zhang Y, Barendse W, Kemper K, et al. A multi-trait, meta-analysis for detecting pleiotropic polymorphisms for stature, fatness and reproduction in beef cattle. PLoS Genet. 2014;10:e1004198.
31. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559–75.
32. Erbe M, Hayes B, Matukumalli L, Goswami S, Bowman P, Reich C, et al. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. J Dairy Sci. 2012;95:4114–29.
33. Garrick DJ, Taylor JF, Fernando RL. Deregressing estimated breeding values and weighting information for genomic regression analyses. Genet Sel Evol. 2009;41:55.
34. MacLeod I, Bowman P, Vander Jagt C, Haile-Mariam M, Kemper K, Chamberlain A, et al. Exploiting biological priors and sequence variants enhances QTL discovery and genomic prediction of complex traits. BMC Genomics. 2016;17:144.
35. Hayes BJ, Bowman PJ, Chamberlain AC, Verbyla K, Goddard ME. Accuracy of genomic breeding values in multi-breed dairy cattle populations. Genet Sel Evol. 2009;41:51.
36. Duenk P, Bijma P, Calus MP, Wientjes YC, van der Werf JH. The impact of non-additive effects on the genetic correlation between populations. G3 (Bethesda). 2020;10:783–95.
37. Goddard M, Jighly A, Benhajali H, Jorjani H, Liu Z. SNPMace—a meta-analysis to estimate SNP effects by combining results from multiple countries. Interbull Bull. 2018;54:13–8.

38. Nguyen TT, Bowman PJ, Haile-Mariam M, Nieuwhof GJ, Hayes BJ, Pryce JE. Short communication: Implementation of a breeding value for heat tolerance in Australian dairy cattle. J Dairy Sci. 2017;100:7362–7.
39. Byrne T, Santos B, Amer P, Martin-Collado D, Pryce J, Axford M. New breeding objectives and selection indices for the Australian dairy industry. J Dairy Sci. 2016;99:8146–67.
40. Xiang R, van den Berg I, MacLeod IM, Daetwyler HD, Goddard ME. Effect direction meta-analysis of GWAS identifies extreme, prevalent and shared pleiotropy in a large mammal. Commun Biol. 2020;3:88.
41. Xiang R, MacLeod IM, Daetwyler HD, de Jong G, O'Connor E, Schrooten C, et al. Genome-wide fine-mapping identifies pleiotropic and functional variants that predict many traits across global cattle populations. Nat Commun. 2021;12:860.
42. de Las Heras-Saldana S, Lopez BI, Moghaddar N, Park W, Park J-E, Chung KY, et al. Use of gene expression and whole-genome sequence information to improve the accuracy of genomic prediction for carcass traits in Hanwoo cattle. Genet Sel Evol. 2020;52:54.
43. Khansefid M, Goddard ME, Haile-Mariam M, Konstantinov KV, Schrooten C, de Jong G, et al. Improving genomic prediction of crossbred and purebred dairy cattle. Front Genet. 2020;11:598580.
44. Ma P, Lund MS, Aamand GP, Su G. Use of a Bayesian model including QTL markers increases prediction reliability when test animals are distant from the reference population. J Dairy Sci. 2019;102:7237–47.
45. Legarra A, Reverter A. Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. Genet Sel Evol. 2018;50:53.
46. Xu S. Theoretical basis of the Beavis effect. Genetics. 2003;165:2259–68.
47. Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc Natl Acad Sci USA. 2003;100:9440–5.

Publisher's Note

Chapter 6: General discussion

6.1. Introduction

Breeding for heat tolerance will continue to be an issue of economic importance over the coming decades because of the effects of rising global temperatures on livestock production, reproduction, health, and wealth-being. In this thesis, several analyses were conducted to enhance our understanding of the genetic aspects of heat tolerance in animals. Ultimately, the new information can be used to minimise the effect of heat stress on livestock performance and welfare, considering the escalating warming climate worldwide.

The research was important for several reasons: 1) the dataset size is much larger than those used in earlier comparable studies in dairy cattle. The dataset used in Chapter 3 (i.e., analysis of $G \times E$) comprised of ~0.5 million cows and ~7 million test-day milk records (milk, fat, and protein yield) while $> 30,000$ and $> 40,000$ cattle with 15 million SNPs were used in Chapter 4 (i.e., GWAS) and Chapter 5 (i.e., genomic predictions), respectively. This dataset is about 10-fold larger than those used so far to study heat tolerance with around 5,000 cattle and less than 1 million SNPs, e.g., the most recent similar study in Australia by Wang et al. (2017); 2) high-producing dairy cattle (e.g., Holstein used in this thesis) are excellent and convenient model to find genes associated with heat tolerance because they are prone to heat stress from the additional metabolic heat of milk production; 3) the study animals (i.e., more specifically, Australian dairy cattle) are suited for understanding the genetic basis for which mammals cope with heat stress because they are predominantly kept on pasture with limited management strategies to alleviate heat stress, in contrast to other countries (e.g., the USA, Canada, and Israel), where dairy cattle are kept indoors and fed total mixed rations (TMR); and 4) the study used powerful statistical tools to discover genetic variants and perform genomic predictions for heat tolerance.

The key findings presented in this thesis are as follows:

- The re-ranking (i.e., $G \times E$; **Chapter 3**) of sires for production traits (milk, fat, protein) between the extreme temperature-humidity index (THI) environments was not substantial since the genetic correlations were > 0.80. Therefore, designing separate breeding programs for colder and warmer climates due to $G \times E$ is not recommended

at this stage based on the general guideline by Robertson (1959) that the genetic correlation higher than 0.80 is not an issue in animal breeding.

- The Australian dairy cattle exhibit substantial genetic variation in heat tolerance. Consequently, dairy sires considered to be resilient to heat stress based on the reaction norm of their estimated breeding value (EBV) across THI environments (**Chapter 3**) were identified, which can allow farmers to make informed breeding decisions best suited for their production environments.

- In the GWAS analysis (**Chapter 4**), several new candidate causal variants and genes (*HSF1, REG3A, CLSTN2, ITPR2, GRIA4, NPFFR2, CALCR, GHR*) that contribute to heat tolerance in animals were documented, which expands our knowledge on the genetic control of this trait.

- Specific candidate causal variants and genes related to the neuronal system and neuroactive ligand-receptor interaction functions (**Chapter 4**) were discovered, which provides new opportunities for understanding and improving thermotolerance in animals.

- In the genomic prediction analyses (**Chapter 5**), it was found that prioritising functional sequence variants from GWAS can increase the genomic prediction accuracy of heat tolerance by up to 10% units in some scenarios, which can be leveraged to drive genetic gain for heat tolerance.

- Overall, the findings in this thesis contribute to our understanding of the biology of heat tolerance and identifies specific genetic variants that can be used to minimise the effects of rising global temperatures on animal production, reproduction, and welfare in animals.

In the following discussion, I will cover the above points and provide a general perspective regarding breeding for thermotolerance in cattle. I will also highlight some limitations of the study and suggest further research considerations towards breeding heat-resilient animals (i.e., animals which can maintain productivity under hot weather).

6.2. **G × E due to heat stress and improving heat tolerance in dairy cattle**

G × E exists when the relative performance of different genotypes changes across different environments. G × E that results in re-ranking of genotypes is a growing issue worldwide as the production environments continue to become increasingly variable (seasonal and temporal), driven mainly by climate changes. For example, Xiong et al. (2020) found that the magnitude of re-ranking in wheat breeding has grown by up to 500% because of climate change and selective breeding for high-yielding lines. In addition, in dairy cattle, a few elite sires are heavily used across a wide range of environments both locally (within countries) and internationally (between countries), mainly facilitated by artificial insemination (AI), meaning G × E could be an issue in the dairy industry.

In the Australian context, G × E due to heat stress is relevant because dairy herds are reared under diverse production environments (spanning an entire continent and being the 6th largest country in the world), which are predominantly pasture-based, meaning that animals are under direct exposure to weather elements (temperature, humidity, solar radiation, etc.). Dairy herds are located mainly in the coastal areas of the country with a large proportion of herds concentrated in the Victoria region in the South-East part of the country (Figure 6-1). Given this dairy landscape, an important question is understanding how big is G × E due to heat stress in Australia? And what are the implications to the industry?

Therefore, in Chapter 3, the magnitude of G × E due to heat stress for production traits in the Australian Holsteins was investigated and breeding implications suggested to the industry. The idea and scope of Chapter 3 are not new, as similar studies have been performed in Australia in earlier years for production and fitness traits (fertility and survival) (Hayes et al., 2003, Haile-Mariam et al., 2008). However, this thesis has benefited from a much larger dataset that has grown over the years, providing better insights into G × E and intriguingly how it has potentially changed over time. For example, the earlier G × E work in Australia (Hayes et al., 2003) used ~870k test-day milk records from ~110k cows, while this thesis used about 7 million records from around 0.5 million cows. The key finding of this thesis is that the genetic correlations for milk records between extreme environments (measured by THI) were high (> 0.80), suggesting that the degree of re-ranking is not substantial to necessitate designing separate breeding programs for hotter environments. However, the extent of G × E appears to be increasing in recent years, which I will discuss in the following section.
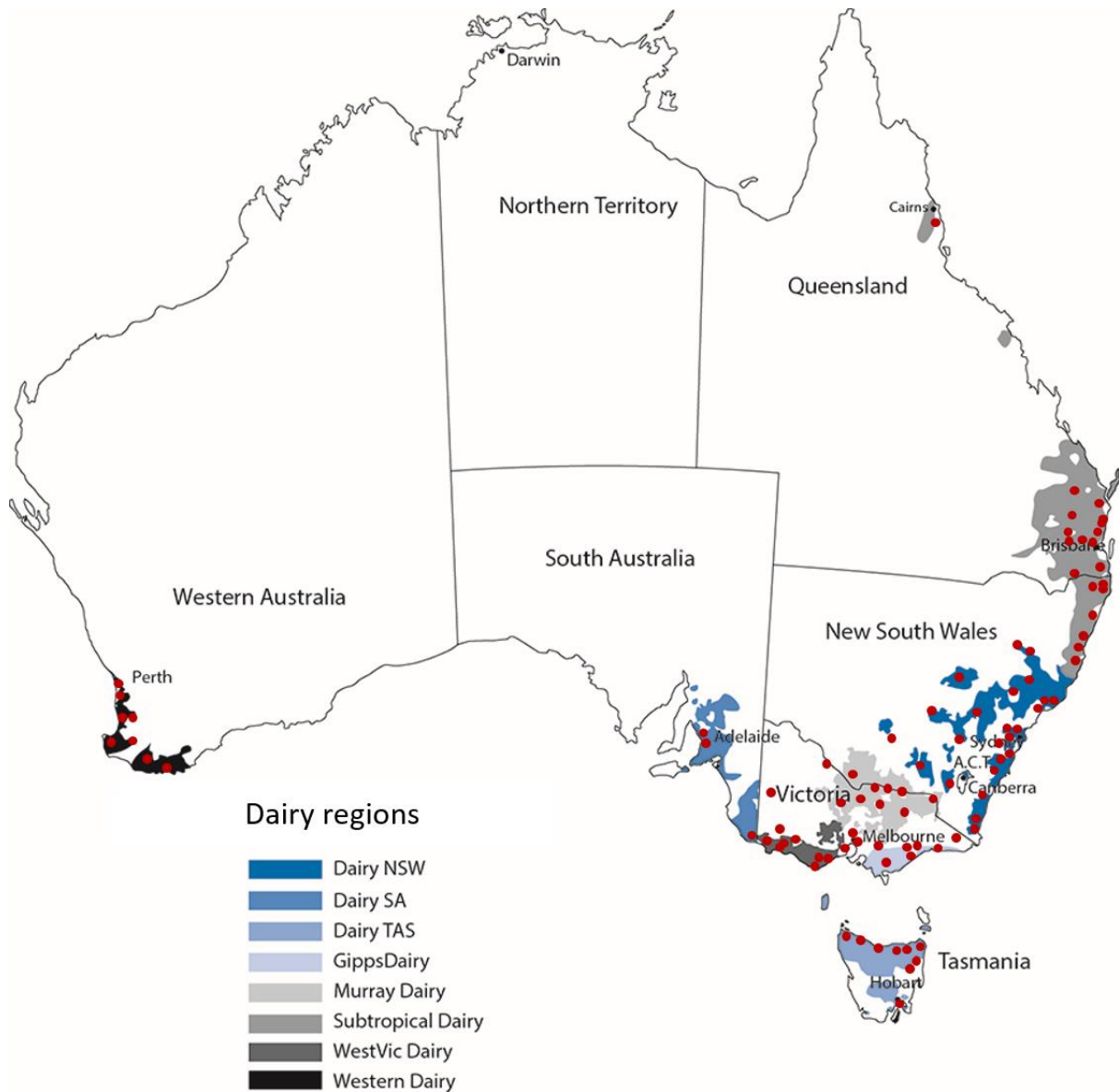
Figure 6-1 Locations of dairy herds (red points) used in this thesis.

While the results from this thesis allay worries regarding the re-ranking of sires at this stage, there are several reasons to consider in further work. As discussed in Chapter 3 of this thesis, most of the data used to estimate $G \times E$ comes from the Victoria region, where dairy cows typically experience moderate heat loads (Nguyen et al., 2016), whereas a small proportion of milk records (approximately 10%) were obtained under high heat load (THI > 70) environments such as Queensland. As such, I expect greater $G \times E$ when larger datasets from hotter climates are available. If indeed substantial $G \times E$ exists, then we will expect

significant re-ranking among the sires, such that it may be necessary to form separate breeding schemes for different production environments, say for colder (e.g., Tasmania) and hotter (e.g., Queensland) climates. Alternatively, G × E could be included in the routine calculation of breeding values, e.g., in Nguyen et al. (2017b), where heat tolerance GEBVs were scaled by dairy regions, given that cows in these regions (8 major dairy regions in Australia) experience varying magnitude of heat stress per year (Nguyen et al., 2016).

6.2.1. **G × E for heat tolerance is increasing in Australia. Should we change the environment to fit the animals or change the animals to fit the changing environments?**

A key finding discussed in Chapter 3 of this thesis is related to the trend of G × E or the re-ranking of sires. Results showed that G × E due to heat stress for production traits has increased over the recent decades, which agrees with earlier reports (Nguyen et al., 2017a). For example, when considering milk yield records, the proportion of sires labelled as heat-sensitive has increased by 7% in recent (2009 – 2017) compared to earlier (2003 – 2008) years. This is a worrying trend showing that dairy cattle are becoming less adapted to the environments, which can be attributed to two main reasons: a) increase in climate changes characterised by frequent and extreme weather, and b) selection emphasize on production traits over the years, considering that heat tolerance is unfavourably associated with milk yield, with correlations of around -0.80 in this thesis. This implies that Australian dairy animals are becoming more specialists, which require more homogeneous and stable environments to maximize performance.

If this increasing G × E trend continues (i.e., dairy animals are becoming less adapted to the environment), then it would be necessary to continuously modify production environments to fit or optimise productivity for animals. For example, several adaptation measures are being sought into pig production systems in many countries worldwide to minimise the effect of heat stress, including designing special housing with good ventilation, reducing stocking density, and modifying or shifting feeding regiments (Schauberger et al., 2019). Similar adaptation measures are being implemented in dairy industries across the world, including Australia, e.g., the Cool Cows program (www.coolcows.com.au), where farmers are provided with the best on-farm practical solutions using shades, sprinklers, or fans to offset heat stress on their farm animals as well as ensuring adequate ventilation for the housed or feedlot cattle, e.g., in North America. A study in New Zealand (Bryant et al., 2006) found that

the performance of Holstein-Friesian genotypes from North America are more sensitive to the level of feeding (i.e., they are specialists) compared to the New Zealand Holsteins or Jerseys (i.e., they are generalist), in part, because cattle populations in these countries are managed differently, as noted above (Bryant et al., 2006). Zwald et al. (2003) investigated $G \times E$ based on milk records from 17 countries, including Australia, and found the heritability of 0.26 and 0.39 for colder and hotter climates, respectively, implying that milk productions in these climates are different traits.

Innovative studies are currently ongoing in some countries aimed at designing special diets for heat-stressed cows that lower core body temperature and allow them to continue feeding under heat stress conditions, e.g., Feeding Cool Cows program in Australia (https://dairyfeedbase.com.au/feeding-cool-cows/). Moreover, complementary permanent genetic solutions have been recommended to better cope with changing environments, such as the genomic selection of heat tolerance in Australia (Nguyen et al., 2017a) or genome editing (Hansen, 2020). Overall, the candidate causal variants discovered in this thesis (discussed in Chapters 4 and 5) can help to breed dairy animals that fit well to warmer climates.

6.2.2. **Should we consider resilient indicators for heat stress? And how are these traits defined?**

Resilience is defined as the ability of an animal to recover quickly following exposure to a perturbation (s) or the ability to be minimally affected by a disturbance (Colditz and Hine, 2016), as illustrated in Figure 6-2. Since $G \times E$ due to heat stress is becoming a growing issue in the agricultural sector worldwide, it is relevant to consider resilience indicators in the breeding goals to help accelerate the genetic improvement for this trait and other functional traits (e.g., health and fertility). This is a growing area of research, particularly in developing indicator traits for resilience that encompasses different aspects of animal wellbeing (weather, pathogens, diseases, social perturbations, etc.) (e.g., Berghof et al., 2019). This is becoming increasingly possible with the availability of big data sets facilitated by the advancement in phenotyping technologies. Research in Australia showed that milk yield in heat-tolerant cows returned to the baseline level after six days following heat challenge in climate chambers compared to 9 days in heat-susceptible cows (Garner et al., 2016).

Depending on the trait, resilient animals are expected to have smaller values on the following components compared to the average population: variance (i.e., closer to zero), autocorrelation (i.e., an indicator of the recovery period as illustrated in Figure 6-2), skewness of deviations, and slopes (Scheffer et al., 2018, Berghof et al., 2019). Poppe et al. (2020) reported low to moderate genetic correlations among three resilience indicators (variance, autocorrelation, skewness) that describe deviations of milk yield from a lactation curve, with heritabilities ranging from 0.01 (skew) to 0.24 (variance). These authors also found a favourable genetic correlation of variance trait with health, longevity, and fertility traits and concluded that it is the best indicator trait to breed for resilience. More recently, Tsartsianidou et al. (2021) developed three resilience phenotypes to account for seasonal changes (autumn, winter, and spring) in milk yield traits for sheep, with heritability ranging from 0.03 to 0.17. Unlike the work of Tsartsianidou et al. (2021) in which modelling spanned low (<10 ^{o}C) and high (>25 ^{o}C) heat environments, this thesis focused on resilience to heat stress (THI > 60, equivalent to the ambient temperature of 22 ^{o}C and relative humidity of 15%; see Figure 2-3). Future work in Australia should also aim at investigating resilience to cold weather conditions, given that a sizable proportion of milk records (16%) in this thesis were collected at low ambient temperatures (< 5 ^{o}C). However, given that the optimum temperature needed to maximise health and productivity in dairy cattle ranges from 5 to around 25 ^{o}C (Kadzere et al., 2002), cold stress is currently not considered as an issue in Australia, unlike other countries such as Siberia (Igoshin et al., 2019).
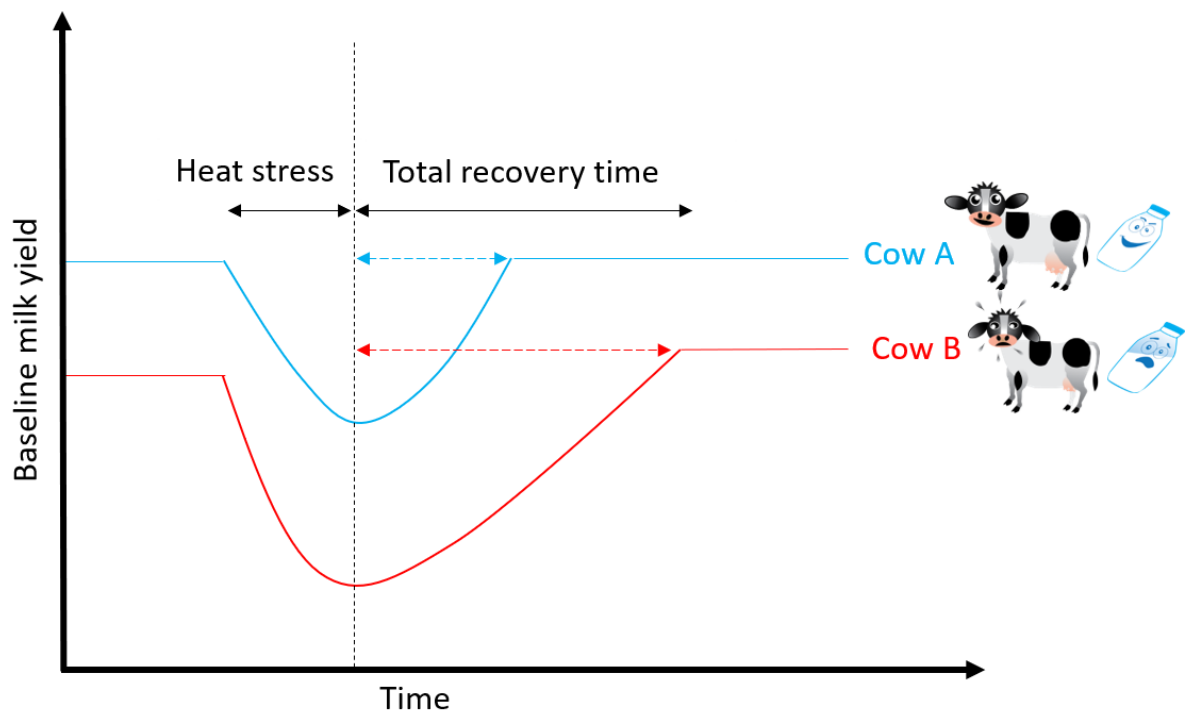
Figure 6-2 Illustration of heat stress and recovery period between two cows: Cow **A** is more resilient to heat than Cow **B** because its milk yield returns to the baseline more quickly after exposure to heat stress or other environmental stressors (e.g., disease and parasites).

6.2.3. **Should we consider other traits beside milk production in the study of heat tolerance?**

In Chapter 3, I focused the $G \times E$ analysis on milk production (milk, fat, and protein) traits for Holsteins dairy cattle using THI as the environmental descriptor for heat stress. I underline the importance of considering other traits given that the profit indices in Australia incorporate many traits of economic importance to farmers such as production, fitness, health as well as efficiency traits (Byrne et al., 2016). I expect greater $G \times E$ for reproduction and fertility traits compared to the estimates for heat tolerance in relation to production traits in this study, in part, due to the genetic architecture of fertility or trait definition. Previous evidence, e.g., Haile-Mariam et al. (2008) found low genetic correlations (< 0.50) between extreme THI environments for fertility traits in Australian dairy cattle. Studies show that fertility traits are sensitive to heat stress. For example, in Australia, Talukder et al. (2015) found that dairy cows calving in summer (under heat stress) are about 43% less likely to conceive than cows calving in autumn (colder season). Also, work at the University of Georgia (Ravagnolo and Misztal,

2002) and in Australia (Morton et al., 2007, Haile-Mariam et al., 2008) have found that the non-return rate (NRR) and conception rate linearly decrease with increasing THI. These traits (conception rates and non-return rate) can be used to investigate the effect of heat stress on reproduction in dairy cattle. Some studies in the USA (e.g., Sigdel et al., 2020) have already tried to uncover the genetic basis underlying cow conception under thermal stress.

6.2.4. **Why should we consider more suitable heat-stress model (s) for the industry?**

As in previous work in Australia (Hayes et al., 2003, Nguyen et al., 2016), the THI (a single value that combines ambient temperature and relative humidity) was used to quantify environmental heat load to the dairy cows throughout this PhD work. Since this index misses relevant pieces of information which contribute to heat stress in animals, such as wind speed and solar radiation, it is possible that the heat-load on animals may have been underestimated in this study and in previous work which used THI in Australia. This is important as dairy cattle in Australia are mostly kept outdoors on pasture; thus, direct solar radiation is likely a key risk factor for these dairy cows. In the G $\times$ E analysis (Chapter 3), I defined a value of THI = 60 as the threshold beyond which milk yield in the study animals begins to decline due to heat stress following the work of (Hayes et al., 2003). I expect this threshold value to change if a more suitable THI model (s) for the Australian dairy system is defined in the future.

While THI variants (including those that incorporate solar radiation and wind speed) have been tested to work well for housed dairy cattle in subtropical environments (Dikmen and Hansen, 2009), information is lacking for cattle kept on pasture. However, the development of heat load indices that are better suited for various production conditions is an ongoing area of research. Lees et al. (2018) developed a dairy heat load index (DHLI) that incorporates panting scores, ambient temperature, relative humidity, solar radiation, and wind speed. These authors transformed the DHLI to have units ranging from 0 to 100; where DHLI = 0 and DHLI = 100, indicates no panting (i.e., no heat stress) and a maximum panting (i.e., high heat stress) within a herd, respectively. However, these researchers found that this index account for about 50% of the panting score variation observed within a herd. In another study, Wang et al. (2018) used data from two sets of experimental cows [one in housed climate chambers at the University of Arizona-USA and the other under outdoor conditions at the University of California-Davis-USA] and developed an index called Equivalent Temperature Index for Cattle (ETIC) which combines the effects ambient temperature, relative humidity, wind speed, and solar radiation

and their interactions. These authors found that the ETIC model outperforms previous THI models in quantifying heat load on dairy animals under outdoor conditions. There are currently no empirical studies to compare the suitability of THI models versus other models that incorporate solar radiation for assessing heat stress from an Australian dairy perspective. This is necessary to recommend suitable heat stress predictive model (s) for the industry, considering that THI values have been and will continue to be relevant for guiding farm management decisions during hot weather and being crucial for research purposes. Besides, it is critical to define a suitable measure of heat stress, which I will discuss in more detail in the subsequent sections.

6.2.5. **Why should we consider heat tolerance in the profit index? And what are the potential challenges?**

The Australian dairy industry currently provides two economic selection indices: (Balanced Performance Index (**BPI**) and Health Weighted Index (**HWI**) in line with farmer preferences for trait improvements (Byrne et al., 2016), neither includes heat tolerance EBVs. The results in this thesis (Chapter 3) show that the dairy cattle in Australia are becoming less fit to the environment, with the proportion of sires labelled as heat-sensitive increasing by 7% in recent (2009 – 2017) compared to earlier (2003 – 2008) years, which supports previous work in Australia that the genetic merit of heat tolerance for Holsteins and Jersey cattle has been declining over the years at a rate of 0.3%/yr (Nguyen et al., 2017a). Also, the warming environment is gradually sweeping across all dairying regions in Australia, with projections showing that heat stress will become a big issue in the coming decades (BOM, 2020). For example, the annual average temperature in the Gippsland dairying region (colder dairy region) is projected to rise by 2.6 °C in 2070, meaning dairy animals will experience more severe hot days and long warmer dry spells (BOM and CSIRO, 2015). Therefore, it is urgent to act now to mitigate the negative consequences of heat stress in the future on productivity, reproduction, and welfare by selecting for heat tolerance alongside other traits instead of stand-alone HT-ABVs, which are currently available in the Australian industry (Nguyen et al., 2017a), or those called 'Augmented BPI' where HT-ABVs are added to the main profit index, separately (i.e., BPI + HT-ABV) (Nguyen et al., 2017b).

However, like other complex traits, e.g., temperament, possible challenges of considering thermoregulation parameters in the profit indices exists. First, it is still a non-trivial

task to define economic weights for use in the selection indices. However, it is possible to derive economic values for resilience to heat by considering the cost associated with lack of heat tolerance, such as production losses and labour costs of managing and treating non-resilient animals (Berghof et al., 2019). The fact that selecting for heat tolerance can contribute positively to animal welfare (Brito et al., 2020) is another motivation to consider this trait in the profit index. Even though it is obvious that resilience has economic merit, the fact that management costs of heat-stressed animals currently vary disproportionately among different regions, say in Gippsland (colder climate) versus Queensland (warmer climate), makes it difficult to standardize economic weights across regions. This has been one of the major issues impeding motivation of incorporating heat tolerance in the Australian main profit index (i.e., BPI). However, the fact that climate warming is projected to become a big issue across Australia (as noted above) and many regions of the world (including the USA) provides a compelling reason to prepare now since multi-trait genetic selection is cumulative and takes an extended number of years to build a heat-tolerant animal population.

DataGene (an organization responsible for genomic evaluation in Australia; https://datagene.com.au/) provides farmers with the *Good Bulls Guide* tool to filter the best bulls that match their breeding goals. To improve heat tolerance and other traits, farmers use a two-step approach to first filter bulls based on BPI and then on heat tolerance. For example, a farmer can look for a bull with a high BPI value, say 350 (meaning this bull is expected to be $350 more profitable than an average bull) and HT-ABV of 105 (meaning this bull is 5% more tolerant to hot/humid conditions than the Australian average score of 100; see Figure 6-4). Recent data from DataGene shows an encouraging upward trend for heat tolerance following the release of HT-ABVs in 2017 (see Figure 2-1). However, the uptake of HT-ABV is perhaps low since this trait is not included in the main Australian index (i.e., BPI), meaning we expect relatively lower genetic progress than if heat tolerance is selected simultaneously with other traits in the BPI. Notably, whether the upward genetic trend of heat tolerance seen in Figure 2-1 is related to the release of breeding HT-ABV in 2017 is early to say at this stage but is nonetheless encouraging to the industry.

For example, using recently published BPI and HT-ABVs data from the DataGene (https://datagene.com.au/; released in August 2021), the correlated response of heat tolerance to selection on the BPI is -0.144 unit decrease of heat tolerance, calculated following formula

described by Nguyen et al. (2016): $R_{HT,BPI} = r_{HT,BPI} * (\sigma_{HT}/\sigma_{BPI}) * 10$, where $R_{HT,BPI} =$ response to selection; $r_{HT,BPI} =$ correlation between HT-ABV and BPI; $(\sigma_{HT}/\sigma_{BPI}) =$ the ratio of genetic standard deviations; the formula is scaled by 10, equivalent to approximately 1 year of genetic improvement (Nguyen et al., 2016). We can expect larger value (i.e., > -0.144 unit) if HT-ABV is included in the BPI. If selection is on the Australian Selection Index [i.e., an index used to rank animals based on the most profitable combination of milk production EBVs according to the way Australian farmers are paid for milk, fat, and protein yield], then the decline is higher at -0.441 units compared to selection based on BPI at -0.144 units (as indicated above). This was expected, in part, because of two possible reasons: a) the correlation between [HT-ABV versus ASI] is stronger (-0.72) than [HT-ABV versus BPI; -0.50], and b) BPI includes fertility and health traits that are positively correlated with HT-ABV (Nguyen et al. 2016).

6.2.6. Can we define a more suitable trait for heat tolerance? And can we use this trait as the breeding objective?

The milk decline traits (i.e., slopes) used in this thesis and other previous work (Nguyen et al., 2016) as proxy for heat tolerance has been criticised for several main reasons: 1) it does not fully capture the effect of heat stress in cattle and 2) it is unfavourably correlated with milk volume with estimates in this thesis of around -0.80, and 3) milk production traits have been already included in the Australian economic indices (Byrne et al., 2016), implying that heat tolerance may have been partly captured in these indices. While this trait (milk decay under heat stress) definition is economically more understandable to farmers and somewhat straightforward to factor in the profit indices, e.g., as described by Nguyen et al. (2017a), it is perhaps not the most suitable choice as breeding objective for heat tolerance, considering the above reasons.

Alternatively, several other indicators of thermoregulation in animals could be used as the breeding objective, such as those related to core body temperature (e.g., rectal, ruminal, or vaginal temperature), heat production (e.g., feed consumption and fermentation), or latent/sensible heat loss (e.g., skin, or cutaneous temperature, sweating rate, respiratory rate, etc.). Among these heat stress indicators, measures of core body temperature, e.g., rectal temperature, are often considered as 'gold standard for heat tolerance with the heritability estimate of 0.17 ± 0.13 (Dikmen et al., 2012) in Holstein cattle. Although this trait could be

considered as a breeding objective for heat tolerance, it is questionable by the fact that it is ambiguous as to what it means, for example, to breed for lower rectal temperature (RT) given that the core body temperature in livestock and other mammals are typically tightly controlled within a very narrow range (Finch, 1986, Gourdine et al., 2017). In addition, the genetic correlations between RT and milk production traits are positive (Dikmen et al., 2012, Luo et al., 2021), meaning that selection for lower RT could undesirably impact milk yield. By taking these factors into account, perhaps a more attractive alternative is to consider traits that capture heat dissipation efficiency, such that high-ranked heat-tolerant animals are defined as those with the superior genetic ability to remove metabolic heat from the core body into the environment. In addition, this will allow continued genetic progress of milk production since high-producing animals can efficiently dispel heat from their core bodies that could have otherwise accumulated due to increased rumen fermentation. This idea is not new and was proposed recently by Brito et al. (2020) as one way to breed for heat tolerance and welfare in livestock.

There is evidence that between-breed differences in heat dissipation exist, e.g., Finch (1986), who observed that *Bos taurus* beef cattle are more superior at dissipating heat than *Bos indicus* beef cattle. Also, within-breed difference in heat transfer exists, e.g., Garner et al. (2016), who found that the mean skin temperatures of heat-tolerant Holstein cows are significantly higher than those for heat-susceptible Holstein cows. Srikanth et al. (2017) subjected 10 Holstein calves to heat stress in the experimental chambers and monitored their rectal and skin temperature during the day (Figure 6-3). These authors found that the rectal temperature for animal ID33 (i.e., one of the study calves) increased rapidly to 41.3 ^{o}C with relatively lower skin temperature, suggesting that it has poor heat dissipation ability compared to other calves (Figure 6-3). Given that calves were investigated in this study, I expect greater genetic variation in thermoregulation for lactating cows due to the higher metabolic heat associated with lactation. In other studies, Dikmen et al. (2008) found that slick-haired cattle can control their core body temperature via superior thermoregulatory mechanisms compared to non-slick Holsteins with a relatively lower drop in their milk yield. Given that even a marginal rise in body temperature has serious negative consequences on cell integrity and metabolic functions, selecting animals that can tightly constrain their body temperature through superior heat dissipation ability can yield the greatest advantage on productivity (Finch, 1986).
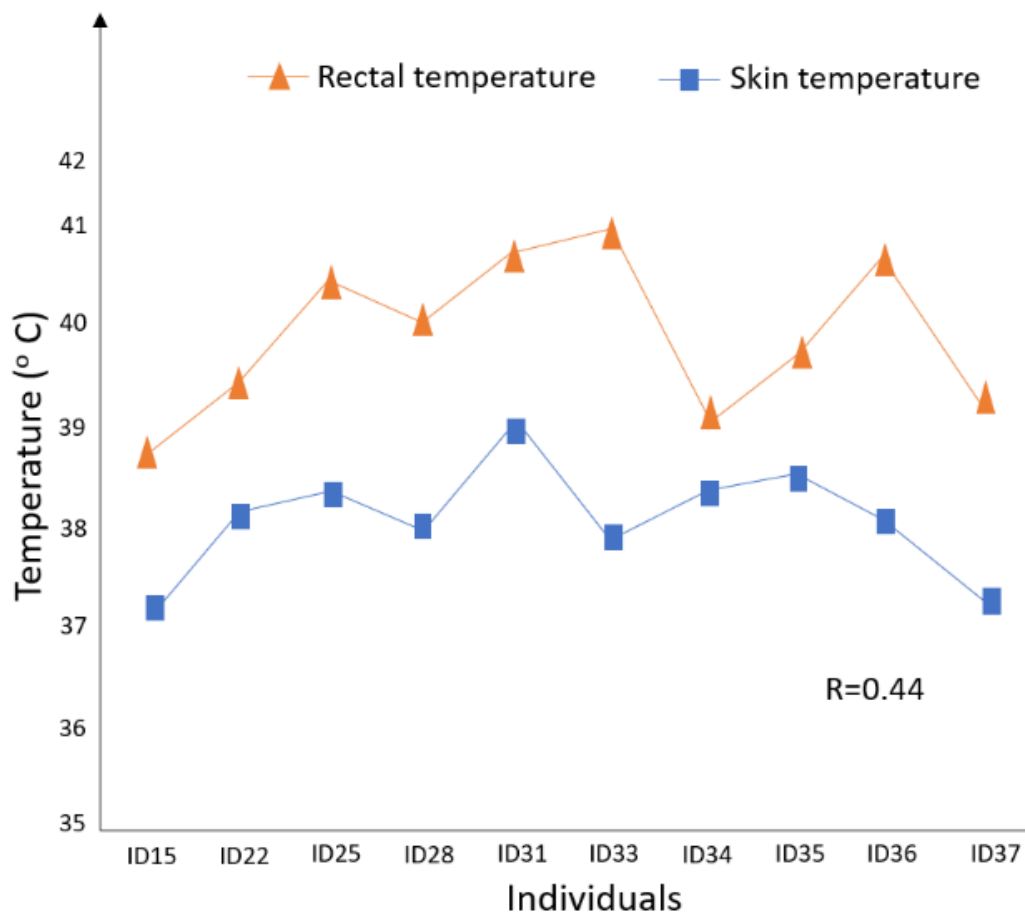
Figure 6-3 Skin and rectal temperature measures for 10 Holstein calves and correlation (R=0.44) after heat stress (19:02 hours); adapted from Srikanth et al. (2017).

Thermal circulation index (TCI) described by Curtis (1983) quantifies the transfer of heat from the core of the body to the skin surface and then to the environment under steady-state thermal conditions computed as follows: $TCI = (T_s - T_a)/(RT - T_s)$, where T_s is the average skin temperature; RT is the core or rectal temperature, and T_a is the ambient temperature. This TCI trait could be used as a proxy for heat dissipation efficiency through the skin, which accounts for up to 85% of total heat loss (Maia and Loureiro, 2005). Besides T_a (ambient temperature), it may be more suitable to calculate TCI based on ETI (Wang et al., 2018), which (as discussed earlier) combines key weather elements, including ambient temperature, relative humidity, solar radiation, and wind speed and is, perhaps, a better predictor (compared to THI) of the environmental heat load on the animals, especially for the pasture-based system in Australia. Looking at Figure 6-3, calf ID33 and ID36 are expected to have large TCI values

because of the relatively high RT and lower T_s, meaning that they are more susceptible to heat stress and may not be an ideal candidate for hotter environments. Using the TCI trait as the selection target for heat tolerance and welfare could be attractive to farmers because it is likely to drive genetic improvement for this trait, with a possible small impact on production traits, although future rigorous studies are needed to support this hypothesis.

As it is, the TCI trait seems to describe the biology of heat tolerance (i.e., heat dissipation efficiency) which, according to the producer perspective, may not be an economically appealing target. However, given producers seeks to maintain production and reproduction under heat stress, the economic value for this trait (TCI) can be defined as the loss of milk production associated with the unit increase in core body temperature of the animal, such that animals with large TCI values are expected to experience greater decline in milk production or fertility under heat stress conditions. This implies that future work is needed to estimate the rate of milk decline or fertility associated with an increase in core body temperature or TCI values related to failure in efficient thermoregulation. Nevertheless, I still expect lingering doubts on the suitability of TCI as the target breeding objective for heat tolerance over, say milk decays (i.e., milk slopes used in this thesis), in part, because of the high cost of obtaining RT and T_s measurements at this stage. However, as high-throughput sensor technologies continue to improve (Koltes et al., 2018), which can facilitate collection of large thermotolerance phenotypes at potentially lower cost, I expect the TCI trait to be a relevant target for heat tolerance in the future because it is likely to best describe heat-tolerance and can allow genetic progress to be made in productivity (Finch, 1986).

6.2.7. **How can we select for heat tolerance – selection criteria?**

The accuracy of the genomic prediction relies on several factors including a) the heritability of a trait, b) effective population size, c) size of the reference population, d) marker density, and e) the architecture of the trait (Daetwyler et al., 2008, Meuwissen, 2009, Hayes et al., 2010, Gonzalez-Recio et al., 2014). With low heritability estimates for thermoregulation traits, e.g., the rectal temperature ($h^2 = 0.17$ in Holsteins; Dikmen et al. (2012)) – a component needed to calculate TCI (described above) – over 20,000 animals of a female reference set would be required to achieve moderate genomic reliabilities of ~0.40 (Gonzalez-Recio et al., 2014). While we have seen a rapid evolution in high-throughput sensor technologies in recent years (Koltes et al., 2018), it is still costly and logistically challenging to build a sufficient

reference population with phenotypic measurements (i.e., rectal and skin temperature) needed to compute animal TCI (i.e., breeding of objective trait as proposed above) values for genomic evaluations.

However, as phenotyping technologies continue to improve and are validated for commercial use (Koltes et al., 2018), an alternative approach at this stage is to have a dedicated genomic reference population with high-quality phenotypes (i.e., RT and T_s). Then, other predictor traits which can be collected cost-effectively in large quantities are used as the selection criteria for heat tolerance. For example, clinical mastitis is a hard-to-measure trait where somatic cell count (SCC) and other traits (e.g., udder depth) have been used as indicator traits for the selection of this economically important trait since large datasets for SCC can be obtained from routine milk recordings (Martin et al., 2018). Considering that TCI is a novel trait proposed in this thesis as the breeding objective for heat tolerance, a lot of future work is required, such as understanding the trait, collection of enough data, calculating genetic parameters (e.g., heritability) for this trait, and correlations with other economic traits, (production, fertility, and health trait, etc.).

In this regard, opportunities are emerging to obtain potentially inexpensive large phenotypic measures from Mid-infrared (MIR) predicted milk biomarkers as the proxy for heat tolerance. Hammami et al. (2015) found that the MIR-predicted traits (e.g., C18:1 *cis*-9) decline with increasing THI in heat-stressed cows. van den Berg et al. (2021) found high genetic correlations (values close to 1.0) between measured serum urea and milk MIR-predicted serum urea. This high genetic correlation means that MIR predicted urea can be used to improve the accuracy of genomic prediction of serum urea. Heat stress increases milk urea in dairy cows, possibly due to elevated deamination of amino acids, increased metabolism of muscle tissues, or reduced feed intake that often occurs under hot weather (Cowley et al., 2015, Gao et al., 2017). Therefore, using phenotypes such as MIR-predicted C18:1 *cis*-9 [which was found to be most sensitive to heat stress by Hammami et al. (2015)] could be used as alternative and inexpensive selection criteria for heat tolerance or used directly to quantify heat tolerance. Moreover, MIR profiles can be used as suitable management biomarkers for predicting animals that are likely to suffer most from heat stress under hot conditions (Hammami et al., 2015). For example, recent work in Australia by Ho and Pryce (2020) found substantial improvement in

predicting cows that are likely to conceive following first insemination when including MIR data in the prediction models.

Also, milk yield decline (proxy for heat tolerance trait used in this thesis) can be used as selection criteria trait for heat tolerance once the breeding objective (i.e., TCI proposed above) has been measured. Results in this thesis (Chapter 4) suggest that heat tolerance milk traits (i.e., milk, fat, and protein yield slopes) are regulated somewhat differently in heat-stressed cows. Comparatively, the decline (per day) in milk, fat, and protein yield ($\pm$ standard errors) for Holstein cows was -0.01 $\pm$ 0.0003 Kg, -0.0007 $\pm$ 0.01 Kg, and -0.0013 $\pm$ 0.008 Kg per unit increase in THI above 60, respectively. These estimates are similar to those reported by Bryant et al. (2007) for Holstein cows in New Zealand, with milk solids (fat and protein yield) reduction (per day) greater than 0.001 Kg/unit increase in THI (averaged for 3 days) exceeding 68. If we assume an increase of THI by 15 units per day in Australia (i.e., THI = 60 to THI = 75), then we expect a yield reduction (per day) of 0.58% (milk), 1.07% (fat), and 2.29% (protein), considering that the corresponding estimates for daily average milk, fat, and protein in this thesis were 25.85 (8.19) Kg, 0.98 (0.30) Kg, and 0.85 (0.26), respectively (values in brackets are the standard deviations for each trait). These percentage estimates suggest that protein yield is more sensitive to heat than milk or fat yield likely, most likely due to heavy selection pressure on this trait over the years.

Currently, the heat-tolerance breeding value (HT-ABV; released to the industry in 2017, Nguyen et al. (2017a)) combines milk, fat and, protein yield slope traits using economic weights of -0.10, 1.79, and 6.92, respectively – equivalent to the weights for milk production traits used in the Australian selection indices (Byrne et al., 2016, Nguyen et al., 2017a). The economic weights should be revised to facilitate more genetic gains to be realised in both heat tolerance and production traits. Future work could focus on estimating the correlation between milk yield slope traits versus other "gold standard" heat tolerance traits, e.g., core body temperature. So that milk yield slope traits (milk, fat, and protein) that correlate strongly with thermoregulation traits can be given more weight in the heat tolerance selection index (i.e., HT-ABV). Furthermore, studies could explore the correlation between milk decay traits versus MIR biomarkers (e.g., C18:1 *cis*-9) that are potentially better predictors of heat stress (Hammami et al., 2015) to inform the appropriate weightings for HT-ABV currently used in the Australian industry.

6.3. **Leveraging whole-genome sequence data to understand the biology of thermal stress and testing the benefits of using prioritised variants to improve heat tolerance**

Characterising causal variants and pathways underpinning the genetic basis for heat tolerance in cattle is at the infancy stage, but it is gaining increased attention due to global warming, as demonstrated by a recent surge in published studies (Hayes et al., 2009, Dikmen et al., 2013, Macciotta et al., 2017, Sigdel et al., 2019) (reviewed in Chapter 2; section 2.9). In addition, big genomic datasets are increasingly becoming available thanks to the international consortiums, such as the 1000 Bull Genome Project, which currently hosts whole-genome sequence data for cattle from over 40 collaborators around world (Hayes and Daetwyler, 2019). Results in Chapter 3 (i.e., G × E interaction; discussed above) clearly showed a substantial genetic variation for heat tolerance in the Australian dairy cattle, which motivated an association study (GWAS) using imputed whole-genome sequence data aimed at understanding genetic variants that make dairy animals vary in thermotolerance. Consequently, I covered the key findings from the association analyses in Chapter 4 of this thesis.

The heat-tolerance phenotypes used in the GWAS (discussed in Chapter 4) were derived from milk records (milk, fat, and proteins), defined as the rate of decline in milk yield traits (i.e., slopes) with increase in THI. Other studies in the USA (e.g., Sigdel et al., 2019) have also used similar traits in searching for heat tolerance variants in cattle. However, using slope traits from milk records presents a challenging task in disentangling genes for heat tolerance and milk production because these traits are highly correlated (estimates of about -0.80), implying that they are largely regulated by the same genes. Indeed, I found that the strong signals (i.e., QTLs) for heat tolerance overlap with well-known genes for milk production, e.g., DGAT1, with candidate variants showing opposing effect direction. This means that selecting for heat tolerance genes could negatively impact milk production. I did a follow-up post-GWAS conditional analyses (described in Chapter 4), which confirmed that the overlapping QTLs are important for both heat tolerance and milk production. These findings have implications for breeding, considering that farmers desire to improve thermotolerance while at the same time trying to increase milk production – a key driver for dairy profitability.

Nonetheless, genomic selection tools offer promising avenues to improve both heat tolerance and milk production. This is evidenced by fertility traits, which deteriorated considerably over the past decades, in part, because of the selection for high milk production,

especially from the second half of the 20th century (Lucy, 2001). However, we have seen a dramatic positive change in both production and fertility traits in dairy cattle within a short period following the introduction of genomic selection tools around 2009 with, for example, sire generation interval dropping from 7 to 2.5 years and the rate of genetic gain for yield traits almost doubling (García-Ruiz et al., 2016).

6.3.1. Can we use the genes identified in this thesis to select for heat tolerance without significantly impacting the genetic progress in milk production?

The published literature is replete showing that selection for heat tolerance could inevitably impact genetic progress in milk production (e.g., Carabaño et al., 2019). The results in Chapter 4 provide several biological insights of thermal stress that can be leveraged to minimise heat stress while maintaining productivity in high-yielding dairy cattle. For example, metabolic adaptations are key biological mechanisms for heat tolerance. In high-yielding dairy cattle, e.g., Holsteins breeds, elevated metabolic-heat is a major proteotoxic stress that impacts milk production with reductions of up to 40% (Kadzere et al., 2002, West, 2003). As such, understanding the genetic basis underlying metabolic adaptations could allow breeding for heat tolerance while maintaining high productivity. Several promising candidate genes for heat tolerance identified in Chapter 4 (*ACLY, PDHA2, MDH1, SUCLG2, PCK1*) are associated with the citrate (Krebs) cycle, which is a crucial metabolic hub in the oxidation of carbohydrates and fatty acids (Belhadj Slimen et al., 2016). Heat stress disturbs the metabolism of carbohydrates – a major source of energy for maintenance and production in animals, for two main reasons: 1) reduced dry matter intake b) altered post-absorptive metabolism (Wheelock et al., 2010). Although studies are still conflicting, some reports show that fatty acids are not mobilized under heat stress, as evidenced by unaltered basal NEFA (associated with negative energy balance) in heat-stressed cows (Rhoads et al., 2010, Wheelock et al., 2010). This is partly related to the inability of heat-stressed cows to employ the 'glucose sparing' effect, such that the adipose tissue is not mobilized to generate NEFA which, in part, explains drastic milk decline (Rhoads et al., 2010, Baumgard and Rhoads Jr, 2013). In contrast, the 'glucose sparing' effect is often enlisted in early lactation (i.e., a stage of high energy demand resulting in negative energy balance) to maintain milk production in dairy cows (Baumgard and Rhoads Jr, 2013).

A fundamental question is whether selecting for the above candidate genes (linked with metabolism) would contribute to heat tolerance without substantially reducing milk production. This remains an open question because it is not clearly understood how these genes are regulated under acute and chronic heat stress conditions to impact animal productivity. Compared to acute stress, the genetic aspects of chronic heat stress are still poorly understood. Recent research in camel somatic cells suggests that acute and chronic heat stress are somewhat controlled differently with the former associated with the increased heat shock proteins and DNA repair enzymes, while the latter heat-response mechanism is linked to altered cell architecture, proteomics, and cytoskeletal proteins (Saadeldin et al., 2020). Therefore, I think that the candidate genes and pathways in this thesis provide interesting insights into the genetic basis for chronic (long time heat exposure) or recurrent heat stress – a characteristic of Australian seasonal summers. Future work is needed to confirm this. Notably, the candidate causal genes in this thesis (Chapter 4) did not overlap with those reported in comparable work in Australia (Hayes et al., 2009, Wang et al., 2017) and in the USA (Sigdel et al., 2019), most likely because they used smaller datasets (typically $< 5,000$ animals while around 30,000 was used in this thesis) and low-resolution SNP sets (50k or 600 SNP set versus 15 million SNPs used in this thesis). Importantly, the candidate causal variants discovered in this thesis were confirmed to be relevant for heat tolerance in an independent validation set and via genomic prediction (as discussed in Chapter 5). Moreover, the conditional analyses in this thesis (Chapter 4) confirmed that the top GWAS hits/signals are in fact associated with the biology of thermal stress in dairy cattle.

Under hot conditions, heat-stressed dairy cows employ various behavioural strategies to regulate internal metabolic heat-production such as lowering feed intake, spending less time grazing and more time standing, resting in shade, and drinking more water (Kadzere et al., 2002). While these behavioural and physiological adjustments implies that milk decline in heat-stressed dairy cows is inevitable, the genetic tools provide opportunities to minimise such losses to the lowest possible level. For example, finding alternative ways, e.g., those that help dairy cows minimize the accumulation of toxic reactive oxygen species (ROS) and the onset of metabolic heat, especially in warmer months, and in an energy-efficient way is critical to breeding heat-tolerant yet high-yielding animals. Heat stress causes overproduction of ROS, which can cause oxidative stress and subsequent apoptosis or cell death (Belhadj Slimen et al.,

2016). There is evidence that efficient scavenging of toxic ROS from cellular metabolism during various environmental stresses is a signature of increased adaptation in plants (Sharma et al., 2012). In this regard, exploring the appropriate nutrition (e.g., Feeding Cool Cows program in Australia; covered earlier) to suppress core body temperature or those that favour lower heat increment during hot weather is gaining increased attention. However, genetic aspects to achieve these goals (lower ROS, core body temperature, or heat increment) in heat-stressed animals remains largely unexplored. A recent study (Atta et al., 2020) suggests that supplementing heat-stressed rats with Pycnogenol (generic French pine bark extract) stimulates genes related to antioxidant activity allowing them to reverse heat-induced ROS damage in testicular and brain tissues. It will be illuminating to understand the role of metabolic-related genes identified in this study to breed more productive and heat-tolerant animals. Moreover, it may help in devising novel ways for improving the nutritional management of heat-stressed dairy cows.

6.3.2. Can we identify other genes that are not strongly linked to milk yield to breed for heat tolerance?

It is also worthwhile exploring other genetic features that allow dairy cows to better regulate or dissipate heat efficiently under hot weather, such as those related to morphology (coat colour, coat length, hair thickness, etc.), physiological (e.g., cardiovascular, and respiration system), and cellular (cell repair, fluidity, stability, etc.) functions. For example, research has shown that Holstein cows with SLICK hair coats are more efficient at regulating core body temperature with the lower decline in milk yield than wild-type cows under heat stress conditions (Dikmen et al., 2008). Causal mutations for the SLICK phenotype have been mapped to PRLR (prolactin receptor) gene in chromosome 20 at ~39 Mb in the Senepol cattle (Olson et al., 2003, Littlejohn et al., 2014). As expected, I did not detect any significant SNP ($p < 1E-05$) within the PRLR gene across all GWAS analyses because the study population (Holsteins) lacks the causal mutation for the SLICK hair phenotype. However, future efforts should aim at introgressing the causal mutation for this gene to the study population to better cope with heat stress, such as those implemented for Holstein cattle in Puerto Rico, and the USA (Hansen, 2020).

So far, Holstein cows that are primarily heterozygous for SLICK genotype have been confirmed to possess superior thermotolerance ability over wild-type cows (Dikmen et al.,

2008, Dikmen et al., 2014). However, a crossbreeding program was initiated in the New Zealand to build homozygous SLICK bulls with up to 75% New Zealand dairy genetic background (Davis et al., 2017). On top of these efforts, it would be interesting to see if additional benefits could be achieved when breeding for homozygous SLICK bulls + high HT-ABV (i.e., HT-ABV + SLICK genotype). Notably, the heat tolerance capacity of Australian dairy cattle is ranked based on genomic estimated breeding values (HT-ABV), such that animals ranked high for HT-ABV are considered more tolerant to heat than the average population (Nguyen et al., 2017b; Figure 6-4). Top ranking cows for HT-ABV were found to maintain lower core body temperature and experience less milk yield decline under heat stress, which is thought to be related to their efficient heat dissipation mechanisms and energy metabolism (Garner et al., 2016). Overall, we can except more benefits from a breeding program that aims to build dairy bulls with high HT-ABV values and carry homozygous SLICK genotype. However, extensive performance data is needed to confirm this hypothesis.
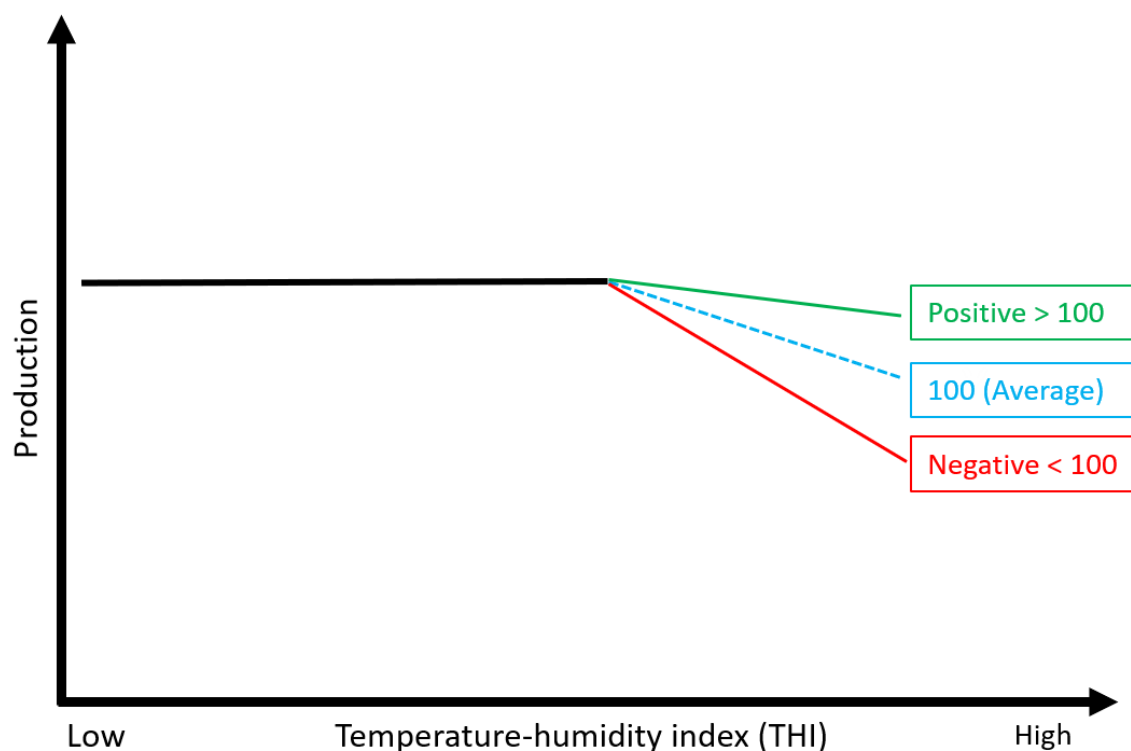
Figure 6-4 Representation of heat tolerance breeding values (HT GEBVs) that were released to the Australian dairy industry in 2017 (Nguyen et al., 2017a). The daughters of a bull with HT GEBV above

average (positive) are expected to be more tolerant to heat than the daughters of an average bull and vis versa for bulls with below average HT GEBVs.

6.3.3. Genes related to nervous system are important for heat tolerance. How can we leverage these findings to benefit farmers?

Another highlight in this thesis covered in Chapter 4 is identifying candidate causal mutations underlying heat tolerance that implicate nervous system mechanisms (the neuroactive ligand-receptor interaction and glutamatergic synapse gene categories were overrepresented). The fact that the genes related to the nervous system were strongly implicated for determining heat tolerance in this thesis is new, and consistent with this pathway playing the most critical function of controlling body temperature by connecting the internal and external environment of animals (Nakamura and Morrison, 2008). Interestingly, the most promising genes in the neuronal pathways could be relevant for feeding and metabolic homeostasis in cattle during thermal stress, which are novel findings that may help manage and breed heat-tolerant animals. The genes identified are linked to neuroendocrine functions, involved in a cascade of hormonal responses such as secretion of growth hormone, insulin, serotonin, prolactin, adrenaline, renin and thyroxine, and corticosteroids associated with milk synthesis (Bernabucci et al., 2010, Rhoads et al., 2010). Altering the activity of these hormones has consequences on feed intake and metabolism, which consequently impacts milk yield (Rhoads et al., 2010).

Since depressed feed intake is a major contributor to milk decline in heat-stressed cows, it would be interesting to see if manipulating candidate genes identified in the neuronal pathways (e.g., *NPFFR2* gene) induces hypothermia and stimulates feed intake. For example, a recent work by Laible et al. (2020) attempted to lighten the skin color for Holstein through gene-editing in a bid to minimise absorption of solar radiation during heat stress (see Figure 2-6). Also, research in mice by Zhang et al. (2018) suggests that the deletion of the *NPFFR2* gene results in impaired diet-induced thermogenesis and energy metabolism. Future in-depth interrogations of these genes, such as transcriptomic and biochemical profiles under different environmental conditions, are warranted. Ultimately, this may provoke innovative ways for managing thermal stress, such as designing specific agonist or antagonist compounds that can be used as feed supplements for dairy animals during heat stress conditions. For example, Ractopamine and Zilpaterol β-agonists used to enhance weight gain and feed efficiency in

livestock, have been developed and approved for commercial use as feed additives in some countries, such as the USA, Canada, and Japan (Centner et al., 2014, Niño et al., 2017).

If similar compounds for minimising heat stress are developed in the future, one possible way to test their benefits is to compare with other known naturally occurring compounds with opposite effects on thermoregulation (i.e., those promoting heat stress). An excellent example is Ergovaline – an ergot alkaloid often found in endophyte-infected species (e.g., Ryegrass and Tall fescue), which is toxic to cattle due to its dopaminergic effects causing disturbances on animal physiology. A number of studies have found an association between Ergovaline and predisposition to heat stress in livestock characterised by increased core body temperature, excessive panting, shade seeking (since Ergovaline promotes vasoconstriction thus limiting heat dissipation), decreased feed intake, weight gain, and milk production (see review by Klotz and Nicol (2016)). In fact, "fescue toxicosis" is a big issue in the USA compromising feed intake and productivity in cattle and sheep during summer seasons when animals are fed endophyte-contaminated diets (Paterson et al., 1995, Klotz and Nicol, 2016). I believe that the development of compounds for use in reducing heat stress in livestock are less likely to face significant regulatory hurdles compared to, for example, Ractopamine or Zilpaterol (discussed above) because they can help to improve animal welfare (health problems, burns, hunger, thirsty, frustration, aggression, etc.), which is becoming a growing issue worldwide due to global warming. Indeed, a recent survey in Brazil (Yunes et al., 2021) found that the public was more receptive to gene-editing aimed to minimise heat stress than gene-editing geared towards improving muscle growth in animals.

6.3.4. Prioritised sequence variants increase the accuracy of genomic predictions. How can the industry benefit or implement these markers?

It is not enough to merely discover candidate variants controlling heat tolerance without demonstrating their relevance in animal breeding and in other species. Therefore, in Chapter 5, I tested whether adding sets of prioritised sequence variants from GWAS into the standard-industry 50k SNP array enhances the prediction accuracy for heat tolerance in dairy cattle. This is relevant because the genetic improvement for a trait is linearly related to the accuracy of estimated breeding values (EBVs), selection intensity, genetic variation and is inversely proportional to the generation cycle (Schaeffer, 2006). Even a smaller lift in prediction

accuracy is valuable to the wider industry with respect to the economic efficiency of breeding programmes.

In Chapter 5, I tested different prediction scenarios to see the margin of increase in the accuracy for heat tolerance. The discovery of sequence variants for heat tolerance benefited from using a large sample size and genotype dataset (~15 million SNPs) of only Holstein cows (N = 20,623) or a combined set of Holsteins + Jersey cows (N = 25,766) in single-trait and meta-GWAS analyses. This means increased power and precision of GWAS to discover multiple candidate causal variants functionally linked to the biology of thermal stress in animals. Consequently, I found gains in accuracy of up to 10% in some scenarios when the pre-selected set of sequence markers (~9,000 SNPs) were added to the standard-industry 50k SNP panel. However, the gain in prediction accuracies varied considerably depending on the scenario tested.

Overall, the results confirmed that pre-selected sequence variants are relevant for heat tolerance and can be used to drive the genetic gain for this trait. There are several possible practical ways to leverage these pre-selected variants for the dairy industry such as: a) including them in the standard 50k SNP arrays used for routine genomic evaluation b) use these variants to design customized SNP panels, e.g., Xiang et al. (2021), which I will discuss in subsequent sections.

6.3.5. **Incorporate prioritised SNPs for heat tolerance in the standard-industry 50k SNP (ST-50K) custom panel?**

The Australian dairy industry currently uses 50k SNP array in routine genomic evaluations implemented by DataGene (https://datagene.com.au/); Melbourne, Australia). This organisation receives genotypes and animal information from various authorised Genomic Service Providers (GSPs). Usually, these GSPs supply genotypes for animals from various lower density SNP chips (only those accepted by DataGene) which are then imputed to the standard 50k SNP array. To do this, DataGene has a special reference set of animals for imputation with real standard 50k genotypes. Building and optimising such robust reference set and SNP chips (e.g., 50k) requires considerable work and resources including SNP discovery, filtering high-quality SNPs (including a high polymorphism, MAF, and call rate), and deploying onto a genotyping assay. Therefore, while this thesis has discovered important

sequence variants for heat tolerance, we should expect a lag of time before an 'imputation reference' set of animals is assembled with real genotypes for these variants. Notably, the routine evaluation of HT-ABV in Australia is currently based on the estimated SNP effects from the standard 50k SNP array that were computed following previous work (Nguyen et al., 2016, Nguyen et al., 2017a).

6.3.6. Incorporate prioritised SNPs for heat tolerance into the combined set of standard-industry 50k SNP (ST-50K) and XT-50K SNP custom panel?

An alternative way to benefit from the heat tolerance variants from this study is to impute them in the 'imputation reference' set of animals, such that the SNPs used for routine imputation in the industry comprise both the real genotypes and imputed sequence variants for heat tolerance instead of only real genotypes. While this could enable faster and cost-effective way for utilizing heat tolerance variants from this thesis in the industry, the fact that some imputed variants in the 'imputation reference' set are associated with an imputation error may be less appealing to the breeders. Besides, most candidate causal variants for heat tolerance discovered in this thesis (see Supplementary Figure S5) are associated with low minor allele frequencies (i.e., they are rare variants), meaning that they are likely to be imputed with large error. However, stringent imputation quality checks can be applied to retain high-quality candidate causal variants for a trait in the 'imputation reference' set. In this regard, there is a proposal in Australia to build 'imputation reference' set of animals that combines real genotype variants for the standard 50k SNP (ST-50K) and XT-50K array (Iona Macleod; personal communication; Figure 6-5). The latter panel (XT-50K) was recently developed by Xiang et al. (2021). The new imputation panel (ST-50K + XT-50K) will contain imputed variants from XT-50K that are missing in the ST-50K and vice versa (Iona MacLeod; personal communication). The new combined panel may provide a convenient way to augment 'refined' set of heat tolerance variants discovered in this thesis for routine genomic evaluation. Therefore, I recommend future tests to see added benefits of integrating heat tolerance 'refined' set of variants into the new combined imputation panel (i.e., ST-50K + XT-50K; Figure 6-5).
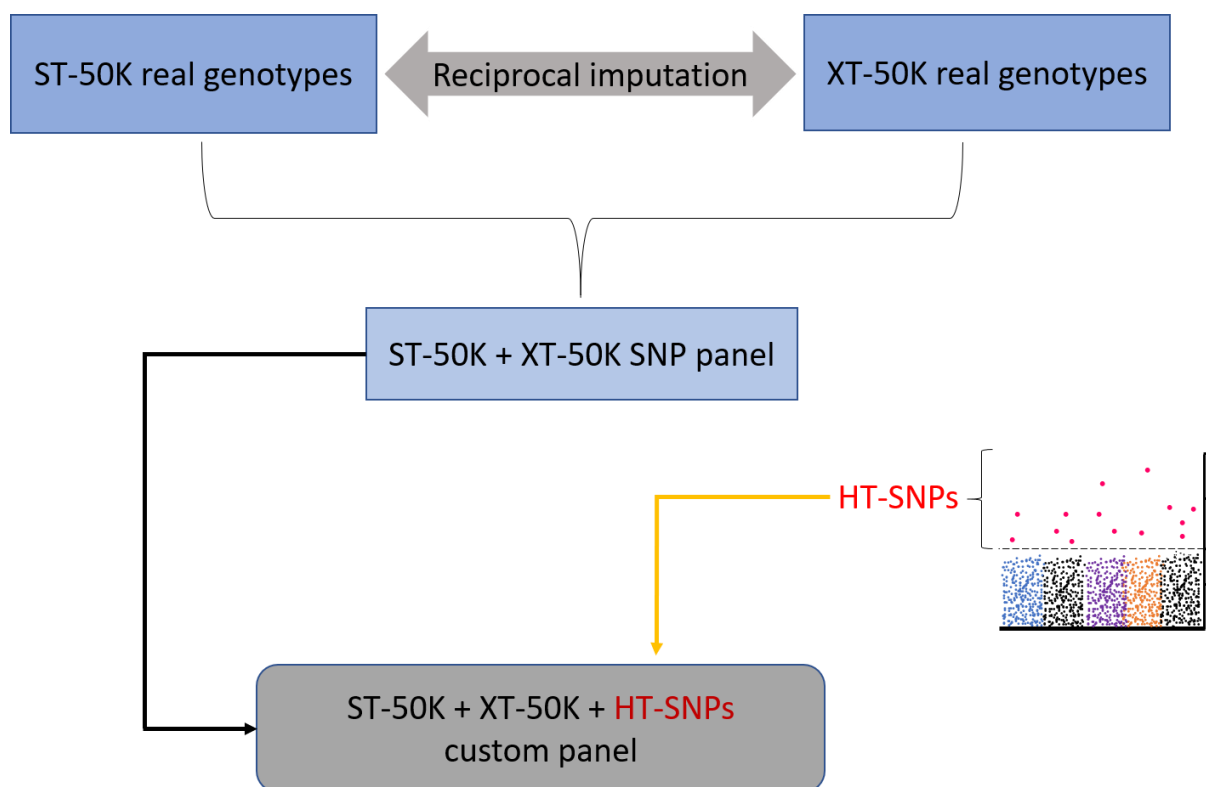
Figure 6-5 Incorporating pre-selected 'refined' set of heat-tolerance SNPs (HT-SNPs) into the combined set of real genotypes from standard-industry 50k (ST-50K) and XT-50K SNP panel developed recently by Xiang et al. (2021).

6.3.7. **Use pre-selected markers to design special SNP panels?**

With the increasing availability of whole-genome sequence data in recent years, there has been an evolving trend towards screening causal mutations, or variants in high LD with causative mutations for multiple traits and using them to design customized SNP panels for driving genetic improvements in livestock (e.g., Liu et al., 2021, Xiang et al., 2021). As noted earlier, Xiang et al. (2021) developed a custom SNP array (called 'XT-50K' array) that includes potential causative mutations discovered from dairy cattle for 34 traits, representing milk production, fertility, type, and management. Since heat tolerance traits were not part of the 34 traits analysed, I recommend that the pre-selected 'refined' core set of variants from this thesis be incorporated in the XT-50K SNP panel in the future. This could enable simultaneous improvement of heat tolerance with other traits contributing to farm profitability (i.e., 'win for all'). Also, it could help circumvent the possible issues related to low imputation accuracy on genomic predictions if the pre-selected variants are imputed in the industry 50k SNP array since most of these variants are characterised by low minor allele frequency (see Chapter 5;

Figure S5). Furthermore, and as demonstrated by (Khansefid et al., 2020), I expect sustained genomic predictions over many generations for heat tolerance when using the XT-50K array since the prioritised markers are closer to the causal mutations. These researchers found that using the XT-50K array yields a consistent and superior accuracy of predictions in crossbred cows than the standard 50K, or HD SNP panels – crossbred cows represent "more distant relationships or many generations".

In their work, Xiang et al. (2021) prioritised sequence variants for designing XT-50K panel based on the extent of pleiotropy and functionality across 34 traits. The magnitude of pleiotropy was estimated based on multi-trait meta-GWAS, while functionality was assessed using Function-And-Evolutionary Trait Heritability (FAETH) score (Xiang et al., 2019). Prioritised markers included top-ranked variants with low multi-trait p-value and high FAETH score, which were found to increase genetic variances and prediction accuracies for economic traits than the bottom-ranked variants. Since pre-selected variants in this thesis did not consider FAETH scores, I expect some selected variants lacking significant functionality across multiple traits should be weeded out, such that a smaller 'refined SNP set' remains for incorporating into the future re-designed XT-50K array.

If the objective is to obtain a 'refined' set of SNPs that are beneficial across breeds, then it would be necessary to further prune the selected variants for their effect direction and the LD. More importantly, it would be crucial to do further rigorous tests to ensure that the selected 'refined SNP set' yields added benefits when they are included in the XT-50K array. Finally, since various biological mechanisms contribute to variations in heat tolerance (morphological, physiological, behavioural, etc.), it is imperative to continue searching for candidate causal variants underlying this trait for inclusion in the custom SNP arrays. Ultimately, this may facilitate rapid genetic progress for thermotolerance while maintaining milk productivity in farm animals.

While including 'refined SNP set' for heat tolerance in the custom panel (e.g., XT-50K) could be the most ideal strategy for the industry to benefit from the outcome of this thesis, we would still expect to see a lag of time, say several years before this is implemented. This is because re-designing and optimizing such fixed custom arrays is time-consuming and costly, requiring the re-assembly of an 'imputation reference' set of animals for routine genomic

evaluations. In addition, combining SNPs for heat tolerance with SNPs for other traits (e.g., based on meta-analysis) when re-designing such custom chips is likely to 'dilute' or diminish the effects of major SNPs for heat tolerance. Furthermore, increasing the number of SNPs in the custom SNP chips (i.e., adding heat tolerance markers) also increases the genotyping costs, which may impede uptake by breeders. Therefore, alternative more suitable options need to be sought taking advantage of emerging genotyping technologies.

6.3.8. **Capitalize on the new sequencing technologies to speed up the uptake of prioritised variants for heat tolerance in the industry?**

The next-generation sequencing technology is rapidly evolving; thus, increasing efficiency and reducing genotyping costs. For example, the low-pass or low-coverage genome sequencing [also called 'skim sequencing' in which a depth of 1x or less of the genome is sequenced] is emerging as a potential revolutionary tool for genotyping (1000 Genomes Project Consortium, 2015). Recent work in humans (Martin et al., 2021, Rubinacci et al., 2021) and cattle (Snelling et al., 2020) have shown high (> 0.90) concordance of variants in the standard SNP arrays versus imputed genotyped calls for individuals that had been sequenced at low coverage (0.5-1x) based on a haplotype reference panel. This implies that skim sequencing could offer a competitive cost-effective alternative in the foreseeable future; thus, replacing the "gold standard" SNP arrays, which have been popular in the market for over a decade. In addition, skim sequencing is also intuitive in that it can minimise the inherent ascertainment bias of SNP arrays and allows the detection of rare variants for a trait (Rubinacci et al., 2021). Overall, such technological advancements are intriguing, as they could allow screening many individuals at an affordable cost to identify heat-tolerant genotypes without the need to continuously re-design customized SNP arrays. However, even as this new development unfolds, it would be crucial at this point to test the value of the candidate variants for heat tolerance discovered in this thesis when they are added to the custom SNP panels such as the XT-50K array (Xiang et al., 2021).

6.3.9. **Consider reference population to further improve genomic prediction for heat tolerance?**

Most genomic predictions analyses in Chapter 5 used a reference set of only Holstein bulls. However, using a multi-breed reference set (Holsteins + Jersey bulls), in which pre-selected variants were from only Holsteins yielded a consistent increase in the accuracies for

most prediction scenarios (see Chapter 5; Figure 4). Also, the bias of predictions (calculated as the regression coefficient of the slope phenotypes on the GEBV in the validation sets) decreases (i.e., closer to 1) when using a multi-breed compared to a single-breed reference set. These findings emphasise the importance of having a large reference population that represents multiple breeds for training prediction equations for heat tolerance. In recent work, Khansefid et al. (2020) found that equalizing breeds in the reference set instead of the Holstein-dominated set increases the accuracy and reduces the bias of predictions. Similarly, van den Berg et al. (2020) found that the reliability of genomic prediction highly depends on the choice of the animals in the reference set. These authors reported that including few closely related Holsteins in the reference set (instead of just increasing the size of Holsteins) increased the reliabilities of prediction in the Australian Red dairy cattle. Work is underway in Australia to enlarge the size of the reference population through dedicated genotyping of cows with high-quality phenotypes under a project called Ginfo (Pryce et al., 2018). As such, future work is required to see if updating the reference set that was used to estimate the SNP effects for HT ABVs in 2017 by Nguyen et al. (2017a) increases the reliability of heat tolerance. In line with this, I recommend testing the added benefits of using the updated reference set in addition to considering the 'refined set' of pre-selected sequence markers discovered in this thesis.

6.4. **Concluding remarks**

Overall, this thesis has made novel discoveries that increase our knowledge of the genetic basis and biology of thermal stress, which may open new avenues for minimising the effects of heat stress in animals, considering escalating warming climate worldwide. The study found substantial genetic variation for heat tolerance among Australian dairy cattle, with some re-ranking occurring among the bulls when comparing their performance in colder and hotter environments, although the level of re-ranking is not high enough (genetic correlation > 0.80) to necessitate forming separate breeding schemes at this stage. However, the study found that Australian dairy animals are becoming less adapted to their production environments (evidenced by increasing $G \times E$ or re-ranking), which is a concerning trend that requires routine monitoring to prevent future economic losses as the climate gets warmer. The dairy sires identified as heat-resilient, or generalists could be used by farmers in warmer environments to optimise performance and can also help reverse the growing $G \times E$ trend in Australia.

The research also leveraged the availability of a large dataset (~30,000 cows and ~15 million SNPs) to discover genetic variants (based on GWAS) that make Australian dairy cattle differ in their response to heat stress. Interestingly, the results revealed specific candidate causal variants and genes related to the nervous (neuroactive ligand-receptor interaction and glutamatergic synapse) and metabolic functions (citrate cycle) as key hallmarks for thermotolerance in animals including 61 candidate causal variants for heat tolerance in genomic sites highly conserved across 100 vertebrate species. In follow-up research (Chapter 5), candidate variants identified from GWAS were found to be relevant for heat tolerance and can be used to increase the accuracy of prediction when they are added into the industry 50k SNP panel, with a margin of gain ranging from 0.1% to 10% units depending on the prediction scenario, which is critical for driving the genetic improvement for this trait.

6.5. Future research considerations

6.5.1. Further refine the prioritised SNP set…

While the results of this thesis are encouraging showing that we can increase the prediction accuracy of heat tolerance by using a small set of pre-selected sequence variants (~2, 000 to 9,000 SNPs), additional work is still needed to obtain a 'refined SNP set' for this trait that can be incorporated in standard industry SNP panels or the custom SNP chips, e.g., XT-50K array (Xiang et al. 2021). Specifically, the pre-selected SNPs from this thesis need to be pruned for LD and MAF, check their effect direction (especially, if there are to be used across multiple breeds), and ensure that the pruned set is well-spaced across the genome in relation to other SNPs in custom arrays. Also, check the FEATH scores as described by Xiang et al. (2019), if the objective is to rank the pre-selected variants based on functional and evolutionary significance across multiple traits. This is important considering that Australian selection indices include multiple traits of economic importance, such as production, fertility, and type traits (Byrne et al., 2016). Furthermore, rigorous tests are needed to ensure that the 'refined SNP set' yields added benefits when they are incorporated into the custom SNP panels, e.g., the XT-50K array (Xiang et al., 2021).

6.5.2. Heat-tolerance trait definition and selection criteria…

Throughout this thesis, I focused analyses of heat tolerance based on milk production traits (i.e., milk, fat, and protein yield slope traits). While these traits are economically attractive as breeding targets for heat stress with respect to producer perspective, other novel

traits that better captures heat tolerance need to be investigated. As in Brito et al. (2020), this thesis proposes traits associated with thermoregulation called thermal circulation index (TCI) which quantifies heat dissipation efficiency, such that animals with large TCI values are expected to experience a greater decline in productivity under heat stress conditions relative to the average population. However, future work is needed to explore the suitability of this trait as the breeding target for heat tolerance. If this trait is established to be a suitable target for heat tolerance, then inexpensive selection criteria traits can be explored, such as milk decays (i.e., milk slope straits) or MIR-predicted milk biomarkers. Also, it is important to consider other traits for the genetic selection of heat tolerance (fertility, health, conformation, energy balance, etc.) since heat tolerance is a highly complex trait involving multi-faceted biological processes (physiological, behavioral, morphological, etc.).

6.5.3. **Explore resilience indicators for multiple traits…**

Another way to breed for heat tolerance is to consider resilience indicators in genomic selection. This is a developing but growing area of research that warrants more and deeper exploration. With increasing climate change and variability, I foresee the potential development of a breeding value called 'general resilience' in the future, which encompasses multiple traits that contribute to farm profitability, including heat-tolerance. So that the top-ranking animals for general resilience breeding value are expected to have smaller phenotypic variance or deviation relative to the average population and can recover more quickly following one or multiple environmental disturbances (heat, diseases, parasites, nutrition, precipitation, management, variability in climate, etc.). Thanks to the advancement in phenotyping technologies, this is increasingly becoming possible with the availability of big data for genetic evaluations.

6.5.4. **New traits and comparative studies…**

One of the main challenges to breeding for resilience is possible negative (undesirable) effects on milk production – a key driver for farm profitability. In this thesis, I spent a considerable amount of work trying to separate heat-resilient genes and those controlling milk yield traits. This proved to be a non-trivial task since the heat-tolerance traits (milk yield slopes) used in the GWAS are strongly negatively correlated (~0.80) with the level milk yield, meaning that they are controlled, to a large extent, by the same genes. Besides, this correlation estimate means that high-producing animals have a higher reduction in their yield as the temperature-

humidity increases and vis versa for lower producing animals. Ideally, farmers would desire to have animals that are both high-producing and heat-tolerant. As such, the results in this thesis and previous work (e.g., Nguyen et al., 2016) suggest that striking the optimum balance between these two traits is challenging but not insurmountable. One possible way to deal with this issue is to explore other heat-tolerance or resilience indicators that are potentially less correlated with milk production (e.g., measures of core body temperature, respiratory rate, heart rate, milk biomarkers, etc.). Recent development in sensor technology has provided an opportunity to capture large heat tolerance phenotypes at an individual level and in real-time (Koltes et al., 2018), which may facilitate us to better understand the biology of this trait. Besides defining new phenotypes, the heat tolerance milk yield slope traits could be considered in the breeding goals and the negative correlation accounted for in the selection indices. Moreover, to better understand the genetic control of heat tolerance, comparative studies within and between species need to be sought, e.g., dairy vs beef breeds or even small ruminants such as goats and sheep.

6.5.5. **More suitable model (s) for quantifying heat stress in Australian dairy industry…**

It is also crucial to investigate alternative more accurate ways of quantifying the magnitude of environmental heat load to the animals. In this thesis, and following previous work in Australia (Hayes et al., 2016, Nguyen et al., 2016), I calculated the rate of milk decline as a function of THI (i.e., heat tolerance or slope traits). Since the THI model used (Yousef, 1985) does not consider wind speed and solar radiation, it is possible that the magnitude of heat on the animals may have been underestimated, given that Australian dairy herds are predominantly reared outdoors on pasture where they experience direct effect of these weather elements, which contrast with some other countries, e.g., the USA where dairy cows are mostly confined indoors and fed total mixed rations. Therefore, future research is warranted to recommend the best suitable THI model (s) for the industry.

6.5.6. **Follow-up work on key biological pathways for heat-tolerance…**

This thesis highlights the relevance of genes related to the nervous (neuroactive ligand-receptor interaction and glutamatergic synapse) and metabolic (citrate or Krebs cycle) functions as key drivers for thermotolerance in dairy animals. However, while several studies have linked these pathways to thermal stress in livestock (e.g., Kim et al., 2017, Garner et al., 2020), this thesis provides new insights into the molecular aspects (i.e., candidate causal

mutations) underpinning these biological features, which require further in-depth interrogations.

6.5.7. **Larger sample size and international collaborations…**

The results in this thesis (Chapter 4) point to a highly polygenic nature of heat tolerance characterised by many causal variants with small effects. As such, an even larger dataset (~ 30,000 cows were used in this thesis) with tens of thousands of individuals such as those typically used in humans (Wood et al., 2014) would be required in future work to detect causal variants with very small effects and the effects of rare causal variants. Sharing data via international collaborations is one critical avenue to achieve the large sample size needed for mapping and validating heat-tolerant causal variants. Alternatively, a more feasible approach is to perform a meta-analysis of GWAS results on heat tolerance traits from different countries, e.g., meta-GWAS for cattle stature (Bouwman et al., 2018).

Moreover, these collaborations may open promising platforms to compare and pinpoint specific genetic variants that make adapted breeds (say those from warm climates, e.g., Zebu) differ from temperate breeds (e.g., Holsteins) in their thermotolerance. In this regard, some promising progress has been made in mapping a thermotolerance gene, referred to as the 'SLICK' gene from the Senepol cattle, and work is now ongoing in several countries to transfer specific alleles from this gene to Holstein dairy cattle via crossbreeding or gene editing (Hansen, 2020). Nonetheless, the specific allele combination or the percentage of adapted breed genetic background required to benefit from heterosis or hybrid vigour advantage for resilience in crossbreds while optimizing productivity remains unknown, which warrants future research. Also, it is possible that crossbred cattle carrying the SLICK genotype and are ranked high for heat tolerance GEBVs have superior thermotolerance ability, which require future interrogations to confirm this hypothesis. Overall, the genetic tools offer promising and long-term prospects of improving resilience aspects in animals, which is crucial in addressing the double challenge of increasing animal production, even more, to feed a growing population while coping with the effects of rising global temperatures and ever-changing production environments. This thesis work contributes to this overarching goal.

6.6. **References**

1000 Genomes Project Consortium. 2015. A global reference for human genetic variation. Nature 526(7571):68.

Atta, M. S., F. A. Farrag, E. A. Almadaly, H. A. Ghoneim, A. S. Hafez, S. K. Al Jaouni, S. A. Mousa, and H. Ali. 2020. Transcriptomic and biochemical effects of pycnogenol in ameliorating heat stress-related oxidative alterations in rats. Journal of Thermal Biology 93:102683.

Baumgard, L. H. and R. P. Rhoads Jr. 2013. Effects of heat stress on postabsorptive metabolism and energetics. Annual Review of Animal Biosciences 1(1):311-337.

Belhadj Slimen, I., T. Najar, A. Ghram, and M. Abdrrabba. 2016. Heat stress effects on livestock: molecular, cellular and metabolic aspects, a review. Journal of Animal Physiology and Animal Nutrition 100(3):401-412.

Berghof, T. V., M. Poppe, and H. A. Mulder. 2019. Opportunities to improve resilience in animal breeding programs. Frontiers in Genetics 9:692.

Bernabucci, U., N. Lacetera, L. H. Baumgard, R. P. Rhoads, B. Ronchi, and A. Nardone. 2010. Metabolic and hormonal acclimation to heat stress in domesticated ruminants. Animal 4(7):1167-1183.

BOM. 2020. State of the Climate 2020. http://www.bom.gov.au/state-of-the-climate/australias-changing-climate.shtml.

BOM and CSIRO. 2015. Climate-ready Victoria: Gippsland. https://www.climatechange.vic.gov.au/__data/assets/pdf_file/0021/60744/Gippsland.pdf.

Bouwman, A. C., H. D. Daetwyler, A. J. Chamberlain, C. H. Ponce, M. Sargolzaei, F. S. Schenkel, G. Sahana, A. Govignon-Gion, S. Boitard, and M. Dolezal. 2018. Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. Nature Genetics 50(3):362-367.

Brito, L. F., H. R. Oliveira, B. R. McConn, A. P. Schinckel, A. Arrazola, J. N. Marchant-Forde, and J. S. Johnson. 2020. Large-scale phenotyping of livestock welfare in commercial production systems: a new frontier in animal breeding. Frontiers in Genetics 11.

Bryant, J., N. López-Villalobos, J. Pryce, C. Holmes, and D. Johnson. 2006. Reaction norms used to quantify the responses of New Zealand dairy cattle of mixed breeds to nutritional environment. New Zealand Journal of Agricultural Research 49(4):371-381.

Bryant, J., N. López-Villalobos, J. Pryce, C. Holmes, and D. Johnson. 2007. Quantifying the effect of thermal environment on production traits in three breeds of dairy cattle in New Zealand. New Zealand Journal of Agricultural Research 50(3):327-338.

Byrne, T., B. Santos, P. Amer, D. Martin-Collado, J. Pryce, and M. Axford. 2016. New breeding objectives and selection indices for the Australian dairy industry. Journal of Dairy Science 99(10):8146-8167.

Carabaño, M. J., M. Ramón, A. Menéndez-Buxadera, A. Molina, and C. Díaz. 2019. Selecting for heat tolerance. Animal Frontiers 9(1):62-68.

Centner, T. J., J. C. Alvey, and A. M. Stelzleni. 2014. Beta agonists in livestock feed: Status, health concerns, and international trade. Journal of Animal Science 92(9):4234-4240.

Colditz, I. G. and B. C. Hine. 2016. Resilience in farm animals: biology, management, breeding and implications for animal welfare. Animal Production Science 56(12):1961-1983.

Cowley, F., D. Barber, A. Houlihan, and D. Poppi. 2015. Immediate and residual effects of heat stress and restricted intake on milk protein and casein composition and energy metabolism. Journal of Dairy Science 98(4):2356-2368.

Curtis, S. E. 1983. Environmental management in animal agriculture. Iowa State University Press.

Daetwyler, H. D., B. Villanueva, and J. A. Woolliams. 2008. Accuracy of predicting the genetic risk of disease using a genome-wide approach. PloS ONE 3(10):e3395.

Davis, S., R. Spelman, and M. Littlejohn. 2017. Breeding and genetics symposium: Breeding heat tolerant dairy cattle: The case for introgression of the "slick" prolactin receptor variant into Bos taurus dairy breeds. Journal of Animal Science 95(4):1788-1800.

Dikmen, S., E. Alava, E. Pontes, J. Fear, B. Dikmen, T. Olson, and P. Hansen. 2008. Differences in thermoregulatory ability between slick-haired and wild-type lactating Holstein cows in response to acute heat stress. Journal of Dairy Science 91(9):3395-3402.

Dikmen, S., J. Cole, D. Null, and P. Hansen. 2012. Heritability of rectal temperature and genetic correlations with production and reproduction traits in dairy cattle. Journal of Dairy Science 95(6):3401-3405.

Dikmen, S., J. B. Cole, D. J. Null, and P. J. Hansen. 2013. Genome-wide association mapping for identification of quantitative trait loci for rectal temperature during heat stress in Holstein cattle. PloS ONE 8(7):e69202.

Dikmen, S. and P. Hansen. 2009. Is the temperature-humidity index the best indicator of heat stress in lactating dairy cows in a subtropical environment? Journal of Dairy Science 92(1):109-116.

Dikmen, S., F. Khan, H. Huson, T. Sonstegard, J. Moss, G. Dahl, and P. Hansen. 2014. The SLICK hair locus derived from Senepol cattle confers thermotolerance to intensively managed lactating Holstein cows. Journal of Dairy Science 97(9):5508-5520.

Finch, V. 1986. Body temperature in beef cattle: its control and relevance to production in the tropics. Journal of Animal Science 62(2):531-542.

Gao, S., J. Guo, S. Quan, X. Nan, M. S. Fernandez, L. Baumgard, and D. Bu. 2017. The effects of heat stress on protein metabolism in lactating Holstein cows. Journal of Dairy Science 100(6):5040-5049.

García-Ruiz, A., J. B. Cole, P. M. VanRaden, G. R. Wiggans, F. J. Ruiz-López, and C. P. Van Tassell. 2016. Changes in genetic selection differentials and generation intervals in US

Holstein dairy cattle as a result of genomic selection. Proceedings of the National Academy of Sciences 113(28):E3995-E4004.

Garner, J., A. Chamberlain, C. Vander Jagt, T. Nguyen, B. Mason, L. Marett, B. Leury, W. Wales, and B. Hayes. 2020. Gene expression of the heat stress response in bovine peripheral white blood cells and milk somatic cells in vivo. Scientific reports 10(1):1-12.

Garner, J., M. Douglas, S. O. Williams, W. Wales, L. Marett, T. Nguyen, C. Reich, and B. Hayes. 2016. Genomic selection improves heat tolerance in dairy cattle. Scientific Reports 6(1):1-9.

Gonzalez-Recio, O., M. Coffey, and J. Pryce. 2014. On the value of the phenotypes in the genomic era. Journal of Dairy Science 97(12):7905-7915.

Gourdine, J.-L., N. Mandonnet, M. Giorgi, and D. Renaudeau. 2017. Genetic parameters for thermoregulation and production traits in lactating sows reared in tropical climate. Animal 11(3):365-374.

Haile-Mariam, M., M. Carrick, and M. Goddard. 2008. Genotype by environment interaction for fertility, survival, and milk production traits in Australian dairy cattle. Journal of Dairy Science 91(12):4840-4853.

Hammami, H., J. Vandenplas, M.-L. Vanrobays, B. Rekik, C. Bastin, and N. Gengler. 2015. Genetic analysis of heat stress effects on yield traits, udder health, and fatty acids of Walloon Holstein cows. Journal of Dairy Science 98(7):4956-4968.

Hansen, P. J. 2020. Prospects for gene introgression or gene editing as a strategy for reduction of the impact of heat stress on production and reproduction in cattle. Theriogenology 154:190-202.

Hayes, B., M. Carrick, P. Bowman, and M. Goddard. 2003. Genotype×environment interaction for milk production of daughters of Australian dairy sires from test-day records. Journal of Dairy Science 86(11):3736-3744.

Hayes, B. J., P. J. Bowman, A. J. Chamberlain, K. Savin, C. P. Van Tassell, T. S. Sonstegard, and M. E. Goddard. 2009. A validated genome wide association study to breed cattle adapted to an environment altered by climate change. PloS ONE 4(8):e6676.

Hayes, B. J. and H. D. Daetwyler. 2019. 1000 bull genomes project to map simple and complex genetic traits in cattle: applications and outcomes. Annual Review of Animal Biosciences 7:89-102.

Hayes, B. J., H. D. Daetwyler, and M. E. Goddard. 2016. Models for genome×environment interaction: examples in livestock. Crop Science 56(5):2251-2259.

Hayes, B. J., J. Pryce, A. J. Chamberlain, P. J. Bowman, and M. E. Goddard. 2010. Genetic architecture of complex traits and accuracy of genomic prediction: coat colour, milk-fat percentage, and type in Holstein cattle as contrasting model traits. PLoS Genetics 6(9):e1001139.

Ho, P. and J. Pryce. 2020. Predicting the likelihood of conception to first insemination of dairy cows using milk mid-infrared spectroscopy. Journal of Dairy Science 103(12):11535-11544.

Igoshin, A. V., Yurchenko, A. A., Belonogova, N. M., Petrovsky, D. V., Aitnazarov, R. B., Soloshenko, V. A., ... & Larkin, D. M. (2019). Genome-wide association study and scan for signatures of selection point to candidate genes for body temperature maintenance under the cold stress in Siberian cattle populations. BMC genetics 20(1):5-14.

Kadzere, C., M. Murphy, N. Silanikove, and E. Maltz. 2002. Heat stress in lactating dairy cows: a review. Livestock Production Science 77(1):59-91.

Khansefid, M., M. E. Goddard, M. Haile-Mariam, K. V. Konstantinov, C. Schrooten, G. de Jong, E. G. Jewell, E. O'Connor, J. E. Pryce, and H. D. Daetwyler. 2020. Improving Genomic Prediction of Crossbred and Purebred Dairy Cattle. Frontiers in Genetics 11.

Kim, J.-M., K.-S. Lim, M. Byun, K.-T. Lee, Y.-r. Yang, M. Park, D. Lim, H.-H. Chai, H.-T. Bang, and J. Hwangbo. 2017. Identification of the acclimation genes in transcriptomic responses to heat stress of White Pekin duck. Cell Stress and Chaperones 22(6):787-797.

Klotz, J. and A. Nicol. 2016. Ergovaline, an endophytic alkaloid. 1. Animal physiology and metabolism. Animal Production Science 56(11):1761-1774.

Koltes, J. E., D. A. Koltes, B. E. Mote, J. Tucker, and D. S. Hubbell III. 2018. Automated collection of heat stress data in livestock: new technologies and opportunities. Translational Animal Science 2(3):319-323.

Laible, G., S.-A. Cole, B. Brophy, S. Leath, S. Jivanji, M. D. Littlejohn, and D. N. Wells. 2020. Holstein Friesian dairy cattle edited for diluted coat color as adaptation to climate change. bioRxiv [preprint].

Lees, J., A. Lees, and J. Gaughan. 2018. Developing a heat load index for lactating dairy cows. Animal Production Science 58(8):1387-1391.

Littlejohn, M. D., K. M. Henty, K. Tiplady, T. Johnson, C. Harland, T. Lopdell, R. G. Sherlock, W. Li, S. D. Lukefahr, and B. C. Shanks. 2014. Functionally reciprocal mutations of the prolactin signalling pathway define hairy and slick cattle. Nature Communications 5(1):1-8.

Liu, Z., C. Sun, Y. Yan, G. Li, X. C. Li, G. Wu, and N. Yang. 2021. Design and evaluation of a custom 50K Infinium SNP array for egg-type chickens. Poultry Science 100(5):101044.

Lucy, M. 2001. Reproductive loss in high-producing dairy cattle: where will it end? Journal of Dairy Science 84(6):1277-1293.

Luo, H., L. F. Brito, X. Li, G. Su, J. Dou, W. Xu, X. Yan, H. Zhang, G. Guo, and L. Liu. 2021. Genetic parameters for rectal temperature, respiration rate, and drooling score in Holstein cattle and their relationships with various fertility, production, body conformation, and health traits. Journal of Dairy Science 104(4):4390-4403.

Macciotta, N., S. Biffani, U. Bernabucci, N. Lacetera, A. Vitali, P. Ajmone-Marsan, and A. Nardone. 2017. Derivation and genome-wide association study of a principal component-based measure of heat tolerance in dairy cattle. Journal of Dairy Science 100(6):4683-4697.

Maia, A. and C. B. Loureiro. 2005. Sensible and latent heat loss from the body surface of Holstein cows in a tropical environment. International Journal of Biometeorology 50(1):17-22.

Martin, A. R., E. G. Atkinson, S. B. Chapman, A. Stevenson, R. E. Stroud, T. Abebe, D. Akena, M. Alemayehu, F. K. Ashaba, and L. Atwoli. 2021. Low-coverage sequencing cost-effectively detects known and novel variation in underrepresented populations. The American Journal of Human Genetics 108(4):656-668.

Martin, P., H. Barkema, L. Brito, S. Narayana, and F. Miglior. 2018. Symposium review: novel strategies to genetically improve mastitis resistance in dairy cattle. Journal of Dairy Science 101(3):2724-2736.

Meuwissen, T. H. 2009. Accuracy of breeding values of 'unrelated' individuals predicted by dense SNP genotyping. Genetics Selection Evolution 41(1):1-9.

Morton, J., W. Tranter, D. G. Mayer, and N. Jonsson. 2007. Effects of environmental heat on conception rates in lactating dairy cows: critical periods of exposure. Journal of Dairy Science 90(5):2271-2278.

Nakamura, K. and S. F. Morrison. 2008. A thermosensory pathway that controls body temperature. Nature Neuroscience 11(1):62-71.

Nguyen, T. T., P. J. Bowman, M. Haile-Mariam, G. J. Nieuwhof, B. J. Hayes, and J. E. Pryce. 2017a. Implementation of a breeding value for heat tolerance in Australian dairy cattle. Journal of Dairy Science 100(9):7362-7367.

Nguyen, T. T., P. J. Bowman, M. Haile-Mariam, J. E. Pryce, and B. J. Hayes. 2016. Genomic selection for tolerance to heat stress in Australian dairy cattle. Journal of Dairy Science 99(4):2849-2862.

Nguyen, T. T., B. J. Hayes, and J. E. Pryce. 2017b. A practical future-scenarios selection tool to breed for heat tolerance in Australian dairy cattle. Animal Production Science 57(7):1488-1493.

Niño, A. M., R. H. Granja, A. C. Wanschel, and A. G. Salerno. 2017. The challenges of ractopamine use in meat production for export to European Union and Russia. Food Control 72:289-292.

Olson, T., C. Lucena, C. Chase Jr, and A. Hammond. 2003. Evidence of a major gene influencing hair length and heat tolerance in Bos taurus cattle. Journal of Animal Science 81(1):80-90.

Paterson, J., C. Forcherio, B. Larson, M. Samford, and M. Kerley. 1995. The effects of fescue toxicosis on beef cattle productivity. Journal of Animal Science 73(3):889-898.

Poppe, M., R. Veerkamp, M. Van Pelt, and H. Mulder. 2020. Exploration of variance, autocorrelation, and skewness of deviations from lactation curves as resilience indicators for breeding. Journal of Dairy Science 103(2):1667-1684.

Pryce, J., T. Nguyen, M. Axford, G. Nieuwhof, and M. Shaffer. 2018. Symposium review: Building a better cow—The Australian experience and future perspectives. Journal of Dairy Science 101(4):3702-3713.

Ravagnolo, O. and I. Misztal. 2002. Effect of heat stress on nonreturn rate in Holsteins: fixed-model analyses. Journal of Dairy Science 85(11):3101-3106.

Rhoads, M., J. Kim, R. J. Collier, B. Crooker, Y. Boisclair, L. Baumgard, and R. Rhoads. 2010. Effects of heat stress and nutrition on lactating Holstein cows: II. Aspects of hepatic growth hormone responsiveness. Journal of Dairy Science 93(1):170-179.

Robertson, A. 1959. The sampling variance of the genetic correlation coefficient. Biometrics 15(3):469-485.

Rubinacci, S., D. M. Ribeiro, R. J. Hofmeister, and O. Delaneau. 2021. Efficient phasing and imputation of low-coverage sequencing data using large reference panels. Nature Genetics 53(1):120-126.

Saadeldin, I. M., A. A.-A. Swelum, M. Elsafadi, A. Mahmood, A. Osama, H. Shikshaky, M. Alfayez, A. N. Alowaimer, and S. Magdeldin. 2020. Thermotolerance and plasticity of camel somatic cells exposed to acute and chronic heat stress. Journal of Advanced Research 22:105-118.

Schaeffer, L. 2006. Strategy for applying genome-wide selection in dairy cattle. Journal of Animal Breeding and Genetics 123(4):218-223.

Schauberger, G., C. Mikovits, W. Zollitsch, S. J. Hörtenhuber, J. Baumgartner, K. Niebuhr, M. Piringer, W. Knauder, I. Anders, and K. Andre. 2019. Global warming impact on confined livestock in buildings: efficacy of adaptation measures to reduce heat stress for growing-fattening pigs. Climatic Change 156(4):567-587.

Scheffer, M., J. E. Bolhuis, D. Borsboom, T. G. Buchman, S. M. Gijzel, D. Goulson, J. E. Kammenga, B. Kemp, I. A. van de Leemput, and S. Levin. 2018. Quantifying resilience of humans and other animals. Proceedings of the National Academy of Sciences 115(47):11883-11890.

Sharma, P., A. B. Jha, R. S. Dubey, and M. Pessarakli. 2012. Reactive oxygen species, oxidative damage, and antioxidative defense mechanism in plants under stressful conditions. Journal of Botany 2012.

Sigdel, A., R. Abdollahi-Arpanahi, I. Aguilar, and F. Peñagaricano. 2019. Whole genome mapping reveals novel genes and pathways involved in milk production under heat stress in US Holstein cows. Frontiers in Genetics 10:928.

Sigdel, A., L. Liu, R. Abdollahi-Arpanahi, I. Aguilar, and F. Peñagaricano. 2020. Genetic dissection of reproductive performance of dairy cows under heat stress. Animal Genetics.

Snelling, W. M., J. L. Hoff, J. H. Li, L. A. Kuehn, B. N. Keel, A. K. Lindholm-Perry, and J. K. Pickrell. 2020. Assessment of imputation from low-pass sequencing to predict merit of beef steers. Genes 11(11):1312.

Srikanth, K., E. Lee, A. Kwan, Y. Lim, J. Lee, G. Jang, and H. Chung. 2017. Transcriptome analysis and identification of significantly differentially expressed genes in Holstein calves subjected to severe thermal stress. International journal of biometeorology 61(11):1993-2008.

Talukder, S., P. Celi, K. Kerrisk, S. Garcia, and N. K. Dhand. 2015. Factors affecting reproductive performance of dairy cows in a pasture-based, automatic milking system research farm: a retrospective, single-cohort study. Animal Production Science 55(1):31-41.

Tsartsianidou, V., V. V. Kapsona, E. Sánchez-Molano, Z. Basdagianni, M. J. Carabaño, D. Chatziplis, G. Arsenos, A. Triantafyllidis, and G. Banos. 2021. Understanding the seasonality of performance resilience to climate volatility in Mediterranean dairy sheep. Scientific Reports 11(1):1-11.

van den Berg, I., P. Ho, T. Luke, M. Haile-Mariam, S. Bolormaa, and J. Pryce. 2021. The use of milk mid-infrared spectroscopy to improve genomic prediction accuracy of serum biomarkers. Journal of Dairy Science 104(2):2008-2017.

van den Berg, I., I. MacLeod, C. Reich, E. Breen, and J. Pryce. 2020. Optimizing genomic prediction for Australian Red dairy cattle. Journal of Dairy Science 103(7):6276-6298.

Wang, T., Y.-P. P. Chen, I. M. MacLeod, J. E. Pryce, M. E. Goddard, and B. J. Hayes. 2017. Application of a Bayesian non-linear model hybrid scheme to sequence data for genomic prediction and QTL mapping. BMC Genomics 18(1):618.

Wang, X., H. Gao, K. G. Gebremedhin, B. S. Bjerg, J. Van Os, C. B. Tucker, and G. Zhang. 2018. A predictive model of equivalent temperature index for dairy cattle (ETIC). Journal of Thermal Biology 76:165-170.

West, J. 2003. Effects of heat-stress on production in dairy cattle. Journal of Dairy Science 86(6):2131-2144.

Wheelock, J., R. Rhoads, M. VanBaale, S. Sanders, and L. Baumgard. 2010. Effects of heat stress on energetic metabolism in lactating Holstein cows. Journal of Dairy Science 93(2):644-655.

Wood, A. R., T. Esko, J. Yang, S. Vedantam, T. H. Pers, S. Gustafsson, A. Y. Chu, K. Estrada, Z. Kutalik, and N. Amin. 2014. Defining the role of common variation in the genomic and biological architecture of adult human height. Nature Genetics 46(11):1173-1186.

Xiang, R., I. M. MacLeod, H. D. Daetwyler, G. de Jong, E. O'Connor, C. Schrooten, A. J. Chamberlain, and M. E. Goddard. 2021. Genome-wide fine-mapping identifies pleiotropic and

functional variants that predict many traits across global cattle populations. Nature Communications 12(1):1-13.

Xiang, R., I. Van Den Berg, I. M. MacLeod, B. J. Hayes, C. P. Prowse-Wilkins, M. Wang, S. Bolormaa, Z. Liu, S. J. Rochfort, and C. M. Reich. 2019. Quantifying the contribution of sequence variants with regulatory and evolutionary significance to 34 bovine complex traits. Proceedings of the National Academy of Sciences 116(39):19398-19408.

Xiong, W., M. P. Reynolds, J. Crossa, T. Payne, U. Schulthess, K. Sonder, N. Addimando, R. P. Singh, K. Ammar, and B. Gerard. 2020. Climate change has increased genotype-environment interactions in wheat breeding. https://repository.cimmyt.org/handle/10883/21039.

Yousef, M. K. 1985. Stress physiology in livestock. Volume I. Basic principles. CRC press.

Yunes, M. C., Z. Osório-Santos, M. A. von Keyserlingk, and M. J. Hötzel. 2021. Gene Editing for Improved Animal Welfare and Production Traits in Cattle: Will This Technology Be Embraced or Rejected by the Public? Sustainability 13(9):4966.

Zhang, L., C. K. Ip, I.-C. J. Lee, Y. Qi, F. Reed, T. Karl, J. K. Low, R. F. Enriquez, N. J. Lee, and P. A. Baldock. 2018. Diet-induced adaptive thermogenesis requires neuropeptide FF receptor-2 signalling. Nature Communications 9(1):1-19.

Zwald, N., K. Weigel, W. Fikse, and R. Rekaya. 2003. Identification of factors that cause genotype by environment interaction between herds of Holstein cattle in seventeen countries. Journal of Dairy Science 86(3):1009-1018.

Appendices

Appendix 1: "Genotype by environment interaction for heat tolerance in Australian Holstein dairy cattle." In *Proceedings for the Association Advancement of Animal Breeding Genetics***. Vol. 23, pp. 39-42. 2019.**

Please note that this conference article has been formatted following the style of the conference proceeding

GENOTYPE BY ENVIRONMENT INTERACTION FOR HEAT TOLERANCE IN AUSTRALIAN HOLSTEIN DAIRY CATTLE

E.K. Cheruiyot[1,2], M. Haile-Mariam[1], T.T.T. Nguyen[1], B.G. Cocks[1,2] and J.E. Pryce[1,2]

[1]Agriculture Victoria, Department of Jobs, Precincts and Regions, Bundoora, VIC 3083, Australia
[2]School of Applied Systems Biology, La Trobe University, Bundoora, VIC 3083, Australia

SUMMARY

Genomic breeding values for heat tolerance in dairy cattle were first released in Australia in December 2017 to select animals with better tolerance to heat stress. It is also important to identify animals which perform well in a wide range of temperature and humidities, given the large seasonal and geographical variation in Australia. The aim of this study was to investigate the magnitude of genotype by environment interactions for heat tolerance in Australian Holsteins. A total of 2.5 million test-day milk yield records from 823,055 cows and 6,615 sires were included in the analysis. The heritability estimates at 5th and 95th percentile of temperature-humidity index (THI) were: 0.27 and 0.21, 0.21 and 0.14, and 0.19 and 0.14 for milk, protein and fat yield, respectively. The genetic correlations at the extreme THI values, that is THI = 60 and THI = 75 (equivalent to the temperature and relative humidity of around 20 °C and 45 and, 31 °C and 50, respectively) were: 0.87, 0.84, and 0.86 for milk, protein and fat, respectively. A re-ranking among sires was observed in different environments. These results could allow farmers to make decisions on whether to select sires which are best suited to specific environments, or those that are consistent across a range of environments.

INTRODUCTION

The desire to breed for robustness in the dairy industry is intensifying, driven in part by climate change. One of the key components of robustness is genotype by environment interactions (G × E), which refers to the change in performance or a change in the ranking of animals in different environments. In Australia, dairying is carried out in a wide range of production systems and climatic conditions suggesting that reranking of genotypes may occur.

Various studies have demonstrated the presence of G × E due to heat stress in dairy cattle as reviewed by Carabaño *et al.* (2017). Previous studies in Australia using test-day records reported evidence of G × E for production traits due to heat stress for Australian Holsteins (Hayes *et al.* 2003; Haile-Mariam *et al.* 2008). These studies used first parity or whole lactation data.

Genetic selection for production traits in Australian dairy cattle has resulted in considerable genetic gains. However, this may have led to increased sensitivity to heat stress in dairy animals (Carabaño *et al.* 2017) and possibly increased G × E because of an unfavourable genetic correlation between heat tolerance and milk production traits (Ravagnolo *et al.* 2000). Nguyen *et al.* (2017) noted a declining genetic trend for heat tolerance in Australian Holstein and Jersey dairy cattle at a rate of 0.3%/year. This declining trend coupled with increasing temperature and frequency of heat events suggests the importance of revisiting the magnitude of changes in animal performance at different environmental temperature and humidities. The objective of this study was to investigate G × E for heat tolerance using test-day milk yield records in combination with temperature and humidity data from publicly available weather stations over a 15-year period.

MATERIALS AND METHODS

Test-day data. First lactation milk, protein and fat yield data (consisting of 6.6 million records for Holstein cows between 2003 to 2017) were obtained from DataGene (DataGene Ltd., Melbourne,

Australia). Data editing was as follows: (1) tests < 5d or > 305d days in milk (DIM) and herd test days with less than 10 cows were removed; (2) sires with daughters in less than 2 herds and herds using fewer than 2 sires were excluded and (3) only cows with at least 4 records were retained for analyses. The final dataset comprised 5.2 million records for 823,055 cows and 6,615 sires from 3,732 herds. The pedigree for these data included up to 15 generations.

Climate data. Climate data included hourly dry bulb and dew point temperature and relative humidity obtained from the Bureau of Meteorology (Melbourne, Australia) for 163 weather stations in Australia from 2003 to 2017. The pairwise distances between herds were calculated from geographical coordinates and assigned to the nearest weather station. Hourly temperature-humidity indexes (THI) for each weather station were calculated as follows (Yousef 1985): $THI = T_{db} + (0.36\, T_{dp}) + 41.2)$, where T_{db} = hourly dry bulb temperature (°C); T_{dp} is dew point temperature (°C) and T_{dp} = (237.3b)/ (1.0 − b), where b = [log (RH/100.0) + (17.27T_{db})/(237.3 + T_{db})]/17.27, and RH = relative humidity. The THI values were then averaged for 24 hours to get the daily THI. The daily THI on the test day, 1, 2, 3, and 4th day before test day were then averaged and assigned to the test-day records.

Milk yield traits in Australia have been reported to begin declining at THI > 60 (Hayes *et al.* 2003; Nguyen *et al.* 2016). Therefore, the THI threshold was set at 60 in this study (i.e., if THI < 60 then THI = 60). A small proportion (0.004%) of tests obtained at THI ≥ 75 were given a value of 75. This was to avoid unexpected trajectories as possible artefacts, which are often related to fitting polynomials with few extreme data points.

Statistical analysis. A univariate random regression sire model was applied to the data as follows: $y_{ijk} = \mu + HTD_i + YS_j + \sum_{n=1}^{3} A_n X_n + \sum_{n=1}^{8} D_n Z_n + \sum_{n=1}^{2} P_n T_n + \sum_{n=0}^{1} S_{kn} W_n + e_{ijk}$, where y_{ijk} is yield of milk in litres, fat or protein in kg from the *i*th herd test day, *j*th year season of calving, and daughter of the *k*th sire; μ is the intercept; HTD_i is the effect of the *i*th herd test day; YS_j is the effect of the *j*th year season of calving; X_n, Z_n and P_n are the *n*th-order Legendre polynomials corresponding to age on day of test, DIM at test, and THI, respectively; A_n, D_n and T_n are the fixed regression coefficient of traits on age at test, DIM at test, and THI, respectively; S_{kn} is a random regression coefficient on THI for the *k*th sire; W_n is either the intercept ($n = 0$) or slope solution ($n = 1$) for heat load index (THI) for cows and sires; and e_{ijk} is the vector of residual effects. The following (co)variance structure was assumed: $Var\,(S) = \begin{bmatrix} s_0 \\ s_1 \end{bmatrix} = \begin{bmatrix} \mathbf{A}\,\sigma^2_{s_0} & \mathbf{A}\sigma_{s_0 s_1} \\ \mathbf{A}\,\sigma_{s_0 s_1} & \mathbf{A}\,\sigma^2_{s_1} \end{bmatrix}$, where **A** is the relationship matrix for sires constructed from pedigree data; s_0, s_1 are the intercept and slope for sires; $\sigma^2_{s_0}$, $\sigma_{s_0 s_1}$, $\sigma^2_{s_1}$ are (co)variance for sire effects of THI. Heterogeneous error variance was modelled for 10 DIM intervals over a lactation (DIM = 5–30, 31–60, 61–90, 91–120, 121–150, 151–180, 181–210, 211–240, 241–270, and 271–300) as follows: $Var\,(e) = \mathbf{R} = \text{diag}\{\mathbf{I}\sigma^2_{e_1}, \mathbf{I}\sigma^2_{e_2} \dots \mathbf{I}\sigma^2_{e_{10}}\}$, where $\sigma^2_{e_1}$, $\sigma^2_{e_2} \dots \sigma^2_{e_{10}}$ represents error variances and **I** is the identity matrix. (Co)variance components were estimated using ASREML version 4.2 (Gilmour *et al.* 2015).

Calculation of genetic parameters. Additive genetic variances for sires were extracted from the diagonal elements of the covariance $\hat{\mathbf{G}}$ matrix calculated as $\hat{\mathbf{G}} = 4 * \boldsymbol{\Phi} Var(\hat{\mathbf{S}}) \boldsymbol{\Phi}'$, where $\boldsymbol{\Phi}$ is the matrix of Legendre polynomial functions for THI; $\hat{\mathbf{S}}$ is the sire (co)variance matrix. The genetic correlations were obtained from transforming the covariance $\hat{\mathbf{G}}$ matrix to a correlation matrix. The heritability as a function of THI was calculated as $h^2_i = \frac{4 * \hat{\sigma}^2_{s(i)}}{\hat{\sigma}^2_{s(i)} + \hat{\sigma}^2_e}$, where $\hat{\sigma}^2_{s(i)}$ is sire variance at *i* THI and $\hat{\sigma}^2_e$ is the average residual variance over the lactation. The estimated breeding value (EBV) for the sire *i* along the THI trajectory was calculated as $EBV_i = \varnothing_j * \hat{a}'_i$, where $\hat{a}'_i$ is the vector of estimated random regression coefficients for the slope and intercept for sire *i*; $\varnothing_j$ is the vector of Legendre polynomials evaluated at THI *j*. To examine the changes in performance along the THI trajectory, we estimated EBVs for sires with more than 1000 daughters with yield records.

RESULTS AND DISCUSSION

Table 1 shows genetic variances and heritability estimates at the 5th, 50th and 95th percentiles of THI. The genetic variance and heritability estimates decrease with increasing THI values. The heritability was greater for milk yield at the 5th and 95th percentiles (0.27 and 0.21) compared to protein yield (0.21 and 0.14) and fat yield (0.19 and 0.14).

Table 1. Additive genetic variances and heritabilities for milk, fat and protein yields at the 5th, 50th and 95th percentiles of the temperature-humidity index (THI)

	Additive genetic variance			Heritability		
	5th	50th	95th	5th	50th	95th
Milk (kg)	4.55	3.86	3.54	0.27	0.23	0.21
Fat (kg)	0.005	0.004	0.003	0.19	0.17	0.14
Protein (kg)	0.004	0.003	0.002	0.21	0.17	0.14

At the extremes of the trajectory of THI (i.e., THI 60 vs 75), the genetic correlations were 0.87, 0.84, 0.86 for milk, protein and fat, respectively (Figure 1). In the previous study, Hayes *et al.* (2003) reported smaller G × E estimates for milk (0.94), protein (0.92) and fat (0.90). Greater G × E in our study is likely in part due to increased sensitivity to heat stress in study population following continued selection for production traits over the years or a slight difference between the analyses; Hayes *et al.* (2003) included a random regression coefficient on THI for cows in their models.
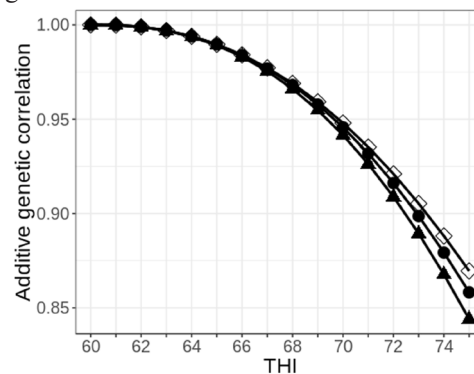

Figure 1. Additive genetic correlations for milk (□), protein (▲) and fat (●) yields at temperature-humidity index (THI) = 60 and those at THI up to 75

Reranking exists among sires, as seen from the differences in the reaction norms of EBVs for fat yield (Figure 2). Two groups of sires were identified based their EBVs at thermoneutral (THI = 60) and heat stress (THI = 75) conditions. The first group (shown in gray) are sires with above-average EBVs at THI = 60 and smaller EBVs at THI = 75 (i.e., environmentally sensitive sires). Daughters of these sires will likely produce less under heat stress conditions and therefore can be used in regions with the consistently low heat load. A more controlled environment, such as the provision of shade and diets designed to reduce core body temperature will be necessary if their daughters are to perform optimally under high heat load conditions.

The second group (Figure 2; shown in black) are sires with above-average and stable EBVs (i.e., resilient or robust sires); their performances are comparatively consistent and are well suited for variable environments. If the objective is to breed for robustness or resilience, then these sires are ideal candidates for selection. Australian dairying is predominantly pasture-based characterised by

an array of factors including weather conditions which vary considerably between years as well as seasonal variability in feed quantity and quality feeds. Under these conditions and considering current trends towards extensive exchange of sires between regions or export of sires to other countries, it would be more beneficial to select for robust sires.

This study only considered first lactation data. Greater reranking is expected with later lactations due to relatively higher sensitivity to heat stress associated with greater milk yield in multiparous cows (Carabaño *et al.* 2017). This will be investigated in further studies.
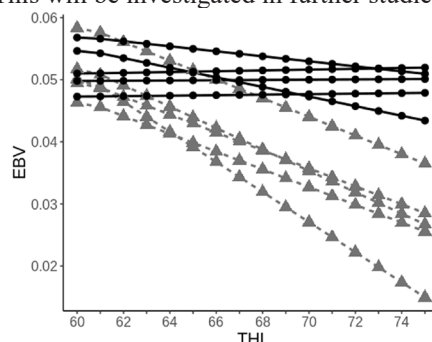
Figure 2. Estimated breeding values (reaction norms) along the THI for a sample of 10 sires with over 1000 daughters with fat yield records; the gray lines (▲) represent sires with above-average EBV at the thermoneutral conditions (THI = 60) and smaller EBV at heat stress conditions (THI = 75) whereas the black lines (●) are sires with above-average and stable EBVs

CONCLUSION

The results from this study indicate G × E due to heat stress exists at extreme THI for all the milk traits studied. The differences observed in the reaction norms (i.e., EBVs along the trajectory of THI) among the sires suggest that genetic variation in sire sensitivity to heat stress exist, which can be used to select animals that perform optimally in different environments.

REFERENCES

Carabaño M.J., Ramón M., Díaz C., Molina A., Pérez-Guzmán M.D. and Serradilla J.M. (2017) *J. Anim. Sci*. **95**: 1813.

Gilmour, A.R., Gogel, B.J., Cullis, B.R., Welham, S. and Thompson R. (2015) ASReml user guide release 4.1 structural specification. Hemel hempstead, VSN international ltd.

Haile-Mariam M., Carrick M.J. and Goddard M.E. (2008) *J. Dairy Sci.* **91**: 4840.

Hayes B.J., Carrick M., Bowman P. and Goddard M.E. (2003) *J. Dairy Sci.* **86**: 3736.

Nguyen T.T.T., Bowman P.J., Haile-Mariam M., Nieuwhof G.J., Hayes B.J. and Pryce J.E. (2017) *J. Dairy Sci.* **100**: 7362.

Nguyen T.T.T., Bowman P. J., Haile-Mariam M., Pryce J.E. and Hayes B.J. (2016) *J. Dairy Sci.* **99**: 2849.

Ravagnolo O., Misztal I. and Hoogenboom G. (2000) *J Dairy Sci* **83**: 2120.

Yousef, M.K. (1985) Stress physiology in livestock. Volume I. Basic principles. CRC press.

Appendix 2: "**Identification and validation of candidate genes for heat tolerance in Australian Holstein dairy cattle.**" **In** *Annual Meeting of the American Dairy Science Association*: **Journal of Dairy Science. Vol. 103, pp. 61-6. 2020.**

Please note that this conference article has been formatted following the style of the conference proceeding

Identification and validation of candidate genes for heat tolerance in Australian Holstein dairy cattle

E. K. Cheruiyot[1,2], M. Haile-Mariam[1], B. G. Cocks[1,2], I. MacLeod[1], and J. E. Pryce[1,2]

[1]Agriculture Victoria Research, Department of Jobs, Precincts and Regions, Bundoora, VIC 3083, Australia

[2]School of Applied Systems Biology, La Trobe University, Bundoora, VIC 3083, Australia

Heat tolerance is the ability of an animal to maintain production and reproduction levels under hot conditions and is now a trait of economic relevance in dairy systems worldwide. Understanding the genetic basis for heat tolerance is an important part of the strategy to breed for cattle adapted to warmer environments. We sought to identify and validate candidate genes involved in heat tolerance in Australian Holstein cattle using 50K SNP chip and imputed whole-genome sequences (WGS) in two datasets: cows (n = 20,623) and bulls (n = 1,622). The WGS was imputed using Run 7 of the 1000 Bull Genome Project based on the ARS-UCD1.2 reference genome. After quality checks, approximately 45K and 15 million SNPs remained for analysis from the 50K chip and the WGS, respectively. The heat tolerance phenotypes that were used for this study were derived from test-day milk, protein and fat yield data of cows that calved between 2003 to 2017 and represent the response of cows to heat stress. For the three milk traits, the slope which measures change in milk yield due to variability in temperature-humidity index from reaction norm models calculated for individual animals after accounting for known fixed effect was used as phenotype in association analysis using GCTA software. We used cows for discovery and bulls that were not sires of the cows for validation. Our results point to the polygenic nature of heat tolerance, with no variants surpassing a significance threshold of $p < 1.0 \times 10^{-6}$ for 50K SNP data. While we detected some associations when the threshold was set at $p < 1.0 \times 10^{-3}$, there were high false discovery rates. In contrast, we detected multiple variants with $p < 1.0 \times 10^{-6}$ across the genome using WGS, suggesting greater power compared to 50K SNP data. Across our cow and bull validation set, we confirmed several candidate genes, including some that have been reported previously. Our results are preliminary and could be useful to enhance the reliability of heat tolerance genomic breeding values of Australian Holstein cattle, which is currently about 38%. This will be investigated in a further study.

Key words: heat tolerance, SNPs, dairy cattle, WGS, milk yields

Appendix 3: "Using selected sequence variants to improve genomic prediction of heat tolerance in dairy cattle." In *Proceedings for the Association Advancement of Animal Breeding Genetics*. Vol. 24, pp. 50-53. 2021.

Please note that this conference manuscript has been formatted in the style of the conference proceeding.

USING SELECTED SEQUENCE VARIANTS TO IMPROVE GENOMIC PREDICTION OF HEAT TOLERANCE IN DAIRY CATTLE

E.K. Cheruiyot[1,2], M. Haile-Mariam[1], B.G. Cocks[1,2], I.M. MacLeod[1] and J.E. Pryce[1,2]

[1]Agriculture Victoria, Department of Jobs, Precincts and Regions, Bundoora, VIC 3083, Australia
[2]School of Applied Systems Biology, La Trobe University, Bundoora, VIC 3083, Australia

SUMMARY

Genomic breeding values for heat tolerance were first developed and released to the Australian dairy industry in 2017, to allow farmers to select animals that better tolerate hot and humid conditions. It is desirable to improve the reliability of these genomic predictions to help accelerate the genetic improvement for this trait. Whole-genome sequence data may contain causative mutations, or variants in high linkage disequilibrium with causal mutations for traits. This study investigated the potential improvements in the accuracy of genomic prediction for heat tolerance when adding informative markers to the 50k industry SNP panel used routinely by DataGene for Australian dairy genomic evaluations. We selected informative sequence variants from a genome-wide association study (GWAS) of heat tolerance phenotypes of 20,623 Holstein cows (each cow with ~15 million imputed sequence variants) and augmented the 50k SNP panel with these SNPs for genomic prediction using a Holstein bull reference (N = 3,323) and Holstein cow validation set (N = 8,484). The accuracy of genomic prediction of heat tolerance for reduction in milk, fat, and protein yield under hot and humid conditions increased by 0.1%, 4%, and 6% units, respectively when informative markers were integrated with 50k SNP data. Since genetic gain is linearly related to EBV accuracy, this lift in accuracy is important for driving the genetic improvement of heat tolerance.

INTRODUCTION

Heat tolerance is the ability of an animal to maintain production and reproductive performance under hot and humid conditions. The desire to breed for heat tolerance is growing worldwide due to the increasing effect of global warming on animal production. Considerable research has been conducted so far in many countries, including Australia, where the first breeding values for heat tolerance were released to the dairy industry in 2017 (Nguyen *et al*. 2017).

Since genetic gain is linearly related to the accuracy of estimated breeding values (EBVs), even a small lift in the accuracy of the heat tolerance EBV is important to the dairy industry. Besides increasing the size of the reference population, one way to boost the accuracy is to increase the density of markers used for genomic predictions. However, increasing the marker set from lower density SNP panels to whole-genome sequence have, in most cases, yielded limited, or no appreciable increase in the accuracies for various traits in cattle (e.g., Van Binsbergen *et al*. 2015). A promising alternative, in which a boost of accuracy has been realized in previous reports (e.g., Moghaddar *et al*. 2019), has been to augment standard industry SNP panels (i.e., 50k or 600K arrays) with a small set of informative or causal mutations for a trait. To fully maximize predictions, this approach requires careful selection of informative markers. Thanks to the 1000 Bull Genomes project (Hayes and Daetwyler 2019), it is now possible to use this sequence database to impute genotyped animals up to whole genome sequence. This may facilitate accurate selection of highly informative variants for use in genomic predictions, especially for complex traits such as heat tolerance.

In this study, we selected informative variants for heat tolerance from a genome-wide association study (GWAS) using milk production records of 20,623 Holstein cows, each having over 15 million

imputed sequence variants. We then investigated the accuracy of prediction when sets of these selected variants were added to the standard industry 50k SNP array, by training the prediction in a bull reference set, and validating it in an independent set of Holstein cows.

MATERIALS AND METHODS

Phenotypes. The phenotypes used in this study were obtained from DataGene (DataGene Ltd., Melbourne, Australia; https://datagene.com.au/) and included test-day milk, fat, and protein yields for Holstein cows and bulls, collected from dairy herds between 2003 and 2017 that were matched with climate data (daily temperature and humidity) obtained from weather stations across Australia's dairying regions. The distribution of dairy herds and weather stations, data filtering, and the calculation of environmental covariate (i.e., temperature-humidity index or **THI**) used in this work were described in our earlier studies (Nguyen *et al.* 2016, Cheruiyot *et al.* 2020).

Calculation of heat tolerance phenotypes for cows and bulls. The rate of decline (slope) in milk, fat, and protein yield due to heat stress events was estimated using reaction norm models as described by Cheruiyot *et al.* 2020. In these models, data on milk, fat, or protein yield were adjusted for fixed effects, including herd test day, year season of calving, parity, age at calving, jointly for parity and DIM, and jointly for stage of lactation and THI. Random effects fitted in the model included a random regression on a linear orthogonal polynomial of THI, where the intercept represents the level of mean milk yield and the linear component represents the change in milk yield (slope) due to heat stress for each cow (i.e., trait deviations (**TD**)) and a residual term. Slope solutions for each bull's daughters were averaged to obtain slope traits for bulls (i.e., daughter trait deviations (**DTD**)).

Genotypes and study design. Two genotype data sets were available: 50k SNP array and ~15 million imputed whole-genome sequence variants. The number of Holstein animals with genotypes and heat tolerance phenotypes were 29,107 ♀/3,323 ♂. We split the Holstein cows into two: 1) QTL discovery set (N = 20,623; comprising older cows born before 2013) for selecting informative markers for heat tolerance, and 2) genomic prediction validation set (N = 1,223; young cows born after 2012). We used Holstein bulls as a training set for genomic prediction. We ensured that none of the cows in the QTL discovery set were daughters of the bulls in the training set to avoid parent-daughter pairs between the two datasets to minimise close genetic relationships.

QTL discovery analysis and selection of informative SNPs. We performed single-trait GWAS analysis to test associations between individual SNP and cows' slope traits (milk, fat, and protein) using GCTA software (Yang *et al.* 2011). The models used for analyses are described by Cheruiyot *et al.* (https://www.biorxiv.org/content/10.1101/2021.02.04.429719v1.full).

Following the GWAS, we selected informative variants defined as 'top SNPs' for each slope trait as follows: for SNPs passing the GWAS threshold of $-log_{10}(p\ value) = 2$; we chose the most significant SNP from within each 100 kb window and sliding 50 kb to the next window along each chromosome. We then removed one SNP of any pair of the selected 'top SNPs' in strong LD ($r^2 > 0.95$).

Genomic prediction. We used BayesR (Erbe *et al.* 2012) to estimate prediction accuracies for 50k SNP panel and compared the resulting accuracies with those obtained from adding 'top SNPs' to the 50k SNP set (i.e., 50k + 'top SNPs') using BayesRC method (MacLeod *et al.* 2016). The BayesR model fitted to the training bulls (N = 3,323) for 42,572 variants from 50k SNP panel was: $y = X\beta + Zg + Wv + e$, where **y** = vector of heat tolerance slope phenotypes; **X** = design matrix; **β** = vector of fixed effect solutions; **Z** = design matrix relating phenotypes to GBV; g = vector of GBV $\sim N(0, I\sigma_g^2)$, where σ_g^2 is the additive genetic variance for the trait; **W** = design matrix of SNP genotypes; **v** = vector of SNP effects, modelled to have four possible normal distributions corresponding to zero, small, medium and large effects, respectively; **e** = vector of residual errors

$N(0, \mathbf{E}\sigma_e^2)$, where $\mathbf{E}$ is a diagonal matrix calculated as $diag(1/w_i)$, with w_i being a weighting factor for ith sire calculated based on the available number records following Garrick *et al.* (2009).

We then used the BayesRC method to analyse 50k + 'top SNPs' dataset; an extension of the BayesR model that allows pre-allocation of variants to 2 or more classes (MacLeod et al., 2016) and hence a different posterior mixture distribution within each class if the class is enriched for informative SNPs. In our case the SNPs from 50k array (42,572) were allocated to class I and the selected 'top SNPs' to a separate class II, because the latter may be enriched with causal mutations for heat tolerance. For both BayesR and BayesRC models, we performed five MCMC replicate chains, each with 40,000 iterations of which 20,000 were discarded as burn-in for all the traits. We ran the analysis for 2 random validation sets of 600, and 623 Holstein cows.

Calculating accuracy of genomic prediction. For each of the three validation cow sets (described above), the accuracy of prediction was calculated as: $Accuracy(Val_i) = \frac{r_{GBV,phen}}{\sqrt{h^2}}$, where Val_i = Holstein cow validation set; $r_{GBV}, phen$ = correlation of GBV and phenotypes (i.e., slope traits); h^2 = genomic heritability calculated for each trait using variance component estimates of Holstein cows (N = 29,107) for 50k SNP array (45,504 SNPs) data based on –reml option of GCTA software (Yang *et al*. 2011). The bias of prediction was assessed as the regression coefficient of the phenotypes (pre-corrected for fixed effects) on the GBV for animals in the validation set.

RESULTS AND DISCUSSION

In this study, we used a large dataset of Holstein cows (N = 20,623) to select informative markers from a GWAS and then tested them for increased genomic prediction of heat tolerance phenotypes.

The genomic heritability estimates (± standard errors) for the heat tolerance milk (**HTMYslope**), fat (**HTFYslope**) and protein (**HTPYslope**) yield slope traits from Holstein cows that used to calculate the accuracy of predictions were 0.23 ± 0.01, 0.21 ± 0.01, and 0.20 ± 0.01, respectively. The number of informative markers for heat tolerance (i.e., 'top SNPs') selected from GWAS (p < 0.01) was highest for HTPYslope (9,633) followed by HTFYslope (9,352), and HTMYslope (9,207) traits. Similarly, the total number of markers used in the BayesRC analyses (i.e., 50k + top SNPs) were 51,750, 51,894, 52,168, for HTMYslope, HTFYslope and HTPYslope traits, respectively. We chose a cut-off of p < 0.01, which is comparatively relaxed, to capture both markers with small and large effect sizes for heat tolerance.

Figure 1 shows the accuracy and bias of genomic predictions in the Holstein validation cows. For the BayesR model using only 50k SNP data, we found the highest accuracy of prediction for HTFYslope (0.49 ± 0.01), followed by HTMYslope (0.49 ± 0.01) and HTPYslope (0.39 ± 0.01). The bias across all study traits was > 1.0 (Figure 1) indicating 'deflation' or under prediction, meaning less variance among predicted than observed values.

When the selected 'top SNPs' were added to the standard 50k SNP array and analysed using the BayesRC model, we found a consistent increase in the prediction accuracy across all the traits with values of 0.001, 0.04, and 0.06 for HTMYslope, HTFYslope and HTPYslope traits, respectively (Figure 1). This increase in accuracy is notable for HTFYslope and HTFYslope traits and likely to be associated with the pre-selected markers (potentially functionally linked with heat tolerance) and the method used (BayesRC). The bias of prediction for BayesRC was comparable that for BayesR. In this study, we investigated the potential benefits of sequence variants selected from a single breed (Holsteins) on the accuracy of genomic predictions for the same breed (within-breed prediction). The value of sequence variants selected in across-breed population (combined Holsteins and Jersey) on genomic prediction of other breeds (Jersey and crossbred cattle) will be investigated in a further study.
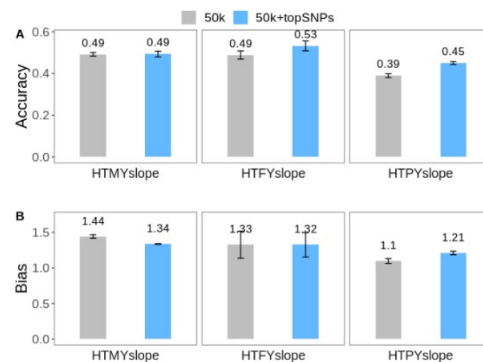
Figure 1. Accuracy of genomic prediction in Holsteins cows based on BayesR (50k; grey) and BayesRC (50k+topSNPs; blue) models for heat tolerance milk (HTMYslope), fat (HTFYslope), and protein (HTPYslope) yield slope traits. Vertical lines are the standard errors of prediction estimated from 2 random validation sets of 600, and 623 Holstein cows

CONCLUSION

Overall, our results show that the accuracy of genomic prediction for reduction in milk, fat, and protein yields under hot and humid conditions can be improved by 0.1%, 4%, and 6% units, respectively when selected informative sequence variants are added to the industry-implemented 50k SNP panel.

ACKNOWLEDGEMENTS

REFERENCES

Cheruiyot, E. K., T. T. T. Nguyen, M. Haile-Mariam, B. G. Cocks, M. Abdelsayed, and J. E. Pryce. (2020) *J Dairy Sci* **103**:2460.

Erbe, M., B. Hayes, L. Matukumalli, S. Goswami, P. Bowman, C. Reich, B. Mason, and M. Goddard. (2012) *J Dairy Sci* **95**:4114.

Garrick, D. J., J. F. Taylor, and R. L. Fernando. (2009) *Genet Sel Evol* **41**:55.

Gilmour, A., B. Gogel, B. Cullis, S. Welham, and R. Thompson. (2015) ASReml user guide release 4.1 structural specification. Hemel hempstead: VSN International Ltd.

Hayes, B. J. and H. D. Daetwyler. (2019) *Annu Rev Anim Biosci* **7**:89.

MacLeod, I., P. Bowman, C. Vander Jagt, M. Haile-Mariam, K. Kemper, A. Chamberlain, C. Schrooten, B. Hayes, and M. Goddard. (2016) *BMC genomics* **17**:144.

Moghaddar, N., M. Khansefid, J. H. van der Werf, S. Bolormaa, N. Duijvesteijn, S. A. Clark, A. A. Swan, H. D. Daetwyler, and I. M. MacLeod. (2019) *Genet Sel Evol* **51**:72.

Nguyen, T. T., P. J. Bowman, M. Haile-Mariam, J. E. Pryce, and B. J. Hayes. (2016) *J Dairy Sci* **99**:2849.

Nguyen, T. T., P. J. Bowman, M. Haile-Mariam, G. J. Nieuwhof, B. J. Hayes, and J. E. Pryce. (2017) *J Dairy Sci* **100**:7362.

Yang, J., S. H. Lee, M. E. Goddard, and P. M. Visscher. (2011) *Am J Hum Genet* **88**:76.

Van Binsbergen, R., M. P. Calus, M. C. Bink, F. A. van Eeuwijk, C. Schrooten, and R. F. Veerkamp. (2015) *Genet Sel Evol* **47**:1.