Comparing the methylomes of two genealogically linked Russian wheat aphid biotypes using whole genome bisulfite sequencing

by

Pieter Herodus du Preez

Thesis presented in partial fulfilment of the requirements for the degree of Magister Scientiae

at

Stellenbosch University

Department of Genetics, Faculty of Natural Sciences

Supervisor: Professor Anna-Maria Botha-Oberholster

Co-supervisor: Mister Nicolaas Francois Visser Burger

March 2020

Stellenbosch University https://scholar.sun.ac.za

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: March 2020

Copyright © 2020 Stellenbosch University
All rights reserved

Stellenbosch University https://scholar.sun.ac.za

Abstract

Diuraphis noxia (Kurdjumov, Hemiptera: Aphididae), commonly known as the Russian wheat aphid (RWA), is an economically important cereal pest. Although it does not spread any plant viruses, the severe symptoms caused by RWA feeding poses a significant threat to world wheat production. Commercial wheat cultivars, resistant to RWA, have been developed and are effective at preventing yield losses. However, new, more virulent RWA biotypes (morphologically similar aphid populations, with the ability to successfully feed on previously resistant cultivars) are continuously emerging, leading to a breakdown in resistance. The molecular mechanisms driving biotypification (the development of new biotypes) have not been identified yet. It has been proposed that an epigenetic modification, such as DNA methylation, might a possible means whereby biotypification might occur. The aim of this study was to explore the possible link between levels of DNA methylation and virulence in the RWA, by performing a whole genome bisulfite sequencing (WGBS) analyses on the South African biotypes SA1 and SAM. Together, SA1 and SAM form a good model for the study of virulence as they are closely related, yet at opposite ends of the virulence scale. The overall trends in RWA DNA methylation, observed in this study, correlates with what has previously been reported in insects: genic bodies, especially exons, are the most methylated regions in the genome, with most of the methylation occurring at CpG sites. The ratio of observed to expected CpG sites in a region has been used to infer levels of methylation, as increased methylation has been correlated to a decrease in CpG abundancy; In this study, however, no correlation was found between CpG abundancy and DNA methylation level. This technique, therefore, is not applicable for insect genes. Using the generated WGBS data, 148 genes were found to be differentially methylated between the two biotypes. The relative expression of five of these genes, which were selected based on gene ontology and the degree of differential methylation, along with that of DNA methyltransferase 3 (*DNMT3*) and ten-eleven translocation enzyme (*TET*), was quantified and compared between biotypes at 0, 6, and 48-hours after performing host-shifts from susceptible to resistant host plants. DNMT3 is the enzyme responsible for the establishment of DNA methylation, while the TET enzyme catalyses the first step of the process of demethylation. The time points were selected to correlate with defence responses elicited from the host plants. While no clear pattern could be observed in the differences in relative expression of differentially methylated genes between biotypes, or within a biotype between time points, a major increase in expression of both *DNMT3* and *TET* 6-hours after performing a host shift to the resistant cultivar Tugela *Dn5* was observed in SAM. This seems to indicate an increased rate of methylation, demethylation, or both methylation and demethylation in SAM, while under stress, which might contribute to its increased ability to respond to, and overcome, plant defence responses.

Stellenbosch University https://scholar.sun.ac.za

Opsomming

Diuraphis noxia (Kurdjumov, Hemiptera: Aphididae), algemeen bekend as die Russiese koringluis (RKL), is ‘n ekonomies relevante graanpes. Alhoewel dit geen plantvirusse versprei nie, bedreig die ernstige simptome van RKL-voeding steeds wêreld graanproduksie. Kommersiële koring kultivars met RKL-weerstand bestaan en is effektief in die verhoeding van opbrengsverliese. Nuwe, meer virulente RKL-biotipes (populasies wat morfologies identies is, maar wat suksesvol kan voed op voorheen weerstandige kultivars) word egter ewig aangetref end dit lei tot ‘n afbreek in weerstand. Die molekulêre meganismes wat biotipifisering (die ontwikkeling van nuwe biotipes) dryf, is nog nie geïdentifiseer nie. Daar is voorgestel dat ‘n epigenetiese modifisering, soos DNA-metilering, ‘n moontlike fasiliteerder van biotipifisering mag wees. Die doel van hierdie studie was om die moontlike verhouding tussen die vlakke van DNA-metilering en virulensie in die RKL te ondersoek, deur heel genoom bisulfiet volgorde bepaling op die Suid-Afrikaanse biotipes SA1 en SAM toe te pas.SA1 en SAM vorm saam ‘n goeie model vir die bestudering van virulensie, aangesien hulle naby verwant is, maar tog op teenoorgestelde ente van die virulensieskaal val. Die algehele eienskappe van DNA-metilering in die RKL stem ooreen met wat tot dusver in insekte gerapporteer is: die hoogste vlakke van metilering word in geniese streke, veral in eksone, aangetref en meeste van die metilering verskyn by CpG setels. Die verhouding tussen die verwagte en waargenome hoeveelheid CpG setels in ‘n streek word soms gebruik vir ‘n beraming van metilering vlakke, aangesien ‘n toename in metilering al gekorreleer is met ‘n afname in CpG-rykheid; in hierdie studie is geen korrelasie egter tussen die vlakke van DNA-metilering en CpG-rykheid waargeneem nie. Dus blyk dit nie of die hierdie tegniek bruikbaar is vir insekgene nie. Deur gebruik te maak van die gegenereerde volgorde data is 148 gene geïdentifiseer as differensieel gemetileer tussen die twee biotipes. Die relatiewe ekspressie van vyf van hierdie gene, geselekteer gebaseer op ontologie en die graad van metilering differensiasie, sowel as die van ‘DNA methyltransferase 3’ (DNMT3) en ‘ten-eleven translocation enzyme (TET) is gekwantifiseer 0, 6, en 48-uur na ‘n gasheerwisseling van ‘n vatbare, tot ‘n weerstandige plant. DNMT3 is die ensiem wat verantwoordelik is vir die stigting van DNA-metileringspatrone, terwyl die TET ensiem die eerste stap in die proses van demetilering kataliseer. Die tydstippe is gekies om ooreen te stem met die verdedigingsresponse van die gasheerplant. Geen patroon is waargeneem in die verskil in relatiewe ekspressie van differensieel gemetileerde gene tussen biotipes, of vir een biotipe tussen tydstippe nie. Daar is egter ‘n drastiese verhoging in die expressive van beide DNMT3 en TET waargeneem in biotipe SAM 6-uur na ‘n verskuiwing na Tugela Dn5. Dit dui moontlik op ‘n verhoogde tempo van metilering, demetilering, of beide metilering en demetilering in SAM gedurende stres, wat moontlik bydra tot SAM se verhoogde vermoë om te reageer op plantverdedigings en dit te oorkom.

Stellenbosch University https://scholar.sun.ac.za

Acknowledgements

I would like to express my deepest and sincerest gratitude toward the following people and organizations, without whom this dissertation would probably never have been completed.

First and foremost, to Professor Anna-Maria Botha-Oberholster, for being a kind and understanding, yet driving mentor, as well as for funding in the form of an NRF grand holder-linked bursary. Thank you for keeping me on my toes and pressuring me to perform, not the way I wanted, but the way you knew I could.

To my co-supervisor Francois Burger, for all your help in the laboratory and the growth rooms and for all you do, outside of your job description. Thank you for the support and morale boosting you and Doctor Vittorio Nicolis provided.

To everyone else in the Cereal Genomics group, for providing a professional, productive environment, when appropriate, and a fun and stimulating one the rest of the time.

To Stellenbosch University and especially the Department of Genetics, for providing the infrastructure and a friendly, welcoming environment in which to conduct this study.

To my mother, for whom I have been a continues source of worry and fret, throughout my education. Thank you for everything you do and sacrifice for me, I do not say it nearly enough.

To my fiancé, Lizé-Mari Montgomery, for keeping me sane with your continued support. Thank you for forcing me to grow as a person, even when I did not want to, and for doing the right thing, even when everything else is going wrong.

This dissertation is dedicated to my grandfather (Hercules Albertus du Preez, 1931 – 2019), who told me countless times: "Study. Study as long as you have the capacity to do so; no one can ever take away your education."

Stellenbosch University https://scholar.sun.ac.za

Preface

The findings obtained and presented in this dissertation are the outcomes of a study conducted between January 2018 and December 2019 under the co-supervision of Professor Anna-Maria Botha and Mister Nicolaas Francois Visser Burger, in the Department of Genetics at Stellenbosch University.

Stellenbosch University https://scholar.sun.ac.za

Table of contents

Stellenbosch University https://scholar.sun.ac.za

Stellenbosch University https://scholar.sun.ac.za

List of figures

Stellenbosch University https://scholar.sun.ac.za

Stellenbosch University https://scholar.sun.ac.za

Stellenbosch University https://scholar.sun.ac.za

Stellenbosch University https://scholar.sun.ac.za

List of tables

Stellenbosch University https://scholar.sun.ac.za

List of abbreviations

µM	Micromolar
3'	Downstream
5'	Upstream
5caC	5-carboxylcytosine
5fC	5-formylcytosine
5ghmC	5-glucosylhydroxymethylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
A	Adenine

Stellenbosch University https://scholar.sun.ac.za

AFLP	Amplified fragment length polymorphisms
Asp	Asparagine
ATP	Adenosine triphosphate
BER	Base excision repair
BMV	Brome mosaic virus
bp	Base pairs
BYDV	Barley yellow dwarf virus
C	Cytosine
cDNA	Complimentary deoxyribonucleic acid
CpG	Dinucleotide: cytosine followed by guanine
CTCF	CCCTC-binding factor
DBM/OVS	Difference between mean
df	Degrees of freedom
Dn	*Diuraphis noxia*
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
DML	Differentially methylated locus
DMR	Differentially methylated regions
DMG	Differentially methylated gene
DSS	Dispersion shrinkage for sequencing data
ELISA	Enzyme-linked immunosorbent assay
et al.	*et alia* (and others)
G	Guanine
GTPase	Guanosine triphosphatase
h	Hours
H	any non-guanine base
i.e.	*id est* (that is)
kg	Kilogram
MeCP2	Methyl-CpG-binding protein 2
Me-DIP	Methylated DNA immunoprecipitation
mM	Millimolar
mRNA	Messenger ribonucleic acid
MSAP	Methylation sensitive amplified polymorphism
NA	Not applicable

Stellenbosch University https://scholar.sun.ac.za

O/E	Ratio of observed over expected sites
PE	Paired-ended
p-value	Probability value
Q	Phred quality score
R	Resistant
RAPD	random amplified polymorphic DNA
RFLP	Restriction fragment length polymorphism
RNA	Ribonucleic acid
RSSFL	Restriction site-specific fluorescent labelling
RWA	Russian wheat aphid
S	Susceptible
SA1 *or* RWASA1	South African Russian wheat aphid biotype 1
SA2 *or* RWASA2	South African Russian wheat aphid biotype 2
SA3 *or* RWASA3	South African Russian wheat aphid biotype 3
SA4 *or* RWASA4	South African Russian wheat aphid biotype 4
SA5 *or* RWASA5	South African Russian wheat aphid biotype 5
SAM	South African mutant Russian wheat aphid biotype
SE	Single-ended
SNP	Single nucleotide polymorphism
SSR	simple short repeats
T	Thiamine
Tab-seq	Ten-elven translocation enzyme-assisted bisulfite sequencing
TET	Ten-eleven-translocation enzyme
TpG	Dinucleotide: thiamine followed by guanine
tRNA	Transfer ribonucleic acid
US1 *or* RWA1	American Russian wheat aphid biotype 1
US2 *or* RWA2	American Russian wheat aphid biotype 2
US3 *or* RWA3	American Russian wheat aphid biotype 3
US4 *or* RWA4	American Russian wheat aphid biotype 4
US5 *or* RWA5	American Russian wheat aphid biotype 5
US6 *or* RWA6	American Russian wheat aphid biotype 6
US7 *or* RWA7	American Russian wheat aphid biotype 7
US8 *or* RWA8	American Russian wheat aphid biotype 8
USA *or* US	United States of America

Stellenbosch University https://scholar.sun.ac.za

w/w	weight per weight
WGBS	Whole genome bisulfite sequencing
x	Times
X^2	Chi-square
%	Percent
°C	Degrees Celsius

Stellenbosch University https://scholar.sun.ac.za

Chapter 1: Introduction

Wheat is an economically important food crop, with the average person consuming 66.7 kg every year (Food and Agriculture Organization of the United Nations, 2018). Wheat production is however under threat by the small, green, cereal pest, *Diuraphis noxia* (Kurdjumov, Hemiptera: Aphididea) commonly known as the Russian wheat aphid (RWA). Unlike many other aphids, the RWA is not known to transmit any plant viruses (Fouché et al., 1984). Symptoms from RWA feeding (which include: necrotic spots, chlorotic streaking, leaf rolling, and head trapping) are severe and can lead to significant yield reductions, of up to 84%, even at initial infestation levels as low as 2.7% (du Toit and Walters, 1981). The United States of America (USA) and South Africa have been the two countries most affected by RWA infestation (Basky, 2003; Botha et al., 2005).

Sources of resistance were found independently in these countries, today, these are now known as *Diuraphis noxia* resistant genes (*Dn*-genes). Fourteen such genes have been described: *Dn1* to *Dn9*, *Dnx* and *Dny* (Botha et al., 2005; Jankielsohn, 2011), as well as *Dn2414, Dn626580* and *Dn2401* (Fazel-Najafabadi et al., 2015; Peng et al., 2007; Valdez et al., 2012). Many of these genes have been incorporated into commercial wheat cultivars (du Toit, 1989; Nkongolo et al., 1991; Quick et al., 1996). However, new biotypes, morphologically similar aphid populations that can successfully feed on previously resistant wheat cultivars, have emerged and are once again a threat to wheat production. A total of eight US biotypes (Haley et al., 2004; Randolph et al., 2009; Weiland et al., 2008) and five South African biotypes, the latest of which only emerged early in 2019 (Jankielsohn, 2019, 2011; Tolmay et al., 2007), have been reported. The only resistance gene which currently deters all naturally occurring biotypes is *Dn7* (Haley et al., 2004; Jankielsohn, 2019). Wheat cultivars containing the *Dn7* gene unfortunately delivers sticky dough with reduced strength and an intolerance to overmixing, this is attributed to the co-translocation of the gene encoding for secalin, a highly undesirable protein in wheat, from the rye R1S chromosome. Due to these poor dough traits, *Dn7* is not useful in commercial cultivars (Dhaliwal et al., 1987).

With the systematic breakdown of resistance to the RWA, it is becoming increasingly important to understand the mechanisms that enable the development of new, more resistant biotypes (Botha, 2013). Studies into biotypification are hampered however, as the genealogy of the naturally occurring biotypes is unknown. It is therefore unclear whether each new biotype developed directly from known, local biotypes, or whether they might have been co-introduced (Shufran and Payton Miller, 2009; van Zyl and Botha, 2008). To overcome this, a laboratory contained biotype, SAM (South African mutant), was selective bred from biotype SA1. This was achieved by maintaining

Stellenbosch University https://scholar.sun.ac.za

biotype SA1 on wheat containing the *Dn1* resistance gene. Continuous selective pressure exerted over the course of 78 generations (van Zyl and Botha, 2008) lead to the development of the SAM biotype which is hypervirulent and undeterred all of the known RWA resistance genes. Together, biotypes SA1 and SAM serve as a useful model to study biotypification and virulence, given their position on opposite extremes of the virulence scale, as well as the known genealogy between the two biotypes (Breeds et al., 2018; van Zyl and Botha, 2008).

Studies using various techniques, such as random amplified polymorphic DNA (RAPD), mitochondrial sequencing and simple short repeats(SSR) have shown very little genetic variation within and between RWA biotypes (Shufran and Payton Miller, 2009; Weiland et al., 2008; Weng et al., 2007). Biotypes SA1 and SAM show a genetic variation as low as 0.0008% in coding regions, according to single nucleotide polymorphism (SNP) analyses (Burger and Botha, 2017). Given the discrepancy between the genetic variation and virulence in RWA biotypes, an epigenetic modification might better explain the differential virulence between biotypes (Breeds et al., 2018). Epigenetic modifications can effect a heritable change in phenotype, without altering the DNA sequence, by affecting gene expression (Russo et al., 1996). One of these modifications, DNA methylation, has been observed in aphids, including RWA, and is believed to influence the development of new biotypes (Breeds et al., 2018; Pasquier et al., 2014).

The link between DNA methylation and virulence was first proposed when it was found that putative effector proteins, proteins excreted through the salivary glands of aphids, were differentially methylated between two US biotypes (Gong et al., 2012). This prompted an investigation into the methylation differences between the South African biotypes SA1 and SAM (Breeds et al., 2018). While the study showed promising results, the methylation sensitive amplified polymorphism (MSAP) and restriction site-specific fluorescent labelling (RSSFL) techniques employed did not provide the necessary positional information to identify differentially methylated loci. For this application, a technique such as whole genome bisulfite sequencing (WGBS), which delivers methylation data on a per base pair resolution would be required (Li and Tollefsbol, 2011).

The aims of the current study were, therefore, to first, provide valuable data for the study of DNA methylation in RWA and secondly, to evaluate the proposed link between the levels of methylation and gene expression. To this end, the objectives for Chapter 3 were to obtain the WGBS data for both biotypes and to align it to a reference genome of biotype SAM (GenBank ID: GCA_001465515.1, BioProject: PRJNA297165), with a good coverage and high quality. While the objectives for Chapter 4 were to use the WGBS data to identify differentially methylated genes between the biotypes and to quantify the relative expression of these genes, at 6-hours and 48-hours (coinciding with the

Stellenbosch University https://scholar.sun.ac.za

Hypersensitive response and Systemic acquired resistance) (Burger et al., 2017) after performing host shifts from a susceptible to a resistant host plant, in order to investigate the influence of the levels of DNA methylation on gene expression in RWA. As well as quantifying the relative expression of two genes, DNA methyl transferase 3 (*DNMT3*) and ten-eleven translocase (*TET*), responsible, respectively, for the addition and removal of DNA methylation, and lastly, to investigating the validity of using the ratio of observed to expected CpG sites ($CpG_{O/E}$) to infer methylation levels.

1.1: Thesis layout

Chapter 2 provides background information on the RWA from current literature regarding its morphology, the feeding symptoms presented by host plants, known biotypes, and current methods of control. The current knowledge on DNA methylation is also relayed, as it is contextually important for the research presented in Chapter 3 and Chapter 4.

Chapter 3 details the WGBS sequencing of RWA biotypes SAM and SA1, as well as the steps and programs used to extract the methylation information from the WGBS reads.

Appendix A contains all the commands used to process the outputs of the Bismark pipeline into the tables presented in Chapter 3.

In Chapter 4 the WGBS data is used to identify differentially methylated genes between the biotypes. The relative expression levels of a subset of these genes, along with that of key methylation genes were quantified. The validity of a technique used to infer methylation levels from sequence information is also evaluated.

Appendix B contains the outputs from DSS-single (used to identify differential methylated genes) and Blast2GO (used to assign gene ontologies based on similarity).

Chapter 5 is a summary of the main findings of the current study, as well as recommendations for future research into DNA methylation and RWA virulence

1.2: Research outputs

All raw and processed data files were uploaded to the NCBI's GEO database, under the accession number: GSE119504

du Preez, P.H., Burger, N.F.V., Botha, A-M., 2018 Investigating the methylome landscape of two genealogy-linked *Diuraphis noxia* biotypes, at the 23rd Biannual Plant Resistance to Insects Workshop, 7-9 March, Rothamsted Research, United Kingdom, an oral presentation by P.H. du Preez.

Stellenbosch University https://scholar.sun.ac.za

du Preez, P.H., Burger, N.F.V., Botha, A-M., 2018 Investigating the methylome landscape of two genealogy-linked *Diuraphis noxia* biotypes, at the Legumes for Life Trilateral Workshop, 2-4 May, Stellenbosch University, South Africa, an oral presentation by P.H. du Preez.

du Preez, P.H., Burger, N.F.V., Botha, A-M., 2018 Comparing the methylomes of two genealogically linked Russian wheat aphid biotypes, at the joint South African Genetics Society and South African Society of Bioinformatics Congress, 16-18 October, Golden Gate Resort, South Africa, a poster presentation by P.H. du Preez

du Preez, P.H., Breeds, K., Burger, N.F.V., Burger, Swiegers, H.W., Truter, C., Botha, A-M., *In press* DNA methylation and demethylation are regulated by functional DNA methyltransferases and DnTET enzymes in *Diuraphis noxia*

1.3: References

Basky, Z., 2003, Biotypic and pest status differences between Hungarian and South African populations of Russian wheat aphid, *Diuraphis noxia* (Kurdjumov) (Homoptera: Aphididae), *Pest Management Science*, 59(10), 1152–1158.

Botha, A-M., Li, Y. & Lapitan, N.L.V., 2005, Cereal host interactions with Russian wheat aphid: A review, *Journal of Plant Interactions*, 1(4), 211–222.

Botha, A-M., 2013, A coevolutionary conundrum: The arms race between *Diuraphis noxia* (Kurdjumov) a specialist pest and its host *Triticum aestivum* (L.), *Arthropod-Plant Interactions*, 7(4), 359–372.

Breeds, K., Burger, N.F.V. & Botha, A-M., 2018, New insights into the methylation status of virulent *Diuraphis noxia* (Hemiptera: Aphididae) biotypes, *Journal of Economic Entomology*, 111(3), 1395–1403.

Burger, N.F.V. & Botha, A-M., 2017, Genome of Russian wheat aphid an economically important cereal aphid, *Standarts in Genomic Sciences*, 12(1), 90.

du Toit, F., 1989, Inheritance of resistance in two *Triticum aestivum* Lines to Russian Wheat Aphid (Hornoptera: Aphididae), *Journal of Economic Entomology*, 82(4), 1251–1253.

du Toit, F. & Walters, M.C., 1981, Damage assessment and economic threshold values for the chemical control of thr Russian wheat aphid, *Diuraphis noxia* (Mordvilko) on winter wheat, *Technical communication, Department of Agriculture, Republic of South Africa*, (191), 58–62.

Fazel-Najafabadi, M., Peng, J., Peairs, F.B., Simkova, H., Kilian, A. & Lapitan, N.L.V., 2015, Genetic mapping of resistance to *Diuraphis noxia* (Kurdjumov) biotype 2 in wheat (*Triticum aestivum L.*) accession CI2401, *Euphytica*, 203(3), 607–614.

Food and Agriculture Organization of the United Nations, 2018, Food outlook, *Biannual Report on Global Food Markets*, 13.

Fouché, A., Verhoeven, R.L., Hewitt, P.H., Kriel, C.F., de Jager, J. & Walters, M.C., 1984, Russian aphid (*Diuraphis noxia*) feeding damage on wheat, related cereals and a *Bromus* grass species, *Technical communication, Department of Agriculture, Republic of South Africa*, (191), 22–33.

Stellenbosch University https://scholar.sun.ac.za

Gong, L., Cui, F., Sheng, C., Lin, Z., Reeck, G., Xu, J. & Kang, L., 2012, Polymorphism and methylation of four genes expressed in salivary glands of Russian wheat aphid (Homoptera: Aphididae), *Journal of Economic Entomology*, 105(1), 232–241.

Haley, S.D., Peairs, F.B., Walker, C.B., Rudolph, J.B. & Randolph, T.L., 2004, Occurrence of a new Russian wheat aphid biotype in Colorado, *Crop Science*, 44(5), 1589–1592.

Jankielsohn, A., 2011, Distribution and diversity of Russian wheat aphid (Hemiptera: Aphididae) biotypes in South Africa and Lesotho., *Journal of economic entomology*, 104(5), 1736–1741.

Jankielsohn, A., 2019, New Russian wheat aphid biotype found in Free State, *Grain SA Mini Focus, Pest Control on Winter Cereals*.

Li, Y. & Tollefsbol, T.O., 2011, DNA methylation detection: bisulfite genomic sequencing analysis, *Epigenetics Protocols*, pp. 11–21.

Nkongolo, K.K., Quick, J.S., Limin, A.E. & Fowler, D.B., 1991, Sources and inheritance of resistance to Russian wheat aphid in *Triticum* species *amphiploids* and Triticum *tauschii*, *Canadian Journal of Plant Science*, 71(3), 703–708.

Pasquier, C., Clément, M., Dombrovsky, A., Penaud, S., Rocha, M. Da, Rancurel, C., Ledger, N., Capovilla, M. & Robichon, A., 2014, Environmentally selected aphid variants in clonality context display differential patterns of methylation in the genome, *PloS One*, 9(12) e115022.

Peng, J., Wang, H., Haley, S., Peairs, F. & Lapitan, N., 2007, Molecular mapping of the Russian wheat aphid resistance gene *Dn2414* in wheat, *Crop Science*, 47(6), 2418–2429.

Quick, J.S., Ellis, G.E., Normann, R.M., Stromberger, J.A., Shanahan, J.F., Peairs, F.B., Rudolph, J.B. & Lorenz, K., 1996, Registration of "Halt" wheat, *Crop Science*, 36(1) 209–210.

Randolph, T.L., Peairs, F., Weiland, A., Rudolph, J.B. & Puterka, G.J., 2009, Plant responses to seven Russian wheat aphid (Hemiptera: Aphididae) biotypes found in the United States, *Journal of Economic Entomology*, 102(5), 1954–1959.

Russo, V.E.A., Martienssen, R.A. & Riggs, A.D., 1996, *Epigenetic mechanisms of gene regulation*, Cold Spring Harbor Laboratory Press.

Shufran, K. & Payton Miller, T., 2009, Limited genetic variation within and between Russian wheat aphid (Hemiptera: Aphididae) biotypes in the United States, *Journal of Economic Entomology*, 102(1), 440–445.

Tolmay, V., Lindeque, R. & Prinsloo, G., 2007, Preliminary evidence of a resistance-breaking biotype of the Russian wheat aphid, *Diuraphis noxia* (Kurdjumov) (Homoptera: Aphididae), in South Africa, *African Entomology*, 15(1), 228–230.

Valdez, V.A., Byrne, P.F., Lapitan, N.L. V, Peairs, F.B., Bernardo, A., Bai, G. & Haley, S.D., 2012, Inheritance and genetic mapping of Russian wheat aphid resistance in Iranian wheat landrace accession PI 626580, *Crop Science*, 52(2), 676–682.

van Zyl, R.A. & Botha, A-M., 2008, Eliciting proteins from *Diuraphis noxia* biotypes differ in size and composition, *18th Biennial International Plant Resistance to Insects Workshop*, Fort Collins, Colorado.

Weiland, A.A., Peairs, F.B., Randolph, T.L., Rudolph, J.B., Haley, S.D. & Puterka, G.J., 2008, Biotypic diversity in Colorado Russian wheat aphid (Hemiptera: Aphididae) populations, *Journal of Economic Entomology*,

Stellenbosch University https://scholar.sun.ac.za

101(2), 569–574.

Weng, Y., Azhaguvel, P., Michels, G.J. & Rudd, J.C., 2007, Cross-species transferability of microsatellite markers from six aphid (Hemiptera: Aphididae) species and their use for evaluating biotypic diversity in two cereal aphids, *Insect Molecular Biology*, 16(5), 613–622.

Stellenbosch University https://scholar.sun.ac.za

Chapter 2: Literature review

2.1: Background

The Russian wheat aphid (RWA), *Diuraphis noxia* (Kurdjumov), is a small green aphid in the order Hemiptera (true bugs). The aphid is indigenous to the surrounding regions of the Fertile Crescent and was first observed in South Africa in 1978, where it was easily distinguished from other aphids in the area by its spindle-shaped body, a characteristic double tail, caused by a protrusion above the cauda, extremely short antennae and siphunculi which are not prominent with the naked eye (Figure 2.1). As its name implies, RWA's host of preference is wheat, but it also feeds on barley and triticale and can overwinter on species of *Bromus* grass. While RWA can feed on rye and oats, these crops do not seem to be favoured hosts and are usually clear of infestation under field conditions (Walters et al., 1980).

Symptoms caused by the feeding of RWA on susceptible plants include purple leaf discolouration, longitudinal chlorotic streaking, necrotic lesions, leaf rolling and head trapping (Walters et al., 1980) (Figure 2.1). Initially, the severe symptoms caused by RWA infestation were thought to be the result of an infection by either barley yellow dwarf virus (BYDV) or brome mosaic virus (BMV), as many other aphid species serve as vectors, transferring viruses from one plant to another (Von Wechmar and Rybicki, 1981). This was disproven however, by co-infesting leaves with RWA and two aphid species know to be vectors for BYDV and BMV, *Rhopalosiphum padi* and *Schizaphis graminum*. After separating the three species and moving each to fresh plants, the streaking and rolling symptoms were only duplicated in plants infested with RWA, not in those infested with either *R. padi* or *S. graminum*. This experiment showed that RWA feeding symptoms could not be attributed to a viral agent, but was a direct result of feeding, possibly through the injection of an enzyme into plant cells (Fouché et al., 1984).

Early studies showed that depending on the severity of infestation and the growth stage of the host plants at time of infestation, yield losses as high as 91% can occur. Preventing these yield losses with insecticides could be costly as single sprays only prevented a significant loss in yield when applied before growth stage 7 of wheat (appearance of the second node). Early detection and treatment was necessary, as spraying after growth stage 9 (the start of flowering) could not prevent significant losses, even if the plant was kept aphid free until harvest (Du Toit and Walters, 1981). To elevate the financial and managerial inputs of chemical control, the release of wheat cultivars resistant to RWA

Stellenbosch University https://scholar.sun.ac.za

became a priority. This was especially the case after the detection of RWA in the United States of America (USA) in 1986, where severe economic losses were reported (Stoetzel, 1987).

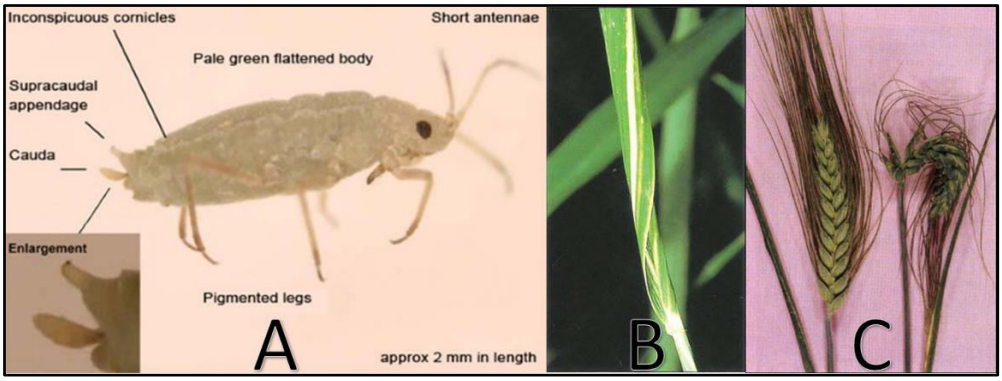

Figure 2.1: (A) An image of RWA illustrating key identifying features (Nicholas et al., 2015), **(B)** a leaf showing chlorotic streaking and rolling, symptoms of a susceptible reaction to RWA feeding (González et al., 1992) **(C)** a normal wheat awn, and one trapped by a flag leaf as a result of RWA feeding, known as head trapping (Nicholas et al., 2015).

Resistance breeding efforts started showing results independently in South Africa and the USA: For the first time, wheat showing resistance to RWA was described as genotype PI 137739 and PI 262660 in South Africa (Du Toit, 1989), followed by the discovery of a resistance gene in the SQ24 cultivar of a goatgrass, *Aegilops tauschii*, (Nkongolo et al., 1991) and the resistant wheat genotype PI 372129 (Quick et al., 1991). These genes were collectively called *Dn* genes (*Diuraphis noxia* resistance genes) and respectively numbered *Dn1* to *Dn4* chronologically in the order in which they were described. Crosses were made to incorporate these resistant genes into cultivars favoured for their agronomic traits. This lead to the release of the first commercial RWA resistant cultivars, Tugela-DN to South African growers, containing *Dn1* (Purchase et al., 1996), and Halt to growers in the US, containing *Dn4* (Quick et al., 1996).

The incorporation of resistance genes mitigated yield losses until typical symptoms of a susceptible reaction to RWA were reported in the spring of 2003 in the USA. This was a cause for concern, as approximately 25% of the wheat, planted in the USA that year, were RWA resistant cultivars (Colorado Agricultural Statistics Service, 2004). Isolates were collected from the field and screened on nine cultivars, of which the reaction to the original RWA found in the USA was known. Plant reactions were scored for leaf rolling, on a scale of one to three, and for overall plant damage, on a scale of one to nine. The new isolates elicited more severe reactions than the original isolates, both for leaf rolling (3.0 vs. 2.0, p-value = 0.003) and for overall plant damage (8.7 vs. 4.7, p-value < 0.001) and were confirmed to be isolates of a new biotype. A further sixteen cultivars were used to test the level of resistance of nine known resistance genes, *Dn1* – *Dn9*, against the new biotype. The only cultivar showing strong resistance was one containing the resistance gene *Dn7* (Haley et al., 2004). In only three years following the discovery of this new biotype, six new biotypes developed, bringing

Stellenbosch University https://scholar.sun.ac.za

the total number of RWA biotypes in the USA to eight (RWA1 – RWA8) by 2006 (Randolph et al., 2009; Weiland et al., 2008) (Table 2.1).

Table 2.1: The virulence profile (**S** = susceptible, **R** = resistant) of the eight Russian wheat aphid biotypes found in North America on nine resistance genes (Puterka et al., 2014).

Resistance gene	**Biotype**							
	RWA1	RWA2	RWA3	RWA4	RWA5	RWA6	RWA7	RWA8
Dn1	R	S	S	S	S	S	S	R
Dn2	R	S	S	S	S	S	S	R
Dn3	R	S	S	S	S	S	S	R
Dn4	R	S	S	S	S	R	S	R
Dn5	R	S	S	S	S	S	S	R
Dn6	R	S	R	R	R	R	R	R
Dn7	R	R	R	R	R	R	R	R
Dn8	S	S	S	S	S	S	S	S
Dn9	S	S	S	S	S	S	S	R

A biotype is defined by the reaction it elicits from a differential set of wheat lines, containing a range of *Dn* genes. Reactions are scored either as resistant or susceptible to a biotype, based on factors like seedling death, chlorotic streaking, and leaf rolling. The larger the number of susceptible reactions a biotype elicits from the wheat lines in a differential set, the more virulent that biotype is scored. Since biotypes are morphologically similar, a screening as used in the study mentioned above, is the only way to differentiate between biotypes (Botha et al., 2010; Smith et al., 1992).

Although RWA was reported in South Africa almost a decade earlier than in the USA, the rate of biotypification was much slower, as the first new South African biotype, RWA SA2, was only described in 2007, after a noted loss of resistance and growing RWA populations in 2005 (Tolmay et al., 2007). Two additional biotypes were identified in 2009 and 2011, RWA SA3 and RWA SA4 respectively (Jankielsohn, 2011). Early in 2019 a new biotype was identified in the Eastern Free State, denoted RWA SA5, and is the most virulent South African biotype found in field conditions described to date (Table 2.2) as it is able to overcome all of the known resistance genes, except for *Dn7* (Jankielsohn, 2019).

It is interesting to note that while each new South African biotype was more virulent than any of the previously reported biotypes (Jankielsohn, 2019), this is not the case in the USA, as RWA2 is the most virulent biotype, while the last biotype described, RWA8, is the least virulent of all (Puterka et al., 2014).

Stellenbosch University https://scholar.sun.ac.za

Table 2.2: The virulence profile (**S** = susceptible, **R** = resistant) of the five Russian wheat aphid biotypes found in South Africa on eleven resistance genes (Jankielsohn, 2019).

Resistance gene	Biotype				
	RWASA1	RWASA2	RWASA3	RWASA4	RWASA5
Dn1	R	S	S	S	S
Dn2	R	S	S	S	S
Dn3	S	S	S	S	S
Dn4	R	R	S	S	S
Dn5	R	R	R	S	S
Dn7	R	R	R	R	R
Dn8	R	S	S	S	S
Dn9	R	S	S	S	S
Dnx2006	R	R	R	R	S
Dny2006	R	S	S	S	S

2.2: SAM: A synthetic model

In addition to the five naturally occurring South African RWA biotypes, a laboratory contained biotype, SAM (South African Mutant), also exists. It was developed by maintaining biotype SA1 on resistant Tugela-DN wheat for 78 generations (van Zyl and Botha, 2008). The SAM biotype is able to overcome all of the known *Dn*-genes, and is therefore the most virulent biotype that has been described to date (Botha, 2013).

Cytogenetic analyses have revealed that the diploid RWA genome consist of five chromosomes, of which the X chromosome is haploid in males, giving the RWA an XX/XO sex determination system (Novotná et al., 2011). The genomes of biotype SAM and SA1, the progenitor of SAM, have previously been sequenced (Burger and Botha, 2017) (GenBank ID: GCA_001465515.1, BioProject: PRJNA297165) with an estimated genome size of between 593 and 623 Mb, based on k-mer frequency analyses. In the current draft assembly, genic regions comprise 16.67% of the genome and have been divided into 31 885 protein coding genes through *ab initio* gene calling. The RWA has a percentage GC content of only 29.5%, making it the most GC poor insect genome sequenced to date, with the closely related pea aphid, *Acyrthosiphon pisum,* in second place at 29.6% (Burger and Botha, 2017).

The positions of biotypes SA1 and SAM on the opposite extremes of the virulence scale, along with the knowledge that SAM developed directly from SA1, make these two biotypes a good model for the study of RWA virulence and biotypification.

An extensive study on plant-insect interactions had been performed using these two biotypes as a virulent *versus* non-virulent model (Burger et al., 2017). In the study, both biotypes were cultivated

Stellenbosch University https://scholar.sun.ac.za

on their preference host and transferred to a differential set of sixteen different wheat lines. The aphids fed on the new host, consisting of various cultivar background and *Dn*-gene combinations, and were sampled at 4 and 48-hour intervals. These sampling times were selected as they respectively coincided with the wheat hypersensitivity response (Martin et al., 2003), and the production of secondary metabolites (as part of systematic acquired resistance) (Botha et al., 1998). RNA was extracted from the heads of aphids (to enrich for salivary gland transcripts), converted to cDNA and used to compare expression levels, using cDNA-amplified fragment length polymorphisms (cDNA-AFLP). A change in host had a substantial effect on the transcription levels within and between biotypes, with 42.63% of the transcripts showing a significant deviation from 0-hour controls (p-value < 0.05). Most of these transcripts were up-regulated in SAM and down-regulated in SA1. Further, the changes in transcription in SA1 correlated strongly with the genotype of the host plant, while changes in SAM transcription correlated strongly with the sampling times. This indicates a greater ability in the SAM biotype to alter expression based on plant defensive responses within a short time frame (Burger et al., 2017). How this altered expression is regulated is still uncertain though, as comparisons between the draft genomes of SA1 and SAM revealed limited variation. To explain the results observed in the study by Burger et al. (2017), a regulatory mechanism is required that is temporal, amenable to rapid modification and heritable.

2.3: DNA methylation

DNA methylation is an epigenetic modification, which is defined as a modification that can exact heritable changes in phenotype without a corresponding change in genotype, by altering gene expression (Russo et al., 1996). The reaction is catalysed by the DNA methyltransferase (DNMT) family of enzymes and involves the covalent addition of a methyl group to the 5-carbon position of a cytosine molecule, forming 5-methylcytosine (5mC), with S-adenosyl-L-methionine as the methyl donor (Okano et al., 1998) (Figure 2.2).

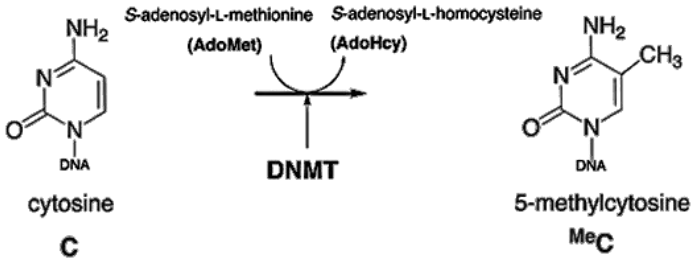

Figure 2.2: The methylation of a cytosine base by DNMT, using S-adenosyl-L-methionine as a methyl donor (Moison et al., 2013).

Stellenbosch University https://scholar.sun.ac.za

Insect DNA methylation appears in three different sequence contexts based on the nucleotides immediately following a methylated cytosine base: CpG, CHG and CHH (where H denotes any non-G base) (Gong et al., 2012). The CpG and CHG contexts are referred to as being symmetrical, as each methylation site will consist of a cytosine base in both the top and bottom strand, whereas only one strand will contain a cytosine base in the asymmetrical CHH context. Most studies to date have focused on the CpG context. This is likely due to the absence of the other contexts in the methylation profile of humans, where methylation has been studied most extensively. Furthermore, some of the techniques used to study methylation are only capable of detecting differences in CpG methylation specifically (refer to section 2.3.6).

Insects studied thus far have presented levels of CpG methylation between 0% (for some species in the orders Diptera, Lepidoptera, and Hymenoptera) and 14% (for some species in the order Blattodea). The order Blattodea shows much higher levels of CpG methylation than any of the other orders, while the Hemiptera are the second most methylated order and only show levels as high as 4% (figure 2.4) (Bewick et al., 2017).

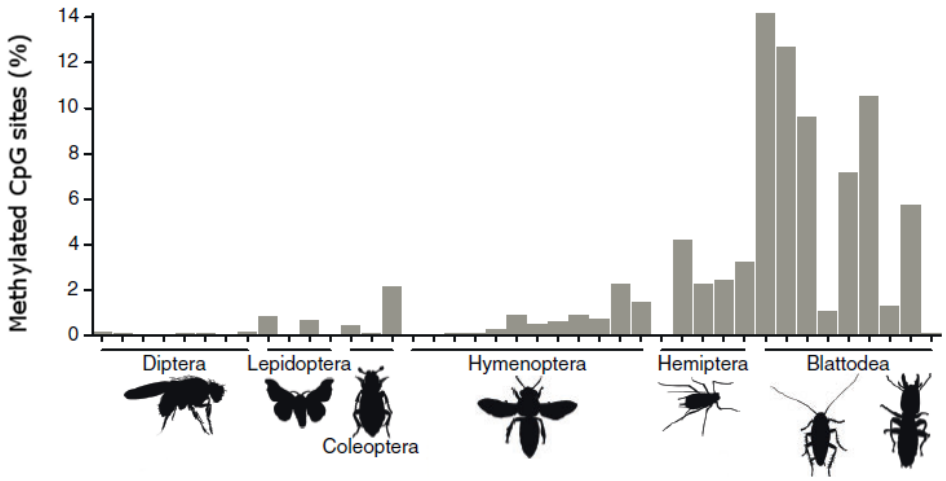

Figure 2.3: The percentage of CpG sites methylated across the genomes of six insect orders (Bewick et al., 2017).

The distribution of CpG methylation across invertebrate genomes is highly conserved and has mostly been observed in genic regions (Suzuki and Bird, 2008). Within these genic regions exons are methylated to a greater extent than introns, with exons located closer to the 5'-end of a gene showing the highest level of methylation (Elango et al., 2009), while intronic methylation is usually localised near the intron boundaries (Feng et al., 2010). Genes that are ubiquitously expressed throughout cell types tend to be methylated, while genes that are differentially expressed between cell types tend to be unmethylated. Glastad et al. (2014) noted, however, that direct comparisons of methylation and tissue specific expression are lacking and that fine scale methylation studies will be

Stellenbosch University https://scholar.sun.ac.za

required to corroborate these findings. The observed trends in insect DNA methylation are illustrated in Figure 2.4.

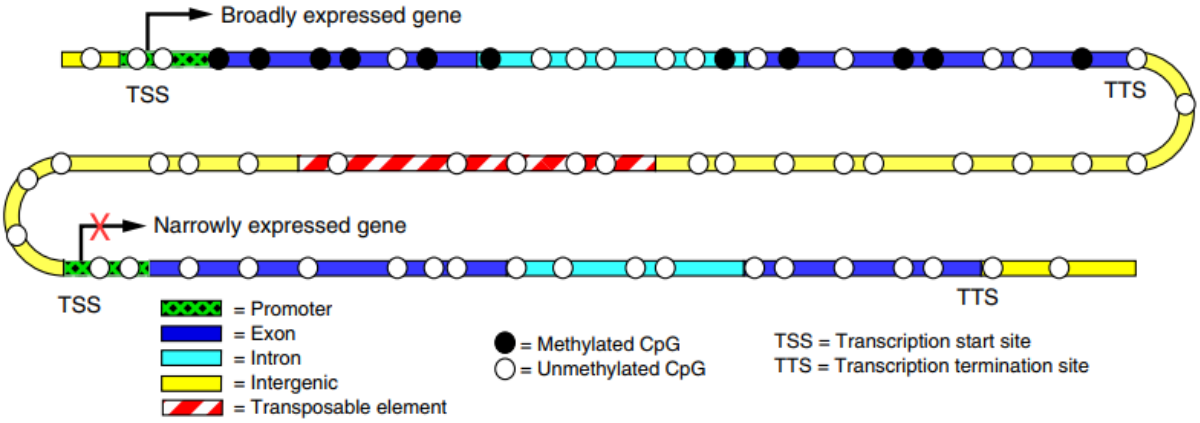

Figure 2.4: The distribution of methylated CpG sites across the insect genome. Genes tend to be methylated, while intergenic regions tend to be unmethylated. Broadly expressed genes tend to be more methylated than narrowly expressed genes. Methylation levels are higher in exons than in introns. Exonic methylation is higher in exons close to the 5' end of the gene, while intronic methylation is highest at the intron boundaries (Glastad et al., 2014).

2.3.1: The methylation machinery

As mentioned earlier, DNA methylation is catalysed by the DNMT family of enzymes, which can be divided into three functional classes, DNMT1, DNMT2 and DNMT3.

DNMT1

DNMT1 is important for the maintenance of methylation patterns (Yoder et al., 1997). During DNA replication, the cytosine bases incorporated in the newly synthesised DNA are unmethylated. When replication occurs at a fully methylated site (i.e. a di- or trinucleotide locus with a methylated cytosine base on both the top and bottom strand) the original strand will be methylated, but the newly synthesised strand will be unmethylated. Such a site, where methylation is only present on one of the strands is referred to as a hemimethylated site (Deobagkar et al., 1990). DNMT1 has a high affinity for these hemimethylated sites and will bind them preferentially, methylating the newly synthesised strand and maintaining the methylation state of the site (Yoder et al., 1997).

DNMT2

Less is known about the DNMT2 class. Interestingly, in addition to DNA methyltransferase activity, it also has RNA methyltransferase activity, albeit specific to only the 38 cytosine bases of the anticodon loop of $tRNA^{Asp}$ (Goll et al., 2006).

Stellenbosch University https://scholar.sun.ac.za

DNMT3

Enzymes belonging to the DNMT3 class have a high affinity for unmethylated DNA, and possess *de novo* methylation activity, making this the class of transferases primarily responsible for the establishment and addition of methylation patterns (Okano et al., 1998). This, along with the responsiveness of DNMT3 activity to environmental stimuli (Goll and Bestor, 2005), suggests that these enzymes play an important role during development and perhaps adaptation to new environmental conditions (Standage et al., 2016; Zhang et al., 2015).

In the past, it was believed that DNMT3 and at least one DNMT1 is required for a functional DNA methylation system (Glastad et al., 2014). This was found to be incorrect in a study by Bewick et al. (2017), when the evolution of the DNMT genes were investigated across twelve insect orders. As seen in Table 2.3, DNMT3 is order-poor and can only be found in four of the investigated insect orders (Bewick et al., 2017). *Drosophila melanogaster* and *Anopheles gambiae* both only contain DNMT2 (Marhold et al., 2004), yet methylation patterns can still be found in these insects, albeit at very low levels (less than 0.002%) (Bewick et al., 2017; Panikar et al., 2015). Therefore, at least in insects, a DNA methylation system can function even with only a single class of DNMT.

Table 2.3: The presence of the three insect DNA methyltransferases across twelve insect orders (Bewick et al., 2017).

Order	DNMT1a	DNMT1b	DNMT2	DNMT3
Diptera	✗	✗	✓	✗
Lepidoptera	✗	✓	✓	✗
Trichoptera	✗	✓	✓	✗
Coleoptera	✗	✓	✓	✓
Hymenoptera	✓	✗	✓	✓
Phthiraptera	✓	✓	✓	✗
Hemiptera	✓	✓	✓	✓
Thysanoptera	✓	✓	✓	✗
Blattodea	✓	✓	✓	✓
Orthoptera	✓	✓	✓	✗
Ephemeroptera	✓	✓	✓	✗
Odonata	✓	✓	✓	✗

Vertebrate genomes contain genes that code for all three of the DNMT classes, with DNMT3 represented by two subclasses, DNMT3a and DNMT3b (Cheng and Blumenthal, 2008). For

Stellenbosch University https://scholar.sun.ac.za

invertebrate genomes, there is no such standard, as different orders can have different complements of the *DNMT* genes (Lyko and Maleszka, 2011). In a study by Bewick et al. (2017), the evolution of the *DNMT* genes were investigated across twelve insect orders. From this study, it was found that in insects only a single DNMT3 homolog is present while two DNMT1 homologs exist, namely DNMT1a and DNMT1b (Bewick et al., 2017) as summarised in Table 2.3. As RWA falls under the order Hemiptera, it likely contains functional copies of all the insect *DNMT* genes.

2.3.2: Differences between vertebrate and invertebrate methylation

In the aforementioned study by Bewick et al. (2017) it was found that in insects only a single DNMT3 homolog is present while two DNMT1 homologs exist, namely DNMT1a and DNMT1b, and that not all insect orders possess all four of these genes (Bewick et al., 2017). Vertebrate genomes however, contain genes that code for all three of the DNMT classes, with DNMT3 represented by two subclasses, DNMT3a and DNMT3b and a single DNMT1 (Cheng and Blumenthal, 2008).

Along with the differences in available methylation machinery between vertebrates and invertebrates, there are also differences in the distribution, context, and levels of methylation. Vertebrate genomes are globally methylated as methylation can be seen in intergenic areas and transposons, as well as intergenic areas, including the promotor and untranslated regions (Glastad et al., 2014). The CpG context of methylation is the only one found in vertebrate genomes, except in very specific examples, such as embryonic stem cells and prostate cancer cells (Truong et al., 2013). This is not the case in insects, as RWA has been found to be able to methylate cytosine bases in all three contexts (Gong et al., 2012). Levels of methylation in insects are very low (when compared to other organisms (Glastad et al., 2011). In mammals for example, between 70% and 80% of CpG sites are methylated (Li and Zhang, 2014).

2.3.3: Factors that affect methylation patterns

As mentioned earlier, DNMT3 is responsive to environmental stimuli which is important for adaptation (Goll and Bestor, 2005). This also holds true for insects, as several biotic and abiotic factors have been shown to elicit a phenotypic response, driven by methylation. These include photoperiod, ambient temperature, the amino acid composition of a host plant, and the population density of aphids during infestation (Kim et al., 2018; Pasquier et al., 2014; Pegoraro et al., 2016).

Symbioses

All aphids house the bacterial endosymbiont, *Buchnera aphidicola*, in specialised bacteriocyte cells that are regulated by the aphid. In the case of the pea aphid, it has been shown that the aphid and bacterial amino acid pathways are intimately integrated and thereby able to control the production

Stellenbosch University https://scholar.sun.ac.za

of essential amino acids necessary for the aphid's survival (Nakabachi et al., 2005). A recent study demonstrated that depending on host-plant amino acid composition, key aphid-*Buchnera* symbiosis regulatory genes within aphid bacteriocytes cells display altered methylation profiles as well as alternative splicing and differential expression (Kim et al., 2018).

Not only are the methylomes of pests affected during infestation, host methylation patterns can also be altered. *Cotesia plutellae*, a larval endoparisitoid of the diamondback moth, *Plutella xylostella*, significantly reduces the expression of host *DNMT* genes during parasitism. The lack of methylation machinery greatly affects gene expression and is believed to favour successful parasitism (Kumar and Kim, 2017).

Temperature

The pea aphid is an example of an insect with considerable phenotypic plasticity, as it appears as winged and wingless morphs with the body of the aphid also readily changing colour between pink, green and white, depending on environmental conditions such as temperature and population density (Pasquier et al., 2014). Under optimal conditions, carotene production is triggered, leading to the production of pink aphids. When population levels exceed a threshold, carotene production ceases, leading to the formation of a white colour, which is also attainable through injecting pink aphids with inhibitors of DNMTs (Dombrovsky et al., 2009). Maintaining the aphid at low temperatures (8°C) results in green pigmentation. High throughput bisulfite sequencing and transcriptome sequencing revealed an increase in methylation and expression for genes coding for proteins involved in aspartate/glutamate metabolism, phosphoinoside metabolism, lipid metabolic process, ATP synthesis coupled to electron transport, mitochondrial electron chain, lipid and carbohydrate transport and GTPase activity. It was suggested that methylation helps the aphid adapt to colder temperatures (Pasquier et al., 2014).

Sex

Whole genome bisulfite sequencing of male and asexual female pea aphids have also shown differences in methylation patterns between sexes. In 79% of the CpG sites that show at least a 15% difference in methylation level, the site is hypermethylated in females. Furthermore, haploid male X chromosomes are hypermethylated, compared to the hypomethylated state of the rest of the male genome and the diploid X chromosomes of females which were also hypomethylated (Mathers et al., 2019).

Stellenbosch University https://scholar.sun.ac.za

Photoperiod

Wasps in the genus *Nasonia* are acutely acclimatised to temperate zones. During winter the larvae need to enter a state of developmental arrest, termed diapause, in order to survive cold winters. Diapause is induced in the offspring when females of this genus are exposed to short photoperiods during autumn (Saunders, 1965). To study this process, the *Nasonia vitripennis* genome was sequenced, revealing a full complement of DNA methylation genes (excluding DNMT1b) (Bewick et al., 2017). The methylation levels were assessed to investigate if DNA methylation is associated with the photoperiod response (Werren et al., 2010). It was found that changes in methylation patterns across the genome, coincided with changes in photoperiod. Furthermore, the response to photoperiod was abolished by either knocking down *DNMT1a* in the wasp or the addition of a DNA methylation inhibitor, namely 5-aza-2'-deoxycytidine (Pegoraro et al., 2016).

2.3.4: The function of DNA methylation in insects

While these studies show a strong correlation between methylation and phenotypes, the mechanism through which DNA methylation might influence gene regulation is still poorly understood (Pegoraro et al., 2016). Promotor methylation is a well-known mechanism of gene silencing in vertebrates (Bird, 2002), methylation in insects however, is much more abundant in exons, rather than promotor regions (Elango et al., 2009), which would limit the amount of regulation exerted by promotor methylation. Although the mechanism is not as well understood as in the case of promotor methylation, genic methylation is also thought to influence transcription to some extent as a strong correlation between highly transcribed and highly methylated genes were found in the Florida carpenter ant, *Camponotus floridanus* (Glastad et al., 2015).

Genic methylation has also been shown to prevent spurious RNA polymerase II entry and ultimately prevent the translation of aberrant proteins (Neri et al., 2017), and to play a role in alternative splicing (Lyko and Maleszka, 2011). Depending on which DNA binding proteins are functional within an organism genic methylation can either promote or prevent exon inclusion. Human CCCTC-binding factor (CTCF) binds to unmethylated DNA and promotes RNA polymerase II pausing and exon inclusion (Yan et al., 2015), while human methyl-CpG-binding protein 2 (MeCP2) binds preferentially to methylated DNA yet also promotes exon inclusion (Maunakea et al., 2013). Homologues for both of these proteins have since been found in *Drosophila* (Cukier et al., 2008; Holohan et al., 2007), but no significant increase in methylated MeCP2 binding sites were found. This was attributed to the overall low level of methylation of the Drosophila genome and the fact that the affinity for methylated DNA by MeCP2 is only about three times more than its affinity for unmethylated DNA (Cukier et al., 2008). In an insect species showing higher levels of methylation, such as a member of

Stellenbosch University https://scholar.sun.ac.za

the order Blattodea, the preferential binding of methylated DNA by MeCP2 may be more pronounced.

In a study by Hunt et al. (2010), a correlation was found between high levels of DNA methylation and slowly evolving, ubiquitously expressed genes in the honeybee, *Apis mellifera*, and the pea aphid, *A. pisum*. This suggests that DNA methylation may play a role in functional conservation of genes (Hunt et al., 2010), although this contradicts the finding that methylated DNA is more prone to deamination and leads to mutations (Duncan and Miller, 1980).

2.3.5: Demethylation

Although DNA methylation is a heritable trait (Russo et al., 1996), it is reversible through a process called demethylation, wherein methylated cytosine bases are returned to an unmodified state. There are two routes of demethylation, an active and a passive route (Chen and Riggs, 2011). Passive demethylation occurs due to the failure of DNMT1 to maintain methylation patterns. If, after replication, DNMT1 does not bind to, and methylate, a temporarily hemimethylated site before another round of replication is completed, both the methylated, as well as unmethylated strand will serve as a template for replication, resulting in one unmethylated and one hemimethylated double strand (Chen and Riggs, 2011). Active demethylation, on the other hand, is a process which comprises of many enzymatic reactions, the first of which is catalysed by a ten-eleven-translocation enzyme (TET), resulting in 5-hydroxymethylcytosine (5hmC) through oxidation (Tahiliani et al., 2009). Mammals possess three TET enzymes (TET1 to TET3) while insects are believed to only have one functional TET homologue (Wojciechowski et al., 2014). A TET homologue has not been found in RWA, but the presence of 5hmC in four RWA biotypes indicates that a functional TET enzyme is present (Breeds et al., 2018).

After the initial oxidization to 5hmC, TET can catalyse another oxidization to deliver 5-formylcytosine (5fC), and another to oxidise 5fC into 5-carboxylcytosine (5caC). In addition to these oxidised forms of 5mC being susceptible to the same route of passive demethylation, they can also serve as the substrate for the enzyme Thymine-DNA glycosylase (TDG), which converts modified cytosine to an abasic site. These abasic sites get replaced by unmodified cytosine bases during base excision repair (BER) (Kohli and Zhang, 2013). The active pathway of DNA demethylation is outlined in Figure 2.5.

Stellenbosch University https://scholar.sun.ac.za

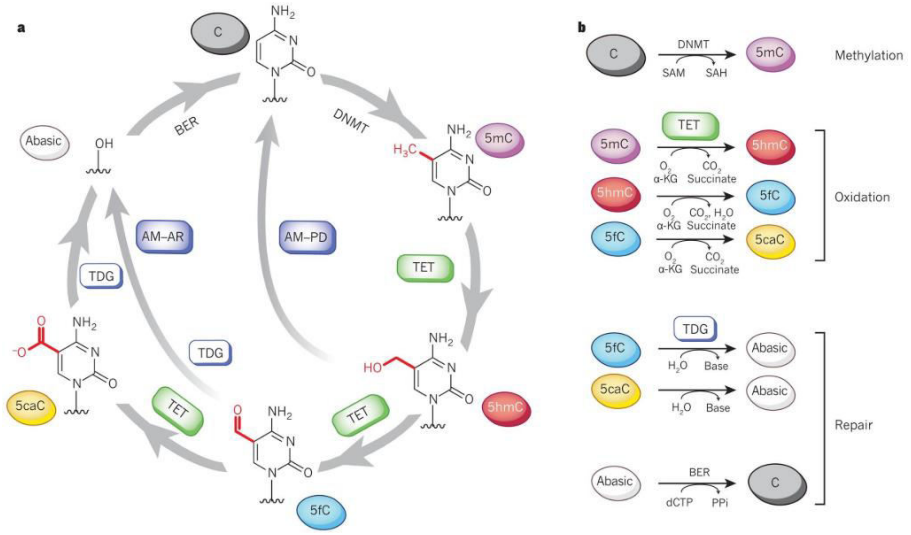

Figure 2.5: The active pathway of DNA demethylation, through the TET and TDG enzymes, and the process of BER (Kohli and Zhang, 2013).

At first 5hmC was only considered as an intermediary in the active demethylation pathway, but upon the discovery of proteins in the mouse genome that bind specifically to 5hmC, and a possible correlation between 5hmC and exon inclusion in honey bee brain cells, it has been studied as an epigenetic mark in its own right (Cingolani et al., 2013; Spruijt et al., 2013).

2.3.6: Methods of measuring DNA methylation

Methylation is an epigenetic mark and does not alter the genetic sequence. As such, specialised techniques are required to investigate methylation patterns (Dupont et al., 2009). These techniques can be divided into three main categories: techniques that are reliant on methylation sensitive restriction enzymes, techniques that utilise methylation sensitive antibodies, and techniques that are based on the chemical modification of methylated DNA (Jin et al., 2010).

Restriction enzymes

Pairs of restriction enzymes that share the same recognition site, but bind differentially depending on the methylation state of the recognition site, are called isoschizomers (McClelland et al., 1994). Isoschizomers, specifically MspI and HpaII with recognition sites CCGG, were the first tools utilised in the study of DNA methylation and can be used in conjunction with techniques such as Restriction fragment length polymorphism (RFLP) (Vos et al., 1995) or AFLP (Reyna-Lopez et al., 1997). When combining isoschizomers with the AFLP technique, it is referred to as methylation-sensitive amplified polymorphism (MSAP). Two disadvantages of this method are that methylation data can only be generated for loci containing the recognition site of the restriction enzymes, and that positional information cannot be obtained. Advantages of using the MSAP method is that no prior knowledge

Stellenbosch University https://scholar.sun.ac.za

of the genome of the organism in question is required, and that genome composition and methylation level does not impact on its use.

Antibodies

The first step in antibody-based methods of methylation detection is the enrichment of methylated DNA, using methylated DNA immunoprecipitation (Me-DIP) (Weber et al., 2005). This can then be followed by sequencing reactions, or quantification through enzyme-linked immunosorbent assay (ELISA) or microarrays (Fouse et al., 2010) to establish the overall levels of methylation. The disadvantages of these methods are that context specific and positional information cannot be obtained. Advantages of utilising methylation binding antibodies include detection of all the methylation contexts, and that no prior knowledge is required about the genome of the organism in question, as with restriction enzyme-based methods. The level of methylation in South African RWA biotypes have been previously calculated using Me-DIP by Breeds et al. (2018), and it was reported that there were no significant differences between South African biotypes. These results may be due to the lowered sensitivity of methylation binding antibodies which struggle to differentiate between samples with equally low levels of methylation (Okitsu & Hsieh, 2015).

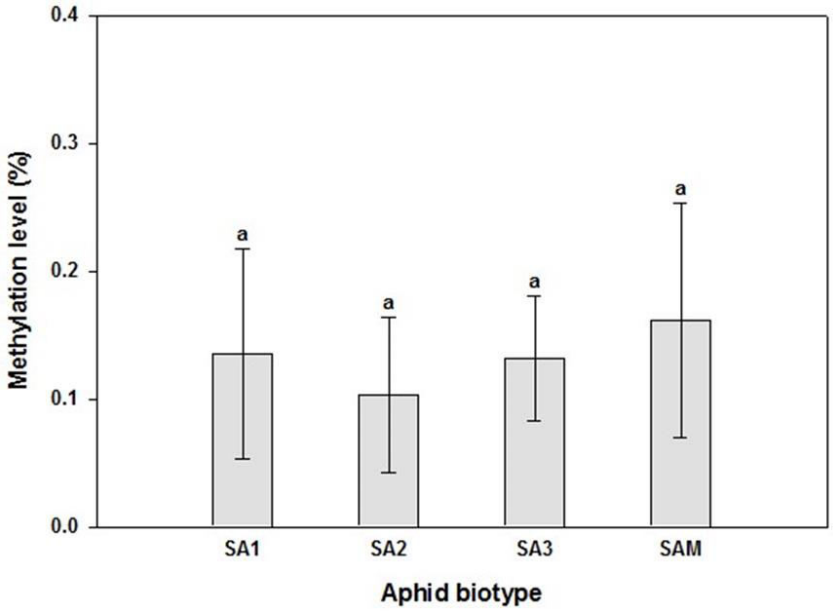

Figure 2.6: The level of DNA methylation in four South African RWA biotypes as calculated using Me-DIP (Breeds et al., 2018).

Chemical treatments

Chemical treatment of DNA with sodium bisulphite leads to the deamination of cytosine bases. When this is followed by a desulphonation step, the resulting bases are chemically identical to uracil (Darst et al., 2010) (Figure 2.7). Methyl groups temporarily protects cytosine bases from deamination, methylated DNA therefore, requires a longer exposure to bisulphite ions in order to be

Stellenbosch University https://scholar.sun.ac.za

deaminated. This time delay in deamination between methylated and unmethylated cytosine can be exploited to elucidate the methylation status of cytosine bases in a technique called bisulfite sequencing (Li and Tollefsbol, 2011). The three main steps involved in bisulfite sequencing are: determining the nucleotide sequence of a DNA sample through Sanger sequencing, bisulfite treating that same sample, and determining the nucleotide sequence of the bisulfite treated sample. When the control sequence is compared to the bisulfite treated sequence, any cytosine bases that have been converted to thymine in the bisulfite treated sequence would indicate unmethylated cytosine bases. Any cytosine bases present in the bisulfite treated sequence would indicate the presence of methylation which protected the site from deamination.

Over exposure of DNA to bisulfite carries a high risk of degredading the DNA. Treatments should therefore be long enough to ensure maximal conversion, without generating sequencing artifacts due to degraded DNA template (Leontiou et al., 2015). The disadvantages of this method if used for whole genome study includes it being computationally intensive and requiring a reference genome of the organism in question. The advantage of this method is that the position and methylation context of all cytosine bases in a genome can be generated at per base pair resolution, making it the most comprehensive method for detecting DNA methylation (Table 2.4).

Table 2.4: The types of data that can be obtained using the different methods of investigating DNA methylation, and whether they are suitable for species for which no reference genome is available.

Criteria	**Method**		
	MSAP	Antibodies	Bisulfite sequencing
Level	YES	YES	YES
Context	NO	NO	YES
Strand	NO	NO	YES
Position	NO	NO	YES
Suitabile without reference genome	YES	YES	NO

The efficieny of the cytosine to thymine conversion can be calculated by adding non-methylated control DNA to samples. Completely unmethylated control DNA should be converted at every cytosine position if the treatment is 100% efficient.

Stellenbosch University https://scholar.sun.ac.za

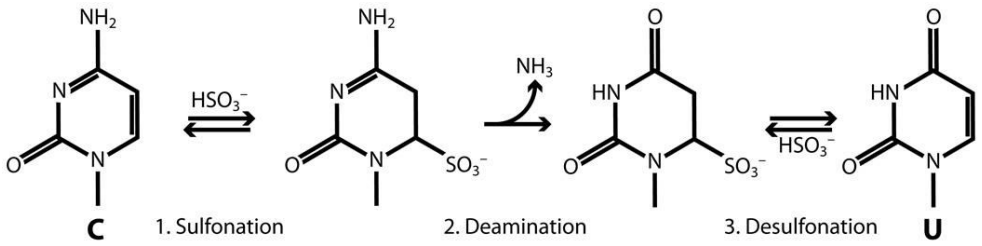

Figure 2.7: The process of bisulfite conversion (Darst et al., 2010).

Lambda phage DNA is commonly added as a control to samples, as viruses are unable to methyalte DNA, at a ratio of 0.1 – 0.5% (w/w) of the sample DNA (Grehl et al., 2018). A treatment efficiency of 95 – 98% can usualy be expected with minimal degredation (Warnecke et al., 2002), while an efficiency of 99.5 – 99.8% can be achieved by extending treatment duration (Grunau et al., 2001).

2.3.7: Indirect detection using the ratio of observed to expected CpG sites

It has also been shown by Suzuki et al. (2007), that the ratio of observed over expected CpG sites $CpG_{O/E}$ in a gene is inversely proportionate to the total number of methylated CpG sites within the gene (Suzuki and Bird, 2008). This is due to methylated CpG sites being prone to abolishment by cytosine-to-thymine mutations, which will lead to a depletion of CpG sites over time in highly methylated regions. The observed number of CpG sites in methylated genes are therefore often lower than expected, $CpG_{O/E} < 1.0$ (Suzuki and Bird, 2008). There are several explanations for the prevalence of deamination in methylated cytosine bases.

Firstly, the spontaneous deamination of methylated cytosine, *in vivo*, is ten times more likely than that of cytosine bases that are not methylated (Duncan and Miller, 1980; Zhang and Mathews, 1994). Secondly, the deamination of unmodified cytosine produces uracil, which due to it not naturally occurring in DNA, is effectively detected and corrected by DNA repair mechanisms (Krokan et al., 2002), while the deamination of methylated cytosine bases produces thymine (Morgan et al., 2004).The repair of the resulting thymine-guanidine mismatches, is much more error prone (Poole et al., 2001). Lastly, in the absence of S-adenosylmethionine which is the methyl group donor during methylation, DNMT enzymes are able to catalyse the deamination on the C4 position of nucleotides (Yebra and Bhagwat, 1995). Although unlikely, this means that CpG sites targeted for methylation might be deaminated to TpG instead (Shen et al., 1992).

Stellenbosch University https://scholar.sun.ac.za

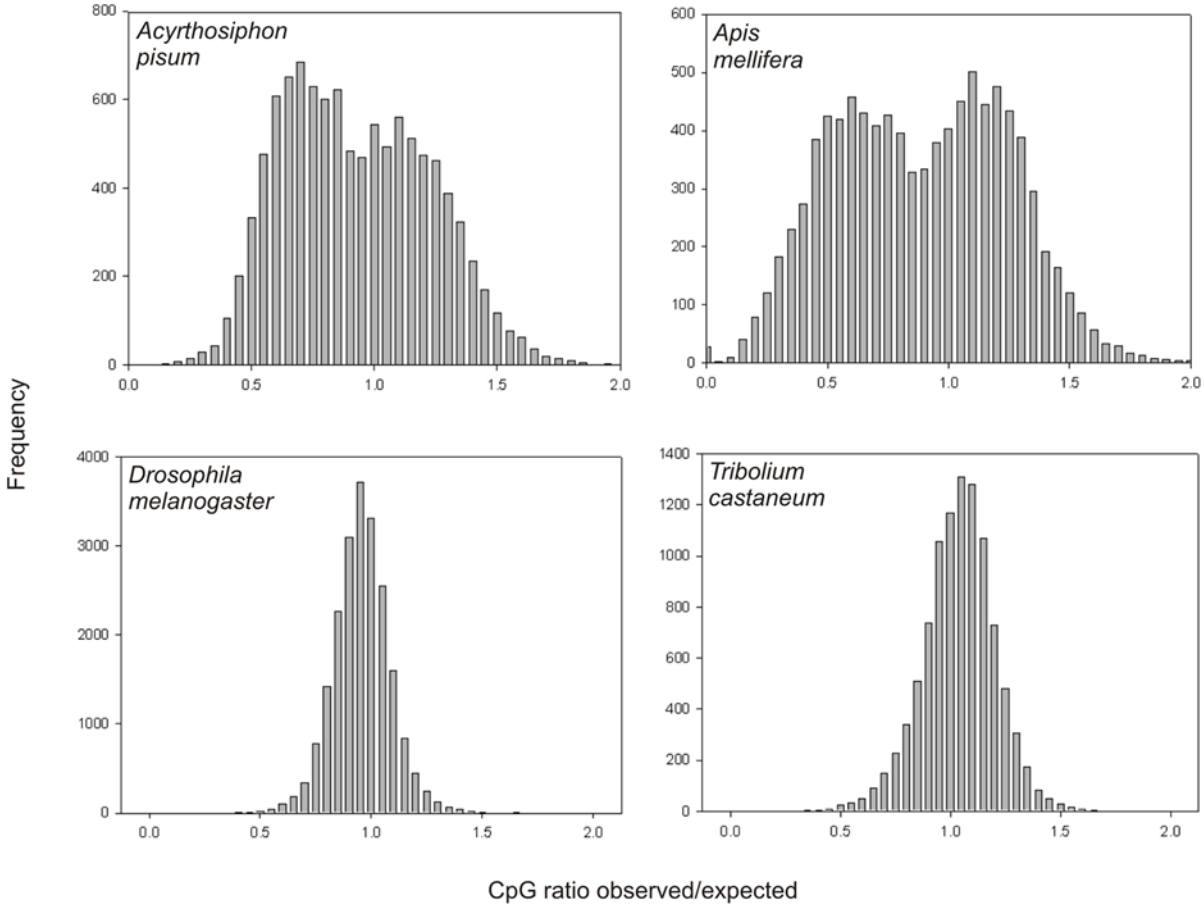

Figure 2.8: Density graphs showing the ratio of observed to expected CpG sites per gene in four insect species as calculated by the International Aphid Genomics Consortium (International Aphid Genomics Consortium, 2006).

The International Aphid Genomics Consortium (2006) calculated the $CpG_{O/E}$ for four insect species, using RefSeq sequence data to obtain gene sets, their results are shown in Figure 2.8. Two of these insects, the pea aphid and the honey bee, are known to have all of the insect *DNMT* genes, and display DNA methylation in genic areas (Walsh et al., 2010; Wang et al., 2006), while the fruit fly and red flour beetle, *Tribolium castaneum*, have incomplete *DNMT* sets and very limited DNA methylation (International Aphid Genomics Consortium, 2006). In species lacking a functional methylation system, the expected number of CpG sites per gene could be observed (i.e. the $CpG_{O/E}$ distribution peaked at about 1.0). The number of observed CpG sites in the gene sequences of species lacking a functional methylation system deviated from the expected and the $CpG_{O/E}$ distribution had two local maxima: one between 0.5 and 0.1 and another between 1.0 and 1.5. It was proposed that these two maxima respectively indicate a group of genes that are highly methylated and a group of genes that are not highly methylated (International Aphid Genomics Consortium, 2006).

Stellenbosch University https://scholar.sun.ac.za

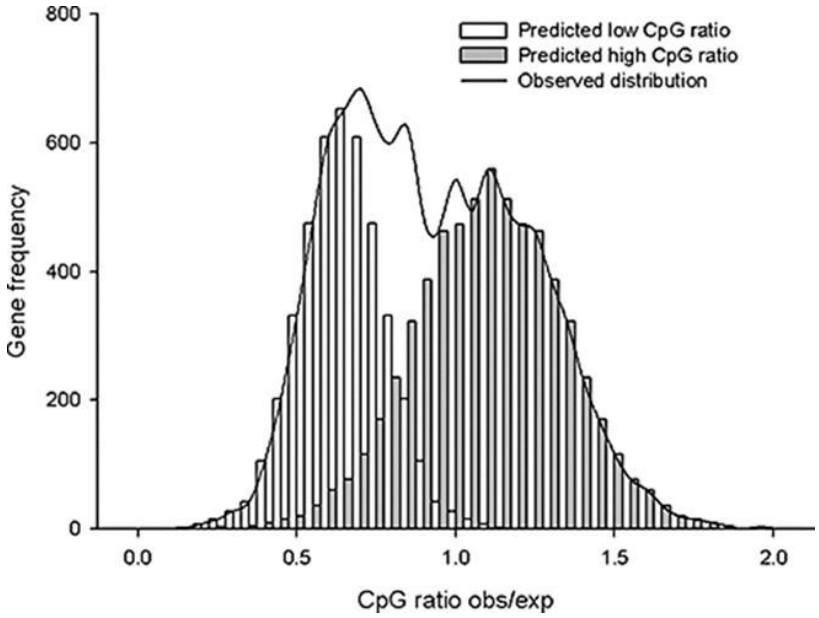

Figure 2.9: Obtained from Walsh et al. (2010), where it was attempted to describe the observed distribution of CpG sites per gene in the pea aphid genome, as the sum of two predicted distributions.

In order to prove the bimodality of the pea aphid distribution, Walsh et al. (2010) tested whether the observed distribution differed significantly from the sum of two predicted best fit distributions, one representing methylated genes, with low observed to expected ratios, and one representing genes that are not methylated, containing the expected number of methylated genes (Walsh et al., 2010) (Figure 2.9), they found however, that the observed distribution differed significantly from the combined distribution (X^2 = 51.0, df = 37, p-value < 0.05) this method could therefore not serve as a proof for bimodality. The only indicator they found for bimodality is a kurtosis value of -0.60, as negative kurtosis values have been associated with bimodal distributions. A much more appropriate value, the bimodality coefficient, can be calculated, which takes the sample size and skewness into account, along with the kurtosis value (Pfister et al., 2013).

2.3.8: Detecting hydroxymethylation

A drawback of all of the methods, described above, to detect methylation is that they cannot readily distinguish between 5mC and 5hmC (Kurdyukov and Bullock, 2016). However, variations on each of these methods have been described, facilitating the detection of 5hmC specifically.

Modified restriction enzyme-based methodes

The restriction enzyme Msp1, has a 5'-CC|GG-3' recognition site, and is methylation insensitive, meaning it restricts at both methylated, as well as unmodified CpG sites (Waalwijk and Flavell, 1978). The addition of a glycosyl group to the central cytosine base blocks the action of Msp1, in order to take advantage of this, DNA can be treated with T4 glycosyltransferase, which selectively glycosylates 5hmC, but not 5mC, leading to restriction only at hydroxymethylated CpG. Another

Stellenbosch University https://scholar.sun.ac.za

restriction enzyme, MspJ1, does not however restrict at unmodified cytosine bases, only at 5mC or 5hmC (Horton et al., 2014). Comparing the digestion of Msp1 after glycosylation with MspJ1 reveals hydroxymethylated CpG sites (Cohen-Karni et al., 2011).

Members of the PvuRts1 restriction enzyme family, such as AbaS1, can restrict at 5hmC as well as 5-glucosylhydroxymethylcytosine (5ghmC), but the addition of potassium acetate significantly inhibits the restriction at 5hmC. AbaS1 can therefore be used to find hydroxymethylated CpG sites, without the use of a second restriction enzyme (Wang et al., 2011).

Modified antibody-based methods

Two affinity-based methods are also available for 5hmC enrichment: J-binding protein, a DNA binding protein derived from *Crithidia fasciculate*, has affinity for 5hmC treated with β-glycosyltransferase (Heidebrecht et al., 2012). Antibodies specific for 5hmC can also be used in hydroxymethylated DNA immunoprecipitation (hmeDIP) (Nestor and Meehan, 2014). This method has already been used to determine the levels of hydroxymethylation in the four South-African RWA biotypes, the obtained are shown in Figure 2.10, SA2 showed significantly less hydroxymethylation than did SA1 and SA3, while SAM showed significantly higher levels (p-value < 0.05) (Breeds et al., 2018).

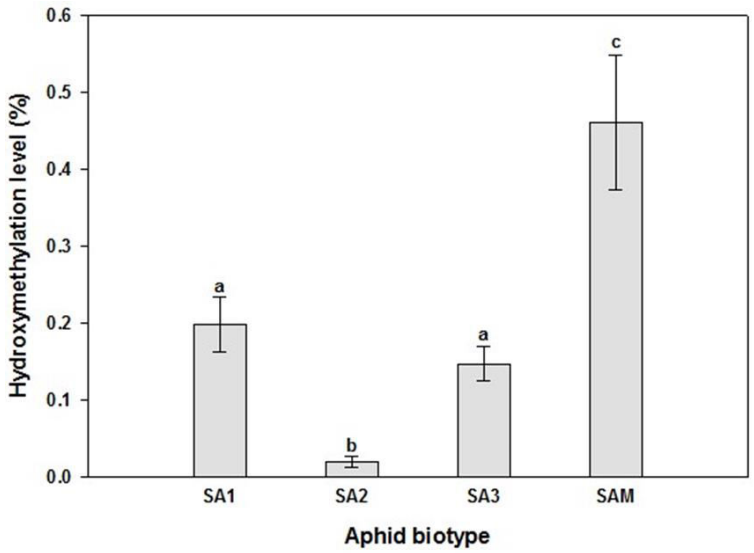

Figure 2.10: Hydroxymethylation levels in the four South-African RWA biotypes, letters indicate significant differences (Breeds et al., 2018).

Modified chemical treatment-based methods

TET-assisted bisulfite sequencing (Tab-seq) is a variation of bisulfite sequencing developed to map 5hmC. During this method, T4 glycosyltransferase is used to glycosylate 5hmC before treating the DNA with a TET enzyme. The glycosyl group prevents TET from binding 5ghmC, while 5mC is oxidised

Stellenbosch University https://scholar.sun.ac.za

to 5hmC followed by 5fC, and then by 5caC, 5ghmC remains unchanged. Treatment with Terminal deoxynucleotidyl transferase then converts 5caC and any remaining 5fC to unmodified cytosine (Ito et al., 2011). If bisulfite sequencing as explained earlier is now performed, any cytosine bases that remain unchanged after bisulfite treatment can be interpreted as 5hmC (Yu et al., 2012).

2.3.9: Analysing bisulfite sequencing data

The first step in analysing whole genome bisulfite sequencing is mapping of the sequenced reads back onto a reference genome. While frequently used alignment programs, such as Bowtie2 (Langmead and Salzberg, 2012) are available, they have not been optimised to align reads from bisulfite sequencing experiments. These reads will contain many polymorphisms, as cytosine bases are converted to thiamine after bisulfite treatment.

The Bismark software program (Krueger and Andrews, 2011) is a whole genome methylation calling program which is used together with an optimised implementation of the alignment program Bowtie 2 to extract methylation data from sequence reads of bisulfite treated samples. This program outcompetes other methylation calling programs such as BS Seeker (Chen et al., 2010), as it is able to run more instances of the alignment program Bowtie2 when analysing reads from a directional library. It supports single and paired-end reads, and the insert size and mapping parameters can be adjusted (Krueger and Andrews, 2011), while its simplistic output format also makes downstream manipulation of data easy.

Table 2.5: A comparison of the Bismark and BS Seeker methylation calling programs (Krueger and Andrews, 2011).

Feature	Program	
	Bismark	BS Seeker
Bowtie instances (directional/non-directional)	4/4	2/4
Single-end (SE)/paired-end (PE) support	Yes/Yes	Yes/No
Variable read length (SE/PE)	Yes/Yes	No/NA
Adjustable insert size (PE)	Yes	NA
Uses basecall qualities for FastQ mapping	Yes	No
Adjustable mapping parameters	5	2
Directional/non-directional library support	Yes/Yes	Yes/Yes

The Bismark pipeline can be broken into three distinct steps:

Stellenbosch University https://scholar.sun.ac.za

Genome preparation

The reference genome, as well as sequenced reads are independently converted into two different versions, one where all C's are converted to T's and one where all G's are converted to A's. These converted versions will also serve as reference genomes during alignment. Unmethylated DNA will have fewer mismatches when mapped to a converted reference as the reads and converted reference will both be enriched for thiamine. A G-to-A converted reference genome version is also included, as the complementary strand conversion of cytosine to thiamine is guanine to adenine, thus allowing using reads from either strand to detect methylation.

Alignment

Bismark invokes the Bowtie2 program to map the converted sequence reads (G-to-A and C-to-T) independently to the converted versions (G-to-A and C-to-T) of the reference genome. This results in four parallel outputs from which the best alignment can be determined, as an example, in Figure 2.11 thread (1) produces the best alignment as minimal mismatches occur.

Methylation extraction

By comparing the sequence reads with the original reference sequence, the methylation state of each cytosine base can be determined. In the example in Figure 2.11 only C-to-T substitutions are presented, but G-to-A substitutions are also applicable.

Stellenbosch University https://scholar.sun.ac.za

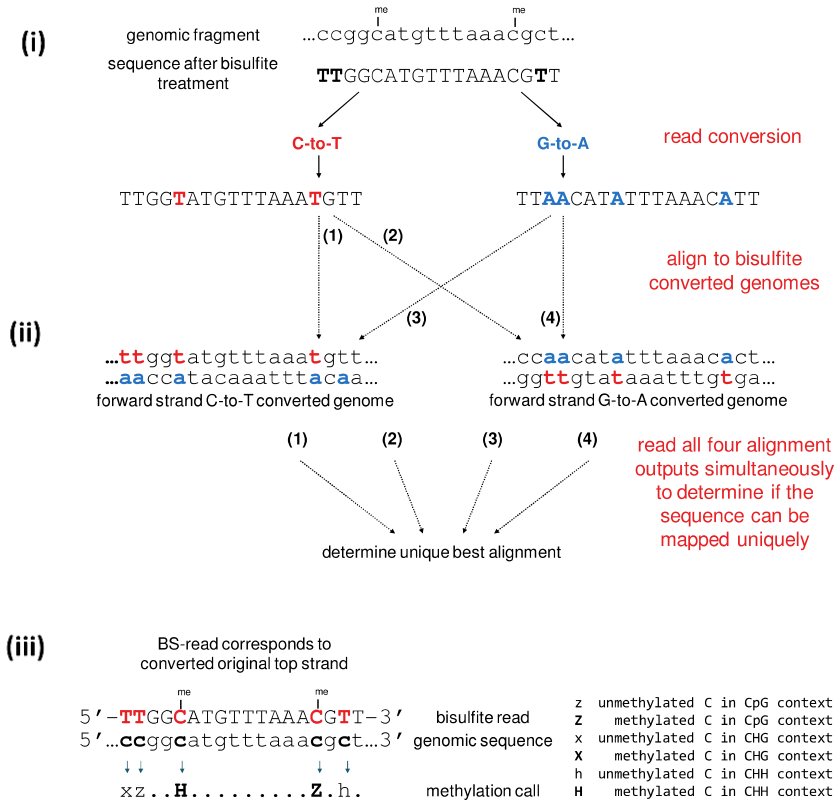

Figure 2.11: A basic outline of the three steps involved in the pipeline employed by the Bismark program, Image sourced from Krueger and Andrews (2011).

Before extracting the methylation data, M-bias plots for the uniquely mapped reads can be generated. M-bias plots are line graphs, showing the distribution of methylated calls across reads. This can be used to identify regions of reads that show a bias, either for methylated or unmethylated calls, which can then be excluded from further analysis (Krueger and Andrews, 2011).

After resolving methylation data to a per-base level, individual methylation sites can be used to identify regions which are rich in sites showing methylation differences. These are referred to as differentially methylated regions (DMRs), and are particularly important when investigating methylation-mediated gene regulation (Condon et al., 2018).

2.3.10: Differentially methylated regions

Various programs have been developed to test for differentially methylated regions, however for most of the tests applied by the various programs, replicates are required. Programs that do not require biological replicates do not include biologicaly significant data, such as spatial correlation and read depth (Wu et al., 2015).

A method incorporating spatial correlation, read depth and biological variation has been developed, specifically for experiments where only a single biological replicate of each sample is available, DSS-single (Dispersion Shrinkage for Sequencing data with single replicates). The bisulfite data is first

Stellenbosch University https://scholar.sun.ac.za

modelled as a beta-binomial distribution (Anscombe, 1950), a simple moving average is applied, and the dispersion parameters are estimated through an empirical Bayes procedure (Efron and Morris, 1972). This generates "pseudo-replicates" by incorporating the variance at nearby CpG sites. Finally a simple chi-squared test (Wald and Wolfowitz, 2019) is applied with a null hypothesis of no difference between samples (Wu et al., 2015).

2.4: Conclusion

The aim of this study was to generate whole-genome bisulfite sequencing (WGBS) data for Biotypes SA1 and SAM, and to process it into a useful format, which will serve as the first site specific overview of DNA methylation in the RWA; as well as to demonstrate how this data can be used to design future studies, by searching for differentially methylated genes and testing for a corresponding difference in transcription.

Burger et al. (2017) showed that moving RWA from one host plant to another resulted in major changes in transcription for both biotypes SA1 and SAM. It was observed that biotype SAM showed a greater ability to alter transcription in response to plant defenses. To investigate the role that DNA methylation might play in this difference in responsiveness between SA1 and SAM, host shifts will be performed prior to RNA extraction.

The Bismark methylation calling program was used as it outcompetes the other popular program, BS Seeker. Simple command line text manipulation was used to process the user-friendly output from Bismark into the final represented data.

With only a single biological repeat available for each biotype, the DSS-Single package was used to identify DMR's, as it was designed with that purpose in mind. Primers were designed for the most significant DMR's and transcription levels were quantified using quantitative PCR (qPCR).

The methylation data as well as DMR's were also compared with $CpG_{O/E}$ data to test the feasibility of using $CpG_{O/E}$ to infer methylation in RWA.

2.5: References

Anscombe, F.J., 1950, Sampling theory of the negative binomial and logarithmic series, *Biometrika*, 37(3), 358–382.

Bewick, A.J., Vogel, K.J., Moore, A.J. & Schmitz, R.J., 2017, Evolution of DNA methylation across insects, *Molecular Biology and Evolution*, 34(3), 654–665.

Bird, A., 2002, DNA methylation patterns and epigenetic memory, *Genes & Development*, 16(1), 6–21.

Botha, A-M., Swanevelder, Z.H. & Lapitan, N.L. V., 2010, Transcript profiling of wheat genes expressed during feeding by two different biotypes of *Diuraphis noxia*, *Environmental Entomology*, 39(4), 1206–1231.

Stellenbosch University https://scholar.sun.ac.za

Botha, A-M., 2013, A coevolutionary conundrum: The arms race between *Diuraphis noxia* (Kurdjumov) a specialist pest and its host *Triticum aestivum* (L.), *Arthropod-Plant Interactions*, 7(4), 359–372.

Botha, A-M., Nagel, M.A.C., van der Westhuizen, A.J. & Botha, F.C., 1998, Chitinase isoenzymes in near-isogenic wheat lines challenged with Russian wheat aphid, exogenous ethylene, and mechanical wounding, *Botanical Bulletin of Academia Sinica*, 39.

Breeds, K., Burger, N.F.V. & Botha, A.-M., 2018, New insights into the methylation status of virulent *Diuraphis noxia* (Hemiptera: Aphididae) biotypes, *Journal of Economic Entomology*, 111(3), 1395–1403.

Burger, N.F. V, Venter, E. & Botha, A-M., 2017, Profiling *Diuraphis noxia* (Hemiptera: Aphididae) transcript expression of the biotypes SA1 and SAM feeding on various *Triticum aestivum* varieties, *Journal of Economic Entomology*, 110(2), 692–701.

Burger, N.F.V. & Botha, A.-M., 2017, Genome of Russian wheat aphid an economically important cereal aphid, *Standards in Genomic Sciences*, 12(1), 90.

Chen, P.-Y., Cokus, S.J. & Pellegrini, M., 2010, BS Seeker: Precise mapping for bisulfite sequencing, *BMC Bioinformatics*, 11(1), 203.

Chen, Z.X. & Riggs, A.D., 2011, DNA methylation and demethylation in mammals, *Journal of Biological Chemistry*, 286(21), 18347–18353.

Cheng, X. & Blumenthal, R.M., 2008, Mammalian DNA methyltransferases: a structural perspective, *Structure*, 16(3), 341–350.

Cingolani, P., Cao, X., Khetani, R.S., Chen, C.C., Coon, M., Sammak, A., Bollig-Fischer, A., Land, S., Huang, Y., Hudson, M.E., Garfinkel, M.D., Zhong, S., Robinson, G.E. & Ruden, D.M., 2013, Intronic non-CG DNA hydroxymethylation and alternative mRNA splicing in honey bees, *BMC Genomics*, 14(1), 66.

Cohen-Karni, D., Xu, D., Apone, L., Fomenkov, A., Sun, Z., Davis, P.J., Kinney, S.R.M., Yamada-Mabuchi, M., Xu, S., Davis, T., Pradhan, S., Roberts, R.J. & Zheng, Y., 2011, The MspJI family of modification-dependent restriction endonucleases for epigenetic studies, *Proceedings of the National Academy of Sciences of the United States of America*, 108(27), 11040–11045.

Colorado Agricultural Statistics Service, 2004, *Winter Wheat Varieties - 2004 Crop*.

Condon, D.E., Tran, P. V, Lien, Y.-C., Schug, J., Georgieff, M.K., Simmons, R.A. & Won, K.-J., 2018, Defiant: (DMRs: easy, fast, identification and ANnoTation) identifies differentially Methylated regions from iron-deficient rat hippocampus, *BMC Bioinformatics*, 19(1), 31.

Cukier, H.N., Perez, A.M., Collins, A.L., Zhou, Z., Zoghbi, H.Y. & Botas, J., 2008, Genetic modifiers of MeCP2 function in *Drosophila*, *PLoS Genetics*, 4(9), e1000179.

Darst, R.P., Pardo, C.E., Ai, L., Brown, K.D. & Kladde, M.P., 2010, Bisulfite Sequencing of DNA, *Current Protocols in Molecular Biology*, 91(1), 7–9.

Deobagkar, D.D., Liebler, M., Graessmann, M. & Graessmann, A., 1990, Hemimethylation of DNA prevents chromatin expression, *Proceedings of the National Academy of Sciences of the United States of America*, 87(5), 1691–1695.

Stellenbosch University https://scholar.sun.ac.za

Dombrovsky, A., Arthaud, L., Ledger, T.N., Tares, S. & Robichon, A., 2009, Profiling the repertoire of phenotypes influenced by environmental cues that occur during asexual reproduction, *Genome Research*, 19(11), 2052–2063.

du Toit, F., 1989, Inheritance of resistance in two *Triticum aestivum* lines to Russian wheat aphid (Hornoptera: Aphididae), *Journal of Economic Entomology*, 82(4), 1251–1253.

du Toit, F. & Walters, M.C., 1981, Damage assessment and economic threshold values for the chemical control of the Russian wheat aphid, *Diuraphis noxia* (Mordvilko) on winter wheat, *Technical communication, Department of Agriculture, Republic of South Africa*, (191), 58–62.

Duncan, B.K. & Miller, J.H., 1980, Mutagenic deamination of cytosine residues in DNA, *Nature*, 287(5782), 560–561.

Dupont, C., Armant, D.R. & Brenner, C.A., 2009, Epigenetics: definition, mechanisms and clinical perspective, *Seminars in Reproductive Medicine*, 27(5), 351–357.

Efron, B. & Morris, C., 1972, Limiting the risk of Bayes and empirical Bayes estimators—Part II: The empirical Bayes case, *Journal of the American Statistical Association*, 67(337), 130–139.

Elango, N., Hunt, B.G., Goodisman, M.A.D. & Yi, S.V., 2009, DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, *Apis mellifera*, *Proceedings of the National Academy of Sciences of the United States of America*, 106(27), 11206–11211.

Feng, S., Cokus, S.J., Zhang, X., Chen, P.Y., Bostick, M., Goll, M.G., Hetzel, J., Jain, J., Strauss, S.H., Halpern, M.E., Ukomadu, C., Sadler, K.C., Pradhan, S., Pellegrini, M. & Jacobsen, S.E., 2010, Conservation and divergence of methylation patterning in plants and animals, *Proceedings of the National Academy of Sciences of the United States of America*, 107(19), 8689–8694.

Fouché, A., Verhoeven, R.L., Hewitt, P.H., Kriel, C.F., de Jager, J. & Walters, M.C., 1984, Russian aphid (*Diuraphis noxia*) feeding damage on wheat, related cereals and a *Bromus* grass species, *Technical Communication, Department of Agriculture, Republic of South Africa*, (191), 22–33.

Fouse, S.D., Nagarajan, R.O. & Costello, J.F., 2010, Genome-scale DNA methylation analysis, *Epigenomics*, 2(1), 105–117.

Glastad, K.M., Hunt, B.G. & Goodisman, M.A.D., 2014, Evolutionary insights into DNA methylation in insects, *Current Opinion in Insect Science*, 1, 25–30.

Glastad, K.M., Hunt, B.G. & Goodisman, M.A.D., 2015, DNA methylation and chromatin organization in insects: Insights from the ant *Camponotus floridanus*, *Genome Biology and Evolution*, 7(4), 931–942.

Glastad, K.M., Hunt, B.G., Yi, S. V & Goodisman, M.A.D., 2011, DNA methylation in insects: On the brink of the epigenomic era, *Insect Molecular Biology*, 20(5), 553–565.

Goll, M.G. & Bestor, T.H., 2005, Eukaryotic cytosine methyltransferases, *Annual Review of Biochemistry*, 74, 481–514.

Goll, M.G., Kirpekar, F., Maggert, K.A., Yoder, J.A., Hsieh, C.-L., Zhang, X., Golic, K.G., Jacobsen, S.E. & Bestor, T.H., 2006, Methylation of tRNA-Asp by the DNA methyltransferase homolog *Dnmt2*, *Science*, 311(5759), 395–398.

Stellenbosch University https://scholar.sun.ac.za

Gong, L., Cui, F., Sheng, C., Lin, Z., Reeck, G., Xu, J. & Kang, L., 2012, Polymorphism and methylation of four genes expressed in salivary glands of Russian wheat aphid (Homoptera: Aphididae), *Journal of Economic Entomology*, 105(1), 232–241.

González, D., Summers, C.G. & Qualset, C.O., 1992, Russian wheat aphid: natural enemies, resistant wheat offer potential control, *California Agriculture*, 46(1), 32–34.

Grehl, C., Kuhlmann, M., Becker, C., Glaser, B. & Grosse, I., 2018, How to design a whole-genome bisulfite sequencing experiment, *Epigenomes*, 2(4), 21.

Grunau, C., Clark, S.J. & Rosenthal, A., 2001, Bisulfite genomic sequencing: Systematic investigation of critical experimental parameters, *Nucleic Acids Research*, 29(13), 65.

Haley, S.D., Peairs, F.B., Walker, C.B., Rudolph, J.B. & Randolph, T.L., 2004, Occurrence of a new Russian wheat aphid biotype in Colorado, *Crop Science*, 44(5), 1589–1592.

Heidebrecht, T., Fish, A., Von Castelmur, E., Johnson, K.A., Zaccai, G., Borst, P. & Perrakis, A., 2012, Binding of the J-binding protein to DNA containing glucosylated hmU (base J) or 5-hmC: Evidence for a rapid conformational change upon DNA binding, *Journal of the American Chemical Society*, 134(32), 13357–13365.

Holohan, E.E., Kwong, C., Adryan, B., Bartkuhn, M., Herold, M., Renkawitz, R., Russell, S. & White, R., 2007, CTCF genomic binding sites in *Drosophila* and the organisation of the bithorax complex, *PLoS Genetics*, 3(7), 112.

Horton, J.R., Wang, H., Mabuchi, M.Y., Zhang, X., Roberts, R.J., Zheng, Y., Wilson, G.G. & Cheng, X., 2014, Modification-dependent restriction endonuclease, MspJI, flips 5-methylcytosine out of the DNA helix, *Nucleic Acids Research*, 42(19), 12092–12101.

Hunt, B.G., Brisson, J.A., Yi, S. V. & Goodisman, M.A.D., 2010, Functional conservation of DNA methylation in the pea aphid and the honeybee, *Genome Biology and Evolution*, 2(1), 719–728.

International Aphid Genomics Consortium, 2006, Insights into social insects from the genome of the honeybee *Apis mellifera*, *Nature*, 443(7114), 931–949.

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C. & Zhang, Y., 2011, Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine, *Science*, 333(6047), 1300–1303.

Jankielsohn, A., 2011, Distribution and diversity of Russian wheat aphid (Hemiptera: Aphididae) biotypes in South Africa and Lesotho, *Journal of Economic Entomology*, 104(5), 1736–1741.

Jankielsohn, A., 2019, New Russian wheat aphid biotype found in Free State, *Grain SA Mini Focus, Pest Control on Winter Cereals*.

Jin, B., Li, Y. & Robertson, K.D., 2011, DNA methylation: Superior or subordinate in the epigenetic hierarchy?, *Genes & Cancer*, 2(6), 607–617.

Jin, S.-G., Kadam, S. & Pfeifer, G.P., 2010, Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine., *Nucleic Acids Research*, 38(11), e125.

Kim, D., Minhas, B.F., Li-Byarlay, H. & Hansen, A.K., 2018, Key transport and ammonia recycling genes involved in aphid symbiosis respond to host-plant specialization, *Genes, Genomes, Genetics*, 8(7), 2433–2443.

Stellenbosch University https://scholar.sun.ac.za

Kohli, R.M. & Zhang, Y., 2013, TET enzymes, TDG and the dynamics of DNA demethylation, *Nature*, 502(7472), 472–479.

Krokan, H.E., Drablos, F. & Slupphaug, G., 2002, Uracil in DNA - occurrence, consequences and repair, *Oncogene*, 21(58), 8935–8948.

Krueger, F. & Andrews, S.R., 2011, Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications, *Bioinformatics*, 27(11), 1571–1572.

Kumar, S. & Kim, Y., 2017, An endoparasitoid wasp influences host DNA methylation, *Scientific Reports*, 7, 43287.

Kurdyukov, S. & Bullock, M., 2016, DNA methylation analysis: Choosing the right method, *Biology*, 5(1), 3.

Langmead, B. & Salzberg, S.L., 2012, Fast gapped-read alignment with Bowtie 2, *Nature Methods*, 9(4), 357–359.

Leontiou, C.A., Hadjidaniel, M.D., Mina, P., Antoniou, P., Ioannides, M. & Patsalis, P.C., 2015, Bisulfite conversion of DNA: Performance comparison of different kits and methylation quantitation of epigenetic biomarkers that have the potential to be used in non-invasive prenatal testing, *PloS One*, 10(8), e0135058.

Li, E. & Zhang, Y., 2014, DNA Methylation in Mammals, *Cold Spring Harbor Perspectives in Biology*, 6(5), a019133.

Li, Y. & Tollefsbol, T.O., 2011, DNA methylation detection: Bisulfite genomic sequencing analysis, *Epigenetics Protocols*, pp. 11–21.

Lyko, F. & Maleszka, R., 2011, Insects as innovative models for functional studies of DNA methylation, *Trends in Genetics*, 27(4), 127–131.

Marhold, J., Rothe, N., Pauli, A., Mund, C., Kuehle, K., Brueckner, B. & Lyko, F., 2004, Conservation of DNA methylation in dipteran insects, *Insect Molecular Biology*, 13(2), 117–123.

Martin, G.B., Bogdanove, A.J. & Sessa, G., 2003, Understanding the functions of plant disease resistance proteins, *Annual Review of Plant Biology*, 54(1), 23–61.

Mathers, T.C., Mugford, S.T., Percival-Alwyn, L., Chen, Y., Kaithakottil, G., Swarbreck, D., Hogenhout, S.A. & van Oosterhout, C., 2019, Sex-specific changes in the aphid DNA methylation landscape, BioRxiv, 286320.

Maunakea, A.K., Chepelev, I., Cui, K. & Zhao, K., 2013, Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition, *Cell Research*, 23(11), 1256–1269.

McClelland, M., Nelson, M. & Raschke, E., 1994, Effect of site-specific modification on restriction endonucleases and DNA modification methyltransferases, *Nucleic Acids Research*, 22(17), 3640–3659.

Moison, C., Guieysse-Peugeot, A.-L. & Arimondo, P.B., 2013, DNA methylation in cancer, *Atlas of Genetics and Cytogenetics in Oncology and Haematology*.

Morgan, H.D., Dean, W., Coker, H.A., Reik, W. & Petersen-Mahrt, S.K., 2004, Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: Implications for epigenetic reprogramming, *The Journal of Biological Chemistry*, 279(50), 52353–52360.

Nakabachi, A., Shigenobu, S., Sakazume, N., Shiraki, T., Hayashizaki, Y., Carninci, P., Ishikawa, H., Kudo, T. & Fukatsu, T., 2005, Transcriptome analysis of the aphid bacteriocyte, the symbiotic host cell that harbors

Stellenbosch University https://scholar.sun.ac.za

an endocellular mutualistic bacterium, *Buchnera*, *Proceedings of the National Academy of Sciences of the United States of America*, 102(15), 5477–5482.

Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., Maldotti, M., Anselmi, F. & Oliviero, S., 2017, Intragenic DNA methylation prevents spurious transcription initiation., *Nature*, 543(7643), 72–77.

Nestor, C.E. & Meehan, R.R., 2014, Hydroxymethylated DNA immunoprecipitation (hmeDIP), *Methods in Molecular Biology*, 1094, 259–267.

Nicholas, A., Puterka, G. & Gopurenko, D., 2015, *Diuraphis noxia* Russian wheat aphid, *National Diagnostic Protocol*, 1(28).

Nkongolo, K.K., Quick, J.S., Limin, A.E. & Fowler, D.B., 1991, Sources and inheritance of resistance to Russian wheat aphid in *Triticum* Species *amphiploids* and *Triticum tauschii*, *Canadian Journal of Plant Science*, 71(3), 703–708.

Novotná, J., Havelka, J., Starý, P., Koutecký, P. & Vítková, M., 2011, Karyotype analysis of the Russian wheat aphid, *Diuraphis noxia* (Kurdjumov) (Hemiptera: Aphididae) reveals a large X-chromosome with rRNA and histone gene families, *Genetica*, 139(3), 281–289.

Okano, M., Xie, S. & Li, E., 1998, Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases, *Nature Genetics*, 19(3), 219–220.

Okitsu, C. Y., & Hsieh, C. L. (2015). Sensitivity and specificity of immunoprecipitation of DNA containing 5-Methylcytosine, *BMC Research Notes*, 8(1), 102.

Panikar, C.S., Rajpathak, S.N., Abhyankar, V., Deshmukh, S. & Deobagkar, D.D., 2015, Presence of DNA methyltransferase activity and CpC methylation in *Drosophila melanogaster*, *Molecular Biology Reports*, 42(12), 1615–1621.

Pasquier, C., Clément, M., Dombrovsky, A., Penaud, S., Rocha, M. Da, Rancurel, C., Ledger, N., Capovilla, M. & Robichon, A., 2014, Environmentally selected aphid variants in clonality context display differential patterns of methylation in the genome, *PloS One*, 9(12), e115022.

Pegoraro, M., Bafna, A., Davies, N.J., Shuker, D.M. & Tauber, E., 2016, DNA methylation changes induced by long and short photoperiods in *Nasonia*, *Genome Research*, 26(2), 203–210.

Pfister, R., Schwarz, K.A., Janczyk, M., Dale, R. & Freeman, J.B., 2013, Good things peak in pairs: A note on the bimodality coefficient, *Frontiers in Psychology*, 4, 700.

Poole, A., Penny, D. & Sjoberg, B.M., 2001, Confounded cytosine! Tinkering and the evolution of DNA, *Nature Reviews. Molecular Cell Biology*, 2(2), 147–151.

Purchase, J.L., Le Roux, J. & Hatting, H., 1996, Progress of the Southern African Regional Wheat Evaluation and Improvement Nursery (SARWEIN) from 1974 to 1993, *The Ninth Regional Wheat Workshop: For Eastern, Central and Southern Africa*, Addis Ababa, Ethiopia, October 2-6, 1995.

Puterka, G.J., Nicholson, S.J., Brown, M.J., Cooper, W.R., Peairs, F.B. & Randolph, T.L., 2014, Characterization of eight Russian wheat aphid (Hemiptera: Aphididae) biotypes using two-category resistant-susceptible plant responses, *Journal of Economic Entomology*, 107(3), 1274–1283.

Quick, J.S., Ellis, G.E., Normann, R.M., Stromberger, J.A., Shanahan, J.F., Peairs, F.B., Rudolph, J.B. & Lorenz, K., 1996, Registration of “Halt” wheat, *Crop Science*, 36(1), 209–210.

Stellenbosch University https://scholar.sun.ac.za

Quick, J.S., Nkongolo, K.K., Meyer, W., Pearis, F.B. & Weaver, B., 1991, Russian wheat aphid reaction and agronomic and quality traits of a resistant wheat, *Crop Science*, 31(1), 50–53.

Randolph, T.L., Peairs, F., Weiland, A., Rudolph, J.B. & Puterka, G.J., 2009, Plant responses to seven Russian wheat aphid (Hemiptera: Aphididae) biotypes found in the United States, *Journal of Economic Entomology*, 102(5), 1954–1959.

Reyna-Lopez, G.E., Simpson, J. & Ruiz-Herrera, J., 1997, Differences in DNA methylation patterns are detectable during the dimorphic transition of fungi by amplification of restriction polymorphisms, *Molecular & General Genetics*, 253(6), 703–710.

Russo, V.E.A., Martienssen, R.A. & Riggs, A.D., 1996, *Epigenetic mechanisms of gene regulation*, Cold Spring Harbor Laboratory Press.

Saunders, D.S., 1965, Larval diapause of maternal origin: Induction of diapause in *Nasonia vitripennis* (Walk.) (Hymenoptera: Pteromalidae), *Journal of Experimental Biology*, 42(3), 495–508.

Shen, J.C., Rideout, W.M. & Jones, P.A., 1992, High frequency mutagenesis by a DNA methyltransferase, *Cell*, 71(7), 1073–1080.

Smith, M.C., Schotzko, D.J., Zemetra, R.S. & Souza, E.J., 1992, Categories of resistance in plant introductions of wheat resistant to the Russian wheat aphid (Homoptera: Aphididae), *Journal of Economic Entomology*, 85(4), 1480–1484.

Spruijt, C.G., Gnerlich, F., Smits, A.H., Pfaffeneder, T., Jansen, P.W.T.C., Bauer, C., Munzel, M., Wagner, M., Muller, M., Khan, F., Eberl, H.C., Mensinga, A., Brinkman, A.B., Lephikov, K., Muller, U., Walter, J., Boelens, R., van Ingen, H., Leonhardt, H., Carell, T. & Vermeulen, M., 2013, Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives, *Cell*, 152(5), 1146–1159.

Standage, D.S., Berens, A.J., Glastad, K.M., Severin, A.J., Brendel, V.P. & Toth, A.L., 2016, Genome, transcriptome and methylome sequencing of a primitively eusocial wasp reveal a greatly reduced DNA methylation system in a social insect, *Molecular Ecology*, 25(8), 1769–1784.

Stoetzel, M.B., 1987, Information on and Identification of *Diuraphis noxia* (Homoptera: Aphididae) and Other Aphid Species Colonizing Leaves of Wheat and Barley in the United States, *Journal of Economic Entomology*, 80(3), 696–704.

Suzuki, M.M. & Bird, A., 2008, DNA methylation landscapes: Provocative insights from epigenomics, *Nature Reviews Genetics*, 9(6), 465.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R. & Aravind, L., 2009, Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1, *Science*, 324(5929), 930–935.

Tolmay, V., Lindeque, R. & Prinsloo, G., 2007, Preliminary evidence of a resistance-breaking biotype of the Russian wheat aphid, *Diuraphis noxia* (Kurdjumov) (Homoptera: Aphididae), in South Africa, *African Entomology*, 15, 228–230.

Truong, M., Yang, B., Wagner, J., Desotelle, J. & Jarrard, D.F., 2013, Analysis of promoter non-CG methylation in prostate cancer, *Epigenomics*, 5(1), 65–71.

Stellenbosch University https://scholar.sun.ac.za

van Zyl, R.A. & Botha, A-M., 2008, Eliciting proteins from Diuraphis noxia biotypes differ in size and composition, *18th Biennial International Plant Resistance to Insects Workshop*, Fort Collins, Colorado.

von Wechmar, M.B. & Rybicki, E.P., 1981, Aphid transmission of three viruses causes Freestate streak disease, *South African Journal of Science*, 77(11), 488–492.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J. & Kuiper, M., 1995, AFLP: A new technique for DNA fingerprinting, *Nucleic Acids Research*, 23(21), 4407–4414.

Waalwijk, C. & Flavell, R.A., 1978, MspI, an isoschizomer of HpaII which cleaves both unmethylated and methylated HpaII sites, *Nucleic Acids Research*, 5(9), 3231–3236.

Wald, A. & Wolfowitz, J., 2019, Statistical tests based on permutations of the observations, *The Annals of Mathematical Statistics*, 15(4), 358–372.

Walsh, T.K., Brisson, J.A., Robertson, H.M., Gordon, K., Jaubert-Possamai, S., Tagu, D. & Edwards, O.R., 2010, A functional DNA methylation system in the pea aphid, *Acyrthosiphon pisum*, *Insect Molecular Biology*, 12(2).

Walters, M.C., Penn, F., du Toit, T.C. & Aalberberg, K., 1980, The Russian wheat aphid, *Farming in South Africa Leaflet Series*, (G.3), 1–6.

Wang, H., Guan, S., Quimby, A., Cohen-Karni, D., Pradhan, S., Wilson, G., Roberts, R.J., Zhu, Z. & Zheng, Y., 2011, Comparative characterization of the PvuRts1I family of restriction enzymes and their application in mapping genomic 5-hydroxymethylcytosine, *Nucleic Acids Research*, 39(21), 9294–9305.

Wang, Y., Jorda, M., Jones, P.L., Maleszka, R., Ling, X., Robertson, H.M., Mizzen, C.A., Peinado, M.A. & Robinson, G.E., 2006, Functional CpG methylation system in a social insect, *Science*, 314(5799), 645–647.

Warnecke, P.M., Stirzaker, C., Song, J., Grunau, C., Melki, J.R. & Clark, S.J., 2002, Identification and resolution of artifacts in bisulfite sequencing, *Methods*, 27(2), 101–107.

Weber, M., Davies, J.J., Wittig, D., Oakeley, E.J., Haase, M., Lam, W.L. & Schubeler, D., 2005, Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells, *Nature Genetics*, 37(8), 853–862.

Weiland, A.A., Peairs, F.B., Randolph, T.L., Rudolph, J.B., Haley, S.D. & Puterka, G.J., 2008, Biotypic diversity in Colorado Russian wheat aphid (Hemiptera: Aphididae) populations, *Journal of Economic Entomology*, 101(2006), 569–574.

Werren, J.H., Richards, S., Desjardins, C.A., Niehuis, O., Gadau, J. & Colbourne, J.K., 2010, Functional and evolutionary insights from the genomes of three parasitoid *Nasonia*, *Science*, 327(5963), 343–348.

Wojciechowski, M., Rafalski, D., Kucharski, R., Misztal, K., Maleszka, J., Bochtler, M. & Maleszka, R., 2014, Insights into DNA hydroxymethylation in the honeybee from in-depth analyses of TET dioxygenase, *Open Biology*, 4(8), 140110.

Wu, H., Xu, T., Feng, H., Chen, L., Li, B., Yao, B., Qin, Z., Jin, P. & Conneely, K.N., 2015, Detection of differentially methylated regions from whole-genome bisulfite sequencing data without replicates, *Nucleic Acids Research*, 43(21), 141.

Yan, H., Bonasio, R., Simola, D.F., Liebig, J., Berger, S.L. & Reinberg, D., 2015, DNA methylation in social insects: How epigenetics can control behavior and longevity, *Annual Review of Entomology*, 60, 435–452.

Stellenbosch University https://scholar.sun.ac.za

Yebra, M.J. & Bhagwat, A.S., 1995, A cytosine methyltransferase converts 5-methylcytosine in DNA to thymine, *Biochemistry*, 34(45), 14752–14757.

Yoder, J.A., Soman, N.S., Verdine, G.L. & Bestor, T.H., 1997, DNA (cytosine-5)-methyltransferases in mouse cells and tissues. Studies with a mechanism-based probe, *Journal of Molecular Biology*, 270(3), 385–395.

Yu, M., Hon, G.C., Szulwach, K.E., Song, C.-X., Zhang, L., Kim, A., Li, X., Dai, Q., Shen, Y., Park, B., Min, J.-H., Jin, P., Ren, B. & He, C., 2012, Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome, *Cell*, 149(6), 1368–1380.

Zhang, J., Xing, Y., Li, Y., Yin, C., Ge, C. & Li, F., 2015, DNA methyltransferases have an essential role in female fecundity in brown planthopper, *Nilaparvata lugens*, *Biochemical and Biophysical Research Communications*, 464(1), 83–88.

Zhang, X. & Mathews, C.K., 1994, Effect of DNA cytosine methylation upon deamination-induced mutagenesis in a natural target sequence in duplex DNA, *The Journal of Biological Chemistry*, 269(10), 7066–7069.

Stellenbosch University https://scholar.sun.ac.za

Chapter 3: Whole genome bisulfite sequencing of two Russian wheat aphid biotypes

3.1: Introduction

The Russian wheat aphid (*Diuraphis noxia*, Kurdjumov) is a major pest of wheat and barley. Crop losses due to Russian wheat aphid (RWA) feeding in the wheat producing regions of South-Africa were first recorded in 1978 (Walters et al., 1980). Resistant cultivars were developed to combat the new threat, with the resistance factors that these cultivars possess being referred to as *Dn* genes (du Toit, 1989). Resistance has steadily been breaking down however, as four new biotypes have since been recorded (SA2 to SA5), each more virulent than the last (Jankielsohn, 2019). RWA biotypes are not defined based on genotypic variation, or on geographical location, but based on the response elicited on a range of cultivars, each containing a different *Dn*-gene (Tolmay et al., 2007). Biotypes that can successfully infest a larger number of cultivars during such a differential screening are deemed more virulent.

A laboratory contained biotype, SAM also exits, which is more virulent than any of the naturally occurring biotypes. SAM is the result of a selective breeding program wherein SA1 was maintained on resistant Tugela *Dn*1 wheat for 78 generations (van Zyl, 2007). Given that SA1 and SAM are positioned on opposite extremes of the virulence scale and that SA1 is the known progenitor of SAM, they can be considered a very useful model for the study of biotypification in aphids. The genealogical link between SA1 and SAM is particularly helpful, as it has not been elucidated whether the other naturally occurring biotypes developed from SA1 or were introduced to South-Africa independently (van Zyl and Botha, 2008).

Epigenetic modifications, mechanisms that modify phenotype without altering the genomic sequence of an organism, have been shown to be involved in various aphid traits, such as winged morph production, insect resistance, and pigmentation (Hick et al., 1996; Ishikawa and Miura, 2009). Given the low level of genetic variation between SA1 and SAM, 0.0008% in coding sequence (Burger and Botha, 2017), the possibility that an epigenetic modification, specifically DNA methylation, might play a role in RWA virulence has been proposed (Breeds et al., 2018).

DNA methylation is a chemical modification to cytosine bases, specifically the addition of a methyl group to the 5-carbon position (Hotchkiss, 1948). Methylation has been studied most extensively in mammals, in which the methylation occurs almost exclusively on cytosine bases in a CpG context (Truong et al., 2013). In plant and insect models non-CpG contexts, such as CHG and CHH, are methylated, but at lower levels than that of the CpG context (Feng et al., 2010; Gong et al., 2012).

Stellenbosch University https://scholar.sun.ac.za

Another difference between mammalian and insect methylation is that insect methylation is more abundant in coding regions, while non-coding, promoter regions are the primary methylation target in mammals (Bewick et al., 2017). Methylation can influence gene expression and exon inclusion (Glastad et al., 2014; Maunakea et al., 2013) and has been a driving force for divergence, as highly methylated regions are prone to mutation (Holliday and Grigg, 1993).

A previous study on the levels of methylation in SA1 and SAM showed higher methylation in the more virulent biotype SAM, although the difference was not statistically significant(Breeds et al., 2018) the methods used however, do not provide positional information and can thus not be used for the identification of differentially methylated genes or regions. The objective of this study was to generate and process whole genome bisulfite sequencing data for biotypes SA1 and SAM into a useful format, and to give an overview of the methylation of each. The processed methylation data could be valuable in future studies to investigate the role that methylation might play in the observed difference in virulence between these biotypes.

3.2: Materials and methods

3.2.1: Host plant cultivation

All cultivars were planted in 12.5 cm pots filled with light expanded clay aggregate (LECA) medium. The plants were well watered with a modified Hoagland's solution (Hoagland and Arnon, 1950); with final concentrations of 6.4 mM KNO_3, 4 mM $Ca(NO_3)_2.4H_2O$, 2 mM $NH_4H_2PO_4$, 2 mM $MgSO_4.7H_2O$, 45 µM $FeCl_2.4H_2O$, 201 µM EDTA, 5 µM SiO_2, 0.5 µM $MnCl_2.4H_2O$, 0.2 µM $ZnSO_4.7H_2O$, 0.2 µM $CuSO_4.5H_2O$, 4.6 µM H_3BO_3, 0.1 µM $(NH_4)_6Mo_7O_{24}.4H_2O$, and grown inside a temperature controlled growth room, at 20 ± 1°C, on a twelve-hour day-night cycle, under a combination of fluorescent and LED grow lights, with a plant density of eight plants per pot.

3.2.2: Aphid rearing

Colonies of parthenogenetic, apterous, female aphids of South African RWA biotypes SA1, and SAM, expressing different levels of virulence, were separately established in BugDorm cages (MegaView Science Education Services) in an insectary with the following conditions: 22.5 ± 2.5°C, 40% relative humidity, and continuous artificial lighting from high pressure sodium lamps. The Russian wheat aphid biotypes were maintained on near isogenic wheat lines Tugela, RWA susceptible, and Tugela *Dn1*, RWA resistant, with SA1 maintained on the susceptible wheat line, and SAM on wheat containing the *Dn1* resistance gene. To avoid any differences in methylation as an artefact of the biotypes feeding on different wheat varieties, both biotypes were transferred to the susceptible cultivar, SST362, one month prior to DNA extraction.

Stellenbosch University https://scholar.sun.ac.za

3.2.3: Whole genome bisulfite sequencing

The high throughput nature of bisulfite sequencing, as well as the ability of this technique to study the methylation of cytosine bases in any sequence context, along with the availability of the SAM reference genome (GenBank ID GCA_001465515.1; BioProject PRJNA297165), made bisulfite sequencing the ideal method to assess the differences in methylation between these two biotypes.

3.2.3.1: DNA extraction.

A total of 100 apterous female aphids of South African RWA biotypes SA1 and SAM were used for DNA extraction. Aphids were flash frozen in liquid nitrogen and grinded with a micro-pestle. The Qiagen DNeasy Blood & Tissue Kit™, employing the manufacturer's recommended protocol for insect DNA extraction, was then used to extract high quality genomic DNA from both biotypes. DNA quantity was assessed through use of a Qubit 2.0 fluorometer (Thermo Fisher Scientific) and integrity through gel electrophoresis with a 2% agarose gel in TAE buffer (40mM tris base, 20mM acetic acid and 1mM ethylenediaminetetraacetic acid) at 80 v for 90 minutes. A single sample of each biotype was submitted to Macrogen, Inc. (South Korea), who performed the library preparation and sequencing.

3.2.3.2: Library preparation and sequencing

The DNA samples were treated with the EZ DNA Methylation Lightning kit™ (Zymo Research), which is used for bisulfite conversion, while denatured in a thermal cycler at 98°C for 8 minutes. Using a TruSeq DNA Methylation Library Kit™ (Illumina), 5' tags were generated through random priming, followed by selective 3' tagging. Illumina P7 and P5 adapters were ligated through amplification to the 5' and 3' ends respectively. The Illumina HiSeq X platform was used to sequence the bisulfite treated samples.

3.2.4: Trimming

The fastq files, obtained from the HiSeq X sequencing, were analysed for quality using FastQC (Andrews, 2010). After inspecting the adapter content, per base sequence content, and per base sequence quality, Trimmomatic (Bolger et al., 2014) was used to remove adapter sequences and trim the paired-end reads for quality. A slidingwindow over 15 base pairs was used to trim for a quality score of 20, along with a headcrop of 10, which removes the first 10 base pairs of each read as these show biased sequence contributions (Figure 3.5 and Figure 3.6). The Illuminaclip parameter was used to search for and remove adapter sequences from the reads. After trimming all reads were filtered for a minimum read length of 40 bp.

Stellenbosch University https://scholar.sun.ac.za

3.2.5: Bismark pipeline

The Bismark software program (Krueger and Andrews, 2011) was used to analyse the methylation status of the trimmed and filtered sequence reads. This program was used because of the easy usability of output files downstream, as well as its ability to discriminate between top and bottom strand as well as cytosine context. The pipeline consists of three distinct steps, outlined in Figure 3.2. The steps proceed as follows:

(i) Genome preparation. The SAM biotype reference genome (GenBank ID GCA_001465515.1; BioProject PRJNA297165) and the obtained HiSeq X reads were independently converted to two versions; one where all C's are converted to T's, and one where all G's are converted to A's. As control DNA, the RWA endosymbiont *Buchnera aphidicola* (NCBI accession number: NZ_CP013259.1) was used.

For the two samples to be comparable, their conversion efficiencies need to be calculated. The conversion efficiency of a sample refers to the percentage of unmethylated cytosine bases that have been converted to thymine bases in the trimmed reads. DNA, known to be unmethylated is required for the calculation. Any remaining cytosine bases in the trimmed reads, that align uniquely to the known unmethylated DNA, is a result of incomplete bisulfite conversion. From this it follows that the conversion efficiency of a sample is equal to one minus the percentage that cytosine bases comprise of the composition of trimmed reads, aligning uniquely to the unmethylated DNA.

The known unmethylated DNA is referred to as control DNA and is added to samples before submission for bisulfite sequencing. Control DNA usually comprises between 0.1% and 0.5% of the total DNA, lambda-phage DNA is commonly used as a control. No control DNA was added to samples for the purposes of this study. Instead, the genome of an endosymbiont *Buchnera aphidicola*, was used as a control as bacteria lack 5mC methyaltion. Therefore, the genome preparation step was repeated for the *B. aphidicola* genome hosted in RWA (accession number: NZ_CP013259.1)

(**ii**) **Alignment.** Bismark invokes the Bowtie2 (Langmead and Salzberg, 2012) program, a next generation sequence alignment program, to map the converted sequence reads (G-to-A and C-to-T) independently to the converted versions (G-to-A and C-to-T) of the reference genome. This results in four parallel outputs from which the best alignment can be determined, as an example, in Figure 3.2 thread (1) produces the best alignment as no mismatches occur. For Bowtie2 to allow an alignment, the alignment score needs to be higher than a set threshold. The threshold is calculated as: $y = mL + c$, where L represents the read length, *m* and *c* are configurable but are both set to a default of -0.6. When Bowtie2 is invoked by Bismark however, the values of *m* and *c* are reduced to -0.2 and 0 respectively. This level of stringency results in a mapping efficiency of only 37.7%, setting the threshold

Stellenbosch University https://scholar.sun.ac.za

back to the default for Bowtie2 results in a mapping efficiency of 55.7%, while reducing the stringency further, with the value of *m* to -0.8 and *c* to -0.6, brings the mapping efficiency up to 63.1%, which can be considered as high for bisulfite sequencing (Chatterjee et al., 2012).

It was important to ensure erroneous mapping would not occur as a result of lower the stringency too much. As can be seen in Figure 3.1 however, a linear decrease in stringency caused a linear increase in the number of reads mapped, as well as the number of CpG sites and the number of methylated CpG sites discovered. If the stringency were too lax, an exponential increase in mapped reads would be expected.

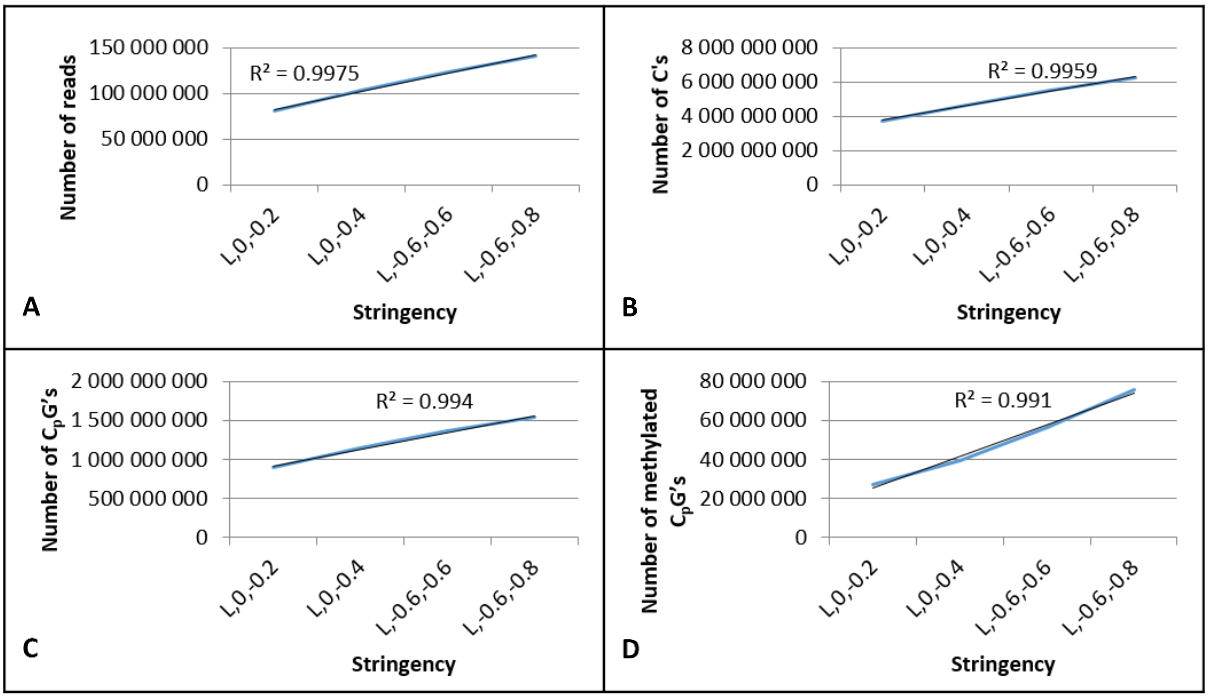

Figure 3.1: Linear regressions drawn between the stringency values used during alignment with Bowtie2 and **(A)** the number of reads mapped, **(B)** the number of cytosine bases detected, **(C)** the number of C_pG dinucleotides detected, **(D)** the number of these C_pG sites that were methylated. L,0,-0.2 is the default stringency used during the Bismark pipeline, L,-0.6,-0.6 is the native Bowtie2 default, and L,-0.6,-0.8 was the stringency used.

(**iii**) **Methylation extraction.** By comparing the sequence reads with the corresponding genomic sequence, the methylation state of each cytosine base is determined. In the example in Figure 3.2 only C-to-T substitutions are presented, but G-to-A substitutions are also applicable. Before extracting the methylation data, M-bias plots for the uniquely mapped reads were generated. M-bias plots are line graphs showing the distribution of methylated calls across reads. This is useful, as it can be used to identify regions of reads that show a bias, either for methylated or unmethylated calls. These biased regions can then be excluded from further analysis. Based on the M-bias plots (Figure 3.7 and Figure 3.8), the first 5 base pairs and the last 15 base pairs of full-length reads were excluded, as these regions showed a bias for methylated calls.

Stellenbosch University https://scholar.sun.ac.za

The information for all cytosine bases, for both strands, in all three available contexts, with a coverage of at least 10 during mapping were extracted. All parameters specified during the methylation extraction step can be found in Appendix A (Table A1).

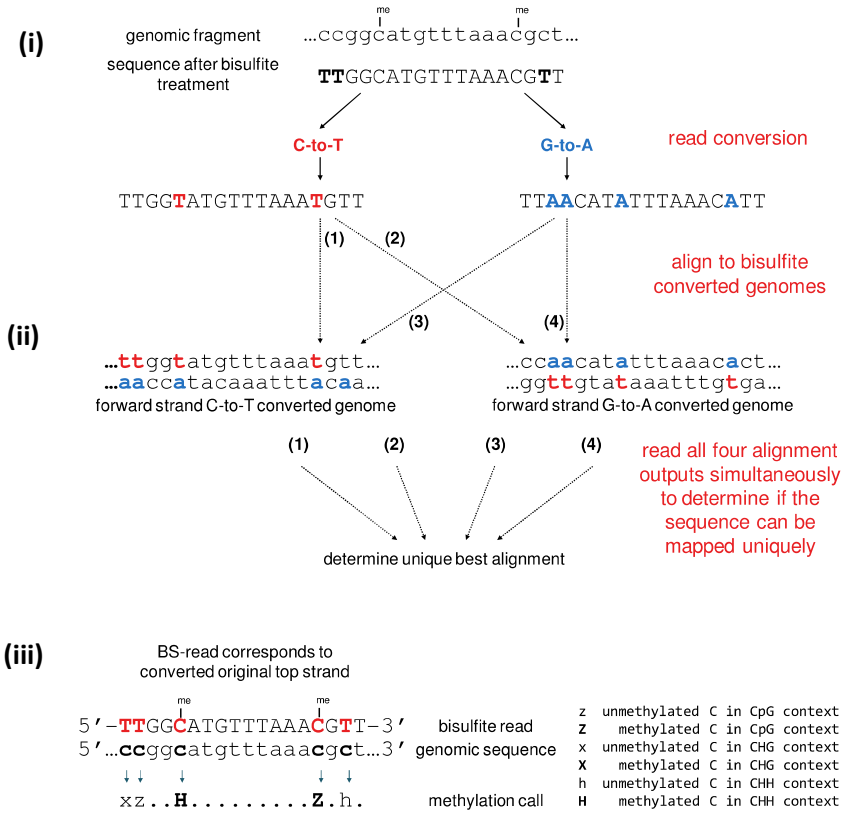

Figure 3.2: A basic outline of the three steps involved in the pipeline employed by the Bismark program. Image sourced from Krueger and Andrews (2011).

3.3: Results

3.3.1: Whole genome bisulfite sequencing

DNA from two Russian wheat aphid (RWA) biotypes were submitted for whole genome bisulfite sequencing: SA1, the least virulent South-African biotype, and SAM, a laboratory contained biotype, expressing a higher virulence than any of the naturally occurring South-African biotypes (Swanevelder et al., 2010). Biotype SAM was developed by applying continuous selective pressure on biotype SA1, in the form of force feeding on Tugela *Dn1*, a wheat cultivar resistant to RWA. A single sample of each biotype was submitted, both containing DNA from 100 apterous, pathogenic, female aphids. Prior to sequencing on the Illumina HiSeq X platform, bisulfite conversion of the DNA samples was performed, using the EZ DNA Methylation-Gold Kit (Zymo Research), and adapter sequences, Illumina P5 and P7,

Stellenbosch University https://scholar.sun.ac.za

were ligated as part of library preparation with the TruSeq DNA Methylation Kit (Illumina). Data from the sequencing report accompanying the sequence data files are presented in Table 3.1.

Table 3.1: Data contained in the sequencing report after whole genome bisulfite sequencing of two Russian wheat aphid biotypes on the Illumina HiSeq X platform.

Biotype	SA1	SAM
Paired-end reads produced	234,970,631	285,766,353
Bases sequenced	70,961,130,562	86,305,968,606
GC content	24.9%	25.8%
Q30	91.2%	91.3%

The GC content for biotype SAM is slightly higher, which is an early indicator of a higher level of methylation, as bisulfite treatment converts unmethylated cytosine bases to thymine.

3.3.2: Trimming

FastQC was used to visualise sequence quality (Andrews, 2010). Data loss during trimming was low, as more than 90% of the bases already had a Phred quality score of at least 30 assigned and did not require extensive trimming for qualsty. Regardless, Trimmomatic (Bolger et al., 2014) was used to perform a ten base pair sliding window trim for a Phred quality score of 30, along with removal of the relevant adapter sequences and the first ten base pairs of each read, which showed irregular per base sequence content (Figure 3.5 and Figure 3.6). Trimmed reads were filtered for a minimum read length of 40.

Stellenbosch University https://scholar.sun.ac.za

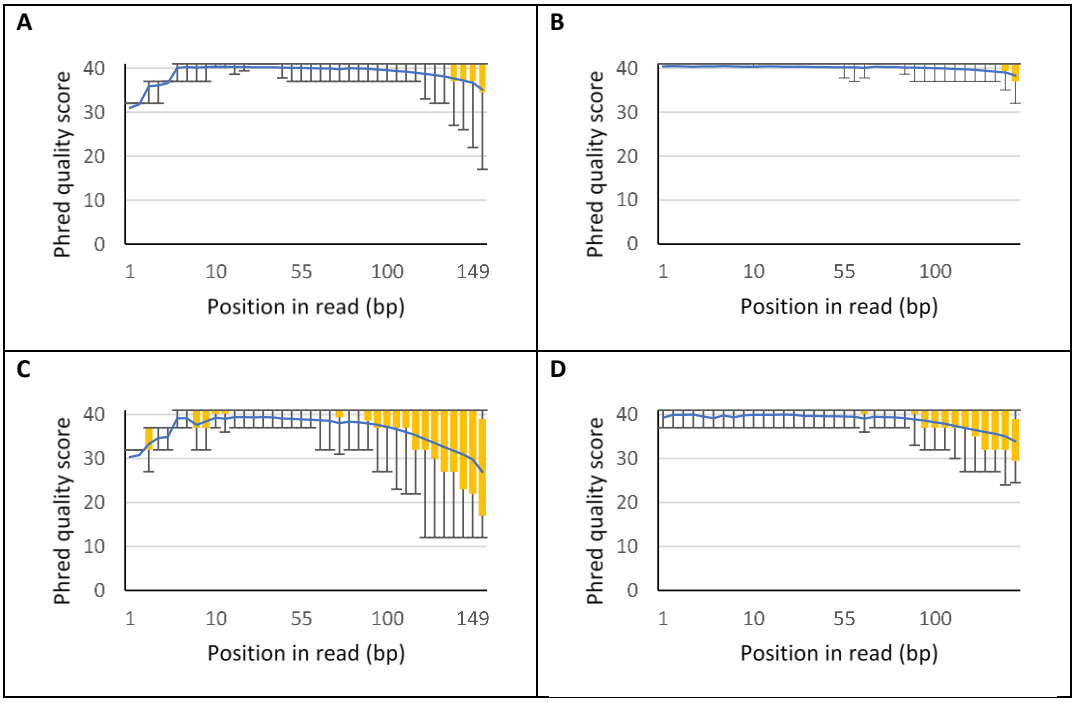

Figure 3.3: The per base sequence quality scores of SA1 reads, before (**A** & **C**) and after (**B** & **D**) trimming. First strand reads are shown in **A** and **B**, while second strand reads are shown in **C** and **D**. FastQC was used to extract the Phred quality scores from the fastq files. The blue line represents the mean, the yellow boxes and the grey handles indicates the upper and lower quartiles, and the 10th and 90th percentiles respectively.

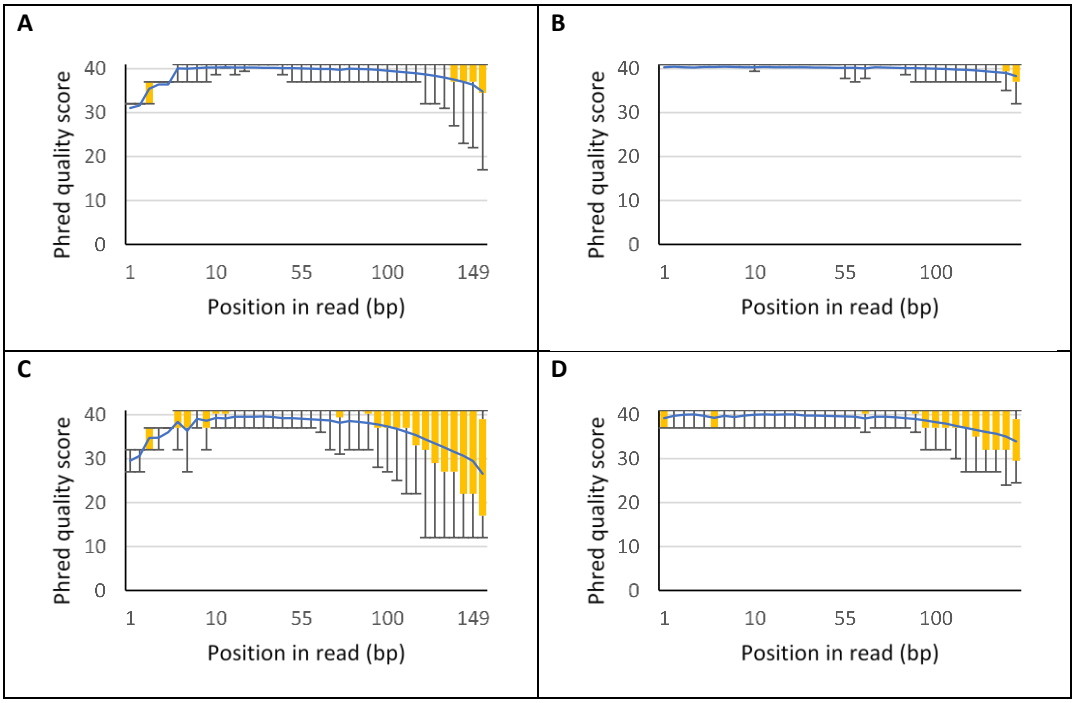

Figure 3.4: The per base sequence quality scores of SAM reads, before (**A** & **C**) and after (**B** & **D**) trimming. First strand reads are shown in A and B, while second strand reads are shown in C and D. FastQC was used to extract the Phred quality scores from the fastq files. The blue line represents the mean, the yellow boxes and the grey handles indicates the upper and lower quartiles, and the 10th and 90th percentiles respectively.

Stellenbosch University https://scholar.sun.ac.za

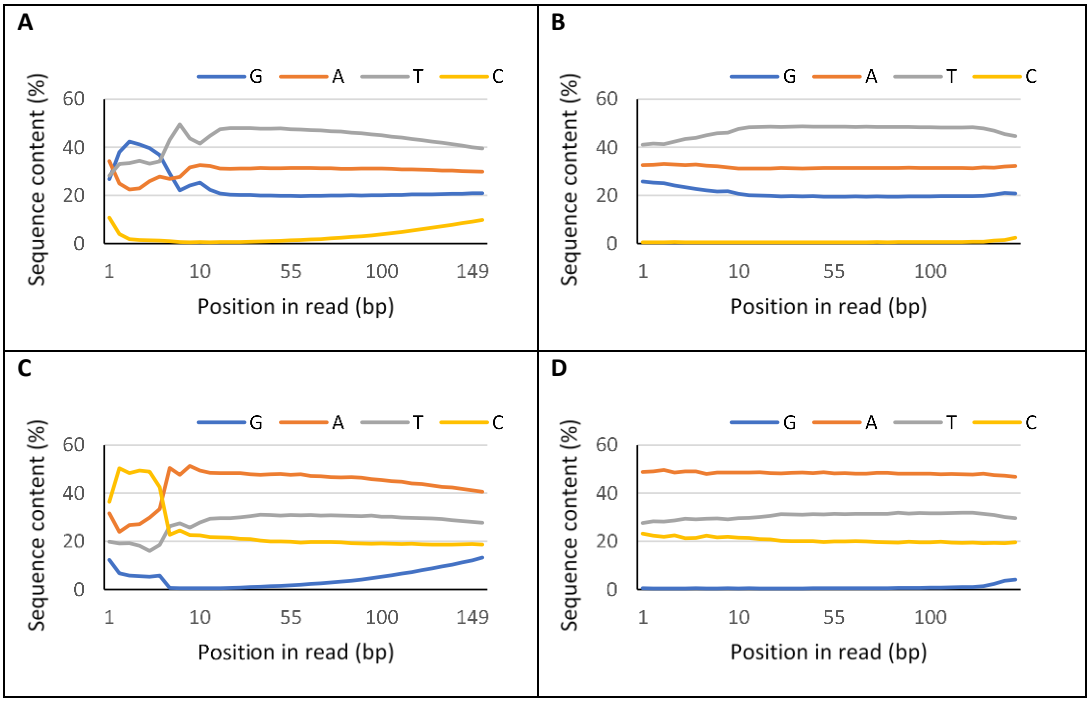

Figure 3.5: The per base sequence contribution of SA1 reads, before (**A** & **C**) and after (**B** & **D**) trimming. First strand reads are shown in A and B, while second strand reads are shown in C and D. FastQC was used to extract the sequence contributions from the fastq files.

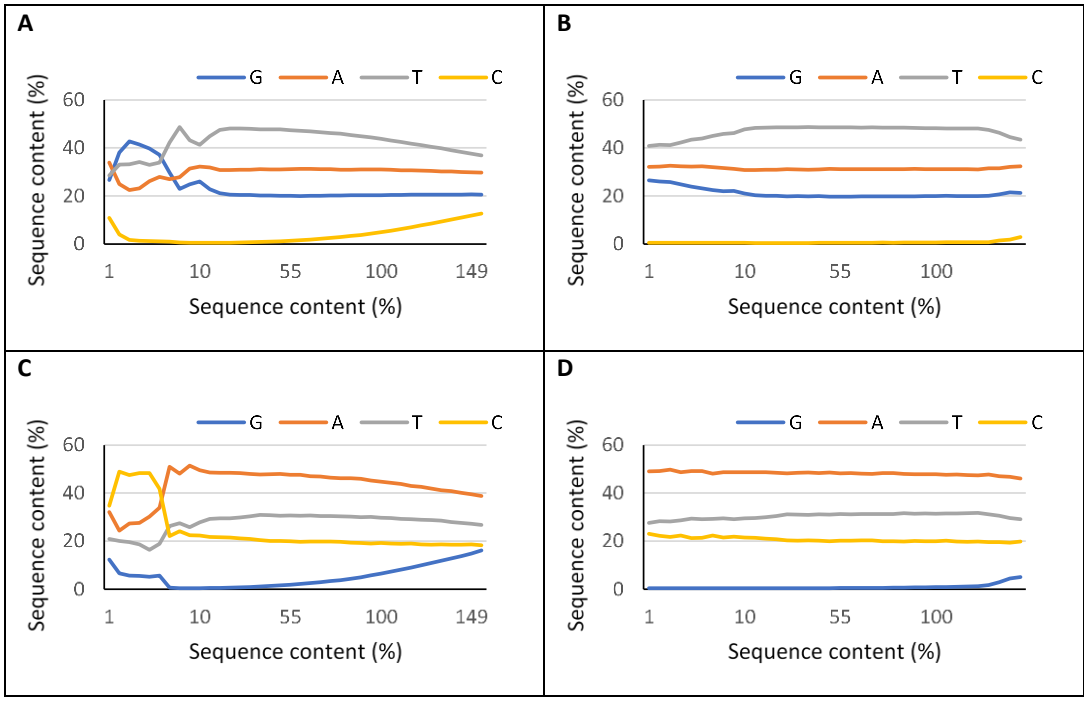

Figure 3.6: The per base sequence contribution of SAM reads, before (**A** & **C**) and after (**B** & **D**) trimming. First strand reads are shown in A and B, while second strand reads are shown in C and D. FastQC was used to extract the sequence contributions from the fastq files.

Stellenbosch University https://scholar.sun.ac.za

Table 3.2: Read data for each RWA biotype before and after trimming.

	Before trimming		After trimming	
Biotype	SA1	SAM	SA1	SAM
Number of paired end reads	234,970,631	285,766,353	218,231,083	266,159,417
Bases covered	70,961,130562	86,305,968,606	54,923,549,873	65,503,189,833
Read length	150	150	40-141	40-141
GC content	24.9%	25.8%	20.7%	21.2%
Q30	91.2%	91.3%	96.0%	96.1%

After trimming at least 96% of all reads had an assigned quality score of at least 30 (Table 3.2) and the per base sequence contribution was uniform (Figure 3.5 and Figure 3.6). The reduced GC content of the reads after trimming is most likely due to the removal of adapter sequences, which were much more GC-rich than the RWA DNA.

3.3.3: Methylation data

Bismark (Krueger and Andrews, 2011), a Bowtie2 (Langmead and Salzberg, 2012) dependant, methylation sensitive alignment program was used to map the bisulfite treated reads to a reference genome and score the methylation status for every cytosine base called. An alignment efficiency of 63.1% was achieved for biotype SA1, with 133 043 439 uniquely mapped read pairs, the alignment efficiency for biotype SAM was 63.7%, with 162 915 229 uniquely mapped read pairs.

Bismark provides the user with the option to output the methylation data grouped into three different contexts, the di-nucleotide or tri-nucleotide sites where cytosine methylation can occur are referred to as a context. The available contexts are as follows: Two symmetrical contexts, CpG and CHG, and an asymmetrical CHH context, where H is any non-G base.

The program also outputs graphs, referred to as M-bias plots, showing the per base methylation levels across reads. These graphs are useful for identifying areas of reads with a bias for either methylated, or unmethylated calls. Theoretically the distribution of methylation should be uniform across the entire read, however the last 15 base pairs of the trimmed forward reads clearly show a bias for methylated calls (Figure 3.7 and Figure 3.8) and were excluded from further analysis.

Stellenbosch University https://scholar.sun.ac.za

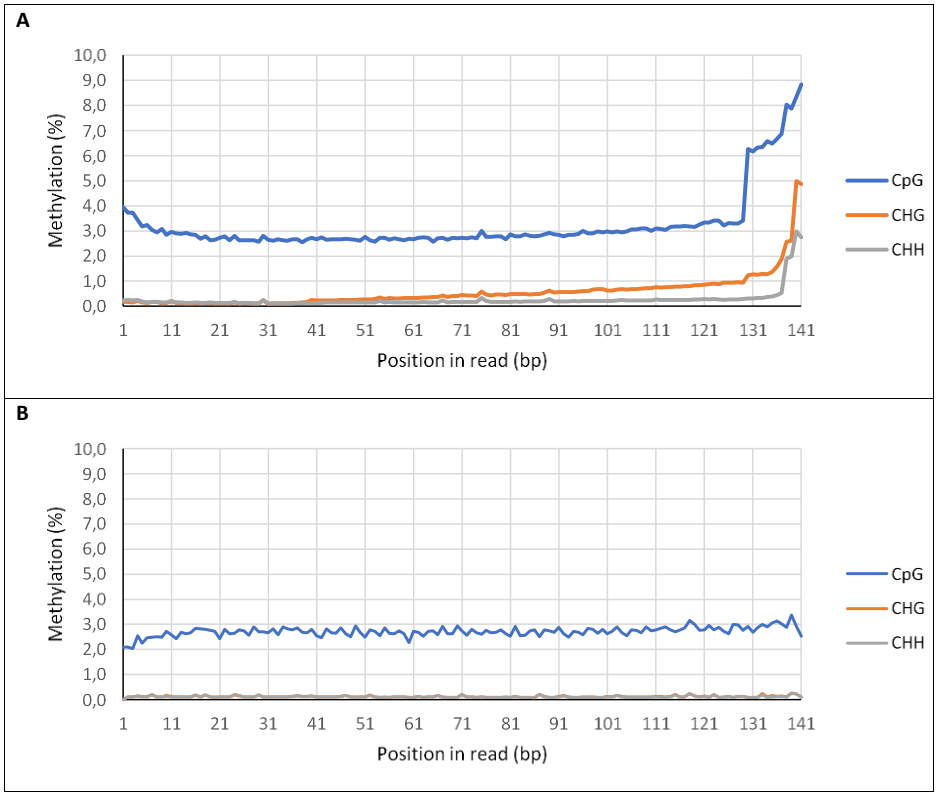

Figure 3.7: The M-bias plots of forward **(A)** and reverse **(B)** reads of biotype SA1. Showing the level of CpG, CHG and CHH methylation of each base pair, averaged over all reads. Data exported from the Bismark program.

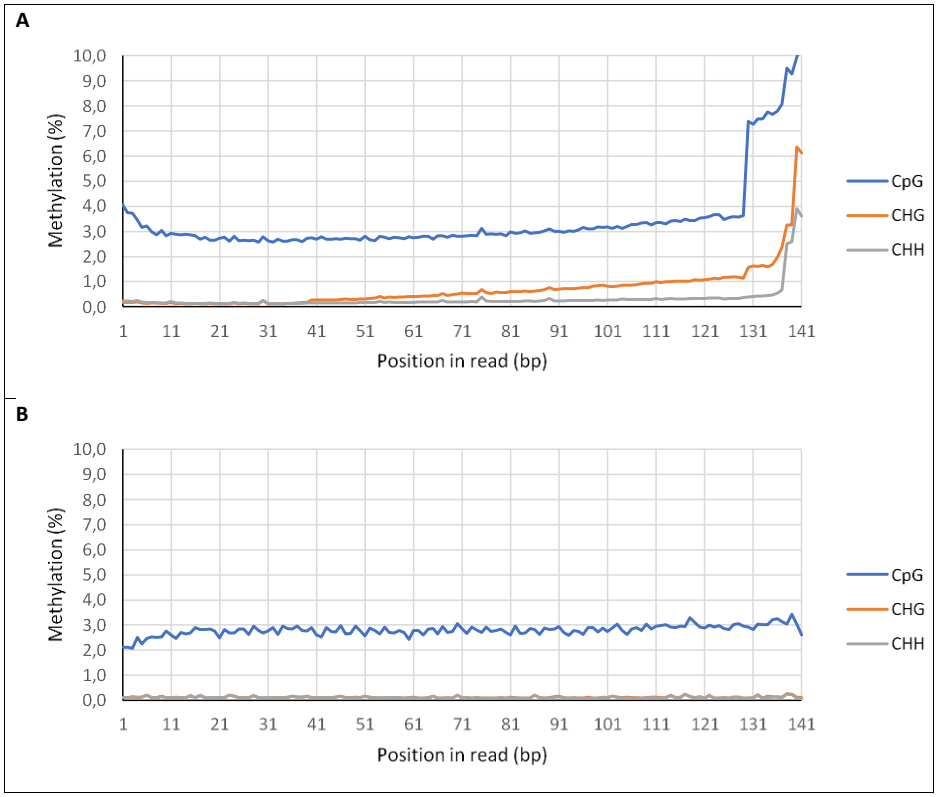

Figure 3.8: The M-bias plots of forward **(A)** and reverse **(B)** reads of biotype SAM. Showing the level of CpG, CHG and CHH methylation of each base pair, averaged over all reads. Data exported from the Bismark program.

Stellenbosch University https://scholar.sun.ac.za

As a single sample for each biotype was used for the analyses, it is important to know if any of the observed differences are the result of a difference in the efficiency of bisulfite conversion of the samples prior to sequencing. All the cytosine bases in DNA lacking any methylation should be converted to thiamine. Any unconverted cytosine bases in unmethylated, treated DNA is a result of incomplete bisulfite conversion and can be used to calculate the treatment efficiency.

The treated reads were mapped to the genome of *Buchnera aphidicola* (accession: CP013259.1), an endosymbiont found within RWA. In bacteria methylation occurs on adenine, not cytosine (Casadesús and Low, 2006), as such the bacterial DNA can be used as an unmethylated control to calculated conversion efficiency. Efficiencies of 99.664% for SA1 and 99.618% for SAM were achieved. There was a 0.046% difference in bisulfite conversion efficiency between the SA1 and SAM samples which will have to be corrected for during methylation analysis.

After extracting the methylation data in all three available contexts (Table 3.3), command line text manipulation was used to further group calls into genic and intergenic regions (Table 3.4), exons and introns (Table 3.5), and top and bottom strands (Table 3.6). All commands used to process the Bismark outputs and compile tables can be found in Appendix A.

Table 3.3: The percentage methylation of the CpG, CHG and CHH contexts of each biotype, as calculated from the number of methylated and total calls, obtained by aligning the trimmed and filtered bisulfite treated reads to the SAM reference genome with Bismark.

Biotype	SA1	SA1	SA1	SAM	SAM	SAM
Context	CpG	CHG	CHH	CpG	CHG	CHH
Total	877502203	572304669	1954826830	1008746940	661091134	2260042456
Methylated	22624905	769939	2333186	26308447	949527	2838605
Methylation (%)	2.57833	0.134533	0.119355	2.60803	0.14363	0.1256

Table 3.4: The percentage methylation of genes and intergenic regions of each biotype, as calculated from the number of methylated and total calls, obtained by aligning the trimmed and filtered bisulfite treated reads to the SAM reference genome with Bismark.

Biotype	SA1	SA1	SAM	SAM
Region	Genic	Intergenic	Genic	Intergenic
Total	1514526611	1890107091	1754514206	2175366324
Methylated	17548941	8179089	20574939	9521640
Methylation (%)	1.15871	0.432732	1.17269	0.437703

Table 3.5: The percentage methylation of exons and introns of each biotype, as calculated from the number of methylated and total calls, obtained by aligning the trimmed and filtered bisulfite treated reads to the SAM reference genome with Bismark.

Biotype	SA1	SA1	SAM	SAM
Region	Exonic	Intronic	Exonic	Intronic
Total	833010919	372838760	965430348	435754074
Methylated	11391036	2625923	13317965	3087380
Methylation (%)	1.36745	0.704305	1.37948	0.708514

Stellenbosch University https://scholar.sun.ac.za

Table 3.6: The percentage methylation of the top and bottom strand of each biotype, as calculated from the number of methylated and total calls, obtained by aligning the trimmed and filtered bisulfite treated reads to the SAM reference genome with Bismark.

Biotype	SA1	SA1	SAM	SAM
Strand	Top	Bottom	Top	Bottom
Total	1684379026	1720254676	1947144605	1982735925
Methylated	12794632	12933398	15037106	15059473
Methylation (%)	0.759605	0.75183	0.772264	0.75953

When considering the total calls, biotype SAM shows slightly higher levels of methylation than SA1 in all three available contexts (Table 3.3), however, none of the contexts show a difference larger than the 0.0462% difference in conservation efficiency. Splitting the calls into genic and intergenic regions, rather than context shows larger differences between biotypes. The distribution of methylation is similar between biotypes, the CpG context, despite being only the second most abundant context contains most of the total methylated calls, while the least number of methylated calls can be found within the most abundant context, CHH. Although the intergenic regions contain a high number of methylated calls, these regions comprise a much greater part of the genome and as such the percentage of calls methylated in intergenic regions is low. Similarly, exons show a higher percentage of methylation than introns, even though the introns comprise a larger part of the genes.

The strand specific information from the Bismark outputs were used to determine whether any hemimethylated sites are present. When one strand in a symmetrical context, such as CpG or CHG, is methylated, but not the other strand, this is referred to as a hemimethylated site. For the calculation of the number of hemimethylated sites, only sites that were covered at least ten times on both strands were considered. The CHH context was not included as it is not a symmetrical context of methylation.

Table 3.7: The percentage of strands methylated and hemimethylated in the top and bottom strand of each biotype, as calculated from the number of sites containing methylated calls, obtained by aligning the trimmed and filtered bisulfite treated reads to the SAM reference genome with Bismark.

Biotype	SA1				SAM			
Context	CpG		CHG		CpG		CHG	
Strand	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom
Total sites	1807699	1807699	1462221	1462221	2015914	2015914	1642447	1642447
Methylated sites	251213	258548	171899	143916	293366	305189	207287	173447
Hemimethylated sites	152180	159515	148809	120826	177801	189624	177563	143723
Sites methylated (%)	13.8968	14.3026	11.756	9.84229	14.5525	15.139	12.6206	10.5603
Sites hemimethylated (%)	8.41844	8.8242	10.1769	8.26318	8.81987	9.40635	10.8109	8.75054

Hemimethylation occurred at a higher frequency in the CpG context than in the CHG context. The distribution of hemimethylation between strands is mostly equal with only a slightly higher percentage of the top strand sites hemimethylated when compared to the bottom strand sites. This is the case in both biotypes and in both contexts (Table 3.7).

Stellenbosch University https://scholar.sun.ac.za

3.4: Discussion

The methylation data can only be used as a basis for comparison between biotypes if a good coverage across the genome can be obtained, with confidence that sequencing artifacts and biases have been addressed. An average coverage of 97 times per cytosine base was achieved for biotype SA1 and 105 times for biotype SAM. The higher coverage in the SAM biotype resulted from the slightly higher mapping efficiency of the SAM reads during alignment, as compared to the SA1 reads. As the SAM reference genome was used during alignment, the higher mapping efficiency in SAM was to be expected. The reduced coverage in SA1 is inconsequential however, as coverages between five times and fifteen times are acceptable for the detection of differentially methylated regions (DMRs), depending on the size of regions and the magnitude of difference (Ziller et al., 2015).

During trimming, all adapter content and aberrant nucleotide contributions were removed. Phred quality scores were high for the trimmed reads used during alignment, with more than 96% of the reads scored above 30, relating to a accuracy of 99.9% (Ewing et al., 1998). The stringency of Bowtie2 was set to fall in a minimum score range, where an increase in minimum score was linearly correlated to the number of reads uniquely mapped. This helped in reducing misalignment. Fifteen bases on the 5'-end of forward reads were excluded from analysis, as this region showed a bias for methylated calls. Control DNA comprised 1.1% of the total trimmed reads, well above the required level of 0.1% (Grehl et al., 2018). A conversion efficiency of 99.90% was achieved during library preparation, which is as high as can be expected without sacrificing sequence quality (Grunau et al., 2001). The difference in the conversion efficiency was very low between samples, 0.010%. All the above increase confidence that any differences in methylation between samples are not the result of artifacts or biases but represent true differences.

In a previous study by Breeds et al. (2018), the global methylation levels of biotype SA1 and SAM were calculated using antibodies specific to methylated cytosine bases. In such an approach, context and position information is ignored. Comparison between the results of antibody-based methods and whole genome bisulfite sequencing methods are thus only appropriate if the same limitations are enforced *in silico*. Using simply the number of methylated calls compared to unmethylated calls, irrespective of context or position, a methylation level of 0.76% was calculated for SA1 and 0.77% for SAM. This is higher than the 0.14% and 0.16% calculated for SA1 and SAM, respectively, using methylation specific antibodies. Due to the low levels of methylation in RWA however, the amount of methylated DNA was at the 0.2 ng/well minimum detectable by the kit.

Restriction enzymes HpaII and MspI, with recognition site CCGG (McClelland et al., 1994), were also used in the Breeds et al. (2018) study to calculate the global level of methylation in RWA. Only

Stellenbosch University https://scholar.sun.ac.za

methylation in the CpG context can be detected using this method, as such CpG data from the WGBS results were isolated, regardless of position, for comparison. Methylation levels of 2.22% for SA1 and 2.85% for SAM were calculated using HpaII and MspI banding, while 2.58% and 2.61% of calls were methylated for SA1 and SAM, respectively. The Results from the WGBS and restriction enzyme banding are similar. All CpG sites were evaluated *in silico*, whereas only CpG sites that form part of the CCGG recognition site of HpaII and MspI would influence the banding pattern. This might be the reason for the slight discrepancy in methylation, as calculated by the two different methods, as the Breeds et al. (2018) study only measured methylation in genomic sites that contained the CCGG pattern. Bisulfite sequencing on other Hemipteran species have shown similar CpG methylation levels between 2.24% and 4.21% (Bewick et al., 2017).

Previous studies have observed a marked decrease in non-CpG methylation, when compared to CpG methylation (Feng et al., 2010; Kunert et al., 2003; Pradhan and Adams, 1995). The WGBS results of RWA supports this observation, as non-CpG methylation was much lower than CpG methylation, in both biotypes. In SA1, 0.13% CHG and 0.12% CHH methylation was detected, with slightly higher levels in SAM (0.14% CHG and 0.13% CHH). Bisulfite sequencing has been used to determine levels of non-CpG methylation in the honeybee (*Apis mellifera*) (Feng et al., 2010). Their results show 0.93% CpG, 0.26% CHG and 0.17%CHH methylation. Interestingly, while non-CpG in RWA is lower than in the honeybee, the level of CpG methylation of RWA is more than double that of the honeybee.

Another previous finding in agreement with the WGBS results. Is that methylation in coding regions of insect genomes is more abundant than in non-coding regions (Bewick et al., 2017). In both biotypes the level of genic methylation is more than two and a half times that of intergenic regions. This holds true within genes as well, as the coding regions, exons, are methylated at twice the level of the introns, which are non-coding.

Hemimethylation was detected from the strand specific information of the WGBS data in both symmetrical contexts. Hemimethylation is impossible in the nonsymmetrical CHH context, as there are not cytosine bases in both the top and bottom strand of the codon. It is very interesting to note that CpG hemimethylation is higher in the bottom strand, while CHG hemimethylation is higher in the top strand. This would seem to indicate that DNMT1 is less efficient in transferring CpG methylation to the top strand and CHG methylation to the bottom strand, or perhaps that the differing modes of synthesis between leading and lagging strands (Okazaki et al., 1968) are influencing the discrepancy. Some hemimethylation specific to cell types will undoubtably have been missed by the WGBS approach, as a locus might for example be methylated on the top strand in brain cells and on the

Stellenbosch University https://scholar.sun.ac.za

bottom strand in the gut lining. The strandedness of the detected hemimethylation sites are therefore conserved across all cell types and should prove to be very interesting loci for further study.

The conversion efficiency of the bisulfite treatment is very high for both biotypes, with an inefficiency of 0.082% in SA1 and 0.093% in SAM. As such, even the low levels of non-CpG and intergenic methylation cannot be ascribed simply to incomplete treatment, but do in fact represent true methylation. Biotype SAM is more methylated than biotype SA1, in all contexts, in both coding and non-coding regions. While the argument might be made that without replicates it cannot be showed that this is not the result of a higher inefficiency in the SAM sample bisulfite treatment, the results are substantiated by both the antibody- and restriction enzyme-based methods applied in a previous study (Breeds et al., 2018). It is therefore clear that there is a difference in methylation between RWA biotypes and that methylation is higher in the more virulent biotype, SAM.

Many questions remain for future studies: Did an increase in methylation provide SAM with an advantage over SA1, perhaps by altering gene expression, either by upregulating the expression of a virulence factor, or by downregulating an elicitor? Also, methylation sensitive transcription factors have been identified in mammalian species (Ziller et al. 2013), are there functional homologs in the RWA?

3.5: References

Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data.

Bewick, A.J., Vogel, K.J., Moore, A.J. & Schmitz, R.J., 2017, Evolution of DNA methylation across insects, *Molecular Biology and Evolution*, 34(3), 654–665.

Bolger, A.M., Loshe, M. & Usadel, B., 2014, Trimmomatic: A flexible trimmer for Illumina sequence data, *Bioinformatics*, *30*(15), 2114-2120.

Breeds, K., Burger, N.F.V. & Botha, A-M., 2018, New insights into the methylation status of virulent *Diuraphis noxia* (Hemiptera: Aphididae) biotypes, *Journal of economic entomology*, 111(3), 1395–1403.

Burger, N.F.V. & Botha, A-M., 2017, Genome of Russian wheat aphid an economically important cereal aphid, *Standarts in Genomic Sciences*, 12(1), 90.

Casadesús, J. & Low, D., 2006, Epigenetic gene regulation in the bacterial world, *Microbiology and Molecular BiologyRreviews*, 70(3), 830–856.

Chatterjee, A., Stockwell, P.A., Rodger, E.J. & Morison, I.M., 2012, Comparison of alignment software for genome-wide bisulphite sequence data, *Nucleic Acids Research*, 40(10), e79–e79.

du Toit, F., 1989, Inheritance of resistance in two *Triticum aestivum* lines to Russian wheat aphid (Homoptera: Aphididae), *Journal of Economic Entomology*, 82(4), 1251–1253.

Ewing, B., Hillier, L., Wendl, M.C. & Green, P., 1998, Base-calling of automated sequencer traces using phred. I. Accuracy assessment, *Genome research*, 8(3), 175–185.

Feng, S., Cokus, S.J., Zhang, X., Chen, P.-Y., Bostick, M., Goll, M.G., Hetzel, J., Jain, J., Strauss, S.H., Halpern, M.E.,

Stellenbosch University https://scholar.sun.ac.za

Ukomadu, C., Sadler, K.C., Pradhan, S., Pellegrini, M. & Jacobsen, S.E., 2010, Conservation and divergence of methylation patterning in plants and animals, *Proceedings of the national academy of sciences of the United States of America*, 107(19), 8689–8694.

Glastad, K.M., Hunt, B.G. & Goodisman, M.A.D., 2014, Evolutionary insights into DNA methylation in insects, *Current Opinion in Insect Science*, 1, 25–30.

Gong, L., Cui, F., Sheng, C., Lin, Z., Reeck, G., Xu, J. & Kang, L., 2012, Polymorphism and methylation of four genes expressed in salivary glands of Russian wheat aphid (Homoptera: Aphididae)., *Journal of Economic Entomology*, 105(1), 232–241.

Grehl, C., Kuhlmann, M., Becker, C., Glaser, B. & Grosse, I., 2018, How to design a whole-genome bisulfite sequencing experiment, *Epigenomes*, 2(4), 21.

Grunau, C., Clark, S.J. & Rosenthal, A., 2001, Bisulfite genomic sequencing: systematic investigation of critical experimental parameters, *Nucleic acids research*, 29(13), 65.

Hick, C.A., Field, L.M. & Devonshire, A.L., 1996, Changes in the methylation of amplified esterase DNA during loss and reselection of insecticide resistance in peach-potato aphids, *Myzus persicae*, *Insect Biochemistry and Molecular Biology*, 26(1), 41–47.

Hoagland, D.R. & Arnon, D.I., 1950, The water-culture method for growing plants without soil, *California Agricultural Experiment Station Circular*, 347(347), 1–32.

Holliday, R. & Grigg, G.W., 1993, DNA methylation and mutation, *Mutation Research*, 285(1), 61–67.

Hotchkiss, R.D., 1948, The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography, *The Journal of Biological Chemistry*, 175(1), 315–332.

Ishikawa, A. & Miura, T., 2009, Differential regulations of wing and ovarian development and heterochronic changes of embryogenesis between morphs in wing polyphenism of the vetch aphid, *Evolution and Development*, 11(6), 680–688.

Jankielsohn, A., 2019, New Russian wheat aphid biotype found in Free State, *Grain SA Mini Focus, Pest Control on Winter Cereals*.

Krueger, F. & Andrews, S.R., 2011, Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications, *Bioinformatics*, 27(11), 1571–1572.

Kunert, N., Marhold, J., Stanke, J., Stach, D. & Lyko, F., 2003, A Dnmt2-like protein mediates DNA methylation in *Drosophila*, *Development*, 130(21), 5083–5090.

Langmead, B. & Salzberg, S.L., 2012, Fast gapped-read alignment with Bowtie 2, *Nature Methods*, 9(4), 357–359.

Maunakea, A.K., Chepelev, I., Cui, K. & Zhao, K., 2013, Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition, *Cell Research*, 23(11), 1256–1269.

McClelland, M., Nelson, M. & Raschke, E., 1994, Effect of site-specific modification on restriction endonucleases and DNA modification methyltransferases, *Nucleic Acids Research*, 22(17), 3640–3659.

Okazaki, R., Okazaki, T., Sakabe, K., Sugimoto, K. & Sugino, A., 1968, Mechanism of DNA chain growth. I. Possible discontinuity and unusual secondary structure of newly synthesized chains, *Proceedings of the National Academy of Sciences of the United States of America*, 59(2), 598–605.

Pradhan, S. & Adams, R.L., 1995, Distinct CG and CNG DNA methyltransferases in *Pisum sativum*, *The Plant*

Stellenbosch University https://scholar.sun.ac.za

Journal, 7(3), 471–481.

Swanevelder, Z.H., Surridge, A.K.J., Venter, E. & Botha, A-M., 2010, Limited endosymbiont variation in *Diuraphis noxia* (Hemiptera: Aphididae) Biotypes From the United States and South Africa, *Journal of Economic Entomology*, 103(3), 887–897.

Tolmay, V., Lindeque, R. & Prinsloo, G., 2007, Preliminary evidence of a resistance-breaking biotype of the Russian wheat aphid, *Diuraphis noxia* (Kurdjumov) (Homoptera: Aphididae), in South Africa, *African Entomology*, 15(1), 228–230.

Truong, M., Yang, B., Wagner, J., Desotelle, J. & Jarrard, D.F., 2013, Analysis of promoter non-CG methylation in prostate cancer, *Epigenomics*, 5(1), 65–71.

van Zyl, R.A., 2007, Elucidation of possible virulence factors present in Russian wheat aphid (*Diuraphis noxia*) biotypes saliva, *MSc*, (March), University of Pretoria.

van Zyl, R.A. & Botha, A-M., 2008, Eliciting proteins from *Diuraphis noxia* biotypes differ in size and composition, *18th Biennial International Plant Resistance to Insects Workshop*, Fort Collins, Colorado.

Walters, M., Penn, F., du Toit, F., Botha, T. & Aalbersberg, K., 1980, The Russian Wheat Aphid. Farming in South Africa, *Africa Leaflet Series: Wheat*, (3), 1–6.

Ziller, M.J., Gu, H., Müller, F., Donaghey, J., Tsai, L.T.Y., Kohlbacher, O., de Jager, P.L., Rosen, E.D., Bennett, D.A., Bernstein, B.E. and Gnirke, A., 2013, Charting a dynamic DNA methylation landscape of the human genome, *Nature*, 500(7463), 477.

Ziller, M.J., Hansen, K.D., Meissner, A., Martin, J., Biology, R., Unit, P., Hospital, M.G. & Hospital, M.G., 2015, Coverage recommendations for methylation analysis by whole genome bisulfite sequencing, *Nature Methodes*, 12(3), 230–232.

Stellenbosch University https://scholar.sun.ac.za

Appendix A

Table A1: All parameters used during the methylation extraction step of the Bismarck pipeline, as well as their function.

Parameter	Function
-p	States that reads are paired end.
--no_overlap	Avoids scoring methylated calls in duplicate if forward and reverse reads overlap.
--ignore 5 ; ignore_3prime 5	Disregard methylation calls of the first 5 and last 15 bp of the forward reads.
--no_header	Removes headers from output files.
--cutoff 10	C's with a coverage of less than 10x are excluded from the cytosine report.
--gazillion	Reference genome consists of many contigs and has not been resolved to complete chromosomes.

Combining cytosine reports

```
awk '{gsub(/[contig_prefix]/,"")} $4+$5>=1 {$1=$1+0 ; t=$4 ; t2=$5 ; $4=t+t2 ; $5=t ; print"1",$0}' >  temp1; awk '{[contig_prefix]/,"")} $4+$5>=1 {$1=$1+0 ; t=$4 ; t2=$5 ; $4=t+t2 ; $5=t ; print"M",$0}' > temp2 ; cat temp1 temp2 gff | sort -n -k2 -k3 > out ; rm temp*
```

Replaces contig names with numbers, swaps the methylated and unmethylated columns, with total and methylated counts, adds an identifier to first column, "1" for SA1 and "M" for SAM. Writes these changes to temp file, concatenates them, along with a modified gff file with an "x" in field 6 and then deletes temp files.

Assign gene numbers

```
z=0 ; awk -v z="$z" 'FILENAME=="gff" {contig[$1]=$2 ; start[$1]=$3 ; stop[$1]=$4 ; next} {if ($6=="x") {z+=1} else if ($2==contig[z] && $3>=start[z] && $3<=stop[z]) {print "gene",z,$0} else {print "intergenic",z,$0}}' gff in > out
```

Uses contig number as well as start and stop positions of genes to assign gene number to all methylation sites. Sites that do not fall within genes are labelled as intergenic, with a number, which indicates the gene preceeding it. The modified gff file is used to increase variable z, when field 6 = "x".

Stellenbosch University https://scholar.sun.ac.za

Split genic regions to exons and introns

```
z=0 ; awk -v z="%z" 'FILENAME=="gff.exon" {gene[$2]=$3 ; region[$1]=$4 ; number [$1]=$5 ; start[$1]=$6 ; stop[$1]=$7 ; next} {if ($3=="gff") z+=1 ; else if ($2=gene[z] && $5>=start[z] && $5<=stop[z]) print$1,$2,region[z],number[z],$3,$4,$5,$6,$7,$8,$9,$10}'
```

Prepare output tables

```
awk 'BEGIN{OFS="\t"}
##### $1=Genic/Intergenic
##### $2=gene number
##### $3=SA1/SAM
##### $4=Contig
##### $5=Position
##### $6=Strand
##### $7=Total calls
##### $8=Methylated calls
##### $9=Context
##### $10=Codon
###TOTALS###
$3==1 {T1+=$7 ; M1+=$8}
$3=="M" {TM+=$7 ; MM+=$8}
###GENIC VS INTERGENIC###
$3==1 && $1=="gene" {T1g+=$7 ; M1g+=$8}
$3==1 && $1=="intergenic" {T1i+=$7 ; M1i+=$8}
$3=="M" && $1=="gene" {TMg+=$7 ; MMg+=$8}
$3=="M" && $1=="intergenic" {TMi+=$7 ; MMi+=$8}
###CONTEXT###
$3==1 && $9=="CG" {T1cg+=$7 ; M1cg+=$8}
$3==1 && $9=="CHG" {T1chg+=$7 ; M1chg+=$8}
$3==1 && $9=="CHH" {T1chh+=$7 ; M1chh+=$8}
$3=="M" && $9=="CG" {TMcg+=$7 ; MMcg+=$8}
$3=="M" && $9=="CHG" {TMchg+=$7 ; MMchg+=$8}
$3=="M" && $9=="CHH" {TMchh+=$7 ; MMchh+=$8}
###TOP VS BOTTOM###
$3==1 && $6=="+" {T1t+=$7 ; M1t+=$8}
$3==1 && $6=="-" {T1b+=$7 ; M1b+=$8}
$3=="M" && $6=="+" {TMt+=$7 ; MMt+=$8}
$3=="M" && $6=="-" {TMb+=$7 ; MMb+=$8}
###GENIC VS INTERGENIC && CONTEXT###
$3==1 && $1=="gene" && $9=="CG" {T1g_cg+=$7 ; M1g_cg+=$8}
$3==1 && $1=="gene" && $9=="CHG" {T1g_chg+=$7 ; M1g_chg+=$8}
$3==1 && $1=="gene" && $9=="CHH" {T1g_chh+=$7 ; M1g_chh+=$8}
$3==1 && $1=="intergenic" && $9=="CG" {T1i_cg+=$7 ; M1i_cg+=$8}
$3==1 && $1=="intergenic" && $9=="CHG" {T1i_chg+=$7 ; M1i_chg+=$8}
$3==1 && $1=="intergenic" && $9=="CHH" {T1i_chh+=$7 ; M1i_chh+=$8}
$3=="M" && $1=="gene" && $9=="CG" {TMg_cg+=$7 ; MMg_cg+=$8}
$3=="M" && $1=="gene" && $9=="CHG" {TMg_chg+=$7 ; MMg_chg+=$8}
$3=="M" && $1=="gene" && $9=="CHH" {TMg_chh+=$7 ; MMg_chh+=$8}
```

Stellenbosch University https://scholar.sun.ac.za


```
$3=="M" && $1=="intergenic" && $9=="CG" {TMi_cg+=$7 ; MMi_cg+=$8}
$3=="M" && $1=="intergenic" && $9=="CHG" {TMi_chg+=$7 ; MMi_chg+=$8}
$3=="M" && $1=="intergenic" && $9=="CHH" {TMi_chh+=$7 ; MMi_chh+=$8}
####################PRINT####################
###TOTALS###
END{print"Biotype\tSA1\tSAM\nTotal",T1,TM"\nMethylated",M1,MM"\nMethylation
(%)",(M1/T1)*100,(MM/TM)*100,
###GENIC VS INTERGENIC###
"\n\nBiotype\tSA1\tSA1\tSAM\tSAM\nRegion\tGenic\tIntergenic\tGenic\tIntergenic\nTotal
",T1g,T1i,TMg,TMi"\nMethylated",M1g,M1i,MMg,MMi"\nMethylation
(%)",(M1g/T1g)*100,(M1i/T1i)*100,(MMg/TMg)*100,(MMi/TMi)*100,
###CONTEXT###
"\n\nBiotype\tSA1\tSA1\tSA1\tSAM\tSAM\tSAM\nContext\tCpG\tCHG\tCHH\tCpG\tCHG\tC
HH\nTotal",T1cg,T1chg,T1chh,TMcg,TMchg,TMchh"\nMethylated",M1cg,M1chg,M1chh,MM
cg,MMchg,MMchh"\nMethylation
(%)",(M1cg/T1cg)*100,(M1chg/T1chg)*100,(M1chh/T1chh)*100,(MMcg/TMcg)*100,(MMch
g/TMchg)*100,(MMchh/TMchh)*100,
###TOP VS BOTTOM###
"\n\nBiotype\tSA1\tSA1\tSAM\tSAM\nStrand\tTop\tBottom\tTop\tBottom\nTotal",T1t,T1b,
TMt,TMb"\nMethylated",MTt,MTb,MMt,MMb"\nMethylation
(%)",(M1t/T1t)*100,(M1b/T1b)*100,(MMt/TMt)*100,(MMb/TMb)*100,
###GENIC VS INTERGENIC && CONTEXT###
"\n\nBiotype\tSA1\tSA1\tSA1\tSA1\tSA1\tSA1\tSAM\tSAM\tSAM\tSAM\tSAM\tSAM\nRegi
on\tGenic\tGenic\tGenic\tIntergenic\tIntergenic\tIntergenic\tGenic\tGenic\tGenic\tInterge
nic\tIntergenic\tIntergenic\nContext\tCpG\tCHG\tCHH\tCpG\tCHG\tCHH\tCpG\tCHG\tCHH\
tCpG\tCHG\tCHH\nTotal",T1g_cg,T1g_chg,T1g_chh,T1i_cg,T1i_chg,T1i_chh,TMg_cg,TMg_ch
g,TMg_chh,TMi_cg,TMi_chg,TMi_chh"\nMethylated",M1g_cg,M1g_chg,M1g_chh,M1i_cg,M
1i_chg,M1i_chh,MMg_cg,MMg_chg,MMg_chh,MMi_cg,MMi_chg,MMi_chh"\nMethylation
(%)",(M1g_cg/T1g_cg)*100,(M1g_chg/T1g_chg)*100,(M1g_chh/T1g_chh)*100,(M1i_cg/T1i_
cg)*100,(M1i_chg/T1i_chg)*100,(M1i_chh/T1i_chh)*100,(MMg_cg/TMg_cg)*100,(MMg_ch
g/TMg_chg)*100,(MMg_chh/TMg_chh)*100,(MMi_cg/TMi_cg)*100,(MMi_chg/TMi_chg)*1
00,(MMi_chh/TMi_chh)*100}'
```

Get Top and Bottom for hemimethylation

```
awk '{if ($6=="-" && $9=="CG") print$1,$2,$3,$4,$5-1,$6,$7,$8,$9,$10 ; else {if ('($6=="-" &&
$9=="CHG") print$1,$2,$3,$4,$5-2,$6,$7,$8,$9,$10 ; else print}}' |

awk 'BEGIN{OFS="\t"} {ph[$1"\t"$2"\t"$3"\t"$4"\t"$5]="empty"} $6=="+"
{toptotal[$1"\t"$2"\t"$3"\t"$4"\t"$5]=$7 ; topmeth[$1"\t"$2"\t"$3"\t"$4"\t"$5]=$8 ;
context[$1"\t"$2"\t"$3"\t"$4"\t"$5]=$9"\t"$10} $6=="-"
{bottomtotal[$1"\t"$2"\t"$3"\t"$4"\t"$5]=$7 ;
bottommeth[$1"\t"$2"\t"$3"\t"$4"\t"$5]=$8} END{for (all in ph) {print all, toptotal[all],
topmeth[all], bottomtotal[all], bottommeth[all], context[all]}}'
```

Prepare hemimethylation tables

```
##### $1=genic/intergenic
##### $2=gene number
##### $3=SA1/SAM
```

Stellenbosch University https://scholar.sun.ac.za


```
##### $4=contig
##### $5=position
##### $6=Top total
##### $7=Top meth
##### $8=Bottom total
##### $9=Bottom meth
##### $10=Context
##### $11=Codon

awk '
BEGIN{OFS="\t"}
$3==1 && $10=="CG" {Tt1cg+=1}
$3==1 && $10=="CG" && $7>=1 {Tm1cg+=1}
$3==1 && $10=="CHG" {Tt1hg+=1}
$3==1 && $10=="CHG" && $7>=1 {Tm1hg+=1}
$3==1 && $10=="CG" {Bt1cg+=1}
$3==1 && $10=="CG" && $9>=1 {Bm1cg+=1}
$3==1 && $10=="CHG" {Bt1hg+=1}
$3==1 && $10=="CHG" && $9>=1 {Bm1hg+=1}
$3=="M" && $10=="CG" {Ttmcg+=1}
$3=="M" && $10=="CG" && $7>=1 {Tmmcg+=1}
$3=="M" && $10=="CHG" {Ttmhg+=1}
$3=="M" && $10=="CHG" && $7>=1 {Tmmhg+=1}
$3=="M" && $10=="CG" {Btmcg+=1}
$3=="M" && $10=="CG" && $9>=1 {Bmmcg+=1}
$3=="M" && $10=="CHG" {Btmhg+=1}
$3=="M" && $10=="CHG" && $9>=1 {Bmmhg+=1}
$3==1 && $10=="CG" && $7>=1 && $9==0 {hT1cg+=1}
$3==1 && $10=="CG" && $7==0 && $9>=1 {hB1cg+=1}
$3==1 && $10=="CHG" && $7>=1 && $9==0 {hT1hg+=1}
$3==1 && $10=="CHG" && $7==0 && $9>=1 {hB1hg+=1}
$3=="M" && $10=="CG" && $7>=1 && $9==0 {hTmcg+=1}
$3=="M" && $10=="CG" && $7==0 && $9>=1 {hBmcg+=1}
$3=="M" && $10=="CHG" && $7>=1 && $9==0 {hTmhg+=1}
$3=="M" && $10=="CHG" && $7==0 && $9>=1 {hBmhg+=1}

END{print"Biotype\tSA1\tSA1\ttSA1\tSA1\tSAM\tSAM\tSAM\tSAM\nContext\tCpG\tCpG\tC
HG\tCHG\tCpG\tCpG\tCHG\tCHG\nStrand\tTop\tBottom\tTop\tBottom\tTop\tBottom\tTop
\tBottom\nTotal sites",Tt1cg,Bt1cg,Tt1hg,Bt1hg,Ttmcg,Btmcg,Ttmhg,Btmhg"\nMethylated
sites",Tm1cg,Bm1cg,Tm1hg,Bm1hg,Tmmcg,Bmmcg,Tmmhg,Bmmhg"\nHemimethylated
sites",hT1cg,hB1cg,hT1hg,hB1hg,hTmcg,hBmcg,hTmhg,hBmhg"\nSites methylated
(%)",(Tm1cg/Tt1cg)*100,(Bm1cg/Bt1cg)*100,(Tm1hg/Tt1hg)*100,(Bm1hg/Bt1hg)*100,(Tmm
cg/Ttmcg)*100,(Bmmcg/Btmcg)*100,(Tmmhg/Ttmhg)*100,(Bmmhg/Btmhg)*100"\nSites
hemimethylated
(%)",(hT1cg/Tt1cg)*100,(hB1cg/Bt1cg)*100,(hT1hg/Tt1hg)*100,(hB1hg/Bt1hg)*100,(hTmcg/
Ttmcg)*100,(hBmcg/Btmcg)*100,(hTmhg/Ttmhg)*100,(hBmhg/Btmhg)*100}'
```

Stellenbosch University https://scholar.sun.ac.za

Chapter 4: Comparing the expression of *DNMT3* and *TET*, as well as that of the differentially methylated genes between Russian wheat aphid biotypes SA1 and SAM

4.1: Introduction

Epigenetic modifications, such as DNA methylation, have the ability to influence an organisms phenotype, without altering the genotype, by exacting changes in splicing and gene expression (Russo et al., 1996). Various environmental factors have been shown to influence the expression of DNA methyl transferase-3 (DNMT3), the enzyme which establishes methylation patterns by catalyzing *de novo* methylation (Goll and Bestor, 2005; Okano et al., 1998). These include temperature, photoperiod, diet, symbiotic relationships and infestation density (Kim et al., 2018; Pasquier et al., 2014; Werren et al., 2010). This responsiveness of methylation to the environment, along with the phenotypic plasticity it imparts, independently of genotypic variation, makes DNA methylation an effective mechanism for organisms to adapt to changing environments within a single generation (Zhang et al., 2015). Functional DNA methylation systems have been found in some insect species and changes in methylation have been associated with changes in phenotype, mostly by observing correlations, but also through direct manipulation (Pegoraro et al., 2016).

Several methods have been developed to detect DNA methylation. There are three groups of molecular technologies utilized in these methods: methylation sensitive isoschizomers (McClelland et al., 1994), methylation sensitive DNA probes (Weber et al., 2005) and chemical modification (Darst et al., 2010). Currently, a method based on the chemical modification of unmethylated cytosine bases by bisulfite ions prior to library preparation, known as bisulfite sequencing, is regarded as the gold standard for the detection and subsequent study of DNA methylation (Li and Tollefsbol, 2011). Bisulfite sequencing can be used to generate data on the level, sequence context, strandedness and position of cytosine methylation, whereas the other two methods can only be used to quantify the levels of methylation in a relative fashion.

It has been shown by Suzuki and Bird (2008) that highly methylated genes can be inferred using only sequence information. They observed that the frequency of CpG sites in a highly methylated genes are lower than would be expected, based on the GC content of the gene (Suzuki and Bird, 2008). This is caused by the spontaneous deamination of methylated cytosine bases, which renders these bases chemically identical to thiamine (Morgan et al., 2004). The repair mechanism of the resulting thymine:guanidine mismatches is error prone and the guanidine is often replaced by an adenine,

Stellenbosch University https://scholar.sun.ac.za

rather than the thymine being replaced by cytosine (Poole et al., 2001). Over time CpG sites are abolished by this process of deamination and erroneous repair, leading to a lowered ratio of observed to expected CpG sites ($CpG_{O/E}$).

Unlike many insect orders studied to date, species belonging to the order Hemiptera possess a full complement of insect methylation machinery (Bewick et al., 2017). These species are, therefore, intuitively used as models to study DNA methylation. While studies on insect DNA methylation are not abundant, the pea aphid (*Acyrthosiphon pisum*) has been a popular Hemipteran model and, of all insects, its methylation profile is arguably the best studied (Dombrovsky et al., 2009; Hunt et al., 2010; Mathers et al., 2019; Walsh et al., 2010). The International Aphid Genomics Consortium (2006) demonstrated the link between a functional methylation system and decreased $CpG_{O/E}$ in insects by comparing the $CpG_{O/E}$ distribution of the pea aphid to two insect species (the fruit fly, *Drosophila melanogaster,* and the red flour beetle, *Tribolium castaneum*) that have very limited DNA methylation. In these two species the $CpG_{O/E}$ distribution showed a single peak at 1.0 (i.e. no depletion of CpG sites), while a bimodal distribution of $CpG_{O/E}$ could be observed for the pea aphid. Theoretically implying a group of methylated genes and a group of unmethylated genes (International Aphid Genomics Consortium, 2006).

The pea aphid hosts an obligate endosymbiont, *Buchnera aphidicola*, in a mutualistic interaction. The aphid is not able to synthesize essential amino acids and is therefore unable to feed on nutrient deficient plant sap. By incorporating the amino acid pathways of *B. aphidicola* an integrated system is formed, enabling the production of essential amino acids and broadening the host range of the aphid (Nakabachi et al., 2005). In a recent study, DNA methylation has been implicated in the regulation of this mutualistic plant-aphid interaction. In this study, a single strain of *A. pisum* was divided into six groups, three of which were maintained on a nutrient rich, preferential host (fava), while the other three were maintained on a nutrient poor host (alfalfa). After dissecting the aphids, body cells and bacteriocytes were harvested for DNA and RNA extractions. WGBS and RNA-seq was used to compare DNA methylation and transcription, respectively, between cell types as well as between the six host-plant groups. Using a multi-response permutation procedure (MRPP) it was showed that methylation levels differed significantly between bacteriocytes and body cell and that the methylation differences are more pronounced between bacteriocytes (delta = 0.2759) than between body cells (delta = 0.1772). Fourteen of the 441 genes that were both differential methylated and differentially expressed have been identified as key regulators of the aphid-*Buchnera* interaction. From these results, the researchers motivated that DNA methylation might play a key role in the integration of the aphid and bacterial amino acid pathways and therefor in the adaptability of the aphid to feed on nutrient poor hosts (Kim et al., 2018). Feeding on non-preferential host plants have similarly been shown to cause

Stellenbosch University https://scholar.sun.ac.za

major changes in transcription in another hemipteran species, the Russia wheat aphid (RWA) (Burger et al., 2017).

Burger et al. (2017) used two South African RWA biotypes to investigate the effect of feeding on preferential and non-preferential host plants on the transcriptome of a virulent (SAM) and non-virulent biotype (SA1). Both biotypes were reared on their respective host of preference (SA1 on Tugela and SAM on Tugela-DN), after which they were moved onto sixteen wheat lines with differing genotypic backgrounds and resistance. RNA was sampled at three time points (0h, 6h and 48h after host shifts) and converted to cDNA to quantify relative expression using amplified fragment length polymorphism (cDNA-AFLP). Feeding on different host plants had a major effect on the transcriptome of RWA as 42.63% of the transcripts deviated significantly from the 0h controls, in terms of level of transcription. Changes in transcription in biotype SA1 correlated strongly with the genotype of the host plant whereas the changes in the transcription in biotypes SAM correlated with the sampling times. The researchers concluded that this indicates a greater ability of biotype SAM to adapt quickly to the defense responses initiated by plants at 6h and 48h after infestation (Burger et al., 2017).

This study was conducted to explore the possibility of DNA methylation playing a role in the adaptation of RWA to non-preferential hosts as it does in the pea aphid, albeit a different preference criterion (resistance *versus* lack of nutrients). A difference in expression of genes that are differentially methylated between SA1 (low virulence) and SAM (highly virulent biotype) would serve as a positive indication that this might be the case. Therefore, differentially methylated genes were identified and filtered based on gene ontologies, with respect to molecular function. The relative expression of these genes, as well as that of two key methylation genes, *DNMT3* and ten-eleven translocase (*TET*) were quantified at 0h, 6h and 48h after being transferred from a preferential host (Tugela) to a non-preferential host (Tugela-DN). The enzymes code for by *DNMT3* and *TET* are respectively responsible for the addition and removal of DNA methylation (Okano et al., 1998; Tahiliani et al., 2009).

4.2: Materials and methods

4.2.1: Detection of differentially methylated genes

The R-package DSS-single (Wu et al., 2015) was used to calculate which genes are significantly differentially methylated (P-value < 0.05) between SA1 and SAM from the WGBS data reported in Chapter 3. For the analysis, only genic CpG loci, with at least a ten times coverage in both biotypes were considered. This amounted to 2,521,410 CpG sites. A test for differentially methylated loci was conducted with the DMLTest function. This test is a three-step process, in which the mean methylation levels are calculated, before estimating the dispersion levels with a built in “smoothing” algorithm, which generates pseudo-replicates, allowing for single sample comparisons. The default smoothing

Stellenbosch University https://scholar.sun.ac.za

span of 500 base pairs was used. A Wald test (Wald, 1943) was then performed, using the null hypothesis that there is no difference in methylation between samples. The CallDMR function was used to identify differentially methylated regions using information from the differentially methylated loci, such as the number of CpG sites in a region and the percentage of sites in a region scored as significant (p-value < 0.05). All the above information was used to calculate a combined score for each region, referred to as an area statistic, which can be used to sort regions based on the degree of differentiation in CpG methylation (Wu et al., 2015). The Blast2GO suite (Conesa et al., 2005) was used to search for the gene ontologies of genes containing a differentially methylated region. Combined GO graphs were created within the suite for biological processes, cellular location and molecular function.

4.2.2: Observed over expected cytosine ratios

Using the SAM biotype reference genome, the observed over expected number of cytosine bases were calculated. This is commonly seen for the CpG context, denoted as $CpG_{O/E}$ (Hunt et al., 2010; Kocher et al., 2015), with the data on all three contexts of cytosine methylation available from the Bismarck pipeline, the other, often overlooked $CHG_{O/E}$ and $CHH_{O/E}$ ratios were also included in this study. The calculations for each of the contexts were as such:

$$C_pG_{O/E} = \frac{F_{C_pG}}{F_C . F_G}$$

$$CHG_{O/E} = \frac{F_{CAG} + F_{CTG} + F_{CCG}}{3(F_C . F_{1-G} . F_G)}$$

$$CHH_{O/E} = \frac{F_{CAA} + F_{CAT} + F_{CAC} + F_{CTA} + F_{CTT} + F_{CTG} + F_{CGA} + F_{CGT} + F_{CGG}}{9(F_C . 2F_{1-G})}$$

Where F represents the frequency of the subscripted nucleotide/dinucleotide/trinucleotide, in the reference genome. As a genome is not available for biotype SA1 however, the calculations were only performed for biotype SAM.

A Pearson's correlation coefficient (Havlicek and Peterson, 1976) as well as a Welch's t-test (Welch, 1947) was used to assess the relationship between the $CpG_{O/E}$ and the level of methylation of each gene for which sequence and methylation data is available. The Welch's t-test was performed with a null hypothesis of no difference in the mean $CpG_{O/E}$ of hypermethylated and hypomethylated genes in biotype SAM.

4.2.3: Host plant cultivation

All cultivars were planted in 12.5 cm pots filled with light expanded clay aggregate (LECA) medium. The plants were well watered throughout with a modified Hoagland's solution (Hoagland and Arnon,

Stellenbosch University https://scholar.sun.ac.za

1950); with a final concentrations of 6.4 mM KNO_3, 4 mM $Ca(NO_3)_2.4H_2O$, 2 mM $NH_4H_2PO_4$, 2 mM $MgSO_4.7H_2O$, 45 µM $FeCl_2.4H_2O$, 201 µM EDTA, 5 µM SiO_2, 0.5 µM $MnCl_2.4H_2O$, 0.2 µM $ZnSO_4.7H_2O$, 0.2 µM $CuSO_4.5H_2O$, 4.6 µM H_3BO_3, 0.1 µM $(NH_4)_6Mo_7O_{24}.4H_2O$, inside a temperature controlled growth room, at 20 ± 1°C, on a twelve-hour day-night cycle, under a combination of fluorescent and LED grow lights, with a plant density of eight plants per pot.

4.2.4: Aphid rearing

Colonies of parthenogenetic, apterous, female aphids of South African RWA biotypes SA1, and SAM, expressing different levels of virulence, were separately established in BugDorm cages (MegaView Science Education Services) in an insectary with the following conditions: 22.5 ± 2.5°C, 40% relative humidity, and continuous artificial lighting from high pressure sodium lamps. The Russian wheat aphid biotypes were maintained on near isogenic wheat lines Tugela, RWA susceptible, and Tugela *Dn1*, RWA resistant, with SA1 maintained on the susceptible wheat line, and SAM on wheat containing the *Dn1* resistance gene. To avoid any differences in methylation as an artefact of the biotypes feeding on different wheat varieties, both biotypes were transferred to the susceptible cultivar, SST362, one month prior to performing host-shift experiments.

4.2.5: Expression of differentially methylated genes

A sub-set of genes found to be differentially methylated were selected based on gene ontologies and the area statistic (Wu et al., 2015). The relative expression of these genes, as well as that of two key methylation genes, *DNMT3* and *TET* (Table 4.1), were quantified to investigate rates of methylation and demethylation, as well as the response of differentially methylated genes to stress.

4.2.5.1: Host-shifts and sampling

Pure colonies of SA1 and SAM were founded on the susceptible wheat cultivar SST356. One month prior to performing host shifts, both biotypes were moved to the susceptible cultivar Tugela. The aphid populations were divided and moved onto two resistant cultivars, Tugela *Dn1* and Tugela *Dn5*. Biotype SAM has overcome the resistance present in these cultivars, while biotype SA1 is avirulent (Breeds et al., 2018).

Three samplings were performed: The first before performing host shifts and the other two at 6 and 48 hours after performing host shifts. With each sampling, 30 adult, apterous aphids were collected from each population, in triplicate (n=3). Aphids were flash frozen with liquid nitrogen during sample collection.

Stellenbosch University https://scholar.sun.ac.za

4.2.5.2: RNA extractions and cDNA synthesis

The frozen aphids were ground with micro-pestles and RNA was extracted using RNeasy Mini Kit (Qaigen), following the manufacturers recommended protocol for insect material. RNA integrity was verified using agarose gel electrophoresis as described in Chapter 3. All gels consisted of 2% agarose and 3% commercial bleachthe bleach denatures the secondary structure of RNA and of RNases and therefore provides consistent migration while preventing RNA degradation (Aranda et al., 2012). cDNA synthesis was performed using SensiFAST cDNA Synthesis Kit (Bioline), with 200 ng of input RNA. First and second strand synthesis occurred during incubation steps of 10 minutes at 25°C and 15 minutes at 42°C, respectively, followed by an inactivation step at 85°C for 5 minutes. The quantity of cDNA in each sample was assessed through use of a Qubit 2.0 fluorometer (Thermo Fisher Scientific).

4.2.5.3: Real-time PCR amplification

A five point, two times serial dilution of a zero-hour SA1 sample was used to generate quantification standards. The relative expression of five differentially methylated genes, as well as *DNMT3* and *TET* were calculated using Pfaffl's mathematical model (Pfaffl, 2001) for each time point (0h, 6h and 48h). A CFX96 Real-Time System (Bio-Rad) was used to perform the real-time PCR analysis. Each reaction started with a denaturation step at 95 °C for 3 minutes, followed 40 cycles of amplification, consisting of a denaturation step at 95 °C for 10 seconds, an annealing step at the relevant temperature for each primer set (Table 4.1) for 30 seconds, and an extension step at 72 °C for 30 seconds. Primer3 (Untergasser et al., 2012) was used to design primers with low or no hairpin formation and self-dimerization temperatures. A melt curve analysis was also performed for each reaction, to verify the absences of non-specific amplification: The incubation temperature was increased in 5 seconds intervals, 0.5 °C at a time, from 65 °C to 95 °C. Amplification plots and melt curves for each of the reactions are included in Appendix B (Figure B1 to Figure B42). Two reference genes, *L27* and *L32* were used for normalisation, as these genes were found to be stably expressed on various host cultivars (Shakesby et al., 2009; Sinha and Smith, 2014).

4.2.5.4: Statistical analysis

A single analysis of variance (ANOVA) was used, with alpha set at 0.05, to test the null hypothesis of no difference between the mean relative expression for different biotypes, time points or cultivars. Where the null hypothesis was rejected and a difference between the means was observed, a Fisher's LSD test was performed, comparing each treatment pairwise with every other treatment.

Stellenbosch University https://scholar.sun.ac.za

Table 4.1: The sequences and annealing temperatures of nine primer pairs, used during real-time PCR analysis, as well as the description of the target genes.

Primer	Sequence	Tm (°C)	Product size (pb)	Target gene description
DMG1_F	CCACCGATGACTGAAAACTGG	59	376	Low-density lipo receptor, [*Pediculus humanus corporis*]
DMG1_R	ATTCGGTAACATGTTCACTTGCA			
DMG2_F	GCACTTCTAGGATTTATTCTTGCTCT	59	518	Plasma membrane calcium-transporting ATPase 2 isoform X1 [*Bombus terrestris*]
DMG2_R	CAGTCTTCACCAGGTTTAAATATACCA			
DMG3_F	GACTGGAAAACCGTCACCCA	60	251	Basement membrane-specific heparan sulfate proteoglycan core
DMG3_R	CTCGATCTTCCACAGGTCCG			
DMG4_F	GGTGATTGAGGTTTTACAGACTG	56	472	Autophagy-related 13 homolog isoform X1
DMG4_R	AGAAAACGAGCCGAATATTGG			
DMG5_F	TTAACCACATCAGGCGCCAT	55	195	Gastrula zinc finger -like isoform X1 [*Diaphorina citri*]
DMG5_R	CGCCTGTGTGAATACGAATATGAC			
DNMT3_F	GGCTTTTGAAACAAGTGCTGC	54	120	DNA methyltransferase 3
DNMT3_R	AACCGGCTTCTTTGTTGGAC			
TET_F	GGCACCCAAAGTACATCCGA	57	244	Ten-eleven translocation enzyme
TET_R	GCGTGTAGTTCCTGTTTTGC			
L27_F	ACCAGCACGATTTTACCAGATTTC	54	90	60S ribosomal protein L27
L27_R	CGTAGCCTGCCCTCGTGTA			
L32_F	CGTCTTCGGACTCTGTTGTCAA	55	98	60S ribosomal protein L32
L32_R	CAAAGTGATCGTTATGACAAACTCAA			

4.3: Results

4.3.1: Differentially methylated genes

A total of 148 genes were found to be differentially methylated between the two biotypes (p-value < 0.05). Most of the differentially methylated genes were hypermethylated in biotype SA1 (89 vs 59). The gene showing the highest difference in methylation was methylated 4.6 times higher in biotype SA1 than in biotype SAM (78% vs 17%). The output data from DSS-single as well as the Blast2GO searches for all 148 genes can be found in Appendix B (Table B1 and Table B2).

Third level combined GO graphs of the hypermethylated genes in each biotype were exported from the Blast2GO suite independently and have been summarised as bar charts in Figure 4.1 to Figure 4.2.

Stellenbosch University https://scholar.sun.ac.za

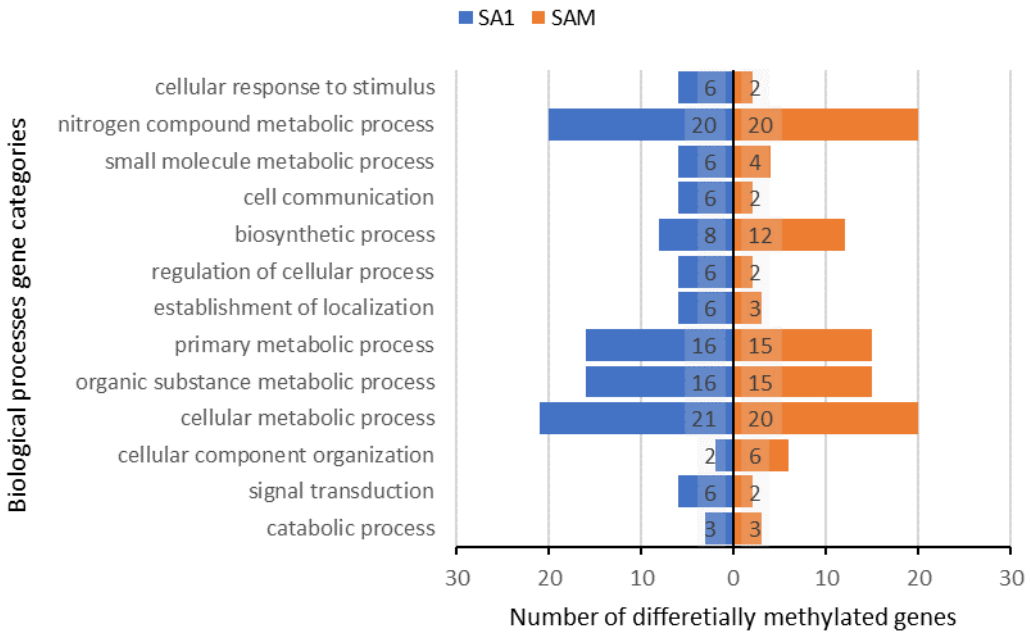

Figure 4.1: The number of genes hypermethylated (p-value < 0.05) in each biotype, divided into the ontological categories of biological processes that the gene products have been assigned to by BLAST2GO (Conessa et al., 2014).

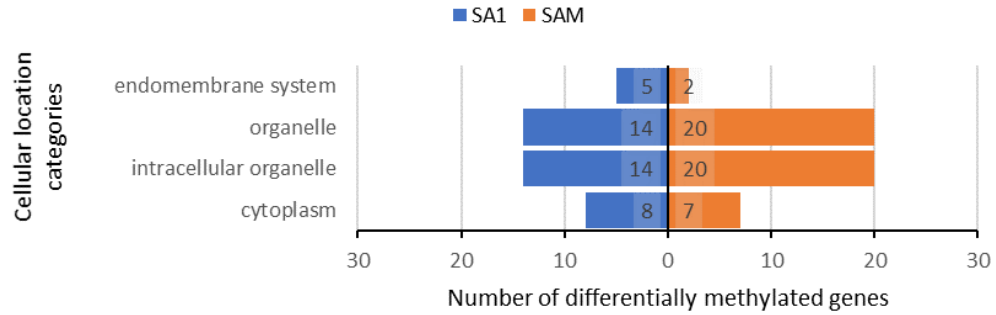

Figure 4.2: The number of genes hypermethylated (p-value < 0.05) in each biotype, divided into the ontological categories of cellular location that the gene products have been assigned to by BLAST2GO (Conessa et al., 2014).

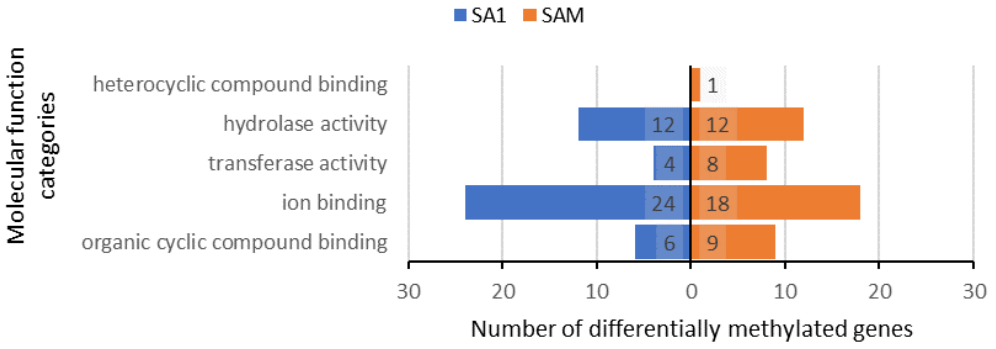

Figure 4.3: The number of genes hypermethylated (p-value < 0.05) in each biotype, divided into the ontological categories of molecular functions that the gene productshave been assigned to by BLAST2GO (Conessa et al., 2014).

Stellenbosch University https://scholar.sun.ac.za

The number of hypermethylated genes in each biotype per category is very similar, for example, in the highly represented “cellular metabolic process” category, a total of 41 genes are differentially methylated, 21 are hypermethylated in SA1 and 20 are hypermethylated in SAM. Similarly, in the underrepresented “catabolic process” category, with a total of 6 differentially methylated genes, 3 are hypermethylated in each biotype (Figure 4.1). While 30 more genes were hypermethylated in SA1 than were hypermethylated in SAM, many of the genes that were hypermethylated in SA1 were not successfully assigned ontologies. Had these genes been assigned ontologies, the discrepancy between the number of genes hypermethylated in each category would have been larger between the biotypes.

4.3.2: Observed over expected ratios of methylation sites

Areas with high levels of methylation, compared to the rest of the genome, have been associated with a lower number of observed CpG sites than would be expected. As such, areas with a lower than expected occurrence of CpG sites have been used to infer hypermethylated regions from sequence data (Suzuki et al., 2007). The observed number of CpG, CHG and CHH sites in each gene (based on the reference SAM genome) was divided by the expected number of sites, calculated using the GC content and gene length (Figure 4.4). The O/E ratio for the reverse sequences of the possible methylation sites i.e. GpC, GHC and HHG were also included to be used as a control.

Stellenbosch University https://scholar.sun.ac.za

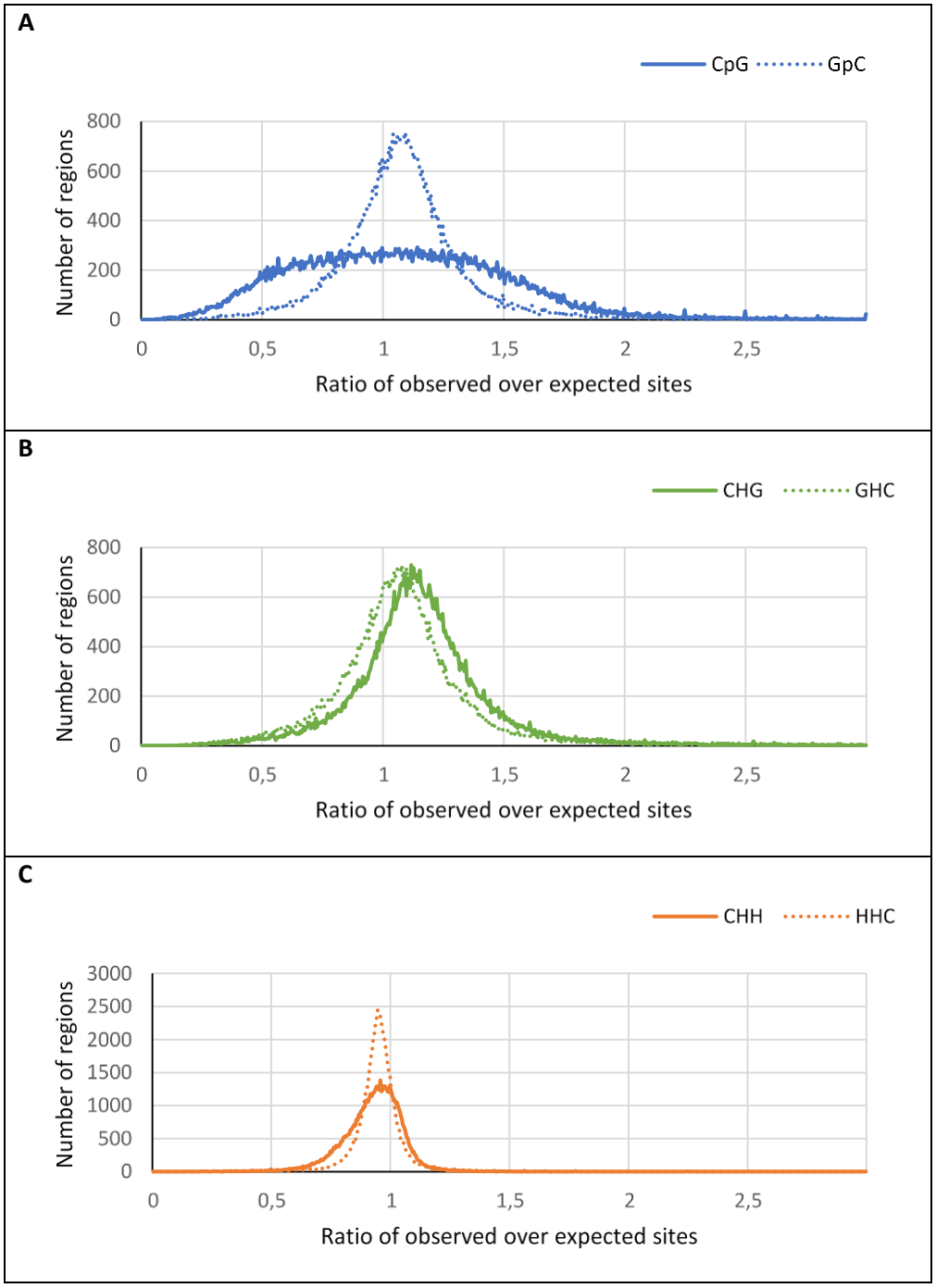

Figure 4.4: Line graphs comparing the ratio of observed potential methylation sites over expected potential methylation sites per gene, in all three available contexts, to the ratio of observed over expected number of control sites in the same gene. The reverse of potential methylation sites was used as a control. The number of expected potential methylation sites, as well as control sites within a gene were calculated using the individual GC content of each gene.

For all three contexts the observed over expected ratio of control sites peak close to a value of 1.0. This indicates they appear at the expected frequency, across the genome, regardless of GC content. The CpG context deviated most clearly from the expected occurrence per gene. The distribution of $CpG_{O/E}$ had a very wide peak and a large portion of the curve falls between 0.5 and 1.0, indicating that many genes show less CpG sites than would be expected. The number of possible methylation sites in

Stellenbosch University https://scholar.sun.ac.za

the other two contexts do not seem to be diminished, the CHG context, in fact, peaks to the right of 1.0, which would suggest that there are more CHG sites than expected (Figure 4.4).

Only the occurrence of CpG sites seemed to be deviating from the expected, as such it was the only context investigated further, by calculating the O/E ratios for genes and intergenic regions (Figure 4.5), as well as exons and introns (Figure 4.6).

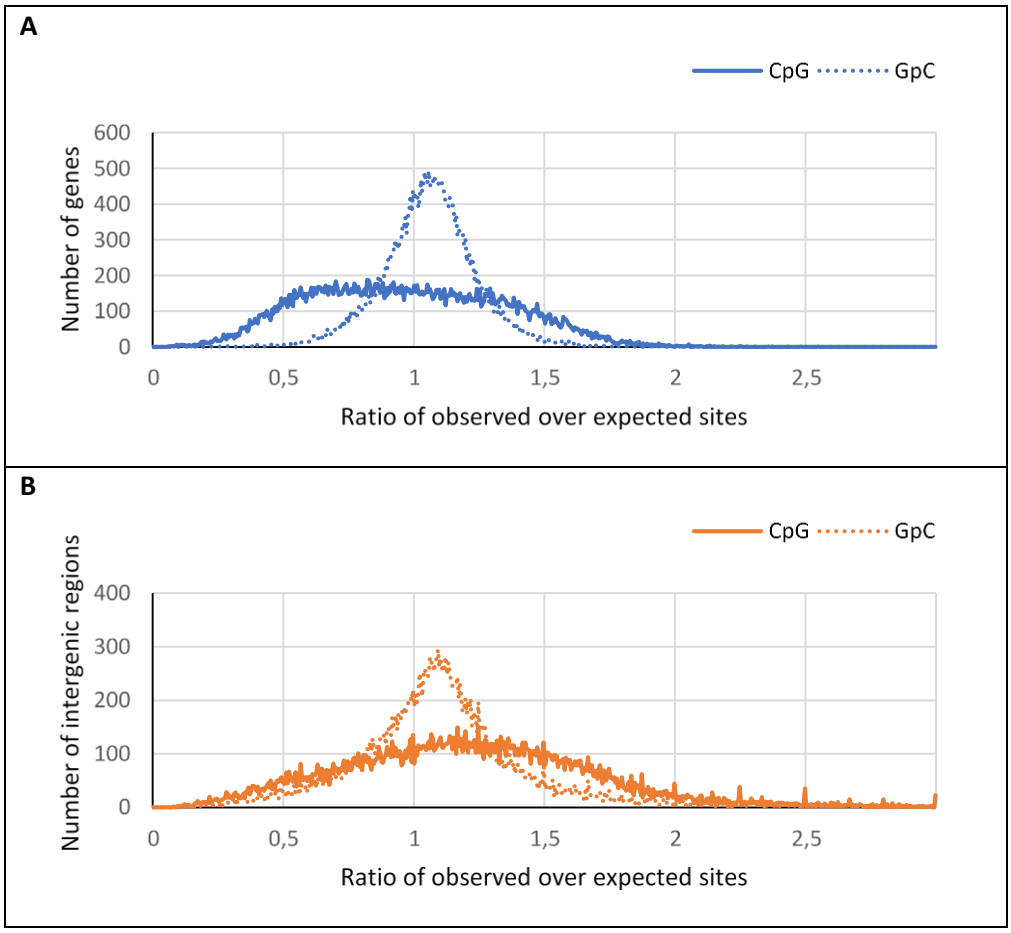

Figure 4.5: Line graphs comparing the ratio of observed CpG sites over the expect number of CpG sites in **(A)** genic and **(B)** intergenic regions to the ratio of observed over expected control sites in the same gene or intergenic region. The reverse of CpG sites, GpC sites, were used as a control. The expected number of CpG sites, as well as GpC sites within a gene or intergenic region were calculated using the GC content of each gene or intergenic region.

In the intergenic regions, the GpC control sites appear at the expected frequency and peak at an O/E ratio of 1:1. The under representation of CpG sites apparent within genes cannot be seen in intergenic regions, as the O/E ratios for intergenic regions seem to be skewed towards the right, indicating a higher than expected occurrence (Figure 4.5).

Stellenbosch University https://scholar.sun.ac.za

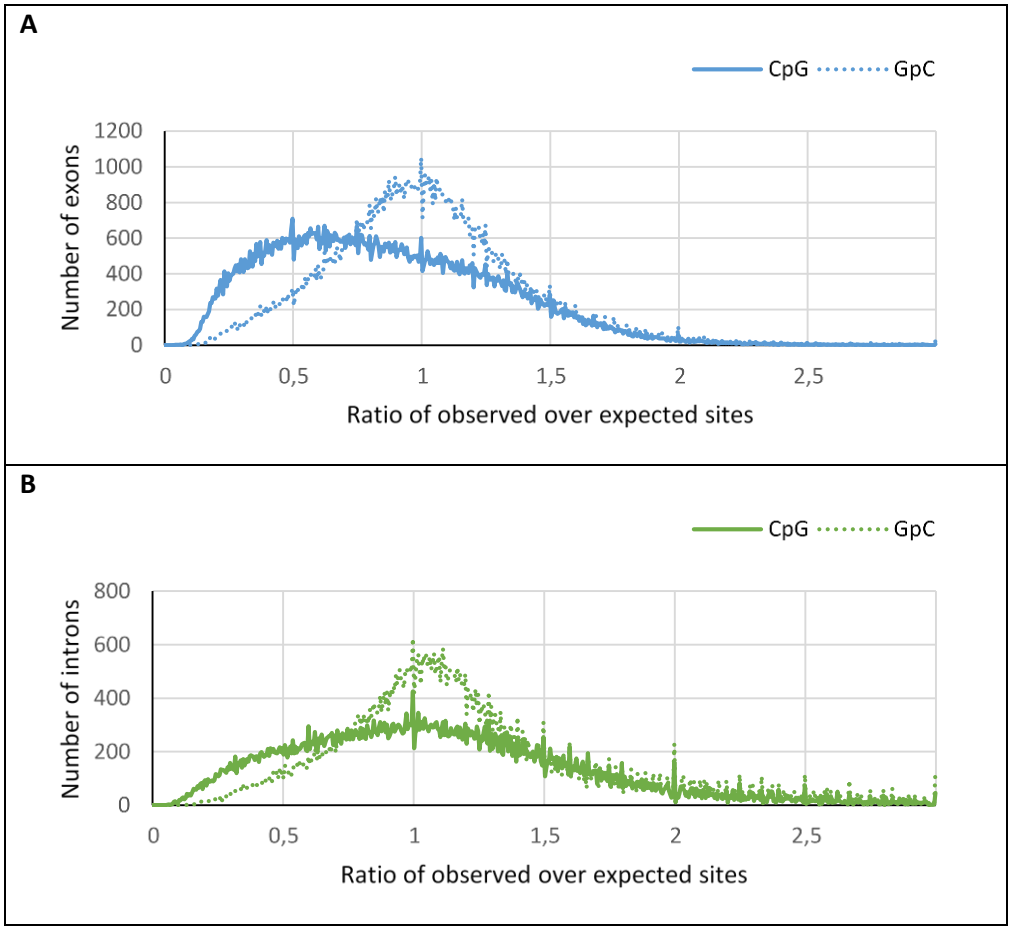

Figure 4.6: Line graphs comparing the ratio of observed CpG sites over the expect number of CpG sites in **(A)** exons and **(B)** introns to the ratio of observed over expected control sites in the same exon or intron. The reverse of CpG sites, GpC, was used as a control. The number of CpG sites, as well as GpC sites within a gene were calculated using the GC content of each gene.

Dividing genic regions into introns and exons does not change the ratio of observed over expected GpC sites used as controls. The O/E exonic CpG sites peak at about 0.5, while the intronic CpG O/E peaks at the expected value of 1.0. Therefore, less CpG sites appear within exons than would be expected, while the frequency of CpG sites within introns is not diminished (Figure 4.6).

Figure 4.7 displays the CpG data in a box and whisker plot, which shows that the mean for the distribution of control GpC sites are greater than one for all four of the considered regions. The box with the lowest mean $CpG_{O/E}$ of all is the one representing the exonic regions, it is low enough to bring the mean for genic regions below 1.0. Intergenic regions have the highest mean $CpG_{O/E}$ and is the only region where the mean $CpG_{O/E}$ is higher than that of the $GpC_{O/E}$.

Stellenbosch University https://scholar.sun.ac.za

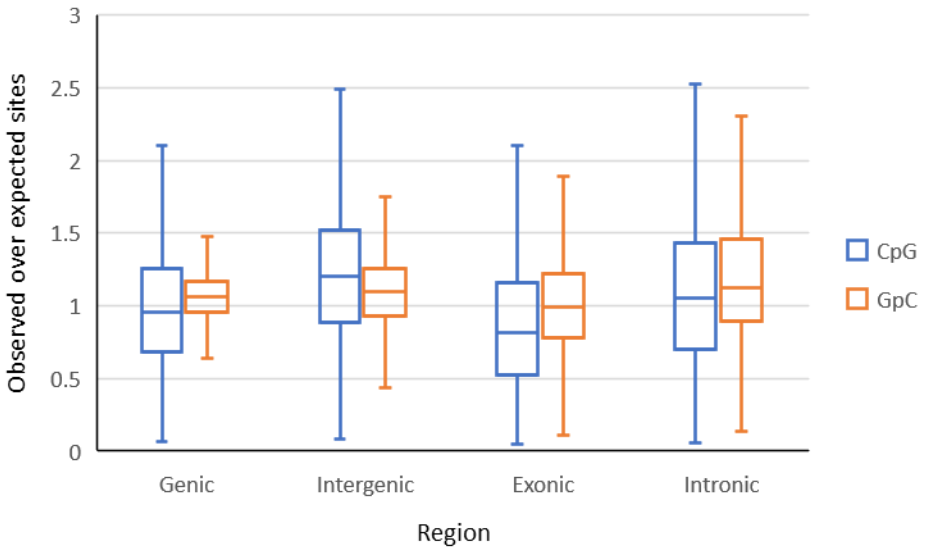

Figure 4.7: The observed over expected number of CpG and GpC sites in genic and intergenic regions. The boundaries of the boxes indicated the upper and lower quartile and the handles indicate the 10th and 90th percentiles.

The ratio of the difference between means (DBM) and overall visible spread (OVS) was used to compare the CpG boxes with the GpC boxes (Table 4.2). For large sample size a ratio of 10% implies a true difference between means.

Table 4.2: The difference between means (DBM) to overall visible spread (OVS) ratio of $CpG_{O/E}$ compared to $GpC_{O/E}$ box and whisker plots of genic and intergenic regions.

Region	**Genic**	**Intergenic**	**Exonic**	**Intronic**
DBM/OVS	18.71%	16.49%	25.89%	10.04%

The Pearson's correlation coefficient for CpG methylation level and $CpG_{O/E}$ was calculated as 0.20, which is considered to be a negligible correlation. A scatter plot using these two criteria as variables does not seem to imply any relationship (Figure 4.8). A p-value of 0.32 was calculated using Welch's t-test, the null hypothesis of no difference between the $CpG_{O/E}$ of hypermethylated and hypomethylated genes in biotype SAM cannot be rejected.

Stellenbosch University https://scholar.sun.ac.za

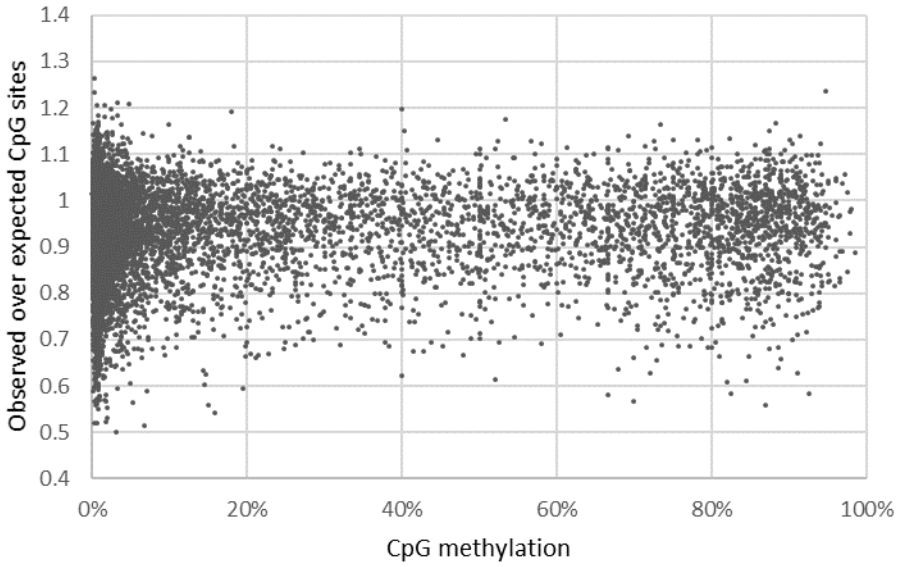

Figure 4.8: A scatter plot showing the relationship between CpG methylation and the $CpG_{O/E}$ of all genes for which methylation and sequence data is available. Pearson's correlation coefficient = 0.20.

4.3.3: Expression of differentially methylated genes

The expression of 5 genes that are deferentially methylated between biotypes SA1 and SAM, as well as that of two key genes in the methylation system, *DNMT3* and *TET*, were relatively quantified against the expression *L27* and *L32*. While the expression patterns for the 5 differentially methylated genes were very similar between biotypes SA1 and SAM (Figure 4.9 to 4.13), there was a significant difference in the expression of both genes involved in methylation (Figure 4.14 and 4.15). At 6-hours after transferring the population to Tugela *Dn5* there is a major increase in relative expression of *DNMT3* and *TET* in biotype SAM. Biotype SA1 does also show an increase in the expression of these proteins, but not nearly to the same extent.

Stellenbosch University https://scholar.sun.ac.za

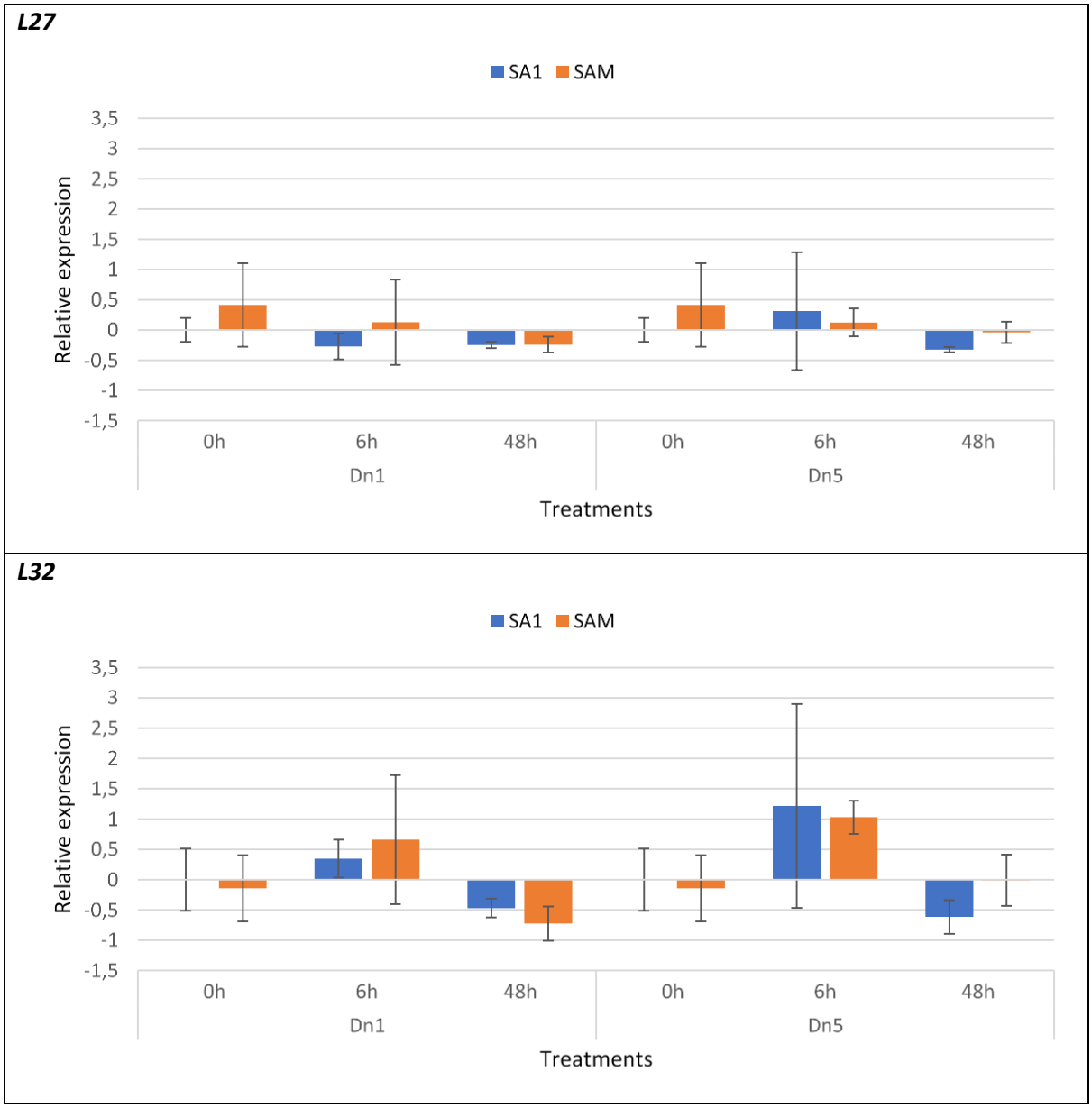

Figure 4.9: The expression of DMG1 (Low-density lipo receptor) in RWA biotypes SA1 and SAM, at 0, 6, and 48-hours after performing host-shifts from Tugela to Tugela *Dn1* and Tugela *Dn5*, relative to the expression of *L27* and *L32* of SA1 at 0-hours. The null hypothesis of no difference in mean expression relative to *L32* was rejected. Results of the ANOVA (Table B2) as well as a matrix showing all the statistically significant differences between treatments (Table B3), calculated using Fisher's LSD, can be found in appendix B.

Stellenbosch University https://scholar.sun.ac.za

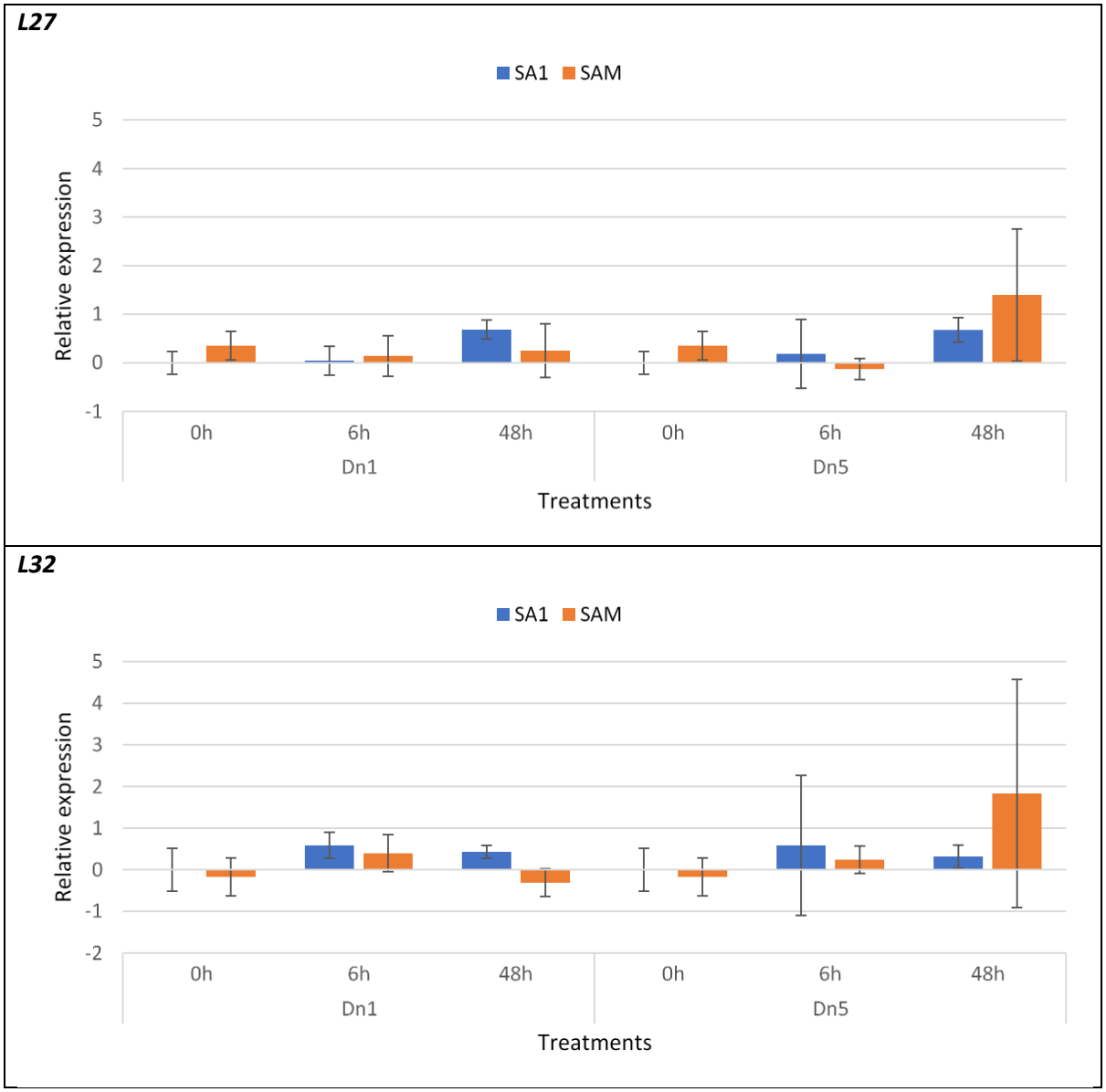

Figure 4.10: The expression of DMG2 (Plasma membrane calcium-transporting ATPase 2 isoform X1) in RWA biotypes SA1 and SAM, at 0, 6, and 48-hours after performing host-shifts from Tugela to Tugela *Dn1* and Tugela *Dn5*, relative to the expression of *L27* and *L32* of SA1 at 0-hours. The null hypothesis of no difference in mean expression could not be rejected.

Stellenbosch University https://scholar.sun.ac.za

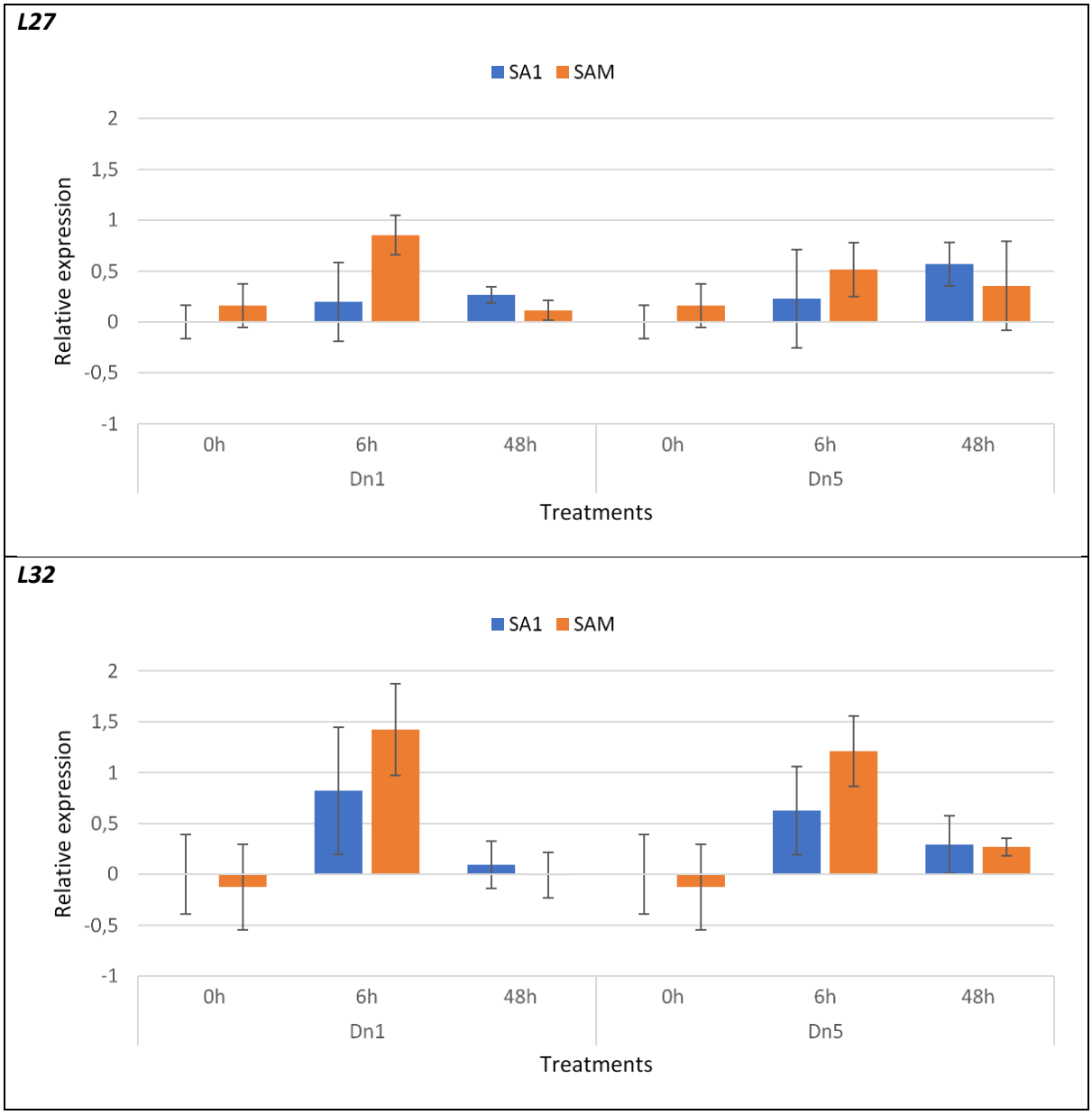

Figure 4.11: The expression of DMG3 (Basement membrane-specific heparan sulfate proteoglycan core) in RWA biotypes SA1 and SAM, at 0, 6, and 48-hours after performing host-shifts from Tugela to Tugela *Dn1* and Tugela *Dn5*, relative to the expression of *L27* and *L32* of SA1 at 0-hours. The null hypothesis of no difference in mean expression relative to *L32* was rejected. Results of the ANOVA (Table B4) as well as a matrix showing all the statistically significant differences between treatments (Table B5), calculated using Fisher's LSD, can be found in appendix B.

Stellenbosch University https://scholar.sun.ac.za

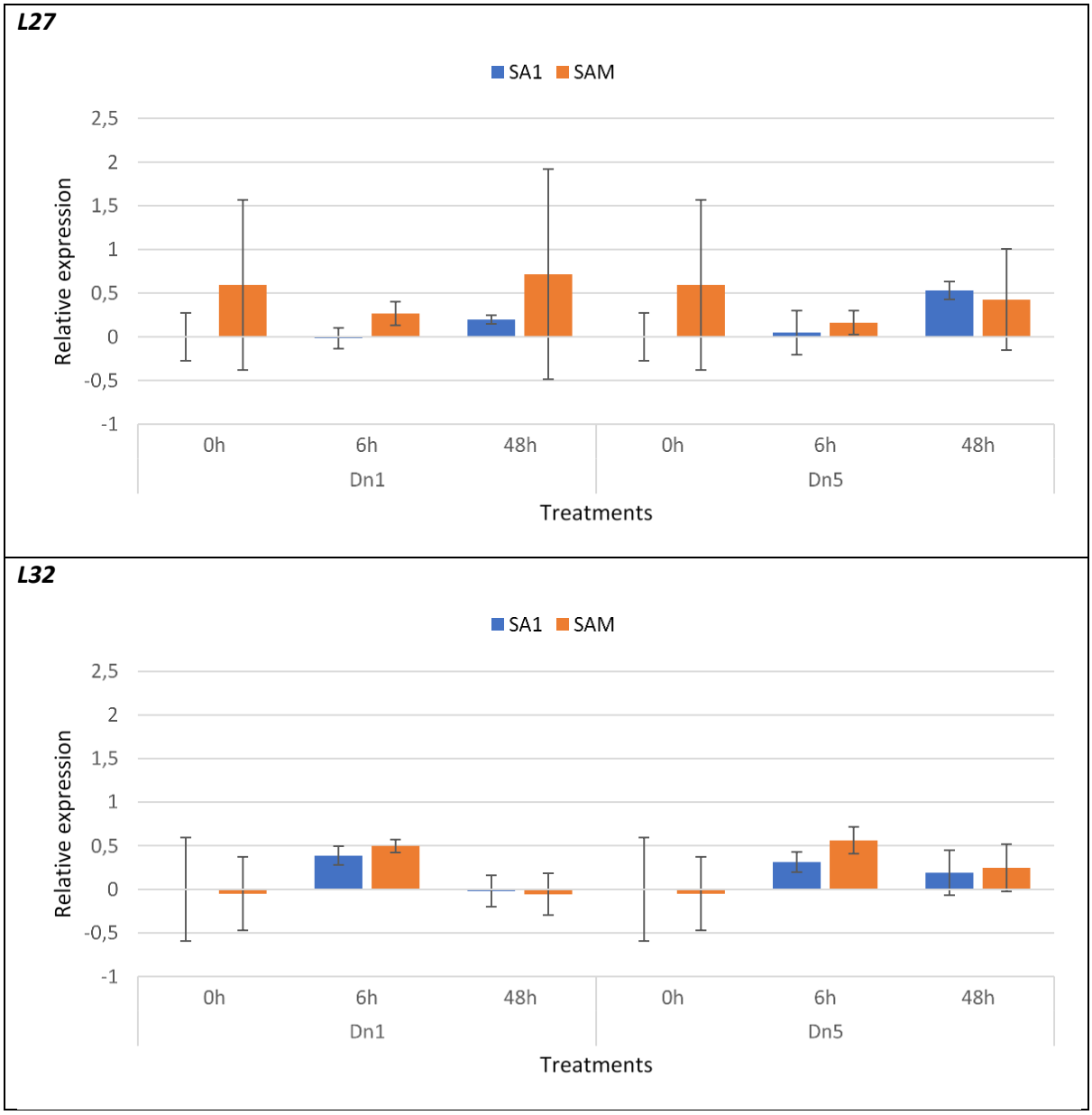

Figure 4.12: The expression of DMG4 (Autophagy-related 13 homolog isoform X1) in RWA biotypes SA1 and SAM, at 0, 6, and 48-hours after performing host-shifts from Tugela to Tugela *Dn1* and Tugela *Dn5*, relative to the expression of *L27* and *L32* of SA1 at 0-hours. The null hypothesis of no difference in mean expression could not be rejected.

Stellenbosch University https://scholar.sun.ac.za

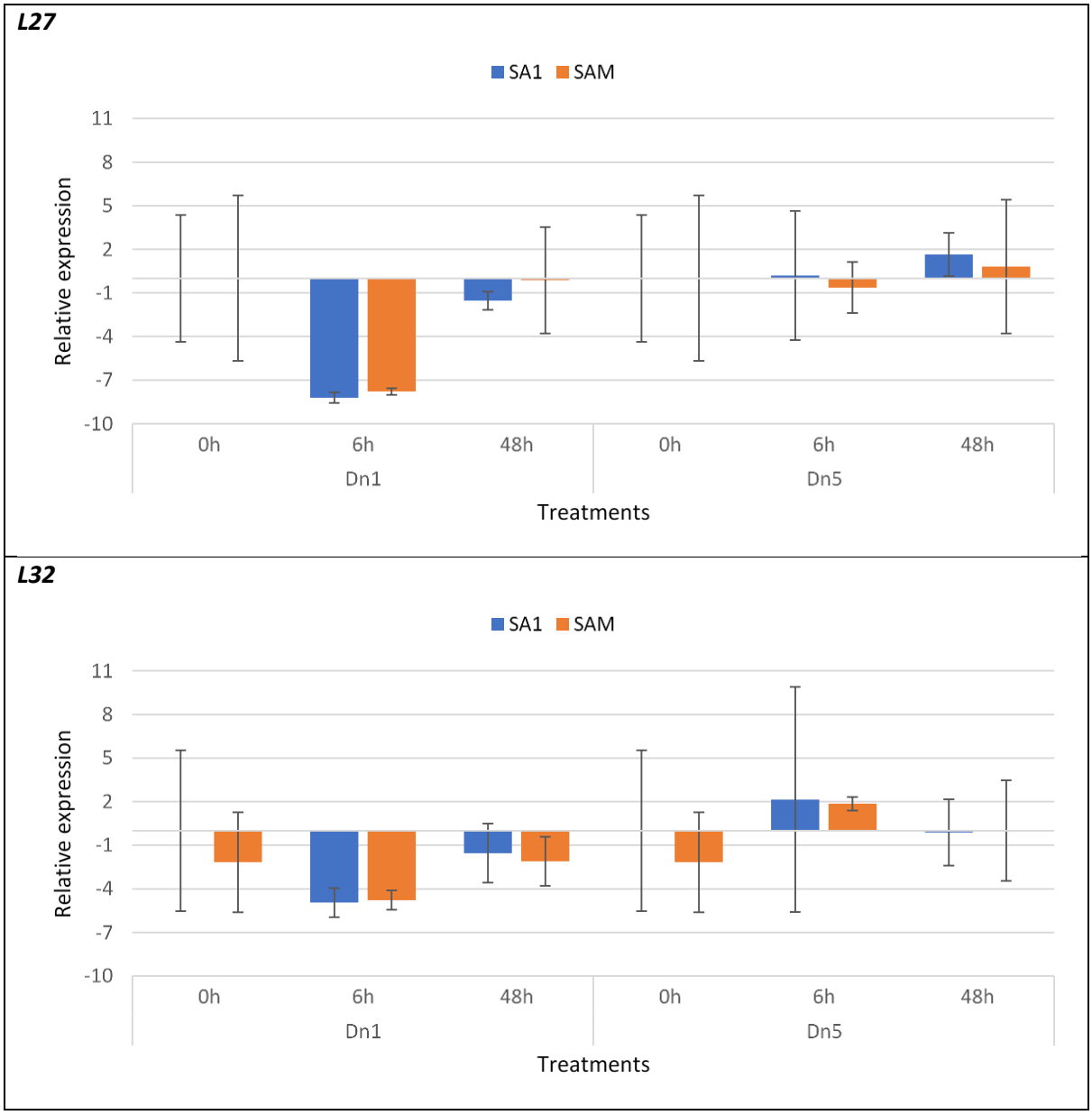

Figure 4.13: The expression of DMG5 (Gastrula zinc finger -like isoform X1) in RWA biotypes SA1 and SAM, at 0, 6, and 48-hours after performing host-shifts from Tugela to Tugela *Dn1* and Tugela *Dn5*, relative to the expression of *L27* and *L32* of SA1 at 0-hours. The null hypothesis of no difference in mean expression relative to both *L27* and *L32* was rejected. Results of the ANOVA's (Table B6 and Table B8) as well as matrices showing all the statistically significant differences between treatments (Table B7 and Table B9), calculated using Fisher's LSD, can be found in appendix B.

Stellenbosch University https://scholar.sun.ac.za

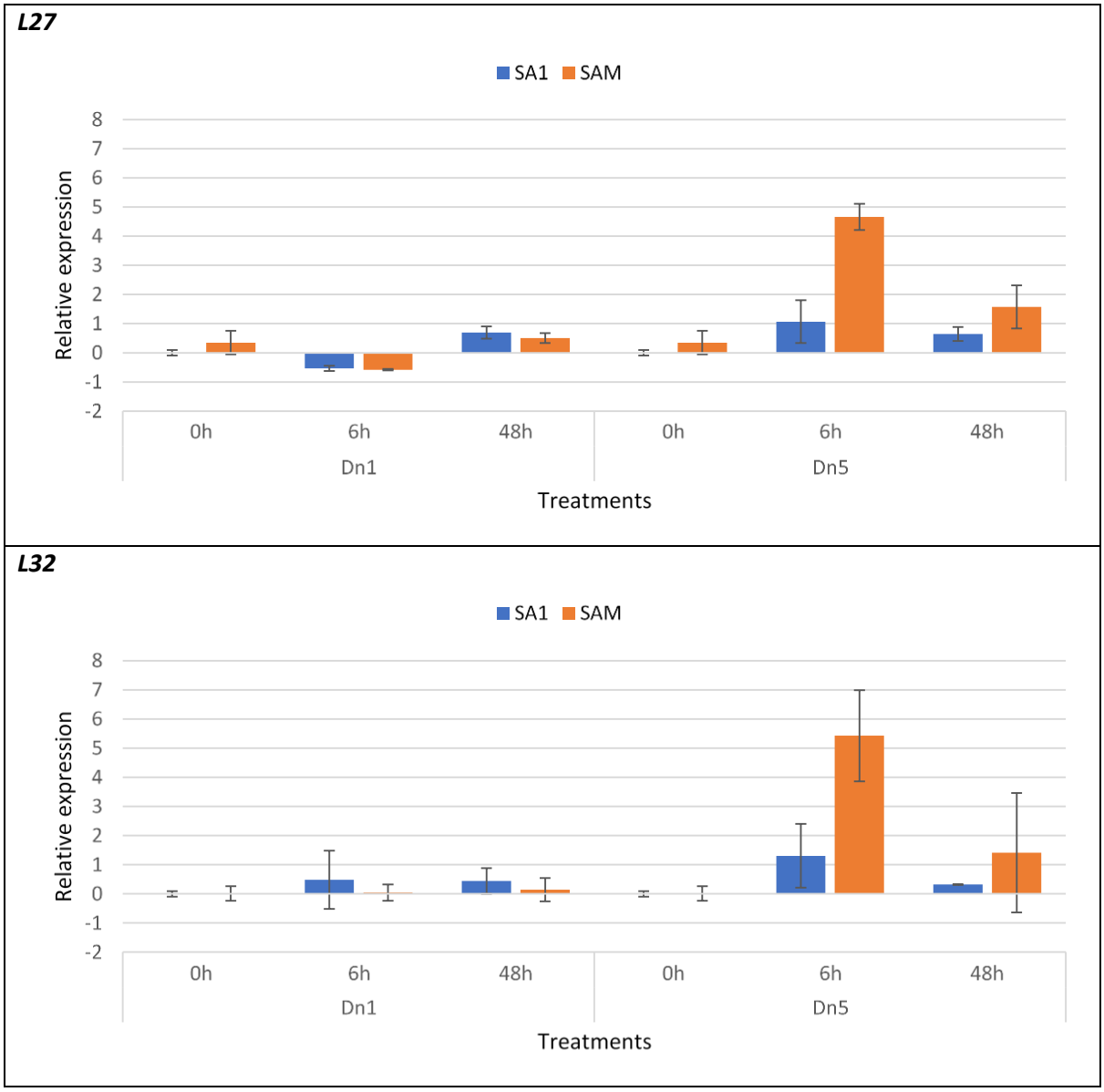

Figure 4.14: The expression of *DNMT3* in RWA biotypes SA1 and SAM, at 0, 6, and 48-hours after performing host-shifts from Tugela to Tugela *Dn1* and Tugela *Dn5*, relative to the expression of *L27* and *L32* of SA1 at 0-hours. The null hypothesis of no difference in mean expression relative to both *L27* and *L32* was rejected. Results of the ANOVA's (Table B10 and Table B12) as well as matrices showing all the statistically significant differences between treatments (Table B11 and Table B13), calculated using Fisher's LSD, can be found in appendix B.

Stellenbosch University https://scholar.sun.ac.za

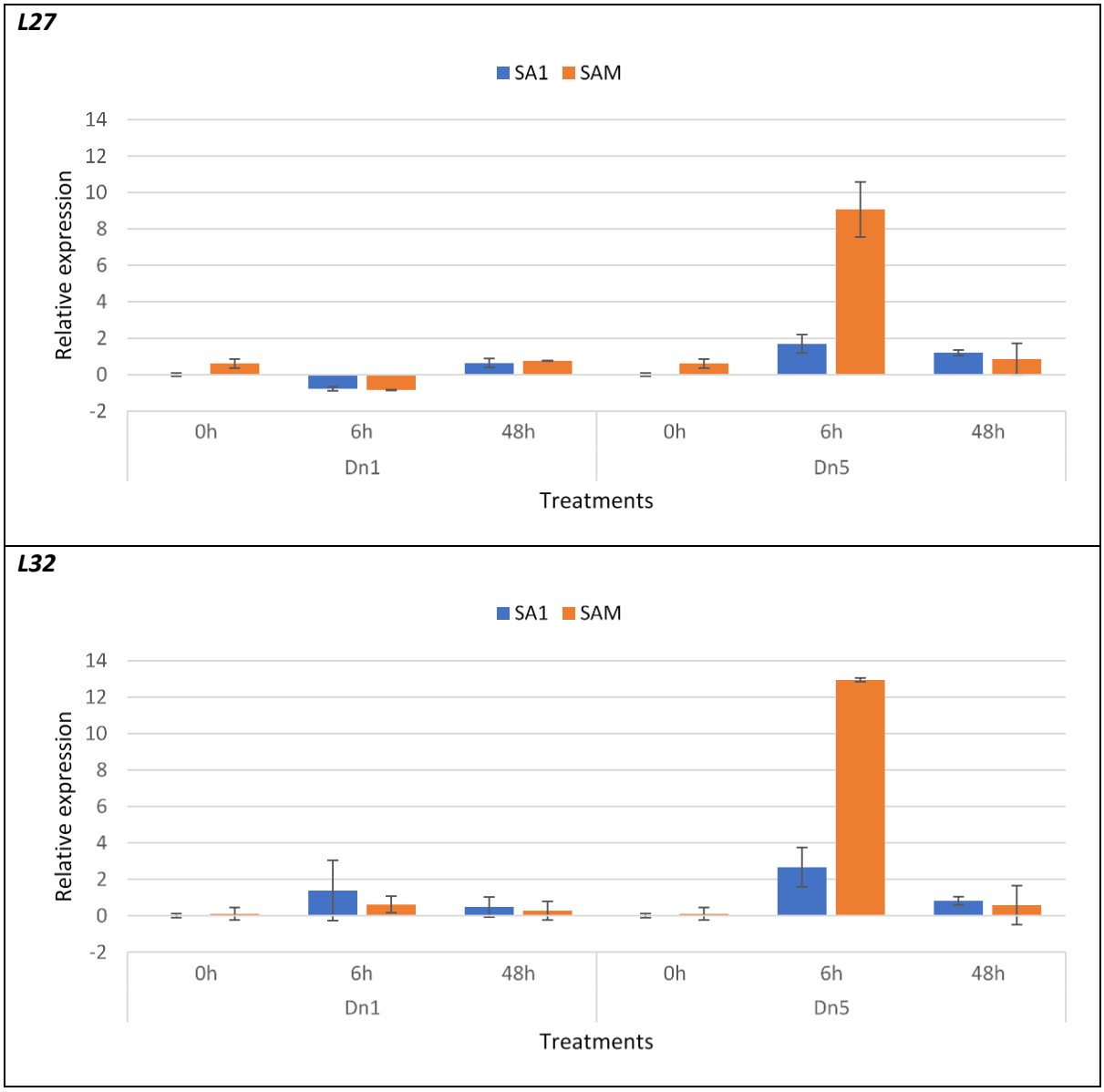

Figure 4.15: The expression of *TET* in RWA biotypes SA1 and SAM, at 0, 6, and 48-hours after performing host-shifts from Tugela to Tugela *Dn1* and Tugela *Dn5*, relative to the expression of *L27* and *L32* of SA1 at 0-hours. The null hypothesis of no difference in mean expression relative to both *L27* and *L32* was rejected. Results of the ANOVA's (Table B14 and Table B16) as well as matrices showing all the statistically significant differences between treatments (Table B15 and Table B17), calculated using Fisher's LSD, can be found in appendix B.

Stellenbosch University https://scholar.sun.ac.za

4.4: Discussion

4.4.1: Observed over expected methylation sites

The relationship between increased DNA methylation and the lower than expected occurrence of CpG sites than would be expected, given gene length and GC content, observed by Suzuki and Bird (2008), does seem to be evident in RWA. When comparing the $CpG_{O/E}$ ratios shown above to the bisulfite sequencing data, represented in Chapter 3, it is clear that regions with high levels of methylation, contain fewer CpG sites than expected, and *vice versa* (Figure 4.7 and Table 3.2 to Table 3.5). Intergenic regions had a high $CpG_{O/E}$ ratio, higher even than the control $GpC_{O/E}$ ratio, the average $CpG_{O/E}$ ratio for introns was lower, but still greater than 1.0, while the average $CpG_{O/E}$ ratio for exons were greatly diminished. This correlates to the overall methylation levels of these regions, as methylation in intergenic regions are the lowest, followed by introns, with exonic regions being methylated to a much greater extent. The number of CHG and CHH sites were not diminished, this might be because the same, error prone mechanism does not repair these mismatches, or because the levels of CpG and CHH methylation are simply too low to influence deamination.

There were significant differences between the $CpG_{O/E}$ ratio and $GpC_{O/E}$ ratio for all the regions investigated, as the ratio of the difference between the means and the overall visible spread for each was greater than 10%. Exonic regions however deviated the most from the expected, with a DBM/OVS ratio of 25.89%. While the DBM/OVS is higher for intergenic regions than for introns, the $CpG_{O/E}$ ratio, as mentioned, is higher than the $GpC_{O/E}$ ratio in intergenic regions. It can therefore be argued that even with a lower DBM/OVS ratio, the introns deviated from the expected more than the intergenic regions, as the occurrence of CpG sites were significantly increased in intergenic regions, rather than decreased. The DBM/OVS ratio was used to compare the $CpG_{O/E}$ and $GpC_{O/E}$ ratios because it intrinsically ignores outliers. With such an AT rich genome (70.5%), which contain many genes shorter than 1 kbp, many outliers arise when comparing the observed to expected occurrence of CpG sites. This can be seen in Figure 4.6, with some genes having a $CpG_{O/E}$ ratio as high as 3.0. Using the DBM/OVS eliminates these outliers without using an arbitrary cutoff based on gene length. Another advantage to the DBM/OVS ratio is that it is robust with large sample sizes. More than 35 000 genes, and a much larger number of exons and introns are being investigated, with such a large sample size, most statistical test would score a very small difference as significant, while the DBM/OVS ratio remains unbiased.

When comparing the levels of methylation to the $CpG_{O/E}$ ratio, on a gene by gene basis, the relationship does not hold. Only a negligible Pearson's correlation coefficient could be calculated between the two variables, and no significant difference could be found between the means of the

Stellenbosch University https://scholar.sun.ac.za

$CpG_{O/E}$ of hypermethylated and hypomethylated genes in biotype SAM. A bimodal distribution, as was reported for the pea aphid (International Aphid Genomics Consortium, 2006; Walsh et al., 2010) of $CpG_{O/E}$ was not observed for RWA. This could be due to the low levels of methylation and the fact that the CG content is already very low in the RWA, the levels of methylation and the GC content of the pea aphid are very comparable to that of RWA, however. A more plausible explanation is that the bin size of 0.1, was simply not a small enough increment to smooth out the curve. In the current study, a bin size of 0.005 was used, which smoothed out the distribution to a single, wide peak. This explanation is strengthened by the lack of statistical evidence of bimodality in the $CpG_{O/E}$ curve of the pea aphid (Walsh et al., 2010). For these reasons, it seems clear that the $CpG_{O/E}$ ratio alone is not sufficient to infer increased methylation in the RWA. Future studies into insect DNA methylation should rely on more direct methods of detecting methylation, such as bisulfite sequencing or MSAP, depending on budget and the research question.

4.4.2: Expression of differentially methylated and key methylation genes

While there were significant differences in the expression of some differentially methylated genes between time points, there was no differences in expression of these genes between biotypes. These results are not encouraging to the idea of DNA methylation being involved in the differentiation between the biotypes while stressed. It is however important to note, that only zero-hour methylation data was used to identify differentially methylated genes, and that the methylation state at 6h and 48h after the host shifts were performed are not known. Since DNA methylation is a temporal mechanism, employed to react to stress, the levels during stress are would probably be more relevant for use in future studies than the basal levels.

The method used to detect differentially methylated genes also relied on a “smoothing” algorithm which employs “pseudo-replication”. This algorithm is not recommended for reduced representation bisulfite sequencing (Wu et al., 2015). While whole genome bisulfite sequencing, and not reduced representation bisulfite sequencing was performed in this study, the low levels of DNA methylation, along with the AT richness of the genome results in data which closely imitates that of reduced representation bisulfite sequencing: High coverages of CpG sites appearing sporadically across the genome. Incorporating replicates for the whole genome bisulfite sequencing would allow for the use of more accurate scoring of differentially methylated genes. Replicate sequencing has been performed and will be included in future studies, the data was not available for this study however, as such the algorithm was the most reliable option available.

While no major differences were found in the expression of differentially methylated genes between biotypes, both *DNMT3* and *TET* are expressed at a much higher level by biotypes SAM than SA1 at six-

Stellenbosch University https://scholar.sun.ac.za

hours after host shifts, on Tugela *Dn5*. This implies a higher rate of both methylation and demethylation in biotype SAM in response to stress. This corroborates the findings of Burger et al. (2017) that biotype SAM possesses a greater ability respond in a timely manner to defence responses of host plants.

The difference in the expression of the methylation machinery between the two biotypes is a very encouraging result and provides evidence that DNA methylation might play a role in the difference in virulence between biotypes. The large number of differentially methylated genes, as high as 40 in categories such as “nitrogen compound metabolic process” and “cellular metabolic process”, also seem to indicate a difference in methylation strategy between the biotypes.

4.5: References

Aranda, P.S., LaJoie, D.M. & Jorcyk, C.L., 2012, Bleach gel: a simple agarose gel for analyzing RNA quality, *Electrophoresis*, 33(2), 366–369.

Bewick, A.J., Vogel, K.J., Moore, A.J. & Schmitz, R.J., 2017, Evolution of DNA methylation across insects, *Molecular Biology and Evolution*, 34(3), 654–665.

Botha, A-M., Li, Y. & Lapitan, N.L. V., 2005, Cereal host interactions with Russian wheat aphid: A review, *Journal of Plant Interactions*, 1(4), 211–222.

Breeds, K., Burger, N.F.V. & Botha, A-M., 2018, New insights into the methylation status of virulent *Diuraphis noxia* (Hemiptera: Aphididae) biotypes, *Journal of Economic Entomology*, 111(3), 1395–1403.

Burger, N.F. V, Venter, E. & Botha, A.-M., 2017, Profiling *Diuraphis noxia* (Hemiptera: Aphididae) Transcript Expression of the Biotypes SA1 and SAM Feeding on Various *Triticum aestivum* Varieties, *Journal of Economic Entomology*, 110(2), 692–701.

Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M. & Robles, M., 2005, Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research, *Bioinformatics*, 21(18), 3674–3676.

Darst, R.P., Pardo, C.E., Ai, L., Brown, K.D. & Kladde, M.P., 2010, Bisulfite sequencing of DNA, *Current Protocols in Molecular Biology*, 91(1), 7–9

Dombrovsky, A., Arthaud, L., Ledger, T.N., Tares, S. & Robichon, A., 2009, Profiling the repertoire of phenotypes influenced by environmental cues that occur during asexual reproduction, *Genome Research*, 19(11), 2052–2063.

du Toit, F., 1989, Inheritance of resistance in two *Triticum aestivum* lines to Russian wheat aphid (Hornoptera: Aphididae), *Journal of Economic Entomology*, 82(4), 1251–1253.

du Toit, F. & Walters, M.C., 1981, Damage assessment and economic threshold values for the chemical control of thr Russian wheat aphid, *Diuraphis noxia* (Mordvilko) on winter wheat, *Technical communication, Department of Agriculture, Republic of South Africa*, (191), 58–62.

Fouché, A., Verhoeven, R.L., Hewitt, P.H., Kriel, C.F., de Jager, J. & Walters, M.C., 1984, Russian aphid (*Diuraphis noxia*) feeding damage on wheat, related cereals and a *Bromus* grass species, *Technical communication, Department of Agriculture, Republic of South Africa*, (191), 22–33.

Stellenbosch University https://scholar.sun.ac.za

Goll, M.G. & Bestor, T.H., 2005, Eukaryotic cytosine methyltransferases., *Annual Review of Biochemistry*, 74, 481–514.

Havlicek, L.L. & Peterson, N.L., 1976, Robustness of the Pearson correlation against violations of assumptions, *Perceptual and Motor Skills*, 43(3_suppl), 1319–1334.

Hoagland, D.R. & Arnon, D.I., 1950, The water-culture method for growing plants without soil, *California Agricultural Experiment Station Circular*, 347(347), 1–32.

Hunt, B.G., Brisson, J.A., Yi, S. V. & Goodisman, M.A.D., 2010, Functional conservation of DNA methylation in the pea aphid and the honeybee, *Genome Biology and Evolution*, 2(1), 719–728.

International Aphid Genomics Consortium, 2006, Insights into social insects from the genome of the honeybee Apis mellifera, *Nature*, 443(7114), 931–949.

Jankielsohn, A., 2019, New Russian wheat aphid biotype found in Free State, *Grain SA Mini Focus, Pest Control on Winter Cereals*.

Kim, D., Minhas, B.F., Li-Byarlay, H. & Hansen, A.K., 2018, Key transport and ammonia recycling genes involved in aphid symbiosis respond to host-plant specialization, *Genes, Genomes, Genetics*, 8(7), 2433–2443.

Kocher, S.D., Li, C., Yang, W., Tan, H., Yi, S. V., Yang, X., Hoekstra, H.E., Zhang, G., Pierce, N.E. & Yu, D.W., 2015, Erratum to: The draft genome of a socially polymorphic halictid bee, *Lasioglossum albipes* [Genome Biol., 14, R142, (2013)], *Genome Biology*, 16(1).

Li, Y. & Tollefsbol, T.O., 2011, DNA methylation detection: Bisulfite genomic sequencing analysis, *Epigenetics Protocols*, pp. 11–21.

Mathers, T.C., Mugford, S.T., Percival-Alwyn, L., Chen, Y., Kaithakottil, G., Swarbreck, D., Hogenhout, S.A. & van Oosterhout, C., 2019, Sex-specific changes in the aphid DNA methylation landscape, *BioRxiv, 286302*.

McClelland, M., Nelson, M. & Raschke, E., 1994, Effect of site-specific modification on restriction endonucleases and DNA modification methyltransferases, *Nucleic Acids Research*, 22(17), 3640–3659.

Morgan, H.D., Dean, W., Coker, H.A., Reik, W. & Petersen-Mahrt, S.K., 2004, Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming, *The Journal of Biological Chemistry*, 279(50), 52353–52360.

Nakabachi, A., Shigenobu, S., Sakazume, N., Shiraki, T., Hayashizaki, Y., Carninci, P., Ishikawa, H., Kudo, T. & Fukatsu, T., 2005, Transcriptome analysis of the aphid bacteriocyte, the symbiotic host cell that harbors an endocellular mutualistic bacterium, Buchnera, *Proceedings of the national academy of sciences of the United States of America*, 102(15), 5477–5482.

Nkongolo, K.K., Quick, J.S., Limin, A.E. & Fowler, D.B., 1991, Sources and inheritance of resistance to Russian wheat aphid in *Triticum* Species *amphiploids* and *Triticum tauschii*, *Canadian Journal of Plant Science*, 71(3), 703–708.

Okano, M., Xie, S. & Li, E., 1998, Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases, *Nature Genetics*, 19(3), 219–220.

Pasquier, C., Clément, M., Dombrovsky, A., Penaud, S., Rocha, M. Da, Rancurel, C., Ledger, N., Capovilla, M. & Robichon, A., 2014, Environmentally selected aphid variants in clonality context display differential patterns of methylation in the genome, *PloS One*, 9(12), e115022.

Stellenbosch University https://scholar.sun.ac.za

Pegoraro, M., Bafna, A., Davies, N.J., Shuker, D.M. & Tauber, E., 2016, DNA methylation changes induced by long and short photoperiods in Nasonia, *Genome Research*, 26(2), 203–210.

Pfaffl, M.W., 2001, A new mathematical model for relative quantification in real-time RT–PCR, *Nucleic Acids Research*, 29(9), e45–e45.

Poole, A., Penny, D. & Sjoberg, B.M., 2001, *Confounded cytosine! Tinkering and the evolution of DNA*, *Nature Reviews. Molecular cell biology*, 2(2), 147–151.

Russo, V.E.A., Martienssen, R.A. & Riggs, A.D., 1996, *Epigenetic mechanisms of gene regulation*, Cold Spring Harbor Laboratory Press.

Shakesby, A., Wallace, I.S., Isaacs, H. V, Pritchard, J., Roberts, D.M. & Douglas, A.E., 2009, A water-specific aquaporin involved in aphid osmoregulation, *Insect biochemistry and molecular biology*, 39(1), 1–10.

Sinha, D.K. & Smith, C.M., 2014, Selection of reference genes for expression analysis in *Diuraphis noxia* (Hemiptera: Aphididae) fed on resistant and susceptible wheat plants, *Scientific reports*, 4, 5059.

Smith, M.C., Schotzko, D.J., Zemetra, R.S. & Souza, E.J., 1992, Categories of Resistance in Plant Introductions of Wheat Resistant to the Russian Wheat Aphid (Homoptera: Aphididae), *Journal of Economic Entomology*, 85(4), 1480–1484.

Suzuki, M.M. & Bird, A., 2008, DNA methylation landscapes: provocative insights from epigenomics, *Nature Reviews Genetics*, 9(6), 465.

Suzuki, M.M., Kerr, A.R.W., de Sousa, D. & Bird, A., 2007, CpG methylation is targeted to transcription units in an invertebrate genome, *Genome Research*, 625–631.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R. & Aravind, L., 2009, Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1, *Science*, 324(5929), 930–935.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M. & Rozen, S.G., 2012, Primer3 - new capabilities and interfaces, *Nucleic acids research*, 40(15), 1289–1291.

Wald, A., 1943, Tests of statistical hypotheses concerning several parameters when the number of observations is large, *Transactions of the American Mathematical society*, 54(3), 426–482.

Walsh, T.K., Brisson, J.A., Robertson, H.M., Gordon, K., Jaubert-Possamai, S., Tagu, D. & Edwards, O.R., 2010, A functional DNA methylation system in the pea aphid, *Acyrthosiphon pisum*, *Insect Molecular Biology*, 19, 215-228.

Walters, M., Penn, F., du Toit, F., Botha, T. & Aalbersberg, K., 1980, The Russian Wheat Aphid. Farming in South *Africa, Africa Leaflet series: Wheat G3:*, 1–6.

Weber, M., Davies, J.J., Wittig, D., Oakeley, E.J., Haase, M., Lam, W.L. & Schubeler, D., 2005, Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells, *Nature Genetics*, 37(8), 853–862.

Welch, B.L., 1947, The generalisation of students problems when several different population variances are involved., *Biometrika*, 34(1–2), 28–35.

Werren, J.H., Richards, S., Desjardins, C.A., Niehuis, O., Gadau, J. & Colbourne, J.K., 2010, Functional and evolutionary insights from the genomes of three parasitoid *Nasonia species*, *Science*, 327(5963), 343–348.

Stellenbosch University https://scholar.sun.ac.za

Wu, H., Xu, T., Feng, H., Chen, L., Li, B., Yao, B., Qin, Z., Jin, P. & Conneely, K.N., 2015, Detection of differentially methylated regions from whole-genome bisulfite sequencing data without replicates, *Nucleic Acids Research*, 43(21), e141.

Zhang, J., Xing, Y., Li, Y., Yin, C., Ge, C. & Li, F., 2015, DNA methyltransferases have an essential role in female fecundity in brown planthopper, *Nilaparvata lugens*, *Biochemical and Biophysical Research Communications*, 464(1), 83–88.

van Zyl, R.A. & Botha, A-M., 2008, Eliciting proteins from *Diuraphis noxia* biotypes differ in size and composition, *18th Biennial International Plant Resistance to Insects Workshop*, Fort Collins, Colorado.

Stellenbosch University https://scholar.sun.ac.za

Appendix B

Table B1: The output data from DSS-single on 148 genes that are differentially methylated (p-value < 0.05) between biotypes SA1 and SAM

Gene	Contig	Start	End	Length	#CG	Methylation (SA1)	Methylation (SAM)	Differential methylation	Area statistic
10752	420	123833	125912	2080	80	0.31	0.01	0.29	281.54
6742	173	232313	236569	4257	54	0.39	0.11	0.28	167.04
14528	973	10166	10622	457	74	0.66	0.42	0.24	156.51
9140	295	157605	159573	1969	48	0.26	0.04	0.22	137.77
13527	767	34911	35977	1067	49	0.29	0.06	0.23	129.97
19223	7605	2583	3154	572	47	0.14	0.43	-0.29	-125.40
2940	48	325414	326521	1108	33	0.66	0.34	0.32	97.49
270	4	612254	616440	4187	34	0.24	0.03	0.21	90.45
5941	140	306752	307789	1038	25	0.43	0.12	0.31	77.09
3367	58	93449	93947	499	19	0.53	0.13	0.40	69.06
8970	287	43970	44804	835	28	0.36	0.12	0.23	65.88
2232	32	275444	275759	316	16	0.35	0.05	0.29	56.28
9610	328	163100	163214	115	28	0.43	0.19	0.24	55.45
7383	201	223823	224210	388	17	0.16	0.01	0.15	44.65
10413	395	70896	71202	307	12	0.05	0.39	-0.34	-42.33
21017	12609	51	806	756	17	0.01	0.14	-0.13	-39.52
6235	152	94817	95219	403	11	0.63	0.28	0.35	35.70
12334	581	88187	88392	206	8	0.80	0.31	0.49	33.27
743	10	60694	60987	294	12	0.09	0.40	-0.30	-33.23
17788	4349	508	664	157	8	0.15	0.61	-0.46	-30.42
27086	38114	687	865	179	15	0.13	0.39	-0.26	-30.34
9894	353	110873	111088	216	6	0.49	0.95	-0.46	-29.65
33751	105594	47	484	438	14	0.49	0.77	-0.28	-29.51
13557	773	48430	48742	313	9	0.00	0.23	-0.23	-27.31
9305	305	141221	141551	331	10	0.23	0.04	0.19	23.71

Stellenbosch University https://scholar.sun.ac.za

1867	26	376349	376423	75	6	0.24	0.69	-0.45	-23.63
8970	287	45112	45429	318	11	0.26	0.09	0.17	23.12
9823	347	19603	19739	137	10	0.43	0.17	0.26	22.59
10079	370	67751	67853	103	11	0.48	0.24	0.24	21.93
23212	20550	336	424	89	4	0.78	0.17	0.61	21.79
1411	20	378877	379192	316	9	0.26	0.04	0.22	21.49
18466	5751	942	1332	391	4	0.55	0.92	-0.37	-21.02
8689	270	117657	117849	193	6	0.60	0.24	0.37	20.56
10247	385	54380	54688	309	8	0.49	0.77	-0.28	-19.84
8510	259	184970	185132	163	7	0.55	0.18	0.37	19.47
21431	14099	73	454	382	7	0.76	0.98	-0.22	-19.34
21818	15443	677	934	258	8	0.04	0.21	-0.17	-19.24
5639	128	332573	332769	197	5	0.29	0.00	0.29	18.78
9140	295	153500	153680	181	8	0.14	0.02	0.12	18.76
8656	267	168514	168808	295	8	0.18	0.03	0.15	17.70
6391	158	149770	150036	267	8	0.05	0.23	-0.19	-17.55
10779	423	38992	39257	266	8	0.31	0.13	0.18	17.37
12262	572	94413	95017	605	4	0.45	0.10	0.35	17.08
8078	234	111983	112042	60	6	0.21	0.00	0.21	16.87
8970	287	46029	46212	184	7	0.24	0.07	0.17	16.83
16430	2054	1082	1341	260	6	0.55	0.25	0.30	16.51
6036	144	77373	77706	334	6	0.65	0.92	-0.27	-16.41
10769	422	83691	83794	104	6	0.01	0.24	-0.22	-15.66
4243	81	328322	328880	559	4	0.88	0.60	0.28	15.63
3382	58	250722	250893	172	7	0.42	0.16	0.26	15.61
17545	3932	3104	3303	200	5	0.41	0.11	0.31	15.57
8813	278	52882	53558	677	4	0.52	0.14	0.38	15.53
1592	22	617678	617740	63	5	0.52	0.18	0.34	15.27
10050	368	10831	10883	53	5	0.93	0.64	0.29	15.27
664	9	366317	366424	108	4	0.20	0.59	-0.39	-14.99

Stellenbosch University https://scholar.sun.ac.za

16406	2025	1484	1632	149	7	0.22	0.48	-0.26	-14.86
6466	161	130363	130443	81	5	0.26	0.00	0.26	14.55
12951	670	38722	38890	169	4	0.86	0.46	0.40	14.43
17453	3762	303	561	259	5	0.88	0.58	0.30	14.32
8503	259	141980	142349	370	6	0.03	0.18	-0.14	-14.23
5941	140	307895	307953	59	7	0.73	0.47	0.26	14.06
13766	807	7755	7843	89	6	0.35	0.13	0.22	13.87
19636	8637	401	626	226	4	0.35	0.08	0.27	13.84
1503	21	694833	695161	329	6	0.43	0.18	0.24	13.80
5214	112	277415	277720	306	4	0.74	0.98	-0.23	-13.38
15737	1374	9508	10148	641	9	0.27	0.47	-0.20	-13.28
8064	234	17222	17484	263	6	0.57	0.84	-0.27	-13.16
221	4	46381	46735	355	5	0.21	0.03	0.18	13.16
10524	403	34551	34692	142	6	0.17	0.00	0.17	12.99
12103	554	60168	62663	2496	4	0.70	0.93	-0.23	-12.93
4946	102	348312	348448	137	6	0.16	0.37	-0.21	-12.84
23153	20320	1326	1548	223	4	0.68	0.95	-0.28	-12.83
3943	73	93989	94114	126	6	0.21	0.05	0.17	12.81
12945	670	4945	5136	192	5	0.85	0.59	0.26	12.77
9404	311	176462	176649	188	5	0.30	0.07	0.23	12.57
506	7	311233	311327	95	4	0.01	0.24	-0.22	-12.45
17111	3105	3351	3881	531	5	0.93	0.75	0.18	12.37
14546	978	8674	8968	295	4	0.54	0.19	0.36	12.28
11282	468	68786	68845	60	6	0.48	0.27	0.21	12.23
3327	57	186403	186536	134	5	0.73	0.39	0.34	12.23
1366	19	697427	697596	170	5	0.74	0.94	-0.20	-12.15
14828	1045	34363	34462	100	6	0.13	0.00	0.13	12.11
18743	6427	2339	2455	117	5	0.67	0.37	0.30	12.04
15015	1092	12437	12533	97	6	0.50	0.20	0.31	11.97
27513	40684	744	972	229	5	0.96	0.80	0.16	11.91

Stellenbosch University https://scholar.sun.ac.za

4551	90	185742	186066	325	4	0.98	0.83	0.15	11.91
10898	433	124719	124806	88	4	0.48	0.16	0.32	11.77
665	9	372983	373122	140	5	0.65	0.90	-0.25	-11.61
16488	2128	5161	5222	62	4	0.92	0.67	0.25	11.43
5930	140	105571	105708	138	4	0.00	0.17	-0.17	-11.36
6136	148	280785	280914	130	4	0.16	0.47	-0.31	-11.32
8342	251	22826	22950	125	5	0.72	0.47	0.25	11.17
9824	347	24061	24284	224	5	0.31	0.58	-0.27	-11.08
5941	140	308204	308374	171	5	0.79	0.56	0.23	11.05
16397	2011	1507	1636	130	4	0.03	0.22	-0.19	-11.03
11834	521	55708	55825	118	4	0.95	0.72	0.23	10.99
2359	34	425949	426037	89	5	0.32	0.13	0.19	10.95
12025	546	5953	6173	221	4	0.91	0.64	0.27	10.94
1642	23	330079	330335	257	4	0.00	0.14	-0.14	-10.87
25019	28129	56	376	321	4	0.87	0.65	0.22	10.80
11478	489	57335	57983	649	4	0.87	0.66	0.21	10.80
7236	194	56761	56847	87	5	0.00	0.23	-0.23	-10.76
23492	21668	460	674	215	4	0.66	0.93	-0.26	-10.64
13425	752	24781	25011	231	5	0.41	0.67	-0.26	-10.56
19745	8911	2213	2436	224	4	0.58	0.95	-0.37	-10.19
30297	59603	713	917	205	4	0.73	0.92	-0.19	-10.11
810	10	868504	868564	61	4	0.97	0.79	0.18	9.99
4587	91	172502	172752	251	4	0.41	0.72	-0.31	-9.94
15070	1109	21488	22142	655	4	0.67	0.86	-0.19	-9.92
11943	534	100244	100514	271	4	0.45	0.25	0.20	9.88
13858	823	54623	55376	754	4	0.25	0.52	-0.27	-9.84
16050	1600	2007	2062	56	4	0.22	0.00	0.22	9.77
6769	174	95134	95210	77	5	0.13	0.00	0.13	9.68
6772	174	123776	123869	94	4	0.33	0.04	0.29	9.64
10862	429	86099	86292	194	4	0.36	0.12	0.24	9.55

Stellenbosch University https://scholar.sun.ac.za

4342	84	39604	39665	62	4	0.80	0.40	0.40	9.52
3655	64	386256	387120	865	5	0.74	0.87	-0.13	-9.49
973	13	680956	681100	145	4	0.94	0.74	0.20	9.47
13781	810	57484	57556	73	4	0.61	0.85	-0.25	-9.44
12641	624	31361	31514	154	4	0.88	0.66	0.22	9.25
2654	42	210575	210757	183	4	0.13	0.00	0.13	9.18
11439	486	29265	29775	511	4	0.75	0.89	-0.14	-9.17
4661	93	118268	118414	147	4	0.82	0.60	0.22	9.09
20661	11588	2046	2275	230	4	0.20	0.04	0.16	8.99
23876	23160	1563	1737	175	4	0.99	0.86	0.13	8.92
1007	14	319115	319169	55	4	0.21	0.46	-0.25	-8.81
16419	2040	3682	6744	3063	4	0.22	0.35	-0.12	-8.79
8664	268	53466	53624	159	4	0.74	0.93	-0.19	-8.65
3369	58	104870	105037	168	4	0.97	0.83	0.14	8.62
7535	208	93187	93427	241	4	0.18	0.01	0.16	8.50
6517	163	118206	118350	145	4	0.63	0.38	0.25	8.48
12148	559	20547	22467	1921	4	0.76	0.57	0.19	8.38
15998	1546	4091	4333	243	4	0.20	0.42	-0.22	-8.37
14653	1002	40895	40993	99	4	0.00	0.09	-0.09	-8.34
12119	556	32859	32912	54	4	0.83	0.60	0.22	8.18
6974	182	62917	62992	76	4	0.09	0.32	-0.23	-8.18
1299	18	458448	458543	96	4	0.13	0.36	-0.23	-7.87
493	7	157841	159362	1522	6	0.76	0.84	-0.08	-7.29
14020	858	45592	46249	658	6	0.05	0.16	-0.11	-7.15
12318	580	46262	49153	2892	4	0.67	0.80	-0.14	-5.97
4406	86	63447	63748	302	4	0.68	0.80	-0.13	-5.44
2983	49	206798	210772	3975	9	0.16	0.22	-0.06	-4.04
4810	98	146703	148206	1504	4	0.77	0.79	-0.03	-3.98
28349	45571	469	1319	851	10	0.85	0.83	0.02	2.90
10210	381	106172	107597	1426	4	0.81	0.75	0.07	2.41

Stellenbosch University https://scholar.sun.ac.za

16226	1812	3754	6279	2526	4	0.88	0.86	0.02	-0.17
7302	197	240033	240690	658	4	0.08	0.07	0.01	0.11
7544	208	213464	214205	742	4	0.89	0.91	-0.02	-0.09

Table B2: The Blast2GO matches for 148 genes found to be differentially methylated (p-value < 0.05) between biotypes SA1 and SAM, using DSS-single

Gene	Description	Length	#Hits	e-Value
10752	Basement membrane-specific heparan sulfate proteoglycan core	3532	20	0
6742	low-density lipo receptor, [Pediculus humanus corporis]	1738	20	0
14528	gastrula zinc finger -like isoform X1 [Diaphorina citri]	290	20	1.8E-171
9140	split ends isoform X3	2083	20	0
13527	plasma membrane calcium-transporting ATPase 2 isoform X1 [Bombus terrestris]	1208	20	0
19223	F-BAR domain only 2 isoform X1	308	20	0
2940	low-density lipo receptor-related 2	4541	20	0
270	autophagy-related 13 homolog isoform X1	461	20	0
5941	epidermal growth factor receptor kinase substrate 8-like isoform X1 [Apis florea]	725	20	0
3367	PREDICTED: uncharacterized protein LOC100165836	420	20	0
8970	microtubule-actin cross-linking factor 1 isoform X12	5597	20	0
2232	cubilin-like	2999	20	0
9610	ATP-binding cassette sub-family G member 1-like [Orussus abietinus]	622	20	0
7383	---NA---	237	0	0
10413	probable RNA-directed DNA polymerase from transposon BS	1050	20	0
21017	cytochrome b561-like [Acyrthosiphon pisum]	92	20	1.49E-49
6235	solute carrier family 22 member 21-like isoform X1	877	20	2.5E-153
12334	cyclin-L1 isoform X1	352	20	0
743	general transcription factor II-I repeat domain-containing 2-like	320	20	1.41E-51
17788	nuclear receptor coactivator 2 isoform X1 [Microplitis demolitor]	762	20	0
27086	EF-hand domain-containing 1-like	193	20	2.7E-113
9894	ATP-binding cassette sub-family D member 3	609	20	0
33751	PREDICTED: trichohyalin-like	77	1	8.72E-06
13557	DNA-directed RNA polymerase, mitochondrial	1394	20	0

Stellenbosch University https://scholar.sun.ac.za

9305	CCR4-NOT transcription complex subunit 2 isoform X1 [Microplitis demolitor]	573	20	0
1867	PREDICTED: uncharacterized protein LOC100161026 isoform X1	598	20	0
8970	microtubule-actin cross-linking factor 1 isoform X12	5597	20	0
9823	mitogen-activated kinase kinase kinase 15 isoform X2	1345	20	0
10079	dedicator of cytokinesis 3 isoform X1	1679	20	0
23212	serine-rich adhesin for platelets-like isoform X1	128	14	1.07E-58
1411	Zinc finger BED domain-containing 4, partial	953	20	1.05E-92
18466	downstream neighbor of son homolog	507	20	0
8689	3-hydroxy-3-methylglutaryl-coenzyme A reductase	1101	20	0
10247	lisH domain-containing -like	362	1	8.96E-56
8510	PREDICTED: uncharacterized protein LOC105253459	602	20	8.05E-72
21431	uncharacterized membrane DDB_G0293934-like [Acyrthosiphon pisum]	213	1	5.6E-108
21818	AAEL003102-PA [Aedes aegypti]	203	20	1.8E-129
5639	zinc finger MYM-type 1-like	586	20	0
9140	split ends isoform X3	2083	20	0
8656	RNA-directed DNA polymerase from mobile element jockey-like	300	20	9.04E-32
6391	ester hydrolase C11orf54 homolog	327	20	2.2E-129
10779	phosphatase 1 regulatory subunit 12B-like, partial	429	20	7.71E-95
12262	AC9 transposase, partial	412	20	1.8E-134
8078	chorion peroxidase [Culex quinquefasciatus]	953	20	0
8970	microtubule-actin cross-linking factor 1 isoform X12	5597	20	0
16430	AF4 FMR2 family member 4 isoform X1 [Halyomorpha halys]	893	20	0
6036	puff-specific Bx42	600	20	1.9E-149
10769	beta- isoform X1	528	20	0
4243	exocyst complex component 2	851	20	0
3382	nucleoporin p54	567	20	5.55E-58
17545	PREDICTED: uncharacterized protein LOC103308576	434	20	5.2E-101
8813	transposase [Carassius auratus]	952	20	1.6E-143
1592	PREDICTED: uncharacterized protein LOC100571031	874	20	3.5E-153
10050	cytochrome P450 307a1	822	20	0

Stellenbosch University https://scholar.sun.ac.za

664	glutamyl aminopeptidase	675	20	6.3E-102
16406	abnormal spindle	1320	20	0
6466	plexin domain-containing 2	523	20	7E-118
12951	galactosylceramide sulfotransferase-like	280	20	1.49E-63
17453	thyroid adenoma-associated homolog	722	20	0
8503	Transposable element Tc3 transposase, partial	198	20	7.9E-41
5941	epidermal growth factor receptor kinase substrate 8-like isoform X1 [Apis florea]	725	20	0
13766	endonuclease-reverse transcriptase	518	20	8.31E-69
19636	zinc finger 862	713	20	0
1503	---NA---	235	0	0
5214	phosphatase 1 regulatory subunit 16A isoform X1	509	20	0
15737	reversion-inducing cysteine-rich , partial	486	20	0
8064	lissencephaly-1 homolog	411	20	0
221	tigger transposable element-derived 4-like	574	20	1E-114
10524	PREDICTED: uncharacterized protein LOC103309687	313	20	4E-101
12103	tumor susceptibility gene 101	680	20	0
4946	nuclear receptor coactivator 3 isoform X1	1494	20	0
23153	kelch domain-containing 10 homolog	170	20	3.8E-111
3943	amyloid beta A4 precursor -binding family B member 1-like isoform X1 [Cimex lectularius]	739	20	0
12945	histone-lysine N-methyltransferase 2C-like isoform X2	3002	20	0
9404	RNA-binding fusilli isoform X1	844	20	0
506	dystrophin, isoform D-like isoform X1	1239	20	0
17111	golgi-specific brefeldin A-resistance guanine nucleotide exchange factor 1	529	20	0
14546	kinesin-related 4-like isoform X1	540	20	0
11282	PREDICTED: uncharacterized protein CG43867 isoform X1 [Acyrthosiphon pisum]	1313	20	0
3327	zinc finger MYM-type 1-like	586	20	0
1366	peptide chain release factor 1-like, mitochondrial isoform X1	401	20	0
14828	lysosomal acid phosphatase-like	111	20	1.25E-27
18743	aubergine-like [Acyrthosiphon pisum]	437	20	0
15015	AAEL008862-PA [Aedes aegypti]	1030	20	0

Stellenbosch University https://scholar.sun.ac.za

27513	sister chromatid cohesion DCC1	62	20	3.05E-29
4551	cylicin-1-like [Acyrthosiphon pisum]	1462	20	0
10898	chorion peroxidase-like	123	20	1.55E-26
665	acyl- dehydrogenase family member 9, mitochondrial	566	20	0
16488	multidrug resistance-associated 7	962	20	0
5930	twitchin isoform X3	8759	20	0
6136	small integral membrane 20	391	20	0
8342	PREDICTED: uncharacterized protein LOC103311737, partial	413	20	4.44E-79
9824	micronuclear linker histone poly -like	702	4	0
5941	epidermal growth factor receptor kinase substrate 8-like isoform X1 [Apis florea]	725	20	0
16397	inositol 1,4,5-trisphosphate receptor isoform X1	1329	20	0
11834	bromodomain adjacent to zinc finger domain 1A isoform X2	1316	20	0
2359	endonuclease-reverse transcriptase	235	20	6.08E-42
12025	aurora kinase A	426	20	0
1642	probable isoaspartyl peptidase L-asparaginase CG7860 isoform X4 [Cimex lectularius]	1154	20	0
25019	zinc finger HIT domain-containing 2	160	20	8.5E-99
11478	chondroitin sulfate proteoglycan 4 [Polistes canadensis]	2583	20	0
7236	nuclear factor erythroid 2-related factor 1 isoform X1	271	20	2.4E-129
23492	acetyl- acetyltransferase, mitochondrial	236	20	2.7E-163
13425	telomerase Cajal body 1	385	20	0
19745	39S ribosomal L39, mitochondrial	245	20	2.6E-172
30297	Mediator of RNA polymerase II transcription subunit 22	106	20	1.32E-68
810	ACYPI000079 [Acyrthosiphon pisum]	358	20	0
4587	Zinc finger BED domain-containing 4	196	5	2.4E-15
15070	Histone-lysine N-methyltransferase trithorax	2171	20	0
11943	dna-mediated transposase	705	20	1.11E-91
13858	myrosinase 1-like	496	20	0
16050	Down syndrome cell adhesion molecule	1187	20	0
6769	ATP synthase subunit b, mitochondrial	425	20	3.24E-71
6772	zinc finger MYM-type 1-like	641	20	0

Stellenbosch University https://scholar.sun.ac.za

10862	major facilitator superfamily domain-containing 9-like	527	20	0
4342	tuberin isoform X1	1830	20	0
3655	DNA replication factor Cdt1	738	20	0
973	sickie-like isoform X4	904	20	0
13781	CCR4-NOT transcription complex subunit 3 isoform X1 [Cimex lectularius]	703	20	0
12641	PAX-interacting 1-like	1586	20	0
2654	[Nematostella vectensis]	411	20	1.24E-97
11439	phosphatase 1L	456	20	0
4661	UPF0568 C14orf166 homolog	249	20	6.8E-172
20661	PREDICTED: uncharacterized protein LOC103308903	389	20	6.72E-68
23876	methyl- -binding domain 4, partial	126	20	5.5E-71
1007	PREDICTED: LOW QUALITY PROTEIN: uncharacterized protein LOC100160261	1985	20	0
16419	carboxypeptidase D-like	278	20	5.1E-170
8664	Fanconi anemia group J	85	20	5.85E-42
3369	longitudinals lacking , isoforms H M V-like [Papilio polytes]	413	20	0
7535	zinc finger MYM-type 1-like	396	20	1.3E-105
6517	guanylate cyclase 32E-like isoform X1 [Bombus terrestris]	1371	20	0
12148	PREDICTED: LOW QUALITY PROTEIN: uncharacterized protein KIAA0195	1305	20	0
15998	zinc finger BED domain-containing 5-like	276	20	2.6E-106
14653	DNA-binding RFX2-like	556	20	0
12119	ubiquitin carboxyl-terminal hydrolase 4-like isoform X2	853	20	0
6974	THAP domain-containing	778	20	3.04E-97
1299	sel-1 homolog 1 [Megachile rotundata]	767	20	0
493	pelota	394	20	0
14020	MATH and LRR domain-containing PFE0570w-like	509	20	3.28E-72
12318	tyrosine- kinase hopscotch	1108	20	0
4406	DNA ligase 1 isoform X1 [Halyomorpha halys]	741	20	0
2983	PREDICTED: uncharacterized protein LOC105679549 isoform X2	262	7	9.25E-09
4810	PQ-loop repeat-containing 1 isoform X5 [Halyomorpha halys]	297	20	1.5E-146
28349	box C D snoRNA 1	149	20	3.11E-98

Stellenbosch University https://scholar.sun.ac.za

10210	neuropathy target esterase sws isoform X3	1164	20	0
16226	myosin heavy chain 95F isoform X2	1234	20	0
7302	polyamine-modulated factor 1-binding 1-like	77	1	1.45E-13
7544	ankyrin repeat domain-containing 12-like isoform X3	666	20	0

Stellenbosch University https://scholar.sun.ac.za

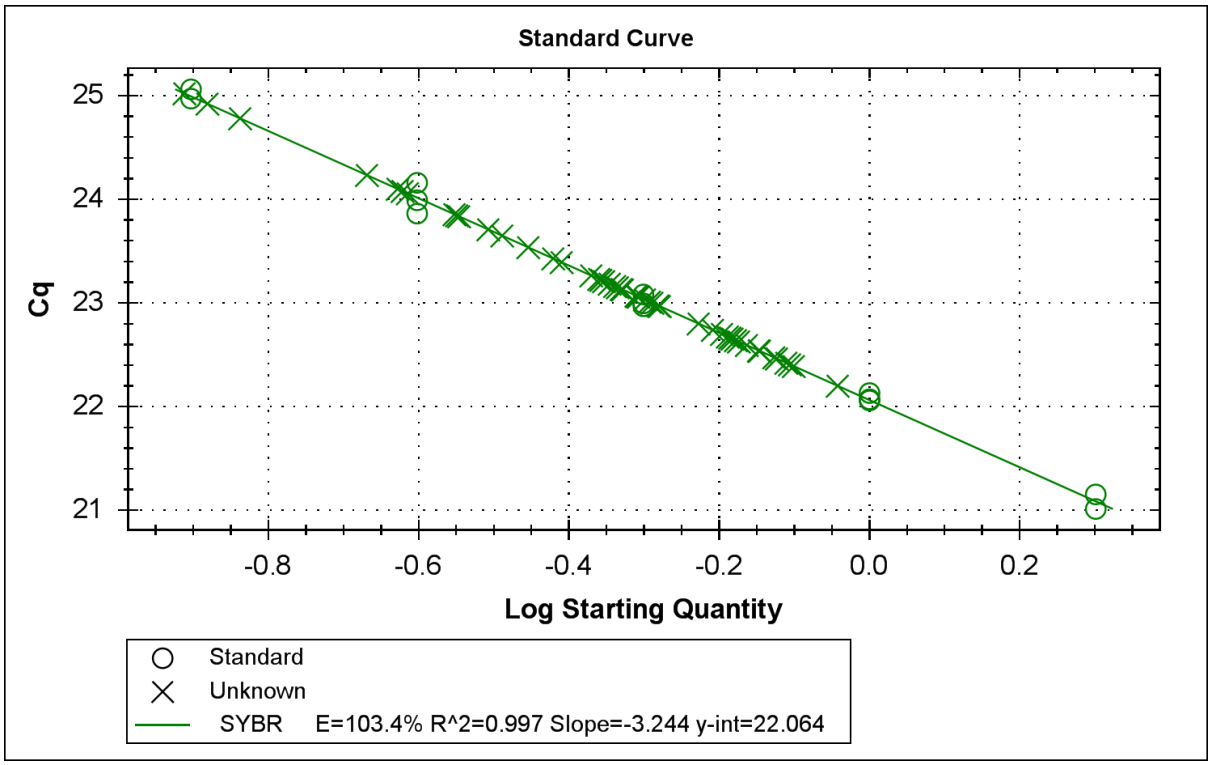

Figure B1: Standard curve of DMG1, for 0h and 48h samples of biotypes SA1 and SAM

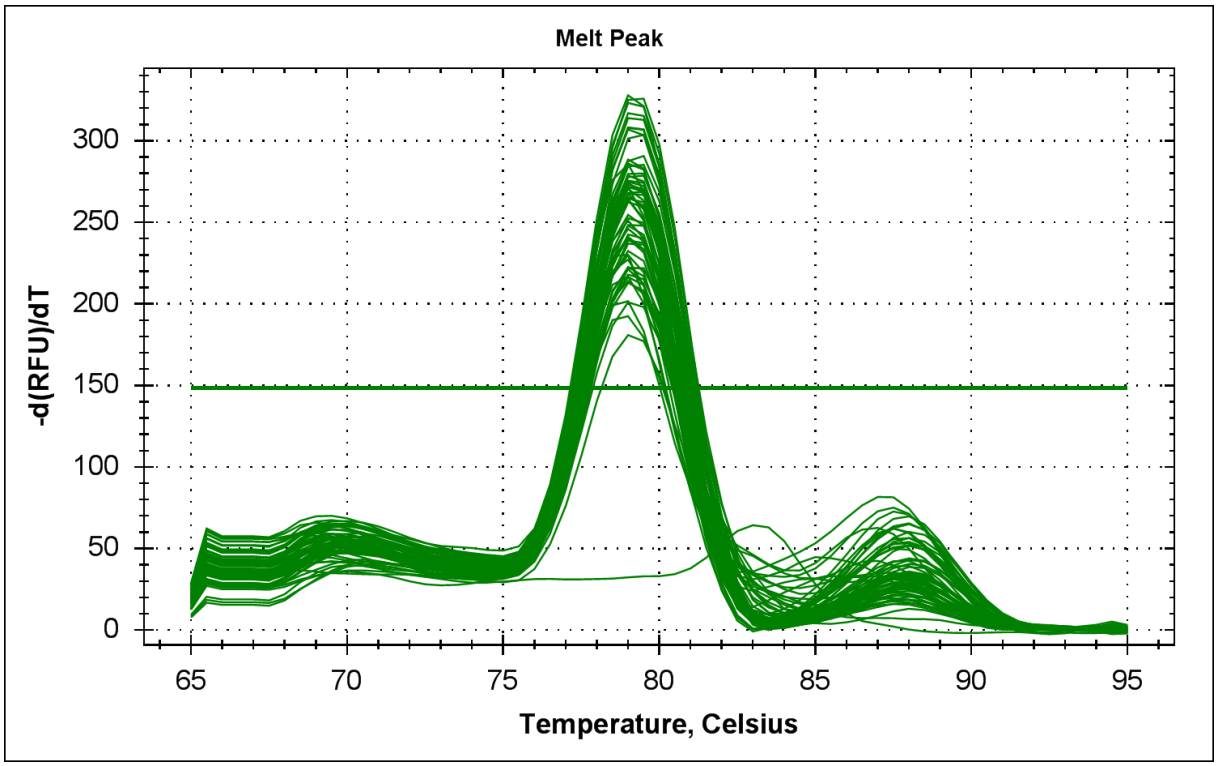

Figure B2: Melt peaks of DMG1, for 0h and 48h samples of biotypes SA1 and SAM

Stellenbosch University https://scholar.sun.ac.za

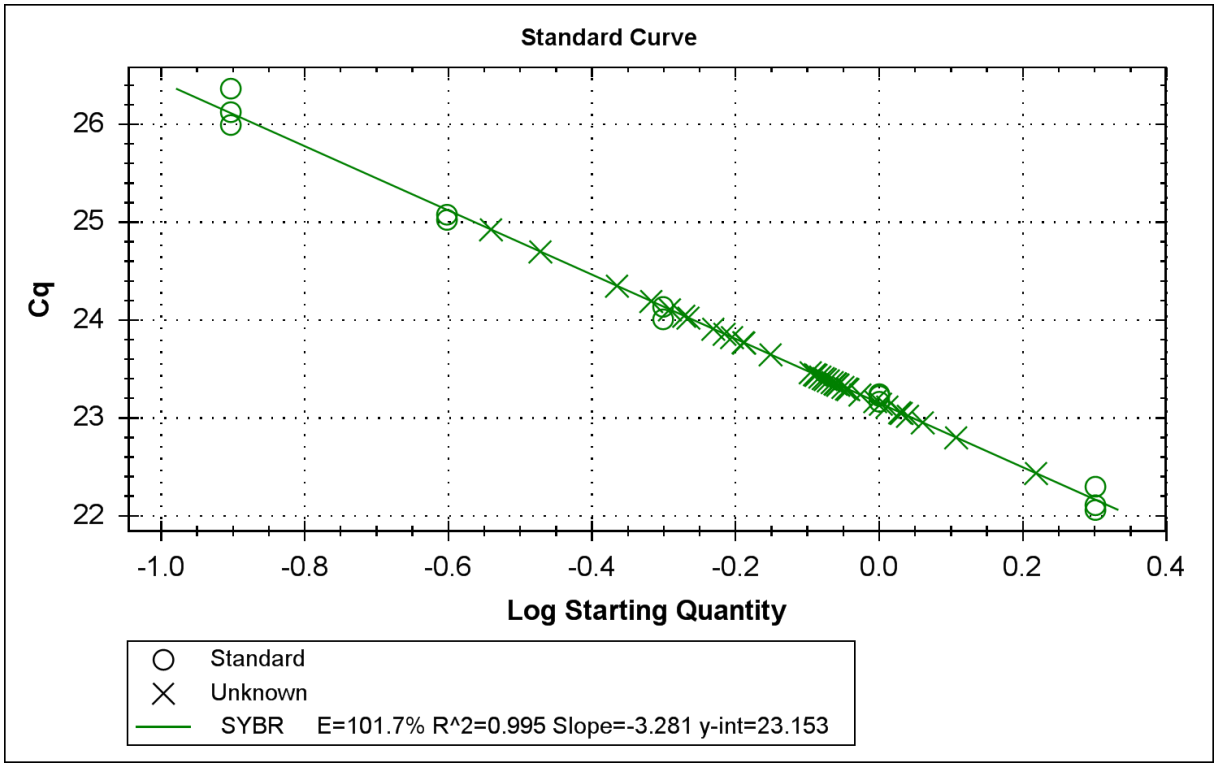

Figure B3: Standard curve of DMG1, for 6h samples of biotypes SA1 and SAM

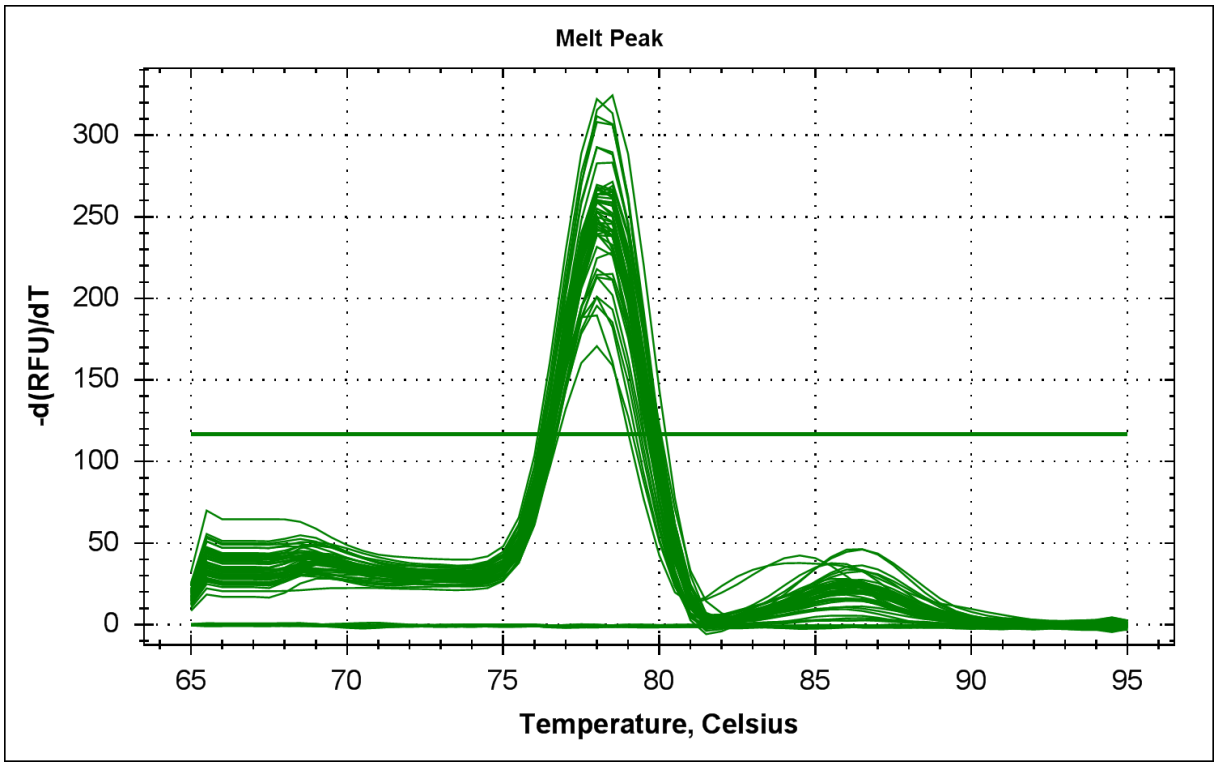

Figure B4: Melt peaks of DMG1, for 6h samples of biotypes SA1 and SAM

Stellenbosch University https://scholar.sun.ac.za

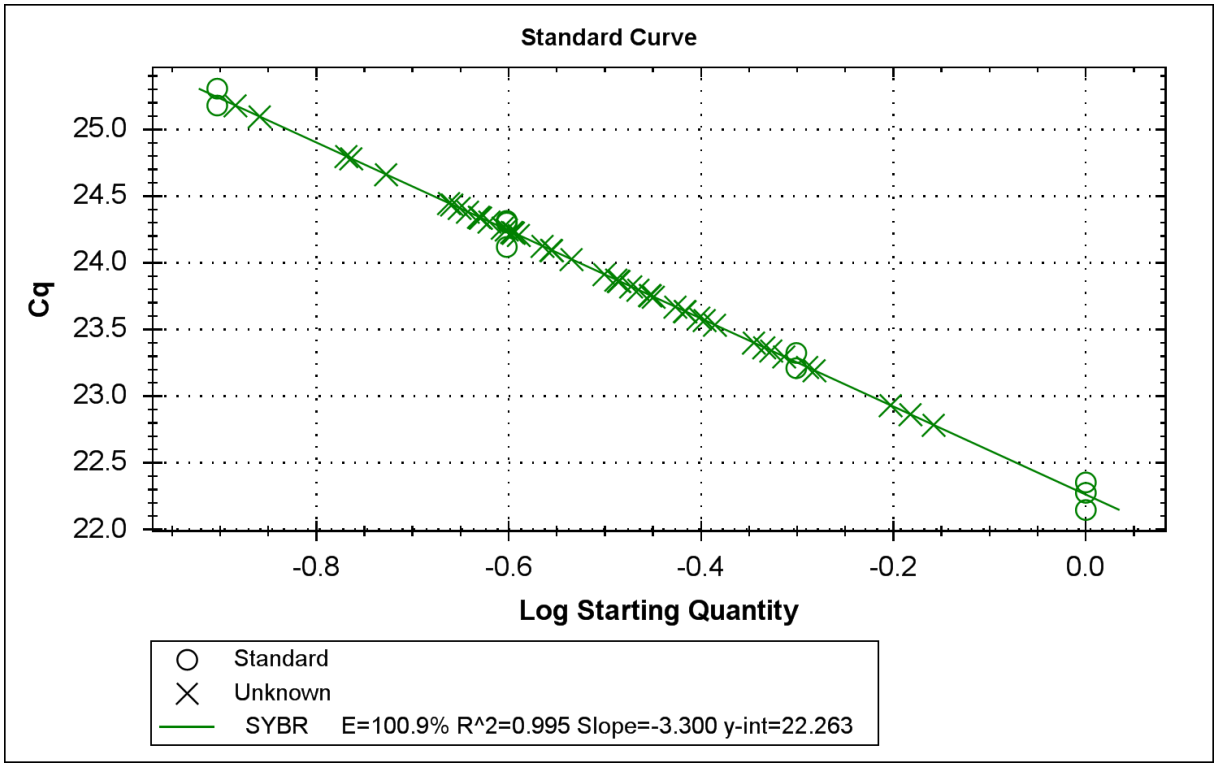

Figure B5: Standard curve of DMG2, for 0h and 48h samples of biotypes SA1 and SAM

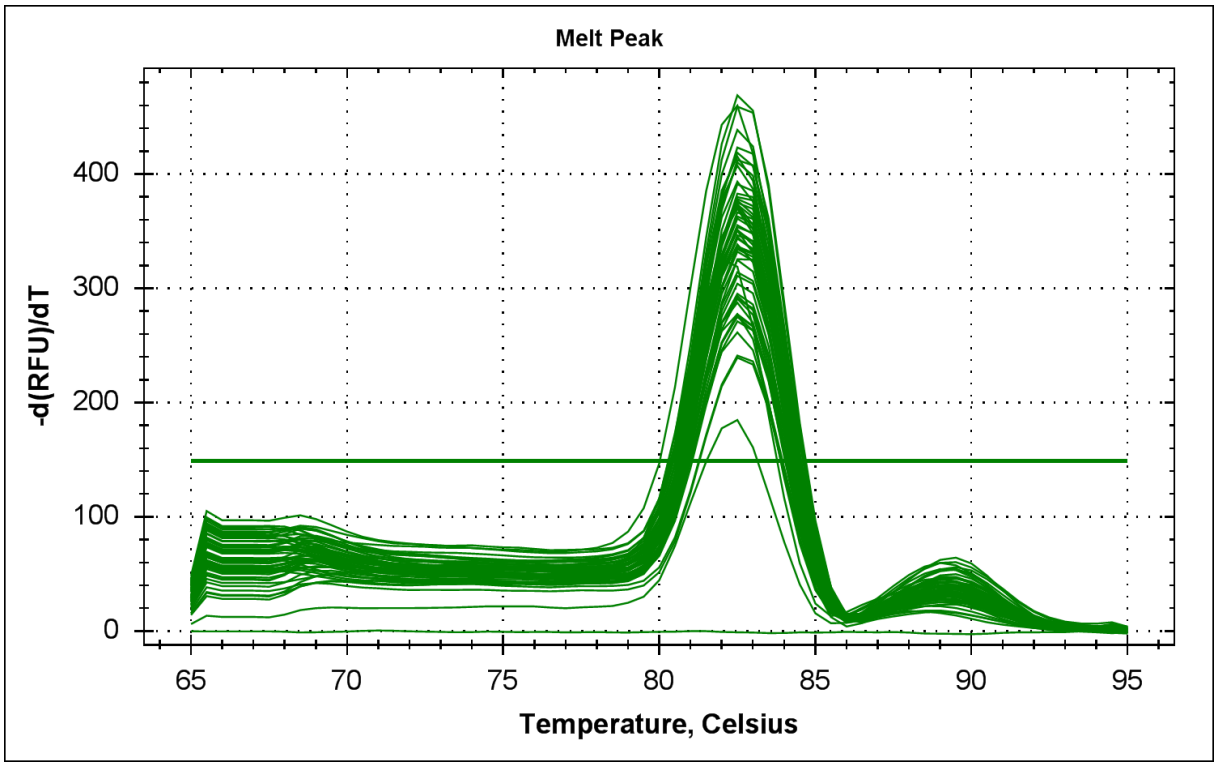

Figure B6: Melt peaks of DMG2, for 0h and 48h samples of biotypes SA1 and SAM

Stellenbosch University https://scholar.sun.ac.za

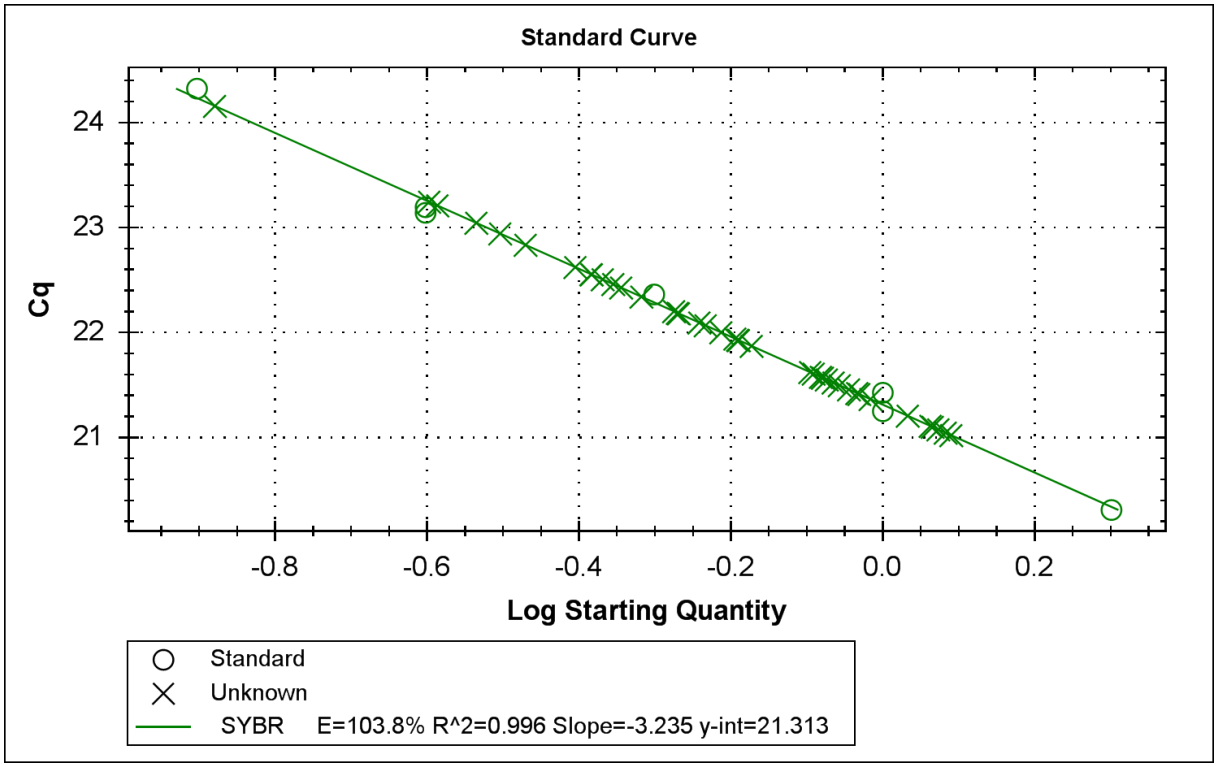

Figure B7: Standard curve of DMG2, for 6h samples of biotypes SA1 and SAM

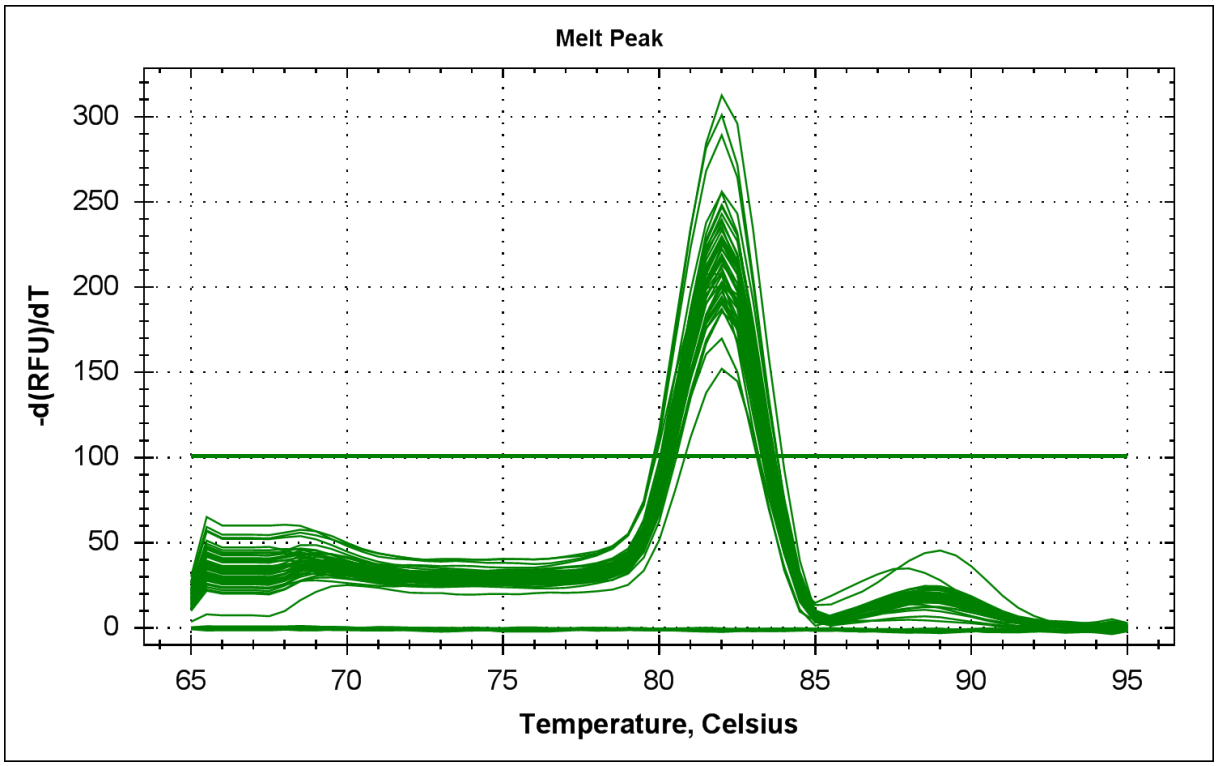

Figure B8: Melt peaks of DMG2, for 6h samples of biotypes SA1 and SAM

Stellenbosch University https://scholar.sun.ac.za

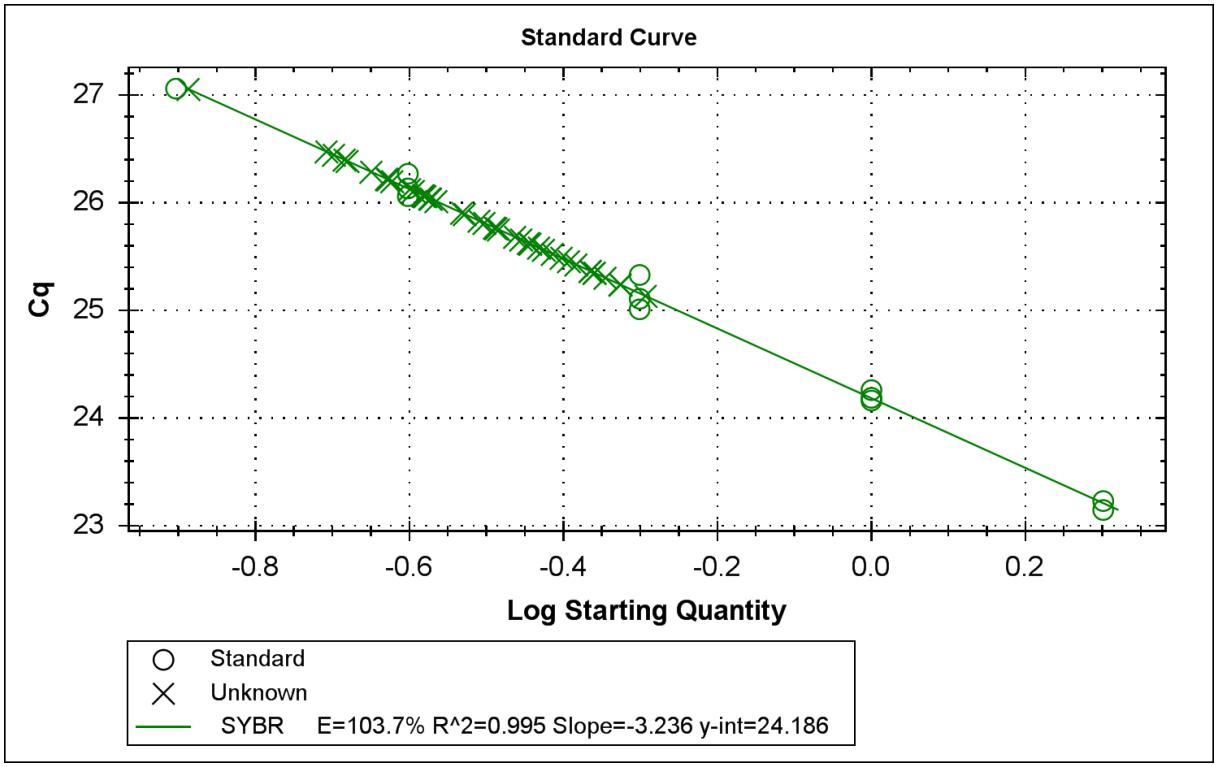

Figure B9: Standard curve of DMG3, for 0h and 48h samples of biotypes SA1 and SAM

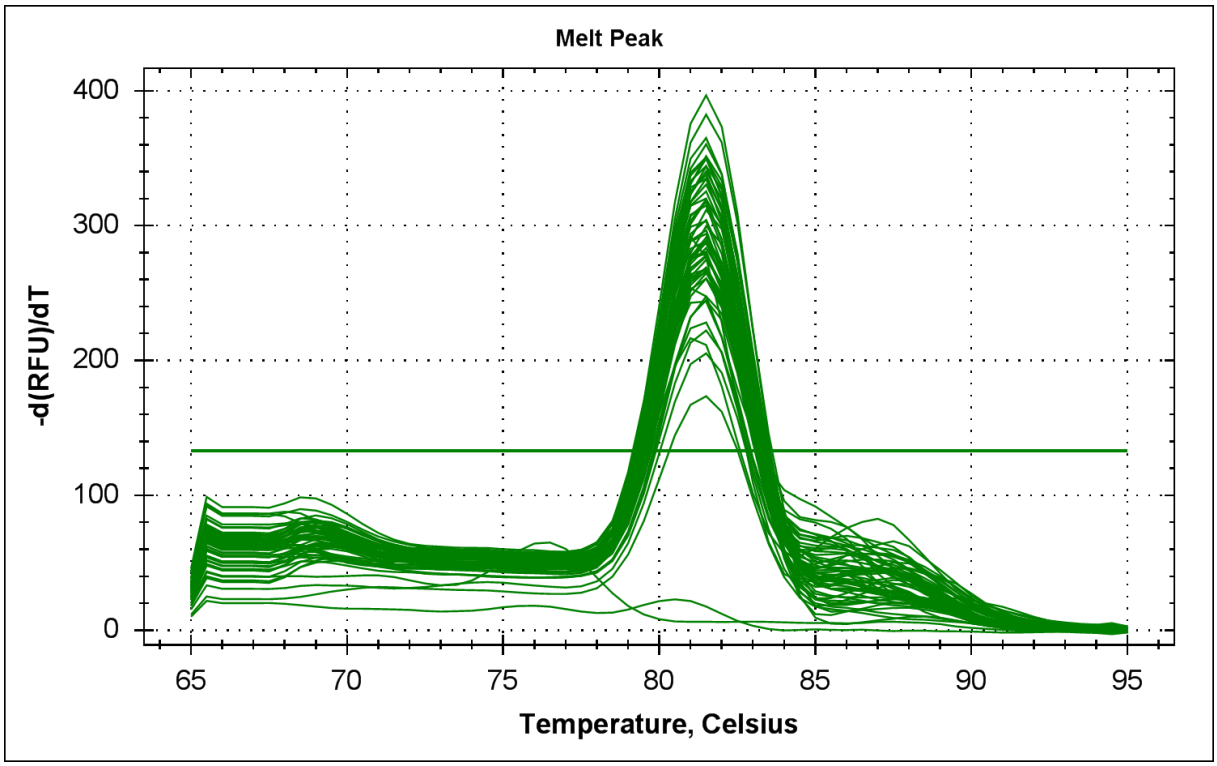

Figure B10: Melt peaks of DMG3, for 0h and 48h samples of biotypes SA1 and SAM

Stellenbosch University https://scholar.sun.ac.za

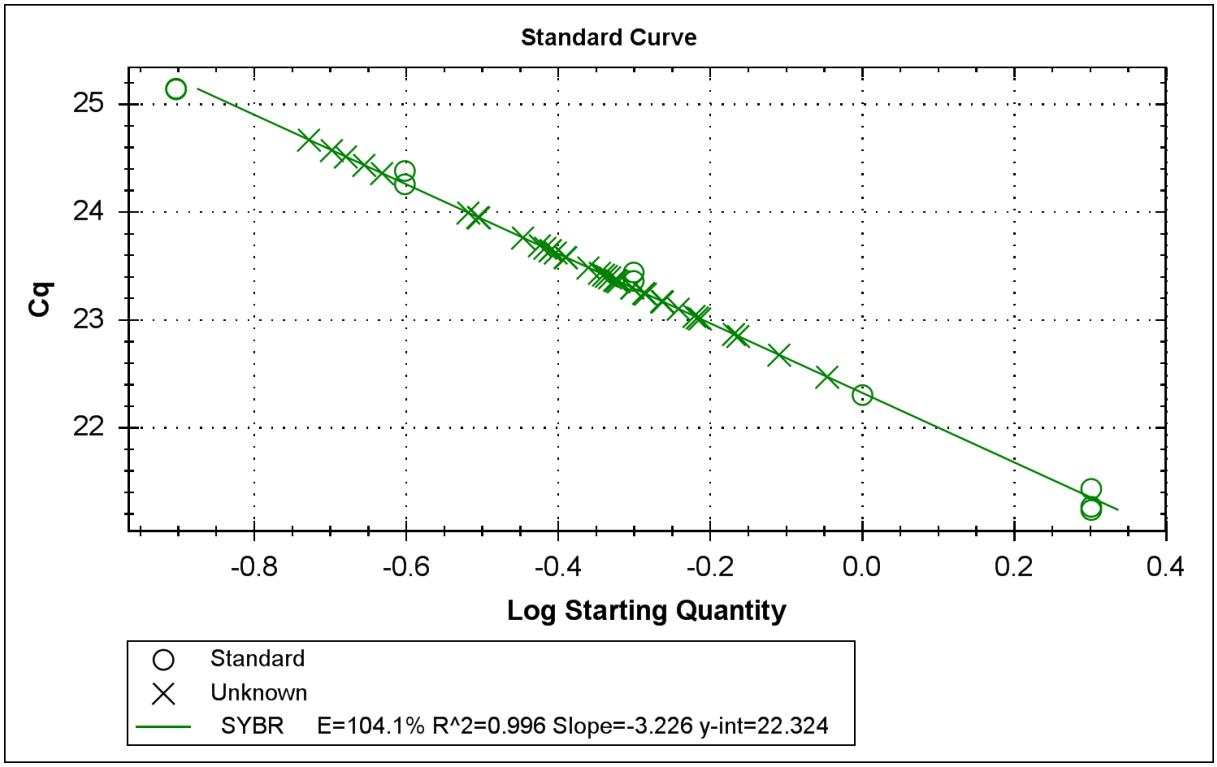

Figure B11: Standard curve of DMG3, for 6h samples of biotypes SA1 and SAM

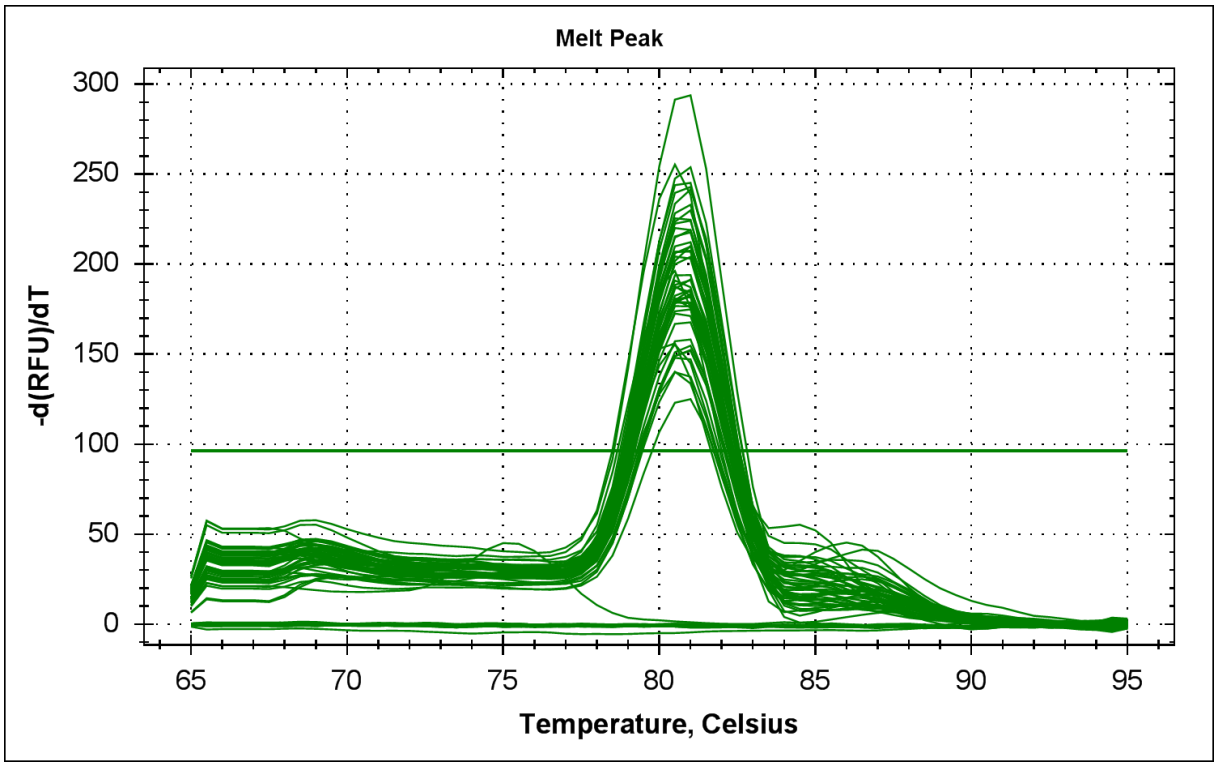

Figure B12: Melt peaks of DMG3, for 6h samples of biotypes SA1 and SAM

Stellenbosch University https://scholar.sun.ac.za

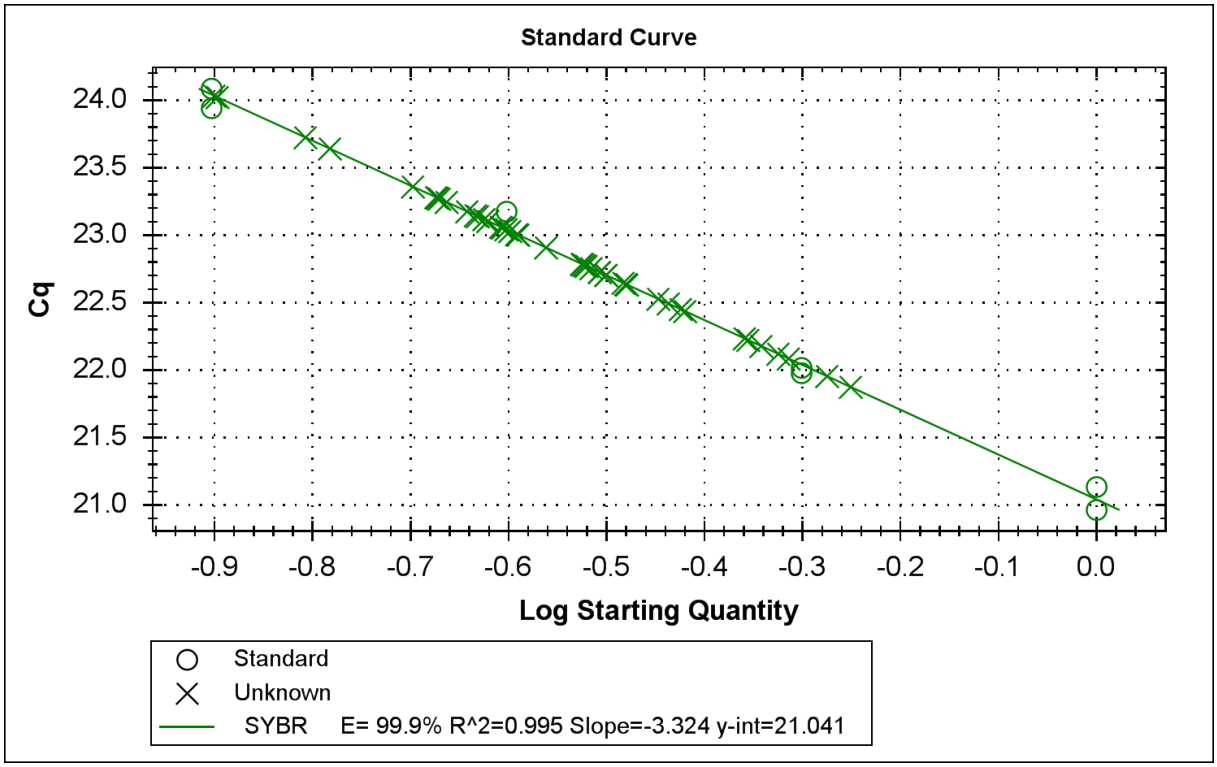

Figure B13: Standard curve of DMG4, for 0h and 48h samples of biotypes SA1 and SAM

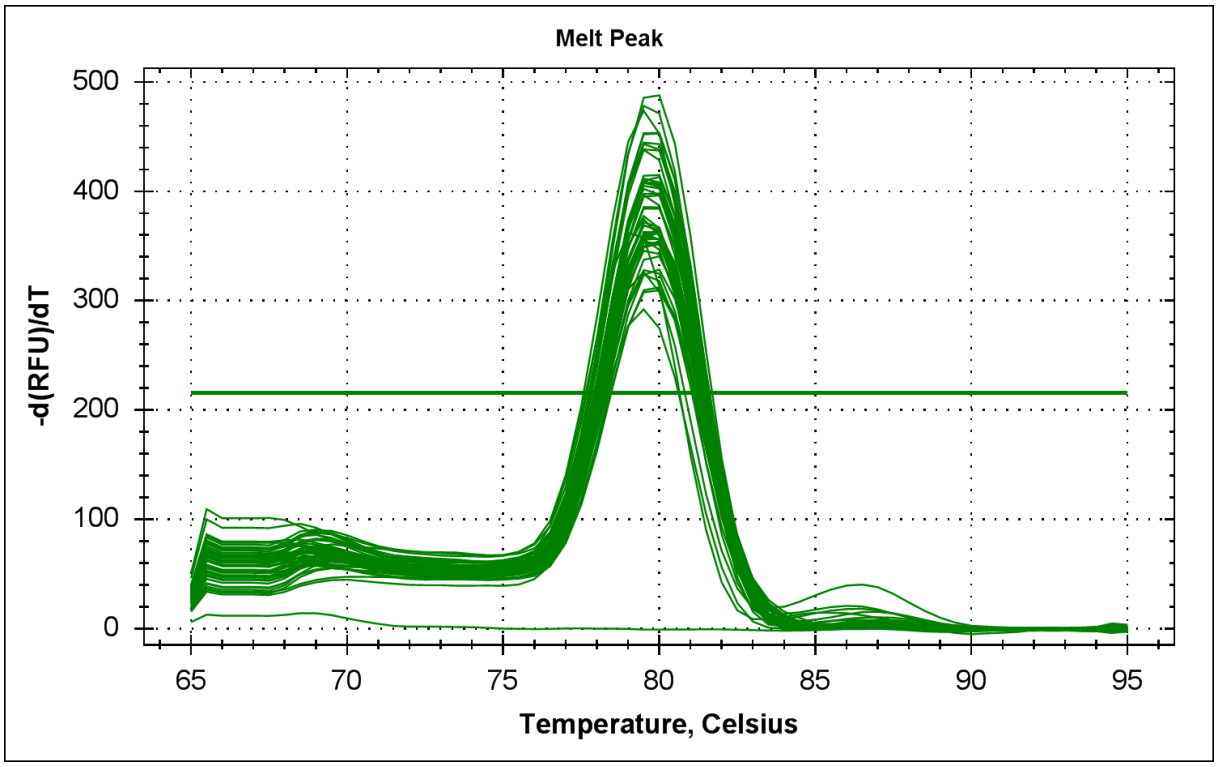

Figure B14: Melt peaks of DMG4, for 0h and 48h samples of biotypes SA1 and SAM

Stellenbosch University https://scholar.sun.ac.za

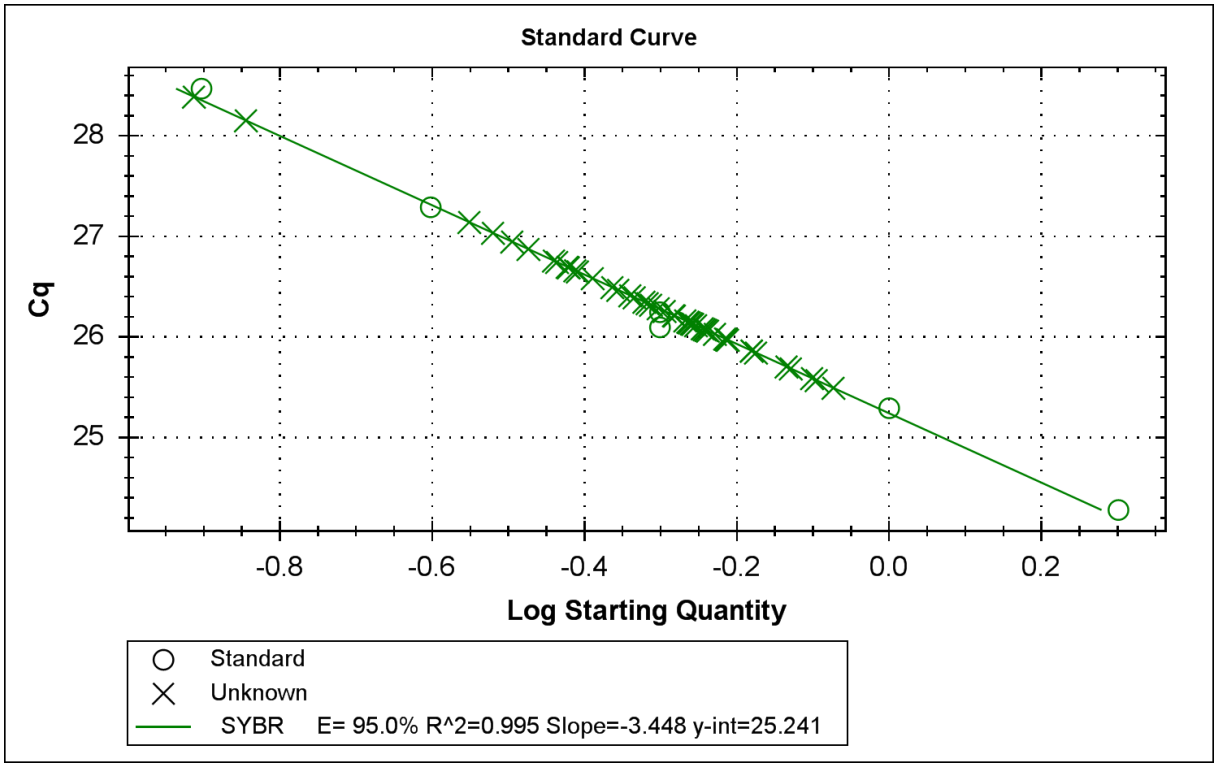

Figure B15: Standard curve of DMG4, for 6h samples of biotypes SA1 and SAM

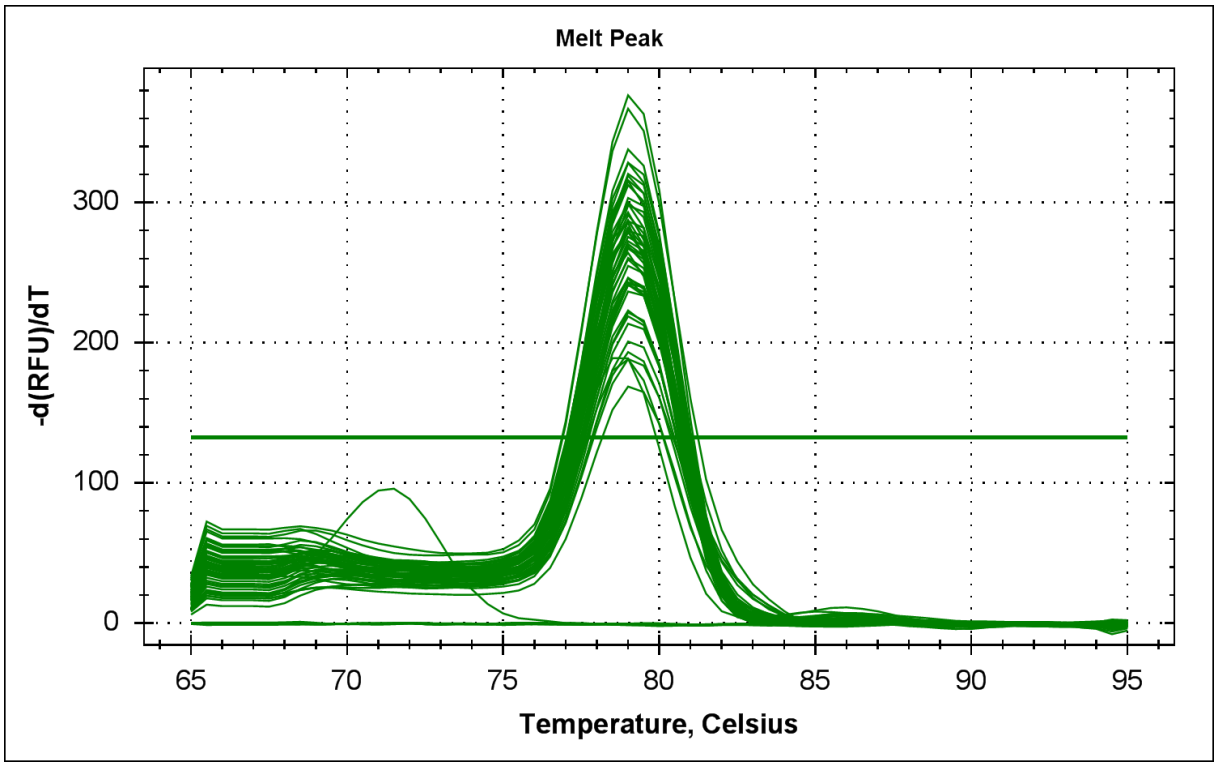

Figure B16: Melt peaks of DMG4, for 6h samples of biotypes SA1 and SAM

Stellenbosch University https://scholar.sun.ac.za

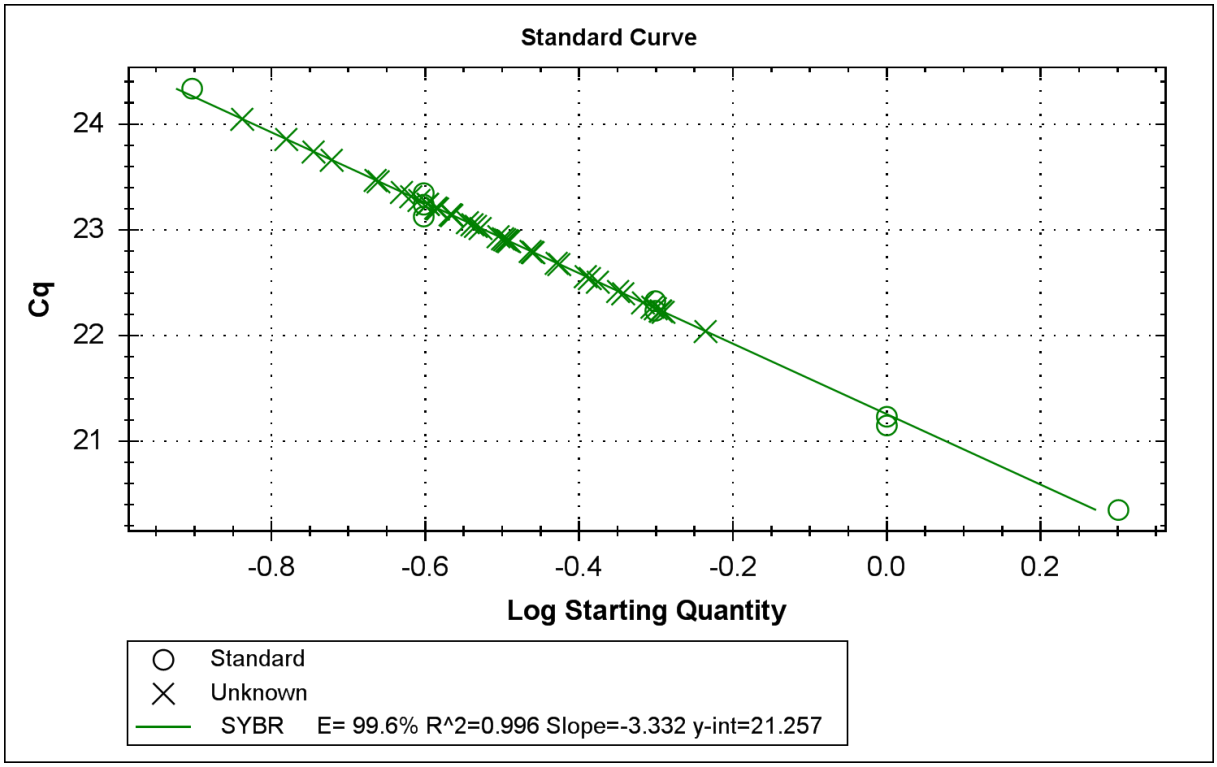

Figure B17: Standard curve of DMG5, for 0h and 48h samples of biotypes SA1 and SAM

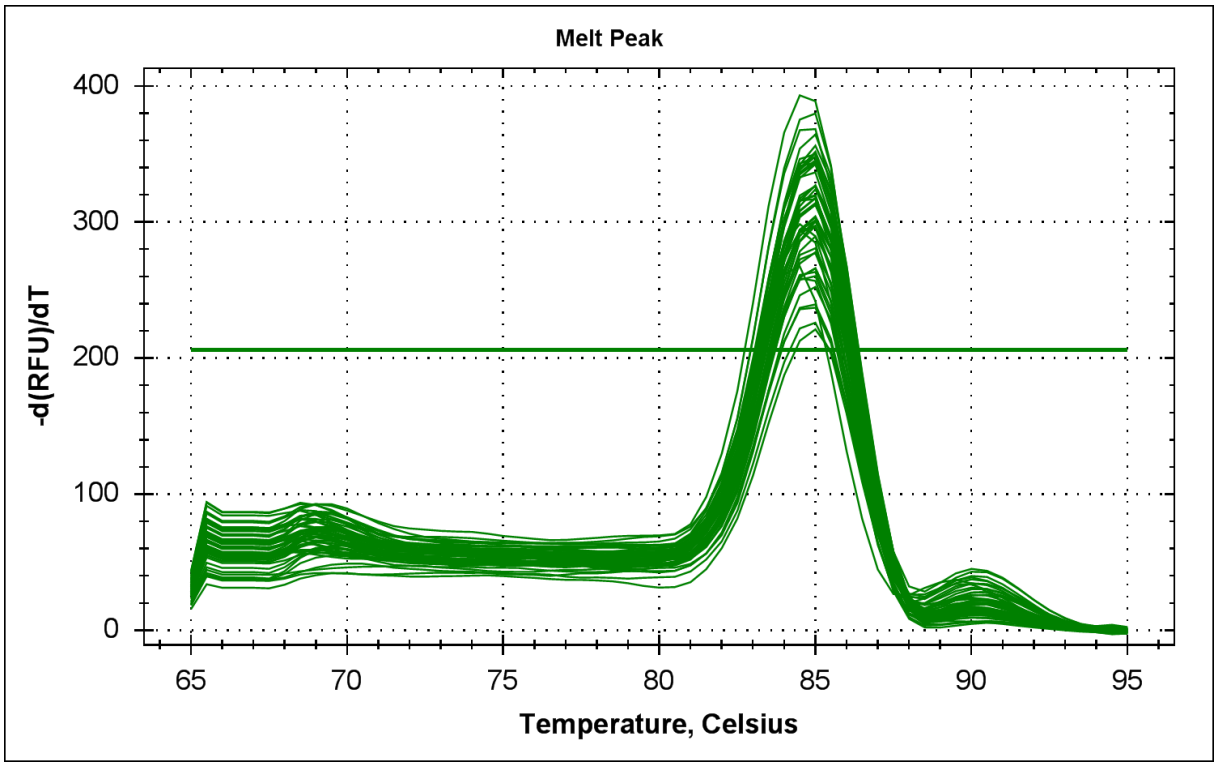

Figure B18: Melt peaks of DMG5, for 0h and 48h samples of biotypes SA1 and SAM

Stellenbosch University https://scholar.sun.ac.za

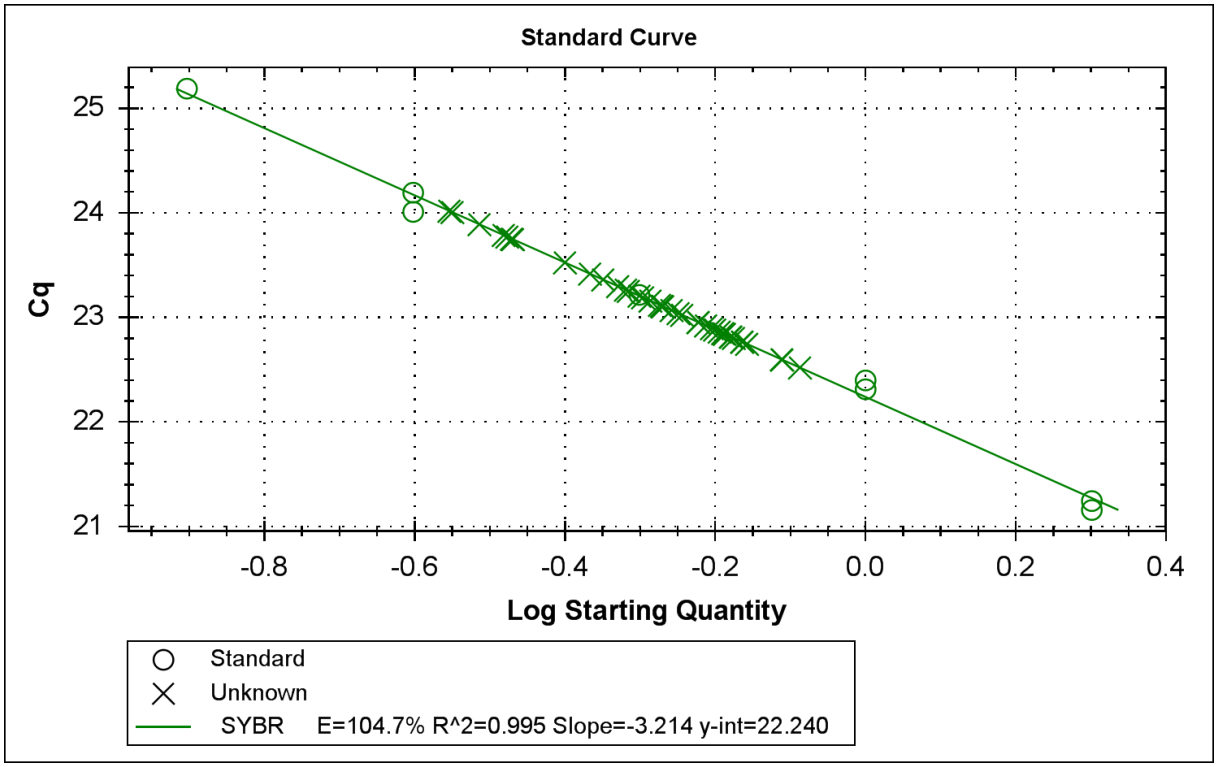

Figure B19: Standard curve of DMG5, for 6h samples of biotypes SA1 and SAM

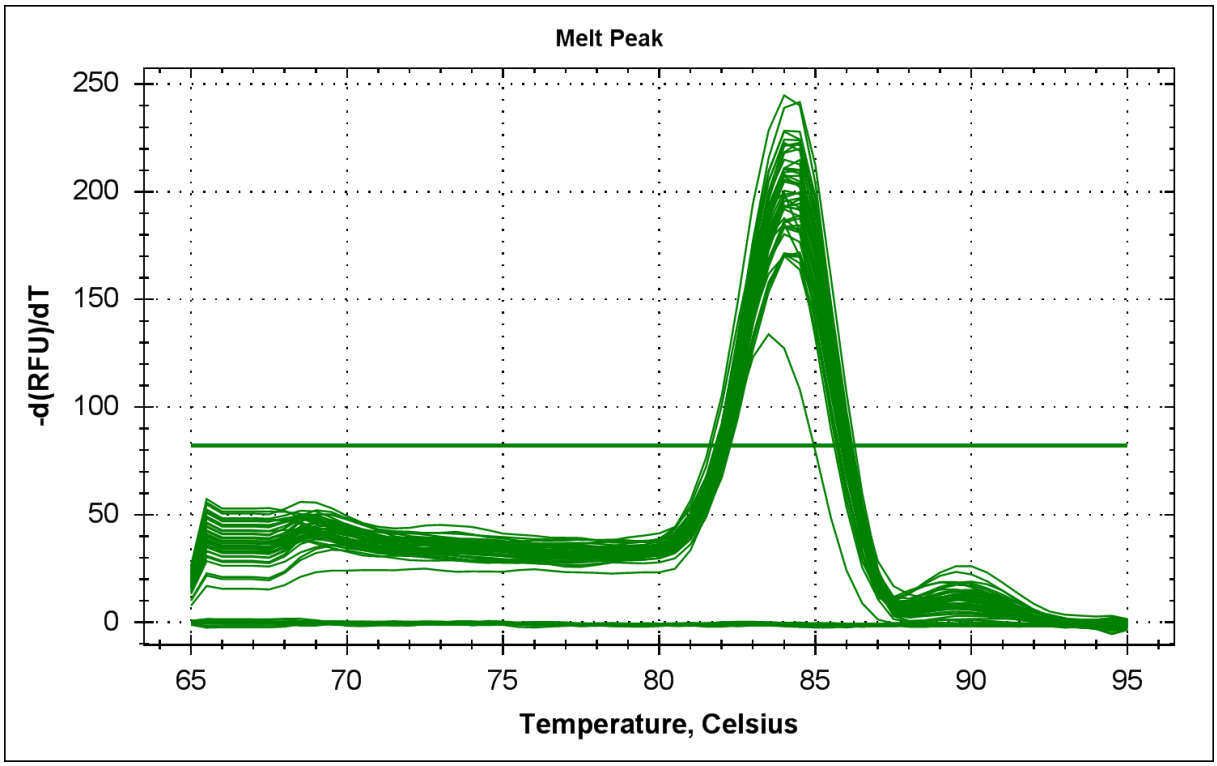

Figure B20: Melt peaks of DMG5, for 6h samples of biotypes SA1 and SAM

Stellenbosch University https://scholar.sun.ac.za

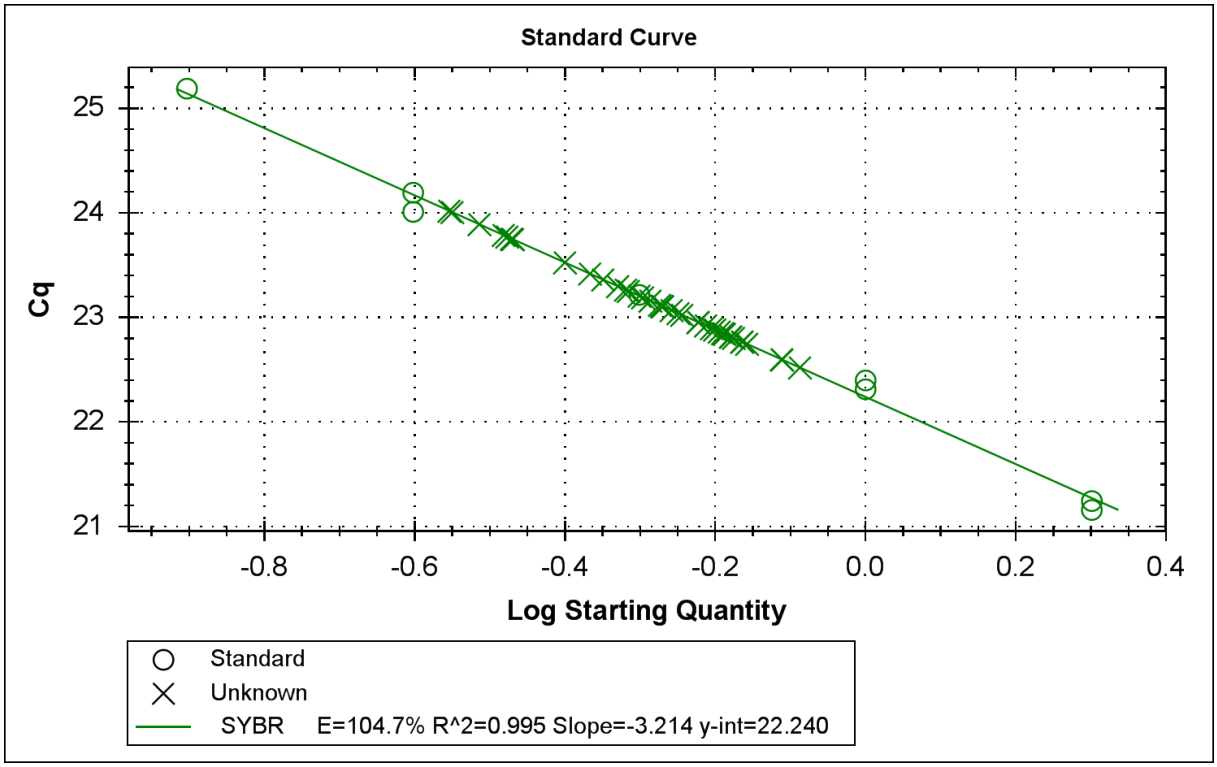

Figure B21: Standard curve of L27, for 0h and 48h samples of biotypes SA1 and SAM

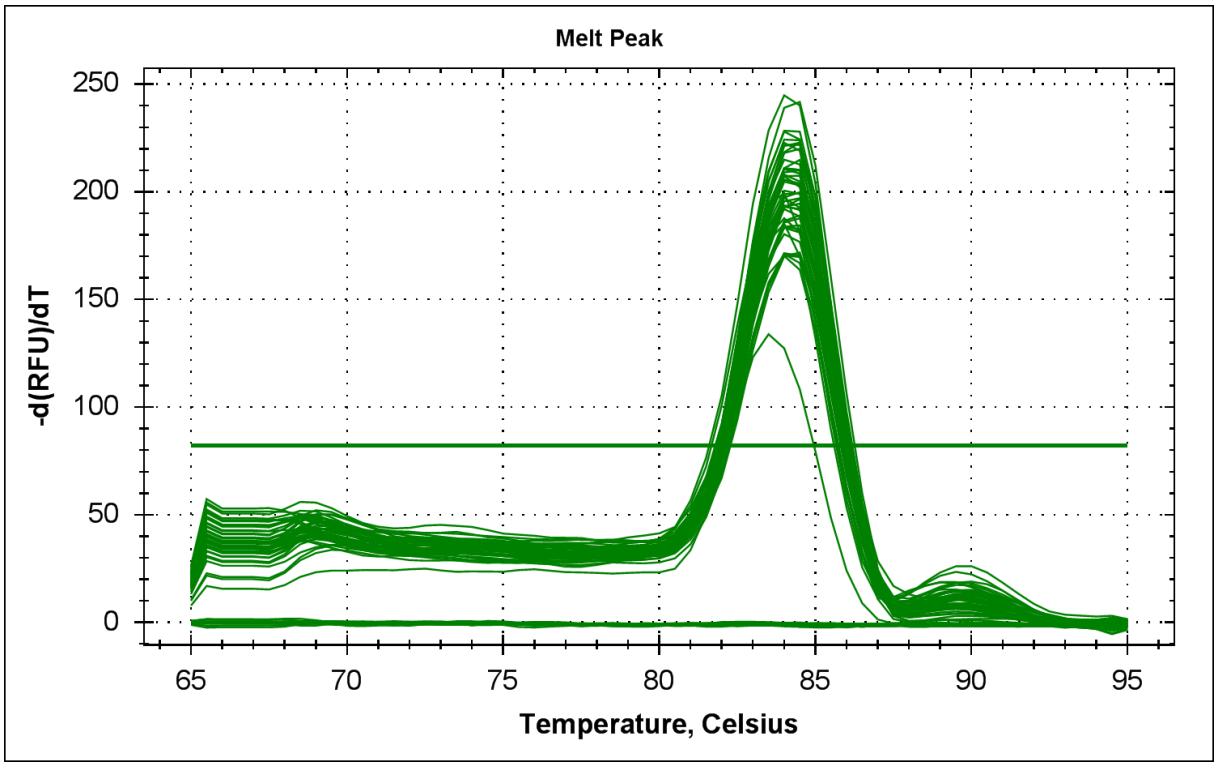

Figure B22: Melt peaks of L27, for 0h and 48h samples of biotypes SA1 and SAM

Stellenbosch University https://scholar.sun.ac.za

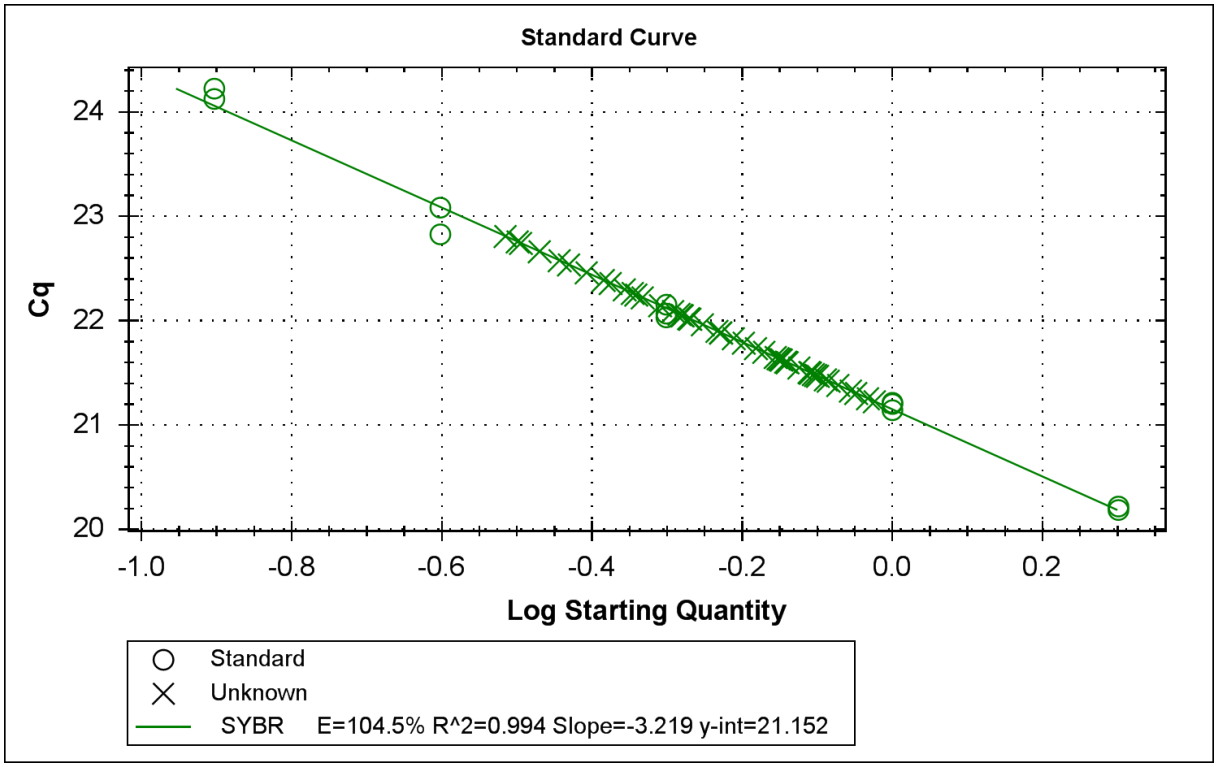

Figure B23: Standard curve of L27, for 6h samples of biotypes SA1 and SAM

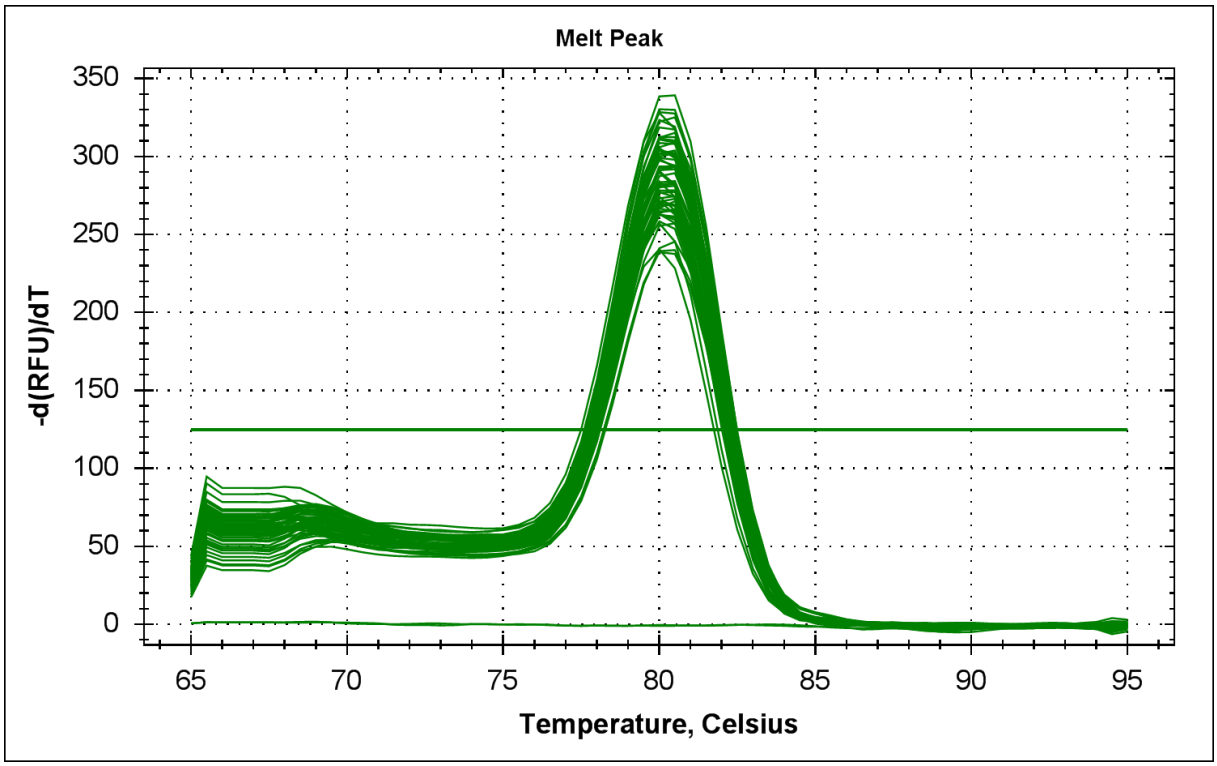

Figure B24: Melt peaks of L27, for 6h samples of biotypes SA1 and SAM

Stellenbosch University https://scholar.sun.ac.za

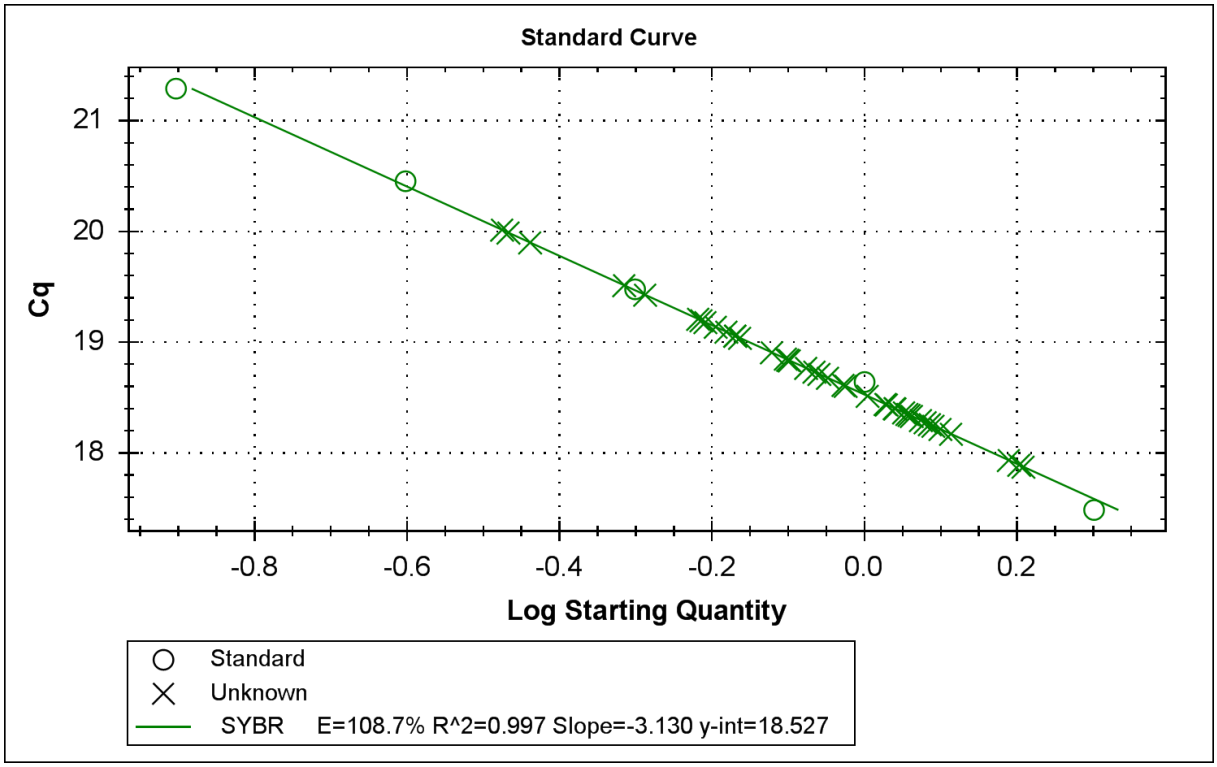

Figure B25: Standard curve of L27, for 6h samples of biotypes SA1 and SAM, on Tugela-*Dn1*

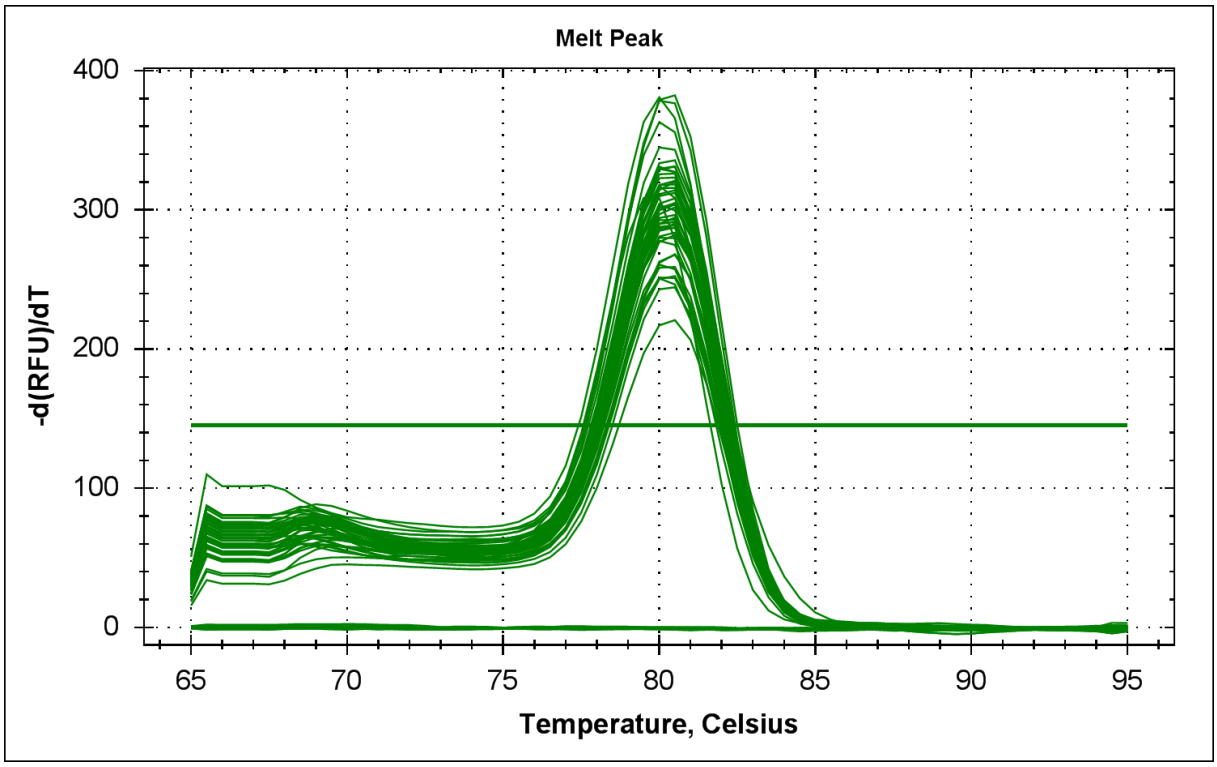

Figure B26: Melt peaks of L27, for 6h samples of biotypes SA1 and SAM, on Tugela-*Dn1*

Stellenbosch University https://scholar.sun.ac.za

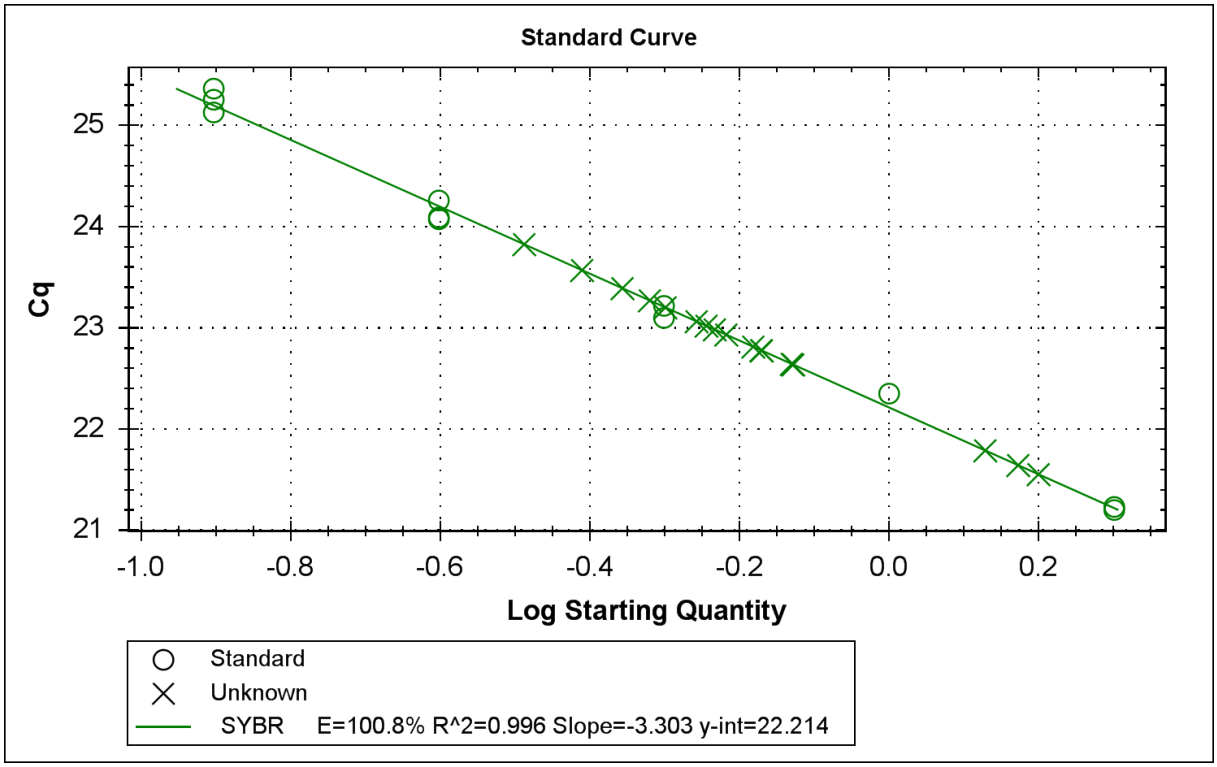

Figure B27: Standard curve of L32, for 0h and 48h samples of biotypes SA1 and SAM

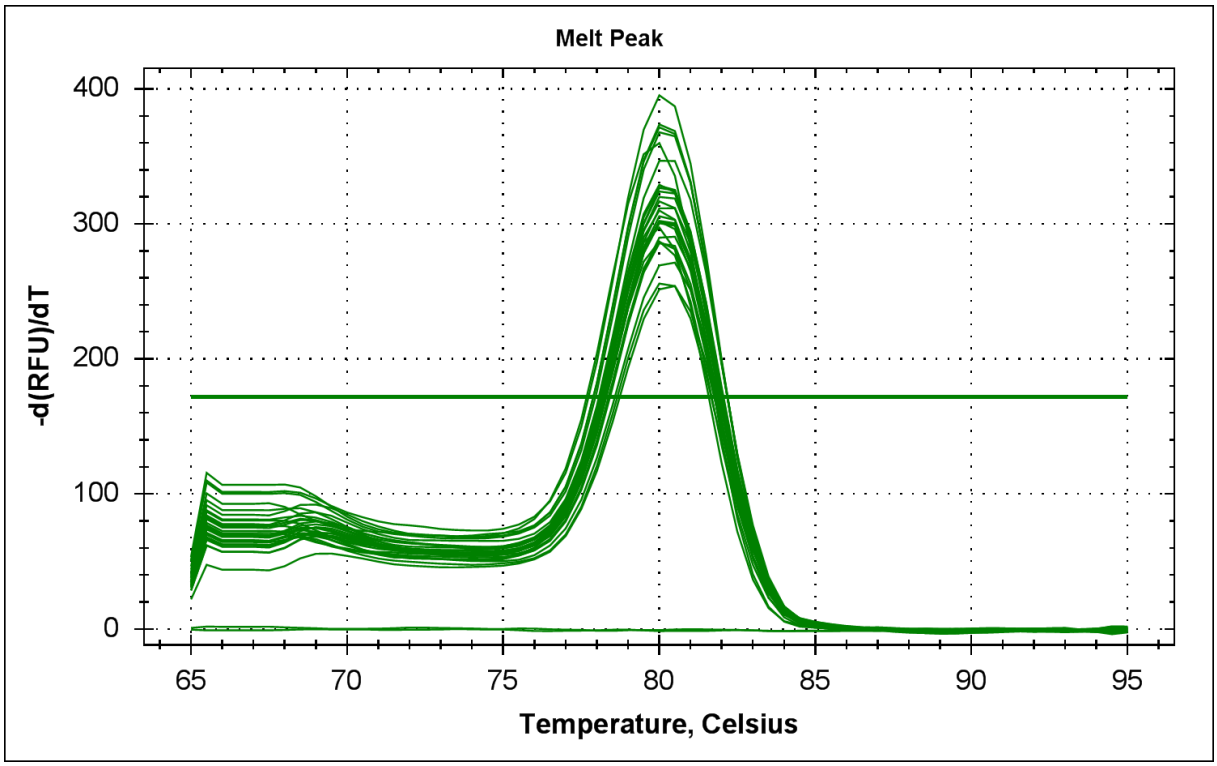

Figure B28: Melt peaks of L32, for 0h and 48h samples of biotypes SA1 and SAM

Stellenbosch University https://scholar.sun.ac.za

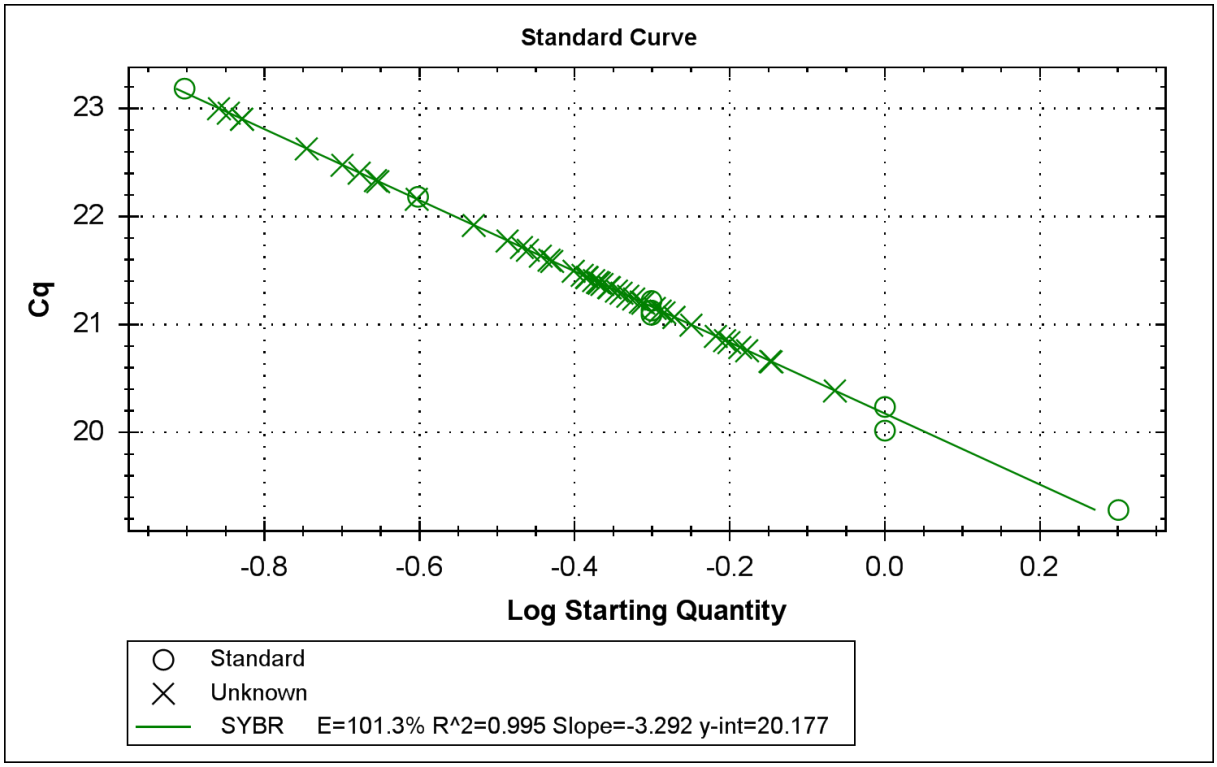

Figure B29: Standard curve of L32, for 6h samples of biotypes SA1 and SAM

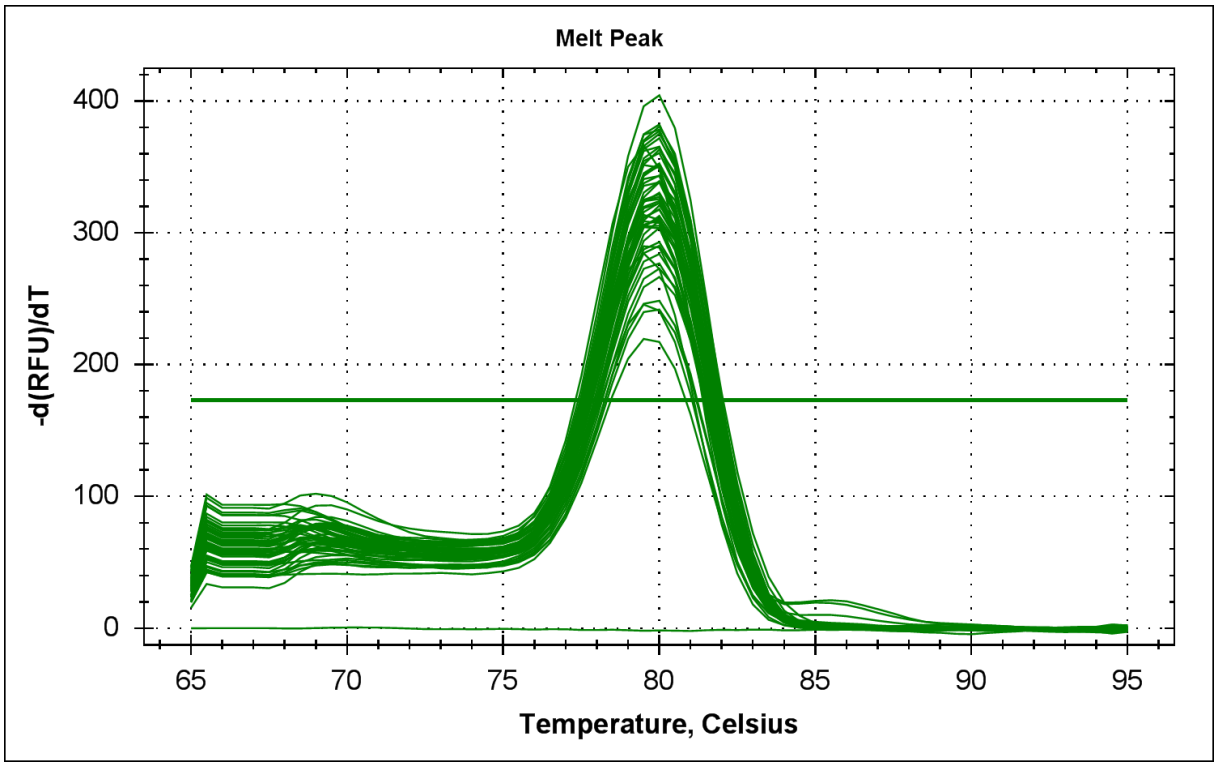

Figure B30: Melt peaks of L32, for 6h samples of biotypes SA1 and SAM

Stellenbosch University https://scholar.sun.ac.za

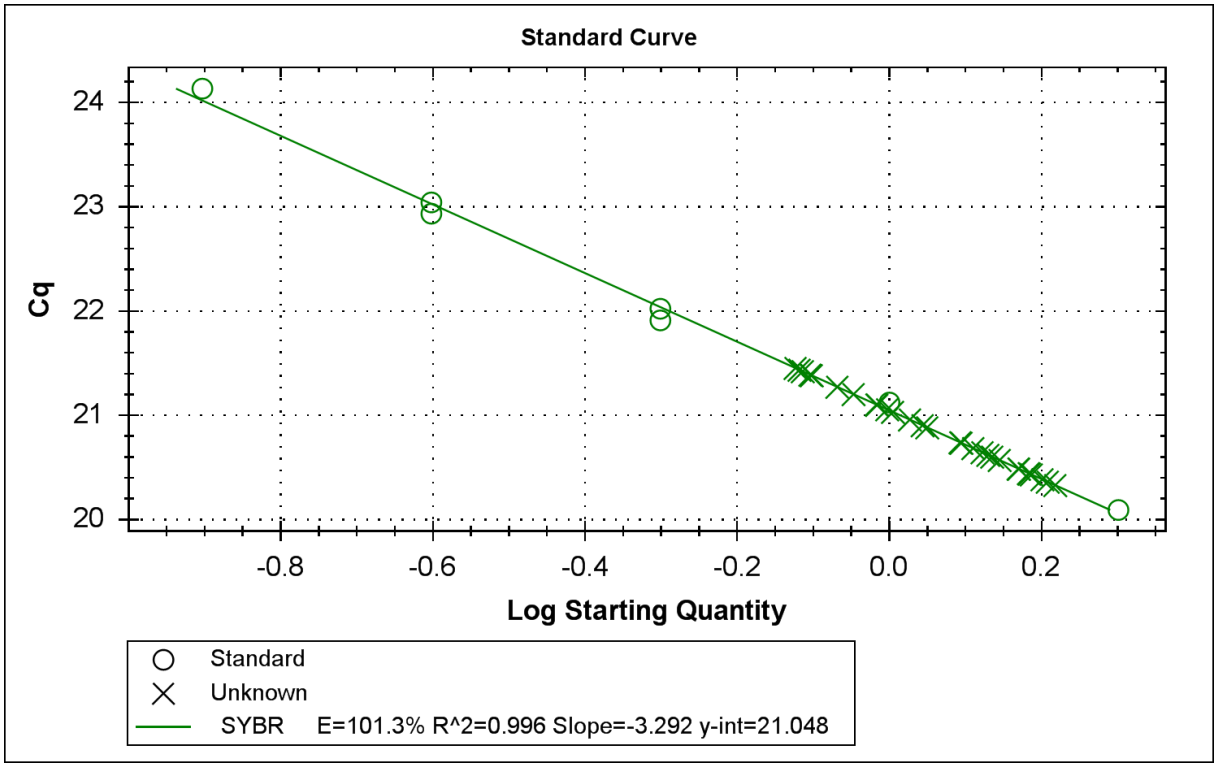

Figure B31: Standard curve of L32, for 6h samples of biotypes SA1 and SAM, on Tugela-*Dn1*

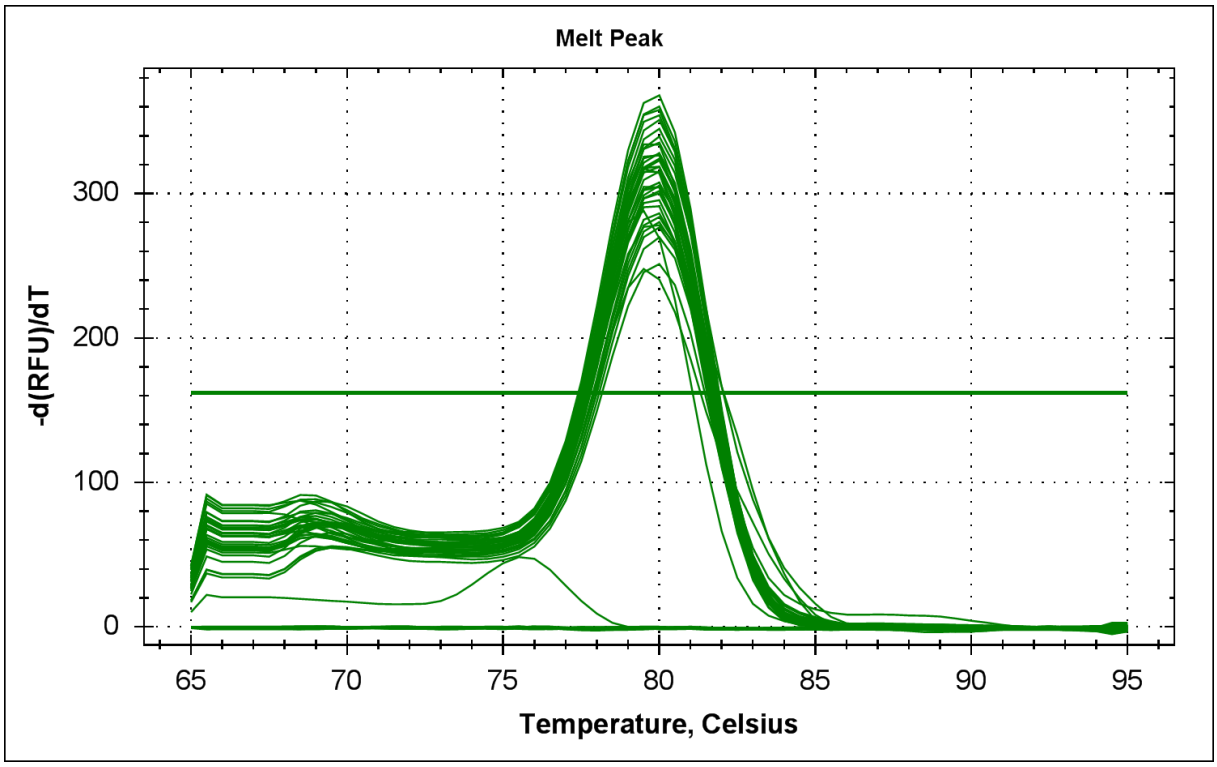

Figure B32: Melt peaks of L32, for 6h samples of biotypes SA1 and SAM, on Tugela-*Dn1*

Stellenbosch University https://scholar.sun.ac.za

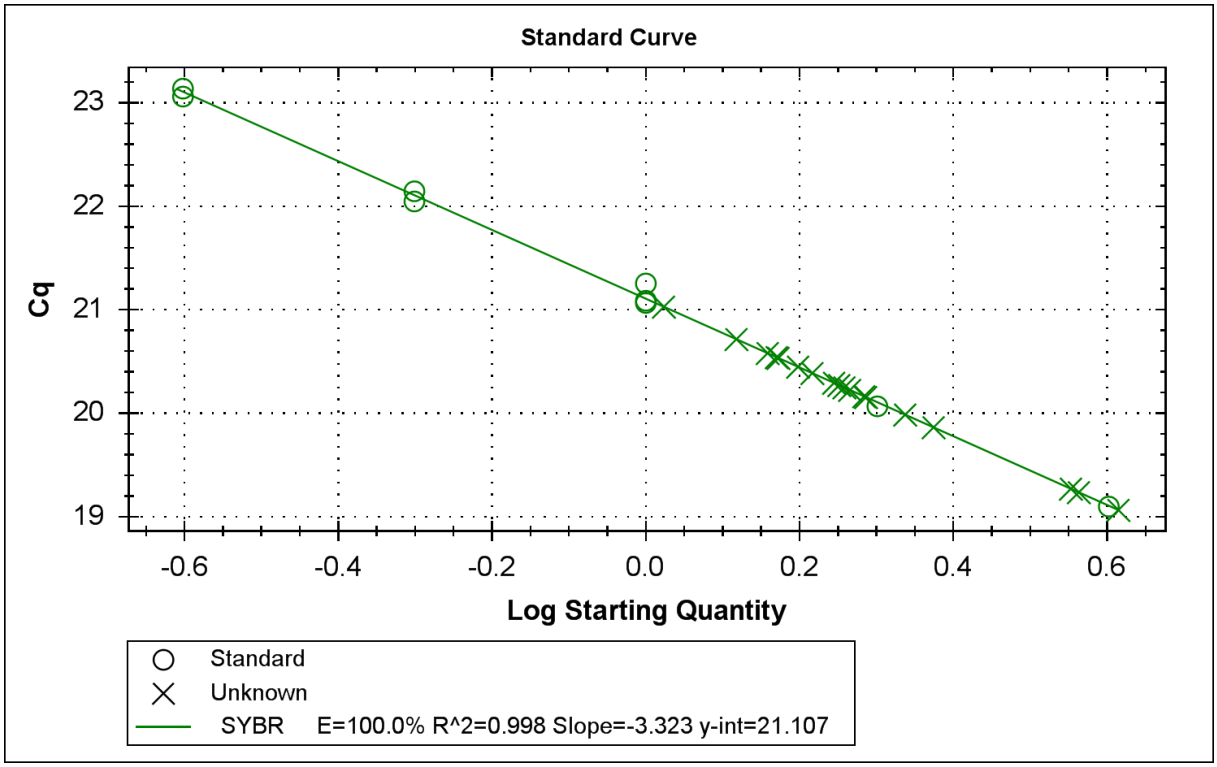

Figure B33: Standard curve of DNMT3, for 0h and 48h samples of biotypes SA1 and SAM

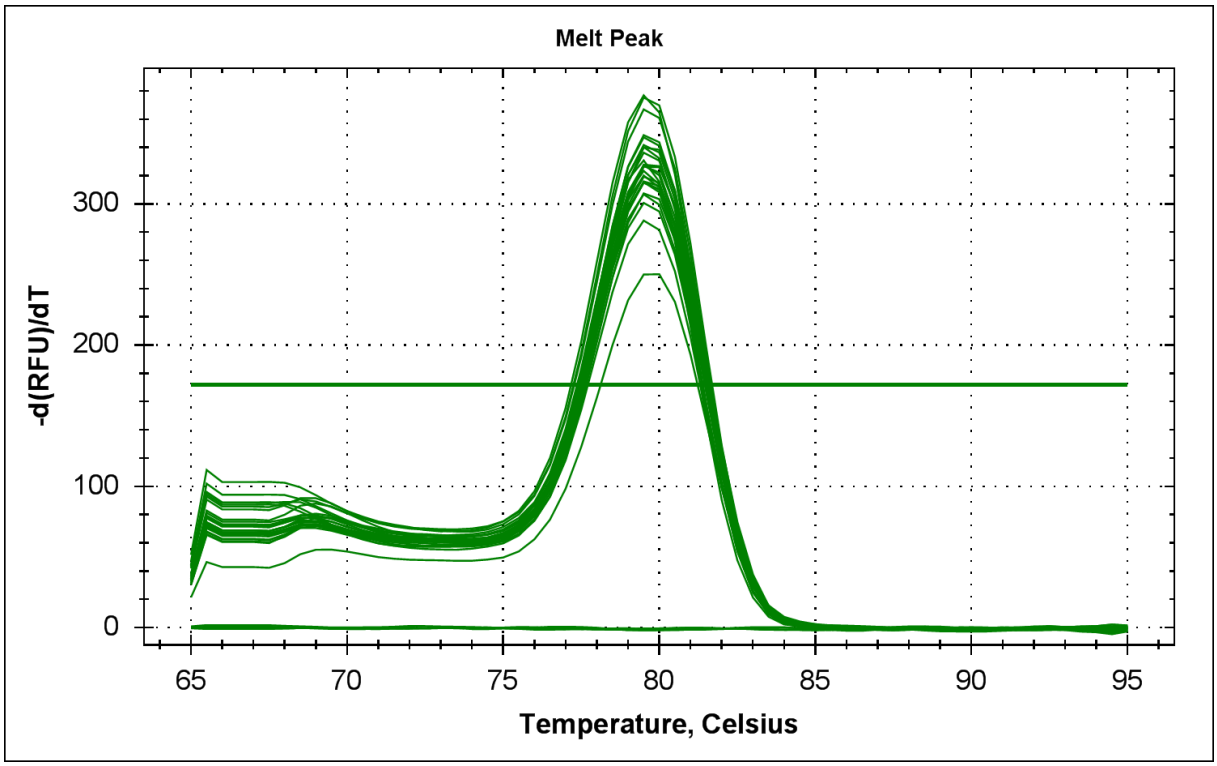

Figure B34: Melt peaks of DNMT3, for 0h and 48h samples of biotypes SA1 and SAM

Stellenbosch University https://scholar.sun.ac.za

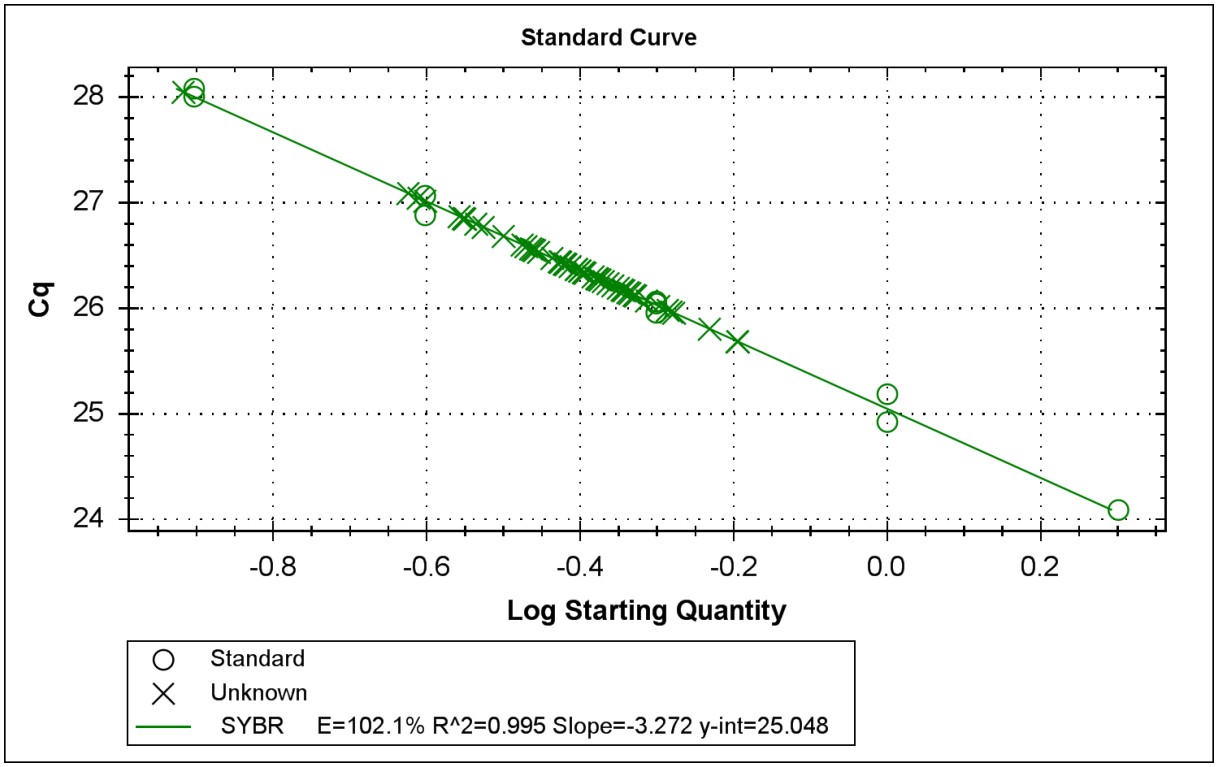

Figure B35: Standard curve of DNMT3, for 6h samples of biotypes SA1 and SAM

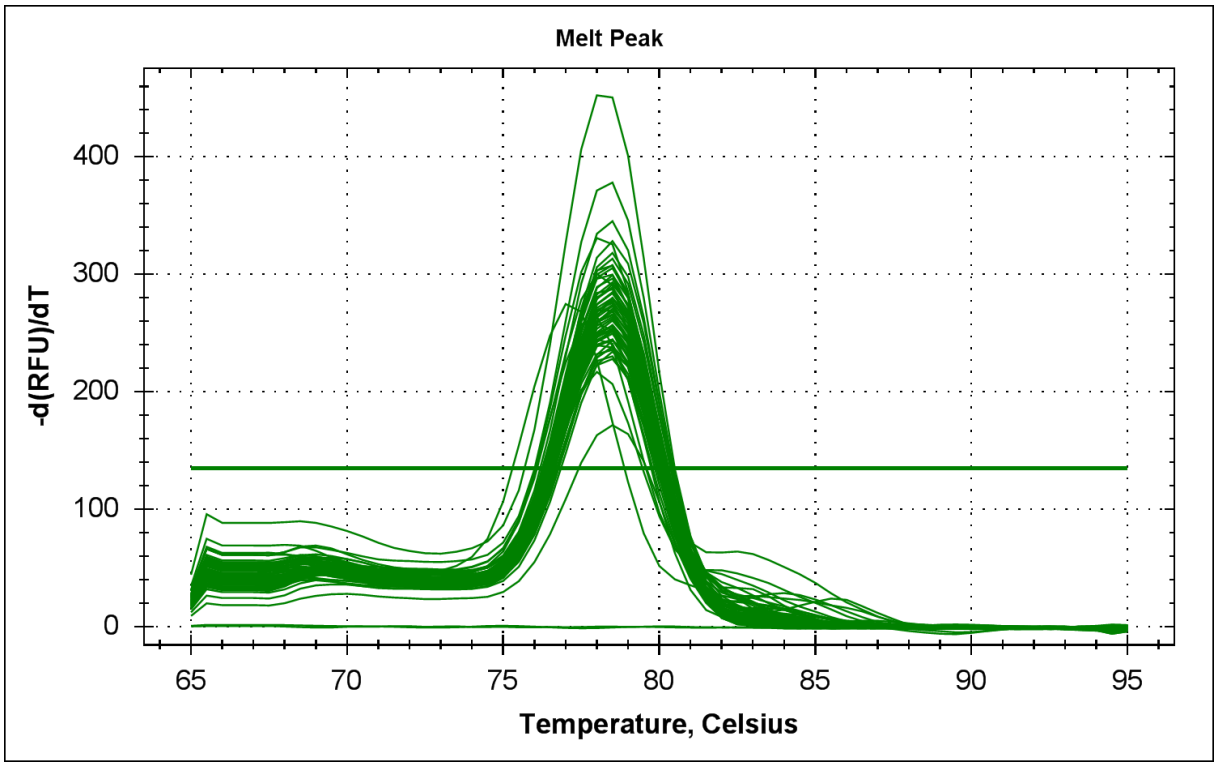

Figure B36: Melt peaks of DNMT3, for 6h samples of biotypes SA1 and SAM

Stellenbosch University https://scholar.sun.ac.za

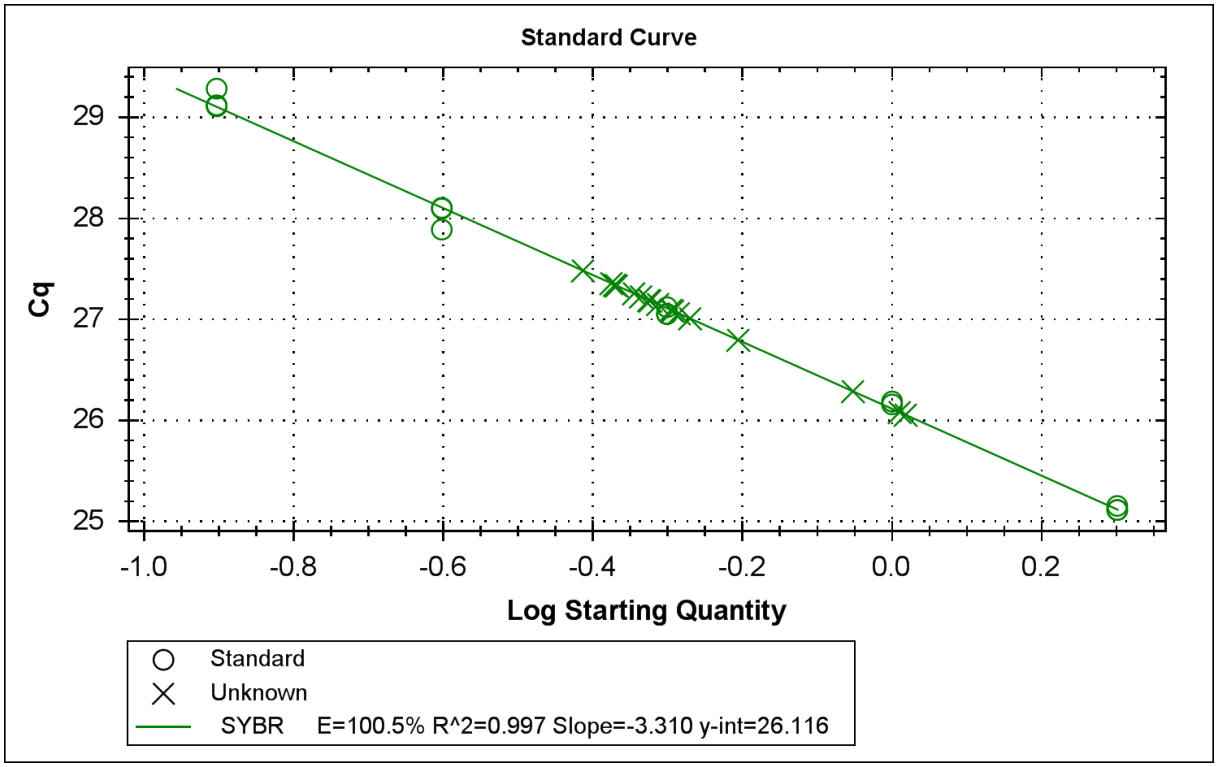

Figure B37: Standard curve of TET, for 0h and 48h samples of biotypes SA1 and SAM

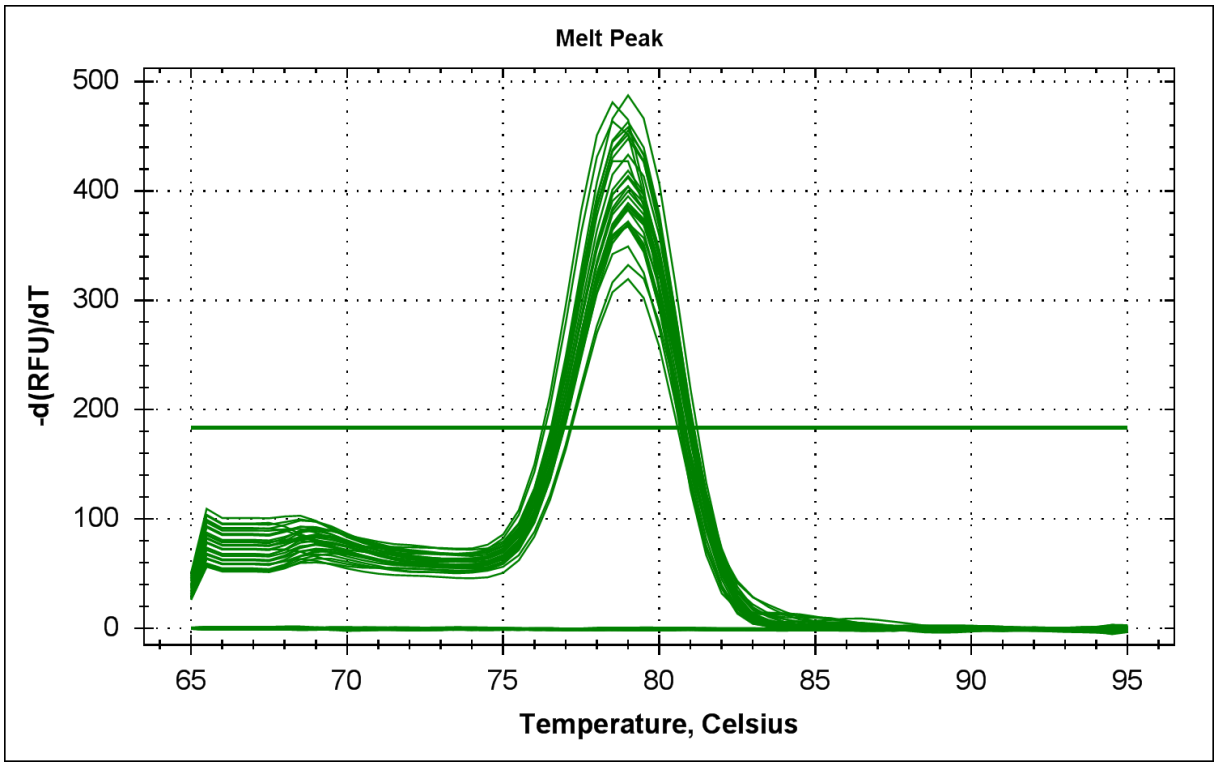

Figure B38: Melt peaks of TET, for 0h and 48h samples of biotypes SA1 and SAM

Stellenbosch University https://scholar.sun.ac.za

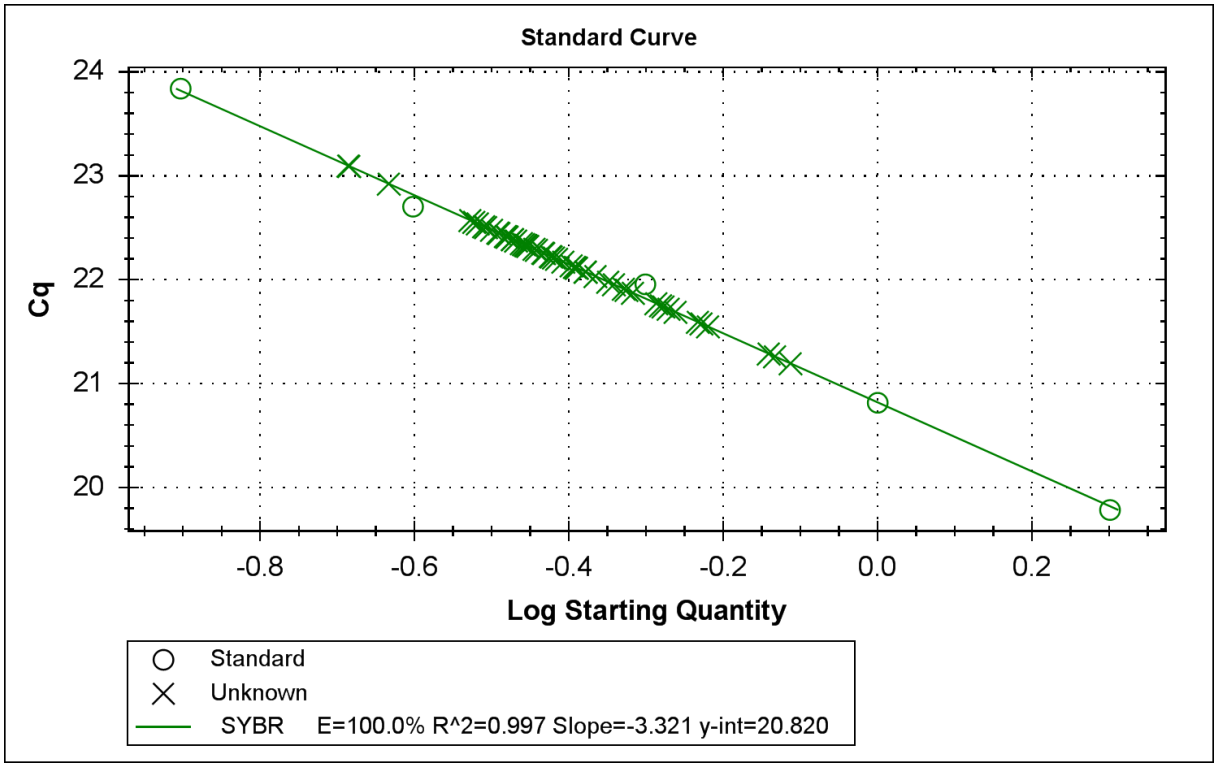

Figure B39: Standard curve of TET, for 6h samples of biotypes SA1 and SAM

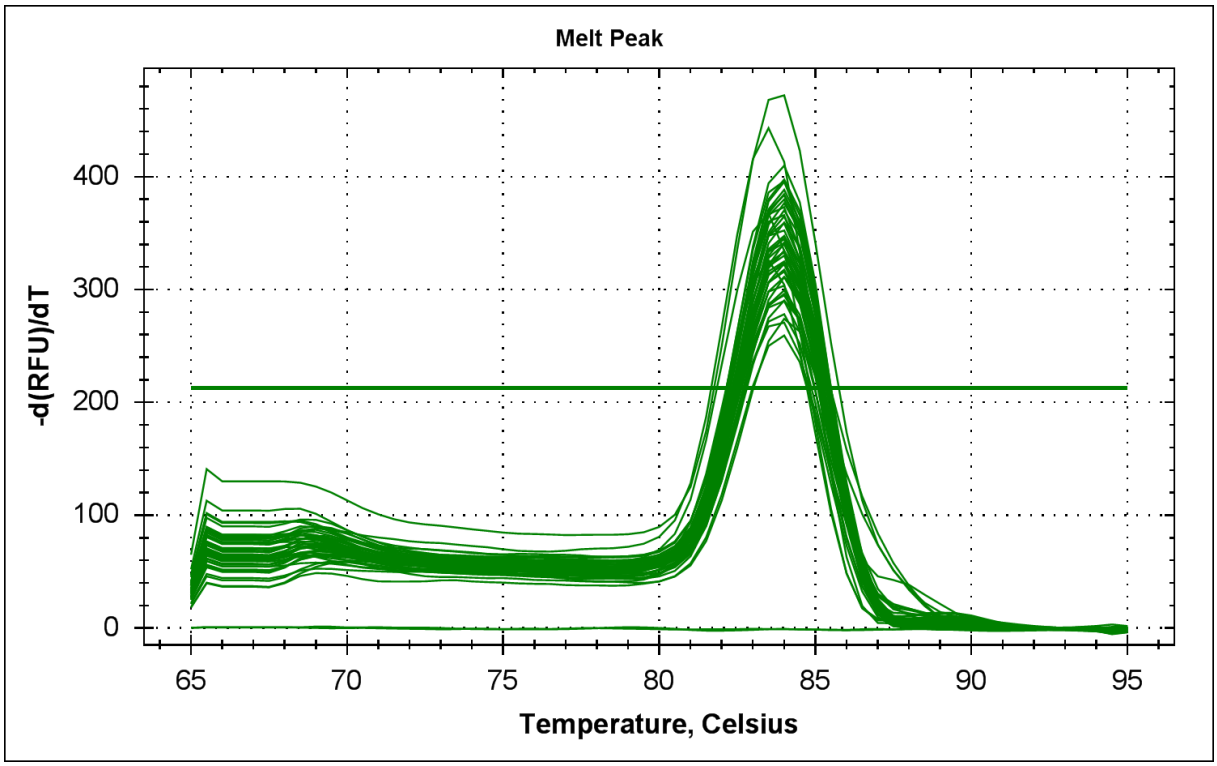

Figure B40: Melt peaks of TET, for 6h samples of biotypes SA1 and SAM

Stellenbosch University https://scholar.sun.ac.za

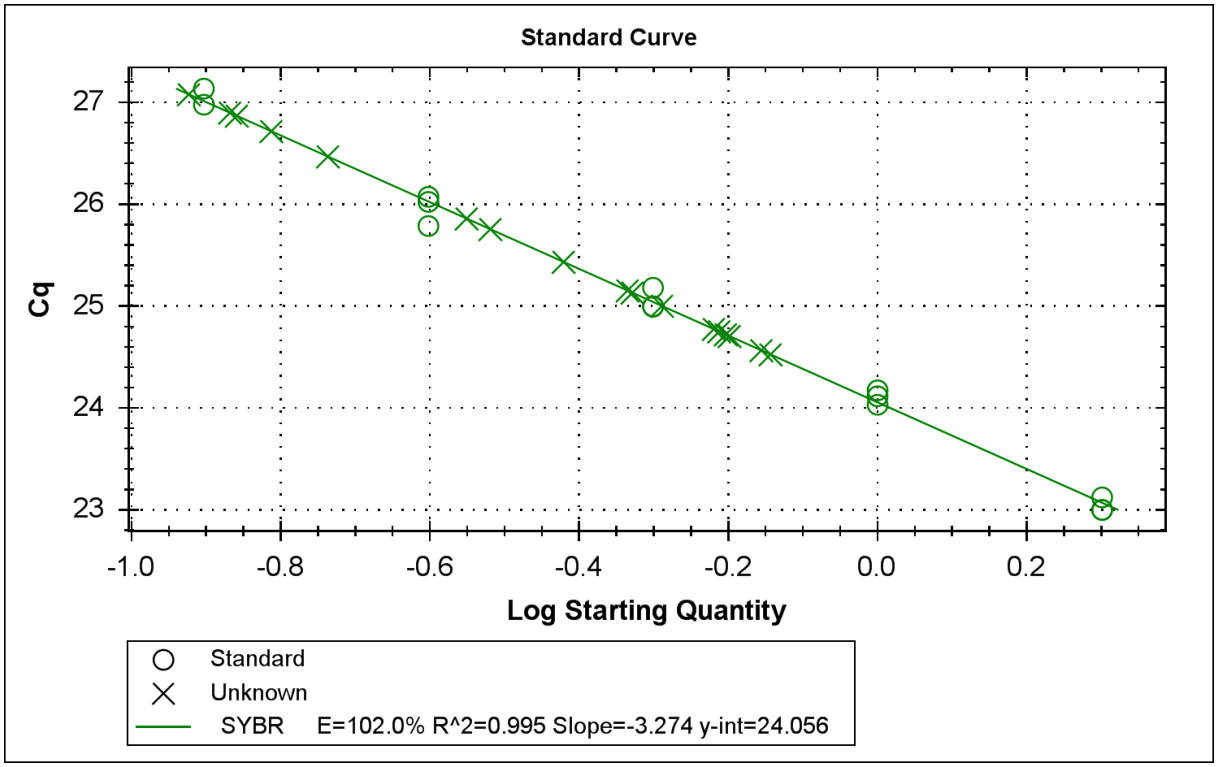

Figure B41: Standard curve of TET, for 6h samples of biotypes SA1 and SAM, on Tugela-*Dn1*

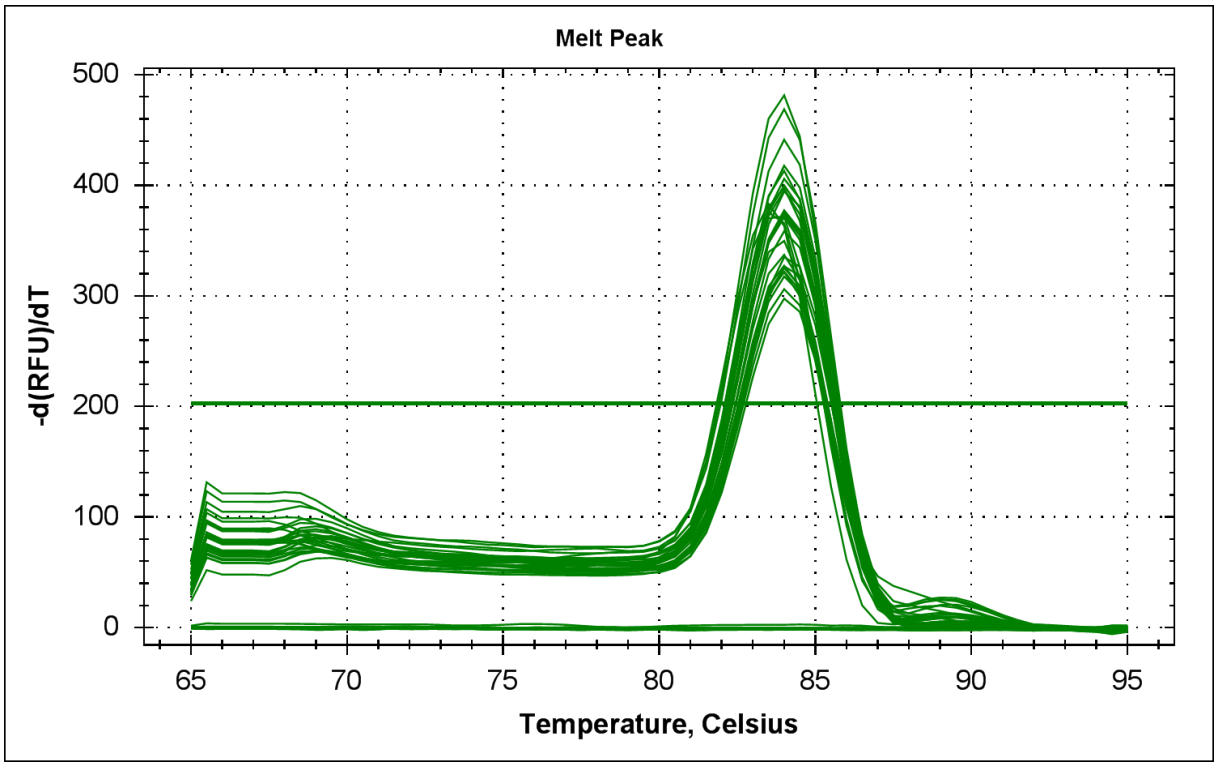

Figure B42: Melt peaks of TET, for 6h samples of biotypes SA1 and SAM, on Tugela-*Dn1*

Stellenbosch University https://scholar.sun.ac.za

Chapter 5: Conclusion

5.1: Summary

The production of wheat, an economically important food crop, is negatively impacted by the Russian wheat aphid (RWA), a cereal pest (du Toit and Walters, 1981). While wheat cultivars resistant to feeding from RWA (du Toit, 1989; Nkongolo et al., 1991; Quick et al., 1996) have been developed, new RWA biotypes have emerged that are able to successfully feed on previously resistant cultivars (Botha, 2013; Haley et al., 2004; Jankielsohn, 2019). With the systematic breakdown of resistance to RWA, it is becoming increasingly important to understand the mechanism of adaptation and biotypification. Biotype SAM is a laboratory contained biotype that has been selectively bred from the least virulent naturally occurring South African biotype, SA1 (van Zyl and Botha, 2008). Together, these two biotypes form an excellent model for the study of RWA virulence, as they are at the opposite ends of the virulence scale, yet they are genealogically closely related (Breeds et al., 2018; van Zyl and Botha, 2008). The genetic differences between these two biotypes are very small, with only a 0.0008% difference in protein-coding sequences (Burger and Botha, 2017).

The discovery of differential methylation in putative effector proteins, between two US biotypes (Gong et al., 2012), as well as preliminary evidence of methylation differences between the South African biotypes SA1 and SAM (Breeds et al., 2018) has prompted the current study, in which whole genome bisulfite sequencing (WGBS) was performed on biotypes SA1 and SAM.

After trimming and filtering 55.0 Gbp and 65.5 Gbp of sequence data, with a Q30 of 96%, was obtained for biotype SA1 and SAM, respectively. An alignment efficiency of 61.3% was achieved for biotype SA1, with a slightly higher efficiency of 61.7% achieved for biotype SAM. A slight increase in efficiency for biotype SAM was expected, as a SAM reference genome was used for alignment. The bisulfite conversion efficiencies were high, 99.664% and 99.618%, for biotypes SA1 and SAM respectively. All available metrices indicated that the WGBS data is of good quality and roughly a 100x coverage was achieved.

The overall trends in DNA methylation in the RWA is in concordance with what has been previously reported for insects, these include: The CpG context being the most highly methylated context, with much lower levels of methylation in the CHG and CHH contexts and increased methylation in the genic regions, compared to intergenic regions, furthermore, within genes the methylation is mostly localized to the exons (Bewick et al., 2017; Glastad et al., 2014). Interestingly, the majority of

Stellenbosch University https://scholar.sun.ac.za

hemimethylated CpG sites occurred on the bottom strand, while more of the hemimethylated CHG sites occurred on the top strand. To the best of my knowledge this is the first time such a discrepancy has been reported.

A total of 148 genes were found to be differentially methylated between the biotypes, with 89 hypermethylated in biotype SA1 and 59 hypermethylated in biotype SAM. While the number of genes hypermethylated per ontological category was similar between biotypes for most categories, some categories were overrepresented by one biotype or the other. The "cellular response to stimulus", "cell communication", "regulation of cellular process", and "signal transduction" categories, for example, had three times the number of hypermethylated genes in SA1 than in SAM, while the "heterocyclic compound binding" category was not represented by SA1 at all.

No discernable pattern was observed in the differences in relative expression of a subset of these genes between the two biotypes, or within a single biotype at 6-hours and 48-hours after performing host shifts from a susceptible to a resistant host plant. There were however major differences in the relative expression of both *DNA methyltransferase 3* (*DNMT3*) and *ten-elven translocase* (*TET*), the gene coding for the enzymes which are, respectively, responsible for the establishment and abolishment of DNA methylation. At 6-hours after host shifts to Tugela *Dn5* the relative expression of *DNMT3* was upregulated nine times and the relative expression of *TET* was upregulated thirteen times in biotype SAM. This likely translates to a much greater rate of methylation and demethylation in biotype SAM and might explain the apparent greater ability of biotype SAM to respond to the defense responses initiated by host plants (Burger et al., 2017).

The ratio of observed to expected CpG sites ($CpG_{O/E}$), was lower for regions showing high levels of methylation (genic regions, especially exons), this also is in line with what has previously been reported (International Aphid Genomics Consortium, 2006). The seemingly bimodal distribution observed in $CpG_{O/E}$ distribution of the pea aphid (Walsh et al., 2010), was not observed for the RWA, however. The correlation between DNA methylation and $CpG_{O/E}$ appears to be negligible at best in the RWA. The inference of methylation using the abundance of CpG sites is therefore not advisable for future studies.

5.2: Future work

The observation that CpG hemimethylation is biased towards the bottom strand, while CHG hemimethylation is biased towards the top strand should be investigated. It might be due to a preference of DNMT1, the enzyme responsible for the maintenance of DNA methylation, for top stranded or bottom stranded hemimethylation sites, or an artifact created during DNA replication.

Stellenbosch University https://scholar.sun.ac.za

With the low levels of DNA methylation in the Russian wheat aphid, detection and investigation of differentially methylated genes, or even differentially methylated exons/introns, might not be sufficient. It might be prudent to score sites as methylated or unmethylated and conduct a study akin to a single nucleotide polymorphism (SNP) analysis, perhaps by combining the methylation data with an RNA-sequencing experiment to identify correlations between methylated sites and transcription levels.

5.3: References

Bewick, A.J., Vogel, K.J., Moore, A.J. & Schmitz, R.J., 2017, Evolution of DNA methylation across insects, *Molecular biology and evolution*, 34(3), 654–665.

Botha, A-M., 2013, A coevolutionary conundrum: The arms race between *Diuraphis noxia* (Kurdjumov) a specialist pest and its host *Triticum aestivum* (L.), *Arthropod-Plant Interactions*, 7(4), 359–372.

Breeds, K., Burger, N.F.V. & Botha, A-M., 2018, New insights into the methylation status of virulent *Diuraphis noxia* (Hemiptera: Aphididae) biotypes, *Journal of Economic Entomology*, 111(3), 1395–1403.

Burger, N.F. V, Venter, E. & Botha, A-M., 2017, Profiling *Diuraphis noxia* (Hemiptera: Aphididae) transcript expression of the biotypes SA1 and SAM feeding on various *Triticum aestivum* varieties, *Journal of Economic Entomology*, 110(2), 692–701.

Burger, N.F.V. & Botha, A.-M., 2017, Genome of Russian wheat aphid an economically important cereal aphid, *Standarts in Genomic Sciences*, 12(1), 90.

du Toit, F., 1989, Inheritance of resistance in two *Triticum aestivum* lines to Russian wheat aphid (Hornoptera: Aphididae), *Journal of Economic Entomology*, 82(4), 1251–1253.

du Toit, F. & Walters, M.C., 1981, Damage assessment and economic threshold values for the chemical control of thr Russian wheat aphid, *Diuraphis noxia* (Mordvilko) on winter wheat, *Technical communication, Department of Agriculture, Republic of South Africa*, (191), 58–62.

Glastad, K.M., Hunt, B.G. & Goodisman, M.A.D., 2014, Evolutionary insights into DNA methylation in insects, *Current Opinion in Insect SCience*, 1, 25–30.

Gong, L., Cui, F., Sheng, C., Lin, Z., Reeck, G., Xu, J. & Kang, L., 2012, Polymorphism and methylation of four genes expressed in salivary glands of Russian wheat aphid (Homoptera: Aphididae), *Journal of Economic Entomology*, 105(1), 232–241.

Haley, S.D., Peairs, F.B., Walker, C.B., Rudolph, J.B. & Randolph, T.L., 2004, Occurrence of a new Russian wheat aphid biotype in Colorado, *Crop Science*, 44(5), 1589–1592.

International Aphid Genomics Consortium, 2006, Insights into social insects from the genome of the honeybee Apis mellifera, *Nature*, 443(7114), 931–949.

Jankielsohn, A., 2019, New Russian wheat aphid biotype found in Free State, *Grain SA Mini Focus, Pest Control on Winter Cereals*.

Nkongolo, K.K., Quick, J.S., Limin, A.E. & Fowler, D.B., 1991, Sources and inheritance of resistance to Russian wheat aphid in *Triticum* species *amphiploids* and *Triticum tyauschii*, *Canadian Journal of Plant Science*, 71(3), 703–708.

Stellenbosch University https://scholar.sun.ac.za

Quick, J.S., Ellis, G.E., Normann, R.M., Stromberger, J.A., Shanahan, J.F., Peairs, F.B., Rudolph, J.B. & Lorenz, K., 1996, Registration of "Halt" wheat, *Crop Science*, 36(1).

van Zyl, R.A. & Botha, A-M., 2008, Eliciting proteins from *Diuraphis noxia* biotypes differ in size and composition, *18th Biennial International Plant Resistance to Insects Workshop*.

Walsh, T.K., Brisson, J.A., Robertson, H.M., Gordon, K., Jaubert-Possamai, S., Tagu, D. & Edwards, O.R., 2010, A functional DNA methylation system in the pea aphid, *Acyrthosiphon pisum*, *Insect molecular biology*, 12(2).