

**IDENTIFICATION OF ESSENTIAL GENE TARGETS IN CASSAVA *BEMISIA*
TABACI, FOR EFFECTIVE WHITEFLY MANAGEMENT USING RNA
INTERFERENCE**

KAWEESI TADEO

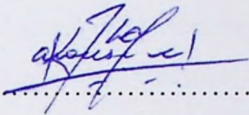
**A thesis submitted in partial fulfillment of the requirement of University of Greenwich for
the Degree of Doctor of Philosophy**

December 2019

DECLARATION

"I certify that this work has not been accepted in substance for any degree and is not concurrently being submitted for any degree other than of Doctor of Philosophy being studied at the University of Greenwich. I also declare that this work is the result of my own investigation except where otherwise identified by reference and that I have not plagiarized the work of others"

Kaweesi Tadeo

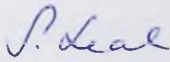


Date

19/12/19

(Student)

Prof. Susan Seal

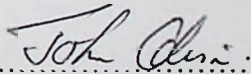


Date

19/12/19

(First supervisor)

Prof. John Colvin

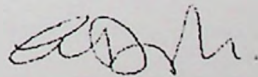


Date

20/12/19

(Second supervisor)

Prof. Angela Douglas



Date ...19 December 2019.....

(Third supervisor)

ACKNOWLEDGEMENTS

I thank and praise the Almighty God, with whom is no variableness, neither shadow of turning, for His love, care and protection that has been invaluable during this PhD.

I am deeply indebted to my supervisors Prof. Angela Douglas, Prof. John Colvin and Prof. Susan Seal who have shared their knowledge, guidance and skills throughout this programme. I also thank Dr. Paul Visendi and Dr. Sophie Bouvaine for the valuable comments and recommendations. This programme has been well supported through a financial grant to the African Cassava Whitefly project, funded by the Bill & Melinda Gates Foundation, of which I am very grateful.

I convey my heartfelt thanks to Prof. Angela Douglas for the training at Douglaslab, Cornell University. I wish to thank Dr. Nana Ankrah, Dr Paul Visendi, Dr. Alysha Booth, Dr. Daniel G. Vassão, Dr Sharon Van Brunschot, Dr Sophie Bouvaine, Dr Moritz Bömer and Dr Habibu Mugerwa for the support in different molecular biology, bioinformatics and metabolic modelling techniques. I am indebted to Mr Noah Clarke of Douglaslab for his support and encouragement. Thanks to Dr Robert Kawuki and Dr Yona Baguma for the lessons I learnt earlier, that formed a foundation for my PhD study. In a special way, I also thank Mr. Dudley Farman for the quick response in the purchase of various material and laboratory reagents. I also remember Evie Keane for the freebies that made student life in the “tower” exciting.

I would like to thank my colleagues Ajith, Annet, Hajar, Jackie, Steve, Antony, Gerald, Hannah, Claudia, Jane, Polo, Sumesh, Siji, Hasting, Joachim, Goshi, Freddy and John for the encouragement and great times in the tower. Thank you Mr Apollo Kabuye and your family for your love, support and all the tours we made – you gave me a glimpse of England. Special thanks to all the members of Chatham Seventhday Adventist church for all the support and encouragement – you made each weekend special.

Lastly, I would like to express my profound thanks to my mother Mrs Catherine Lukyamuzi and my father, Mr Joseph Lukyamuzi for all the care, love and support they have rendered through the years – You spent wisely. To Doreen my wife (Okusembyayo – Kakodyo), I am greatly indebted. Thank you for the support and taking care of the boys. Thank you, Joe, Jay and Josh, for the patience 😊.

DEDICATION

To Doreen my wife and our sons Joel Semanda, Jason Katongole and Joshua Lwasa

**The more we discover the intricate details of nature, the more we marvel.... How
unsearchable are your ways – O God, the creator of the universe**

ABSTRACT

Cassava (*Manihot esculenta* Crantz), as a food security crop in Africa is devastated by many biotic factors which include pests and diseases. Of these, diseases caused by cassava brown streak viruses and cassava mosaic viruses, transmitted by whiteflies, *Bemisia tabaci* are causing the highest yield losses in Sub-Saharan Africa. The aim of this study was to understand the molecular basis of key processes that underpin the capacity of whiteflies to utilize phloem-sap and thereby identify gene targets for novel whitefly control strategies. This study was founded on a principle that phloem-sap feeders require osmoregulation genes to maintain fluid homeostasis and metabolic interaction with endosymbionts for the provisioning of essential amino acids, that are deficient in the phloem-sap.

Guts and bacteriocytes from a cassava *B. tabaci* population (SSA1-SG1) were dissected, their RNA extracted and sequenced to compare transcripts with enriched expression in the gut or bacteriocytes relative to the wholebody. RNAseq analysis revealed a total of 15 genes encoding alpha-glucosidase, family 13 with sugar processing domain (Amy, IPR006047) and enriched expression in the gut. Of these, only one gene (ENSSSA1UGT002057) encoded for a protein that had a specific molecular signature that influence sucrose preference. This gene was selected as it encodes the hydrolysis of sucrose to fructose and glucose. It is, therefore, a suitable gene target against SSA1-SG1. In addition to sugar hydrolysis, *B. tabaci* maintains fluid homeostasis within the gut through water recycling. A water-specific aquaporin encoded by ENSSSA1UGT022145 (Ssa03484), an *AQP1* gene in SSA1-SG1 was identified. This was differentially expressed in the whitefly gut where it is considered to mediate water recycling from the distal to the proximal end of the gut, thereby maintaining fluid homeostasis within the gut.

RNAseq analysis of bacteriocytes dissected from SSA1-SG1, revealed three sets of genes. These were, (i) aminotransferases, (ii) amino acid transporters and (iii) horizontally transferred genes of either bacterial or fungal origin that influence amino acid provisioning in *Bemisia tabaci* SSA1-SG1. In addition to RNAseq analysis, genome-scale metabolic reconstruction unraveled the shared-metabolic interaction between SSA1-SG1 and its primary endosymbiont, *Portiera aleyrodidarum*. Application of constraint based metabolic modeling to this two-compartment model of cassava whitefly (SSA1-SG1) revealed 270 indispensable reactions, which, upon knockdown are predicted to cause devastating effects on both growth and survival of cassava

whitefly. A total of four essential symbiosis genes (*AAT*, *BCAT*, *argH* and *lysA*), both of intrinsic origin or horizontally transferred genes mediating terminal reactions in the phenylalanine, isoleucine, leucine, valine, arginine and lysine biosynthesis pathways were selected and validated. These are proposed as candidate gene targets that can be used for future development of cassava management control options such as RNA interference.

TABLE OF CONTENTS

DECLARATION.....	ii
ACKNOWLEDGEMENTS	iii
DEDICATION.....	iv
ABSTRACT.....	vi
LIST OF TABLES.....	xi
LIST OF FIGURES	xii
ABBREVIATIONS.....	xvi
INTRODUCTION.....	1
LITERATURE REVIEW	8
2.1 Global economic importance of the whitefly, <i>Bemisia tabaci</i> (<i>sensu lato</i>)	8
2.2 Systematics of <i>Bemisia tabaci</i>	9
2.2.1 <i>B. tabaci</i> species associated with African cassava production and cassava diseases.....	13
2.3 Management of <i>Bemisia tabaci</i> populations	15
2.4 RNAi and its application in insect control	17
2.5 Potential essential gene targets for controlling cassava whiteflies	20
2.6 Osmoregulation genes in phloem-sap feeders.....	22
2.6.1 Glycoside hydrolase family 13 enzymes and sucrose assimilation in whitefly	22
2.6.2 Major intrinsic proteins and water recycling in the <i>B. tabaci</i> gut	24
2.7 Symbiosis and amino-acid provisioning in plant phloem-sap feeders	26
2.7.1 Primary and secondary symbionts in phloem-sap feeders	26
2.7.2 Shared metabolic pathways between symbionts and phloem-sap feeders	27
2.7.3 Amino-acid transporters and amino-acid provisioning in phloem-sap feeders.....	30
IDENTIFICATION OF KEY OSMOREGULATION AND SYMBIONT GENES AS TARGETS FOR CONTROLLING CASSAVA WHITEFLIES (<i>BEMISIA TABACI</i>) USING RNA INTERFERENCE.....	32
3.1 INTRODUCTION	32
3.2 MATERIALS AND METHODS.....	34
3.2.1 Dissection of guts and bacteriocytes from cassava <i>B. tabaci</i>	34
3.2.2 RNA extraction from dissected guts and bacteriocytes	35
3.2.3 rRNA depletion and RNAseq library preparation.....	35
3.2.4 RNA sequencing and RNAseq analysis.....	36
3.2.5 Identification of osmoregulation and symbiosis genes	36
3.2.6 Phylogenetic analysis.....	37
3.2.7 Validation of selected genes using real-time quantitative PCR	38

3.3	RESULTS	40
3.3.1	RNAseq analysis for gut and bacteriocytes samples from cassava <i>B. tabaci</i>	40
3.3.2	Identification of transcripts enriched in the cassava <i>B. tabaci</i> (SSA1-SG1) gut and bacteriocytes	40
3.3.3	Gene ontology categorization for the differentially expressed transcripts in <i>B. tabaci</i> gut	45
3.3.4	Categorization of transcripts based on superfamilies and functional domains	49
3.3.5	Identification of genes for sucrose metabolism and water recycling in cassava <i>B. tabaci</i> (SSA1-SG1) gut	52
3.3.6	Structural analysis for residues of the hallmarks of the water-specific aquaporins	55
3.3.7	Identification of α -glucosidase proteins with both the catalytic residue and predicted signal peptide	58
3.3.8	Identification of symbiosis genes in cassava <i>B. tabaci</i> (SSA1-SG1)	63
3.3.9	Gene ontology categorization for the differentially expressed transcripts in cassava <i>B. tabaci</i> (SSA1-SG1) bacteriocyte	68
3.3.10	Identification of potential symbiosis genes targets in cassava <i>B. tabaci</i> (SSA1-SG1)	72
3.3.11	Horizontally transferred genes of bacterial origin expressed in cassava (SSA1-SG1) <i>B. tabaci</i> species	75
3.3.12	Validation of expression of selected genes using real-time quantitative PCR	83
3.4	DISCUSSION	85
IDENTIFICATION OF ESSENTIAL GENES TARGETS IN CASSAVA <i>B. TABACI</i>, THROUGH GENOME-SCALE METABOLIC MODELING		90
4.1	INTRODUCTION	90
4.2	MATERIALS AND METHODS	92
4.2.1	Dissection of bacteriocytes from cassava <i>B. tabaci</i>	92
4.2.2	Identification of metabolic genes for both SSA1-SG1 and <i>Portiera aleyrodidarum</i>	92
4.2.2	Genome-scale reconstruction of <i>Portiera</i> and SSA1-SG1 metabolic models	93
4.2.3	Conversion of the metabolic reconstruction into a mathematical model	95
4.2.4	Analysis and evaluation of the models	96
4.3	RESULTS	97
4.3.1	Assembly and functional annotation of <i>Portiera</i> genome from <i>B. tabaci</i> SSA1-SG1	97
4.3.2	Comparison of <i>Portiera</i> from <i>B. tabaci</i> SSA1-SG1, MEAM1 and MED	100
4.3.3	Single compartment models of SSA1-SG1- <i>Portiera</i> and <i>B. tabaci</i> SSA1-SG1	100
4.3.4	Two-compartment models of <i>B. tabaci</i> SSA1-SG1 and <i>Portiera aleyrodidarum</i>	102
4.3.5	<i>In silico</i> prediction of essential amino acid synthesis rate in <i>B. tabaci</i> SSA1-SG1	110
4.3.6	Essential gene targets in different essential amino acid biosynthesis pathways in SSA1-SG1	114

4.4	DISCUSSION	118
EFFICACY OF RNA INTERFERENCE AGAINST OSMOREGULATION GENES IN THE MANAGEMENT OF <i>BEMISIA TABACI</i>		122
5.1	INTRODUCTION	122
5.2	MATERIALS AND METHODS	124
5.2.1	Plant material and insects.....	124
5.2.2	Evaluating <i>B. tabaci</i> performance on tomato transgenic plants.....	124
5.2.3	Analysis of <i>SUC1</i> and <i>AQP1</i> gene expression in MEAM1 <i>B. tabaci</i>	125
5.2.4	Honeydew collection.....	126
5.2.5	Analysis of sugars in honeydew samples from whiteflies	127
5.2.6	Quantification of higher sugars in honeydew samples.....	128
5.2.7	Determination of osmotic pressure of <i>B. tabaci</i> hemolymph.....	129
5.3	RESULTS	130
5.3.1	Sucrose hydrolysis in whiteflies feeding on different sources of sucrose.....	130
5.3.2	Comparison of sucrose hydrolysis in aphids and whiteflies	132
5.3.3	<i>B. tabaci</i> performance on different transgenic tomato lines	133
5.3.4	Analysis of <i>SUC1</i> and <i>AQP1</i> gene expression in <i>B. tabaci</i> feeding on different transgenic tomato plants.....	137
		140
5.3.5	Effect of RNA interference on <i>B. tabaci</i> development.....	140
5.3.6	Effect of RNAi against <i>SUC1</i> and <i>AQP1</i> genes on <i>B. tabaci</i> osmotic pressure	142
5.3.7	Sucrose hydrolysis in whiteflies feeding on different transgenic tomato lines.....	145
5.3.8	Effect of RNA interference on sugar composition in <i>B. tabaci</i> honeydew	147
5.3.9	Effect of RNAi on the concentration of different sugar categories in honeydew samples	148
5.3.10	Comparison of diet sugar utilization of whiteflies and aphids.....	149
5.4	DISCUSSION	154
GENERAL DISCUSSION AND CONCLUSIONS.....		160
APPENDIX.....		168
REFERENCES.....		184

LIST OF TABLES

Table 3.1: Primers used in the validation of gene expression of selected genes	39
Table 3.2: Mapping quality of RNAseq reads from SSA1-SG1 whole-body, bacteriocytes and guts samples.....	40
Table 3.3: Summary statistics of RNAseq analysis for genes expressed in <i>B. tabaci</i> gut and bacteriocytes	41
Table 3.4: Candidate osmoregulation genes selected from differentially expressed genes in cassava <i>B. tabaci</i> (SSA1-SG1) gut	53
Table 3.5: NPA motif and ar/R filter of selected aquaporin protein sequences of cassava <i>B. tabaci</i> (SSA1-SG1).....	56
Table 3.6: Genes encoding proteins with characteristic catalytic site residue and a signal peptide for α -glucosidase enzyme	60
Table 3.7: Potential gene targets selected from differentially expressed genes in cassava <i>B. tabaci</i> bacteriocytes	74
Table 3.8: Horizontally transferred genes of bacterial origin expressed in cassava <i>B. tabaci</i> (SSA1-SG1) bacteriocytes	77
Table 3.9: Horizontally transferred genes of fungal origin expressed in cassava <i>B. tabaci</i> (SSA1-SG1) bacteriocytes	80
 Table 4.1: Comparison of genomic features of <i>Portiera</i> genome from MEAM1, MED and SSA1-SG1..	97
Table 4.2: Missing protein/function in <i>Portiera</i> genome from MEAM1, MED and SSA1-SG1	100
Table 4.3: Comparison of genome-scale metabolic models of <i>B. tabaci</i> SSA1-SG1 and MEAM1	101
Table 4.4: Accessory essential genes mediating critical reactions of essential amino acid biosynthesis in <i>B. tabaci</i> (SSA1-SG1) bacteriocytes simulated using FBA, MOMA and ROOM	116
Table 4.5: Essentiality of genes mediating synthesis of metabolic inputs in the synthesis of essential amino acids	117
Table 4.6: List of essential reactions in a two-compartment metabolic model of SSA1-SG1	179
 Table 5.1: Standards required for glucose assay	128

LIST OF FIGURES

Figure 2.1: Evolutionary relationship of <i>B. tabaci</i> based on Bayesian analysis	11
Figure 2.2: Mechanism of RNAi in the cell.....	19
Figure 2.3: Structure of water-specific aquaporin.	25
Figure 3.1: Dissected bacteriocytes and gut from the abdomen of <i>B. tabaci</i> (SSA1-SG1)	34
Figure 3.2: Differential gene expression in cassava <i>B. tabaci</i> (SSA1-SG1) gut.....	41
Figure 3.3: Differential gene expression in cassava <i>B. tabaci</i> (SSA1-SG1) gut.....	42
Figure 3.4: Tag distribution for the transcripts enriched in the cassava <i>B. tabaci</i> (SSA1-SG1) gut	43
Figure 3.5: Species distribution of all the NCBI top blast-hits for the transcripts enriched in cassava <i>B. tabaci</i> (SSA1-SG1) gut	44
Figure 3.6: Gene ontology (GO) categorization for biological process, molecular function and cellular component of the transcripts enriched in the cassava <i>B. tabaci</i> (SSA1-SG1) gut.	46
Figure 3.7: Direct GO count for biological process of transcripts enriched in cassava <i>B. tabaci</i> gut..	47
Figure 3.8: Gene ontology (GO) categorization for Biological process of the transcripts enriched in the cassava <i>B. tabaci</i> (SSA1-SG1) gut..	48
Figure 3.9: InterProScan family distribution of the transcripts enriched in the cassava <i>B. tabaci</i> (SSA1-SG1) gut.....	50
Figure 3.10: InterProScan domain distribution of the transcripts enriched in the cassava <i>B. tabaci</i> (SSA1-SG1) gut.....	51
Figure 3.11: Phylogenetic relationship of all Aquaporin genes of cassava <i>B. tabaci</i> (SSA1-SG1) and experimentally validated Aquaporins	57
Figure 3.12: Multiple sequence alignment of selected α -glucosidase protein sequences.....	59
Figure 3.13: Phylogenetic relationships of α -glucosidases with both predicted signal peptides and catalytic site residue and experimentally validated <i>Sucrase</i> gene in <i>Acyrtosiphon pisum</i>	62
Figure 3.14: Differential gene expression in cassava <i>B. tabaci</i> (SSA1-SG1) gut. Volcano plot showing genes enriched in the bacteriocyte.	64
Figure 3.15: Differential gene expression in cassava <i>B. tabaci</i> (SSA1-SG1) gut.....	65
Figure 3.16: Tag distribution for the transcripts enriched in the cassava <i>B. tabaci</i> (SSA1-SG1) bacteriocyte.....	66
Figure 3.17: Top-hit species distribution for the differentially expressed transcripts in cassava <i>B. tabaci</i> (SSA1-SG1) bacteriocyte.....	67

Figure 3.18: Gene ontology (GO) categorization for the transcripts enriched in the cassava <i>B. tabaci</i> (SSA1-SG1) bacteriocyte.....	69
Figure 3.19: Sequence distribution of transcripts enriched in the cassava <i>B. tabaci</i> (SSA1-SG1) bacteriocyte and the biological processes they mediate.....	70
Figure 3.20: Direct gene ontology count for transcripts enriched in the cassava <i>B. tabaci</i> (SSA-SG1) bacteriocyte.....	71
Figure 3.21: Quantitative real-time PCR gene expression of selected genes in the cassava whitefly (SSA1-SG1) gut and bacteriocytes relative to whole-body.....	84
Figure 4.1: Sequence distribution for biological processes encoded by genes in <i>Portiera aleyrodidarum</i> genome from <i>B. tabaci</i> SSA1-SG1	98
Figure 4.2: Direct GO counts for biological processes encoded by genes in <i>Portiera aleyrodidarum</i> genome from <i>B. tabaci</i> SSA1-SG1	99
Figure 4.3: Identification of secondary symbionts co-existing with <i>Portiera aleyrodidarum</i> in the cassava <i>B. tabaci</i> (SSA1-SG1) bacteriocyte.....	103
Figure 4.4: Phylogenetic relationship of the 16S V ₆ -V ₇ region of selected <i>B. tabaci</i> endosymbionts. ...	104
Figure 4.5: Biosynthetic pathway for two essential amino acids (i) lysine and (ii) Arginine.....	106
Figure 4.6: Biosynthesis pathway for four essential amino acids synthesis pathway (i) L-Leucine, (ii) L-Valine, (iii) L-Threonine and (iv) L-Isoleucine	107
Figure 4.7: Biosynthetic pathway for histidine and methionine	108
Figure 4.8: Biosynthesis pathway for two essential amino acids. (i) Phenylalanine and (ii) Tryptophan.	109
Figure 4.9: In silico prediction of production rates of essential amino acid of the two-compartment metabolic model of <i>B. tabaci</i> SSA1-SG1 and <i>Portiera</i>	111
Figure 4.10: Robustness analysis measuring sensitivity of the metabolic object function (growth rate of <i>B. tabaci</i> SSA1-SG1) to the quantitative flux levels through terminal reactions for (A) arginine, (B) lysine, (C) histidine, (D) leucine, (E) isoleucine and (F) valine.	112
Figure 4.11: Robustness analysis measuring sensitivity of the metabolic object function (growth rate of <i>B. tabaci</i> SSA1-SG1) to the quantitative flux levels through terminal reactions for (A) phenylalanine (B) methionine (C) threonine and (D) tryptophan.....	113
Figure 4.12: Robustness of the metabolic network “SSA1-SG1_ <i>Portiera</i> ” iKT420 showing the effect of single reaction deletion on the growth rate of <i>B. tabaci</i> SSA1-SG1 and <i>Portiera</i> mutant relative to the wild-type, simulated using flux balance analysis.....	115

Figure 5.1: Experimental setup of the efficacy study consisting of a single row non-replicated plot with five tomato plants for each treatment.....	125
Figure 5.2: Glucose standard curve used to calculate the amount of glucose in different honeydew samples.....	131
Figure 5.3: Sucrose hydrolysis in whiteflies feeding on different sucrose sources	131
Figure 5.4: Comparison of sucrose hydrolysis in the aphid <i>Aphis gossypii</i> and the whitefly <i>B. tabaci</i> MEAM1 feeding on both tobacco and artificial diet with 0.75 M sucrose	133
Figure 5.5: Percentage mortality of whiteflies feeding on different tomato plants with dsRNA against <i>RNase</i> , <i>GFP</i> , <i>AQP1</i> and <i>SUC1</i> in the first experiment.	134
Figure 5.6: Clip cages used in the experiment.	135
Figure 5.7: Percentage mortality of whiteflies <i>B. tabaci</i> MEAM1 feeding on different transgenic tomato plants with dsRNA against <i>RNase</i> , <i>GFP</i> , <i>AQP1</i> and <i>SUC1</i> after six days in the second experiment.	136
Figure 5.8: Relative gene expression of <i>AQP1</i> genes in <i>B. tabaci</i> MEAM1 that survived after six days of feeding on different stably transformed tomato plants in the first experiment.	137
Figure 5.9: Relative gene expression of <i>SUC1</i> genes in <i>B. tabaci</i> MEAM1 that survived after six days of feeding on different stably transformed tomato plants in the first experiment.	138
Figure 5.10: Gene expression of <i>AQP1</i> genes in <i>B. tabaci</i> MEAM1 that survived after six days of feeding on different stably transformed tomato plants in the second experiment.	139
Figure 5.11: Gene expression of <i>SUC1</i> genes in <i>B. tabaci</i> MEAM1 that survived after six days of feeding on different stably transformed tomato plants in the second experiment.	140
Figure 5.12: Effect of RNA interference on the development of <i>B. tabaci</i> MEAM1 feeding on different transgenic tomato plants in the first experiment.	142
Figure 5.13: Standard curve used to estimate haemolymph osmotic pressure for each analyzed <i>B. tabaci</i> sample	143
Figure 5.14: Haemolymph osmotic pressure for <i>B. tabaci</i> feeding on different stably transformed tomato plants.....	144
Figure 5.15: Sucrose hydrolysis as a physiological change resulting from whiteflies feeding on different stably transformed tomato plants containing dsRNA against <i>SUC1</i>	146
Figure 5.16: The molar ratios of different sugars in honeydew samples collected from whiteflies feeding on different transgenic tomato plants.....	148
Figure 5.17: Sugar concentration (hexose units per mg of honeydew) collected from whiteflies feeding on different transgenic tomato plants.....	149

Figure 5.18: Comparison of molar ratios of sugars in honeydew samples collected from aphids and whiteflies feeding on tobacco and artificial diet..	152
Figure 5.19: Comparison of sugar concentration in different honeydew samples collected from aphids and whiteflies feeding on both tobacco and artificial diet..	153

ABBREVIATIONS

Aa_trans	Amino acid transporter, transmembrane domain
AA_Permcase_2	Amino acid/Polyamine transporter 1
AA_Permcase_C	Cationic amino acid transporter, C-terminal
Aminotran_1_2	Aminotransferase, class I/II
Aminotran_5	Aminotransferase class V domain
Aminotrans_IV	Aminotransferase class IV
<i>AQP1</i>	Aquaporin 1 gene
ar/R	Aromatic/arginine region
Asp_trans	Aspartate aminotransferase
BIB	Big brain protein
bp	Base pair
CBSD	Cassava brown streak disease
CBSV	<i>Cassava brown streak virus</i>
CMD	Cassava mosaic disease
CMGs	Cassava mosaic geminiviruses
DRIP	Drosophila integral protein
dsRNA	Double-stranded RNA
EAA	Essential amino acids
EC	Enzyme Commission
Eglps	Entomoglyceroporin
FBA	Flux balance analysis
FVA	Flux variability analysis
<i>GFP</i>	Green fluorescent protein
GH-13	Glycosyl hydrolase family 13
HTG	Horizontally transferred genes
M	Molar
MEAM1	Middle East-Asia Minor I
MED	Mediterranean <i>B. tabaci</i> species
<i>mtCOI</i>	Mitochondrial Cytochrome C Oxidase I
NCBI	National Center for Biotechnology Information
PAR	Photo active radiation
POT_fam	Proton-dependent oligopeptide transporter family
PTR2	Proton-dependent oligopeptide transporter
qPCR	quantitative polymerase chain reaction
RAPD	Random Amplified polymorphic DNA
RdRP	RNA-dependent RNA polymerase
RNases	Ribonucleases
RISC	RNA-Induced Silencing Complex
RNAi	RNA interference
RNAseq	RNA sequence
<i>RPL13</i>	60S ribosomal protein L13a
SDF	Sodium:dicarboxylate symporter
SSA1	Sub-Saharan Africa 1
SSA2	Sub-Saharan Africa 2
SSA1-SG1	Sub-Saharan Africa 1-subgroup 1

SSA1-SG2	Sub-Saharan Africa 1-subgroup 2
SSA1-SG3	Sub-Saharan Africa 1-subgroup 3
SSA1-SG4	Sub-Saharan Africa 1-subgroup 4
SSA1-SG5	Sub-Saharan Africa 1-subgroup 5
SID-1	Systemic interference defective-1
siRNA	Small interfering double-stranded RNA
<i>SUC1</i>	Sucrase 1 gene
<i>SUC2</i>	Sucrase 2 gene
UCBSV	<i>Uganda cassava brown streak virus</i>

INTRODUCTION

Cassava (*Manihot esculenta* Crantz), is one of the most important root crops in the tropics (Pujol *et al.*, 2002). It is widely grown in Asia, Latin America and Africa. In Africa, the crop was introduced in the 17th century by Portuguese explorers and traders from Brazil (Olsen, 2004) and since then, its production has continued to grow making it the second most important staple food in Africa after maize (*Zea mays* L.) (Norton, 2014). Its tubers have a dry matter of 30-40% and it is mainly used as a source of carbohydrates, either for human consumption or in starch production (Teye *et al.*, 2014). As a source of human nutrition, it is currently estimated to support over 800 million people in the tropics (Burns *et al.*, 2010). Cassava leaves are also consumed as vegetables in many areas, because they contain high amounts of vitamins (thiamine (B₁), vitamin C, niacin (B₃) and riboflavin (B₂)), minerals (iron, zinc, calcium, potassium) and proteins (USDA¹, 2018). Cassava has a comparative advantage in marginal environments, because it tolerates adverse soil conditions and drought (El-Sharkawy, 2005). Its harvest time also ranges from 8-24 months after planting (Salcedo *et al.*, 2010), which allows for piecemeal harvesting, so it is considered an excellent food security crop. In addition, there has been a high demand for cassava in recent years, especially from China and Thailand (Hersey *et al.*, 2017). This is because cassava starch has a high-energy value, low amylose content and its gluten-free property gives it a comparative advantage over other forms of starch (Witczak *et al.*, 2016). These attributes make cassava a potential cash crop that could increase farm income, thereby reducing poverty levels in the rural and urban areas.

In eastern and southern Africa, the production of cassava as a food security and cash crop has been affected by two on-going and devastating pandemics, cassava mosaic disease (CMD) and cassava brown streak disease (CBSD). CMD can be caused by at least eight different species of geminiviruses and is prevalent in all cassava growing countries in Africa (Legg *et al.*, 2014a; Maruthi *et al.*, 2004b). Typically, CMD is characterized by a chlorotic mosaic distortion in leaf shape, a reduction in leaf size, coupled with general stunting of plant growth (Fauquet and Fargette, 1990). The CMD pandemic that affected cassava production in 11 countries in East and Central Africa in the 1990s (Otim-Nape *et al.*, 2000) was associated with a synergistic interaction between

¹ <https://ndb.nal.usda.gov/ndb/search/list>

geminiviruses (Harrison *et al.*, 1997; Zhang *et al.*, 2001) and an increase in cassava whitefly, *Bemisia tabaci* (*sensu lato*) populations in cassava producing areas (Zhou *et al.*, 1997; Legg, 1999; Zhang *et al.*, 2000; Colvin *et al.*, 2004; Thresh and Cooter, 2005). A total yield loss of US\$1.9-2.7 billion annually was estimated for all affected countries (Scholthof *et al.*, 2011).

Since 2000, the re-emergence of CBSD in Uganda (Alicai *et al.*, 2007) has aggravated the problem. CBSD has been reported amongst the most serious threats to world food security (Pennisi, 2010). This disease is caused by viruses classified into either of two ipomovirus species, namely *Cassava brown streak virus* (CBSV) and *Uganda cassava brown streak viruses* (UCBSV) that cause up to 100% yield loss in susceptible cassava genotypes (Winter *et al.*, 2010). More virus species causing CBSD have been recently reported (Ndunguru *et al.*, 2015b). In the early 2000s, it was estimated that CBSD caused annual yield losses of ~US \$6-7 million in Malawi (Gondwe *et al.*, 2003) and US\$16.5 million in Tanzania (Kanju *et al.*, 2007). More recently, a total of US\$75 million loss was estimated for all eight countries of East and Central Africa affected by CBSD (Manyong *et al.*, 2012). It has been predicted, with the increasing disease pressure due to the emergence and spread of UCBSV, that this estimate is likely to rise (Ndunguru *et al.*, 2015a).

Both CMVs and (U)CBSVs are transmitted by whiteflies within the *B. tabaci* species complex (Maruthi *et al.*, 2005; Legg *et al.*, 2011). *B. tabaci* is a cryptic species complex made up of more than 39 morphologically indistinguishable species (Mugerwa *et al.*, 2018). These species have been distinguished most recently, based on a molecular marker consisting of the partial fragment of the mitochondrial cytochrome C oxidase I (mtCO1) gene and using a 3.5% species divergence boundary (Dinsdale *et al.*, 2010). Of these putative *B. tabaci* species, Sub-Saharan Africa 1 (SSA1), Sub-Saharan Africa 2 (SSA2) and Sub-Saharan Africa 10 (SSA10) utilize cassava in East Africa (Mugerwa *et al.*, 2018).

Transect collection of whiteflies occurring 'ahead', 'at' and 'behind' of the CMD epidemic front that devastated cassava production in Uganda in the 1990s, revealed two whiteflies species SSA2 ('Invasive', Ug2) and SSA1 ('Local', Ug1) (Legg *et al.*, 2002). These two whitefly species were found to be sympatric on cassava and were reported to interbreed, producing fertile F1 offspring (Maruthi *et al.*, 2004a). Recent surveys have revealed that one species (SSA1-SG1) in particular has increased in abundance and is currently prevalent in most East African cassava growing areas affected by CMD and CBSD (Legg *et al.*, 2014c; Mugerwa *et al.*, 2019).

Colvin *et al.*, (2006) showed that the spread of both viral diseases and whiteflies is a complex problem, because the whitefly population increased significantly more rapidly on virus-infected cassava plants, compared to healthy plants. This was attributed to an increase in amino acids such as tryptophan, glutamine, tyrosine and asparagine in the phloem-sap of the diseased plants. These discoveries show clearly that efforts to control cassava viral diseases and their whitefly vectors need to be co-ordinated, because simply focusing on one aspect of the problem is unlikely to provide a sustainable solution. In comparison to efforts made to develop CMD-resistant cassava varieties, however, only limited efforts have been made to combat cassava whiteflies in sub-Saharan Africa. Indeed, many 'elite' CMD-resistant cassava genotypes are highly susceptible to cassava whiteflies (Omongo *et al.*, 2012), which has contributed to the "superabundant" cassava whitefly phenomenon in East Africa (Legg *et al.*, 2014b; Legg *et al.*, 2014c).

Since the early 1990s, several consortia have been formed and a number of projects funded to manage whitefly problems in different cropping systems (Oliveira *et al.*, 2001). In late the 1990s, a consortium composed of scientists from the International Center for Tropical Agriculture (CIAT), the *International Institute of Tropical Agriculture* (IITA), the World Vegetable Center (AVRDC), the International Potato Center (CIP) and International Centre of Insect Physiology and Ecology (ICIPE) was formed to understand whiteflies and whitefly-borne viruses in the tropics (Anderson *et al.*, 2005). Through this collaboration, the Tropical Whitefly IPM project (TWP) studies were conducted on whiteflies in different cropping systems in eastern and southern Africa, Central America, Mexico, Caribbean and Latin America (Morales, 2007). Its major objective was to build a knowledge base on whiteflies, for subsequent whitefly and whitefly-borne virus disease management. These and other efforts contributed to the management of CMD pandemic in East Africa in the 1990s (Otim-Nape *et al.*, 2000). Due to the more recent emergence of CBSD (Alicai *et al.*, 2007) and the increasing whitefly populations in sub-Saharan Africa (Legg *et al.*, 2014c), various studies were initiated under the African cassava whitefly project². The goal of these studies was/is to understand: (i) African whiteflies systematics, (ii) factors driving cassava whitefly outbreaks and superabundance, (iii) the genetic basis of whitefly resistance in cassava, (iv) African cassava whitefly genetics. Insights from these studies, especially on whitefly biology and interaction between whiteflies and other organisms in the ecosystem will facilitate the

² <http://www.cassavawhitefly.org/>

development of efficient management strategies for whiteflies, not only for those that utilize cassava but also for whiteflies in different cropping systems.

Several methods have been applied to manage whitefly population in different cropping systems. These include: (i) use of pesticides (Houndete *et al.*, 2010), (ii) cultural practices (Hilje *et al.*, 2001), (iii) application of biological control methods (Otoidobiga *et al.*, 2003) and (iv) host-plant resistance (Ariyo *et al.*, 2005). The “super-abundant” cassava whitefly populations in sub-Saharan Africa may require additional novel methods to provide a sustainable solution. Post-transcriptional control mechanisms like RNA interference (RNAi) have been used in the control of some pests in different crop production systems. For example, RNAi has been used successfully in the down-regulation of specific gene expression in Western corn rootworm (*Diabrotica virgifera virgifera* Le Conte) (Baum *et al.*, 2007), bollworm (*Helicoverpa armigera* Hübner) (Mao *et al.*, 2011), aphids (*Myzus persicae* Sulzer) (Pitino *et al.*, 2011) and whitefly (*B.tabaci*, MEAM1) (Thakur *et al.*, 2014). It is worth noting that these studies were proof-of-concept and therefore detailed studies are required to evaluate the efficacy of using RNAi in the management of whiteflies under field conditions. The success of RNAi as a control strategy depends on the development of a double-stranded RNA (dsRNA) construct, systemic delivery of dsRNA within the body of the target pest and identification and selection of suitable gene targets (Price and Gatehouse, 2008).

The search for accessory essential genes in *B. tabaci* is of paramount importance especially as new methods like RNAi are being developed for whitefly management. The term “accessory” used here refers to critical genes required in specific biological systems, or for the survival of specific cell types (Juhas *et al.*, 2012). Whitefly species feed solely on plant phloem-sap throughout their life cycle, so this presents an opportunity to identify accessory essential genes for whitefly management (Cohen, 2013). Phloem-sap poses two major nutritional problems that whiteflies have to overcome, described as the “nitrogen and sugar barriers” (Douglas, 2006). The nitrogen barrier to phloem-sap utilization relates to its low nitrogen quality, consisting of a ratio of essential amino acid to non-essential amino acid of 1:4-1:20, compared to 1:1 in animal protein. This means, therefore, that unless this deficiency is rectified, the growth and fecundity of phloem-sap feeding insects is greatly limited. An additional obstacle presented by a diet of plant phloem-sap is ‘the sugar barrier’. The high sugar concentration in phloem sap, of up to and often exceeding 1 M

sucrose and a resultant osmotic pressure 2-5 times greater than the osmotic pressure of the insect body (*ibid*), results in water transfer from the body fluids to the gut content resulting in the osmotic collapse of the insect (Jing *et al.*, 2016).

Previous studies conducted on other phloem-sap feeders showed that these insects possess physiological and molecular mechanisms that enable them to overcome the sugar barrier (Ashford *et al.*, 2000; Karley *et al.*, 2005; Jing *et al.*, 2016). These studies showed that hemipteran insects possess sugar processing enzymes for example α -glucosidase that transform ingested sucrose to constituent monosaccharides (fructose and glucose) and through transglucosidation, to oligosaccharides like trehalulose, making sap movement within the insect possible without causing shriveling or dehydration. Furthermore, phloem-sap feeders have both genes and structural adaptations that allow water recycling from the distal to the proximal end of the midgut to ensure fluid homeostasis (Cristofolletti *et al.*, 2003; Shakesby *et al.*, 2009; Mathew *et al.*, 2011). These two sets of genes (that mediate sugar processing and water recycling) are highly enriched in the gut, so are henceforth referred to as osmoregulation genes.

The nitrogen barrier, on the other hand, is overcome through shared metabolic interactions with endosymbionts. Phloem-sap feeders possess both primary and secondary endosymbionts, that co-operate metabolically with the host to produce essential nutrients and amino acids that are essential for survival, growth and reproduction (Gunduz & Douglas, 2009). This cooperative metabolism has been revealed in aphids (Hansen & Moran, 2011) and whiteflies (MEAM1) (Luan *et al.*, 2015; Ankrah *et al.*, 2017). Genomic insights into essential amino acid biosynthesis revealed that there are critical points in this system that can be targeted for the management of these insect pests. First, the genes that are required for nutrient exchange across the symbiosomal membrane (Douglas, 2014), second, the horizontally acquired genes whose function lies in essential amino-acid biosynthesis pathways (Husnik *et al.*, 2013; Luan *et al.*, 2015) and genes that deemed essential after perturbing a genome-scale metabolic model. These genes, henceforth called symbiosis genes, are required for provisioning or biosynthesis of essential nutrients and amino acids within the whitefly. Symbiosis genes as used in this thesis refer to genes that encode for different proteins that are involved in the essential amino acid biosynthesis and shared metabolic interaction between the whitefly and endosymbionts in the bacteriocytes. Advances in molecular techniques like genome-scale metabolic reconstruction coupled with constraint-based modeling offer

opportunities to explore the interaction between whitefly and its endosymbionts and to identify essential gene targets for pest management. In the recent past, a three-compartment model for *Bemisia tabaci* MEAM1, and its bacterial symbionts *Portiera aleyrodidarum* and *Hamiltonella defense* was reconstructed (Ankrah *et al.*, 2017). It cannot be assumed that other species within the *B. tabaci* species complex will have exactly the same biochemical pathways, so these processes also merit investigation in other whitefly species, to develop specific and effective management options for different whitefly species. This study therefore was initiated to identify critical gene targets in cassava whiteflies that can be used as RNAi targets for cassava whitefly management. The research questions addressed in the following chapters are:

1. What are the key osmoregulation genes that encode enzymes that hydrolyze sucrose in the cassava whitefly (SSA1-SG1) gut? (Chapter 3)
2. What are the key symbiosis genes that are enriched in the cassava whitefly (SSA1-SG1) bacteriocyte? (Chapter 3)

Identification of essential osmoregulation and symbiosis gene targets was achieved by determining differential expression in the whitefly gut and bacteriocytes compared to the whole-insect body through RNAseq analysis. Bacteriocytes are special cells where the primary endosymbiont resides, a site of essential amino acid biosynthesis in whiteflies. Therefore, gene(s) enriched in the gut and bacteriocytes were selected and analyzed to determine their function. Genes selected were based on their unique features, such as domains and motifs that are related to either osmoregulation function or essential amino-acid biosynthesis or transport.

3. What are the essential metabolic genes of cassava *B. tabaci* (SSA1-SG1) that co-operate with the primary endosymbiont? (Chapter 4)

Genome-scale metabolic reconstruction coupled with constraint-based modeling was applied to identify key gene targets in essential amino acid biosynthesis and provisioning. Potential target genes include genes that encode for amino acid transporters, horizontally transferred genes and proteins that interact with intermediate precursors from endosymbionts for essential amino-acid biosynthesis in the whitefly bacteriocytes. The metabolic model was constrained to identify genes (essential genes), which limit growth when they are perturbed.

4. What is the efficacy of RNA interference for the management of MEAM1 whiteflies? (Chapter 5)

Efficacy of RNAi was studied by evaluating the performance of MEAM1 whiteflies on stably transformed tomato plants containing double-stranded RNA against aquaporin and the α -glucosidase (Sucrase) gene of MEAM1 *B. tabaci*. The tomato plants were transformed based on findings by Luo *et al.* (2017). The specific aim of this activity was to gain insights into how to develop an effective RNAi system to manage cassava whiteflies.

LITERATURE REVIEW

2.1 Global economic importance of the whitefly, *Bemisia tabaci* (*sensu lato*)

Members of *B. tabaci* species complex (Hemiptera: Sternorrhyncha: Aleyrodidae: Aleyrodidae) rose to international prominence, due to their impact on agriculture as pests and vectors of plant viruses (Jones, 2003). Crop damage caused by *B. tabaci* species, either directly or indirectly, has been estimated in billions of USA dollars. For example, in areas affected by cassava mosaic virus diseases, total annual losses have been estimated at US\$1.9-2.7 billion. (Scholthof *et al.*, 2011). No study has yet been done to compile the estimates of damage caused by *B. tabaci* on crops globally, but the overall damage is enormous, given that some *B. tabaci* species such as MEAM1 have a wide host range of more 48 plant families (Malka *et al.*, 2018). The plant families affected most by *B. tabaci* include the *Solanaceae*, *Malvaceae*, *Convolvulaceae*, *Brassicaceae*, *Asteraceae*, *Cucurbitaceae*, *Euphorbiaceae* and *Fabaceae* (Hillocks *et al.*, 2016). In tropical and sub-tropical countries, *B. tabaci* affects major food and cash crops, such as cassava (*Manihot esculenta*), sweet potato (*Ipomea batatas* L) tomato (*Lycopersicon esculentum* M), eggplant (*Solanum melongena* L), peppers (*Capsicum spp*), cotton (*Gossypium hirsutum* L) and tobacco (*Nicotiana tabacum* L) (Anderson *et al.*, 2005).

Whiteflies damage plants both directly and indirectly. Direct feeding causes a reduction in plant vigour, leaf chlorosis, general plant stunting and in some crops, it causes phytotoxic disorders like leaf silverying in cucurbits (Brown & Bird, 1992). When at high populations, whiteflies produce significant honeydew that acts as a substrate for fungal and sooty mold growth on the leaf surface, which reduces photosynthetic processes within the plant (Omongo *et al.*, 2012). In addition, members of the *B. tabaci* species complex collectively transmits more than 110 plant viruses, some of which cause crop diseases that have been listed in the top 10 most economically damaging globally (Brown, 1994; Jones, 2003; Lowe *et al.*, 2004; Maruthi *et al.*, 2005; Scholthof *et al.*, 2011). These include *Cotton leaf crumple virus*, *Squash leaf curl virus*, *African cotton leaf curl virus*, cassava mosaic begomoviruses and cassava brown streak ipomoviruses.

In East Africa, whitefly contributed to the cassava mosaic disease pandemic in the 1990s, which caused devastating effects in cassava growing areas (Colvin *et al.*, 2004) and to the more recent spread of cassava brown streak disease (Pati *et al.*, 2015).

2.2 Systematics of *Bemisia tabaci*

The identification, classification and naming of whitefly populations within the *B. tabaci* species complex has changed gradually since the first description of *Aleyrodes tabaci* by Panayiotis Gennadius (Gennadius, 1889). *B. tabaci* systematics has been influenced by increasing knowledge about the pest, aided by the development of robust diagnostic tools. Whitefly taxonomy was traditionally based on morphological characters of the 4th instar (Gill, 1992). However, these were later found to be inconsistent and plastic (Mound, 1963), so were unsuitable for distinguishing or classifying *B. tabaci* species (Rosell *et al.*, 1997). Mohanty and Basu (1986) also reported that morphological characters, especially of the 4th instar, were highly affected by the environment and physical factors such as humidity and temperature.

In the late 1950s, the biotype or host-race naming system for *B. tabaci* whitefly populations was initiated, naming whitefly populations based on the specific adaptation(s) to host plants. Bird (1957) conducted a study on the whitefly-transmitted virus of *Jatropha gossypifolia* L and observed that one whitefly population had a specific adaptation and transmitted virus exclusively to and from *J. gossypifolia*. This population, therefore, was named the *Jatropha* biotype. Another study by Burban *et al.* (1992) classified two whitefly populations as “cassava biotype” and “okra biotype” based on specific adaptations to different hosts. This evidence of differential host utilization also revealed a significant underlying genetic structure within the *B. tabaci* group. The genetic structure within *B. tabaci* was later confirmed by studying 18 *B. tabaci* populations collected from different geographic locations (Bedford *et al.*, 1994).

In addition to host-range utilization, other biological characteristics were considered and included in the subsequent biotype delimitation studies. These included: (i) insecticide resistance (Byrne *et al.*, 2000; Denholm *et al.*, 2008), (ii) capacity to transmit viruses especially begomoviruses (Simon *et al.*, 2003), (iii) capacity to induce physiological changes like silver leafing (Costa & Brown, 1991) and (iv) esterase profiles (Brown and Coats, 1992). These studies led to the identification of 36 biotypes, namely: Cv, Cassava, A, G (India), Sida, AN, D, Okra, T, B, ZHJ2, S, G (Guatemala), B2, PCG-2, Q, BR, SY, C, T, F, E, ZHJ1, PCG-1, NA, P,M, *Jatropha* K, H, L, J, PK1, L (De Barro *et al.*, 2011).

The naming of *B. tabaci* species based on morphology and allozymic characters initiated a debate as to whether *B. tabaci* was a group of cryptic species or a species complex (Bellows *et al.*, 1994; Boykin, 2014). To answer this question, different molecular markers were developed to aid species delimitation. These included, 18S rDNA (Campbell *et al.*, 1993), random amplification of polymorphic DNA (RAPDs) (DeBarro & Driver, 1997), *mtCOI* (Frohlich *et al.*, 1999), cleaved amplified polymorphic sequences (CAPs) (Khasdan *et al.*, 2005), PCR-RFLP (Rabello *et al.*, 2008) and cDNA AFLP (Guo *et al.*, 2012). Of these, partial *mtCOI* gene-markers were used to reconstruct the first phylogeography of *B. tabaci*, proving that genetic, biochemical and behavioral polymorphism existed within *B. tabaci* and therefore that *B. tabaci* was a group of highly cryptic sibling species (Frohlich *et al.*, 1999).

Acceptance of the evidence that *B. tabaci* comprises multiple cryptic species has been slow (Frohlich *et al.*, 1999). Several additional phylogenetic studies have since been conducted, however, to study the genetic structure of the most economically damaging biotypes of *B. tabaci* (Abdullahi *et al.*, 2003; Viscarret *et al.*, 2003; Maruthi *et al.*, 2004; Rekha *et al.*, 2005; De la Rua *et al.*, 2006). These studies used either parsimony or neighbor-joining methods, which could not resolve the global phylogenetic relationships between the biotypes of *B. tabaci*. As a result, Boykin *et al.* (2007) used Bayesian methods to determine the global relationship of *B. tabaci* based on 366 partial *mtCOI* sequences. This revealed 12 major *B. tabaci* genetic groups, which were: *B. tabaci* (Italy), *B. tabaci* (Mediterranean), *B. tabaci* (Mediterranean/Asia Minor/Africa), *B. tabaci* (New World), *B. tabaci* (Sub-Saharan Africa non-silver leafing), *B. tabaci* (Uganda sweet potato), *B. tabaci* (China), *B. tabaci* (Australia), *B. tabaci* (Asia 1), *B. tabaci* (Indian Ocean), *B. tabaci* (Sub-Saharan Africa silver leaf) and *B. tabaci* (Asia II). Dinsdale *et al.* (2010) subsequently refined the global analysis of *B. tabaci*, also based on partial *mtCOI* sequences, by using quantifiable bounds to describe the genetic structure of the *B. tabaci* group. This provided a phylogenetic framework with the assignment of putative species. Based on a 3.5% genetic species boundary, a total of 24 putative cryptic species were proposed. These included: Mediterranean (Q, J, L, Sub-Saharan Africa silver leaf), New World (A, C, D, F, Jatropha, N, R, Sida), Asian II-2 (ZHJ1), Asian II-7 (Cv), Asian II- 8, China 2, Asian II 1 (K, P, ZHJ2), Uganda, Asia 1 (H, M, NA), Australia/Indonesia, Australia (AN), China 1 (ZHJ3), Indian Ocean (MS), Middle East Asia Minor (B₁, B₂), Middle East-Asia Minor 2, Asia II-4, Asia II-5 (G), Asia II 6, Italy (T), Sub-Saharan Africa 1, Sub-Saharan Africa 2 (S), Sub-Saharan Africa 3 and Sub-Saharan Africa 4 (Figure 2.1).

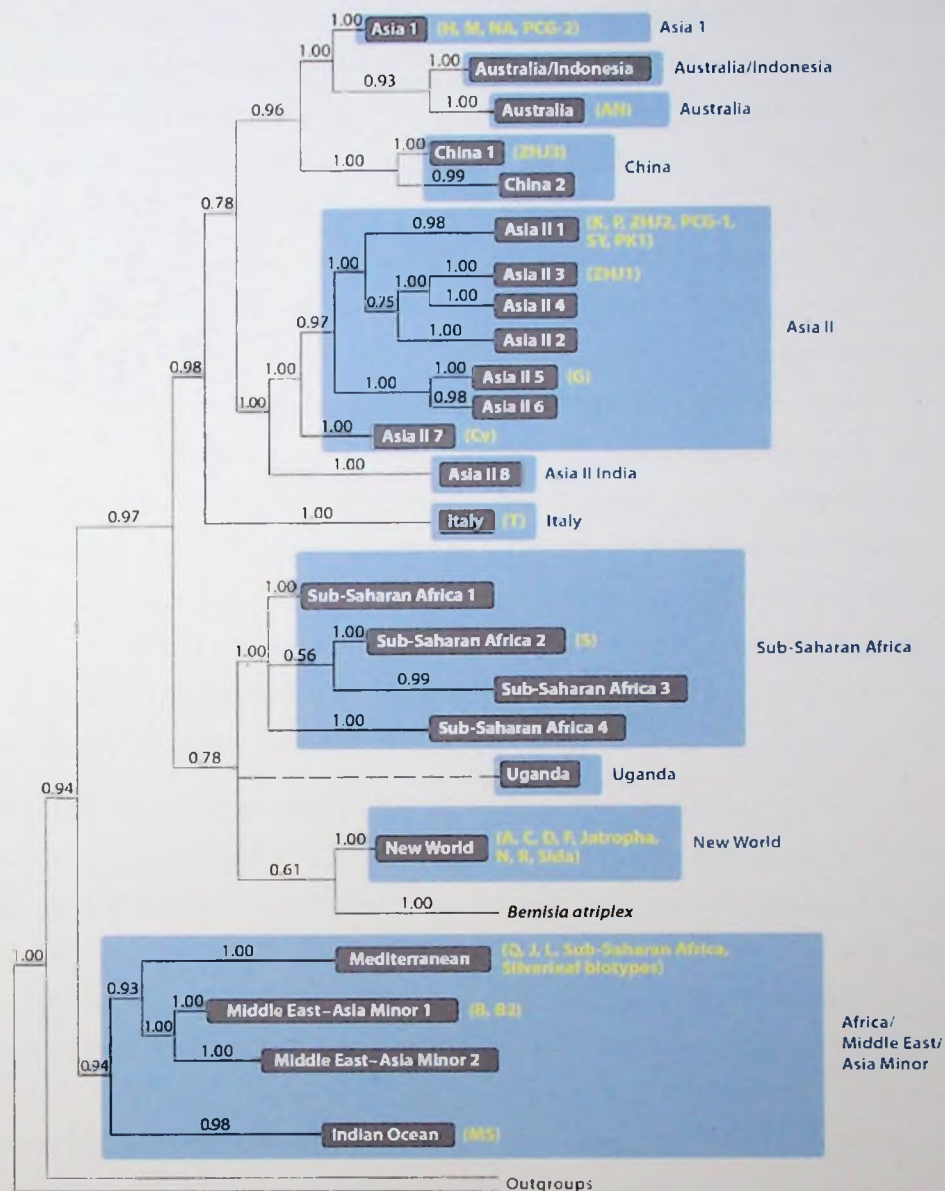


Figure 2.1: Evolutionary relationship of *B. tabaci* based on Bayesian analysis (Dinsdale *et al.*, 2010). Black boxes indicate 24 putative species-level groups based on 3.5% sequence divergence; blue boxes represents 11 high-level (genetic groups) separated by 11% sequence divergence. Biotypes are presented in the parenthesis. The figure is reproduced from De Barro *et al.* (2011)

After the refinement of the global mtCO1 phylogenetic relationships of *B. tabaci* (De Barro *et al.*, 2011), additional reports of new species from different regions were published. Four new *B. tabaci* species (Asia II-9, Asia II-10, China 3 and Asia III), for example, were reported in China (Hu *et al.*, 2011), one from India (Asia I – India) and Argentina (New World 2) (Alemandri *et al.*, 2012; Chowda-Reddy *et al.*, 2012). In 2013, the total number of *B. tabaci* putative species including a new species JPL earlier reported in Japan by Ueda *et al.* (2009) had risen to 31 (Lee *et al.*, 2013). This study also revealed that six *B. tabaci* species (Australia, Mediterranean, New World, Sub Saharan Africa 1, Asia II-7 and Asia II-1) had a genetic divergence higher than 3.5%. Therefore, these authors expressed a need to increase the 3.5% genetic boundary to 4.0%, to get a more realistic species delimitation. In addition, Firdaus *et al.* (2013) urged that more genes should be included in the analysis to resolve species delimitation amongst *B. tabaci* species. This is because the investigation of individual genes cannot provide an accurate description of genome-wide DNA sequence divergence. This was illustrated by the inclusion of additional sequences described in Ueda *et al.* (2009) and resulted in two new groups Japan 1 and Japan 2. Other groups reported included Asia II-11, Sub-Saharan Africa 5, Asia IV and Asia II-12. To date, more than 39 cryptic *B. tabaci* species have been proposed (Mugerwa *et al.*, 2018).

In addition to phylogenetic methods, mating compatibility studies have been conducted on many *B. tabaci* populations to test their biological species status. This is because phylogenetic methods can provide a molecular-identification framework, but the application of the biological species concept is central to the identification of species (Mayr, 1942; Suatoni *et al.*, 2006; Liu *et al.*, 2012). Speciation is influenced by evolutionary processes which include (i) mutation, (ii) migration, (iii) natural selection and (iv) genetic drift. These processes may affect the genetics, physiology, morphology, behaviour and development of the affected organism within the population (De Queiroz, 2007). These aspects may cause reproductive isolation, which renders some members of the population unable to interbreed. Therefore, defining species based on their ability to interbreed is more robust than using only molecular markers.

Several reciprocal mating studies have been done to determine the extent of reproductive isolation between genetic groups within the *B. tabaci* species complex (Bedford *et al.*, 1994; Maruthi *et al.*, 2001; Maruthi *et al.*, 2004; Zang and Liu, 2007; Sun *et al.*, 2011; Wang *et al.*, 2011; Liu *et al.*, 2012). The principle behind these studies is that two whitefly populations are reproductively

incompatible, if reciprocal crosses do not result in female progeny (Byrne & Devonshire, 1996; Xu *et al.*, 2010). The relationship between phylogenetically and biologically identified cryptic species within *B. tabaci* remains to be determined, because there are examples of species showing mtCO1 genetic divergence greater than 8% yet mating successfully (Qin *et al.*, 2016) and incompatibility shown between species with less than 3.5% mtCO1 sequence divergence (Vyskočilová *et al.*, 2018).

More studies on *B. tabaci* species delimitation are still on-going in different cropping systems. Figure 2.1 shows that some lineages within *B. tabaci* are evolving rapidly which makes precise species delimitation especially at 3.5% mtCO1 sequence divergence difficult. Combination of both molecular and mating studies, increasing the genetic species boundary and including more genes will improve species delimitation of closely-related *B. tabaci* species.

2.2.1 *B. tabaci* species associated with African cassava production and cassava diseases

The term “superabundant” was first used for cassava *B. tabaci* to describe those populations in the cassava growing areas in Uganda (Legg *et al.*, 2014a; Legg *et al.*, 2011), where the insects were so numerous that they gave rise to symptoms of physical damage. The CMD pandemic that devastated cassava production causing 100% yield loss in the 1990s was associated with unusually high whitefly populations (Colvin *et al.*, 2004). Legg *et al.* (2002) identified two distinct *B. tabaci* mtCO1 genotypes called Uganda 1 (Ug 1) and Uganda 2 (Ug2) with a partial mtCO1 sequence divergence of approximately 7.8%. These were categorized subsequently by Dinsdale *et al.* (2010) as Sub-Saharan Africa 1 (SSA1) and Sub-Saharan Africa 2 (SSA2), respectively. SSA2 was proposed to be an invader species with close relatives in West Africa. It was claimed to be the predominant population behind the “front” of the CMD epidemic, while SSA1 was restricted primarily to locations “ahead” of the “front” (Legg *et al.*, 2002). Based on these observations, Legg *et al.* (2002) hypothesized that SSA2 was one of the main causes of the CMD pandemic in Uganda. This interpretation of the data, however, was not supported by Maruthi *et al.* (2001), who conducted a study to compare mating compatibility, life-history traits and RAPD-PCR variation of whitefly population in pandemic and non-pandemic zones. Results from this study revealed that populations collected from behind and ahead of the CMD front had similar RAPD-PCR banding patterns, no statistically significant difference in both fecundity and nymphal development and

were not reproductively isolated. They concluded, therefore, that the CMD pandemic in East Africa was unlikely associated with a distinct and invasive genetic group of cassava *B. tabaci*.

Recent surveys in cassava producing areas, especially those affected by both CMD and CBSD, revealed that the SSA1, in particular a sub-clade called SSA1-SG1 has continued to maintain high populations and has hence described as “super-abundant” (Legg *et al.*, 2014b), while the SSA2 population has become rare (Sseruwagi *et al.*, 2005; Mugerwa *et al.*, 2012). Studies are still ongoing to determine the factors that influenced the shift in the population dynamics of these two *B. tabaci* populations.

In addition to the two whitefly species associated with CMD pandemic, a total of 11 putative species belonging to the sub-Saharan Africa mtCOI high level genetic group clade have been reported. These have been termed SSA3 to SSA13 (Abdullahi *et al.*, 2003; Berry *et al.*, 2004; Esterhuizen *et al.*, 2013; Mugerwa *et al.*, 2018). Of these, five reported on cassava include: SSA1, SSA2, SSA5, SSA8 and SSA10 (Boykin *et al.*, 2013; Esterhuizen *et al.*, 2013; Mugerwa *et al.*, 2018). SSA1 comprises five subgroups, which include Sub-Saharan Africa 1-subgroup 1 (SSA1-SG1), Sub-Saharan Africa 1-subgroup 2 (SSA1-SG2), Sub-Saharan Africa 1-subgroup 3 (SSA1-SG3), Sub-Saharan Africa 1-subgroup 4 (SSA1-SG4) and Sub-Saharan Africa 1-subgroup 5 (SSA1-SG5) (Ghosh *et al.*, 2015). Some of these whitefly populations exhibit significant biological differences in fecundity and survivorship and Mugerwa *et al.* (2019), for example, reported that SSA1-SG1 produced a significantly higher number of eggs and resulting adults, compared to SSA1-SG3.

To combat cassava viral diseases and their *B. tabaci* vector, efficient and durable management options are required. Previous studies have shown that species within the *B. tabaci* complex respond differently to management interventions (Frewin *et al.*, 2014; Naveen *et al.*, 2017), so before designing interventions, a detailed study needs to be carried out on the species of interest to design effective options that are well-tailored to that species. In this thesis, therefore, SSA1-SG1 was studied to identify essential gene targets for the management of their abundant populations in the cassava-producing areas of East Africa.

2.3 Management of *Bemisia tabaci* populations

Several approaches have been taken to manage whitefly populations in different cropping systems. These include cultural, chemical, biological and the use of host-plant resistance. There has also been increasing interest in the application of molecular and biotechnological tools and strategies for whitefly control.

Cultural practices such as intercropping, crop rotation, use of host-free periods and physical exclusion strategies have been applied in horticultural crop systems (Hilje *et al.*, 2001). The principle behind these activities is to create an unfavourable environment for the growth of whitefly populations. Asiimwe *et al.* (2007b) reported high *B. tabaci* egg mortality due to dislodgement resulting from a combined effect of environmental factors (wind and rain) and physical force. Dislodgement, therefore, can be facilitated by crop arrangement and spacing. These approaches have been unsuccessful in African cassava production, however, because cultural methods require communal cooperation between farmers, which has been difficult to achieve in many cassava-producing areas. In addition, the *B. tabaci* species and sub-groups that utilize cassava have a wide range of alternate hosts. SSA1, for example, can utilize and colonize many non-cassava plant species. These include: *Euphorbia heterophylla* L, *Manihot glaziovii* Mull. Arg, *Jatropha gossypifolia* L, *Phaseolus vulgaris* L, *Solanum melongena* L, *Sida acuta* Burm, *Vernonia amygdalina* Delile, *Commelina benghalensis* L, *Cucurbita moschata* Duchesne, *Solanum lycopersicon* Mill, *Abelmoschus esculenta* L and *Aspilia Africana* C.D. Adams (Sseruwagi *et al.*, 2006; Mugerwa, 2018). These alternate host-plants, therefore, facilitate survival of whitefly populations, even during periods when cassava is unavailable.

A few studies have been conducted on the efficiency of using biological control methods in the field management of whitefly species. Under field condition, more than half of immature nymph and egg mortality is due to dislodgement, parasitism and predation (Asiimwe *et al.*, 2007b). Beneficial insects, therefore, have been used in the development of biological management options for whiteflies in different cropping systems. A classic example of the use of biological control is provided by the management of citrus woolly whitefly, *Aleurothrixus floccosus* Maskell (Umeh & Adeyemi, 2011). In East Africa, the most common parasitoids of *B. tabaci* belong to family Aphelinidae. These include *Eretmocerus mundus* Mercet, *Encarsia lutea*, *Encarsia adrianae* and *Encarsia sophia* (Otim *et al.*, 2005). Surveys conducted in Uganda and other cassava producing

areas have shown that *Aphelinid* parasitoids of genera *Eretmocerus* and *Encarsia* occur in all whitefly-infested areas (Gerling *et al.*, 1990; Otim *et al.*, 2005). All these parasitoids have different mechanisms of attacking whitefly. *E. mundus* lays its eggs on the leaf underneath the *B. tabaci* nymph allowing the wasp larva to penetrate the whitefly nymph causing death, while *Encarsia* species oviposit directly within the whitefly nymph (Gelman *et al.*, 2005). The ability of the natural enemy to develop faster and maintain a high population at a time when the prey population is still high as well as having efficient means of ovipositing and developing successfully within the host are the key factors required for successful biological control of the pest. In addition, *Serangium n* sp. (*Col.*, *Coccinellidae*) has been also reported as an effective natural enemy of cassava *B. tabaci* (Asiimwe *et al.*, 2007a). Results from this study showed that a single *Serangium n* sp. larva consumed 1000 nymphs during its entire development. Integration of biological control strategies with other management options such as whitefly-resistant cassava varieties will not only make these methods effective, but also sustainable.

Breeding for host-plant resistance is potentially one of the most efficient method of managing cassava whitefly populations, especially among resource-constrained farmers. Limited studies have been done in screening for resistance, understanding inheritance and the genetic architecture of *B. tabaci* resistance in cassava. A number of genotypes resistant to South American whitefly species (*Aleurotrachelus socialis* Bondar) have been identified (Bellotti & Arias, 2001). These include: MEcu72, MEcu64, CG489-31 and CG489-34. To verify whether this resistance was effective against African cassava *B. tabaci*, Omongo *et al.* (2012) evaluated MEcu72 along with some selected Ugandan genotypes. MEcu72 and several Ugandan genotypes like Ofumba chai, Nabwire 1 and Mercury showed the most promising level of resistance. More studies are required to characterize this resistance in order to develop breeding options/designs for *B. tabaci* resistance.

Smith (2005) defined host-plant resistance as the sum of all constitutive genetically inherited qualities that may include mechanisms like antibiosis and antixenosis. Through antibiosis, the plant as a host poses an adverse effect on insect biology (larval mortality, disturbance of life cycle and reduction in fertility of the insects), while antixenosis relates to non-preference for oviposition or/and feeding leading to low oviposition on resistant genotypes. Antixenosis is due to the presence of morphological traits (plant hairs/trichomes and waxy layer) and/or constitutively or induced secondary metabolites that affect insect behaviour causing mild to lethal effects on different

genotypes or indirectly by attracting natural enemies (Mithoefer & Boland, 2012). Secondary metabolites effective against *B. tabaci* include acyl sugars reported in *Solanum pennellii* (Liedl *et al.*, 1995) and 7-epi Zingiberene in *Solanum habrochaitus* (Bleeker *et al.*, 2012). There is no published information yet available on the mechanism of resistance against *B. tabaci* and secondary metabolites that repel or deter whiteflies from feeding or ovipositing on cassava. There is, therefore, a need to screen cassava genotypes to identify resistant or tolerant genotypes with these mechanisms, study the genetic architecture of these traits in cassava and also evaluate the U/CBSV transmission efficiency of different whitefly species on different cassava genotypes. The development of modern techniques like genome-wide association analysis and genomic selection can aid in the study of resistance to whitefly in cassava as a highly heterozygous crop.

In addition, use of reverse genetics tools or strategies such as RNAi may result in more efficient approaches for dealing with superabundant population of cassava *B. tabaci*. The success of these molecular tools depends on the effective design of constructs and selection of gene targets. In this study, both the identity of candidate gene targets in cassava whiteflies and the efficacy of RNAi against key osmoregulation genes in whiteflies are evaluated.

2.4 RNAi and its application in insect control

Management of whiteflies in different cropping systems require species-specific management options that have been well-designed to manage the pest of interest without disrupting the ecosystem. Genetic tools for developing species-specific pest management options include RNAi and CRISPR/Cas9 (Bi *et al.*, 2016). RNAi is a post-transcriptional gene silencing mechanism that leads to downregulation of specific gene(s) through a sequence-specific RNA degradation process. It was first discovered in the nematode species *Caenorhabditis elegans* (Fire *et al.*, 1998). Since then, several studies on gene expression, gene function analysis and gene perturbation, have been reported. RNAi has been demonstrated as a proof of principle in the control of cyst nematodes (Klink *et al.*, 2009; Sindhu *et al.*, 2009), chewing insects including bollworm (Mao *et al.*, 2011), corn rootworm (Baum *et al.*, 2007; Baum, 2008), planthoppers (Chen *et al.*, 2010; Zha *et al.*, 2011), phloem-sap feeding insects like aphids (Pitino *et al.*, 2011) and whitefly (Malik *et al.*, 2016; Thakur *et al.*, 2014; Upadhyay *et al.*, 2011). In 2007, Monsanto and DOW AgroSciences collaborated to develop SmartStaxTM, the first ever eight-gene-stacked corn. Combination of

multiple modes of insect protection (RNAi and stacked Bt toxins) in SmartStaxTM was successful in controlling a broad spectrum of both above-ground and underground pests.

Functional RNAi machinery should consist of two major components, which include both the core components (Argonaute protein, Dicer enzymes and RNA binding factors) and a systemic component (Siomi & Siomi, 2009). The RNAi mechanism is made up of three steps namely: initiation, guiding and cleavage. In the initiation step, the RNAi mechanism is triggered by the presence of double-stranded RNA (dsRNA) that is recognized by RNase III endonuclease (Dicer). The RNase III nuclease cleaves the dsRNA into small interfering double-stranded RNA (siRNA) fragments with 3' overhangs of 2 or 3 nucleotides, 5-phosphate and 3-hydroxyl termini making up a total of 21-25 base pairs (Hamilton & Baulcombe, 1999), with siRNA composed of two strands called, (i) sense (passenger strand) and (ii) antisense (guide strand) (Jinek & Doudna, 2009). The duplex is unwound to load the guide strand into the RNA-induced silencing complex (RISC), while the passenger strand is discarded. The RISC component consists of two factors: RNA-binding factors and argonaute 'slicers' that together with the Dicer, form the core component of RNAi machinery (Lee *et al.*, 2004). The RNA-binding factor then guides the RISC component to identify a specific RNA transcript homologous to the guide strand of siRNA (Elbashir *et al.*, 2001). Once identified, the RISC attaches to the target mRNA, allowing the RNase H, in an argonaute protein complex to cleave it into a single site 10-11 nucleotide upstream from the 5' end of the siRNA (Figure 2.2) causing down-regulation of a gene (Hammond, 2005).

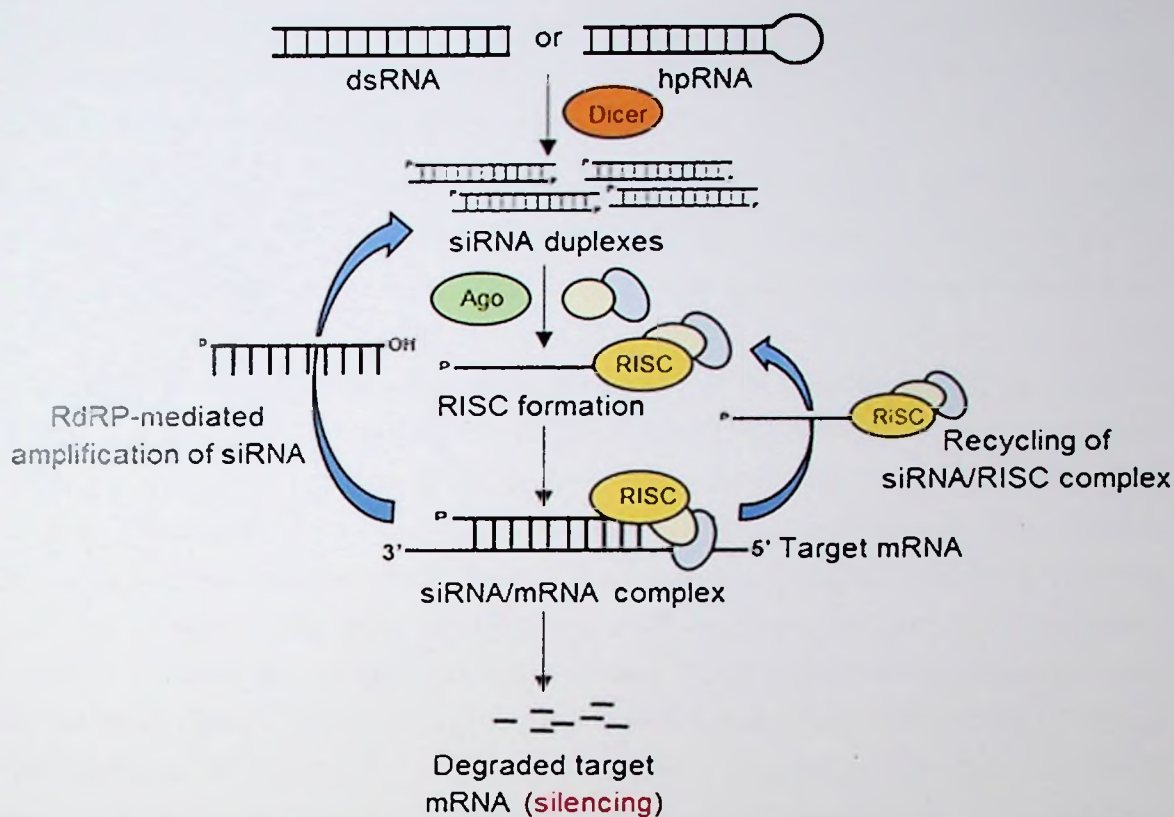


Figure 2.2: Mechanism of RNAi in the cell. The three steps include initiation, guiding and cleavage. RdRP that is responsible for amplification of the RNAi signal from one cell to another is missing in many insects. Figure reproduced from Majumdar *et al.*, 2017.

In addition to the core RNA machinery, some organisms exhibit a systemic component that amplifies the dsRNA signal from one cell to another and from one tissue to another within the organism (Siomi & Siomi, 2009). This is more efficient in plants and nematodes, especially *C. elegans*. These possess RNA-dependent RNA polymerase (RdRP), an enzyme responsible for the amplification of the RNAi signal from one cell to another, leading to a persistent and systemic RNA effect (Pak & Fire, 2007). Sijen *et al.* (2001) proposed that RdRP replicates siRNA as an epigenetic agent permitting their spread through the organism. Several studies conducted on insects have shown that RdRP is lacking in insects (Huvenne & Smagghe, 2010), even though the systemic spread of RNAi has been demonstrated in some beetles (Powell *et al.*, 2017). It was

hypothesized, therefore, that efficient systemic RNAi does not exist in some insects. Efforts to identify orthologs of RdRP in insects have been initiated, especially as RNAi is becoming a tool of choice for pest management in various cropping systems.

Proteins like systemic interference defective-1 (SID-1) have been reported, that act as a dsRNA channel, facilitating uptake of dsRNA (environmental RNAi) (Winston *et al.*, 2002). Though first reported in *C. elegans*, putative insect orthologs for the SID-1 gene have been described (Dong & Friedrich, 2005; Xu & Han, 2008). Earlier studies showed that SID-1 encodes for 11 helix transmembrane channel proteins that facilitate uptake of dsRNA, thereby mediating a systemic RNA effect (Jose & Hunter, 2007). Homologs of SID-1 have been identified in aphids (Xu & Han, 2008). Efforts to identify and characterize major components of the siRNA machinery like R2D2, Dicer 2, Argonaute 2 and SID-1 in whitefly has been successful (Upadhyay *et al.*, 2013). Results from this study showed that whitefly SID-1 had high amino acid sequence similarity (49-50%) with that of aphids, suggesting that whiteflies could also have systemic RNAi machinery. However, in the red flour beetle, *Tribolium castaneum* Herbst., SID-1-like genes are apparently not required for systemic RNAi, suggesting that this SID-1-like gene may have a different function than dsRNA uptake (Tomoyasu *et al.*, 2008). Other studies have demonstrated that insect SID-1 gene was homologous to TAG130 and CUP-1 genes that encode for dietary cholesterol uptake and transport in *C. elegans* (Wyder *et al.*, 2007; Valdes *et al.*, 2012). Wolfrum *et al.* (2007) exploited this ability by using cholesterol-conjugated siRNA to silence gene expression *in vivo*. These results demonstrated that SID-1 gene is a sterol transporter, unless a lipoplex technology for RNAi (Santel *et al.*, 2006) is applied, it may not be effective in mediating movement of dsRNA or siRNA across membranes. Studies are needed to characterize siRNA machinery in both aphids and whiteflies to accelerate the development of insect-pest RNAi-based control strategies. In the meantime, identification of genes differentially expressed in insect body-parts like the gut or salivary glands will increase the efficacy of RNAi in the control of insect pests. This is because silencing such genes may not require systemic RNAi machinery.

2.5 Potential essential gene targets for controlling cassava whiteflies

The search for essential genes in different organisms has been increasing since the emergence of synthetic biology and genetic engineering. Essential genes are those that are required for normal

growth and development in a given organism under a given condition, whose disruption usually results in fitness disadvantage and/or lethality (Juhas *et al.*, 2011). Essential genes have been characterized into two categories: (i) the core essential genes, which are universally required for every living cell and (ii) accessory essential genes that are critical in the specific biological system or for the survival of a specific cell type (Juhas *et al.*, 2012). Due to the specificity, targeting accessory essential genes is important to minimize effects on non-target organisms. Identification of accessory essential genes in cassava whiteflies is of outstanding interest especially in the struggle to combat cassava-virus diseases and the causal viruses that are transmitted by cassava-colonising *B. tabaci*.

Insect physiology includes seven important biological systems, which are: (i) nervous, (ii) respiratory, (iii) digestive, (iv) reproductive, (v) circulatory, (vi) endocrine and (vii) muscular, which are governed by numerous proteins or enzymes. A number of genes encoding these proteins or key enzymes have been targeted as control targets in many insect-management programs. Kola *et al.* (2015) reviewed many studies on insect management and catalogued key essential genes that have been targeted especially through RNAi. These include: (i) Vacuolar ATPase, (ii) Aminopeptidase N, (iii) Acetylcholinesterase, (iv) Vitellogenin, (v) Cytochrome P450 enzyme, (vi) Hydroxy-3-methyl glutaryl coenzyme A reductase, (vii) Arginine kinase, (viii) Chitin synthase, (ix) Catalase, (x) Tryptophan oxygenase, (xi) Allatostatin, (xii) Glutathione-S-transferase, (xiii) Hormone receptor gene and (xiv) Trehalose phosphate synthase. In whiteflies, several genes have been selected and used to assess the efficacy of RNAi. These genes varied in their efficiency showing that selection of genes has a very strong impact on the success of RNAi (Grover *et al.*, 2019). Selection of these genes as a target depends on the effect of gene perturbation and efficacy of RNAi on the physiology and survival of the insect. In phloem-sap feeders, the digestive/osmoregulation system is the most important system to target for control of these insects. This is because hemipterans such as whiteflies and aphids feed solely on phloem-sap, a nitrogen deficient and sugar-rich diet (Douglas, 2006). Osmoregulation and symbiosis, therefore, are critical for the survival of the insect. The genes that encode for proteins that mediate reactions in these systems are categorized as essential genes and are enriched in the gut and bacteriocytes, respectively. In this study, both structural and functional analyses were conducted to identify the

accessory essential genes as potential gene targets for control of phloem-sap feeders, especially African cassava whiteflies.

2.6 Osmoregulation genes in phloem-sap feeders

Whiteflies feed solely on plant-phloem sap that is highly concentrated with sucrose, often exceeding 1 M. This results in an osmotic pressure that is two to five times greater than the osmotic pressure of the insect's haemolymph (Douglas, 2006). This imbalance would be expected to cause a continuous flow of water down the osmotic gradient from the insect haemocoel to the gut content, thus causing the body to dehydrate and insect to die. Since the osmotic pressure exerted by solutes is determined by their molarity and not their weight, phloem-sap feeders overcome the osmotic challenge by transforming sucrose into oligosaccharides of lower molarity (Rhodes *et al.*, 1997) and water cycling. Osmoregulation has been studied in detail in the aphid (*Acyrtosiphon pisum* Harris) (Ashford *et al.*, 2000; Cristofolletti *et al.*, 2003). Based on these studies, aphids possess sugar-transforming enzymes, sucrase (α -glucosidase) especially in their proximal intestines that converts sucrose in the presence of water to α -D-glucose and β -D-fructose. The role of α -glucosidase was confirmed when aphids were fed on acarbose, an α -glucosidase inhibitor that inhibited sucrase activity in the gut homogenates (Karley *et al.*, 2005). In whitefly, this enzyme has been identified as α -D-glucoside glucohydrolase (sucrase-transglucosidase) belonging to the GH-13 family (Jing *et al.*, 2016).

2.6.1 Glycoside hydrolase family 13 enzymes and sucrose assimilation in whitefly

The glycoside hydrolase (GH) superfamily is a large group of carbohydrate-active enzymes that hydrolyze glycoside bonds between two or more carbohydrates (Cantarel *et al.*, 2009). Glucohydrolase family 13 formally known as an α -amylase family is the largest family of glycoside hydrolases and is composed of hydrolases, transglucosidases and isomerases (MacGregor *et al.*, 2001). The GH-13 enzymes show sequence similarity and a conserved structural core composed of three domains called A, B and C. Domain A folds as a $(\beta/\alpha)_8$ -barrel (TIM-barrel) (Abad *et al.*, 2002), domain B is a loop inserted between the helix α_3 and β_3 of the $(\beta/\alpha)_8$ -barrel (Janecek, 1997) and domain C, a C-terminal extension (Ramasubbu *et al.*, 1996). Enzymes in GH-13 contain a catalytic triad that includes an aspartate (Asp 206) as a catalytic

nucleophile, glutamate (Glu 230) that acts as a general acid/base and another aspartate (Asp 297) that acts in stabilizing the transition state (Uitdehaag *et al.*, 1999). In addition to the catalytic triad, enzymes in GH-13 also contain an arginine residue. These enzymes use a double displacement catalytic mechanism to break down the glycosyl moiety bond in heterogeneous substrates.

Alpha-glucosidases are categorized into three types, I, II and III, based on substrate recognition (Nakai *et al.*, 2005). The α -glucosidase family 13 type I has been observed in bacteria and insects (Chiba, 1997). They specifically recognize the α -glycosyl moiety and hydrolyze heterogeneous substrates such as sucrose. Three enzymes hydrolyze sucrose to fructose and glucose. These include sucrose α -glucosidase (EC 3.2.1.48), β -fructofuranosidase (*Invertase*) (EC 3.2.1.26) and α -glucosidase (EC 3.2.1.20) (KEGG database)³. Of these, transcripts for α -glucosidase (EC 3.2.1.20) have been identified and reported to be enriched in the whitefly gut of MEAM1 (Jing *et al.*, 2016). Monosaccharides arising from these transformations undergo various modifications. In the pea aphid, for example, fructose is assimilated across the gut walls, while glucose is transformed through gut transglucosidase activity into oligosaccharides that are released as honeydew (Ashford *et al.*, 2000). Both the sucrase and transglucosidase activities contribute to the reduction of osmotic pressure of ingested phloem sap in the gut.

The biochemical processes of sugar utilization have been investigated in *Bemisia argentifolii* (= *B. tabaci* MEAM1) (Salvucci, 2000). The difference, though, is the final sugar product in honeydew. Salvucci *et al.* (1997) showed that the honeydew of whitefly feeding on high dietary sucrose contained elevated levels of the disaccharide trehalulose (1-0- α -D-glucopyranosyl-D-fructose) and diminutive amounts of oligosaccharides. Byrne *et al.* (2003) estimated a total of 30% of the total carbohydrates found in whitefly honeydew to be trehalulose. Analysis of whitefly extract treated with bromoconduritol, an inhibitor of α -glucosidase activity, showed inhibition of both trehalulose synthase and α -glucosidase activity (Salvucci, 2000), thus confirming the role of trehalulose synthase in sugar transformation in whitefly. These results showed that the transglucosidase reaction varies according to the mode of action, producing different products. It is important, therefore, to study the mode of action of transglucosidase to identify key enzymes that are important in sugar transformation and transport within cassava whitefly *B. tabaci* species.

³ https://www.genome.jp/kegg-bin/show_pathway?ec00500

2.6.2 Major intrinsic proteins and water recycling in the *B. tabaci* gut

Phloem-sap feeders also overcome the osmotic challenge through water cycling in the gut through a specific water-channel protein, aquaporin (Campbell *et al.*, 2008; Shakesby *et al.*, 2009). Aquaporin, a major intrinsic protein of 28kDa was first described by Preston *et al.* (1992). The major intrinsic protein (MIP) family comprises more than 800 members, categorized into three classes, which are: (i) the aquaporins that conduct water and ions, (ii) aquaglyceroporins, that allow passage of water and (iii) uncharged molecules and glycerol-facilitators that facilitate glycerol transport (Van Ekert *et al.*, 2016). In whitefly, *B. tabaci* (MEAM1), aquaporin 1 (BtDRIP/BtAQP1) has been reported to transport water (Mathew *et al.*, 2011). This water-specific aquaporin has been also reported in *Aedes aegypti* L (Drake *et al.*, 2015), in *Acyrtosiphon pisum* (Shakesby *et al.*, 2009) and *Drosophila melanogaster* Meigen (Kaufmann *et al.*, 2005).

In general, the aquaporin channel forms a constriction pore halfway into the lipid layer (hourglass structure), comprised of six membrane alpha-helical domains connected by five extramembrane loops (Figure 2.3). In the centre of the pore is an asparagine-proline-alanine motif that forms a ring (Jung *et al.*, 1994). This pore is restricted by four residues (phenylalanine, histidine, cysteine and arginine) that comprise the Ar/R restriction site forming a principal filter of the channel, regulating the flow of water through the channel. Results from earlier studies demonstrated that BtAQP1 is located primarily in the filter chamber and anterior ileum of the hindgut (Mathew *et al.*, 2011). A detailed description of the filter chamber has been reported by Ghanim *et al.* (2001), who demonstrated that a filter chamber is a highly specialized organ that allows passage of water down the osmotic gradient in the whitefly gut. Aquaporins, therefore, regulate osmoregulation by facilitating the osmotic movement of water from the distal gut to the proximal midgut ensuring equilibrium of the osmotic pressure between the haemolymph and the gut lumen.

Major intrinsic proteins are classified into functional categories based on the identity of amino acids in the aromatic/arginine (ar/R) region, which determines the size and selectivity of the MIP pore. Based on the crystal structure of the ar/R, member proteins like aquaporins are categorized as water-specific proteins and non-water specific proteins (aquaglyceroporins and glycerol facilitators). Water selectivity in water-specific aquaporins is achieved by a specialized structure of the ar/R residue that comprises four amino acids: (i) Phe at helix 2, (ii) His at helix 5, (iii) Cys

at LE₁ and (iv) Arg at LE₂, forming a narrow constriction region of 2.8 angstrom (1 angstrom $\approx$ 0.1nm) (Sui *et al.*, 2001; Thomas *et al.*, 2002). The amino acid at LE₁ in ar/R filter of water-specific aquaporin varies in different insect species. For example in *B. tabaci* MEAM1, BtDRIP/BtAQP1, the conserved water selective ar/R selectivity filter, comprised of Phe 71, His 198, Ala 207 and Arg 213 (Van Ekert *et al.*, 2016). Non-water selective aquaporins (aquaglyceroporins) such as Aquaporin 3 (AQP3), Aquaporin 7 (AQP7) and Aquaporin 9 (AQP9) contain different amino acids in the ar/R residue. These include Try at helix 2, Gly at helix 5, Try at LE₁ and Arg at LE₂, forming a constriction region of 3.4 angstroms that allowed transport of uncharged dissolved molecules like CO₂, glycerol, ammonia and urea at a lower rate (*ibid*).

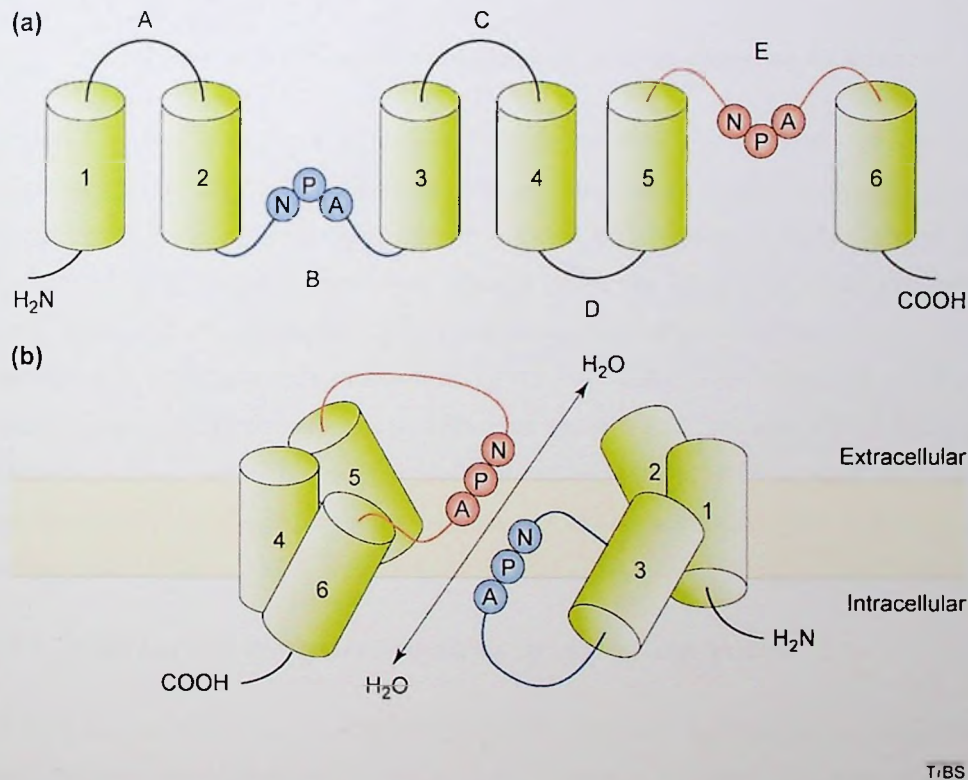


Figure 2.3: Structure of water-specific aquaporin. (a) The structural layout of AQP1 monomer comprised of six membrane-spanning helices connected by five loops. (b) Functional layout of the AQP1 monomer forming a water channel. The figure was reproduced from Zeuthen (2001).

Classical insect aquaporins have been further classified into four major subfamilies that are: (i) Pyrocoelia rufa integral protein (PRIP)-like (Lee *et al.*, 2001), (ii) Drosophila integral protein (DRIP)-like (Kaufmann *et al.*, 2005), (iii) the big brain protein (BIB) (Campbell *et al.*, 2008) and (iv) entomoglyceroporin (EgIps) (Finn *et al.*, 2015). DRIPs and PRIPs contain sequence motifs like the water-specific channel comprised of asparagine-proline-alanine (NPA) boxes both in the loop B (cytoplasmic) and loop E (extracellular) that define the water channel. The water-selective aquaporin in *B. tabaci* MEAM1 (BtAQP1) belongs to the DRIP subfamily (Van Ekert *et al.*, 2016). It is important to identify and functionally characterize aquaporins in other *B. tabaci* species, so in this study, the water-specific aquaporin was identified as a gene target for the management of superabundant cassava *B. tabaci* (SSA1-SG1) using RNAi.

2.7 Symbiosis and amino-acid provisioning in plant phloem-sap feeders

Phloem-sap feeding insects cannot synthesize ten amino acids that are required for protein synthesis. These include: (i) histidine, (ii) methionine, (iii) isoleucine, (iv) phenylalanine, (v) leucine, (vi) lysine, (vii) threonine, (viii) tryptophan (ix) valine and (x) arginine. The problem is aggravated by the fact that phloem sap of most plants has a low concentration of amino acids (0.1-2 mg ml⁻¹), yet phytophagous insects require amino acids for their growth, reproduction and development (Sandstrom & Pettersson, 1994). The reduction in the concentration of even one essential amino acid, therefore, can affect the growth and reproduction of the insect drastically (Douglas, 2006). To overcome this problem, whiteflies and other phloem-sap feeders depend on endosymbionts for the provisioning of essential amino acids (Douglas & Prosser, 1992; Gunduz & Douglas, 2009).

2.7.1 Primary and secondary symbionts in phloem-sap feeders

Whiteflies, like other phloem-sap feeders, possess both obligate intracellular (primary symbionts) and facultative (secondary symbionts). Primary symbionts are vertically transmitted via the female ovary through a process called transovarial transmission (Luan *et al.*, 2016) and are restricted to the cytoplasm of a specialized insect cell known as a bacteriocyte, which acts as a medium of interaction between the host and symbiont (Baumann, 2005; Heddi *et al.*, 2005). Their major role is to provide several essential amino-acids or metabolites for intermediate reactions for essential

amino-acid biosynthesis in the host. Examples of primary symbionts include: (i) *Buchnera aphidicola* in aphids, (ii) *Tremblaya princeps* in mealybugs, (iii) *Carsonella ruddi* in psyllids and (iv) *Portiera aleyrodidarum* in the *B. tabaci* species complex (Thao & Baumann, 2004a).

In addition, *B. tabaci* species have a diversity of secondary symbionts which include *Wolbachia* (*Rickettsiales*) (Ghosh *et al.*, 2015), *Fritschea* (*Chlamydiales*), *Arsenophonus* (*Enterobacteriales*), *Hamiltonella* (*Enterobacteriales*) (Thao & Baumann, 2004b), *Cardinium* (*Bacteroidetes*) and *Rickettsia* (*Rickettsiales*) (Gottlieb *et al.*, 2006). Studies on Asian whitefly species have also revealed a secondary symbiont called *Hemipterophilus* (*Rickettsiales*) (Bing *et al.*, 2013). Secondary endosymbionts can be vertically transmitted as well as horizontally transmitted at a longer evolutionary time-scale, so occupy different locations within the host, for example, in the sheath cell of bacteriocytes, guts, and reproductive organs (Engelstadter & Hurst, 2009). In some cases, *Hamiltonella*, *Arsenophonus*, *Wolbachia*, *Cardinium* and *Rickettsia* have been found together with a primary symbiont in the bacteriocytes within the whitefly (Gottlieb *et al.*, 2008). This fosters co-operation with the primary symbiont in essential amino-acid biosynthesis (Ankrah *et al.*, 2017)

Secondary symbionts play various roles in their host, which include: (i) ability to withstand heat stress (Brumin *et al.*, 2011), (ii) host range expansion (Henry *et al.*, 2013), (iii) resistance to natural enemies (Mahadav *et al.*, 2008), (iv) resistance to pathogens (Hedges *et al.*, 2008), (v) insecticide resistance (Ghanim & Kontsedalov, 2009) and (vi) more efficient virus transmission (Gottlieb *et al.*, 2010). Some secondary symbionts such as *Rickettsia* have also been reported to benefit their whitefly hosts by increasing their fitness by influencing reproduction, especially as a female-biased sex ratio, fecundity and faster development time (Asiimwe *et al.*, 2014). Furthermore, secondary symbionts contribute to the nutritional requirements of their host. For example, *Hamiltonella defensa* in aphids has been reported to supply essential vitamins except for pantothenate (B5) and thiamine (B1) (Degnan *et al.*, 2009) and *Heteropsylla cabana* Crawford complements *Carsonella eucalypti* in the biosynthesis of tryptophan in psyllids (Sloan & Moran, 2012b).

2.7.2 Shared metabolic pathways between symbionts and phloem-sap feeders

Several studies have also demonstrated that phloem-sap feeders especially aphids and whiteflies have shared metabolic pathways (Russell *et al.*, 2013; Luan *et al.*, 2015), through which the host

synthesizes amino acids and other metabolic products from either dietary precursors like sucrose or metabolites from primary endosymbionts (Douglas, 1998). Analysis of whitefly gene expression revealed that the whitefly genome interacts with *Portiera aleyrodidarum* (henceforth referred to as *Portiera*) in amino acid and carotenoid biosynthesis (Upadhyay *et al.*, 2015). The *Portiera* genome is highly reduced (357 kb) with conserved genes encoding for mostly essential amino acids and carotenoids (Santos-Garcia *et al.*, 2012; Rao *et al.*, 2015). The majority of its metabolism genes (50-60%) contribute to the synthesis of the nine essential amino acids (Luan *et al.*, 2015). *Portiera* has genes that encode for proteins involved in the terminal reaction of threonine and tryptophan biosynthesis, while it lacks the last reactions in the biosynthesis of seven essential amino acid (histidine, lysine, valine, isoleucine, leucine, phenylalanine and arginine) (Sloan & Moran, 2012a; Luan *et al.*, 2015). These amino acids and metabolic intermediates are transported from the bacteriocyte to the whitefly. *Portiera* also lacks almost all the genes required for the synthesis of vitamins and cofactors, however these genes are found in some secondary endosymbionts.

The whitefly host also contributes to this interaction. Some critical amino-acid biosynthesis genes, for example, for serine and proline, are missing in both the *Portiera* and *Hamiltonella* genomes (Rao *et al.*, 2015). However, analysis of whitefly transcriptome sequences revealed that the missing reactions are performed by the whitefly itself (Wang *et al.*, 2010). This phenomenon has also been also reported in the aphid-*Buchnera* interaction (Russell *et al.*, 2013). This contribution facilitates symbiosis in at least two ways, which are: (i) the whitefly possesses genes for biosynthesis of non-essential amino acids like glutamate, glutamine, asparagine, serine, aspartate, alanine, glycine, proline, tyrosine and cysteine. These amino acids are produced in sufficient quantities to meet the needs of both the whitefly and the bacterial symbionts. (ii) Amino acids can then be transformed into essential amino acids through transamination, a process in which the amino group from an amino acid is transferred to an alpha-keto acid, thereby generating a new amino acid from an existing one. An example is amino-acid glutamate, which is a precursor of a number of essential amino acids, such as threonine, isoleucine, valine, leucine and lysine (Sasaki & Ishikawa, 1995). Glutamate is converted from glutamine, a dominant amino acid in the phloem-sap (Sandstrom & Pettersson, 1994; Valle *et al.*, 1998). In aphids, *in silico* modeling of flux through a metabolic network showed that four non-essential amino acids (aspartate, glutamate,

serine and glutamine) are taken up by the primary symbiont, *Buchnera aphidicola*, as a precursor of essential amino acids (Gunduz & Douglas, 2009). The contribution of whitefly (MEAM1) to the *Portiera* symbiont relationship has been reported by Luan *et al.* (2015) and Ankrah *et al.* (2017), with inputs like pyruvate and aspartate.

The lack of essential genes in the host and its primary symbiont raises a question of how the carbon skeleton of missing essential amino acids are produced in the whitefly. This is achieved by interacting with secondary endosymbionts, especially those that co-exist with *Portiera* in the bacteriocyte (Chiel *et al.*, 2009). *Hamiltonella defensa*, for example, contains genes that are lacking in both the whitefly and *Portiera* for the biosynthesis of lysine (*dapB*) (Luan *et al.*, 2015). Recent metabolic modeling studies on the MEAM1-*Portiera*-*Hamiltonella* interaction, however, showed that *Hamiltonella* produces lysine and threonine, but only in sufficient quantities to meet its own nutritional requirements (Ankrah *et al.*, 2017). In addition, it also contributes multiple vitamins and co-factors like biotin (B7), folic acid (B9), riboflavin (B2), pyridoxine (B6) to the whitefly-symbiont relationship (Skaljac *et al.*, 2010; Rollat-Farnier *et al.*, 2015). More studies are required to study the contribution of different secondary symbionts to the nutritional demands of different whitefly species. This will enable the selection of more efficient gene targets for controlling *B. tabaci* species.

The contribution of the whitefly host in the symbiotic relationship has been facilitated by both nitrogen recycling (Upadhyay *et al.*, 2015) and lateral acquisition of genes from bacteria (Luan *et al.*, 2015). These horizontally transferred genes are integrated into the whitefly genome, thereby compensating for gene loss in the bacterial symbiont and contributing to essential amino-acid biosynthesis (Husnik *et al.*, 2013; Sloan *et al.*, 2014). Therefore, as the bacterial symbionts lose genes due to relaxed functional constraints on redundant genes (Andersson *et al.*, 1998), the host acquires new genes coding for critical reactions, allowing the host to alter metabolite flux through the shared metabolic pathway, thereby retaining control over essential amino-acid production.

The crosstalk between host and symbionts is also facilitated by amino-acid transporters (amino-acid polyamine organocation transporters and the amino-acid transporters). Orthologs of these genes have been reported in the citrus mealybug, *Planococcus citri* Risso (Price *et al.*, 2011). Studies are required to identify these transporters in whitefly especially in the cassava *B. tabaci*. In general, identification of these key genes that (i) control essential amino-acid biosynthesis, (ii)

transport amino acids and metabolic intermediates and (iii) facilitate nitrogen recycling in whitefly will facilitate the development of effective biotechnological control strategies to combat whitefly.

2.7.3 Amino-acid transporters and amino-acid provisioning in phloem-sap feeders

The distribution of essential amino acids from the endosymbionts to the host tissues and amino acids from the host to the endosymbiont across the symbiosome membrane is mediated by amino-acid transporters (Douglas, 2014; Duncan *et al.*, 2014). A symbiosome membrane is a special membrane that separates the endosymbiont and the host in the bacteriocyte. Key amino-acid transporters reported in eukaryote cells belong to two gene families: the amino acid/ auxin permease (AAP) family (transporter classification (TC) # 2.A.18) and the amino acid polyamine organocation (APC) family (TC # 2.A.3) (Castagna *et al.*, 1997; Saier *et al.*, 2016). APC amino-acid transporters act as antiporter, uniporter and symporter, therefore, they catalyze uptake as well as the exchange of amino acids across the membrane, while the AAP family uses the symport mechanism for single or multiple amino acids across a lipid membrane (Saier, 2000). Amino-acid transporters have been further classified into different systems depending on the substrate specialty (profile of amino-acids transported) and the transport mechanism. Based on substrate profile, amino-acid transporters interact with amino acids depending on the polarity and electrochemical properties of their side chains. Therefore, the polar cationic amino acids (lysine and arginine) are mediated by the y^+L transport system, which in presence of Na^+ , also mediate the neutral amino acids (leucine, methionine and glutamine) (Vanwinkle *et al.*, 1985). Other transport systems for cationic amino acids include; $B^{0,+}$ (lysine, arginine, alanine and valine), $b^{0,+}$ (lysine, arginine and leucine) and y^+ (lysine, arginine and ornithine) (Deves *et al.*, 1992; Dye *et al.*, 2004). In addition, transport of neutral amino acids is also mediated by system L, system A (glycine, isoleucine, leucine, methionine, phenylalanine and proline), system ASC (alanine, threonine, serine and cysteine), system T (tryptophan, phenylalanine and tyrosine) and system N (glutamine, asparagine and histidine), while polar anionic amino acids are transported via X^-c (Cystine & glutamate), system X^-_{AG} (aspartate and glutamate) and system X^-_A (Aspartate) (Bridges *et al.*, 2012).

In aphids, the dynamic recruitment of amino acid transporters to the aphid/*Buchnera* interface has been studied (Duncan *et al.*, 2014). Limited information is available on the recruitment of amino-acid transporters on the whitefly/*Portiera* interface. Identification of key amino acid transporters

enriched in the bacteriocytes is very important especially as the search for accessory essential genes in whiteflies has become relevant. Recently, a total of 14 putative APCs and 25 putative AAAPs were identified in *B. tabaci* MEAM1 (Xia *et al.*, 2017), however, this study used whole body samples of four different life stages of whitefly to screen for amino-acid transporters. To identify amino-acid transporters that play a significant role at the whitefly/*Portiera* interface, amino-acid transporters enriched in the bacteriocytes should be identified and characterized.

In summary, the impact of *B. tabaci* on the economy and food security of Sub-Saharan Africa warrants an urgent response to manage the damage caused by cassava whiteflies. In this study, the two critical physiological systems of the whitefly, osmoregulation and symbiosis were investigated to identify accessory, essential-gene targets for the management of cassava *B. tabaci* SSA1-SG1 through RNAi. Genes of interest were: (i) gene(s) that encode for proteins that mediate water recycling in the gut of SSA1-SG1 whiteflies, (ii) genes that encode for enzymes that mediate sucrose hydrolysis and transglucosidation and (iii) symbiosis genes that are enriched in the bacteriocytes (transaminases, amino-acid transporters and horizontally-transferred genes).

IDENTIFICATION OF KEY OSMOREGULATION AND SYMBIONT GENES AS TARGETS FOR CONTROLLING CASSAVA WHITEFLIES (*BEMISIA TABACI*) USING RNA INTERFERENCE.

3.1 INTRODUCTION

Bemisia tabaci (*sensu lato*) feed solely on phloem-sap, which presents physiological challenges (Section 2.6 & 2.7). Understanding how phloem-sap feeders overcome these physiological challenges offers opportunities for identifying key gene targets and designing effective pest management strategies for phloem-sap feeders. The specialized food source of *B. tabaci* and other phloem-sap feeders makes them vulnerable to genetic perturbations especially when physiological systems, for example, those that mediate osmoregulation and amino acid provisioning are targeted (Douglas, 2006). Osmoregulation genes are required for the maintenance of fluid homeostasis within the insect body cavity. Due to the high osmotic pressure exerted by the sugar-rich diet ingested in the gut, osmoregulation genes, sugar processing and water recycling genes have been reported to be enriched in the insect gut (Karley *et al.*, 2005; Jing *et al.*, 2016). Symbiosis genes, on the other hand, are required for amino acid biosynthesis and provisioning between *B. tabaci* as a host and the symbionts in the bacteriocytes. The shared metabolic pathways between *B. tabaci* and its primary endosymbionts have been studied in *B. tabaci* MEAM1 (see Section 2.7.2). Potential symbiosis genes that play a significant role in the provisioning of amino acid especially in host-symbiont relationships include (i) amino acid transporters, (ii) amino acid transaminase and (iii) horizontally transferred genes in the bacteriocytes (Douglas, 2014; Luan *et al.*, 2015).

Advances in molecular techniques especially next-generation sequencing technologies, are allowing the study of these physiological processes at high resolution. RNA sequencing (RNAseq), for example, has been used to study physiological processes at the transcriptional level in different organisms, allowing quantification and comparison of gene expression which facilitates identification of essential genes (Wang *et al.*, 2009). In this study, the differential expression of genes in the cassava *B. tabaci* (SSA1-SG1) gut and bacteriocytes was analyzed to identify essential gene targets for the management of *B. tabaci* SSA1-SG1 population using RNA interference. Functional analysis of the genes enriched in gut and bacteriocytes was done to select the best gene

targets, whose disruption would potentially have a detrimental effect on *B. tabaci* SSA1-SG1 survival.

3.2 MATERIALS AND METHODS

3.2.1 Dissection of guts and bacteriocytes from cassava *B. tabaci*

Genes enriched in the gut and bacteriocytes were identified by comparing gene expression either (i) in the *B. tabaci* gut and (ii) bacteriocyte with the whole-body. The cassava *B. tabaci* population (SSA1-SG1) used in this study was derived from a single virgin female and male originally collected from Namulonge, Uganda, an area with superabundant *B. tabaci* populations and a high incidence of both Cassava brown streak disease and Cassava mosaic disease. The *B. tabaci* population was maintained at the Natural Resources Institute, the University of Greenwich, on aubergine (*Solanum melongena* cv. Black Beauty) under insectary conditions (60% relative humidity, 28°C and 12h light:12h dark regime) with 500 $\mu\text{mol m}^{-2}\text{s}^{-1}$ PAR. light intensity. Approximately 300 guts and more than 10000 bacteriocytes were dissected in phosphate buffered saline (PBS, 1X) from 300 and 500 female adult whiteflies respectively, belonging to cassava *B. tabaci* species SSA1-SG1 using fine pins and a stereomicroscope. Both guts and bacteriocytes were maintained in RNAlater™ stabilization solution (Invitrogen™) to preserve their RNA quality. A total of 50 female adult whiteflies were also collected and immediately stored at -80°C for analysis of gene expression in wholebody.

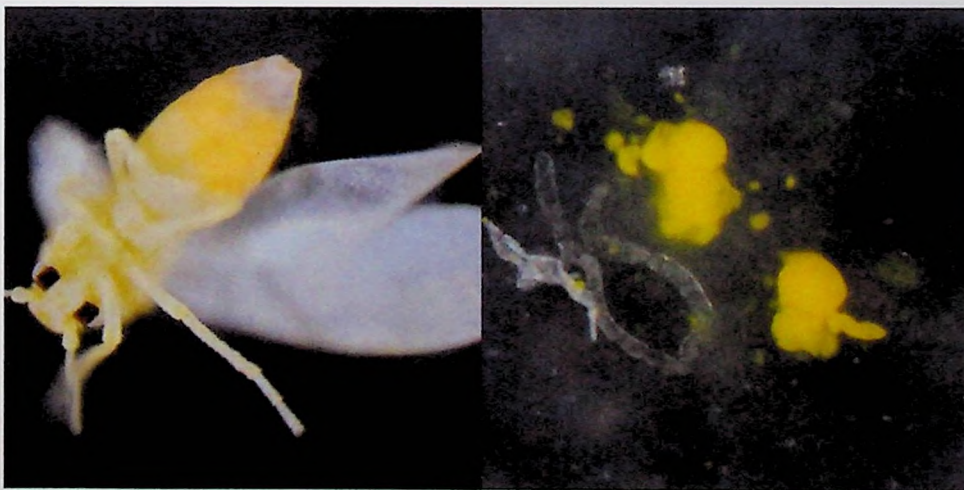


Figure 3.1: Dissected bacteriocytes and gut from the abdomen of *B. tabaci* (SSA1-SG1)

3.2.2 RNA extraction from dissected guts and bacteriocytes

Total RNA was extracted from *B. tabaci* guts, bacteriocytes and whole bodies using TRIzol™ reagent (Invitrogen™), with slight modifications made in the manufacturer's instructions to ensure purity and quantity of RNA for downstream analysis for low tissue amounts. Briefly, the *B. tabaci* whole body, guts and bacteriocytes were homogenized with 500 µl TRIzol™. Homogenized samples were incubated at room temperature for 5 minutes followed by adding 100 µl chloroform. The mixture was then incubated for 3 minutes at room temperature and centrifuged at 12,000g for 15 minutes at 4°C to separate the aqueous phase. The aqueous phase was then transferred to a clean tube and the second cleanup was done by adding 100 µl of chloroform. The mixture was vortexed for 15 seconds then incubated at room temperature for 3 minutes and thereafter centrifuged at 12,000g for 10 minutes at 4°C. Phase separation was repeated and 1 µl of linear acrylamide was added to the aqueous phase to improve the precipitation of RNA. Ice cold isopropanol (250 µl) was then added and mixed by pipetting. The mixture was further incubated at room temperature for 10 minutes and then centrifuged at 12000g for 15 minutes at 4°C. The supernatant was removed and discarded using a pipette and the RNA pellet washed twice with 500 µl of 75% ethanol. Finally, the RNA pellet was air dried for 5-10 minutes then dissolved in DNase/RNase free water followed by incubating at 60°C for 10 minutes to ensure all the RNA had dissolved. Total RNA was further treated with ezDNase™ enzyme (Invitrogen™) according to the manufacturer's instructions, to remove contaminating genomic DNA. In brief, 1 µl 10X ezDNase buffer and 1 µl ezDNase™ enzyme were added to 8 µl RNA and mixed gently. RNA and ezDNase™ mixture were then incubated for 2 minutes at 37°C followed by 5 minutes at 55°C in presence of 10 mM DTT to inactivate the enzyme. RNA quantity and integrity were confirmed using a Qubit fluorometer (Life Technologies).

3.2.3 rRNA depletion and RNAseq library preparation

Total RNA was depleted of cytoplasmic and mitochondrial rRNA using the Ribo-Zero Gold rRNA removal epidemiology kit (Illumina, San Diego, USA) according to manufacturer's instruction. For the whole-body samples, 1 µg total RNA was used for rRNA depletion, due to limited material only 367 ng total RNA was used for the bacteriocyte RNA sample and 230 ng for the guts sample.

RNAseq libraries were prepared with the NEBNext Ultra II RNA Library Prep Kit (New England Biolabs Inc), according to the manufacturer's instructions. For the whole-body samples, 15 ng rRNA-depleted RNA was used as input while for the bacteriocyte and gut samples, all the rRNA-depleted RNA (not quantified) was used as input. The final libraries were quantified with a Qubit HS DNA assay and the size distribution was determined on a Fragment Analyzer instrument (Agilent Technologies, Inc. CA, USA). Equimolar amounts of each library were pooled for sequencing.

3.2.4 RNA sequencing and RNAseq analysis

The RNAseq libraries were sequenced and >40 million single-end 75 nt reads were generated on the NEXTseq500 sequencing platform at the Cornell Genomics Facility (Biotechnology Resource Center). Raw reads were filtered to remove low quality reads and adaptor sequences with Cutadapt (`-m 50 -q 20 -a AGATCGGAAGAGCACACGTCTGAACTCCAG --match-read-wildcards`) (Martin, 2011). The cleaned RNAseq reads were aligned against the SSA1-SG1 genome using HISAT2 (Kim *et al.*, 2015), which runs Bowtie2 in the background. The accepted hits were used, firstly to develop gene annotation file containing gene models based on RNAseq aligned reads for guts, bacteriocytes and whole-body using cufflinks and cuffmerge version 2.2.1 (Trapnell *et al.*, 2010). Secondly, the accepted hits were assembled and aligned reads counted using featureCounts (Liao *et al.*, 2014). Differential gene expression in cassava *B. tabaci* (SSA1-SG1) gut and bacteriocytes relative to the whole-body was performed based feature counts using edgeR (Robinson *et al.*, 2010) with false discovery rate controls based on the Benjamini & Hochberg, (1995) method and trimmed mean of M-values normalization to account for library size variation between the samples. Up-regulated genes with log2 fold difference (logFC) based on general linear model (likelihood ratio test) greater than one in *B. tabaci* guts and bacteriocytes were selected for further analysis.

3.2.5 Identification of osmoregulation and symbiosis genes

Nucleotide sequences for the differentially expressed genes in both *B. tabaci* gut and bacteriocytes were extracted using Samtools faidx tool (Li *et al.*, 2009). Blastx search (Altschul *et al.*, 1997) with a cutoff E-value of 1e-3 against NCBI protein reference non-redundant database (nr),

mapping and annotation of these sequences were done using Blast2GO suite (Conesa *et al.*, 2005). The suite was also used to generate and characterize gene ontology terms based on BLAST output. Gene ontology annotation was characterized into three groups, namely molecular function, biological process and cellular component. Selected sequences with metabolic functions were further translated to protein sequences using the ExPASy molecular biology server of the Swiss Institute of Bioinformatics (Gasteiger *et al.*, 2003). The translated protein sequences with Amy domain (PF00128, IPR006047) for glycosyl hydrolases (family 13), aquaporin domain (PF00230, IPR000425) for osmoregulation genes and amino acid transporter domains (PF01490, IPR013057), amino acid/polyamine transporter I (IPR002293) and amino acid permease domain (IPR004841, IPR013612) for symbiont genes were selected from the blast2go output. The selected sequences were further analyzed using SignalP 4.0 server (Petersen *et al.*, 2011), NCBI conserved domain database (Marchler-Bauer *et al.*, 2015), HMMER V.3 (Eddy, 2011) and InterProScan (Zdobnov & Apweiler, 2001) to confirm the functional domain for sucrase, aquaporin and amino acid transport. The selected sequences were also analyzed using NCBI Batch web CD-search [tool](#)⁴ to determine if the conserved domains were complete or incomplete at N-terminus or C-terminus.

Furthermore, a candidate-based search for horizontally transferred genes (HTGs) in SSA1-SG1 was conducted. Sequences of HTGs identified in MEAM1 (Luan *et al.*, 2015; Chen *et al.*, 2016), were identified from the whitefly genome database⁵. A reciprocal blast against SSA1-SG1 genome and transcriptome was done to identify genes that encode for HTGs in SSA1-SG1 and their expression level in the bacteriocytes.

3.2.6 Phylogenetic analysis

To determine the conservation of osmoregulation function of the selected genes, phylogenetic analyses were done to compare selected sequences to osmoregulation genes in other *B. tabaci* species and aphids (Jing *et al.*, 2016). Protein sequences were aligned by ClustalW (Larkin *et al.*, 2007) implemented in MEGA X using default parameters. The best-fit amino acid substitution model to describe the substitution pattern was selected based on Bayesian information criterion

⁴ <https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi>

⁵ <http://www.whiteflygenomics.org>

(BIC) scores, computed using MEGA X (Kumar *et al.*, 2016; Kumar *et al.*, 2018). The best amino acid substitution model with the lowest BIC score was selected and used in the phylogenetic reconstruction implemented in Bayesian Evolutionary Analysis Sampling Trees (BEAST version 1.10.2) (Suchard *et al.*, 2018) with four gamma categories and strict molecular clock model at 1.0 clock rate. Yule process model (Gernhard, 2008) with uniform birth rate and inverse gamma as the gamma shape was used to provide prior on the tree. Markov Chain Monte Carlo (MCMC) were set with a chain length of 10,000,000. The generated tree was visualized using FigTree V1.4.3. The phylogenetic trees were used to study the relationship of selected genes with experimentally validated osmoregulation genes in other *B. tabaci* and aphid species (Price *et al.*, 2007; Shakesby *et al.*, 2009; Mathew *et al.*, 2011) to select the most suitable gene target.

3.2.7 Validation of selected genes using real-time quantitative PCR

Expression of the selected genes was validated using real-time quantitative PCR. In brief, the total RNA for whitefly gut, bacteriocytes whole-body samples were normalized to ensure an equal amount of starting RNA template (120 ng). Normalized RNAs were then treated with 1 µl of the ezDNaseTM enzyme (InvitrogenTM) for 2 minutes at 37°C to remove genomic DNA. The RNA sample was further incubated at 55°C for 5 minutes in the presence of 10 mM DTT to inactivate the enzyme. First-strand cDNA synthesis was carried out using SuperScriptTM III Reverse Transcriptase (InvitrogenTM) according to the manufacturer's instructions. A three-step thermocycling protocol was used with polymerase activation and DNA denaturation at 95°C for 3 minutes followed by 40 cycles of both denaturation at 95°C and annealing/extension at 55.2°C for 45 seconds, performed on C1000 Thermocycler with CFX96 Real-time detection system (Bio-Rad), using 60S ribosomal protein L13a (RPL13) and β-tubulin as the internal controls. A total of 20 µl of reaction mix containing 10 µl of iQTM SYBR[®] Green Supermix (2x), 1 µl of forward and reverse primer, 1 µl of cDNA and 7 µl of water was used. Melting curve analysis at the end of the reaction was done by increasing temperature from 55-95°C in increments of 0.5°C every 5 seconds to assess the dissociation characteristics of double-stranded DNA during heating.

Table 3.1: Primers used in the validation of gene expression of selected genes

Target gene	Primer sequence	Amplicon length (bp)	Tm (°C)
ENSSSA1UGT022145 (AQP1)	F - TTGTTTCGCAAGTTTGCCGT	90	59.83
	R - GACTGATTGACGCCCTGGAT		59.82
ENSSSA1UGT002057 (SUC1)	F - AACACTGCGAATAGCGCATC	86	59.35
	R - CGCCACTCTAGATGTTCGCA		60.18
ENSSSA1UGT002066 (SUC2)	F - CGGTAAGGTCTGAAACTGCGAT	110	60.86
	R - GTTTTGCTAGATGTGCAAGGCA		58.76
Diaminopimelate decarboxylase (<i>LysA</i>)	F - ACTACAATTCTCGCCCTCGC	86	60.18
	R - GTCATCAAAGGTCTCCCGCC		60.74
Arginosuccinate lyase (<i>ArgH</i>)	F - AAGCTCTGGTGTAAGGCACA	98	59.24
	R - AGGATGTCTTGGGTCGCTTC		59.75
Branched-chain-amino-acid aminotransferase (<i>BCAT</i>)	F - CGTCCAGAGTCAGTGGCAA	93	59.63
	R - GTTCATGGCTCCGGCTTCAG		61.37
Aspartate aminotransferase (<i>AAT</i>)	F - GGTCTTACCAGTTGTGCGAA	95	59.97
	R - AGAATGGAGCCGAACCCAAG		60.04
60S ribosomal protein L13a (<i>RPL13A</i>)	F - CATTCCACTACAGAGCTCCA	101	60.00
	R - TTTCAGGTTTCGGATGGCTT		60.00
β -Tubulin (β - <i>Tub</i>)	F - TGTCAGGAGTAACGACGTGTTTG	150	60.00
	R - TTCGGGAACGGTAAGTGCTC		60.00
4-hydroxy- tetrahydrodipicolinate reductase (<i>dapB</i>)	F - TGGTAAAAGACTACCAGGCGAA	118	59.37
	R - AGCTTGGTGTTTACAGCTGAGAG		60.81

3.3 RESULTS

3.3.1 RNAseq analysis for gut and bacteriocytes samples from cassava *B. tabaci*

A total of 144,377,532 sequencing reads were generated from guts, bacteriocytes and whole-body samples of cassava *B. tabaci* (SSA1-SG1) (Table 3.2). These samples had a high percentage mapping ranging from 78.0 to 83.8%. Of these, the bacteriocyte sample had the highest number of reads mapping on to the reference genome with a percentage mapping of 83.84%.

Table 3.2: Mapping quality of RNAseq reads from SSA1-SG1 whole-body, bacteriocytes and guts samples

Sample	Mapping to SSA1-SG1 genome		
	Mapped reads	Unmapped reads	Percentage mapping (%)
Whole-body	37,170,765	10,475,500	78.01
Bacteriocyte	39,660,655	7,644,558	83.84
Guts	38,603,036	10,823,018	78.10
Total	115,434,450	28,943,076	

3.3.2 Identification of transcripts enriched in the cassava *B. tabaci* (SSA1-SG1) gut and bacteriocytes

RNAseq analysis to identify genes with enriched expression in the gut and bacteriocytes relative to the whole-body revealed 3178 and 3212 transcripts that are differentially expressed (both up-regulated and down-regulated) in the gut and bacteriocytes respectively. From these, only transcripts with enriched expression ($\text{LogFC} > 1$) in the gut (1316 transcripts) and bacteriocyte (1909 transcripts) were selected for further analysis (Table 3.3, Figure 3.2). A considerable number of transcripts had a high fold difference (greater than 3) (Figure 3.3, A). A blastx search against the NCBI non-redundant (nr) database was done to identify the selected transcripts enriched in the cassava *B. tabaci* gut and bacteriocyte. All gut transcripts selected for further analysis (1316 transcripts) had NCBI blast hits (Figure 3.4). The species distribution analysis of all blastx top-hits revealed that 97.6% of the blastx top hits belonged to *Bemisia tabaci*. A few blast top hits

belonged to other species which included; *Cryptotermes secundus* Hill, *Myzus persicae* Sulzer, *Halyomorpha halys* Stal and *Bactrocera tryoni* Froggatt (Figure 3.5).

Table 3.3: Summary statistics of RNAseq analysis for genes expressed in *B. tabaci* gut and bacteriocytes

Description	Comparison	
	Gut/Whole-body	Bacteriocyte/Whole-body
Number of total transcripts tested	10027	10092
Number of transcripts with enriched expression	3178	3212
Number of transcripts with LogFC >1	1316	1909
Number of transcripts with LogFC <-1	1862	1303

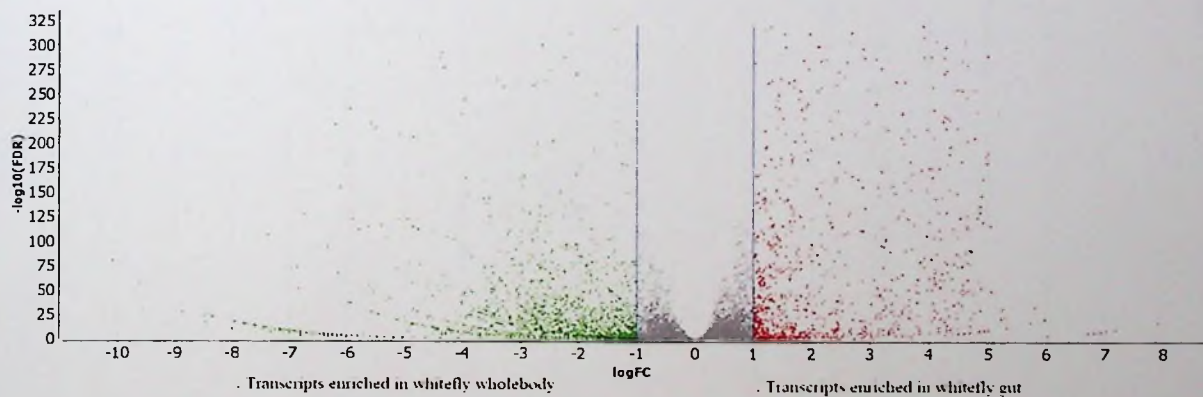


Figure 3.2: Differential gene expression in cassava *B. tabaci* (SSA1-SG1) gut. Volcano plot showing transcripts enriched in the *B. tabaci* SSA1-SG1 gut.

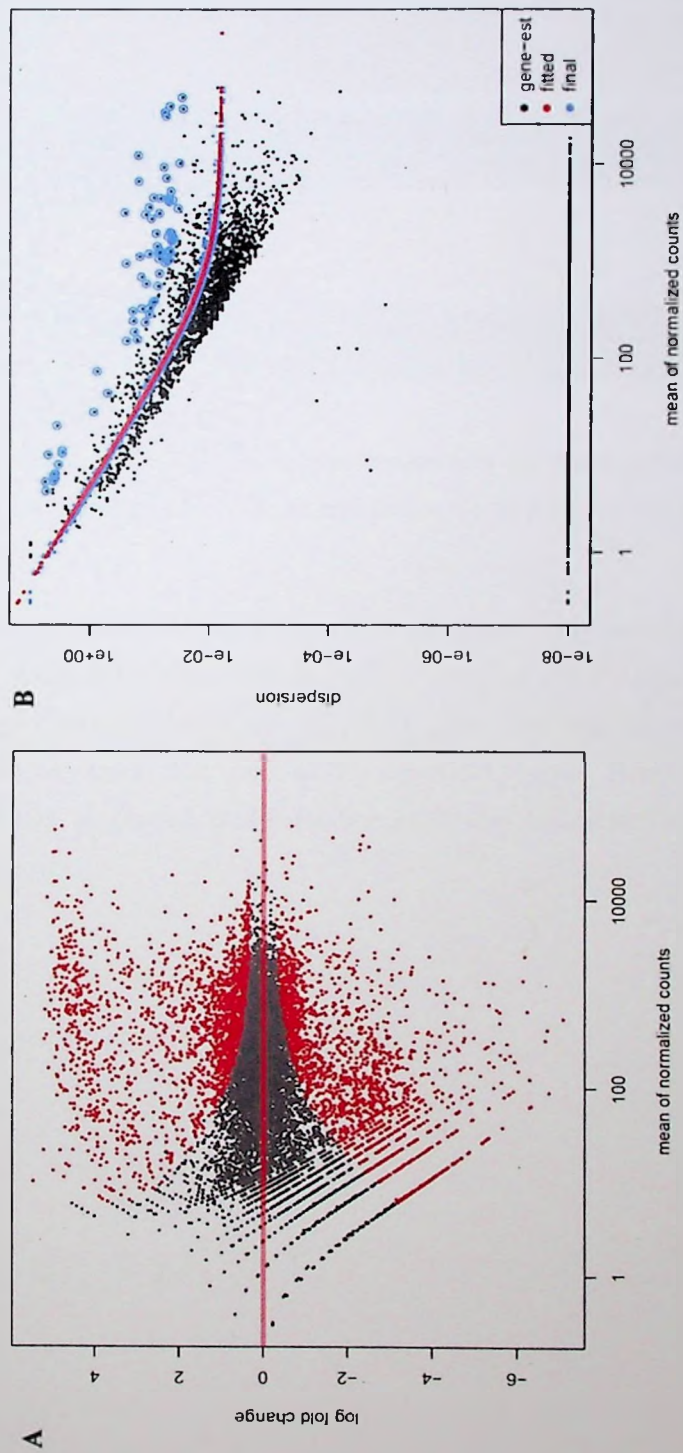


Figure 3.3: Differential gene expression in cassava *B. tabaci* (SSA1-SG1) gut. A) MA-plot showing a global view of differential expression in the *B. tabaci* gut compared to whole-body. B) Dispersion estimates in gene expression in the gut.

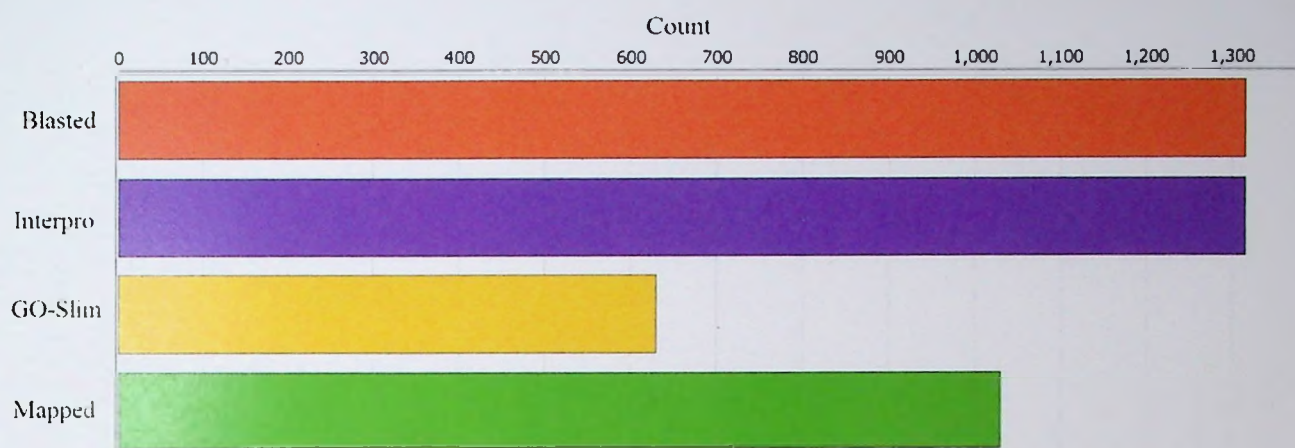


Figure 3.4: Tag distribution for the transcripts enriched in the cassava *B. tabaci* (SSA1-SG1) gut. A number of sequences with hits in the NCBI non-redundant database and InterproScan database.

In addition, a total of 1316 sequences were successfully scanned against Interpro signatures to identify functional domains, Interpro IDs, gene ontology IDs and names. Gene ontology (GO) mapping of protein blast hits revealed that only 1030 sequences had functional labels. Of these, only 630 sequences had specific GO annotation terms. Based on these functional labels and Interpro IDs, functional osmoregulation genes enriched in the cassava *B. tabaci* (SSA1-SG1) gut were selected.

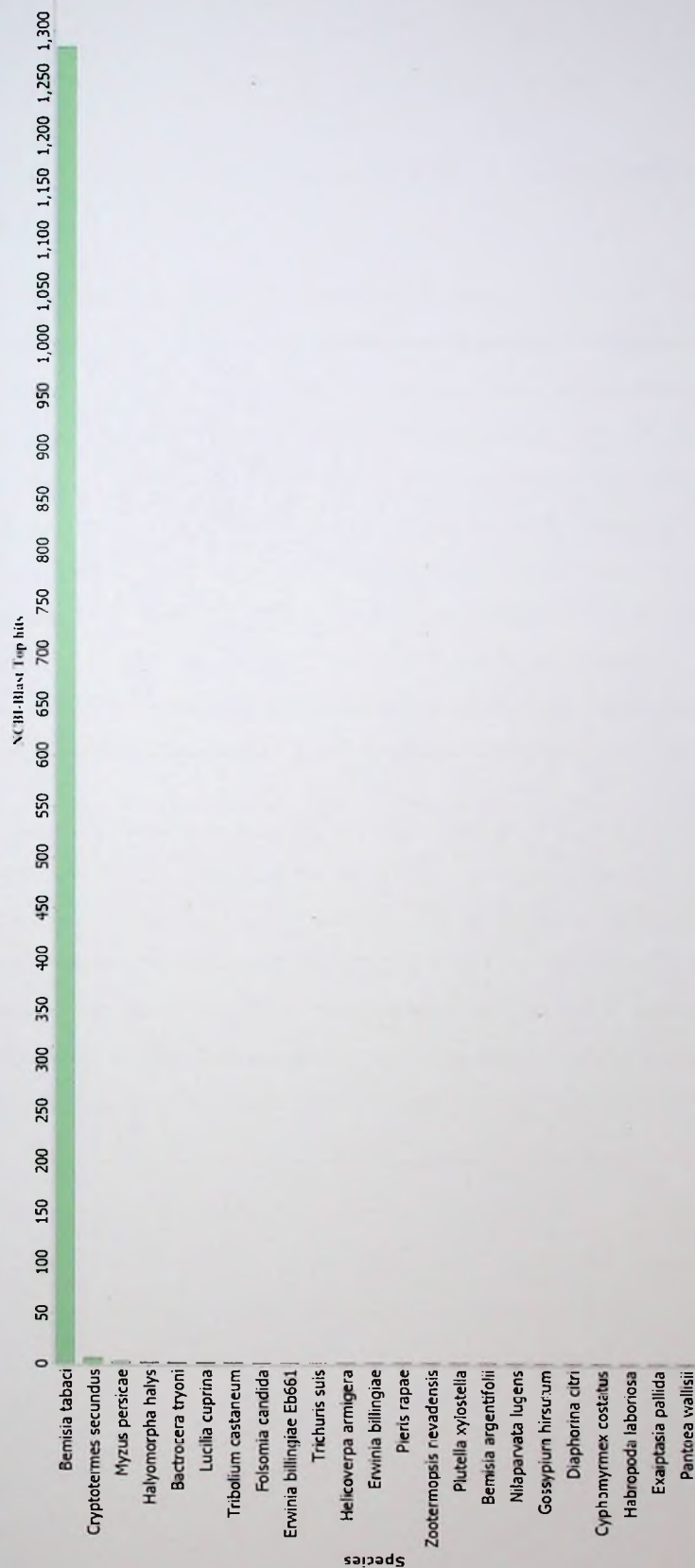


Figure 3.5: Species distribution of all the NCBI top blast-hits for the transcripts enriched in cassava *B. tabaci* (SSA1-SG1) gut.

3.3.3 Gene ontology categorization for the differentially expressed transcripts in *B. tabaci* gut

Annotation of differentially expressed transcripts in the cassava *B. tabaci* (SSA1-SG1) gut revealed roles and localization of these transcripts in terms of GO terms. The transcripts were characterized into three categories (i) cellular component (ii) molecular function and (iii) biological process. Based on the cellular component, the highest number of transcripts were expressed in the membrane (23%) and membrane part (20%). At the molecular level, a high percentage of transcripts (56% and 35%) in the *B. tabaci* (SSA1-SG1) gut encoded for proteins that mediate catalytic and binding activities respectively. These genes were mostly involved in metabolic and cellular biological processes (Figure 3.6). In summary, most of the genes enriched in cassava *B. tabaci* (SSA1-SG1) gut encode for catalytic, binding and transporter activities especially for metabolic processes and they are more expressed in the membrane and intracellular component of the cassava *B. tabaci* (SSA1-SG1) gut.

Direct GO count was also done to gain deeper insight into the biological roles of the transcripts enriched the cassava *B. tabaci* gut. The count showed that the highest number of transcripts encode for transmembrane transport proteins (95 transcripts), oxidation-reduction process (87 transcripts) and metabolic process (62 transcripts). This analysis also revealed a sucrose metabolic process as one of the top 15 biological processes in the cassava *B. tabaci* (SSA1-SG1) gut (Figure 3.7). In summary, a large percentage of transcripts (43%) enriched in the cassava *B. tabaci* (SSA1-SG1) gut encode for proteins that are involved in metabolic processes (Figure 3.7 & Figure 3.8)

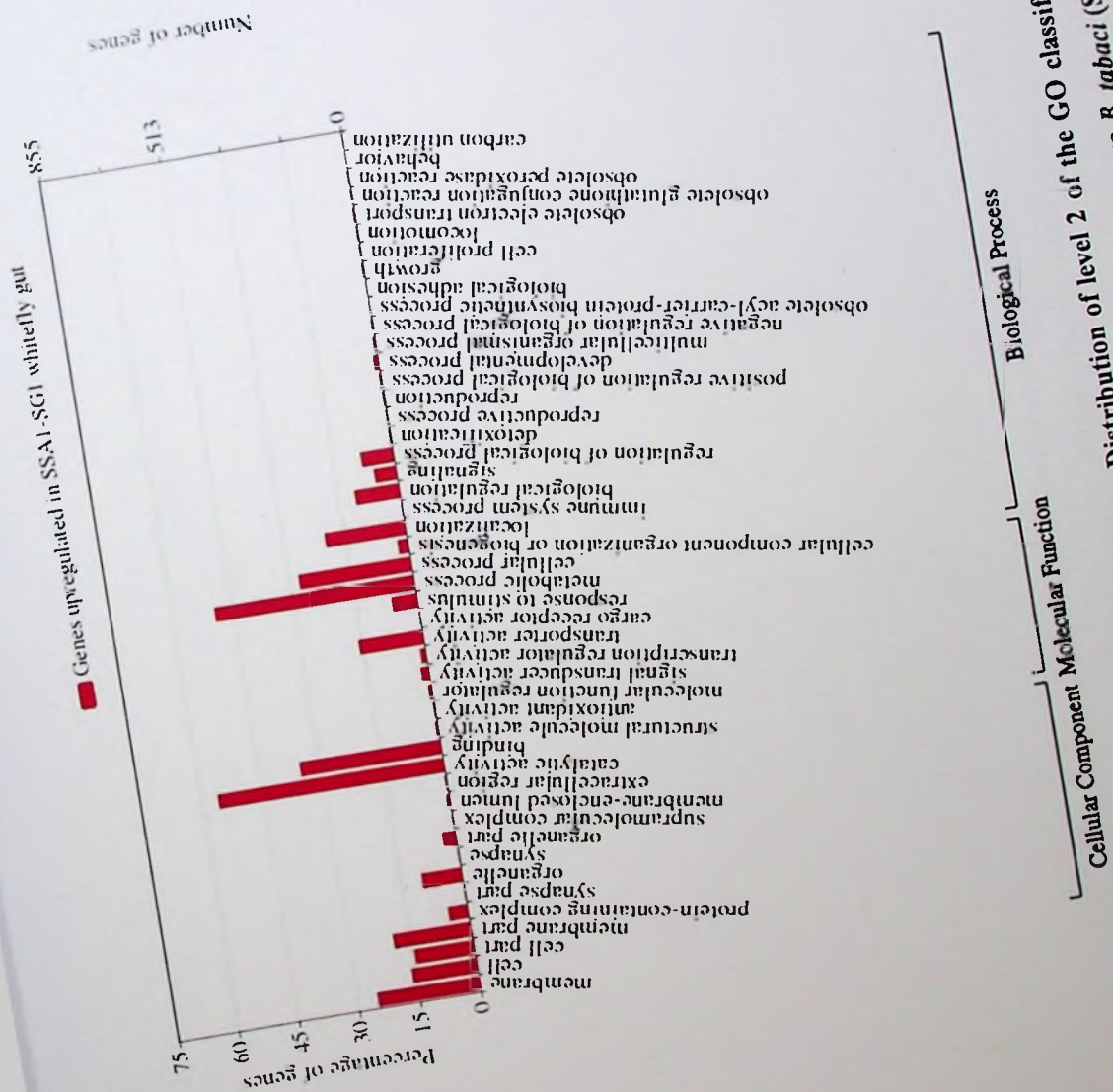


Figure 3.6: Gene ontology (GO) annotation. Distribution of level 2 of the GO classification for biological process, molecular function and cellular component of the transcripts enriched in the cassava *B. tabaci* (SSA1-SG1) gut.

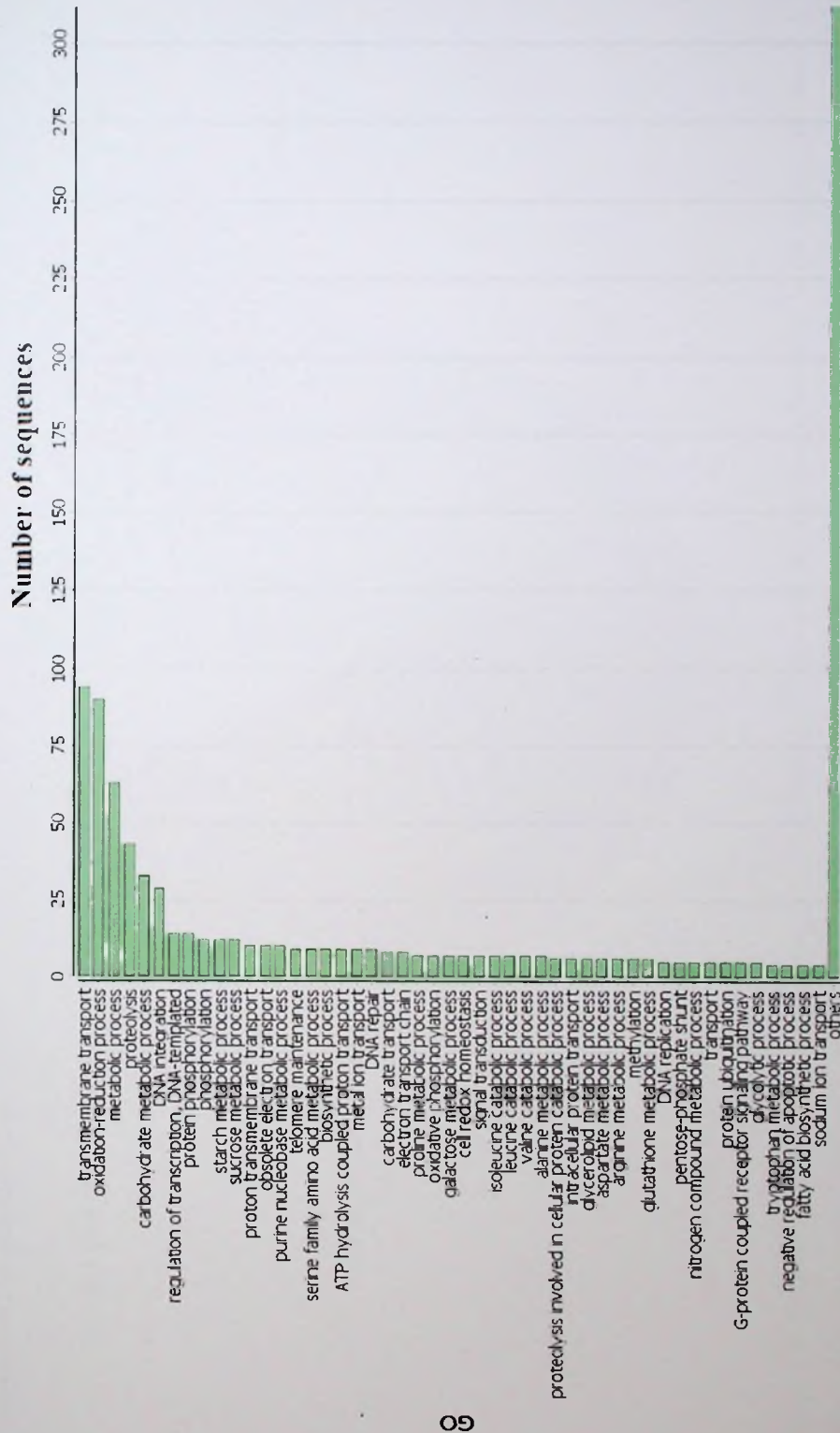


Figure 3.7: Direct GO count for biological process of transcripts enriched in cassava *B. tabaci* gut. Sucrose metabolic process was listed in the top 15 biological processes encoded by the sequences that are enriched in the cassava whitefly gut.

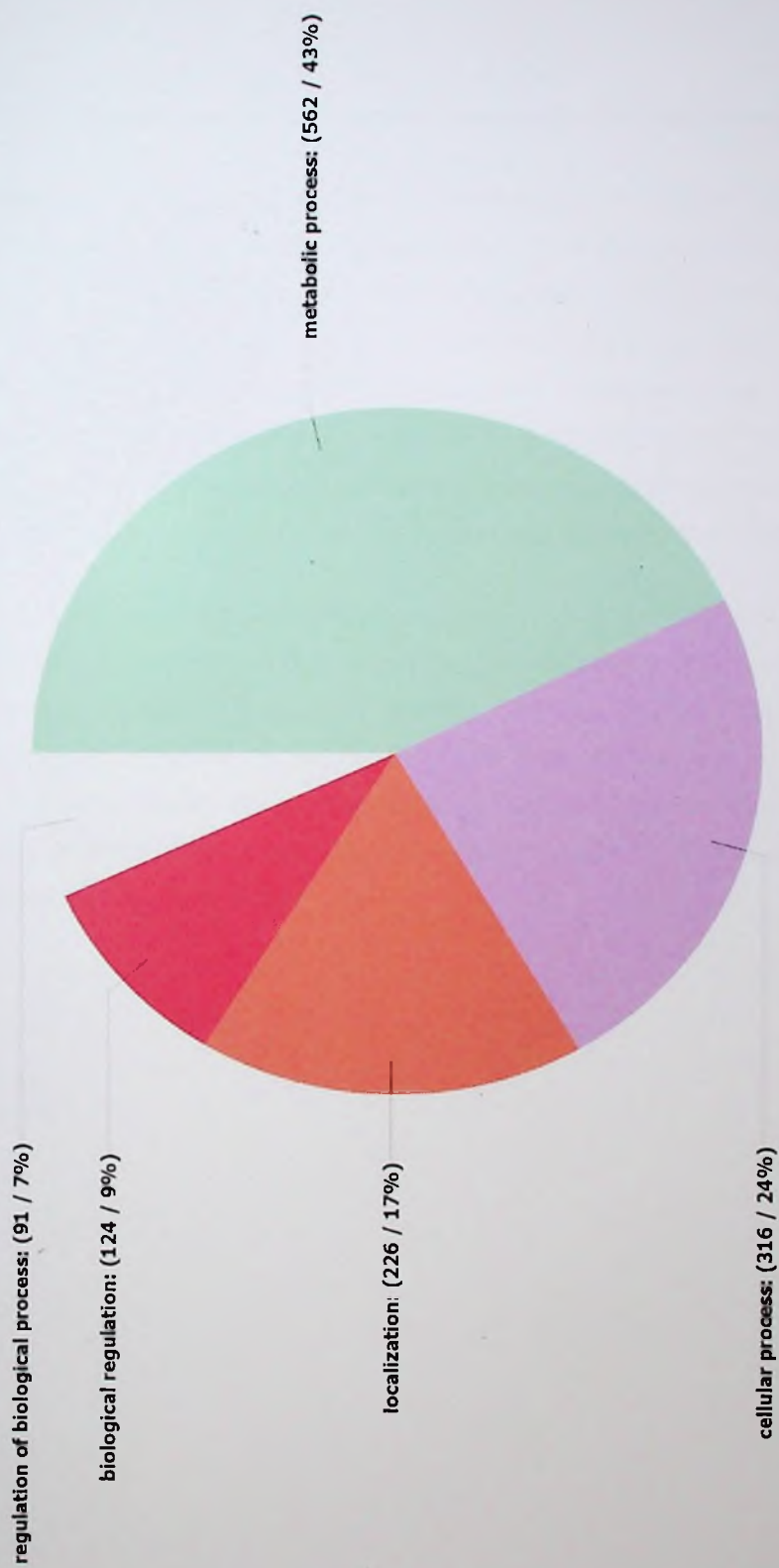


Figure 3.8: Gene ontology (GO) categorization, level 2 for Biological process of the transcripts enriched in the cassava *B. tabaci* (SSA1-SG1) gut. The largest proportion of transcripts enriched in the whitefly gut encode for metabolic processes.

3.3.4 Categorization of transcripts based on superfamilies and functional domains

Based on InterProScan IDs of the transcripts enriched in cassava *B. tabaci* (SSA1-SG1) gut, a total of 19 major superfamilies were characterized. The highest number of transcripts belonged to Major Facilitator Superfamily (MFS) transporter superfamily (140 transcripts) and major facilitator, sugar transporter-like (100 transcripts). Of the top eight superfamilies with the highest abundance of transcripts, three relate to metabolism and in particular, sugar metabolic processes. These include: MFS transporter superfamily (IPR036259), major facilitator-sugar transport like (IPR005828), UDP-glucuronosyl/UDP-glucosyltransferase (IPR002213), Alpha/Beta hydrolase fold (IPR029058) and glycoside hydrolase superfamily (IPR017853) (Figure 3.9)

An InterProScan domain distribution analysis for all transcripts enriched in the gut revealed the common and functional domains in the cassava whitefly gut. Based on this analysis, glycosyl hydrolase family 13, catalytic domain (IPR006047), was identified among the top five InterProScan domain for the transcripts enriched in cassava *B. tabaci* (SSA1-SG1) gut (Figure 3.10). Other InterProScan domains that relate to metabolism include; Major facilitator superfamily domain (IPR020846), peptidase C1A, papain C-terminal (IPR000668) and amino acid transporter, transmembrane (IPR013057).



Figure 3.9: InterProScan family distribution of the transcripts enriched in the cassava *B. tabaci* (SSA1-SG1) gut

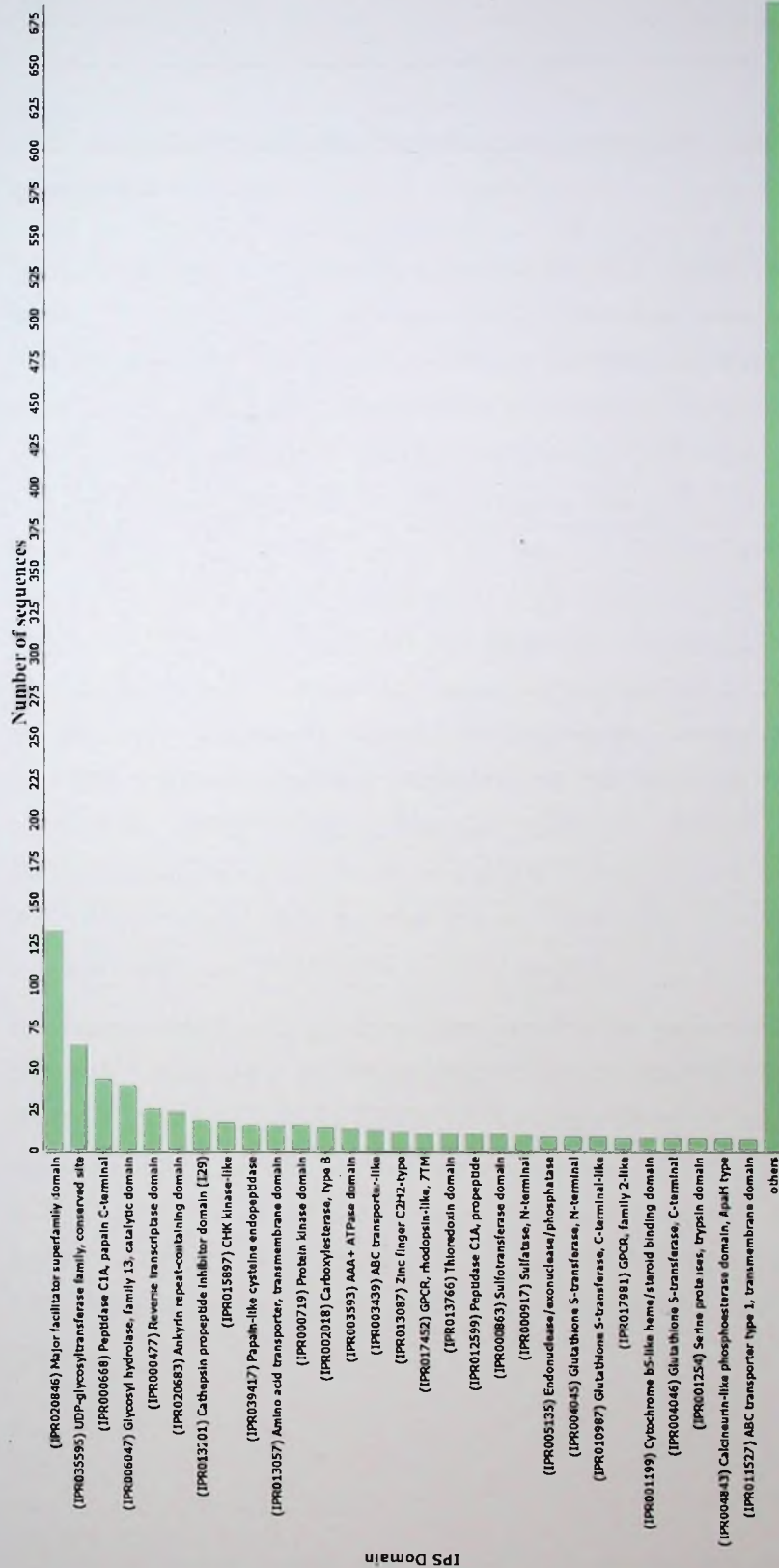


Figure 3.10: InterProScan domain distribution of the transcripts enriched in the cassava *B. tabaci* (SSA1-SG1) gut.

3.3.5 Identification of genes for sucrose metabolism and water recycling in cassava *B. tabaci* (SSA1-SG1) gut

Sucrose hydrolysis is mediated by α -glucosidase (EC 3.2.1.20), belonging to glycosyl hydrolase family 13 (IPR006047) (Jing *et.al.*, 2016). Transcripts with glycosyl hydrolase family 13 (IPR006047) catalytic domain were selected for functional analysis. A total of 24 transcripts encoding for sugar processing enzymes belonging to glycosyl hydrolase, family 13 were identified and characterized. These genes had high expression in the *B. tabaci* gut relative to the whole-body with log2 Counts Per Million (logCPM) ranging from 1.50 – 11.86 and fold difference ranging from 4.41 – 52.71-fold. Of these, gene ENSSSA1UGT025983 (Ssa01347), ENSSSA1UGT002057 (Ssa05154), ENSSSA1UGT024045 (Ssa07143), Ssa12230, ENSSSA1UGT006139 (Ssa02431), ENSSSA1UGT025983 (Ssa01347) and Ssa04510 had the highest expression with logCPM ranging from 10.18 – 11.86 and fold difference ranging from 17.15 – 29.65-fold in the *B. tabaci* gut relative to whole-body (Table 3.3). These genes corresponded to Bta07453, Bta05386, Bta12682, Bta15649, Bta03439, Bta04298 and Bta14419, orthologs of α -glucosidase gene in MEAM1 *B. tabaci* (*B. tabaci* Genome Database)⁶. Another gene, ENSSSA1UGT008267 (Ssa04743) had a relatively low expression (logCPM of 1.50) but with the highest fold differences (52.71-fold) in the *B. tabaci* gut relative to the whole-body. This gene corresponded to Bta06059, an α -glucosidase gene in MEAM1 *B. tabaci* species.

For water recycling in the gut, three genes encoding for major intrinsic protein (IPR000425) were identified. These were (i) ENSSSA1UGT022145 (Ssa03484⁷), (ii) ENSSSA1UGT021259 (Ssa02238) and (iii) ENSSSA1UGT023163 (Ssa08000) which has been identified through NCBI blastx as *AQP1*, *AQP4* and *AQP12* respectively. Of the three major intrinsic proteins, ENSSSA1UGT022145 had the highest expression in cassava *B. tabaci* (SSA1-SG1) gut with a logCPM of 8.49 and fold difference of 32.67-fold in *B. tabaci* gut relative to the whole-body (Table 3.4). This gene corresponded to Bta01973, a water specific *AQP1* gene in MEAM1 *B. tabaci* species.

⁶ <http://whiteflygenomics.org>

⁷ Genes in parenthesis are corresponding gene in the African cassava whitefly genome (Chen *et al.*, 2019)

Table 3.4: Candidate osmoregulation genes selected from differentially expressed genes in cassava *B. tabaci* (SSA1-SG1) gut

Gene ID***	Description	Domain		logCPM	Fold enrichment		Gene**
		Name	Interpro		Gut	P-value	
ENSSSA1UGT025983	Maltase A3-like	GH-13*	IPR006047	11.86	25.11	0.01	Bta07453
Ssa12250	Maltase A3-like	GH-13*	IPR006047	11.25	17.15	0	Bta15649
ENSSSA1UGT024045	Maltase 2-like	GH-13*	IPR006047	11.23	25.11	5e-278	Bta12682
ENSSSA1UGT002057	Maltase 2-like	GH-13*	IPR006047	11.05	21.71	0.01	Bta05386
Ssa04510	Maltase A3-like	GH-13*	IPR006047	10.44	29.65	0.01	Bta14419
ENSSSA1UGT025983	Maltase 2-like	GH-13*	IPR006047	10.9	27.67	0	Bta04298
ENSSSA1UGT006139	Maltase A3-like	GH-13*	IPR006047	10.18	24.93	4.6e-116	Bta03439
ENSSSA1UGT009050	Maltase A3-like	GH-13*	IPR006047	9.96	23.92	1.7e-57	Bta08425
ENSSSA1UGT002261	Maltase A3-like	GH-13*	IPR006047	9.93	23.42	1.5e-75	Bta05396
ENSSSA1UGT002066	Maltase A3-like	GH-13*	IPR006047	9.43	25.99	6.8e-79	Bta03818
ENSSSA1UGT013702	Maltase A3-like	GH-13*	IPR006047	9.12	20.25	0.01	Bta06458
Ssa04512	Maltase A3-like	GH-13*	IPR006047	9.04	17.27	4.9e-306	Bta14422
ENSSSA1UGT001313	Maltase A3-like	GH-13*	IPR006047	8.99	11.00	0.01	Bta07452
ENSSSA1UGT028740	Maltase A3-like	GH-13*	IPR006047	8.14	30.06	1.3e-166	Bta07377
ENSSSA1UGT021237	Maltase A3-like	GH-13*	IPR006047	7.35	25.46	0.01	Bta07764

GH-13* - (Glycosyl hydrolase - Family 13)
Gene** - Ortholog in Bemisia tabaci (MEAM1)
Gene ID*** - (scaffold_xxxxxxF_transcript_id_xxxx)

Table 3.4 continued

Gene ID***	Description	Domain/Family		logCPM	Fold enrichment		P-value	Gene**
		Name	Interpro		Gut			
ENSSSA1UGT006406	Maltase A3-like	GH-13*	IPR006047	7.17	7.36	0.01	Bta01478	
ENSSSA1UGT005553	Maltase A3-like	GH-13*	IPR006047	7.12	22.63	0	Bta08427	
ENSSSA1UGT001811	Maltase A3-like	GH-13*	IPR006047	5.07	20.25	9.0e-192	Bta09696	
ENSSSA1UGT031126	Maltase A3-like	GH-13*	IPR006047	4.74	17.27	1.8e-146	Bta10022	
ENSSSA1UGT021697	Maltase	GH-13*	IPR006047	4.35	28.25	3.3e-127	Bta09633	
ENSSSA1UGT000627	Maltase A3-like	GH-13*	IPR006047	3.25	4.41	1.2e-24	Bta13914	
ENSSSA1UGT003669	Maltase A3-like	GH-13*	IPR006047	2.71	8.51	3.2e-11	Bta11358	
ENSSSA1UGT010695	Maltase A3-like	GH-13*	IPR006047	2.04	28.25	1.7e-23	Bta05340	
ENSSSA1UGT008267	Maltase A3-like	GH-13*	IPR006047	1.50	52.71	1.2e-19	Bta06059	
ENSSSA1UGT022145	Aquaporin 1	Major intrinsic protein	IPR000425	8.49	32.67	0.01	Bta01973	
ENSSSA1UGT021259	Aquaporin 4	Major intrinsic protein	IPR000425	5.17	2.20	3.9e-84	Bta07505	
ENSSSA1UGT023163	Aquaporin 12	Aquaporin 11/12	IPR016697	3.31	2.07	2.2e-08	Bta14320	

GH-13* - (Glycosyl hydrolase - Family 13)
Gene** - Ortholog in *Bemisia tabaci* (MEAM1)
Gene ID** - (scaffold_XXXXXXF_transcript_id_XXXXX)

3.3.6 Structural analysis for residues of the hallmarks of the water-specific aquaporins

In addition to sequence identity, the selected sequences for aquaporins were further analyzed to identify hallmarks of the water-specific aquaporins: the aromatic/arginine (ar/R) region and the asparagine-proline-alanine (NPA) boxes in loop B and Loop E. This analysis revealed that gene ENSSSA1UGT022145 had a similar structure of ar/R and NPA as *BtAQPI*, an experimentally validated water-specific gut aquaporin gene in *B. tabaci* MEAM1 (Mathew *et.al.*, 2011). Gene ENSSSA1UGT022145 ar/R filter comprised of phenylalanine at position 71, histidine at position 198, alanine at position 207 and arginine at position 213 which is characteristic to water-specific aquaporin. Its NPA had an asparagine at position 91 in loop B and position 210 in loop E (Table 3.5).

Both ENSSSA1UGT021259 (Ssa02238) and ENSSSA1UGT023163 (Ssa08000) showed variation in their ar/R filter, with ENSSSA1UGT021259 having a leucine at position 196 instead of histidine at position 198 and serine at position 205 instead of alanine at position 207. It also showed differences in its NPA structure with asparagine at position 88 in loop B and at position 208 in loop E. This variation has been reported to be due to a mutation in the ar/R selectivity filter of water-selective channel, forming entomoglyceroporins (EgIps), which are more closely related to the classical aquaporin 4-type channel (Finn *et al.*, 2015). Gene ENSSSA1UGT023163 was truncated, therefore it lacked both ar/R region and NPA boxes.

Phylogenetic analysis based on Bayesian approach showed that ENSSSA1UGT022145 clustered with *BtAQPI*, indicating that ENSSSA1UGT022145 is the water specific *AQPI* gene enriched in the cassava *B. tabaci* (SSA1-SG1) gut (Figure 3.11). Both ENSSSA1UGT021259 and ENSSSA1UGT023163 clustered with APA28759.1 and APA28762.1 respectively. This analysis confirmed that ENSSSA1UGT021259 (Ssa02238) is an entomoglyceroporins (*AQP4*) while ENSSSA1UGT023163 is an aquaporin 7 (*BtAQPI2L*).

Table 3.5: NPA motif and ar/R filter of selected aquaporin protein sequences of cassava *B. tabaci* (SSA1-SG1)

Gene ID	Gene	Ar/R Filter	NPA signature motif		
			Loop B	Loop E	
ENSSSA1UGT022145	Bta01973	F (71) H (198) A (207) R (213)	NPA (91)	NPA (210)	
ENSSSA1UGT021259	Bta07505	F (68) L (196) S (205) R (211)	NPA (88)	NPA (208)	
ENSSSA1UGT023163	Bta14320	-	-	-	
BtAQPI*		F (71) H (198) A (207) R (213)	NPA (91)	NPA (210)	

*Experimentally validated water-specific gut aquaporin gene of *B. tabaci* (Mathew *et al.*, 2011)

The number in parenthesis is the position of the respective amino acid in Ar/R filter and asparagine of the NPA motif for each aquaporin protein sequence

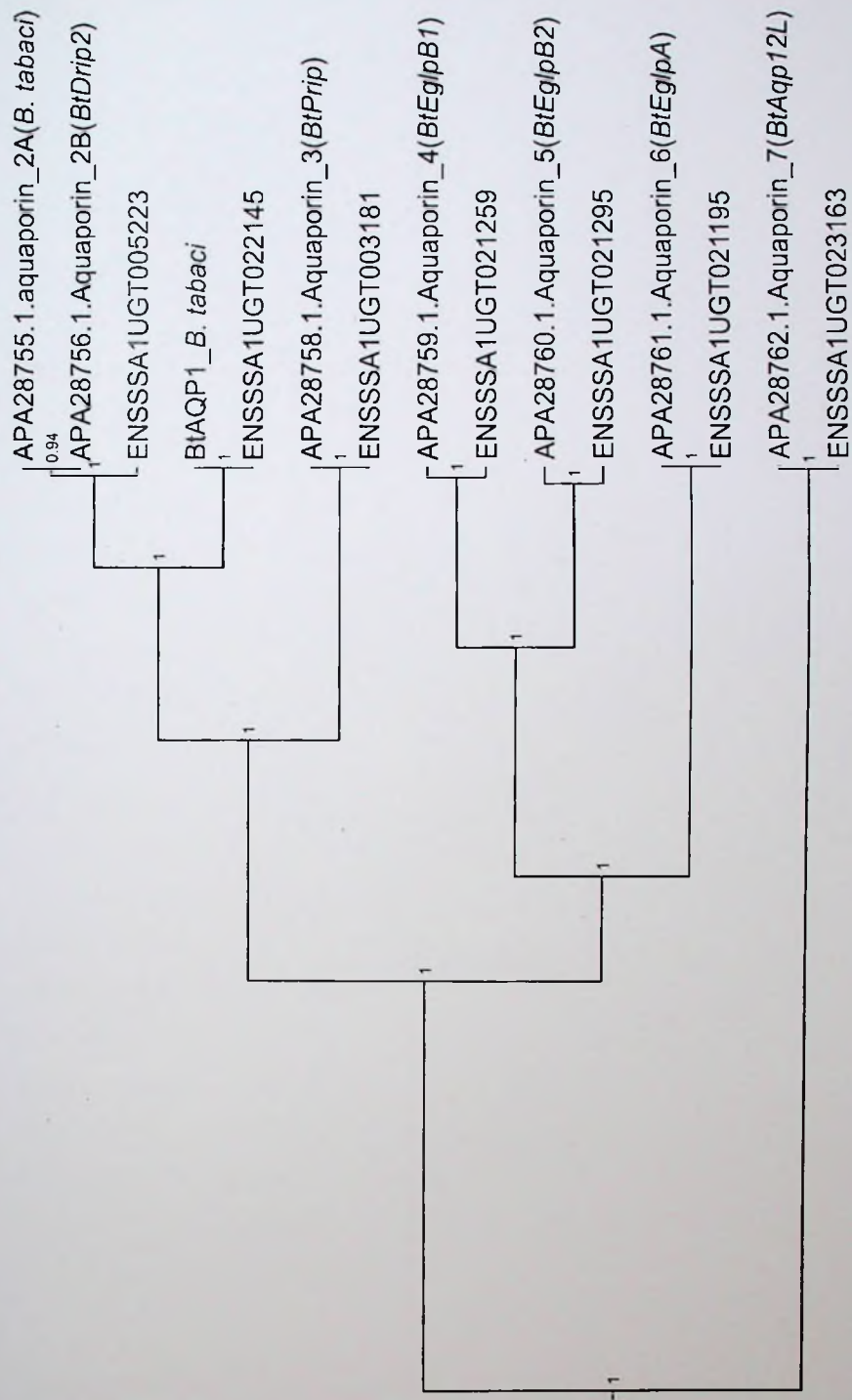


Figure 3.11: Phylogenetic relationship of all aquaporin genes of cassava *B. tabaci* (SSA1-SG1) and experimentally validated aquaporins (Van Ekert *et al.*, 2016). Phylogenetic analysis was done using Bayesian method implemented in Bayesian Evolutionary Analysis Sampling Trees (BEAST version 1.10.2). Scalebar represent branch length, an inference of genetic difference between the analyzed sequences (number of substitutions per site)

3.3.7 Identification of α -glucosidase proteins with both the catalytic residue and predicted signal peptide

A total of 24 protein sequences belonging to glycosyl hydrolase family 13 were identified using sequence similarity methods. Analysis of amino acid residues was done to identify protein sequences with both predicted signal peptide and catalytic site residue, characteristic to α -glucosidase of GH-13. Of the 24 α -glucosidase protein sequences selected, only 14 had both a predicted signal peptide and catalytic site residue. These included: (i) Ssa04510 (Bta14419⁸), (ii) ENSSSA1UGT005553 (Bta08427), (iii) ENSSSA1UGT002057 (Bta05386), (iv) ENSSSA1UGT010695 (Bta05340), (v) ENSSSA1UGT021697 (Bta09633), (vi) ENSSSA1UGT008267 (Bta06059), (vii) ENSSSA1UGT000627 (Bta13914), (viii) ENSSSA1UGT021237 (Bta07764), (ix) ENSSSA1UGT002066 (Bta03818), (x) Ssa12230 (Bta15649), (xi) ENSSSA1UGT006139 (Bta03439), (xii) Ssa04512 (Bta14422), (xiii) ENSSSA1UGT013702 (Bta06458) and (xiv) ENSSSA1UGT009050 (Bta08425). All these protein sequences had aspartic acid as the nucleophile and glutamic acid as the proton donor located at varying positions in each protein sequence except for one sequence ENSSSA1UGT013702 (Bta06458) which had glutamic acid at position 229 and arginine as a proton donor at position 287 (Table 3.6).

The phylogenetic analysis for the selected protein sequences with both predicted signal peptide and catalytic site residue also revealed three clusters (Figure 3.13), which may relate to the substrate (sucrose, maltose and other sugars) specificity of these enzymes. A total of three genes clustered with an experimentally validated sucrase gene (Price *et al.*, 2007). These included; Ssa12230, ENSSSA1UGT002057 and ENSSSA1UGT002066. All these had a very high level of expression with fold difference ranging between 17.15 – 25.99 in the cassava *B. tabaci* (SSA1-SG1) gut relative to whole-body. Based on NCBI conserved protein domain family, the three sequences; Ssa12230, ENSSSA1UGT002057 (Ssa05154) and ENSSSA1UGT002066 (Ssa05164) along with the experimentally validated sucrase had an AmyAc_Maltase (cd11328) catalytic domain. These are potential α -glucosidase genes in cassava *B. tabaci* (SSA1-SG1) that hydrolyze sucrose to glucose and fructose.

⁸ Gene in parenthesis represent a corresponding ortholog in *B. tabaci* MEAM1

3.3.8 Identification of sucrose hydrolase in cassava *B. tabaci* (SSA1-SG1) gut

All selected protein sequences had variation in seven amino acid residues of the partial sequence of the conserved region II of α -glucosidase family 13. A conserved region II is a region on the protein sequence where the catalytic site of alpha-glucosidase begins. The protein sequences analyzed exhibited a sequence motif DAXxxxx except for ENSSSA1UGT013702 (Ssa11469) which had a different sequence motif ETxxxxx. Partial sequences of conserved region II of all the selected protein sequences were analyzed and compared with that of the experimentally validated sucrose to identify the enzyme that specifically hydrolyzes sucrose in the whitefly gut. Only one protein sequence ENSSSA1UGT002057 (Ssa05154) had the closest configuration of the seven amino acid (DAVPYLF) to that of the experimentally validated *sucrase/transglucosidase* of the pea aphid (DAVNYLF) (Price *et al.*, 2007) (Table 3.6, Figure 3.12)

ENSSSA1UGT028740 (Ssa12486)	PMVDLGVDITDFRAVDPIFGTMDDFKRVVQGIHERGMRIILDVFPNHTSNTCEWFKLSER	140
NP_001119607.1_Aphid	PQNDFGYDISDYKEIDPIYGSMDFERMRDEFHKGKIKVLLDFVPNHTSDEHEWFKSIK	148
ENSSSA1UGT002066 (Ssa05164)	PMADFGYDISDFRAIEPMFGTMDFERLKRFLHKNGLKMIILDVFPNHTSDEHDFKKSVA	141
ENSSSA1UGT002057 (Ssa05154)	PMADFGYDISNYTEINEIFGNLDDLHLREKLHARGVKLIILDVFPNHTSNEHPWFLNSVK	180
Ssa12230 (Bta15649)	PGLDLGYDISNYTRIDPLFGTLEDLEYLRDALHKAGIRLLLDVFPNHTSDEHPWFQWSIK	147

ENSSSA1UGT028740 (Ssa12486)	RIEPTYNFYVWQDPKQLNDSHWYPPNNWRSKFA-DSAWTSEVRKQYYHFAEQQNDVN	199
NP_001119607.1_Aphid	KIEPFSDYYVWKDPIRDVHGNNTPPSNWLGVFNSGSAWEWNEERQYYLHQFQVKQPDNL	208
ENSSSA1UGT002066 (Ssa05164)	RVDPTYDYNNVNGKMSNGTIQPPNNWLSYF-GGSAWTWNEKRKQYYLHQFHPKQPDNL	200
ENSSSA1UGT002057 (Ssa05154)	KIDPTYNYVWRDPIVDANGKSPNNWLSIFNSGSAWEWNEERQYYLHQFAVQQPDNL	240
Ssa12230 (Bta15649)	KVDPTYNYVWKAQM-VNGQRKPPTNWISFEG-CSMWKWNDEKQYYFLHQFHKQPDNL	205

Conserved Region II		
ENSSSA1UGT028740 (Ssa12486)	YRDPHAHEEMQSVLRFWLDMGVDGFRVDAVEYLYEATHLMDEPLRENFNGNISWEALSH	259
NP_001119607.1_Aphid	YRNPVSREEIKNTLLYLWLGKVDGFRFDVAVNYLFEEDLADEPKSNKIGYLDTDYDSLTH	268
ENSSSA1UGT002066 (Ssa05164)	YRNPVQEMKDVLTLYWMDKGVDFRMDAVMTIMEDIKLDEPLSGKTDVLPDDEEYLNH	260
ENSSSA1UGT002057 (Ssa05154)	YNCKELVEEMKEVLRFWLSHGVDGFRMDAVPYLFEDSLFRNEPYVDENIKNSKMYDSLTH	300
Ssa12230 (Bta15649)	YECKELVQEMDVRFWLDRGVDGLRLDAINFMFEDPLWRDQELLAPDLP-NFFLSYNR	264

ENSSSA1UGT028740 (Ssa12486)	IYTHQPNENAEILQSWRALDEYAQKDG-QARLLCTESYNPPKILEQQLGNDTHPGAQIP	318
NP_001119607.1_Aphid	TSTLDQPEYTSIVRQWRQMLDSYRTREK-KTKFMVCEYSPFDKTLLYGNSNPGAHP	327
ENSSSA1UGT002066 (Ssa05164)	IYTRDQPGTYEILKQFRKHLDDYSRKTH-TVKEMATEAYSNLSTMKYYGTDKNPGAHT	319
ENSSSA1UGT002057 (Ssa05154)	IYSMDRPKTYEMIGQFRKVLDDFAKEHNTPPKVLSEAYTSLNYTMKYGDADKNPGSHMP	360
Ssa12230 (Bta15649)	SRMDQPATYEMVTRPREIFDSYSKKDG-RTRVMTEAYTTLDKTLGFYQYNGKPGAHP	323

Figure 3.12: Multiple sequence alignment of selected α -glucosidase protein sequences. The highlighted region is the partial sequence of conserved region II of α -glucosidase family 13. The function and substrate preference of α -glucosidase enzyme is determined by the amino acid residue at position 4 and 5 of the conserved region II. Protein sequences were aligned using Clustal Omega⁹ software

⁹ <https://www.ebi.ac.uk/Tools/msa/clustalo/>

Table 3.6: Genes encoding proteins with characteristic catalytic site residue and a signal peptide for α -glucosidase enzyme

Gene ID	Gene*	Fold difference	Signal peptide	D-Score**	Cleavage site	Nucleophile	Proton Donor	Conserved region II
ENSSSA1UGT008267	Bta06059	52.71	+	0.760	A(19)F	D(224)	E(289)	DAPGWLM
ENSSSA1UGT028740	Bta07377	30.06	+	0.768	S(20)R	D(227)	E(295)	DAVEYLY
Ssa04510	Bta14419	29.65	+	0.474	S(20)H	D(230)	E(297)	DAVNHL
ENSSSA1UGT010695	Bta05340	28.25	+	0.739	G(21)H	D(183)	E(250)	DAARRHFF
ENSSSA1UGT021697	Bta09633	28.25	+	0.774	C(20)N	D(227)	E(295)	DAPEFIF
ENSSSA1UGT025983*	Bta04298	27.67	-	0.195	-	D(19)	E(86)	DAVTYMY
ENSSSA1UGT002066	Bta03818	25.99	+	0.843	C(16)R	D(228)	E(296)	DAVMTIM
ENSSSA1UGT021237	Bta07764	25.46	+	0.512	E(20)E	D(257)	E(289)	DAVQILF
ENSSSA1UGT025983*	Bta07453	25.11	-	0.103	-	D(65)	E(132)	DAVTYMY
ENSSSA1UGT024045	Bta12682	25.11	-	0.119	-	D(288)	E(344)	DAVPHLI
ENSSSA1UGT006139	Bta03439	24.93	+	0.833	A(22)Q	D(227)	E(232)	DAAKWLF
ENSSSA1UGT009050	Bta08425	23.92	+	0.825	G(20)I	D(231)	E(290)	DAVPWLY
ENSSSA1UGT002261	Bta05396	23.43	-	0.329	-	D(244)	E(312)	DAIKHLV
ENSSSA1UGT005553	Bta08427	22.63	+	0.860	G(20)V	D(227)	E(286)	DAVVCLY
ENSSSA1UGT002057	Bta05386	21.71	+	0.743	A(24)V	D(268)	E(337)	DAVPYLF
ENSSSA1UGT013702	Bta06458	20.25	+	0.765	G(20)G	E(229)	R(287)	ETVSYLF

* Ortholog in *Bemisia tabaci* (MEAM1) ** Discrimination score (D-score) for a best-predicted signal peptide for predicted proteins of GH13 family

*** Catalytic site residue in each GH13 sequence. The position of the Nucleophile residue of the partial sequence of conserved region II is indicated by the number in the parentheses

Presence or lack of signal peptide denoted as + or - respectively

Table 3.6 continued

Gene ID	Gene****	Fold difference	Signal peptide	D-Score**	Cleavage site	Nucleophile	Proton Donor	Catalytic site residue***
ENSSSA1UGT001811	Bta09696	20.25	+	0.742	A(25)Q	-	-	-
Ssa04512	Bta14422	17.27	+	0.799	Q(23)S	D(228)	E(296)	DALKHVV
ENSSSA1UGT031126*	Bta10022	17.27	-	0.147	-	-	-	-
Ssa12230	Bta15649	17.15	+	0.693	E(21)F	D(233)	E(300)	DAINFMF
ENSSSA1UGT001313	Bta07452	11.00	-	0.428	-	D(227)	E(294)	DAVAVLF
ENSSSA1UGT003669*	Bta11358	8.51	+	0.810	A(20)D	-	-	-
ENSSSA1UGT006406	Bta01478	7.36	-	0.105	-	D(281)	E(369)	DAVHTMF
ENSSSA1UGT000627	Bta13914	4.41	+	0.805	S(17)T	D(220)	E(288)	DAIPILF
NP_001119607.1_Aphid	-	-	+	0.781	S(21)E	D(236)	E(304)	DAVNVLF

* Partial sequence ** Discrimination score (D-score) for a best-predicted signal peptide for predicted proteins of GH13 family

*** Catalytic site residue in each GH13 sequence. The position of the residue is indicated by the number in the parentheses. **** Ortholog in *Bemisia tabaci* (MEAM1)

Presence or lack of signal peptide denoted as + or - respectively

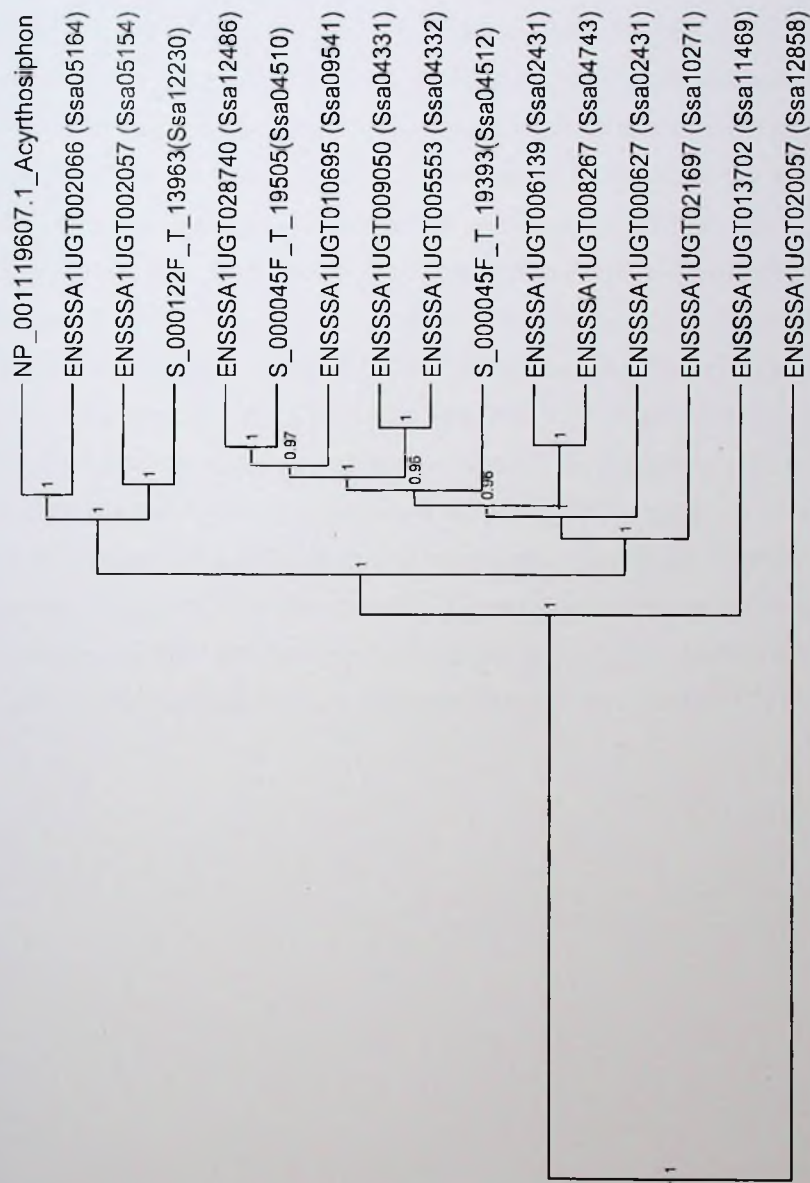


Figure 3.13: Phylogenetic relationships of α -glucosidases protein sequences with both predicted signal peptides and catalytic site residue and experimentally validated Sucrase gene in *Acyrthosiphon pisum* (Price *et.al.*, 2007). Genes in parenthesis are the orthologs of alpha-glucosidase family 13 in MEAM1 *B. tabaci* species. Phylogenetic analysis was done using Bayesian method implemented in Bayesian Evolutionary Analysis Sampling Trees (BEAST version 1.10.2). Scalebar represent branch length, an inference of genetic difference between the analyzed sequences (number of substitutions per site).

3.3.8 Identification of symbiosis genes in cassava *B. tabaci* (SSA1-SG1)

In *B. tabaci*, biosynthesis of essential amino acids requires interaction between the host and endosymbionts in the bacteriocyte. RNAseq analysis was conducted to identify critical genes that facilitate the shared metabolic interaction in *B. tabaci* SSA1-SG1. A total of 3212 transcripts were differentially expressed in both bacteriocyte and cassava *B. tabaci* (SSA1-SG1) whole-body (Table 3.3). Of these, a total of 1909 transcripts that are enriched in the cassava *B. tabaci* (SSA1-SG1) bacteriocyte were selected for further analysis. Both volcano and MA-plot showed a wide dispersion showing a high level of differences in gene expression between bacteriocytes and whole-body samples (Figure 3.14, Figure 3.15). Blastx search against NCBI non-redundant (nr) database was done to identify the transcripts enriched in the cassava *B. tabaci* (SSA1-SG1) bacteriocyte. Only one sequence did not have NCBI-blast hit (Figure 3.16). Top-hit species distribution analysis revealed that a large percentage (93%) of the top hits belonged to *Bemisia tabaci* (Figure 3.17). A small number of hits belonged to *Candidatus Portiera aleyrodidarum*, *Candidatus P. aleyrodidarum* BT-B-HRs, *Candidatus P. aleyrodidarum* MED and *Candidatus P. aleyrodidarum* BT-QVLC, resident endosymbiont in bacteriocytes of different *B. tabaci* species. Furthermore, the analysis was done to identify the functional signatures of the transcripts that are enriched in the *B. tabaci* bacteriocyte. A total of 1909 sequences were successfully scanned against Interpro signatures to identify and characterize functional domains and GO terms. The blasted protein sequences with hits were then mapped against extensively curated gene ontology annotated proteins. Of the 1909 Interpro scanned transcripts, only 1419 had functional labels and 817 had specific GO annotation terms at an annotation score cutoff of 55 (Figure 3.18).

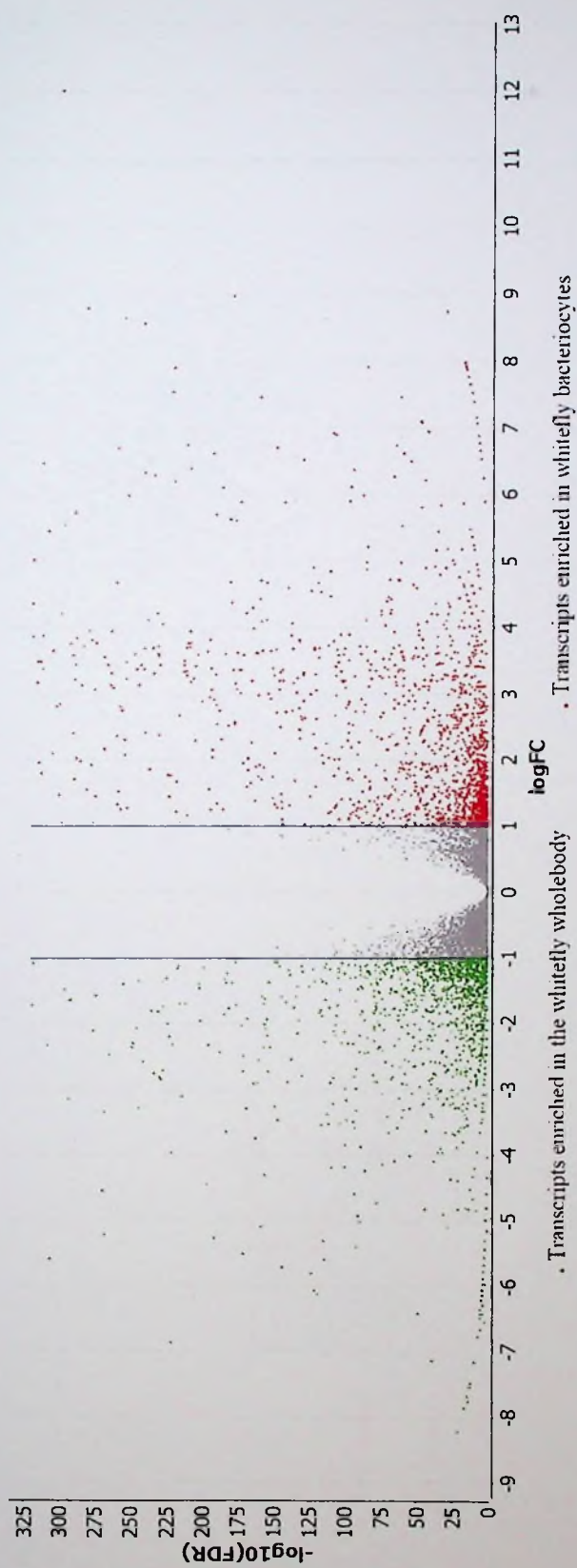


Figure 3.14: Differential gene expression in cassava *B. tabaci* (SSA1-SG1) gut. Volcano plot showing genes enriched in the bacteriocyte.

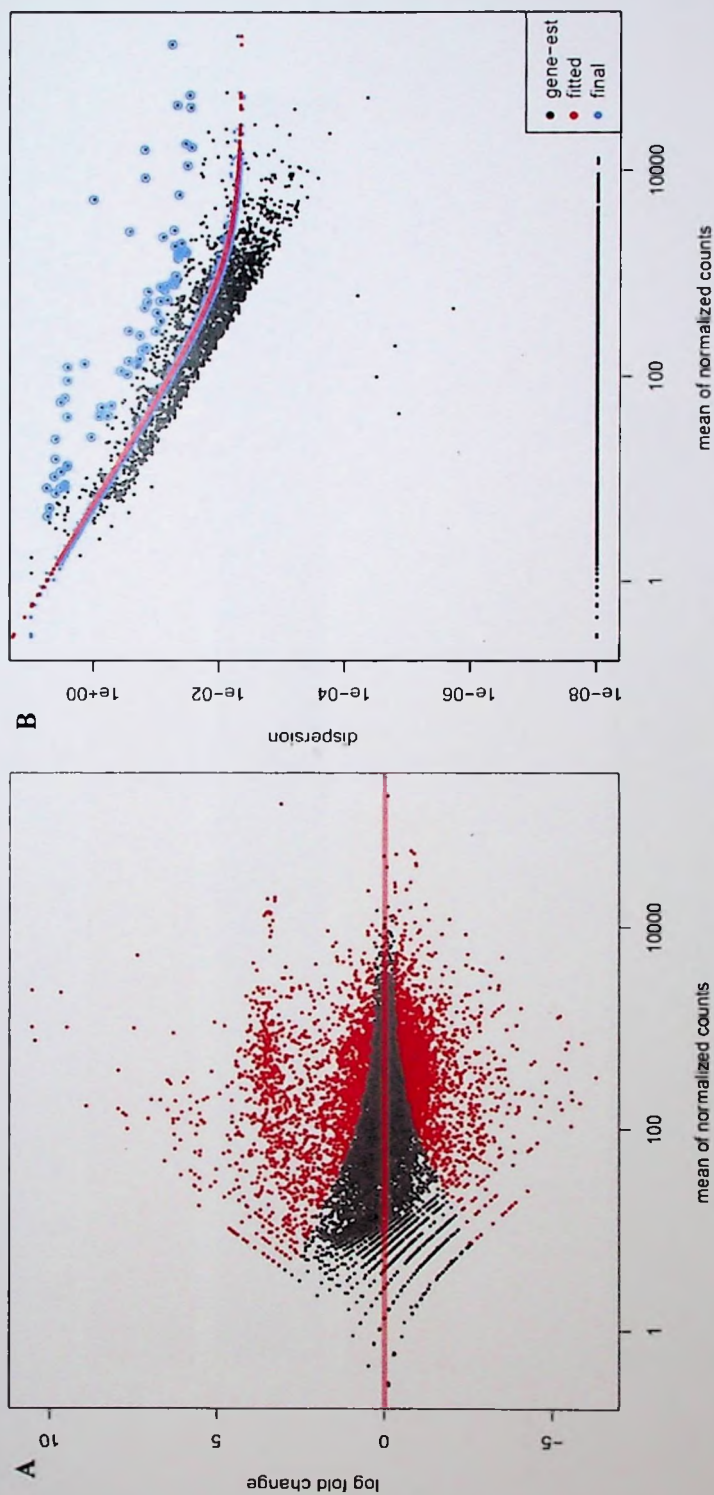


Figure 3.15: Differential gene expression in cassava *B. tabaci* (SSA1-SG1) gut. A) MA-plot showing a global view of differential expression in the *B. tabaci* bacteriocytes compared to whole-body. B) Dispersion estimates of gene expression in cassava *B. tabaci* bacteriocytes.

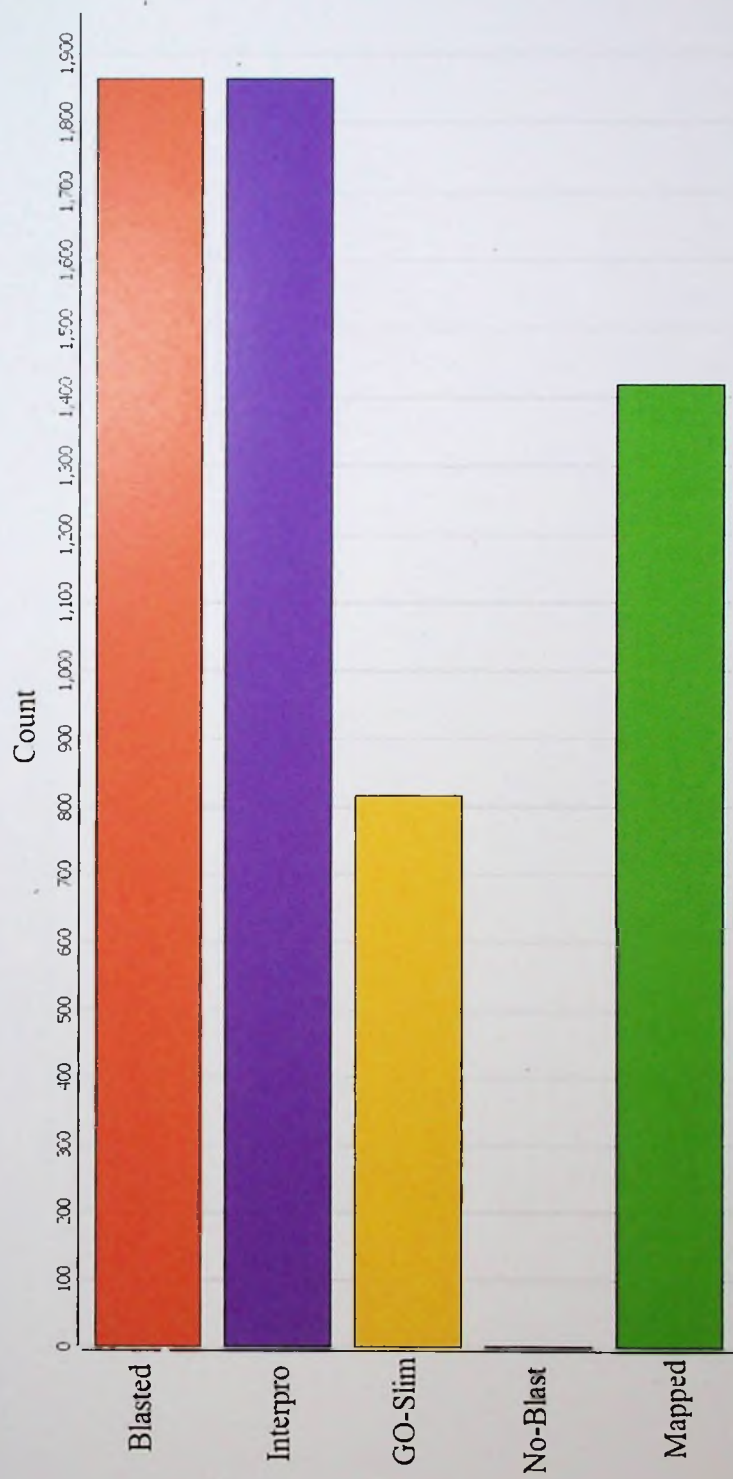


Figure 3.16: Tag distribution for the transcripts enriched in the cassava *B. tabaci* (SSA1-SG1) bacteriocyte. Number of sequences blasted against NCBI non-redundant database and InterproScan database.

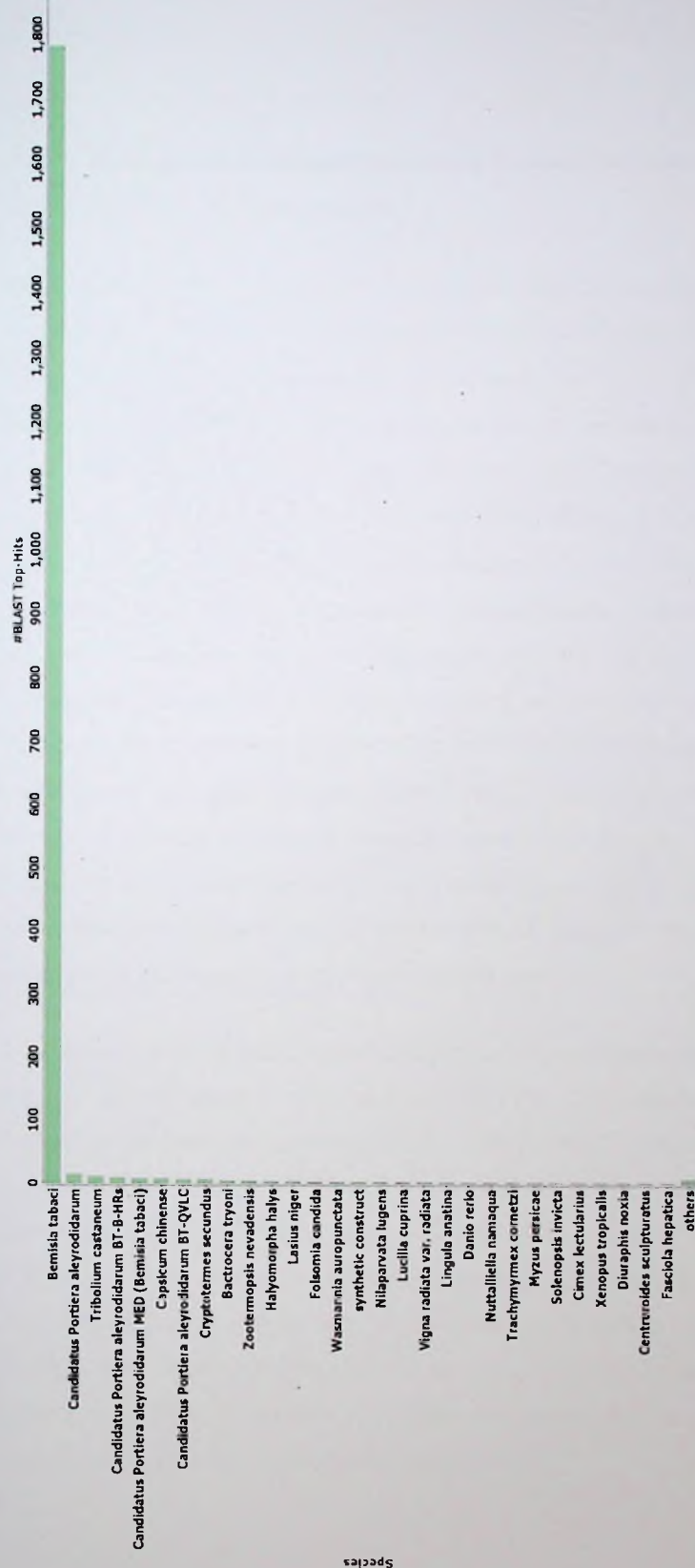


Figure 3.17: Top-hit species distribution for the differentially expressed transcripts in cassava *B. tabaci* (SSA1-SG1) bacteriocyte blasted against NCBI non-redundant database.

3.3.9 Gene ontology categorization for the differentially expressed transcripts in cassava *B. tabaci* (SSA1-SG1) bacteriocyte

Gene ontology classification for transcripts enriched in cassava *B. tabaci* (SSA1-SG1) bacteriocyte was based on three aspects: molecular level activities performed by gene products (molecular function), biological processes and the location relative to cellular structure (cellular component). Based on the cellular component, many transcripts were expressed in the intracellular component (204 transcripts), intracellular part (158 transcripts), intracellular organelle (125 transcripts) and membrane-bounded organelle (100 transcripts) (Figure 3.18). At the molecular level, most of the transcripts enriched in the *B. tabaci* bacteriocyte encoded for ion binding (203 transcripts), hydrolase activity (109 transcripts), oxidoreductase activity (94 transcripts) and transferase activity (68 transcripts). Several transcripts also showed molecular signatures for transmembrane transporter activities, catalytic activity-acting on protein and protein binding. These are important functions for amino acid biosynthesis and host-symbiont interaction in the bacteriocyte. Gene ontology for biological process showed that the largest percentage of transcripts are involved in metabolic processes (organic substance metabolic process, primary metabolic process, cellular metabolic process, nitrogen compound metabolic process). Other biological processes that relate to metabolism include regulation of cellular process, biosynthetic process, small molecule metabolism and regulation of metabolic process.

Biological processes were further categorized based on sequence distribution of the transcript that are enriched in cassava *B. tabaci* bacteriocyte. This categorization revealed that the top five biological processes in cassava *B. tabaci* (SSA1-SG1) bacteriocyte relate to protein metabolism (amino acid biosynthesis). These include cellular nitrogen compound metabolic process (26%), biosynthetic process (22%), transport (14%) and cellular protein modification process (8%) (Figure 3.19). In addition, direct gene ontology count for the transcripts enriched in cassava *B. tabaci* bacteriocyte also showed several transcripts related to protein metabolic processes which include: biosynthetic process (104 transcripts), cellular nitrogen compound metabolic process (94 transcripts), cellular protein modification process (42 transcripts) and cellular amino acid metabolic process (37 transcripts) (Figure 3.20).

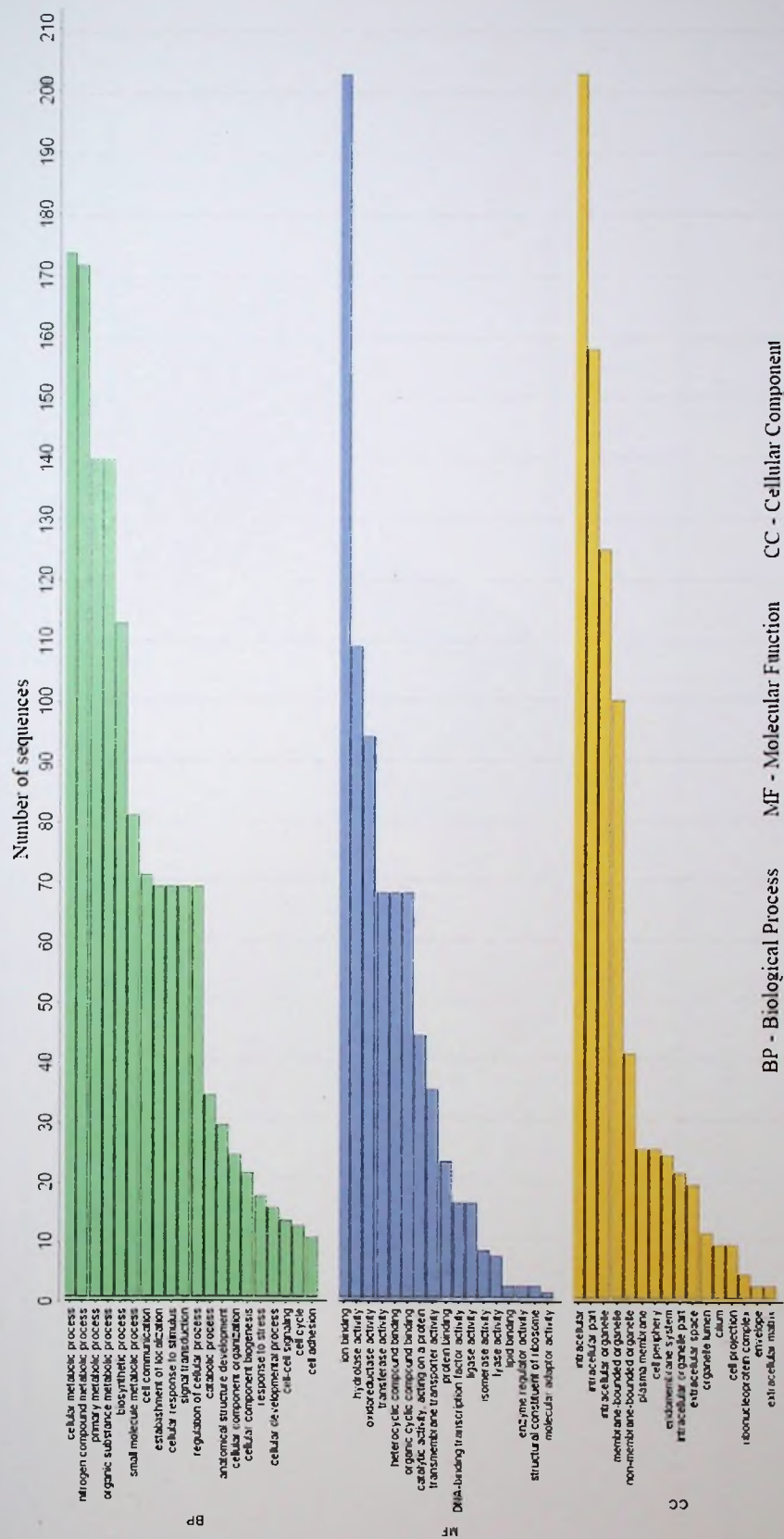


Figure 3.18: Gene ontology (GO) categorization, level 3 for the transcripts enriched in the cassava *B. tabaci* (SSA1-SG1) bacteriocyte.



Figure 3.19: Distribution of transcripts enriched in the bacteriocyte as assigned to different categories of biological processes in the cassava *B. tabaci* (SSA1-SG1) bacteriocyte.

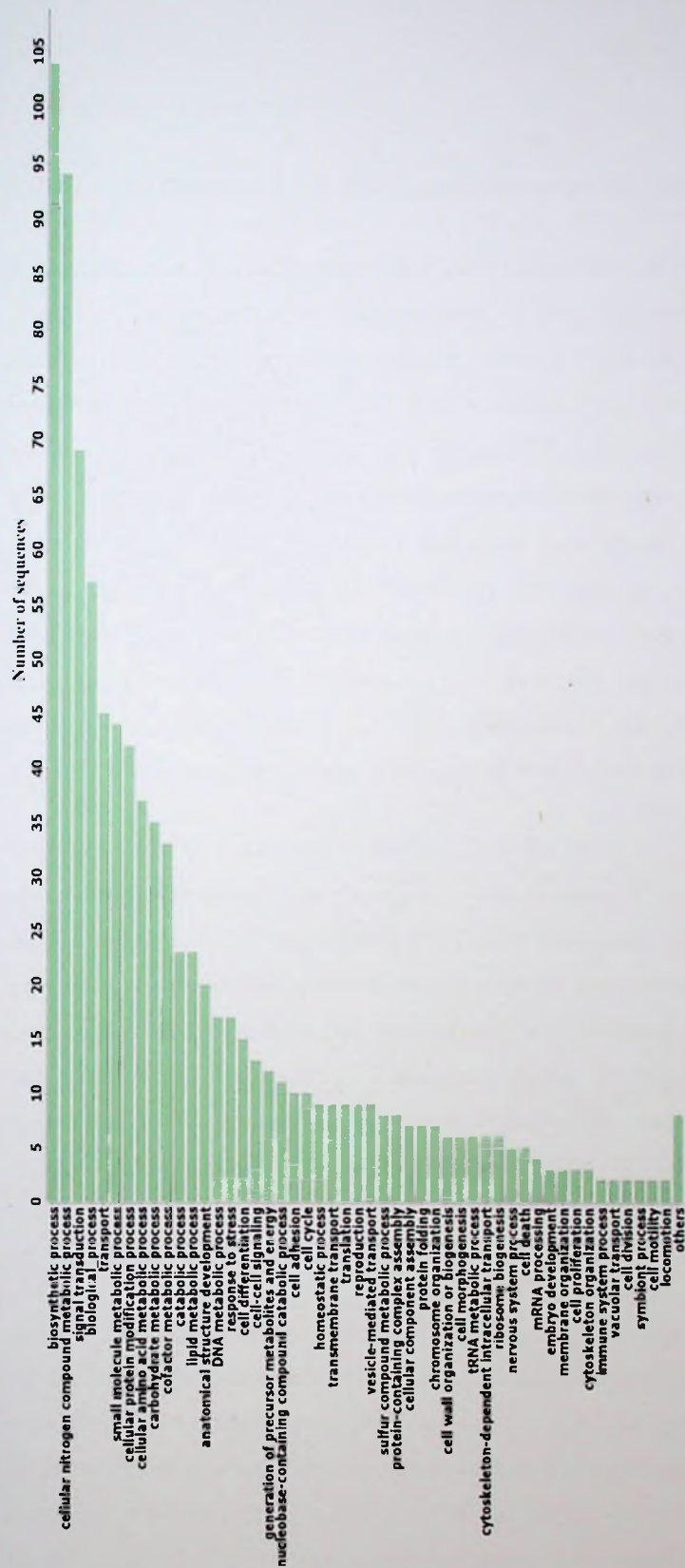


Figure 3.20: Direct gene ontology count for transcripts enriched in the cassava *B. tabaci* (SSA-SG1) bacteriocyte.

3.3.10 Identification of potential symbiosis genes targets in cassava *B. tabaci* (SSA1-SG1)

Transcripts enriched in *B. tabaci* SSA1-SG1 bacteriocyte were analyzed to identify proteins that facilitate essential amino acid biosynthesis, as potential symbiont gene targets. These included (i) aminotransferases, (ii) amino acid transporters, (iii) horizontally transferred genes and (iv) receptors of intermediate metabolites for essential amino acid biosynthesis across the symbiosome membrane, a special membrane that separates the endosymbiont and host in the bacteriocyte. A total of 20 genes encoding for amino acid transporters and aminotransferases were identified and selected. Amino acid transporters belonged to different families which included (i) Amino acid/polyamine transporter 1 (IPR002293), (ii) cationic amino acid transporter, C-terminal (IPR029485), (iii) proton-dependent oligopeptide transporter (IPR000109) and (iv) sodium: dicarboxylate symporter (IPR001991). Amino acid transferases belonged to (i) kynurenine--oxoglutarate transaminase 3, KAT_III (IPR034612), (ii) aminotransferase class 4 (IPR001544) and (iii) aminotransferase class V domain (IPR00192) (Table 3.7).

Genes (ENSSSA1UGT021004 (Ssa04473) and ENSSSA1UGT023671 (Ssa10714)) belonging to aminotransferase family had the highest gene expression among the selected genes in cassava *B. tabaci* bacteriocyte. These were identified as branched-chain-amino acid aminotransferase and kynurenine--oxoglutarate transaminase 3 with an expression of log (base 2) Counts Per Million (logCPM) of 8.89 and 6.79 and fold difference of 8.41 and 4.75-fold in bacteriocytes relative to the whole-body respectively. These two genes corresponded to Bta10673 and Bta05157 respectively, two aminotransferase orthologs in MEAM1 *B. tabaci*. Other aminotransferase included phosphoserine aminotransferase (ENSSSA1UGT011663 (Ssa14433)), which showed a relatively higher gene expression (log CPM of 7.30 and fold difference of 6.16) in the bacteriocyte compared to other selected genes.

Among the amino acid transporters, the highest number of potential gene targets (6 genes) among the selected belonged to amino acid transporter, transmembrane domain (IPR013057). Six targets belonging to amino acid/auxin permease family were also identified as amino acid permease, these amino acid transporters are more specialized in transporting cation amino acids (lysine, arginine, leucine, methionine and glutamine) from the cell to the extracellular matrix (Deves *et al.*, 1992; Dye *et al.*, 2004). These include, cationic amino acid transporter 3 (ENSSSA1UGT002031), high

affinity cationic AA transporter 1 (ENSSSA1UGT020942 and ENSSSA1UGT030948) and y⁺L amino acid transporter 2 (ENSSSA1UGT020386, ENSSSA1UGT011941 and ENSSSA1UGT019654). Of these, a y⁺L amino acid transporter 2, ENSSSA1UGT011941 (Ssa06643) had a relatively high expression with a logCPM of 7.39 and fold difference of 8.59-fold relative to whole-body. This gene corresponded to Bta11657, a y⁺L amino acid transporter 2 gene in MEAM1 *B. tabaci*.

Another set of amino acid transporter identified contained an amino acid transporter, transmembrane domain (Aa_trans, IPR013057), which is unique for the amino acid polyamine organocation superfamily. These amino acids included (i) transmembrane protein 104 homolog (ii) three, proton-coupled amino acid transporter 1, (iii) proton-coupled amino acid transporter 4 and (iv) sodium-coupled neutral AA transporter 9 (Table 3.7). Other amino acid transporters identified included excitatory amino acid transporter 1 and peptide transporter family 1 belonging to proton-dependent oligopeptide transporter family.

In summary, based on the level of gene expression in the cassava *B. tabaci* bacteriocyte, eight genes (ENSSSA1UGT011663, ENSSSA1UGT021004, ENSSSA1UGT017527, ENSSSA1UGT023671, ENSSSA1UGT011941, ENSSSA1UGT030948, ENSSSA1UGT002031 and ENSSSA1UGT011431) can be selected as potential targets whose disruption may cause significant effects on growth and development of cassava *B. tabaci* (SSA1-SG1).

Table 3.7: Potential gene targets selected from differentially expressed genes in cassava *B. tabaci* bacteriocytes

Gene ID	NCBI-BLAST		Domain		LogCPM	Fold differences	P-value	Gene**
	Hit	Name	Interpro					
ENSSSA1UGT021004	Branched-chain-amino-acid aminotransferase	Aminotrans_4	IPR001544	8.89	6.79	0	Bta10673	
ENSSSA1UGT023671	Kynurenine--oxoglutarate transaminase 3	KAT_III	IPR034612	8.41	4.75	0	Bta05157	
ENSSSA1UGT011663	Phosphoserine aminotransferase	Aminotran_5	IPR000192	7.30	6.16	3.7e-85	Bta04363	
ENSSSA1UGT011941	γ -L amino acid transporter 2	AA_Permcase_2	IPR002293	7.39	8.59	0	Bta11657	
ENSSSA1UGT030948	High affinity cationic AA transporter 1	AA_Permcase_2	IPR002293	6.57	3.19	1.01	Bta01499	
ENSSSA1UGT002031	Cationic amino acid transporter 3	AA_Permcase_C	IPR029485	6.36	4.07	1.0e-105	Bta11613	
ENSSSA1UGT019654	γ -L amino acid transporter 2	AA_Permcase_2	IPR002293	5.98	3.21	3.5e-258	Bta02775	
ENSSSA1UGT020386	γ -L amino acid transporter 2	AA_Permcase_2	IPR002293	5.96	5.24	0	Bta08771	
ENSSSA1UGT020942	High affinity cationic AA transporter 1	AA_Permcase_2	IPR002293	5.18	12.25	0	Bta13409	
ENSSSA1UGT011431	Transmembrane protein 104 homolog	Aa_trans	IPR013057	7.41	5.32	0	Bta03456	
ENSSSA1UGT000759	Proton-coupled amino acid transporter 1	Aa_trans	IPR013057	4.83	4.49	1.4e-202	Bta08775	
ENSSSA1UGT001543	Proton-coupled amino acid transporter 4	Aa_trans	IPR013057	4.74	3.25	5.4e-127	Bta08693	
ENSSSA1UGT024971	Proton-coupled amino acid transporter 1	Aa_trans	IPR013057	4.88	4.78	1.3e-222	Bta01726	
ENSSSA1UGT000813	Sodium-coupled neutral AA transporter 9	Aa_trans	IPR013057	2.66	2.47	7.8e-21	Bta11166	
ENSSSA1UGT025011	Proton-coupled amino acid transporter 1	Aa_trans	IPR013057	0.61	3.62	9.6e-48	Bta01718	
ENSSSA1UGT017527	vesicular glutamate transporter 3	Major facilitator	IPR020846	6.95	11.97	0	Bta06520	
ENSSSA1UGT008430	Neutral and basic amino acid transport protein rBAT	GH 13	IPR006589	6.26	2.92	2.2e-246	Bta10614	
ENSSSA1UGT021057	excitatory amino acid transporter 1	PTR2	IPR000109	5.02	7.09	4.3e-317	Bta11487	
ENSSSA1UGT023619	peptide transporter family 1	PTR2	IPR000109	3.40	8.99	0	Bta04879	
ENSSSA1UGT012586	Na-dependent excitatory AA transporter glt-6	SDF	IPR001991	2.01	4.02	5.4e-28	Bta09439	

**ortholog in *Helicoverpa tabaci* (ME:AM1)

3.3.11 Horizontally transferred genes of bacterial origin expressed in cassava (SSA1-SG1)

B. tabaci species

Previous studies by Luan *et.al.* (2015) showed that horizontally transferred genes (HTGs) are important for essential amino acid biosynthesis in *B. tabaci*. Candidate based search of HTGs previously identified in MEAM1 *B. tabaci* species (Chen *et al.*, 2016) was done to identify HTGs expressed in cassava (SSA1-SG1) *B. tabaci* bacteriocytes. A reciprocal blast revealed a total of 43 HTGs of bacterial origin in SSA1-SG1 bacteriocytes. These genes had a varying level of gene expression in the bacteriocytes relative to the whole-body ranging from log CPM of -0.42-11.16 and fold difference ranging from -4.73-13.89-fold.

HTGs contributing to lysine, arginine and proline biosynthesis in the bacteriocytes had relatively high gene expression compared to other genes identified. These included (i) 2OG-Fe (II) oxygenase (logCPM 11.16, 1.00-fold), (ii) 4-hydroxy-tetrahydrodipicolinate reductase (logCPM 7.88, 5.45-fold), (iii) allophanate hydrolase (logCPM 7.04, 1.13-fold), (iv) diaminopimelate decarboxylase (LysA) (logCPM 6.89, 13.89 fold), (v) arginosuccinate lyase (ArgH) (logCPM 6.68, 1.79 fold) and (vi) arginosuccinate synthase (ArgG) (logCPM 5.95, 2.10 fold). These genes corresponded to Bta20012, Bta20020, Bta13949, Bta03593, Bta00063 and Bta00062, all HTGs in MEAM1 *B. tabaci* species (Table 3.8). Other genes involved in amino acid biosynthesis in the bacteriocyte include (i) an Acetyltransferase, (ii) amidinotransferase and (iii) histidine ammonia-lyase (Table 3.7).

HTGs that are important in vitamin B (biotin) biosynthesis pathway also exhibited a relatively high gene expression in *B. tabaci* bacteriocytes. These included (i) adenosylmethionine--8-amino-7-oxononanoate aminotransferase (*bioA*) (logCPM 6.69, 7.11 fold), (ii) ATP-dependent dethiobiotin synthetase (*bioD*) (logCPM 6.69, 7.11-fold), (iii) adenosylmethionine--8-amino-7-oxononanoate aminotransferase (*bioA*) (logCPM 5.16, 1.50 fold), (iv) biotin synthase (*bioB*) (logCPM 5.21, 1.41-fold) and (v) ATP-dependent dethiobiotin synthetase (*bioD*) (logCPM 5.16, -1.50-fold), corresponding to Bta01937, Bta1938, Bta00841, Bta09725 and Bta00840, HTGs in MEAM1. SSA1-SG1 contained two copies of *bioA* and *bioD* like what was reported in MEAM1 (Chen *et al.*, 2016). Gene 000024F_arrow:4385697-4418725 (*bioA*) has a lower expression (1.50 compared to 7.11-fold of the second *bioA*) in the bacteriocyte. This gene corresponded to

Bta00841, a truncated (Pseudogene) in MEAM1. For *bioD*, Bta01938 was reported as a pseudogene in MEAM1, however the corresponding gene in SSA1-SG1, ENSSSA1UGT012203 (Ssa01088) had a high expression (logCPM 6.69, 7.11-fold) in the bacteriocyte. Other HTGs with enriched gene expression in bacteriocytes and contributing to various pathways were cyclopropane-fatty-acyl-phospholipid synthase (logCPM 9.57, 11.07-fold), crossover junction endodeoxyribonuclease *rusA* (logCPM 7.89, 1.36-fold), BH1803 protein (logCPM 7.75, 1.54-fold), leucine rich repeat containing proteins (logCPM 7.75, 1.54 fold), ribosome recycling factor (logCPM 6.48, 10.77-fold) and L-galactonate dehydratase (logCPM 5.49, 1.9-fold).

A search for possible sources of all detected HTGs revealed several bacterial genera which included: *Rickettsia*, *Planctomycetes*, *Klebsiella*, *Pantoea*, *Curtobacterium*, *Pseudomonas*, *Rahnella*, *Wolbachia*, *Cardinium*, *Photorhabdus*, *Bacillus*, *Sodalis*, *Cronobacter*, *Rhizobium*, *Methylobacterium*, *Paraburkholderia*, *Burkholderia*, *Pectobacterium* and *Mucilaginibacter*.

In addition, a total of 38 HTGs of fungal origin were also identified in the SSA1-SG1 *B. tabaci* bacteriocyte sequences. These genes had lower expression in the bacteriocytes compared to HTGs of bacterial origin identified in this study. Of the 38 genes identified, only four genes had enriched expression within the SSA1-SG1 bacteriocytes. These encoded for gamma-glutamyltranspeptidase (logCPM 3.02, 3.80-fold), aromatic peroxygenase (logCPM 7.17, 7.42-fold), major royal jelly-related protein (logCPM 4.73, 2.09-fold) and squalene synthase (logCPM 6.43, 2.20-fold) (Table 3.9). Function analysis revealed that gene, ENSSSA1UGT000931 (Ssa05748) encoding for gamma-glutamyltranspeptidase could be a potential target. This is because gamma-glutamyltranspeptidase has been reported to be key in the transfer of amino acids across cellular membranes (Taniguchi and Ikeda, 1998)

Table 3.8: Horizontally transferred genes of bacterial origin expressed in cassava *B. tabaci* (SSA1-SG1) bacteriocytes

Gene ID	Description	Pathway	Gene expression		Possible origin: Bacteria	Gene***
			LogCPM	Fold difference		
ENSSSA1UGT010158	2OG-Fe (II) oxygenase	Arg. Proline	11.16	1.00	<i>Pantoea</i>	Bta20012
ENSSSA1UGT008257	4-hydroxy-tetrahydronicotinate reductase <i>dapB</i>	L-lysine	7.88	5.45	<i>Rickettsia</i>	Bta20020
ENSSSA1UGT007134	Diaminopimelate decarboxylase (<i>Lys4</i>)	L-lysine	6.89	13.89	<i>Planctomyces</i>	Bta03593
ENSSSA1UGT012203	ATP-dependent dehydrobiotin synthetase (<i>BioD</i>)	Biotin	6.69	7.11	<i>Wolbachia</i>	Bta01938
ENSSSA1UGT003467	Arginosuccinate lyase (<i>ArgH</i>)	L-arginine	6.68	1.79	<i>Erwinia</i>	Bta00063
ENSSSA1UGT016543	Ribosome recycling factor		6.48	10.77	Not well defined	Bta15019
ENSSSA1UGT003597	Arginosuccinate synthase (<i>ArgG</i>)	L-arginine	5.95	2.10	<i>Pantoea Erwinia</i>	Bta00062
ENSSSA1UGT003262	Phenazine biosynthesis-like domain-containing protein		5.45	-1.51	<i>Pseudomonas</i>	Bta02987
ENSSSA1UGT031701	Biotin synthase (<i>BioB</i>)	Biotin	5.21	1.41	Not well defined	Bta09725
ENSSSA1UGT012203	ATP-dependent dehydrobiotin synthetase (<i>BioD</i>)	Biotin	5.16	-1.50	<i>Wolbachia</i>	Bta00840
ENSSSA1UGT025651	Chorismate mutase (<i>CM</i>)	shikimate	5.01	2.57	<i>Rahnella</i>	Bta15103
ENSSSA1UGT004633	Leucine-rich repeat containing proteins		4.71	-1.87	<i>Rickettsia</i>	Bta13776
ENSSSA1UGT009294	SCP-like extracellular protein		4.27	1.32	<i>Bacillus</i>	Bta15713
ENSSSA1UGT017373	ATP-dependent DNA helicase Q-like 3		4.04	1.24	<i>Photorhabdus</i>	Bta00002
ENSSSA1UGT006636	Squalene-hopene cyclase		3.77	-4.73	<i>Burkholderia</i>	Bta02625
ENSSSA1UGT028316	Squalene-hopene cyclase		3.72	1.05	<i>Paraburkholderia</i>	Bta07125
ENSSSA1UGT008515	Ribosomal RNA subunit methyltransferase A		3.65	8.17	<i>Methylobacterium</i>	Bta03791
ENSSSA1UGT002124	3-methyl-2-oxobutanoate hydroxymethyltransferase	Pantothenate	3.57	30.50	<i>Pseudomonas</i>	Bta05339

Table 3.8 continued

Gene ID	Description	Pathway	Gene expression		Possible origin: Bacteria	
			LogCPM	Fold difference	Genus	Gene***
ENSSSA1UGT016889	Tryptophan-tRNA ligase		3.32	-1.31	<i>Pectobacterium</i>	Bta06820
ENSSSA1UGT001349	Ribonuclease H		3.04	6.12	<i>Wolbachiae</i>	Bta09186
ENSSSA1UGT005350	Urea amidolyase		2.86	-2.15	<i>Pseudomonas</i>	Bta04508
ENSSSA1UGT018122	Squalene-hopene cyclase		0.22	2.96	<i>Paraburkholderia</i>	Bta07024
ENSSSA1UGT013675	Diaminopimelate epimerase (<i>DapH</i>)	L-lysine	0.13	1.57	<i>Klebsiella</i>	Bta06657
Ssa09979	Cyclopropane-fatty-acyl-phospholipid synthase		9.57	11.07	<i>Cronobacter</i>	Bta06818
Ssa10244	Crossover junction endodeoxyribonuclease <i>rusA</i>		7.89	1.36	<i>Sodalis</i>	Bta03871
Ssa12239	Leucine-rich repeat containing proteins		7.06	1.15	<i>Rickettsia</i>	Bta13226
Ssa03692	Allophanate hydrolase	L-arginine	7.04	1.13	<i>Pantoea</i>	Bta13949
Ssa01553	Adenosylmethionine--8-amino-7-oxononanoate aminotransferase (<i>Bio4</i>)	Biotin	6.69	7.11	<i>Cardinium</i>	Bta01937
Ssa09927	Methyltransferase		6.31	1.11	<i>Pseudomonas</i>	Bta20014
Ssa08027	4,5-DOPA dioxygenase extradiol	Betalain	6.27	-1.42	<i>Pantoea</i>	Bta02812
Ssa03128	L-galactonate dehydratase		5.49	1.39	<i>Pseudomonas</i>	Bta03200
Ssa12718	Cyclopropane-fatty-acyl-phospholipid synthase		5.36	1.08	<i>Cronobacter</i>	Bta11912
Ssa01553	adenosylmethionine--8-amino-7-oxononanoate aminotransferase (<i>Bio4</i>)	Biotin	5.16	1.50	<i>Wolbachia</i>	Bta00841
Ssa00937	Acetyltransferase	L-lysine	5.14	0.86	<i>Klebsiella</i>	Bta04431
Ssa08712	Amidohydrolase		4.92	-1.14	<i>Mucilaginibacter</i>	Bta14802

Table 3.8 continued

Gene ID	Description	Pathway	Gene expression		Possible origin: Bacteria	
			LogCPM	Fold difference	Genus	Gene***
Ssa02297	SCP-like extracellular		4.27	1.32	<i>Bacillus</i>	Bta15712
Ssa10201	tRNA-splicing ligase RtcB		4.17	6.61	<i>Cardinium</i>	Bta14500
Ssa15043	Histidine ammonia-lyase	L-histidine	3.48	-1.04	<i>Pantoea</i>	Bta08776
Ssa00739	Amidinotransferase	Thr,Ser,Arg	2.90	-3.83	<i>Curtobacterium</i>	Bta04987
Ssa00740	Amidinotransferase	Thr,Ser,Arg	2.90	-3.83	<i>Curtobacterium</i>	Bta04988
Ssa10534	Leucine-rich repeat containing proteins		2.82	-1.46	Not well defined	Bta11397
Ssa00993	Cyclopropane-fatty-acyl-phospholipid synthase		0.47	1.37	Not well defined	Bta04469
Ssa12577	Oxidoreductase, 2OG-Fe (II) oxygenase		-0.42	1.78	<i>Pantoea</i>	Bta15191

Gene *** - Ortholog in ME/AMI

Table 3.9: Horizontally transferred genes of fungal origin expressed in cassava *B. tabaci* (SSA1-SG1) bacteriocytes

Gene ID	Description	Gene expression		Gene**
		LogCPM	Fold change	
ENSSSA1UGT005001	Major royal jelly-related protein	8.04	-1.37	Bta10729
ENSSSA1UGT003831	NFX1-type zinc finger-containing protein 1	7.53	-1.08	Bta08945
ENSSSA1UGT023534	cryptochrome	6.83	1.37	Bta20021
ENSSSA1UGT009368	Squalene synthase	6.43	2.20	Bta11043
ENSSSA1UGT002200	Serine/threonine dehydratase serine racemase	5.40	-1.02	Bta13117
ENSSSA1UGT027113	Adenosine deaminase	5.37	1.49	Bta03315
ENSSSA1UGT002367	MYND finger family protein	5.03	-2.43	Bta00932
ENSSSA1UGT016597	Major royal jelly-related protein	4.73	2.09	Bta14896
ENSSSA1UGT001581	Major royal jelly-related protein	4.54	1.63	Bta07455
ENSSSA1UGT000931	Gamma-glutamyltranspeptidase	3.02	3.80	Bta03557
ENSSSA1UGT003831	NFX1-type zinc finger-containing protein 1	2.73	-1.38	Bta08944
ENSSSA1UGT011235	Gamma-glutamyltranspeptidase	0.19	4.92	Bta05545
ENSSSA1UGT009478	Glucosylceramidase, putative	0.16	1.19	Bta02658
ENSSSA1UGT007947	Major royal jelly-related protein	0.16	-1.92	Bta10760
ENSSSA1UGT007941	Major royal jelly-related protein	-0.89	3.29	Bta10761
Ssa04524	2OG-Fe (II) oxygenase superfamily protein	7.18	-1.04	Bta14373
Ssa04526	2OG-Fe (II) oxygenase superfamily protein	7.17	-1.36	Bta14374

Table 3.9: Continued

Gene ID	Description	Gene expression		
		LogCPM	Fold change	Gene***
Ssa12366	Aromatic peroxxygenase	7.17	7.42	Bta05753
Ssa12366	Aromatic peroxxygenase	5.98	1.26	Bta15602
Ssa10936	Aromatic peroxxygenase	4.34	-33.68	Bta00041
Ssa04914	MYND finger family protein	3.77	-7.98	Bta05475
Ssa15364	Aromatic peroxxygenase	3.67	9.23	Bta11373
Ssa09482	Major royal jelly-related protein	3.53	-42.90	Bta05467
Ssa15384	MYND finger family protein	3.05	-7.06	Bta07433
Ssa09478	Major royal jelly protein	2.49	-35.17	Bta05463
Ssa02002	Zinc finger MYND-type protein	2.26	-3.27	Bta10703
Ssa14298	Uracil phosphoribosyltransferase	2.09	-1.51	Bta14185
Ssa14166	Dextranase	2.06	-1.51	Bta02250
Ssa06839	Aminotransferase family protein (LolT), putative	1.73	1.12	Bta02178
Ssa03924	Galactose oxidase	1.61	-18.38	Bta01072
Ssa09479	Major royal jelly protein	1.37	-622.58	Bta05465
Ssa06267	Aromatic peroxxygenase	1.29	-14.49	Bta04808
Ssa10522	Late sexual development protein	0.44	3.03	Bta12908

Table 3.9: Continued

Gene ID	Description	Gene expression		Gene***
		LogCPM	Fold change	
Ssa10745	Major royal jelly protein	0.33	-288.36	Bta10736
Ssa07646	Aromatic peroxxygenase	0.80	124.02	Bta07721
Ssa03925	Aromatic peroxxygenase	0.05	-5.27	Bta01073
Ssa02113	Galactose oxidase	-0.56	4.77	Bta05093
Ssa02112	Galactose oxidase	-0.56	4.77	Bta05092
Gene *** - Ortholog in MEAM1				

3.3.12 Validation of expression of selected genes using real-time quantitative PCR.

A total of eight genes were selected as critical gene targets with enriched expression in either the whitefly gut or the bacteriocyte compared to the whole-body. These were, (i) ENSSSA1UGT022145: Aquaporin 1 (*AQP1*) for water recycling in the gut, (ii) ENSSSA1UGT002057: Alpha-glucosidase (*SUC1*) for sucrose hydrolysis, (iii) ENSSSA1UGT002066: alpha-glucosidase (*SUC2*) for sucrose hydrolysis, (iv) ENSSSA1UGT007134: diaminopimelate decarboxylase (*LysA*), essential in lysine biosynthesis, (v) ENSSSA1UGT003467: arginosuccinate lyase (*ArgH*), essential in arginine biosynthesis, (vi) ENSSSA1UGT021004: branched-chain-amino-acid aminotransferase (*BCAT*), critical in isoleucine, leucine and valine biosynthesis, (vii) ENSSSA1UGT008852: aspartate aminotransferase (*AAT*), essential in phenylalanine biosynthesis and (viii) ENSSSA1UGT000364: 4-hydroxy-tetrahydronicotinate reductase (*dapB*), essential in lysine biosynthesis. The selected genes with enriched expression in the bacteriocyte were selected from both differentially expressed gene set and horizontally transferred genes (Chapter 3) and their essentiality confirmed through constraint-based metabolic modeling of the genome-scale model of cassava whitefly (SSA1-SG1) (Chapter 4).

This analysis revealed that, of the eight selected genes, seven had enriched expression either in the cassava whitefly gut (*AQP1*, *SUC1* and *SUC2*) or bacteriocyte (*ArgH*, *BCAT*, *dapB* and *LysA*) compared to the whole-body (Figure 3.21). The expression of *AAT* was more enriched in the whole-body compared to the bacteriocyte. Results from RNAseq analysis and RT-qPCR were similar with fold differences in the gut ranging from 22.30 to 26.54-folds for RT-qPCR compared to 21.71 to 32.67-folds from RNAseq analysis. In the bacteriocyte, gene expression for the selected gene ranged from 0.57 to 10.07-folds for RT-qPCR compared to 1.79 to 13.89-folds for RNAseq analysis.

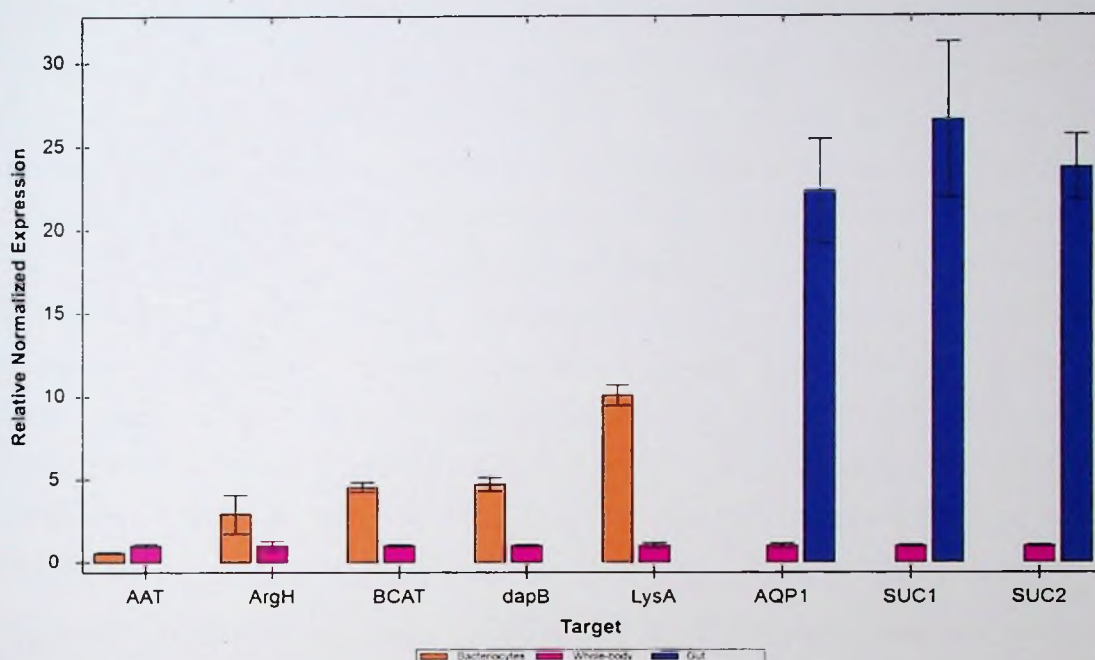


Figure 3.21: Quantitative real-time PCR gene expression of selected genes in the cassava whitefly (SSA1-SG1) gut and bacteriocytes relative to whole-body. Gene ENSSSA1UGT022145 (*AQP1*), ENSSSA1UGT002057 (*SUC1*) and ENSSSA1UGT002066 (*SUC2*) were selected as critical osmoregulation genes with enriched expression in the whitefly gut while ENSSSA1UGT008852 (*AAT*), ENSSSA1UGT003467 (*ArgH*), ENSSSA1UGT021004 (*BCAT*), ENSSSA1UGT000364 (*dapB*) and ENSSSA1UGT007134 (*LysA*) were selected as essential symbiosis genes that mediate terminal reaction in the phenylalanine, arginine, leucine, isoleucine, valine and lysine biosynthesis pathways. The relative gene expression of the selected genes was normalized relative to the relative expression of two reference genes (*ribosomal protein L13a* and β -*tubulin*) using normalized expression $\Delta\Delta C_q$ method (Livak & Schmittgen, 2001).

3.4 DISCUSSION

The superabundant populations of *B. tabaci* in cassava growing areas in eastern, central and southern African (Legg *et al.*, 2014a; Legg *et al.*, 2014b) and the increasing effects of viruses (CBSVs and begomoviruses) they transmit (Maruthi *et al.*, 2005) require more effective management options. Precision methods such as RNAi that target specific genes are proposed to offer a useful approach to effectively manage the *B. tabaci* population (Grover *et al.*, 2019). This work presented here was initiated to identify essential gene targets for managing the superabundant *B. tabaci* population in cassava growing areas. This study was built on a premise that co-downregulation of essential genes with different molecular but linked physiological functions would act synergistically to suppress growth of *B. tabaci*, thereby providing a tool to manage their populations in agricultural production systems. Both alpha-glucosidase family 13 and aquaporin, genes with enriched expression in the *B. tabaci* (SSA1-SG1) gut, have similar physiological role of maintaining fluid homeostasis within the insect body cavity. These genes were selected as osmoregulation genes to target for the management of whitefly population.

Aquaporins, a major intrinsic protein in insects serve as channels in the transfer of water across membranes (Campbell *et al.*, 2008). They enable insects and other organisms to adapt to extreme conditions and to utilize special diets. The *B. tabaci* (SSA1-SG1) genome, for example, has a total of seven aquaporin genes (Figure 3.9), however, only three have enriched expression in the *B. tabaci* gut. Structural analysis of the highly expressed aquaporin gene ENSSSA1UGT022145 (Ssa03484) revealed that protein translated from this gene had two NPA motifs, conserved ar/R constriction residue and a putative conserved mercury-sensitive cysteine residue located two residues from the first NPA. These characteristics revealed that ENSSSA1UGT022145 is a water-selective channel, a classical grade of aquaporin (DRIP) that moderate movement of water within the gut. *AQP1* has been reported to be highly expressed in the filter chamber of the *B. tabaci* (Mathew *et al.*, 2011) where water is rapidly transferred between the distal and proximal gut regions (Cicero *et al.*, 1995). Phylogenetic analysis and reciprocal blast analysis of the identified aquaporin genes in SSA1-SG1 and experimentally validated aquaporin genes in MEAM1 (Mathew *et al.*, 2011; Jing *et al.*, 2016) showed that both ENSSSA1UGT022145 and Bta01973 (*AQP1* in MEAM1) had high sequence similarity, sharing 100% amino acid identity. This implies that the *AQP1*, a water-specific aquaporin in both MEAM1 and SSA1-SG1 is highly conserved and under purifying selection. The high

similarity of water-specific *AQP1* in MEAM1 and SSA1-SG1 means that RNAi strategies targeting *AQP1* gene will be effective in managing the population of both MEAM1 and SSA1-SG1, two economically important *B. tabaci* species.

Whiteflies also reduce osmotic pressure of the ingested phloem-sap in the gut by hydrolyzing sucrose to glucose and fructose and transforming excess glucose to oligosaccharides thereby lowering the molarity of ingested phloem sap sugars and osmotic pressure in the gut (Salvucci *et al.*, 1997; Salvucci, 2000). In this study, the search for enzymes that mediate sugar conversion in cassava whiteflies yielded many α -glucosidases (EC 3.2.1.20) belonging to glycosidase hydrolase family 13. Whiteflies possess many α -glucosidases (35-48 genes) (Supplementary Table 6.3) though a few possess both signal peptide and catalytic site residues. This may imply that some α -glucosidase genes are not functional or have become pseudogenes. In this study, a total of 14 α -glucosidase with enriched expression in the SSA1-SG1 gut had both signal peptide and catalytic site residue. Many of these enzymes show sequence similarity and a conserved structure core with variation in amino acid residues in the conserved regions.

Differences in the amino acid residues especially in the conserved region II influences substrate preference of α -glucosidase in the gut. Conserved region II is the region on the α -glucosidase amino acid sequence that includes key residues of the catalytic site of α -glucosidase. Ngiwsara *et al.* (2012) showed that the arrangement of the first five amino acid residue in conserved region II influenced preference of enzyme to either sucrose or maltose and the nature of glucosidic linkage formed (regioselectivity) during transglucosylation. Mutation analysis revealed that proline (P) in the fourth position of the conserved region II followed by tyrosine (Y) influenced the preference of sucrose to maltose. Since the predominant sugar in phloem sap is sucrose (Hijaz & Killiny, 2014), both phylogenetic analysis of α -glucosidase and structural analysis of conserved region II was done to identify the α -glucosidase enzyme that hydrolyzes sucrose in SSA1-SG1 *B. tabaci*. Three enzymes clustered with the experimentally validated α -glucosidase that hydrolyzes sucrose (Price *et al.*, 2007). Of these, only one (ENSSSA1UGT002057) had proline and tyrosine in the 4th and 5th position of the conserved region II (DAVPYLF) respectively (Figure 3.10). The role of tyrosine (Y) at position 5 of the conserved region II relates to sucrose preference while proline at position 4 relates to the formation of α -1,4-glucosidic linkage during transglucosylation process of the enzyme (Ngiwsara *et al.*, 2012). These characteristics make ENSSSA1UGT002057 the best α -glucosidase gene target for managing SSA1-SG1 *B. tabaci*. A reciprocal blast of the gene

against the *B. tabaci* genome (whiteflygenomics.org) database¹⁰ revealed a 97% amino acid sequence similarity with α -glucosidase gene Bta05386. The high sequence similarity of this gene in both MEAM1 and SSA1-SG1 makes it a good gene target in the management of the two *B. tabaci* species.

Recently, studies have been published on the application of RNAi to downregulate osmoregulation genes in MEAM1 (Raza *et al.*, 2016; Luo *et al.*, 2017), though these two studies show contrasting results on the effect of RNAi on whitefly performance. The discrepancy in these results may be attributed to the different plants used for dsRNA delivery and the sucrase gene sequence used in the development of dsRNA constructs. Luo *et al.* (2017) used sequence for α -glucosidase gene, Bta03818 while Raza *et al.* (2016) used sequence for α -glucosidase gene, Bta07452, which lacked a signal peptide (Table 3.5). The absence of a signal peptide in Bta07452 means that the enzyme is non-functional, or the gene is a pseudogene.

RNAseq analysis to study metabolic interaction between cassava *B. tabaci* (SSA1-SG1) and *Portiera* revealed three sets of genes in this study with highly enriched expression in the bacteriocytes. These were (i) amino acid transferases, (ii) amino acid transporters and (iii) horizontally transferred genes. The gene with the highest expression in the SSA1-SG1 bacteriocyte (ENSSSA1UGT021004) encoded for branched-chain amino acid aminotransferase, that mediates the terminal reaction in leucine, isoleucine and valine biosynthesis pathways. These three amino acids interact with transmembrane domains of membranous proteins and also play a crucial role in determining the structure of globular proteins (Brosnan & Brosnan, 2006). Gene essentiality analysis of the genome-scale metabolic model of cassava *B. tabaci* (SSA1-SG1) and *Portiera* (Chapter 4) revealed that this gene is indispensable for the survival of both the *B. tabaci* and its primary symbiont. These three characteristics (i) high expression in the bacteriocyte, (ii) mediating terminal reaction of the three essential amino acids and (iii) being critically essential for the growth of both *B. tabaci* and its primary endosymbiont, makes this gene, one of the best symbiosis gene targets for the control of cassava *B. tabaci*.

HTGs on the other hand offers a selection advantage to recipient organism (Hotopp *et al.*, 2007; Zakharov, 2016). HTGs have been reported in phloem sap feeders, providing recipients with new functionalities (Nikoh & Nakabachi, 2009; Nikoh *et al.*, 2010; Husnik *et al.*, 2013). In cassava *B. tabaci* (SSA1-SG1), a total of 48 and 45 genes of bacterial and fungal origin

¹⁰ <http://www.B.tabaci/genomics.org/cgi-bin/bta/blast.cgi>

respectively were identified in this study. The expression values of these genes, especially of bacterial origin, show that the transferred genes are active and functional, mediating reactions in lysine, arginine, threonine, serine, proline, histidine and biotin synthesis. These innovative gene transfers (Husnik & McCutcheon, 2018) provide cassava *B. tabaci* as a host with new functionality in the synthesis of essential amino acids. The gain of new gene functions in the shared metabolic network leads to gene redundancy of some genes which in turn has led to the loss of essential genes and genome reduction of the primary endosymbiont (Andersson & Kurland, 1998). One example is gene *dapB* that encode for 4-hydroxy-tetrahydronicotinate reductase (EC 1.17.1.8). This gene is missing in the genome of SSA1-SG1-*Portiera* but present in the genome of MEAM1-*Portiera* and MED-*Portiera*, though as pseudogenes (Luan *et al.*, 2015) (Chapter 4). Though missing in *Portiera*, the loss of this gene has been compensated for by acquiring a horizontally transferred *dapB* into cassava *B. tabaci* genome, exhibiting a high expression in the bacteriocyte. The acquisition of such genes has enabled *B. tabaci* as the host to alter metabolite flux in the shared metabolic pathways which allow the host to control the synthesis of essential amino acid.

Another way *B. tabaci* control synthesis and provisioning of the essential amino acid is by retaining HTGs that mediate terminal reaction in various amino acid biosynthesis pathways through purifying selection. For example, *lysA* that encodes for diaminopimelate decarboxylase that mediates terminal reaction in lysine synthesis and *argH* that encodes for arginosuccinate lyase that mediates the terminal reaction in arginine synthesis. Gene deletion analysis revealed that *dapB*, *dapF*, *lysA* and *argH* are indispensable genes and therefore critical for the growth and survival of cassava *B. tabaci* (SSA1-SG1) (Chapter 4). Other HTGs that mediate terminal reaction was *bioB* which mediates the final reaction in the biotin synthesis. Biotin is essential in the catabolism of leucine, valine and isoleucine (Massey *et al.*, 1976). In summary, results in this study revealed three important insights, first, SSA1-SG1 and MEAM1 have similar HTGs mediating similar essential amino acid biosynthesis reactions in the bacteriocytes (Luan *et al.*, 2015). Secondly, some HTGs are indispensable and mediate critical (terminal) reactions in some essential amino acid biosynthesis pathway, which makes them good gene targets for the management of *B. tabaci* population and lastly, the absence of *dapB* in SSA1-SG1-*Portiera* genome imply that SSA1-SG1 and its *Portiera* are co-evolving faster than MEAM1 and MED (Chapter 4).

The transfer of metabolites between the *B. tabaci* and its primary endosymbiont among phloem-sap feeding insects is facilitated by amino acid transporters (Duncan *et al.*, 2014,

Douglas, 2014). In this study, a total of 16 amino-acid transporters of intrinsic (host) origin and one horizontally transferred gene of fungal origin (GGT1 encoding for gamma-glutamyltransferase) (Sweiry *et al.*, 1995; Taniguchi and Ikeda, 1998) were identified. Fewer amino acid transporters were identified in this study compared to those reported in MEAM1 (Xia *et al.*, 2017). These amino acid transporters have enriched expression in the whitefly (SSA1-SG1) bacteriocyte. Amino-acid transporters in the bacteriocyte catalyze uptake as well as exchange of amino acid across the symbiosome membrane which sustain the shared metabolic interaction between the whitefly and *Portiera*. It is worth noting that there is a widespread redundancy in the amino acid specificity across the different amino acid transporters i.e, the same amino acid can be a substrate of different transporters in different categories (Scalise *et al.*, 2014). It is therefore important that more studies are conducted to characterize the amino-acid transporters identified in this study to identify the best gene target for managing *B. tabaci* population in cassava growing areas.

In conclusion, essential gene targets encoding for proteins with different molecular but linked physiological functions have been identified. Several factors have been considered, which include: (i) substrate specificity for sugar processing enzymes (ii) water specificity for water recycling mechanism in the gut (iii) reaction mediated in the essential amino acid biosynthesis pathways in the bacteriocyte and (iv) indispensability of the genes in the amino acid biosynthesis. Based on these factors, seven genes have been selected as critical genes in cassava *B. tabaci* (SSA1-SG1). These include: (i) ENSSSA1UGT022145: aquaporin (*AQP1*), (ii) ENSSSA1UGT002057: alpha glucosidase (*SUC1*) (iii) ENSSSA1UGT002066: alpha glucosidase (*SUC2*) (iv) *dapB*, *dapF* and *lysA* for lysine biosynthesis, (iv) branched chain amino-acid transferase for isoleucine, leucine, phenylalanine and valine biosynthesis and *argH* for arginine biosynthesis. The expression of these genes has been validated using RT-qPCR (Figure 3.20). It is envisaged that stacking these essential genes with similar physiological functions in combinatorial RNAi will lead to the management of super abundant whiteflies in cassava growing areas in Sub-Saharan Africa.

IDENTIFICATION OF ESSENTIAL GENES TARGETS IN CASSAVA *B. TABACI*, THROUGH GENOME-SCALE METABOLIC MODELING

4.1 INTRODUCTION

Whiteflies and other phloem-sap feeding insects depend on endosymbionts for essential amino acid and B vitamin biosynthesis (Section 2.5). This interaction demonstrates the interdependence among organisms for growth and survival (Douglas, 2016). The exploitation of this interdependency in the control of *B. tabaci* populations requires a comprehensive knowledge of both *B. tabaci* and endosymbiont genetics and physiology. The increasing availability of genomic resources of different *B. tabaci* species and their respective endosymbionts offers opportunities to study this interaction at the systems level. Genome-scale stoichiometric models coupled with constraints-based modeling techniques like flux balance analysis and flux variability analysis (Varma & Palsson, 1994) offers a complete overview of the gene-protein stoichiometric network and flow of flux, that can facilitate studying the interaction between *B. tabaci* and symbionts (section 2.5.2). Furthermore, other techniques, for example minimization of metabolic adjustment (MOMA) (Segre *et al.*, 2002) and regulatory on/off minimization (ROOM) (Shlomi *et al.*, 2005) also offer ways to predict the outcome of genetic perturbation thereby allowing identification of essential genes. Genome-scale metabolic models have been used to identify essential gene targets for pharmaceutical applications (Hu *et al.*, 2007; Jamshidi & Palsson, 2007) and also for efficient production of amino acids (Lee *et al.*, 2007; Park *et al.*, 2007; Pharkya *et al.*, 2003) in different organisms.

The interactions of the primary endosymbionts (*Portiera aleyrodidarum*) and secondary endosymbiont (*Hamiltonella defensa*) in *B. tabaci* MEAM1 has been investigated using genome-scale metabolic modeling (Ankrah *et al.*, 2017). This study showed that *B. tabaci* (MEAM1) interact with resident endosymbionts in the biosynthesis of essential amino acids in the bacteriocyte, though *Hamiltonella* acted as a nutritional “parasite”, retaining most or all its essential amino acids and B vitamins that it is capable of synthesizing. The cryptic nature of *B. tabaci* species complex and varying prevalence of endosymbionts in different *B. tabaci* species necessitates that essential amino acid biosynthesis and protein provisioning in *B. tabaci* is studied in different species to identify key symbiosis genes for effective *B. tabaci* control in different

cropping systems. Recent studies on the prevalence of endosymbionts in cassava whiteflies showed that cassava *B. tabaci* (SSA1-SG1) did not possess *Hamiltonella defensa* (Ghosh *et al.*, 2015). Its therefore important to identify the secondary endosymbionts that co-exist with *Portiera* in the bacteriocytes of SSA1-SG1, ascertain their metabolic role especially in amino acid provisioning to the *B. tabaci* and determine whether shared metabolic patterns are conserved in different *B. tabaci* species.

This study was initiated therefore to study the interaction between cassava *B. tabaci* (SSA1-SG1) and its primary and secondary endosymbionts in the bacteriocyte using genome-scale metabolic modeling, to identify key gene targets responsible for amino acid provisioning in cassava *B. tabaci* SSA1-SG1. SSA1-SG1 is a superabundant *B. tabaci* species that has been associated with the spread of cassava brown streak disease and cassava mosaic disease epidemics in cassava producing areas, especially in East Africa (Legg *et al.*, 2014; Legg *et al.*, 2015). It was hypothesized that identification of key symbiosis genes in SSA1-SG1 and downregulating them using RNAi in novel management tools will lead to the effective suppression of cassava *B. tabaci* SSA1-SG1 and the cassava diseases (CMD and CBSD) they transmit.

4.2 MATERIALS AND METHODS

4.2.1 Dissection of bacteriocytes from cassava *B. tabaci*

The cassava *B. tabaci* population (SSA1-SG1) used in this study was derived from a single virgin female and male originally collected from Namulonge, Uganda, an area with superabundant *B. tabaci* populations and a high incidence of both Cassava Brown Streak Disease and Cassava Mosaic Disease. The *B. tabaci* population was maintained at the Natural Resources Institute, the University of Greenwich on aubergine (*Solanum melongena* cv. Black Beauty) under insectary conditions (60% relative humidity, 28°C and 12h light:12h dark regime) with 500 $\mu\text{mol m}^{-2} \text{s}^{-1}$ PAR. light intensity. Bacteriocytes were dissected from *B. tabaci* SSA1-SG1 adult female flies as described in section 3.2.1. It is worth noting that the bacteriocyte sample used in this chapter was the same sample used in Chapter 3. RNA extraction, rRNA depletion, library preparation and sequencing were done as described in section 3.2.2, 3.2.3 and 3.3.4 respectively.

4.2.2 Identification of metabolic genes for both SSA1-SG1 and *Portiera aleyrodidarum*

B. tabaci SSA1-SG1 reads were mapped using blastn to a *Portiera aleyrodidarum* reference assembly from NCBI, REFseq id NC_018677.1. Reads for the two SSA1-SG1-*Portiera* were mapped and the blast mappings were parsed and paired, taking the best hits in each case using the highest bit scores. Where there was a conflict, resulting from a read mapping equally to two locations because of the same bit score, the longest match length was picked. Matching read ids were then used to retrieve the original fastq reads which were paired and assembled using the Sassy assembler¹¹ as described in (Kazakoff *et al.*, 2012; Visendi *et al.*, 2016). The generated *Portiera* genome was then uploaded to RAST (Rapid Annotation Using Subsystem Technology) server (Aziz *et al.*, 2008) and annotated using RAST annotation scheme with a gene caller based on release70 FIGfam version. Metabolic genes with their respective enzyme commission numbers (EC numbers) were used for the reconstruction of the genome-scale metabolic model of *Portiera*.

For *B. tabaci* (SSA1-SG1), the coding sequences were obtained from the draft genome currently undergoing annotation. A Blastx search (Altschul *et al.*, 1997) with a cutoff E-value of 1e-3 against

¹¹ <https://github.com/minillanim/SaSSY/blob/master/src/Assembler.h>

NCBI protein reference non-redundant database (nr), mapping and annotation for all these sequences was done using Blast2GO suite (Conesa *et al.*, 2005). Coding sequences with EC numbers were used to construct the genome-scale metabolic reconstruction for cassava *B. tabaci* (SSA1-SG1).

4.2.2 Genome-scale reconstruction of *Portiera* and SSA1-SG1 metabolic models

Single compartment models for *Portiera* and *B. tabaci* (SSA1-SG1) were reconstructed following the procedures reported by Thiele & Palsson (2010). In brief, Enzyme Commission (EC) numbers for coding sequences of both *Portiera* and SSA1-SG1 were used to identify enzymes and pathways for the selected metabolic genes in Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa & Goto, 2000). Reconstruction assembly was assembled manually in a pathway-pathway manner based on KEGG pathways using an Excel spreadsheet. Genes associated with each EC number were noted and used in constructing gene-protein-reaction (GPR) based on Boolean rules. Distinct proteins catalyzing the different enzymatic reaction to produce the same metabolite or product were considered as isozyme, therefore, an “OR” rule was applied while proteins from different transcripts acting together to catalyze a given reaction (complex reaction) were assigned with the “AND” rule. The selected enzyme was further analyzed using the BiGG Models, a knowledgebase of genome-scale metabolic network reconstructions (King *et al.*, 2016) to determine the exact reaction stoichiometry of the biochemical reaction, chemical formula, changes of metabolites and reaction direction and reversibility. Other databases used to verify reaction stoichiometry of the biochemical reactions include EcoCyc (Keseler *et al.*, 2013) and Braunschweig enzyme database (BRENDA) (Chang *et al.*, 2015).

On completion, the reconstruction assembly for both *Portiera* and *B. tabaci* consisted of (i) reaction code, (ii) reaction equation, (iii) genes, (iv) gene-protein-reaction (GPR) association, (v) subsystem, (vi) reaction thermodynamics (reversibility), (vii) reaction constraints, (viii) confidence score, (ix) EC number for each reaction, (x) Biomass objective function and (xi) metabolite list with both charged and uncharged formula. Refining and manual curation of the model were done by (i) analyzing and verifying substrate and cofactor usage for each reaction, (ii) adding sink and demand reaction, (iii) verifying subcellular localization for each reaction and (iv) adding transport and exchange reactions which represent interaction and communication between

Portiera and SSA1-SG1 *B. tabaci*. A pseudo reaction for biomass generation was developed as the objective function for the model as described by (Feist & Palsson, 2010; Ankrah *et al.*, 2017). Briefly, each transcript expressed in cassava whitefly was translated and the number of each amino acid residue for each protein translated from each transcript was calculated. Transcript abundance for each transcript was multiplied by the total number of amino acid residues for each protein translated from each transcript to determine the total amount of each amino acid translated from each transcript. This was then used to calculate the total amount of each amino acid expressed in the cassava whitefly. The amino acid quantity for each amino acid was then used to determine the fractional contribution of each amino acid as a ratio of each amino acid to the total amino acid translated in the cassava whitefly. This fractional contribution was converted into fractional proportion of each amino acid ($\text{Mmol}_{\text{monomer}}/\text{gDw}$) required to fulfil the total protein requirement (Mmol/gDw) for cassava whitefly. These fractional proportions were then used to construct the objective function for the survival of cassava whitefly (SSA1-SG1) (Supplementary table 6.5). Stoichiometric coefficients in the objective function describe the relative abundance of essential building blocks, focusing on amino acid synthesis (amino acids, ATP and water) that enables *Portiera* and *B. tabaci* to grow and survive. This reaction was selected as the objective in the calculation of optimal flux distribution for the *Portiera* and *B. tabaci* SSA1-SG1 model.

The two single compartment models generated (*iKT90* and *iKT330*) were combined to form a two-compartment model of *B. tabaci* SSA1-SG1 and *Portiera*, that was used for analysis and identification of essential symbiosis genes. The reactions in single models were renamed with a prefix SSA1_Por for *Portiera* model (*iKT90*) and SSA1_Bt for the *B. tabaci* SSA1-SG1 model (*iKT330*). Metabolites in *iKT90* and *iKT330* were localized in [p] and [c] compartments respectively. Transport reactions were assigned to ensure that dead-end metabolites from each compartment are exchanged between the two compartments (*Portiera* and bacteriocyte) and import essential metabolites into the compartment. Exchange reactions were also included so that the metabolites or products produced by the two compartments are exported to the external environment (*B. tabaci* haemolymph).

4.2.3 Conversion of the metabolic reconstruction into a mathematical model

The curated metabolic reconstruction was converted into the genome-scale metabolic model, a stoichiometric matrix using Constraint-based reconstruction and analysis (COBRA) toolbox (Schellenberger *et al.*, 2011) running within MATLAB environment (Mathworks Inc., Natick, MA) using the Gurobi solver¹². The stoichiometric matrix enables computation of metabolic capabilities and quantitative prediction of cellular behaviour using constraint-based approaches. Constraints implemented on individual reaction fluxes included defining lower and upper bounds. Bounds define the space of allowable flux distribution that estimate the rate at which a metabolite is produced or consumed by a reaction. In the *Portiera* model, bounds for exchange reactions were set to zero and 1000 mmol gDw⁻¹ h⁻¹ for lower and upper bound respectively to prevent metabolite uptake while leaving the rate of secretion of desired metabolite unconstrained. For irreversible reactions, the lower bound was constrained to 0 mmol gDw⁻¹ h⁻¹ while reversible reactions were constrained to -1000 mmol gDw⁻¹ h⁻¹. In the *B. tabaci* metabolic model, RNAseq expression data were used to set the upper bounds. For reaction mediated by a single gene or many genes with 'OR' Boolean rule, the highest value of the expression (log₂CPM) was used, while the mean of expression values for all the genes was used for the reactions with 'AND' Boolean rule. Another constraint considered was a mass balance which assumed that each metabolite produced or transported into the cell is balanced by the amount of metabolite consumed within or secreted from the cell represented by.

$$S \cdot v = 0$$

$$\text{With } v^{lb} \leq v \leq v^{ub}$$

S corresponds to the stoichiometric matrix of size $m \times n$ for a system with m compounds and n reactions and v to the array of metabolic fluxes with v^{lb} as the lower bound and v^{ub} as the upper bound

¹² <http://www.gurobi.com/>

4.2.4 Analysis and evaluation of the models

Several parameters of the model were assessed. These included; the optimal steady state flux profile and essential reaction/genes in the *Portiera* and *Portiera/B. tabaci* model that would best predict the amino acid requirement for growth of *Portiera* and *B. tabaci*. The model was analyzed using flux balance analysis (Varma & Palsson, 1994) to determine the steady state flux profile that optimizes the objective function and to simulate gene essentiality of the two-compartment model. This also allowed identification of network gaps, blocked reactions or dead-end metabolite and determine whether the model generates the precursor metabolites for reactions.

The amino acid requirement for growth of *Portiera* and *B. tabaci* (SSA1-SG1) were determined using flux balance analysis (FBA) with simulations performed under aerobic conditions with a maximum oxygen uptake rate of $20 \text{ mmol gDW}^{-1}\text{h}^{-1}$, fructose and glucose as carbon sources and ammonia as the nitrogen source. The objective function for the two-compartment model was developed to optimize the amino acid requirement for both *Portiera* and *B. tabaci*. FBA was also used to determine the effect of gene/reaction deletion on the metabolic model. Each reaction was perturbed, one at a time, by constraining both the upper and lower bound to zero to restrict it from carrying flux. Gene or reaction knock out were implemented and growth simulated using *SingleGeneDeletion* or *SingleReactionDelection* function of COBRA Toolbox. A reaction was classified as essential or critical if when removed from the model, the resultant value of biomass flux is zero (the network does not support growth) or non-essential if perturbation caused no effect on the network (Network supports growth). Genes mediating reactions in essential amino acid biosynthesis were selected from the set of identified essential genes as potential gene targets for controlling cassava *B. tabaci* through RNAi. Minimization of metabolic adjustment (MOMA) (Segre *et al.*, 2002) and regulatory on/off minimization (ROOM) (Shlomi *et al.*, 2005) were also used to confirm the essential genes selected.

4.3 RESULTS

4.3.1 Assembly and functional annotation of *Portiera* genome from *B. tabaci* SSA1-SG1

A single compartment metabolic model for SSA1-SG1-*Portiera* was reconstructed using the assembled SSA1-SG1-*Portiera* genome. The assembled *Portiera* genome from cassava *B. tabaci* SSA1-SG1 was comparable in size, percentage GC content and number of coding sequences to *Portiera* genome from *B. tabaci* species MEAM1 and MED. It had a size of 347165 bp compared to 352068 bp and 357461 bp of *Portiera* from MEAM1 and MED respectively (Table 4.1).

Table 4. 1: Comparison of genomic features of *Portiera* genome from MEAM1, MED and SSA1-SG1

Description	MEAM1	MED	SSA1-SG1
Size (bp)	352068	357461	347165
GC content (%)	26.2	26.1	26.2
Number of coding sequences	282	290	291
Number of subsystems	66	71	65
Number of RNAs	36	36	36
Number of unique hypothetical proteins	23*	27**	31***

*present in MEAM1 but missing in SSA1-SG1 **present in MED but missing in SSA1-SG1 ***present in SSA1-SG1 but missing in MEAM1

A total of 291 coding sequences were annotated using both RAST and Blast2GO to determine the function of the genes in the SSA1-SG1 *Portiera* genome. Sequence distribution of genes in the SSA1-SG1-*Portiera* genome revealed that the highest percentage of genes encoded for gene translation (21%), ribosome biogenesis (15%) and alpha-amino acid biosynthetic processes (13%) (Figure 4.1). Direct Gene Ontology (GO) count of the biological processes for the coding sequences confirmed the contribution of *Portiera* in the biosynthesis and metabolism of amino acids in the bacteriocyte. Of the 20 top direct GO counts of biological processes, 10 related to amino acids biosynthesis and metabolic processes (Figure 4.2).



Figure 4.1: Distribution of sequences as assigned to different categories of biological processes encoded by genes in *Portiera aleyrodidarum* genome from *B. tabaci* SSA1-SG1

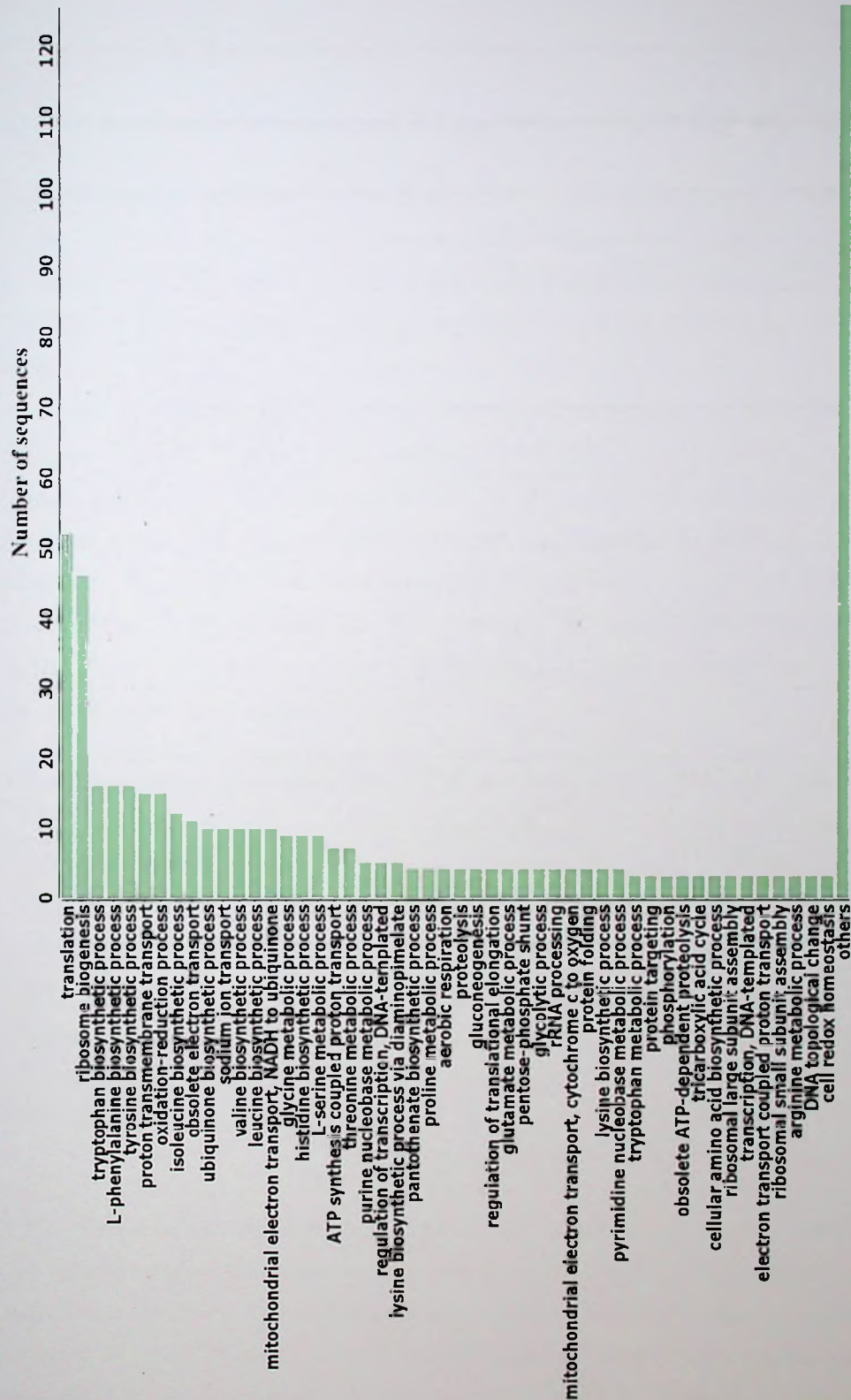


Figure 4.2: Direct GO counts for biological processes encoded by genes in *Portiera aleyrodidarum* genome from *B. tabaci* SSA1-SG1

4.3.2 Comparison of *Portiera* from *B. tabaci* SSA1-SG1, MEAM1 and MED

Although *Portiera* genome from MEAM1, MED and SSA1-SG1 were comparable in size, GC content, number of coding sequences and RNAs, comparative functional analysis was done to identify protein/function missing in the three *Portiera* genomes. A total of six unique proteins were identified, all of which were missing in SSA1-SG1-*Portiera* (Table 4.2). These included (i) 4-hydroxy-tetrahydrodipicolinate reductase *DapB* (EC 1.17.1.8), (ii) inner membrane protein translocase component YidC, (iii) tRNA uridine 5-carboxymethylaminomethyl enzyme (GidA), (iv) cytochrome O ubiquinol oxidase subunit 1, (v) GTPase and tRNA-U34 5-formylation enzyme (TrmE) and (vi) SSU ribosomal protein S6p. Of these, only one metabolic enzyme (4-hydroxy-tetrahydrodipicolinate reductase--EC 1.17.1.8) was identified in *Portiera* from MEAM1 and MED but missing in *Portiera* from cassava *B. tabaci* SSA1-SG1 (Table 4.2). Previous study by (Luan et al., 2015) showed that the metabolic enzyme (4-hydroxy-tetrahydrodipicolinate reductase--EC 1.17.1.8) that is missing in SSA1-SG1-*Portiera* is encoded by *dapB*, a key gene in the lysine biosynthesis pathway.

Table 4.2: Missing protein/function in *Portiera* genome from MEAM1, MED and SSA1-SG1

Unique protein/function	MEAM1	MED	SSA1-SG1
4-hydroxy-tetrahydrodipicolinate reductase (EC 1.17.1.8)	+	+	-
Inner membrane protein translocase component YidC	-	+	-
tRNA uridine 5-carboxymethylaminomethyl enzyme (GidA)	-	+	-
Cytochrome O ubiquinol oxidase subunit 1	+	+	-
GTPase and tRNA-U34 5-formylation enzyme TrmE	-	+	-
SSU ribosomal protein S6p	-	+	-

4.3.3 Single compartment models of SSA1-SG1-*Portiera* and *B. tabaci* SSA1-SG1

The reconstructed single compartment metabolic model for SSA1-SG1-*Portiera* comprised of 76 intercellular reactions, 150 metabolites and a total of 90 genes hence supporting a metabolic network, *iKT90*. The model had an optimal solution based on flux balance analysis of 6.70, showing that the flux through the reconstructed metabolic network maximizes growth rate,

resulting in a predicted exponential growth rate of 6.70 h^{-1} . The symbiosis system in both SSA1-SG1 (*iKT90*) and MEAM1 (*iNA94*) is quite conserved. Comparison of both single compartment model of *Portiera* from both SSA1-SG1 and MEAM1 showed that the two models are quite similar in the number of intercellular reactions, genes and metabolites (Table 4.3).

Table 4.3: Comparison of genome-scale metabolic models of *B. tabaci* SSA1-SG1 and MEAM1

Description	<i>B. tabaci</i> SSA1-SG1			<i>B. tabaci</i> MEAM1		
	<i>iKT90</i>	<i>iKT330</i>	<i>iKT420</i> **	<i>iNA94</i>	<i>iNA332</i>	<i>iNA774</i> ***
Number of metabolic genes	90	330	420	94	332	774
Number of intracellular reactions	76	233	310	76	236	774
Number of metabolites	150	251	402	148	253	550
Flux balance optimal solution (h^{-1})	6.70	25.38	0.31	12.02	26.55	0.20
Total protein (mmol/gDw)			6.17			6.20
Total cost for protein synthesis			26.58			26.71

** Two compartment model for *B. tabaci* SSA1-SG1 and *Portiera*

*** Three-compartment model for *B. tabaci* MEAM1, *Portiera* and *Hamiltonella* (Ankrah *et al.*, 2017)

The primary role of *Portiera* in the bacteriocyte relates to amino acid biosynthesis. The reconstructed metabolic model for *Portiera* was analyzed to determine essential amino acid provisioning and the genes mediating critical reaction in the essential amino acid biosynthesis in the cassava whitefly bacteriocyte. *Portiera* as a primary symbiont in *B. tabaci* SSA1-SG1 contributes final/terminal reactions for the synthesis of three essential amino acids. These include (i) threonine, (ii) methionine and (iii) tryptophan (Figure 4.6, Figure 4.7 & Figure 4.8). Once synthesized, the three essential amino acids are exported to the bacteriocyte, where they are made available to the *B. tabaci*. Furthermore, *Portiera* also contributes to the synthesis of metabolites for intermediate reactions of other essential amino acids. For example, it possesses genes that encode for proteins required for chorismate biosynthesis, which is a precursor of aromatic amino acids (phenylalanine, tryptophan and tyrosine) (Figure 4.8).

The bulk of reactions of the biosynthesis pathways of seven essential amino acids (arginine, histidine, lysine, phenylalanine, isoleucine, valine and leucine) are mediated by genes in *Portiera*.

Metabolites produced in these reactions are exported across a symbiosome membrane into the bacteriocyte for amino acid biosynthesis. Metabolites (intermediate precursors) exported across the symbiosome membrane include, (i) N(Omega)-(L-Arginino) succinate for arginine biosynthesis, (ii) L-Histidinol phosphate for histidine biosynthesis, (iii) LL-2,6-Diaminoheptanedioate for lysine biosynthesis, (iv) phenylpyruvate for phenylalanine biosynthesis, (v) (5)-3-methyl-2-oxopentanoate for isoleucine biosynthesis, (vi) 3-methyl-2-oxobutanoate for valine biosynthesis and (vii) 3-carboxy-4-methyl-2-oxopentanoate for leucine biosynthesis (Figure 4.5, 4.6, 4.7 & 4.8).

In addition, the single compartment model of *Portiera* revealed that *Portiera* lacks genes that encode for metabolites required in the biosynthesis of non-essential amino acids (alanine, asparagine, aspartic acid, cysteine, glutamic acid, glutamine, glycine, proline, serine and tyrosine). This implies that *Portiera* depends on the *B. tabaci* SSA1-SG1 for the provision of non-essential amino acids. It also requires eight host-derived metabolites for the biosynthesis of essential amino acids. These include, (i) 2, 3, 4, 5-tetrahydrodipicolinate (Figure 4.5), (ii) L-aspartate (Figure 4.6), (iii) alpha-D-ribose 5-phosphate (Figure 4.7), (iv) 5-methyltetrahydrofolate (Figure 4.7), (v) L-glutamine, and (vi) L-serine (Figure 4.8). These metabolites are imported across the symbiosome membrane by *Portiera* as precursors in different essential amino acid biosynthesis pathways.

4.3.4 Two-compartment models of *B. tabaci* SSA1-SG1 and *Portiera aleyrodidarum*

4.3.4.1 Identification of endosymbionts in the SSA1-SG1 bacteriocyte

Secondary endosymbionts in the bacteriocytes also make a contribution to essential amino acid biosynthesis, albeit little (Ankrah *et al.*, 2017). Therefore, a preliminary study was done to identify endosymbionts co-existing in the SSA1-SG1 bacteriocyte. Modified primers for 16S rRNA (907Fmod and 1237R) (Jing *et al.*, 2014) were used to amplify the V6-V7 region of the all endosymbionts in the bacteriocytes. Prior to this analysis, *In silico* analysis of PacBio reads of SSA1-SG1 (used for assembly of SSA1-SG1 genome) revealed traces of *Wolbachia*, therefore both wholebody and bacteriocyte samples of SSA1-SG1 were compared with *B. tabaci* MED_AS1 and a control, which were both positive for *Wolbachia* and other endosymbionts. Digestion of 16S rRNA amplicon of the V6-V7 region using *MwoI* restriction enzyme showed that MED_AS1 had both *Portiera* (restriction bands of 331kb) and *Wolbachia* (restriction band of 171kb) while *B.*

tabaci SSA1-SG1 had only *Portiera* in its bacteriocyte (Figure 4.3). Phylogenetic analysis of digested 16S rRNA amplicon of the V6-V7 region from SSA1-SG1 bacteriocytes confirmed that the dominant endosymbiont was *Portiera* (Figure 4.4). Based on these results, a two-compartment model for *B. tabaci* SSA1-SG1 and *Portiera* was constructed to identify the essential gene targets in the different essential amino acid biosynthetic pathways.

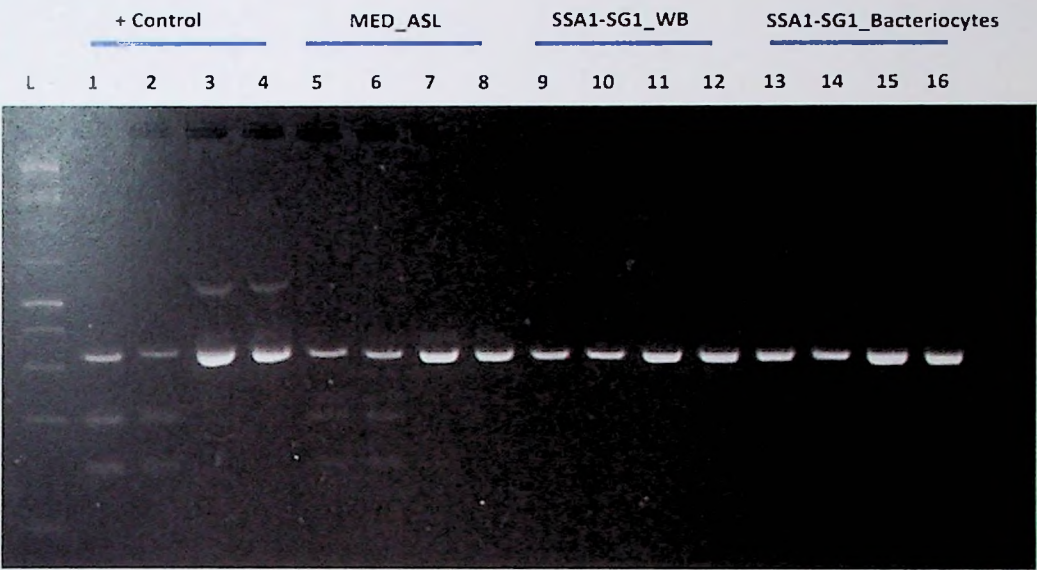


Figure 4.3: Identification of secondary symbionts co-existing with *Portiera aleyrodidarum* in the cassava *B. tabaci* (SSA1-SG1) bacteriocyte. Restriction pattern of endosymbionts based on 16S for V₆-V₇ region generated using *Mwo*I restriction enzyme and visualized using a 1 kb ladder (Quick-load, New England BioLabs Inc). Lane 1 and 2, positive control for *Wolbachia* - digested, lane 3 and 4, positive control for *Wolbachia* - undigested PCR product, lane 5 and 6, MED_AS1 – digested, lane 7 and 8, MED - undigested PCR product, lane 9 and 10, SSA1-SG1 whole-body - digested, lane 11 and 12, SSA1-SG1 whole-body - undigested PCR product, lane 13 and 14 SSA1-SG1 bacteriocytes – digested and lane 15 and 16 SSA1-SG1 bacteriocyte - undigested PCR products.

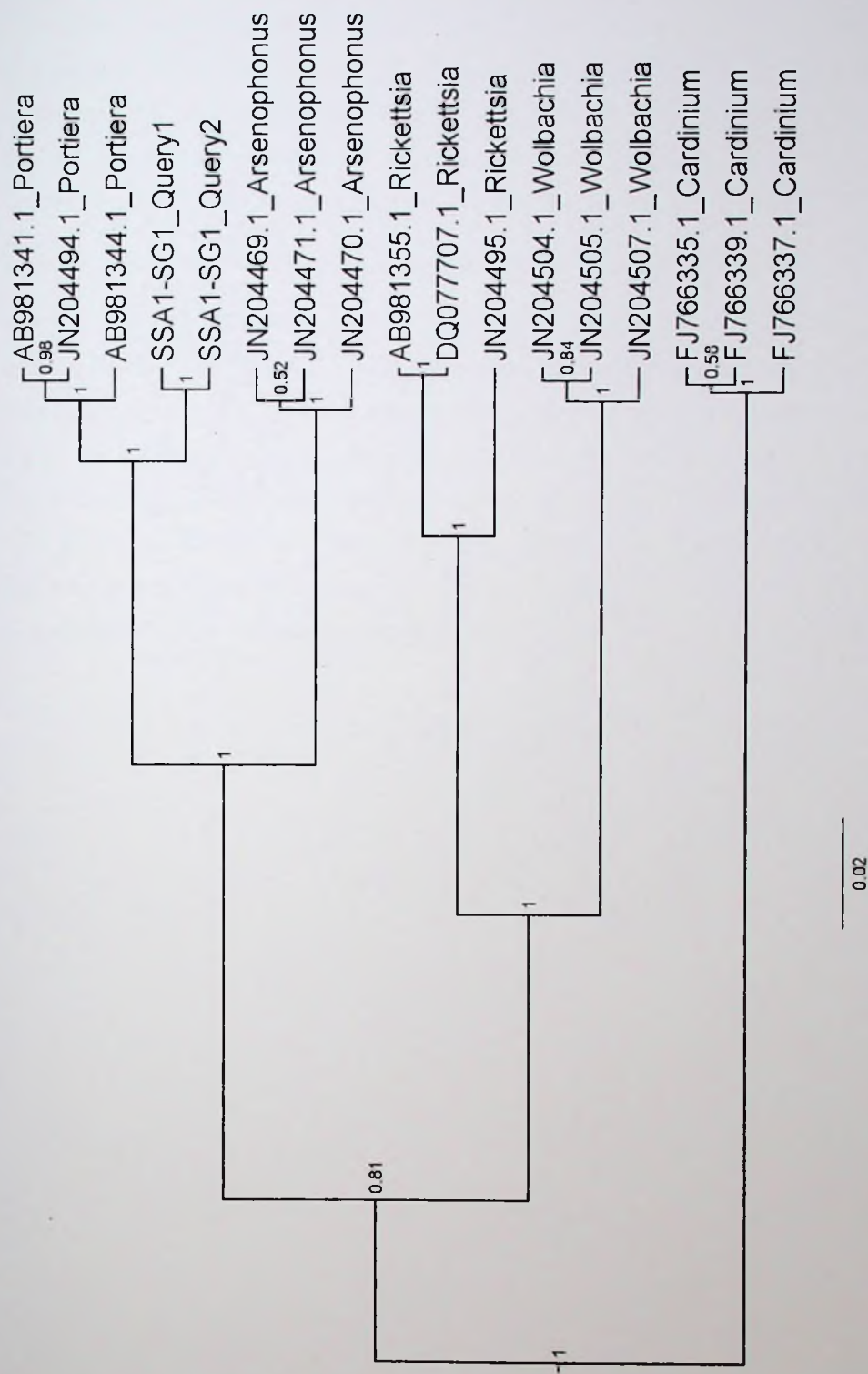


Figure 4.4: Phylogenetic relationship of the 16S V6-V7 region of selected *B. tabaci* endosymbionts. Identity of the two SSA1-SG1 query DNA sequences (Figure 13B, lane 15-16). Phylogenetic analysis was conducted using Bayesian methods implemented in Bayesian Evolutionary Analysis Sampling Trees (BEAST version 1.10.2) (Suchard *et al.*, 2018).

Therefore, a two-compartment model (*iKT420*) was reconstructed combining both the single compartment model for *Portiera* (*iKT90*) and for *B. tabaci* (*iKT330*). The naming of the two-compartment model (*iKT420*) for cassava whitefly (SSA1-SG1) followed the standard system biology procedure of naming metabolic models with *i* representing *insilico* metabolic model containing 420 metabolic genes (Table 4.3) generated by Kaweesi Tadeo (KT).

4.3.4.2 Shared metabolic interaction between *Portiera* and SSA1-SG1.

Terminal reactions in the essential amino acid biosynthesis pathway are important because they influence the provisioning of amino acid. Genome-scale metabolic reconstruction of both *B. tabaci* SSA1-SG1 and its primary endosymbiont revealed that the terminal reaction of isoleucine, leucine, valine, histidine and phenylalanine are all mediated by genes of intrinsic origin in the host bacteriocyte. For example, for branched-chain amino acids, one enzyme, branched-chain amino acid transaminases (EC 2.6.1.42) encoded by *BCAT*, mediate the conversion of 3-methyl-2-oxobutanoate to valine, 4-methyl-2-oxopentanoate to leucine and (S)-3-Methyl-2-oxopentanoate to isoleucine (Figure 4.6). other genes include, *hisD* (EC 1.1.1.23) and *aspC* (EC 2.6.1.58).

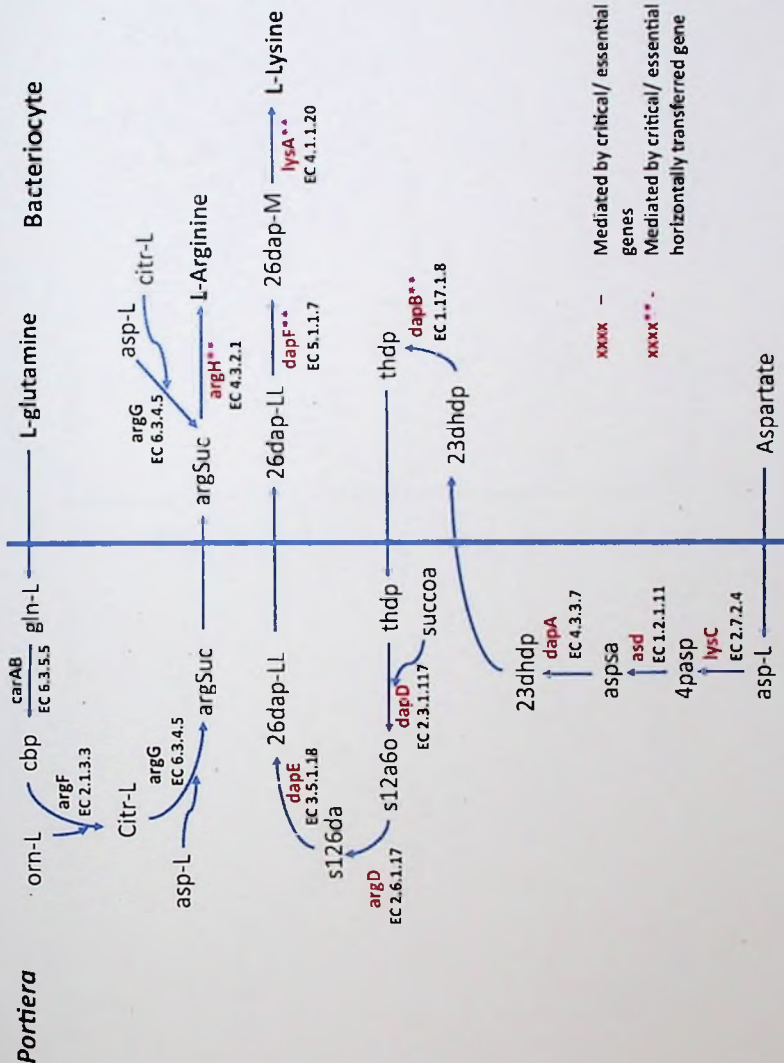


Figure 4.5: Biosynthetic pathway for two essential amino acids (i) lysine comprising of asp-L (Aspartate), 4pas (4-Phospho-L-aspartate), aspA (L-Aspartate 4-semialdehyde), 23dhdp (2,3-Dihydrodipicolinate), thdp (2,3,4,5-Tetrahydrodipicolinate), s12a60 (N-Succinyl-2-L-amino-6-oxoheptanedioate), s126da (N-Succinyl-L-2,6-diaminoheptanedioate), 26dap-LL (LL-2,6-Diaminoheptanedioate), 26dap-M (Meso-2,6-Diaminoheptanedioate) and (ii) Arginine with the following metabolites; gln-L (L-Glutamine), cbp (Carbamoyl phosphate), orn-L (L Ornithine), Citr-L (L-Citrulline) and argSuc (N(omega)-(L-Arginino)succinate)

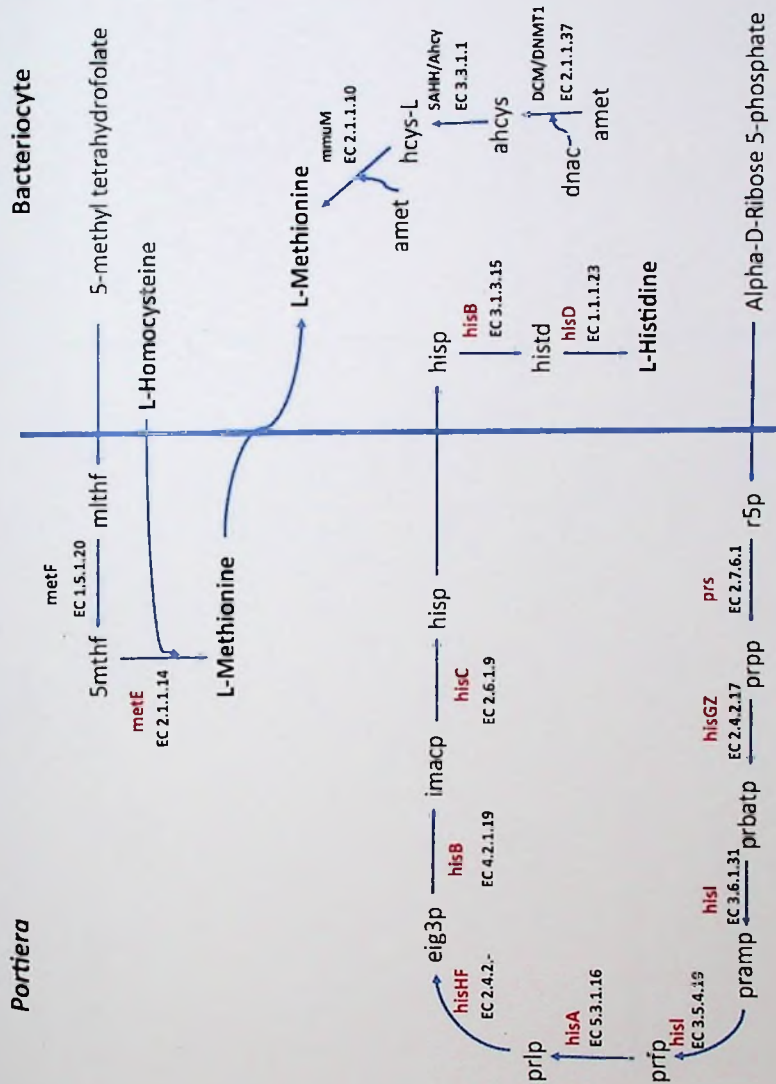


Figure 4.7: Biosynthetic pathway for histidine and methionine with essential genes highlighted in red. Metabolites produced in (i) histidine pathway include: r5p (Alpha-D-Ribose 5-phosphate), prpp (5-Phospho-alpha-D-ribose 1-diphosphate), prbatp (1-(5-Phosphoribosyl)-ATP), pramp (5-Phospho-beta-D-ribosylamine), prfp (1-(5-Phosphoribosyl)-5-[(5-phosphoribosylamino)methylideneamino]imidazole-4-carboxamide), prlp (5-[(5-phospho-1-deoxyribulos-1-ylamino)methylideneamino]-1-(5-phosphoribosyl)imidazole-4-carboxamide), eig3p (D-erythro-1-(Imidazol-4-yl)glycerol 3-phosphate), inacp (3-(Imidazol-4-yl)-2-oxopropyl phosphate), hisp (L-Histidinol phosphate), histd (L-Histidinol) (ii) Methionine pathway: mlthf (5,10-Methylenetetrahydrofolate), 5mthf (5-Methyltetrahydrofolate), amct (S-Adenosyl-L-methionine), ahcys (S-Adenosyl-L-homocysteine) and hcs-L (L-Homocysteine).

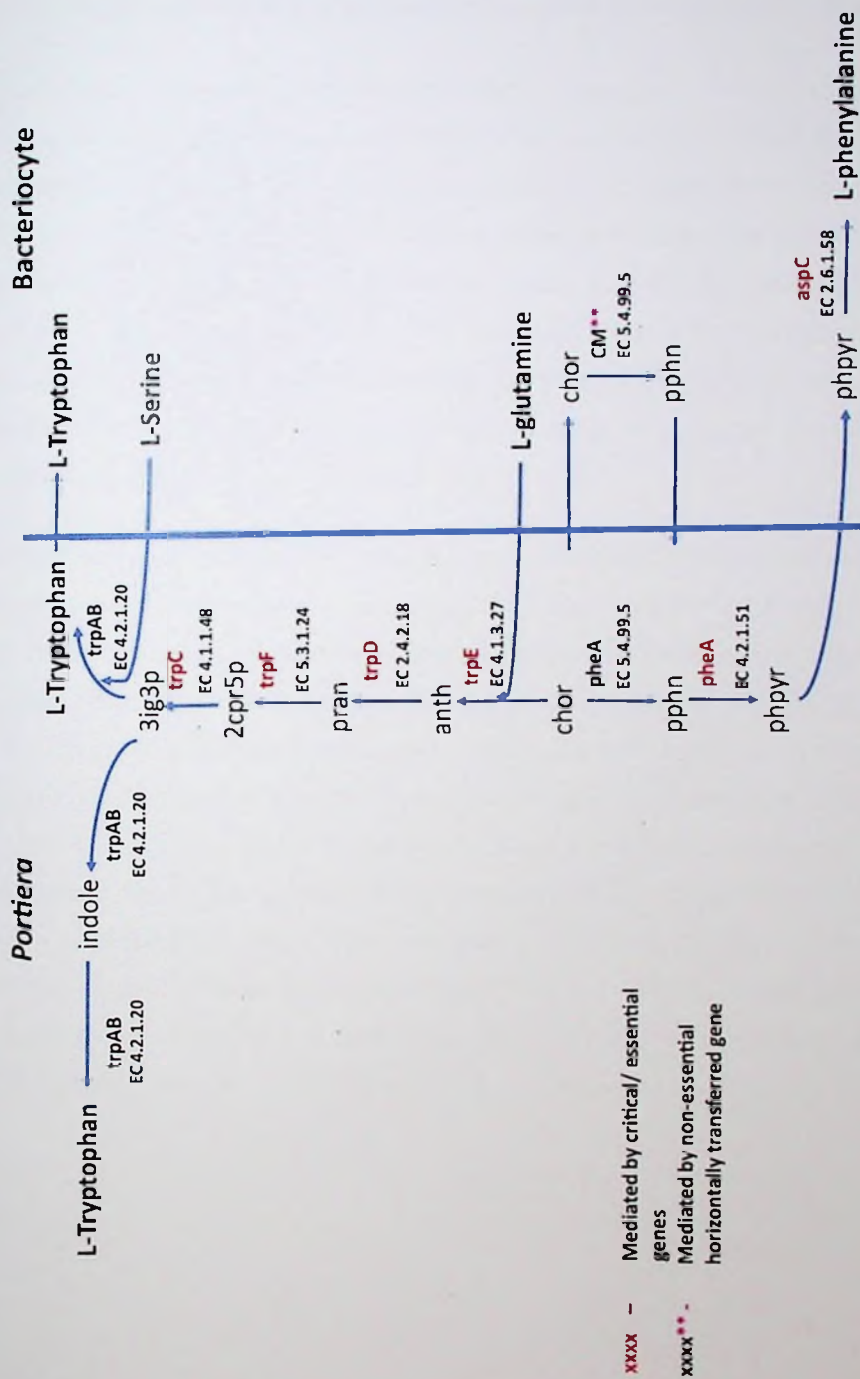


Figure 4.8: Biosynthesis pathway for two essential amino acids. (i) Phenylalanine pathway involving chor (Chorismate), pphn (Prephenate), phpvr (Phenylpyruvate) and (ii) Tryptophan pathway consisting of the following metabolites: 3ig3p (C'-3-Indolyl)-glycerol 3-phosphate), anth (Anthranilate), pran (N-(5-Phospho-D-ribose) anthranilate), indole and 2cpr5p (1-(2-Carboxyphenylamino)-1-deoxy-D-ribose 5-phosphate)

4.3.5 *In silico* prediction of essential amino acid synthesis rate in *B. tabaci* SSA1-SG1

Flux balance analysis was used to predict the *in-silico* production rate of essential amino acids by the two-compartment genome-scale metabolic model of *B. tabaci* SSA1-SG1. The predicted production rate of different essential amino acids in *B. tabaci* SSA1-SG1 varied from 0.05 mmol gDW⁻¹h⁻¹ to 0.21 mmol gDW⁻¹h⁻¹. Three essential amino acids had a relatively high rate of production in the SSA1-SG1 bacteriocytes. These were (i) phenylalanine with 0.21 mmol gDW⁻¹h⁻¹, (ii) leucine with 0.17 mmol gDW⁻¹h⁻¹ and (iii) lysine with 0.16 mmol gDW⁻¹h⁻¹ (Figure 4.9). Some essential amino acids, for example, histidine, methionine, tryptophan and threonine were produced at a relatively low rate. Three of these amino acids (methionine, Threonine and tryptophan) are produced mainly by *Portiera*.

Robustness analysis to determine the systemic effect of varying metabolite flux through terminal reactions of different essential amino acids on the objective (growth of *B. tabaci* SSA1-SG1) revealed five reactions that are sensitive to fluctuation in the maximum allowable flux. These were (i) SSA1-Bt-HISTD for histidine synthesis (ii) SSA1-Por-TRPS1 for tryptophan synthesis (iii) SSA1-Bt-DAPDC for lysine synthesis (iv) SSA1-Bt-LEUTAI for leucine synthesis and (v) SSA1-Bt-ARGSL for arginine, all with maximum allowable flux ranging from 9.5 – 24 mmol gDW⁻¹h⁻¹ (Figure 4.10 & Figure 4.11). The most crucial point in this robustness analysis is the flux at which the optimal flux balance solution is attained and the flux at which the model begins to deviate from the optimal solution, which is the optimum growth rate (Figure 4.10 and Figure 4.11). Some reaction especially terminal reactions within *Portiera* (SSA1-Por-METS, for methionine, SSA1-Por-THRS for threonine and SSA1-Por-TRPS1 for tryptophan) exhibit different patterns, with relatively very low flux required to attain the optimal solution (Figure 4.11 B, C and D)



Figure 4.9: *In silico* prediction of production rates of essential amino acid of the two-compartment metabolic model of *B. tabaci* SSA1-SG1 and *Portiera*.

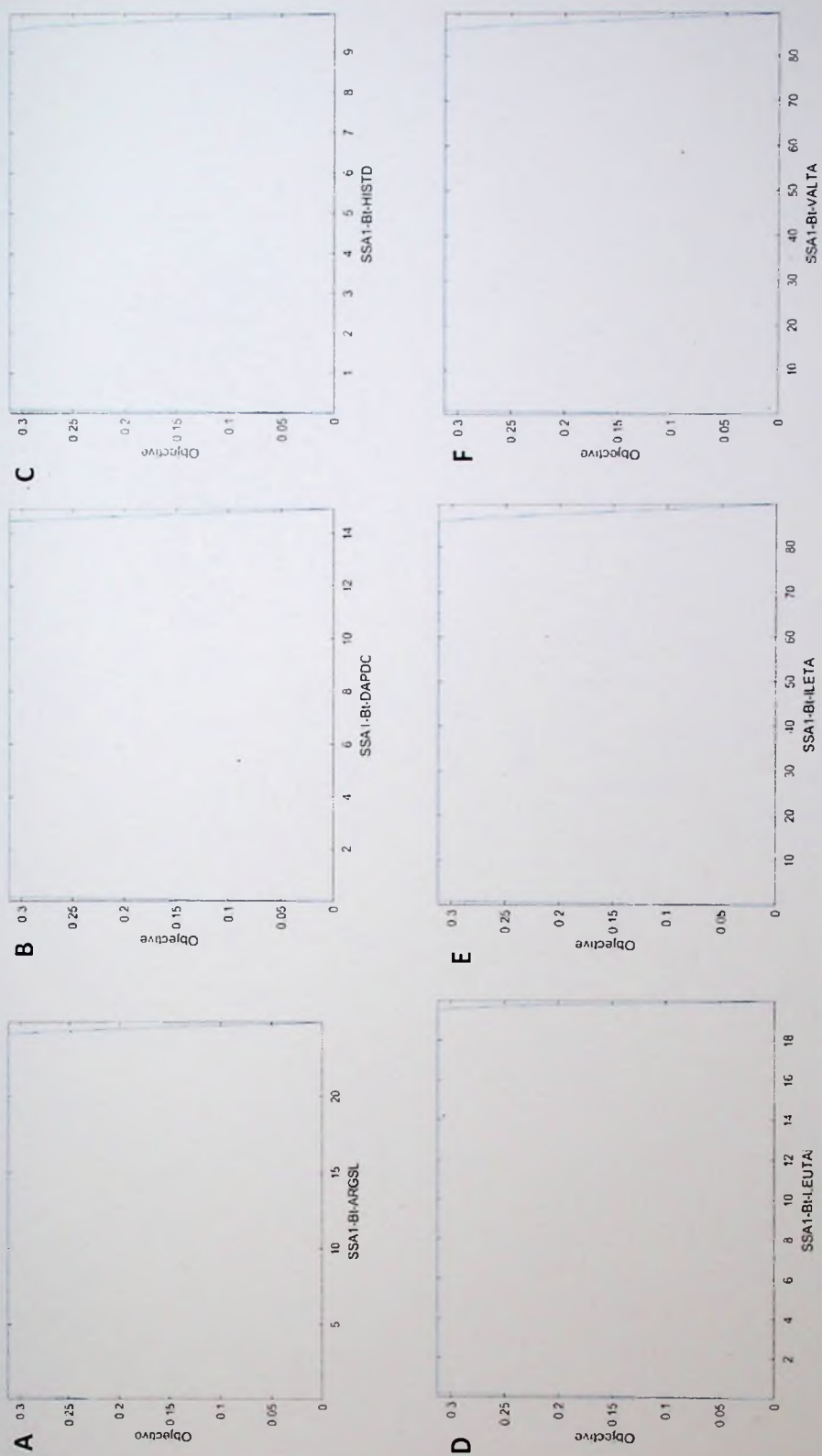


Figure 4.10: Robustness analysis measuring sensitivity of the metabolic object function (growth rate of *B. tabaci* SSA1-SG1) to the quantitative flux levels through terminal reactions for (A) arginine biosynthesis, (B) lysine biosynthesis, (C) histidine biosynthesis, (D) leucine biosynthesis, (E) isoleucine synthesis and (F) valine biosynthesis.

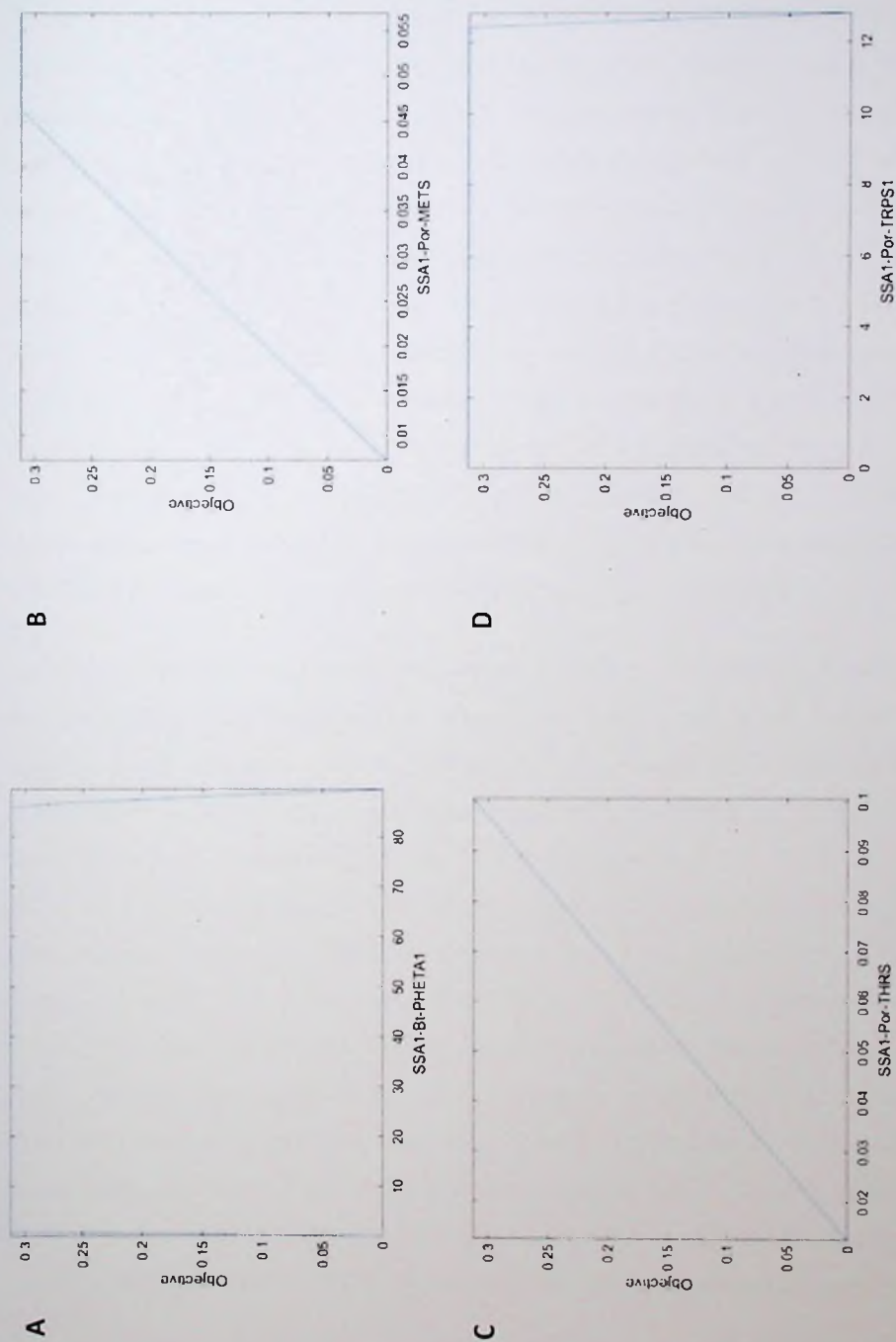


Figure 4.11: Robustness analysis measuring sensitivity of the metabolic object function (growth rate of *B. tabaci* SSA1-SG1) to the quantitative flux levels through terminal reactions for (A) phenylalanine biosynthesis (B) methionine biosynthesis (C) threonine biosynthesis and (D) tryptophan biosynthesis.

4.3.6 Essential gene targets in different essential amino acid biosynthesis pathways in SSA1-SG1

Genome-scale reconstruction of both *B. tabaci* SSA1-SG1 and *Portiera* demonstrated an interdependence of both cassava whitefly and its primary endosymbiont. Analysis of the shared metabolic interaction coupled with constraint-based metabolic modeling offered opportunities to identify and select critical/essential symbiosis genes that can be used as gene targets in the management of cassava whiteflies. Gene essentiality was simulated by constraining each reaction, one at a time, from carrying flux and the effect of this perturbation was assessed by comparing the growth rate of the perturbed network to that of wild type (Figure 4.12). Based on the growth rate, a total of 270 reactions were characterized as essential. Of these, 129 reactions were mediated by genes in *Portiera* while 141 reactions were mediated by genes in *B. tabaci* SSA1-SG1 (Supplementary table 6.6). Essential amino acid biosynthesis pathways in *B. tabaci* SSA1-SG1 had a varying number of indispensable genes. Histidine pathway had the highest number of indispensable genes (10 genes) (Figure 4.7) followed by lysine and leucine biosynthesis pathway with 9 and 6 essential genes respectively (Figure 4.5 & Figure 4.6).

Single gene deletion also revealed that seven of the ten biosynthesis pathway had essential genes mediating their terminal reactions. Therefore, these genes were characterized as accessory essential genes in the provisioning of essential amino acids in the cassava *B. tabaci* SSA1-SG1. These include, (i) *argH* (EC 4.3.2.1) for arginine biosynthesis, (ii) *lysA* (EC 4.1.1.20) for lysine biosynthesis, (iii) Branched-chain amino acid transaminase (EC 2.6.1.42) for isoleucine, valine and leucine biosynthesis, (iv) *hisD* (EC 1.1.1.23) for histidine and (v) *aspC* (EC 2.6.1.58) for phenylalanine biosynthesis. Of these, two genes (*argH* and *lysA*) are horizontally transferred genes of bacterial origin while three genes (*BCAT*, *hisD* and *aspC*) are of intrinsic origin (Table 4.4). All the selected reactions except PHETA1 for phenylalanine biosynthesis were mediated by the protein encoded by a single gene. This is contrary to the genes that mediate metabolic inputs (metabolites that are exported across the symbiosome membrane) in different essential amino acid biosynthesis pathways (Table 4.5). Therefore, three factors, (i) indispensability of the gene, (ii) reaction mediated by a single gene and (iii) gene mediating a terminal reaction were considered in the selection of the best candidate gene targets for the management of cassava whitefly SSA1-SG1 (Table 4.4).

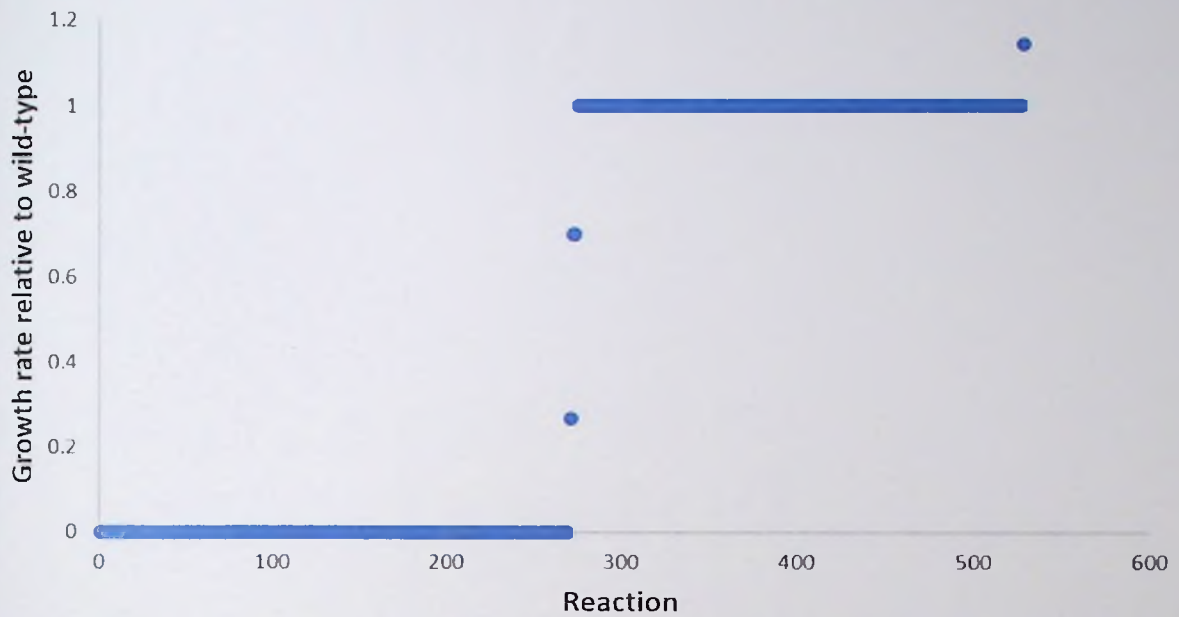


Figure 4.12: Robustness of the metabolic network “SSA1-SG1_*Portiera*” iKT420 showing the effect of single reaction deletion on the growth rate of *B. tabaci* SSA1-SG1 and *Portiera* mutant relative to the wild-type, simulated using flux balance analysis. A total of 270 reaction deletions affected growth, therefore, were deemed indispensable/essential for the growth and survival of cassava *B. tabaci* SSA1-SG1

Table 4.4: Accessory essential genes mediating critical reactions of essential amino acid biosynthesis in *B. tabaci* (SSA1-SG1) bacteriocytes simulated using FBA, MOMA and ROOM

Reaction	Gene name	Enzyme	EC number	Origin*	Flux	Number of isoforms	Pathway
DHDPRy	dapB	Dihydrodipicolinate reductase	1.17.1.8	HTG	0.16	1	Lysine
DAPE	dapF	Diaminopimelate epimerase	5.1.1.7	HTG	0.16	1	Lysine
DAPDC	lysA	Diaminopimelate decarboxylase	4.1.1.20	HTG	0.16	1	Lysine
ARGSL	argH	Argininosuccinate lyase	4.3.2.1	HTG	0.11	1	Arginine
PHETA1	aspC	Phenylalanine transaminase	2.6.1.58	Intrinsic	0.21	2	Phenylalanine
LEUTAI	BCAT	Leucine transaminase	2.6.1.42	Intrinsic	0.17	1	Leucine
ILETA	BCAT	Isoleucine transaminase	2.6.1.42	Intrinsic	0.11	1	Isoleucine
VALTA	BCAT	Valine transaminase	2.6.1.42	Intrinsic	0.15	1	Valine
HISTP	hisB	Histidinol phosphatase	3.1.3.15	Intrinsic	0.05		Histidine
HISTD	hisD	Histidinol dehydrogenase	1.1.1.23	Intrinsic	0.05		Histidine

* HTG – Horizontally transferred genes.

Intrinsic – Host genes

Flux – the amount of substrates/products produced in a certain reaction per unit time (mmol gDW⁻¹ h⁻¹)

Table 4.5: Essentiality of genes mediating synthesis of metabolic inputs in the synthesis of essential amino acids

Metabolite	Reaction	Gene	Enzyme	Essentiality	EC No	No of genes	Pathway
mlthf	GHMT2r	GlyA	Glycine hydroxymethyl transferase	Essential	2.1.2.1	2	Methionine
	GLYCL	GcvT	Glycine cleavage system	Non-essential	2.1.2.10	12	Methionine
	MTHFD		Methylenetetrahydrofolate dehydrogenase	Essential	1.5.1.5	2	Methionine
amet	METAT	metK	Methionine adenosyltransferase	Essential	2.5.1.6	1	Methionine
r5p	NRSPRT	PNP	Nicotinate D – ribonucleoside: orthophosphate ribosyltransferase	Non-essential	2.4.2.1	2	Histidine
	RP1	RpiA	Ribose-5-phosphate isomerase	Essential	5.3.1.6	1	Histidine
asp-L	TKT1	TKtA	Transketolase	Non-essential	2.2.1.1	2	Histidine
	ASPTA	AspC	Aspartate transaminase	Essential	2.6.1.1	3	Lysine, Threonine
	ASNN	AnsA	L-asparaginase	Non-essential	3.5.1.1	6	Lysine, Threonine
gln-L	GLNS	GlnA	Glutamine synthase	Essential	6.3.1.2	3	Arginine, Tryptophan

4.4 DISCUSSION

An obligatory relationship between *B. tabaci* and *Portiera aleyrodidarum* is essential for the survival of whiteflies on the nutritionally impoverished diet of plant phloem sap. This study was initiated to identify accessory essential gene targets that can be used to disrupt this relationship, for the management of the superabundant whitefly population in cassava growing areas. A two-compartment metabolic model of *B. tabaci* SSA1-SG1 and its primary endosymbiont *Portiera* was constructed to represent the shared metabolic interaction between the host and the bacteria in the biosynthesis of essential amino acid. In MEAM1, a three-compartment metabolic model of *Hamiltonella defensa*, *Portiera aleyrodidarum* and *B. tabaci* MEAM1 has been constructed (Ankrah *et al.*, 2017). In this study, *Hamiltonella defensa* was not detected in SSA1-SG1 population tested, neither was any other secondary endosymbiont detected, either in the whole body or in the bacteriocyte. Several studies have been conducted on the prevalence and diversity of endosymbionts in different whitefly populations (Asiimwe *et al.*, 2014; Ghosh *et al.*, 2015; Tajebe *et al.*, 2015). Key findings on cassava *B. tabaci* SSA1-SG1 were, (i) none of the SSA1-SG1 whiteflies tested had *Hamiltonella defensa* and (ii) a relatively large percentage (38%) of the tested SSA1-SG1 whiteflies were completely free of secondary endosymbiont (Ghosh *et al.*, 2015; Tajebe *et al.*, 2015). Therefore, *Portiera* as a sole endosymbiont in the SSA1-SG1 population the was tested and its host exhibited a full capacity for the biosynthesis of all essential amino acids required for growth and survival of the cassava whitefly.

Comparison of reactions mediated in both *Portiera* and bacteriocyte of SSA1-SG1 (iKT420) and MEAM1 (iNA774) showed that symbiosis in these two whitefly species is conserved with slight differences in the simulated growth rate and *in silico* production rate of different essential amino acids. The difference in these two outputs may be attributed to the impact of the third partner (*Hamiltonella defensa*) in the bacteriocyte, a nutritional parasite in the MEAM1-bacterial symbiosis (Ankrah *et al.*, 2017). Co-habitation of different endosymbionts in the same bacteriocytes generates constraints on both the host and primary endosymbiont (Gottlieb *et al.*, 2008), which affects the overall performance of either the host or the primary endosymbiont in the symbiotic relationship. In this case, both high growth and amino acid production rate as observed in this study may have a strong influence on the fitness and population dynamics of *B. tabaci*.

On the other hand, the two-partner model generated in this study contained only three final reactions in the biotin biosynthesis pathway mediated enzymes encoded by *BioA*, *BioD* and *BioB*, horizontally transferred genes in the bacteriocyte. One question remains, however, as to where *B. tabaci* SSA1-SG1 obtains 8-amino-7-oxononanoate, a precursor of 7,8-diaminononanoate, required for the last two reactions in Biotin synthesis? In MEAM1, the secondary endosymbiont, *Hamiltonella defensa* has a complete biotin biosynthesis pathway including reaction mediated enzymes encoded by *BioA*, *BioD* and *BioB*. However, constraint-based modeling revealed that it contributes little or nothing to B vitamins nutrition of its host (Ankrah *et al.*, 2017). These results, therefore, imply (i) *B. tabaci* is slowly gaining an ability to synthesize B vitamins through horizontal gene transfer, (ii) there could be other secondary endosymbionts that are contributing to biotin biosynthesis but the molecular diagnostic methods used in this study as well as in other studies (Ghoshi *et al.*, 2015; Tajebe *et al.*, 2015) are not robust enough to capture all the endosymbionts in *B. tabaci* SSA1-SG1. Therefore to answer this question, more robust methods such as visualization using either electron microscope, confocal microscope or Fluorescence in situ hybridization (FISH) (Kliot *et al.*, 2014) are recommended.

Interaction between *B. tabaci* and their respective endosymbionts has evolved over time. Shared metabolic interaction between these partners shows that whiteflies are slowly gaining genes that are required for the biosynthesis of essential amino acids within the bacteriocytes. These innovative metabolic abilities have been facilitated not only by the symbiotic relationship with bacterial partners but also by lateral transfer of genes of bacterial origin. In this study, two pathways; L-lysine (Figure 4.5) and L-phenylalanine (Figure 4.8) demonstrate host compensation in the bacteriocyte. First, *B. tabaci* SSA1-SG1 acquired a bacterial gene *dapB* (EC 1.17.1.8) that encodes for 4-hydroxy-tetrahydridipicolinate reductase in its lysine biosynthesis pathway. In SSA1-SG1-*Portiera*, *dapB* is missing but the gene is still present in *Portiera* genome of both MEAM1 and MED (Table 4.2). Previous studies by Luan *et al.*, (2015) showed that *dapB* is present in the MEAM1-*Portiera* genome as a pseudogene with significantly reduced expression in the bacteriocyte compared to other genes in the same metabolic pathway. These results imply, (i) SSA1-SG1 is co-evolving with *Portiera* faster than MEAM1 and MED, or (ii) the relaxed functional constraint due to gene redundancy in the shared metabolic interaction in the bacteriocyte leads to loss of essential genes especially by the endosymbiont (Andersson & Kurland, 1998; Rao *et al.*, 2015). Therefore, the loss of *dapB* in *Portiera* has enabled the horizontally acquired *dapB*

gene in the bacteriocyte to take over the role of synthesizing 2,3,4,5-Tetrahydrodipicolinate from 2,3-Dihydrodipicolinate rendering the transferred gene indispensable in the biosynthesis of lysine in SSA1-SG1 bacteriocyte. Second, a similar compensation process is still on-going in phenylalanine biosynthesis pathway. *B. tabaci* SSA1-SG1 acquired *CM* (EC 5.4.99.5) encoding chorismate mutase, an enzyme that converts chorismate to prephenate, a precursor of phenylpyruvate (Figure 4.8). Essentiality analysis revealed that *CM* is dispensable. This is because there are two pathways that can lead to prephenate synthesis, one mediated within *Portiera* and the other in the bacteriocyte. Speculatively, due to this redundancy, it can be predicted that *pheA* (EC 5.4.99.5), a gene that mediates prephenate biosynthesis within *Portiera* will undergo pseudogenization, making *CM* in the bacteriocyte an essential gene in the biosynthesis of phenylalanine.

Other potential targets in the SSA1-SG1 bacteriocyte are genes encoding enzymes that mediate terminal reactions for the biosynthesis of lysine and arginine amino acids. Both *argH* and *lysA* are horizontally transferred genes of bacterial origin. These genes have been subjected to selection pressure (purifying selection) within the bacteriocyte that has rendered them indispensable in the biosynthesis of arginine and lysine respectively. This acquisition of essential genes by the host within the bacteriocyte especially those that mediate terminal reactions enable the host not only to alter metabolite flux but also to control both the synthesis and allocation of essential amino acids between the host and the endosymbiont.

The control of amino acid biosynthesis, allocation of essential amino acid within the bacteriocyte and the less redundancy of genes encoding proteins mediating terminal reactions (Table 4.4) makes these genes ideal targets for the management of cassava whiteflies. Biological systems usually maintain a buffering mechanism against genetic perturbation through redundancy or feedback regulation. In phloem-sap feeders, maintenance of these mechanisms seems to be costly for the host to maintain especially for reactions that are mediated within the bacteriocyte. In *B. tabaci* SSA1-SG1, the terminal reaction of six essential amino acids in the bacteriocytes are mediated by a single indispensable gene (Table 4.4). These terminal reactions were for, (i) arginine synthesis, (ii) lysine synthesis, (iii) valine synthesis, (iv) isoleucine synthesis, (v) leucine synthesis and (vi) histidine synthesis. Of these, the terminal reactions of three branched chain amino acids (Isoleucine, valine and leucine) are mediated by a single gene (*BCAT*, EC 2.6.1.42) (Figure 4.6).

The essentiality and multifunctionality of *BCAT* makes this gene one of the best gene target in the management of the super abundant whitefly population in cassava growing areas. Branched-chain amino acids play vital roles not only in protein synthesis but also in the regulation of energy homeostasis and nutritional metabolism (glucose uptake and fatty acid oxidation) (Nie *et al.*, 2018). Therefore, it can be hypothesized that knockdown of *BCAT* will have a significant effect on the survival of SSA1-SG1. Several studies have demonstrated by using single amino acid deletion technique that lack of even one essential amino acid can have detrimental effects on the insect especially in the larval stage (Zografou *et al.*, 1998; Wilkinson & Douglas, 2003; Chang, 2004).

Another potential target gene is *hisD*, a gene mediating the synthesis of histidine from L-histidinol in the bacteriocyte. Histidine is a very active and versatile amino acid that is often key residue in enzymatic reactions and protein-protein interaction (Uchida, 2003; Liao *et al.*, 2013). Though the model predicted this reaction to be essential, the gene encoding for the enzyme that mediates the synthesis of histidinol dehydrogenase was not detected in both *B. tabaci* transcriptome and *Portiera* genome. More studies are required to identify and isolate the gene that encodes for histidinol dehydrogenase. Down-regulation of this gene either through RNAi will have a detrimental effect on the growth and survival of *B. tabaci* SSA1-SG1.

In conclusion, four genes have been selected as the potential gene targets to disrupt six essential amino acid biosynthesis pathways in the *B. tabaci* SSA1-SG1 bacteriocyte. These include (i) *BCAT* for valine, isoleucine and leucine biosynthesis, (ii) *lysA* for lysine biosynthesis, (iii) *argH* for arginine biosynthesis and (iv) *hisD* for histidine biosynthesis. Factors considered in the selection of these gene include (i) essentiality of the gene, (ii) genes that encode for a protein that mediate terminal reactions and (iii) the terminal reactions that are mediated by the protein encoded by a single gene. It is envisaged that targeting the selected symbiosis genes will not only control the superabundant SSA1-SG1 population but also manage the spread of cassava viral diseases they transmit.

EFFICACY OF RNA INTERFERENCE AGAINST OSMOREGULATION GENES IN THE MANAGEMENT OF *BEMISIA TABACI*

5.1 INTRODUCTION

Despite previous efforts to manage the whitefly pest problems in Sub-Saharan Africa (Section 2.3), they have not been effective and some populations, for example SSA1-SG1, have become so numerous and widespread that they have been termed “super-abundant” (Legg *et al.*, 2014). It is evident, therefore, that new methods especially those based on sequence-specific suppression of gene expression such as RNA interference (RNAi) should be explored to deal with the *B. tabaci* problem in Sub-Saharan Africa (Grover *et al.*, 2019). This technology has been applied in the management of different insect pests in a range of cropping systems, with varying degrees of success (Scott *et al.*, 2013). Key factors that influence the success of RNAi as a pest control strategy depends on the: (i) development of dsRNA constructs, (ii) the efficiency of the dsRNA delivery system, (iii) cellular uptake and amplification of delivered dsRNA and (iv) efficiency of dsRNA processing to siRNA.

In *B. tabaci*, several studies provide proof of concept data that demonstrate the efficacy of RNAi in controlling whiteflies (Ghanim *et al.*, 2007; Upadhyay *et al.*, 2011; Thakur *et al.*, 2014; Chen *et al.*, 2015; Malik *et al.*, 2016). These studies varied in the genes targeted as well as the dsRNA delivery method. For example, Ghanim *et al.* (2007) used injection as a dsRNA delivery method, which demonstrated that RNAi was effective on *B. tabaci* with 70% gene knockdown. However, the method is not practical in the management of whitefly populations in farmers’ fields. Thakur *et al.* (2014) and Upadhyay *et al.* (2011) demonstrated direct feeding of dsRNA as an efficient alternative method. The high success reported in these studies can be attributed to the nature of the insect midgut. The midgut of most insects is made of a perimicrovillar membrane, making it one of the sensitive portions of the insect body with an active interface in contact with the physical environment. The perimicrovillar membrane is made up of microvilli, stem cells and endocrine cells that provide channel and endocytosis apparatus for dsRNA uptake (Hakim *et al.*, 2010; Nardi *et al.*, 2019). Important processes, for example, digestion and osmoregulation, take place within the gut. Therefore, to increase the efficiency of RNAi in *B. tabaci*, essential genes with enriched expression within the gut are promising targets to select. This is because the reduction of gene

expression of these genes is less dependent on the systemic spread of dsRNA in the insect body (Scott *et al.*, 2013; Yu *et al.*, 2013)

Management of whiteflies either in farmers' fields or in the glasshouse requires an efficient dsRNA delivery system. Therefore, the efficacy of RNAi must be evaluated using *in planta* delivery systems. Tzin *et al.* (2015) used an *in planta* delivery system based on transient expression of dsRNA to evaluate the efficacy of RNAi against osmoregulation genes in phloem-sap feeders. More studies are required to explore the use of stably transformed plants in the control of phloem-sap feeders. In this study, stably transformed transgenic plant, with phloem-specific promoters to facilitate the expression of dsRNA in phloem tissues were used to study the efficacy of RNAi in the control of whiteflies.

The effectiveness of dsRNA-mediated gene silencing can also be influenced by: (i) concentration of dsRNA (Miller *et al.*, 2012), (ii) mode and frequency of administration and (ii) the target gene selected (Ghanim *et al.*, 2007; Wang *et al.*, 2016; Luo *et al.*, 2017; Mamta & Rajam, 2017). Of these, gene selection is the most important factor. An ideal gene target should: (i) produce mRNA pools with low turnover (Larsson *et al.*, 2010) and (ii) be consistently expressed throughout the relevant life stages. Examples of genes-targeted in controlling insects using RNAi include, (i) aquaporins in aphids (Shakesby *et al.*, 2009), (ii) immunoresponsive midgut-expressed gene (*TsetseEP*) in tsetse fly (Walshe *et al.*, 2009), (iii) digestive and caste-regulatory genes in *Reticulitermes flavipes* (Zhou *et al.*, 2008), (iv) molting genes in *Tribolium castaneum* (Konopova & Jindra, 2007), (v) olfactory receptors gene (*orco*) in *Apolygus locorum* (Zhou *et al.*, 2014) and (vi) *CYP6AE14* in *Helicoverpa armigera* (Mao *et al.*, 2011). In *B. tabaci*, two candidate osmoregulation genes; *Aquaporin 1 (AQPI)*, a major intrinsic protein and *Sucrase 1 (SUC1)* an alpha glucoside (family 13) have been identified in MEAM1 (Jing *et al.*, 2016). These genes are required to maintain fluid homeostasis (Mathew *et al.*, 2011; Jing *et al.*, 2016; Raza *et al.*, 2016). In this study, the efficacy of RNAi against *AQPI* and *SUC1*, genes with different molecular but linked physiological function was evaluated. Tomato plants (*Solanum lycopersicum* L) stably transformed with dsRNA against these genes were used to determine both the effect of these transgenic plants on survival and physiology of whiteflies.

5.2 MATERIALS AND METHODS

5.2.1 Plant material and insects

The *B. tabaci* species, Middle East-Asia Minor (MEAM1) (GenBank accession no KM507785) was collected originally from Ithaca, New York. The colony had been maintained at Cornell University on tomato (*Solanum lycopersicum* cv Florida Lanai) at $25\pm 2^{\circ}\text{C}$ with 12:12h L:D at $350\ \mu\text{mol m}^{-2}\text{s}^{-1}$ PAR. Six types of stably transformed tomato plant (cv Florida Lanai) generated by Luo *et.al*, (2017) were obtained from the Douglas laboratory, Department of Entomology, Cornell University. These plants had the following transformations: (i) dsRNA against RNase (dsRNase1 & 2) (NCBI Accession KX390872 and KX390873 respectively), (ii) dsRNA against GFP (dsGFP-dsRNA), (iii) dsRNA against sucrose (dsSUC) (NCBI Accession KX390871), (iv) dsRNA against aquaporin (NCBI Accession KX390870) and sucrose (dsAQP+dsSUC), (v) dsRNA against RNase 1 & 2, aquaporin and sucrose (dsRNase+dsAQP+dsSUC) and (vi) empty vector.

Due to the limited number of transformed tomato plants per line (T_1), tomato seeds from different stably transformed plants were generated and treated with Rapidase (12ml/L of water) for 1.5-2 hours. The treated seeds were established to generate seedlings from each line (second generation plants, T_2) for the experiment. Seedlings were tested for the presence of target dsRNAs to ensure that only stably transformed plants were used. The plants were grown on compost (Cornell mix) supplemented with a water-soluble fertilizer and maintained in a climate-controlled greenhouse at $25\pm 2^{\circ}\text{C}$, 60% relative humidity and a 12:12h L:D for 6 weeks.

5.2.2 Evaluating *B. tabaci* performance on tomato transgenic plants

After 6 weeks, the transgenic tomato plants were transferred to the climate-controlled chamber ($25\pm 2^{\circ}\text{C}$, 60% relative humidity and a 12:12h L:D with a light intensity of $350\ \mu\text{mol m}^{-2}\text{s}^{-1}$ PAR). The experiment was set up as a non-replicated single row, with each row consisting of five plants per transgenic line and two clip cages were attached to each plant. A tomato leaf containing only 3rd instar nymphs and pupae was placed in a cage containing a clean tomato plant, at two days before the experimental setup to ensure that only newly emerged whiteflies were used. Wet cotton wool was placed on the leaf stalk to avoid quick desiccation. After two days, all the emerged whiteflies had settled on the leaves and were acclimatized. Only whiteflies in courtship (pairs)

were collected in glass vials. This was done to minimize stress or strain on whiteflies prior to the experiment. A total of 10 newly emerged whiteflies (5 males :5 females) were transferred into each cage. Data on insect survival was taken daily for six days.



Figure 5.1: Experimental setup of the efficacy study consisting of a single row non-replicated plot with five tomato plants for each treatment, established in the climate-controlled growth chamber at Cornell University, USA.

5.2.3 Analysis of *SUC1* and *AQP1* gene expression in MEAM1 *B. tabaci*

The effect of RNAi on the expression of both *SUC1* and *AQP1* genes was analyzed using quantitative real-time PCR (qRT-PCR). Total RNAs were extracted from MEAM 1 *B. tabaci* that survived after six days of feeding on the transgenic plants as described in Section 5.2.2. Total RNA for all samples were normalized to ensure an equal amount of starting RNA template (100 ng). Normalized RNAs were then treated with 1 μ l of the ezDNaseTM enzyme (InvitrogenTM) for 2 minutes at 37°C to remove genomic DNA. The RNA samples were further incubated at 55°C for 5 minutes in the presence of 10 mM DTT to inactivate the enzyme. First-strand cDNA synthesis was carried out using SuperScriptTM II Reverse Transcriptase (InvitrogenTM) according to manufacturer's instructions. A three-step thermocycling protocol was used with polymerase

activation and DNA denaturation at 95°C for 3 minutes followed by 40 cycles of both denaturation at 95°C and annealing/extension at 55.2°C for 45 seconds, performed on a C1000 Thermo cycler with CFX96 Real-time detection system (Bio-Rad), with 60S ribosomal protein L13a (*RPL13*) as the internal control. A total of 20 µl reaction mix containing 10 µl iQ™ SYBR® Green Supermix (2x), 0.5µl of forward and reverse primer, 1 µl of cDNA and 8µl of water was used. Melting curve analysis at the end of the reaction was done by increasing temperature from 55-95°C in increments of 0.5°C every 5 seconds to assess the dissociation characteristics of dsDNA during heating.

5.2.4 Honeydew collection

The honeydew analysis was done to determine the physiological impact of RNAi on sugar transformation within the *B. tabaci* gut. Honeydew was collected from whiteflies feeding on transgenic plants, wild-type tomato, *Arabidopsis thaliana* (L.) Heynth cv Columbia and artificial diet. An artificial diet with 0.75 M sucrose was made following a modified protocol by Prosser & Douglas. (1992) and Douglas *et al* (2001). A total of 200 whiteflies were placed in each cage and honeydew was collected after 48 h of feeding. Each cage was placed at the same level (fourth branch) of the plant. Honeydew was collected on aluminum foil placed at the base of the cage. Before the collection, the aluminum foil was washed with 80% methanol to remove any contaminants before being placed in an oven at 60°C for 12 h. The weight of the aluminium foil (X_0) was taken before it was used in the cage (X_0) using an ultrabalance (Toledo MX5). The weight of honeydew (y) was assessed by determining the difference in the weight of aluminum foil (X_0) and the weight of aluminium foil after *B. tabaci* depositing honeydew (X_1) (Equation 1). Honeydew was then washed off the aluminum foil using 300-500 µl of deionized water. Aluminium foil was dried and weighed again (X_2) to determine the amount of honeydew washed off the aluminum foil (Equation 2). Honeydew samples were then stored at -20°C for sugar analysis. A negative control was included (foil from an empty clip cage) to correct for weight due to any other factors like moisture.

Amount of honeydew collected	$Y_1 = X_1 - X_0$	Equation 1
------------------------------	-------------------	------------

Amount of honeydew stored for analysis	$Y_2 = X_1 - X_2$	Equation 2
--	-------------------	------------

5.2.5 Analysis of sugars in honeydew samples from whiteflies

5.2.5.1 Sucrose assays

The amount of sucrose in the honeydew sample signified the activity of *Sucrase* (net hydrolysis of sucrose) that ensures osmotic balance within *B. tabaci*. Therefore, in this study, honeydew samples were analyzed to quantify the amount of sucrose in honeydew produced by *B. tabaci* feeding on different stably transformed tomato plants. Sucrose was hydrolyzed *in vitro* to its constituent monosaccharides, α -glucose and β -fructose by β -fructofuranosidase (*Invertase*, *Sucrase*, EC 3.2.1.26). Each sample was divided into two aliquots. One of the aliquots was treated with *Invertase* while the other was not. In principle, if the honeydew sample contains sucrose, the amount of glucose in the aliquoted sample treated with *Invertase* should be higher than the aliquoted sample that has not been treated with *Invertase*. A total of 20 μ l honeydew was hydrolyzed to completion using 4 μ l (0.4 U) *Invertase* (Sigma I-4504) in 50 mmol l⁻¹ sodium acetate buffer, pH 4.5 at 37°C for 30 min and the glucose produced was determined by the glucose method described below

5.2.5.2 Glucose assay

Glucose assays were conducted using Sigma Glucose Oxidase kit (GAGO20). In brief, the kit contains a glucose oxidase assay reagent which is a mixture of glucose oxidase reagent and O-dianisidine reagent and a glucose assay standard. Glucose standards were made according to Table 5.1. The aliquoted honeydew samples were normalized by adding 4 μ l water to 20 μ l honeydew. Then 5 μ l of the normalized sample, glucose standard and honeydew sample treated with *Invertase* were added to microtiter, flat-bottom 96 well plate (Laboratory Product Sales Inc, Rochester, New York) in triplicates for each sample. This was followed by adding 150 μ l assay reagent to each sample. The samples were then incubated at 37°C for 30 minutes. After incubation, 150 μ l of dilute sulphuric acid (12 N H₂SO₄) was added to each sample and then absorbance was read at 540 nm using a Bio-RAD xMark™ microplate spectrophotometer. A regression equation from the glucose standard curve was used to compute the content of glucose in each sample (Figure 5.2). The amount of sucrose in each sample was calculated using the formula

$$Y_s = X_4 - X_3 \dots\dots\dots \text{Equation 3}$$

where Y_s is the amount of sucrose in the honeydew sample, X_4 is the amount of glucose in honeydew samples treated with *Invertase* and X_3 is the amount of free glucose in samples that were not treated with *Invertase*. The sucrose hydrolysis rate was then computed using the formula;

$$\text{Sucrose hydrolysis} = \frac{\text{Sucrose in diet} - \text{Sucrose in honeydew}}{\text{Sucrose in diet}} \dots\dots\dots 4$$

Analysis of variance on the amount of sucrose in honeydew samples collected from different transgenic lines and sucrose hydrolysis rate was used to determine the effect of dsRNAs on the sugar transformation within the *B. tabaci* gut.

Table 5. 1: Standards required for glucose assay

µg glucose in 5 µl sample	Glucose stock* (µl)	Buffer (µl)
1.0	200	800
0.9	180	820
0.8	160	840
0.7	140	860
0.6	120	880
0.5	100	900
0.4	80	920
0.3	60	940
0.2	40	960
0.1	20	980

*Glucose stock concentration – 1.0 mg/ml of glucose in 0.1 benzoic acid

5.2.6 Quantification of higher sugars in honeydew samples

Each honeydew sample collected from different transgenic plants, tobacco plants and artificial diet was divided into aliquots. One of the aliquots was used to measure sucrose content using enzymatic methods (*Invertase*) while the other aliquot was sent to the Max Planck Institute for Chemical Ecology (Jena, Germany) for analysis of disaccharides and higher sugars. Disaccharides and higher sugars were analyzed using liquid chromatography-mass spectrometry. Both percentage molar ratio and sugar concentration for different sugar categories in each honeydew samples were computed. Sugar category analyses included were fructose, glucose, disaccharides, trisaccharides, tetrasaccharides and pentasaccharides. Sugar concentration was computed by calculating the hexose units (nmol/mg) for all the sugar categories measured for each milligram of honeydew.

Assimilation efficiencies of both whiteflies and aphids feeding on an artificial diet with known sucrose concentration (0.75M) were then computed using the formula:

$$\text{Assimilation efficiency} = \left(\frac{\text{Sugar concentration in diet} - \text{Sugar concentration in honeydew}}{\text{Sugar concentration in diet}} \right) \times 100$$

5.2.7 Determination of osmotic pressure of *B. tabaci* hemolymph

Haemolymph osmotic pressure of *B. tabaci* feeding on different transgenic tomato plants was measured to assess the physiological consequence of RNAi against the osmoregulation genes (aquaporin and sucrase). Third instar nymphs collected from different transgenic tomato plants (SUC1, dsAQP+SUC1, dsRNase+dsAQP+dsSUC, Empty vector, Wild-type and dsRNase) were used to measure the impact of RNAi on the osmotic pressure in whiteflies. This was done using a freezing point depression method, measured on the Otago™ nanoliter osmometer. Haemolymph was collected from individual third instar nymphs. Each nymph was placed in water saturated non-drying immersion oil (Cargille Laboratories, New Jersey, USA) on a microscope glass slide. Each nymph was pierced in-between the eyes using a very fine dissection pin to obtain the haemolymph. The capillary tube containing non-drying immersion oil A (Cargille laboratories, New Jersey, USA) was used to transfer collected haemolymph to non-drying immersion oil B (Cargille Laboratories, New Jersey, USA) on the cooling block of the nanoliter osmometer.

The sample in non-drying immersion oil B on the cooling block was frozen to -20°C then the melting temperature of the sample was noted. A standard curve was done using 100 mOSm/kg, 500 mOSm/kg, 900 mOSm/kg, 1500 mOSm/kg and 2000 mOSm/kg osmolarity linear set (Advanced Instrument Inc, Norwood Massachusetts, USA). A linear equation from the standard curve was used to determine the haemolymph osmotic pressure for each sample (Figure 5.11). Analysis of variance was used to determine whether the dsRNAs in each test plant had a significant impact on the osmotic pressure of *B. tabaci*.

5.3 RESULTS

5.3.1 Sucrose hydrolysis in whiteflies feeding on different sources of sucrose

Before evaluating the efficacy of RNAi on the survival of whiteflies, preliminary studies were conducted to understand sucrose hydrolysis within whiteflies feeding on different sources of sucrose, the role of *SUC1* gene in osmoregulation and the suitability of tomato plants (*Solanum lycopersicum*) as source for *B. tabaci* (MEAM1). Honey dew was collected from whiteflies feeding on four sugar sources. These were: (i) tomato (*Solanum lycopersicum* cv. Florida Lanai), (ii) tobacco (*Nicotiana tabacum* L), (iii) *Arabidopsis thaliana* and (iv) artificial diet with known sucrose concentration (0.75 M). Honeydew samples from whiteflies feeding on arabidopsis and tomato had significantly higher sucrose content (0.001 M and 0.001 M, respectively) compared to the honeydew samples from whiteflies feeding on tobacco and artificial diet (0.5 mM and 0.3 mM respectively) (Figure 5.3). Though honeydew samples from whiteflies feeding on arabidopsis had significantly higher sucrose, it had significantly lower free glucose compared to honeydew samples from whiteflies feeding on tomato, tobacco and artificial diet. Honeydew samples from whiteflies feeding on tomato, tobacco and artificial diet had a relatively similar amount of free glucose, that was not statistically different from each other.

Sugar analysis from honeydew samples of whiteflies feeding on an artificial diet with known sucrose concentration (0.75 M) revealed that whiteflies hydrolyzed 97% ingested sucrose (Equation 4) to monosaccharides; fructose and glucose. Differences therefore in sucrose hydrolysis in whiteflies feeding on different sucrose sources are attributed to the sucrose sources (tobacco, tomato, *Arabidopsis* and artificial diet) rather than the insect physiology.

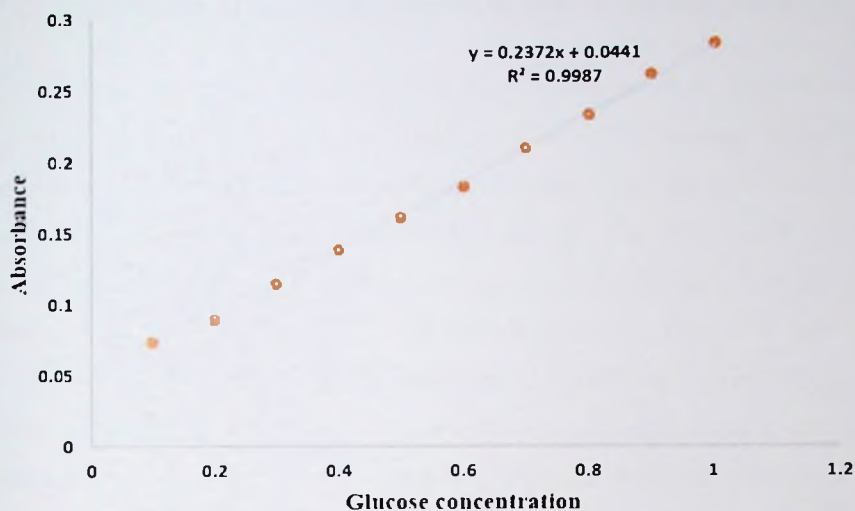


Figure 5.2: Glucose standard curve used to calculate the amount of glucose in different honeydew samples

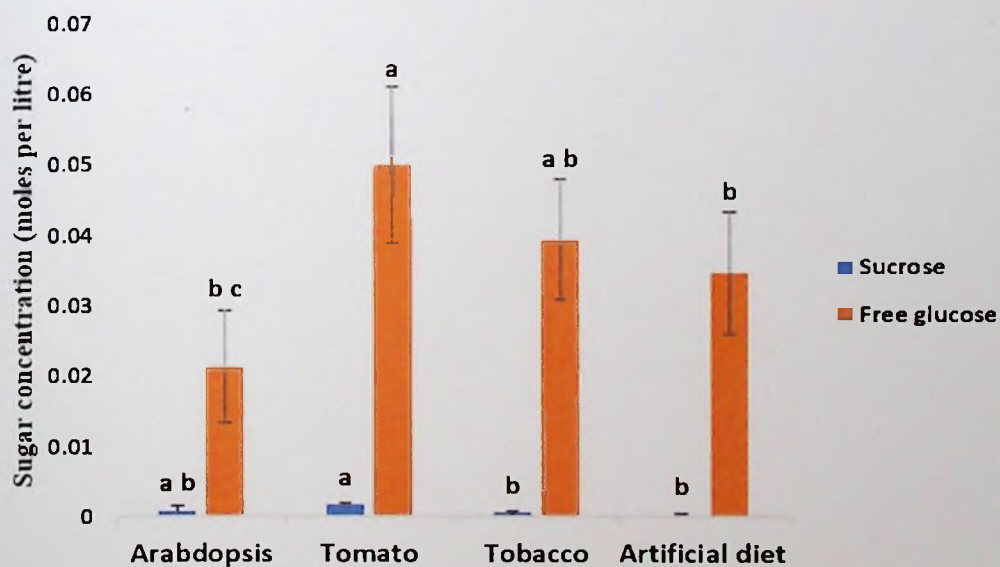


Figure 5.3: Sucrose hydrolysis in whiteflies feeding on different sucrose sources. Amount of sucrose and free glucose in honeydew collected from whiteflies feeding on Arabidopsis, tomato, tobacco and artificial diet with 0.75 M sucrose. Treatments sharing the same compact letter display are not significantly different at 5% significance level.

5.3.2 Comparison of sucrose hydrolysis in aphids and whiteflies

Honeydew samples from aphids (in this case, *Myzus persicae* Sulzer) and whiteflies (*B. tabaci* MEAM1) feeding on tobacco and artificial diet (0.75 M) were analyzed to compare sucrose hydrolysis in aphids and whiteflies. Tobacco plants were used in preference to tomato because tomato plants were not a suitable host for *Myzus persicae*. There was no significant difference in the amount of sucrose in the honeydew samples from both aphids and whiteflies feeding on tobacco and artificial diet. Nearly all the sucrose (0.74 M out of 0.75 M) was hydrolyzed by both the aphids and whiteflies. Sugar hydrolysis of honeydew samples from an artificial diet with known sucrose concentration (0.75 M) revealed that free glucose in honeydew from whiteflies was significantly higher than that from aphids, showing that the differences in sugar assimilation and isomerization is attributed to insect physiology rather than the sugar sources (Figure 5.4)

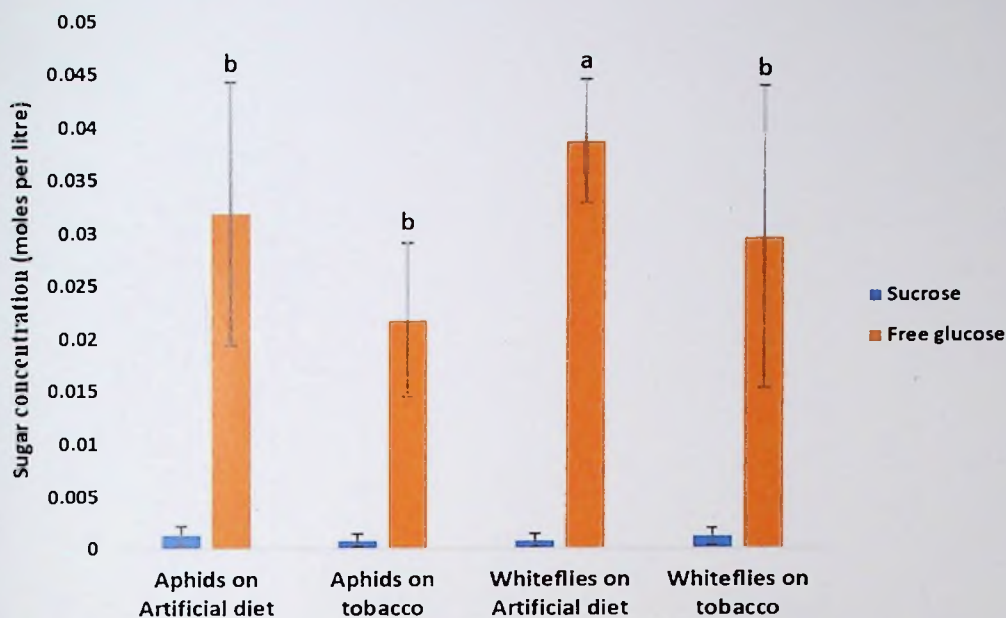


Figure 5.4: Comparison of sucrose hydrolysis in the aphid *Aphis gossypii* and the whitefly *B. tabaci* MEAM1 feeding on both tobacco and artificial diet with 0.75 M sucrose. Error bars represent standard deviation. Treatments sharing the same compact letter display are not significantly different at 5% significance level.

5.3.3 *B. tabaci* performance on different transgenic tomato lines

The efficacy of RNAi in the control of *B. tabaci* MEAM1 population on tomato plants was determined by evaluating the survival of *B. tabaci* on stably transformed tomato plants during a span of six days. Tomato plants were stably transformed to down-regulate the expression of either or both *AQP1* and *SUC1* genes to disrupt osmoregulation within *B. tabaci*, thereby affecting their survival. Analysis of mortality among the whiteflies feeding on different transgenic tomato using generalized linear model (negative binomial) revealed that there was no significant difference ($P_{0.05} = 0.27$) in the percentage mortality recorded in whiteflies feeding on all the tested tomato transgenic plants. A low mortality ranging from 2-12% was observed for all the tested transgenic tomato lines with the highest mortality on plants with dsGFP construct (12%) followed by plants with dsSUC construct (10%) and dsRNase+dsAQP+dsSUC construct (7%) (Figure 5.5).

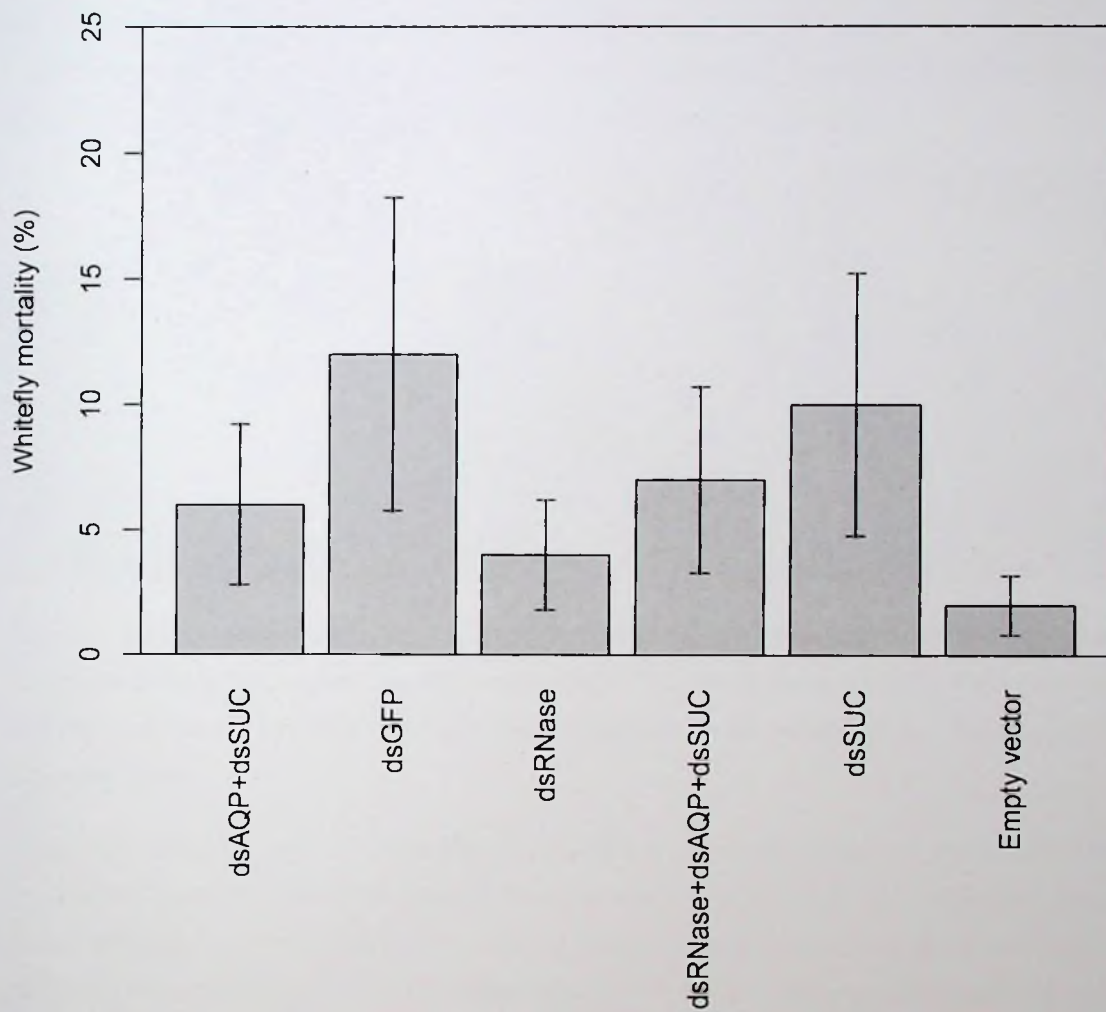


Figure 5.5: Percentage mortality of whiteflies feeding on different tomato plants with dsRNA against *RNase*, *GFP*, *AQP1* and *SUC1* in the first experiment.

Based on these results, the experiment was re-established with slight modification in the experimental methods (Section 5.2.2) to validate the results obtained in the first experiment. Modification in the experimental methods included,

- (i) use of new clip cages that allow in enough light (Figure 5.6B). This was based on the fact that light intensity influences expression of genes in plants (Petrillo *et al.*, 2014).
- (ii) whiteflies were also immobilized using carbon dioxide before transferring them into the clip cages. This modified methodology is henceforth referred to as the second method.

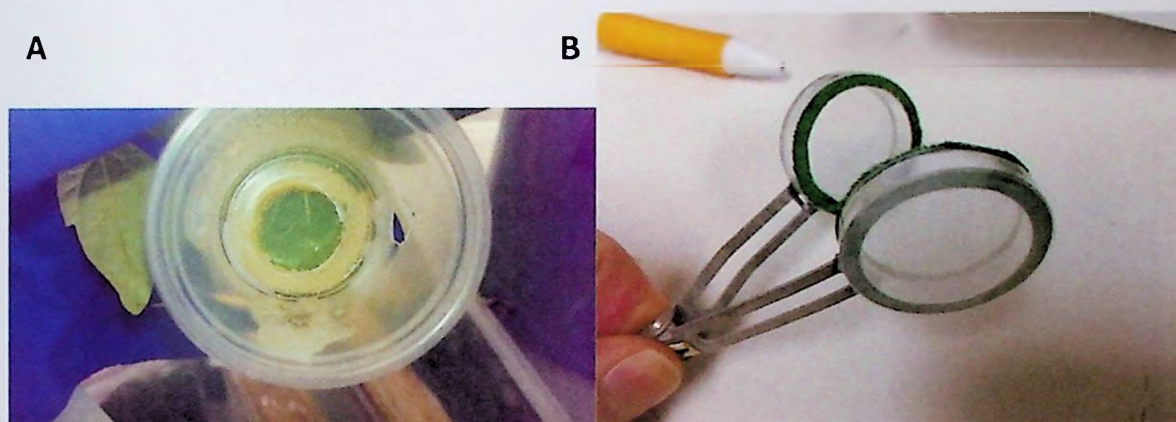


Figure 5.6: Clip cages used in the experiment A) clip cage in the first experimental setup, b) clip cages used in the second experimental setup. Clip B allows in light on both sides of the clip cage, to ensure that the cross-section of the leaf where whiteflies are feeding receives adequate light.

Percentage mortality of the whiteflies feeding on the tested tomato transgenic plants in the second experiment ranged from 24-53%. Analysis of variance of whitefly mortality on different transgenic tomato plants revealed a statistically significant effect in whitefly mortality in whiteflies feeding on tomato plants with dsAQP+dsSUC construct ($P_{0.05} = 0.007$), with percentage mortality of 53%, based on generalized linear model (negative binomial) test. This performance was almost twice that recorded on tomato plants with dsSUC and dsRNase+dsAQP+dsSUC construct (Figure 5.7).

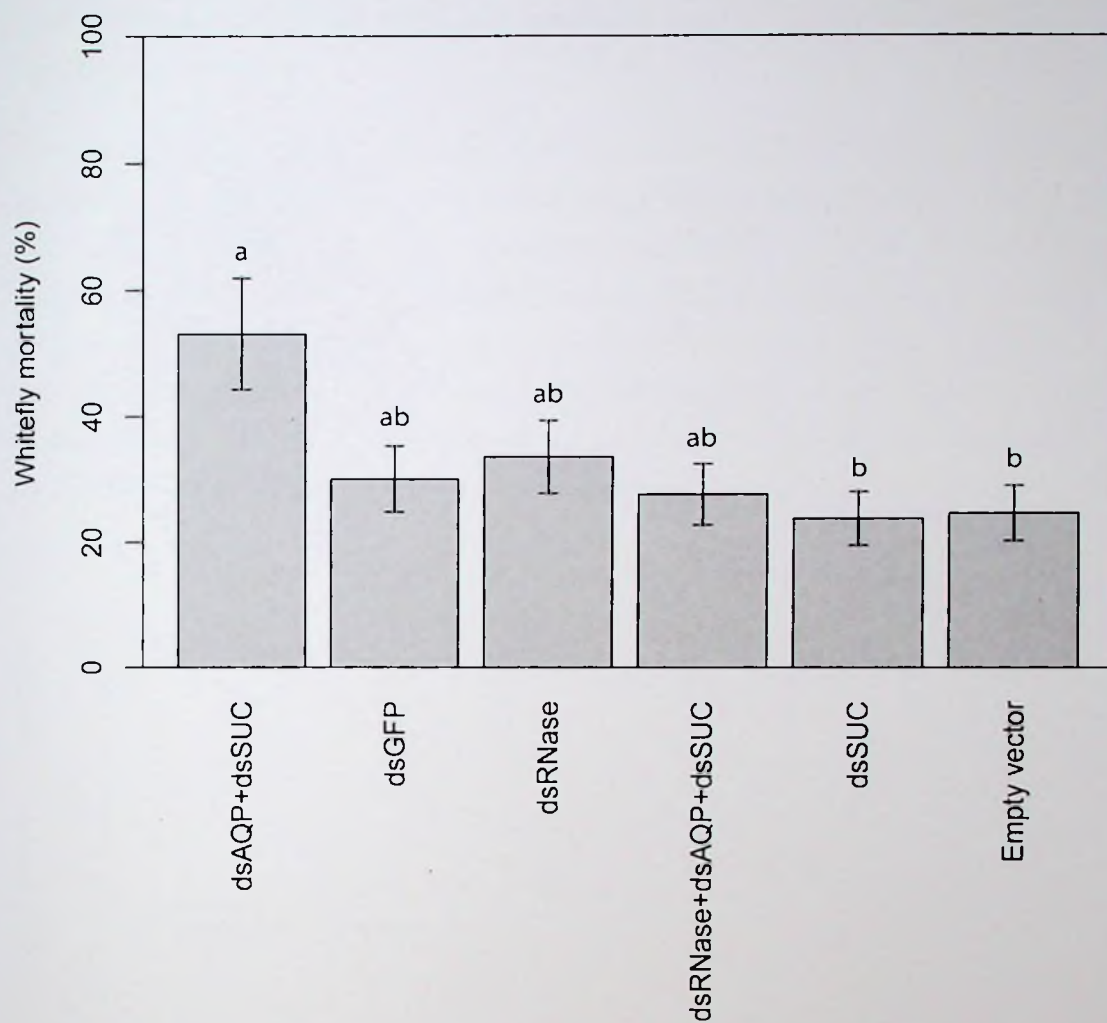


Figure 5.7: Percentage mortality of whiteflies *B. tabaci* MEAM1 feeding on different transgenic tomato plants with dsRNA against *RNase*, *GFP*, *AQP1* and *SUC1* after six days in the second experiment. Both plants with empty vector construct and wildtype tomato were used as controls. Treatments sharing the same compact letter display are not significantly different at 5% significance level.

5.3.4 Analysis of *SUC1* and *AQP1* gene expression in *B. tabaci* feeding on different transgenic tomato plants

Expression of both *AQP1* and *SUC1* gene in the surviving whiteflies after six days of feeding was quantified using RT-qPCR. In the first experiment, transgenic lines transformed to knockdown *SUC1* gene (dsSUC, dsAQP+dsSUC and dsRNase+dsAQP+dsSUC) had a minimal effect on the expression of *SUC1* gene in whiteflies feeding on these tomato plants, with incomplete knockdown not exceeding 22%. Downregulation of *SUC1* was observed in whiteflies feeding on tomato transgenic plants containing dsSUC construct and dsAQP+dsSUC construct (Figure 5.8). Tomato plants containing dsAQP+dsSUC construct were able to downregulate the expression of both *AQP1* and *SUC1* expression by 0.15 ± 0.11 fold (15%) and 0.22 ± 0.10 fold (22%), respectively.

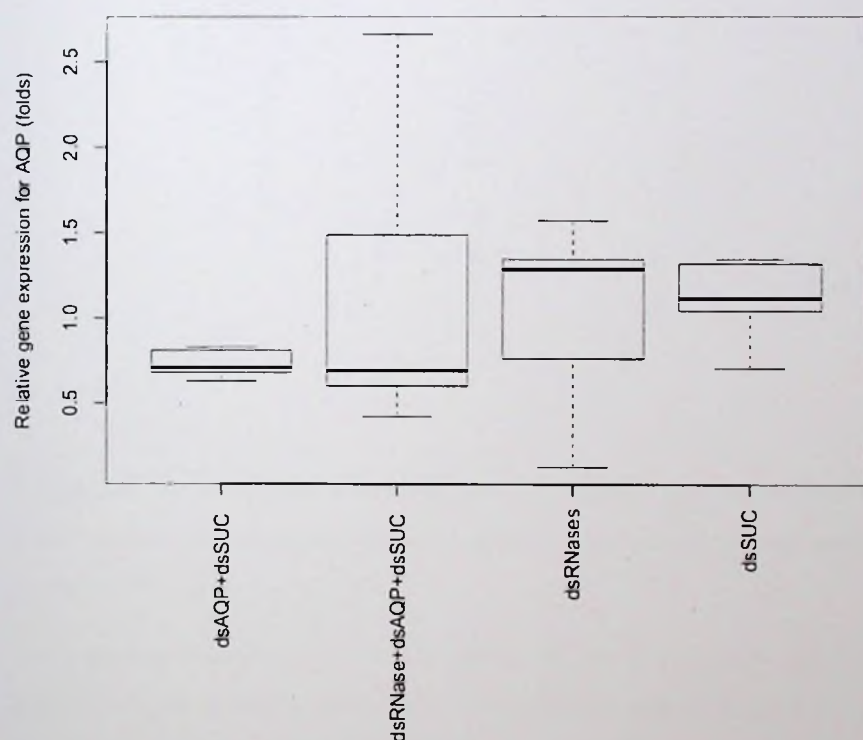


Figure 5.8: Relative gene expression of *AQP1* genes in *B. tabaci* MEAM1 that survived after six days of feeding on different stably transformed tomato plants in the first experiment.

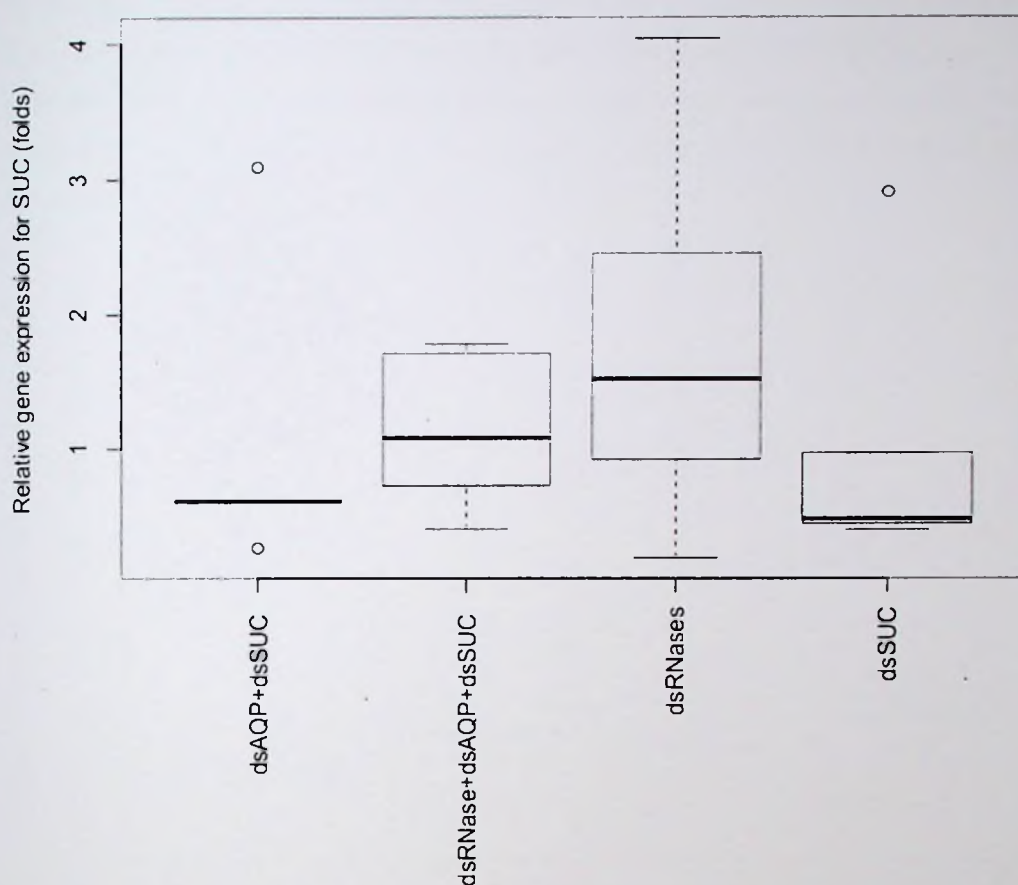


Figure 5.9: Relative gene expression of *SUC 1* genes in *B. tabaci* MEAM1 that survived after six days of feeding on different stably transformed tomato plants in the first experiment.

In the second experiment, there was no significant difference in the gene expression of *SUC1* in whiteflies feeding on tomato plants with dsRNase construct, dsSUC construct, dsAQP+dsSUC construct and dsRNase+dsAQP+dsSUC construct. However, plants with dsAQP+dsSUC construct had the lowest gene expression of *SUC1* of 0.08-fold compared to 0.04 fold and 1.22 fold in whiteflies feeding on tomato plants with dsSUC construct and dsRNase+dsAQP+dsSUC construct, respectively. Furthermore, no significant effect was observed in *AQP1* gene expression

in whiteflies feeding on the tested transgenic tomato line (Empty vector, dsRNase construct, dsSUC construct, dsAQP+dsSUC construct and dsRNase+dsAQP+dsSUC construct). However, whiteflies feeding on plants with the dsAQP+dsSUC construct had a reduction in AQP1 expression (0.09 ± 0.28 fold) compared to an increase in the gene expression of 1.56 ± 0.48 fold (56%) recorded on the empty vector. (Figure 5.9).

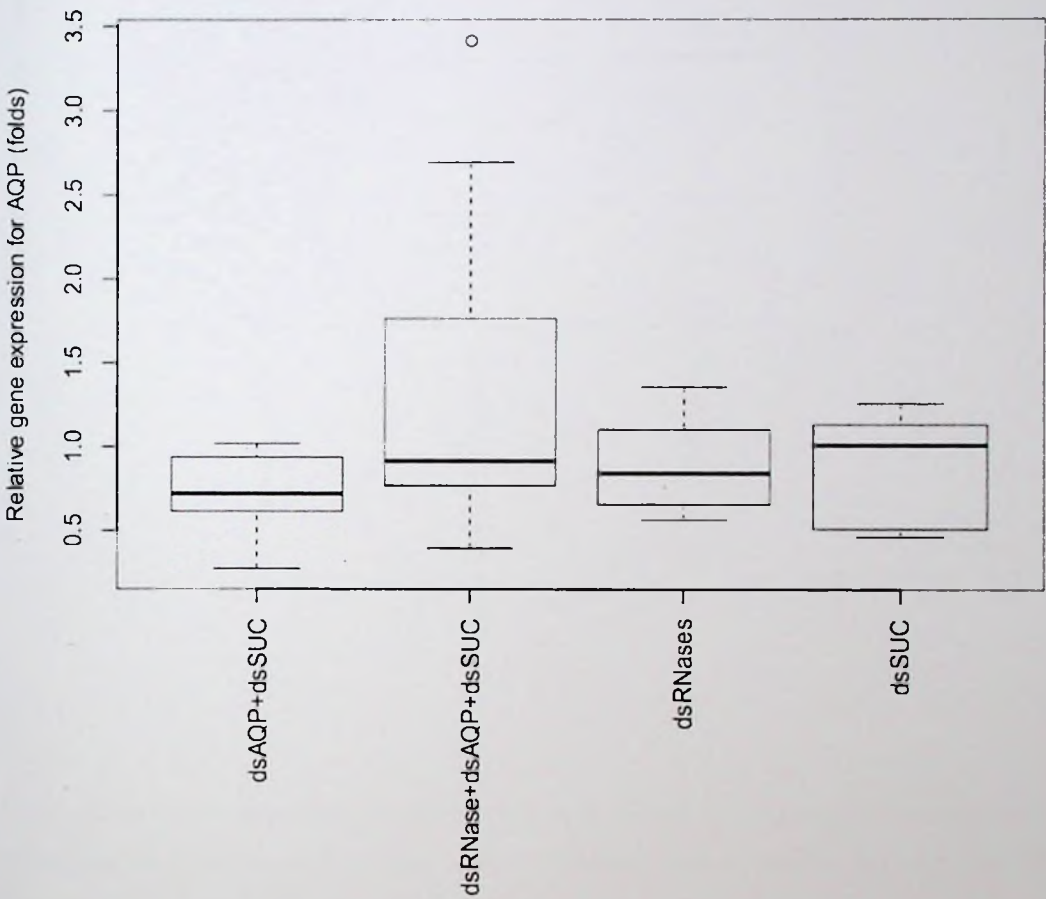


Figure 5.10: Gene expression of *AQP1* genes in *B. tabaci* MEAM1 that survived after six days of feeding on different stably transformed tomato plants in the second experiment.

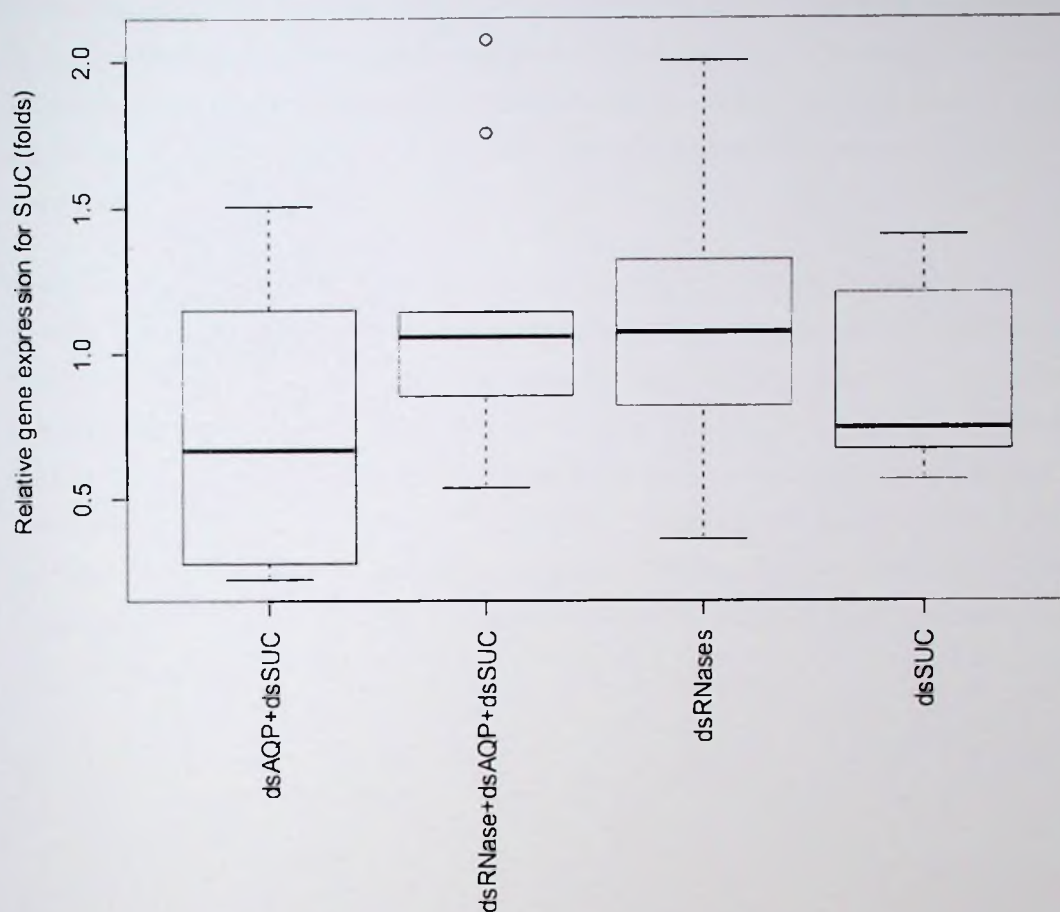


Figure 5.11: Gene expression of SUC1 genes in *B. tabaci* MEAM1 that survived after six days of feeding on different stably transformed tomato plants in the second experiment. Plants with empty vector construct were used as control

5.3.5 Effect of RNA interference on *B. tabaci* development

In the first experiment, adult whiteflies feeding on different transgenic tomato plants did not exhibit significant effects of RNAi within the six days as hypothesized. Crawlers and nymphs arising from these whiteflies, therefore, were monitored to determine whether RNAi was effective

on the next generation. Crawlers (first instar nymphs of whiteflies) emerging from eggs deposited by the five female whiteflies in each cage were counted. These crawlers were monitored through second, third and fourth instar stages. The number of fourth instar nymphs was compared to the crawlers to determine the survival rate, which gave an inference on the effect of RNAi on the development of whiteflies.

Analysis of variance revealed that there was no significant effect on the number of fourth instar nymphs compared to the crawlers on each test transgenic line. The number of crawlers ranged from 108.2–138.6 with the highest number of crawlers recorded on plants with dsRNase+dsAQP+dsSUC construct (138.6) followed by those with dsAQP+dsSUC construct (133.8). The transgenic plants with the lowest number of crawlers also had the highest level of instar mortality between crawler and 4th instar stage. Plants with the dsGFP-dsRNA construct had the highest instar mortality (32.8%) followed by plants with dsRNase construct (26.4%) and empty vector (20.2%). On average, 70% of the emerged whiteflies on each transgenic line developed to the adult stage (Figure 5.10).

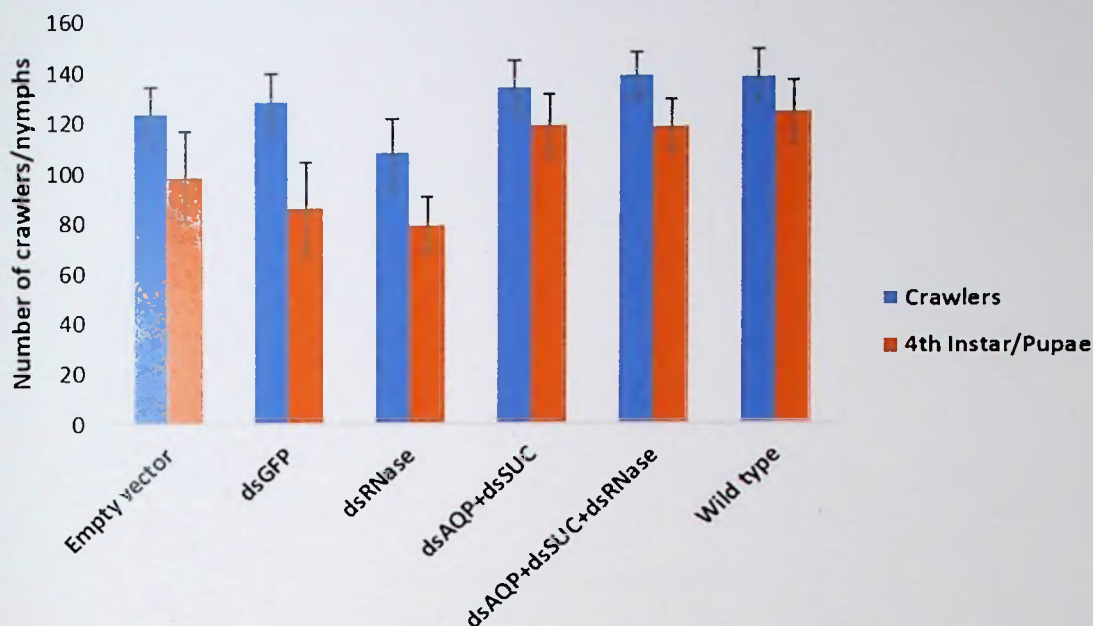


Figure 5.12: Effect of RNA interference on the development of *B. tabaci* MEAM1 feeding on different transgenic tomato plants in the first experiment. One-way ANOVA revealed that there was no significant difference between the number of 3rd instar nymphs and crawlers on different tomato transgenic plants.

5.3.6 Effect of RNAi against *SUC1* and *AQP1* genes on *B. tabaci* osmotic pressure

Knock down of osmoregulation gene(s) in phloem-sap feeding insects is expected to increase osmotic pressure within the insect haemolymph. Therefore, haemolymph samples from third instar nymphs feeding on different transgenic tomato lines, empty vector and wild-type tomato plants in the first experiment were analyzed. Results revealed that there was no significant difference ($P_{0.05} = 1$) in haemolymph osmotic pressure of nymphs feeding on all transgenic tomato lines compared to nymphs feeding on empty vector and wild-type tomato plants. Haemolymph osmotic pressure from nymphs feeding on all tested transgenic tomato lines ranged from 1.322 – 1.375 MPa. Nymphs feeding on transgenic plants containing double-stranded RNA against *SUC1* and *AQP1* genes (dsAQP+dsSUC and dsRNase+dsAQP+dsSUC) had relatively higher osmotic pressure (1.375 MPa and 1.373 MPa respectively) compared to nymphs feeding on other transgenic lines, empty vector and wild-type tomato plants (Figure 5.12)

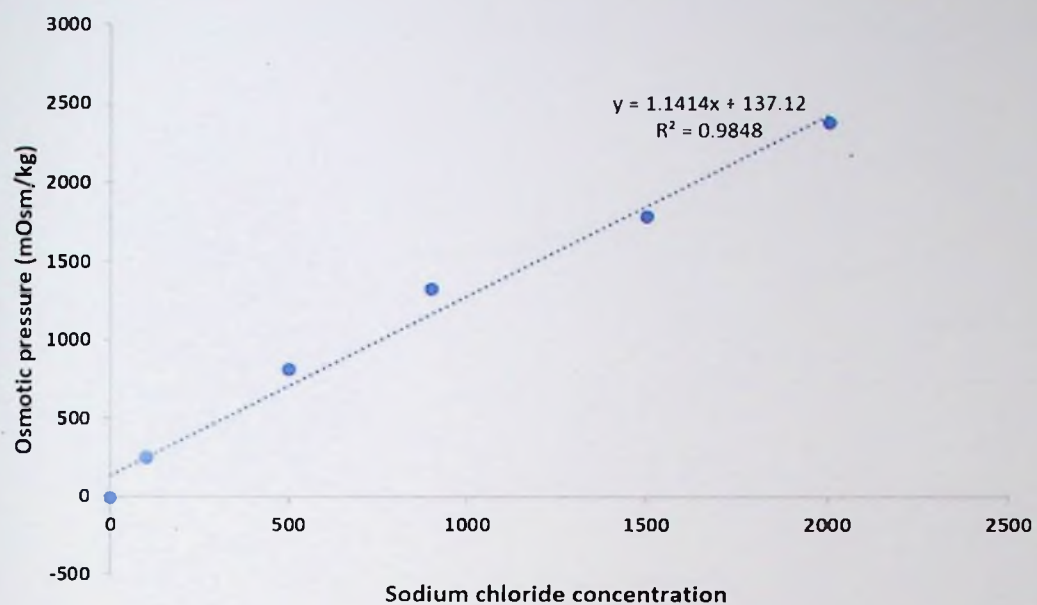


Figure 5.13: Standard curve used to estimate haemolymph osmotic pressure for each analyzed *B. tabaci* sample

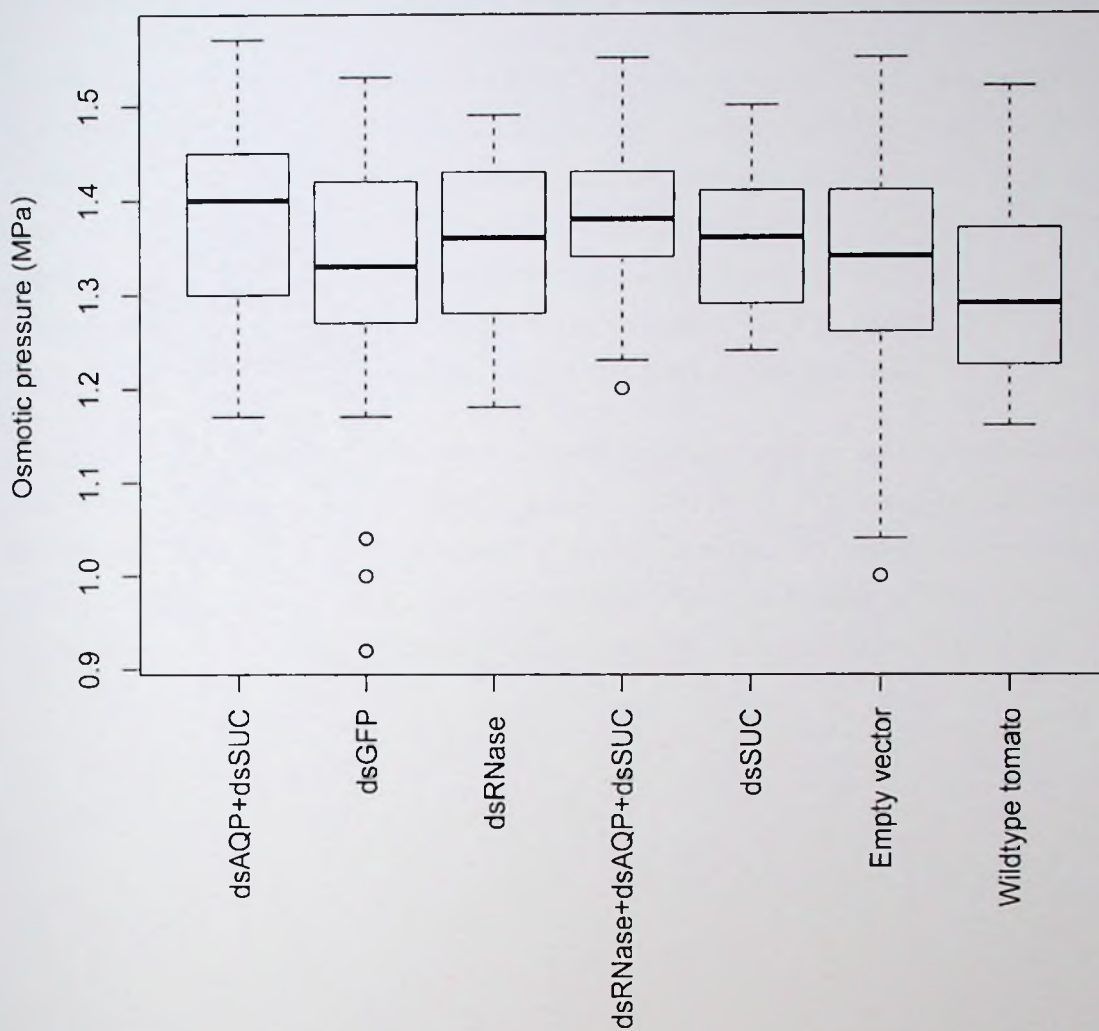


Figure 5.14: Haemolymph osmotic pressure for *B. tabaci* feeding on different stably transformed tomato plants. Osmotic pressure in this study was used as a physiological index of *B. tabaci* osmoregulation to study the effect of feeding whiteflies on different stably transformed tomato plants with different dsRNA against osmoregulation genes in the *B. tabaci*.

5.3.7 Sucrose hydrolysis in whiteflies feeding on different transgenic tomato lines

The two preliminary studies (Section 5.3.1 and 5.3.2) showed a general trend in sucrose hydrolysis in *B. tabaci*. It was hypothesized that knockdown of *SUC1* would reverse this trend, in other words, knockdown of *SUC1* should leave a significant amount of sucrose unhydrolyzed. Sugar analysis of honeydew samples from whiteflies feeding on different transgenic tomato plants with dsRNA against *SUC1* (dsSUC) revealed that a significant amount of sucrose remained unhydrolyzed in tomato transgenic line dsSUC (6 M) compared to transgenic line with dsAQP+dsSUC construct and dsRNase+dsAQP+dsSUC construct with 0.4 mM and 0.8 mM respectively.

A reverse trend was observed for free glucose, with significantly lower free glucose in honeydew from whiteflies feeding on transgenic line dsSUC (20 mM) followed by plants with the dsAQP+dsSUC and dsRNase+dsAQP+dsSUC constructs with 30 M and 40 M respectively. In summary, sugar analysis revealed that as more dsRNAs were stacked together with dsSUC, sucrose hydrolysis efficiency increased, yielding more free glucose (Figure 5.13).

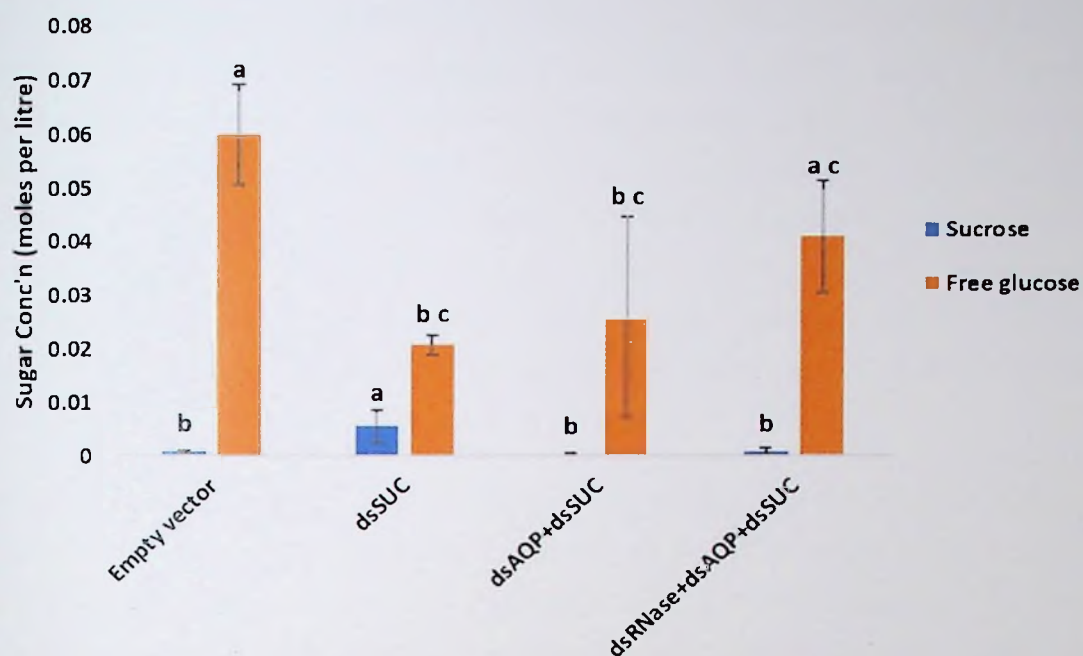


Figure 5.15: Sucrose hydrolysis as a physiological change resulting from whiteflies feeding on different stably transformed tomato plants containing dsRNA against *SUC1*. Treatments sharing the same compact letter display are not significantly different at 5% significance level.

5.3.8 Effect of RNA interference on sugar composition in *B. tabaci* honeydew

Liquid chromatography-mass spectrometry sugar analysis was also conducted to determine the percentage molar ratio of monosaccharides, disaccharides and higher sugars in honeydew samples collected from whiteflies feeding on different transgenic tomato plants. There was no significant difference in the molar ratio of glucose, disaccharides and pentasaccharides in honeydew collected from whiteflies feeding on all tested transgenic plants, compared to honeydew collected from whiteflies feeding on plants with empty vector. Based on the percentage molar ratio, *B. tabaci* honeydew samples tested had more disaccharides in all samples compared to other sugar categories, with the highest amount of disaccharide recorded in honeydew collected from plants with dsRNase construct.

An inverse relation between the number of hexose units (degree of polymerization) and the molar ratio was observed for the oligosaccharides in the honeydew samples analyzed. As the number of hexose units increases, the percentage molar ratio of oligosaccharide reduces, therefore in all honeydew samples, the amount of trisaccharides was higher than tetrasaccharides and pentasaccharides. Analysis of variance of percentage molar ratio among the tested honeydew samples showed that honeydew from whiteflies feeding on plants with dsSUC had a significantly higher percentage of trisaccharides ($P_{0.05} = 0.03$) and tetrasaccharides ($P_{0.05} = 2.21 \times 10^{-6}$) with $14.03 \pm 2.44\%$ and $4.01 \pm 0.26\%$ respectively compared to all honeydew samples from whiteflies feeding on all other tested transgenic plants and plant with the empty vector. Furthermore, honeydew from whiteflies feeding on tomato plants with the dsSUC construct had a significantly lower amount of fructose ($P_{0.05} = 0.006$) compared to other transgenic plants and control plant (Figure 5.14).

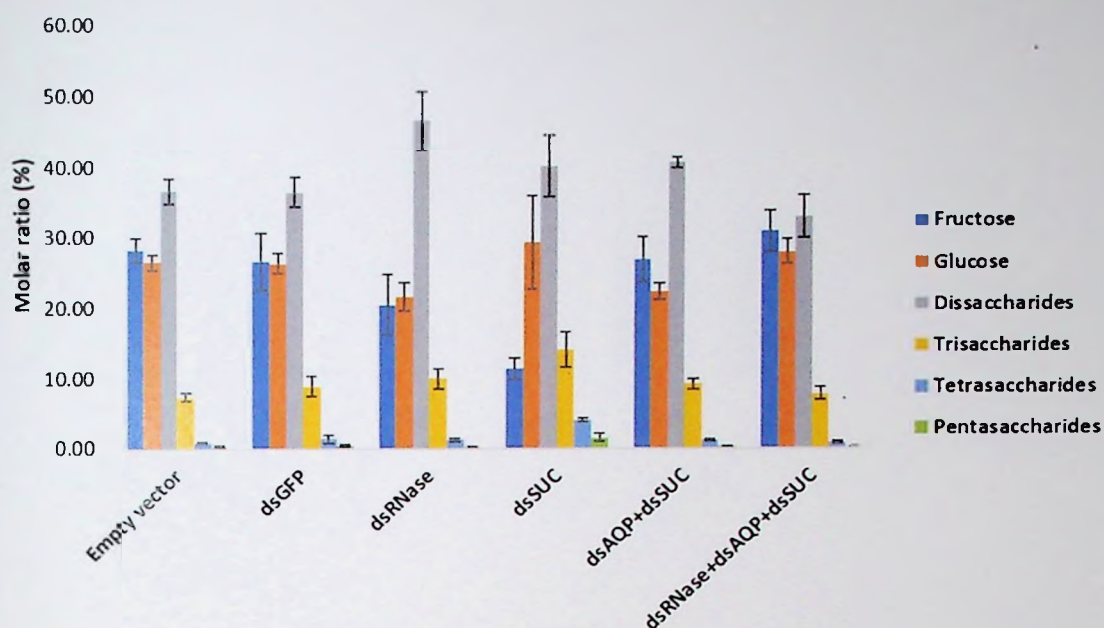


Figure 5.16: The molar ratios of different sugars in honeydew samples collected from whiteflies feeding on different transgenic tomato plants. Error bars represent standard deviations.

5.3.9 Effect of RNAi on the concentration of different sugar categories in honeydew samples

Analysis of concentration of honeydew samples from whiteflies feeding on different transgenic tomato plants revealed that there was no significant difference in concentration of fructose ($P_{0.05} = 0.05$), glucose ($P_{0.05} = 0.26$) and higher sugars in the honeydew samples tested. Honeydew samples from whiteflies feeding on tomato plants with dsSUC construct had relatively few hexose units per mg of honeydew for fructose, glucose and disaccharides but had relatively more hexose units per mg of honeydew for tetrasaccharides and pentasaccharides (Figure 5.15).

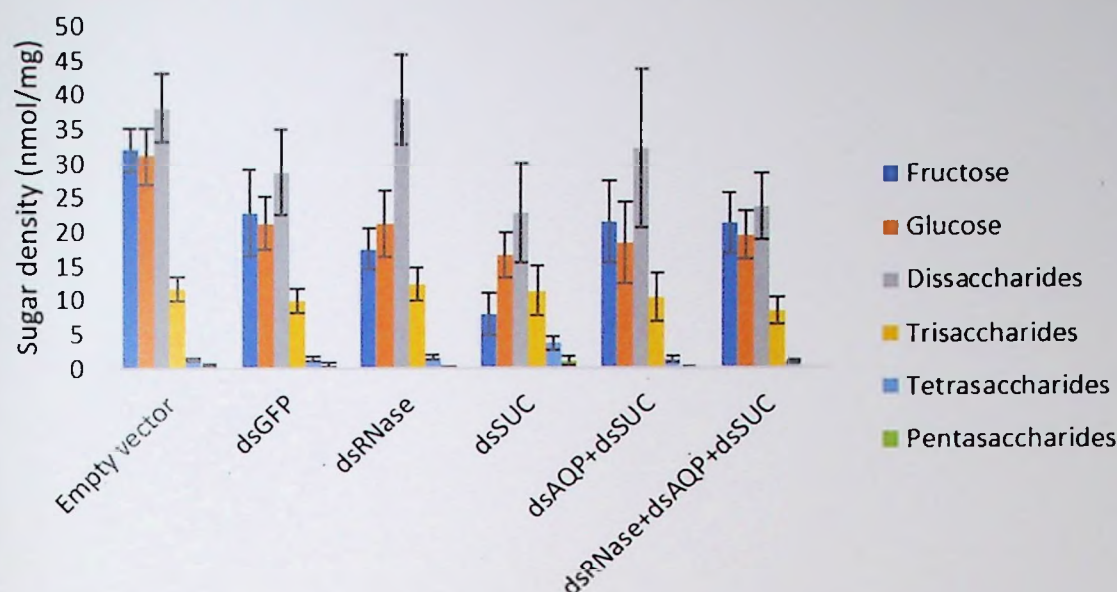


Figure 5.17: Sugar concentration (hexose units per mg of honeydew) collected from whiteflies feeding on different transgenic tomato plants. Error bars represent standard deviation.

5.3.10 Comparison of diet sugar utilization of whiteflies and aphids

5.3.10.1 Assimilation efficiency of whiteflies and aphids feeding on artificial diet

A comparison of sugar concentration of honeydew from both *B. tabaci* and *Aphis gossypii* feeding on an artificial diet with 0.75M sucrose concentration showed that *B. tabaci* honeydew had significantly lower hexose units per mg of honeydew (313.1 ± 51.1 nmol/mg) compared to *A. gossypii* (612.3 ± 67.7 nmol/mg). Results also showed that *B. tabaci* MEAM1 assimilated more dietary sugar than aphids with an assimilation rate of 79.1% compared to 59.2% for aphids (Table 5.2). On tobacco, the sugar concentration in honeydew is much lower for both whiteflies and aphids. Sugar concentration pattern for honeydew from tobacco was similar to that of artificial diet, with significantly lower hexose units (100.9 ± 12.6 nmol/mg) in *B. tabaci* honeydew, compared to that in aphid honeydew (445.8 ± 99.2 nmol/mg).

Table 5.2: Sugar concentration per mg of honeydew collected from the whitefly *B. tabaci* MEAM1 and the aphids *A. gossypii* feeding on artificial diet

Insect	Sugar concentration (Hexose units nmol/mg)		Assimilation efficiency (%) **
	Artificial diet (0.75M)	Tobacco	
<i>B. tabaci</i> (MEAM1)	313.1±51.1	100.9±12.6	79.1
<i>A. gossypii</i> (Aphid)	612.3±67.7	445.8±99.2	59.2
<i>p-value</i>	0.002	0.003	

** Assimilation efficiency based on sugar concentration of honeydew from aphids and whiteflies feeding on an artificial diet with known sugar concentration

5.3.10.2 Comparison of molar ratio and concentration of sugars in aphid and *B. tabaci* honeydew samples

Comparison of the molar ratios of sugars in honeydew collected from aphids and whiteflies feeding on tobacco and an artificial diet with known sucrose concentration revealed that whiteflies feeding on tobacco produced significantly higher amounts of monosaccharides; fructose and glucose with percentage molar ratio of 23.26 ± 1.89 and 23.32 ± 1.42 , respectively compared to honeydew from aphids (14.41 ± 2.71 and 9.42 ± 1.82 , respectively). On the contrary, aphids produced significantly higher amount of higher sugars (tetrasaccharides and pentasaccharides), with percentage molar ratio of 10.26 ± 3.92 and 2.61 ± 1.01 , respectively compared to 1.14 ± 0.15 and 0.09 ± 0.02 respectively in *B. tabaci* honeydew.

Analysis of honeydew collected from aphids and whiteflies feeding on an artificial diet with the same concentration of sucrose (0.75 M) revealed that there was no significant difference in the percentage molar ratio of fructose, glucose and trisaccharides in honeydew samples. Honeydew from whiteflies had a significantly higher amount of disaccharide, almost 50% compared to 28.9% in the honeydew from aphids. On the other hand, honeydew from aphids was much richer in higher sugars than that from whiteflies. Honeydew samples from aphids feeding on an artificial diet had the significantly higher amount of tetrasaccharides and pentasaccharides (16.5% and 5.6% respectively) compared to 4.1% and 0.4%, respectively from *B. tabaci* honeydew samples (Figure 5.16). These results show that *B. tabaci* honeydew is composed of more monosaccharides and disaccharides (higher percentage of monomers), while aphids produce a high percentage of oligomers.

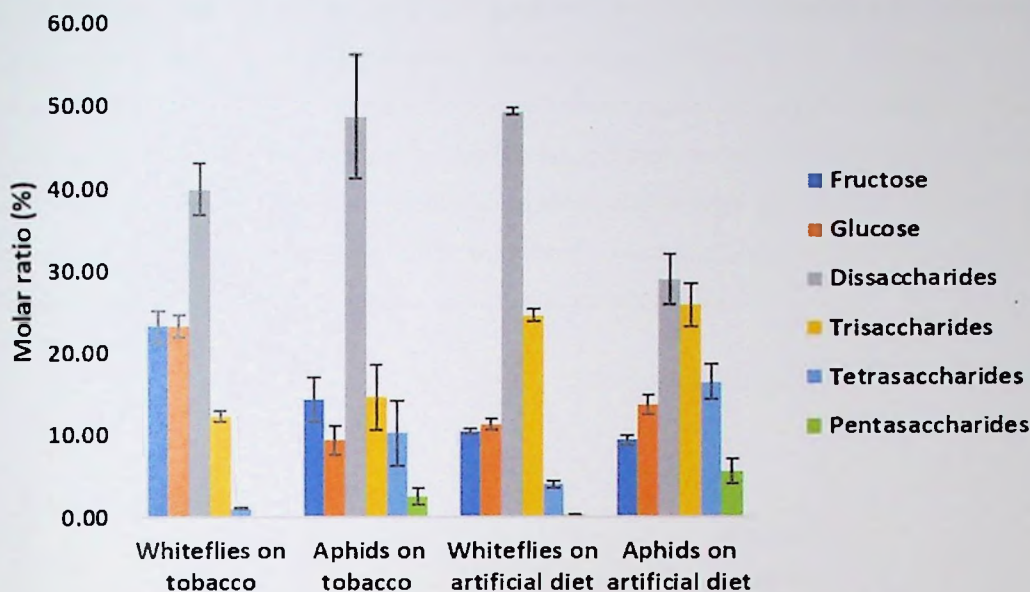


Figure 5.18: Comparison of molar ratios of sugars in honeydew samples collected from aphids and whiteflies feeding on tobacco and artificial diet. Error bars represent standard deviation.

In addition, analysis of sugar concentration in honeydew samples collected from both aphids and whiteflies feeding on tobacco and an artificial diet showed that honeydew collected from aphids was much richer in sugars compared to honeydew from whiteflies (Figure 5.17). Pairwise comparison of monosaccharides, disaccharides and higher sugars of aphid and whitefly honeydew sugars showed no significant difference in the concentration of fructose and trisaccharides per mg of honeydew, in both aphids and whiteflies honeydew samples from tobacco. Whiteflies feeding on tobacco produced honeydew with a significantly higher number of hexose units per mg of honeydew for glucose (11.87 ± 1.52), compared to 6.74 ± 1.18 in aphid honeydew sample. On the contrary, honeydew samples from aphids feeding on tobacco had significantly higher concentration for disaccharide (110.6 nmol/mg), tetrasaccharides (23.37 nmol/mg) and pentasaccharides (0.07 nmol/mg) compared to 19.18 per mg, 1.08 per mg and 0.02 per mg respectively, in *B. tabaci* honeydew.

On the artificial diet, the pattern of sugar concentration for monosaccharide, disaccharide and oligosaccharide changed, showing that the source of sugar influences the concentration of sugars in the resulting honeydew. Comparison of the concentration of different sugars in honeydew samples collected from whiteflies and aphids feeding on the same artificial diet with 0.75M sucrose showed that only tetrasaccharides in aphid honeydew sample were significant, with 46.2 nmol/mg compared to 8.64 nmol/mg in *B. tabaci* honeydew. In summary, aphids produce honeydew richer in sugars than the equivalent honeydew samples from whiteflies.

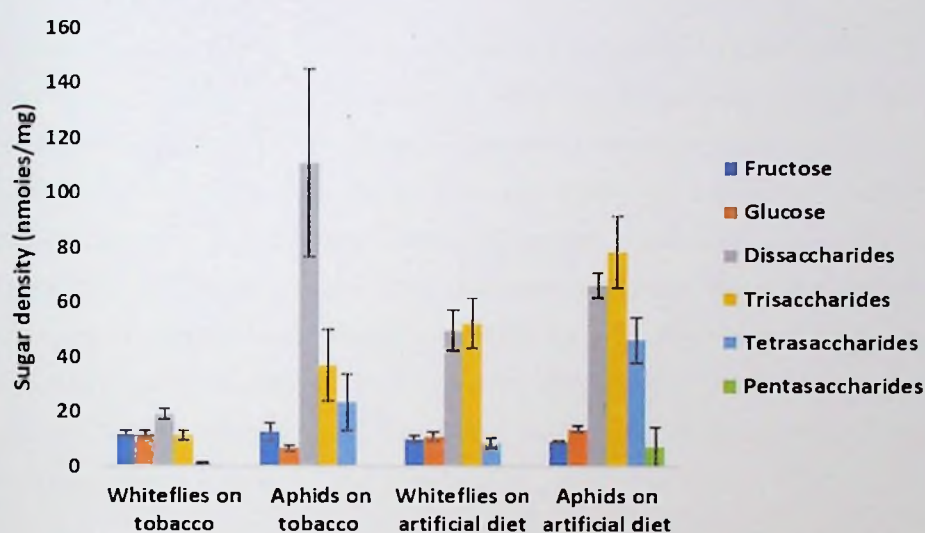


Figure 5.19: Comparison of sugar concentration in different honeydew samples collected from aphids and whiteflies feeding on both tobacco and artificial diet. Error bars represent standard deviation.

5.4 DISCUSSION

RNAi is a promising strategy in the management of *B. tabaci* in different cropping systems (Grover *et al.*, 2019). The rapidly increasing population of cassava whiteflies in cassava growing areas, commonly associated with the *B. tabaci* with *mtCOI* sequence classified as SSA1-SG1 requires species-specific control methods that are environmentally friendly. In this study, the efficacy of using RNAi as a post-transcriptional gene silencing strategy was evaluated on the *B. tabaci* MEAM1 whitefly species.

Two experiments were conducted to evaluate the impact of different stably transformed tomato plants on the survival and development of whiteflies. In the first, no significant difference was observed in the mortality of whiteflies feeding on all transgenic plants tested, with low percentage mortality not exceeding 12%. When the experiment was repeated, a higher and significant mortality occurred for whiteflies feeding on tomato plants with the dsAQP+dsSUC construct (Figure 5.7). Furthermore, there was, evidence of genetic, physiological effects on osmotic pressure and honeydew composition, especially for whiteflies feeding on transgenic plants with the dsAQP+dsSUC construct. For example, dsAQP+dsSUC construct plants had reduced expression of both *AQP1* and *SUC1* in both first and second experiment, although the reduction was not significant (Figure 5.8 and Figure 5.9).

The minimal phenotypic and physiological effect observed especially in whiteflies feeding on dsRNase+dsAQP+dsSUC construct imply that there were underlying factors that affected the expression of dsRNA and the downregulation of the expression of the osmoregulation genes targeted in this study. In several studies, these factors have been investigated in different pests and include: (i) incomplete activation of RNAi machinery within the insect leading to low or no expression of core RNAi genes (Dicer-2, Argonaute 2 and R2D2) (Belles, 2010), (ii) insufficient concentration and length of dsRNA fragment (Miller *et al.*, 2012), (iii) delivery of dsRNA into insect cell and (iv) non-specific degradation of dsRNA within the target insect (Christiaens *et al.*, 2014; Wang *et al.*, 2016). To understand the factors that affected the efficiency of RNAi in *B. tabaci* in this study, two possibilities are considered which are the: (i) fate of dsRNA in the whitefly and (ii) fate of dsRNA in the stably transformed plants.

Terenius *et al.* (2011) reviewed more than 148 studies on RNA interference especially in the control of lepidopteran insects to understand clearly the factors that limit the efficacy of RNAi and their implications on experimental design. These studies revealed that the major problem influencing the efficacy of RNAi in insects relates to the fate of dsRNA in the insect. Insects possess endonucleases that breakdown dsRNA before it triggers the RNA interference machinery within the insect. Similar conclusions have been made in several RNAi studies, for example, on (i) tarnished plant bug (Allen & Walker, 2012), (ii) silkworm (Liu *et al.*, 2012) and (iii) desert locusts (Wynant *et al.*, 2014).

In *B. tabaci*, the molecular basis of effective RNAi has been investigated (Luo *et al.* 2017). Based on this study, the effectiveness of RNAi was influenced by non-specific endonucleases (*RNase1* and *RNase2*), that degrade the dsRNA. In this study, tomato plants were stably transformed with phloem-specific promoters; *Galactinol synthase* (*Cucumis melo*, *CmGAS*) (Ayre *et al.*, 2003) and *Sucrose-H⁺ symporter* (*Arabidopsis thaliana*, *AtSUC2*) (Dunoyer *et al.*, 2007), and dsRNA to knock down expression of both RNases (*RNase1* and *RNase2*) and the selected osmoregulation genes (*SUC1* and *AQP1*) in *B. tabaci*. Therefore, with a premise that the dsRNAs were protected within the whiteflies especially those feeding on dsRNase+dsAQP+dsSUC construct, it can be concluded that factors that affected this study relate mostly to the fate of dsRNA in the stably transformed plants. Results in this study imply either (i) whiteflies did not obtain enough dsRNA (low concentration of dsRNA) from the transgenic plants' phloem sap to activate the RNAi machinery, or (ii) it required more time to observe the phenotypic effects resulting from the downregulation of the osmoregulation genes.

There are many factors that influence the expression of dsRNA within the plant, which influences the concentration and availability of dsRNA within the phloem-sap of the transgenic plant and the delivery of dsRNA from the plants. Therefore, more studies are required to investigate the fate of dsRNA in plants and ascertain these factors, especially in cassava as progress is made towards RNAi application at field level. Possible factors include: (i) position-effects (Hanin & Paszkowski, 2003), especially when the transgene is inserted in or near heterochromatin region or region of repetitive DNA (Prols & Meyer, 1992), (ii) cis- or trans-inactivation of transgene, common when multiple transgenes driven by single promoter are inserted into the plant either at the same or different locus (Linn *et al.*, 1990; Meyer *et al.*, 1992) and (iii) methylation (Li *et al.*, 2001) which

renders the promoter inaccessible to the transcription initiation complex causing transgene silencing. (Finnegan & McElroy, 1994). One or all these factors could have influenced the expression of dsRNA in the transformed plants used in this study. More studies and approaches are required to improve the efficacy of RNA interference in control of whiteflies.

Several approaches have been suggested and implemented to improve transgene expression in plants (Butaye *et al.*, 2005). Implementing these approaches while developing stably transformed plants against essential *B. tabaci* genes is more likely to improve the efficacy of RNAi in the control of *B. tabaci* as a vector and pest. These approaches include the application of a matrix attachment region, a non-transcribed AT-rich DNA sequence to flank the transgene prior to transformation thereby (i) shielding the gene from encroaching heterochromatin (Phivan *et al.*, 1990; Spiker & Thompson, 1996; Allen *et al.*, 2000; Nowak *et al.*, 2001) and (ii) increasing transgene expression (Allen *et al.*, 1996; Levin *et al.*, 2005; Wang *et al.*, 2007). In this study, matrix attachment regions were not used. Previous studies demonstrated that transgene expression in stably transformed tomato plants was enhanced by using a matrix attachment region (Chinn & Comai, 1996). Secondly, position-effects arising from the random integration of the transgene into the genome can be controlled through site-specific recombination (Srivastavai *et al.*, 2004), for example, *FLP* recombinase (Buchholz *et al.*, 1998) and *Cre* recombinase (Gilbertson, 2003). These systems allow the introduction of the transgene at a specific or favorable locus within a plant genome, which can improve transgene expression. The *Cre-loxP* system has been widely used especially in site-specific recombination for genetic engineering in plants (Lyznik *et al.*, 2003).

The nature and quantity of sugar in honeydew samples collected from whiteflies feeding on different transgenic plants and osmotic pressure of whitefly haemolymph were used as to make inference about the physiological effect of RNAi on *B. tabaci*. It was hypothesized that downregulating *SUC1* expression would influence sucrose hydrolysis within *B. tabaci* thereby increasing osmotic pressure in the haemolymph leading to the death of the whitefly. Sugar analysis of honeydew samples collected from whiteflies feeding on transgenic plants with dsSUC revealed that a significant amount of sucrose remained unhydrolyzed, resulting in lower free glucose compared to plants with the dsAQP+dsSUC and dsRNase+dsAQP+dsSUC constructs (Figure 5.13). In addition, analysis of higher sugars in honeydew collected from different transgenic plants showed that RNAi did not cause a significant effect on both the molar ratio and the sugar

concentration of different sugars in the honeydew, except for that collected from plants with dsSUC. These results imply that: (i) there was a knockdown of *SUC1* gene in dsSUC (Figure 5.8 and Figure 5.9). The gene knockdown was partial, however, so the effect was not sufficiently large to cause an increase in osmotic pressure (Figure 5.12) and whitefly death. (ii) As more dsRNAs are stacked together (combinatorial RNAi), the efficiency of single dsRNA in downregulating a target gene (in this case, *SUC1*) is reduced. Tzin *et al.* (2015) evaluated the effect of both single and combinatorial RNAi against gut osmoregulatory genes in green peach aphid (*Myzus persicae*) and potato/tomato psyllid (*Bactericera cockerelli* Sule). The effect was particularly pronounced for insects treated with RNAi against multiple genes (combinatorial RNAi) compared to single genes. The difference, however, between this study and that of Tzin *et al.*, (2015) is the nature of plant transformation. In the latter, transgenic plants were transiently transformed while plants in this study were stably transformed. More studies are required to investigate the effect of stacking dsRNA on the efficacy of RNAi especially in stably transformed plants.

The whiteflies and aphids used in this study were fed on the same dietary source (artificial diet and tobacco) to compare how two different phloem-sap feeders species utilize dietary sugar. An artificial diet with known sucrose concentration was used to estimate both hydrolysis rate and assimilation efficiency of the whiteflies and aphids tested. Nearly all the sucrose (0.74 M out of 0.75 M) in the artificial diet was hydrolyzed by both whiteflies and aphids. However, a significant difference in the assimilation efficiency was observed, indicating that the whiteflies studied assimilate more dietary sugar than the aphids used in this study. This suggests that whiteflies may either have a higher metabolic rate compared to the aphids or the whiteflies ingest less sugar per unit weight compared to aphids. On the other hand, sugar density/concentration of honeydew obtained from whiteflies and aphids feeding on tobacco were lower for both whiteflies and aphids compared to that of honeydew from the artificial diet (Figure 5.4). This could either be due to the low insect preference to tobacco compared to an artificial diet which influences the feeding rate of these insects or low amount of sucrose in tobacco phloem-sap. Studies by Salvucci *et al.*, (1997) showed that the amount and type of carbohydrates excreted by whiteflies in their honeydew depends on the concentration of sucrose in the diet.

Phloem-sap feeders hydrolyze all the sucrose to maintain fluid homeostasis within the insect gut and body cavity. The most dominant sugar category in the honeydew collected from whiteflies and

aphids was disaccharides. Previous studies reported that the dominant sugar in *B. tabaci* (*Bemisia argentifolli* = MEAM1) honeydew was a disaccharide, trehalulose (Salvucci *et al.*, 1997; Byrne *et al.*, 2003). Other oligosaccharides include bemisiose (trisaccharide), maltosucrose (tetrasaccharide) and diglucomelezitose (pentasaccharides) (Hendrix *et al.*, 1992; Hendrix & Wei, 1994; Wei *et al.*, 1997). Phloem-sap feeders hydrolyze sucrose to fructose and glucose and use the fructose for the insect dietary needs (Ashford *et al.*, 2000). The amount of fructose and glucose used for metabolism depends on the metabolic needs of the insect. Excess monosaccharides, most preferably the glucose moiety of sucrose is used through transglucosidation to synthesize disaccharides and higher sugars, which are osmotically neutral with the insect haemolymph (Wilkinson *et al.*, 1997). This process is conserved in different phloem-sap feeders, though with different sugar conversion efficiencies. In this study, a difference was also observed in sugar conversion between whiteflies and aphids. Aphids convert more sugar to long-chain oligosaccharides especially trisaccharides, tetrasaccharides and pentasaccharides compared to whiteflies. This discrepancy between whiteflies and aphid's conversion efficiency may either be due to a high rate of food movement within the *B. tabaci* gut compared to aphid or the difference in sugar conversion resulting from different isomerization products in both whiteflies and aphids. Whiteflies possess monomeric enzymes, sucrose isomerases, which catalyzes the isomerization of excess sucrose into trehalulose and isomaltulose (Salvucci, 2003). This could explain the difference in the assimilation efficiency between whiteflies and aphids.

In conclusion, results in this study have demonstrated that whiteflies, like other phloem-sap feeders, require inherent physiological abilities to hydrolyze sucrose to its constituent monomers to maintain their fluid homeostasis. Different phloem feeders assimilate sugars differently, so preliminary studies are required to understand sugar assimilation and identify essential genes encoding for key proteins required in these processes. In Chapter 3, osmoregulation genes (aquaporin 1 and alpha-glucosidase (family 13)) for cassava whitefly (SSA1-SG1) were identified. These genes show high similarity with those that mediate osmoregulation in MEAM1. The efficacy of RNAi against these osmoregulation genes using stably transformed tomato plants was tested as described above. Results showed that there were both physiological and genetic effects, especially in whiteflies feeding on stably transformed tomato plants with the dsAQP+dsSUC construct, although the effect was minimal. Further experiments are required, therefore, to

ascertain the factors that limited the downregulation of the target osmoregulation genes in the whitefly. The starting point would be with the stably transformed tomato plants and to use a micro-scale phloem-sap-extraction sensor device (Ono *et al.*, 2017) to collect pure phloem-sap. The collected phloem-sap would then be tested for presence and quantity of dsRNAs using either ion pair reverse phase chromatography and mass spectrometry (Nwokeoji *et al.*, 2017) or liquid chromatography/mass spectrometry (LC-MS analysis) (Kung *et al.*, 2018) to determine the amount of dsRNA in the phloem-sap, which would infer the level of expression of specific dsRNA in the transformed tomato plants. This would provide strong evidence of whether or not the limiting factors are related to plants or to the insect. Identifying and addressing these current knowledge gaps will enable future researchers to enhance the effectiveness of RNAi in the management of whitefly populations in different cropping systems.

GENERAL DISCUSSION AND CONCLUSIONS

Cassava as a food security and an increasingly economic crop in the tropics especially in Sub Saharan Africa is threatened by devastating plant viruses (CMVs and CBSVs) transmitted by *B. tabaci* (Section 1.1). The resurgence of whiteflies in different areas and crops have prompted farmers to use readily available management options (Section 2.3) to combat these whitefly pest problems. These methods, for example, the application of insecticides have not been effective in the long term, especially when used in isolation. This is because insecticides not only disrupt the ecosystem but some whitefly species (such as *Bemisia tabaci* MED) possess insecticide resistance genes that confer resistance to insecticide groups such as the neonicotinoids (Alon *et al.*, 2008; Wang *et al.*, 2010; Kontsedalov *et al.*, 2012; Horowitz & Ishaaya, 2014).

There is an increasing demand, therefore, for new methods that are both environmentally friendly and can selectively inhibit or downregulate target gene expression with high specificity and fidelity for use in integrated whitefly management programs. This study was initiated, therefore, to identify gene targets that can be used to manage whitefly population using RNAi, a sequence-specific and environmentally friendly approach to pest management.

For RNAi to be effective in the management of whitefly populations in African cassava growing regions, one of the most important factors to consider is the selection of the gene target. In chapter 2 section 5 (Section 2.5) of this thesis, the categories of essential genes in different insects were reviewed. In Chapter 3, both osmoregulation and symbiosis genes in cassava *B. tabaci* SSA1-SG1 were identified. The selection of these two sets of genes was because whiteflies, like other phloem-sap feeders require special genes to: (i) hydrolyse sucrose in the phloem-sap to maintain optimum osmotic pressure within the gut and also for their nutritional needs, (ii) ensure water recycling within the gut to maintain fluid haemostasis within the body cavity and (iii) interact with endosymbionts within the bacteriocytes for essential amino acids biosynthesis (Douglas, 2006).

Among the osmoregulation genes identified, the candidate genes selected for sucrose hydrolysis belonged to α -glucosidase family 13. Alpha-glucosidase family 13 is a multi-gene family composed of a total of 48, 35 and 47 genes in MEAM1, MED¹³ and SSA1-SG1 *B. tabaci* species

¹³ ftp://parrot.genomics.cn/gigadb/pub/10.5524/100001_101000/100286/

respectively (Supplementary Table 6.1, 6.2 & 6.3). Selection of a candidate gene as a gene target from a multi-gene family poses a challenge. For example, in previous studies by Raza *et al.* (2016) and Vyas *et al.* (2017), an α -glucosidase (NCBI accession KF377803.1) was selected and used as an RNAi target in the management of whiteflies. This gene corresponded to Bta07452, an α -glucosidase in MEAM1. However, functional analysis revealed that this gene lacks both signal peptide and catalytic site residue (Supplementary Table 6.1). This indicates that this gene was non-functional or a pseudogene. The presence of a signal peptide is important because signal peptides are required for the enzyme to function as an extracellular enzyme in the gut lumen (von Heijne, 1998). This example also demonstrates the importance of the need of a detailed analysis to select the best gene target from a multi-gene family.

In Chapter 3, a detailed analysis starting with: (i) RNAseq analysis to identify genes with enriched gene expression either in the gut or bacteriocyte, (ii) functional analysis of each selected genes to identify those that encode for proteins with functional domains, (iii) selecting genes with functional domains of interest, (iv) phylogenetic analysis to compare the selected genes with experimentally validated genes and (v) analysis of conserved regions especially region II to determine the substrate specificity or preference of the candidate enzyme.

Through this detailed approach, two candidate genes were selected as gene targets for disrupting sucrose hydrolysis in SSA1-SG1. One of these genes was, ENSSSA1UGT002066 (Ssa05164), which corresponded to NCBI accession KX390871.1, an α -glucosidase sequence selected by Jing *et al.* (2016) and used by Luo *et al.* (2017). This sequence corresponded to Bta03818, an α -glucosidase in MEAM1, which contains both signal peptide and catalytic site residue (Supplementary Table 6.1) and was highly similar to the experimentally validated sucrase (Price *et al.*, 2007) (Figure 3.13). However, this gene lacked a proline and tyrosine at position four and five within the conserved region II of the α -glucosidase, that determines sucrose preference or specificity of the α -glucosidase enzymes in the honey bee (Ngiwsara *et al.*, 2012). The best gene target identified using this detailed approach was ENSSSA1UGT002057 (Ssa05154), an ortholog of Bta05386, an α -glucosidase in MEAM1. This gene had all the characteristics of *sucrase I* (*SUCI*) which included: (i) high expression in the gut (Figure 3.20), (ii) possessed both signal peptide and a catalytic site residue and (iii) protein residues for sucrose preference/specificity within the conserved region II of α -glucosidase (Figure 3.12). Studies are on-going to validate

whether proline and tyrosine at position four and five within the conserved region II determines preference or specificity of α -glucosidase to sucrose in *B. tabaci*. Interestingly, both transcriptomic and RT-qPCR gene expression analysis of the selected genes (ENSSSA1UGT002057 and ENSSSA1UGT002066) showed that the selected genes had enriched expression in *B. tabaci* gut (Figure 3.21), so can be used as gene targets in an RNAi system to manage *B. tabaci* populations.

As mentioned above, *SUC1* enzyme in *B. tabaci* facilitates both osmoregulation and digestion through hydrolysis of sucrose to glucose and fructose. Both the hydrolysis and assimilation rate for sucrose in whitefly was quantified (Table 5.3). Sugar analysis in honeydew samples collected from whiteflies showed that 97% of sucrose was hydrolyzed to glucose and fructose (Section 5.3.5) showing the importance of these selected genes in whitefly osmoregulation and digestion. In aphids (*Acyrtosiphon pisum* Harris), osmoregulation has been studied in detail (Ashford *et al.*, 2000; Cristofolletti *et al.*, 2003).

In this study, we compared assimilation efficiency of both aphids (*Myzus persicae*) and *B. tabaci* (MEAM1), showing that *B. tabaci* has a higher assimilation efficiency compared to aphids. These insights show that phloem-sap feeders hydrolyze sucrose the same way but have differences in assimilation efficiencies and transglucosidation which related to insect physiology especially the enzymes involved. Understanding these differences is important in selection of gene targets and also the designing of management options for different pests.

B. tabaci also regulate osmoregulation by facilitating the osmotic movement of water from the distal gut to the proximal midgut ensuring fluid haemostasis and equilibrium of the osmotic pressure between the haemolymph and the gut lumen. The osmoregulation gene ENSSSA1UGT022145 (Ssa03484) (*AQPI*), therefore was identified and was of particular interest, because of its role in water recycling in the gut of SSA1-SG1. Functional analysis showed that this gene encodes for a protein with a conserved fingerprint (asparagine-proline-alanine motif) that ensures water selectivity and an aromatic/arginine region as a proton filter. These conserved features give the *AQPI* a two-stage filter that ensures water selectivity (de Groot & Grubmuller, 2001). Pairwise comparison with the corresponding ortholog in MEAM1 (Supplementary figure 6.2) showed that this gene shared a 98% similarity to Bta01973, an *AQPI* in MEAM1 (Mathew *et al.*, 2011). Downregulation of this gene within the *B. tabaci* gut will disrupt the water recycling from the distal gut inducing malfunctioning of osmoregulation in the whiteflies.

B. tabaci require essential amino acids for their growth and survival, which are deficient in the phloem-sap that they utilize (Section 2.7). In Chapter 3 and Chapter 4, the essential amino acid provisioning in cassava whitefly SSA1-SG1 was studied. RNAseq analysis revealed three sets of genes with enriched expression in the SSA1-SG1 bacteriocytes. These include, (i) amino acid transaminase, (ii) amino acid transporters and (iii) horizontally transferred genes. Genome-scale metabolic reconstruction coupled with constraint-based metabolic modeling showed that amino acid transaminase and horizontally transferred genes played a critical role in essential amino acid biosynthesis. Kidwell, (1993) defined horizontal gene transfer as the non-sexual transmission of genetic material across a species boundary. In Chapter 3, horizontally transferred genes of *B. tabaci* SSA1-SG1 of either bacterial or fungal origin were identified (Table 3.8 and 3.9).

Once introduced within the system, the horizontally transferred genes cause functional redundancy that makes one copy of the gene to undergo relaxed selection (McCutcheon & Moran, 2012) while the other copy may either undergo degenerative mutation that leads to a pseudogene (nonfunctionalization) or the two genes may undergo mutation that reduces their expression to a level of a single indigenous gene (subfunctionalization) (Lynch & Force, 2000). In chapter 4, results showed that shared metabolic interaction within *B. tabaci* favour nonfunctionalization, especially of the endosymbiont gene while preserving the horizontally transferred genes within the whitefly (Table 4.2, Figure 4.5).

These phenomena have led to innovation transfers (Husnik & McCutcheon, 2018) in which *B. tabaci* are slowly gaining genetic ability or new functionalities especially in the biosynthesis of essential amino acids and B vitamins. This is very important especially in the search for critical genes to select as gene target in novel pest control approaches. Results presented here further showed that these horizontally transferred genes are placed at critical points within the essential amino acid biosynthesis pathways. For example, genes such as *argH* and *lysA* mediate terminal reactions of arginine and lysine respectively (Figure 4.5), which gives the whitefly as the host control over essential amino provisioning within the bacteriocytes. Constraint-based metabolic modeling (Chapter 4) showed that many of these innovative transfers within the whitefly encode for proteins that are essential for the development and survival of whitefly.

Among the amino acid transferase, branched-chain amino acid transferase (*BCAT*) was identified as an indispensable gene of intrinsic origin encoding for a protein that mediated the terminal

reaction of branched-chain amino acids (BCAA) (isoleucine, leucine and valine) within the cassava *B. tabaci* (SSA1-SG1) bacteriocyte. The mediation of the terminal reaction of the three essential amino acids and its indispensability makes *BCAT* one of the best symbiosis gene targets. BCAA are very important, not only in protein synthesis, but also in metabolic and regulatory roles. No studies have been done so far to investigate the role of branched-chain amino acids in *B. tabaci*, however, BCAA especially isoleucine have been reported to mediate glucose uptake and metabolism (Doi *et al.*, 2003; Liu *et al.*, 2017; Zhang *et al.*, 2017). It can be envisaged that downregulation of both the selected α -glucosidase genes (*SUC1* and *SUC2*) and *BCAT* will not only not disrupt osmoregulation but will also affect essential amino acid and energy production within cassava *B. tabaci*.

Today, more than 39 morphologically indistinguishable species making up the *B. tabaci* cryptic species complex have been reported (Section 2.2). Results from this study revealed that the key osmoregulation and symbiosis genes selected as target genes are highly homologous/similar between the *B. tabaci* species examined. Homology search in the whitefly genome database revealed that the selected genes in SSA1-SG1 are highly similar compared to corresponding orthologs in MEAM1. For example, ENSSSA1UGT002057 shares a 96% similarity with Bta05386. On a broader scale, a total of 49 pairwise d_N/d_S comparisons was done to determine the selection pressure acting on different α -glucosidase genes in MED, MEAM1 and SSA1-SG1. Only two comparisons (4 genes), all within SSA1-SG1 had d_N/d_S greater than 1 (Supplementary Table 6.4), showing that the α -glucosidase in the three *B. tabaci* species are not only duplicated but are highly conserved. Therefore, the non-synonymous mutations within these genes were deleterious (except for the four genes) and therefore removed by purifying selection for a sustained function of the orthologs.

For protein provisioning, the structure of the metabolic models of SSA1-SG1 and MEAM1 (Table 4.3) shows that the metabolic interaction between the host (*B. tabaci*) and its primary endosymbiont (*Portiera*) within these two economically important *B. tabaci* species is highly similar. Furthermore, essential genes involved in this interaction are highly similar for example, *argH* (ENSSSA1UGT003467 (Ssa15329)) shared 95% similarity with Bta00063, an arginine succinate lyase in MEAM1, *lysA* (ENSSSA1UGT007134 (Ssa05826)) shared 96% similarity with Bta03593, a diaminopimelate decarboxylase in MEAM1 and *BCAT* (ENSSSA1UGT021004

(Ssa04473)) had 97% similarity with Bta10673, branched-chain amino acid aminotransferase in MEAM1. Therefore, RNAi strategies designed to downregulate *argH*, *BCAT* and *lysA* should be effective in the management of both SSA1-SG1 and MEAM1.

The high conservation of critical genes in different whitefly species means that RNAi transgenic plants can be produced using these highly conserved whitefly gene sequences to control more than one whitefly species at a time. Also, that, the conserved region among paralogs (duplicated genes within a given species) can be used to downregulate multiple genes copies simultaneously. This would, in turn, make the development of control strategies of different whitefly species cost-effective.

For other genes however, a comparison of gene content in the three *B. tabaci* species (MED, MEAM1 and SSA1-SG1) revealed that the cassava whitefly SSA1-SG1 is evolving faster than MED and MEAM1. This conclusion is based on two observations, firstly, in Chapter 4, it was observed that primary endosymbiont of SSA1-SG1 (SSA1-SG1-*Portiera*) has already purged itself of a gene (*dapB*) due to functional redundancy resulting from the introduced *dapB*, a horizontally transferred gene in the bacteriocyte, yet this gene is still present in the *Portiera* genome of MED and MEAM1 as a pseudogene (Table 4.2) (Luan *et al.*, 2015). Secondly, of the three *B. tabaci* species, only SSA1-SG1 had four α -glucosidase genes under positive selection (Supplementary Table 6.4). Therefore, the non-synonymous mutations within these four genes were fixed which may have strongly influenced the function and fitness of cassava whitefly. These and many other changes within SSA1-SG1 may partly explain the high relative fitness of SSA1-SG1 observed in the field (Mugerwa *et al.*, 2019), though more work is needed to confirm this conclusion.

Now that essential genes in cassava *B. tabaci* (SSA1-SG1) have been identified, efforts should shift to the development and implementation of an efficient RNAi delivery system for controlling *B. tabaci* especially in farmer-preferred cassava varieties. In Chapter 5, the efficacy of RNAi in the downregulation of *AQP1* and *SUC1* was evaluated on *B. tabaci* MEAM1. Downregulation of *AQP1* and *SUC1* was effective in whiteflies feeding on tomato transgenic plants with the dsAQP+dsSUC construct (Figure 5.8 and Figure 5.9) although the downregulation in gene expression was insufficient to cause significant phenotypic effect on whitefly survival. Development of efficient dsRNA delivery system such as stably transformed plants that can be used in the management of *B. tabaci* population in both the field and greenhouses is of paramount

importance. The most important factors to consider in this process include, (i) the development of constructs that are effective in the down-regulation of the desired gene(s), (ii) development of an efficient dsRNA delivery system that delivers dsRNA in the required concentration and (iii) protecting the dsRNA against endonucleases within the whitefly body cavity (Section 5.1).

There has been great progress in understanding the fate of dsRNA within the whitefly and measures to mitigate the degradation of dsRNA by the endonuclease (Luo *et al.*, 2017). What remains to be understood, however are the different factors that influence expression of dsRNA within stably transformed plants and the impact of stacking dsRNA (combinatorial RNAi) on the efficacy of RNAi in the management of *B. tabaci*. In Chapter 5, the efficacy of RNAi in the management of *B. tabaci* was evaluated. Results showed that there was a limited effect on whiteflies especially those feeding on stably transformed tomato plants (dsRNase+dsAQPI+dsSUC) on the development and survival of whiteflies. This may be because either: (i) the whiteflies did not obtain the dsRNA from the transgenic plants, (ii) the concentration of dsRNA obtained from the plants was limited and therefore, did not activate the RNAi machinery or (iii) the stacking of different dsRNA in a single construct (combinatorial RNAi) affected the efficiency of dsRNA in downregulating a target gene (Figure 5.13). Therefore, more studies are required to determine: (i) the factors that influence dsRNA expression especially in stably transformed plants, (ii) the effective lethal dose (LC₅₀), (iii) the whitefly stage that is most vulnerable to RNAi and (iv) the impact of combining different dsRNA (combinatorial RNAi) on the efficiency of dsRNA expression in stably transformed plants.

In conclusion, the identification of essential genes within *B. tabaci* as gene targets in the management of cassava whiteflies using RNAi is the first critical step in the development of an effective RNAi technology. A total of three osmoregulation genes (*AQPI*, *SUC1* and *SUC2*) and five symbiosis genes (*AAT*, *argH*, *BCAT*, *dapB* and *lysA*) have been selected and validated as potential gene targets. These genes show high similarity in both SSA1-SG1 and MEAM1, which offers an opportunity to design an RNAi system that is effective in the management of these two economically important *B. tabaci* species. More work is required to evaluate the impact of targeting these genes singly or in combination on the development and survival of *B. tabaci*. The best gene or gene combination causing the highest mortality of *B. tabaci* should be used in the development of stably transformed plants for use in the integrated whitefly management programs

in cassava growing areas. Management of cassava whiteflies will not only reduce the direct damage caused by high whitefly population but will also reduce the spread of viral diseases (CBSD and CMD) transmitted by these whiteflies. This technology, therefore, has the long-term potential to prevent future cassava viral-disease pandemics.

APPENDIX

Supplementary table 6.1: Genes for MEAM1 *B. tabaci* species encoding proteins with characteristic catalytic site residue and signal peptide for α -glucosidase enzyme

Gene ID	Signal peptide	D-Score*	Cleavage site	Catalytic site residue**	
				Nucleophile	Proton Donor
Bta07764	+	0.532	E(20)E	D(227)	E(289)
Bta03994	+	0.804	T(22)Q	D(236)	E(325)
Bta03992	+	0.819	T(23)E	D(226)	E(297)
Bta03991	+	0.649	A(22)D	D(225)	E(296)
Bta06327	+	0.839	A(19)Q	D(230)	E(451)
Bta03439	+	0.791	A(22)Q	D(227)	E(292)
Bta15649	+	0.450	A(24)E	D(233)	E(300)
Bta05386	+	0.700	A(24)V	D(268)	E(337)
Bta11977	+	0.800	G(17)V	D(229)	E(284)
Bta11978	+	0.783	A(19)Q	D(223)	E(277)
Bta11358	+	0.809	A(20)D	D(227)	E(295)
Bta12678	+	0.749	A(23)D	D(225)	E(279)
Bta08425	+	0.802	G(20)V	D(231)	E(290)
Bta08427	+	0.807	S(18)E	D(227)	E(287)
Bta03818	+	0.712	C(16)R	D(228)	E(296)
Bta05518	+	0.727	C(20)E	D(230)	E(298)
Bta14422	+	0.799	Q(23)S	D(228)	E(296)
Bta12578	+	0.742	A(19)L	D(226)	E(294)
Bta04306	+	0.648	G(21)S	D(228)	E(299)
Bta06458	+	0.765	G(20)G	E(229)	R(287)
Bta08431	+	0.845	G(20)V	D(185)	E(244)
Bta09696	+	0.742	A(25)Q	D(231)	E(318)
Bta04042	+	0.610	G(22)R	-	-
Bta13913	-	0.103	-	-	-
Bta07453	-	0.211	-	D(243)	E(310)

* Discrimination score (D-score) for best-predicted signal peptide for predicted proteins of GH13 family

** Catalytic site residue in each GH13 sequence. The position of the residue is indicated by the number in the parentheses.

Presence or lack of signal peptide denoted as + or - respectively

Supplementary Table 6.1 continued

Gene ID	Signal peptide	D-Score*	Cleavage site	Catalytic site residue**	
				Nucleophile	Proton Donor
Bta05340	+	0.739	G(21)H	D(183)	E(250)
Bta08425	+	0.802	G(20)V	D(231)	E(290)
Bta08426	+	0.811	G(20)V	D(231)	E(290)
Bta07452	-	0.416	-	-	-
Bta10022	-	0.369	-	D(233)	E(300)
Bta01478	-	0.105	-	D(281)	E(369)
Bta05396	-	0.349	-	D(244)	E(312)
Bta05397	-	0.232	-	D(253)	E(321)
Bta02810	-	0.101	-	D(202)	E(269)
Bta11979	-	0.358	-	D(261)	E(315)
Bta12624	-	0.249	-	D(226)	E(294)
Bta12683	-	0.144	-	G(275)	D(331)
Bta12682	-	0.118	-	D(288)	E(344)
Bta10152	-	0.110	-	D(216)	E(284)
Bta04421	-	0.152	-	D(280)	E(342)
Bta14419	-	0.425	-	D(230)	E(296)
Bta04298	-	0.154	-	D(241)	E(308)
Bta04297	-	0.106	-	D(280)	E(348)
Bta03112	-	0.135	-	D(166)	D(233)
Bta13238	-	0.199	-	D(235)	E(302)
Bta13239	-	0.099	-	D(157)	E(220)
Bta09634	+	0.763	G(20)E	-	-
Bta06059	-	0.098	-	D(147)	E(212)

* Discrimination score (D-score) for best-predicted signal peptide for predicted proteins of GH13 family

** Catalytic site residue in each GH13 sequence. The position of the residue is indicated by the number in the parentheses.

Supplementary Table 6.2 continued

Gene ID	Gene*	Signal peptide	D-Score**	Cleavage site	Catalytic site residue***	
					Nucleophile	Proton Donor
BTA002583.1	Bta11977	+	0.800	G(17)V	D(229)	E(284)
BTA004330.1	Bta06059	+	0.746	S(20)Q	D(218)	E(254)
BTA004410.2	Bta12683	+	0.712	G(30)Y	D(250)	E(306)
BTA006577.1	Bta08426	+	0.820	G(20)V	D(231)	E(290)
BTA028567.1	Bta09633	-	0.112	-	D(122)	E(190)
BTA000982.1	Bta03112	-	0.135	-	D(167)	D(234)
BTA000120.2	Bta05386	-	0.193	-	D(191)	E(260)
BTA002581.1	Bta11979	-	0.357	-	D(261)	E(242)
BTA007696.1	Bta09696	-	0.118	-	D(156)	E(242)
BTA011266.1	Bta14419	-	0.113	-	S(156)	E(244)
BTA019527.1	Bta04297	-	0.113	-	D(200)	E(268)
BTA020828.1	Bta07453	-	0.104	-	D(277)	E(344)
BTA021845.1	Bta03818	-	0.116	-	D(209)	E(277)
BTA001570.1	Bta05518	-	0.223	-	D(268)	E(436)
BTA003227.1	Bta06441	-	0.127	-	D(285)	E(353)
BTA011369.1	Bta11975	-	0.179	-	D(258)	E(314)
BTA007986.1	Bta13236	-	0.100	-	D(146)	E(214)

* Corresponding gene in *Bemisia tabaci* (MEAM1). ** Discrimination score (D-score) for best-predicted signal peptide for predicted proteins of GH13 family

*** Catalytic site residue in each GH13 sequence. The position of the residue is indicated by the number in the parentheses. Presence or lack of signal peptide denoted as + or - respectively

Supplementary table 6.3: Genes for SSA1-SG1 *B. tabaci* species encoding proteins with characteristic catalytic site residue and signal peptide for α -glucosidase enzyme

Gene ID	Gene*	Signal peptide	D-Score**	Cleavage site	Catalytic site residue***	
					Nucleophile	Proton Donor
ENSSSA1UGT002057	Bta05386	+	0.675	A(24)V	D(268)	E(337)
ENSSSA1UGT001313	Bta07452	+	0.799	A(20)L	D(227)	E(294)
ENSSSA1UGT031126	Bta10022	+	0.844	A(20)S	D(233)	E(300)
ENSSSA1UGT006139	Bta03439	+	0.711	A(34)Q	D(239)	E(304)
ENSSSA1UGT006406	Bta01478	+	0.868	A(20)E	D(223)	E(311)
ENSSSA1UGT001578	Bta06327	+	0.730	R(21)E	D(230)	E(451)
ENSSSA1UGT021674	Bta09634	+	0.693	G(20)E	D(226)	E(294)
ENSSSA1UGT023876	Bta04421	+	0.440	A(21)N	D(230)	E(292)
Ssa04512	Bta14422	+	0.731	Q(31)S	D(236)	E(314)
Ssa04510	Bta14419	+	0.507	S(20)H	D(230)	E(297)
ENSSSA1UGT002084	Bta03112	+	0.770	C(23)S	D(235)	D(302)
ENSSSA1UGT011453	Bta04306	+	0.703	G(21)S	D(228)	E(299)
ENSSSA1UGT025983	Bta04298	+	0.578	A(36)L	D(241)	E(308)
ENSSSA1UGT026079	Bta04297	+	0.744	S(16)L	D(265)	E(333)
ENSSSA1UGT013702	Bta06458	+	0.730	G(20)G	E(229)	E(286)
Ssa12230	Bta15649	+	0.634	A(24)K	D(233)	E(300)
Ssa09110	Bta12624	+	0.753	S(25)K	D(227)	E(295)
ENSSSA1UGT017917	Bta04382	+	0.846	A(25)A	D(229)	E(297)

Supplementary Table 6.3 continued

Gene ID	Gene*	Signal peptide	D-Score**	Cleavage site	Catalytic site residue***	
					Nucleophile	Proton Donor
ENSSSA1UGT008267	Bta06059	+	0.785	A(19)F	D(224)	E(289)
ENSSSA1UGT004067	Bta11978	+	0.789	A(19)Q	D(232)	E(287)
ENSSSA1UGT003959	Bta11977	+	0.800	G(17)V	D(229)	E(284)
ENSSSA1UGT009050	Bta08425	+	0.647	G(31)I	D(242)	E(301)
ENSSSA1UGT005553	Bta08427	+	0.840	S(18)E	D(227)	E(288)
ENSSSA1UGT021138	Bta03991	+	0.865	H(20)A	D(226)	E(297)
ENSSSA1UGT023763	Bta05518	+	0.676	C(20)E	D(230)	E(298)
ENSSSA1UGT002066	Bta03818	+	0.854	C(24)R	D(236)	E(304)
ENSSSA1UGT010695	Bta05340	+	0.842	A(21)H	D(225)	E(292)
ENSSSA1UGT023957	Bta12678	+	0.823	A(23)D	D(225)	E(279)
ENSSSA1UGT002611	Bta05397	+	0.762	C(16)E	D(226)	E(294)
ENSSSA1UGT002066	Bta03818	+	0.530	C(28)R	D(240)	E(308)
ENSSSA1UGT004887	Bta12578	+	0.757	A(19)L	D(234)	E(302)
ENSSSA1UGT001811	Bta09696	+	0.786	A(25)Q	-	-
ENSSSA1UGT003669	Bta11358	-	0.253	-	D(182)	E(250)
ENSSSA1UGT010695	Bta05340	-	0.038	-	D(146)	E(213)
ENSSSA1UGT001534	Bta07453	-	0.119	-	D(65)	E(132)
ENSSSA1UGT028904	Bta06441	-	0.039	-	D(145)	E(213)

Supplementary Table 6.3 continued

Gene ID	Gene*	Signal peptide	D-Score**	Cleavage site	Catalytic site residue***	
					Nucleophile	Proton Donor
ENSSSA1UGT004229	Bta10152	-	0.094	-	D(271)	E(338)
ENSSSA1UGT026329	Bta07764	-	0.284	-	D(61)	E(122)
ENSSSA1UGT004300	Bta11979	-	0.215	-	D(274)	E(328)
ENSSSA1UGT003837	Bta11975	-	0.142	-	D(262)	E(318)
ENSSSA1UGT005585	Bta08427	-	0.104	-	-	-
ENSSSA1UGT010695	Bta05340	+	0.871	A(21)H	-	-
ENSSSA1UGT028904	Bta06441	-	0.239	-	D(252)	E(320)
ENSSSA1UGT010664	Bta05342	-	0.276	-	D(245)	E(312)
ENSSSA1UGT024045	Bta12682	-	0.119	-	D(288)	E(344)
ENSSSA1UGT024045	Bta12683	-	0.127	-	G(192)	D(248)
ENSSSA1UGT022820	Bta02810	-	0.184	-	-	-

* Corresponding gene in *Bemisia tabaci* (MEAM1) ** Discrimination score (D-score) for best-predicted signal peptide for predicted proteins of GH13 family

*** Catalytic site residue in each GH13 sequence. The position of the residue is indicated by the number in the parentheses

Presence or lack of signal peptide denoted as + or - respectively



Supplementary figure 6.1: Phylogenetic relationship of all the alpha-glucosidase in MED, MEAM1 and SSA1-SG1. Phylogenetic analysis was conducted using Bayesian methods implemented in Bayesian Evolutionary Analysis Sampling Trees (BEAST version 1.10.2).

Supplementary Table 6.4 continued

Ortholog 1	Ortholog 2	S-sub	N-sub	d_s	d_N	d_N/d_s	Selection pressure
Bta08427	ENSSSA1UGT005553	462.9	1280.1	5.40	4.06	0.75	Negative selection
BTA018902.1	Bta07764	72.9	1286.1	0.009	0.006	0.63	Negative selection
Bta04042	Bta04421	83.8	1377.2	1.03	0.32	0.31	Negative selection
Bta04421	ENSSSA1UGT023876	488.2	1242.8	35.93	3.99	0.11	Negative selection
BTA009935.1	Bta11358	67.2	1738.8	0.014	0.003	0.19	Negative selection
BTA030093.1	Bta12578	74.7	1722.3	0.19	0.05	0.26	Negative selection
Bta12578	ENSSSA1UGT004887	83.9	1713.1	0.12	0.03	0.23	Negative selection
BTA017167.1	Bta05340	76.3	1549.7	0.98	0.23	0.24	Negative selection
S_000197F_8848	Bta05340	75.0	1554.0	0.15	0.04	0.25	Negative selection
BTA01716.1	Bta13236	86.6	1698.4	1.27	0.26	0.21	Negative selection
BTA011261.1	Bta14422	94.5	1678.5	0.22	0.04	0.18	Negative selection
Ssa04512	Bta14422	453.9	1355.1	5.28	4.34	0.82	Negative selection
BTA017162.1	ENSSSA1UGT008267	65.2	1476.8	0.12	0.03	0.23	Negative selection
Bta03439	ENSSSA1UGT006139	438.2	1262.8	4.57	4.23	0.93	Negative selection
BTA002582	Bta11978	78.0	1680.0	0.05	0.02	0.40	Negative selection
Bta11978	ENSSSA1UGT004067	72.0	1677.0	0.25	0.06	0.24	Negative selection
BTA002583.1	Bta11977	71.2	1215.8	0.015	0.014	0.97	Negative selection
ENSSSA1UGT003959	Bta11977	54.7	1754.3	0.08	0.02	0.21	Negative selection
BTA004410.2	Bta11977	80.7	1707.3	1.17	0.35	0.29	Negative selection
BTA029782.1	Bta12678	106.3	1654.7	0.000	0.002	0	Neutral mutation
ENSSSA1UGT023957	Bta12678	97.6	1747.4	0.10	0.02	0.23	Negative selection
ENSSSA1UGT013702	Bta06458	532.2	1219.8	29.19	5.94	0.20	Negative selection

S-sub – Number of synonymous substitution sites
substitution sites

N-sub – Number of non-synonymous

d_N Non-synonymous substitution rate

d_s Synonymous substitution rate

Supplimentnary table 6.5: Quantification of amino acid molar fraction per gram of dry weight (gDW) required for the growth and survival of cassava whitefly (SSA1-SG1). These molar fractions were used in the formulation of the objection function

Cellular content of Protein	Amino acid	Amino acid quantity	Fractional contribution	Molecular weight monomer	Molecular weight of water	gmonomer/ Molprotein	gprotein/ olprotein	gmonomer/ gprotein	Molmonomer/ gDW	Mmolmonomer /gDW
0.706	Ala	1726608.13	0.055	71.079	18.015	2.927	94.914	0.031	0.000306	0.3063453
0.706	Cys	646104.13	0.021	103.145	18.015	1.757	94.914	0.019	0.000127	0.1267213
0.706	Asp	1591917.70	0.051	114.080	18.015	4.886	94.914	0.051	0.000319	0.3185518
0.706	Glu	2044502.37	0.065	128.107	18.015	7.191	94.914	0.076	0.000418	0.4175088
0.706	Phe	1421647.17	0.045	147.177	18.015	5.866	94.914	0.062	0.000296	0.2964647
0.706	Gly	1633612.67	0.052	57.052	18.015	2.038	94.914	0.021	0.000266	0.2656792
0.706	His	819461.36	0.026	137.142	18.015	3.119	94.914	0.033	0.000169	0.1691449
0.706	Ile	1725229.40	0.055	113.160	18.015	5.244	94.914	0.055	0.000345	0.3447024
0.706	Lys	2221089.11	0.071	129.183	18.015	7.888	94.914	0.083	0.000454	0.4541873
0.706	Leu	2861305.10	0.091	113.160	18.015	8.697	94.914	0.092	0.000572	0.5716914
0.706	Met	544967.19	0.017	131.199	18.015	1.970	94.914	0.021	0.000112	0.1117168
0.706	Asn	1575751.58	0.050	114.104	18.015	4.837	94.914	0.051	0.000315	0.3153293
0.706	Pro	1847872.94	0.059	97.117	18.015	4.670	94.914	0.049	0.000358	0.3576701
0.706	Gln	1169190.85	0.037	128.131	18.015	4.113	94.914	0.043	0.000239	0.2387683
0.706	Arg	1757998.58	0.056	157.197	18.015	7.816	94.914	0.082	0.00037	0.3698623
0.706	Ser	2897404.99	0.093	87.078	18.015	6.393	94.914	0.067	0.000546	0.5461054
0.706	Thr	1704271.69	0.054	101.105	18.015	4.524	94.914	0.048	0.000333	0.3328353
0.706	Val	1812503.73	0.058	99.133	18.015	4.697	94.914	0.049	0.000352	0.3524473
0.706	Trp	309510.50	0.010	186.214	18.015	1.663	94.914	0.018	6.64E-05	0.0664295
0.706	Tyr	995747.63	0.032	163.176	18.015	4.617	94.914	0.049	0.00021	0.2104863
Total		31306696.83				94.914				6.1726477

Supplementary table 6.6: List of essential reactions in a two-compartment metabolic model of SSA1-SG1

Reaction	Flux (mmol gDW ⁻¹ h ⁻¹)	Reaction	Flux (mmol gDW ⁻¹ h ⁻¹)
SSA1_Por_OCBT	0.111152	SSA1_Por_OMCDC	0.16531
SSA1_Por_METS	0.046402	SSA1_Por_PDH	0.165312
SSA1_Por_PRPPS	0.116569	SSA1_Por_AKGDH	0.209024
SSA1_Por_ATPPRT	0.049133	SSA1_Por_ATPS4rpp	3.89506
SSA1_Por_PRATPP	0.049133	SSA1_Por_CYTBO3_4pp	0.493353
SSA1_Por_PRAMPC	0.049133	SSA1_Por_NADH16pp	0.493353
SSA1_Por_PRMIC1	0.049133	SSA1_Por_PSY1	4.93E-06
SSA1_Por_IG3PS	0.049133	SSA1_Por_PSY2	4.93E-06
SSA1_Por_IGPDH	0.049133	SSA1_Por_PDS1	4.93E-06
SSA1_Por_HSTPT	0.049133	SSA1_Por_PDS2	4.93E-06
SSA1_Por_ASPK	0.260487	SSA1_Por_ZCAROTDH1	4.93E-06
SSA1_Por_ASAD	0.260487	SSA1_Por_ZCAROTDH2	2.47E-06
SSA1_Por_DHDPS	0.159891	SSA1_Por_LCYBZC	2.47E-06
SSA1_Por_THDPS	0.159891	SSA1_Por_HEMEOS	2.47E-06
SSA1_Por_SDPTA	0.159891	SSA1_Por_LIPOS	2.47E-06
SSA1_Por_SDPDS	0.159891	SSA1_Por_MTHFR2	0.046402
SSA1_Por_HSD	0.100596	SSA1_Por_DPR	2.91E-06
SSA1_Por_HSK	0.100596	SSA1_Por_CLPNS160pp	0.000318
SSA1_Por_THRS	0.100596	SSA1_Por_DM_LIPOPB	2.47E-06
SSA1_Por_DDPA	0.277102	SSA1_Por_DM_GLYC	0.000318
SSA1_Por_DHQS	0.277102	SSA1_Por_O2tex	0.246684
SSA1_Por_DHQTi	0.277102	SSA1_Por_H2Otex	-2.83424
SSA1_Por_SHK3Dr	0.277102	SSA1_Por_CO2tcy	1.77185
SSA1_Por_SHKK	0.277102	SSA1_Por_UP_2DHP	2.91E-06
SSA1_Por_PSCVT	0.277102	SSA1_Por_UP_PYR	0.992473
SSA1_Por_CHORS	0.277102	SSA1_Por_UP_2OBUT	0.113294
SSA1_Por_ANS	0.067436	SSA1_Por_UP_E4P	0.277102
SSA1_Por_ANPRT	0.067436	SSA1_Por_UP_PEP	0.554204
SSA1_Por_PRAli	0.067436	SSA1_Por_UP_R5P	0.116569
SSA1_Por_IGPS	0.067436	SSA1_Por_UP_PG160	0.000636
SSA1_Por_PPNDH	0.209666	SSA1_Por_UP_GGDP	9.87E-06
SSA1_Por_ACHBS	0.113294	SSA1_Por_UP_FRDP	2.47E-06
SSA1_Por_KARA2	0.113294	SSA1_Por_UP_OCTAPB	2.47E-06
SSA1_Por_DHAD2	0.113294	SSA1_Por_UP_PYDX5P	2.47E-06
SSA1_Por_ACLS	0.310706	SSA1_Por_UP_THMPP	2.47E-06
SSA1_Por_KARA1	0.310706	SSA1_Por_UP_FMNH2	2.47E-06
SSA1_Por_DHAD1	0.310706	SSA1_Por_UP_THDP	0.159891
SSA1_Por_IPPS	0.16531	SSA1_Por_UP_MLTHF	0.046402
SSA1_Por_IPPMIb	0.16531	SSA1_Por_UP_COA	0.049135
SSA1_Por_IPPMIa	0.16531	SSA1_Por_UP_PHEME	2.47E-06
SSA1_Por_IPMD	0.16531	SSA1_Por_UP_NADP	2.61E-05

Supplementary table 6.6: Continued

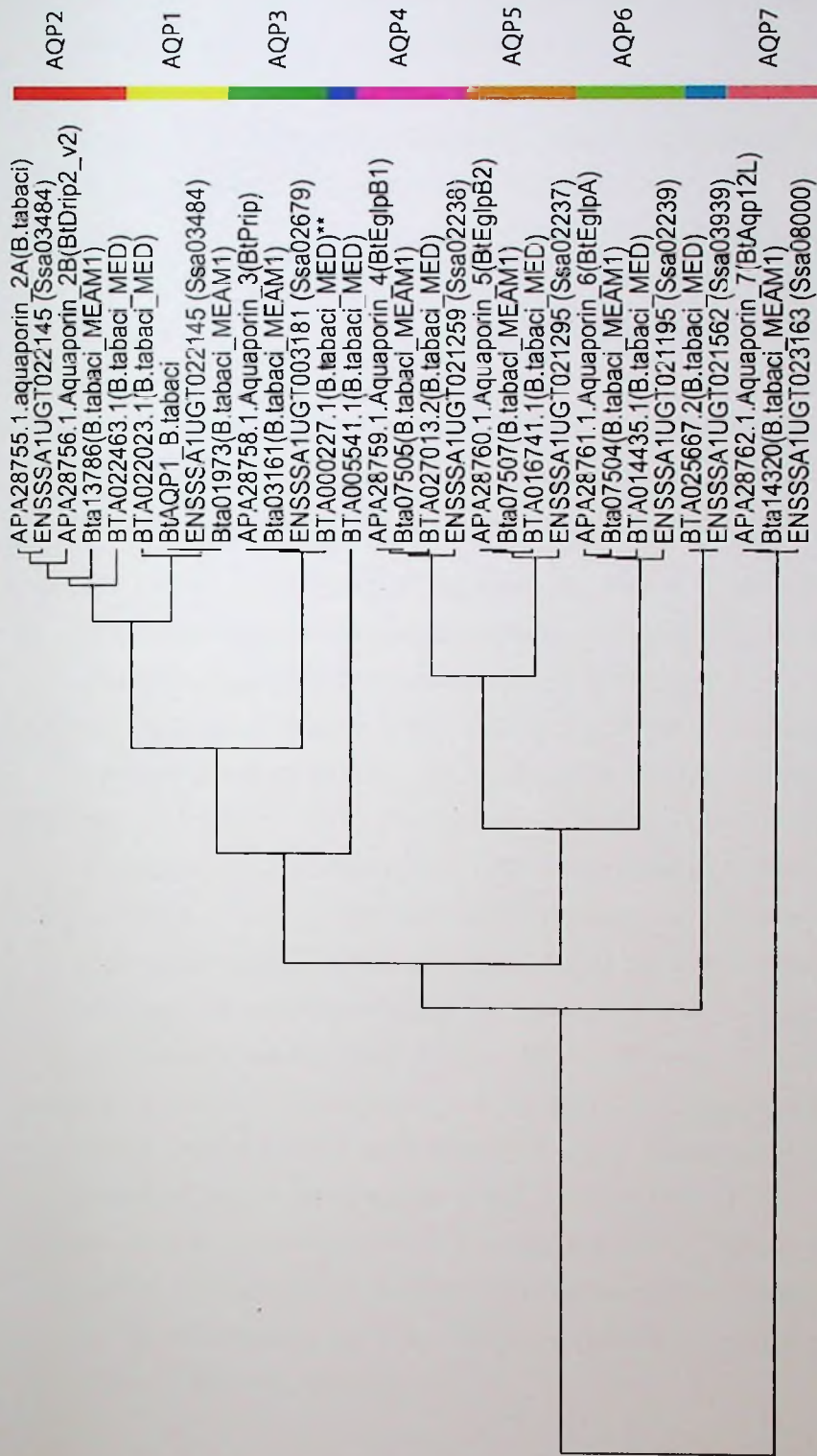
Reaction	Flux (mmol gDW ⁻¹ h ⁻¹)	Reaction	Flux (mmol gDW ⁻¹ h ⁻¹)
SSA1_Por_UP_NADH	0.000109	SSA1_Por_EF_26DAP-LL	0.159891
SSA1_Por_UP_DATP	0.001809	SSA1_Por_EF_ARGSUC	0.082103
SSA1_Por_UP_DCTP	0.000643	SSA1_Por_EF_HISP	0.049133
SSA1_Por_UP_DGTP	0.000643	SSA1_Por_EF_TRP-L	0.065903
SSA1_Por_UP_DTTP	0.001809	SSA1_Por_EF_THR-L	0.088147
SSA1_Por_UP_ADP	0.255983	SSA1_Por_EF_MET-L	0.039207
SSA1_Por_UP_CTP	0.006216	SSA1_Bt_ASPTA	0.817173
SSA1_Por_UP_GTP	0.010014	SSA1_Bt_ASNSI	0.135612
SSA1_Por_UP_UTP	0.006709	SSA1_Bt_DHDPRy	0.159891
SSA1_Por_UP_AMET	4.93E-06	SSA1_Bt_DAPE	0.159891
SSA1_Por_UP_HCYS-L	0.046402	SSA1_Bt_DAPDC	0.159891
SSA1_Por_UP_ORN	0.111152	SSA1_Bt_GLNS	0.842491
SSA1_Por_UP_ASN-L	0.013436	SSA1_Bt_CBPS	0.015376
SSA1_Por_UP_ASP-L	0.356567	SSA1_Bt_P5CR	998.112
SSA1_Por_UP_CYS-L	0.017419	SSA1_Bt_P5CDr	0.217739
SSA1_Por_UP_GLN-L	0.235647	SSA1_Bt_GSC	0.111152
SSA1_Por_UP_GLY	0.015866	SSA1_Bt_ORNTAr	0.111152
SSA1_Por_UP_PRO-L	0.007604	SSA1_Bt_ARGSL	0.111152
SSA1_Por_UP_SER-L	0.083706	SSA1_Bt_METAT	4.93E-06
SSA1_Por_UP_TYR-L	0.006368	SSA1_Bt_CYSTS	0.113294
SSA1_Por_UP_ARG-L	0.010652	SSA1_Bt_CYSTGL	0.113294
SSA1_Por_UP_HIS-L	0.004045	SSA1_Bt_HISTP	0.049133
SSA1_Por_UP_ILE-L	0.021221	SSA1_Bt_HISTD	0.049133
SSA1_Por_UP_LEU-L	0.02112	SSA1_Bt_ILETA	0.113294
SSA1_Por_UP_LYS-L	0.021646	SSA1_Bt_LEUTAi	0.16531
SSA1_Por_UP_PHE-L	0.007694	SSA1_Bt_PHETA1	0.209666
SSA1_Por_UP_VAL-L	0.01781	SSA1_Bt_PHETHPTOX	0.104398
SSA1_Por_EF_SUCC	0.159891	SSA1_Bt_VALTA	0.145393
SSA1_Por_EF_G3P	0.067436	SSA1_Bt_GTPCI	0.104398
SSA1_Por_EF_PPI	0.283837	SSA1_Bt_PTHPS	0.104398
SSA1_Por_EF_PANT-R	2.91E-06	SSA1_Bt_SPR	0.104398
SSA1_Por_EF_PHPYR	0.209666	SSA1_Bt_GHMT2r	0.238214
SSA1_Por_EF_3MOB	0.145396	SSA1_Bt_MTHFD	0.195203
SSA1_Por_EF_3MOP	0.113294	SSA1_Bt_MTHFC	0.195203
SSA1_Por_EF_4MOP	0.16531	SSA1_Bt_RBFK	2.47E-06
SSA1_Por_EF_SUCCOA	0.049133	SSA1_Bt_ANNA	0.000135
SSA1_Por_EF_DAD-5	4.93E-06	SSA1_Bt_DNGAL	0.000135
SSA1_Por_EF_AMP	0.198622	SSA1_Bt_NADK	2.61E-05
SSA1_Por_EF_THF	0.046402	SSA1_Bt_MOHMT	2.91E-06
SSA1_Por_EF_AICAR	0.049133	SSA1_Bt_PANTS	2.91E-06
SSA1_Por_EF_23DHDP	0.159891	SSA1_Bt_PNTK	2.91E-06

Supplementary table 6.6: Continued

Reaction	Flux (mmol gDW ⁻¹ h ⁻¹)	Reaction	Flux (mmol gDW ⁻¹ h ⁻¹)
SSA1_Bt_PPNCL2	2.91E-06	SSA1_Bt_GRTT	1.23E-05
SSA1_Bt_PPCDC	2.91E-06	SSA1_Bt_FRTT	9.87E-06
SSA1_Bt_PTPATi	2.91E-06	SSA1_Bt_HCO3E	1.10492
SSA1_Bt_DPCOAK	2.91E-06	SSA1_Bt_CAT	1.23E-06
SSA1_Bt_ALASm	1.97E-05	SSA1_Bt_ASPCT	0.015376
SSA1_Bt_PPBNGS	9.87E-06	SSA1_Bt_DHORTS	0.015376
SSA1_Bt_HMBS	2.47E-06	SSA1_Bt_DHORD2	0.015376
SSA1_Bt_UPP3S	2.47E-06	SSA1_Bt_ORPRT	0.015376
SSA1_Bt_UPPDC1	2.47E-06	SSA1_Bt_ORPDC	0.015376
SSA1_Bt_CPPPGO	2.47E-06	SSA1_Bt_UMPK	0.015376
SSA1_Bt_PPPGO	2.47E-06	SSA1_Bt_ATUD	0.013567
SSA1_Bt_FCLT	2.47E-06	SSA1_Bt_ATCM	2.91E-06
SSA1_Bt_GLUK	1.88671	SSA1_Bt_ATCD	-0.00064
SSA1_Bt_GAPD	3.92792	SSA1_Bt_ATDCD	0.000643
SSA1_Bt_PGK	3.92792	SSA1_Bt_ATDGD	0.000643
SSA1_Bt_PGM	3.92792	SSA1_Bt_ATGD	0.173724
SSA1_Bt_ENO	3.92792	SSA1_Bt_NDPK	0.001809
SSA1_Bt_PYK	3.37372	SSA1_Bt_ATDTD	0.001809
SSA1_Bt_PDH	0.666063	SSA1_Bt_PRDPAR	0.125234
SSA1_Bt_PCt	0.978394	SSA1_Bt_PPRGL	0.125234
SSA1_Bt_CS	0.665922	SSA1_Bt_FPGFT	0.125234
SSA1_Bt_ACONTa	0.665922	SSA1_Bt_PRFGS	0.125234
SSA1_Bt_ACONTb	0.665922	SSA1_Bt_PRFGCL	0.125234
SSA1_Bt_SUCOAS	0.049113	SSA1_Bt_PRAIC	0.125234
SSA1_Bt_SUCDi	0.209004	SSA1_Bt_PRAIS	0.125234
SSA1_Bt_FUM	0.504702	SSA1_Bt_AIAL	0.125234
SSA1_Bt_MDH	0.504702	SSA1_Bt_FPAIF	0.174367
SSA1_Bt_PGL	0.531102	SSA1_Bt_ICH	0.174367
SSA1_Bt_RPI	0.440854	SSA1_Bt_IMPD	0.115055
SSA1_Bt_CYOO6m	1.66454	SSA1_Bt_GMPS(glu)	0.115055
SSA1_Bt_CYOR(q9)m	3.32908	SSA1_Bt_AGPT	0.115055
SSA1_Bt_TRDR	0.004903	SSA1_Bt_DGOTO	0.000643
SSA1_Bt_PPA	0.565679	SSA1_Bt_ADSS	0.059312
SSA1_Bt_ACACTlr	4.69E-05	SSA1_Bt_AAL(fum)	0.059312
SSA1_Bt_HMGCOAS	4.69E-05	SSA1_Bt_ATAM	0.537788
SSA1_Bt_HMGCOARx	4.69E-05	SSA1_Bt_DAOTO	0.001809
SSA1_Bt_MEVK1x	4.69E-05	SSA1_Bt_DCDT	0.000643
SSA1_Bt_PMEVKx	4.69E-05	SSA1_Bt_TMDS	0.001809
SSA1_Bt_DPMVDx	4.69E-05	SSA1_Bt_DTMPK	0.001809
SSA1_Bt_IPDDI	1.23E-05	SSA1_Bt_DM_ALA-B	2.91E-06
SSA1_Bt_DMATT	1.23E-05	SSA1_Bt_DM_PG160	0.000636

Supplementary table 6.6: Continued

Reaction	Flux (mmol gDW ⁻¹ h ⁻¹)	
SSA1_Bt_DM_OCTAPB	2.47E-06	
SSA1_Bt_DM_DHBPT	0.104398	
SSA1_Bt_O2tex	2.01563	
SSA1_Bt_Pitex	-0.34944	
SSA1_Bt_NH4tex	-3.69554	
SSA1_Bt_UP_hcys-L	0.159695	
SSA1_Bt_UP_thmpp	2.47E-06	
SSA1_Bt_UP_ribflv	2.47E-06	
SSA1_Bt_UP_pdx5p	2.47E-06	
SSA1_Bt_UP_nicrnt	0.000135	
SSA1_Bt_EF_dad-5	4.93E-06	
SSA1_Bt_EF_pppi	0.104398	
Bt_GROW	0.312566	
EX_pppi(e)	0.104398	
EX_dad-5(e)	4.93E-06	
EX_o2(e)	-2.01563	
EX_h(e)	3.43033	
EX_pi(e)	-0.34944	
EX_nh4(e)	-3.69554	
EX_hcys-L(e)	-0.1597	
EX_thmpp(e)	-2.47E-06	
EX_ribflv(e)	-2.47E-06	
EX_pdx5p(e)	-2.47E-06	
EX_nicrnt(e)	-0.00013	



Supplementary figure 6.2: Phylogenetic relationship of all aquaporin genes of MED, MEAM1 and SSA1-SG1 *B. tabaci* species and experimentally validated aquaporins (Van Ekert *et al.*, 2016). Protein sequence with ** was a partial sequence. Phylogenetic analysis was conducted using Bayesian methods implemented in Bayesian Evolutionary Analysis Sampling Trees (BEAST version 1.10.2).

REFERENCES

- Abad, M.C., Binderup, K., Rios-Steiner, J., Arni, R.K., Preiss, J., Geiger, J.H. (2002) The X-ray crystallographic structure of Escherichia coli branching enzyme. *Journal of Biological Chemistry*, **277**(44), 42164-42170.
- Abdullahi, I., Winter, S., Atiri, G.I., Thottappilly, G. (2003) Molecular characterization of whitefly, Bemisia tabaci (Hemiptera : Aleyrodidae) populations infesting cassava. *Bulletin of Entomological Research*, **93**(2), 97-106.
- Alemandri, V., De Barro, P., Bejerman, N., Arguecello Caro, E.B., Dumon, A.D., Mattio, M.F., Rodriguez, S.M., Truol, G. (2012) Species Within the Bemisia tabaci (Hemiptera: Aleyrodidae) Complex in Soybean and Bean Crops in Argentina. *Journal of Economic Entomology*, **105**(1), 48-53.
- Alicai, T., Omongo, C.A., Maruthi, M.N., Hillocks, R.J., Baguma, Y., Kawuki, R., Bua, A., Otim-Nape, G.W., Colvin, J. (2007) Re-emergence of cassava brown streak disease in Uganda. *Plant Disease*, **91**(1), 24-29.
- Allen, G.C., Hall, G., Michalowski, S., Newman, W., Spiker, S., Weissinger, A.K., Thompson, W.F. (1996) High-level transgene expression in plant cells: Effects of a strong scaffold attachment region from tobacco. *Plant Cell*, **8**(5), 899-913.
- Allen, G.C., Spiker, S., Thompson, W.F. (2000) Use of matrix attachment regions (MARs) to minimize transgene silencing. *Plant Molecular Biology*, **43**(2-3), 361-376.
- Allen, M.L. and Walker, W.B. (2012) Saliva of Lygus lineolaris digests double stranded ribonucleic acids. *Journal of Insect Physiology*, **58**(3), 391-396.
- Alon, M., Alon, F., Nauen, R., Morin, S. (2008) Organophosphates' resistance in the B-biotype of Bemisia tabaci (Hemiptera: Aleyrodidae) is associated with a point mutation in an ace1-type acetylcholinesterase and overexpression of carboxylesterase. *Insect Biochemistry and Molecular Biology*, **38**(10), 940-949.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J.H., Zhang, Z., Miller, W., Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, **25**(17), 3389-3402.
- Andersson, S.G.E., Zomorodipour, A., Andersson, J.O., Sicheritz-Ponten, T., Alsmark, U.C.M., Podowski, R.M., Naslund, A.K., Eriksson, A.S., Winkler, H.H., Kurland, C.G. (1998) The genome sequence of Rickettsia prowazekii and the origin of mitochondria. *Nature*, **396**(6707), 133-140.

- Andersson, S.G.E. and Kurland, C.G. (1998) Reductive evolution of resident genomes. *Trends in Microbiology*, **6**(7), 263-268.
- Ankrah, N.Y.D., Luan, J., Douglas, A.E. (2017) Cooperative Metabolism in a Three-Partner Insect-Bacterial Symbiosis Revealed by Metabolic Modeling. *Journal of Bacteriology*, **199**(15).
- Ariyo, O.A., Dixon, A.G.O., Atiri, G.I. (2005) Whitefly *Bemisia tabaci* (Homoptera : Aleyrodidae) infestation on cassava genotypes grown at different ecozones in Nigeria. *Journal of Economic Entomology*, **98**(2), 611-617.
- Ashford, D.A., Smith, W.A., Douglas, A.E. (2000) Living on a high sugar diet: the fate of sucrose ingested by a phloem-feeding insect, the pea aphid *Acyrtosiphon pisum*. *Journal of Insect Physiology*, **46**(3), 335-341.
- Asiimwe, P., Ecaat, J.S., Guershon, M., Kyamanywa, S., Gerling, D., Legg, J.P. (2007a) Evaluation of *Serangium* n. sp (Col., Coccinellidae), a predator of *Bemisia tabaci* (Hom., Aleyrodidae) on cassava. *Journal of Applied Entomology*, **131**(2), 76-80.
- Asiimwe, P., Ecaat, J.S., Otim, M., Gerling, D., Kyamanywa, S., Legg, J.P. (2007b) Life-table analysis of mortality factors affecting populations of *Bemisia tabaci* on cassava in Uganda. *Entomologia Experimentalis Et Applicata*, **122**(1), 37-44.
- Asiimwe, P., Kelly, S.E., Hunter, M.S. (2014) Symbiont infection affects whitefly dynamics in the field. *Basic and Applied Ecology*, **15**(6), 507-515.
- Ayre, B.G., Keller, F., Turgeon, R. (2003) Symplastic continuity between companion cells and the translocation stream: Long-distance transport is controlled by retention and retrieval mechanisms in the phloem. *Plant Physiology*, **131**(4), 1518-1528.
- Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., Zagnitko, O. (2008) The RAST server: Rapid annotations using subsystems technology. *BMC Genomics*, **9**.
- Baum, J.A. (2008) AGRO 121-Control of coleopteran insect pests through RNA interference. *Abstracts of Papers of the American Chemical Society*, **236**.
- Baum, J.A., Bogaert, T., Clinton, W., Heck, G.R., Feldmann, P., Ilagan, O., Johnson, S., Plaetinck, G., Munyikwa, T., Pleau, M., Vaughn, T., Roberts, J. (2007) Control of coleopteran insect pests through RNA interference. *Nature Biotechnology*, **25**(11), 1322-1326.

- Baumann, P. (2005) Biology of bacteriocyte-associated endosymbionts of plant sap-sucking insects. *Annual Review of Microbiology*, **59**, 155-189.
- Bedford, I.D., Briddon, R.W., Brown, J.K., Rosell, R.C., Markham, P.G. (1994) Geminivirus-transmission and biological characterization of *Bemisia tabaci* (gennadius) biotypes from different geographic regions. *Annals of Applied Biology*, **125**(2), 311-325.
- Belles, X. (2010) Beyond Drosophila: RNAi In Vivo and Functional Genomics in Insects. *Annual Review of Entomology*, **55**, 111-128.
- Bellotti, A.C. and Arias, B. (2001) Host plant resistance to whiteflies with emphasis on cassava as a case study. *Crop Protection*, **20**(9), 813-823.
- Bellows, T.S., Perring, T.M., Gill, R.J., Headrick, D.H. (1994) Description of a species of *Bemisia* (homoptera, aleyrodidae). *Annals of the Entomological Society of America*, **87**(2), 195-206.
- Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate - a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-Methodological*, **57**(1), 289-300.
- Berry, S.D., Fondong, V.N., Rey, C., Rogan, D., Fauquet, C.M., Brown, J.K. (2004) Molecular evidence for five distinct *Bemisia tabaci* (Homoptera : Aleyrodidae) geographic haplotypes associated with cassava plants in sub-Saharan Africa. *Annals of the Entomological Society of America*, **97**(4), 852-859.
- Bi, H.L., Xu, J., Tan, A.J., Huang, Y.P. (2016) CRISPR/Cas9-mediated targeted gene mutagenesis in *Spodoptera litura*. *Insect Science*, **23**(3), 469-477.
- Bing, X.L., Yang, J., Zchori-Fein, E., Wang, X.W., Liu, S.S. (2013) Characterization of a newly discovered symbiont of the whitefly *Bemisia tabaci* (Hemiptera: Aleyrodidae). *Applied and Environmental Microbiology*, **79**(2), 569-575.
- Bird, J (1957). A whitefly-transmitted mosaic of *Jatropha gossypifolia*, 22nd edition. Agricultural Experiment Station, University of Puerto Rico.
- Bleeker, P.M., Mirabella, R., Diergaarde, P.J., VanDoorn, A., Tissier, A., Kant, M.R., Prins, M., de Vos, M., Haring, M.A., Schuurink, R.C. (2012) Improved herbivore resistance in cultivated tomato with the sesquiterpene biosynthetic pathway from a wild relative. *Proceedings of the National Academy of Sciences of the United States of America*, **109**(49), 20124-20129.
- Boykin, L.M. (2014) *Bemisia tabaci* nomenclature: lessons learned. *Pest Management Science*, **70**(10), 1454-1459.

- Boykin, L.M., Armstrong, K.F., Kubatko, L., De Barro, P. (2012) Species delimitation and global biosecurity. *Evolutionary bioinformatics online*, **8**, 1-37.
- Boykin, L.M., Bell, C.D., Evans, G., Small, I., De Barro, P.J. (2013) Is agriculture driving the diversification of the *Bemisia tabaci* species complex (Hemiptera: Sternorrhyncha: Aleyrodidae)? Dating, diversification and biogeographic evidence revealed. *BMC Evolutionary Biology*, **13**.
- Boykin, L.M., Kinene, T., Wainaina, J.M., Savill, A., Seal, S., Mugerwa, H., Macfadyen, S., Tay, W.T., De Barro, P., Kubatko, L., Alicai, T., Omongo, C.A., Tairo, F., Ndunguru, J., Sseruwagi, P. (2018) Review and guide to a future naming system of African *Bemisia tabaci* species. *Systematic Entomology*, **43**(3), 427-433.
- Boykin, L.M., Shatters, R.G., Jr., Rosell, R.C., McKenzie, C.L., Bagnall, R.A., De Barro, P., Frohlich, D.R. (2007) Global relationships of *Bemisia tabaci* (Hemiptera : Aleyrodidae) revealed using Bayesian analysis of mitochondrial COI DNA sequences. *Molecular Phylogenetics and Evolution*, **44**(3), 1306-1319.
- Bridges, R.J., Natale, N.R., Patel, S.A. (2012) System xc- cystine/glutamate antiporter: an update on molecular pharmacology and roles within the CNS. *British Journal of Pharmacology*, **165**(1), 20-34.
- Brosnan, J.T. and Brosnan, M.E. (2006) Branched-chain amino acids: Enzyme and substrate regulation. *Journal of Nutrition*, **136**(1), 207S-211S.
- Brown, J.K. (1994) Current status of *Bemisia tabaci* as a plant pest and virus vector in agroecosystems worldwide. FAO (Food and Agriculture Organization of the United Nations) *Plant Protection Bulletin*, **42**(1-2), 3-32.
- Brown, J.K. and Bird, J. (1992) Whitefly-transmitted geminiviruses and associated disorders in the america and the caribbean basin. *Plant Disease*, **76**(3), 220-225.
- Brown, J.K. and Coats, S. (1992) Biotypic characterization of *Bemisia tabaci* populations based on esterase profiles, DNA fingerprinting, virus transmission and bioassay to key host plant species. *Phytopathology*, **82**(10), 1104-1104.
- Brumin, M., Kontsedalov, S., Ghanim, M. (2011) Rickettsia influences thermotolerance in the whitefly *Bemisia tabaci* B biotype. *Insect Science*, **18**(1), 57-66.
- Buchholz, F., Angrand, P.O., Stewart, A.F. (1998) Improved properties of FLP recombinase evolved by cycling mutagenesis. *Nature Biotechnology*, **16**(7), 657-662.
- Burban, C., Fishpool, L.D.C., Fauquet, C., Fargette, D., Thouvenel, J.C. (1992) Host-associated biotypes within west african populations of the whitefly *Bemisia-tabaci*

- (genn), (hom, aleyrodidae). *Journal of Applied Entomology-Zeitschrift Fur Angewandte Entomologie*, **113**(4), 416-423.
- Burns A, Gleadow R, Cliff J, Zacarias A, Cavagnaro T (2010). Cassava: The drought, war and famine crop in a changing world. *Sustain*, **2**: 3572-3607.
- Butaye, K.M.J., Cammue, B.P.A., Delaure, S.L., De Bolle, M.F.C. (2005) Approaches to minimize variation of transgene expression in plants. *Molecular Breeding*, **16**(1), 79-91.
- Byrne, D.N., Hendrix, D.L., Williams, L.H. (2003) Presence of trehalulose and other oligosaccharides in hemipteran honeydew, particularly Aleyrodidae. *Physiological Entomology*, **28**(2), 144-149.
- Byrne, F.J. and Devonshire, A.L. (1996) Biochemical evidence of haplodiploidy in the whitefly *Bemisia tabaci*. *Biochemical Genetics*, **34**(3-4), 93-107.
- Byrne, F.J., Gorman, K.J., Cahill, M., Denholm, I., Devonshire, A.L. (2000) The role of B-type esterases in conferring insecticide resistance in the tobacco whitefly, *Bemisia tabaci* (Genn). *Pest Management Science*, **56**(10), 867-874.
- Campbell, B.C., Duffus, J.E., Baumann, P. (1993) Determining whitefly species. *Science*, **261**(5126), 1333-1333.
- Campbell, E.M., Ball, A., Hoppler, S., Bowman, A.S. (2008) Invertebrate aquaporins: a review. *Journal of Comparative Physiology B-Biochemical Systemic and Environmental Physiology*, **178**(8), 935-955.
- Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Research*, **37**, D233-D238.
- Castagna, M., Shayakul, C., Trotti, D., Sacchi, V.F., Harvey, W.R., Hediger, M.A. (1997) Molecular characteristics of mammalian and insect amino acid transporters: Implications for amino acid homeostasis. *Journal of Experimental Biology*, **200**(2), 269-286.
- Chang, A., Schomburg, I., Placzek, S., Jeske, L., Ulbrich, M., Xiao, M., Sensen, C.W., Schomburg, D. (2015) BRENDA in 2015: exciting developments in its 25th year of existence. *Nucleic Acids Research*, **43**(D1), D439-D446.
- Chang, C.L. (2004) Effect of amino acids on larvae and adults of *Ceratitidis capitata* (Diptera : Tephritidae). *Annals of the Entomological Society of America*, **97**(3), 529-535.

- Chen, J., Zhang, D., Yao, Q., Zhang, J., Dong, X., Tian, H., Zhang, W. (2010) Feeding-based RNA interference of a trehalose phosphate synthase gene in the brown planthopper, *Nilaparvata lugens*. *Insect Molecular Biology*, **19**(6), 777-786.
- Chen, X., Li, L., Hu, Q., Zhang, B., Wu, W., Jin, F., Jiang, J. (2015) Expression of dsRNA in recombinant *Isaria fumosorosea* strain targets the TLR7 gene in *Bemisia tabaci*. *Bmc Biotechnology*, **15**.
- Chen, W.B., Hasegawa, D.K., Kaur, N., Kliot, A., Pinheiro, P.V., Luan, J.B., Stensmyr, M.C., Zheng, Y., Liu, W.L., Sun, H.H., Xu, Y.M., Luo, Y., Kruse, A., Yang, X.W., Kontsedalov, S., Lebedev, G., Fisher, T.W., Nelson, D.R., Hunter, W.B., Brown, J.K., Jander, G., Cilia, M., Douglas, A.E., Ghanim, M., Simmons, A.M., Wintermantel, W.M., Ling, K.S., Fei, Z.J. (2016) The draft genome of whitefly *Bemisia tabaci* MEAM1, a global crop pest, provides novel insights into virus transmission, host adaptation, and insecticide resistance. *BMC Biology*, **14**, 110.
- Chen W, Wosula E.N, Hasegawa D.K, Casinga C, Shirima R.R, Fiaboe K.K.M, Hanna R, Fosto A, Goergen G, Tamò M, Mahuku G, Murithi H.M, Tripathi L, Mware B, Kumar L.P, Ntawuruhunga P, Moyo C, Yomeni M, Boahen S, Edet M, Awoyale W, Wintermantel W.M, Ling KS, Legg J.P, Fei Z (2019) Genome of the African cassava whitefly *Bemisia tabaci* and distribution and genetic diversity of cassava-colonizing whiteflies in Africa. *Insect Biochemical Molecular Biology*. 110:112-120.
- Chiba, S. (1997) Molecular mechanism in alpha-glucosidase and glucoamylase. *Bioscience Biotechnology and Biochemistry*, **61**(8), 1233-1239.
- Chiel, E., Zchori-Fein, E., Inbar, M., Gottlieb, Y., Adachi-Hagimori, T., Kelly, S.E., Asplen, M.K., Hunter, M.S. (2009) Almost There: Transmission routes of bacterial symbionts between Trophic Levels. *Plos One*, **4**(3).
- Chinn, A.M. and Comai, L. (1996) The heat shock cognate 80 gene of tomato is flanked by matrix attachment regions. *Plant Molecular Biology*, **32**(5), 959-968.
- Chowda-Reddy, R.V., Kirankumar, M., Seal, S.E., Muniyappa, V., Valand, G.B., Govindappa, M.R., Colvin, J. (2012) *Bemisia tabaci* phylogenetic groups in India and the relative transmission efficacy of tomato leaf curl Bangalore virus by an indigenous and an exotic population. *Journal of Integrative Agriculture*, **11**(2), 235-248.
- Christiaens, O., Swevers, L., Smagghe, G. (2014) DsRNA degradation in the pea aphid (*Acyrtosiphon pisum*) associated with lack of response in RNAi feeding and injection assay. *Peptides*, **53**, 307-314.

- Cicero, J.M., Hiebert, E., Webb, S.E. (1995) The alimentary canal of *Bemisia-tabaci* and *trialeurodes abutilonea* (Homoptera, Sternorrhynchi) - histology, ultrastructure and correlations to function. *Zoomorphology*, **115**(1), 31-39.
- Cohen, E. (2013) Water homeostasis and osmoregulation as targets in the control of insect pests. Target receptors in the control of insect pests: *Pt I*, **44**, 1-61.
- Colvin, J., Omongo, C.A., Maruthi, M.N., Otim-Nape, G.W., Thresh, J.M. (2004) Dual begomovirus infections and high *Bemisia tabaci* populations: two factors driving the spread of a cassava mosaic disease pandemic. *Plant Pathology*, **53**(5), 577-584.
- Colvin, J., Omongo, C.A., Govindappa, M.R., Stevenson, P.C., Maruthi, M.N., Gibson, G., Seal, S.E., Muniyappa, V. (2006) Host-plant viral infection effects on arthropod-vector population growth, development and behaviour: Management and epidemiological implications. *Plant Virus Epidemiology*, **67**, 419-452.
- Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., Robles, M. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, **21**(18), 3674-3676.
- Costa, H.S. and Brown, J.K. (1991) Variation in biological characteristics and esterase patterns among populations of *Bemisia-tabaci* and the association of one population with silverleaf symptom induction. *Entomologia experimentalis et applicata*, **61**(3), 211-219.
- Cristofolletti, P.T., Ribeiro, A.F., Deraison, C., Rahbe, Y., Terra, W.R. (2003) Midgut adaptation and digestive enzyme distribution in a phloem feeding insect, the pea aphid *Acyrtosiphon pisum*. *Journal of Insect Physiology*, **49**(1), 11-24.
- De Barro, P.J., Liu, S.-S., Boykin, L.M., Dinsdale, A.B. (2011) *Bemisia tabaci*: A statement of species status. *Annual Review of Entomology*, Vol 56, **56**, 1-19.
- De Barro, P.J. and Driver, F. (1997) Use of RAPD PCR to distinguish the B biotype from other biotypes of *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae). *Australian Journal of Entomology*, **36**, 149-152.
- de Groot, B.L. and Grubmuller, H. (2001) Water permeation across biological membranes: Mechanism and dynamics of aquaporin-1 and GlpF. *Science*, **294**(5550), 2353-2357.
- De la Rua, P., Simon, B., Cifuentes, D., Martinez-Mora, C., Cenis, J.L. (2006) New insights into the mitochondrial phylogeny of the whitefly *Bemisia tabaci* (Hemiptera : Aleyrodidae) in the Mediterranean Basin. *Journal of Zoological Systematics and Evolutionary Research*, **44**(1), 25-33.

- De Queiroz, K. (2007) Species concepts and species delimitation. *Systematic Biology*, **56**(6), 879-886
- Degnan, P.H., Yu, Y., Sisneros, N., Wing, R.A., Moran, N.A. (2009) *Hamiltonella defensa*, genome evolution of protective bacterial endosymbiont from pathogenic ancestors. *Proceedings of the National Academy of Sciences of the United States of America*, **106**(22), 9063-9068.
- Denholm, I., Gorman, K., Williamson, M. (2008) Insecticide resistance in *Bemisia tabaci*: A global perspective. *Journal of Insect Science*, **8**, 16-16.
- Deves, R., Chavez, P., Boyd, C.A.R. (1992) Identification of a new transport-system (y+L) in human erythrocytes that recognizes lysine and leucine with high-affinity. *Journal of Physiology-London*, **454**, 491-501.
- Dinsdale, A., Cook, L., Riginos, C., Buckley, Y.M., De Barro, P. (2010) Refined global analysis of *Bemisia tabaci* (Hemiptera: Sternorrhyncha: Aleyrodidae: Aleyrodidae) Mitochondrial Cytochrome Oxidase 1 to identify species level genetic boundaries. *Annals of the Entomological Society of America*, **103**(2), 196-208.
- Doi, M., Yamaoka, I., Fukunaga, T., Nakayama, M. (2003) Isoleucine, a potent plasma glucose-lowering amino acid, stimulates glucose uptake in C2C12 myotubes. *Biochemical and Biophysical Research Communications*, **312**(4), 1111-1117.
- Dong, Y. and Friedrich, M. (2005) Nymphal RNAi: systemic RNAi mediated gene knockdown in juvenile grasshopper. *BMC Biotechnology*, **5**.
- Douglas, A.E. (1998) Nutritional interactions in insect-microbial symbioses: Aphids and their symbiotic bacteria Buchnera. *Annual Review of Entomology*, **43**, 17-37.
- Douglas, A.E., Minto, L.B., Wilkinson, T.L. (2001) Quantifying nutrient production by the microbial symbionts in an aphid. *Journal of Experimental Biology*, **204**(2), 349-358.
- Douglas, A.E. (2006) Phloem-sap feeding by animals: problems and solutions. *Journal of Experimental Botany*, **57**(4), 747-754.
- Douglas, A.E. (2014) Molecular dissection of nutrient exchange at the insect-microbial interface. *Current Opinion in Insect Science*, **4**, 23-28.
- Douglas, A.E. and Prosser, W.A. (1992) Synthesis of the essential amino-acid tryptophan in the pea aphid (*Acyrtosiphon-pisum*) symbiosis. *Journal of Insect Physiology*, **38**(8), 565-568.
- Douglas, A.E. (2016) How multi- partner endosymbioses function. *Nature Reviews Microbiology*, **14**(12), 731-743.

- Drake, L.L., Rodriguez, S.D., Hansen, I.A. (2015) Functional characterization of aquaporins and aquaglyceroporins of the yellow fever mosquito, *Aedes aegypti*. *Scientific Reports*, **5**.
- Duncan, R.P., Husnik, F., Van Leuven, J.T., Gilbert, D.G., Davalos, L.M., McCutcheon, J.P., Wilson, A.C.C. (2014) Dynamic recruitment of amino acid transporters to the insect/symbiont interface. *Molecular Ecology*, **23**(6), 1608-1623.
- Dunoyer, P., Himber, C., Ruiz-Ferrer, V., Alioua, A., Voinnet, O. (2007) Intra- and intercellular RNA interference in *Arabidopsis thaliana* requires components of the microRNA and heterochromatic silencing pathways. *Nature Genetics*, **39**(7), 848-856.
- Dye, J.F., Vause, S., Johnston, T., Clark, P., Firth, J.A., D'Souza, S.W., Sibley, C.P., Glazier, J.D. (2004) Characterization of cationic amino acid transporters and expression of endothelial nitric oxide synthase in human placental microvascular endothelial cells. *Faseb Journal*, **18**(1), 125-127.
- Eddy, S.R. (2011) Accelerated Profile HMM Searches. *Plos Computational Biology*, **7**(10).
- Elbashir, S.M., Lendeckel, W., Tuschl, T. (2001) RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes & Development*, **15**(2), 188-200.
- El-Sharkawy, M.A. (2005) How can calibrated research-based models be improved for use as a tool in identifying genes controlling crop tolerance to environmental stresses in the era of genomics - from an experimentalist's perspective. *Photosynthetica*, **43**(2), 161-176.
- Engelstadter, J. and Hurst, G.D.D. (2009) The ecology and evolution of microbes that manipulate host reproduction. *Annual Review of Ecology Evolution and Systematics*, **40**, 127-149.
- Esterhuizen, L.L., Mabasa, K.G., van Heerden, S.W., Czosnek, H., Brown, J.K., van Heerden, H., Rey, M.E.C. (2013) Genetic identification of members of the *Bemisia tabaci* cryptic species complex from South Africa reveals native and introduced haplotypes. *Journal of Applied Entomology*, **137**(1-2), 122-135.
- Fauquet, C. and Fargette, D. (1990) African cassava mosaic-virus - etiology, epidemiology, and control. *Plant Disease*, **74**(6), 404-411.
- Feist, A.M., Palsson, B.O. (2010) The biomass objective function. *Current Opinion Microbiology* **13**(3) 344-9.

- Finn, R.N., Chauvigne, F., Stavang, J.A., Belles, X., Cerda, J. (2015) Insect glycerol transporters evolved by functional co-option and gene replacement. *Nature Communications*, **6**.
- Finnegan, J. and McElroy, D. (1994) Transgene inactivation - plants fight back. *Bio-Technology*, **12**(9), 883-888.
- Firdaus, S., Vosman, B., Hidayati, N., Supena, E.D.J., Visser, R.G.F., van Heusden, A.W. (2013) The *Bemisia tabaci* species complex: Additions from different parts of the world. *Insect Science*, **20**(6), 723-733.
- Fire, A., Xu, S.Q., Montgomery, M.K., Kostas, S.A., Driver, S.E., Mello, C.C. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, **391**(6669), 806-811.
- Frewin, A.J., Scott-Dupree, C., Murphy, G., Hanner, R. (2014) Demographic Trends in Mixed *Bemisia tabaci* (Hemiptera: Aleyrodidae) Cryptic Species Populations in Commercial Poinsettia Under Biological Control- and Insecticide-Based Management. *Journal of Economic Entomology*, **107**(3), 1150-1155.
- Frohlich, D.R., Torres-Jerez, I., Bedford, I.D., Markham, P.G., Brown, J.K. (1999) A phylogeographical analysis of the *Bemisia tabaci* species complex based on mitochondrial DNA markers. *Molecular Ecology*, **8**(10), 1683-1691.
- Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D., Bairoch, A. (2003) ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*, **31**(13), 3784-3788.
- Gelman, D.B., Gerling, D., Blackburn, M.B., Hu, J.S. (2005) Host-parasite interactions between whiteflies and their parasitoids. *Archives of Insect Biochemistry and Physiology*, **60**(4), 209-222.
- Gennadius, P. (1889). Disease of the tobacco plantations in the Trikonía. The aleurodid of tobacco. *Ellenike Georgia* 5:1-3.
- Gerling, D., Orion, T., Delarea, Y. (1990) Eretmocerus penetration and immature development - a novel-approach to overcome host immunity. *Archives of Insect Biochemistry and Physiology*, **13**(3-4), 247-253.
- Gernhard, T. (2008) The conditioned reconstructed process. *Journal of Theoretical Biology*, **253**(4), 769-778.
- Ghanim, M. and Kontsedalov, S. (2009) Susceptibility to insecticides in the Q biotype of *Bemisia tabaci* is correlated with bacterial symbiont densities. *Pest Management Science*, **65**(9), 939-942.

- Ghanim, M., Kontsedalov, S., Czosnek, H. (2007) Tissue-specific gene silencing by RNA interference in the whitefly *Bemisia tabaci* (Gennadius). *Insect Biochemistry and Molecular Biology*, **37**(7), 732-738.
- Ghanim, M., Rosell, R.C., Campbell, L.R., Czosnek, H., Brown, J.K., Ullman, D.E. (2001) Digestive, salivary and reproductive organs of *Bemisia tabaci* (Gennadius) (Hemiptera : Aleyrodidae) B type. *Journal of Morphology*, **248**(1), 22-40.
- Ghosh, S., Bouvaine, S., Maruthi, M.N. (2015) Prevalence and genetic diversity of endosymbiotic bacteria infecting cassava whiteflies in Africa. *BMC Microbiology*, **15**.
- Gibson, R.W., Legg, J.P., OtimNape, G.W. (1996) Unusually severe symptoms are a characteristic of the current epidemic of mosaic virus disease of cassava in Uganda. *Annals of Applied Biology*, **128**(3), 479-490.
- Gilbertson, L. (2003) Cre-lox recombination: Cre-ative tools for plant biotechnology. *Trends in Biotechnology*, **21**(12), 550-555.
- Gill, R.J. (1992) A review of the sweet-potato whitefly in southern california. *Pan-Pacific Entomologist*, **68**(2), 144-152.
- Gondwe F.M.T., Mahungu N.M., Hillocks R.J., Raya M.D., Moyo C.C., Soko M.M., F.P. Chipungu, Benesi I.R.M (2003). Economic losses experienced by small-scale farmers in Malawi due to cassava brown streak virus. p. 28–35. In J.P. Legg and R.J. Hillocks (ed.) Cassava Brown Streak Disease: Past, Present and Future. Proceedings of an International Workshop, Mombasa, Kenya, 20–30 October 2002. National Resources International Limited, Aylesford, the United Kingdom.
- Gottlieb, Y., Ghanim, M., Chiel, E., Gerling, D., Portnoy, V., Steinberg, S., Tzuri, G., Horowitz, A.R., Belausov, E., Mozes-Daube, N., Kontsedalov, S., Gershon, M., Gal, S., KatZir, N., Zchori-Fein, E. (2006) Identification and localization of a Rickettsia sp in *Bemisia tabaci* (Homoptera : Aleyrodidae). *Applied and Environmental Microbiology*, **72**(5), 3646-3652.
- Gottlieb, Y., Ghanim, M., Gueguen, G., Kontsedalov, S., Vavre, F., Fleury, F., Zchori-Fein, E. (2008) Inherited intracellular ecosystem: symbiotic bacteria share bacteriocytes in whiteflies. *Faseb Journal*, **22**(7), 2591-2599.
- Gottlieb, Y., Zchori-Fein, E., Mozes-Daube, N., Kontsedalov, S., Skaljic, M., Brumin, M., Sobol, I., Czosnek, H., Vavre, F., Fleury, F., Ghanim, M. (2010) The transmission efficiency of Tomato yellow leaf curl virus by the whitefly *Bemisia tabaci* Is correlated

- with the presence of a specific symbiotic bacterium species. *Journal of Virology*, **84**(18), 9310-9317.
- Grover, S., Jindal, V., Banta, G., Taning, C.N.T., Smagghe, G., Christiaens, O. (2019) Potential of RNA interference in the study and management of the whitefly, *Bemisia tabaci*. *Archives of Insect Biochemistry and Physiology*, **100**(2).
- Gunduz, E.A. and Douglas, A.E. (2009) Symbiotic bacteria enable insect to use a nutritionally inadequate diet. *Proceedings of the Royal Society B-Biological Sciences*, **276**(1658), 987-991.
- Guo, X.-j., Rao, Q., Zhang, F., Luo, C., Zhang, H.-y., Gao, X.-w. (2012) Diversity and genetic differentiation of the whitefly *Bemisia tabaci* species complex in China Based on mtCOI and cDNA-AFLP Analysis. *Journal of Integrative Agriculture*, **11**(2), 206-214.
- Hakim, R.S., Baldwin, K., Smagghe, G. (2010) Regulation of Midgut Growth, Development, and Metamorphosis. *Annual Review of Entomology*, **55**, 593-608.
- Hamilton, A.J. and Baulcombe, D.C. (1999) A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science*, **286**(5441), 950-952.
- Hammond, S.M. (2005) Dicing and slicing - The core machinery of the RNA interference pathway. *Febs Letters*, **579**(26), 5822-5829.
- Hanin, M. and Paszkowski, J. (2003) Plant genome modification by homologous recombination. *Current Opinion in Plant Biology*, **6**(2), 157-162.
- Hansen, A.K. and Moran, N.A. (2011) Aphid genome expression reveals host-symbiont cooperation in the production of amino acids. *Proceedings of the National Academy of Sciences of the United States of America*, **108**(7), 2849-2854.
- Harrison, B.D., Zhou, X., Otim-Nape, G.W., Liu, Y., Robinson, D.J. (1997) Role of a novel type of double infection in the geminivirus-induced epidemic of severe cassava mosaic in Uganda. *Annals of Applied Biology*, **131**(3), 437-448.
- Heddi, A., Vallier, A., Anselme, C., Xin, H., Rahbe, Y., Wackers, F. (2005) Molecular and cellular profiles of insect bacteriocytes: mutualism and harm at the initial evolutionary step of symbiogenesis. *Cellular Microbiology*, **7**(2), 293-305.
- Hedges, L.M., Brownlie, J.C., O'Neill, S.L., Johnson, K.N. (2008) Wolbachia and virus Protection in Insects. *Science*, **322**(5902), 702-702.
- Hendrix, D.L. and Wei, Y.A. (1994) Bemisiase - An unusual trisaccharide in bemisia honeydew. *Carbohydrate Research*, **253**, 329-334.

- Hendrix, D.L., Wei, Y.A., Leggett, J.E. (1992) Homopteran honeydew sugar composition is determined by both the insect and plant-species. *Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology*, **101**(1-2), 23-27.
- Henry, L.M., Peccoud, J., Simon, J.C., Hadfield, J.D., Maiden, M.J.C., Ferrari, J., Godfray, H.C.J. (2013) Horizontally transmitted symbionts and host colonization of ecological niches. *Current Biology*, **23**(17), 1713-1717.
- Hershey C (Ed). (2017). Achieving sustainable cultivation of cassava Volume 2: Genetics, *breeding, pests and diseases* . Burleigh Dodds, 322 p. (Burleigh Dodds Series in Agricultural Science)
- Hijaz, F. and Killiny, N. (2014) Collection and Chemical Composition of Phloem Sap from *Citrus sinensis* L. Osbeck (Sweet Orange). *Plos One*, **9**(7).
- Hilje, L., Costa, H.S., Stansly, P.A. (2001) Cultural practices for managing *Bemisia tabaci* and associated viral diseases. *Crop Protection*, **20**(9), 801-812.
- Hillocks, R., Maruthi, M., Kulembeka, H., Jeremiah, S., Alacho, F., Masinde, E., Ogendo, J., Arama, P., Mulwa, R., Mkamilo, G., Kimata, B., Mwakanyamale, D., Mhone, A., Benesi, I. (2016) Disparity between leaf and root symptoms and crop losses associated with Cassava brown streak disease in four countries in Eastern Africa. *Journal of Phytopathology*, **164**(2), 86-93.
- Horowitz, A.R. and Ishaaya, I. (2014) Dynamics of biotypes B and Q of the whitefly *Bemisia tabaci* and its impact on insecticide resistance. *Pest Management Science*, **70**(10), 1568-1572.
- Hotopp, J.C.D., Clark, M.E., Oliveira, D., Foster, J.M., Fischer, P., Torres, M.C., Giebel, J.D., Kumar, N., Ishmael, N., Wang, S.L., Ingram, J., Nene, R.V., Shepard, J., Tomkins, J., Richards, S., Spiro, D.J., Ghedin, E., Slatko, B.E., Tettelin, H., Werren, J.H. (2007) Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science*, **317**(5845), 1753-1756.
- Houndete, T.A., Ketoh, G.K., Hema, O.S.A., Brevault, T., Glitho, I.A., Martin, T. (2010) Insecticide resistance in field populations of *Bemisia tabaci* (Hemiptera: Aleyrodidae) in West Africa. *Pest Management Science*, **66**(11), 1181-1185.
- Hu, W.Q., Sillaots, S., Lemieux, S., Davison, J., Kauffman, S., Breton, A., Linteau, A., Xin, C.L., Bowman, J., Becker, J., Jiang, B., Roemer, T. (2007) Essential gene identification and drug target prioritization in *Aspergillus fumigatus*. *PLOS Pathogens*, **3**(3).

- Hu, J., De Barro, P., Zhao, H., Wang, J., Nardi, F., Liu, S.-S. (2011) An extensive field survey combined with a phylogenetic analysis reveals rapid and widespread invasion of two alien whiteflies in China. *Plos One*, **6**(1).
- Husnik, F. and McCutcheon, J.P. (2018) Functional horizontal gene transfer from bacteria to eukaryotes. *Nature Reviews Microbiology*, **16**(2), 67-79.
- Husnik, F., Nikoh, N., Koga, R., Ross, L., Duncan, R.P., Fujie, M., Tanaka, M., Satoh, N., Bachtrog, D., Wilson, A.C.C., von Dohlen, C.D., Fukatsu, T., McCutcheon, J.P. (2013) Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. *Cell*, **153**(7), 1567-1578.
- Huvenne, H. and Smagghe, G. (2010) Mechanisms of dsRNA uptake in insects and potential of RNAi for pest control: A review. *Journal of Insect Physiology*, **56**(3), 227-235.
- Jamshidi, N. and Palsson, B.O. (2007) Investigating the metabolic capabilities of *Mycobacterium tuberculosis* H37Rv using the in silico strain iNJ661 and proposing alternative drug targets. *BMC Systems Biology*, **1**.
- Janecek, S. (1997) alpha-amylase family: Molecular biology and evolution. *Progress in Biophysics & Molecular Biology*, **67**(1), 67-97.
- Jinek, M. and Doudna, J.A. (2009) A three-dimensional view of the molecular machinery of RNA interference. *Nature*, **457**(7228), 405-412.
- Jing, X.F., Wong, A.C.N., Chaston, J.M., Colvin, J., McKenzie, C.L., Douglas, A.E. (2014) The bacterial communities in plant phloem-sap-feeding insects. *Molecular Ecology*, **23**(6), 1433-1444.
- Jing, X., White, T.A., Luan, J., Jiao, C., Fei, Z., Douglas, A.E. (2016) Evolutionary conservation of candidate osmoregulation genes in plant phloem sap-feeding insects. *Insect Molecular Biology*, **25**(3), 251-258.
- Jones, D.R. (2003) Plant viruses transmitted by whiteflies. *European Journal of Plant Pathology*, **109**(3), 195-219.
- Jose, A.M. and Hunter, C.P. (2007) Transport of sequence-specific RNA interference information between cells. *Annual Review of Genetics*, **41**, 305-330.
- Juhas, M., Eberl, L., Church, G.M. (2012) Essential genes as antimicrobial targets and cornerstones of synthetic biology. *Trends in Biotechnology*, **30**(11), 601-607.
- Juhas, M., Eberl, L., Glass, J.I. (2011) Essence of life: essential genes of minimal genomes. *Trends in Cell Biology*, **21**(10), 562-568.

- Jung, J.S., Preston, G.M., Smith, B.L., Guggino, W.B., Agre, P. (1994) Molecular-structure of the water channel through aquaporin chip - the hourglass model. *Journal of Biological Chemistry*, **269**(20), 14648-14654.
- Kanehisa, M. and Goto, S. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, **28**(1), 27-30.
- Kanju E.E., Masumba E., Masawe M., Tollano S., Muli B., Zacarias A., Mahungu N (2007) Breeding cassava for brown streak resistance: Regional cassava development strategy based on farmer and consumer preference. *Proceedings of the 13th ISTRC symposium* pp 95-101
- Karley, A.J., Ashford, D.A., Minto, L.M., Pritchard, J., Douglas, A.E. (2005) The significance of gut sucrase activity for osmoregulation in the pea aphid, *Acyrtosiphon pisum*. *Journal of Insect Physiology*, **51**(12), 1313-1319.
- Kaufmann, N., Mathai, J.C., Hill, W.G., Dow, J.A.T., Zeidel, M.L., Brodsky, J.L. (2005) Developmental expression and biophysical characterization of a *Drosophila melanogaster* aquaporin. *American Journal of Physiology-Cell Physiology*, **289**(2), C397-C407.
- Kazakoff, S.H., Imelfort, M., Edwards, D., Koehorst, J., Biswas, B., Batley, J., Scott, P.T., Gresshoff, P.M. (2012) Capturing the Biofuel Wellhead and Powerhouse: The Chloroplast and Mitochondrial Genomes of the Leguminous Feedstock Tree *Pongamia pinnata*. *PLOS One*, **7**(12).
- Keseler, I.M., Mackie, A., Peralta-Gil, M., Santos-Zavaleta, A., Gama-Castro, S., Bonavides-Martinez, C., Fulcher, C., Huerta, A.M., Kothari, A., Krummenacker, M., Latendresse, M., Muniz-Rascado, L., Ong, Q., Paley, S., Schroder, I., Shearer, A.G., Subhraveti, P., Travers, M., Weerasinghe, D., Weiss, V., Collado-Vides, J., Gunsalus, R.P., Paulsen, I., Karp, P.D. (2013) EcoCyc: fusing model organism databases with systems biology. *Nucleic Acids Research*, **41**(D1), D605-D612.
- Khasdan, V., Levin, I., Rosner, A., Morin, S., Kontsedalov, S., Maslenin, L., Horowitz, A.R. (2005) DNA markers for identifying biotypes B and Q of *Bemisia tabaci* (Hemiptera : Aleyrodidae) and studying population dynamics. *Bulletin of Entomological Research*, **95**(6), 605-613.
- Kidwell, M.G. (1993) Lateral Transfer in natural-populations of eukaryotes. *Annual Review of Genetics*, **27**, 235-256.

- Kim, D., Landmead, B., Salzberg, S.L. (2015) HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, **12**(4), 357-U121.
- King, Z.A., Lu, J., Drager, A., Miller, P., Federowicz, S., Lerman, J.A., Ebrahim, A., Palsson, B.O., Lewis, N.E. (2016) BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Research*, **44**(D1), D515-D522.
- Klink, V.P., Kim, K.-H., Martins, V., MacDonald, M.H., Beard, H.S., Alkharouf, N.W., Lee, S.-K., Park, S.-C., Matthews, B.F. (2009) A correlation between host-mediated expression of parasite genes as tandem inverted repeats and abrogation of development of female *Heterodera glycines* cyst formation during infection of *Glycine max*. *Planta*, **230**(1), 53-71.
- Kliot, A., Kontsedalov, S., Lebedev, G., Brumin, M., Cathrin, P.B., Marubayashi, J.M., Skaljic, M., Belausov, E., Czosnek, H., Ghanim, M. (2014) Fluorescence in situ Hybridizations (FISH) for the localization of viruses and endosymbiotic bacteria in plant and insect tissues. *Jove-Journal of Visualized Experiments*(84).
- Kola, V.S.R., Renuka, P., Madhav, M.S., Mangrauthia, S.K. (2015) Key enzymes and proteins of crop insects as candidate for RNAi based gene silencing. *Frontiers in Physiology*, **6**.
- Konopova, B. and Jindra, M. (2007) Juvenile hormone resistance gene Methoprene-tolerant controls entry into metamorphosis in the beetle *Tribolium castaneum*. *Proceedings of the National Academy of Sciences of the United States of America*, **104**(25), 10488-10493.
- Kontsedalov, S., Abu-Moch, F., Lebedev, G., Czosnek, H., Horowitz, A.R., Ghanim, M. (2012) *Bemisia tabaci* Biotype Dynamics and Resistance to Insecticides in Israel During the Years 2008-2010. *Journal of Integrative Agriculture*, **11**(2), 312-320.
- Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K. (2018) MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, **35**(6), 1547-1549.
- Kumar, S., Stecher, G., Tamura, K. (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for bigger datasets. *Molecular Biology and Evolution*, **33**(7), 1870-1874.
- Kung, A.W., Kilby, P.M., Portwood, D.E., Dickman, M.J. (2018) Quantification of dsRNA using stable isotope labeling dilution liquid chromatography/mass spectrometry. *Rapid Communications in Mass Spectrometry*, **32**(7), 590-596.

- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G. (2007) Clustal W and clustal X version 2.0. *Bioinformatics*, **23**(21), 2947-2948.
- Larsson, E., Sander, C., Marks, D. (2010) mRNA turnover rate limits siRNA and microRNA efficacy. *Molecular Systems Biology*, **6**.
- Lee, K.S., Kim, S.R., Lee, S.M., Lee, K.R., Sohn, H.D. and Jin, B.R. (2001) Molecular cloning and expression of a cDNA encoding the aquaporin homologue from the firefly, *Pyrocoelia rufa*. *Korean J Entomol* **31**: 269–279.
- Lee, W., Park, J., Lee, G.-S., Lee, S., Akimoto, S.-i. (2013) Taxonomic status of the *Bemisia tabaci* complex (Hemiptera: Aleyrodidae) and reassessment of the number of its constituent species. *Plos One*, **8**(5).
- Lee, Y.S., Nakahara, K., Pham, J.W., Kim, K., He, Z.Y., Sontheimer, E.J., Carthew, R.W. (2004) Distinct roles for *Drosophila* dicer-1 and dicer-2 in the siRNA/miRNA silencing pathways. *Cell*, **117**(1), 69-81.
- Lee, K.H., Park, J.H., Kim, T.Y., Kim, H.U., Lee, S.Y. (2007) Systems metabolic engineering of *Escherichia coli* for L-threonine production. *Molecular Systems Biology*, **3**.
- Legg, J.P. (1999) Emergence, spread and strategies for controlling the pandemic of cassava mosaic virus disease in east and central Africa. *Crop Protection*, **18**(10), 627-637.
- Legg, J., Somado, E.A., Barker, I., Beach, L., Ceballos, H., Cuellar, W., Elkhoury, W., Gerling, D., Helsen, J., Hershey, C., Jarvis, A., Kulakow, P., Kumar, L., Lorenzen, J., Lynam, J., McMahon, M., Maruthi, G., Miano, D., Mtunda, K., Natwuruhunga, P., Okogbenin, E., Pezo, P., Terry, E., Thiele, G., Thresh, M., Wadsworth, J., Walsh, S., Winter, S., Tohme, J., Fauquet, C. (2014a) A global alliance declaring war on cassava viruses in Africa. *Food Security*, **6**(2), 231-248.
- Legg, J.P., Shirima, R., Tajebe, L.S., Guastella, D., Boniface, S., Jeremiah, S., Nsami, E., Chikoti, P., Rapisarda, C. (2014b) Biology and management of *Bemisia tabaci* whitefly vectors of cassava virus pandemics in Africa. *Pest Management Science*, **70**(10), 1446-1453.
- Legg, J.P., French, R., Rogan, D., Okao-Okuja, G., Brown, J.K. (2002) A distinct *Bemisia tabaci* (Gennadius) (Hemiptera : Sternorrhyncha : Aleyrodidae) genotype cluster is associated with the epidemic of severe cassava mosaic virus disease in Uganda. *Molecular Ecology*, **11**(7), 1219-1229.

- Legg, J.P., Jeremiah, S.C., Obiero, H.M., Maruthi, M.N., Ndyetabula, I., Okao-Okuja, G., Bouwmeester, H., Bigirimana, S., Tata-Hangy, W., Gashaka, G., Mkamilo, G., Alicai, T., Kumar, P.L. (2011) Comparing the regional epidemiology of the cassava mosaic and cassava brown streak virus pandemics in Africa. *Virus Research*, **159**(2), 161-170.
- Legg, J.P., Sseruwagi, P., Boniface, S., Okao-Okuja, G., Shirima, R., Bigirimana, S., Gashaka, G., Herrmann, H.-W., Jeremiah, S., Obiero, H., Ndyetabula, I., Tata-Hangy, W., Masembe, C., Brown, J.K. (2014b) Spatio-temporal patterns of genetic change amongst populations of cassava *Bemisia tabaci* whiteflies driving virus pandemics in East and Central Africa. *Virus Research*, **186**, 61-75.
- Legg, J.P., Lava Kumar, P., Makeshkumar, T., Tripathi, L., Ferguson, M., Kanju, E., Ntawuruhunga, P., Cuellar, W. (2015) Cassava virus diseases: biology, epidemiology, and management. *Advances in virus research*, **91**, 85-142.
- Levin, J.S., Thompson, W.F., Csinos, A.S., Stephenson, M.G., Weissinger, A.K. (2005) Matrix attachment regions increase the efficiency and stability of RNA-mediated resistance to Tomato Spotted Wilt Virus in transgenic tobacco. *Transgenic Research*, **14**(2), 193-206.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., Genome Project Data, P. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**(16), 2078-2079.
- Li, X.G., Zhu, Z., Feng, D.J., Chang, T.J., Liu, X. (2001) Influence of DNA methylation on transgene expression. *Chinese Science Bulletin*, **46**(15), 1300-1304.
- Liao, S.M., Du, Q.S., Meng, J.Z., Pang, Z.W., Huang, R.B. (2013) The multiple roles of histidine in protein interactions. *Chemistry Central Journal*, **7**.
- Liedl, B.E., Lawson, D.M., White, K.K., Shapiro, J.A., Cohen, D.E., Carson, W.G., Trumble, J.T., Mutschler, M.A. (1995) Acylsugars of wild tomato *lycopersicon-pennellii* alters settling and reduces oviposition of *Bemisia-argentifolii* (homoptera, aleyrodidae). *Journal of Economic Entomology*, **88**(3), 742-748.
- Liao, Y., Smyth, G.K., Shi, W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **30**(7), 923-930.
- Linn, F., Heidmann, I., Saedler, H., Meyer, P. (1990) Epigenetic changes in the expression of the maize *al* gene in *Petunia-hybrida* - role of numbers of integrated gene copies and state of methylation. *Molecular & General Genetics*, **222**(2-3), 329-336.

- Liu, J.H., Swevers, L., Iatrou, K., Huvenne, H., Smaghe, G. (2012) Bombyx mori DNA/RNA non-specific nuclease: Expression of isoforms in insect culture cells, subcellular localization and functional assays. *Journal of Insect Physiology*, **58**(8), 1166-1176.
- Liu, S.-s., Colvin, J., De Barro, P.J. (2012) Species concepts as applied to the whitefly *Bemisia tabaci* systematics: how many species are there? *Journal of Integrative Agriculture*, **11**(2), 176-186.
- Livak, K.J. and Schmittgen, T.D. (2001) Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the $2^{-\Delta\Delta CT}$ Method. *Methods*, **25**, 402-408.
- Lowe, S., Browne, M., Boudjelas, S., De Poorter, M (2004). 100 of the world's worst invasive alien species. A selection from the global invasive species database. Invasive Species Spec. Group World Conservation. *Union* 1-12.
- Luan, J.B., Chen, W.B., Hasegawa, D.K., Simmons, A.M., Wintermantel, W.M., Ling, K.S., Fei, Z.J., Liu, S.S., Douglas, A.E. (2015) Metabolic coevolution in the bacterial symbiosis of whiteflies and related plant sap-feeding insects. *Genome Biology and Evolution*, **7**(9), 2635-2647.
- Luan, J.B., Shan, H.W., Isermann, P., Huang, J.H., Lammerding, J., Liu, S.S., Douglas, A.E. (2016) Cellular and molecular remodelling of a host cell for vertical transmission of bacterial symbionts. *Proceedings of the Royal Society B-Biological Sciences*, **283**(1833).
- Luo, Y.A., Chen, Q.G., Luan, J.B., Chung, S.H., Van Eck, J., Turgeon, R., Douglas, A.E. (2017) Towards an understanding of the molecular basis of effective RNAi against a global insect pest, the whitefly *Bemisia tabaci*. *Insect Biochemistry and Molecular Biology*, **88**, 21-29.
- Liu, Y.X., Dong, W.B., Shao, J., Wang, Y.B., Zhou, M.Y., Sun, H.P. (2017) Branched-Chain Amino Acid Negatively Regulates KLF15 Expression via PI3K-AKT Pathway. *Frontiers in Physiology*, **8**.
- Lynch, M. and Force, A. (2000) The probability of duplicate gene preservation by subfunctionalization. *Genetics*, **154**(1), 459-473.
- Lyznik, L.A., Gordon-Kamm, W.J., Tao, Y. (2003) Site-specific recombination for genetic engineering in plants. *Plant Cell Reports*, **21**(10), 925-932.
- MacGregor, E.A., Janecek, S., Svensson, B. (2001) Relationship of sequence and structure to specificity in the alpha-amylase family of enzymes. *Biochimica et biophysica acta-protein structure and molecular enzymology*, **1546**(1), 1-20.

- Mahadav, A., Gerling, D., Gottlieb, Y., Czosnek, H., Ghanim, M. (2008) Parasitization by the wasp *Eretmocerus mundus* induces transcription of genes related to immune response and symbiotic bacteria proliferation in the whitefly *Bemisia tabaci*. *Bmc Genomics*, **9**.
- Majumdar, R., Rajasekaran, K., Cary, J.W. (2017) RNA Interference (RNAi) as a Potential Tool for Control of Mycotoxin Contamination in Crop Plants: Concepts and Considerations. *Frontiers in Plant Science*, **8**.
- Malik, H.J., Raza, A., Amin, I., Scheffler, J.A., Scheffler, B.E., Brown, J.K., Mansoor, S. (2016) RNAi-mediated mortality of the whitefly through transgenic expression of double-stranded RNA homologous to acetylcholinesterase and ecdysone receptor in tobacco plants. *Scientific Reports*, **6**.
- Malka, O., Santos-Garcia, D., Feldmesser, E., Sharon, E., Krause-Sakate, R., Delatte, H., van Brunschot, S., Patel, M., Visendi, P., Mugerwa, H., Seal, S., Colvin, J., Morin, S. (2018) Species-complex diversification and host-plant associations in *Bemisia tabaci*: A plant-defence, detoxification perspective revealed by RNA-Seq analyses. *Molecular Ecology*, **27**(21), 4241-4256.
- Manyong V. M., Maeda C., Kanju, E., Legg, J. P. (2012). Economic damage of cassava brown streak disease in sub-Saharan Africa. In *Tropical Root and Tuber Crops and the Challenges of Globalization and Climate Change*, 11th ISTRC-AB Symposium, 4–8 October 2010, pp. 61–68. Edited by R. U. Okechukwu & P. Ntawuruhunga. Kinshasa, Democratic Republic of Congo.
- Mao, Y.-B., Tao, X.-Y., Xue, X.-Y., Wang, L.-J., Chen, X.-Y. (2011) Cotton plants expressing CYP6AE14 double-stranded RNA show enhanced resistance to bollworms. *Transgenic Research*, **20**(3), 665-673.
- Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S.N., Chitsaz, F., Geer, L.Y., Geer, R.C., He, J., Gwadz, M., Hurwitz, D.I., Lanczycki, C.J., Lu, F., Marchler, G.H., Song, J.S., Thanki, N., Wang, Z.X., Yamashita, R.A., Zhang, D.C., Zheng, C.J., Bryant, S.H. (2015) CDD: NCBI's conserved domain database. *Nucleic Acids Research*, **43**(D1), D222-D226.
- Martin, M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, **17**(1), 10-12
- Maruthi, M.N., Colvin, J., Seal, S. (2001) Mating compatibility, life-history traits and RAPD-PCR variation in *Bemisia tabaci* associated with the cassava mosaic disease pandemic in East Africa. *Entomologia experimentalis et applicata*, **99**(1), 13-23.

- Maruthi, M.N., Colvin, J., Thwaites, R.M., Banks, G.K., Gibson, G., Seal, S.E. (2004a) Reproductive incompatibility and cytochrome oxidase I gene sequence variability amongst host-adapted and geographically separate *Bemisia tabaci* populations (Hemiptera : Aleyrodidae). *Systematic Entomology*, **29**(4), 560-568.
- Maruthi, M.N., Seal, S., Colvin, J., Briddon, R.W., Bull, S.E. (2004b) East African cassava mosaic Zanzibar virus - a recombinant begomovirus species with a mild phenotype. *Archives of Virology*, **149**(12), 2365-2377.
- Maruthi, M.N., Hillocks, R.J., Mtunda, K., Raya, M.D., Muhanna, M., Kiozia, H., Rekha, A.R., Colvin, J., Thresh, J.M. (2005) Transmission of Cassava brown streak virus by *Bemisia tabaci* (Gennadius). *Journal of Phytopathology*, **153**(5), 307-312.
- Massey, L.K., Sokatch, J.R., Conrad, R.S. (1976) Branched-chain amino-acid catabolism in bacteria. *Bacteriological Reviews*, **40**(1), 42-54.
- Mathew, L.G., Campbell, E.M., Yool, A.J., Fabrick, J.A. (2011) Identification and characterization of functional aquaporin water channel protein from alimentary tract of whitefly, *Bemisia tabaci*. *Insect Biochemistry and Molecular Biology*, **41**(3), 178-190.
- Mayr, E (1942). Systematics and the origin of species, from the viewpoint of a zoologist. Harvard University Press
- McCutcheon, J.P. and Moran, N.A. (2012) Extreme genome reduction in symbiotic bacteria. *Nature Reviews Microbiology*, **10**(1), 13-26.
- Meyer, P., Linn, F., Heidmann, I., Meyer, H., Niedenhof, I., Saedler, H. (1992) Endogenous and environmental-factors influence 35s promoter methylation of a maize A1 gene construct in transgenic petunia and its color phenotype. *Molecular & General Genetics*, **231**(3), 345-352.
- Miller, S.C., Miyata, K., Brown, S.J., Tomoyasu, Y. (2012) Dissecting Systemic RNA Interference in the Red Flour Beetle *Tribolium castaneum*: Parameters Affecting the Efficiency of RNAi. *Plos One*, **7**(10).
- Mithoefer, A. and Boland, W. (2012) Plant defense against herbivores: chemical aspects. *Annual Review of Plant Biology*, Vol 63, **63**, 431-450.
- Mohanty, A. K. & A. N. Basu. (1986). Effect of host plants and seasonal factors on intraspecific variations in pupal morphology of the whitefly vector, *Bemisia tabaci* (Gennadius). *Journal of Entomological Research* **10**(1):19-26.
- Morales, F.J. (2007) Tropical Whitefly IPM Project. *Advances in Virus Research*, Vol 69, **69**, 249-311.

- Mound, L.A. (1963). Host-correlated variation in *Bemisia tabaci* (Gennadius) (Homoptera:Aleyrodidae). *Proc. R. Entomol. Soc. Lond. Ser. Gen. Entomol.* 38, 171–180.
- Mugerwa, H., Rey, M.E.C., Alicai, T., Ateka, E., Atuncha, H., Ndunguru, J., Sseruwagi, P. (2012) Genetic diversity and geographic distribution of *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae) genotypes associated with cassava in East Africa. *Ecology and Evolution*, 2(11), 2749-2762.
- Mugerwa, H., Seal, S., Wang, H.-L., Patel, M.V., Kabaalu, R., Omongo, C.A., Alicai, T., Tairo, F., Ndunguru, J., Sseruwagi, P., Colvin, J. (2018) African ancestry of New World, *Bemisia tabaci*-whitefly species. *Scientific reports*, 8(1), 2734-2734.
- Mugerwa, H (2018) Molecular characterization of cassava-colonising population of whitefly, *Bemisia tabaci*. Ph.D. Thesis, University of Greenwich
- Mugerwa H., Rey M.E.C., Tairo F., Ndunguru J., Sseruwagi P (2019) Two sub-Saharan Africa 1 populations of *Bemisia tabaci* exhibit distinct biological differences in fecundity and survivorship on cassava. *Crop Protection* 117, 7–14
- Nakai, H., Okuyama, M., Kim, Y.M., Saburi, W., Wongchawalit, J., Mori, H., Chiba, S., Kimura, A. (2005) Molecular analysis of alpha-glucosidase belonging to GH-family 31. *Biologia*, 60, 131-135.
- Nardi, J.B., Miller, L.A., Mark Bee, C (2019) Interfaces between microbes and membranes of host epithelial cells in hemipteran midguts. *Journal of Morphology*. 1–15.
- Naveen, N.C., Chaubey, R., Kumar, D., Rebijith, K.B., Rajagopal, R., Subrahmanyam, B., Subramanian, S. (2017) Insecticide resistance status in the whitefly, *Bemisia tabaci* genetic groups Asia-I, Asia-II-1 and Asia-II-7 on the Indian subcontinent. *Scientific Reports*, 7.
- Ndunguru, J., Sseruwagi, P., Tairo, F., Stomeo, F., Maina, S., Djikeng, A., Kehoe, M., Boykin, L.M. (2015a) Analyses of twelve new whole genome sequences of Cassava Brown Streak Viruses and Ugandan Cassava Brown Streak Viruses from East Africa: Diversity, Supercomputing and Evidence for Further Speciation (vol 10, e0139321, 2015). *Plos One*, 10(10).
- Ndunguru, J., Sseruwagi, P., Tairo, F., Stomeo, F., Maina, S., Djinkeng, A., Kehoe, M., Boykin, L.M. (2015b) Analyses of twelve new whole genome sequences of Cassava Brown Streak Viruses and Ugandan Cassava Brown Streak Viruses from East Africa: Diversity, Supercomputing and Evidence for Further Speciation. *Plos One*, 10(10).

- Ngiwsara, L., Iwai, G., Tagami, T., Sato, N., Nakai, H., Okuyama, M., Mori, H., Kimura, A. (2012) Amino acids in conserved region II are crucial to substrate specificity, reaction velocity, and regioselectivity in the transglucosylation of honeybee GH-13 alpha-Glucosidases. *Bioscience Biotechnology and Biochemistry*, **76**(10), 1967-1974.
- Nie, C.X., He, T., Zhang, W.J., Zhang, G.L., Ma, X. (2018) Branched chain amino acids: beyond nutrition metabolism. *International Journal of Molecular Sciences*, **19**(4).
- Nikoh, N., McCutcheon, J.P., Kudo, T., Miyagishima, S., Moran, N.A., Nakabachi, A. (2010) Bacterial genes in the aphid genome: Absence of functional gene transfer from *Buchnera* to its host. *Plos Genetics*, **6**(2).
- Nikoh, N. and Nakabachi, A. (2009) Aphids acquired symbiotic genes via lateral gene transfer. *BMC Biology*, **7**.
- Norton R (2014) Global starch market outlook and feedstock economics. Cassava World. Africa 2014. Centre for Management Technology (CMT), Lusaka
- Nowak, W., Gawłowska, M., Jarmolowski, A., Augustyniak, J. (2001) Effect of nuclear matrix attachment regions on transgene expression in tobacco plants. *Acta Biochimica Polonica*, **48**(3), 637-646.
- Nwokeoji, A.O., Kung, A.W., Kilby, P.M., Portwood, D.E., Dickman, M.J. (2017) Purification and characterisation of dsRNA using ion pair reverse phase chromatography and mass spectrometry. *Journal of Chromatography A*, **1484**, 14-25.
- Oliveira, M.R.V., Henneberry, T.J., Anderson, P. (2001) History, current status, and collaborative research projects for *Bemisia tabaci*. *Crop Protection*, **20**(9), 709-723.
- Olsen, K.M. (2004) SNPs, SSRs and inferences on cassava's origin. *Plant Molecular Biology*, **56**(4), 517-526.
- Omongo, C.A., Kawuki, R., Bellotti, A.C., Alicai, T., Baguma, Y., Maruthi, M.N., Bua, A., Colvin, J. (2012) African cassava whitefly, *Bemisia tabaci*, resistance in African and South American cassava genotypes. *Journal of Integrative Agriculture*, **11**(2), 327-336.
- Ono, A., Izumi, R., Ishizuka, H., Terao, K., Takao, H., Ichihashi, R., Kobayashi, T., Kataoka, I., Shimokawa, F., Ieee. (2017) Highly pure phloem-sap-extraction sensor device for direct component analysis of nutrition in plant shoots. 2017 19th International Conference on Solid-State Sensors, Actuators and Microsystems (Transducers), 1604-1607.

- Otim, M., Legg, J., Kyamanywa, S., Polaszek, A., Gerling, D. (2005) Occurrence and activity of *Bemisia tabaci* parasitoids on cassava in different agro-ecologies in Uganda. *Biocontrol*, **50**(1), 87-95.
- Otoidobiga, L.C., Vincent, C., Stewart, R.K. (2003) Susceptibility of field populations of adult *Bemisia tabaci* Gennadius (Homoptera : Aleyrodidae) and *Eretmocerus* sp (Hymenoptera : Aphelinidae) to cotton insecticides in Burkina Faso (West Africa). *Pest Management Science*, **59**(1), 97-106.
- Otim-Nape, G.W., Bua, A., Thresh, J. M., Baguma, Y., Ogwal, S., Semakula, G. N., Acola, G., Byabakama, B., Colvin, J., Cooter, R. J., and Martin, A. (2000). Cassava mosaic virus disease in Uganda: The current pandemic and approaches to control. University of Greenwich, PSTC28.
- Pak, J. and Fire, A. (2007) Distinct populations of primary and secondary effectors during RNAi in *C-elegans*. *Science*, **315**(5809), 241-244.
- Park, J.H., Lee, K.H., Kim, T.Y., Lee, S.Y. (2007) Metabolic engineering of *Escherichia coli* for the production of L-valine based on transcriptome analysis and in silico gene knockout simulation. *Proceedings of the National Academy of Sciences of the United States of America*, **104**(19), 7797-7802.
- Pati, B.L., Legg, J.P., Kanju, E., Fauquet, C.M. (2015) Cassava brown streak disease: a threat to food security in Africa. *Journal of General Virology*, **96**, 956-968.
- Petersen, T.N., Brunak, S., von Heijne, G., Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, **8**(10), 785-786.
- Petrillo, E., Herz, M.A.G., Barta, A., Kalyna, M., Kornblihtt, A.R. (2014) Let there be light: Regulation of gene expression in plants. *RNA Biology*, **11**(10), 1215-1220.
- Pennisi, E. (2010) Armed and Dangerous. *Science*, **327**(5967), 804-805.
- Pharkya, P., Burgard, A.P., Maranas, C.D. (2003) Exploring the overproduction of amino acids using the bilevel optimization framework OptKnock. *Biotechnology and Bioengineering*, **84**(7), 887-899.
- Phivan, L., Vonkries, J.P., Ostertag, W., Stratling, W.H. (1990) The chicken lysozyme-5' matrix attachment region increases transcription from a heterologous promoter in heterologous cells and dampens position effects on the expression of transfected genes. *Molecular and Cellular Biology*, **10**(5), 2302-2307.

- Pitino, M., Coleman, A.D., Maffei, M.E., Ridout, C.J., Hogenhout, S.A. (2011) Silencing of aphid genes by dsRNA feeding from plants. *Plos One*, **6**(10).
- Powell, M.E., Bradish, H.M., Gatehouse, J.A., Fitches, E.C. (2017) Systemic RNAi in the small hive beetle *Aethina tumida* Murray (Colcoptera: Nitidulidae), a serious pest of the European honey bee *Apis mellifera*. *Pest Management Science*, **73**(1), 53-63.
- Preston, G.M., Carroll, T.P., Guggino, W.B., Agre, P. (1992) Appearance of water channels in xenopus oocytes expressing red-cell chip28 protein. *Science*, **256**(5055), 385-387.
- Price, D.R.G., Duncan, R.P., Shigenobu, S., Wilson, A.C.C. (2011) Genome expansion and differential expression of amino acid transporters at the aphid/buchnera symbiotic Interface. *Molecular Biology and Evolution*, **28**(11), 3113-3126.
- Prols, F. and Meyer, P. (1992) The methylation patterns of chromosomal integration regions influence gene activity of transferred dna in petunia-hybrida. *Plant Journal*, **2**(4), 465-475.
- Prosser, W.A. and Douglas, A.E. (1992) A Test of the hypotheses that nitrogen is upgraded and recycled in an aphid (*Acyrtosiphon-pisum*) symbiosis. *Journal of Insect Physiology*, **38**(2), 93-99.
- Pujol, B., Gigot, G., Laurent, G., Pinheiro-Kluppel, M., Elias, M., Hossaert-McKey, M., McKey, D. (2002) Germination ecology of cassava (*Manihot esculenta* Crantz, Euphorriaceae) in traditional agroecosystems: Seed and seedling biology of a vegetatively propagated domesticated plant. *Economic Botany*, **56**(4), 366-379.
- Qin, L., Pan, L.-L., Liu, S.-S. (2016) Further insight into reproductive incompatibility between putative cryptic species of the Bemisia tabaci whitefly complex. *Insect science*, **23**(2), 215-224.
- Rabello, A.R., Queiroz, P.R., Simoes, K.C.C., Hiragi, C.O., Lima, L.H.C., Oliveira, M.R.V., Mehta, A. (2008) Diversity analysis of *Bemisia tabaci* biotypes: RAPD, PCR-RFLP and sequencing of the ITS1 rDNA region. *Genetics and Molecular Biology*, **31**(2), 585-590.
- Ramasubbu, N., Paloth, V., Luo, Y.G., Brayer, G.D., Levine, M.J. (1996) Structure of human salivary alpha-amylase at 1.6 angstrom resolution: Implications for its role in the oral cavity. *Acta Crystallographica Section D-Biological Crystallography*, **52**, 435-446.
- Rao, Q., Rollat-Farnier, P.A., Zhu, D.T., Santos-Garcia, D., Silva, F.J., Moya, A., Latorre, A., Klein, C.C., Vavre, F., Sagot, M.F., Liu, S.S., Mouton, L., Wang, X.W. (2015) Genome

- reduction and potential metabolic complementation of the dual endosymbionts in the whitefly *Bemisia tabaci*. *BMC Genomics*, **16**.
- Raza, A., Malik, H.J., Shafiq, M., Amin, I., Scheffler, J.A., Scheffler, B.E., Mansoor, S. (2016) RNA interference based approach to down regulate osmoregulators of whitefly (*Bemisia tabaci*): Potential technology for the control of whitefly. *Plos One*, **11**(4).
- Rekha, A.R., Maruthi, M.N., Muniyappa, V., Colvin, J. (2005) Occurrence of three genotypic clusters of *Bemisia tabaci* and the rapid spread of the B biotype in south India. *Entomologia experimentalis et applicata*, **117**(3), 221-233.
- Rhodes, J.D., Croghan, P.C., Dixon, A.F.G. (1997) Dietary sucrose and oligosaccharide synthesis in relation to osmoregulation in the pea aphid, *Acyrtosiphon pisum*. *Physiological Entomology*, **22**(4), 373-379.
- Robinson, M.D., McCarthy, D.J., Smyth, G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**(1), 139-140.
- Rollat-Farnier, P.A., Santos-Garcia, D., Rao, Q., Sagot, M.F., Silva, F.J., Henri, H., Zchori-Fein, E., Latorre, A., Moya, A., Barbe, V., Liu, S.S., Wang, X.W., Vavre, F., Mouton, L. (2015) Two host clades, two bacterial arsenals: evolution through gene losses in facultative endosymbionts. *Genome Biology and Evolution*, **7**(3), 839-855.
- Rosell, R.C., Bedford, I.D., Frohlich, D.R., Gill, R.J., Brown, J.K., Markham, P.G. (1997) Analysis of morphological variation in distinct populations of *Bemisia tabaci* (Homoptera: Aleyrodidae). *Annals of the Entomological Society of America*, **90**(5), 575-589.
- Russell, C.W., Bouvaine, S., Newell, P.D., Douglas, A.E. (2013) Shared Metabolic Pathways in a Coevolved Insect-Bacterial Symbiosis. *Applied and Environmental Microbiology*, **79**(19), 6117-6123.
- Saier, M.H. (2000) Families of transmembrane transporters selective for amino acids and their derivatives. *Microbiology-Sgm*, **146**, 1775-1795.
- Saier, M.H., Reddy, V.S., Tsu, B.V., Ahmed, M.S., Li, C., Moreno-Hagelsieb, G. (2016) The Transporter Classification Database (TCDB): recent advances. *Nucleic Acids Research*, **44**(D1), D372-D379.
- Salcedo, A., Del Valle, A., Sanchez, B., Ocasio, V., Ortiz, A., Marquez, P., Siritunga, D. (2010) Comparative evaluation of physiological post-harvest root deterioration of 25 cassava (*Manihot esculenta*) accessions: visual vs. hydroxycoumarins fluorescent accumulation analysis. *African Journal of Agricultural Research*, **5**(22), 3138-3144.

- Salvucci, M.E. (2003) Distinct sucrose isomerases catalyze trehalulose synthesis in whiteflies, *Bemisia argentifolii*, and *Erwinia rhapontici*. *Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology*, **135**(2), 385-395.
- Salvucci, M.E. (2000) Effect of the alpha-glucosidase inhibitor, bromoconduritol, on carbohydrate metabolism in the silverleaf whitefly, *Bemisia argentifolii*. *Archives of Insect Biochemistry and Physiology*, **45**(3), 117-128.
- Salvucci, M.E., Wolfe, G.R., Hendrix, D.L. (1997) Effect of sucrose concentration on carbohydrate metabolism in *Bemisia argentifolii*: Biochemical mechanism and physiological role for trehalulose synthesis in the silverleaf whitefly. *Journal of Insect Physiology*, **43**(5), 457-464.
- Sandstrom, J. and Pettersson, J. (1994) Amino-acid-composition of phloem sap and the relation to intraspecific variation in pea aphid (*Acyrtosiphon-pisum*) performance. *Journal of Insect Physiology*, **40**(11), 947-955.
- Santel, A., Aleku, M., Keil, O., Endruschat, J., Esche, V., Fisch, G., Dames, S., Löffler, K., Fechtner, M., Arnold, W., Giese, K., Klippel, A., Kaufmann, J. (2006) A novel siRNA-lipoplex technology for RNA interference in the mouse vascular endothelium. *Gene Therapy*, **13**(16), 1222-1234.
- Santos-Garcia, D., Farnier, P.A., Beitia, F., Zchori-Fein, E., Vavre, F., Mouton, L., Moya, A., Latorre, A., Silva, F.J. (2012) Complete genome sequence of "*Candidatus Portiera aleyrodidarum*" BT-QVLC, an obligate symbiont that supplies amino acids and carotenoids to *Bemisia tabaci*. *Journal of Bacteriology*, **194**(23), 6654-6655.
- Sasaki, T. and Ishikawa, H. (1995) Production of essential amino-acids from glutamate by mycetocyte symbionts of the pea aphid, *Acyrtosiphon-pisum*. *Journal of Insect Physiology*, **41**(1), 41-46.
- Scalise, M., L. Pochini, M. Galluccio and C. Indiveri (2014). Amino acid transporters in drug discovery. *Current Research in Drug Discovery*, **1**: 1-16.
- Schellenberger, J., Que, R., Fleming, R.M.T., Thiele, I., Orth, J.D., Feist, A.M., Zielinski, D.C., Bordbar, A., Lewis, N.E., Rahmanian, S., Kang, J., Hydeke, D.R., Palsson, B.O. (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nature Protocols*, **6**(9), 1290-1307.
- Scholthof, K.-B.G., Adkins, S., Czosnek, H., Palukaitis, P., Jacquot, E., Hohn, T., Hohn, B., Saunders, K., Candresse, T., Ahlquist, P., Hemenway, C., Foster, G.D. (2011) Top 10 plant viruses in molecular plant pathology. *Molecular Plant Pathology*, **12**(9), 938-954.

- Scott, J.G., Michel, K., Bartholomay, L.C., Siegfried, B.D., Hunter, W.B., Smagghe, G., Zhu, K.Y., Douglas, A.E. (2013) Towards the elements of successful insect RNAi. *Journal of Insect Physiology*, **59**(12), 1212-1221.
- Secker, A.E., Bedford, I.D., Markham, P.G., De Williams, M.E.C (1998). Squash, a reliable field indicator for the presence of the B biotype of tobacco whitefly, *Bemisia tabaci*. Presented at the Brighton Crop Protection Conference: Pests and Diseases, Brighton, UK.
- Segre, D., Vitkup, D., Church, G.M. (2002) Analysis of optimality in natural and perturbed metabolic networks. *Proceedings of the National Academy of Sciences of the United States of America*, **99**(23), 15112-15117.
- Shakesby, A.J., 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.
- Shlomi, T., Berkman, O., Ruppin, E. (2005) Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proceedings of the National Academy of Sciences of the United States of America*, **102**(21), 7695-7700.
- Sijen, T., Fleenor, J., Simmer, F., Thijssen, K.L., Parrish, S., Timmons, L., Plasterk, R.H.A., Fire, A. (2001) On the role of RNA amplification in dsRNA-triggered gene silencing. *Cell*, **107**(4), 465-476.
- Simon, B., Cenis, J.L., Beitia, F., Khalid, S., Moreno, I.M., Fraile, A., Garcia-Arenal, F. (2003) Genetic structure of field populations of begomoviruses and of their vector *Bemisia tabaci* in Pakistan. *Phytopathology*, **93**(11), 1422-1429.
- Sindhu, A.S., Maier, T.R., Mitchum, M.G., Hussey, R.S., Davis, E.L., Baum, T.J. (2009) Effective and specific in planta RNAi in cyst nematodes: expression interference of four parasitism genes reduces parasitic success. *Journal of Experimental Botany*, **60**(1), 315-324.
- Siomi, H. and Siomi, M.C. (2009) On the road to reading the RNA-interference code. *Nature*, **457**(7228), 396-404.
- Skaljac, M., Zanic, K., Ban, S.G., Kontsedalov, S., Ghanim, M. (2010) Co-infection and localization of secondary symbionts in two whitefly species. *BMC Microbiology*, **10**.
- Sloan, D.B. and Moran, N.A. (2012a) Endosymbiotic bacteria as a source of carotenoids in whiteflies. *Biology Letters*, **8**(6), 986-989.

- Sloan, D.B. and Moran, N.A. (2012b) Genome Reduction and Co-evolution between the Primary and Secondary Bacterial Symbionts of Psyllids. *Molecular Biology and Evolution*, **29**(12), 3781-3792.
- Sloan, D.B., Nakabachi, A., Richards, S., Qu, J.X., Murali, S.C., Gibbs, R.A., Moran, N.A. (2014) Parallel histories of horizontal gene transfer facilitated extreme reduction of endosymbiont genomes in sap-feeding insects. *Molecular Biology and Evolution*, **31**(4), 857-871.
- Smith, C.M. (2005) Plant resistance to arthropods: molecular and conventional approaches. plant resistance to arthropods: *Molecular and Conventional Approaches*.
- Sseruwagi, P., Legg, J.P., Maruthi, M.N., Colvin, J., Rey, M.E.C., Brown, J.K. (2005) Genetic diversity of *Bemisia tabaci* (Gennadius) (Hemiptera : Aleyrodidae) populations and presence of the B biotype and a non-B biotype that can induce silverleaf symptoms in squash, in Uganda. *Annals of Applied Biology*, **147**(3), 253-265.
- Sseruwagi, P., Maruthi, M.N., Colvin, J., Rey, M.E.C., Brown, J.K., Legg, J.P. (2006) Colonization of non-cassava plant species by cassava whiteflies (*Bemisia tabaci*) in Uganda. *Entomologia experimentalis et applicata*, **119**(2), 145-153.
- Srivastava, V., Ariza-Nieto, M., Wilson, A.J. (2004) Cre-mediated site-specific gene integration for consistent transgene expression in rice. *Plant Biotechnology Journal*, **2**(2), 169-179.
- Spiker, S. and Thompson, W.F. (1996) Nuclear matrix attachment regions and transgene expression in plants. *Plant Physiology*, **110**(1), 15-21.
- Suatoni, E., Vicario, S., Rice, S., Snell, T., Caccone, A. (2006) An analysis of species boundaries and biogeographic patterns in a cryptic species complex: The rotifer - *Brachionus plicatilis*. *Molecular Phylogenetics and Evolution*, **41**(1), 86-98.
- Suchard, M.A., Lemey, P., Baele, G., Ayres, D.L., Drummond, A.J., Rambaut, A. (2018) Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution*, **4**(1).
- Sui, H.X., Han, B.G., Lee, J.K., Walian, P., Jap, B.K. (2001) Structural basis of water-specific transport through the AQP1 water channel. *Nature*, **414**(6866), 872-878.
- Sun, D.B., Xu, J., Luan, J.B., Liu, S.S. (2011) Reproductive incompatibility between the B and Q biotypes of the whitefly *Bemisia tabaci* in China: genetic and behavioural evidence. *Bulletin of Entomological Research*, **101**(2), 211-220.

- Sweiry, J.H., Sastre, J., Vina, J., Elsasser, H.P., Mann, G.E. (1995) A role for gamma-glutamyl-transpeptidase and the amino-acid-transport system X(C)(-) in cystine transport by a human pancreatic duct cell-line. *Journal of Physiology-London*, **485**(1), 167-177.
- Tajebe, L.S., Guastella, D., Cavalieri, V., Kelly, S.E., Hunter, M.S., Lund, O.S., Legg, J.P., Rapisarda, C. (2015) Diversity of symbiotic bacteria associated with *Bemisia tabaci* (Homoptera: Aleyrodidae) in cassava mosaic disease pandemic areas of Tanzania. *Annals of Applied Biology*, **166**(2), 297-310.
- Taniguchi, N. and Ikeda, Y. (1998) gamma-glutamyl transpeptidase: Catalytic mechanism and gene expression. *Advances in Enzymology*, Vol 72: *Amino Acid Metabolism*, Pt A, **72**, 239-+.
- Terenius, O., Papanicolaou, A., Garbutt, J.S., Eleftherianos, I., Havenne, H., Kanginakudru, S., Albrechtsen, M., An, C.J., Aymeric, J.L., Barthel, A., Bebas, P., Bitra, K., Bravo, A., Chevalieri, F., Collinge, D.P., Crava, C.M., de Maagd, R.A., Duvic, B., Erlandson, M., Faye, I., Felfoldi, G., Fujiwara, H., Futahashi, R., Gandhe, A.S., Gatehouse, H.S., Gatehouse, L.N., Giebertowicz, J.M., Gomez, I., Grimmeliikhuijzen, C.J.P., Groot, A.T., Hauser, F., Heckel, D.G., Hegedus, D.D., Hrycaj, S., Huang, L.H., Hull, J.J., Iatrou, K., Iga, M., Kanost, M.R., Kotwica, J., Li, C.Y., Li, J.H., Liu, J.S., Lundmark, M., Matsumoto, S., Meyering-Vos, M., Millichap, P.J., Monteiro, A., Mrinal, N., Niimi, T., Nowara, D., Ohnishi, A., Oostra, V., Ozaki, K., Papakonstantinou, M., Popadic, A., Rajam, M.V., Saenko, S., Simpson, R.M., Soberon, M., Strand, M.R., Tomita, S., Toprak, U., Wang, P., Wee, C.W., Whyard, S., Zhang, W.Q., Nagaraju, J., Ffrench-Constant, R.H., Herrero, S., Gordon, K., Swelters, L., Smagghe, G. (2011) RNA interference in Lepidoptera: An overview of successful and unsuccessful studies and implications for experimental design. *Journal of Insect Physiology*, **57**(2), 231-245.
- Teye, E., Asare, A. P., Amoah, R. S., & Tetteh, J. P. (2014). Determination of the dry matter content of cassava (*Manihot esculenta*, Crantz) tubers using specific gravity method. *ARPN Journal of Agriculture and Biological Science*.
- Thakur, N., Upadhyay, S.K., Verma, P.C., Chandrashekar, K., Tuli, R., Singh, P.K. (2014) Enhanced Whitefly Resistance in Transgenic Tobacco Plants Expressing Double Stranded RNA of v-ATPase A Gene. *Plos One*, **9**(3).
- Thao, M.L. and Baumann, P. (2004a) Evolutionary relationships of primary prokaryotic endosymbionts of whiteflies and their hosts. *Applied and Environmental Microbiology*, **70**(6), 3401-3406.

- Thao, M.L.L. and Baumann, P. (2004b) Evidence for multiple acquisition of *Arsenophonus* by whitefly species (Sternorrhyncha : Aleyrodidae). *Current Microbiology*, **48**(2), 140-144.
- Thiele, I. and Palsson, B.O. (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols*, **5**(1), 93-121.
- Thomas, D., Bron, P., Ranchy, G., Duchesne, L., Cavalier, A., Rolland, J.P., Raguenes-Nicol, C., Hubert, J.F., Haase, W., Delamarche, C. (2002) Aquaglyceroporins, one channel for two molecules. *Biochimica et biophysica acta-bioenergetics*, **1555**(1-3), 181-186.
- Thresh, J.M. and Cooter, R.J. (2005) Strategies for controlling cassava mosaic virus disease in Africa. *Plant Pathology*, **54**(5), 587-614.
- Tomoyasu, Y., Miller, S.C., Tomita, S., Schoppmeier, M., Grossmann, D., Bucher, G. (2008) Exploring systemic RNA interference in insects: a genome-wide survey for RNAi genes in *Tribolium*. *Genome Biology*, **9**(1).
- Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., Pachter, L. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*, **28**(5), 511-U174.
- Tzin, V., Yang, X.W., Jing, X.F., Zhang, K., Jander, G., Douglas, A.E. (2015) RNA interference against gut osmoregulatory genes in phloem-feeding insects. *Journal of Insect Physiology*, **79**, 105-112.
- Uchida, K. (2003) Histidine and lysine as targets of oxidative modification. *Amino Acids*, **25**(3-4), 249-257.
- Ueda, S., Kitamura, T., Kijima, K., Honda, K.I., Kanmiya, K. (2009) Distribution and molecular characterization of distinct Asian populations of *Bemisia tabaci* (Hemiptera: Aleyrodidae) in Japan. *Journal of Applied Entomology*, **133**(5), 355-366.
- Uitdehaag, J.C.M., Mosi, R., Kalk, K.H., van der Veen, B.A., Dijkhuizen, L., Withers, S.G., Dijkstra, B.W. (1999) X-ray structures along the reaction pathway of cyclodextrin glycosyltransferase elucidate catalysis in the alpha-amylase family. *Nature Structural Biology*, **6**(5), 432-436.
- Umeh, V.C. and Adeyemi, A. (2011) Population dynamics of the woolly whitefly *Aleurothrixus floccosus* (Maskell) on sweet orange varieties in Nigeria and association of *A. floccosus* with the entomopathogenic fungi *Aschersonia* spp. *Fruits*, **66**(6), 385-392.

- Upadhyay, S.K., Chandrashekar, K., Thakur, N., Verma, P.C., Borgio, J.F., Singh, P.K., Tuli, R. (2011) RNA interference for the control of whiteflies (*Bemisia tabaci*) by oral route. *Journal of Biosciences*, **36**(1), 153-161.
- Upadhyay, S.K., Dixit, S., Sharma, S., Singh, H., Kumar, J., Verma, P.C., Chandrashekar, K. (2013) siRNA Machinery in Whitefly (*Bemisia tabaci*). *Plos One*, **8**(12).
- Upadhyay, S.K., Sharma, S., Singh, H., Dixit, S., Kumar, J., Verma, P.C., Chandrashekar, K. (2015) Whitefly Genome Expression Reveals Host-Symbiont Interaction in Amino Acid Biosynthesis. *Plos One*, **10**(5).
- Valdes, V.J., Athie, A., Salinas, L.S., Navarro, R.E., Vaca, L. (2012) CUP-1 Is a Novel Protein Involved in Dietary Cholesterol Uptake in *Caenorhabditis elegans*. *Plos One*, **7**(3).
- Valle, E.M., Boggio, S.B., Heldt, H.W. (1998) Free amino acid composition of phloem sap and growing fruit of *Lycopersicon esculentum*. *Plant and Cell Physiology*, **39**(4), 458-461.
- Van Ekert, E., Chauvigne, F., Finn, R.N., Mathew, L.G., Hull, J.J., Cerda, J., Fabrick, J.A. (2016) Molecular and functional characterization of *Bemisia tabaci* aquaporins reveals the water channel diversity of hemipteran insects. *Insect Biochemistry and Molecular Biology*, **77**, 39-51.
- Vanwinkle, L.J., Christensen, H.N., Campione, A.L. (1985) Na⁺-dependent transport of basic, zwitterionic and bicyclic amino-acids by a broad-scope system in mouse blastocysts. *Journal of Biological Chemistry*, **260**(22), 2118-2123.
- Varma, A. and Palsson, B.O. (1994) Metabolic flux balancing - basic concepts, scientific and practical use. *Bio-Technology*, **12**(10), 994-998.
- Viscarret, M.M., Torres-Jerez, I., De Manero, E.A., Lopez, S.N., Botto, E.E., Brown, J.K. (2003) Mitochondrial DNA evidence for a distinct new world group of *Bemisia tabaci* (Gennadius) (Hemiptera : Aleyrodidae) indigenous to Argentina and Bolivia and presence of the Old World B biotype in Argentina. *Annals of the Entomological Society of America*, **96**(1), 65-72.
- Visendi, P., Berkman, P.J., Hayashi, S., Golicz, A.A., Bayer, P.E., Ruperao, P., Hurgobin, B., Montenegro, J., Chan, C.K.K., Stankova, H., Batley, J., Simkova, H., Dolezel, J., Edwards, D. (2016) An efficient approach to BAC based assembly of complex genomes. *Plant Methods*, **12**.
- Von Heijne, G. (1998) Protein transport - Life and death of a signal peptide. *Nature*, **396**(6707), 111-+.

- Vyskocilova, S., Tay, W.T., van Brunschot, S., Seal, S., Colvin, J. (2018) An integrative approach to discovering cryptic species within the *Bemisia tabaci* whitefly species complex. *Scientific Reports*, **8**.
- Walshe, D.P., Lehane, S.M., Lehane, M.J., Haines, L.R. (2009) Prolonged gene knockdown in the tsetse fly *Glossina* by feeding double stranded RNA. *Insect Molecular Biology*, **18**(1), 11-19.
- Wang, T.Y., Xue, L.X., Hou, W.H., Yang, B.S., Chai, Y.R., Ji, X.A., Wang, Y.F. (2007) Increased expression of transgene in stably transformed cells of *Dunaliella salina* by matrix attachment regions. *Applied Microbiology and Biotechnology*, **76**(3), 651-657.
- Wang, Z., Gerstein, M., Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, **10**(1), 57-63.
- Wang, P., Sun, D.-B., Qiu, B.-L., Liu, S.-S. (2011) The presence of six cryptic species of the whitefly *Bemisia tabaci* complex in China as revealed by crossing experiments. *Insect Science*, **18**(1), 67-77.
- Wang, X.W., Luan, J.B., Li, J.M., Bao, Y.Y., Zhang, C.X., Liu, S.S. (2010) De novo characterization of a whitefly transcriptome and analysis of its gene expression during development. *BMC Genomics*, **11**.
- Wang, K.X., Peng, Y.C., Pu, J., Fu, W.X., Wang, J.L., Han, Z.J. (2016) Variation in RNAi efficacy among insect species is attributable to dsRNA degradation in vivo. *Insect Biochemistry and Molecular Biology*, **77**, 1-9.
- Wei, Y.A., Hendrix, D.L., Nieman, R. (1997) Diglucomelezitose, a novel pentasaccharide in silverleaf whitefly honeydew. *Journal of Agricultural and Food Chemistry*, **45**(9), 3481-3486.
- Wilkinson, T.L. and Douglas, A.E. (2003) Phloem amino acids and the host plant range of the polyphagous aphid, *Aphis fabae*. *Entomologia Experimentalis et Applicata*, **106**(2), 103-113.
- Wilkinson, T.L., Ashford, D.A., Pritchard, J., Douglas, A.E. (1997) Honeydew sugars and osmoregulation in the pea aphid *Acyrtosiphon pisum*. *Journal of Experimental Biology*, **200**(15), 2137-2143.
- Winter, S., Koerbler, M., Stein, B., Pietruszka, A., Paape, M., Butgereitt, A. (2010) Analysis of cassava brown streak viruses reveals the presence of distinct virus species causing cassava brown streak disease in East Africa. *Journal of General Virology*, **91**, 1365-1372.

- Winston, W.M., Molodowitch, C., Hunter, C.P. (2002) Systemic RNAi in *C-elegans* requires the putative transmembrane protein SID-1. *Science*, **295**(5564), 2456-2459.
- Witczak, M., Ziobro, R., Juszczak, L., Korus, J. (2016) Starch and starch derivatives in gluten-free systems - A review. *Journal of Cereal Science*, **67**, 46-57.
- Wolfrum, C., Shi, S., Jayaprakash, K.N., Jayaraman, M., Wang, G., Pandey, R.K., Rajeev, K.G., Nakayama, T., Charrise, K., Ndungo, E.M., Zimmermann, T., Kotliansky, V., Manoharan, M., Stoffel, M. (2007) Mechanisms and optimization of in vivo delivery of lipophilic siRNAs. *Nature Biotechnology*, **25**(10), 1149-1157.
- Wyder, S., Kriventseva, E.V., Schroder, R., Kadowaki, T., Zdobnov, E.M. (2007) Quantification of ortholog losses in insects and vertebrates. *Genome Biology*, **8**(11).
- Wynant, N., Santos, D., Verdonck, R., Spit, J., Van Wielendaele, P., Vanden Broeck, J. (2014) Identification, functional characterization and phylogenetic analysis of double stranded RNA degrading enzymes present in the gut of the desert locust, *Schistocerca gregaria*. *Insect Biochemistry and Molecular Biology*, **46**, 1-8.
- Xia, J.X., Yang, Z.Z., Gong, C., Xie, W., Pan, H.P., Guo, Z.J., Zheng, H.X., Yang, X., Sun, X.D., Kang, S., Yang, F.S., Wu, Q.J., Wang, S.L., Cong, B., Teng, X.F., Zhang, Y.J. (2017) Genome-wide identification and expression analysis of amino acid transporters in the whitefly, *Bemisia tabaci* (Gennadius). *International Journal of Biological Sciences*, **13**(6), 735-747.
- Xu, J., De Barro, P.J., Liu, S.S. (2010) Reproductive incompatibility among genetic groups of *Bemisia tabaci* supports the proposition that the whitefly is a cryptic species complex. *Bulletin of Entomological Research*, **100**(3), 359-366.
- Xu, W. and Han, Z. (2008) Cloning and phylogenetic analysis of sid-1-like genes from aphids. *Journal of Insect Science*, **8**.
- Yu, N., Christiaens, O., Liu, J., Niu, J., Cappelle, K., Caccia, S., Huvenne, H., Smagghe, G. (2013) Delivery of dsRNA for RNAi in insects: an overview and future directions. *Insect Science*, **20**(1), 4-14.
- Zakharov, I.A. (2016) Horizontal gene transfer into the genomes of insects. *Russian Journal of Genetics*, **52**(7), 702-707.
- Zang, L.-S. and Liu, S.-S. (2007) A comparative study on mating behaviour between the B biotype and a non-B biotype of *Bemisia tabaci* (Hemiptera : Aleyrodidae) from Zhejiang, China. *Journal of Insect Behavior*, **20**(2), 157-171.

- Zdobnov, E.M. and Apweiler, R. (2001) InterProScan - an integration platform for the signature-recognition methods in InterPro. *Bioinformatics*, **17**(9), 847-848.
- Zeuthen, T. (2001) How water molecules pass through aquaporins. *Trends in Biochemical Sciences*, **26**(2), 77-79.
- Zha, W., Peng, X., Chen, R., Du, B., Zhu, L., He, G. (2011) Knockdown of midgut genes by dsRNA-transgenic plant-mediated RNA interference in the hemipteran insect *Nilaparvata lugens*. *Plos One*, **6**(5).
- Zhang, X.S., Holt, J., Colvin, J. (2000) A general model of plant-virus disease infection incorporating vector aggregation. *Plant Pathology*, **49**(4), 435-444.
- Zhang, X.S., Holt, J., Colvin, J. (2001) Synergism between plant viruses: a mathematical analysis of the epidemiological implications. *Plant Pathology*, **50**(6), 732-746.
- Zhang, S.H., Zeng, X.F., Ren, M., Mao, X.B., Qiao, S.Y. (2017) Novel metabolic and physiological functions of branched chain amino acids: a review. *Journal of Animal Science and Biotechnology*, **8**.
- Zhou, X.P., Liu, Y.L., Calvert, L., Munoz, C., OtimNape, G.W., Robinson, D.J., Harrison, B.D. (1997) Evidence that DNA-A of a geminivirus associated with severe cassava mosaic disease in Uganda has arisen by interspecific recombination. *Journal of General Virology*, **78**, 2101-2111.
- Zhou, X., Wheeler, M.M., Oi, F.M., Scharf, M.E. (2008) RNA interference in the termite *Reticulitermes flavipes* through ingestion of double-stranded RNA. *Insect Biochemistry and Molecular Biology*, **38**(8), 805-815.
- Zhou, Y.-L., Zhu, X.-Q., Gu, S.-H., Cui, H.-h., Guo, Y.-Y., Zhou, J.-J., Zhang, Y.-J. (2014) Silencing in *Apolygus lucorum* of the olfactory coreceptor Orco gene by RNA interference induces EAG response declining to two putative semiochemicals. *Journal of Insect Physiology*, **60**, 31-39.
- Zografou, E.N., Tsiropoulos, G.J., Margaritis, L.H. (1998) Survival, fecundity and fertility of *Bactrocera oleae*, as affected by amino acid analogues. *Entomologia Experimentalis et Applicata*, **87**(2), 125-132.